Markov Field Types and Tilings

Yuliy Baryshnikov
Dept. Math. & Electrical Eng.
Univ. Illinois at Urbana-Champaign
Email: ymb@illinois.edu

Jaroslaw Duda
Center for Science of Information
Purdue University
Email: dudaj@purdue.edu

Wojciech Szpankowski
Department of Computer Science
Purdue University
West Lafayette, IN, USA
Email: spa@cs.purdue.edu

***Abstract*—The method of types is one of the most popular technique in information theory and combinatorics. However, it was never thoroughly studied for Markov fields. Markov fields can be viewed as models for systems involving a large number of variables with local dependencies and interactions. These local dependencies can be captured by a shape of interactions (locations that contribute the next probability transition). Shapes marked by symbols from a finite alphabet are called tiles. Two assignments in a Markov filed have the same type if they have the same empirical distribution or they can be tiled by the same number of *tile types*. Our goal is to study the growth of the number of Markov field types or the number of tile types. This intricate and important problem was left open for too long.**

I. Introduction

Method of types [3], [4], [6], [13], [17] is used in myriad of applications from the minimax redundancy [6] to simulation of information sources [10]. However, thus far it was mostly (if not exclusively) studied for *one-dimensional* Markov processes [7] and general one-dimensional stationary ergodic processes [13]. Here we investigate types of *Markov fields* [2], [16] that find applications ranging from sensor networks [12], to image processing, to information retrieval.

In order to gently introduce Markov fields and their types, we start with a one-dimensional Markov chain over a finite alphabet $\mathcal{A} = \{1, 2, \dots, m\}$. We shall follow notation and the combinatorial approach introduced in [7]. Let us write $x^n = x_1 \dots x_n \in \mathcal{A}^n$ for a sequence of length n generated by a Markov source. For Markov sources of order $r = 1$ we have two equivalent representations for the probability $P(x^n)$:

$$P(x^n) = P(x_1) \prod_{i=2}^{n} P(x_i|x_{i-1}) = P(x_1) \prod_{t \in T} p_t^{k(t)} \quad (1)$$

with $\sum_{x_1} P(x_1) = 1$ where $t = (ij) \in \mathcal{A}^2 =: T$ and $p_{ij} = p_t$ is the transition probability from $i \in \mathcal{A}$ to $j \in \mathcal{A}$. The frequency vector $k(t) := k(x^n, t)$ counts the number of pairs $t = (ij)$ in the sequence x^n. Similarly, for one-dimensional Markov sources of order r we define $t = (j_1, \dots, j_{r+1}) \in \mathcal{A}^{r+1} =: T$ as the $(r+1)$ tuples of the underlying Markov source and $k(t)$ as the number of t in x^n.

Two interesting questions arise: For a given vector count $\{k(t)\}_{t \in T}$ how many sequences x^n realize it (i.e., have the same $\{k(t)\}_{t \in T}$), and how many distinct vector counts $\{k(t)\}_{t \in T}$ (i.e., distinct empirical distributions) are there? In the language of Markov sources, we identify the vector count $\{k(t)\}_{t \in T}$ as a *Markov type*. The number of sequences of a given Markov type was first addressed by Whittle [17] and then re-established by analytic method in [6]. A precise evaluation of the number of Markov types was left open until it was recently discussed in [7] (see also [9] for tree models).

In this paper, we consider Markov fields [2] and count the number of distinct (empirical) distributions, that is, the number of *Markov field types*. We start our discussion with some general remarks and definitions. We shall follow notation from [2]. Throughout we assume that the underlying process, taking values in a finite alphabet $\mathcal{A} = \{1, 2, \dots, m\}$, is defined on a finite d dimensional box $\mathcal{I}_N = I_{n_1} \times I_{n_2} \times .. \times I_{n_d}$, with $I_n := \{0, 1, .., n-1\}$ and $N := n_1 \cdot n_2 \cdot \dots \cdot n_d$. Elements of $\mathcal{I}_N$ are called *locations*. A random field on $\mathcal{I}_N$ is a collection of random variables $X = \{X(\ell)\}_{\ell \in \mathcal{I}_N}$ with values in $\mathcal{A}^N$. We also write $X(S) = \{X(\ell),\ \ell \in S\}$ for some $S \subset \mathcal{I}_N$.

