Distribution of values of bounded generalized polynomials

V. Bergelson* and A. Leibman*

September 20, 2006

Abstract

A *generalized polynomial* is a real-valued function which is obtained from conventional 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 $u: \mathbb{Z}^d \longrightarrow \mathbb{R}^l$ 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 $\mathcal{S} \subset \mathbb{R}^l$ (with respect to the Borel measure on $\mathcal{S}$ that is the image of the Lebesgue measure under the piecewise polynomial function defining $\mathcal{S}$). 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 of Furstenberg-Weiss.

Table of Contents

*Supported by NSF grants DMS-0345350 and DMS-06007181. Leibman was also supported by the Sloan foundation.

0. Introduction and formulation of main results

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 conventional polynomials of one or several variables and applying in arbitrary 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 $n \geq 1$,

$$\mathrm{GP}_n = \mathrm{GP}_{n-1} \cup \big\{ v+w : v,w \in \mathrm{GP}_{n-1} \big\} \cup \big\{ vw : v,w \in \mathrm{GP}_{n-1} \big\} \cup \big\{ [v] : v \in \mathrm{GP}_{n-1} \big\}.$$

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*.

In this paper we will mainly deal with GP mappings of integer vector argument, that is, with GP mappings $\mathbb{Z}^d \longrightarrow \mathbb{R}^l$.

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. Note that if one identifies $\mathbb{R}/\mathbb{Z}$ with $[0,1)$, there there is no distinction between $\langle\!\langle p \rangle\!\rangle$ and $p \bmod 1$, so that expressions like $[p_1]^2 \langle\!\langle p_2[p_3]+p_4 \rangle\!\rangle^3 (\bmod 5)$ are generalized polynomials as well.

0.3. Clearly, generalized polynomials form an algebra, and the composition of two generalized polynomial mappings is a generalized polynomial mapping.

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

Theorem. *Given a (conventional) polynomial $p(n)=\sum_{i=0}^{k} a_i n^i$ such that at least one among the coefficients $a_1,\ldots,a_k$ 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 on $[0,1]$. In particular, for any $\varepsilon>0$ there exists $n \in \mathbb{N}$ such that $\langle\!\langle p(n) \rangle\!\rangle < \varepsilon$.*

0.5. 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 on $[0,1]$. They are, however, uniformly distributed on $[0,1]$ with respect to the measure $\frac{dx}{2\sqrt{x}}$.

(2) The sequence $\langle\!\langle -\sqrt{2}n[\sqrt{2}n]\rangle\!\rangle$, $n \in \mathbb{N}$, is dense and uniformly distributed on $[0,1]$ 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 subsection 3.6 below.) On the other hand, one can show that the sequence $\langle\!\langle -\sqrt[3]{2}n[\sqrt[3]{2}n]\rangle\!\rangle$, $n \in \mathbb{N}$, is uniformly distributed on $[0,1]$ with respect to the standard Lebesgue measure. (This fact is a special case of Proposition 5.3 in [Hå2].)

(3) The sequence $\langle\!\langle an\rangle\!\rangle\langle\!\langle bn\rangle\!\rangle$, $n \in \mathbb{N}$, is uniformly distributed on $[0,1]$ 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 on $\left[0,\frac{1}{3}\right] \cup \left[\frac{2}{3},1\right]$ 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, with frequency $1-\langle\!\langle a\rangle\!\rangle$ and $\langle\!\langle a\rangle\!\rangle$ respectively; in other words, $u(n)$ is uniformly distributed on $[0,1]$ with respect to the measure $(1-\langle\!\langle a\rangle\!\rangle)\delta_0 + \langle\!\langle a\rangle\!\rangle\delta_1$. (The generalized polynomial $u(n)$, often called nowadays *Beatty sequence*, appears already in the work of astronomer J. Bernoulli III (see [Mar]), and is found, under different names, in a variety of mathematical contexts, from symbolic dynamics to theory of mathematical games.)

0.6. 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 the values of a generalized polynomial.

0.7. A more general version of Theorem 0.4, 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* $S = \bigcup_{i=1}^{k} S_i$.

0.8. When S consists of several components, that is, when $k \geq 2$, we say that a sequence is uniformly distributed on S if it is uniformly distributed on the components S_i of S with respect to the Haar measures μ_{S_i}, or more precisely, is uniformly distributed on S with respect to a measure $\mu_S = \sum_{i=1}^{k} \alpha_i \mu_{S_i}$, with $\alpha_1,\dots,\alpha_k \in (0,1)$. Here is an example. Let a be an irrational number, and 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: $S_0, S_1, S_1, S_0, S_1, S_1, \ldots$, and is uniformly distributed on $S_0 \cup S_1$ with respect to the probability measure $\mu_S = \frac{1}{3}\mu_{S_0} + \frac{2}{3}\mu_{S_1}$ where μ_{S_i} denotes the normalized Lebesgue measure on S_i, $i = 0, 1$.

0.9. A frequently cited special case of the above theorem concerns the situation where the components of p, the polynomials $p_1, \ldots, p_l$, are rationally independent. In this case the sequence $\{\tilde{p}(n)\}_{n\in\mathbb{N}}$ is uniformly distributed on $\mathbb{T}^l$. 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.

0.10. 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 0.7 above as sections of K by a finite system of parallel planes. One can now rephrase Theorem 0.7 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 want to bring a couple of examples demonstrating some peculiarities of distribution of vector-valued generalized polynomials.

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

(1) The values of the GP 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 $S = \big\{(x, x^2),\ x \in [0,1]\big\}$ in $\mathbb{R}^2$ and uniformly distributed on S with respect to the measure dx.

(2) The values of $u(n) = \big(\langle\!\langle an \rangle\!\rangle, [2\langle\!\langle bn \rangle\!\rangle](2\langle\!\langle an \rangle\!\rangle^2 - 1) - \langle\!\langle an \rangle\!\rangle^2 + 1\big)$, $n \in \mathbb{Z}$, are dense and uniformly distributed with respect to the measure dx on the union of two intersecting parabola segments $\big\{(x, x^2),\ x \in [0,1]\big\}$ and $\big\{(x, 1 - x^2),\ x \in [0,1]\big\}$.

0.11. While the examples in 0.5 and 0.10 indicate that too direct a generalization of Weyl's theorem cannot be hoped for, it turns out that the values of any bounded generalized polynomial $u\colon \mathbb{Z}^d \longrightarrow \mathbb{R}^l$ are uniformly distributed, in a manner to be made precise, on a *piecewise polynomial surface* (see subsection 0.24 below). We will now discuss the ideas behind the proof of this fact. Let us return for a moment to Theorem 0.4. 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 on $[0,1]$, that is, for any interval $[b,c] \subseteq [0,1]$ one has $\frac{1}{N} \cdot \#\{n \le N : a_n \in [b,c]\} \underset{N\to\infty}{\longrightarrow} c - b$;

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

(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 prove the uniform distribution of the sequence $\{\langle\!\langle p(n) \rangle\!\rangle\}_{n\in\mathbb{N}}$, 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 efficiently applied it in his work. See [vdC].)

0.12. A different approach to the proof of Theorem 0.4, which might be called dynamical, deals with a special class of affine maps of a torus. This approach was introduced by Furstenberg in [F1] and [F2] (see also [H] and [C] for a similar treatment), and can be described as follows. Let $p(n) = a_0 + a_1 n + a_2 n^2 \ldots + a_k n^k = b_0 + b_1 n + b_2 \binom{n}{2} + \ldots + b_k \binom{n}{k} \in \mathbb{R}[n]$. Consider the following affine transformation, called *a skew product*, of the k-dimensional torus $\mathbb{T}^k = \mathbb{R}^k/\mathbb{Z}^k$:

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

Let $y = (0, \ldots, 0, b_0) \in \mathbb{T}^k$; one can check by induction on n that $(T^n y)_k = p(n) (\mathrm{mod}\, 1)$, $n \in \mathbb{Z}$. One can now use the known properties of the dynamical system $(\mathbb{T}^k, T)$ in order to characterize the behavior of the sequence $\{\langle\!\langle p(n)\rangle\!\rangle\}_{n\in\mathbb{Z}}$. In particular, if a_k is irrational the system $(\mathbb{T}^k, T)$ is uniquely ergodic (with the unique T-invariant measure being the Lebesgue measure on $\mathbb{T}^k$), which implies (the one-dimensional version of) Weyl's theorem.

0.13. 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 $[0,1]$ *is generated by* a uniquely ergodic dynamical system (X, T, μ) (where X is a compact metric space, T is a homeomorphism $X \longrightarrow X$, and μ is a unique T-invariant measure on X) in the sense that for some Riemann integrable function $f\colon X \longrightarrow \mathbb{R}$ and a point $x \in X$ one has $a_n = f(T^n x)$, then, as a consequence of unique ergodicity, one will have for any function $h \in C(\mathbb{R})$

$$\begin{aligned}\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} h(a_n) &= \lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} h(f(T^n x)) \\ &= \int_X h(f(x))\, d\mu = \int_{\mathbb{R}} h\, d\nu,\end{aligned} \tag{0.2}$$

where $\nu = f_*(\mu)$. Note that, due to the unique ergodicity of T, formula (0.2) holds for the *uniform* Cesàro averages $\frac{1}{N-M}\sum_{n=M}^{N-1} h(a_n)$ (rather than for the more traditional averages $\frac{1}{N}\sum_{n=1}^{N} h(a_n)$); this means that the sequence $\{a_n\}_{n\in\mathbb{Z}}$ is *well distributed* (rather than *uniformly distributed*) with respect to the measure ν on $[0,1]$. (See [F3] and [Wa] for discussion of basic properties of unique ergodicity, and [KN] for more information on well distribution.) The phenomenon of well distribution of sequences generated by a uniquely ergodic measure preserving systems takes place for actions of any amenable group; in this paper we will mainly deal with $\mathbb{Z}^d$-actions.

0.14. The following example shows how a generalized polynomial can be generated by a uniquely ergodic dynamical system. Let $u(n) = \langle\!\langle an[bn]\rangle\!\rangle$, $n \in \mathbb{Z}$, where $a, b \in \mathbb{R}$; we are going to obtain the generalized polynomial u "dynamically". Let G be the group of 4×4 upper triangular matrices with unit diagonal, $G = \left\{ \begin{pmatrix} 1 & a_{1,2} & a_{1,3} & a_{1,4} \\ 0 & 1 & a_{2,3} & a_{2,4} \\ 0 & 0 & 1 & a_{3,4} \\ 0 & 0 & 0 & 1 \end{pmatrix},\ a_{i,j} \in \mathbb{R} \right\}$, 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, $g(g'\Gamma) = (gg')\Gamma$, $g, g' \in G$. 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 $g^n = \begin{pmatrix} 1 & -an & n & 0 \\ 0 & 1 & 0 & bn \\ 0 & 0 & 1 & abn \\ 0 & 0 & 0 & 1 \end{pmatrix}$, $n \in \mathbb{Z}$. Define a transformation T of X by $Tx = gx$, $x \in X$. 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 $T^n x$ "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 $(T^n x)_{1,4}$ of the point $T^n x$ is just $\langle\!\langle an[bn]\rangle\!\rangle$, and we have obtained u dynamically as $u(n) = (T^n x)_{1,4}$, $n \in \mathbb{Z}$. (X, T) is not a uniquely ergodic system, and the sequence $\{T^n x\}_{n\in\mathbb{Z}}$ is not dense in X; let $Y = \overline{\{T^n x\}}_{n\in\mathbb{Z}} \subset X$. One can show that Y is a submanifold of X, and that the action of T on Y is uniquely ergodic. (This can be shown directly, but also follows from the general theory, see [Le] or [L2].) Thus, u is generated by the uniquely ergodic system $(Y, T|_Y)$. This implies that the sequence $\{u(n)\}_{n\in\mathbb{Z}}$ is well distributed with respect to a certain Borel measure ν on $[0, 1]$. (Namely, $\nu = (x_{1,4})_*(\mu_Y)$, where μ_Y is the unique T-invariant measure on Y.)

0.15. In the example above, the group G of upper triangular matrices with unit diagonal 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.

0.16. Let us note that the skew product transformation (0.1) of the torus $\mathbb{T}^k$, which was utilized in subsection 0.12 to generate the generalized polynomial $\langle\!\langle p\rangle\!\rangle=\left\langle\!\!\left\langle b_0+b_1x+b_2\binom{x}{2}+\ldots+b_k\binom{x}{k}\right\rangle\!\!\right\rangle$, can also be viewed as a nilrotation. Indeed, let G be the group of upper triangular matrices with unit diagonal $\begin{pmatrix}1 & a_{1,2} & a_{1,3} & \ldots & a_{1,k+1}\\ & 1 & a_{2,3} & \ldots & a_{2,k+1}\\ & & \ddots & & \vdots\\ & \mathbf{O} & & 1 & a_{k,k+1}\\ & & & & 1\end{pmatrix}$ with $a_{i,j}\in\mathbb{Z}$ for $1\leq i<j\leq k$ and $a_{i,k+1}\in\mathbb{R}$ 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 formula (0.1).

0.17. 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 $\inf_{g\in G}\operatorname{dist}(gx,gy)$ is positive.) 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]).

0.18. We are now going to formulate a theorem that establishes a connection between bounded generalized polynomials and nilsystems. But first we need to introduce the notion of a piecewise polynomial function on a nilmanifold. Given a connected nilmanifold X, one can define a bijective *coordinate mapping* $\tau\colon X\longrightarrow[0,1)^k$ (see the formal definition in subsection 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 mapping $h\colon B\longrightarrow\mathbb{R}^l$ from a set $B\subseteq\mathbb{R}^k$ is *piecewise polynomial* if there is a partition $B=\mathcal{L}_1\cup\ldots\cup\mathcal{L}_r$ and polynomial mappings $P_1,\ldots,P_r\colon\mathbb{R}^k\longrightarrow\mathbb{R}^l$ such that each $\mathcal{L}_j$ is defined by a system of polynomial inequalities and $h|_{\mathcal{L}_j}=P_j$, $j=1,\ldots,r$. We say that a mapping $f\colon X\longrightarrow\mathbb{R}^l$ is piecewise polynomial if the mapping $f\circ\tau^{-1}\colon[0,1)^k\longrightarrow\mathbb{R}^l$ is piecewise polynomial. This definition does not depend on the choice of a coordinate system on X (see [L4]). We say that a mapping of a non-connected nilmanifold X is piecewise polynomial if it is piecewise polynomial on every connected component of X. A piecewise polynomial mapping may be discontinuous, but it is clearly Riemann integrable. (A function on a compact metric space X equipped with a finite measure is Riemann integrable iff it is bounded and continuous almost everywhere in X.)

0.19. Theorem A. (i) *For any nilmanifold X, any action φ of $\mathbb{Z}^d$ by nilrotations on X, any piecewise polynomial mapping $f\colon X\longrightarrow\mathbb{R}^l$, and any point $x\in X$, the mapping*

$u(n)=f(\varphi(n)x)$, $n\in\mathbb{Z}^d$, is a GP mapping.
(ii) *For any bounded GP mapping $u\colon\mathbb{Z}^d\longrightarrow\mathbb{R}^l$ there exists a nilmanifold X, an ergodic action φ of $\mathbb{Z}^d$ by nilrotations on X, a piecewise polynomial mapping $f\colon X\longrightarrow\mathbb{R}^l$, and a point $x\in X$ such that $u(n)=f(\varphi(n)x)$, $n\in\mathbb{Z}^d$.*

In other words, any mapping that is generated by a nilsystem and a piecewise polynomial mapping is a (bounded) GP mapping, and any bounded GP mapping is generated by an ergodic nilsystem and a piecewise polynomial mapping.

0.20. Remarks. (1) It is important to emphasize that the piecewise polynomial mapping f appearing in the formulation of Theorem A may be discontinuous, and that this (rather mild) discontinuity of f is inevitable: not every bounded generalized polynomial is of the form $u(n)=f(T^nx)$ where T is a nilrotation, $x\in X$, and $f\in C(X)$. Moreover, not every bounded generalized polynomial can be represented as $u(n)=f(T^nx)$ where T is a (continuous) 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, for example, [F3], p. 160), and thus the sequence $f(T^nx)$ with $f\in C(X)$ cannot have nonrecurrent values, whereas some generalized polynomials may (see examples in subsection 3.4). The same argument shows that not every bounded generalized polynomial is representable as $f(T^nx)$ where T is a continuous uniquely ergodic transformation of a compact space X, $f\in C(X)$, and the unique T-invariant measure μ on X is such that $\mathrm{supp}(\mu)=X$.[1] (It is not hard to show that under these conditions the system (X,T) is minimal; see, for example, [Wa], Theorem 6.17.) Finally, not all bounded generalized polynomials without isolated values are representable as $f(T^nx)$ where T is distal and f is continuous; the simplest example of such a polynomial is $u(n)=\langle\!\langle [an]b\rangle\!\rangle$ (see [Hå1]).

(2) Also, not all bounded generalized polynomials can be obtained by using a skew product transformation of a torus (like in the example discussed in 0.12 above), and a Riemann integrable (not necessarily continuous) function thereon. 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 T 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 $u(n)=f(T^nx)$, $n\in\mathbb{Z}$, and let $\tilde{f}=e^{2\pi if}$. Then $\tilde{f}(T^nx)=e^{2\pi ian[bn]}$, $n\in\mathbb{Z}$. One can show that for any character χ on X one has $\chi(T^nx)=e^{2\pi ip(n)}$, where p is a polynomial. Using the method described in subsection 3.6 below one can check that for any ordinary polynomial p the sequence $\langle\!\langle an[bn]-p(n)\rangle\!\rangle$, $n\in\mathbb{N}$, is uniformly distributed on $[0,1]$. Hence, $\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^N e^{2\pi i(an[bn]-p(n))}=0$. Since T is uniquely ergodic (this follows from Proposition 3.10 in [F3]), the sequence $T^n(x)$ is uniformly distributed on X, and so

$$\int_X\tilde{f}\cdot\overline{\chi}\,d\mu=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^N\tilde{f}(T^nx)\cdot\overline{\chi(T^nx)}\,d\mu=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^N e^{2\pi ian[bn]}e^{-2\pi ip(n)}=0.$$

[1] On the other hand, it follows from Theorem 4.2.2 in [Hå1] that every bounded generalized polynomial can be obtained with the help of a uniquely ergodic system if the condition $\mathrm{supp}(\mu)=X$ is dropped.

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

0.21. In order to formulate corollaries of Theorem A we need to introduce some terminology. *A Følner sequence* in $\mathbb{Z}^d$ is a sequence $\{\Phi_N\}_{N=1}^{\infty}$ 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 (not necessarily nested) 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 $\mathbb{Z}^d$. When saying 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.

