Distribution of values of bounded generalized polynomials

V. Bergelson* and A. Leibman*

August 12, 2005

Abstract

A *generalized polynomial* is a real-valued function constructible from ordinary polynomials by the use of the operations of addition, multiplication and taking the integer part; a *generalized polynomial mapping* is a vector-valued mapping whose coordinates are generalized polynomials. We show that any bounded generalized polynomial mapping $u\colon \mathbb{Z}^d \longrightarrow \mathbb{R}^l$ has a representation $u(n) = f(\varphi(n)x)$, $n \in \mathbb{Z}^d$, where f is a piecewise polynomial function on a compact nilmanifold X, $x \in X$ and φ is an ergodic $\mathbb{Z}^d$-action by translations on X. This fact is used to show that the sequence $u(n)$, $n \in \mathbb{Z}^d$, is well distributed on a piecewise polynomial surface $\mathcal{P} \subset \mathbb{R}^l$ (with respect to the Borel measure on $\mathcal{P}$ which is the image of the Lebesgue measure under the piecewise polynomial function defining $\mathcal{P}$). As corollaries we also obtain a von Neumann-type ergodic theorem along generalized polynomials and a result on diophantine approximations extending the work of van der Corput and Furstenberg-Weiss.

0. Introduction

0.1. The main object of study in this paper is the class GP of *generalized polynomials*, namely the class of functions which is generated by starting with ordinary polynomials of one or several variables and applying in orbitrary order the operations of taking the integer part (sometimes called bracket function, or floor function), addition and multiplication. We will denote the integer part of a number $a \in \mathbb{R}$ or, more generally, of a vector $a \in \mathbb{R}^l$, by $[a]$, and the fractional part of a, $a-[a]$, by $\langle\!\langle a \rangle\!\rangle$. Accordingly, given a real or a vector-valued function f, the functions $[f]$ and $\langle\!\langle f \rangle\!\rangle$ are defined by $[f](x) = [f(x)]$ and $\langle\!\langle f \rangle\!\rangle = f - [f]$.

The following description presents the class GP in a more formal way. For a fixed $d \in \mathbb{N}$ let GP_0 denote the ring of polynomial mappings from either $\mathbb{Z}^d$ or $\mathbb{R}^d$ to $\mathbb{R}$, and let $\mathrm{GP} = \bigcup_{n=1}^{\infty} \mathrm{GP}_n$ where, for $n \geq 1$,

$$\mathrm{GP}_n = \mathrm{GP}_{n-1} \cup \bigl\{u+v : u,v \in \mathrm{GP}_{n-1}\bigr\} \cup \bigl\{uv : u,v \in \mathrm{GP}_{n-1}\bigr\} \cup \bigl\{\pm[u] : u \in \mathrm{GP}_{n-1}\bigr\}.$$

Finally, let us call vector-valued generalized polynomials $u = (u_1,\ldots,u_l)\colon \mathbb{Z}^d \longrightarrow \mathbb{R}^l$, or $\mathbb{R}^d \longrightarrow \mathbb{R}^l$, with $u_1,\ldots,u_l \in \mathrm{GP}$, *generalized polynomial mappings*, or *GP-mappings*.

---

*Supported by NSF grants DMS-0070566 and DMS-0345350. Leibman was also supported by the Sloan foundation.

0.2. Examples. If p_i are ordinary polynomials of one or several variables, then $[p_1]$, $p_1[p_2]$, $p_1 + p_2[p_3]$, $\big[[[p_1]p_2 + p_3][p_4]p_5 + p_6\big] + p_7[p_8]^3$ are generalized polynomials. The fractional part $\langle\!\langle p_1 \rangle\!\rangle = p_1 - [p_1]$, the values of a polynomial modulo 1, $p_1 (\mathrm{mod}\, 1)$ (which is the same as $\langle\!\langle p_1 \rangle\!\rangle$ if $\mathbb{R}/\mathbb{Z}$ is identified with $[0,1)$), and combined expressions such as $[p_1]^2 \langle\!\langle p_2[p_3] + p_4 \rangle\!\rangle^3 (\mathrm{mod}\, 5)$ are generalized polynomials as well.

0.3. Generalized polynomials of a special type are featured in the following classical result due to H. Weyl ([We]).

Theorem. *Given a (conventional) polynomial $p(t) = \sum_{i=0}^{k} a_i t^i$ such that at least one among the coefficients $a_1, \dots, a_k$ is irrational, the sequence of values $\{\langle\!\langle p(n) \rangle\!\rangle\}_{n\in\mathbb{N}}$ of the generalized polynomial $\langle\!\langle p \rangle\!\rangle$ is uniformly distributed in $[0,1]$. In particular, for any $\varepsilon > 0$ there exists $n \in \mathbb{N}$ such that $\langle\!\langle p(n) \rangle\!\rangle < \varepsilon$.*

0.4. The following examples demonstrate various distribution phenomena which one encounters when dealing with bounded generalized polynomials $u\colon \mathbb{Z} \longrightarrow \mathbb{R}$:

Examples. Let a and b be rationally independent irrational numbers.

(1) The values of the generalized polynomial $u(n) = \langle\!\langle an \rangle\!\rangle^2$ are dense but not uniformly distributed in $[0,1]$. They are, however, uniformly distributed in $[0,1]$ with respect to the measure $\frac{dx}{2\sqrt{x}}$.

(2) The sequence $\big\langle\!\big\langle -\sqrt{2}n[\sqrt{2}n] \big\rangle\!\big\rangle$, $n \in \mathbb{N}$, is dense and uniformly distributed in $[0,1]$ with respect to the measure which is equal to $\frac{dx}{2\sqrt{2x}}$ on $\left[0, \frac{1}{2}\right]$ and to $\frac{dx}{2\sqrt{2x-1}}$ on $\left[\frac{1}{2}, 1\right]$. (See section 1.28 below.) On the other hand, one can show that the sequence $\big\langle\!\big\langle -\sqrt[3]{2}n[\sqrt[3]{2}n] \big\rangle\!\big\rangle$, $n \in \mathbb{N}$, is uniformly distributed in $[0,1]$ with respect to the standard Lebesgue measure. (This fact is a special case of Proposition 1.2 in [Hå3].)

(3) The sequence $\langle\!\langle an \rangle\!\rangle \langle\!\langle bn \rangle\!\rangle$, $n \in \mathbb{N}$, is uniformly distributed in $[0,1]$ with respect to the measure $-\log x\, dx$. (This follows from the fact that the vector-valued sequence $\big(\langle\!\langle an \rangle\!\rangle, \langle\!\langle bn \rangle\!\rangle\big)$ is uniformly distributed in the square $[0,1]^2$.)

(4) The sequence $\frac{2}{3}\langle\!\langle an \rangle\!\rangle + \frac{1}{3}[2\langle\!\langle an \rangle\!\rangle]$, $n \in \mathbb{N}$, is uniformly distributed in $\left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$ with respect to the (normalized) Lebesgue measure.

(5) For the sequence $u(n) = [2\langle\!\langle an \rangle\!\rangle]\langle\!\langle bn \rangle\!\rangle$, $n \in \mathbb{N}$, the set $\mathcal{Z} = \big\{n \in \mathbb{N} : u(n) = 0\big\}$ has density $1/2$, and the sequence of the nonzero values of u, $\big\{u(n),\ n \notin \mathcal{Z}\big\}$, is uniformly distributed on the interval $[0,1]$ with respect to the standard Lebesgue measure.

(6) The sequence $u(n) = [(n+1)a] - [na] - [a]$, $n \in \mathbb{N}$, takes on only the values 0 and 1. (The generalized polynomial $u(n)$, often called nowadays *Beatty sequence*, appears already in the work of astronomer J. Bernoulli III ([Mar]), and is found, under different names, in variety of mathematical situations, from symbolic dynamics to theory of mathematical games.)

0.5. The examples above indicate that a generalized polynomial can have quite intricate distributional properties. Given a bounded generalized polynomial u, one would like at least to know whether the sequence $\{u(n)\}_{n\in\mathbb{Z}}$ has some regular behavior. In particular, one would like to know the answer to the following recalcitrant question posed in [BHå]:

Question. *Is it true that* $\lim\limits_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}e^{2\pi i u(n)}$ *exists for any generalized polynomial* u*?*

A general result which we obtain in this paper (Theorem B below) not only implies that the answer to this question is positive, but also gives a description of the measure which, so to say, governs the law of distribution of the sequence of values of a generalized polynomial.

0.6. A more general version of Theorem 0.3, also obtained in [We], deals with vector-valued generalized polynomials of the special form $p \bmod 1 = (p_1 \bmod 1, \dots, p_l \bmod 1)\colon \mathbb{Z} \longrightarrow \mathbb{T}^l = \mathbb{R}^l/\mathbb{Z}^l$, where $p = \big(p_1, \dots, p_l\big)\colon \mathbb{Z} \longrightarrow \mathbb{R}^l$ is a polynomial mapping.

Theorem. (Cf. [We], Theorem 18) *Let* $p\colon \mathbb{Z} \longrightarrow \mathbb{R}^l$ *be a polynomial mapping and let* $\tilde{p} = p \bmod 1\colon \mathbb{Z} \longrightarrow \mathbb{T}^l$ *be the corresponding generalized polynomial obtained by reduction modulo 1. There exist (disjoint, parallel and isomorphic) subtori* $S_1, \dots, S_k$ *in* $\mathbb{T}^l$ *such that the sequence* $\{\tilde{p}(n)\}_{n\in\mathbb{N}}$ *is uniformly distributed on* $\bigcup_{i=1}^k S_i$*. (More pricesely, let* $i_n = i$ *when* $\tilde{p}(n) \in S_i$, $n \in \mathbb{N}$*; then the sequence* $\{i_n\}_{n\in\mathbb{N}}$ *is periodic, and if* r *is the period of this sequence and* $r_i = \#\big\{1 \le n < r : \tilde{p}(n) \in S_i\big\}$, $i = 1, \dots, k$, *then* $\{\tilde{p}(n)\}_{n\in\mathbb{N}}$ *is uniformly distributed on* $\bigcup_{i=1}^k S_i$ *with respect to the measure* $\sum_{i=1}^k \frac{r_i}{r}\lambda_i$*, where* λ_i *denotes the normalized Lebesgue measure on* S_i, $i = 1, \dots, k$*.)*

0.7. The following illustrative example corresponds to $l = 3$, $k = 2$, $r_1 = 1$ and $r_2 = 2$ in the above theorem. Let a be an irrational number; consider the sequence $\tilde{p}(n) = \big(\frac{1}{3}n^2 \bmod 1, na \bmod 1, n^2 a \bmod 1\big)$, $n \in \mathbb{N}$, in $\mathbb{T}^3$. Let S_0, S_1 be the two-dimensional tori defined by $S_0 = \{0\} \times \mathbb{T}^2$ and $S_1 = \big\{\frac{1}{3}\big\} \times \mathbb{T}^2$. The sequence $\{\tilde{p}(n)\}_{n\in\mathbb{N}}$ visits the tori S_0, S_1 in the following order: $S_0, S_1, S_1, S_0, S_1, S_1, \dots$, and is uniformly distributed on $S_0 \cup S_1$ with respect to the probability measure $\frac{1}{3}\lambda_0 + \frac{2}{3}\lambda_1$ where λ_i denotes the normalized Lebesgue measure on S_i, $i = 0, 1$.

0.8. A frequently cited special case of the above theorem concerns the situation where the components of p, the polynomials $p_1, \dots, p_l$, are rationally independent. In this case the sequence $\{\tilde{p}(n)\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{T}^l$. From our perspective, the case when p_i are rationally dependent is more significant since it contains in embryonic form certain elements of a general theorem pertaining to *arbitrary* generalized polynomials.

0.9. Identifying the torus $\mathbb{T}^l$ with the unit cube $K = [0,1)^l$ (and not distinguishing between $p \bmod 1$ and $\langle\!\langle p \rangle\!\rangle$) allows one to view the subtori appearing in the formulation of Theorem 0.6 above as sections of K by a finite system of parallel planes. One can now rephrase Theorem 0.6 by saying that the sequence $\{\langle\!\langle p(n) \rangle\!\rangle\}_{n\in\mathbb{N}}$ is uniformly distributed on a bounded piecewise linear surface in $\mathbb{R}^l$. The main goal of this paper is to obtain a version of this fact for general GP-mappings. But first we bring some examples which demonstrate new phenomena that occur when dealing with vector-valued generalized polynomials.

Examples. Let a and b be rationally independent irrational numbers.

(1) The values of the generalized polynomial mapping $u(n) = \big(\langle\!\langle an \rangle\!\rangle, \langle\!\langle an \rangle\!\rangle^2\big)$, $n \in \mathbb{Z}$, are dense on the parabola segment $\big\{(x, x^2),\ x \in [0,1]\big\}$ in $\mathbb{R}^2$ (and uniformly distributed with respect to the measure whose projection on the first coordinate is the standard Lebesgue

measure on $[0,1]$).

(2) The values of $u(n) = \Big(⟪an⟫, [2⟪bn⟫](2⟪an⟫^2 - 1) - ⟪an⟫^2 + 1\Big)$, $n \in \mathbb{Z}$, are dense (and uniformly distributed) on the union of two intersecting parabola segments $\big\{(x, x^2),\ x \in [0,1]\big\}$ and $\big\{(x, 1-x^2),\ x \in [0,1]\big\}$.

0.10. While the examples in 0.4 and 0.9 indicate that too direct a generalization of Weyl's theorem cannot be hoped for, it turns out that values of any bounded generalized polynomial $u\colon \mathbb{Z}^d \longrightarrow \mathbb{R}^l$ are uniformly distributed on a piecewise polynomial surface $\mathcal{P}$ (see Theorems 0.25 and 1.25 below). We will now discuss the ideas behind the proof of this fact. Let us return for a moment to Theorem 0.3. There are essentially two known approaches to the proof of this theorem. The original approach of Weyl in [We] can be described as follows. First Weyl establishes the equivalence of the following conditions for a sequence $\{a_n\}_{n\in\mathbb{N}}$ in $[0,1]$:

(i) $\{a_n\}$ is uniformly distributed in $[0,1]$, that is, for any interval $[b,c] \subseteq [0,1]$ one has $\frac{1}{N} \cdot \#\{n \le N : a_n \in [b,c]\} \underset{N\to\infty}{\longrightarrow} b - c$;

(ii) for any Riemann integrable function f on $[0,1]$ one has $\frac{1}{N}\sum_{n=1}^{N} f(a_n) \underset{N\to\infty}{\longrightarrow} \int_0^1 f\,dx$;

(iii) For any $m \in \mathbb{Z} \setminus \{0\}$, $\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i m a_n} \underset{N\to\infty}{\longrightarrow} 0$.

To establish the uniform distribution of the sequence $\{⟪p(n)⟫\}_{n\in\mathbb{N}}$, Weyl uses the fact that if for any $m \in \mathbb{N}$ the sequence $\{a_{n+m} - a_n\}_{n\in\mathbb{N}}$ is uniformly distributed modulo 1, then the sequence $\{a_n\}_{n\in\mathbb{N}}$ is also uniformly distributed modulo 1. Since after finitely many applications of the difference operator $D_m p(n) = p(n+m) - p(n)$ the situation is reduced to the case of linear polynomials, for which the condition (iii) above is easily verified, the result follows. (The difference trick described above is usually called *van der Corput's difference theorem* in honor of van der Corput, who educed and successfully applied it in his work. See [vdC].)

0.11. A different approach to the proof of Theorem 0.3, which may be called dynamical, deals with a special class of affine maps of a torus. This approach was introduced by Furstenberg in [F2] and [F1] (see also [H] and [C] for a similar treatment) and can be described as follows. Let $p(x) = a_0 + a_1 x + a_2 x^2 \ldots + a_k x^k = b_0 + b_1 x + b_2\binom{x}{2} + \ldots + b_k\binom{x}{k} \in \mathbb{R}[x]$. Consider the following affine transformation, called *a skew product*, of the k-dimensional torus $\mathbb{T}^k = \mathbb{R}^k/\mathbb{Z}^k$:

$$g(y_1, y_2, \ldots, y_k) = \big(y_1 + b_k, y_2 + y_1 + b_{k-1}, \ldots, y_k + y_{k-1} + b_1\big).$$

Let $y = (0, \ldots, 0, b_0) \in \mathbb{T}^k$; one can check by induction on n that $(g^n y)_k = p(n) (\mathrm{mod}\, 1)$, $n \in \mathbb{Z}$. Now one can use the known properties of the dynamical system $(\mathbb{T}^k, g)$ in order to characterize the behavior of the sequence $\{⟪p(n)⟫\}_{n\in\mathbb{Z}}$. In particular, if a_k is irrational, then the system $(\mathbb{T}^k, g)$ is uniquely ergodic (with the unique g-invariant measure being the Lebesgue measure on $\mathbb{T}^k$), which implies (the one-dimensional version of) the Weyl theorem.

0.12. Let us now return to generalized polynomials. While various modifications of the technique based on the van der Corput difference theorem allow one to treat successfully some special classes of generalized polynomials which are uniformly distributed with respect to the Lebesgue measure (see [Hå1], [Hå2], [Hå3]), it seems not to be applicable in the situations where the distribution law is not known in advance or is complicated. On the other hand, the dynamical approach has much greater range of applicability. Indeed, if a sequence $\{a_n\}_{n\in\mathbb{N}}$ in $[0,1]$ *comes from* a uniquely ergodic dynamical system (X,T,μ) in the sense that for some Riemann integrable function $\varphi\colon X \longrightarrow \mathbb{R}$ and a point $x \in X$ one has $a_n = \varphi(T^n x)$ then, as a consequence of unique ergodicity, one will have for any function $f \in C([0,1])$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(a_n) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(\varphi(T^n x)) = \int_X f(\varphi(x))\,d\mu.$$

(See [F3] and [Wa] for treatment of basic properties of unique ergodicity.)

0.13. The following example shows how the dynamical apporach can be applied to generalized polynomials. Let $u(n) = \langle\!\langle an[bn]\rangle\!\rangle$, $n \in \mathbb{Z}$, where $a,b \in \mathbb{R}$; we will obtain the generalized polynomial u "dynamically". Let G be the group of 4×4 upper triangular matrices, $G = \left\{\begin{pmatrix} 1 & \alpha_{1,2} & \alpha_{1,3} & \alpha_{1,4} \\ 0 & 1 & \alpha_{2,3} & \alpha_{2,4} \\ 0 & 0 & 1 & \alpha_{3,4} \\ 0 & 0 & 0 & 1 \end{pmatrix},\ \alpha_{i,j} \in \mathbb{R}\right\}$, and let $\Gamma = \left\{\begin{pmatrix} 1 & m_{1,2} & m_{1,3} & m_{1,4} \\ 0 & 1 & m_{2,3} & m_{2,4} \\ 0 & 0 & 1 & m_{3,4} \\ 0 & 0 & 0 & 1 \end{pmatrix},\ m_{i,j} \in \mathbb{Z}\right\}$. Then $X = G/\Gamma$ is a compact manifold, on which the group G naturally acts by left translations, $g(a\Gamma) = (ga)\Gamma$. The elements of X can be identified with matrices $x = \begin{pmatrix} 1 & x_{1,2} & x_{1,3} & x_{1,4} \\ 0 & 1 & x_{2,3} & x_{2,4} \\ 0 & 0 & 1 & x_{3,4} \\ 0 & 0 & 0 & 1 \end{pmatrix}$ where $x_{i,j} \in [0,1)$; we will call $x_{i,j}$, $1 \le i < j \le 4$, *the coordinates* of x. Note that while the coordinate functions $x_{i,j}$ are not continuous on X, the set of points of discontinuity of each of these functions has measure 0 and therefore, each $x_{i,j}$ is Riemann integrable.

Let $g = \begin{pmatrix} 1 & -a & 1 & 0 \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & ab \\ 0 & 0 & 0 & 1 \end{pmatrix} \in G$; one checks that $g^n = \begin{pmatrix} 1 & -an & n & 0 \\ 0 & 1 & 0 & bn \\ 0 & 0 & 1 & abn \\ 0 & 0 & 0 & 1 \end{pmatrix}$, $n \in \mathbb{Z}$. The matrix g acts on X by left translation, $g(x\Gamma) = (gx)\Gamma$. Let $x = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\Gamma \in X$; in order to write the sequence $g^n x$ "in coordinates" on X, we have to find, for each $n \in \mathbb{Z}$, a matrix $\gamma_n \in \Gamma$ such that $g^n\gamma_n$ has all its entries in $[0,1)$. Multiplying g^n by $\begin{pmatrix} 1 & -[-an] & -n & 0 \\ 0 & 1 & 0 & -[bn] \\ 0 & 0 & 1 & -[abn] \\ 0 & 0 & 0 & 1 \end{pmatrix}$ we get $\begin{pmatrix} 1 & \langle\!\langle -an\rangle\!\rangle & 0 & \xi_n \\ 0 & 1 & 0 & \langle\!\langle bn\rangle\!\rangle \\ 0 & 0 & 1 & \langle\!\langle abn\rangle\!\rangle \\ 0 & 0 & 0 & 1 \end{pmatrix}$, where $\xi_n = an[bn] - n[abn]$. Finally, multiplying this matrix by $\begin{pmatrix} 1 & 0 & 0 & -[\xi_n] \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ we obtain $\begin{pmatrix} 1 & \langle\!\langle -an\rangle\!\rangle & 0 & \langle\!\langle an[bn]\rangle\!\rangle \\ 0 & 1 & 0 & \langle\!\langle bn\rangle\!\rangle \\ 0 & 0 & 1 & \langle\!\langle abn\rangle\!\rangle \\ 0 & 0 & 0 & 1 \end{pmatrix}$. Thus, the $(1,4)$-coordinate of the sequence $\{g^n x\}_{n\in\mathbb{Z}}$ on X is just $\{\langle\!\langle an[bn]\rangle\!\rangle\}_{n\in\mathbb{Z}}$.

0.14. In the example above, the group G of upper triangular matrices is a nilpotent Lie group, Γ is a uniform subgroup of G, and X is, therefore, *a compact nilmanifold*. It turns out that the class of dynamical systems which are generated by translations on nilmanifolds provides the adequate framework for the study of generalized polynomials. In this paper the term *nilmanifold* will stand for a compact homogeneous space $X = G/\Gamma$ where G is a nilpotent, not necessarily connected, Lie group and Γ is a discrete subgroup of G. The group G acts on X by left translations, or, as we will often say, by *nilrotations*. We will use the term *nilsystem* to denote any dynamical system of the form (X, H) where $X = G/\Gamma$ is a (compact) nilmanifold and H is a subgroup of G acting on X by nilrotations.

0.15. Let us note that the skew product transformation of $\mathbb{T}^k$

$$g(y_1, y_2, \ldots, y_k) = \big(y_1 + b_k, y_2 + y_1 + b_{k-1}, \ldots, y_k + y_{k-1} + b_1\big), \tag{0.1}$$

which was utilized in 0.11 above in order to dynamically obtain the generalized polynomial $\langle\!\langle p \rangle\!\rangle = \big\langle\!\!\big\langle b_0 + b_1 x + b_2 \binom{x}{2} + \ldots + b_k \binom{x}{k} \big\rangle\!\!\big\rangle$, can also be viewed as a nilrotation. Indeed, let G be the group of upper triangular matrices $\begin{pmatrix} 1 & \alpha_{1,2} & \alpha_{1,3} & \ldots & \alpha_{1,k+1} \\ & 1 & \alpha_{2,3} & \ldots & \alpha_{2,k+1} \\ & & \ddots & & \vdots \\ & \mathbf{O} & & 1 & \alpha_{k,k+1} \\ & & & & 1 \end{pmatrix}$ with $\alpha_{i,j} \in \mathbb{Z}$ for $1 \leq i < j \leq k$ and $\alpha_{i,k+1} \in \mathbb{R}$ for $1 \leq i \leq k$, and let Γ be the subgroup of G consisting of the matrices with integer entries. Then G is a nilpotent (non-connected) Lie group with $X = G/\Gamma \simeq \mathbb{T}^k$, and the system defined on X by the nilrotation by the element $g = \begin{pmatrix} 1 & 1 & 0 & 0 & \ldots & b_k \\ & 1 & 1 & 0 & \ldots & b_{k-1} \\ & & & \ddots & & \vdots \\ & \mathbf{O} & & & 1 & b_1 \\ & & & & & 1 \end{pmatrix} \in G$ is isomorphic to the dynamical system on $\mathbb{T}^k$ defined by (0.1).

0.16. Nilsystems have some remarkable properties which will be relied upon in this paper. First, they are known to be *distal*, see [AGH], [K1], [K2]. (An action of a group G on a compact metric space is said to be distal if for any distinct points x and y of this space $\inf_{g \in G} \operatorname{dist}(g^n x, g^n y)$ is positive.) Now, if a group of homeomorphisms of a compact space X acts distally, then X is a disjoint union of minimal sets, which are orbit closures of points of X. While not every distal minimal system is uniquely ergodic, the minimal components of nilsystems are (see [Le] or [L2]).

0.17. We are going now to formulate a theorem that establishes a connection between generalized polynomials and nilsystems. But first we need to introduce the notion of a piecewise polynomial function on a nilmanifold. Given a compact connected nilmanifold X, one can define *a coordinate mapping* $\tau\colon X \longrightarrow [0,1)^k$ (see the formal definition in section 1.5 below). While the mapping τ is not continuous, its inverse τ^{-1} is. (This is clear in the case $X = \mathbb{T}^k$, where $\tau\colon \mathbb{T}^k \longrightarrow [0,1)^k$ is the standard coordinate mapping, and is analogous in the general case.) Let us say that a function h on $[0,1)^k$ is *piecewise polynomial* if there is a partition $[0,1)^k = \mathcal{L}_1 \cup \ldots \cup \mathcal{L}_r$ and polynomials $P_1, \ldots, P_r$ on $\mathbb{R}^k$ such that each $\mathcal{L}_j$ is defined by a system of polynomial equations and $h|_{\mathcal{L}_j} = P_j$, $j = 1, \ldots, r$. Let us say that a function f on X is piecewise polynomial if $f \circ \tau^{-1}$ is a piecewise polynomial function on $[0,1)^k$. This definition does not depend on the choice of a coordinate system on X

(see [L4]). We say that a function on a non-connected compact nilmanifold X is piecewise polynomial if it is piecewise polynomial on every connected component of X. A piecewise polynomial function does not have to be continuous, but it is clearly Riemann integrable.

0.18. Theorem A. *Any bounded generalized polynomial u on $\mathbb{Z}^d$ can be obtained dynamically in the following way: there exists a (not necessarily connected) compact nilmanifold X, an ergodic action φ of $\mathbb{Z}^d$ by nilrotations on X, a piecewise polynomial function f on X and a point $x \in X$ such that $u(n) = f(\varphi(n)x)$ for all $n \in \mathbb{Z}^d$.*

0.19. Remarks. (1) Not all bounded generalized polynomials can be obtained by using distal (in particular, nilpotent) systems and continuous functions thereon, that is, not every bounded generalized polynomial is of the form $u(n) = f(T^n x)$ where T is a distal transformation of a compact metric space X, $x \in X$ and $f \in C(X)$. Indeed, all points in a distal system are recurrent (see [F3]), and thus the sequence $f(T^n x)$ cannot have nonrecurrent values, whereas some generalized polynomials can (see examples in 0.26). Bounded generalized polynomials without isolated values may also not be representable in the form $f(T^n x)$ with distal T and continuous f; the simplest example of such a polynomial is provided by $u(n) = \langle\!\langle [\alpha n]\beta \rangle\!\rangle$, see [Hå1].

(2) Also, not all bounded generalized polynomials can be obtained by using skew product transformations on a torus (like in the example discussed in 0.11 above) and Riemann integrable functions. Indeed, consider the generalized polynomial $u(n) = \langle\!\langle an[bn] \rangle\!\rangle$, where a and b are rationally independent irrational numbers. Let X be a torus with the standard measure μ and let g be an ergodic skew product transformation of X. Assume that there exist a Riemann integrable function f on X and a point $x \in X$ such that $u(n) = f(g^n x)$, $n \in \mathbb{Z}$, and let $\tilde{f} = e^{2\pi i f}$. Then $\tilde{f}(g^n x) = e^{2\pi i an[bn]}$, $n \in \mathbb{Z}$. For any character χ on X one has $\chi(g^n x) = e^{2\pi i p(n)}$, where p is a conventional polynomial. Using the method described in section 1.28 below (or any other method) one can check that for any conventional polynomial p the sequence $\langle\!\langle an[bn] - p(n) \rangle\!\rangle$, $n \in \mathbb{N}$, is uniformly distributed in $[0,1]$. Hence, $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i (an[bn]-p(n))} = 0$. Since g is uniquely ergodic, the sequence $g^n(x)$ is uniformly distributed in X, and thus

$$\int_X \tilde{f} \cdot \overline{\chi}\, d\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \tilde{f}(g^n x) \cdot \overline{\chi(g^n x)}\, d\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i an[bn]} e^{-2\pi i p(n)} = 0.$$

Hence, $\tilde{f}$ is orthogonal to all characters on X, which contradicts the completeness of the system of characters on X.

0.20. In order to formulate and discuss corollaries of Theorem A we have to introduce some notation. Let u be a generalized polynomial; we will call the ordinary polynomials occurring in the expression for v *the polynomials involved in u.* More precisely, the set $I(u)$ of polynomials involved in u is

$\{u\}$ (the set whose only element is u) if u is an ordinary polynomial;

$I(u_1) \cup I(u_2)$ if $u = u_1 + u_2$ or $u = u_1 u_2$;

$I(v)$ if $u = \pm[v]$;

in view of the inductive definition in 0.1, $I(u)$ is defined for any representation of any generalized polynomial. It is important to stress that the set $I(u)$ depends on the expression representing u in terms of the ordinary polynomials; when speaking of polynomials involved in u we will always mean a concrete representation of u.

Example. The polynomials involved in $p_1\big[p_2+[p_3]p_4\big]$ are p_1, p_2, p_3 and p_4.

When $u=(u_1,\ldots,u_l)$ is a GP-mapping, the set of polynomials involved in u is $I(u)=I(u_1)\cup\ldots\cup I(u_l)$.

0.21. *The weight*, $\mathrm{wgt}(u)$, of a generalized polynomial u reflects the complexity of the "bracket pattern" in u. If u is a *simple* generalized polynomial, meaning that it has a representation in terms of the ordinary polynomials that does not contain the "+" or "−" sign, and is not of the form $[w]$, then $\mathrm{wgt}(u)$ is the number of brackets in u; otherwise $\mathrm{wgt}(u)$ is the maximum of the weights of the "simple components" of u. Formally,

$\mathrm{wgt}(u)=1$ if u is an ordinary polynomial;

$\mathrm{wgt}(u)=\max\{\mathrm{wgt}(u_1),\mathrm{wgt}(u_2)\}$ if $u=u_1+u_2$;

$\mathrm{wgt}(u)=\mathrm{wgt}(u_1)+\mathrm{wgt}(u_2)$ if $u=u_1u_2$;

$\mathrm{wgt}(u)=\mathrm{wgt}(v)$ if $u=\pm[v]$.

The weight of a GP-mapping $u=(u_1,\ldots,u_l)$ is defined as $\mathrm{wgt}(u)=\max\{\mathrm{wgt}(u_i)\}_{i=1}^{l}$.

Examples. If P_i are ordinary polynomials, then $\mathrm{wgt}(P_1)=\mathrm{wgt}([P_1])=1$, $\mathrm{wgt}\big(P_1[P_2]\big)=\mathrm{wgt}\big(P_1[P_2]+P_3\big)=2$, $\mathrm{wgt}\big(P_1[P_2[P_3]]\big)=\mathrm{wgt}\big(P_1[P_2][P_3]\big)=\mathrm{wgt}\big(P_1[P_2[P_3]+P_4]+P_5[P_6]\big)=3$, $\mathrm{wgt}\big(P_1[P_2[P_3]+P_4][P_5]+P_6\big)=4$.

0.22. We define the set of coefficients of a (conventional) polynomial mapping $\mathcal{P}=(P_1,\ldots,P_l)\colon\mathbb{R}^s\longrightarrow\mathbb{R}^l$ as the union of the sets of coefficients of polynomials $P_1,\ldots,P_l$. *The degree* of $\mathcal{P}$ is the maximum of the degrees of $P_1,\ldots,P_l$.