There are many ways to define Markov fields. We adopt here the so called "unilateral scanning" approach that is popular in some applications such as image processing [2]. We now postulate that the underlying d dimensional process X visits locations $\ell_1, \dots, \ell_N$ in that order taking values $X(\ell_j)$. The *Markov property* implies that the probability $P(x(\ell_1), \dots, x(\ell_N)) := P(X(\ell_1) = x(\ell_1), \dots, X(\ell_N) = x(\ell_N))$ depends *only* on the local past visits (see Figure 1). More precisely, let $S \subset \mathbb{Z}^d$, where $\mathbb{Z}$ is the set of integers, be a shape, and also $S_{(j)} = S \setminus \{\ell_j\}$. For the homogeneous Markov field X we require that

$$P(X(\ell_j) = x(\ell_j) | X(\ell_{j-1}) = x(\ell_{j-1}), \dots X(\ell_1) = x(\ell_1)) = \quad (2)$$
$$P(X(\ell_j) = x(\ell_j) | X(S_{(j)}) = x(S_{(j)})).$$

Let us now define a tile $t :\ S \to \mathcal{A}$ as marked shape S with symbols from the alphabet $\mathcal{A}$. Notice that the set of all tiles T has the cardinality $D := |\mathcal{A}|^{|S|} = m^{|S|}$. By (2), we can write the joint probability $P(x^N)$ as [2] (provided shapes are translational invariant)

$$P(x^N) = P(x_0) \prod_{j=1}^{N} P(x(\ell_j) | x(S_{(j)}) = P(x_0) \prod_{t \in T} p_t^{k(t)} \quad (3)$$

where $k(t)$ counts the number of tiles t which can be viewed as values of the process $(x(\ell_j), x(S_{(j)})$ on $\mathcal{A}^S$. We should point out that (3) can be even better justified by appealing to Hammersley-Clifford Theorem [11] and Gibbs distribution. However, we leave this for the journal version of the paper.

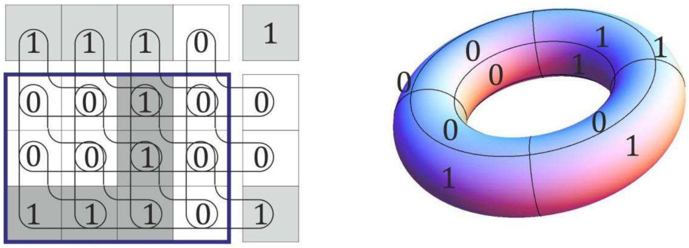

Figure 1. Box of size 3×4 and its corresponding torus over $\mathcal{A} = \{0, 1\}$.

Finally, it is now easy to see that the number of (empirical) distributions is fully characterized by the count vector $\{k(t),\ t \in T\}$ and the initial probability. The latter is ignored in the cyclic representation of the underlying Markov field discussed next.

We first re-formulate our question in terms of *counting tiles* [1]. For dimension $d \in \mathbb{N}$ and $\mathbf{n} = (n_1, n_2, .., n_d)$ define the *torus*

$$\mathcal{O}_{\mathbf{n}} = I_{n_1} \times I_{n_2} \times .. \times I_{n_d} \subset \mathbb{Z}^d, \quad N := n_1 \cdot n_2 \cdot ... \cdot n_d$$

which is a *cyclic* multidimensional box with periodic boundary conditions, or equivalently as $\mathbb{Z}^d$ modulo $(n_1, n_2, .., n_d)$. For $d = 2$ in Figure 1 we show the box $\mathcal{I}_{3\times 4}$ (with L shape S of dependency) and its corresponding cyclic representation.

In this conference version we mostly deal with *cyclic fields* (see Figure 1) that can be viewed as functions from torus $\mathcal{O}_{\mathbf{n}}$ to $\mathcal{A}$, that is, the underlying Markov field is defined on

$$\mathcal{X}^{\mathbf{n}} := \{x^{\mathbf{n}} : \mathcal{O}_{\mathbf{n}} \to \mathcal{A}\}.$$

Our approach through tilings has the advantage of allowing us to introduce a general formulation of Markov field types. We first recall that for one-dimensional Markov chains, we study correlations between consecutive positions represented by $t = (ij)$, as in (1). For Markov fields defined on $\mathcal{O}_{\mathbf{n}}$, and a general *shape* $S \subset \mathbb{Z}^d$, we replace the pairs (ij) of the one-dimensional case by tiles t which are marked shapes by symbols from $\mathcal{A}$. Now, the tile count $k(t)$ is a function $k : T \to \mathbb{N}$ enumerating the number of t occurrences in the underlying field, that is,

$$k(t) \equiv k_S(t) = |\{\mathbf{s} \in \mathcal{O}_{\mathbf{n}}\ :\ x|_{S+\mathbf{s}} = t\}| \tag{4}$$

where the tile t is properly shifted by $\mathbf{s}$ and $f|_B$ denotes a function restricted to a smaller domain B.