0.22. Let ω be a mapping of $\mathbb{Z}^d$ to a compact metric space X endowed with a finite nonzero Borel measure μ. We will say that the (multiparameter) sequence $\{\omega(n)\}_{n\in\mathbb{Z}^d}$ *is well distributed on* X with respect to μ if for any open set $U \subseteq X$ with $\mu(\partial U) = 0$ one has $\mathcal{D}(\omega^{-1}(U)) = \mu(U)/\mu(X)$. When 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 $\lim\limits_{N\to\infty} \frac{1}{|\Phi_N|}\sum_{n\in\Phi_N} f(\omega(n)) = \int_X f\,d\mu$.

0.23. Let a set $\mathcal{L} \subseteq \mathbb{R}^s$, with nonempty interior, be defined by a system of polynomial inequalities, and let $\mathcal{P}$ be a polynomial mapping $\mathbb{R}^s \longrightarrow \mathbb{R}^l$. We will call $\mathcal{S} = \mathcal{P}(\mathcal{L})$ a (parameterized) *polynomial surface in* $\mathbb{R}^l$. Let λ be the Lebesgue measure on $\mathbb{R}^s$; we will denote by $\mu_{\mathcal{S}}$ the normalized measure $\mathcal{P}_*(\lambda)$ on $\mathcal{S}$, which is defined by $\mu_{\mathcal{S}}(A) = \lambda(\mathcal{P}^{-1}(A)\cap\mathcal{L})/\lambda(\mathcal{L})$ for Borel sets A in $\mathbb{R}^l$. A *piecewise polynomial surface* $\mathcal{S}$ is a finite (not necessarily disjoint) union of polynomial surfaces, $\mathcal{S} = \bigcup_{i=1}^k \mathcal{S}_i$, endowed with a measure $\mu_{\mathcal{S}}$ of the form $\mu_{\mathcal{S}} = \sum_{i=1}^k \alpha_i\mu_{\mathcal{S}_i}$ for some $\alpha_1,\ldots,\alpha_k > 0$.

0.24. We are now in position to formulate a corollary of Theorem A pertaining to well distribution of bounded generalized polynomials. In order to keep the technicalities to the minimum, we give here a somewhat simplified version of a more comprehensive theorem to be found in subsection 3.1 below.

Theorem B. *Let $u\colon \mathbb{Z}^d \longrightarrow \mathbb{R}^l$ be a bounded GP mapping. There exists a bounded piecewise polynomial surface $\mathcal{S}$ such that $u(n) \in \mathcal{S}$ for almost all $n \in \mathbb{Z}^d$ and the sequence $\{u(n)\}_{n\in\mathbb{Z}^d}$ is well distributed on $\mathcal{S}$ with respect to $\mu_{\mathcal{S}}$.*

0.25. In particular, we have

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 is equal to $\int f\,d\mu_{\mathcal{S}}$.*

0.26. The following special case of Corollary 0.25 gives the affirmative answer to the question formulated in 0.6:

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.*

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

0.27. From Corollary 0.26 one can deduce, with the help of the spectral theorem, the following two generalizations of the classical von Neumann's ergodic theorem. (For proofs see subsections 4.1 and 4.2 below.)

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.*

0.28. 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.*

0.29. We will now formulate one more corollary of Theorem B, which deals with the existence of *invariant means* (also called *Banach limits*) on the algebra $\mathfrak{B}$ of bounded generalized polynomials $\mathbb{Z}^d\longrightarrow\mathbb{R}$. It follows from Theorem B that for any $u\in\mathfrak{B}$ the number $L(u)=\lim\limits_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{n\in\Phi_N}f(u(n))$ does not depend on the choice of a Følner sequence $\{\Phi_N\}$. This fact implies that all Banach limits agree on u (see [Lo] or [Su].) Consequently, we have the following result.

Proposition. *There exists a unique invariant mean on the algebra $\mathfrak{B}$ of bounded generalized polynomials $\mathbb{Z}^d\longrightarrow\mathbb{R}$. In other words, there exists a unique linear functional $L\colon\mathfrak{B}\longrightarrow\mathbb{R}$ having the following properties:*
(i) *for any $m\in\mathbb{Z}^d$, $L(u_m)=L(u)$, where $u_m(n)=u(n+m)$, $n\in\mathbb{Z}^d$;*
(ii) *$L(u)\geq 0$ if $u\geq 0$;*
(iii) *$L(1)=1$.*

Let us also remark that the analogous fact holds for the algebra generated by functions of the form $f\circ u$, where u is a bounded generalized polynomial $\mathbb{Z}^d\longrightarrow\mathbb{R}^l$ and $f\in C(\mathbb{R}^l)$.

0.30. While Theorem B utilizes the unique ergodicity of (ergodic) nilrotations, the fact that nilrotations are also distal provides additional information about the character of distribution of GP mappings on piecewise polynomials surfaces. Given an infinite sequence $E=\{n_1,n_2,\ldots\}$ (of not necessarily distinct elements) in $\mathbb{Z}^d$, let $\mathrm{FS}(E)$ denote the set of finite sums of distinct elements of E: $\mathrm{FS}(E)=\big\{\sum_{i\in F}n_i : F\subset\mathbb{N},\ 0<|F|<\infty\big\}$. Sets of the form $\mathrm{FS}(E)$ are called in ergodic theory *IP sets* and are intrinsically connected with recurrence properties of distal systems (see [F3] and [B]). A set $P\subseteq\mathbb{Z}^d$ is called *an IP* set* if it has a nontrivial intersection with any IP set in $\mathbb{Z}^d$. One can show that any IP* set P is syndetic, that is, has the property that the union of finitely many shifts of P covers $\mathbb{Z}^d$. In fact, the property of IP*-ness is quite a bit stronger than that of syndeticity. For instance, while the intersection of two syndetic sets may be empty, the intersection of any finite family of IP* sets is again an IP* set. (See [F3], Lemma 9.5.) A set Q is called *IP$^*_+$* if it is a "shifted" IP* set, that is, is of the form $n+P$ where P is an IP* set. While IP$^*_+$ sets do not have the filter property (the intersection of two IP$^*_+$ sets may be empty), they still have some "regularity" properties and form a smaller class than that of general syndetic sets. (See [B] for examples of syndetic sets which are not IP^*_+.) The relevance of IP^* and IP^*_+ sets to the distal systems is revealed by the following theorem:

Theorem. (Cf. [B], Theorems 3.8, 3.9.) *Let φ be a $\mathbb{Z}^d$-action by homeomorphisms of a compact metric space X. The action φ is distal iff for any $x \in X$ and any open neighborhood W of x the set $\{n : \varphi(n)x \in W\}$ is an IP* set. If the system (X, φ) is minimal (that is, the orbit $\{\varphi(n)x\}_{n\in\mathbb{Z}^d}$ of every point $x \in X$ is dense in X), then the action of φ is distal iff for any $x \in X$ and any open $W \subseteq X$ the set $\{n : \varphi(n)x \in W\}$ is IP$^*_+$.*

0.31. Let u be a bounded GP mapping and let $\mathcal{S}$ be the piecewise polynomial surface on which the values of u are well distributed. It follows from Theorem B that for any nonempty open set $W \subseteq \mathcal{S}$ the set $u^{-1}(W) = \{n \in \mathbb{Z}^d : u(n) \in W\}$ is syndetic. From the distality of nilsystems we will deduce the following enhancement of this fact:

Theorem C. *For any nonempty open set $W \subseteq \mathcal{S}$, $u^{-1}(W)$ is an IP$^*_+$ set.*

0.32. Let us say that a value $u(n) \in \mathbb{R}^l$ of u is *IP*-recurrent* if for any neighborhood W of $u(n)$ the set $u^{-1}(W)$ is an IP^* set, and is *IP$^*_+$-recurrent* if for any neighborhood W of $u(n)$ the set $u^{-1}(W)$ is an IP^*_+ set. It now follows from Theorems B and C that almost all values of u are IP^*_+-recurrent. (Or, more precisely, $u(n)$ is an IP^*_+-recurrent value of u for almost all $n \in \mathbb{Z}^d$.)

0.33. For a given polynomial mapping u, Theorem C gives no information about whether a concrete value of u is recurrent. 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 occurring in the representation of 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 with the natural projection $\mathbb{R}^l \longrightarrow \mathbb{T}^l$. Then $0 \in \mathbb{T}^l$ is an IP*-recurrent value of $\tilde{u}$. (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.)*

(The expression "polynomials occurring in the representation of u" refers to the polynomials occurring in the representation of the coordinates of u; for example, polynomials occurring in the representation of $u = \big([2[p_1] + p_2]p_3, [p_4][p_5], 3[p_6][p_7]\big)$ are $p_1, \dots, p_7, 2, 3$. Below we will use the term "polynomials *involved* in u"; see a formal definition in 2.9.)

0.34. We will now briefly discuss an interesting Diophantine application of Theorem D. The following theorem was obtained in [vdC]:

Theorem. *Let $u_i\colon \mathbb{Z}^{d+i-1} \longrightarrow \mathbb{R}$, $i = 1, \dots, k$, be polynomials without constant term. For any $\delta > 0$, the set of $n \in \mathbb{Z}^d$ for which there exist $m_1, \dots, m_k \in \mathbb{Z}$ such that*

$$\big|u_1(n) - m_1\big| < \delta,\ \big|u_2(n, m_1) - m_2\big| < \delta,\ \dots,\ \big|u_k(n, m_1, \dots, m_{k-1}) - m_k\big| < \delta \qquad (0.3)$$

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

Furstenberg and Weiss proved in [FW] that the set of $n \in \mathbb{Z}^d$ for which the system (0.3) has a solution is IP^*. This fact was further enhanced and generalized in [BHåM]. We will derive from Theorem D yet another generalization of Furstenberg-Weiss' theorem:

Theorem. *Let $u_i\colon \mathbb{Z}^{d+i-1} \longrightarrow \mathbb{R}$, $i = 1,\dots,k$, be generalized polynomials such that all ordinary polynomials occurring in the representation of u_i have zero constant term. For any $\delta > 0$, the set of $n \in \mathbb{Z}^d$ for which there exist $m_1,\dots,m_k \in \mathbb{Z}$ satisfying (0.3) is an IP* set.*

0.35. In conclusion, we would like to say a few words about bounded generalized polynomials of continuous argument. We do believe that all the results above extend to this case. We, however, cannot prove this here because of the absence in the literature of the continuous version of Theorem 2.3, which is an essential ingredient in our proofs. A version of Theorem 2.3 where the well distribution is replaced by the uniform distribution follows from the results in [Sh1]; this allows one to obtain a continuous version of Theorem B, which we will presently formulate. For a measurable set $E \subseteq \mathbb{R}^d$ let us write $\mathcal{D}_B(E) = \alpha$ if $\lim\limits_{r\to\infty} \lambda(E\cap B_r)/\lambda(B_r) = \alpha$, 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 $\omega\colon \mathbb{R}^d \longrightarrow X$ is a mapping to a topological space X equipped with a nonzero finite Borel measure μ, let us say that ω is *ball-uniformly distributed on X* if for any open set U in X with $\mu(\partial U) = 0$ one has $\mathcal{D}_B(\omega^{-1}(U)) = \mu(U)/\mu(X)$.

Theorem B$_c$. *Let $u\colon \mathbb{R}^d \longrightarrow \mathbb{R}^l$ be a bounded GP mapping. There exists a bounded piecewise polynomial surface $\mathcal{S}$ such that u is ball-uniformly distributed on $\mathcal{S}$ with respect to $\mu_{\mathcal{S}}$.*

0.36. The goal of this subsection is to help the reader to navigate through this – sometimes quite entangled – paper. In the course of proving Theorem A and in order to derive its corollaries we will formulate various modifications of Theorems A, B, B$_c$, etc. The following diagram describes logical connections between the major theorems and indicates the subsections where they appear:

$$
\begin{array}{ccccccccc}
 & & & & \mathrm{A}(0.19) & \Rightarrow & \mathrm{B}(0.24) & & \\
 & & & & \uparrow & & \uparrow & & \\
 & & & & \mathrm{A}^*(1.7) & & \mathrm{B}^*(3.1) & & \\
 & & & & \uparrow & & \uparrow & & \\
\mathrm{A}^{***}(10.4) & \rightarrow & \mathrm{A}_1^{**}(1.14) & \rightleftarrows & \mathrm{A}^{**}(1.17) & \Rightarrow & \mathrm{B}^{**}(2.12) & \Rightarrow & \mathrm{C}(0.31) \\
 & & & & \Downarrow & \searrow & & & \\
 & & & & \mathrm{D}(0.33) & & \mathrm{B}_c(0.35) & \rightarrow & \mathrm{B}_c^*(4.11)
\end{array}
$$

where "$P \rightarrow Q$" means that Q is a special case of P and "$P \Rightarrow Q$" means that Q is derivable from P.

Here is a brief description of the structure of the paper.

In Section 1 we introduce coordinates on a nilmanifold and present another version of Theorem A, Theorem A*, which says that any GP mapping is generated with the help of a coordinate mapping of a connected nilmanifold and a sequence of *polynomial* transformations thereof. We then formulate an extension of Theorem A*, Theorem A**, which (i) deals with families of functions more general than that of generalized polynomials, and (ii) ties *the complexity* of a GP mapping with the nilpotency class of the nilsystem that generates it. The (long and difficult) proof of Theorem A** is self-contained, and we postpone it until the last sections of the paper, first focusing on applications of this theorem.

In Section 2 we describe how subnilmanifolds look in coordinates on a nilmanifold, and use this information to derive from Theorem A a technical version of Theorem B, Theorem B**. Theorem B** is used in Section 3 to obtain Theorem B*, a refinement of Theorem B that contains some additional information about the distribution of the values of a bounded GP mapping u on a piecewise polynomial surface $\mathcal{S}$. In particular, it connects the degrees and the coefficients of the polynomials that define $\mathcal{S}$ with the complexity of u and, respectively, with the constant terms of the polynomials occurring in the representation of u. We then discuss *exceptional values* of GP mappings, and provide an instructive example of computation of the distribution of the values of a generalized polynomial.

In Section 4 we derive the rest of the results formulated in the introduction; in particular, we prove Theorems C and D.

Sections 5–10 are devoted to the proof of Theorem A**; in this proof we use the nilpotent group of upper triangular matrices with unit diagonal. In Sections 5 and 6 we reduce the problem to an algebraic one, namely, to proving that any generalized polynomial can be produced by applying special algebraic operations to entries of an appropriately chosen upper triangular matrix. (The general algebraic version of Theorem A, Theorem A***, is formulated in Section 10.)

In Sections 7–9 we deal with *elementary* generalized polynomials (the generalized polynomials produced from the conventional polynomials by using only multiplication and the bracket operation (and no addition or subtraction)). The structure of an elementary generalized polynomial can be described by *a tree* (an oriented cycle-free graph), and we use a rather cumbersome induction over the set of trees to show that any elementary generalized polynomial can be "read off", modulo "smaller" elementary generalized polynomials, from an upper triangular matrix.

In Section 10 we conclude the proof of Theorem A***, passing from elementary to arbitrary generalized polynomials.

Acknowledgments. We thank H. Furstenberg and the anonymous referee for useful comments and suggestions.

1. Coordinates on a nilmanifold and a reformulation of Theorem A

1.1. Let G be a nilpotent Lie group of nilpotency class D and let Γ be a discrete uniform subgroup of G. The compact homogeneous space $X = G/\Gamma$ is called *a nilmanifold of nilpotency class* D. We will assume that G is connected and simply-connected, which will suffice for our goals.

1.2. We will list here some facts about connected simply-connected nilpotent Lie groups; for more details see [Mal].

For any $g \in G$ there exists a unique one-parameter subgroup $\{g^t\}_{t\in\mathbb{R}}$ in G such that $g^1 = g$. Let $G = G_1 \supseteq G_2 \supseteq \ldots \supseteq G_{\mathrm{D}} \supseteq G_{\mathrm{D}+1} = \{\mathbf{1}_G\}$ be the lower central series of

G; then, for each j, G_j/G_{j+1} is a finite dimensional $\mathbb{R}$-vector space. G has *a Malcev basis* compatible with Γ, that is, an ordered set $\{e_1,\ldots,e_k\}\subset\Gamma$ such that, for a certain sequence $1=k_1<\ldots<k_D$ of positive integers, the (images of) the elements $e_{k_j},\ldots,e_{k_{j+1}-1}$ form a basis in G_j/G_{j+1} for every $j=1,\ldots,D$. If $\{e_1,\ldots,e_k\}$ is a Malcev basis in G, every element $g\in G$ is uniquely representable in the form $g=e_1^{a_1}\ldots e_k^{a_k}$ where *the coordinates* $a_1,\ldots,a_k$ are real numbers, with $g\in\Gamma$ iff $a_1,\ldots,a_k\in\mathbb{Z}$.

For every $i\in\{1,\ldots,k\}$ let $D_i\in\mathbb{N}$ be such that $e_i\in G_{D_i}\setminus G_{D_i+1}$.

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

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

and

$$g^t=e_1^{a_1t}\ldots e_i^{a_it+q_i(a_1,\ldots,a_{i-1},t)}\ldots e_k^{a_kt+q_k(a_1,\ldots,a_{k-1},t)},\ t\in\mathbb{R},\tag{1.2}$$

where, for each $i=2,\ldots,k$, 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 $i=2,\ldots,k$ one has $\deg p_i,\deg q_i\leq D_i$. Moreover, $\deg p_i\bigl(a_1^{D_1},\ldots,a_k^{D_k},b_1^{D_1},\ldots,b_k^{D_k}\bigr)\leq D_i$. It follows that if $S_1,\ldots,S_k,R_1,\ldots,R_k$ are polynomials with $\deg S_i$, $\deg R_i\leq D_i$ for all $i=1,\ldots,k$, then $\deg p_i\bigl(S_1,\ldots,S_k,R_1,\ldots,R_k\bigr)\leq D_i$, $i=1,\ldots,k$. (See [L1]; in the terminology of [L1] the multiplication in G is a continuous polynomial mapping of lc-degree $\leq(1,2,\ldots,D)$.)

1.5. *The coordinate mapping* $\tilde{\tau}\colon G\longrightarrow\mathbb{R}^k$, $g=e_1^{a_1}\ldots e_k^{a_k}\mapsto(a_1,\ldots,a_k)$, *is a diffeomorphism satisfying* $\tilde{\tau}(\Gamma)=\mathbb{Z}^k$. "The cube" $Q=\tilde{\tau}^{-1}\bigl([0,1)^k\bigr)\subset G$ 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 $g\gamma_{i-1}=e_1^{x_1}\ldots e_{i-1}^{x_{i-1}}e_i^{b_i}\ldots e_k^{b_k}$ with $x_1,\ldots,x_{i-1}\in[0,1)$, put $\gamma_i=\gamma_{i-1}e_i^{-[b_i]}$. Then $g\gamma_i=g\gamma_{i-1}e_i^{-[b_i]}=e_1^{x_1}\ldots e_{i-1}^{x_{i-1}}e_i^{x_i}e_{i+1}^{c_{i+1}}\ldots e_k^{c_k}$ with $x_i=b_i-[b_i]\in[0,1)$. For $\gamma=\gamma_k$ we therefore have $g\gamma=e_1^{x_1}\ldots e_k^{x_k}$ with $x_1,\ldots,x_k\in[0,1)$.