0.23. Let $P_1,\ldots,P_l,R_1,\ldots,R_m$ be polynomials $\mathbb{R}^s\longrightarrow\mathbb{R}$ such that the set $\mathcal{L}=\big\{y\in\mathbb{R}^s: 0\le R_i(y)<1,\ i=1,\ldots,m\big\}$ has a nonempty interior. By abuse of language, we will identify the mapping $\mathcal{P}=(P_1,\ldots,P_l)_{|\mathcal{L}}\colon\mathcal{L}\longrightarrow\mathbb{R}^l$ with its range $\mathcal{P}(\mathcal{L})$ and call it a (parametrized) *polynomial surface in* $\mathbb{R}^k$. Let λ be the Lebesgue measure on $\mathbb{R}^s$; we will denote by $\mu_{\mathcal{P}}$ the normalized measure $\mathcal{P}_*(\lambda)$ on $\mathbb{R}^l$, which is defined by $\mu_{\mathcal{P}}(A)=\lambda(\mathcal{P}^{-1}(A))/\lambda(\mathcal{L})$ for A being a Borel set in $\mathbb{R}^l$. By definition, *the coefficients* of the polynomial surface $\mathcal{P}$ are the coefficients of the polynomials $P_1,\ldots,P_l,R_1,\ldots,R_m$, and *the degree* of $\mathcal{P}$ is the maximal degree of $P_1,\ldots,P_l,R_1,\ldots,R_m$.

0.24. *A Følner sequence* in $\mathbb{Z}^d$ is a sequence $\{\Phi_N\}_{N\in\mathbb{N}}$ of finite subsets of $\mathbb{Z}^d$ such that, for any $n\in\mathbb{Z}^d$, $\frac{|(\Phi_N+n)\triangle\Phi_N|}{|\Phi_N|}\longrightarrow 0$ as $N\to\infty$. (A standard example of a Følner sequence is provided by a sequence of cubes of increasing size in $\mathbb{Z}^d$.) We will say that a set $E\subseteq\mathbb{Z}^d$ *has density* α and write $\mathcal{D}(E)=\alpha$ if $\lim\limits_{N\to\infty}\frac{|E\cap\Phi_N|}{|\Phi_N|}=\alpha$ for every Følner sequence $\{\Phi_N\}_{N=1}^{\infty}$ in $\mathbb{Z}^d$.

Remark. Notice that one often uses the expression "density" for a limit along a specific Følner sequence. For example, for $E\subseteq\mathbb{Z}$, the sequence of intervals $[-N,N]$, $N\in\mathbb{N}$, is frequently used. We use a stronger notion of density requiring the limit to be the same along *any* Følner sequence.

When we say that a statement holds for *almost all* elements of $\mathbb{Z}^d$ we mean that this statement holds for all elements of $\mathbb{Z}^d$ but a subset of zero density.

Let $E \subseteq \mathbb{Z}^d$ with $\mathcal{D}(E) > 0$ and let ω be a mapping from E to a topological space X endowed with a finite (nonzero) Borel measure μ. We will call the set $\omega(E)$ parametrized by elements of E *a sequence* and say that $\omega(E)$ is *well distributed in* X (with respect to μ) if for any open ball $U \subseteq X$ one has $\mathcal{D}(\omega^{-1}(U))/\mathcal{D}(E) = \mu(U)/\mu(X)$. This is the case, for any Riemann integrable function f on X and any Følner sequence $\{\Phi_N\}_{N=1}^{\infty}$ in $\mathbb{Z}^d$ one has $\lim\limits_{N\to\infty} \frac{1}{|E\cap\Phi_N|} \sum_{n\in E\cap\Phi_N} f(\omega(n)) = \int_X f\,d\mu$.

Remark. Note that the notion of well distribution is a more restricitve one than the more common "uniform distribution".

0.25. We are now in position to formulate our main result.

Theorem B. *Let* $u\colon \mathbb{Z}^d \longrightarrow \mathbb{R}^l$ *be a bounded GP-mapping. There exist bounded polynomial surfaces* $\mathcal{P}_1,\ldots,\mathcal{P}_k \subset \mathbb{R}^l$ *of degree* $\leq \mathrm{wgt}(u)$ *and a partition* $\mathbb{Z}^d = \mathcal{Z}_* \cup \bigcup_{j=1}^k \mathcal{Z}_j$ *such that* $\mathcal{D}(\mathcal{Z}_*) = 0$ *and for every* $j \in \{1,\ldots,k\}$ *the sequence* $u(\mathcal{Z}_j)$ *is well distributed on* $\mathcal{P}_j$ *with respect to* $\mu_{\mathcal{P}_j}$. *The coefficients of* $\mathcal{P}_1,\ldots,\mathcal{P}_k$ *belong to the ring generated over* $\mathbb{Q}$ *by the constant terms of the polynomials involved in* u.

When the set $\mathcal{Z}_*$ in the assertion of Theorem B is fixed, we will call the values of u at the points of $\mathcal{Z}_*$ *exceptional*, and the other values of u *regular*. The theorem then says that the regular values of any generalized polynomial mapping u lie and are well distributed on a piecewise polynomials surface (whereas the exceptional values, which do not affect the distributional behavior of u, are out of our control).

0.26. Here are some examples of generalized polynomials with exceptional values.

Examples. (1) Let a be an irrational number and let $u(n) = \big[1 - \langle\!\langle an \rangle\!\rangle\big]$. Then $u(n) = 0$ for all $n \neq 0$ and $u(0) = 1$ is an exceptional value of u.

(2) Let $a \in \mathbb{R}$ be such that the set $S_a = \big\{n \in \mathbb{N} : 0 < \langle\!\langle an \rangle\!\rangle < \frac{1}{n}\big\}$ is infinite. (For instance, $a = \sum_{n=1}^{\infty} 2^{-(2^n-1)}$ works since, as it is easy to check, $2^{2^n-1} \in S_a$ for all $n \in \mathbb{N}$.) Let b be any irrational number. Define $u(n) = \big\langle\!\big\langle \big[1 - \langle\!\langle [\langle\!\langle an \rangle\!\rangle n] b \rangle\!\rangle\big] an \big\rangle\!\big\rangle$, $n \in \mathbb{N}$. Then $u(n) = \langle\!\langle an \rangle\!\rangle < \frac{1}{n}$ for $n \in S_a$ and $u(n) = 0$ for $n \notin S_a$. The regular values $u(n)$, $n \in \mathbb{N} \setminus S_a$, of u are all equal to 0 whereas the exceptional values $u(n)$, $n \in S_a$, form a sequence converging to 0.

(3) In the notation of the preceding example, let now $u(n) = \big\langle\!\big\langle \big[1 - \langle\!\langle [\langle\!\langle an \rangle\!\rangle n] b \rangle\!\rangle\big] cn \big\rangle\!\big\rangle$, $c \in \mathbb{R}$. One can show that, varying the parameter c, one may achieve any a priori given distribution (with respect to any a priori chosen Følner sequence) of the sequence of exceptional values $u(n)$, $n \in S_a$, in $[0,1]$.

0.27. We will now derive some corollaries of Theorem B.

Corollary. *Let* $u\colon \mathbb{Z}^d \longrightarrow \mathbb{R}^l$ *be a bounded GP-mapping. For any* $f \in C(\mathbb{R}^l)$ *and any Følner sequence* $\{\Phi_N\}_{N=1}^{\infty}$ *in* $\mathbb{Z}^d$, $\lim\limits_{N\to\infty} \frac{1}{|\Phi_N|}\sum_{n\in\Phi_N} f(u(n))$ *exists and equals* $\sum_{j=1}^{k}\bigl(\mathcal{D}(\mathcal{Z}_j)\int_{\mathcal{P}_j} f\,d\mu_{\mathcal{P}_j}\bigr)$.

0.28. The following special case of Corollary 0.27 gives a positive answer to the question posed in 0.5:

Corollary. *For any generalized polynomial* $u\colon \mathbb{Z}^d \longrightarrow \mathbb{R}$ *and any Følner sequence* $\{\Phi_N\}_{N=1}^{\infty}$ *in* $\mathbb{Z}^d$, $\lim\limits_{N\to\infty} \frac{1}{|\Phi_N|}\sum_{n\in\Phi_N} e^{2\pi i u(n)}$ *exists.*

(The point here is that the generalized polynomial u is not assumed to be bounded, but this does not matter in view of the identity $e^{2\pi i u(n)} = e^{2\pi i \langle\!\langle u(n)\rangle\!\rangle}$.)

0.29. From Corollary 0.28 and the spectral theorem one can deduce the following generalization of von Neumann's ergodic theorem:

Corollary. *Let* $U_1^t,\ldots,U_k^t$, $t\in\mathbb{R}$, *be commuting unitary flows on a Hilbert space* $\mathcal{H}$ *and let* $u_1,\ldots,u_k$ *be generalized polynomials* $\mathbb{Z}^d \longrightarrow \mathbb{R}$. *For any Følner sequence* $\{\Phi_N\}_{N=1}^{\infty}$ *in* $\mathbb{Z}^d$ *the sequence* $\frac{1}{|\Phi_N|}\sum_{n\in\Phi_N} U_1^{u_1(n)}\cdots U_k^{u_k(n)}$ *is convergent in the strong operator topology.*

Proof. An application of the spectral theorem reduces the problem to the case where $\mathcal{H} = L^2(\Omega)$ for some measure space Ω and $(U_j^t g)(x) = e^{2\pi i f_j(x)t} g(x)$, $g\in L^2(\Omega)$, $j=1,\ldots,k$, $x\in\Omega$, where f_j are measurable real-valued functions on Ω. Then, for any $g\in L^2(\Omega)$ and $x\in\Omega$,

$$\Bigl(\prod_{j=1}^{k} U_j^{u_j(n)} g\Bigr)(x) = \Bigl(\prod_{j=1}^{k} e^{2\pi i u_j(n) f_j(x)}\Bigr) g(x) = e^{2\pi i \sum_{j=1}^{k} u_j(n) f_j(x)} g(x) = e^{2\pi i u_x(n)} g(x),$$

where $u_x(n) = \sum_{j=1}^{k} f_j(x) u_j(n)$, $n\in\mathbb{Z}^d$. By Corollary 0.28, the sequence $\frac{1}{|\Phi_N|}\sum_{n\in\Phi_N} e^{2\pi i u_x(n)} g(x)$ converges pointwise on Ω, and thus in $\mathcal{H} = L^2(\Omega)$. ■

0.30. The following corollary can be derived in a similar fashion:

Corollary. *Let* $U_1,\ldots,U_k$ *be commuting unitary operators on a Hilbert space and let* $u_1,\ldots,u_k$ *be generalized polynomials* $\mathbb{Z}^d \longrightarrow \mathbb{Z}$. *For any Følner sequence* $\{\Phi_N\}_{N=1}^{\infty}$ *in* $\mathbb{Z}^d$ *the sequence* $\frac{1}{|\Phi_N|}\sum_{n\in\Phi_N} U_1^{u_1(n)}\cdots U_k^{u_k(n)}$ *is convergent in the strong operator topology.*

0.31. We will now show that the converse of Theorem B also takes place, namely that for, in some sense, "any" piecewise polynomial surface in $\mathbb{R}^l$ there exists a generalized polynomial mapping whose values are well distributed on this surface:

Theorem. *Let* $\bigcup_{i=1}^{k}\mathcal{L}_i$ *be a partition of a cube* $Q = [0,1]^s$ *where each* $\mathcal{L}_i$, $i = 1,\ldots,k$, *is defined by several polynomial inequalities and has a nonempty interior, and let* $\mathcal{P}_1,\ldots,\mathcal{P}_k\colon \mathbb{R}^s \longrightarrow \mathbb{R}^l$ *be polynomial mappings. Then there exists a GP-mapping* $u\colon \mathbb{Z} \longrightarrow \mathbb{R}^l$ *and a partition* $\mathbb{Z} = \bigcup_{i=1}^{k}\mathcal{Z}_i$ *such that for each* $i = 1,\ldots,k$ *the sequence* $\{u(n)\}_{n\in\mathcal{Z}_i}$ *is well distributed on the polynomial surface* $\mathcal{P}_i(\mathcal{L}_i)$.

Proof. Let us start with the case of a single polynomial surface, $k = 1$. Let $v\colon \mathbb{Z} \longrightarrow \mathbb{R}^s$ be a GP-mapping whose values are well distributed in the cube $Q = [0,1]^s$ (say, $v(n) = \big(\langle\!\langle a_1 n\rangle\!\rangle, \ldots, \langle\!\langle a_s n\rangle\!\rangle\big)$, $n \in \mathbb{Z}$, where $a_1, \ldots, a_s$ are rationally independent irrational numbers), and let $\mathcal{P}$ be a polynomial mapping $\mathbb{R}^s \longrightarrow \mathbb{R}^l$. Then the values of the GP-mapping $u = \mathcal{P}_1 \circ v$ are located on the polynomial surface $\mathcal{P}_1(Q)$ in $\mathbb{R}^l$, and are well distributed on this surface with respect to the image $\mathcal{P}_*(\lambda)$ of the Lebesgue measure λ on Q.

Let now $k \geq 2$. Choose a GP-mapping $v\colon \mathbb{Z} \longrightarrow \mathbb{R}^s$ such that the values of v are well distributed in Q. For a polynomial $R\colon \mathbb{R}^s \longrightarrow \mathbb{R}^l$ with $\big|R_{|Q}\big| < M$ define $v_R(n) = -\big[R(v(n))/M\big]$ and $v'_R(n) = 1 + \big[-R(v(n))/M\big]$, $n \in \mathbb{Z}$; then $v_R(n) = 1$ if $R(v(n)) < 0$, $v_R(n) = 0$ if $R(v(n)) \geq 0$, $v'_R(n) = 1$ if $R(v(n)) \leq 0$ and $v'_R(n) = 0$ if $R(v(n)) > 0$. Now, if $\mathcal{L}_i$ is defined by the inequalities $R_1(x), \ldots, R_k(x) < 0$, $R'_1(x), \ldots, R'_l(x) \leq 0$, put $v_i(n) = \prod_{j=1}^k v_{R_j}(n) \prod_{j=1}^l v'_{R'_j}(n)$, $n \in \mathbb{Z}$; then $v_i(n) = 1$ if $v(n) \in \mathcal{L}_i$ and $v_i(n) = 0$ otherwise. Finally, define $u(n) = \sum_{i=1}^k v_i(n)\mathcal{P}_i(v(n))$, $n \in \mathbb{Z}$, then $u(n) = \mathcal{P}_i(v(n))$ if $v(n) \in \mathcal{L}_i$. For each $i \in \{1, \ldots, k\}$ let $\mathcal{Z}_i = \big\{n \in \mathbb{Z} : v(n) \in \mathcal{L}_i\big\}$; then the sequence $\{v(n)\}_{n \in \mathcal{Z}_i}$ is well distributed in $\mathcal{L}_i$, and so, $\{u(n)\}_{n \in \mathcal{Z}_i}$ is well distributed on the polynomial surface $\mathcal{P}_i(\mathcal{L}_i)$. ∎

0.32. To clarify the source of exceptional values of a generalized polynomial mapping, let us now assume that, in the notation of Theorem 0.31, the sets $\mathcal{L}_1, \ldots, \mathcal{L}_r$ have nonempty interiors whereas $\mathcal{L}_{r+1}, \ldots, \mathcal{L}_k$ have empty interiors. Then, if the values of a GP-mapping v are well distributed in Q, the (possibly empty) set $\mathcal{Z}_* = \big\{n \in \mathbb{Z} : v(n) \in \bigcup_{i=r+1}^k \mathcal{L}_i\big\}$, has zero density in $\mathbb{Z}$. Let, again, $u(n) = \sum_{i=1}^k v_i(n)\mathcal{P}_i(v(n))$, $n \in \mathbb{Z}$. Then for each $i \in \{1, \ldots, r\}$ the sequence $\{u(n)\}_{n \in \mathcal{Z}_i}$ is well distributed on the polynomial surface $\mathcal{P}_i(\mathcal{L}_i)$, but one has no information about the distribution of $\{u(n)\}_{n \in \mathcal{Z}_*}$.

0.33. Given a mapping u from a set S to a topological space X, we will say that a point x of X is *a limit point* for u if for any neighborhood W of x the set $u^{-1}(W)$ is infinite. It follows from Theorem B that every regular value of a bounded generalized polynomial mapping u is a limit point for u.

Actually, a stronger fact holds. The set of finite sums of distinct elements of a sequence in $\mathbb{Z}^d$ is called *an IP-set*; a subset of $\mathbb{Z}^d$ that has nonempty intersection with any IP-set is called *an IP*-set.* IP*-sets are “regular” and “large”; in particular, any IP*-set is syndetic, that is, has bounded gaps. (See [F3], Ch. 9.) A “shifted” IP*-set, that is, a set of the form $n + E$ where $n \in \mathbb{Z}^d$ and E is an IP*-set, is called *an* IP^*_+*-set.* Given a mapping $u\colon \mathbb{Z}^d \longrightarrow X$, we say that $x \in X$ is *an IP*-limit* for u if the preimage $u^{-1}(W)$ of any neighborhood W of x is an IP*-set, and that x is *an* IP^*_+*-limit* for u if the preimage $u^{-1}(W)$ of any neighborhood W of x is an IP^*_+-set.

0.34. Theorem C. *Let u be a bounded GP-mapping. Then $u(n)$ is an IP^*_+-limit for u for almost all $n \in \mathbb{Z}^d$. (In other words, there is a set $\mathcal{Z}_* \subset \mathbb{Z}^d$ of zero density such that for any $n \in \mathbb{Z}^d \setminus \mathcal{Z}_*$ and any $\varepsilon > 0$ the set $\big\{m \in \mathbb{Z}^d : |u(m) - u(n)| < \varepsilon\big\}$ is an IP^*_+-set.)*

Theorem C can be interpreted as describing good recurrence properties of generalized polynomials: if a generalized polynomial “visits” a point then almost surely it will visit any neighborhood of this point quite regularly.

0.35. For a given polynomial mapping u Theorem C gives no information about what values of u are limit points for u. This gap is partly filled by the following theorem.

Theorem D. *Let u be a GP-mapping $\mathbb{Z}^d \longrightarrow \mathbb{R}^l$ such that all polynomials involved in u have zero constant term, and let $\tilde{u} = u (\mathrm{mod}\, 1)$ viewed as a mapping to the torus $\mathbb{T}^l = \mathbb{R}^l/\mathbb{Z}^l$, that is, let $\tilde{u}$ be the composition of u and of the natural projection $\mathbb{R}^l \longrightarrow \mathbb{T}^l$. Then $0 \in \mathbb{T}^l$ is an IP*-limit for $\tilde{u}$. (In other words, for any $\varepsilon > 0$ the set $\{n \in \mathbb{Z}^d : \|u(n)\| < \varepsilon\}$, where $\|x\|$ is the distance from $x \in \mathbb{R}^l$ to $\mathbb{Z}^l$, is an IP_+^*-set.)*

0.36. The following theorem was obtained in [vdC]:

Theorem. *Let $p_i \colon \mathbb{Z}^{d+i-1} \longrightarrow \mathbb{R}$, $i = 1, \ldots, k$, be polynomials without constant term. Then for any $\varepsilon > 0$, the set of solutions $(n, m_1, \ldots, m_k) \in \mathbb{Z}^{d+k}$ of the system*

$$\begin{gathered} \big|p_1(n) - m_1\big| < \delta \\ \big|p_2(n, m_1) - m_2\big| < \delta \\ \vdots \\ \big|p_k(n, m_1, \ldots, m_{k-1}) - m_k\big| < \delta \end{gathered} \tag{0.2}$$

is syndetic in $\mathbb{Z}^{d+k}$.

0.37. It was proved [FW] that the set of solutions of the system (0.2) is IP*. This fact was further generalized in [BHåM]. We will now derive from Theorem D another generalization of Theorem 0.36.

Theorem. *Let $u_i \colon \mathbb{Z}^{d+i-1} \longrightarrow \mathbb{R}$, $i = 1, \ldots, k$, be generalized polynomials such that all ordinary polynomials involved therein have zero constant term. Then for any $\delta > 0$, the set of $(n_1, \ldots, n_d) \in \mathbb{Z}^d$ for which there exist $m_1, \ldots, m_k \in \mathbb{Z}$ satisfying*

$$\begin{gathered} \big|u_1(n) - m_1\big| < \delta \\ \big|u_2(n, m_1) - m_2\big| < \delta \\ \vdots \\ \big|u_k(n, m_1, \ldots, m_{k-1}) - m_k\big| < \delta \end{gathered} \tag{0.3}$$

is an IP-set.*

Proof. Put $[u]^1 = [u]$ and $[u]^{-1} = -[-u]$. Define

$$\begin{aligned} &v_1(n) = u_1(n),\ n \in \mathbb{Z}^d, \\ &v_2^{\epsilon_1}(n) = u_2\big(n, [v_1(n)]^{\epsilon_1}\big),\ n \in \mathbb{Z}^d,\ \epsilon_1 \in \{0,1\}, \\ &v_3^{\epsilon_1,\epsilon_2}(n) = u_3\Big(n, [v_1(n)]^{\epsilon_1}, [v_2^{\epsilon_1}(n)]^{\epsilon_2}\Big),\ n \in \mathbb{Z}^d,\ \epsilon_1, \epsilon_2 \in \{0,1\}, \\ &\quad\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ &v_k^{\epsilon_1,\ldots,\epsilon_k}(n) = u_k\Big(n, [v_1(n)]^{\epsilon_1}, [v_2^{\epsilon_1}(n)]^{\epsilon_2}, \ldots, [v_{k-1}^{\epsilon_1,\ldots,\epsilon_{k-2}}(n)]^{\epsilon_{k-1}}\Big),\ n \in \mathbb{Z}^d,\ \epsilon_1, \ldots, \epsilon_k \in \{0,1\}. \end{aligned}$$

By Theorem 0.35, for any $\delta > 0$, the set of $n \in \mathbb{Z}^d$ for which $\mathrm{dist}\big(v_i^{\epsilon_1,\ldots,\epsilon_{i-1}}(n) \bmod 1, 0\big) < \delta$ for all $i = 1, \ldots, k$ and $\epsilon_1, \ldots, \epsilon_k \in \{0,1\}$ is an IP*-set. For any such n we now construct a solution of (0.3) in the following way.

We have either $\langle\!\langle u_1(n) \rangle\!\rangle \overset{\delta}{\approx} 0$ or $\langle\!\langle u_1(n) \rangle\!\rangle \overset{\delta}{\approx} 1$; in the first case we put $\epsilon_1 = 1$, in the second case we put $\epsilon_1 = -1$. Define $m_1 = [u_1(n)]^{\epsilon_1}$; then, in both cases, $|u_1(n) - m_1| < \delta$.

We now have $v_2^{\epsilon_1}(n) = u_2(n, m_1)$ and either $\langle\!\langle u_2(n, m_1) \rangle\!\rangle \overset{\delta}{\approx} 0$ or $\langle\!\langle u_2(n, m_1) \rangle\!\rangle \overset{\delta}{\approx} 1$; in the first case we put $\epsilon_2 = 1$, in the second case we put $\epsilon_2 = -1$. Define $m_2 = [u_2(n, m_1)]^{\epsilon_2}$; then, in both cases, $|u_2(n, m_1) - m_2| < \delta$.

Next, we have $v_3^{\epsilon_1, \epsilon_2}(n) = u_2(n, m_1, m_2)$ and either $\langle\!\langle u_3(n, m_1, m_2) \rangle\!\rangle \overset{\delta}{\approx} 0$ or $\langle\!\langle u_3(n, m_1, m_2) \rangle\!\rangle \overset{\delta}{\approx} 1$; in the first case we put $\epsilon_3 = 1$, in the second case we put $\epsilon_3 = -1$. Define $m_3 = [u_2(n, m_1, m_2)]^{\epsilon_3}$; then, in both cases, $|u_3(n, m_1, m_2) - m_3| < \delta$. And so on, inductively. ■

0.38. One can show that the above results extend to generalized polynomials of continuous argument. As an example, we bring here a continuous version of Theorem B which can be obtained using the result from [Sh1]. For a measurable set $E \subseteq \mathbb{R}^d$ let us write $\mathcal{D}_B(E) = \alpha$ if for any sequence $r_n \longrightarrow \infty$ one has $\lim_{n \to \infty} \lambda(E \cap B_{r_n}) / \lambda(B_{r_n}) = \alpha$, where λ is the Lebesgue measure in $\mathbb{R}^d$ and $B_r \subset \mathbb{R}^d$ is the ball of radius r centered at 0. If $E \subseteq \mathbb{R}^d$ with $\mathcal{D}_B(E) > 0$ and ω is a mapping from E to a topological space X equipped with a nonzero finite Borel measure μ, let us say that $\omega(E)$ is *ball-uniformly distributed in* X if for any open ball U in X one has $\mathcal{D}_B(\omega^{-1}(U)) / \mathcal{D}_B(E) = \mu(U)/\mu(X)$.

Theorem B$'$. *Let $u\colon \mathbb{R}^d \longrightarrow \mathbb{R}^l$ be a bounded GP-mapping. There exist bounded polynomial surfaces $\mathcal{P}_1, \ldots, \mathcal{P}_k \subset \mathbb{R}^l$ of degree $\leq \mathrm{wgt}(u)$ and a partition $\mathbb{R}^d = \mathcal{Z}_* \cup \bigcup_{j=1}^k \mathcal{Z}_j$ such that $\mathcal{D}_B(\mathcal{Z}_*) = 0$ and for every $j \in \{1, \ldots, k\}$ the set $u(\mathcal{Z}_j)$ is ball-uniformly distributed on $\mathcal{P}_j$ with respect to $\mu_{\mathcal{P}_j}$. The coefficients of $\mathcal{P}_1, \ldots, \mathcal{P}_k$ belong to the ring generated over $\mathbb{Q}$ by the constant terms of the polynomials involved in u.*

0.39. The rest of the paper consists of two sections. In Section 1 we introduce coordinates on a nilmanifold, describe how sub-nilmanifolds look in these coordinates and use a modification of Theorem A, Theorem A*, to prove Theorems B, B$'$, C and D. The most difficult part of the paper is Section 2, which is devoted to proving Theorem A*. This section is purely algebraic; we deal there with the nilpotent group of upper triangular matrices over an arbitrary ring, and prove an algebraic generalization of Theorem A*, formulated in 2.41.

1. Coordinates on nilmanifolds and properties of generalized polynomial mappings

We start this section with a summary of some classical facts about nilpotent Lie groups and nilmanifolds. For more details see [Mal].

1.1. From now on we will only deal with connected nilpotent Lie groups, which will suffice for our goals. Let G be a connected simply-connected nilpotent Lie group of nilpotency class δ (which means that in the *lower central series* $G_1 = G$, $G_2 = [G_1, G]$, $G_3 = [G_2, G]$, $\ldots$ the subgroup $G_{\delta+1}$ is trivial) and let Γ be a discrete cocompact subgroup of G. The compact homogeneous space $X = G/\Gamma$ is called *a nilmanifold*, and we will say that X has *nilpotency class* δ.

1.2. For any $g \in G$ there exists a unique one-parameter subgroup $\{g^t\}_{t\in\mathbb{R}}$ in G such that $g^1 = g$. G has *a basis* compatible with Γ, that is, an ordered set $\{e_1,\dots,e_k\} \subset \Gamma$ having the following properties: $\{e_1,\dots,e_k\}$ generates Γ; every element $g \in G$ is uniquely representable in the form $g = e_1^{a_1}\dots e_k^{a_k}$ where *the coordinates* $a_1,\dots,a_k$ are real numbers; if D_l is the group generated by $\{e_l^t,\dots,e_k^t\}_{t\in\mathbb{R}}$, $l = 1,\dots,k$, and $D_{k+1} = \{\mathbf{1}_G\}$, then for any $1 \le i < j \le k$, $[D_i, D_j] \subseteq D_{j+1}$; if $G = G_1 \supseteq G_2 \supseteq \dots \supseteq G_\delta \supseteq G_{\delta+1} = \{\mathbf{1}_G\}$ is the lower central series of G, then $\{e_1,\dots,e_k\} \cap G_l$ is a basis in G_l, $l = 1,\dots,\delta$.

For every $i \in \{1,\dots,k\}$ let δ_i denote the integer for which $e_i \in G_{\delta_i} \setminus G_{\delta_i+1}$.

1.3. In the coordinates $(a_1,\dots,a_k)$ the multiplication in G is given by polynomial formulas: if $g = e_1^{a_1}\dots e_i^{a_i}\dots e_k^{a_k}$ and $h = e_1^{b_1}\dots e_i^{b_i}\dots e_k^{b_k}$, then

$$gh = e_1^{a_1+b_1}\dots e_i^{a_i+b_i+p_i(a_1,\dots,a_{i-1},b_1,\dots,b_{i-1})}\dots e_k^{a_k+b_k+p_k(a_1,\dots,a_{k-1},b_1,\dots,b_{k-1})} \tag{1.1}$$

and

$$g^t = e_1^{a_1 t}\dots e_i^{a_i t+q_i(a_1,\dots,a_{i-1},t)}\dots e_k^{a_k t+q_k(a_1,\dots,a_{k-1},t)},\ t \in \mathbb{R}, \tag{1.2}$$

where, for each $i = 2,\dots,k$, p_i is a polynomial in $2(i-1)$ variables with rational coefficients which takes integer values on $\mathbb{Z}^{2(i-1)}$ and q_i is a polynomial in i variables with rational coefficients which takes integer values on $\mathbb{Z}^i$. (See [Mal].)

1.4. For each $i = 2,\dots,k$ one has $\deg p_i, \deg q_i \le \delta_i$. Moreover, $\deg p_i\big(a_1^{\delta_1},\dots,a_k^{\delta_k}, b_1^{\delta_1},\dots,b_k^{\delta_k}\big) \le \delta_i$. It follows that if $S_1,\dots,S_k,R_1,\dots,R_k$ are polynomials with $\deg S_i$, $\deg R_i \le \delta_i$ for all $i = 1,\dots,k$, then $p_i\big(S_1,\dots,S_k,R_1,\dots,R_k\big) \le \delta_i$, $i = 1,\dots,k$. (See [L1]; in the terminology of [L1] the multiplication in G is a continuous polynomial mapping of lc-degree $\le (1,2,\dots,\delta)$.)

1.5. *The coordinate mapping* $\tilde\tau\colon G \longrightarrow \mathbb{R}^k$, $g = e_1^{a_1}\!\dots e_k^{a_k} \mapsto (a_1,\dots,a_k)$, is a diffeomorphism satisfying $\tilde\tau(\Gamma) = \mathbb{Z}^k$. "The cube" $Q = \tilde\tau^{-1}\big([0,1)^k\big) \subset G$ is the fundamental domain for X, which means that for any $g \in G$ there exists a unique $\gamma \in \Gamma$ such that $\tilde\tau(g\gamma) \in [0,1)^k$. Indeed, put $\gamma_0 = \mathbf{1}_G$, and if $\gamma_{i-1} \in \Gamma$ is such that $g\gamma_i = e_1^{x_1}\!\dots e_{i-1}^{x_{i-1}} e_i^{b_i}\!\dots e_k^{b_k}$ with $x_1,\dots,x_{i-1} \in [0,1)$, put $\gamma_i = \gamma_{i-1}e_i^{-[b_i]}$. Then $g\gamma_i = g\gamma_{i-1}e_i^{-[b_i]} = e_1^{x_1}\!\dots e_{i-1}^{x_{i-1}} e_i^{x_i} e_{i+1}^{c_{i+1}}\!\dots e_k^{c_k}$ with $x_i = b_i - [b_i] \in [0,1)$. For $\gamma = \gamma_k$ we therefore have $g\gamma = e_1^{x_1}\!\dots e_k^{x_k}$ with $x_1,\dots,x_k \in [0,1)$.

For $g \in G$ define $\chi(g) = g\gamma \in Q$ and $\tau(g) = \tilde\tau(\chi(g)) = (x_1,\dots,x_k) \in [0,1)^k$. The mapping $\tau\colon G \longrightarrow [0,1)^k$ factors to a one-to-one mapping $X \longrightarrow [0,1)^k$, which is a diffeomorphism on $\tau^{-1}\big((0,1)^k\big)$ but is discontinuous at the points of $\tau^{-1}\big([0,1)^k \setminus (0,1)^k\big)$. τ transfers (the completion of) the Haar measure on X to the Lebesgue measure on $[0,1)^k$. For $1 \le i \le k$ let τ_i be the i-th coordinate of τ. We will refer to $\tau = (\tau_1,\dots,\tau_k)$ as to *a coordinate mapping* of X or *a coordinate system* on X.

1.6. We will now formulate a modification of Theorem A, which will be easier to prove. The idea is to obtain generalized polynomials as values of a coordinate function along the orbit of a point of a nilmanifold under a "polynomial" action of $\mathbb{Z}^d$ instead of a conventional action.

A polynomial mapping $\omega: \mathbb{Z}^d \longrightarrow G$ to a nilpotent group is a mapping of the form $\omega(n) = g_1^{p_1(n)} \cdots g_r^{p_r(n)}$, $n \in \mathbb{Z}^d$, where $g_1, \ldots, g_r \in G$ and $p_1, \ldots, p_r$ are polynomials.

