ELSEVIER

Automatic Proofs for Formulae Enumerating Proper Polycubes

Gill Barequet [1]

Dept. of Computer Science
The Technion—Israel Inst. of Technology
Haifa 32000, Israel

Mira Shalah [2]

Dept. of Computer Science
The Technion—Israel Inst. of Technology
Haifa 32000, Israel

Abstract

We develop a general framework for computing formulae enumerating polycubes of size n which are proper in $n-k$ dimensions (spanning all $n-k$ dimensions), for a fixed value of k. Besides the fundamental importance of knowing the number of these simple combinatorial objects, such formulae are central in the literature of statistical physics in the study of percolation processes and the collapse of branched polymers. We re-affirm the already-proven formulae for $k \leq 3$, and prove rigorously, for the first time, that the number of polycubes of size n that are proper in $n-4$ dimensions is $2^{n-7}n^{n-9}(n-4)(8n^8 - 128n^7 + 828n^6 - 2930n^5 + 7404n^4 - 17523n^3 + 41527n^2 - 114302n + 204960)/6$.

[1] Email: barequet@cs.technion.ac.il
[2] Email: mshalah@cs.technion.ac.il

1 Introduction

A d-dimensional polycube of size n is a connected set of n cubes in d dimensions, where connectivity is through $(d-1)$-dimensional faces. Two *fixed* polycubes are considered distinct if they differ in their shapes or orientations. A polycube is called *proper* in d dimensions if the convex-hull of the centers of its cubes is d-dimensional. Following Lunnon [4], we let $\mathrm{DX}(n,d)$ denote the number of fixed polycubes of size n that are proper in d dimensions.

Enumeration of polycubes and computing their asymptotic growth rate are important problems in combinatorics and discrete geometry, originating in statistical physics [3], where they play a fundamental role in the analysis of percolation processes and the collapse of branched polymers. To-date, no formula is known for $A_d(n)$, the number of fixed polycubes of size n in d dimensions, for any fixed value of d. The main interest in DX stems from the fact that $A_d(n)$ can be easily computed using the formula $A_d(n) = \sum_{i=0}^{d} \binom{d}{i} \mathrm{DX}(n,i)$ (Lunnon [4]). In a matrix listing the values of DX, the top-right triangular half and the main diagonal contain only 0s. This gives rise to the question of whether a pattern can be found in the sequences $\mathrm{DX}(n,n-k)$, where $k > 0$ is the ordinal number of the diagonal. Significant progress in estimating λ_d, the asymptotic growth rate of the number of polycubes in d dimensions, has been obtained in the literature of statistical physics, although the computations usually relied on unproven assumptions and on formulae for $\mathrm{DX}(n,n-k)$ interpolated empirically from known values of $A_d(n)$. Peard and Gaunt [7] predicted that for $k > 1$, the diagonal formula $\mathrm{DX}(n,n-k)$ has the pattern $2^{n-2k+1}n^{n-2k-1}(n-k)h_k(n)$, where $h_k(n)$ is a polynomial in n, and conjectured explicit formulae for $h_k(n)$ for $k \leq 6$. Luther and Mertens [5] conjectured a formula for $k = 7$.

It is easy to show that $\mathrm{DX}(n,n-1) = 2^{n-1}n^{n-3}$ (seq. A127670 in [6]). Barequet et al. [2] proved for the first time that $\mathrm{DX}(n,n-2) = 2^{n-3}n^{n-5}(n-2)(2n^2-6n+9)$ (seq. A171860). The proof uses a case analysis of the possible structures of spanning trees of the polycubes, and the various ways in which cycles can be formed in their cell-adjacency graphs. Similarly, Asinowski et al. [1] proved that $\mathrm{DX}(n,n-3) = 2^{n-6}n^{n-7}(n-3)(12n^5-104n^4+360n^3-679n^2+1122n-1560)/3$, again, by counting spanning trees of polycubes, yet the reasoning and calculations were significantly more involved. The inclusion-exclusion principle was applied in order to count correctly polycubes whose cell-adjacency graphs contained certain subgraphs, so-called "distinguished structures." In comparison with the case $k=2$, the number of such structures is substantially higher, and the ways in which they can appear in spanning trees
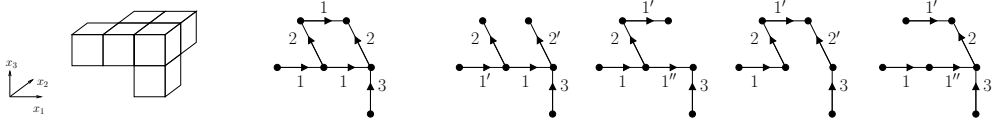
(a) Polycube   (b) Graph   (c) Four spanning trees