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

1.6. Let us note the following fact:

Lemma. *Any piecewise polynomial mapping* $h\colon B\longrightarrow\mathbb{R}^l$ *of a bounded subset* $B\subset\mathbb{R}^k$ *is the restriction on* B *of a GP mapping* $w\colon\mathbb{R}^k\longrightarrow\mathbb{R}^l$.

Proof. Recall that a mapping $h\colon B \longrightarrow \mathbb{R}^l$ is said to be piecewise polynomial if B is partitioned, $B = \mathcal{L}_1 \cup \ldots \cup \mathcal{L}_r$, so that for each $j = 1, \ldots, r$, the set $\mathcal{L}_j$ is defined by a system of polynomial inequalities,

$$\mathcal{L}_j = \bigl\{t \in B : p_{j,1}(t), \ldots, p_{j,n_j}(t) > 0,\ q_{j,1}(t), \ldots, q_{j,m_j}(t) \geq 0\bigr\},$$

and $P_j = h|_{\mathcal{L}_j}$ is a polynomial mapping. Let M be such that $|p_{j,i}(t)|, |q_{j,i}(t)| < M$ for all j, i and all $t \in B$. For any number c with $|c| < M$ one has $-[-c/M] = \begin{cases} 1, & c > 0 \\ 0, & c \leq 0 \end{cases}$ and $1 + [c/M] = \begin{cases} 1, & c \geq 0 \\ 0, & c < 0 \end{cases}$. Thus, if we define a GP mapping w as

$$w = \sum_{j=1}^{r} \Bigl(\prod_{i=1}^{n_j} \bigl(-[-p_{j,i}/M]\bigr)\Bigr)\Bigl(\prod_{i=1}^{m_j} \bigl(1 + [q_{j,i}/M]\bigr)\Bigr) P_j,$$

then, for $t \in B$, $w(t) = P_j$ iff $t \in \mathcal{L}_j$. ∎

It follows that the composition of a bounded GP mapping with a piecewise polynomial mapping is a GP mapping.

1.7. We will now formulate a modification of Theorem A, which, on one hand, is more natural, and on the other hand, will be easier for us to prove. The idea is to obtain a generalized polynomial as 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 $\omega\colon \mathbb{Z}^d \longrightarrow G$ to a nilpotent Lie 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.

Theorem A*. *A mapping $u\colon \mathbb{Z}^d \longrightarrow [0,1)^l$ is a GP mapping iff there exist a connected nilmanifold $X = G/\Gamma$ equipped with a coordinate system $\tau = (\tau_1, \ldots, \tau_k)$, a polynomial mapping $\omega\colon \mathbb{Z}^d \longrightarrow G$, and indices $i_1, \ldots, i_l \in \{1, \ldots, k\}$ such that $u = (\tau_{i_1}, \ldots, \tau_{i_l})\circ\omega$.*

1.8. We will now explain how Theorem A can be derived of Theorem A*. To prove (i), assume that $u(n) = f(\varphi(n)x)$, $n \in \mathbb{Z}^d$, where $f\colon Y \longrightarrow \mathbb{R}^l$ is a piecewise polynomial mapping of a nilmanifold $Y = H/\Lambda$, φ is a homomorphism $\mathbb{Z}^d \longrightarrow H$, and $x \in Y$. Let $\pi\colon H \longrightarrow Y$ be the natural projection and let $g \in H$ be such that $\pi(g) = x$; then the mapping $\omega\colon \mathbb{Z}^d \longrightarrow H$, $\omega(n) = \varphi(n)g$, $n \in \mathbb{Z}^d$, is polynomial. The function f is the composition $f = h\circ\tau$ where τ is a coordinate function on Y and h is a piecewise polynomial function on $\mathbb{R}^k$. By (the if part of) Theorem A*, $v(n) = \tau(\varphi(n)x) = \tau(\omega(n))$ is a GP mapping. Thus, by Lemma 1.6, $u = h\circ v$ is a GP mapping.

To prove (ii), assume that a GP mapping u is represented in the form $u(n) = (\tau_{i_1}, \ldots, \tau_{i_l})(\omega(n))$, $n \in \mathbb{Z}^d$, as in Theorem A*. Let $\pi\colon G \longrightarrow X$ be the natural projection. It is shown in [L3] that one can find another connected nilmanifold $\widetilde{X} = \widetilde{G}/\widetilde{\Gamma}$ with a continuous mapping $\eta\colon \widetilde{X} \longrightarrow X$, a homomorphism $\varphi\colon \mathbb{Z}^d \longrightarrow \widetilde{G}$, and a point $\tilde{x} \in \widetilde{X}$ such that $\pi(\omega(n)) = \eta(\varphi(n)\tilde{x})$ for all $n \in \mathbb{Z}^d$. It is also shown in [L3] (and, as well, follows from the results in [Le] or [Sh2]) that the closure $Y = \overline{\varphi(\mathbb{Z}^d)\tilde{x}}$ of the orbit of

$\tilde{x}$ under the action of φ is a (not necessarily connected) subnilmanifold of $\widetilde{X}$. Hence, $u(n)=(\tau_{i_1},\ldots,\tau_{i_l})(\eta(\varphi(n)\tilde{x}))$, $n\in\mathbb{Z}^d$. Moreover, the action φ is ergodic on Y, the mapping $(\tau_{i_1},\ldots,\tau_{i_l})\circ\eta$ is piecewise polynomial on $\widetilde{X}$, and hence $f=(\tau_{i_1},\ldots,\tau_{i_l})\circ\eta|_Y$ is a piecewise polynomial mapping from Y (see [L4]).

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

When $\mathcal{A}$ is the ring of polynomials $\mathbb{Z}^d\longrightarrow\mathbb{R}$, the $\mathcal{A}$-mappings to a nilpotent Lie group G are just polynomial mappings.

1.10. For $D\in\mathbb{N}$ we define

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

1.11. 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 precisely, $v\in\mathfrak{B}(\mathcal{A})$ if

$v\in\mathcal{A}$,
or $v=v_1+v_2$ where $v_1,v_2\in\mathfrak{B}(\mathcal{A})$,
or $v=v_1v_2$ where $v_1,v_2\in\mathfrak{B}(\mathcal{A})$,
or $v=\pm[w]$ where $w\in\mathfrak{B}(\mathcal{A})$.[2]

We define $\mathfrak{B}^{\mathbf{o}}(\mathcal{A})=\big\{u\in\mathfrak{B}(\mathcal{A}):\mathrm{Ran}(u)\in[0,1)\big\}=\big\{u-[u]:u\in\mathfrak{B}(\mathcal{A})\big\}$.

1.12. *The complexity*, $\mathrm{cmp}(v)$ of $v\in\mathfrak{B}(\mathcal{A})$ is defined by

$\mathrm{cmp}(v)=1$ if $v\in\mathcal{A}$;
$\mathrm{cmp}(v)=\max\{\mathrm{cmp}(v_1),\mathrm{cmp}(v_2)\}$ if $v=v_1+v_2$;
$\mathrm{cmp}(v)=\mathrm{cmp}(v_1)+\mathrm{cmp}(v_2)$ if $v=v_1v_2$;
$\mathrm{cmp}(v)=\mathrm{cmp}(w)$ if $v=\pm[w]$.

(Note that $\mathrm{cmp}(v)$ is not uniquely defined and depends on the representation of v in terms of elements of $\mathcal{A}$. This will not affect our arguments, since we will deal with *concrete representations* of generalized polynomial rather than with polynomials themselves. We refer the reader to Section 6, where a formalism for dealing with representations of generalized polynomials is introduced.)

[2] Here is the clarification of how this definition should be understood. Put $\mathbf{B}_0(\mathcal{A})=\mathcal{A}$; then put $\mathbf{B}_k(\mathcal{A})=\mathbf{B}_{k-1}(\mathcal{A})\cup\big\{v_1+v_2,\ v_1,v_2\in\mathbf{B}_{k-1}(\mathcal{A})\big\}\cup\big\{v_1v_2,\ v_1,v_2\in\mathbf{B}_{k-1}(\mathcal{A})\big\}\cup\big\{\pm[v],\ v\in\mathbf{B}_{k-1}(\mathcal{A})\big\}$ for $k=1,2,\ldots$, and finally, let $\mathfrak{B}(\mathcal{A})=\bigcup_{k=0}^{\infty}\mathbf{B}_k(\mathcal{A})$.

Examples. If $p_i \in \mathcal{A}$, then
$\mathrm{cmp}(p_1) = 1$, $\mathrm{cmp}([p_1]) = 1$, $\mathrm{cmp}\big(p_1[p_2]\big) = 2$, $\mathrm{cmp}\big(p_1[p_2]+p_3\big) = 2$, $\mathrm{cmp}\big(p_1[p_2[p_3]]\big) = 3$, $\mathrm{cmp}\big(p_1[p_2][p_3]\big) = 3$, $\mathrm{cmp}\big(p_1[p_2[p_3]+p_4]+p_5[p_6]\big) = 3$, $\mathrm{cmp}\big(p_1[p_2[p_3]+p_4][p_5]+p_6\big) = 4$.

When $v = (v_1, \ldots, v_l)$ is a GP mapping, we define $\mathrm{cmp}(v) = \max\{\mathrm{cmp}(v_i)\}_{i=1}^{l}$.

1.13. From the definition of complexity we immediately have:

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

1.14. For $D \in \mathbb{N}$ we define $\mathfrak{B}_D(\mathcal{A}) = \big\{v \in \mathfrak{B}(\mathcal{A}) : \mathrm{cmp}(v) \leq D\big\}$ and $\mathfrak{B}^{\mathbf{o}}_D(\mathcal{A}) = \mathfrak{B}_D(\mathcal{A}) \cap \mathfrak{B}^{\mathbf{o}}(\mathcal{A})$. [3]

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

The inclusion $\mathfrak{B}^{\mathbf{o}}_D(\mathcal{A}) \subseteq \mathfrak{N}_D(\mathcal{A})$ of this theorem will be proved in Sections 5 – 10.

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

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

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

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

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 $\tau_i \circ \omega = \tilde{\tau}_i \circ \chi(\omega) = \xi_i$, $i = 1, \ldots, k$. We have to show that $\xi_1, \ldots, \xi_k \in \mathfrak{B}^{\mathbf{o}}_D(\mathcal{A})$.

Assume by induction on i that in the formula (1.3), $\xi_1, \ldots, \xi_{i-1}, \beta_i, \ldots, \beta_k \in \mathfrak{B}(\mathcal{A})$, and, in the notation of subsection 1.2, $\mathrm{cmp}(\xi_j) \leq D_j$, $j = 1, \ldots, i-1$, and $\mathrm{cmp}(\beta_j) \leq D_j$, $j = i, \ldots, k$. Then $\xi_i = \beta_i - [\beta_i] \in \mathfrak{B}^{\mathbf{o}}(\mathcal{A})$, and $\mathrm{cmp}(\xi_i) = \mathrm{cmp}(\beta_i) \leq D_i \leq D$, so $\xi_i \in \mathfrak{B}^{\mathbf{o}}_D(\mathcal{A})$. By 1.3, the functions $\zeta_{i+1}, \ldots, \zeta_k$ in formula (1.4) are given by polynomial expressions in $\xi_1, \ldots, \xi_{i-1}, \beta_i, \ldots, \beta_k$ and $[\beta_i]$, hence $\zeta_{i+1}, \ldots, \zeta_k \in \mathfrak{B}(\mathcal{A})$, and by 1.4 and Lemma 1.13, $\mathrm{cmp}(\zeta_j) \leq D_j$, $j = i+1, \ldots, k$. ■

[3] Let us clarify this definition as well: $\mathfrak{B}_D(\mathcal{A})$ consists of the elements of $\mathfrak{B}(\mathcal{A})$ that *have a representation* with $\mathrm{cmp} \leq D$.

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

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

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

Proof. If $\beta = (\beta_1,\ldots,\beta_l) \in \mathfrak{N}^l_D(\mathcal{A})$ then $\beta_1,\ldots,\beta_l \in \mathfrak{N}_D(\mathcal{A})$ by definition. Assume that for each $j = 1,\ldots,l$ one has $\beta_j \in \mathfrak{N}_D(\mathcal{A})$, that is, assume that there exist a nilmanifold $X_j = G_j/\Gamma_j$ of nilpotency class $\leq D$ 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 $\omega = (\omega_1,\ldots,\omega_l)\colon \mathcal{Z} \longrightarrow G$. Then X is a nilmanifold of nilpotency class $\leq D$, $\omega \in G(\mathcal{A})$, $(\tau_{1,1},\ldots,\tau_{l,k_l})$ is a coordinate system on X, and we have $\beta = (\beta_1,\ldots,\beta_l) = (\tau_{1,i_1},\ldots,\tau_{l,i_l})\circ\omega$. ■

1.17. In light of Lemma 1.16, Theorem A_1^{**} implies its multidimensional extension:

Theorem A**. *For any ring $\mathcal{A}$ of real-valued functions and any $D, l \in \mathbb{N}$, $\mathfrak{N}^l_D(\mathcal{A}) = (\mathfrak{B}^{\mathrm{o}}_D(\mathcal{A}))^l$.*

2. Coordinate representation of a subnilmanifold and primitive GP mappings

2.1. We preserve notation of 1. Let $\pi\colon G \longrightarrow X$ be the natural projection, $\pi(g) = g\Gamma \in X$. Any closed (not necessarily connected) subgroup of G is a simply-connected nilpotent Lie group. *A subnilmanifold 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 a 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 subnilmanifold Y of X is *rational* if it is of the form $Y = \pi(gH)$ with rational $g \in G$.

2.2. We want to remind the reader some of the terminology introduced in Introduction.
(i) We 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 $\mathbb{Z}^d$.
(ii) 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 say that *the sequence $\{\omega(z)\}_{z\in E}$ is well distributed on X* (with respect to μ) if for any open set $U \subseteq X$ with $\mu(\partial U) = 0$ one has $\mathcal{D}(\omega^{-1}(U))/\mathcal{D}(E) = \mu(U)/\mu(X)$.

2.3. 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 $\{\pi(\omega(z))\}_{z\in\mathcal{Z}_i}$ is well distributed on a subnilmanifold Y_i of X with respect to the Haar measure on Y_i. If $\omega(0)=\mathbf{1}_G$, the subnilmanifolds $Y_1,\ldots,Y_m$ are all rational.*

2.4. In light of Theorem A* and Theorem 2.3, in order to describe the distribution of the values of a bounded generalized polynomial we have to determine how a subnilmanifold Y of X looks in coordinates on X. It is shown in [L4] that if Y is a subnilmanifold of X and τ is a coordinate mapping of X, then, up to a subset of Y of zero measure, $\tau(Y)$ is a piecewise polynomial surface. This clearly implies Theorem B. However, if we want to get information about the degree and the coefficients of the polynomials defining this surface we need to study $\tau(Y)$ more carefully.

2.5. Let $\{e_1,\ldots,e_k\}\in\Gamma$ be a Malcev 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 Malcev 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),

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

where, by 1.3, $S_1,\ldots,S_k$ are polynomials on $\mathbb{R}^s$. By 1.4, $\deg S_i\leq D_i$, $i=1,\ldots,k$. 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

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

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. In other words, H appears in coordinates on G as an s-dimensional rational polynomial surface of degree $\leq D$.

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

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

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

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

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

2.7. Now let Y be the subnilmanifold $\pi(gH) \subseteq X$. 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

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

the immersion $\mathcal{R}_Y = \tau \circ \eta^{-1}$ 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" $\hat{\pi} = \tau \circ \pi \circ \tilde{\tau}^{-1}: \mathbb{R}^k \longrightarrow [0,1)^k$:

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

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 $M_\gamma: \mathbb{R}^k \longrightarrow \mathbb{R}^k$ 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 of $\mathcal{R}_Y$ on $\mathcal{L}_j$ 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 $\deg R_i \leq D_i$, $i = 1, \ldots, k$, by 1.4 we have $\deg \mathcal{R}_j \leq D$.

2.8. We arrive at the following result:

Proposition. *Let $Y = \pi(gH)$ be a connected subnilmanifold of a connected nilmanifold X of nilpotency class D and let $\tau: X \longrightarrow [0,1)^k$ and $\eta: Y \longrightarrow [0,1)^s$ be coordinate systems on X and Y. The mapping $\mathcal{R}_Y = \tau \circ \eta^{-1}: [0,1)^s \longrightarrow [0,1)^k$ is piecewise polynomial in the following sense: there are one-to-one polynomial mappings $\mathcal{R}_1, \ldots, \mathcal{R}_N: \mathbb{R}^s \longrightarrow \mathbb{R}^k$ of degree $\leq D$ 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.*

2.9. We now return to generalized polynomials. Let u be (a fixed representation of) a generalized polynomial; the conventional polynomials occurring in the representation of u will be called *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 appears in the given representation as $u_1 + u_2$ or $u_1 u_2$;

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

In view of the constructive definition in 0.1, $I(u)$ is defined for any representation of any generalized polynomial.

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 $I(u)=I(u_1)\cup\ldots\cup I(u_l)$.