1.7. Theorem A*. *A mapping* $u: \mathbb{Z}^d \longrightarrow [0,1)$ *is a generalized polynomial if and only if there exist a connected compact nilmanifold* X *equipped with a coordinate system* $(\tau_1, \ldots, \tau_k)$*, a polynomial mapping* $\omega: \mathbb{Z}^d \longrightarrow G$ *and an index* $i \in \{1, \ldots, k\}$ *such that* $\tau_i(g(n)) = u(n)$*,* $n \in \mathbb{Z}^d$*.*

The "if" part of Theorem A* is almost obvious (see 1.14 below), whereas the proof of the "only if" part is very complicated and occupies Section 2.

1.8. We will now explain how Theorem A follows from Theorem A*. It is shown in [L3] that, given a connected compact nilmanifold $X = G/\Gamma$, a polynomial mapping $\omega: \mathbb{Z}^d \longrightarrow G$ and a point $x \in X$, one can find another connected compact nilmanifold $\widetilde{X} = \widetilde{G}/\widetilde{\Gamma}$ with a continuous mapping $\pi: \widetilde{X} \longrightarrow X$, a homomorphism $\varphi: \mathbb{Z}^d \longrightarrow \widetilde{G}$ and a point $\tilde{x} \in \widetilde{X}$ such that $\omega(n)x = \pi(\varphi(n)\tilde{x})$ for all $n \in \mathbb{Z}^d$. It is also shown in [L3] (and follows from the results in [Le] or [Sh2]) that the closure $\overline{\varphi(\mathbb{Z}^d)\tilde{x}}$ of the orbit of $\tilde{x}$ under the action of φ is a (not necessarily connected) subnilmanifold Y of $\widetilde{X}$ and the sequence $\{\varphi(n)x\}_{n \in \mathbb{Z}^d}$ is well distributed on Y. Moreover, the action φ is ergodic on Y, for any coordinate τ on X, the function $\tau \circ \pi$ is piecewise polynomial on $\widetilde{X}$, and the restriction of $\tau \circ \pi$ is piecewise polynomial on Y (see [L4]).

1.9. As a matter of fact we are going to obtain an extension of Theorem A* which will be applicable to various classes of polynomial mappings to G: continuous polynomial flows, polynomial mappings with zero constant term, etc. We will thus consider a more general situation. Let $\mathcal{A}$ be a ring of real-valued functions on a set $\mathcal{Z}$. We will call *an* $\mathcal{A}$*-mapping* any mapping $\omega: \mathcal{Z} \longrightarrow G$ of the form $\omega(z) = g_1^{\alpha_1(z)} \cdots g_r^{\alpha_r(z)}$, $z \in \mathcal{Z}$, where $g_1, \ldots, g_r \in G$ and $\alpha_1, \ldots, \alpha_r \in \mathcal{A}$. If $\{e_1, \ldots, e_k\}$ is a basis in G, then, since the multiplication in G is polynomial, any $\mathcal{A}$-mapping $\omega: \mathcal{Z} \longrightarrow G$ can be written in terms of this basis: $\omega(z) = e_1^{\alpha'_1(z)} \cdots e_k^{\alpha'_k(z)}$, $z \in \mathcal{Z}$, with $\alpha'_1, \ldots, \alpha'_k \in \mathcal{A}$. We will denote the set of $\mathcal{A}$-mappings to G by $G(\mathcal{A})$.

For $\delta \in \mathbb{N}$ let us define

$$\mathfrak{N}_\delta(\mathcal{A}) = \Big\{\beta: \mathcal{Z} \longrightarrow [0,1) : \begin{array}{l}\text{there exist a nilmanifold } X = G/\Gamma \text{ of nilpotency class} \le \delta \\ \text{equipped with a coordinate system } (\tau_1, \ldots, \tau_k),\ \omega \in G(\mathcal{A}) \\ \text{and } i \in \{1, \ldots, k\} \text{ such that } \beta = \tau_i \circ \omega\Big\}.\end{array}$$

1.10. Let $\mathcal{A}$ be a ring of real-valued functions on a set $\mathcal{Z}$. We will call the minimal algebra of real-valued functions on $\mathcal{Z}$ which contains $\mathcal{A}$ and is closed under the operation of taking the integer part *the bracket extension of* $\mathcal{A}$ and denote it by $\mathfrak{B}(\mathcal{A})$. More exactly, $v \in \mathfrak{B}(\mathcal{A})$ if

$v \in \mathcal{A}$,
or $v = v_1 + v_2$ where $v_1, v_2 \in \mathfrak{B}(\mathcal{A})$,
or $v = v_1 v_2$ where $v_1, v_2 \in \mathfrak{B}(\mathcal{A})$,
or $v = \pm[w]$ where $w \in \mathfrak{B}(\mathcal{A})$.
We define $\mathfrak{B}^{\mathbf{o}}(\mathcal{A}) = \bigl\{u \in \mathfrak{B}(\mathcal{A}) : \mathrm{Ran}(u) \in [0,1)\bigr\} = \bigl\{u - [u] : u \in \mathfrak{B}(\mathcal{A})\bigr\}$.

1.11. *The weight* $\mathrm{wgt}(v)$ of $v \in \mathfrak{B}(\mathcal{A})$ is defined by
$\mathrm{wgt}(v) = 1$ if $v \in \mathcal{A}$;
$\mathrm{wgt}(v) = \max\{\mathrm{wgt}(v_1), \mathrm{wgt}(v_2)\}$ if $v = v_1 + v_2$;
$\mathrm{wgt}(v) = \mathrm{wgt}(v_1) + \mathrm{wgt}(v_2)$ if $v = v_1 v_2$;
$\mathrm{wgt}(v) = \mathrm{wgt}(w)$ if $v = \pm[w]$.
For $\delta \in \mathbb{N}$ we define $\mathfrak{B}_\delta(\mathcal{A}) = \bigl\{v \in \mathfrak{B}(\mathcal{A}) : \mathrm{wgt}(v) \le \delta\bigr\}$ and $\mathfrak{B}^{\mathbf{o}}_\delta(\mathcal{A}) = \mathfrak{B}_\delta(\mathcal{A}) \cap \mathfrak{B}^{\mathbf{o}}(\mathcal{A})$.

1.12. From the definition of weight we immediately have:

Lemma. *Let p be a polynomial in k variables, $n_1, \dots, n_k \in \mathbb{N}$, $\deg p\bigl(x_1^{n_1}, \dots, x_k^{n_k}\bigr) = n$ and $v_1, \dots, v_k \in \mathfrak{B}(\mathcal{A})$ satisfy $\mathrm{wgt}(v_i) \le n_i$, $i = 1, \dots, k$. Then $\mathrm{wgt}\bigl(p(v_1, \dots, v_k)\bigr) \le n$.*

1.13. Theorem A* is a special case of

Theorem $\mathbf{A}_1^*$. *For any ring $\mathcal{A}$ of real-valued functions and any $\delta \in \mathbb{N}$, $\mathfrak{N}_\delta(\mathcal{A}) = \mathfrak{B}^{\mathbf{o}}_\delta(\mathcal{A})$.*

The inclusion $\mathfrak{B}^{\mathbf{o}}_\delta(\mathcal{A}) \subseteq \mathfrak{N}_\delta(\mathcal{A})$ will be proved in Section 2.

1.14. Proof of the inclusion $\mathfrak{N}_\delta(\mathcal{A}) \subseteq \mathfrak{B}^{\mathbf{o}}_\delta(\mathcal{A})$. Let $X = G/\Gamma$ be a nilmanifold of nilpotency class $\le \delta$ with a coordinate system $(\tau_1, \dots, \tau_k)$, and $\omega \in G(\mathcal{A})$; we need to show that $\tau_i \circ \omega \in \mathfrak{B}_\delta(\mathcal{A})$ for all $i = 1, \dots, k$. Let $\{e_1, \dots, e_k\}$ be the basis in G which induces the coordinates $\tau_1, \dots, \tau_k$ on X. Define $\sigma_0 \colon \mathcal{Z} \longrightarrow G$, $\sigma_0 \equiv \mathbf{1}_G$. Assume that $\sigma_{i-1} \colon \mathcal{Z} \longrightarrow \Gamma$ is already defined so that

$$\omega(z)\sigma_{i-1}(z) = e_1^{\xi_1(z)} \dots e_{i-1}^{\xi_{i-1}(z)} e_i^{\beta_i(z)} \dots e_k^{\beta_k(z)},\ z \in \mathcal{Z}, \tag{1.3}$$

with $\xi_1(z), \dots, \xi_{i-1}(z) \in [0,1)$, $z \in \mathcal{Z}$. Define $\sigma_i \colon \mathcal{Z} \longrightarrow \Gamma$ by $\sigma_i(z) = \sigma_{i-1}(z) e_i^{-[\beta_i(z)]}$, $z \in \mathcal{Z}$. Then

$$\omega(z)\sigma_i(z) = \omega(z)\sigma_{i-1}(z) e_i^{-[\beta_i(z)]} = e_1^{\xi_1(z)} \dots e_{i-1}^{\xi_{i-1}(z)} e_i^{\xi_i(z)} e_{i+1}^{\zeta_{i+1}(z)} \dots e_k^{\zeta_k(z)},\ z \in \mathcal{Z}, \tag{1.4}$$

with $\xi_i(z) = \beta_i(z) - [\beta_i(z)] \in [0,1)$, $z \in \mathcal{Z}$.

Now put $\chi(\omega) = \omega\sigma_k$. Then $\mathrm{Ran}(\chi(\omega)) \subseteq Q = \tilde{\tau}^{-1}([0,1)^k)$, so that $\tau_i \circ \omega = \tilde{\tau}_i \circ \chi(\omega)$ and we have $\tau_i \circ \omega = \tilde{\tau}_i \circ \chi(\omega) = \xi_i$, $i = 1, \dots, k$. We have to show that $\xi_1, \dots, \xi_k \in \mathfrak{B}^{\mathbf{o}}_\delta(\mathcal{A})$.

Assume by induction on i that in the formula (1.3), $\xi_1, \dots, \xi_{i-1}, \beta_i, \dots, \beta_k \in \mathfrak{B}(\mathcal{A})$, and, in the notation of 1.2, $\mathrm{wgt}(\xi_j) \le \delta_j$, $j = 1, \dots, i-1$, and $\mathrm{wgt}(\beta_j) \le \delta_j$, $j = i, \dots, k$. Then $\xi_i = \beta_i - [\beta_i] \in \mathfrak{B}^{\mathbf{o}}(\mathcal{A})$, and $\mathrm{wgt}(\xi_i) = \mathrm{wgt}(\beta_i) \le \delta_i \le \delta$, so $\xi_i \in \mathfrak{B}^{\mathbf{o}}_\delta(\mathcal{A})$. By 1.3, the functions $\zeta_{i+1}, \dots, \zeta_k$ in formula (1.4) are given by polynomial expressions in $\xi_1, \dots, \xi_{i-1}, \beta_i, \dots, \beta_k$ and $[\beta_i]$, hence $\zeta_{i+1}, \dots, \zeta_k \in \mathfrak{B}(\mathcal{A})$, and by 1.4 and Lemma 1.12, $\mathrm{wgt}(\zeta_j) \le \delta_j$, $j = i+1, \dots, k$. ∎

1.15. Let us consider now vector-valued functions. For a ring $\mathcal{A}$ of real-valued functions on a set $\mathcal{Z}$ and $\delta, l \in \mathbb{N}$ let us define

$$\mathfrak{N}^l_\delta(\mathcal{A}) = \Big\{\beta: \mathcal{Z} \longrightarrow [0,1)^l : \text{there exist a nilmanifold } X = G/\Gamma \text{ of nilpotency class } \delta \text{ equipped with a coordinate system } (\tau_1, \ldots, \tau_k),\ \omega \in G(\mathcal{A}) \text{ and } i_1, \ldots, i_l \in \{1, \ldots, k\} \text{ such that } \beta = (\tau_{i_1}, \ldots, \tau_{i_l})\circ\omega\Big\}.$$

Lemma. *$\beta = (\beta_1, \ldots, \beta_l) \in \mathfrak{N}^l_\delta(\mathcal{A})$ iff $\beta_j \in \mathfrak{N}_\delta(\mathcal{A})$ for all $j = 1, \ldots, l$.*

Proof. If $\beta = (\beta_1, \ldots, \beta_l) \in \mathfrak{N}^l_\delta(\mathcal{A})$ then $\beta_1, \ldots, \beta_l \in \mathfrak{N}_\delta(\mathcal{A})$ by definition. Assume that for each $j = 1, \ldots, l$ one has $\beta_j \in \mathfrak{N}_\delta(\mathcal{A})$, that is, there exist a nilmanifold $X_j = G_j/\Gamma_j$ of nilpotency class $\leq \delta$ with a coordinate system $(\tau_{j,1}, \ldots, \tau_{j,k_j})$, $\omega_j \in G(\mathcal{A})$ and $i_j \in \{1, \ldots, k_j\}$ such that $\beta_j = \tau_{i_j}\circ\omega_j$. Define $G = G_1 \times \ldots \times G_l$, $\Gamma = \Gamma_1 \times \ldots \times \Gamma_l$, $X = G/\Gamma = X_1 \times \ldots \times X_l$ and $\omega = (\omega_1, \ldots, \omega_l): \mathcal{Z} \longrightarrow G$. Then X is a nilmanifold of nilpotency class $\leq \delta$, $\omega \in G(\mathcal{A})$, $(\tau_{1,1}, \ldots, \tau_{l,k_l})$ is a coordinate system on X, and we have $\beta = (\beta_1, \ldots, \beta_l) = (\tau_{1,i_1}, \ldots, \tau_{l,i_l})\circ\omega$. ■

1.16. We have now the multidimensional extension of Theorem A^*_1:

Theorem A^*_l. *For any ring $\mathcal{A}$ of real-valued functions and any $\delta, l \in \mathbb{N}$, $\mathfrak{N}^l_\delta(\mathcal{A}) = (\mathfrak{B}^o_\delta(\mathcal{A}))^l$.*

1.17. Let G be a connected simply-connected nilpotent Lie group of nilpotency class δ, Γ be a discrete subgroup of G and $X = G/\Gamma$. Let $\pi: G \longrightarrow X$ be the natural projection, $\pi(g) = g\Gamma \in X$. Any closed subgroup of G is then a simply-connected nilpotent Lie group. *A sub-nilmanifold of X* is a closed subset Y of X of the form $Y = \pi(bH) = b\pi(H)$, where H is a connected closed subgroup of G and $b \in G$. Thus, Y is a translate of $\pi(H) = H/(\Gamma \cap H)$ and hence, has the natural structure of a nilmanifold.

An element $g \in G$ is said to be *rational* if $g^n \in \Gamma$ for some $n \in \mathbb{N}$. Given a coordinate system $(\tilde{\tau}_1, \ldots, \tilde{\tau}_k)$ on G, the coordinates of a rational element g of G are rational, $\tilde{\tau}_1(g), \ldots, \tilde{\tau}_k(g) \in \mathbb{Q}$. (See [L4].) We will say that a sub-nilmanifold Y of X is *rational* if it is of the form $Y = \pi(gH)$ with rational $g \in G$.

1.18. When $\mathcal{A}$ is the ring of real-valued polynomials, the $\mathcal{A}$-mappings (see 1.9) to a nilpotent Lie group G are called *polynomial mappings*. The following theorem is proved in [L3] and [L4]:

Theorem. *Let $\omega: \mathbb{Z}^d \longrightarrow G$ be a polynomial mapping. There exists a subgroup $\mathcal{Z}$ of finite index m in $\mathbb{Z}^d$, with cosets $\mathcal{Z}_1 = \mathcal{Z}, \mathcal{Z}_2, \ldots, \mathcal{Z}_m$, such that for each $i = 1, \ldots, m$ the sequence $\big\{\pi(\omega(z)) : z \in \mathcal{Z}_i\big\}$ is well distributed on a sub-nilmanifold Y_i of X with respect to the Haar measure on Y_i. If $\omega(0) = \mathbf{1}_G$, the sub-nilmanifolds $Y_1, \ldots, Y_m$ are all rational.*

1.19. We want now to determine how a sub-nilmanifold Y of X looks in coordinates on X; we will show that, up to a subset of Y of zero measure, it is a union of several polynomial surfaces.

1.20. Let $\{e_1,\ldots,e_k\} \in \Gamma$ be a basis in G and $\tilde{\tau}: G \longrightarrow \mathbb{R}^k$ be the corresponding coordinate mapping. Let H be a closed connected subgroup of G such that $\Gamma \cap H$ is uniform in H, and let $\{c_1,\ldots,c_s\} \subset H \cap \Gamma$ be a basis in H. We have a diffeomorphism $\tilde{\eta}: H \longrightarrow \mathbb{R}^s$, $c_1^{y_1} \cdots c_s^{y_s} \mapsto (y_1,\ldots,y_s)$ with $\tilde{\eta}(\Gamma \cap H) = \mathbb{Z}^s$.

One has $H = \{c_1^{y_1} \cdots c_s^{y_s}\}_{y_1,\ldots,y_s \in \mathbb{R}}$, and by formulas (1.1) and (1.2),

$$H = \left\{e_1^{S_1(y_1,\ldots,y_s)} \ldots e_k^{S_k(y_1,\ldots,y_s)}\right\}_{y_1,\ldots,y_s \in \mathbb{R}},$$

where, by 1.3, $S_1,\ldots,S_k$ are polynomials on $\mathbb{R}^s$. By 1.4, $\deg S_i \le \delta_i$, $i = 1,\ldots,k$. Since $e_1,\ldots,e_s \in \Gamma$, the polynomials $S_1,\ldots,S_k$ take on integer values on $\mathbb{Z}^s$ and hence have rational coefficients. In the commutative diagram

$$\begin{array}{ccc} H & \overset{Id_H}{\longrightarrow} & G \\ \tilde{\eta}\downarrow & & \downarrow\tilde{\tau} \\ \mathbb{R}^s & \overset{\mathcal{R}_H}{\longrightarrow} & \mathbb{R}^k \end{array}$$

the immersion $\mathcal{R}_H = \tilde{\tau} \circ \tilde{\eta}^{-1}: \mathbb{R}^s \longrightarrow \mathbb{R}^k$ is $(S_1,\ldots,S_k)$, and so, is a polynomial mapping with rational coefficients. That is, informally, in coordinates H is an s-dimensional rational polynomial surface of degree $\le \delta$.

1.21. Let $g \in G$, $g = e_1^{b_1} \cdots e_k^{b_k}$; the coset gH can be written as

$$\begin{aligned} gH &= \left\{e_1^{b_1} \ldots e_k^{b_k} \cdot e_1^{S_1(y_1,\ldots,y_s)} \ldots e_k^{S_k(y_1,\ldots,y_s)}\right\}_{y_1,\ldots,y_s \in \mathbb{R}} \\ &= \left\{e_1^{R_1(y_1,\ldots,y_s)} \ldots e_k^{R_k(y_1,\ldots,y_s)}\right\}_{y_1,\ldots,y_s \in \mathbb{R}}, \end{aligned}$$

where, by 1.3 and 1.4, $R_1,\ldots,R_k$ are polynomials with $\deg R_i \le \delta_i$, $i = 1,\ldots,k$, and coefficients in the ring $\mathfrak{R}$ generated by $\mathbb{Q}$ and $b_1,\ldots,b_k$. In the commutative diagram

$$\begin{array}{ccc} H & \overset{g \cdot Id_H}{\longrightarrow} & G \\ \tilde{\eta}\downarrow & & \downarrow\tilde{\tau} \\ \mathbb{R}^s & \overset{\mathcal{R}_{gH}}{\longrightarrow} & \mathbb{R}^k \end{array}$$

the insertion $\mathcal{R}_{gH} = \tilde{\tau} \circ (g \cdot \tilde{\eta}^{-1}) = (R_1,\ldots,R_k): \mathbb{R}^s \times \mathbb{Z}^r \longrightarrow \mathbb{R}^k$ is therefore a polynomial mapping of degree $\le \delta$ with coefficients from $\mathfrak{R}$.

1.22. Now let Y be the sub-nilmanifold $\pi(gH) \subseteq X$. Let $\eta: Y \longrightarrow [0,1)^s$ be the coordinate system on Y which corresponds to $\tilde{\eta}|_H$ and let $\tau: X \longrightarrow [0,1)^k$ be the coordinate system on X corresponding to $\tilde{\tau}$. In the commutative diagram

$$\begin{array}{ccc} Y & \subseteq & X \\ \eta\downarrow & & \downarrow\tau \\ [0,1)^s & \overset{\mathcal{R}_Y}{\longrightarrow} & [0,1)^k \end{array}$$

the insertion $\mathcal{R}_Y = \tau \circ \eta^{-1}$ is the composition of $\hat{\mathcal{R}}_Y = \mathcal{R}_{gE}|_{[0,1)^s}: [0,1)^s \longrightarrow \mathbb{R}^k$ and of "the projection" $\hat{\pi} = \tau \circ \pi \circ \tilde{\tau}^{-1}: \mathbb{R}^k \longrightarrow [0,1)^k$:

$$\begin{array}{ccc} G & \overset{\pi}{\longrightarrow} & X \\ \tilde{\tau}\downarrow & & \downarrow\tau \\ \mathbb{R}^k & \overset{\hat{\pi}}{\longrightarrow} & [0,1)^k \end{array}$$

Let $Q = \tilde{\tau}^{-1}([0,1)^k) \subset G$, then G is represented as the disjoint union $\bigcup_{\gamma\in\Gamma} Q\gamma$. For $\gamma \in \Gamma$ let $C_\gamma = \tilde{\tau}(Q\gamma)$, then $\mathbb{R}^k$ is the disjoint union $\bigcup_{\gamma\in\Gamma} C_\gamma$. Let $M_\gamma\colon \mathbb{R}^k \longrightarrow \mathbb{R}^k$ be defined by $M_\gamma(x) = \tilde{\tau}(\tilde{\tau}^{-1}(x)\gamma)$; by (1.1), the mapping M_γ is polynomial with rational coefficients. Then $C_\gamma = M_\gamma([0,1)^k)$, and $\hat{\pi}|_{C_\gamma} = M_{\gamma^{-1}}|_{C_\gamma}$.

Let $\gamma_1,\ldots,\gamma_N \in \Gamma$ be such that $\hat{\mathcal{R}}_Y([0,1)^s) \subseteq \bigcup_{j=1}^N C_{\gamma_j}$ and let $\mathcal{L}_j = \hat{\mathcal{R}}_Y^{-1}(C_{\gamma_j})$, $j = 1,\ldots,N$. Then $[0,1)^s$ is the disjoint union $\bigcup_{j=1}^N \mathcal{L}_j$. Let $j \in \{1,\ldots,N\}$. The restriction on $\mathcal{L}_j$ of $\mathcal{R}_Y$ is $\mathcal{R}_j = M_{\gamma^{-1}}\circ\hat{\mathcal{R}}_Y|_{\mathcal{L}_j}$, which is a polynomial mapping with coefficients from $\mathfrak{R}$, and $\mathcal{L}_j$ is defined by $\mathcal{L}_j = \mathcal{R}_j^{-1}([0,1)^k) \cap [0,1)^s$. Since the coordinates $R_1,\ldots,R_k$ of $\hat{\mathcal{R}}_Y$ satisfy $\deg R_i \le \delta_i$, $i = 1,\ldots,k$, by 1.4 we have $\deg \mathcal{R}_j \le \delta$.

1.23. We arrive at the following result.

Theorem. *Let $Y = \pi(gH)$ be a connected sub-nilmanifold of a connected nilmanifold X of nilpotency class δ and let $\tau\colon X \longrightarrow [0,1)^k$ and $\eta\colon Y \longrightarrow [0,1)^s$ be coordinate systems on X and Y. The mapping $\mathcal{R}_Y = \tau\circ\eta^{-1}\colon [0,1)^s \longrightarrow [0,1)^k$ is piecewise polynomial in the following sense: there are one-to-one polynomial mappings $\mathcal{R}_1,\ldots,\mathcal{R}_N\colon \mathbb{R}^s \longrightarrow \mathbb{R}^k$ of degree $\le \delta$ such that the sets $\mathcal{L}_j = \mathcal{R}_j^{-1}([0,1)^k) \cap [0,1)^s$, $j = 1,\ldots,N$, partition the cube $[0,1)^s$, and for each $j = 1,\ldots,N$ one has $\mathcal{R}_Y|_{\mathcal{L}_j} = \mathcal{R}_j$. The coefficients of $\mathcal{R}_1,\ldots,\mathcal{R}_N$ are contained in the ring $\mathfrak{R}$ generated by $\mathbb{Q}$ and the coordinates of g.*

1.24. We return now to generalized polynomials. Let $u\colon \mathbb{Z}^d \longrightarrow \mathbb{R}^l$ be a bounded generalized polynomial mapping; we may assume that $\mathrm{Ran}(u) \subseteq [0,1)^l$. Let $\delta = \mathrm{wgt}(u)$ and let $\mathcal{A}$ be the ring of real-valued functions on $\mathbb{Z}^d$ generated by the polynomials involved in u, so that $u \in (\mathfrak{B}^{\mathrm{o}}_\delta(\mathcal{A}))^l$. By Theorem A_l^* in 1.16, $u \in \mathfrak{N}^l_\delta(\mathcal{A})$. This means that there exist a nilmanifold $X = G/\Gamma$ of nilpotency class $\le \delta$ with $\pi\colon G \longrightarrow X$ being the natural projection, a coordinate system $\tau = (\tau_1,\ldots,\tau_k)\colon X \longrightarrow [0,1)^k$, $\omega \in G(\mathcal{A})$ and $n_1,\ldots,n_l \in \{1,\ldots,k\}$ such that $u = (\tau_{n_1},\ldots,\tau_{n_l})\circ\pi\circ\omega$. Let $\rho(x_1,\ldots,x_k) = (x_{n_1},\ldots,x_{n_l})$, then $u = \rho\circ\tau\circ\pi\circ\omega$:

$$u : \mathbb{Z}^d \xrightarrow{\omega} G \xrightarrow{\pi} X \xrightarrow{\tau} [0,1)^k \xrightarrow{\rho} [0,1)^l.$$

Since ω is a polynomial mapping, by Theorem 1.18 there exist a subgroup $\mathcal{Z}$ with cosets $\mathcal{Z}_1 = \mathcal{Z}, \mathcal{Z}_2,\ldots,\mathcal{Z}_m$ in $\mathbb{Z}^d$ and connected sub-nilmanifolds $Y_1,\ldots,Y_m$ of X such that for each $i = 1,\ldots,m$ the sequence $\{\pi(\omega(z)) : z \in \mathcal{Z}_i\}$ is well distributed on Y_i.

Fix $i \in \{1,\ldots,m\}$, and let $\eta_i\colon Y_i \longrightarrow [0,1)^s$ be a coordinate system on Y_i. Then by 1.14, $v_i = \eta_i\circ\pi\circ\omega|_{\mathcal{Z}_i}\colon \mathcal{Z}_i \longrightarrow [0,1)^s$ is a GP-mapping of weight $\le \delta$, and we have $u|_{\mathcal{Z}_i} = \rho\circ\tau\circ\eta_i^{-1}\circ v_i = \rho\circ\mathcal{R}_{Y_i}\circ v_i$:

$$\begin{tikzcd} u|_{\mathcal{Z}_i} : \mathcal{Z}_i \arrow[r, "\pi\circ\omega"] \arrow[dr, "v_i"'] & Y_i \arrow[r, phantom, "\subseteq"] \arrow[d, "\eta_i"] & X \arrow[d, "\tau"] \\ & {[0,1)^s} \arrow[r, "\mathcal{R}_{Y_i}"] & {[0,1)^k} \arrow[d, "\rho"] \\ & & {[0,1)^l}. \end{tikzcd}$$

Since the coordinate mapping η_i maps the Haar measure on Y_i to the Lebesgue measure λ on $[0,1)^s$ and is continuous on an open subset of Y_i of full measure, $v_i(\mathcal{Z}_i)$ is well distributed in $[0,1)^s$. By Theorem 1.23 there exist a partition $[0,1)^s=\bigcup_{j=1}^{N_i}\mathcal{L}_{i,j}$ and polynomial mappings $\mathcal{R}_{i,1},\dots,\mathcal{R}_{i,N_i}\colon\mathbb{R}^s\longrightarrow\mathbb{R}^k$ of degree $\leq\delta$ such that $\mathcal{R}_{Y_i}|_{\mathcal{L}_{i,j}}=\mathcal{R}_{i,j}$ and $\mathcal{L}_{i,j}=\mathcal{R}_{i,j}^{-1}([0,1)^k)\cap[0,1)^s$, $j=1,\dots,N_i$. For $j\in\{1,\dots,N_i\}$ let $\mathcal{Z}_{i,j}=v_i^{-1}(\mathcal{L}_{i,j})\subseteq\mathcal{Z}_i$ and let $\mathcal{P}_{i,j}=\rho\circ\mathcal{R}_{i,j}$; then $\mathcal{P}_{i,j}$ is a polynomial mapping $\mathbb{R}^s\longrightarrow\mathbb{R}^l$ of degree $\leq\delta$ and $u|_{\mathcal{Z}_{i,j}}=\mathcal{P}_{i,j}\circ v_i|_{\mathcal{Z}_{i,j}}$:

$$u|_{\mathcal{Z}_{i,j}}:\mathcal{Z}_{i,j}\xrightarrow{v_i}\mathcal{L}_{i,j}\xrightarrow{\mathcal{P}_{i,j}}[0,1)^l.$$

The coefficients of the polynomials $\mathcal{R}_{i,j}$ (and thus, of $\mathcal{P}_{i,j}$) belong to a certain ring of real numbers which we will now describe. Let $\tilde{\tau}\colon G\longrightarrow\mathbb{R}^k$ be the coordinate mapping of G corresponding to the coordinate system τ on X, and let $\tilde{\tau}\circ\omega(z)=\big(\alpha_1(z),\dots,\alpha_k(z)\big)$, $z\in\mathbb{Z}^d$, where $\alpha_1,\dots,\alpha_k$ are polynomials from $\mathcal{A}$. Then $\tilde{\tau}(\omega(0))=\big(\alpha_1(0),\dots,\alpha_k(0)\big)$, and $\alpha_1(0),\dots,\alpha_k(0)$ belong to the ring $\mathfrak{F}$ generated by $\mathbb{Q}$ and the constant terms of the polynomials involved in u. Define $\omega'(z)=\omega(0)^{-1}\omega(z)$, so that $\omega'(0)=\mathbf{1}_G$. By Theorem 1.18, the components $Y_1',\dots,Y_m'$ of $\overline{\big\{\pi(\omega'(z)):z\in\mathbb{Z}^d\big\}}$ are rational sub-nilmanifolds of X, that is, $Y_i'=\pi(g_iH_i)$ where g_i have rational coordinates. Thus the components $Y_1,\dots,Y_m$ of $\overline{\big\{\pi(\omega(z)):z\in\mathbb{Z}^d\big\}}$ have form $Y_i=\omega(0)Y_i'=\pi\big(\omega(0)g_iH_i\big)$, $i=1,\dots,m$. By Theorem 1.23, the coefficients of $\mathcal{R}_{i,j}$, $i=1,\dots,m$, $j=1,\dots,N_i$, are contained in the ring generated by $\mathbb{Q}$, the coordinates of g_i and the coordinates $\alpha_1(0),\dots,\alpha_k(0)$ of $\tilde{\tau}(\omega(0))$, and so, in $\mathfrak{F}$.