Our goal is to count the number of different ways to tile the torus $\mathcal{O}_{\mathbf{n}}$ or the box $\mathcal{I}_{\mathbf{n}}$. In other words, we would like to partition $\mathcal{X}^{\mathbf{n}}$ into subsets of the same types and count the number of these subsets, not the number of fields of a given type. While tilings and counting them are discussed in many references [1], [8], our problem is distinctly different and we couldn't find any relevant literature. Counting tilings usually means to enumerate the number of different tilings of all types. Here, we count the number of distinct (tiling, vectors) *types* $\{k(t)\}_{t\in T}$, that is, we are interested in the cardinality of

$$\mathcal{P}_{\mathbf{n}}(m, S) = \{\mathbf{k} :\ \exists_{x^{\mathbf{n}} \in \mathcal{X}^{\mathbf{n}}}\ x^{\mathbf{n}} \text{ is of type } \ \mathbf{k}\} \tag{5}$$

where $\mathbf{k} = \{k(t),\ t \in T\}$ is the count vector.

We now briefly discuss our main findings. We shall view the set of frequency counts $\{k(t)\}_{t\in T}$ as a $D := |T| = m^{|S|}$–dimensional vector $\mathbf{k}$ indexed by $t \in T$. Clearly, $k(t) \geq 0$ for all $t \in T$, however, this vector satisfies some additional constraints that have major impact on the cardinality of $\mathcal{P}_{\mathbf{n}}(m, S)$. First of all, the normalization condition

$$\sum_{t\in T} k(t) = N := n_1 \cdot \ldots \cdot n_d \tag{6}$$

is quite obvious for the torus. Moreover, in order to tile a torus the number of tiles "ending" with a subtile $t' : S' \to \mathcal{A}$ for some subshape $S' \subset S$ must be equal to the number of tiles that "begin" with t'. This leads to the *conservation law*

$$\forall_{S' \subset S} \forall_{t' : S' \to \mathcal{A}} \quad k_{S'}(t') - k_{S'+\mathbf{s}}(t') = 0 \tag{7}$$

where t' is properly shifted by $\mathbf{s} \in \mathbb{Z}^d$ subject to $(S' + \mathbf{s}) \subset S$ (see also (8)). The system of equations (7) and the normalization equation (6) over $\mathbf{k} \in \mathbb{N}^D$ constitutes a *linear system of Diophantine equations*. We denote by $\mathcal{F}_{\mathbf{n}} := \mathcal{F}_{\mathbf{n}}(m, S)$ the set of nonnegative integer solutions to (6)-(7). Clearly, $|\mathcal{P}_{\mathbf{n}}| \leq |\mathcal{F}_{\mathbf{n}}|$ since all $\mathbf{k} \in \mathcal{P}_{\mathbf{n}}$ lead to a realizable (periodic) tiling. In the one-dimension case we used analytic approach to enumerate precisely $\mathcal{F}_{\mathbf{n}}$ (see also [14]). Furthermore, for $d = 1$ we show in [7] that $|\mathcal{P}_n| \sim |\mathcal{F}_n|$, however, this does not hold any longer for the multidimensional case where the set of types $\mathcal{P}_{\mathbf{n}}$ is a *proper* subset of $\mathcal{F}_{\mathbf{n}}$.

To analyze the cardinality of $\mathcal{F}_{\mathbf{n}}$ and ultimately $\mathcal{P}_{\mathbf{n}}$ we need to understand the geometry of the D-dimensional count vectors $\mathbf{k}$ as illustrated in Figure 2. In particular, we need to estimate the dimensionality of a subspace on which $\mathcal{F}_{\mathbf{n}}$ and $\mathcal{P}_{\mathbf{n}}$ reside. To accomplish this we shall write the conservation law as $C \cdot \mathbf{k} = \mathbf{0}$ where C is a matrix of coefficients of the conservation laws (7). This allows us to define the cone $\mathscr{C}$ as:

$$\mathscr{C} \equiv \mathscr{C}(m, S) \quad = \quad \{\mathbf{k} \in \mathbb{N}^D\ :\ C \cdot \mathbf{k} = \mathbf{0}\}.$$

(Recall that a set $\mathscr{C}$ is a cone if $\mathbf{k} \in \mathscr{C}$ implies $\lambda\mathbf{k} \in \mathscr{C}$ for any $\lambda > 0$.) We also define $\mathcal{F}_N \equiv \mathcal{F}_N(m, S) = \{\mathbf{k} \in \mathscr{C} : \sum_i k_i = N\}$.

The matrix C is hugely over-determined. In fact, we prove that $\mathcal{F}_{\mathbf{n}}$ lies on a subspace of dimensionality $\mu = D - 1 - \mathrm{rk}(C)$ where $\mathrm{rk}(C)$ is the rank of C (see Theorem 1). For example, for $d = 2$ and a 2×2 square shape we have $\mu = m^4 - 2m^2 + m$, while for a 3×2 rectangular shape we find $\mu = m^6 - m^4 - m^3 + m^2$.