2.10. Let $u\colon\mathbb{Z}^d\longrightarrow\mathbb{R}^l$ be a bounded GP mapping; we will assume that $\mathrm{Ran}(u)\subseteq[0,1)^l$. Let $D=\mathrm{cmp}(u)$ and let $\mathcal{A}$ be the ring of real-valued functions on $\mathbb{Z}^d$ generated by the polynomials involved in u, so that $u\in(\mathfrak{B}^{\mathbf{o}}_D(\mathcal{A}))^l$. By Theorem A**, $u\in\mathfrak{N}^l_D(\mathcal{A})$. This means that there exist a nilmanifold $X=G/\Gamma$ of nilpotency class $\le D$ 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 2.3 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 subnilmanifolds $Y_1,\ldots,Y_m$ 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 1.15, $v_i=\eta_i\circ\pi\circ\omega|_{\mathcal{Z}_i}\colon\mathcal{Z}_i\longrightarrow[0,1)^s$ is a GP mapping of complexity $\le D$, and we have $u|_{\mathcal{Z}_i}=\rho\circ\tau\circ\eta_i^{-1}\circ v_i=\rho\circ\mathcal{R}_{Y_i}\circ v_i$:

$$\begin{array}{llcl}
u|_{\mathcal{Z}_i}:\mathcal{Z}_i & \xrightarrow{\pi\circ\omega} Y_i & \subseteq & X\\
\quad v_i\searrow & \quad\downarrow\eta_i & & \downarrow\tau\\
 & [0,1)^s & \xrightarrow{\mathcal{R}_{Y_i}} & [0,1)^k\\
 & & & \downarrow\rho\\
 & & & [0,1)^l.
\end{array}$$

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, $\{v_i(z)\}_{z\in\mathcal{Z}_i}$ is well distributed on $[0,1)^s$ with respect to λ. By Proposition 2.8 there exist a partition $[0,1)^s=\bigcup_{j=1}^{N_i}\mathcal{L}_{i,j}$ and polynomial mappings $\mathcal{R}_{i,1},\ldots,\mathcal{R}_{i,N_i}\colon\mathbb{R}^s\longrightarrow\mathbb{R}^k$ of degree $\le D$ such that $\mathcal{R}_{Y_i}|_{\mathcal{L}_{i,j}}=\mathcal{R}_{i,j}$ and $\mathcal{L}_{i,j}=\mathcal{R}_{i,j}^{-1}([0,1)^k)\cap[0,1)^s$, $j=1,\ldots,N_i$. For $j\in\{1,\ldots,N_i\}$ 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 $\le D$ 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 $\tilde{\tau}\circ\omega(z)=\big(\alpha_1(z),\ldots,\alpha_k(z)\big)$, $z\in\mathbb{Z}^d$, where $\alpha_1,\ldots,\alpha_k$ are polynomials from $\mathcal{A}$. Then $\tilde{\tau}(\omega(0))=\big(\alpha_1(0),\ldots,\alpha_k(0)\big)$, and $\alpha_1(0),\ldots,\alpha_k(0)$ 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 2.3, the components $Y_1',\dots,Y_m'$ of $\overline{\{\pi(\omega'(z)) : z\in\mathbb{Z}^d\}}$ are rational subnilmanifolds 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 $\overline{\{\pi(\omega(z)) : z\in\mathbb{Z}^d\}}$ have form $Y_i = \omega(0)Y_i' = \pi\bigl(\omega(0)g_iH_i\bigr)$, $i=1,\dots,m$. By Proposition 2.8, 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, are contained in $\mathfrak{F}$.

2.11. Let us say that a GP mapping $v\colon\mathcal{Z}\longrightarrow\mathbb{R}^s$, where $\mathcal{Z}$ is a subgroup of $\mathbb{Z}^d$, is *primitive* if v is representable as a composition $v=\eta\circ\pi\circ\omega$ where $\omega\colon\mathcal{Z}\longrightarrow H$ is a polynomial mapping to a nilpotent Lie group H, $\pi\colon H\longrightarrow Y$ is the projection mapping to a connected nilmanifold $Y=H/\Gamma'$ such that $\pi(\omega(\mathcal{Z}))$ is dense in Y, and $\eta\colon Y\longrightarrow[0,1)^s$ is a coordinate system on Y. If v is primitive, by Theorem 2.3 the sequence $\{\pi(\omega(z))\}_{z\in\mathcal{Z}}$ is well distributed in Y with respect to the Haar measure, and thus the sequence $\{v(z)\}_{z\in\mathcal{Z}}$ is well distributed on $[0,1)^s$ with respect to the Lebesgue measure.

Example. If a, b are rationally independent irrational numbers, the GP mapping $v(n) = \bigl(\langle\!\langle an\rangle\!\rangle, \langle\!\langle bn\rangle\!\rangle, \langle\!\langle -an[bn]\rangle\!\rangle\bigr)$, $n\in\mathbb{Z}$, is primitive. Indeed, let $H=\left\{\begin{pmatrix}1&a_{1,2}&a_{1,3}\\0&1&a_{2,3}\\0&0&1\end{pmatrix},\ a_{i,j}\in\mathbb{R}\right\}$, $\Lambda=\left\{\begin{pmatrix}1&m_{1,2}&m_{1,3}\\0&1&m_{2,3}\\0&0&1\end{pmatrix},\ m_{i,j}\in\mathbb{Z}\right\}$, Y be the connected nilmanifold H/Λ with "the natural" coordinate system $\eta\left(\begin{pmatrix}1&x_{1,2}&x_{1,3}\\0&1&x_{2,3}\\0&0&1\end{pmatrix}\right) = (x_{1,2},x_{1,3},x_{2,3})$ where $\begin{pmatrix}1&x_{1,2}&x_{1,3}\\0&1&x_{2,3}\\0&0&1\end{pmatrix}$ is the representation of $x\in X$ with all $x_{i,j}\in[0,1)$, and define the polynomial mapping $\omega\colon\mathbb{Z}\longrightarrow H$ by $\omega(n)=\begin{pmatrix}1&an&0\\0&1&bn\\0&0&1\end{pmatrix}$, $n\in\mathbb{Z}$. Then one can show that $\pi(\omega(\mathbb{Z}))$ is dense in Y, and we have $v=\eta\circ\pi\circ\omega$.

2.12. We can now summarize the content of subsection 2.10 thusly:

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 $s\in\mathbb{N}$, primitive GP mappings $v_i\colon\mathcal{Z}_i\longrightarrow[0,1)^s$, $i=1,\dots,m$, of complexity $\le\mathrm{cmp}(u)$, 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 $\le\mathrm{cmp}(u)$ 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 $\le\mathrm{cmp}(u)$ with coefficients from $\mathfrak{F}$, such that for $\mathcal{Z}_{i,j}=v_i^{-1}(\mathcal{L}_{i,j})$ one has $u|_{\mathcal{Z}_{i,j}}=\mathcal{P}_{i,j}\circ v_i$, $i=1,\dots,m$, $j=1,\dots,N_i$.*

3. Proof of Theorem B and exceptional values of GP mappings

3.1. Let us recall that *a polynomial surface $\mathcal{S}$* in $\mathbb{R}^l$ is the image under a polynomial mapping $\mathcal{P}\colon\mathbb{R}^s\longrightarrow\mathbb{R}^l$ of a subset $\mathcal{L}$ of $\mathbb{R}^s$ defined in $\mathbb{R}^s$ by a system of polynomial inequalities $0\le\mathcal{R}_j<1$, $j=1,\dots,k$, and having nonempty interior. *The degree* of $\mathcal{S}$ is the maximum of the degrees of $\mathcal{P}$ and of $\mathcal{R}_j$, $j=1,\dots,k$; *the coefficients* of $\mathcal{S}$ are those of

$\mathcal{P}$ and of $\mathcal{R}_j$, $j=1,\ldots,k$. The measure $\mu_{\mathcal{S}}$ on $\mathcal{S}$ is the normalized image of the Lebesgue measure $\lambda|_{\mathcal{L}}$ under $\mathcal{P}$, defined by $\mu_{\mathcal{S}}(A)=\lambda(\mathcal{P}^{-1}(A)\cap\mathcal{L})/\lambda(\mathcal{L})$ for Borel sets $A\subseteq\mathbb{R}^l$. Theorem B** implies the following, more precise version of Theorem B:

Theorem B*. *Let $u\colon\mathbb{Z}^d\longrightarrow\mathbb{R}^l$ be a bounded GP mapping. There exist bounded polynomial surfaces $\mathcal{S}_1,\ldots,\mathcal{S}_k\subset\mathbb{R}^l$ of degree $\leq\operatorname{cmp}(u)$ and a partition $\mathbb{Z}^d=\mathcal{Z}_*\cup\bigcup_{i=1}^k E_i$ such that $\mathcal{D}(\mathcal{Z}_*)=0$ and such that for every $i\in\{1,\ldots,k\}$, $\mathcal{D}(E_i)>0$ and the sequence $\{u(z)\}_{z\in E_i}$ is well distributed on $\mathcal{S}_i$ with respect to $\mu_{\mathcal{S}_i}$. The coefficients of $\mathcal{S}_1,\ldots,\mathcal{S}_k$ 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 GP mapping u lie and are well distributed on a piecewise polynomial surface (whereas the exceptional values, which do not affect the distributional behavior of u, are out of our control).

Proof. We keep the notation of subsection 2.12. Fix $i\in\{1,\ldots,m\}$. For each $j\in\{1,\ldots,N_i\}$ the set $\mathcal{L}_{i,j}\subseteq[0,1)^s$ is defined by a collection of polynomial inequalities, and thus either the interior $\mathcal{L}^{\mathbf{o}}_{i,j}$ of $\mathcal{L}_{i,j}$ is nonempty, $\lambda(\mathcal{L}_{i,j})>0$ and $\{v_i(z)\}_{z\in\mathcal{Z}_{i,j}}$ is well distributed on $\mathcal{L}_{i,j}$ with respect to the Lebesgue measure, or $\mathcal{L}^{\mathbf{o}}_{i,j}$ is empty, $\lambda(\mathcal{L}_{i,j})=0$ and $\mathcal{Z}_{i,j}$ has zero density in $\mathcal{Z}_i$. Let us assume that $\mathcal{L}^{\mathbf{o}}_{i,1},\ldots,\mathcal{L}^{\mathbf{o}}_{i,r_i}\neq\emptyset$ and $\mathcal{L}^{\mathbf{o}}_{r_i+1},\ldots,\mathcal{L}^{\mathbf{o}}_{N_i}=\emptyset$, 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 $j=1,\ldots,r_i$ the sequence $\{u(z)\}_{z\in\mathcal{Z}_{i,j}}$ is well distributed, with respect to $\mu_{\mathcal{S}_{i,j}}$, on the polynomial surface $\mathcal{S}_{i,j}=\mathcal{P}_{i,j}(\mathcal{L}_{i,j})$ of degree $\leq\operatorname{cmp}(u)$. Finally, we put $\mathcal{Z}_*=\bigcup_{i=1}^m\mathcal{Z}_{i,*}$. $\blacksquare$

3.2. Remark. The values of u are well distributed on the piecewise polynomial surface $\mathcal{S}=\bigcup_{\substack{i=1,\ldots,m\\ j=1,\ldots,r_i}}\mathcal{S}_{i,j}=\bigcup_{\substack{i=1,\ldots,m\\ j=1,\ldots,r_i}}\mathcal{P}_{i,j}(\mathcal{L}_{i,j})$; as well, we can take $\mathcal{S}=\bigcup_{\substack{i=1,\ldots,m\\ j=1,\ldots,r_i}}\mathcal{P}_{i,j}(\mathcal{L}^{\mathbf{o}}_{i,j})$. We then see that, in Theorem B, one may assume that $\mathcal{S}=f(V)$ where V is a dense open subset of a nilmanifold Y and f is a piecewise polynomial mapping $Y\longrightarrow\mathbb{R}^l$, continuous on V.

3.3. 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 $w\colon\mathbb{Z}^d\longrightarrow\mathbb{R}$, with $\mathcal{D}(\mathcal{W})=0$.*

Proof. Note that, in the proof of Theorem B* in 3.1, 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 $w|_{\mathcal{Z}_i}=S_i\circ v_i$, $i=1,\ldots,m$. For each i, since S_i is a nonzero polynomial and $\{v_i(z)\}_{z\in\mathcal{Z}_i}$ is well distributed on $[0,1)^s$ with respect to the Lebesgue measure, $w|_{\mathcal{Z}_i}^{-1}(0)$ has zero density in $\mathcal{Z}_i$. $\blacksquare$

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

Examples. (1) Let a be an irrational number and let $u(n)=\bigl[1-\langle\!\langle an\rangle\!\rangle\bigr]$. Then $u(n)=0$ for all $n\neq0$ and $u(0)=1$ is an exceptional value of u.

(2) Let $a\in\mathbb{R}$ be such that the set $S_a=\bigl\{n\in\mathbb{N}:0<\langle\!\langle an\rangle\!\rangle<\frac{1}{n}\bigr\}$ 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 $[0, 1]$.

3.5. A converse of Theorem B also holds, namely, for any piecewise polynomial surface in $\mathbb{R}^l$ whose domain is a cube, or a finite union of cubes, there exists a GP mapping whose values are well distributed on this surface. Indeed, assume that a piecewise polynomial surface $\mathcal{S}$ is defined by a piecewise polynomial function $h: Q \longrightarrow \mathbb{R}^l$, where $Q \subset \mathbb{R}^s$ is a cube, or a finite union of cubes. Choose a GP mapping $v: \mathbb{Z} \longrightarrow Q$ such that the values of v are well distributed on Q with respect to the Lebesgue measure. (Say, if $Q = [0, 1]^s$, we can take $v(n) = \big(\langle\!\langle a_1 n \rangle\!\rangle, \ldots, \langle\!\langle a_s n \rangle\!\rangle\big)$, $n \in \mathbb{Z}$, where $a_1, \ldots, a_s$ are rationally independent irrational numbers.) Define $u = h \circ v$; by Lemma 1.6, u is a GP mapping, and the values of u are well distributed on $\mathcal{S}$.

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

The group $G = \left\{ \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix},\ a, b, c \in \mathbb{R} \right\}$ of 3×3 upper triangular matrices with unit diagonal is a connected simply-connected nilpotent Lie group, and $\Gamma = \left\{ \begin{pmatrix} 1 & m & k \\ 0 & 1 & l \\ 0 & 0 & 1 \end{pmatrix},\ m, k, l \in \mathbb{Z} \right\}$ is a discrete uniform subgroup of G; let $X = G/\Gamma$ and $\pi: G \longrightarrow X$ be the natural projection. Let $e_{1,2} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $e_{1,3} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $e_{2,3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$, then $\{e_{2,3}, e_{1,2}, e_{1,3}\}$ is a Malcev basis in G such that $e_{2,3}^c e_{1,2}^a e_{1,3}^b = \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}$, $a, b, c \in \mathbb{R}$. Thus, in the Malcev basis $\{e_{2,3}, e_{1,2}, e_{1,3}\}$ the coordinates of a matrix $A = \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} \in G$ are $\tilde{\tau}(A) = (c, a, b)$. The fundamental domain in G is $Q = \left\{ \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix},\ a, b, c \in [0, 1) \right\}$, and for a matrix $A = \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix} \in G$ 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 $\tau_3(\omega(n)) = \big\langle\!\big\langle \frac{1}{2}\alpha^2 n^2 - \alpha n [\alpha n] \big\rangle\!\big\rangle = u(n)$, $n \in \mathbb{Z}$.

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, thus $Y = \pi(H)$ is a 1-dimensional subnilmanifold of X. Define the coordinate mapping $\tilde{\eta}: H \longrightarrow \mathbb{R}$ by $\tilde{\eta}\Big(\Big(\begin{smallmatrix}1 & a & \frac{1}{2}a^2\\ 0 & 1 & a\\ 0 & 0 & 1\end{smallmatrix}\Big)\Big) = \frac{1}{2}a$, 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 on $[0,1)$ with respect to the Lebesgue measure, and so, $\pi(\omega(n))$ is well distributed on Y with respect to the Haar measure on 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 $\gamma = \Big(\begin{smallmatrix}1&1&0\\0&1&1\\0&0&1\end{smallmatrix}\Big)$ one has

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

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 $\mathcal{R}_H\big([\frac{1}{2},1)\big) \subset C_\gamma$. 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 of $\mathcal{R}_2$ 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 $u(n) = \mathcal{P}\big(\langle\!\langle \frac{1}{2}\alpha n\rangle\!\rangle\big)$, $n \in \mathbb{Z}$. The sequence $\big\{\frac{1}{2}\alpha n\big\}_{n\in\mathbb{Z}}$ is well distributed on $[0,1)$ with respect to the Lebesgue measure dy; hence, $u(n)$, $n \in \mathbb{Z}$, is well distributed on $[0,1)$ with respect to the measure

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

4. Proofs of Theorems C, D, B$_c$, and of other results from Introduction

4.1. The following theorem (cf. Corollary 0.25 of Introduction) clearly follows from Theorem B*:

Theorem. *Let $u: \mathbb{Z}^d \longrightarrow \mathbb{R}^l$ 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 $\sum_{i=1}^k\big(\mathcal{D}(E_i)\int_{\mathcal{S}_i} f\,d\mu_{\mathcal{S}_i}\big)$.*

4.2. Corollary. (Corollary 0.26 of Introduction) *For any generalized polynomial $u: \mathbb{Z}^d \longrightarrow \mathbb{R}$ 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.*

4.3. The proofs of the following two propositions (Corollaries 0.27 and 0.28 of Introduction) are similar; we confine ourselves to the proof of the first of them:

Proposition. *Let $U_1^t,\dots,U_k^t$, $t\in\mathbb{R}$, be commuting unitary flows on a Hilbert space $\mathcal{H}$ and let $u_1,\dots,u_k$ 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 $(U_j^tg)(x)=e^{2\pi i f_j(x)t}g(x)$, $g\in L^2(\Omega)$, $j=1,\dots,k$, $x\in\Omega$, where f_j are measurable real-valued functions on Ω. Then, for any $g\in L^2(\Omega)$ and $x\in\Omega$,

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

where $u_x(n)=\sum_{j=1}^k f_j(x)u_j(n)$, $n\in\mathbb{Z}^d$. By Corollary 4.2, 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 $\mathcal{H}=L^2(\Omega)$. ■

4.4. Proposition. *Let $U_1,\dots,U_k$ be commuting unitary operators on a Hilbert space and let $u_1,\dots,u_k$ 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.*

4.5. The following proposition can be viewed as a "measure preserving" version of the "unitary" result contained in Proposition 4.3.