1.25. Let us say that a generalized polynomial mapping $v\colon\mathbb{Z}^d\longrightarrow\mathbb{R}^s$ is *regular* if $v(\mathbb{Z}^d)$ is well distributed in $[0,1)^s$. We can now summarize the above in the following theorem:

Theorem B*. *Let $u\colon\mathbb{Z}^d\longrightarrow\mathbb{R}^l$ be a bounded GP-mapping and let $\mathfrak{F}$ be the ring generated by the constant terms of the polynomials involved in u. There exist a subgroup $\mathcal{Z}$ of finite index m in $\mathbb{Z}^d$ with cosets $\mathcal{Z}_1=\mathcal{Z},\mathcal{Z}_2,\dots,\mathcal{Z}_m$, an integer $s\in\mathbb{N}$, regular GP-mappings $v_i\colon\mathcal{Z}_i\longrightarrow[0,1)^s$, $i=1,\dots,m$, of weight $\leq\mathrm{wgt}(u)$, partitions $[0,1)^s=\bigcup_{j=1}^{N_i}\mathcal{L}_{i,j}$, $i=1,\dots,m$, where each $\mathcal{L}_{i,j}$ is defined by polynomial inequalities of degree $\leq\mathrm{wgt}(u)$ with coefficients from $\mathfrak{F}$, and polynomial mappings $\mathcal{P}_{i,j}\colon\mathbb{R}^s\longrightarrow\mathbb{R}^l$, $i=1,\dots,m$, $j=1,\dots,N_i$, of degree $\leq\mathrm{wgt}(u)$ with coefficients from $\mathfrak{F}$, such that for $\mathcal{Z}_{i,j}=v_i^{-1}(\mathcal{L}_{i,j})$ one has $u|_{\mathcal{Z}_{i,j}}=\mathcal{P}_{i,j}\circ v_i$, $i=1,\dots,m$, $j=1,\dots,N_i$.*

1.26. Proof of Theorem B. We keep the notation from 1.25. Fix $i\in\{1,\dots,m\}$. For each $j\in\{1,\dots,N_i\}$ the set $\mathcal{L}_{i,j}\subseteq[0,1)^s$ is defined by a collection of polynomial inequalities, and thus either the interior $\mathcal{L}_{i,j}^{\mathbf{o}}$ of $\mathcal{L}_{i,j}$ is nonempty, $\lambda(\mathcal{L}_{i,j})>0$ and $v_i(\mathcal{Z}_{i,j})$ is well distributed in $\mathcal{L}_{i,j}$, or $\mathcal{L}_{i,j}^{\mathbf{o}}$ is empty, $\lambda(\mathcal{L}_{i,j})=0$ and $\mathcal{Z}_{i,j}$ has zero density in $\mathcal{Z}_i$. Let us assume that $\mathcal{L}_{i,1}^{\mathbf{o}},\dots,\mathcal{L}_{i,r_i}^{\mathbf{o}}\neq\emptyset$ and $\mathcal{L}_{r_i+1}^{\mathbf{o}},\dots,\mathcal{L}_{N_i}^{\mathbf{o}}=\emptyset$, and define $\mathcal{Z}_{i,*}=\mathcal{Z}_{i,r_i+1}\cup\ldots\cup\mathcal{Z}_{i,N_i}$. Then $\mathcal{Z}_{i,*}\cup\bigcup_{j=1}^{r_i}\mathcal{Z}_{i,j}$ is a partition of $\mathcal{Z}_i$, the set $\mathcal{Z}_{i,*}$ has zero density in $\mathcal{Z}_i$ and for each $j=1,\dots,r_i$ the sequence $u(\mathcal{Z}_{i,j})$ is well distributed on the polynomial surface $\mathcal{P}_{i,j}(\mathcal{L}_{i,j})$ of degree $\leq\mathrm{wgt}(u)$. It suffices now to put $\mathcal{Z}_*=\bigcup_{i=1}^m\mathcal{Z}_{i,*}$. ■

1.27. Corollary of the proof. *In the notation of Theorem B, the set $\mathcal{Z}_*$ is contained in the set $\mathcal{W}=w^{-1}(0)$ of zeroes of a generalized polynomial $w\colon\mathbb{Z}^d\longrightarrow\mathbb{R}$. Moreover, $\mathcal{D}(\mathcal{W})=0$.*

Proof. Note that, in the proof of Theorem B in 1.26, for any $i \in \{1,\ldots,m\}$ and $j > r_i$ the set $\mathcal{L}_{i,j}$ is contained in the set of zeroes of a nonzero polynomial $S_{i,j}$ on $\mathbb{R}^s$. Put $S_i = \prod_{j=r_i+1}^{N_i} S_{i,j}$ and define a generalized polynomial w by $w|_{\mathcal{Z}_i} = S_i \circ v_i$, $i = 1,\ldots,m$. Since S_i is a nonzero polynomial and $v_i(\mathcal{Z}_i)$ is well distributed in $[0,1)^s$, ${w|_{\mathcal{Z}_i}}^{-1}(0)$ has zero density in $\mathcal{Z}_i$. ∎

1.28. Let us demonstrate the calculation of the distribution of values of a generalized polynomial by carrying it out on one simple example. Let α be an irrational number; consider the generalized polynomial $u(n) = \langle\!\langle \frac{1}{2}\alpha^2 n^2 - \alpha n[\alpha n] \rangle\!\rangle$, $n \in \mathbb{Z}$. We are going to read u off a nilmanifold.

The group $G = \left\{ \left(\begin{smallmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{smallmatrix}\right),\ a,b,c \in \mathbb{R} \right\}$ of 3×3 upper triangular matrices is a connected simply-connected nilpotent Lie group, and $\Gamma = \left\{ \left(\begin{smallmatrix} 1 & m & k \\ 0 & 1 & l \\ 0 & 0 & 1 \end{smallmatrix}\right),\ m,k,l \in \mathbb{Z} \right\}$ is a discrete uniform subgroup of G; let $X = G/\Gamma$ and $\pi: G \longrightarrow X$ be the natural projection. Let $e_{1,2} = \left(\begin{smallmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$, $e_{1,3} = \left(\begin{smallmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$ and $e_{2,3} = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{smallmatrix}\right)$, then $\{e_{2,3}, e_{1,2}, e_{1,3}\}$ is a basis in G such that $e_{2,3}^c e_{1,2}^a e_{1,3}^b = \left(\begin{smallmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{smallmatrix}\right)$, $a,b,c \in \mathbb{R}$. Thus, in the basis $\{e_{2,3}, e_{1,2}, e_{1,3}\}$ the coordinates of a matrix $A = \left(\begin{smallmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{smallmatrix}\right) \in G$ are $\tilde{\tau}(A) = (c,a,b)$. The fundamental domain in G is $Q = \left\{ \left(\begin{smallmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{smallmatrix}\right),\ a,b,c \in [0,1) \right\}$, and for a matrix $A = \left(\begin{smallmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{smallmatrix}\right) \in G$ the corresponding matrix $\chi(A) \in Q$ with $\pi(A) = \pi(\chi(A))$ is

$$\chi(A) = \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -[c] \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -[a] & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -[b-a[c]] \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \langle\!\langle a \rangle\!\rangle & \langle\!\langle b-a[c] \rangle\!\rangle \\ 0 & 1 & \langle\!\langle c \rangle\!\rangle \\ 0 & 0 & 1 \end{pmatrix}.$$

For the polynomial sequence $\omega(n) = \begin{pmatrix} 1 & \alpha n & \frac{1}{2}\alpha^2 n^2 \\ 0 & 1 & \alpha n \\ 0 & 0 & 1 \end{pmatrix}$ in G we will therefore have $\tau_3(\omega(n)) = \langle\!\langle \frac{1}{2}\alpha^2 n^2 - \alpha n[\alpha n] \rangle\!\rangle = u(n)$, $n \in \mathbb{Z}$.

Consider the subgroup $H = \left\{ \begin{pmatrix} 1 & a & \frac{1}{2}a^2 \\ 0 & 1 & a \\ 0 & 0 & 1 \end{pmatrix},\ a \in \mathbb{R} \right\}$ of G; we have $\omega(n) \in H$ for all $n \in \mathbb{Z}$. $\Gamma \cap H$ is uniform in H and hence $Y = \pi(H)$ is a 1-dimensional sub-nilmanifold of X. Define the coordinate mapping $\tilde{\eta}: H \longrightarrow \mathbb{R}$ by $\tilde{\eta}\left(\begin{pmatrix} 1 & a & \frac{1}{2}a^2 \\ 0 & 1 & a \\ 0 & 0 & 1 \end{pmatrix}\right) = \frac{1}{2}a$, so that $\tilde{\eta}(\Gamma \cap H) = \mathbb{Z}$. The mapping $\mathcal{R}_H = \tilde{\tau} \circ \tilde{\eta}^{-1}: \mathbb{R} \longrightarrow \mathbb{R}^3$ has form $\mathcal{R}_H(y) = (2y, 2y, 2y^2)$. Let $\eta: Y \longrightarrow [0,1)$ be the coordinate mapping corresponding to $\tilde{\eta}$, then the sequence $v(n) = \eta(\pi(\omega(n))) = \langle\!\langle \frac{1}{2}\alpha n \rangle\!\rangle$ is well distributed in $[0,1)$, and so, $\pi(\omega(n))$ is well distributed in Y.

Let $C = [0,1)^3$. Then $\mathcal{R}_H\big([0,\frac{1}{2})\big) \subset C$. Define $\mathcal{R}_1 = \mathcal{R}_H|_{[0,\frac{1}{2})} = (2y, 2y, 2y^2)$. For $\gamma = \left(\begin{smallmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{smallmatrix}\right)$ one has

$$M_\gamma(c,a,b) = \tilde{\tau}\left(\left(\begin{smallmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{smallmatrix}\right) \left(\begin{smallmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{smallmatrix}\right) \right) = (c+1, a+1, b+a),$$

and so

$$C_\gamma = M_\gamma(C) = \big\{ (c,a,b) : 1 \le c < 2,\ 1 \le a < 2,\ a-1 \le b < a \big\},$$

and $\mathcal{R}_H\big(\big[\frac{1}{2},1\big)\big) \subset C_\gamma$. Define $\mathcal{R}_2 = M_{\gamma^{-1}} \circ \mathcal{R}_H|_{[\frac{1}{2},1)}$, $\mathcal{R}_2(y) = (2y-1, 2y-1, 2y^2-2y+1)$. Let $\mathcal{P}_1$, $\mathcal{P}_2$ be the 3-rd coordinates of $\mathcal{R}_1$ and $\mathcal{R}_2$ respectively, $\mathcal{P}_1(y) = 2y^2$ and $\mathcal{P}_2(y) = 2y^2 - 2y + 1$.

We have arrived at the following: the interval $[0,1)$ is partitioned into the pieces $\mathcal{L}_1 = \big[0, \frac{1}{2}\big)$ and $\mathcal{L}_2 = \big[\frac{1}{2}, 1\big)$, a mapping $\mathcal{P}: [0,1) \longrightarrow [0,1)$ is defined by $\mathcal{P}|_{\mathcal{L}_1} = \mathcal{P}_1$ and $\mathcal{P}|_{\mathcal{L}_2} = \mathcal{P}_2$, that is, $\mathcal{P}(y) = \begin{cases} 2y^2, & y \in \big[0, \frac{1}{2}\big) \\ 2y^2 - 2y + 1, & y \in \big[\frac{1}{2}, 1\big) \end{cases}$, and we have $u(n) = \mathcal{P}\big(\big\langle\!\big\langle \frac{1}{2}\alpha n \big\rangle\!\big\rangle\big)$, $n \in \mathbb{Z}$. The sequence $\big\{\frac{1}{2}\alpha n\big\}_{n \in \mathbb{Z}}$, is well distributed in $[0,1)$ with respect to the Lebesgue measure dy; hence, $u(n)$, $n \in \mathbb{Z}$, is well distributed in $[0,1)$ with respect to the measure

$$\mathcal{P}_*(dy) = \begin{cases} \frac{dx}{2\sqrt{2x}}, & x \in \big[0, \frac{1}{2}\big) \\ \frac{dx}{2\sqrt{2x-1}}, & x \in \big[\frac{1}{2}, 1\big). \end{cases}$$

1.29. We now move the discussion to the recurrence properties of generalized polynomials, dealt with in Theorems C and D of the introduction. The following fact is obtained in [L3]:

Proposition. *Let $X = G/\Gamma$ be a nilmanifold, $\pi: G \longrightarrow X$ be the natural projection and $\omega: \mathbb{Z}^d \longrightarrow G$ be a polynomial mapping. Then $\pi(\omega(0))$ is an IP*-limit for $\pi \circ \omega$.*

It follows that $\pi(\omega(z'))$ is an IP^*_+-limit for $\pi \circ \omega$ for any $z' \in \mathbb{Z}^d$; indeed, $\pi(\omega(z'))$ is an IP*-limit for the mapping $\pi \circ \omega'$ where $\omega'(z) = \omega(z + z')$.

1.30. Proof of Theorem C. Let $u: \mathbb{Z}^d \longrightarrow \mathbb{R}^l$ be a bounded GP-mapping; we may assume that $\mathrm{Ran}(u) \subseteq [0,1)^l$. Like in 1.24, find a nilmanifold $X = G/\Gamma$ with the natural projection $\pi: G \longrightarrow X$, a coordinate system $\tau: X \longrightarrow [0,1)^k$, a polynomial mapping $\omega: \mathbb{Z}^d \longrightarrow G$ and $n_1, \ldots, n_l \in \{1, \ldots, k\}$ such that $u = (\tau_{n_1}, \ldots, \tau_{n_l}) \circ \pi \circ \omega$. By Theorem 1.18 there exist pairwise disjoint connected sub-nilmanifolds $Y_1, \ldots, Y_M$ of X such that $\overline{\pi(\omega(\mathbb{Z}^d))} = \bigcup_{i=1}^M Y_i$ and for each $i = 1, \ldots, M$ the sequence $\big\{\pi(\omega(z)) : z \in (\pi \circ \omega)^{-1}(Y_i)\big\}$ is well distributed on Y_i.

Fix $i \in \{1, \ldots, M\}$, let $Y = Y_i$, $\mathcal{Z} = (\pi \circ \omega)^{-1}(Y)$ and let $\eta: Y \longrightarrow [0,1)^s$ be a coordinate system in Y. Then $(\tau_{n_1}, \ldots, \tau_{n_l})|_Y = \mathcal{P}_Y \circ \eta$ and $u|_{\mathcal{Z}} = \mathcal{P}_Y \circ \eta \circ \pi \circ \omega|_{\mathcal{Z}}$,

$$u|_{\mathcal{Z}} : \mathcal{Z} \xrightarrow{\omega} G \xrightarrow{\pi} X \supseteq Y \xrightarrow{\eta} [0,1)^s \xrightarrow{\mathcal{P}_Y} [0,1)^l,$$

with $\mathcal{P}_Y$ of the following form: there is a partition $[0,1)^s = \bigcup_{j=1}^N \mathcal{L}_j$ such that for each $j = 1, \ldots, N$, $\mathcal{L}_j$ is defined by polynomial inequalities and $\mathcal{P}_Y|_{\mathcal{L}_j}$ is a polynomial mapping. The sequence $\pi(\omega(\mathcal{Z}))$ is well distributed in Y and η is continuous on an open set of full measure in Y, thus η is continuous at almost all points of $\pi(\omega(\mathcal{Z}))$. Next, $\eta(\pi(\omega(\mathcal{Z})))$ is well distributed in $[0,1)^s$ and $\mathcal{P}_Y$ is continuous on the union of the interiors of $\mathcal{L}_j$, $j = 1, \ldots, N$, which is an open set of full measure in $[0,1)^s$; thus $\mathcal{P}_Y$ is continuous at almost all points of $\eta(\pi(\omega(\mathcal{Z})))$. Therefore, $\mathcal{P}_Y \circ \eta$ is continuous at almost all points of $\pi(\omega(\mathcal{Z}))$. By Proposition 1.29 and since $Y_1, \ldots, Y_M$ are compact and disjoint, all points of $\pi(\omega(\mathcal{Z}))$ are IP^*_+-limits for $\pi \circ \omega|_{\mathcal{Z}}$. Hence, almost all points of $u(\mathcal{Z}) = \mathcal{P}_Y(\eta(\pi(\omega(\mathcal{Z}))))$ are IP^*_+-limits for $u|_{\mathcal{Z}} = \mathcal{P}_Y \circ \eta \circ \pi \circ \omega|_{\mathcal{Z}}$. ■

1.31. Proof of Theorem D. Let $u: \mathbb{Z}^d \longrightarrow \mathbb{R}^l$ be a GP-mapping such that all the polynomials involved in u have zero constant term and let $\mathcal{A}$ be the ring generated by these polynomials. All the polynomials from $\mathcal{A}$ also vanish at 0. Define $v = u - [u]$, then $\mathrm{Ran}(v) \subseteq [0,1)^l$.

Again, find a nilmanifold $X = G/\Gamma$ with the natural projection $\pi: G \longrightarrow X$, a coordinate system $\tau: X \longrightarrow [0,1)^k$, $\omega \in G(\mathcal{A})$ and $n_1, \ldots, n_l \in \{1, \ldots, k\}$ such that $v = (\tau_{n_1}, \ldots, \tau_{n_l}) \circ \pi \circ \omega$. Since $\omega \in G(\mathcal{A})$, $\omega(0) = \mathbf{1}_G$. Let $o = \pi(\mathbf{1}_G)$, then $\pi \circ \omega(0) = o$.

Let σ be the natural insertion $[0,1)^k \longrightarrow \mathbb{T}^k = \mathbb{R}^k/\mathbb{Z}^k$, that is, the restriction on $[0,1)^k$ of the natural projection $\mathbb{R}^k \longrightarrow \mathbb{T}^k$. Then $\sigma \circ \tau: X \longrightarrow \mathbb{T}^k$ maps o to $0 \in \mathbb{T}^k$ and is continuous at o. Indeed, if a sequence $\{x_j\}_{j=1}^{\infty}$ in X converges to o, then a limit point of $\{\tau(x_j)\}_{j=1}^{\infty}$ may only be a vertex of the cube $[0,1]^k$, and all the vertices of $[0,1]^k$ are mapped by σ to $0 \in \mathbb{T}^k$.

The mapping $\tilde{u} = u(\mathrm{mod}\, 1) = v(\mathrm{mod}\, 1): \mathbb{Z}^d \longrightarrow \mathbb{T}^l$ is the composition of $\sigma \circ \tau \circ \pi \circ \omega$ and the projection $\rho: \mathbb{T}^k \longrightarrow \mathbb{T}^l$, $\rho(y_1, \ldots, y_k) = (y_{n_1}, \ldots, y_{n_l})$:

$$\tilde{u} : \mathbb{Z}^d \xrightarrow{\omega} G \xrightarrow{\pi} X \xrightarrow{\tau} [0,1)^k \xrightarrow{\sigma} \mathbb{T}^k \xrightarrow{\rho} \mathbb{T}^l.$$

By Proposition 1.29, o is an IP*-limit for $\pi \circ \omega$. Hence, $0 \in \mathbb{T}^k$ is an IP*-limit for $\sigma \circ \tau \circ \pi \circ \omega$ and $0 \in \mathbb{T}^l$ is an IP*-limit for $\tilde{u}$. ∎

1.32. Proof of Theorem B′. One just has, in the proof of Theorem B, to switch from $\mathbb{Z}^d$ to $\mathbb{R}^d$ and to substitute the following theorem, which is a special case of a result in [Sh1], for Theorem 1.18.

Theorem. *Let $X = G/\Gamma$ be a nilmanifold, $\pi: G \longrightarrow X$ be the natural projection, and let $\omega: \mathbb{R}^d \longrightarrow G$ be a polynomial mapping. There exists a connected sub-nilmanifold Y of X such that $\pi(\omega(\mathbb{R}^d))$ is ball-uniformly distributed in Y with respect to the Haar measure in Y.*

∎

1.33. We conclude this section with a topological version of Theorem 0.35:

Proposition. *Let $X = G/\Gamma$ be a nilmanifold, $\pi: G \longrightarrow X$ be the natural projection, let $\mathcal{A}$ be the algebra of generalized polynomials on $\mathbb{Z}^d$ and $\omega \in G(\mathcal{A})$, that is, $\omega(z) = g_1^{u_1(z)} \ldots g_r^{u_r(z)}$, $z \in \mathbb{Z}^d$, with $g_1, \ldots, g_r \in G$ and $u_1, \ldots, u_r$ being generalized polynomials. Then $\lim_{N\to\infty} \frac{1}{|\Phi_N|} \sum_{z\in\Phi_N} f(\pi(\omega(z)))$ exists for any $f \in C(X)$ and any Følner sequence $\{\Phi_N\}_{N=1}^{\infty}$ in $\mathbb{Z}^d$.*

Proof. Let $\tau: X \longrightarrow [0,1)^k$ be a coordinate system. By Theorem 1.16, $u = \tau \circ \pi \circ \omega \in \mathfrak{B}(\mathcal{A})^k$, and in the case under consideration $\mathfrak{B}(\mathcal{A}) = \mathcal{A}$. So, $u: \mathbb{Z}^d \longrightarrow [0,1)^k$ is a GP-mapping. Let $f \in C(X)$; since τ^{-1} is continuous, $\hat{f} = f \circ \tau^{-1}$ is a continuous function on $[0,1)^k$. By Corollary 0.27, $\lim_{N\to\infty} \frac{1}{|\Phi_N|} \sum_{z\in\Phi_N} f(\pi(\omega(z))) = \lim_{N\to\infty} \frac{1}{|\Phi_N|} \sum_{z\in\Phi_N} \hat{f}(u(z))$ exists for any Følner sequence $\{\Phi_N\}_{N=1}^{\infty}$ in $\mathbb{Z}^d$. ∎

2. Any generalized polynomial can be read off a nilmanifold

2.1. We remind the reader that $\mathcal{A}$ stands for a ring of real-valued functions on a set $\mathcal{Z}$; $\mathfrak{B}(\mathcal{A})$ is a bracket extension of $\mathcal{A}$, that is, the minimal ring of functions containing $\mathcal{A}$ and closed under the taking-a-bracket operation; $\mathfrak{B}^{\mathbf{o}}(\mathcal{A}) \subset \mathfrak{B}(\mathcal{A})$ consists of functions with range in $[0,1)$; and $\mathfrak{N}(\mathcal{A})$ is the set of functions which can be "read off" a nilmanifold, that is, functions of the form $\tau\circ\omega$ where ω is an $\mathcal{A}$-mapping from $\mathcal{Z}$ to a connected nilpotent group G and τ is a coorinate on the nilmanifold $X = G/\Gamma$.

2.2. To establish the inclusion $\mathfrak{B}^{\mathbf{o}}(\mathcal{A}) \subseteq \mathfrak{N}(\mathcal{A})$ in Theorem A_1^* we will utilize the group of upper triangular matrices. For $d \in \mathbb{N}$ let $M_d = \left\{ \begin{pmatrix} 1 & a_{1,2} & \cdots & a_{1,d} \\ & 1 & \cdots & a_{2,d} \\ & \mathbf{O} & \ddots & \vdots \\ & & & 1 \end{pmatrix},\ a_{i,j} \in \mathbb{R} \right\}$. M_d is a connected simply-connected nilpotent Lie group, and $\Gamma_d = \left\{ \begin{pmatrix} 1 & n_{1,2} & \cdots & n_{1,d} \\ & 1 & \cdots & n_{2,d} \\ & \mathbf{O} & \ddots & \vdots \\ & & & 1 \end{pmatrix},\ n_{i,j} \in \mathbb{Z} \right\}$ is a discrete uniform subgroup of M_d.

We will refer to elements of M_d as to *upper triangular matrices.* Dealing with matrices from M_d we will often ignore their diagonal and under-diagonal entries and therefore assume that their entries are indexed by the pairs (i,j) with $1 \leq i < j \leq d$.

2.3. Let $\mathcal{A}$ be a ring of real-valued functions on a set $\mathcal{Z}$. The set of $\mathcal{A}$-mappings $\mathcal{Z} \longrightarrow M_d$ is then the set $M_d(\mathcal{A}) = \left\{ \begin{pmatrix} 1 & \alpha_{1,2} & \cdots & \alpha_{1,d} \\ & 1 & \cdots & \alpha_{2,d} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix},\ \alpha_{i,j} \in \mathcal{A} \right\}$ of upper-triangular matrices with entries from $\mathcal{A}$. Let $\mathfrak{B}(\mathcal{A})$ be the bracket extension of $\mathcal{A}$; for any matrix $P \in M_d(\mathcal{A})$ one can construct a matrix $\chi(P) \in M_d(\mathfrak{B}(\mathcal{A}))$ which is equal to P modulo Γ_d and takes values in the fundamental domain of M_d. Our goal is to show that for any $u \in \mathfrak{B}(\mathcal{A})$ there exist $d \in \mathbb{N}$, a basis in M_d and a matrix $P \in M_d(\mathcal{A})$ such that the $(1,d)$-coordinate of the matrix $\chi(P)$ in this basis is equal to $u - [u]$.

2.4. For $1 \leq i < j \leq d$, let $E_{i,j}$ be the upper triangular matrix whose only nonzero entry is 1 at the (i,j)-th position:

$$E_{i,j} = \begin{matrix} \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ & 1 & \cdots & 0 & \cdots & 0 \\ & & \ddots & & & \vdots \\ & & & 1 & \cdots & \vdots \\ & & & & \ddots & \vdots \\ & & & & & 1 \end{pmatrix} & \begin{matrix} \\ \\ \\ i \\ \\ \\ \end{matrix} \\ \qquad\qquad j & \end{matrix}$$

The set $\{E_{i,j}\}_{1\leq i<j\leq d}$ is a basis in M_d compatible (see 1.2) with Γ_d, and for $a \in \mathbb{R}$ we have $E_{i,j}^a = \begin{matrix} \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ & 1 & \cdots & 0 & \cdots & 0 \\ & & \ddots & & & \vdots \\ & & & a & \cdots & \vdots \\ & & & & \ddots & \vdots \\ & & & & & 1 \end{pmatrix} & \begin{matrix} \\ \\ \\ i. \\ \\ \\ \end{matrix} \\ \qquad\qquad j & \end{matrix}$

2.5. On first approach it may seem that in the basis $\{E_{i,j}\}_{1\leq i<j\leq d}$ the coordinates of a matrix $\begin{pmatrix} 1 & a_{1,2} & \cdots & a_{1,d} \\ & 1 & \cdots & a_{2,d} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_d$ are its entries $a_{i,j}$, and that the corresponding fundamental

domain for M_d/Γ_d is the set of matrices with all $a_{i,j} \in [0,1)$. However, this is not true, or, more exactly, is only true for a specific ordering of the basis $\{E_{i,j}\}_{1\le i<j\le d}$. Indeed, if an ordering is such that for some $1 \le k < n < l \le d$ the element $E_{k,n}$ of the basis precedes the element $E_{n,l}$, then the (k,l)-entry of the product $\prod\limits_{1\le i<j\le d} E_{i,j}^{a_{i,j}}$ computed with respect to this ordering contains, in addition to $a_{k,l}$, a summand of the form $a_{k,n}a_{n,l}$. Therefore, the coordinates of a matrix in the basis $\{E_{i,j}\}_{1\le i<j\le d}$ are equal to its entries only if the elements of the basis are ordered as follows:

$$\big(E_{d-1,d}, E_{d-2,d-1}, E_{d-2,d}, E_{d-3,d-2}, \dots, E_{2,d}, E_{1,2}, \dots, E_{1,d}\big).$$

Denote the corresponding order by $\prec$, that is, let $(i,j) \prec (k,l)$ if $i > k$, or $i = k$ and $j < l$. The product $\prod\limits_{1\le i<j\le d} E_{i,j}^{a_{i,j}}$ computed with respect to $\prec$ equals $\begin{pmatrix} 1 & a_{1,2} & \dots & a_{1,d} \\ & 1 & \dots & a_{2,d} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix}$.

2.6. The class of elements of $\mathfrak{B}(\mathcal{A})$ which are readable off M_d/Γ_d with respect to the basis $\{E_{i,j}\}_{1\le i<j\le d}$, ordered by the order $\prec$ defined in 2.5, is restricted to *nested* elements, that is, the elements whose representation does not contain products of brackets. Here is a rigorous definition: an element $u \in \mathfrak{B}(\mathcal{A})$ is nested if either $u \in \mathcal{A}$, or $u = \pm[v]$ where v is nested, or $\alpha[v]$ where v is nested and $\alpha \in \mathcal{A}$, or $u = u_1 + u_2$ where u_1, u_2 are nested. (Example: for $\alpha_i \in \mathcal{A}$, $\alpha_1\big[\alpha_2[\alpha_3] + \alpha_4[\alpha_5 + \alpha_6]\big] + \alpha_7[\alpha_8]$ is nested and $\alpha_1[\alpha_2][\alpha_3]$ is not.)

Given a matrix $P = \begin{pmatrix} 1 & \alpha_{1,2} & \dots & \alpha_{1,d} \\ & 1 & \dots & \alpha_{2,d} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_d(\mathcal{A})$, the matrix $\chi(P)$ is computed in the following way: for $1 \le k < l \le d$, if integer-valued functions $m_{i,j} \in \mathfrak{B}(\mathcal{A})$ have already been defined for all $(i,j) \prec (k,l)$, we put $P_{k,l} = P \cdot \big(\prod_{(i,j)\prec(k,l)} E_{i,j}^{m_{i,j}}\big)$ (where the product is computed with respect to $\prec$), $\xi_{i,j}^{k,l}$ be the (i,j)-entry of $P_{k,l}$ and $m_{k,l} = -[\xi_{k,l}^{k,l}]$. Then $\chi(P) = P \cdot \big(\prod_{(i,j)} E_{i,j}^{m_{i,j}}\big)$.

By induction on (k,l) assume that $\xi_{i,j}^{k,l} = \alpha_{i,j}$ for all $j \le k$ and that $\xi_{i,j}^{k,l}$ are nested for all $j > k$. Then

$$\big(P_{k,l}E_{k,l}^{m_{k,l}}\big)_{i,j} = \begin{cases} \xi_{i,j}^{k,l} \text{ if } j \ne l \\ \xi_{i,j}^{k,l} + \xi_{i,k}^{k,l} m_{k,l} = \xi_{i,j}^{k,l} - \alpha_{i,k}[\xi_{k,l}^{k,l}] \text{ if } j = l, \end{cases}$$

which is equal to $\alpha_{i,j}$ if $j \le k$ and which is nested if $j > k$.

This gives us the following proposition:

Proposition. *For $P \in M_d(\mathcal{A})$ all entries of $\chi(P)$ are nested elements of $\mathfrak{B}^{\mathbf{o}}(\mathcal{A})$.*

The converse is also true: any nested element of $\mathfrak{B}^{\mathbf{o}}(\mathcal{A})$ is obtainable as an entry of $\chi(P)$ for a suitable P. We omit the proof.

2.7. For a matrix $P \in M_d$ we will now compute the coordinates of $\chi(P)$ with respect to the basis $\{E_{i,j}^{\epsilon_{i,j}}\}_{1\le i<j\le d}$, $\epsilon_{i,j}=\pm 1$, taken in an arbitrary order $\prec$. Actually, $\prec$ cannot be completely arbitrary, since the elements $E_{i,j}^{\epsilon_{i,j}}$ taken in accordance with $\prec$ must form a basis in M_d in the sense of 1.2. We will say that a linear order $\prec$ on the set $\{(i,j)\}_{1\le i<j\le d}$ is *legal* if $(i,j)\preceq(k,l)$ whenever $i\ge k$ and $j\le l$.

Let $\prec$ be a legal order on $\{(i,j)\}_{1\le i<j\le d}$ and let $\epsilon_{i,j}=\pm 1$, $1\le i<j\le d$. Let $P\in M_d$, $P=\begin{pmatrix} 1 & a_{1,2} & \dots & a_{1,d} \\ & 1 & \dots & a_{2,d} \\ & & \ddots & \vdots \\ & & & 1\end{pmatrix}$; we will call $a_{i,j}$ *the (i,j)-entry* of P. P is representable in the form $P=\prod_{1\le i<j\le d}E_{i,j}^{\epsilon_{i,j}z_{i,j}}$, where the product is being computed in accordance with $\prec$; we will call $z_{i,j}$ *the (i,j)-coordinate of P.*

2.8. We start with finding recurrence formulas connecting the entries $a_{i,j}$ and the coordinates $z_{i,j}$ of P. For indices $(k,l)\preceq(i,j)$ let $\theta_{i,j}^{k,l}$ be the (i,j)-entry of $\prod_{(r,s)\prec(k,l)}E_{r,s}^{\epsilon_{i,j}z_{r,s}}$ and let $\theta_{i,j}=\theta_{i,j}^{i,j}$. Then

$$\theta_{i,j}^{k,l}=\sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}}\theta_{i,n}^{n,j}\theta_{n,j}+\sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}}a_{i,n}\theta_{n,j},$$

$$\theta_{i,j}=\theta_{i,j}^{i,j}=\sum_{(n,j)\prec(i,n)}\theta_{i,n}^{n,j}\theta_{n,j}+\sum_{(n,j)\succ(i,n)}a_{i,n}\theta_{n,j},$$

and $z_{i,j}=\epsilon_{i,j}(a_{i,j}-\theta_{i,j})$.