Our goal, however, is to estimate the cardinality of the number of types $\mathcal{P}_{\mathbf{n}}$, that is, the number of *realizable* tiling types of $\mathcal{X}^{\mathbf{n}}$ or the number of distinct count vectors $\mathbf{k}$. In other words, we need to evaluate the number of lattice points in $\mathcal{P}_{\mathbf{n}}$. We shall see that the closure of the *normalized* set $\hat{\mathcal{P}}_N := \frac{1}{N} \bigcup_{\prod_i n_i = N} \mathcal{P}_{\mathbf{n}}$ is dense in $\hat{\mathcal{F}}_N := \mathcal{F}_N / N$ (see

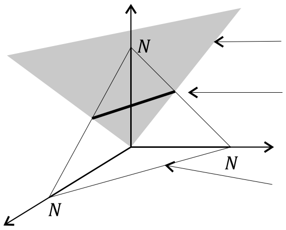

Figure 2. Schematic general picture in $D = 3$ dimensions.

Lemma 2) leading to our main result $|\mathcal{P}_{\mathbf{n}}(m, S)| = \Theta(N^{\mu})$ (see Theorem 4). However, unlike $d = 1$ where we proved $|\mathcal{P}_{\mathbf{n}}| \sim |\mathcal{F}_{\mathbf{n}}|$ in the multidimensional case $\mathcal{F}_{\mathbf{n}}$ is not asymptotically equivalent to $\mathcal{P}_{\mathbf{n}}$ even if the growth of both is the same. We also show in Theorem 5 that the number of types in the box $\mathcal{I}_{\mathbf{n}}$ is much larger, namely $\Theta(N^{D-1}/(\min_i\ n_i)^{\mathrm{rk}(C)})$.

To establish these findings we use tools of discrete, convex, and analytic multidimensional geometry that somewhat resembles the method discussed in [14]. In particular, we apply Ehrhart Theorem [5] to count the number of lattice points in a polytope. This allows us to find the number of nonnegative integer solutions of a linear system of Diophantine equations (i.e., conservation laws) that leads to the enumeration of the Markov field types.

II. Main Results

A. Basic Definitions and Examples

We illustrate our definitions from the introduction with an example.

Example 1: *Markov Field in $d = 2$ with the L-Shape.*
We deal here with $n_1 \times n_2$ rectangle with cyclic boundary conditions: $x_{i,j} = x_{i+n_1,j} = x_{i,j+n_2}$. Let us take the 3×4 torus $\mathcal{O}_{\{3,4\}} = \{0,1,2\} \times \{0,1,2,3\}$. For example, consider the following field over $\mathcal{A} = \{1,2\}$

$$x = \begin{pmatrix} 1121 \\ 1121 \\ 2221 \end{pmatrix}.$$

Because of the cyclic condition we have $x(4,0) = x(0,3) = x(4,3) = x(0,0)$. For the 3 point "L"-like shape: $S = \{(0,0),(0,1),(1,0)\}$, we find $k\begin{pmatrix} 1 \\ 12 \end{pmatrix} = 2$ since this pattern appears at $\mathbf{s} \in \{(3,0),(1,1)\}$ positions. □

We aim at finding the number of (cyclic) Markov field types that is also the number of "periodic" tilings of the underlying torus $\mathcal{O}_{\mathbf{n}}$. Observe that the count vectors $\mathbf{k} \in \mathcal{P}_{\mathbf{n}} := \mathcal{P}_{\mathbf{n}}(m, S)$ satisfy some additional constraints such as the normalization equation (6) and the conservation laws (7) that we discuss next. This will contribute to the reduction of the dimensionality of the space on which $\mathcal{P}_{\mathbf{n}}$ resides.

B. Conservation Laws

Let us start with a new definition. We say that $k_{S'}(t')$ is a restriction of $k_S(t)$ to a smaller subshape $S' \subset S$ if

$$k_{S'}(t') = \sum_{t'':(S \setminus S') \to \mathcal{A}} k_S(t' \cup t'') \qquad t' : S' \to \mathcal{A} \tag{8}$$

where $t' \cup t''$ is concatenation of these tiles.

We now proceed to describe the conservation laws that play crucial role in determining the asymptotics of $|\mathcal{P}_{\mathbf{n}}| := |\mathcal{P}_{\mathbf{n}}(S, m)|$. Observe that a subshape $S' \subset S$ may be translated to a few positions in S; e.g., a single subshape can appear in all $|S|$ positions (but perhaps having different subtile t'). Since the $k(t)$ function is obtained in translationally invariant way, all restrictions to subshapes differing only by a translation have to be identical. We will call the set of all these constrains the *conservation laws* as expressed in (7). As discussed in the introduction, (7) follows from the fact that the number of tiles "ending" with a subtile t' of subshape $S' \subset S$ must be equal to the number of tiles that "begin" with t'. If we treat $k(t)$ as a $D = m^{|S|}$ dimensional vector $\mathbf{k}$, then for any $(S', \mathbf{s}, t')$ triple, we have a single *linear equation* in (7). Let us denote by $C^*(\{(S', \mathbf{s}, t')\})$ the corresponding $1 \times D$ single row of coefficients of a much larger matrix C^*. Thus $k_{S'}(t') - k_{S'+\mathbf{s}}(t') = 0$ can be written as $C^*(\{(S', \mathbf{s}, t')\}) \cdot \mathbf{k} = 0$.