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

Proof. Let $\tau\colon X\longrightarrow[0,1)^k$ be a coordinate system. By Theorem A**, $u=\tau\circ\pi\circ\omega\in\mathfrak{B}(\mathcal{A})^k$, and in the case under consideration $\mathfrak{B}(\mathcal{A})=\mathcal{A}$. So, $u\colon\mathbb{Z}^d\longrightarrow[0,1)^k$ 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 Theorem 4.1, $\lim\limits_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{z\in\Phi_N}f(\pi(\omega(z)))=\lim\limits_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{z\in\Phi_N}\hat f(u(z))$ exists for any Følner sequence $\{\Phi_N\}_{N=1}^\infty$ in $\mathbb{Z}^d$. ■

4.6. We now move the discussion to the recurrence properties of generalized polynomials, dealt with in Theorems C and D of Introduction. Given a mapping u from $\mathbb{Z}^d$ to a topological space X, we will say that a point $x\in X$ is *an IP*-limit of u* if for any neighborhood W of x the set $u^{-1}(W)$ is IP*, and that $x\in X$ is *an IP_+^*-limit of u* if for any neighborhood W of x the set $u^{-1}(W)$ is IP_+^*. The following fact is proved in [L3]:

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

4.7. Corollary. *If $Y=\overline{\pi\circ\omega(\mathbb{Z}^d)}$, then every point of Y is an IP_+^*-limit of $\pi\circ\omega$.*

Proof. For any $z' \in \mathbb{Z}^d$ the point $\pi(\omega(z'))$ is an IP*-limit of the mapping $\pi\circ\omega'$ where $\omega'(z) = \omega(z+z')$, and thus for every open set $W \subseteq Y$ the set $(\pi\circ\omega)^{-1}(W)$ is an IP$^*_+$ set. ∎

4.8. Theorem C. *Let $u{:}\,\mathbb{Z}^d \longrightarrow \mathbb{R}^l$ be a bounded GP mapping and let $\mathcal{S}$ be the piecewise polynomial surface on which the values of u are well distributed. Then for every open set $W \subseteq \mathcal{S}$, $u^{-1}(W)$ is an IP$^*_+$ set. In other words, every point of $\mathcal{S}$ is an IP$^*_+$-limit of u, and in particular, every regular value of u is IP$^*_+$-recurrent.*

Proof. By Theorem A, u is representable in the form $u = f\circ\pi\circ\omega$ where $\omega{:}\,\mathbb{Z}^d \longrightarrow G$ is a polynomial mapping to a nilpotent Lie group G, $\pi{:}\,G \longrightarrow X$ is the projection to a nilmanifold $X = G/\Gamma$, and $f{:}\,X \longrightarrow \mathbb{R}^l$ is a piecewise polynomial mapping. Let $Y = \overline{\pi\circ\omega(\mathbb{Z}^d)}$; then Y has a dense open subset V such that f is continuous on V and $\mathcal{S} = f(V)$. (See Remark 3.2.) If W is an open subset of $\mathcal{S}$ then $U = f^{-1}(W) \cap V$ is a nonempty open subset of Y; by Corollary 4.7, $(\pi\circ\omega)^{-1}(U)$ is an IP$^*_+$ set, and so $u^{-1}(W)$ is an IP$^*_+$ set. ∎

4.9. 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 $\tilde{u}$ 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 of $\tilde{u}$.*

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

Utilizing Theorem A**, we can find a nilmanifold $X = G/\Gamma$ with the natural projection $\pi{:}\,G \longrightarrow X$, a coordinate system $\tau{:}\,X \longrightarrow [0,1)^k$, a mapping $\omega \in G(\mathcal{A})$, and indices $i_1,\ldots,i_l \in \{1,\ldots,k\}$ such that $v = (\tau_{i_1},\ldots,\tau_{i_l})\circ\pi\circ\omega$. 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 of the projection $\rho{:}\,\mathbb{T}^k \longrightarrow \mathbb{T}^l$, $\rho(y_1,\ldots,y_k) = (y_{i_1},\ldots,y_{i_l})$:

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

By Proposition 4.6, o is an IP*-limit of $\pi\circ\omega$. Hence, $0 \in \mathbb{T}^k$ is an IP*-limit of $\sigma\circ\tau\circ\pi\circ\omega$ and $0 \in \mathbb{T}^l$ is an IP*-limit of $\tilde{u}$. ∎

4.10. Theorem. (Theorem 0.34 of Introduction) *Let $u_i{:}\,\mathbb{Z}^{d+i-1} \longrightarrow \mathbb{R}$, $i = 1,\ldots,k$, be generalized polynomials such that all ordinary polynomials involved in u_i have zero constant term. Then for any $\delta > 0$, the set of $n \in \mathbb{Z}^d$ for which there exist $m_1,\ldots,m_k \in \mathbb{Z}$ satisfying*

$$\bigl|u_1(n) - m_1\bigr| < \delta,\ \bigl|u_2(n,m_1) - m_2\bigr| < \delta,\ \ldots,\ \bigl|u_k(n,m_1,\ldots,m_{k-1}) - m_k\bigr| < \delta \qquad (4.1)$$

is an IP* set.

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

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

By Theorem 0.33, 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 $\epsilon_1,\ldots,\epsilon_k\in\{-1,1\}$ is an IP* set. For any such n we now construct a solution of (4.1) 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. ■

4.11. The following is a refinement of Theorem B$_c$:

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

To prove this theorem one just has, in the proof of Theorem B*, to switch from $\mathbb{Z}^d$ to $\mathbb{R}^d$ and to substitute Theorem 2.3 by the following theorem, which is a special case of results in [Sh1].

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

5. Legal orders and reduction formulas

We now proceed to the algebraic part of the paper, which will lead us to the proof of Theorem A$_1^{**}$.

5.1. 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 operation of taking brackets; $\mathfrak{B}^{\mathbf{o}}(\mathcal{A}) \subset \mathfrak{B}(\mathcal{A})$ consists of functions with range in $[0,1)$; and $\mathfrak{N}(\mathcal{A})$ is the set of functions which can be generated by a nilsystem, that is, functions of the form $\tau\circ\omega$ where ω is an $\mathcal{A}$-mapping from $\mathcal{Z}$ to a nilpotent Lie group G and τ is a coordinate on the nilmanifold $X = G/\Gamma$.

5.2. It was shown in subsection 1.15 that $\mathfrak{N}(\mathcal{A}) \subseteq \mathfrak{B}^{\mathbf{o}}(\mathcal{A})$. To establish the inclusion $\mathfrak{B}^{\mathbf{o}}(\mathcal{A}) \subseteq \mathfrak{N}(\mathcal{A})$, stated in Theorem A_1^{**}, we will use the group of upper triangular matrices with unit diagonal. For $d \in \mathbb{N}$ let $M_d = \left\{ \begin{pmatrix} 1 & a_{1,2} & \dots & a_{1,d} \\ & 1 & \dots & a_{2,d} \\ & & \ddots & \vdots \\ \mathbf{O} & & & 1 \end{pmatrix},\ a_{i,j} \in \mathbb{R} \right\}$. M_d is a connected simply-connected nilpotent Lie group, and $\Gamma_d = \left\{ \begin{pmatrix} 1 & n_{1,2} & \dots & n_{1,d} \\ & 1 & \dots & n_{2,d} \\ & & \ddots & \vdots \\ \mathbf{O} & & & 1 \end{pmatrix},\ n_{i,j} \in \mathbb{Z} \right\}$ 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 subdiagonal entries and therefore assume that their entries are indexed by the pairs (i,j) with $1 \le i < j \le d$.

5.3. 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 $M_d(\mathcal{A}) = \left\{ \begin{pmatrix} 1 & \alpha_{1,2} & \dots & \alpha_{1,d} \\ & 1 & \dots & \alpha_{2,d} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix},\ \alpha_{i,j} \in \mathcal{A} \right\}$ of upper-triangular matrices with entries from $\mathcal{A}$. Let $\mathfrak{B}(\mathcal{A})$ be the bracket extension of $\mathcal{A}$; for any matrix $P \in M_d(\mathcal{A})$ there exists a unique 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 Malcev 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 $u - [u]$.

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

$$E_{i,j} = \begin{array}{c@{}l} \begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ & 1 & \dots & 0 & \dots & 0 \\ & & \ddots & & & \vdots \\ & & & 1 & & \vdots \\ & & & & \ddots & \vdots \\ & & & & & 1 \end{pmatrix} & \begin{array}{l} \\ \\ \\ i \\ \\ \\ \end{array} \\ \begin{array}{cccccc} & & & j & & \end{array} & \end{array}$$

The set $\{E_{i,j}\}_{1\le i<j\le d}$ is a Malcev basis in M_d compatible (see 1.2) with Γ_d, and for $a \in \mathbb{R}$ we have $E_{i,j}^a = \begin{array}{c@{}l} \begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ & 1 & \dots & 0 & \dots & 0 \\ & & \ddots & & & \vdots \\ & & & a & & \vdots \\ & & & & \ddots & \vdots \\ & & & & & 1 \end{pmatrix} & \begin{array}{l} \\ \\ \\ i. \\ \\ \\ \end{array} \\ \begin{array}{cccccc} & & & j & & \end{array} & \end{array}$

5.5. At first glance it seems that with respect to the Malcev basis $\{E_{i,j}\}_{1\le i<j\le d}$ the coordinates of a matrix $\begin{pmatrix} 1 & a_{1,2} & \dots & a_{1,d} \\ & 1 & \dots & a_{2,d} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_d$ 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 precisely, is only true for a specific ordering of the Malcev 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 Malcev 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 if $i = k$ and $j < l$. Then the product $\prod\limits_{1\le i<j\le d} E_{i,j}^{a_{i,j}}$ computed with respect to $\prec$ equals $\begin{pmatrix} 1 & a_{1,2} & \dots & a_{1,d} \\ & 1 & \dots & a_{2,d} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix}$.

5.6. The set of elements of $\mathfrak{B}(\mathcal{A})$ which can be obtained with the help of the Malcev basis $\{E_{i,j}\}_{1\le i<j\le d}$ in M_d ordered by the order $\prec$ defined in 5.5, is restricted to *nested* elements, that is, the elements of $\mathfrak{B}(\mathcal{A})$ 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 $\chi(P)$ (that was introduced in subsection 5.3) 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.

5.7. For a matrix $P \in M_d$ we will now compute the coordinates of $\chi(P)$ with respect to the Malcev basis $\{E_{i,j}^{\epsilon_{i,j}}\}_{1\le i<j\le d}$, $\epsilon_{i,j} \in \{-1,1\}$, 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 the order $\prec$ must form a Malcev 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 (simultaneously) $i \ge k$ and $j \le l$.

Let $\prec$ be a legal order on $\{(i,j)\}_{1\le i<j\le d}$ and let $\epsilon_{i,j} \in \{-1,1\}$, $1 \le i < j \le d$. 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 $P = \prod_{1\le i<j\le d} E_{i,j}^{\epsilon_{i,j}z_{i,j}}$, $z_{i,j} \in \mathbb{Z}$, where the product is computed in accordance with the order $\prec$; we will call $z_{i,j}$ *the* (i,j)*-coordinate of* P. (Note that though the integers $z_{i,j}$ in the formula for P may take negative values, the signs $\epsilon_{i,j}$ are not superfluous: the bases $\{E_{i,j}\}$ and $\{E_{i,j}^{\epsilon_{i,j}}\}$ are different, and may produce different generalized polynomials.)

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

5.9. Now let $\chi(P)$ be the matrix in the fundamental domain of M_d corresponding to P, that is, $\chi(P) = P\big(\prod_{1\le i<j\le d} E_{i,j}^{m_{i,j}}\big)$ 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

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

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 $\prod_{(i,j)\prec(k,l)} E_{r,s}^{\epsilon_{i,j} x_{r,s}}$ 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 $\prod_{(r,s)\prec(k,l)} E_{r,s}^{\epsilon_{r,s} x_{r,s}}$ 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 $x_{i,j} = \epsilon_{i,j} \xi_{i,j} - [\epsilon_{i,j} \xi_{i,j}]$.

6. Bracket algebra

6.1. 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 they are elements of an arbitrary commutative ring $\mathcal{A}$. 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 brackets. We will now introduce the necessary algebraic formalism.

6.2. Given a set S, we denote by $\Sigma[S]$ the free commutative ring generated by the set $\{[u],\ u \in S\}$, that is, the commutative ring of formal finite sums of the form $\sum_{i=1}^{l} \pm[u_{i,1}]\dots[u_{i,m_i}]$ with $l \geq 0$, $m_i \in \mathbb{N}$, and $u_{i,j} \in S$, where the cancellation of equal summands appearing with opposite signs is allowed.

For commutative rings R and Q let $R * Q$ be the commutative ring freely generated by R and Q, that is, $R * Q = R \oplus Q \oplus (R \otimes Q)$ 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 $r \otimes q$.

6.3. Let $\mathcal{A}$ be a commutative ring. We are going to construct an algebra $\mathfrak{B}$, which we will call *the bracket algebra over* $\mathcal{A}$. We put $\mathbf{B}_0 = \Sigma[\mathcal{A}]$; if $\mathbf{B}_k$ is already defined, let $\mathbf{B}_k^{\mathbf{b}} = \Sigma[\mathbf{B}_k]$ and $\mathbf{B}_{k+1} = \mathbf{B}_k * \mathbf{B}_k^{\mathbf{b}}$.[4] 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 $[\cdot] \colon \mathbf{B} \longrightarrow \mathbf{B}^{\mathbf{b}}$ 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 $[\cdot] \colon \mathfrak{B} \longrightarrow \mathfrak{B}^{\mathbf{b}}$ is well defined, identical on $\mathfrak{B}^{\mathbf{b}}$ and satisfies $[u+v] = [u] + v$ for any $u \in \mathfrak{B}$, $v \in \mathfrak{B}^{\mathbf{b}}$. The elements of $\mathfrak{B}$ will be called *bracket expressions* over $\mathcal{A}$.

6.4. As an abelian group, $\mathfrak{B}$ is generated by the expressions 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 of the form $[v_1]\dots[v_m]$ where $m \geq 1$, $v_1, \dots, v_m \in \mathfrak{B}$. We will call such expressions *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 $\mathfrak{B}$.

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

[4] Note that elements of $\Sigma[\mathbf{B}_k]$, as defined in subsection 6.2, have no integer coefficients, so that no confusion of elements of $\mathbf{B}_k^{\mathbf{b}}$ with elements of $\mathcal{A} \otimes \mathbf{B}_k^{\mathbf{b}}$ may occur even if $\mathcal{A}$ contains integers. For example, an expression of the form, say, $2[v]$ should not be interpreted as $[v] + [v]$, but only as $2 \otimes [v]$.

6.6. We will say that an expression

$$\sum_{i=1}^{k} a_i[v_{i,1}]\dots[v_{i,m_i}] + \sum_{i=1}^{l} \pm[u_{i,1}]\dots[u_{i,n_i}], \quad \text{with } a_i \in \mathcal{A} \text{ and } u_{i,j}, v_{i,j} \in \mathfrak{B}, \tag{6.1}$$

representing an element of $\mathfrak{B}$ is *reduced* if (i) all $v_{i,j}$, $u_{i,j}$ belong to $\mathfrak{B}^{\mathbf{t}}$ and are represented in the reduced form; (ii) the monomials $v_i = [v_{i,1}]\dots[v_{i,m_i}]$, for $i = 1, \dots, k$, are all different, so that no combining of like terms is possible; and (iii) equal monomials $u_i = [u_{i,1}]\dots[u_{i,n_i}]$, for $i = 1, \dots, l$, have identical signs, so that no cancellation is possible. Every expression $u \in \mathfrak{B}$ is uniquely representable in the reduced form.[5] 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.

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

6.8. We will now transfer the formulas obtained in 5.9 to the "abstract" environment we introduced. 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 $\epsilon = (\epsilon_{i,j})_{1\le i<j\le d}$ with $\epsilon_{i,j} \in \{-1, 1\}$, $1 \le i < j \le d$, *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 $\epsilon = (\epsilon_{i,j})_{1\le i<j\le d}$, and a legal order $\prec$ (see 5.7) on the set $\{(i,j)\}_{1\le i<j\le d}$, we define $\varphi_{i,j}, \psi_{i,j}, \xi_{i,j}$ and $\varphi_{i,j}^{k,l}, \psi_{i,j}^{k,l}, \xi_{i,j}^{k,l} \in$

[5] Here is how the reduction can be done. First, one reduces the expressions for all the elements $v_{i,j}$ and $u_{i,j}$ of $\mathfrak{B}$ appearing in (6.1) (which can be done by induction on the length of the expressions), and takes their $\mathfrak{B}^{\mathbf{b}}$-parts out of the brackets. Next, if a monomial $[v_{i,1}]\dots[v_{i,m_i}]$ appears in (6.1) more than once, the corresponding summands should be combined by adding coefficients. Finally, if a monomial $[u_{i,1}]\dots[u_{i,n_i}]$ appears in (6.1) twice with opposite signs, the corresponding summands should be canceled.

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

$$\begin{aligned}
\varphi_{i,j}^{k,l} &= a_{i,j} - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}} \varphi_{i,n}^{n,j}[\xi_{n,j}]^{\epsilon_{n,j}} - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}} \big(\varphi_{i,n} - [\xi_{i,n}]^{\epsilon_{i,n}}\big)[\xi_{n,j}]^{\epsilon_{n,j}} \\
\varphi_{i,j} &= \varphi_{i,j}^{i,j} = a_{i,j} - \sum_{(n,j)\prec(i,n)} \varphi_{i,n}^{n,j}[\xi_{n,j}]^{\epsilon_{n,j}} - \sum_{(n,j)\succ(i,n)} \big(\varphi_{i,n} - [\xi_{i,n}]^{\epsilon_{i,n}}\big)[\xi_{n,j}]^{\epsilon_{n,j}} \\
\psi_{i,j}^{k,l} &= \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}} \psi_{i,n}^{n,j}\big(\xi_{n,j} - [\xi_{n,j}]^{\epsilon_{n,j}}\big) + \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}} \big(\varphi_{i,n} - [\xi_{i,n}]^{\epsilon_{i,n}}\big)\big(\xi_{n,j} - [\xi_{n,j}]^{\epsilon_{n,j}}\big) \\
\psi_{i,j} &= \psi_{i,j}^{i,j} = \sum_{(n,j)\prec(i,n)} \psi_{i,n}^{n,j}\big(\xi_{n,j} - [\xi_{n,j}]^{\epsilon_{n,j}}\big) + \sum_{(n,j)\succ(i,n)} \big(\varphi_{i,n} - [\xi_{i,n}]^{\epsilon_{i,n}}\big)\big(\xi_{n,j} - [\xi_{n,j}]^{\epsilon_{n,j}}\big) \\
\xi_{i,j}^{k,l} &= \varphi_{i,j}^{k,l} - \psi_{i,j}^{k,l} = a_{i,j} - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}} \xi_{i,n}^{n,j}[\xi_{n,j}]^{\epsilon_{n,j}} - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\prec(i,n)}} \psi_{i,n}^{n,j}\xi_{n,j} \\
&\qquad + \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}} [\xi_{i,n}]^{\epsilon_{i,n}}\xi_{n,j} - \sum_{\substack{(n,j)\prec(k,l)\\(n,j)\succ(i,n)}} \varphi_{i,n}\xi_{n,j} \\
\xi_{i,j} &= \varphi_{i,j} - \psi_{i,j} = a_{i,j} - \sum_{(n,j)\prec(i,n)} \xi_{i,n}^{n,j}[\xi_{n,j}]^{\epsilon_{n,j}} - \sum_{(n,j)\prec(i,n)} \psi_{i,n}^{n,j}\xi_{n,j} \\
&\qquad + \sum_{(n,j)\succ(i,n)} [\xi_{i,n}]^{\epsilon_{i,n}}\xi_{n,j} - \sum_{(n,j)\succ(i,n)} \varphi_{i,n}\xi_{n,j}.
\end{aligned} \tag{6.2}$$

When it is not clear from the context for what 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 (6.2) 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 $s > j$.

6.9. 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 needed to obtain an element $u \in \mathfrak{B}^{\mathbf{t}}$. We will later formulate and prove a stronger statement, Theorem A^{***} in 10.4, from which Theorem A_1^{**} will follow.