2.9. Now let $\chi(P)$ be the matrix in the fundamental domain of M_d corresponding to P, that is, $\chi(P)=P\cdot\big(\prod_{1\le i<j\le d}E_{i,j}^{m_{i,j}}\big)$ with all $m_{i,j}\in\mathbb{Z}$ so that $\chi(P)=\prod_{1\le i<j\le d}E_{i,j}^{\epsilon_{i,j}x_{i,j}}$ with all $x_{i,j}\in[0,1)$. We will compute the coordinates $x_{i,j}$ of $\chi(P)$. For an index (k,l) let

$$P_{k,l}=P\cdot\big(\prod_{(i,j)\prec(k,l)}E_{i,j}^{m_{i,j}}\big),$$

then

$$P_{k,l}=\big(\prod_{(i,j)\prec(k,l)}E_{i,j}^{\epsilon_{i,j}x_{i,j}}\big)E_{k,l}^{\xi_{k,l}}\big(\prod_{(i,j)\succ(k,l)}E_{i,j}^{v_{i,j}^{k,l}}\big)$$

for some $\xi_{k,l}$ and $v_{i,j}^{k,l}$, and one has $m_{k,l}=-\epsilon_{k,l}[\epsilon_{k,l}\xi_{k,l}]$ and $x_{k,l}=\epsilon_{k,l}(\xi_{k,l}+m_{k,l})=\epsilon_{k,l}\xi_{k,l}-[\epsilon_{k,l}\xi_{k,l}]$.

For $(i,j)\succeq(k,l)$ let $\varphi_{i,j}^{k,l}$ be the (i,j)-entry of $P_{k,l}$ and $\varphi_{i,j}=\varphi_{i,j}^{i,j}$. For $(i,j)\prec(k,l)$ the (i,j)-entry of $P_{k,l}$ is $\varphi_{i,j}+m_{i,j}$, thus

$$\varphi_{i,j}^{k,l}=a_{i,j}+\sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}}\varphi_{i,n}^{n,j}m_{n,j}+\sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}}(\varphi_{i,n}+m_{i,n})m_{n,j}$$

and

$$\varphi_{i,j}=a_{i,j}+\sum_{(n,j)\prec(i,n)}\varphi_{i,n}^{n,j}m_{n,j}+\sum_{(n,j)\succ(i,n)}(\varphi_{i,n}+m_{i,n})m_{n,j}.$$

For $(i,j) \succeq (k,l)$ let $\psi_{i,j}^{k,l}$ be the (i,j)-entry of $\prod\limits_{(i,j)\prec(k,l)} E_{r,s}^{\epsilon_{i,j}x_{r,s}}$ and $\psi_{i,j} = \psi_{i,j}^{i,j}$, then $\xi_{i,j} = (\varphi_{i,j} - \psi_{i,j})$. For $(i,j) \prec (k,l)$ the (i,j)-entry of $\prod\limits_{(r,s)\prec(k,l)} E_{r,s}^{\epsilon_{r,s}x_{r,s}}$ is $\varphi_{i,j} + m_{i,j}$, thus

$$\psi_{i,j}^{k,l} = \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}} \psi_{i,n}^{n,j}\epsilon_{n,j}x_{n,j} + \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}} (\varphi_{i,n} + m_{i,n})\epsilon_{n,j}x_{n,j}$$

and

$$\psi_{i,j} = \sum_{(n,j)\prec(i,n)} \psi_{i,n}^{n,j}\epsilon_{n,j}x_{n,j} + \sum_{(n,j)\succ(i,n)} (\varphi_{i,n} + m_{i,n})\epsilon_{n,j}x_{n,j}.$$

For $(k,l) \preceq (i,j)$ we define $\xi_{i,j}^{k,l} = \varphi_{i,j}^{k,l} - \psi_{i,j}^{k,l}$ and compute

$$\begin{aligned}
\xi_{i,j}^{k,l} &= a_{i,j} + \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}} \varphi_{i,n}^{n,j}m_{n,j} + \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}} (\varphi_{i,n} + m_{i,n})m_{n,j} - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}} \psi_{i,n}^{n,j}\epsilon_{n,j}x_{n,j} \\
&\qquad - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}} (\varphi_{i,n} + m_{i,n})\epsilon_{n,j}x_{n,j} \\
&= a_{i,j} + \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}} \left(-\varphi_{i,n}^{n,j}\epsilon_{n,j}[\epsilon_{n,j}\xi_{n,j}] - \psi_{i,n}^{n,j}(\xi_{n,j} - \epsilon_{n,j}[\epsilon_{n,j}\xi_{n,j}])\right) \\
&\qquad + \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}} \left(\varphi_{i,n} - \epsilon_{i,n}[\epsilon_{i,n}\xi_{i,n}]\right)\left(-\epsilon_{n,j}[\epsilon_{n,j}\xi_{n,j}] - (\xi_{n,j} - \epsilon_{n,j}[\epsilon_{n,j}\xi_{n,j}])\right) \\
&= a_{i,j} - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}} \xi_{i,n}^{n,j}\epsilon_{n,j}[\epsilon_{n,j}\xi_{n,j}] - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}} \psi_{i,n}^{n,j}\xi_{n,j} - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}} \varphi_{i,n}\xi_{n,j} \\
&\qquad + \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}} \epsilon_{i,n}[\epsilon_{i,n}\xi_{i,n}]\xi_{n,j}.
\end{aligned}$$

In particular,

$$\xi_{i,j} = a_{i,j} - \sum_{(n,j)\prec(i,n)} \xi_{i,n}^{n,j}\epsilon_{n,j}[\epsilon_{n,j}\xi_{n,j}] - \sum_{(n,j)\prec(i,n)} \psi_{i,n}^{n,j}\xi_{n,j} - \sum_{(n,j)\succ(i,n)} \varphi_{i,n}\xi_{n,j} + \sum_{(n,j)\succ(i,n)} \epsilon_{i,n}[\epsilon_{i,n}\xi_{i,n}]\xi_{n,j}$$

and $x_{i,j} = \epsilon_{i,j}\xi_{i,j} - [\epsilon_{i,j}\xi_{i,j}]$.

2.10. Once the above formulas have been obtained, we find ourselves in a purely algebraic context. The nature of entries of our matrices is no longer important to us, and we may assume that our matrices are defined over an arbitrary commutative ring. Moreover, we prefer to use the abstract algebraic language because we are going to deal with neither numbers nor functions, but with abstract expressions built from the elements of $\mathcal{A}$ by applying the operations of addition, multiplication and "taking the bracket". We will now introduce the necessary algebraic formalism.

2.11. Given a set S we will denote by $\mathbb{Z}[S]$ the free commutative ring generated by the set $\{[s] : s \in S\}$, that is, the ring of formal finite sums of the form $\sum_{i=1}^{l} m_i[s_1]\dots[s_n]$ with $m_i \in \mathbb{Z}$ and $s_1,\dots,s_n \in S$. For commutative rings R and Q let $R * Q$ be the commutative ring freely generated by R and Q, that is, $R * Q = R \oplus Q \oplus (R \otimes_{\mathbb{Z}} Q)$ with multiplication defined by $rq = r \otimes q$, $r_1(r \otimes q) = (r_1 r) \otimes q$, $q_1(r \otimes q) = r \otimes (q_1 q)$ and $(r_1 \otimes q_1)(r \otimes q) = (r_1 r) \otimes (q_1 q)$ for $r, r_1 \in R$, $q, q_1 \in Q$. We will write rq for $r \otimes q$.

Let $\mathcal{A}$ be a commutative ring; we will now construct an algebra $\mathfrak{B}$ which we will call *the bracket algebra over* $\mathcal{A}$. Put $\mathbf{B}_0 = \mathcal{A}$; if $\mathbf{B}_k$ is already defined, define $\mathbf{B}_k^{\mathbf{b}} = \mathbb{Z}[\mathbf{B}_k]$ and $\mathbf{B}_{k+1} = \mathbf{B}_k * \mathbf{B}_k^{\mathbf{b}}$. Let $\mathbf{B}^{\mathbf{b}} = \bigcup_{k=0}^{\infty} \mathbf{B}_k^{\mathbf{b}}$ and $\mathbf{B} = \bigcup_{k=0}^{\infty} \mathbf{B}_k$; a mapping $[\cdot]: \mathbf{B} \longrightarrow \mathbf{B}^{\mathbf{b}}$ is naturally defined. Let I be the ideal in $\mathbf{B}$ generated by the sets $\{[v] - v : v \in \mathbf{B}^{\mathbf{b}}\}$ and $\{[u+v] - [u] - v : u \in \mathbf{B},\ v \in \mathbf{B}^{\mathbf{b}}\}$; we define $\mathfrak{B} = \mathbf{B}/I$ and $\mathfrak{B}^{\mathbf{b}} = \mathbf{B}^{\mathbf{b}}/(I \cap \mathbf{B}^{\mathbf{b}})$. The mapping $[\cdot]: \mathfrak{B} \longrightarrow \mathfrak{B}^{\mathbf{b}}$ is well defined, identical on $\mathfrak{B}^{\mathbf{b}}$ and satisfies $[u+v] = [u] + v$ for any $u \in \mathfrak{B}$, $v \in \mathfrak{B}^{\mathbf{b}}$.

2.12. As an abelian group $\mathfrak{B}$ is spanned by elements of the form $a[v_1]\dots[v_m]$ where $a \in \mathcal{A}$, $m \geq 0$, $v_1,\dots,v_m \in \mathfrak{B}$, and elements of the form $[v_1]\dots[v_m]$ where $m \geq 1$, $v_1,\dots,v_m \in \mathfrak{B}$. We will call such elements *monomials.* The monomials of the form $[v_1]\dots[v_m]$ span $\mathfrak{B}^{\mathbf{b}}$; let $\mathfrak{B}^{\mathbf{t}}$ be the subgroup of $\mathfrak{B}$ spanned by the monomials of the form $a[v_1]\dots[v_m]$ with $a \in \mathcal{A}$. Then $\mathfrak{B} = \mathfrak{B}^{\mathbf{t}} \oplus \mathfrak{B}^{\mathbf{b}}$, and $\mathfrak{B}^{\mathbf{t}}$ is an ideal in $\mathfrak{B}$.

2.13. For $u \in \mathfrak{B}$ we define $\mathbf{t}(u) \in \mathfrak{B}^{\mathbf{t}}$ and $\mathbf{b}(u) \in \mathfrak{B}^{\mathbf{b}}$ so that $u = \mathbf{t}(u) + \mathbf{b}(u)$. From the definition of $\mathfrak{B}^{\mathbf{t}}$ and $\mathfrak{B}^{\mathbf{b}}$ we clearly have:

Lemma. *For $u_1, u_2, u \in \mathfrak{B}$ one has*
$\mathbf{t}(u_1 + u_2) = \mathbf{t}(u_1) + \mathbf{t}(u_2)$, $\quad \mathbf{b}(u_1 + u_2) = \mathbf{b}(u_1) + \mathbf{b}(u_2)$,
$\mathbf{t}(u_1 u_2) = \mathbf{t}(u_1)\,\mathbf{t}(u_2) + \mathbf{t}(u_1)\,\mathbf{b}(u_2) + \mathbf{b}(u_1)\,\mathbf{t}(u_2)$, $\quad \mathbf{b}(u_1 u_2) = \mathbf{b}(u_1)\,\mathbf{b}(u_2)$,
$\mathbf{t}([u]) = 0$ *and* $\mathbf{b}([u]) = [\mathbf{t}(u)] + \mathbf{b}(u)$.

2.14. We say that an expression $\sum_{i=1}^{k}\big(a_i[v_{i,1}]\dots[v_{i,m_i}]\big) + \sum_{i=1}^{l}\big(\pm[u_{i,1}]\dots[u_{i,n_i}]\big)$ is *reduced* if (i) the monomials $([v_{i,1}]\dots[v_{i,m_i}])$, $i = 1,\dots,k$, are all different so that no combining like terms is possible, (ii) equal monomials $([u_{i,1}]\dots[u_{i,n_i}])$, $i = 1,\dots,l$, occur with equal signs so that no cancellations are possible, (iii) all $v_{i,j}$, $u_{i,j}$ belong to $\mathfrak{B}^{\mathbf{t}}$ and are represented in the reduced form. Every element $u \in \mathfrak{B}$ is uniquely representable in the reduced form. When we are free to choose an expression representing an element of $\mathfrak{B}$ we will assume that this is the reduced representation of the element.

2.15. From now on let $\mathcal{A}$ be a ring and let $\mathfrak{B}$ be the bracket algebra over $\mathcal{A}$.

For $u \in \mathfrak{B}$ we define $[u]^1 = [u]$ and $[u]^{-1} = -[-u]$.

2.16. For $d \in \mathbb{N}$ let $M_d(\mathcal{A})$ be the group of upper triangular matrices with entries from $\mathcal{A}$. We will call an upper triangular matrix $\epsilon = (\epsilon_{i,j})_{1 \leq i < j \leq d}$ with $\epsilon_{i,j} = \pm 1$, $1 \leq i < j \leq d$, a *sign matrix*.

Given a matrix $P = \begin{pmatrix} 1 & a_{1,2} & \dots & a_{1,d} \\ & 1 & \dots & a_{2,d} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_d(\mathcal{A})$, a sign matrix $\epsilon = (\epsilon_{i,j})_{1 \leq i < j \leq d}$ and a legal order $\prec$ (see 2.7) on the set $\{(i,j)\}_{1 \leq i < j \leq d}$, we define $\varphi_{i,j}, \psi_{i,j}, \xi_{i,j}$ and $\varphi_{i,j}^{k,l}, \psi_{i,j}^{k,l}, \xi_{i,j}^{k,l} \in$

$\mathfrak{B}$ for $(i,j) \preceq (k,l)$ inductively by

$$
\begin{aligned}
\varphi_{i,j}^{k,l} &= a_{i,j} - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}} \varphi_{i,n}^{n,j}[\xi_{n,j}]^{\epsilon_{n,j}} - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}} \big(\varphi_{i,n} - [\xi_{i,n}]^{\epsilon_{i,n}}\big)[\xi_{n,j}]^{\epsilon_{n,j}}\\
\varphi_{i,j} = \varphi_{i,j}^{i,j} &= a_{i,j} - \sum_{(n,j)\prec(i,n)} \varphi_{i,n}^{n,j}[\xi_{n,j}]^{\epsilon_{n,j}} - \sum_{(n,j)\succ(i,n)} \big(\varphi_{i,n} - [\xi_{i,n}]^{\epsilon_{i,n}}\big)[\xi_{n,j}]^{\epsilon_{n,j}}\\
\psi_{i,j}^{k,l} &= \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}} \psi_{i,n}^{n,j}\big(\xi_{n,j} - [\xi_{n,j}]^{\epsilon_{n,j}}\big) + \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}} \big(\varphi_{i,n} - [\xi_{i,n}]^{\epsilon_{i,n}}\big)\big(\xi_{n,j} - [\xi_{n,j}]^{\epsilon_{n,j}}\big)\\
\psi_{i,j} = \psi_{i,j}^{i,j} &= \sum_{(n,j)\prec(i,n)} \psi_{i,n}^{n,j}\big(\xi_{n,j} - [\xi_{n,j}]^{\epsilon_{n,j}}\big) + \sum_{(n,j)\succ(i,n)} \big(\varphi_{i,n} - [\xi_{i,n}]^{\epsilon_{i,n}}\big)\big(\xi_{n,j} - [\xi_{n,j}]^{\epsilon_{n,j}}\big)\\
\xi_{i,j}^{k,l} = \varphi_{i,j}^{k,l} - \psi_{i,j}^{k,l} &= a_{i,j} - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}} \xi_{i,n}^{n,j}[\xi_{n,j}]^{\epsilon_{n,j}} - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}} \psi_{i,n}^{n,j}\xi_{n,j}\\
&\qquad + \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}} [\xi_{i,n}]^{\epsilon_{i,n}}\xi_{n,j} - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}} \varphi_{i,n}\xi_{n,j}\\
\xi_{i,j} = \varphi_{i,j} - \psi_{i,j} &= a_{i,j} - \sum_{(n,j)\prec(i,n)} \xi_{i,n}^{n,j}[\xi_{n,j}]^{\epsilon_{n,j}} - \sum_{(n,j)\prec(i,n)} \psi_{i,n}^{n,j}\xi_{n,j}\\
&\qquad + \sum_{(n,j)\succ(i,n)} [\xi_{i,n}]^{\epsilon_{i,n}}\xi_{n,j} - \sum_{(n,j)\succ(i,n)} \varphi_{i,n}\xi_{n,j}.
\end{aligned}
\tag{2.1}
$$

When it is not clear from the context for which matrix P, sign matrix ϵ and/or order $\prec$ we are computing the elements $\varphi_{i,j}^{k,l}, \psi_{i,j}^{k,l}, \xi_{i,j}^{k,l}$, we will write $\varphi_{i,j}^{k,l}(P,\epsilon,\prec), \psi_{i,j}^{k,l}(P,\epsilon,\prec)$, $\xi_{i,j}^{k,l}(P,\epsilon,\prec)$.

Notice that in formulas (2.1) the elements $\xi_{i,j}$ are computed in terms of $\xi_{r,s}$ with $i \le r < s \le j$. Therefore, when computing $\xi_{i,j}$ we may restrict ourselves to the submatrix of P indexed by $\{(r,s)\}_{i\le r<s\le j}$. We will say that the (r,s)-entry *does not affect* the (i,j)-entry if $r < i$ or $s > j$.

2.17. Our goal is to prove the following:

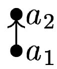
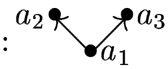
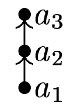
Proposition. *For any $u \in \mathfrak{B}^{\mathbf{t}}$ there exist $d \in \mathbb{N}$, $P \in M_d(\mathcal{A})$, a sign matrix $\epsilon = (\epsilon_{i,j})_{1\le i<j\le d}$ and a legal order $\prec$ on the set $\{(i,j)\}_{1\le i<j\le d}$ such that $\mathbf{t}\big(\xi_{1,d}(P,\epsilon,\prec)\big) = u$.*

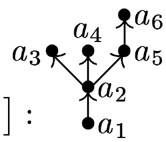
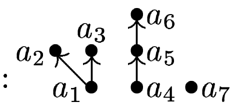
Let us remark that this proposition does not yet imply Theorem A_1^* because it says nothing about the nilpotency class of the group necessary for obtaining an element $u \in \mathfrak{B}^{\mathbf{t}}$. We will later formulate and prove a stronger statement, Theorem A**, from which Theorem A_1^* will follow.

2.18. Let us say that an element $p \in \mathfrak{B}$ is *simple* if it is constructible from elements of $\mathcal{A}$ without using the "+" sign. More exactly, p is simple if $p = a[p_1]\dots[p_m]$ where $a \in \mathcal{A}$, $m \ge 0$ and $p_1,\dots,p_m \in \mathfrak{B}$ are simple, or $p = [p_1]\dots[p_m]$ where $m \ge 1$ and $p_1,\dots,p_m \in \mathfrak{B}$ are simple. We will denote the set of simple elements of $\mathfrak{B}$ by $\mathfrak{S}$.

2.19. The elements of $\mathfrak{S}$ can be described by oriented graphs labeled by elements of $\mathcal{A}$. The following examples illustrate what we mean:

Examples.

$p_1 = a :$  $\quad p_2 = a_1[a_2] :$  $\quad p_3 = a_1[a_2][a_3] :$  $\quad p_4 = a_1[a_2[a_3]] :$ 

$p_5 = a_1\big[a_2[a_3][a_4][a_5[a_6]]\big] :$  $\quad q = \big[a_1[a_2][a_3]\big]\big[a_4[a_5[a_6]]\big][a_7] :$ 

(While we do not base our proofs on this graphic representation of elements of $\mathfrak{S}$, the reader may find it useful for the vizualization of the exposition.)

2.20. We will now subdivide $\mathfrak{S}$ into two subsets, $\mathfrak{S} = \mathfrak{S}^{\mathbf{t}} \cup \mathfrak{S}^{\mathbf{b}}$, where $\mathfrak{S}^{\mathbf{b}} = \mathfrak{S} \cap \mathfrak{B}^{\mathbf{b}}$ and $\mathfrak{S}^{\mathbf{t}} = \mathfrak{S} \cap \mathfrak{B}^{\mathbf{t}}$. Elements of $\mathfrak{S}^{\mathbf{t}}$ are of the form $p = a[p_1]\dots[p_m]$ with $a \in \mathcal{A}$, $m \geq 0$ and $p_1, \dots, p_m \in \mathfrak{S}^{\mathbf{t}}$, and will be referred to as *trees*, with *the root* a and *the branches* $p_1, \dots, p_m$.

Elements of $\mathfrak{S}^{\mathbf{b}}$ are of the form $p = [p_1]\dots[p_m]$ with $m \geq 1$ and $p_1, \dots, p_m \in \mathfrak{S}^{\mathbf{t}}$, and will be referred to as *bushes*, with *the branches* $p_1, \dots, p_m$.

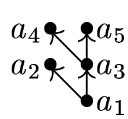
In the examples 2.19, $p_1, \dots, p_5$ are trees and q is a bush.

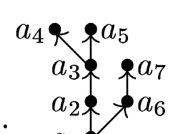
2.21. In the proof of Proposition 2.17 we will use a cumbersome induction based on the structure of the trees representing elements of $\mathfrak{S}$. We will now introduce some parameters of trees and bushes, that is, of elements of $\mathfrak{S}$.

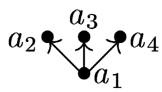
(i) *The weight* $\mathrm{wgt}(p)$ of $p \in \mathfrak{S}$ is the number of its vertices, that is,
$\mathrm{wgt}(p) = 1$ if $p \in \mathcal{A}$;
$\mathrm{wgt}(p) = 1 + \mathrm{wgt}(p_1) + \ldots + \mathrm{wgt}(p_m)$ for $p = a[p_1]\dots[p_m] \in \mathfrak{S}^{\mathbf{t}}$ with $a \in \mathcal{A}$ and $p_1, \dots, p_m \in \mathfrak{S}^{\mathbf{t}}$;
$\mathrm{wgt}(p) = \mathrm{wgt}(p_1) + \ldots + \mathrm{wgt}(p_m)$ for $p = [p_1]\dots[p_m] \in \mathfrak{S}^{\mathbf{b}}$ with $p_1, \dots, p_m \in \mathfrak{S}^{\mathbf{t}}$.

(ii) *The height* $\mathrm{hgt}(p)$ of $p \in \mathfrak{S}$ is
$\mathrm{hgt}(p) = 0$ if $p \in \mathcal{A}$;
$\mathrm{hgt}(p) = 1 + \max\big\{\mathrm{hgt}(p_1), \dots, \mathrm{hgt}(p_m)\big\}$ if $p = a[p_1]\dots[p_m] \in \mathfrak{S}^{\mathbf{t}}$ with $a \in \mathcal{A}$ and $p_1, \dots, p_m \in \mathfrak{S}^{\mathbf{t}}$;
$\mathrm{hgt}(p) = \max\big\{\mathrm{hgt}(p_1), \dots, \mathrm{hgt}(p_m)\big\}$ if $p = [p_1]\dots[p_m] \in \mathfrak{S}^{\mathbf{b}}$ with $p_1, \dots, p_m \in \mathfrak{S}^{\mathbf{t}}$.

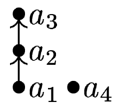
(iii) *The number of branches* $\mathrm{brn}(p)$ for $p \in P$ is defined by
$\mathrm{brn}(p) = 0$ if $p \in \mathcal{A}$,
$\mathrm{brn}(p) = m$ if $p = a[p_1]\dots[p_m] \in \mathfrak{S}^{\mathbf{t}}$ with $a \in \mathcal{A}$ and $p_1, \dots, p_m \in \mathfrak{S}^{\mathbf{t}}$;
$\mathrm{brn}(p) = m$ if $p = [p_1]\dots[p_m] \in \mathfrak{S}^{\mathbf{b}}$ with $p_1, \dots, p_m \in \mathfrak{S}^{\mathbf{t}}$.

Examples. $p_1 = a_1[a_2]\big[a_3[a_4][a_5]\big] :$  $\mathrm{wgt}(p_1) = 5$, $\mathrm{hgt}(p_1) = 2$, $\mathrm{brn}(p_1) = 2$.

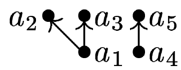
$p_2 = a_1\big[a_2[a_3[a_4][a_5]]\big][a_6[a_7]] :$  $\mathrm{wgt}(p_2) = 7$, $\mathrm{hgt}(p_2) = 3$, $\mathrm{brn}(p_2) = 2$.

$p_3 = a_1[a_2][a_3][a_4]$: 

$\mathrm{wgt}(p_3) = 4$, $\mathrm{hgt}(p_3) = 1$, $\mathrm{brn}(p_3) = 3$.

$q_1 = \big[a_1[a_2[a_3]]\big][a_4]$: 

$\mathrm{wgt}(q_1) = 4$, $\mathrm{hgt}(q_1) = 2$, $\mathrm{brn}(q_1) = 2$.

$q_2 = \big[a_1[a_2][a_3]\big]\big[a_4[a_5]\big]$: 

$\mathrm{wgt}(q_2) = 5$, $\mathrm{hgt}(q_2) = 1$, $\mathrm{brn}(q_2) = 2$.

2.22. The following can be checked directly:

Lemma. (a) *For* $p \in \mathfrak{S}^{\mathbf{t}}$, $\mathrm{wgt}([p]) = \mathrm{wgt}(p)$, $\mathrm{hgt}([p]) = \mathrm{hgt}(p)$ *and* $\mathrm{brn}([p]) = 1$.
(b) *For* $p, q \in \mathfrak{S}^{\mathbf{b}}$ *or* $p \in \mathfrak{S}^{\mathbf{t}}$, $q \in \mathfrak{S}^{\mathbf{b}}$, $\mathrm{wgt}(pq) = \mathrm{wgt}(p) + \mathrm{wgt}(q)$. *For* $p, q \in \mathfrak{S}^{\mathbf{t}}$, $\mathrm{wgt}(pq) = \mathrm{wgt}(p) + \mathrm{wgt}(q) - 1$.
(c) *For* $p, q \in \mathfrak{S}^{\mathbf{b}}$ *or* $p, q \in \mathfrak{S}^{\mathbf{t}}$, $\mathrm{hgt}(pq) = \max\{\mathrm{hgt}(p), \mathrm{hgt}(q)\}$. *For* $p \in \mathfrak{S}^{\mathbf{t}}$, $q \in \mathfrak{S}^{\mathbf{b}}$, $\mathrm{hgt}(pq) = \max\{\mathrm{hgt}(p), 1 + \mathrm{hgt}(q)\}$.
(d) *For* $p, q \in \mathfrak{S}$, $\mathrm{brn}(pq) = \mathrm{brn}(p) + \mathrm{brn}(q)$.

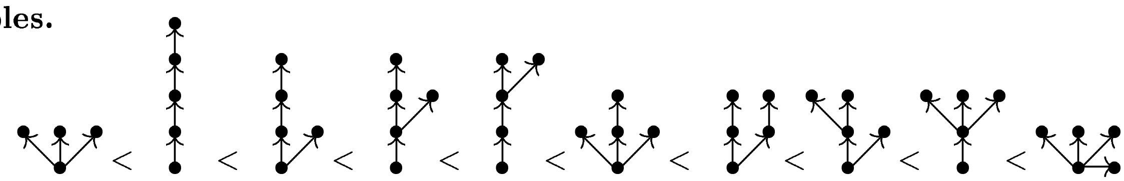
2.23. We now introduce an order on the set of trees $\mathfrak{S}^{\mathbf{t}}$ and, independently, on the set of bushes $\mathfrak{S}^{\mathbf{b}}$; trees will not be comparable with bushes. Strictly speaking, this will be a linear order on the set of non-labeled trees and bushes; for elements $p, q \in \mathfrak{S}^{\mathbf{t}}$ or $\in \mathfrak{S}^{\mathbf{b}}$ having the same graph structure we will assume $p \le q$ and $q \le p$.

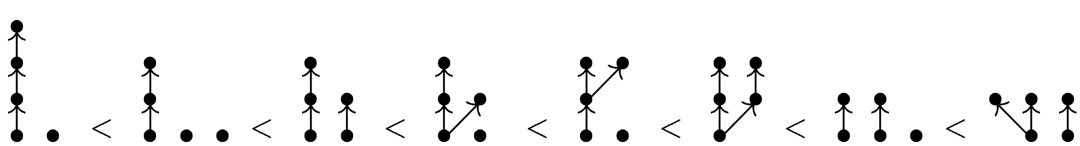
For $p, q \in \mathfrak{S}^{\mathbf{t}}$ or $p, q \in \mathfrak{S}^{\mathbf{b}}$ we will write $p < q$, or $p = o(q)$, if

$$\begin{aligned} &\mathrm{wgt}(p) < \mathrm{wgt}(q) \\ \text{or } &\mathrm{wgt}(p) = \mathrm{wgt}(q) \text{ and } \mathrm{hgt}(p) > \mathrm{hgt}(q) \\ \text{or } &\mathrm{wgt}(p) = \mathrm{wgt}(q) \text{ and } \mathrm{hgt}(p) = \mathrm{hgt}(q) \text{ and } \mathrm{brn}(p) > \mathrm{brn}(q). \end{aligned}$$

If $\mathrm{wgt}(p) = \mathrm{wgt}(q)$, $\mathrm{hgt}(p) = \mathrm{hgt}(q)$ and $\mathrm{brn}(p) = \mathrm{brn}(q) = m$, write $p = a[p_1]\ldots[p_m]$ or $p = [p_1]\ldots[p_m]$, with $a \in \mathcal{A}$ and $p_1, \ldots, p_m \in \mathfrak{S}^{\mathbf{t}}$, so that $p_1 \ge \ldots \ge p_m$, and write $q = b[q_1]\ldots[q_m]$ or $q = [q_1]\ldots[q_m]$, with $b \in \mathcal{A}$ and $q_1, \ldots, q_m \in \mathfrak{S}^{\mathbf{t}}$, so that $q_1 \ge \ldots \ge q_m$. Then $p < q$ if there is i such that $p_1 \le q_1, \ldots, p_{i-1} \le q_{i-1}$ and $p_i < q_i$; in this case we will say that *the list of branches of* p *is smaller than the list of branches of* q.

Examples.

Trees: 

Bushes: 

2.24. We will now obtain several technical lemmas describing the properties of the introduced orders on $\mathfrak{S}^{\mathbf{t}}$ and $\mathfrak{S}^{\mathbf{b}}$.

Lemma. (a) *For* $q \in \mathfrak{S}^{\mathbf{t}}$, $[o(q)] = o([q])$.
(b) *For* $q, r \in \mathfrak{S}$ *with* $\mathrm{hgt}(q) > \mathrm{hgt}(r)$, $o(q)r = o(qr)$.
(c) *For* $q \in \mathfrak{S}$ *and* $r \in \mathfrak{S}^{\mathbf{t}}$, $qo([r]) = o(q[r])$.
(d) *For* $q \in \mathfrak{S}$ *and* $r \in \mathfrak{S}^{\mathbf{t}}$ *with* $\mathrm{hgt}(q) > \mathrm{hgt}(r)$, $o(q)o([r]) = o(q[r])$.

Proof. (a) Let $p = o(q)$, that is, $p \in \mathfrak{B}^{\mathbf{t}}$ with $p < q$. We have $\mathrm{wgt}([q]) = \mathrm{wgt}(q)$, $\mathrm{wgt}([p]) = \mathrm{wgt}(p)$, $\mathrm{hgt}([q]) = \mathrm{hgt}(q)$ and $\mathrm{hgt}([p]) = \mathrm{hgt}(p)$, so if $\mathrm{wgt}(p) < \mathrm{wgt}(q)$ or $\mathrm{wgt}(p) = \mathrm{wgt}(q)$, $\mathrm{hgt}(p) > \mathrm{hgt}(q)$, then $[p] < [q]$. We also have $\mathrm{brn}([q]) = \mathrm{brn}([p]) = 1$, so if both $\mathrm{wgt}(p) = \mathrm{wgt}(q)$ and $\mathrm{hgt}(p) = \mathrm{hgt}(q)$, then $[p] < [q]$ iff $p < q$.