Fig. 1. A polycube P, the corresponding graph $\gamma(P)$, and spanning trees of $\gamma(P)$.

are much more varied. The latter proof provided a better understanding of the difficulties that one would face in applying this technique to higher values of k. The number of distinguished structures grows rapidly, the inclusion relations between them are much more complicated, and the ways in which they are connected by forests are much more varied. As anticipated [1], it is totally impractical to manually achieve a similar proof for $k>3$.

In this paper we create a theoretical set-up for proving the formulae for $\mathrm{DX}(n, n-k)$ for a fixed value of k. Our method fully automates the manual method presented in [1,2]. For this nontrivial generalization, we need a few key observations about polycubes that are proper in $n-k$ dimensions. We also provide a general characterization of distinguished structures, and design algorithms that produce and analyze them automatically. Using our implementation of this method, we find the explicit formula (which has never been proven before) for $\mathrm{DX}(n, n-4)$, stated in the following theorem.

Theorem 1.1 $\mathrm{DX}(n, n-4) = 2^{n-7}n^{n-9}(n-4)(8n^8 - 128n^7 + 828n^6 - 2930n^5 + 7404n^4 - 17523n^3 + 41527n^2 - 114302n + 204960)/6.$

2 Overview of the Method

Denote by $\mathcal{P}_n$ the set of proper polycubes of size n in $n-k$ dimensions. Let $P \in \mathcal{P}_n$, and let $\gamma(P)$ denote the adjacency graph of P constructed as follows: The vertices correspond to the cells of P; two vertices are connected by an edge if their corresponding cells are adjacent; and an edge has label i ($1 \le i \le n-k$) if the corresponding cells have different i-coordinate. The direction of the edge is from the lower to the higher cell. See Figure 1 for an example. Since $P \mapsto \gamma(P)$ is an injection, it suffices to count the graphs obtained from the members of $\mathcal{P}_n$ in this way. We count these graphs by counting their spanning trees. A spanning tree of $\gamma(P)$ has $n-1$ edges labeled by numbers from the set $\{1, 2, \ldots, n-k\}$; All these labels are present because the polycube is proper in $n-k$ dimensions. Hence, $n-k$ edges of the tree are labeled with the labels $1, 2, ..., n-k$, and the remaining $k-1$ edges repeat labels from the same set. There is a bijection between the possibilities of repeated edge labels and
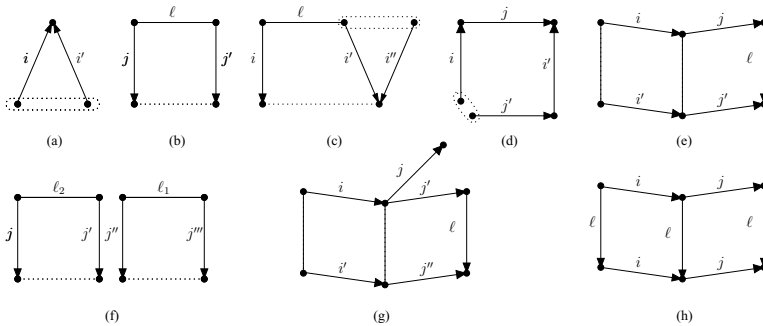
Fig. 2. (a–g) A few distinguished structures for $k = 4$ (note that (f) is disconnected); (h) A cycle structure. A dotted line is drawn between every pair of neighboring cells and around every pair of coinciding cells.