7. Elementary bracket expressions and ordering of trees and bushes

7.1. We will say that a bracket expression $p \in \mathfrak{B}$ is *elementary* if it is constructible from elements of $\mathcal{A}$ without using the addition or subtraction. More precisely, p is elementary if either $p = a[p_1]\dots[p_m]$ where $a \in \mathcal{A}$, $m \ge 0$, and $p_1,\dots,p_m \in \mathfrak{B}$ are elementary, or

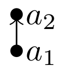
$p = [p_1] \dots [p_m]$ where $m \geq 1$ and $p_1, \dots, p_m \in \mathfrak{B}$ are elementary.[6] We will denote the set of elementary elements of $\mathfrak{B}$ by $\mathfrak{E}$.

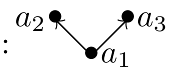
Example. $p_1[p_2[p_3]][p_4]$ is elementary, $p_1[p_2 + p_3]$ and $p_1[p_2] - p_3$ are not.

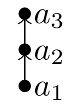
7.2. Elements of $\mathfrak{E}$ can be described by oriented graphs labeled by elements of $\mathcal{A}$. The following examples illustrate what we mean:

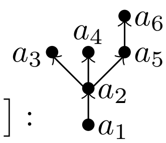
Examples.

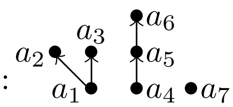
$p_1 = a$: 

$p_2 = a_1[a_2]$: 

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

$p_4 = a_1[a_2[a_3]]$: 

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

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

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

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

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

In the examples 7.2, $p_1, \dots, p_5$ are trees and q is a bush.

7.4. In the proof of Proposition 6.9 we will use a cumbersome induction based on the structure of the trees representing elements of $\mathfrak{E}$. We will now introduce some parameters of trees and bushes, that is, of elements of $\mathfrak{E}$.

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

(ii) *The height* $\operatorname{hgt}(p)$ of $p \in \mathfrak{E}$ is
- $\operatorname{hgt}(p) = 0$ if $p \in \mathcal{A}$;
- $\operatorname{hgt}(p) = 1 + \max\{\operatorname{hgt}(p_i)\}_{i=1}^m$ if $p = a[p_1] \dots [p_m] \in \mathfrak{E}^{\mathbf{t}}$ with $a \in \mathcal{A}$ and $p_1, \dots, p_m \in \mathfrak{E}^{\mathbf{t}}$;
- $\operatorname{hgt}(p) = \max\{\operatorname{hgt}(p_i)\}_{i=1}^m$ if $p = [p_1] \dots [p_m] \in \mathfrak{E}^{\mathbf{b}}$ with $p_1, \dots, p_m \in \mathfrak{E}^{\mathbf{t}}$.

---

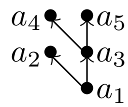
[6] Here is how this definition should be understood: we put $\mathbf{E}_0 = \mathcal{A}$, then $\mathbf{E}_k = \mathbf{E}_{k-1} \cup \Big\{a[p_1] \dots [p_m],\ m \geq 0,\ a \in \mathcal{A},\ p_1, \dots, p_m \in \mathbf{E}_{k-1}\Big\} \cup \Big\{[p_1] \dots [p_m],\ m \geq 1,\ p_1, \dots, p_m \in \mathbf{E}_{k-1}\Big\}$ for $k = 1, 2, \dots$, and finally, $\mathfrak{E} = \bigcup_{k=0}^{\infty} \mathbf{E}_k$; then $\mathfrak{E}$ is the set of elementary elements of $\mathfrak{B}$.

(iii) *The number of branches* $\operatorname{brn}(p)$ for $p \in P$ is defined by

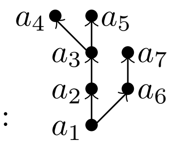
$\operatorname{brn}(p) = 0$ if $p \in \mathcal{A}$,

$\operatorname{brn}(p) = m$ if $p = a[p_1]\dots[p_m] \in \mathfrak{E}^{\mathbf{t}}$ with $a \in \mathcal{A}$ and $p_1, \dots, p_m \in \mathfrak{E}^{\mathbf{t}}$;

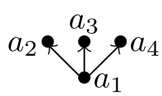
$\operatorname{brn}(p) = m$ if $p = [p_1]\dots[p_m] \in \mathfrak{E}^{\mathbf{b}}$ with $p_1, \dots, p_m \in \mathfrak{E}^{\mathbf{t}}$.

Examples. $p_1 = a_1[a_2]\big[a_3[a_4][a_5]\big]$: 

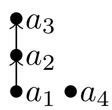
$\operatorname{cmp}(p_1) = 5$, $\operatorname{hgt}(p_1) = 2$, $\operatorname{brn}(p_1) = 2$.

$p_2 = a_1\big[a_2[a_3[a_4][a_5]]\big][a_6[a_7]]$: 

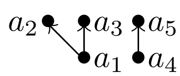
$\operatorname{cmp}(p_2) = 7$, $\operatorname{hgt}(p_2) = 3$, $\operatorname{brn}(p_2) = 2$.

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

$\operatorname{cmp}(p_3) = 4$, $\operatorname{hgt}(p_3) = 1$, $\operatorname{brn}(p_3) = 3$.

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

$\operatorname{cmp}(q_1) = 4$, $\operatorname{hgt}(q_1) = 2$, $\operatorname{brn}(q_1) = 2$.

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

$\operatorname{cmp}(q_2) = 5$, $\operatorname{hgt}(q_2) = 1$, $\operatorname{brn}(q_2) = 2$.

7.5. The following can be checked directly:

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

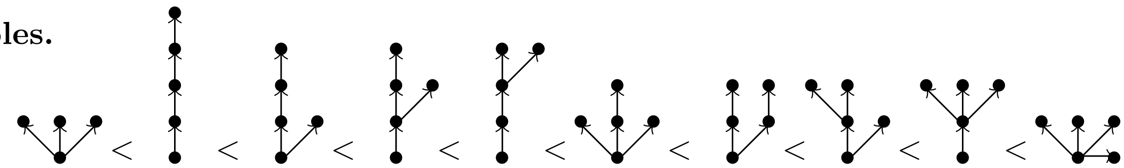
7.6. We now introduce an order on the set of trees $\mathfrak{E}^{\mathbf{t}}$ and, independently, on the set of bushes $\mathfrak{E}^{\mathbf{b}}$; trees will not be comparable with bushes. Strictly speaking, this will be linear orders on the set of non-labeled trees and bushes; for elements $p, q \in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

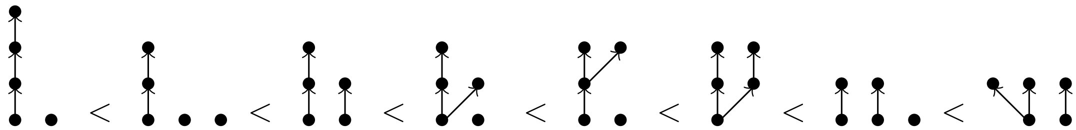
For $p, q \in \mathfrak{E}^{\mathbf{t}}$ or $p, q \in \mathfrak{E}^{\mathbf{b}}$ we will write $p < q$, or $p = o(q)$, if

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

If $\operatorname{cmp}(p) = \operatorname{cmp}(q)$, $\operatorname{hgt}(p) = \operatorname{hgt}(q)$, and $\operatorname{brn}(p) = \operatorname{brn}(q) = m$, write $p = a[p_1]\dots[p_m]$ or $p = [p_1]\dots[p_m]$, with $a \in \mathcal{A}$ and $p_1, \dots, p_m \in \mathfrak{E}^{\mathbf{t}}$, so that $p_1 \ge \dots \ge p_m$, and write $q = b[q_1]\dots[q_m]$ or $q = [q_1]\dots[q_m]$, with $b \in \mathcal{A}$ and $q_1, \dots, q_m \in \mathfrak{E}^{\mathbf{t}}$, so that $q_1 \ge \dots \ge q_m$. Then $p < q$ if there is i such that $p_1 \le q_1, \dots, p_{i-1} \le q_{i-1}$ 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: 

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

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

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

(b) Let $p = o(q)$. (That is, if $q \in \mathfrak{E}^{\mathbf{t}}$ then $p \in \mathfrak{E}^{\mathbf{t}}$ and $p < q$; if $q \in \mathfrak{E}^{\mathbf{b}}$ then $p \in \mathfrak{E}^{\mathbf{b}}$ and $p < q$.) If $\mathrm{cmp}(p) < \mathrm{cmp}(q)$ then $\mathrm{cmp}(pr) < \mathrm{cmp}(qr)$. If $\mathrm{cmp}(p) = \mathrm{cmp}(q)$ and $\mathrm{hgt}(p) > \mathrm{hgt}(q)$, then $\mathrm{cmp}(pr) = \mathrm{cmp}(qr)$ and $\mathrm{hgt}(pr) = \mathrm{hgt}(p) > \mathrm{hgt}(q) = \mathrm{hgt}(qr)$. If $\mathrm{cmp}(p) = \mathrm{cmp}(q)$, $\mathrm{hgt}(p) = \mathrm{hgt}(q)$, and $\mathrm{brn}(p) > \mathrm{brn}(q)$ then $\mathrm{cmp}(pr) = \mathrm{cmp}(qr)$, $\mathrm{hgt}(pr) = \mathrm{hgt}(p) = \mathrm{hgt}(q) = \mathrm{hgt}(qr)$, 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 $\mathrm{cmp}(p) = \mathrm{cmp}(q)$, $\mathrm{hgt}(p) = \mathrm{hgt}(q)$, and $\mathrm{brn}(p) = \mathrm{brn}(q)$ then $\mathrm{cmp}(pr) = \mathrm{cmp}(qr)$, $\mathrm{hgt}(pr) = \mathrm{hgt}(qr)$, 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 $p \in \mathfrak{E}^{\mathbf{b}}$ and $p < [r]$. If $\mathrm{cmp}(p) < \mathrm{cmp}([r])$ then $\mathrm{cmp}(qp) < \mathrm{cmp}(q[r])$ and $qp < q[r]$. Assume that $\mathrm{cmp}(p) = \mathrm{cmp}([r])$, then $\mathrm{cmp}(qp) = \mathrm{cmp}(q[r])$. 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), $o(q)o([r]) = o(o(q)[r]) = o(o(q[r])) = o(q[r])$. ■

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

Proof. Let $p \in \mathfrak{E}^{\mathbf{t}}$, $p \leq qs$. Then $\mathrm{cmp}(pt) \leq \mathrm{cmp}(q[r])$; assume that $\mathrm{cmp}(pt) = \mathrm{cmp}(q[r])$. 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, $\mathrm{cmp}([qs]t) \leq \mathrm{cmp}([q[r]])$, $\mathrm{hgt}([qs]t) \geq \mathrm{hgt}(q) = \mathrm{hgt}([q[r]])$, and $\mathrm{brn}([qs]t) \geq 2 > 1 = \mathrm{brn}([q[r]])$, so $[qs]t < [q[r]]$. By Lemma 7.7(c), $o([qs])t = o([qs]t) < [q[r]]$. ■

7.9. Lemma. *If* $q, r, t \in \mathfrak{E}^{\mathbf{t}}$, $s \in \mathfrak{E}^{\mathbf{b}}$, $r \neq 0$, $\mathrm{hgt}(q) > \mathrm{hgt}(r)$, *and* $\mathrm{cmp}([qs]t) \leq \mathrm{cmp}(q[r])$, *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 7.7(c), $o([qs])t = o([qs]t) < q[r]$. ■

7.10. Lemma. *If* $q, r, t \in \mathfrak{E}^{\mathbf{t}}$, $a \in \mathcal{A}$, $r \neq 0$, $\mathrm{hgt}(q) > \mathrm{hgt}(r)$, *and* $\mathrm{cmp}([q]t) \leq \mathrm{cmp}(a[q[r]])$, *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 $\mathrm{cmp}(t) > 1$ 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 $\mathrm{cmp}(t) = 1$ then $\mathrm{cmp}([q]t) = \mathrm{cmp}(q)+1 < \mathrm{cmp}(q)+\mathrm{cmp}(r)+1 = \mathrm{cmp}(a[q[r]])$, and again $[q]t < a[q[r]]$. By Lemma 7.7(c), $o([q])t = o([q]t) < a[q[r]]$. ■

7.11. Lemma. *If* $q, r, t \in \mathfrak{E}^{\mathbf{t}}$, $s \in \mathfrak{E}^{\mathbf{b}}$, $a \in \mathcal{A}$, $r \neq 0$, $\mathrm{hgt}(q) > \mathrm{hgt}(r)$, $\mathrm{cmp}(s) < \mathrm{cmp}(r)$, *and* $\mathrm{cmp}([qs]t) \leq \mathrm{cmp}(a[q[r]])$, *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 $\mathrm{cmp}(t) > 1$ 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 $\mathrm{cmp}(t) = 1$ then $\mathrm{cmp}([qs]t) = \mathrm{cmp}(q) + \mathrm{cmp}(s) + 1 < \mathrm{cmp}(q) + \mathrm{cmp}(r) + 1 < \mathrm{cmp}(a[q[r]])$, and again $[qs]t < a[q[r]]$. By Lemma 7.7(c), $o([qs])t = o([qs]t) < a[q[r]]$. ■

8. Components of bracket expressions

8.1. We start the proof of Proposition 6.9 by treating first its simplified version. This simplification is achieved by dealing, instead of elements of $\mathfrak{B}$, with a new sort of bracket expressions. Such "new" expressions are obtained from the "old" bracket expressions by treating the bracket mapping $[\cdot]$ as an additive function, that is, by assuming that $[u+v] = [u]+[v]$. This will be done by corresponding to every element u of $\mathfrak{B}$ an unordered list $\mathbf{c}(u)$ of "components of u", which we will write as a formal sum, $\mathbf{c}(u) = p_1 + \ldots + p_k$ with $p_1, \ldots, p_k \in \mathfrak{E}$.

8.2. Let $\mathfrak{S}$ be the set of formal sums of the form $\mathbf{c} = p_1 + \ldots + p_k$ with $p_1, \ldots, p_k \in \mathfrak{E}$; the order of summands in $\mathbf{c}$ is not essential, but no combining of like terms is allowed. For two elements $\mathbf{c} = p_1 + \ldots + p_k$ and $\mathbf{c}_1 = q_1 + \ldots + q_l$ of $\mathfrak{S}$, the sum $\mathbf{c} + \mathbf{c}_1 = p_1 + \ldots + p_m + q_1 + \ldots + q_l \in \mathfrak{S}$ and the product $\mathbf{c}\,\mathbf{c}_1 = p_1q_1 + \ldots + p_1q_l + \ldots + p_mq_1 + \ldots + p_mq_l \in \mathfrak{S}$ are naturally defined. We also define $[\mathbf{c}] = [p_1] + \ldots + [p_m]$.

8.3. Now let $u \in \mathfrak{B}$; we define $\mathbf{c}(u) \in \mathfrak{S}$ in the following way:

if $u=a\in\mathcal{A}$, we put $\mathbf{c}(u)=a$;
if $u\notin\mathcal{A}$ and $u=\sum_{i=1}^{k}a_i[v_{i,1}]\dots[v_{i,m_i}]+\sum_{i=1}^{l}\pm[u_{i,1}]\dots[u_{i,n_i}]$, with $a_i\in\mathcal{A}$ and $u_{i,j},v_{i,j}\in\mathfrak{B}^{\mathbf{t}}$, is the reduced representation of u, we define

$$\mathbf{c}(u)=\sum_{i=1}^{k}a_i[\mathbf{c}(v_{i,1})]\dots[\mathbf{c}(v_{i,m_i})]+\sum_{i=1}^{l}[\mathbf{c}(u_{i,1})]\dots[\mathbf{c}(u_{i,n_i})].$$

When $\mathbf{c}(u)=p_1+\ldots+p_k$, we call $p_1,\ldots,p_k\in\mathfrak{E}$ *the components* of u.

Examples. (1) Let $u=a_1\big[a_2+a_3[a_4][a_5]\big]+a_6[a_7+a_8]-[a_9]\big[a_{10}+a_{11}[a_{12}]\big]$, with $a_i\in\mathcal{A}$. To compute $\mathbf{c}(u)$ we simply "open brackets", and ignore the "$-$" before the $\mathbf{b}$-part of u:

$$\mathbf{c}(u)=a_1[a_2]+a_1\big[a_3[a_4][a_5]\big]+a_6[a_7]+a_6[a_8]+[a_9][a_{10}]+[a_9][a_{11}[a_{12}]]. \tag{8.1}$$

Warning: though the right hand side of (8.1) looks like "a bracket expression", that is, an element of $\mathfrak{B}$, it is just a notation for an unordered list of elements of $\mathfrak{E}$! In what follows, it will be always clear from the context what interpretation of a bracket-like expression is intended.
(2) For $u=\big[a_1+a_2[a_3]\big]-\big[a_1-a_2[a_3]\big]$ with $a_i\in\mathcal{A}$, $\mathbf{c}(u)=[a_1]+[a_2[a_3]]+[a_1]+[-a_2[a_3]]$.

8.4. 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 (1) above,

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

8.5. From Lemma 6.5 one deduces the following:

Lemma. *For any $u_1,u_2,u\in\mathfrak{B}$,*
$\mathbf{c}^{\mathbf{t}}(u_1+u_2)=\mathbf{c}^{\mathbf{t}}(u_1)+\mathbf{c}^{\mathbf{t}}(u_2)$, $\quad\mathbf{c}^{\mathbf{b}}(u_1+u_2)=\mathbf{c}^{\mathbf{b}}(u_1)+\mathbf{c}^{\mathbf{b}}(u_2)$,
$\mathbf{c}^{\mathbf{t}}(u_1u_2)=\mathbf{c}^{\mathbf{t}}(u_1)\,\mathbf{c}^{\mathbf{t}}(u_2)+\mathbf{c}^{\mathbf{t}}(u_1)\,\mathbf{c}^{\mathbf{b}}(u_2)+\mathbf{c}^{\mathbf{b}}(u_1)\,\mathbf{c}^{\mathbf{t}}(u_2)$, $\quad\mathbf{c}^{\mathbf{b}}(u_1u_2)=\mathbf{c}^{\mathbf{b}}(u_1)\,\mathbf{c}^{\mathbf{b}}(u_2)$,
$\mathbf{c}^{\mathbf{t}}([u])=0$ *and, if* $\mathbf{c}^{\mathbf{t}}(u)=p_1+\ldots+p_m$*,* $\mathbf{c}([u])=\mathbf{c}^{\mathbf{b}}([u])=[p_1]+\ldots+[p_m]+\mathbf{c}^{\mathbf{b}}(u)$.

8.6. Given $\mathbf{c}=p_1+\ldots+p_m\in\mathfrak{S}$, we define $\operatorname{cmp}(\mathbf{c})=\max\{\operatorname{cmp}(p_i)\}_{i=1}^{m}$, $\operatorname{hgt}(\mathbf{c})=\min\{\operatorname{hgt}(p_i)\}_{i=1}^{m}$, and $\operatorname{brn}(\mathbf{c})=\min\{\operatorname{brn}(p_i)\}_{i=1}^{m}$.