It is not difficult to see that the matrix C^* of the coefficients corresponding to all conservation laws is hugely over determined with many dependent rows. Our goal is to find matrix $C \equiv C_m(S)$ fulfilling all conservation laws that can be written as $C \cdot \mathbf{k} = \mathbf{0}$. We denote rank of C as $\mathrm{rk}(C)$ which plays major role in our analysis.

In fact, we aim at finding matrix C with independent rows. There are three ways to remove dependencies of C:

1. The normalization equation $\sum_{t':S' \to \mathcal{A}} k_{S'}(t') = N$ eliminates for every S' and $\mathbf{s}$ one equation since summing (7) over all t' we obtain the trivial equation. Thus for every S' and $\mathbf{s}$ we can remove the equation with t' using only the last symbol $m \in \mathcal{A}$ (i.e., t' being constant function $t' = m$).
2. Observe that for a given $S' \subset S$ the conservation laws (7) contain equations between all of its shifted positions $S' + \mathbf{s}$. We can instead choose some fixed position of S' and use equalities only with this position. To accomplish this, let us denote by $\mathcal{S}^0$ the set of all nonempty subsets of S, but having only a single representation of $S' \equiv S' + \mathbf{s}$, which we can formally write as: $\mathcal{S}^0$ is a maximal nonempty set of subsets of S such that $\neg\exists_{S',S'' \in \mathcal{S}^0,\ \mathbf{s} \neq \mathbf{0}}\ S' = S'' + \mathbf{s}$. We will only use these subsets in C.
3. These two reductions are sufficient for small shapes S. However, for larger shapes like 2×2 squares, we also need another reduction that turns out to be the final one. Let $\ell \in \mathbb{Z}^d \setminus S'$ be a position in $\mathcal{O}_{\mathbf{n}}$ for some subshape S'. If the value of the tile t at this position is $i \in \mathcal{A}$, that is, $t(\ell) = i$, then we denote it as $\ell(i)$. Then for $t' : S' \to \mathcal{A}$ we can write

the restriction from $S' \cup \ell$ to S' as:

$$k(t') = k(t' \cup \ell(m)) + \sum_{i=1,..,m-1} k(t' \cup \ell(i)). \quad (9)$$

This formula allows to express (7) for $S' \cup \{\ell\}$ subshape by disregarding symbol m, but requiring also to have all conditions for smaller subshape S'. Using this formula multiple times, we see that by restricting (7) to the smaller alphabet $\{1,..,m-1\}$, we still can deduce all conditions on S' (on complete alphabet $\{1,..,m\}$). Equivalently, having only $k_{S'}(t')$ on all $S' \in \mathcal{S}^0$ and $t' : \{1,..,m-1\} \to \mathcal{A}$, linear equations (9) allow us to deduce the whole k function.

These three restrictions suggest that the vector count $\mathbf{k} \in \mathbb{Z}^D$ resides in a space of dimensionality $\mu = D - \mathrm{rk}(C) - 1$ where $\mathrm{rk}(C)$ is the rank of matrix C. We formally establish it in the theorem below.

Theorem 1. *The following matrix*

$$\begin{aligned} C_m(S) \quad = \quad & C^*\left(\{(S', \mathbf{s}, t') : \; S' \in \mathcal{S}^0, \; (S' + \mathbf{s}) \subset S, \right. \\ & \left. t' : S' \to \{1,..,m-1\}\}\right) \end{aligned}$$

of rank

$$\mathrm{rk}(C) = \sum_{S' \in \mathcal{S}^0} (|\{\mathbf{s} : (S' + \mathbf{s}) \subset S\}| - 1)(m-1)^{|S'|} \quad (10)$$

consists of linearly independent rows of the conservation laws. There are $\mu = D - \mathrm{rk}(C) - 1$ independent coordinates

$$\{k_{S'}(t') : S' \in \mathcal{S}^0, t' : S' \to \{1,..,m-1\}\} \quad (11)$$

of the count vector $\mathbf{k} \in \mathbb{Z}^D$. In particular, for the box shape $S = I_{l_1} \times I_{l_2} \times \ldots \times I_{l_d}$ we find

$$\mu = D - 1 - \mathrm{rk}(C) = \sum_{\mathbf{s} \in \{0,1\}^d} m^{\prod_i (l_i - s_i)} \cdot (-1)^{\sum_i s_i} \quad (12)$$

where $\mathbf{l} = (l_1, \ldots, l_d) \in \mathbb{N}^d$.