the partitions of the integer $k-1$. Specifically, each partition $p=\{a_1, \ldots, a_h\}$ (where $\sum_{i=1}^{h} a_i = k-1$) corresponds to h repeated labels in the spanning tree, such that the ith repeated label appears a_i+1 times. In such a case we say that the tree is "labeled according to p." We denote by $\Pi(m)$ the set of all partitions of the integer m. When we consider a spanning tree of $\gamma(P)$, we distinguish a repeated label i which appears r times by $i, i', ..., i'^{(r-1)}$.

To compute $|\mathcal{P}_n|$, we consider all possible directed edge-labeled trees of size n with the possible repetitions of edge labels, and count only those that represent valid polycubes, then derive the actual number of polycubes.

3 Distinguished Structures

For each directed edge-labeled tree, one can attempt to build the corresponding polycube. In this process there are two types of problems: (a) Cells may coincide (a tree with overlapping cells is invalid, see Figures 2(a,d)); and (b) Two cells which are not connected by a tree edge may be adjacent (such a tree corresponds to a polycube P with cycles in $\gamma(P)$, and, therefore, its spanning tree is not unique, see Figures 2(b,e)). A *distinguished* structure is the union of all paths (edges and incident vertices) that connect two coinciding or adjacent cells. This characterization allows the design of an algorithm for producing $\mathcal{DS}_k$: The set of all distinguished structures in $n-k$ dimensions. We begin with generating all "free trees" (non-isomorphic trees) of size at most the value specified in Lemma 3.1. Then, we process each free tree T of size t by labeling its edges according to every partition $p \in \cup_{i=1}^{k-1}\Pi(i)$ so as to obtain a directed edge-labeled tree T', and then checking (by a DFS traversal) whether T' contains coinciding or neighboring cells. T' is added to $\mathcal{DS}_k$ if it is *not* isomorphic to any structure of size t already in $\mathcal{DS}_k$, and if

at least one of the following conditions holds: (i) T' contains two coinciding or neighboring cells which are connected by a path with $t-1$ edges (see, e.g., Figures 2(a,b,d,e)); (ii) T' is isomorphic to the union of $d_1, ..., d_m \in \mathcal{DS}_k$, such that the isomorphic copies of $d_1, ..., d_m$ in T' cover all its edges (see, e.g., Figures 2(c,g)). Disconnected distinguished structures (see, e.g., Figure 2(f)) are generated by checking if collections of edge-connected structures in $\mathcal{DS}_k$ yield a single disconnected structure labeled according to $p \in \cup_{i=1}^{k-1} \Pi(i)$.

Lemma 3.1 *A connected (resp., disconnected) distinguished structure in $\mathcal{DS}_k$ has at most $3k - 2$ (resp., $4k$) vertices.*

Lemma 3.2 *[1, Lemma 7] [2, Lemma 2] The number of directed trees with n vertices and $n-1$ distinct edge labels is $2^{n-1}n^{n-3}$, for $n \geq 2$.*

Let T_p denote the number of directed trees with n vertices labeled according to $p \in \Pi(k-1)$. Then, $T_p = \pi(p)\binom{n-k}{|p|}2^{n-1}n^{n-3}$.