For $u\in\mathfrak{B}$ we put $\operatorname{cmp}(u)=\operatorname{cmp}(\mathbf{c}(u))$. (This agrees with the definition of cmp given in 1.12.)

8.7. Let $P\in M_d(\mathcal{A})$, $\epsilon=(\epsilon_{i,j})_{1\le i<j\le d}$ be a sign matrix, $\prec$ be a legal order on $\{(i,j)\}_{1\le i<j\le d}$, and let $\varphi_{i,j}^{k,l},\psi_{i,j}^{k,l},\xi_{i,j}^{k,l}\in\mathfrak{B}$ be defined by formulas (6.2). The following lemma can be 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 $\operatorname{cmp}(\varphi_{i,j}^{k,l}),\operatorname{cmp}(\psi_{i,j}^{k,l}),\operatorname{cmp}(\xi_{i,j}^{k,l})\le j-i$.*

8.8. Given $\mathbf{c}=p_1+\ldots+p_m\in\mathfrak{S}$ and $\mathbf{c}_1=q_1+\ldots+q_l\in\mathfrak{S}$, we will write $\mathbf{c}<\mathbf{c}_1$ if $\max\{p_i\}_{i=1}^{m}<\max\{q_j\}_{j=1}^{l}$ and $\mathbf{c}\le\mathbf{c}_1$ if $\max\{p_i\}_{i=1}^{m}\le\max\{q_j\}_{j=1}^{l}$. If $\mathbf{c}<p$ (that is, $p_1,\ldots,p_m<p$), we will also write $\mathbf{c}_1=o(p)$.

9. Tree growing and induction over elementary bracket expressions

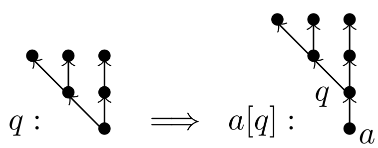
9.1. If $(J_1,\prec_1),\dots,(J_k,\prec_k)$ are linearly ordered sets, then $\big((J_1,\prec_1),\dots,(J_k,\prec_k)\big)$ will stand for the linear order $\prec$ on $J_1\cup\ldots\cup J_k$ which coincides with $\prec_i$ on each J_i and satisfies $J_i\prec J_j$ whenever $i<j$. If J_i is a one-element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

9.2. Given a tree $p\in\mathfrak{E}^{\mathbf{t}}$ we will now construct a matrix $P_p\in M_d(\mathcal{A})$, $d=\mathrm{cmp}(p)+1$, 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 9.4 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 $\mathrm{cmp}(p)=1$, that is, $p=a\in\mathcal{A}$, we define $P_p=\left(\begin{smallmatrix}1 & a\\ & 1\end{smallmatrix}\right)$.

Now let $p\in\mathfrak{E}^{\mathbf{t}}$ with $\mathrm{cmp}(p)>1$ and assume that for all $q\in\mathfrak{E}^{\mathbf{t}}$ with $\mathrm{cmp}(q)<\mathrm{cmp}(p)$ a matrix P_q and an order $\prec_q$ on the set of entries of P_q have been constructed. Let $d=\mathrm{cmp}(p)+1$ and $m=\mathrm{brn}(p)$. Represent $p=a[p_1]\dots[p_m]$ so that $a\in\mathcal{A}$ and $p_1,\dots,p_m\in\mathfrak{E}^{\mathbf{t}}$ 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 $\mathrm{cmp}(q)=d-2$.



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

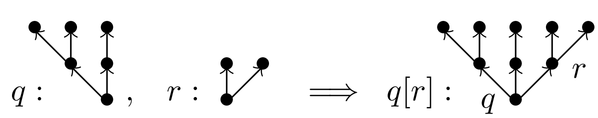
$$P_p=\left(\begin{array}{c|cccc}1 & -a & 0 & \dots & 0\\ \hline & 1 & b_{1,2} & \dots & b_{1,d-1}\\ & & 1 & \dots & b_{2,d-1}\\ & & & \ddots & \vdots\\ & & & & 1\end{array}\right)=\left(\begin{array}{c|c}1 & -a\;0\;\dots\;0\\ \hline & P_q\end{array}\right)\in M_d(\mathcal{A}),$$

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

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

(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 $q=a[p_1]\dots[p_{m-1}]$ and $r=p_m$, then $p=q[r]$ with $\mathrm{hgt}(q)>\mathrm{hgt}(r)$.



Let $d_1 = \mathrm{cmp}(q) + 1$ and $d_2 = \mathrm{cmp}(r) + 1$, then $d = \mathrm{cmp}(p) + 1 = d_1 + d_2 - 1$. 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

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

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.)

9.3. Lemma. *Let $p \in \mathfrak{E}^{\mathbf{t}}$, $d = \mathrm{cmp}(p) + 1$, and $1 < n < d$. 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 $t \in \mathfrak{E}^{\mathbf{t}}$. If $\mathrm{brn}(p) \ge 2$ so that Case 2 takes place, that is, $p = q[r]$ with $q, r \in \mathfrak{E}^{\mathbf{t}}$, $\mathrm{hgt}(q) > \mathrm{hgt}(r)$, $\mathrm{cmp}(q) = d_1$, $\mathrm{cmp}(r) = d_2$, and if $1 \le n < d_2$, then Q is equal to $P_{q[s]}$ and $\prec_p|_Q = \prec_{q[s]}$ for some $s \in \mathfrak{E}^{\mathbf{t}}$.*

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

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

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

$$P_p = \left(\begin{array}{ccccccc} 1 & c_{1,2} & \dots & c_{1,d_2} & 0 & \dots & 0 \\ & 1 & \dots & c_{2,d_2} & 0 & \dots & 0 \\ & & \ddots & \vdots & \vdots & & \vdots \\ & & & 1 & b_{1,2} & \dots & b_{1,d_1} \\ \cline{5-7} & & & & \multicolumn{1}{|c}{1} & \dots & b_{2,d_1} \\ & & & & \multicolumn{1}{|c}{} & \ddots & \vdots \\ & & & & \multicolumn{1}{|c}{} & & 1 \end{array}\right)$$

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

$$P_p = \begin{pmatrix} 1 & c_{1,2} & \dots & c_{1,d_2} & 0 & \dots & 0 \\ & 1 & \dots & c_{2,d_2} & 0 & \dots & 0 \\ & & \ddots & \vdots & \vdots & & \vdots \\ & & & 1 & b_{1,2} & \dots & b_{1,d_1} \\ & & & & 1 & \dots & b_{2,d_1} \\ & & & & & \ddots & \vdots \\ & & & & & & 1 \end{pmatrix}$$

R is a submatrix of P_r and, by induction, $R = P_s$ and $\prec_q|_R = \prec_s$ for some $s \in \mathfrak{E}^{\mathbf{t}}$. 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 $t = q[s]$. ■

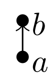
9.4. Proposition. *For any $p \in \mathfrak{E}^{\mathbf{t}}$ one has $\mathbf{c}^{\mathbf{t}}\big(\xi_{1,d}(P_p, \prec_p)\big) = p + o(p)$.*

Proof. We will prove this proposition by induction on $\mathrm{cmp}(p)$. 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 $\mathrm{cmp}(p) = 1$, that is, $p = a \in \mathcal{A}$, we have $P_p = \left(\begin{smallmatrix} 1 & a \\ & 1 \end{smallmatrix}\right)$ and $\xi_{1,2} = \varphi_{1,2} = p$.

$a : \ \bullet a$

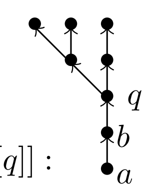
If $\mathrm{cmp}(p) = 2$ then p has form $a[b]$ with $a, b \in \mathcal{A}$.

$a[b]$: 

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 (6.2), $\xi_{1,3} = 0 - \xi_{1,2}^{2,3}[\xi_{2,3}] - \psi_{1,2}^{2,3}\xi_{2,3}$. 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 $p \in \mathfrak{E}^{\mathbf{t}}$, $\mathrm{cmp}(p) \ge 3$; put $d = \mathrm{cmp}(p) + 1$. We consider several cases.

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

$a[b[q]]$: 

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

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

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

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

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

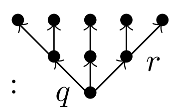
From formulas (6.2),

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

One checks by induction on n that for any $n \in \{3,\ldots,d\}$ and $(k,l) \in I_q$ one has $\varphi^{k,l}_{1,n} = \psi^{k,l}_{1,n} = \xi^{k,l}_{1,n} = 0$. It follows that $\sum_{n=3}^{d-1} \xi^{n,d}_{1,n}[\xi_{n,d}] + \sum_{n=3}^{d-1} \psi^{n,d}_{1,n}\xi_{n,d} = 0$.

We have $\xi^{2,d}_{1,2} = -a$ and $\psi^{2,d}_{1,2} = 0$, 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 7.7(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 $\varphi_{1,d} = -\sum_{n=2}^{d-1} \varphi^{n,d}_{1,n}[\xi_{n,d}] = -\varphi^{2,d}_{1,2}[\xi_{2,d}] = a[\xi_{2,d}] \in \mathfrak{B}^{\mathbf{t}}$, and so, $\mathbf{c}^{\mathbf{b}}(\varphi_{1,d}) = 0$.

Case 2a: $p = q[r]$ where $q, r \in \mathfrak{E}^{\mathbf{t}}$, $\operatorname{cmp}(r) \ge 2$, and $\operatorname{hgt}(q) > \operatorname{hgt}(r)$.

$q[r]$:


Let $d_1 = \operatorname{cmp}(q) + 1$, $d_2 = \operatorname{cmp}(r) + 1$, $P_q = \begin{pmatrix} 1 & b_{1,2} & \ldots & b_{1,d_1} \\ & 1 & \ldots & b_{2,d_1} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{d_1}(\mathcal{A})$, 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

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

Let $I_q = \{(i,j)\}_{d_2\le i<j\le d}$, $I_r = \{(i,j)\}_{1\le i<j\le d_2}$, and $J = \{(i,j)\}_{\substack{1\le i\le d_2-1 \\ d_2+1\le j\le d}}$. 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, in particular, will index the entries of P_q by I_q instead of $\{(i,j)\}_{1\le i<j\le d_1}$. 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 $\varphi^{k,l}_{i,j}(P_p), \psi^{k,l}_{i,j}(P_p), \xi^{k,l}_{i,j}(P_p)$ with $(i,j) \in I_r$ are equal to the corresponding $\varphi^{k,l}_{i,j}(P_r), \psi^{k,l}_{i,j}(P_r)$, $\xi^{k,l}_{i,j}(P_r)$ and the elements $\varphi^{k,l}_{i,j}(P_p), \psi^{k,l}_{i,j}(P_p), \xi^{k,l}_{i,j}(P_p)$ with $(i,j) \in I_q$ are equal to the corresponding $\varphi^{k,l}_{i,j}(P_q), \psi^{k,l}_{i,j}(P_q), \xi^{k,l}_{i,j}(P_q)$. From formulas (6.2) we have

$$\xi_{1,d} = -\sum_{n=d_2+1}^{d-1} \xi^{n,d}_{1,n}[\xi_{n,d}] - \sum_{n=d_2+1}^{d-1} \psi^{n,d}_{1,n}\xi_{n,d} + \sum_{n=2}^{d_2-1} [\xi_{1,n}]\xi_{n,d} - \sum_{n=2}^{d_2-1} \varphi_{1,n}\xi_{n,d} \\ +[\xi_{1,d_2}]\xi_{d_2,d} - \varphi_{1,d_2}\xi_{d_2,d}. \tag{9.1}$$

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 7.7(a), $\mathbf{c}([\xi_{1,d_2}]) = [r] + o([r])$. By Lemma 7.7(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 (9.1) are smaller than p or $[p]$. By Lemma 8.7, the complexity of these terms do not exceed $d - 1 = \mathrm{cmp}(p)$.

We start with the sums $\sum_{n=d_2+1}^{d-1} \xi^{n,d}_{1,n}[\xi_{n,d}]$ and $\sum_{n=d_2+1}^{d-1} \psi^{n,d}_{1,n}\xi_{n,d}$. 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 (6.2) that $\varphi^{k,l}_{i,j} = \xi^{k,l}_{i,j} = c_{i,j}$ and $\psi^{k,l}_{i,j} = 0$. In particular, $\mathrm{cmp}(\varphi^{k,l}_{i,d_2}), \mathrm{cmp}(\psi^{k,l}_{i,d_2}), \mathrm{cmp}(\xi^{k,l}_{i,d_2}) \le 1$ for any $i \in \{1, \ldots, d_2 - 1\}$. 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 $\varphi^{k,l}_{i,j}, \psi^{k,l}_{i,j}, \xi^{k,l}_{i,j}$ with $(i,j) \in J$. By Lemma 8.7 one has $\mathrm{cmp}(\varphi_{n,j}), \mathrm{cmp}(\psi_{n,j}), \mathrm{cmp}(\xi_{n,j}) \le n - j$, and one checks by induction on j that for any $(i,j) \in J$, $\mathrm{cmp}(\varphi^{k,l}_{i,j}), \mathrm{cmp}(\psi^{k,l}_{i,j}), \mathrm{cmp}(\xi^{k,l}_{i,j}) \le j - d_2 + 1$. In particular, for any $n \in \{d_2{+}1, \ldots, d{-}1\}$ one has $\mathrm{cmp}(\xi^{n,d}_{1,n}), \mathrm{cmp}(\psi^{n,d}_{1,n}) \le n{-}d_2{+}1$, and since $d_2 = \mathrm{cmp}(r){+}1 \ge 3$, we obtain $\mathrm{cmp}\big(\xi^{n,d}_{1,n}[\xi_{n,d}]\big), \mathrm{cmp}\big(\psi^{n,d}_{1,n}\xi_{n,d}\big) \le n - d_2 + 1 + d - n < d - 1 = \mathrm{cmp}(p)$.

Now turn to the sums $\sum_{n=2}^{d_2-1} [\xi_{1,n}]\xi_{n,d}$ and $\sum_{n=2}^{d_2-1} \varphi_{1,n}\xi_{n,d}$. Fix $n \in \{2, \ldots, d_2 - 1\}$. By Lemma 9.3 the submatrix of P_p indexed by $\{(i,j)\}_{n \le i < j \le d_2}$ has form $P_{q[s]}$ for some $s \in \mathfrak{E}^{\mathbf{t}}$ with $\mathrm{cmp}(s) < \mathrm{cmp}(r)$, and by induction hypothesis $\mathbf{c}(\xi_{n,d}) = q[s] + o(q[s]) + o([q[s]])$. By Lemma 7.8, $\mathbf{c}([\xi_{1,n}])\,\mathbf{c}^{\mathbf{t}}(\xi_{n,d}) = o(q[r]) = o(p)$ 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 7.8 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 7.9 that $\mathbf{c}^{\mathbf{t}}(\varphi_{1,n})\,\mathbf{c}^{\mathbf{b}}(\xi_{n,d}) = o(q[r])$. By Lemma 8.7, $\mathrm{cmp}(\xi_{n,d}) \le d{-}n$ and $\mathrm{cmp}(\varphi_{1,n}) \le n{-}1$, so $\mathrm{cmp}\big(\mathbf{c}^{\mathbf{t}}(\varphi_{1,n})\,\mathbf{c}^{\mathbf{t}}(\xi_{n,d})\big) \le d{-}n{+}n{-}1{-}1 = d{-}2 < \mathrm{cmp}(p)$. Summarizing, $\mathbf{c}^{\mathbf{t}}(\varphi_{1,n}\xi_{n,d}) = o(q[r]) = o(p)$ and $\mathbf{c}^{\mathbf{b}}(\varphi_{1,n}\xi_{n,d}) = o([q[r]]) = o([p])$.

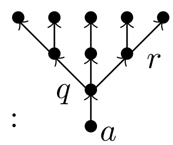
Now consider the term $\varphi_{1,d_2}\xi_{d_2,d}$. Again, we have $\mathrm{cmp}\big(\mathbf{c}^{\mathbf{t}}(\varphi_{1,d_2})\,\mathbf{c}^{\mathbf{t}}(\xi_{d_2,d})\big) < \mathrm{cmp}(p)$. By our induction hypothesis, $\mathbf{c}^{\mathbf{t}}(\xi_{d_2,d}) = q{+}o(q)$, $\mathbf{c}^{\mathbf{b}}(\xi_{d_2,d}) = o([q])$, and $\mathbf{c}^{\mathbf{b}}(\varphi_{1,d_2}) = o([r])$. By Lemma 7.7(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 7.7(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 (6.2),

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

Again, for $n \in \{d_2+1, \ldots, d-1\}$ one has $\operatorname{cmp}(\varphi_{1,n}^{n,d}) \le n - d_2 + 1$ and so, $\operatorname{cmp}(\varphi_{1,n}^{n,d}[\xi_{n,d}]) < d-1 = \operatorname{cmp}(p)$. For $n \in \{2, \ldots, d_2 - 1\}$ one has $\mathbf{c}([\xi_{n,d}]) = [q[s]] + o([q[s]])$, and by Lemma 7.8, $\mathbf{c}^{\mathbf{b}}\big(\alpha_{1,n}[\xi_{n,d}]\big) + \mathbf{c}^{\mathbf{b}}\big([\xi_{1,n}][\xi_{n,d}]\big) = o([q[r]]) = o([p])$. 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 7.7, $\mathbf{c}^{\mathbf{b}}\big(\alpha_{1,n}[\xi_{n,d}]\big) + \mathbf{c}^{\mathbf{b}}\big([\xi_{1,n}][\xi_{n,d}]\big) = [q][r] + o([q][r]) = o([p])$.

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

$a[q[r]]$: 

Let $d_1 = \operatorname{cmp}(q) + 1$, $d_2 = \operatorname{cmp}(r) + 1$, $P_q = \begin{pmatrix} 1 & b_{1,2} & \ldots & b_{1,d_1} \\ & 1 & \ldots & b_{2,d_1} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{d_1}(\mathcal{A})$, 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

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

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 $J = \{(i,j)\}_{\substack{2 \le i \le d_2 \\ d_2+2 \le j \le d}}$. 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 $\prec_p = \big(\big(I_q \setminus \{(d_2+1, d)\}, \prec_q\big), \big(I_r, \prec_r\big), (d_2+1, d), \big(J, \prec_J\big), (1,2), \ldots, (1,d)\big)$.