We now discuss a few examples illustrating reduction of the conservation laws and Theorem 1.

Example 2: *One-dimensional Markov Chain*. Consider now $d = 1$ Markov chain over $\mathcal{A} = \{1, 2\}$. We have four tiles $((11), (21), (12), (22))$ that constitute four coordinates of the vector count $\mathbf{k}$. The normalization condition is

$$k(11) + k(21) + k(12) + k(22) = N.$$

To find the conservation law (7) we choose a one point subshape, that can appear in two different positions: e.g. for $S' = \{(0)\}$ and $\mathbf{s} = (1)$. Clearly, $k(t)$ restricted to a single point leads for $m = 2$ to two conservation equations

$$\begin{aligned} k(11) + k(12) &= k(1*) = k(*1) = k(11) + k(21), \\ k(21) + k(22) &= k(2*) = k(*2) = k(12) + k(22) \end{aligned}$$

where $*$ denotes "don't care" symbol. Summing these two equations we obtain the normalization equation, thus one of them can be removed as linearly dependent leading to only one conservation equation $k(12) - k(21) = 0$ that can be written in the matrix form as $C\mathbf{k} = 0$ or

$$(0, -1, 1, 0) \cdot \mathbf{k} = \mathbf{0}.$$

The vector count $\mathbf{k}$ of the original $D = 4$ dimensional space lies on a $\mu = 2$–dimensional polytope (by the normalization and the conservation laws). Such a vector has two independent coordinates, for example $k(1)$ and $k(11)$ that satisfy (11) for $\mathcal{S}^0 = \{\{(0)\}, \{(0), (1)\}\}$. Then by (9), we can find all other coordinates as follows: $k(2) = 1 - k(1)$, $k(21) = k(12) = k(1) - k(11)$, $k(22) = k(2) - k(12) = 1 - 2k(12)$. $\square$

Example 3: *Markov Field for $d = 2$ with the L Shape*. For the "L"-shape in the $d = 2$ case and $m = 2$, the frequency vector $\mathbf{k}$ has $D = m^3 = 8$ coordinates however, only five of them are independent. The only nontrivial subshape S' appearing in multiple positions is a one point shape but it can appear in 3 different positions. We could use (7) in three different ways but one of these conservative equation is redundant. We can show that the matrix C is in this case:

$$C\mathbf{k} = \begin{pmatrix} 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & -1 & 1 & 0 & 1 & 0 \end{pmatrix} \cdot \mathbf{k} = \mathbf{0}.$$

These two independent conservation laws restrict the space of $\mathbf{k}$ to $\mu + 1$ =6–dimensional cone, and the normalization equation further restricts it to $\mu = 5$ dimensional polytope. $\square$

C. Geometry and Enumeration

We now explore the geometry of the count vector $\mathbf{k} = \{k(t)\}_{t \in T}$ in the $D = m^{|S|}$ space. As discussed before and illustrated in Figure 2, the conservation laws $C\mathbf{k} = \mathbf{0}$ restrict $\mathbf{k}$ to a $D - \mathrm{rk}(C) = \mu + 1$ dimensional cone $\mathscr{C}$ and the normalization equation $\sum_t k(t) = N$ further restricts $\mathbf{k}$ to the polytope $\mathcal{F}_\mathbf{n}$. Formally,

$$\begin{aligned} \mathscr{C} \quad &\equiv \quad \mathscr{C}(m, S) = \{\mathbf{k} \in \mathbb{N}^D \;:\; C_m(S) \cdot \mathbf{k} = \mathbf{0}\}, \quad (14) \\ \mathcal{F} \quad &\equiv \quad \mathcal{F}_N(m, S) = \{\mathbf{k} \in \mathscr{C} : \sum_i k_i = N\}. \quad (15) \end{aligned}$$

We also define the normalized polytope $\hat{\mathcal{F}}(m, S)$ of frequency vectors $\hat{\mathbf{k}}$ as

$$\hat{\mathcal{F}} \equiv \hat{\mathcal{F}}(m, S) = \{\hat{\mathbf{k}} \in \{(\mathbb{R}^+ \cup \{0\})^D \;:\; C \cdot \hat{\mathbf{k}} = \mathbf{0}, \; \sum_i \hat{k}_i = 1\} \quad (16)$$

and then

$$\mathcal{F}_N = \{N\hat{\mathbf{k}} : \hat{\mathbf{k}} \in \hat{\mathcal{F}}, \; N\hat{\mathbf{k}} \in \mathbb{Z}^D\} \quad (17)$$

for the scaled polytope. Finally, the set of all realizable count vectors (Markov types) is then

$$\hat{\mathcal{P}}(m, S) \equiv \hat{\mathcal{P}} = \left\{ \frac{\mathbf{k}}{N} \;:\; \exists_{\mathbf{n} \in \mathbb{N}^d} \; \mathbf{k} \in \mathcal{P}_\mathbf{n}(m, S), \; N = \prod_i n_i \right\}.$$