(b) Let $p = o(q)$. (That is, if $q \in \mathfrak{B}^{\mathbf{t}}$ then $p \in \mathfrak{B}^{\mathbf{t}}$ and $p < q$; if $q \in \mathfrak{B}^{\mathbf{b}}$ then $p \in \mathfrak{B}^{\mathbf{b}}$ and $p < q$.) If $\mathrm{wgt}(p) < \mathrm{wgt}(q)$ then $\mathrm{wgt}(pr) < \mathrm{wgt}(qr)$. If $\mathrm{wgt}(p) = \mathrm{wgt}(q)$ and $\mathrm{hgt}(p) > \mathrm{hgt}(q)$, then $\mathrm{wgt}(pr) = \mathrm{wgt}(qr)$ and $\mathrm{hgt}(pr) = \mathrm{hgt}(p) > \mathrm{hgt}(q) = \mathrm{hgt}(qr)$. If $\mathrm{wgt}(p) = \mathrm{wgt}(q)$, $\mathrm{hgt}(p) = \mathrm{hgt}(q)$ and $\mathrm{brn}(p) > \mathrm{brn}(q)$ then $\mathrm{wgt}(pr) = \mathrm{wgt}(qr)$, $\mathrm{hgt}(pr) = \mathrm{hgt}(p) = \mathrm{hgt}(q) = \mathrm{hgt}(qr)$ and $\mathrm{brn}(pr) = \mathrm{brn}(p) + \mathrm{brn}(r) > \mathrm{brn}(q) + \mathrm{brn}(r) = \mathrm{brn}(qr)$. In all these cases $pr < qr$. If $\mathrm{wgt}(p) = \mathrm{wgt}(q)$, $\mathrm{hgt}(p) = \mathrm{hgt}(q)$ and $\mathrm{brn}(p) = \mathrm{brn}(q)$ then $\mathrm{wgt}(pr) = \mathrm{wgt}(qr)$, $\mathrm{hgt}(pr) = \mathrm{hgt}(qr)$ and $\mathrm{brn}(pr) = \mathrm{brn}(qr)$, and we pass to the branches of pr and qr. Since the list of branches of p is smaller than the list of branches of q, the list of branches of pr is smaller than the list of branches of qr, and so, $pr < qr$.

(c) Let $p \in \mathfrak{B}^{\mathbf{b}}$ and $p < [r]$. If $\mathrm{wgt}(p) < \mathrm{wgt}([r])$ then $\mathrm{wgt}(qp) < \mathrm{wgt}(q[r])$ and $qp < q[r]$. Assume that $\mathrm{wgt}(p) = \mathrm{wgt}([r])$, then $\mathrm{wgt}(qp) = \mathrm{wgt}(q[r])$. In this case $\mathrm{hgt}(p) \geq \mathrm{hgt}([r])$, so $\mathrm{hgt}(qp) \geq \mathrm{hgt}(q[r])$. We have $\mathrm{brn}(q[r]) = \mathrm{brn}(q) + 1 \leq \mathrm{brn}(q) + \mathrm{brn}(p) = \mathrm{brn}(qp)$. If $\mathrm{brn}(p) > 1$ then $\mathrm{brn}(q[r]) < \mathrm{brn}(qp)$ and $qp < q[r]$. Otherwise $p = [s]$ with $s < r$, and the list of branches of $qp = q[s]$ is smaller than the list of branches of $q[r]$.

(d) By (c) and (b), $o(q)o([r]) = o(o(q)[r]) = o(o(q[r])) = o(q[r])$. ■

2.25. Lemma. *If* $q, r \in \mathfrak{S}^{\mathbf{t}}$, $s, t \in \mathfrak{S}^{\mathbf{b}}$, $r \neq 0$, $\mathrm{hgt}(q) > \mathrm{hgt}(r)$ *and* $\mathrm{wgt}(qst) \leq \mathrm{wgt}(q[r])$, *then* $qst < q[r]$, $o(qs)t < q[r]$, $[qs]t < [q[r]]$ *and* $o([qs])t < [q[r]]$.

Proof. Let $p \in \mathfrak{S}^{\mathbf{t}}$, $p \leq qs$. Then $\mathrm{wgt}(pt) \leq \mathrm{wgt}(q[r])$; assume that $\mathrm{wgt}(pt) = \mathrm{wgt}(q[r])$. If $\mathrm{hgt}(p) > \mathrm{hgt}(q)$, then $\mathrm{hgt}(pt) > \mathrm{hgt}(q) = \mathrm{hgt}(q[r])$, and $pt < q[r]$. Let $\mathrm{hgt}(p) = \mathrm{hgt}(q)$. Then it must be $\mathrm{brn}(p) \geq \mathrm{brn}(qs) = \mathrm{brn}(q) + \mathrm{brn}(s)$, and so, $\mathrm{brn}(pt) \geq \mathrm{brn}(q) + \mathrm{brn}(s) + \mathrm{brn}(t) > \mathrm{brn}(q) + 1 = \mathrm{brn}(q[r])$. So, $pt < q[r]$.

Next, $\mathrm{wgt}([qs]t) \leq \mathrm{wgt}([q[r]])$, $\mathrm{hgt}([qs]t) \geq \mathrm{hgt}(q) = \mathrm{hgt}([q[r]])$ and $\mathrm{brn}([qs]t) \geq 2 > 1 = \mathrm{brn}([q[r]])$, so $[qs]t < [q[r]]$. By Lemma 2.24(c), $o([qs])t = o([qs]t) < [q[r]]$. ■

2.26. Lemma. *If* $q, r, t \in \mathfrak{S}^{\mathbf{t}}$, $s \in \mathfrak{S}^{\mathbf{b}}$, $r \neq 0$, $\mathrm{hgt}(q) > \mathrm{hgt}(r)$ *and* $\mathrm{wgt}([qs]t) \leq \mathrm{wgt}(q[r])$, *then* $[qs]t < q[r]$ *and* $o([qs])t < q[r]$.

Proof. $\mathrm{hgt}([qs]t) \geq \mathrm{hgt}(q) + 1 > \mathrm{hgt}(q) = \mathrm{hgt}(q[r])$, so $[qs]t < q[r]$. By Lemma 2.24(c), $o([qs])t = o([qs]t) < q[r]$. ■

2.27. Lemma. *If* $q, r, t \in \mathfrak{S}^{\mathbf{t}}$, $a \in \mathcal{A}$, $r \neq 0$, $\mathrm{hgt}(q) > \mathrm{hgt}(r)$ *and* $\mathrm{wgt}([q]t) \leq \mathrm{wgt}(a[q[r]])$, *then* $[q]t < a[q[r]]$ *and* $o([q])t < a[q[r]]$.

Proof. We have $\mathrm{hgt}([q]t) \geq \mathrm{hgt}(q) + 1 = \mathrm{hgt}(a[q[r]])$. If $\mathrm{wgt}(t) > 1$ then $\mathrm{brn}(t) \geq 1$, so $\mathrm{brn}([q]t) \geq 2 > 1 = \mathrm{brn}(a[q[r]]$ and so, $[q]t < a[q[r]]$. If $\mathrm{wgt}(t) = 1$ then $\mathrm{wgt}([q]t) = \mathrm{wgt}(q) + 1 < \mathrm{wgt}(q) + \mathrm{wgt}(r) + 1 = \mathrm{wgt}(a[q[r]])$, and again $[q]t < a[q[r]]$. By Lemma 2.24(c),

$o([q])t = o([q]t) < a[q[r]]$. ■

2.28. Lemma. *If* $q, r, t \in \mathfrak{S}^{\mathbf{t}}$, $s \in \mathfrak{S}^{\mathbf{b}}$, $a \in \mathcal{A}$, $r \neq 0$, $\mathrm{hgt}(q) > \mathrm{hgt}(r)$, $\mathrm{wgt}(s) < \mathrm{wgt}(r)$ *and* $\mathrm{wgt}([qs]t) \leq \mathrm{wgt}(a[q[r]])$, *then* $[qs]t < a[q[r]]$ *and* $o([qs])t < a[q[r]]$.

Proof. We have $\mathrm{hgt}([qs]t) \geq \mathrm{hgt}(q) + 1 = \mathrm{hgt}(a[q[r]])$. If $\mathrm{wgt}(t) > 1$ then $\mathrm{brn}(t) \geq 1$, so $\mathrm{brn}([qs]t) \geq 2 > 1 = \mathrm{brn}(a[q[r]]$ and $[qs]t < a[q[r]]$. If $\mathrm{wgt}(t) = 1$ then $\mathrm{wgt}([qs]t) = \mathrm{wgt}(q) + \mathrm{wgt}(s) + 1 < \mathrm{wgt}(q) + \mathrm{wgt}(r) + 1 < \mathrm{wgt}(a[q[r]])$, and again $[qs]t < a[q[r]]$. By Lemma 2.24(c), $o([qs])t = o([qs]t) < a[q[r]]$. ■

2.29. We will first prove Proposition 2.17 treating $[\cdot]$ as if it were an additive mapping, $[u+v] = [u]+[v]$; in this case any element of $\mathfrak{B}$ is representable as a sum of simple elements. For every $u \in \mathfrak{B}$ we define a list $\mathbf{c}(u)$ of elements of $\mathfrak{S}$ called *the components* of u. We will write $\mathbf{c}(u)$ additively, as $\mathbf{c}(u) = p_1 + \ldots + p_k$, where the order of the summands is not essential, but no combining like terms is allowed.
For $a \in \mathcal{A}$ we put $\mathbf{c}(a) = a$;
if $u_1, u_2 \in \mathfrak{B}$ with $\mathbf{c}(u_1) = p_1 + \ldots + p_m$ and $\mathbf{c}(u_2) = q_1 + \ldots + q_l$ we define

$$\mathbf{c}(u_1 + u_2) = \mathbf{c}(u_1) + \mathbf{c}(u_2) = p_1 + \ldots + p_m + q_1 + \ldots + q_l;$$
$$\mathbf{c}(u_1 u_2) = \mathbf{c}(u_1)\,\mathbf{c}(u_2) = p_1 q_1 + \ldots + p_1 q_l + \ldots + p_m q_1 + \ldots + p_m q_l;$$
$$\mathbf{c}(\pm[u_1]) = [p_1] + \ldots + [p_m].$$

Example. Let $u = a_1\big[a_2 + a_3[a_4][a_5]\big] + a_6[a_7 + a_8] + [a_9]\big[a_{10} + a_{11}[a_{12}]\big]$. To compute $\mathbf{c}(u)$ we simply "open the brackets":

$$\mathbf{c}(u) = a_1[a_2] + a_1\big[a_3[a_4][a_5]\big] + a_6[a_7] + a_6[a_8] + [a_9][a_{10}] + [a_9][a_{11}[a_{12}]].$$

2.30. We will write $\mathbf{c}^{\mathbf{t}}(u)$ for the "tree part" and $\mathbf{c}^{\mathbf{b}}(u)$ for the "bush part" of $\mathbf{c}(u)$, that is, $\mathbf{c}^{\mathbf{t}}(u) = \mathbf{c}(\mathbf{t}(u))$ and $\mathbf{c}^{\mathbf{b}}(u) = \mathbf{c}(\mathbf{b}(u))$.

Example. In the example above,

$$\mathbf{c}^{\mathbf{t}}(u) = a_1[a_2] + a_1\big[a_3[a_4][a_5]\big] + a_6[a_7] + a_6[a_8] \quad \text{and} \quad \mathbf{c}^{\mathbf{b}}(u) = [a_9][a_{10}] + [a_9][a_{11}[a_{12}]].$$

2.31. From Lemma 2.13 one deduces the following:

Lemma. *For any* $u_1, u_2, u \in \mathfrak{B}$,
$\mathbf{c}^{\mathbf{t}}(u_1 + u_2) = \mathbf{c}^{\mathbf{t}}(u_1) + \mathbf{c}^{\mathbf{t}}(u_2)$, $\quad \mathbf{c}^{\mathbf{b}}(u_1 + u_2) = \mathbf{c}^{\mathbf{b}}(u_1) + \mathbf{c}^{\mathbf{b}}(u_2)$,
$\mathbf{c}^{\mathbf{t}}(u_1 u_2) = \mathbf{c}^{\mathbf{t}}(u_1)\,\mathbf{c}^{\mathbf{t}}(u_2) + \mathbf{c}^{\mathbf{t}}(u_1)\,\mathbf{c}^{\mathbf{b}}(u_2) + \mathbf{c}^{\mathbf{b}}(u_1)\,\mathbf{c}^{\mathbf{t}}(u_2)$, $\quad \mathbf{c}^{\mathbf{b}}(u_1 u_2) = \mathbf{c}^{\mathbf{b}}(u_1)\,\mathbf{c}^{\mathbf{b}}(u_2)$,
$\mathbf{c}^{\mathbf{t}}([u]) = 0$ *and, if* $\mathbf{c}^{\mathbf{t}}(u) = p_1 + \ldots + p_m$, $\mathbf{c}([u]) = \mathbf{c}^{\mathbf{b}}([u]) = [p_1] + \ldots + [p_m] + \mathbf{c}^{\mathbf{b}}(u)$.

2.32. Given an additive list $p_1 + \ldots + p_m$ with $p_1, \ldots, p_m \in \mathfrak{S}$ we define $\mathrm{wgt}(p_1 + \ldots + p_m) = \max\{\mathrm{wgt}(p_i)\}_{i=1}^m$, $\mathrm{hgt}(p_1 + \ldots + p_m) = \min\{\mathrm{hgt}(p_i)\}_{i=1}^m$, and $\mathrm{brn}(p_1 + \ldots + p_m) = \min\{\mathrm{brn}(p_i)\}_{i=1}^m$. For $u \in \mathfrak{B}$ we put $\mathrm{wgt}(u) = \mathrm{wgt}(\mathbf{c}(u))$. (This is in agreement with the definition in 1.11.)

2.33. Let $P \in M_d(\mathcal{A})$, $\epsilon = (\epsilon_{i,j})_{1 \leq i < j \leq d}$ be a sign matrix and $\prec$ be a legal order on $\{(i,j)\}_{1 \leq i < j \leq d}$, and let $\varphi_{i,j}^{k,l}, \psi_{i,j}^{k,l}, \xi_{i,j}^{k,l} \in \mathfrak{B}$ be defined by formulas (2.1). The following lemma is easily proved by induction on $j - i$.

Lemma. *For any $1 \le i < j \le d$ and $1 \le k < l \le d$ with $(k,l) \prec (i,j)$ one has* $\mathrm{wgt}(\varphi^{k,l}_{i,j}), \mathrm{wgt}(\psi^{k,l}_{i,j}), \mathrm{wgt}(\xi^{k,l}_{i,j}) \le j - i$.

2.34. Given two additive lists $p_1 + \ldots + p_m$ and $q_1 + \ldots + q_l$ with $p_1, \ldots, p_m, q_1, \ldots, q_l \in \mathfrak{S}^{\mathbf{t}}$ or $\in \mathfrak{S}^{\mathbf{b}}$, we will write $p_1 + \ldots + p_m < q_1 + \ldots + q_l$ if $\max\{p_i\}_{i=1}^m < \max\{q_j\}_{j=1}^l$ and $p_1 + \ldots + p_m \le q_1 + \ldots + q_l$ if $\max\{p_i\}_{i=1}^m \le \max\{q_j\}_{j=1}^l$. If $p_1 + \ldots + p_m < p$ we will also write $p_1 + \ldots + p_m = o(p)$.

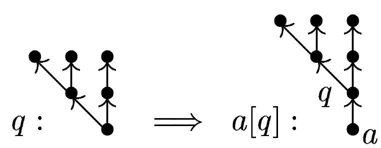
2.35. We will need one more piece of notation. If $(J_1, \prec_1), \ldots, (J_k, \prec_k)$ are linearly ordered sets, then $\big((J_1, \prec_1), \ldots, (J_k, \prec_k)\big)$ will stand for the linear order $\prec$ on $J_1 \cup \ldots \cup J_k$ which coincides with $\prec_i$ on each J_i and satisfies $J_i \prec J_j$ whenever $i < j$. If J_i is a one-element set, $J_i = \{m_i\}$, in this definition we will write m_i instead of $(J_i, \prec_i)$ that is, $\prec = \big((J_1, \prec_1), \ldots, m_i, \ldots, (J_k, \prec_k)\big)$.

2.36. Given a tree $p \in \mathfrak{S}^{\mathbf{t}}$ we will now construct a matrix $P_p \in M_d(\mathcal{A})$, $d = \mathrm{wgt}(p) + 1$, and a legal order $\prec_p$ on $\{(i,j)\}_{1 \le i < j \le d}$ such that p appears as "the principal part" of $\xi_{1,d}(P_p, \prec_p)$ (see Proposition 2.38 below for the exact formulation). Our computations will not be affected by the choice of the signs $\epsilon_{i,j}$, and we will take $\epsilon_{i,j} = 1$ for all i, j.

If $\mathrm{wgt}(p) = 1$, that is, $p = a \in \mathcal{A}$, we define $P_p = \left(\begin{smallmatrix} 1 & a \\ & 1 \end{smallmatrix}\right)$.

Now let $p \in \mathfrak{S}^{\mathbf{t}}$ with $\mathrm{wgt}(p) > 1$ and assume that for all $q \in \mathfrak{S}^{\mathbf{t}}$ with $\mathrm{wgt}(q) < \mathrm{wgt}(p)$ a matrix P_q and an order $\prec_q$ on the set of entries of P_q have been constructed. Let $d = \mathrm{wgt}(p) + 1$ and $m = \mathrm{brn}(p)$. Represent $p = a[p_1] \ldots [p_m]$ so that $a \in \mathcal{A}$ and $p_1, \ldots, p_m \in \mathfrak{S}^{\mathbf{t}}$ satisfy $\mathrm{hgt}(p_1) \ge \ldots \ge \mathrm{hgt}(p_m)$. We distinguish between two cases:

Case 1: $m = 1$ ("extending the trunk"). Put $q = p_1$, then $p = a[q]$ with $\mathrm{wgt}(q) = d - 2$.



Let $P_q = \begin{pmatrix} 1 & b_{1,2} & \ldots & b_{1,d-1} \\ & 1 & \ldots & b_{2,d-1} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{d-1}(\mathcal{A})$. Define

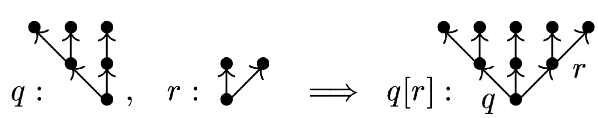
$$P_p = \begin{pmatrix} 1 & -a & 0 & \ldots & 0 \\ & 1 & b_{1,2} & \ldots & b_{1,d-1} \\ & & 1 & \ldots & b_{2,d-1} \\ & & & \ddots & \vdots \\ & & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & -a\ 0\ \ldots\ 0 \\ & P_q \end{pmatrix} \in M_d(\mathcal{A}),$$

shift the order $\prec_q$ so that it is now defined on $I_q = \{(i,j)\}_{2 \le i < j \le d}$ instead of $\{(i,j)\}_{1 \le i < j \le d-1}$, and put

$$\prec_p = \Big(\big(I_q, \prec_q\big), (1,2), (1,3), \ldots, (1,d)\Big).$$

(In plain words, the entries of P_q go first then follow the entries of the first row of P_p.)

Case 2: $m \ge 2$ ("adding a branch"). Put $q = a[p_1] \ldots [p_{m-1}]$ and $r = p_m$, then $p = q[r]$ with $\mathrm{hgt}(q) > \mathrm{hgt}(r)$.



Let $d_1 = \operatorname{wgt}(q) + 1$ and $d_2 = \operatorname{wgt}(r) + 1$, then $d = \operatorname{wgt}(p) + 1 = d_1 + d_2 - 1$. Let $P_q = \begin{pmatrix} 1 & b_{1,2} & \dots & b_{1,d_1} \\ & 1 & \dots & b_{2,d_1} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{d_1}(\mathcal{A})$ and $P_r = \begin{pmatrix} 1 & c_{1,2} & \dots & c_{1,d_2} \\ & 1 & \dots & c_{2,d_2} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{d_2}(\mathcal{A})$. We define

$$P_p = \left(\begin{array}{cccc|cccc} 1 & c_{1,2} & \dots & c_{1,d_2} & 0 & \dots & 0 \\ & 1 & \dots & c_{2,d_2} & 0 & \dots & 0 \\ & & \ddots & \vdots & \vdots & & \vdots \\ \hline & & & \boxed{1} & b_{1,2} & \dots & b_{1,d_1} \\ & & & & 1 & \dots & b_{2,d_1} \\ & & & & & \ddots & \vdots \\ & & & & & & 1 \end{array}\right) = \left(\begin{array}{c|c} P_r & 0 \\ \hline & P_q \end{array}\right) \in M_d(\mathcal{A}).$$

That is, P_q occupies the submatrix of P_p indexed by $I_q = \{(i,j)\}_{d_2 \le i < j \le d}$ and P_r occupies the submatrix of P_p indexed by $I_r = \{(i,j)\}_{1 \le i < j \le d_2}$. Shift the order $\prec_q$ so that it is defined on I_q instead of $\{(i,j)\}_{1 \le i < j \le d_1}$, and let $\prec_J$ be any legal order on $J = \{(i,j)\}_{\substack{1 \le i \le d_2 - 1 \\ d_2 + 1 \le j \le d}}$. We define the order $\prec_p$ on $\{(i,j)\}_{1 \le i < j \le d}$ to be

$$\prec_p = \Big(\big(I_q \setminus \{(d_2, d)\}, \prec_q\big), \big(I_r, \prec_r\big), (d_2, d), \big(J, \prec_J\big) \Big).$$

(That is, first the entries of P_q excluding b_{1,d_1} appear, then follow the entries of P_r, then follow b_{1,d_1}, and finally all other entries of P_p follow.)

2.37. Lemma. *Let $p \in \mathfrak{S}^{\mathbf{t}}$, let $d = \operatorname{wgt}(p) + 1$ and let $1 < n < d$. Then the submatrix Q of P_p indexed by $\{(i,j)\}_{n \le i < j \le d}$ is equal to P_t and $\prec_p|_Q = \prec_t$ for some $t \in \mathfrak{S}^{\mathbf{t}}$. If $\operatorname{brn}(p) \ge 2$ so that Case 2 takes place, that is, $p = q[r]$ with $q, r \in \mathfrak{S}^{\mathbf{t}}$, $\operatorname{hgt}(q) > \operatorname{hgt}(r)$, $\operatorname{wgt}(q) = d_1$, $\operatorname{wgt}(r) = d_2$, and if $1 \le n < d_2$, then Q is equal to $P_{q[s]}$ and $\prec_p|_Q = \prec_{q[s]}$ for some $s \in \mathfrak{S}^{\mathbf{t}}$.*

Proof. In Case 1, that is, when $p = a[q]$, Q is a submatrix of P_q, and we are done by induction on $\operatorname{wgt}(p)$.

$$P_p = \left(\begin{array}{cc|ccc} 1-a & 0 & \dots & & 0 \\ & 1 & b_{1,2} & \dots & b_{1,d-1} \\ \hline & & 1 & \dots & b_{2,d-1} \\ & & & \ddots & \vdots \\ & & & & 1 \end{array}\right),$$

Consider Case 2, where $p = q[r]$, $d_1 = \operatorname{wgt}(q)$, $d_2 = \operatorname{wgt}(r)$. If $n \ge d_2$ then Q is a submatrix of P_q and we are done.

$$P_p = \left(\begin{array}{cccc|ccc} 1 & c_{1,2} & \dots & c_{1,d_2} & 0 & \dots & 0 \\ & 1 & \dots & c_{2,d_2} & 0 & \dots & 0 \\ & & \ddots & \vdots & \vdots & & \vdots \\ & & & 1 & b_{1,2} & \dots & b_{1,d_1} \\ \hline & & & & 1 & \dots & b_{2,d_1} \\ & & & & & \ddots & \vdots \\ & & & & & & 1 \end{array}\right)$$

If $n < d_2$, consider the submatrix R indexed by $\{(i,j)\}_{n \le i < j \le d_2}$.

$$P_p = \left(\begin{array}{c|cccccc} 1 & c_{1,2} & \dots & c_{1,d_2} & 0 & \dots & 0 \\ \hline & 1 & \dots & c_{2,d_2} & 0 & \dots & 0 \\ & & \ddots & \vdots & \vdots & & \vdots \\ & & & 1 & b_{1,2} & \dots & b_{1,d_1} \\ & & & & 1 & \dots & b_{2,d_1} \\ & & & & & \ddots & \vdots \\ & & & & & & 1 \end{array}\right)$$

R is a submatrix of P_r and by induction, $R = P_s$ and $\prec_q|_R = \prec_s$ for some $s \in \mathfrak{S}^{\mathbf{t}}$. The matrix Q and the order $\prec_p|_Q$ are obtained from P_q, $\prec_p$, P_s and $\prec_s$ in the way described in Case 2; hence, $Q = P_t$ and $\prec_p|_Q = \prec_t$ for $t = q[s]$. ■

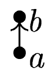
2.38. Proposition. *For any $p \in \mathfrak{S}^{\mathbf{t}}$ one has $\mathbf{c}^{\mathbf{t}}\big(\xi_{1,d}(P_p, \prec_p)\big) = p + o(p)$.*

Proof. We will prove this proposition by induction on $\mathrm{wgt}(p)$. We strengthen our induction hypothesis and will simultaneously be proving that $\mathbf{c}^{\mathbf{b}}\big(\xi_{1,d}(P_p, \prec_p)\big) < [p]$ and $\mathbf{c}^{\mathbf{b}}\big(\varphi_{1,d}(P_p, \prec_p)\big) < [p]$.

If $\mathrm{wgt}(p) = 1$, that is, $p = a \in \mathcal{A}$, we have $P_p = \left(\begin{smallmatrix} 1 & a \\ & 1 \end{smallmatrix}\right)$ and $\xi_{1,2} = \varphi_{1,2} = p$.

$a :$ 

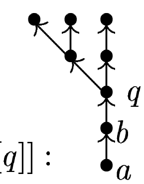
If $\mathrm{wgt}(p) = 2$ then p has form $a[b]$ with $a, b \in \mathcal{A}$.

$a[b] :$ 

P_p, $\prec_p$ are constructed in accordance with Case 1: $P_p = \left(\begin{smallmatrix} 1 & -a & 0 \\ & 1 & b \\ & & 1 \end{smallmatrix}\right)$ and $\prec_p = \big((2,3), (1,2), (1,3)\big)$. By formulas (2.1), $\xi_{1,3} = 0 - \xi_{1,2}^{2,3}[\xi_{2,3}] - \psi_{1,2}^{2,3}\xi_{2,3}$. We have $\xi_{1,2}^{2,3} = -a$, $\xi_{2,3} = b$ and $\psi_{1,2} = 0$, so that $\xi_{1,3} = a[b] = p$. Also, $\varphi_{1,3} = -a_{1,2}^{2,3}[\xi_{2,3}] = a[b]$ and so, $\mathbf{c}^{\mathbf{b}}(\varphi_{1,3}) = 0$.

Now let $p \in \mathfrak{S}^{\mathbf{t}}$, $\mathrm{wgt}(p) \geq 3$; put $d = \mathrm{wgt}(p) + 1$. We consider several cases.

Case 1a: $p = a[b[q]]$ where $a, b \in \mathcal{A}$ and $q \in \mathfrak{S}^{\mathbf{t}}$.

$a[b[q]] :$ 

Let $P_q = \begin{pmatrix} 1 & c_{1,2} & \dots & c_{1,d-1} \\ & 1 & \dots & c_{2,d-2} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{d-2}(\mathcal{A})$, then

$$P_p = \begin{pmatrix} 1 & a_{1,2} & 0 & 0 & \dots & 0 \\ & 1 & a_{2,3} & 0 & \dots & 0 \\ & & 1 & a_{3,4} & \dots & a_{3,d} \\ & & & 1 & \dots & a_{4,d} \\ & & & & \ddots & \vdots \\ & & & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & -a & 0 & 0 & \dots & 0 \\ & 1 & -b & 0 & \dots & 0 \\ & & 1 & c_{1,2} & \dots & c_{1,d-2} \\ & & & 1 & \dots & c_{2,d-2} \\ & & & & \ddots & \vdots \\ & & & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & -a & 0 & 0 \dots 0 \\ & 1 & -b & 0 \dots 0 \\ & & \multicolumn{2}{|c}{P_q} \end{pmatrix} \in M_d(\mathcal{A}).$$

We will identify P_q with the submatrix of P_p indexed by $I_q = \{(i,j)\}_{3 \leq i < j \leq d}$ and shift $\prec_q$ so that it is defined on I_q instead of $\{(i,j)\}_{1 \leq i < j \leq d-2}$. Then

$$\prec_p = \Big(\big(I_q, \prec_q\big), (2,3), \dots, (2,d), (1,2), \dots, (1,d)\Big).$$

The entries $(i,j) \notin I_q$ do not affect the entries from I_q, therefore the elements $\varphi_{i,j}^{k,l}(P_p)$, $\psi_{i,j}^{k,l}(P_p), \xi_{i,j}^{k,l}(P_p)$ with $(i,j) \in I_q$ are equal to the corresponding $\varphi_{i,j}^{k,l}(P_q), \psi_{i,j}^{k,l}(P_q), \xi_{i,j}^{k,l}(P_q)$.

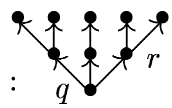
From formulas (2.1),

$$\xi_{1,d} = -\sum_{n=2}^{d-1} \xi_{1,n}^{n,d}[\xi_{n,d}] - \sum_{n=2}^{d-1} \psi_{1,n}^{n,d}\xi_{n,d}.$$

One checks by induction on n that for any $n \in \{3, \dots, d\}$ and $(k,l) \in I_q$ one has $\varphi_{1,n}^{k,l} = \psi_{1,n}^{k,l} = \xi_{1,n}^{k,l} = 0$. It follows that $\sum\limits_{n=3}^{d-1} \xi_{1,n}^{n,d}[\xi_{n,d}] + \sum\limits_{n=3}^{d-1} \psi_{1,n}^{n,d}\xi_{n,d} = 0$.

We have $\xi_{1,2}^{2,d} = -a$ and $\psi_{1,2}^{2,d} = 0$, so, $\xi_{1,d} = a[\xi_{2,d}] \in \mathfrak{B}^{\mathbf{t}}$, and so, $\mathbf{c}^{\mathbf{b}}(\xi_{1,d}) = 0$. By our induction hypothesis, $\mathbf{c}(\xi_{2,d}) = b[q] + o(b[q]) + o([b[q]])$. By Lemma 2.24(a) and (c), $\mathbf{c}^{\mathbf{t}}(\xi_{1,d}) = \mathbf{c}^{\mathbf{t}}(a[\xi_{2,d}]) = a[b[q]] + o(a[b[q]]) = p + o(p)$. We also find that $\varphi_{1,d} = -\sum\limits_{n=2}^{d-1} \varphi_{1,n}^{n,d}[\xi_{n,d}] = -\varphi_{1,2}^{2,d}[\xi_{2,d}] = a[\xi_{2,d}] \in \mathfrak{B}^{\mathbf{t}}$, and so, $\mathbf{c}^{\mathbf{b}}(\varphi_{1,d}) = 0$.

Case 2a: $p = q[r]$ where $q, r \in \mathfrak{S}^{\mathbf{t}}$, $\mathrm{wgt}(r) \geq 2$ and $\mathrm{hgt}(q) > \mathrm{hgt}(r)$.

$q[r]$: 

Let $d_1 = \mathrm{wgt}(q) + 1$, $d_2 = \mathrm{wgt}(r) + 1$, $P_q = \begin{pmatrix} 1 & b_{1,2} & \dots & b_{1,d_1} \\ & 1 & \dots & b_{2,d_1} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{d_1}(\mathcal{A})$ and $P_r = \begin{pmatrix} 1 & c_{1,2} & \dots & c_{1,d_2} \\ & 1 & \dots & c_{2,d_2} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{d_2}(\mathcal{A})$, then

$$P_p = \begin{pmatrix} 1 & a_{1,2} & \dots & a_{1,d_2} & 0 & \dots & 0 \\ & 1 & \dots & a_{2,d_2} & 0 & \dots & 0 \\ & & \ddots & \vdots & \vdots & & \vdots \\ & & & 1 & a_{d_2,d_2+1} & \dots & a_{d_2,d} \\ & & & & 1 & \dots & a_{d_2+1,d} \\ & & & & & \ddots & \vdots \\ & & & & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & c_{1,2} & \dots & c_{1,d_2} & 0 & \dots & 0 \\ & 1 & \dots & c_{2,d_2} & 0 & \dots & 0 \\ & & \ddots & \vdots & \vdots & & \vdots \\ & & & 1 & b_{1,2} & \dots & b_{1,d_1} \\ & & & & 1 & \dots & b_{2,d_1} \\ & & & & & \ddots & \vdots \\ & & & & & & 1 \end{pmatrix} \in M_d(\mathcal{A}).$$

Let $I_q = \{(i,j)\}_{d_2 \leq i < j \leq d}$, $I_r = \{(i,j)\}_{1 \leq i < j \leq d_2}$ and $J = \{(i,j)\}_{\substack{1 \leq i \leq d_2 - 1 \\ d_2+1 \leq j \leq d}}$. We will identify the matrices P_q and P_r with their images in P_p indexed by I_q and I_r respectively, and, in particular, will index the entries of P_q by I_q instead of $\{(i,j)\}_{1 \leq i < j \leq d_1}$. We then have $\prec_p = \big((I_q \setminus \{(d_2, d)\}, \prec_q), (I_r, \prec_r), (d_2, d), (J, \prec_J) \big)$.