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 (6.2), for $n \in \{3, \ldots, d\}$ and $1 \le k < l \le d$ with $(k,l) \preceq_p (1,n)$ we have

$$\begin{aligned}
\varphi_{1,n}^{k,l} &= -\sum_{\substack{m=2,\dots,n-1\\(m,n)\prec_p(k,l)}} \varphi_{1,m}^{m,n}[\xi_{m,n}],\\
\psi_{1,n}^{k,l} &= -\sum_{\substack{m=2,\dots,n-1\\(m,n)\prec_p(k,l)}} \psi_{1,m}^{m,n}\big(\xi_{m,n}-[\xi_{m,n}]\big),\\
\xi_{1,n}^{k,l} &= -\sum_{\substack{m=2,\dots,n-1\\(m,n)\prec_p(k,l)}} \big(\xi_{1,m}^{m,n}[\xi_{m,n}]+\psi_{1,m}^{m,n}\xi_{m,n}\big).
\end{aligned} \tag{9.2}$$

So,

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

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

From formulas (6.2),

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

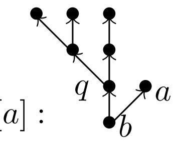
We have $\xi_{1,2}^{2,d}=-a$, $\psi_{1,2}^{2,d}=0$, and by induction hypothesis, $\mathbf{c}(\xi_{2,d})=q[r]+o(q[r])+o([q[r]])$. By Lemma 7.7(b), $\mathbf{c}\big(\xi_{1,2}^{2,d}[\xi_{2,d}]+\xi_{2,d}\psi_{1,2}^{2,d}\big)=a[q[r]]+o(a[q[r]])=p+o(p)$.

We will now show that $\mathbf{c}\big(\sum_{n=3}^{d-1}\xi_{1,n}^{n,d}[\xi_{n,d}]+\sum_{n=3}^{d-1}\psi_{1,n}^{n,d}\xi_{n,d}\big)=o(p)$. First, fix $n\in\{3,\dots,d_2\}$. By Lemma 9.3 and the induction hypothesis, $\mathbf{c}(\xi_{n,d})=q[s]+o(q[s])+o([q[s]])$ for some $s\in\mathfrak{E}^{\mathbf{t}}$ with $\mathrm{cmp}(s)<\mathrm{cmp}(r)$. By Lemma 7.11, $\mathbf{c}(\xi_{1,n}^{n,d})\,\mathbf{c}([\xi_{n,d}])=o(a[q[r]])=o(p)$ and $\mathbf{c}(\psi_{1,n}^{n,d})\,\mathbf{c}^{\mathbf{b}}(\xi_{n,d})=o(a[q[r]])=o(p)$. Also, $\mathrm{cmp}\big(\mathbf{c}(\psi_{1,n}^{n,d})\,\mathbf{c}^{\mathbf{t}}(\xi_{n,d})\big)<\mathrm{cmp}(p)$, so that $\mathbf{c}(\psi_{1,n}^{n,d})\,\mathbf{c}^{\mathbf{t}}(\xi_{n,d})=o(p)$.

Now put $n=d_2+1$. By our induction hypothesis, $\mathbf{c}(\xi_{n,d})=q+o(q)+o([q])$. By Lemma 7.10, $\mathbf{c}\big(\xi_{1,n}^{n,d}[\xi_{n,d}]\big)=o(a[q[r]])=o(p)$ and $\mathbf{c}(\psi_{1,n}^{n,d})\,\mathbf{c}^{\mathbf{b}}(\xi_{n,d}))=o(a[q[r]])=o(p)$. Also, $\mathrm{cmp}\big(\mathbf{c}(\psi_{1,n}^{n,d})\,\mathbf{c}^{\mathbf{t}}(\xi_{n,d})\big)<\mathrm{cmp}(p)$, so that $\mathbf{c}(\psi_{1,n}^{n,d})\,\mathbf{c}^{\mathbf{t}}(\xi_{n,d})=o(p)$.

Finally, let $n\in\{d_2+2,\dots,d-1\}$. In formulas (9.2), 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 $\xi_{1,n}^{k,l}=\psi_{1,n}^{k,l}=0$. In particular, $\xi_{1,n}^{n,d}=\psi_{1,n}^{n,d}=0$. Hence, $\xi_{1,n}^{n,d}[\xi_{n,d}]+\psi_{1,n}^{n,d}\xi_{n,d}=0$.

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

$b[q][a]$: 

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

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

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

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

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 (6.2),

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

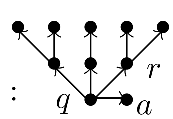
We have checked in Case 1a that $\sum_{n=3}^{d-1} \xi_{1,n}^{n,d}[\xi_{n,d}] + \sum_{n=3}^{d-1} \xi_{n,d}\psi_{1,n}^{n,d} = 0$ and $\mathbf{b}\big(\sum_{n=3}^{d-1} \varphi_{1,n}^{n,d}[\xi_{n,d}]\big) = 0$. Since $\xi_{1,2} = \varphi_{1,2} = a$ and by our induction hypothesis $\mathbf{c}(\xi_{2,d}) = b[q] + o(b[q]) + o([b[q]])$, we obtain by Lemma 7.7(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 $q, r \in \mathfrak{E}^{\mathbf{t}}$ with $\operatorname{hgt}(q) > \operatorname{hgt}(r)$.

$q[r][a]$: 

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

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

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

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 (6.2),

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

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 7.7(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

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

■

10. Conclusion of the proof of Theorem A

10.1. We want to emphasize again that Proposition 6.9 (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 5; we will compute the nilpotency class of certain subgroups of M_d, $d \in \mathbb{N}$.

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

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

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

10.2. Let $\mathcal{T} \subseteq \{(i,j) : 1 \le i < j \le d\}$ 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 $k_1,\dots,k_{m-1} \in \{1,\dots,d\}$ with $(i,k_1),(k_1,k_2),\dots,(k_{m-2},k_{m-1}),(k_{m-1},j) \in \mathcal{T}$. We also define $\mathrm{step}(\mathcal{T}) = \max\{\mathrm{step}_{\mathcal{T}}(i,j) : (i,j) \in \mathcal{T}\}$.

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

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 $(M_{\mathcal{T}})_1 = M_{\mathcal{T}}$, $(M_{\mathcal{T}})_m = [(M_{\mathcal{T}})_{m-1}, M_{\mathcal{T}}]$, $m = 2,3,\dots$, of $M_{\mathcal{T}}$ is generated by $\{E_{i,j} : (i,j) \in \mathcal{T},\ \mathrm{step}_{\mathcal{T}}(i,j) = m\}$. ■

10.3. We will say that an order $\prec$ on the elements of a transitive set $\mathcal{T}$ is *legal* if $(i,j),(j,k) \prec (i,k)$ whenever $(i,j),(j,k),(i,k) \in \mathcal{T}$. This definitions agrees with the definition of a legal order given in 5.7 for the case $\mathcal{T} = \{(i,j) : 1 \le i < j \le d\}$.

10.4. We return to the bracket algebra $\mathfrak{B}$ over a commutative ring $\mathcal{A}$. For $D \in \mathbb{N}$ we define

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

and

$$\mathfrak{M} = \{ u \in \mathfrak{B}^{\mathbf{t}} : u \in \mathfrak{M}_D \text{ with } D = \mathrm{cmp}(u) \}.$$

We will prove the following enhancement of Proposition 6.9, which implies Theorem A_1^{**}:

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

10.5. Lemma. *If $u \in \mathfrak{M}_D$ then $-u \in \mathfrak{M}_D$. 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 $\mathcal{T} \subseteq \big\{(i,j) : 1 \leq i < j \leq d\big\}$ with $(1,d) \in \mathcal{T}$ and $\mathrm{step}(\mathcal{T}) \leq D$, a matrix $P = \begin{pmatrix} 1 & a_{1,2} & a_{1,3} & \dots & a_{1,d} \\ & 1 & a_{2,3} & \dots & a_{2,d} \\ & & \ddots & & \vdots \\ & & & 1 & a_{d-1,d} \\ & & & & 1 \end{pmatrix} \in M_{\mathcal{T}}(\mathcal{A})$, a sign matrix $\epsilon = (\epsilon_{i,j})_{1 \leq i < j \leq d}$, and a legal order $\prec$ on $\mathcal{T}$. Define $P' = \begin{pmatrix} 1 & -a_{1,2} & -a_{1,3} & \dots & -a_{1,d} \\ & 1 & a_{2,3} & \dots & a_{2,d} \\ & & \ddots & & \vdots \\ & & & 1 & a_{d-1,d} \\ & & & & 1 \end{pmatrix} \in M_{\mathcal{T}}(\mathcal{A})$ and a sign matrix $\epsilon' = (\epsilon'_{i,j})_{1 \leq i < j \leq d}$ by $\epsilon'_{i,j} = \begin{cases} -\epsilon_{i,j} & \text{if } i = 1 \\ \epsilon_{i,j} & \text{otherwise,} \end{cases}$ $1 \leq i < j \leq d$. Using the identity $[-u]^{-1} = -[u]^1$ one checks from formulas (6.2) by induction on (i,j) that

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

and

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

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 $-u \in \mathfrak{M}_D$. ■

10.6. Lemma. *If $u \in \mathfrak{M}_{D_1}$ and $v \in \mathfrak{M}_{D_2}$ then $u + v \in \mathfrak{M}_{\max\{D_1, D_2\}}$. 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 $\mathcal{T}_1 \subseteq \big\{(i,j) : 1 \leq i < j \leq d_1\big\}$ with $(1, d_1) \in \mathcal{T}_1$ and $\mathrm{step}(\mathcal{T}_1) \leq D_1$, 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 a legal order $\prec_1$ on $\mathcal{T}_1$, and let $v = \big(\xi_{1,d_2}(Q, \epsilon_2, \prec_2)\big)$ for $d_2 \in \mathbb{N}$, a transitive set $\mathcal{T}_2 \subseteq \big\{(i,j) : 1 \leq i < j \leq d_2\big\}$ with $(1, d_2) \in \mathcal{T}_2$ and $\mathrm{step}(\mathcal{T}_2) \leq D_2$, 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 a legal order $\prec_2$ on $\mathcal{T}_2$. Put $d = d_1 + d_2$

and define

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

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 $(1,d) \in \mathcal{T}$, $\operatorname{step}(\mathcal{T}) = \max\{\operatorname{step}(\mathcal{T}_1), \operatorname{step}(\mathcal{T}_2)\} \le \max\{D_1, D_2\}$, 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 $\mathcal{T}$ 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 (6.2), 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 precisely, 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, $u + v = \mathbf{t}\big(\xi_{1,d}(P)\big) \in \mathfrak{M}_{\max\{D_1,D_2\}}$. ■

10.7. Lemma. *If $u \in \mathfrak{M}_{D_1}$ and $v \in \mathfrak{M}_{D_2}$ then $[u]v + u[v] - uv \in \mathfrak{M}_{D_1+D_2}$. 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 $\mathcal{T}_1 \subseteq \{(i,j) : 1 \le i < j \le d_1\}$ with $(1,d_1) \in \mathcal{T}_1$ and $\operatorname{step}(\mathcal{T}_1) \le D_1$, a matrix $R = \begin{pmatrix} 1 & a_{1,2} & \cdots & a_{1,d_1} \\ & 1 & \cdots & a_{2,d_1} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{\mathcal{T}_1}(\mathcal{A})$, a sign matrix ϵ_1, and a legal order $\prec_1$ on $\mathcal{T}_1$, 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 $\mathcal{T}_2 \subseteq \{(i,j) : 1 \le i < j \le d_2\}$ with $(1,d_2) \in \mathcal{T}_2$ and $\operatorname{step}(\mathcal{T}_2) \le D_2$, a matrix $S = \begin{pmatrix} 1 & b_{1,2} & \cdots & b_{1,d_2} \\ & 1 & \cdots & b_{2,d_2} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \in M_{\mathcal{T}_2}(\mathcal{A})$, a sign matrix ϵ_2, and a legal order $\prec_2$ on $\mathcal{T}_2$. 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 $(1,d) \in \mathcal{T}$, $\operatorname{step}(\mathcal{T}) \le \operatorname{step}(\mathcal{T}_1) + \operatorname{step}(\mathcal{T}_2) \le D_1 + D_2$, 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

$$\Big(\big(\mathcal{T}_1 \setminus \{(1,d_1)\}, \prec_1\big), \big(\mathcal{T}_2 \setminus \{(d_1,d)\}, \prec_2\big), (1,d_1), (d_1,d), \big(\mathcal{T} \cap J, \prec_J\big)\Big)$$

and $\prec'$ to be

$$\Big(\big(\mathcal{T}_1 \setminus \{(1,d_1)\}, \prec_1\big), \big(\mathcal{T}_2 \setminus \{(d_1,d)\}, \prec_2\big), (d_1,d), (1,d_1), \big(\mathcal{T} \cap J, \prec_J\big)\Big).$$

(In plain words, first the entries of R excluding a_{1,d_1} appear, then the entries of S excluding b_{1,d_2} follow, then a_{1,d_1} and b_{1,d_2} follow, 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 (6.2) we now have

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

Hence,

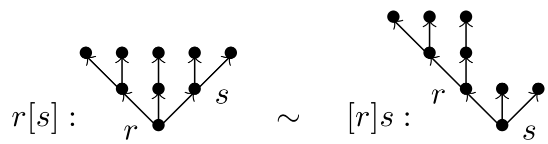
$$\begin{aligned} \mathbf{t}(\xi_{1,d}(P,\prec)) - \mathbf{t}(\xi_{1,d}(P,\prec')) &= \mathbf{t}\big(\xi_{1,d}(P,\prec) - \xi_{1,d}(P,\prec')\big) \\ &= [\mathbf{t}(\xi_{1,d_1})]\,\mathbf{t}(\xi_{d_1,d}) + \mathbf{b}(\xi_{1,d_1})\,\mathbf{t}(\xi_{d_1,d}) + \mathbf{t}(\xi_{1,d_1})[\mathbf{t}(\xi_{d_1,d})]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

Since $\mathbf{t}(\xi_{1,d}(P,\prec)), \mathbf{t}(\xi_{1,d}(P,\prec')) \in \mathfrak{M}_{D_1+D_2}$, by Lemmas 10.5 and 10.6, $[u]v + u[v] - uv = \mathbf{t}(\xi_{1,d}(P,\prec)) - \mathbf{t}(\xi_{1,d}(P,\prec')) \in \mathfrak{M}_{D_1+D_2}$. ■

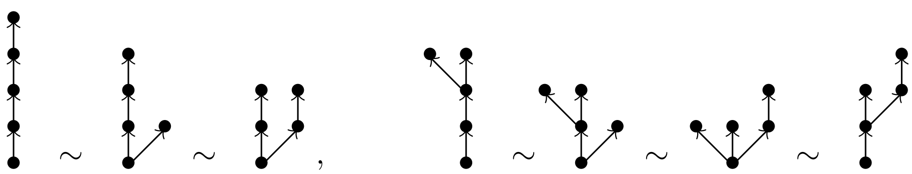
10.8. From Lemmas 10.7, 10.5, and 10.6 we derive the following:

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

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



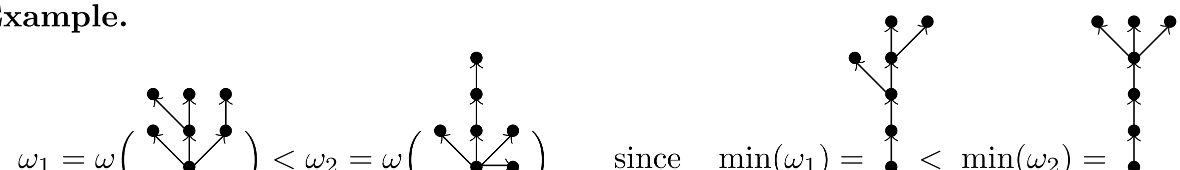
Examples.



10.10. 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* $q_1 \sim p_1, \ldots, q_m \sim p_m$.

10.11. Let Ω be the set of equivalence classes for $\sim$. 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.



For $p \in \mathfrak{E}^{\mathbf{t}}$ we denote by $\omega(p)$ the class in Ω that contains p, and for $u \in \mathfrak{B}^{\mathbf{t}}$ let $\omega(u) = \max\{\omega(p) : p \in \mathbf{c}(u)\}$.

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

The minimal-depth-of-a-plus, "dop", of elements of $\mathfrak{B}^{\mathbf{t}} \setminus \mathfrak{E}^{\mathbf{t}}$ is defined by the following rules: $\operatorname{dop}(u_1 + u_2) = 0$ for $u_1, u_2 \in \mathfrak{B}^{\mathbf{t}}$ and $\operatorname{dop}(a[v_1]\ldots[v_m]) = 1 + \min\{\operatorname{dop}(v_1), \ldots, \operatorname{dop}(v_m)\}$ 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 $\operatorname{nop}(u) = 1$ and $\operatorname{dop}(u) = 3$, for $v = a_1[a_2[a_3[a_4 + a_5]][a_6 + a_7]$ one has $\operatorname{nop}(v) = 2$ and $\operatorname{dop}(v) = 1$.

10.13. 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 9.4 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 $u \in \mathfrak{E}^{\mathbf{t}}$ 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 10.5 and 10.6. 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 $\mathrm{cmp}(v_1), \mathrm{cmp}(v), \mathrm{cmp}(v_1 v) < \mathrm{cmp}(u)$, by our induction hypothesis we have $v_1, v, v_1 v \in \mathfrak{M}$. Thus by Lemma 10.8, $u' = v_1[v] \in \mathfrak{M}$. By Lemma 10.10, $\mathbf{c}(u') = q + q_1 + \ldots + q_k$ 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 $r, s \in \mathfrak{E}^{\mathbf{t}}$, then, since by the induction hypothesis $r, s, rs \in \mathfrak{M}$, Lemma 10.8 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 $u \in \mathfrak{E}^{\mathbf{t}}$, 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 10.6, $u \in \mathfrak{M}$. 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 $\mathrm{cmp}(v_1), \mathrm{cmp}(v), \mathrm{cmp}(v_1 v) < \mathrm{cmp}(u)$, by our induction hypothesis $v_1, v, v_1 v \in \mathfrak{M}$. By Lemma 10.8, $u \in \mathfrak{M}$. ∎

Bibliography

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

[B] V. Bergelson, Minimal idempotents and ergodic Ramsey theory, *Topics in Dynamics and Ergodic Theory*, 8–39, London Math. Soc. Lecture Note Series **310**, Cambridge Univ. Press, Cambridge, 2003.

[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 Knutson 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.

[KN] Kuipers and Niederreiter, *Uniform Distribution of Sequences*, John Wiley, New York, 1974.

[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, *Ergodic Theory and Dynamical Systems* **26** (2006), 787-798.

[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.

[Lo] G. G. Lorentz, A contribution to the theory of divergent sequences, *Acta Math.* **80** (1948), 167-190.

[Mal] A. Malcev, On a class of homogeneous spaces, *Amer. Math. Soc. Transl.* **9** (1962), 276-307.

[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.

[Su] L. Sucheston, Banach limits, *Amer. Math. Monthly* **74** (1967), 308-311.

[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.