Lemma 3.3 *Let $\sigma \in \mathcal{DS}_k$ be composed of $k^* \geq 1$ trees $s_1, \ldots, s_{k^*}$ with a total of n^* vertices and distinct edge labels. The number of occurrences of σ in trees of size n with distinct edge labels is $F_n(\sigma) = (\prod_{i=1}^{k^*} |s_i|)\frac{(n-n^*+k^*-1)!}{(n-n^*)!}n^{n-n^*+k^*-2}$.*

Proof. (Sketch) We proceed by double counting, enumerating in two ways the different sequences of directed edges that can be added to a graph with $n-n^*$ vertices and a structure σ so as to form a rooted tree. One way is to add the edges one by one: There are $\mathcal{N} = \prod_{i=1}^{k^*} |s_i|$ options to choose a root for each component. We begin with a forest with $n-n^*+k^*$ rooted trees. After adding a set of edges forming a rooted forest with i trees, there are $n(i-1)$ choices for the next edge. Therefore, the total number of choices is $\mathcal{N} \prod_{i=2}^{n-n^*+k^*} n(i-1)$. Another way to count these sequences is to start with an unrooted edge-labeled tree containing σ, choose one of its n vertices as a root, and choose one of the $(n-n^*)!$ possible sequences, say, η, label the $n-n^*$ vertices which do not belong to σ according to η, and "shift" each vertex label to the incident edge towards the root. The number of sequences formed this way is $nF_n(\sigma)(n-n^*)!$. Hence, $F_n(\sigma) = \mathcal{N}\frac{(n-n^*+k^*-1)!}{(n-n^*)!}n^{n-n^*+k^*-2}$, as claimed. $\square$

3.1 Inclusion-Exclusion

Let $\mathcal{F}_n(\sigma)$ denote the number of occurrences of σ in *directed* edge-labeled trees of size n. Obviously, $\mathcal{F}_n(\sigma) = 2^{n-n^*+k^*-1}F_n(\sigma)$. Let the distinguished structure $\sigma' \in \mathcal{DS}_k$ be labeled according to $p' \in \cup_{i=1}^{k-1}\Pi(i)$. Let us denote by $\mathcal{O}_p(\sigma')$ the number of occurrences of σ' in directed trees of size n that are labeled according to $p \in \Pi(k-1)$. Computing $\mathcal{O}_p(\sigma')$ involves choosing the $|p|$

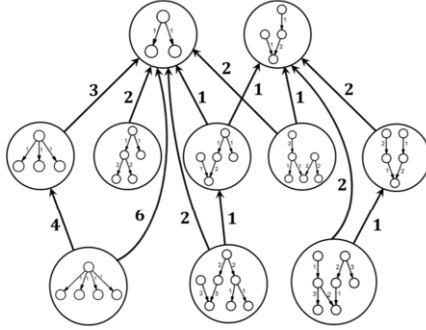Fig. 3. A snapshot of the IE graph for $k = 4$.

repeated labels in the tree and the labels of σ', counting the automorphisms of σ', and multiplying by $\mathcal{F}_n(\sigma')$.

When counting the occurrences of a structure $\sigma \in \mathcal{DS}_k$, other distinguished structures which contain multiple occurrences of σ are counted multiple times. In order to count correctly, we build an inclusion-exclusion graph IE=$(\mathcal{V}, \mathcal{E})$ whose vertices correspon to the structures in $\mathcal{DS}_k$. There is an edge $e = \sigma_1 \rightarrow \sigma_2$ labeled with c if σ_1 contains c occurrences of σ_2. Hence, the roots of IE are all the structures that are not contained in any other structure. Let $\ell(e)$ denote the label of the edge e, $I(\sigma_2)$ denote the set $\{\sigma_1 \in \mathcal{V} : (\sigma_1 \rightarrow \sigma_2) \in \mathcal{E}\}$, and $T_p(\sigma)$ denote the number of trees of size n labeled according to $p \in \Pi(k-1)$ that contain σ but no other structure $\sigma' \in I(\sigma)$ as a subtree. It is easy to prove by induction that $T_p(\sigma_2) = \mathcal{O}_p(\sigma_2) - \sum_{\sigma_1 \in I(\sigma_2)} \ell(\sigma_1 \rightarrow \sigma_2) T_p(\sigma_1)$. Figure 3 shows a subgraph of IE for $k=4$. A simple bottom-up procedure traverses the graph and computes $T_p(u)$ for every vertex $u \in \mathcal{V}$ and every partition $p \in \Pi(k-1)$.