The entries of P_q do not affect the entries of P_r and vice versa. Thus, the elements $\varphi_{i,j}^{k,l}(P_p), \psi_{i,j}^{k,l}(P_p), \xi_{i,j}^{k,l}(P_p)$ with $(i,j) \in I_r$ are equal to the corresponding $\varphi_{i,j}^{k,l}(P_r), \psi_{i,j}^{k,l}(P_r)$, $\xi_{i,j}^{k,l}(P_r)$ and the elements $\varphi_{i,j}^{k,l}(P_p), \psi_{i,j}^{k,l}(P_p), \xi_{i,j}^{k,l}(P_p)$ with $(i,j) \in I_q$ are equal to the corresponding $\varphi_{i,j}^{k,l}(P_q), \psi_{i,j}^{k,l}(P_q), \xi_{i,j}^{k,l}(P_q)$. From formulas (2.1) we have

$$\begin{aligned} \xi_{1,d} = -\sum_{n=d_2+1}^{d-1} \xi_{1,n}^{n,d}[\xi_{n,d}] - \sum_{n=d_2+1}^{d-1} \psi_{1,n}^{n,d}\xi_{n,d} + \sum_{n=2}^{d_2-1} [\xi_{1,n}]\xi_{n,d} - \sum_{n=2}^{d_2-1} \varphi_{1,n}\xi_{n,d} \\ + [\xi_{1,d_2}]\xi_{d_2,d} - \varphi_{1,d_2}\xi_{d_2,d}. \end{aligned} \tag{2.2}$$

By our induction hypothesis, $\mathbf{c}(\xi_{1,d_2}) = r + o(r) + o([r])$ and $\mathbf{c}(\xi_{d_2,d}) = q + o(q) + o([q])$. By Lemma 2.24(a), $\mathbf{c}([\xi_{1,d_2}]) = [r] + o([r])$. By Lemma 2.24(b), (c) and (d), $\mathbf{c}\big([\xi_{1,d_2}]\xi_{d_2,d}\big) = q[r] + o(q[r]) + o([q][r]) = p + o(p) + o([q][r])$. Since $\mathrm{hgt}(q) > \mathrm{hgt}(r)$, we have $\mathrm{hgt}([q][r]) = \mathrm{hgt}(q) = \mathrm{hgt}([q[r]])$; since $\mathrm{brn}([q][r]) > \mathrm{brn}([q[r]])$, we have $[q][r] < [q[r]] = [p]$ and so, $\mathbf{c}\big([\xi_{1,d_2}]\xi_{d_2,d}\big) = p + o(p) + o([p])$.

We will now show that the components of all other terms on the right hand side of (2.2) are smaller than p or $[p]$. By Lemma 2.33, the weights of these terms do not exceed $d-1 = \mathrm{wgt}(p)$.

We start with the sums $\sum\limits_{n=d_2+1}^{d-1} \xi_{1,n}^{n,d}[\xi_{n,d}]$ and $\sum\limits_{n=d_2+1}^{d-1} \psi_{1,n}^{n,d}\xi_{n,d}$. Fix any $(k,l) \in I_q \setminus \{(d_2,d)\}$. Since for $(i,j) \in I_r$ one has $(k,l) \prec_p (i,j)$, we obtain from formulas (2.1) that $\varphi_{i,j}^{k,l} = \xi_{i,j}^{k,l} = c_{i,j}$ and $\psi_{i,j}^{k,l} = 0$. In particular, $\mathrm{wgt}(\varphi_{i,d_2}^{k,l}), \mathrm{wgt}(\psi_{i,d_2}^{k,l}), \mathrm{wgt}(\xi_{i,d_2}^{k,l}) \le 1$ for any $i \in \{1,\dots,d_2-1\}$. Next, since $(k,l) \prec_p (n,j)$ for any $(n,j) \in J$, the entries (n,j) with $n < d_2$ do not participate in formulas for $\varphi_{i,j}^{k,l}, \psi_{i,j}^{k,l}, \xi_{i,j}^{k,l}$ with $(i,j) \in J$. By Lemma 2.33 one has $\mathrm{wgt}(\varphi_{n,j}), \mathrm{wgt}(\psi_{n,j}), \mathrm{wgt}(\xi_{n,j}) \le n-j$, and one checks by induction on j that for any $(i,j) \in J$, $\mathrm{wgt}(\varphi_{i,j}^{k,l}), \mathrm{wgt}(\psi_{i,j}^{k,l}), \mathrm{wgt}(\xi_{i,j}^{k,l}) \le j - d_2 + 1$. In particular, for any $n \in \{d_2+1,\dots,d-1\}$ one has $\mathrm{wgt}(\xi_{1,n}^{n,d}), \mathrm{wgt}(\psi_{1,n}^{n,d}) \le n - d_2 + 1$, and since $d_2 = \mathrm{wgt}(r) + 1 \ge 3$, we obtain $\mathrm{wgt}\big(\xi_{1,n}^{n,d}[\xi_{n,d}]\big), \mathrm{wgt}\big(\psi_{1,n}^{n,d}\xi_{n,d}\big) \le n - d_2 + 1 + d - n < d - 1 = \mathrm{wgt}(p)$.

Now turn to the sums $\sum\limits_{n=2}^{d_2-1} [\xi_{1,n}]\xi_{n,d}$ and $\sum\limits_{n=2}^{d_2-1} \varphi_{1,n}\xi_{n,d}$. Fix $n \in \{2,\dots,d_2-1\}$. By Lemma 2.37 the submatrix of P_p indexed by $\{(i,j)\}_{n \le i < j \le d_2}$ has form $P_{q[s]}$ for some $s \in \mathfrak{S}^{\mathbf{t}}$ with $\mathrm{wgt}(s) < \mathrm{wgt}(r)$, and by induction hypothesis $\mathbf{c}(\xi_{n,d}) = q[s] + o(q[s]) + o([q[s]])$. By Lemma 2.25, $\mathbf{c}([\xi_{1,n}])\,\mathbf{c}^{\mathbf{t}}(\xi_{n,d}) = o(q[r]) = o(p)$ and $\mathbf{c}([\xi_{1,n}])\,\mathbf{c}^{\mathbf{b}}(\xi_{n,d}) = o([q[r]]) = o(p) + o([p])$. We also have by Lemma 2.25 that $\mathbf{c}^{\mathbf{b}}(\varphi_{1,n})\,\mathbf{c}^{\mathbf{t}}(\xi_{n,d}) = o(q[r])$ and $\mathbf{c}^{\mathbf{b}}(\varphi_{1,n})\,\mathbf{c}^{\mathbf{b}}(\xi_{n,d}) = o([q[r]])$, and by Lemma 2.26 that $\mathbf{c}^{\mathbf{t}}(\varphi_{1,n})\,\mathbf{c}^{\mathbf{b}}(\xi_{n,d}) = o(q[r])$. By Lemma 2.33, $\mathrm{wgt}(\xi_{n,d}) \le d-n$ and $\mathrm{wgt}(\varphi_{1,n}) \le n-1$, so $\mathrm{wgt}\big(\mathbf{c}^{\mathbf{t}}(\varphi_{1,n})\,\mathbf{c}^{\mathbf{t}}(\xi_{n,d})\big) \le d - n + n - 1 - 1 = d - 2 < \mathrm{wgt}(p)$. Summarizing, $\mathbf{c}^{\mathbf{t}}(\varphi_{1,n}\xi_{n,d}) = o(q[r]) = o(p)$ and $\mathbf{c}^{\mathbf{b}}(\varphi_{1,n}\xi_{n,d}) = o([q[r]]) = o([p])$.

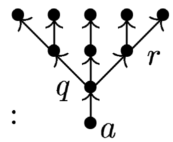
Now consider the term $\varphi_{1,d_2}\xi_{d_2,d}$. Again, we have $\mathrm{wgt}\big(\mathbf{c}^{\mathbf{t}}(\varphi_{1,d_2})\,\mathbf{c}^{\mathbf{t}}(\xi_{d_2,d})\big) < \mathrm{wgt}(p)$. By our induction hypothesis, $\mathbf{c}^{\mathbf{t}}(\xi_{d_2,d}) = q + o(q)$, $\mathbf{c}^{\mathbf{b}}(\xi_{d_2,d}) = o([q])$ and $\mathbf{c}^{\mathbf{b}}(\varphi_{1,d_2}) = o([r])$. By Lemma 2.24(b) and (d), $\mathbf{c}^{\mathbf{b}}(\varphi_{1,d_2})\,\mathbf{c}^{\mathbf{t}}(\xi_{d_2,d}) = o(q[r]) = o(p)$ and $\mathbf{c}^{\mathbf{b}}(\varphi_{1,d_2})\,\mathbf{c}^{\mathbf{b}}(\xi_{d_2,d}) = o([q][r]) = o([p])$. By Lemma 2.24(c), $\mathbf{c}^{\mathbf{t}}(\varphi_{1,d_2})\,\mathbf{c}^{\mathbf{b}}(\xi_{d_2,d}) = \mathbf{c}^{\mathbf{t}}(\varphi_{1,d_2})o\big([q]\big) = o\big(\mathbf{c}^{\mathbf{t}}(\varphi_{1,d_2})[q]\big)$. Since $\mathrm{hgt}\big(\mathbf{c}^{\mathbf{t}}(\varphi_{1,d_2})[q]\big) \ge \mathrm{hgt}(q) + 1 > \mathrm{hgt}(p)$, we have $\mathbf{c}^{\mathbf{t}}(\varphi_{1,d_2})[q] = o(p)$ and so, $\mathbf{c}^{\mathbf{t}}(\varphi_{1,d_2})\,\mathbf{c}^{\mathbf{b}}(\xi_{d_2,d}) = o(p)$. Hence, $\mathbf{c}^{\mathbf{t}}(\varphi_{1,d_2}\xi_{d_2,d}) = o(p)$ and $\mathbf{c}^{\mathbf{b}}(\varphi_{1,d_2}\xi_{d_2,d}) = o([p])$.

It remains to check that $\mathbf{c}^{\mathbf{b}}(\varphi_{1,d}) < [p]$. By formulas (2.1),

$$\varphi_{1,d} = -\sum_{n=d_2+1}^{d-1} \varphi_{1,n}^{n,d}[\xi_{n,d}] - \sum_{n=2}^{d_2} \big(\varphi_{1,n} - [\xi_{1,n}]\big)[\xi_{n,d}].$$

Again, for $n \in \{d_2+1,\dots,d-1\}$ one has $\mathrm{wgt}(\varphi_{1,n}^{n,d}) \le n - d_2 + 1$ and so, $\mathrm{wgt}(\varphi_{1,n}^{n,d}[\xi_{n,d}]) < d - 1 = \mathrm{wgt}(p)$. For $n \in \{2,\dots,d_2-1\}$ one has $\mathbf{c}([\xi_{n,d}]) = [q[s]] + o([q[s]])$, and by Lemma 2.25, $\mathbf{c}^{\mathbf{b}}\big(\alpha_{1,n}[\xi_{n,d}]\big) + \mathbf{c}^{\mathbf{b}}\big([\xi_{1,n}][\xi_{n,d}]\big) = o([q[r]]) = o([p])$. For $n = d_2$ one has $\mathbf{c}([\xi_{n,d}]) = [q] + o([q])$, $\mathbf{c}^{\mathbf{b}}(\alpha_{1,n}) = o([r])$ and $\mathbf{c}([\xi_{1,n}]) = [r] + o([r])$; by Lemma 2.24, $\mathbf{c}^{\mathbf{b}}\big(\alpha_{1,n}[\xi_{n,d}]\big) + \mathbf{c}^{\mathbf{b}}\big([\xi_{1,n}][\xi_{n,d}]\big) = [q][r] + o([q][r]) = o([p])$.

Case 1b: $p = a[q[r]]$ where $a \in \mathcal{A}$ and $q, r \in \mathfrak{S}^{\mathbf{t}}$ with $\mathrm{hgt}(q) > \mathrm{hgt}(r)$.

$a[q[r]]$:


Let $d_1 = \mathrm{wgt}(q) + 1$, $d_2 = \mathrm{wgt}(r) + 1$, $P_q = \begin{pmatrix} 1 & b_{1,2} & \dots & b_{1,d_1} \\ & 1 & \dots & b_{2,d_1} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{d_1}(\mathcal{A})$ and $P_r = \begin{pmatrix} 1 & c_{1,2} & \dots & c_{1,d_2} \\ & 1 & \dots & c_{2,d_2} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{d_2}(\mathcal{A})$, then

$$P_p = \begin{pmatrix} 1 & a_{1,2} & 0 & \dots & 0 & 0 & \dots & 0 \\ & 1 & a_{2,3} & \dots & a_{1,d_2+1} & 0 & \dots & 0 \\ & & 1 & \dots & a_{3,d_2+1} & 0 & \dots & 0 \\ & & & \ddots & \vdots & \vdots & & \vdots \\ & & & & 1 & a_{d_2+1,d_2+2} & \dots & a_{d_2+1,d} \\ & & & & & 1 & \dots & a_{d_2+2,d} \\ & & & & & & \ddots & \vdots \\ & & & & & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & -a & 0 & \dots & 0 & 0 & \dots & 0 \\ & 1 & c_{1,2} & \dots & c_{1,d_2} & 0 & \dots & 0 \\ & & 1 & \dots & c_{2,d_2} & 0 & \dots & 0 \\ & & & \ddots & \vdots & \vdots & & \vdots \\ & & & & 1 & b_{1,2} & \dots & b_{1,d_1} \\ & & & & & 1 & \dots & b_{2,d_1} \\ & & & & & & \ddots & \vdots \\ & & & & & & & 1 \end{pmatrix} \in M_d(\mathcal{A}).$$

Let $I_q = \{(i,j)\}_{d_2+1 \le i < j \le d}$, $I_r = \{(i,j)\}_{2 \le i < j \le d_2+1}$ and $J = \{(i,j)\}_{\substack{2 \le i \le d_2 \\ d_2+2 \le j \le d}}$. We will identify the matrices P_q and P_r with their images in P_p indexed by I_q and I_r respectively, and will index the entries of P_q and P_r by I_q and I_r respectively. We then have $\prec_p = \big((I_q \setminus \{(d_2+1,d)\}, \prec_q), (I_r, \prec_r), (d_2+1,d), (J, \prec_J), (1,2), \dots, (1,d)\big)$.

The entries of P_q do not affect the entries of P_r and vice versa. Thus, the elements $\varphi_{i,j}^{k,l}(P_p), \psi_{i,j}^{k,l}(P_p), \xi_{i,j}^{k,l}(P_p)$ with $(i,j) \in I_r$ are equal to the corresponding $\varphi_{i,j}^{k,l}(P_r), \psi_{i,j}^{k,l}(P_r)$, $\xi_{i,j}^{k,l}(P_r)$ and the elements $\varphi_{i,j}^{k,l}(P_p), \psi_{i,j}^{k,l}(P_p), \xi_{i,j}^{k,l}(P_p)$ with $(i,j) \in I_q$ are equal to the corresponding $\varphi_{i,j}^{k,l}(P_q), \psi_{i,j}^{k,l}(P_q), \xi_{i,j}^{k,l}(P_q)$.

From formulas (2.1), for $n \in \{3, \dots, d\}$ and $1 \le k < l \le d$ with $(k,l) \preceq_p (1,n)$ we have

$$\begin{aligned} \varphi_{1,n}^{k,l} &= -\sum_{\substack{m=2,\dots,n-1 \\ (m,n) \prec_p (k,l)}} \varphi_{1,m}^{m,n}[\xi_{m,n}], \\ \psi_{1,n}^{k,l} &= -\sum_{\substack{m=2,\dots,n-1 \\ (m,n) \prec_p (k,l)}} \psi_{1,m}^{m,n}\big(\xi_{m,n} - [\xi_{m,n}]\big), \\ \xi_{1,n}^{k,l} &= -\sum_{\substack{m=2,\dots,n-1 \\ (m,n) \prec_p (k,l)}} \big(\xi_{1,m}^{m,n}[\xi_{m,n}] + \psi_{1,m}^{m,n}\xi_{m,n}\big). \end{aligned} \tag{2.3}$$

So,

$$\mathbf{b}(\varphi^{k,l}_{1,n}) = -\sum_{\substack{m=2,\dots,n-1\\(m,n)\prec_p(k,l)}} \mathbf{b}(\varphi^{m,n}_{1,m})\,\mathbf{b}([\xi_{m,n}]),$$

$$\mathbf{b}(\psi^{k,l}_{1,n}) = -\sum_{\substack{m=2,\dots,n-1\\(m,n)\prec_p(k,l)}} \mathbf{b}(\psi^{m,n}_{1,m})\big(\mathbf{b}(\xi_{m,n}) - \mathbf{b}([\xi_{m,n}])\big),$$

$$\mathbf{b}(\xi^{k,l}_{1,n}) = -\sum_{\substack{m=2,\dots,n-1\\(m,n)\prec_p(k,l)}} \big(\mathbf{b}(\xi^{m,n}_{1,m})\,\mathbf{b}([\xi_{m,n}]) + \mathbf{b}(\psi^{m,n}_{1,m})\,\mathbf{b}(\xi_{m,n})\big),$$

and by induction on n, $\mathbf{b}(\varphi^{k,l}_{1,n}) = \mathbf{b}(\psi^{k,l}_{1,n}) = \mathbf{b}(\xi^{k,l}_{1,n}) = 0$. Hence, $\varphi^{k,l}_{1,n}, \psi^{k,l}_{1,n}, \xi^{k,l}_{1,n} \in \mathfrak{B}^{\mathbf{t}}$. In particular, $\mathbf{c}^{\mathbf{b}}(\varphi_{1,d}) = 0$.

From formulas (2.1),

$$\xi_{1,d} = -\sum_{n=2}^{d-1} \xi^{n,d}_{1,n}[\xi_{n,d}] - \sum_{n=2}^{d-1} \psi^{n,d}_{1,n}\xi_{n,d}.$$

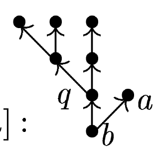
We have $\xi^{2,d}_{1,2} = -a$, $\psi^{2,d}_{1,2} = 0$ and by induction hypothesis, $\mathbf{c}(\xi_{2,d}) = q[r]+o(q[r])+o([q[r]])$. By Lemma 2.24(b), $\mathbf{c}\big(\xi^{2,d}_{1,2}[\xi_{2,d}] + \xi_{2,d}\psi^{2,d}_{1,2}\big) = a[q[r]] + o(a[q[r]]) = p + o(p)$.

We will now show that $\mathbf{c}\big(\sum_{n=3}^{d-1} \xi^{n,d}_{1,n}[\xi_{n,d}] + \sum_{n=3}^{d-1} \psi^{n,d}_{1,n}\xi_{n,d}\big) = o(p)$. First, fix $n \in \{3,\dots,d_2\}$. By Lemma 2.37 and the induction hypothesis, $\mathbf{c}(\xi_{n,d}) = q[s]+o(q[s])+o([q[s]])$ for some $s \in \mathfrak{S}^{\mathbf{t}}$ with $\mathrm{wgt}(s) < \mathrm{wgt}(r)$. By Lemma 2.28, $\mathbf{c}(\xi^{n,d}_{1,n})\,\mathbf{c}([\xi_{n,d}]) = o(a[q[r]]) = o(p)$ and $\mathbf{c}(\psi^{n,d}_{1,n})\,\mathbf{c}^{\mathbf{b}}(\xi_{n,d}) = o(a[q[r]]) = o(p)$. Also, $\mathrm{wgt}\big(\mathbf{c}(\psi^{n,d}_{1,n})\,\mathbf{c}^{\mathbf{t}}(\xi_{n,d})\big) < \mathrm{wgt}(p)$, so that $\mathbf{c}(\psi^{n,d}_{1,n})\,\mathbf{c}^{\mathbf{t}}(\xi_{n,d}) = o(p)$.

Now put $n = d_2+1$. By our induction hypothesis, $\mathbf{c}(\xi_{n,d}) = q + o(q) + o([q])$. By Lemma 2.27, $\mathbf{c}\big(\xi^{n,d}_{1,n}[\xi_{n,d}]\big) = o(a[q[r]]) = o(p)$ and $\mathbf{c}(\psi^{n,d}_{1,n})\,\mathbf{c}^{\mathbf{b}}(\xi_{n,d})) = o(a[q[r]]) = o(p)$. Also, $\mathrm{wgt}\big(\mathbf{c}(\psi^{n,d}_{1,n})\,\mathbf{c}^{\mathbf{t}}(\xi_{n,d})\big) < \mathrm{wgt}(p)$, so that $\mathbf{c}(\psi^{n,d}_{1,n})\,\mathbf{c}^{\mathbf{t}}(\xi_{n,d}) = o(p)$.

Finally, let $n \in \{d_2+2,\dots,d-1\}$. In formulas (2.3), for any $(k,l) \in I_q \setminus \{(d_{2+1},d)\}$, if $(m,n) \prec_p (k,l)$ then it must be $m \geq d_2+1$, and by induction on n we conclude that $\xi^{k,l}_{1,n} = \psi^{k,l}_{1,n} = 0$. In particular, $\xi^{n,d}_{1,n} = \psi^{n,d}_{1,n} = 0$. Hence, $\xi^{n,d}_{1,n}[\xi_{n,d}] + \psi^{n,d}_{1,n}\xi_{n,d} = 0$.

Case 2b: $p = b[q][a]$ where $a, b \in \mathcal{A}$ and $q \in \mathfrak{S}^{\mathbf{t}}$.

$b[q][a]:$ 

Let $P_q = \begin{pmatrix} 1 & c_{1,2} & \dots & c_{1,d-1} \\ & 1 & \dots & c_{2,d-2} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{d-2}(\mathcal{A})$, then

$$P_p = \begin{pmatrix} 1 & a_{1,2} & 0 & 0 & \dots & 0 \\ & 1 & a_{2,3} & 0 & \dots & 0 \\ & & 1 & a_{3,4} & \dots & a_{3,d} \\ & & & 1 & \dots & a_{4,d} \\ & & & & \ddots & \vdots \\ & & & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & a & 0 & 0 & \dots & 0 \\ & 1 & -b & 0 & \dots & 0 \\ & & 1 & c_{1,2} & \dots & c_{1,d-2} \\ & & & 1 & \dots & c_{2,d-2} \\ & & & & \ddots & \vdots \\ & & & & & 1 \end{pmatrix} = \left(\begin{array}{cc|c} 1 & a & 0\ 0\ \dots\ 0 \\ & 1 & -b\ 0\ \dots\ 0 \\ \hline & & P_q \end{array}\right) \in M_d(\mathcal{A}).$$

After redefining $\prec_q$ so that it is now defined on I_q we have

$$\prec_p = \Big((I_q, \prec_q), (2,3), \ldots, (2, d-1), (1,2), (2,d), (1,3), \ldots, (1,d) \Big).$$

The only difference between this case and Case 1a is that we switched the entries $(2,d)$ and $(1,2)$. This change only affects the $(1,d)$-entry of P_p, all other computations remain the same. From formulas (2.1),

$$\begin{aligned} \xi_{1,d} &= -\sum_{n=3}^{d-1} \xi_{1,n}^{n,d}[\xi_{n,d}] - \sum_{n=3}^{d-1} \xi_{n,d}\psi_{1,n}^{n,d} + \big([\xi_{1,2}] - \varphi_{1,2}\big)\xi_{2,d}, \\ \varphi_{1,d} &= -\sum_{n=3}^{d-1} \varphi_{1,n}^{n,d}[\xi_{n,d}] + \big([\xi_{1,2}] - \varphi_{1,2}\big)[\xi_{2,d}]. \end{aligned}$$

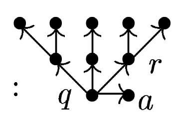
We have checked in Case 1a that $\sum\limits_{n=3}^{d-1} \xi_{1,n}^{n,d}[\xi_{n,d}] + \sum\limits_{n=3}^{d-1} \xi_{n,d}\psi_{1,n}^{n,d} = 0$ and $\mathbf{b}\big(\sum\limits_{n=3}^{d-1} \varphi_{1,n}^{n,d}[\xi_{n,d}]\big) = 0$. Since $\xi_{1,2} = \varphi_{1,2} = a$ and by our induction hypothesis $\mathbf{c}(\xi_{2,d}) = b[q] + o(b[q]) + o([b[q]])$, we obtain by Lemma 2.24(b)

$$\begin{aligned} \mathbf{c}(\xi_{1,d}) = \mathbf{c}\big(([a]-a)\xi_{2,d}\big) &= b[q][a] + o(b[q][a]) + o([b[q]][a]) + ab[q] + o(ab[q]) + o(a[b[q]]) \\ &= p + o(p) + o([p]) \end{aligned}$$

and

$$\mathbf{c}^{\mathbf{b}}(\varphi_{1,d}) = \mathbf{c}^{\mathbf{b}}\big(([a]-a)[\xi_{2,d}]\big) = [a][b[q]] + o([a][b[q]]) = o([p]).$$

Case 2c: $p = q[r][a]$ where $a \in \mathcal{A}$ and $q, r \in \mathfrak{S}^{\mathbf{t}}$ with $\mathrm{hgt}(q) > \mathrm{hgt}(r)$.

$q[r][a]:$ 

Let $d_1 = \mathrm{wgt}(q)+1$, $d_2 = \mathrm{wgt}(r)+1$, $P_q = \begin{pmatrix} 1 & b_{1,2} & \ldots & b_{1,d_1} \\ & 1 & \ldots & b_{2,d_1} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{d_1}(\mathcal{A})$ and $P_r = \begin{pmatrix} 1 & c_{1,2} & \ldots & c_{1,d_2} \\ & 1 & \ldots & c_{2,d_2} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{d_2}(\mathcal{A})$, then

$$P_p = \begin{pmatrix} 1 & a_{1,2} & 0 & \ldots & 0 & 0 & \ldots & 0 \\ & 1 & a_{2,3} & \ldots & a_{1,d_2+1} & 0 & \ldots & 0 \\ & & 1 & \ldots & a_{3,d_2+1} & 0 & \ldots & 0 \\ & & & \ddots & \vdots & \vdots & & \vdots \\ & & & & 1 & a_{d_2+1,d_2+2} & \ldots & a_{d_2+1,d} \\ & & & & & 1 & \ldots & a_{d_2+2,d} \\ & & & & & & \ddots & \vdots \\ & & & & & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & a & 0 & \ldots & 0 & 0 & \ldots & 0 \\ & 1 & c_{1,2} & \ldots & c_{1,d_2} & 0 & \ldots & 0 \\ & & 1 & \ldots & c_{2,d_2} & 0 & \ldots & 0 \\ & & & \ddots & \vdots & \vdots & & \vdots \\ & & & & 1 & b_{1,2} & \ldots & b_{1,d_1} \\ & & & & & 1 & \ldots & b_{2,d_1} \\ & & & & & & \ddots & \vdots \\ & & & & & & & 1 \end{pmatrix} \in M_d(\mathcal{A}).$$

Let $I_q = \{(i,j)\}_{d_2+1\le i<j\le d}$, $I_r = \{(i,j)\}_{2\le i<j\le d_2+1}$ and $J = \{(i,j)\}_{\substack{2\le i\le d_2 \\ d_2+2\le j\le d}}$. After redefining $\prec_q$ and $\prec_r$ so that they are now defined on I_q and I_r respectively, we have $\prec_p = \big(\big(I_q\setminus\{(d_2+1,d)\},\prec_q\big),\big(I_r,\prec_r\big),(1,2),(d_2+1,d),\big(J,\prec_J\big),\ldots,(1,d)\big)$.

The difference between this case and Case 1b is that we switched the entries (d_2+1,d) and $(1,2)$; this change of order only affects the $(1,d)$-entry of P_p, all other computations remain the same. From formulas (2.1),

$$\begin{aligned}\xi_{1,d} &= -\sum_{n=3}^{d-1}\xi_{1,n}^{n,d}[\xi_{n,d}] - \sum_{n=3}^{d-1}\psi_{1,n}^{n,d}\xi_{n,d} + \big([\xi_{1,2}]-\varphi_{1,2}\big)\xi_{2,d},\\ \varphi_{1,d} &= -\sum_{n=3}^{d-1}\varphi_{1,n}^{n,d}[\xi_{n,d}] + \big([\xi_{1,2}]-\varphi_{1,2}\big)[\xi_{2,d}].\end{aligned}$$

We have checked in Case 1b that $\mathbf{c}\big(\sum_{n=3}^{d-1}\xi_{1,n}^{n,d}[\xi_{n,d}]+\sum_{n=3}^{d-1}\psi_{1,n}^{n,d}\xi_{n,d}\big) < a[q[r]] < p$ and $\mathbf{b}\big(\sum_{n=3}^{d-1}\varphi_{1,n}^{n,d}[\xi_{n,d}]\big) = 0$. We have $\xi_{1,2}=\varphi_{1,2}=a$ and $\mathbf{c}(\xi_{2,d}) = q[r]+o(q[r])+o([q[r]])$, thus by Lemma 2.24(b),

$$\begin{aligned}\mathbf{c}\big(([\xi_{1,2}]-\varphi_{1,2})\xi_{2,d}\big) &= q[r][a]+o(q[r][a])+o([q[r]][a])+aq[r]+o(aq[r])+o(a[q[r]])\\ &= p+o(p)+o([p])\end{aligned}$$

and

$$\mathbf{c}^{\mathbf{b}}\big(([\xi_{1,2}]-\varphi_{1,2})[\xi_{2,d}]\big) = [q[r]][a]+o([q[r]][a]) = o([p]).$$

■

2.39. We want to emphasize again that Proposition 2.17 (which has not been proven yet!) does not imply Theorem A_1^* because it says nothing about the nilpotency class of the group necessary for obtaining an element $u\in\mathfrak{B}^{\mathbf{t}}$. Let us return to the notation introduced in the beginning of Section 2; we will compute the nilpotency class of certain subgroups of M_d, $d\in\mathbb{N}$.

Let us say that a set $\mathcal{T}\subseteq\big\{(i,j): 1\le i<j\le d\big\}$ is *transitive* if its satisfies the following condition: $(i,j),(j,k)\in\mathcal{T}$ implies $(i,k)\in\mathcal{T}$. Given a transitive set $\mathcal{T}\subseteq\big\{(i,j): 1\le i<j\le d\big\}$,

$$M_{\mathcal{T}} = \left\{\begin{pmatrix}1 & a_{1,2} & a_{1,3} & \cdots & a_{1,d}\\ & 1 & a_{2,3} & \cdots & a_{2,d}\\ & & \ddots & & \vdots\\ & & & 1 & a_{d-1,d}\\ & & & & 1\end{pmatrix}\in M_d : a_{i,j}=0 \text{ whenever } (i,j)\notin\mathcal{T}\right\}$$

is a connected Lie subgroup of M_d, and $\{E_{i,j}\}_{(i,j)\in\mathcal{T}}$ is a basis in $M_{\mathcal{T}}$. We will now determine the nilpotency class of $M_{\mathcal{T}}$.

2.40. Let $\mathcal{T}\subseteq\big\{(i,j): 1\le i<j\le d\big\}$ be a transitive set. For $(i,j)\in\mathcal{T}$ let us define $\mathrm{step}_{\mathcal{T}}(i,j)$ to be the maximal length of a chain connecting i and j in $\mathcal{T}$, that is, the maximal integer m for which there exist $k_1,\ldots,k_{m-1}\in\{1,\ldots,d\}$ with $(i,k_1),(k_1,k_2),\ldots,(k_{m-2},k_{m-1}),(k_{m-1},j)\in\mathcal{T}$. We also define $\mathrm{step}(\mathcal{T}) = \max\big\{\mathrm{step}_{\mathcal{T}}(i,j):(i,j)\in\mathcal{T}\big\}$.

Lemma. *The nilpotency class of $M_{\mathcal{T}}$ is equal to* $\operatorname{step}(\mathcal{T})$*.*

Proof. We have $[E_{i,j}, E_{k,l}] = \begin{cases} E_{i,l} \text{ if } j = k \\ \mathbf{1} \text{ otherwise.} \end{cases}$ It follows that the m-th member of the lower central series $(M_{\mathcal{T}})_1 = M_{\mathcal{T}}$, $(M_{\mathcal{T}})_m = \big[(M_{\mathcal{T}})_{m-1}, M_{\mathcal{T}}\big]$, $m = 2, 3, \ldots$, of $M_{\mathcal{T}}$ is generated by $\big\{E_{i,j} : (i,j) \in \mathcal{T},\ \operatorname{step}_{\mathcal{T}}(i,j) = m\big\}$. ■

2.41. Let us return to the bracket algebra $\mathfrak{B}$ over a ring $\mathcal{A}$. For $\delta \in \mathbb{N}$ define

$$\mathfrak{M}_\delta = \Big\{u \in \mathfrak{B}^{\mathbf{t}} : \begin{array}{l}\text{there exist } d \in \mathbb{N}, \text{ a transitive set } \mathcal{T} \subseteq \big\{(i,j) : 1 \le i < j \le d\big\} \\ \text{with } (1,d) \in \mathcal{T} \text{ and } \operatorname{step}(\mathcal{T}) \le \delta, \text{ a matrix } P \in M_{\mathcal{T}}(\mathcal{A}), \text{ a sign} \\ \text{matrix } \epsilon = (\epsilon_{i,j})_{1 \le i < j \le d} \text{ and a legal order } \prec \text{ on } \{(i,j)\}_{1 \le i < j \le d} \\ \text{such that } u = \mathbf{t}\big(\xi_{1,d}(P, \epsilon, \prec)\big)\end{array}\Big\}$$

and

$$\mathfrak{M} = \big\{u \in \mathfrak{B}^{\mathbf{t}} : u \in \mathfrak{M}_\delta \text{ with } \delta = \operatorname{wgt}(u)\big\}.$$

We will to prove the following enhancement of Proposition 2.17, which implies Theorem A_1^*:

Theorem A.** $\mathfrak{M} = \mathfrak{B}^{\mathbf{t}}$.