Observe that $\hat{\mathcal{F}}$ is an intersection of a linear subspace with $\{\hat{k}_i \geq 0\}$ half planes for $i = 1, \ldots, D$. From basic convex analysis we then know that $\hat{\mathcal{F}}$ is convex with *extremal points* that algebraically satisfy the original linear conditions

$C \cdot \hat{\mathbf{k}} = \mathbf{0}$ with $(1,1,...,1) \cdot \hat{\mathbf{k}} = 1$ and $\mu = D - \mathrm{rk}(C) - 1$ of D equations $\hat{k}_i = 0$ for $i = 1, \ldots, D$. The number of the extremal points obtained this way is finite and at most $\binom{D}{\mu}$. Therefore, $\hat{\mathcal{F}}$ is a *convex polytope* and these extremal points are in fact *vertices*. We can find them by solving a system of linear equations for each of $\binom{D}{\mu}$ cases and removing those having negative coordinates. For example, for the "L"-shape in the case with $m = 2$ we have 7 vertices of $\mu = 5$ dimensional polytope in $D = 8$ dimensional space.

Furthermore, we shall prove in the journal version that topological closure $\mathrm{cl}(\hat{\mathcal{P}})$ of $\hat{\mathcal{P}}$ is a convex subset of $\hat{\mathcal{F}}$. We formally express it as a lemma below.

Lemma 2. *Topological closure* $\mathrm{cl}(\hat{\mathcal{P}})$ *is a convex subset of* $\hat{\mathcal{F}}$.

The lattice $\mathcal{F}_N$ consists of all integer points inside the polytope $\hat{\mathcal{F}}$ scaled by N factor. Volume of $\mathcal{F}_N$ is of order N^μ, and we expect the number of integer points in $\mathcal{F}_N$ also grows asymptotically as N^μ. This is indeed the case by the Ehrhart Theorem [5] which shows that if $\hat{\mathcal{F}}$ is a convex polytope with rational vertices, then the scaled polytope $\mathcal{F}_N$ has cardinality of order N^μ.

Theorem 3 (Ehrhart, 1967)**.** *If $\hat{\mathcal{F}}$ is a convex polytope with vertices in $\mathbb{Q}^D$, where $\mathbb{Q}$ is the set of rational numbers, then there exist a period $p \in \mathbb{N}$ and real coefficients $c_{i,j}$ such that $c_{\mu,j} \neq 0$ for some j and $|\mathcal{F}_N| = a_{\mu,j} N^\mu + a_{\mu-1,j} N^{\mu-1} + \ldots a_{0,j}$ if $N \equiv j \pmod{p}$ where μ is the dimension of $\mathcal{F}_N$.*

In our case, it is easy to see that $\hat{\mathcal{F}}$ has vertices in $\mathbb{Q}^D$. Indeed, by the construction these vertices are solutions of a system of linear equations with integer coefficients.

But $\mathcal{F}_N$ is only a superset of $\mathcal{P}_N$. Therefore, to establish the growth of $|\mathcal{P}_N|$ we need a matching lower bound. We establish it in the journal version of this paper proving the following main result.

Theorem 4. *Consider the torus $\mathcal{O}_{\mathbf{n}}$. There exists $0 < c^{min} \leq c^{max}$ such that if $n_i \geq 2w_i - 1$ for all i, then*

$$c^{min} N^\mu \leq |\mathcal{P}_{\mathbf{n}}(m, S)| \leq c^{max} N^\mu \tag{18}$$

where the width $\mathbf{w}$ of shape S is the smallest $(w_1, .., w_d) \in \mathbb{N}^d$ such that for some shift $S \subset I_{w_1} \times .. \times I_{w_d}$.

Finally, we consider the box $\mathcal{I}_{\mathbf{n}} = I_{n_1} \times I_{n_2} \times \ldots \times I_{n_d} \subset \mathbb{Z}^d$ and count the number of types $\tilde{\mathcal{P}}_{\mathbf{n}}(m, S)$ in such a box (i.e, non-torus case), It turns out that the number of types significantly increases due to the boundary effect. In particular, instead of $C \cdot \mathbf{k} = \mathbf{0}$ conservation laws, we have $C \cdot \mathbf{k} = \mathbf{b}$ for some $\mathbf{b}$ vectors defining imbalance caused by the boundary effect. For each $\mathbf{b}$ we have μ dimensional polytope of $\Theta(N^\mu)$ size. Furthermore, different $\mathbf{b}$ correspond to shifted polytopes in the remaining $D - 1 - \mu$ dimensions: there are at most $\Theta((N/\min_i \ n_i)^{D-1-\mu})$ of them, therefore we find

$$\Theta(N^\mu) \times \Theta\left(\frac{N}{\min_i \ n_i}\right)^{D-1-\mu}.$$

We formulate this formally in the next theorem.