4 Counting Polycubes

4.1 Trees

Similarly to DX(n, d), let DT(n, d) denote the number of fixed *tree* polycubes of size n that are proper in d dimensions. Every tree polycube gives rise to a unique spanning tree, and every directed tree labeled according to $p \in \Pi(k-1)$ corresponds to a tree polycube in $\mathcal{P}_n$ unless it contains a structure $\sigma \in \mathcal{DS}_k$ as a subtree. (A spanning tree of a tree polycube can neither contain coinciding cells because these are illegal, nor can it contain neighboring cells). Thus, we exclude all the trees that contain any structure $\sigma \in \mathcal{DS}_k$ as a subtree, obtaining that DT$(n, n-k) = \sum_{p \in \Pi(k-1)} \frac{T_p - \sum_{\sigma \in \mathcal{DS}_k} T_p(\sigma)}{\prod_{j=1}^{|p|} p[j]!}$. (The division by $\prod_{j=1}^{|p_i|}(p_i[j]!)$ is because each tree polycube is counted that many times.)

4.2   Nontrees

Let $\mathcal{C}(k)$ denote the set of all cycle structures of polycubes proper in $n{-}k$ dimensions. This set can be found using $\mathcal{DS}_k$: A distinguished structure is a spanning tree of a cycle if it contains only neighboring cells and no coinciding cells. For example, the structure shown in Figure 2(e) is a spanning tree of the cycle shown in Figure 2(h). For any $\mathcal{C}_i \in \mathcal{C}(k)$, let $P_{\mathcal{C}_i}$ denote the number of polycubes $P \in \mathcal{P}_n$ that contain $\mathcal{C}_i$ in $\gamma(P)$. Suppose that a distinguished structure $\sigma \in \mathcal{DS}_k$ has c occurrences in $\mathcal{C}_i$. Then, $P_{\mathcal{C}_i} = \sum_{p \in \Pi(k-1)} \frac{T_p(\sigma)}{c \prod_{j=1}^{|p|} p[j]!}$.
Finally, $\mathrm{DX}(n, n{-}k) = \mathrm{DT}(n, n{-}k) + \sum_{i=1}^{|\mathcal{C}(k)|} P_{\mathcal{C}_i}$.

5   Results

The entire method was implemented in a C++ program, using `Mathematica` for simplifying the final formula. Our results agree completely with the formulae conjectured in the literature of statistical physics. For $k{=}3$, the program found 147 distinguished structures and 13 cycle structures. For $k{=}4$, the program found 8,397 distinguished structures and 179 cycles. The parallel computation took about 15 minutes on a supercomputer with 12 processors and 64 GB of RAM. The program produced data files which document the entire computation, serving as a proof of Theorem 1.1.

References

[1] A. ASINOWSKI, G. BAREQUET, R. BAREQUET, AND G. ROTE, Proper n-cell polycubes in $n{-}3$ dimensions, *J. of Integer Sequences*, 15 (2012), #12.8.4.

[2] R. BAREQUET, G. BAREQUET, AND G. ROTE, Formulae and growth rates of high-dimensional polycubes, *Combinatorica*, 30 (2010), 257–275.

[3] S.R. BROADBENT AND J.M. HAMMERSLEY, Percolation processes: I. Crystals and mazes, *Proc. Cambridge Philosophical Society*, 53 (1957), 629–641.

[4] W.F. LUNNON, Counting multidimensional polyominoes, *The Computer Journal*, 18 (1975), 366–367.

[5] S. LUTHER AND S. MERTENS, Counting lattice animals in high dimensions, *J. of Statistical Mechanics: Theory and Experiment*, 9 (2011), 546–565.

[6] *The On-Line Encyclopedia of Integer Sequences*, available at `http://oeis.org` .

[7] P.J. PEARD AND D.S. GAUNT, $1/d$-expansions for the free energy of lattice animal models of a self-interacting branched polymer, *J. Physics, A: Mathematical and General*, 28 (1995), 6109–6124.