2.42. Lemma. *If $u \in \mathfrak{M}_\delta$ then $-u \in \mathfrak{M}_\delta$. In particular, $u \in \mathfrak{M}$ implies $-u \in \mathfrak{M}$.*

Proof. Let $u = \mathbf{t}\big(\xi_{1,d}(P, \epsilon, \prec)\big)$ for $d \in \mathbb{N}$, a transitive set $\mathcal{T} \subseteq \big\{(i,j) : 1 \le i < j \le d\big\}$ with $(1,d) \in \mathcal{T}$ and $\operatorname{step}(\mathcal{T}) \le \delta$, a matrix $P = \begin{pmatrix} 1 & a_{1,2} & a_{1,3} & \ldots & a_{1,d} \\ & 1 & a_{2,3} & \ldots & a_{2,d} \\ & & \ddots & & \vdots \\ & & & 1 & a_{d-1,d} \\ & & & & 1 \end{pmatrix} \in M_{\mathcal{T}}(\mathcal{A})$, a sign matrix $\epsilon = (\epsilon_{i,j})_{1 \le i < j \le d}$ and a legal order $\prec$ on $\{(i,j)\}_{1 \le i < j \le d}$. Define $P' = \begin{pmatrix} 1 & -a_{1,2} & -a_{1,3} & \ldots & -a_{1,d} \\ & 1 & a_{2,3} & \ldots & a_{2,d} \\ & & \ddots & & \vdots \\ & & & 1 & a_{d-1,d} \\ & & & & 1 \end{pmatrix} \in M_{\mathcal{T}}(\mathcal{A})$ and a sign matrix $\epsilon' = (\epsilon'_{i,j})_{1 \le i < j \le d}$ by $\epsilon'_{i,j} = \begin{cases} -\epsilon_{i,j} \text{ if } i = 1 \\ \epsilon_{i,j} \text{ otherwise,} \end{cases}$ $1 \le i < j \le d$. Using the identity $[-u]^{-1} = -[u]^1$ one checks from formulas (2.1) by induction on (i,j) that

$$\begin{gathered}\varphi^{k,l}_{1,j}(P', \epsilon', \prec) = -\varphi^{k,l}_{1,j}(P, \epsilon, \prec), \quad \psi^{k,l}_{1,j}(P', \epsilon', \prec) = -\psi^{k,l}_{1,j}(P, \epsilon, \prec), \\ \xi^{k,l}_{1,j}(P', \epsilon', \prec) = -\xi^{k,l}_{1,j}(P, \epsilon, \prec) \qquad \text{for any } 1 < j \le d \text{ and } (k,l) \prec (1,j)\end{gathered}$$

and

$$\begin{gathered}\varphi^{k,l}_{i,j}(P', \epsilon', \prec) = \varphi^{k,l}_{i,j}(P, \epsilon, \prec), \quad \psi^{k,l}_{i,j}(P', \epsilon', \prec) = \psi^{k,l}_{i,j}(P, \epsilon, \prec), \\ \xi^{k,l}_{i,j}(P', \epsilon', \prec) = \xi^{k,l}_{i,j}(P, \epsilon, \prec) \qquad \text{for any } 2 \le i < j \le d \text{ and } (k,l) \prec (1,j).\end{gathered}$$

In particular, $\xi_{1,d}(P', \epsilon', \prec) = -\xi_{1,d}(P, \epsilon, \prec)$, so $\mathbf{t}\big(\xi_{1,d}(P', \epsilon', \prec)\big) = -\,\mathbf{t}\big(\xi_{1,d}(P, \epsilon, \prec)\big) = -u$ and $-u \in \mathfrak{M}_\delta$. ■

2.43. Lemma. *If $u \in \mathfrak{M}_{\delta_1}$ and $v \in \mathfrak{M}_{\delta_2}$ then $u + v \in \mathfrak{M}_{\max\{\delta_1, \delta_2\}}$. In particular, $u, v \in \mathfrak{M}$ implies $u + v \in \mathfrak{M}$.*

Proof. Let $u = \mathbf{t}\big(\xi_{1,d}(P, \epsilon_1, \prec_1)\big)$ for $d_1 \in \mathbb{N}$, a transitive set $\mathcal{T}_1 \subseteq \big\{(i,j) : 1 \le i < j \le d_1\big\}$ with $(1,d_1) \in \mathcal{T}_1$ and $\operatorname{step}(\mathcal{T}_1) \le \delta_1$, a matrix $R = \begin{pmatrix} 1 & a_{1,2} & \dots & a_{1,d_1} \\ & 1 & \dots & a_{2,d_1} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{\mathcal{T}_1}(\mathcal{A})$, a sign matrix ϵ_1 and an order $\prec_1$, and let $v = \big(\xi_{1,d_2}(Q, \epsilon_2, \prec_2)\big)$ for $d_2 \in \mathbb{N}$, a transitive set $\mathcal{T}_2 \subseteq \big\{(i,j) : 1 \le i < j \le d_2\big\}$ with $(1,d_2) \in \mathcal{T}_2$ and $\operatorname{step}(\mathcal{T}_2) \le \delta_2$, a matrix $S = \begin{pmatrix} 1 & b_{1,2} & \dots & b_{1,d_2} \\ & 1 & \dots & b_{2,d_2} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{\mathcal{T}_2}(\mathcal{A})$, a sign matrix ϵ_2 and an order $\prec_2$. Put $d = d_1 + d_2$ and define

$$P = \begin{pmatrix}
1 & a_{1,2} & \dots & a_{1,d_1-1} & b_{1,2} & b_{1,3} & \dots & b_{1,d_2-1} & a_{1,d_1}+b_{1,d_2} \\
& 1 & \dots & a_{2,d_1-1} & 0 & 0 & \dots & 0 & a_{2,d_1} \\
& & \ddots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
& & & a_{d_1-2,d_1-1} & 0 & 0 & \dots & 0 & a_{d_1-2,d_1} \\
& & & 1 & 0 & 0 & \dots & 0 & a_{d_1-1,d_1} \\
& & & & 1 & b_{2,3} & \dots & b_{2,d_2-1} & b_{2,d_2} \\
& & & & & 1 & \dots & b_{3,d_2-1} & b_{3,d_2} \\
& & & & & & \ddots & \vdots & \vdots \\
& & & & & & & 1 & b_{d_2-1,d_2} \\
& & & & & & & & 1
\end{pmatrix} \in M_d(\mathcal{A}).$$

That is, R occupies the submatrix of P indexed by $I_R = \{(i,j)\}_{i,j\in\{1,2,\dots,d_1-1,d\}, i<j}$ and S occupies the submatrix of P indexed by $I_S = \{(i,j)\}_{i,j\in\{1,d_1,d_1+1,\dots,d\}, i<j}$; the only common entry of these submatrices is the $(1,d)$-entry, which equals $a_{1,d_1} + b_{1,d_2}$. We will identify R and S with their images in P and redefine ϵ_1, $\prec_1$, $\mathcal{T}_1$, ϵ_2, $\prec_2$ and $\mathcal{T}_2$ accordingly.

Put $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2$; then $(1,d) \in \mathcal{T}$, $\operatorname{step}(\mathcal{T}) = \max\big\{\operatorname{step}(\mathcal{T}_1), \operatorname{step}(\mathcal{T}_2)\big\} \le \max\{\delta_1, \delta_2\}$, and $P \in M_{\mathcal{T}}(\mathcal{A})$. Let $\epsilon = (\epsilon_{i,j})_{1\le i<j\le d}$ be any sign matrix whose restrictions on I_R and I_S coincide with ϵ_1 and ϵ_2 respectively; the "common entry" $\epsilon_{1,d} = 1$. Let $\prec$ be any legal order on $\{(i,j)\}_{1\le i<j\le d}$ such that the restriction of $\prec$ on I_R and on I_S coincides with $\prec_1$ and $\prec_2$ respectively.

When one computes $\xi_{i,j}(P, \epsilon, \prec)$ by formulas (2.1), the entries of R do not affect the entries of S and vice versa, except the common $(1,d)$-entry, which accumulates values from both R and S. More exactly, one checks by induction that if $(i,j) \notin I_R \cup I_S$, then $\varphi^{k,l}_{i,j}(P), \psi^{k,l}_{i,j}(P), \xi^{k,l}_{i,j}(P) = 0$; if $(i,j) \in I_R \setminus \{(1,k)\}$, then $\varphi^{k,l}_{i,j}(P)$, $\psi^{k,l}_{i,j}(P)$ and $\xi^{k,l}_{i,j}(P)$ are equal to the corresponding $\varphi^{k,l}_{i,j}(R)$, $\psi^{k,l}_{i,j}(R)$ and $\xi^{k,l}_{i,j}(R)$; if $(i,j) \in I_S \setminus \{(1,k)\}$, then $\varphi^{k,l}_{i,j}(P)$, $\psi^{k,l}_{i,j}(P)$ and $\xi^{k,l}_{i,j}(P)$ are equal to the corresponding $\varphi^{k,l}_{i,j}(S)$, $\psi^{k,l}_{i,j}(S)$ and $\xi^{k,l}_{i,j}(S)$; and finally, $\xi_{1,d}(P) = \xi_{1,d_1}(R) + \xi_{1,d_2}(S)$. Hence, $u + v = \mathbf{t}\big(\xi_{1,d}(P)\big) \in \mathfrak{M}_{\max\{\delta_1,\delta_2\}}$. ■

2.44. Lemma. *If $u \in \mathfrak{M}_{\delta_1}$ and $v \in \mathfrak{M}_{\delta_2}$ then $[u]v + u[v] - uv \in \mathfrak{M}_{\delta_1+\delta_2}$. In particular, $u, v \in \mathfrak{M}$ implies $[u]v + u[v] - uv \in \mathfrak{M}$.*

Proof. Let $u = \mathbf{t}\big(\xi_{1,d_1}(R, \epsilon_1, \prec_1)\big)$ for $d_1 \in \mathbb{N}$, a transitive set $\mathcal{T}_1 \subseteq \big\{(i,j) : 1 \le i < j \le d_1\big\}$ with $(1,d_1) \in \mathcal{T}_1$ and $\operatorname{step}(\mathcal{T}_1) \le \delta_1$, a matrix $R = \begin{pmatrix} 1 & a_{1,2} & \dots & a_{1,d_1} \\ & 1 & \dots & a_{2,d_1} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{\mathcal{T}_1}(\mathcal{A})$, a sign matrix ϵ_1 and an order $\prec_1$, and let $v = \mathbf{t}\big(\xi_{1,d_2}(S, \epsilon_2, \prec_2)\big)$ for $d_2 \in \mathbb{N}$, a transitive set $\mathcal{T}_2 \subseteq \big\{(i,j) : 1 \le i < j \le d_2\big\}$ with $(1, d_2) \in \mathcal{T}_2$ and $\operatorname{step}(\mathcal{T}_2) \le \delta_2$, a matrix $S = \begin{pmatrix} 1 & b_{1,2} & \dots & b_{1,d_2} \\ & 1 & \dots & b_{2,d_2} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{\mathcal{T}_2}(\mathcal{A})$, a sign matrix ϵ_2 and an order $\prec_2$. Put $d = d_1 + d_2 - 1$ and define

$$P = \begin{pmatrix} 1 & a_{1,2} & \dots & a_{1,d_1} & 0 & \dots & 0 \\ & 1 & \dots & a_{2,d_1} & 0 & \dots & 0 \\ & & \ddots & \vdots & \vdots & & \vdots \\ & & & 1 & b_{1,2} & \dots & b_{1,d_2} \\ & & & & 1 & \dots & b_{2,d_2} \\ & & & & & \ddots & \vdots \\ & & & & & & 1 \end{pmatrix} \in M_d(\mathcal{A}).$$

That is, R occupies the submatrix of P indexed by $I_R = \{(i,j)\}_{1 \le i < j \le d_1}$ and S occupies the submatrix of P indexed by $I_S = \{(i,j)\}_{d_1 \le i < j \le d}$; we will identify R and S with their images in P and redefine ϵ_1, $\prec_1$, $\mathcal{T}_1$, ϵ_2, $\prec_2$ and $\mathcal{T}_2$ accordingly.

Let $\mathcal{T}$ be the transitive subset of $\big\{(i,j) : 1 \le i < j \le d\big\}$ generated by $\mathcal{T}_1$ and $\mathcal{T}_2$, that is, the minimal transitive subset containing $\mathcal{T}_1 \cup \mathcal{T}_2$. Then one checks that $(1,d) \in \mathcal{T}$, $\operatorname{step}(\mathcal{T}) \le \operatorname{step}(\mathcal{T}_1) + \operatorname{step}(\mathcal{T}_2) \le \delta_1 + \delta_2$ and $P \in M_{\mathcal{T}}(\mathcal{A})$. Let ϵ be any sign matrix whose restrictions on I_R and I_S coincide with ϵ_1 and ϵ_2 respectively, except for ϵ_{1,d_1} and $\epsilon_{d_1,d}$, which we put equal to 1. Introduce two different orders $\prec$ and $\prec'$ on $\{(i,j)\}_{1 \le i < j \le d}$ in the following way. Let $\prec_J$ be any legal order on $J = \{(i,j)\}_{\substack{1 \le i \le d_1 - 1 \\ d_1 + 1 \le j \le d}}$; put $\prec$ to be

$$\Big(\big(I_R \setminus \{(1,d_1)\}, \prec_1\big), \big(I_S \setminus \{(d_1,d)\}, \prec_2\big), (1,d_1), (d_1,d), \big(J, \prec_J\big)\Big)$$

and $\prec'$ to be

$$\Big(\big(I_R \setminus \{(1,d_1)\}, \prec_1\big), \big(I_S \setminus \{(d_1,d)\}, \prec_2\big), (d_1,d), (1,d_1), \big(J, \prec_J\big)\Big).$$

(In plain words, first the entries of R excluding a_{1,d_1} appear, then follow the entries of S excluding b_{1,d_2}, then follow a_{1,d_1} and b_{1,d_2}, then all other entries of P follow; $\prec'$ is obtained from $\prec$ by switching the order of a_{1,d_1} and b_{1,d_2}.)

The entries of R do not affect the entries of S, and vice versa. Thus, for both $\prec$ and $\prec'$, the elements $\varphi^{k,l}_{i,j}(P), \psi^{k,l}_{i,j}(P), \xi^{k,l}_{i,j}(P)$ with $(i,j) \in I_R$ are equal to the corresponding $\varphi^{k,l}_{i,j}(R), \psi^{k,l}_{i,j}(R), \xi^{k,l}_{i,j}(R)$, and the elements $\varphi^{k,l}_{i,j}(P), \psi^{k,l}_{i,j}(P), \xi^{k,l}_{i,j}(P)$ with $(i,j) \in I_S$ are equal to the corresponding $\varphi^{k,l}_{i,j}(S), \psi^{k,l}_{i,j}(S), \xi^{k,l}_{i,j}(S)$. The difference between the orders $\prec$ and $\prec'$ only affects the last, $(1,d)$-entry of P. Since with respect to $\prec'$ the entry (d_1,d) precedes $(1,d_1)$ and does not affect it, $\xi^{d_1,d}_{1,d_1}(P,\prec') = \xi_{1,d_1}(P,\prec')$ and $\psi^{d_1,d}_{1,d_1}(P,\prec') = \psi_{1,d_1}(P,\prec')$. From formulas (2.1) we now have

$$\begin{aligned} \xi_{1,k}(P,\prec) - \xi_{1,k}(P,\prec') &= [\xi_{1,d_1}]^{\epsilon_{1,d_1}} \xi_{d_1,d} - \varphi_{1,d_1}\xi_{d_1,d} + \xi_{1,d_1}[\xi_{d_1,d}]^{\epsilon_{d_1,d}} + \psi_{1,d_1}\xi_{d_1,d} \\ &= [\xi_{1,d_1}]\xi_{d_1,d} + \xi_{1,d_1}[\xi_{d_1,d}] - \big(\varphi_{1,d_1} - \psi_{1,d_1}\big)\xi_{d_1,d} = [\xi_{1,d_1}]\xi_{d_1,d} + \xi_{1,d_1}[\xi_{d_1,d}] - \xi_{1,d_1}\xi_{d_1,d}. \end{aligned}$$

Hence,

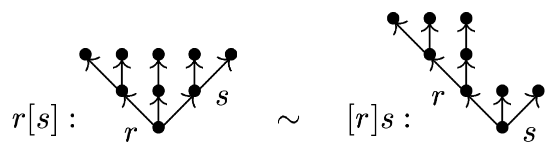
$$\begin{aligned} \mathbf{t}(\xi_{1,d}(P,\prec)) - \mathbf{t}(\xi_{1,d}(P,\prec')) &= \mathbf{t}\big(\xi_{1,d}(P,\prec) - \xi_{1,d}(P,\prec')\big) \\ &= [\mathbf{t}(\xi_{1,d_1})]\,\mathbf{t}(\xi_{d_1,d}) + \mathbf{b}(\xi_{1,d_1})\,\mathbf{t}(\xi_{d_1,d}) + \mathbf{t}(\xi_{1,d_1})[\mathbf{t}(\xi_{d_1,d})] + \mathbf{t}(\xi_{1,d_1})\,\mathbf{b}(\xi_{d_1,d}) \\ &\quad - \mathbf{t}(\xi_{1,d_1})\,\mathbf{t}(\xi_{d_1,d}) - \mathbf{t}(\xi_{1,d_1})]\,\mathbf{b}(\xi_{d_1,d}) - \mathbf{b}(\xi_{1,d_1})]\,\mathbf{t}(\xi_{d_1,d}) \\ &= [u]v + u[v] - uv. \end{aligned}$$

Since $\mathbf{t}(\xi_{1,d}(P,\prec)), \mathbf{t}(\xi_{1,d}(P,\prec')) \in \mathfrak{M}_{\delta_1+\delta_2}$, by Lemmas 2.42 and 2.43, $[u]v + u[v] - uv = \mathbf{t}(\xi_{1,d}(P,\prec)) - \mathbf{t}(\xi_{1,d}(P,\prec')) \in \mathfrak{M}_{\delta_1+\delta_2}$. ■

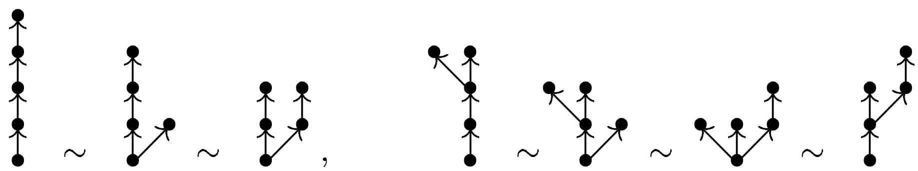
2.45. From Lemmas 2.44, 2.42 and 2.43 we derive the following:

Lemma. *If* $u, v, [u]v, uv \in \mathfrak{M}$ *then* $u[v] \in \mathfrak{M}$.

2.46. Let "$\sim$" be the minimal equivalence relation on the set of trees $\mathfrak{S}^{\mathbf{t}}$ for which $r[s] \sim [r]s$ for any $r, s \in \mathfrak{S}^{\mathbf{t}}$. Graphically, two trees are equivalent if they are obtainable from each other by changing the position of the root vertex:



Examples.

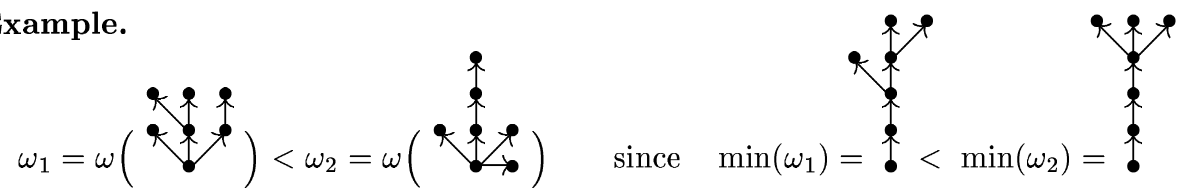


2.47. Lemma. *Let* $u, v \in \mathfrak{B}^{\mathbf{t}}$ *and* $\mathbf{c}([u]v) = p_1 + \ldots + p_m$. *Then* $\mathbf{c}(u[v]) = q_1 + \ldots + q_m$ *with* $q_1 \sim p_1, \ldots, q_m \sim p_m$.

2.48. Let Ω be the set of equivalence classes for $\sim$. For $p \in \mathfrak{S}^{\mathbf{t}}$ we denote by $\omega(p)$ the class in Ω that contains p, and for $u \in \mathfrak{B}^{\mathbf{t}}$ let $\omega(u) = \max\{\omega(p) : p \in \mathbf{c}(u)\}$.

We define an order on Ω in the following way: for $\omega_1, \omega_2 \in \Omega$ we write $\omega_1 < \omega_2$ if $\min(\omega_1) < \min(\omega_2)$.

Example.



2.49. We need more notation. We define *the number-of-pluses*, "nop", of elements of $\mathfrak{B}^{\mathbf{t}}$ in the following way: $\mathrm{nop}(a) = 0$ for $a \in \mathcal{A}$, $\mathrm{nop}\big(a[v_1]\ldots[v_m]\big) = \mathrm{nop}(v_1) + \ldots + \mathrm{nop}(v_m)$ for $a \in \mathcal{A}$ and $v_1, \ldots, v_m \in \mathfrak{B}^{\mathbf{t}}$ and $\mathrm{nop}(u_1 + u_2) = \mathrm{nop}(u_1) + \mathrm{nop}(u_2) + 1$ for $u_1, u_2 \in \mathfrak{B}^{\mathbf{t}}$. Note that $\mathrm{nop}(u) = 0$ implies $u \in \mathfrak{S}^{\mathbf{t}}$.

The minimal-depth-of-a-plus, "dop", of elements of $\mathfrak{B}^{\mathbf{t}} \setminus \mathfrak{S}^{\mathbf{t}}$ is defined by the following rules: $\mathrm{dop}(u_1 + u_2) = 0$ for $u_1, u_2 \in \mathfrak{B}^{\mathbf{t}}$ and $\mathrm{dop}\big(a[v_1]\ldots[v_m]\big) = 1 + \min\{\mathrm{dop}(v_1), \ldots, \mathrm{dop}(v_m)\}$ for $a \in \mathcal{A}$ and $v_1, \ldots, v_m \in \mathfrak{B}^{\mathbf{t}}$.

Example. For $u = a_1[a_2[a_3[a_4 + a_5]]$ one has $\mathrm{nop}(u) = 1$ and $\mathrm{dop}(u) = 3$, for $v = a_1[a_2[a_3[a_4 + a_5]]][a_6 + a_7]$ one has $\mathrm{nop}(v) = 2$ and $\mathrm{dop}(v) = 1$.

2.50. Proof of Theorem A.** We will use induction on Ω; fix $\omega \in \Omega$ and assume that $v \in \mathfrak{M}$ for any $v \in \mathfrak{B}^{\mathbf{t}}$ with $\omega(v) < \omega$.

We will first show that $\omega \cap \mathfrak{M} \neq \emptyset$. We will use induction on $\mathrm{nop}(u)$ and $\mathrm{dop}(u)$ of elements $u \in \mathfrak{M}$ for which $\mathbf{c}(u) = p + p_1 + \ldots + p_k$ with $p \in \omega$ and $\omega(p_1), \ldots, \omega(p_k) < \omega$.

First of all, such an element u exists. Indeed, let p be the minimal element of ω. By Proposition 2.38 there exists $u \in \mathfrak{M}$ such that $\mathbf{c}(u) = p + p_1 + \ldots + p_k$ with $p_1, \ldots, p_k < p$. Since p is the minimal element of ω we have $\omega(p_1), \ldots, \omega(p_k) < \omega$.

If $\mathrm{nop}(u) = 0$ then $u \in \mathfrak{S}^{\mathbf{t}}$ and so, $p = u \in \mathfrak{M} \cap \omega$. Let $\mathrm{nop}(u) > 0$. If $\mathrm{dop}(u) = 0$ then $u = u_1 + u_2$. Assume that $\omega(u_1) = \omega$, then $\omega(u_2) < \omega$. By our induction hypothesis $u_2 \in \mathfrak{M}$, and hence, $u_1 = u - u_2 \in \mathfrak{M}$ by Lemmas 2.42 and 2.43. Since $\mathrm{nop}(u_1) < \mathrm{nop}(u)$, by induction on $\mathrm{nop}(u)$ we have $p \in \mathfrak{M}$.

If $\mathrm{nop}(u) > 0$ and $\mathrm{dop}(u) > 0$ represent $u = a[v_1][v_2] \ldots [v_m]$, with $a \in \mathcal{A}$ and $v_1, \ldots, v_m \in \mathfrak{B}^{\mathbf{t}}$, so that $\mathrm{dop}(v_1) \leq \mathrm{dop}(v_i)$ for $i = 2, \ldots, m$ and so, $\mathrm{dop}(u) = \mathrm{dop}(v_1) + 1$. Define $v = a[v_2] \ldots [v_m]$, then $u = [v_1]v$. Since $\mathrm{wgt}(v_1), \mathrm{wgt}(v), \mathrm{wgt}(v_1 v) < \mathrm{wgt}(u)$, by our induction hypothesis we have $v_1, v, v_1 v \in \mathfrak{M}$. Thus by Lemma 2.45, $u' = v_1[v] \in \mathfrak{M}$. By Lemma 2.47, $\mathbf{c}(u') = q + q_1 + \ldots + q_k$ with $\omega(q) = \omega(p) = \omega$ and $\omega(q_i) = \omega(p_i) < \omega$, $i = 1, \ldots, k$. Since $\mathrm{dop}(u') \leq \mathrm{dop}(v_1) < \mathrm{dop}(u)$, by induction on $\mathrm{dop}(u)$ we have $q \in \mathfrak{M}$.

We will now show that every element of ω belongs to $\mathfrak{M}$. Indeed, if $q \in \mathfrak{M} \cap \omega$ and $q = r[s]$ with $r, s \in \mathfrak{S}^{\mathbf{t}}$, then, since by the induction hypothesis $r, s, rs \in \mathfrak{M}$, Lemma 2.45 states that $[r]s \in \mathfrak{M}$.

Now, let u be an arbitrary element of $\mathfrak{B}^{\mathbf{t}}$ with $\omega(u) = \omega$. We will show by induction on $\mathrm{nop}(u)$ and $\mathrm{dop}(u)$ that $u \in \mathfrak{M}$. If $\mathrm{nop}(u) = 0$ then $u \in \mathfrak{S}^{\mathbf{t}}$, so $u \in \omega$ and $u \in \mathfrak{M}$ is proved. Let $\mathrm{nop}(u) > 0$. If $\mathrm{dop}(u) = 0$ then $u = u_1 + u_2$, by induction on $\mathrm{nop}(u)$ we have $u_1, u_2 \in \mathfrak{M}$ and by Lemma 2.43, $u \in \mathfrak{M}$. Let $\mathrm{nop}(u) > 0$ and $\mathrm{dop}(u) > 0$. Represent $u = [v_1]v$ so that $\mathrm{dop}(u) = \mathrm{dop}(v_1) + 1$. Define $u' = v_1[v]$, then $\omega(u') = \omega(u) = \omega$ and $\mathrm{dop}(u') \leq \mathrm{dop}(v_1) < \mathrm{dop}(u)$. By induction on $\mathrm{dop}(u)$ we have $u' \in \mathfrak{M}$ and since $\mathrm{wgt}(v_1), \mathrm{wgt}(v), \mathrm{wgt}(v_1 v) < \mathrm{wgt}(u)$, by our induction hypothesis $v_1, v, v_1 v \in \mathfrak{M}$. By Lemma 2.45, $u \in \mathfrak{M}$. ■

Bibliography

[AGH] L. Auslander, L. Green and F. Hahn, *Flows on homogeneous spaces*, Annals of Math. Studies, vol. 53, Princeton Univ. Press, 1963.

[BHå] V. Bergelson and I. J. Håland, Sets of recurrence and generalized polynomials, *Convergence in Ergodic Theory and Probability*, Walter de Gruyter, 1996, 91–110.

[BHåM] V. Bergelson, I. J. Håland and R. McCutcheon, Simultaneous Diophantine approximation and VIP systems, *Acta Arith.* **116** (2005,), no. 1, 13-23.

[Br] J. Brezin, Applications of nilmanifold theory to Diophantine approximations, *Proc. Amer. Math. Soc.* **33** (1972), 543-547.

[C] J. Cigler, On a theorem of H. Weyl, *Compositio Math.* **21** (1969), 151-154.

[vdC] J. G. van der Corput, Diophantische Ungleichungen I. Zur Gleichverteilung modulo Eins, *Acta Math.* **56** (1931), 373-456.

[F1] H. Furstenberg, *Stationary processes and prediction theory*, Annals of Math. Studies, vol. 44, Princeton Univ. Press, 1960.

[F2] H. Furstenberg, The structure of distal flows, *Amer. J. Math.* **85** (1963), 477-515.

[F3] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton Univ. Press, 1981.

[FW] H. Furstenberg and B. Weiss, Simultaneous Diophantine approximation and IP-sets, *Acta Arith.* **49** (1988), no. 4, 413–426.

[H] F. J. Hahn, On affine transformations of compact abelian groups, *Amer. J. Math.* **85** (1963), 428-446 (Errata: *Amer. J. Math.* **86** (1964), 463-464).

[Hå1] I. J. Håland, *Uniform distribution of generalized polynomials*, Ph.D. Thesis, The Ohio State University, 1992.

[Hå2] I. J. Håland, Uniform distribution of generalized polynomials, *J. Number Theory* **45** (1993), 327-366.

[Hå3] I. J. Håland, Uniform distribution of generalized polynomials of the product type, *Acta Arith.* **67** (1994), 13-27.

[K1] H. B. Keynes, Topological dynamics in coset transformation groups, *Bull. Amer. Math. Soc.* **72** (1966), 1033-1035.

[K2] H. B. Keynes, A study of the proximal relation in coset transformation groups, *Trans. Amer. Math. Soc.* **128** (1967), 389-402.

[L1] A. Leibman, Polynomial mappings of groups, *Israel J. of Mathematics* **129** (2002), 29–60.

[L2] A. Leibman, Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold, *Ergod. Th. and Dynam. Sys.* **25** (2005), 201-213.

[L3] A. Leibman, Pointwise convergence of ergodic averages for polynomial actions of $\mathbb{Z}^d$ by translations on a nilmanifold, *Ergod. Th. and Dynam. Sys.* **25** (2005), 215-225.

[L4] A. Leibman, Rational sub-nilmanifolds of a compact nilmanifold, preprint. Available at *http://www.math.ohio-state.edu/~leibman/preprints*

[Le] E. Lesigne, Sur une nil-variété, les parties minimales assocées à une translation sont uniquement ergodiques, *Ergod. Th. and Dynam. Sys.* **11** (1991), 379-391.

[Mal] A. Malcev, On a class of homogeneous spaces, *Izvestiya Akad. Nauk SSSR, Ser. Mat.* **13** (1949), 9-32.

[Mar] A. A. Markoff, Sur une question de Jean Bernoulli, *Math. Ann.* **19** (1882), 27-36.

[Sh1] N. Shah, Limit distributions of polynomial trajectories on homogeneous spaces, *Duke Math. J.* **75** (1994), no. 3, 711-732.

[Sh2] N. Shah, Invariant measures and orbit closures on homogeneous spaces for actions of subgroups generated by unipotent elements, *Lie Groups and Ergodic Theory (Mumbai, 1996)*, 229-271, Tata Inst. Fund. Res., Bombay, 1998.

[Wa] P. Walters, *An introduction to ergodic theory*, Graduate Texts in Math., vol. 79, Springer Verlag, 1982.

[We] H. Weyl, Über die Gleichverteilung von Zahlen mod Eins, *Math. Ann.* **77** (1916), 313-352.