Theorem 5. *Consider the box $\mathcal{I}_N$. There exists $0 < \tilde{c}^{min} \leq \tilde{c}^{max}$ such that if $n_i \geq 4w_i - 1$ for all i, then*

$$\tilde{c}^{min} \frac{N^{D-1}}{(\min_i \ n_i)^{\mathrm{rk}(C)}} \leq |\tilde{\mathcal{P}}_{\mathbf{n}}(m, S)| \leq \tilde{c}^{max} \frac{N^{D-1}}{(\min_i \ n_i)^{\mathrm{rk}(C)}} \tag{19}$$

where the width $\mathbf{w}$ of shape S is the smallest $(w_1, .., w_d) \in \mathbb{N}^d$ such that for some shift $S \subset I_{w_1} \times .. \times I_{w_d}$.

We should point out that in [7] for $d = 1$ it was shown that $|\mathcal{F}_N|$ is asymptoticly equivalent to $|\mathcal{P}_N|$, that is, $|\mathcal{P}_N| \sim |\mathcal{F}_N|$ as $N \to \infty$. Generally, this turns out not to be true. However, in some special cases we can say more about $|\mathcal{P}_N|$ provided we have a more precise estimate for $|\mathcal{F}_N|$. Furthermore, the constants appearing in Theorems 4 and 5 are very small, of order $O(1/\mu!)$ or smaller. For example, for $d = 1$ and $m = 5$ it was shown in [7] that the constant is $O(10^{-22})$.

Acknowledgment

This work was supported by NSF Center for Science of Information (CSoI) Grant CCF-0939370, and in addition by Grants from ARPA HR0011-07-1-0002 and ONR N000140810668, NSA Grant H98230-11-1-0141, and NSF Grants DMS-0800568, CCF-0830140, and the MNSW grant DEC-2013/09/B/ST6/02258. W. Szpankowski is also a Visiting Professor at ETI, Gdańsk University of Technology, Poland.

References

[1] F. Ardila and R. Stanley, Tilings, Clay Public Lecture at the IAS/Park City Mathematics Institute, July, 2004;

[2] P. Brémaud, *Markov Chains, Gibbs Fields, Monte Carlo Simulation, and Queues*, Springer-Verlag, New York, 1999.

[3] T. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991.

[4] I. Csziszár, The Method of Types, *IEEE Trans. Information Theory*, 44, 2505-2523, 1998.

[5] E. Ehrhart, Sur un problème de géomŕie diophantienne lináire, *J. reine angew. Math.*, 227, 1-29, 1967.

[6] P. Jacquet and W. Szpankowski, Markov Types and Minimax Redundancy for Markov Sources, *IEEE Trans. Information Theory*, 50, 1393-1402, 2004.

[7] P. Jacquet, C. Knessl, W. Szpankowski, Counting Markov Types, Balance d Matrices, and Eulerian Graphs. *IEEE Transactions on Information Theory* 58(7), 4261-4272, 2012.

[8] P. Kasteleyn, The statistics of dimers on a lattice. I. The number of dimer arrangements on a quadratic lattice, *Physica*, 27 (12), 1961.

[9] A. Martín, G. Seroussi, and M. J. Weinberger, Type classes of tree models, *Proc. ISIT 2007*, Nice, France, 2007.

[10] N. Merhav and M. J. Weinberger, On universal simulation of information sources using training data, *IEEE Trans. Inform. Theory*, 50, 1, 5-20, 2004.

[11] M. Mezard, A. Montanari, *Information, Physics, and Computation* Oxford University Press, 2009.

[12] Y. Rachlin, R. Negi and P. Khosla, Sensing Capacity for Markov Random Fields, *ISIT 2005*, Adelaide, 2005.

[13] G. Seroussi, On Universal Types, *IEEE Trans. Information Theory*, 52, 171-189, 2006.

[14] R. Stanley, *Enumerative Combinatorics*, Vol. II, Cambridge University Press, Cambridge, 1999.

[15] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*, Wiley, New York, 2001.

[16] P.O. Vontobel, Counting in graph covers: a combinatorial characterization of the Bethe entropy function, *IEEE Trans. Inf. Theory*, 59, 6018-6048, 2013.

[17] P. Whittle, Some Distribution and Moment Formulæ for Markov Chain, *J. Roy. Stat. Soc.,* Ser. B., 17, 235-242, 1955.