8

Counting Linear Extensions: Polyhedral Methods

G. R. Brightwell and W. T. Trotter

April 14, 2006

This chapter is the first of three concerned with various aspects of the problem of estimating the number $e(\mathbf{P})$ of linear extensions of a partial order $\mathbf{P}$. In this chapter, we concentrate on the general problem, giving various inequalities relating $e(\mathbf{P})$ to various other parameters of $\mathbf{P}$.

The emphasis will be placed on relating $e(\mathbf{P})$ to the volume of certain convex polytopes. This has proved to be a beautiful and fruitful approach to dealing with $e(\mathbf{P})$, and we will hear its echoes in subsequent chapters.

8.1 Introduction; Elementary Inequalities

Recall that a *linear extension* of a poset $\mathbf{P} = (X, P)$ is a linear order $\prec$ on X extending $<$, i.e., so that $x \prec y$ whenever $x < y$.

For a poset $\mathbf{P}$, let $E(\mathbf{P})$ denote the set of linear extensions of $\mathbf{P}$, and $e(\mathbf{P}) = |E(\mathbf{P})|$ the number of linear extensions of P. Obviously $1 \leq e(\mathbf{P}) \leq n!$, where $n = |X|$, with the bounds achieved in the case of $\mathbf{P}$ a chain and an antichain respectively. If $\mathbf{Q} = (X, Q)$ is another poset on X with $P \subset Q$, then $e(\mathbf{P}) > e(\mathbf{Q})$: the strict inequality follows from Szpilrajn's Theorem, Theorem 9.99.

For some purposes, it is helpful to regard $E(\mathbf{P})$ as a probability space, with each element equally likely. Then we can regard subsets as events, and talk about their probabilities. For instance, we can consider the event that element x is below element y and its probability $\Pr(x \prec y)$. We will use this notion only occasionally in this chapter, but it will be central to the next two.

One motivation for studying $e(\mathbf{P})$ is its connection with sorting. Suppose we are given a set $X = \{x_1, \ldots, x_n\}$, told that it is in some linear order, unknown to us, and asked to find the linear order: this is the *sorting* problem.

In the purest form, all we are allowed to do to accomplish the task is to make sequential queries of the form "Is x_i below x_j?", receiving an (instant and truthful) Yes/No answer. The information we have available to us part-way through a sorting process is in the form of a poset $\mathbf{P}$ on the ground-set X. Now the set of linear orders on X that remain consistent with the information we have is just $E(\mathbf{P})$.

How should we go about choosing our next query? Ideally, we would pick a pair (x_i, x_j) such that about half of the linear extensions of P have x_i above x_j: then, no matter what the answer to the query "Is x_i below x_j?", we will have reduced the number of linear extensions by a factor of about 2. Indeed, it is easy to see that $\log_2 e(\mathbf{P})$ is a lower bound on the number of queries we need, in the worst case or in the average case, to complete the sorting process. Is this a good bound? Can we always find a pair to compare such that x_i is below x_j "about half the time"? We return to these questions in Chapter 10.

For now, let us simply note that there is reason to be interested in methods for bounding or estimating the number of linear extensions of a partial order.

We'll start with a simple yet useful inequality.

Theorem 8.1.1 *For any poset $\mathbf{P} = (X, P)$, with $|X| = n$,*

$$e(\mathbf{P}) \leq \mathrm{width}(\mathbf{P})^n.$$

Proof We proceed by induction on n. The result is certainly true for the one-element poset. Suppose then that we have the inequality for all posets on fewer than n elements.

Given a poset $\mathbf{P}$ on a ground-set X of size n, we partition $E(\mathbf{P})$ according to the bottom element. So, for x maximal in P, let $e(\mathbf{P}; x)$ be the number of linear extensions in which x is bottom, and note that $e(\mathbf{P}; x) = e(\mathbf{P} - x)$.

Thus we have

$$e(\mathbf{P}) = \sum_{x \in \mathrm{Min}(\mathbf{P})} e(\mathbf{P} - x). \tag{8.1}$$

Each poset $\mathbf{P} - x$ has $n-1$ elements, and width at most $\mathrm{width}(\mathbf{P})$, so $e(\mathbf{P} - x) \leq \mathrm{width}(\mathbf{P})^{n-1}$. The number of terms in the sum is $|\,\mathrm{Min}(\mathbf{P})| \leq \mathrm{width}(\mathbf{P})$, so the inequality holds for $\mathbf{P}$ and we are done. $\square$

Another way to look at this is as building a linear extension from the bottom up: at each stage there are at most $\mathrm{width}(\mathbf{P})$ choices, irrespective of what decisions were made earlier.

Fig. 8.1. The five linear extensions of the poset N, and the corresponding units of flow.

Yet another way: partition $\mathbf{P}$ into $w = \mathrm{width}(\mathbf{P})$ chains $C_1, \ldots, C_w$, and encode a linear extension $\prec$: $x_1 \prec x_2 \prec \cdots \prec x_n$ of $\mathbf{P}$ by a string $\sigma_1 \cdots \sigma_n$, where $\sigma_i = j$ if x_i is in C_j. The linear extension can be recovered from the string, and there are w^n possible strings.

This bound can easily be improved; see Exercise 1. But the general form is what matters: if w is fixed, n is large, and $\mathbf{P} = (X, P)$ is an n-element poset of width w, then $e(\mathbf{P})$ is at most w^n, which is much smaller than $n!$, the total number of linear orderings of X.

Indeed, we sometimes want to think of factors that are merely exponentially large (at most C^n for some constant C) as being negligible in this context. In our sorting application, for instance, the number of comparisons required to sort from scratch is well known to be $(1 + o(1))n \log_2 n$, whereas if we are down to C^n possible outcomes then (as we shall confirm in Chapter 10) a further $C'n$ comparisons are all that we need to finish the job.

Our first real result can be thought of as an extension of (8.1); the result and entertaining proof are due to Sidorenko [99].

Theorem 8.1.2 *For any antichain A in a poset $\mathbf{P}$,*

$$e(\mathbf{P}) \leq \sum_{x \in A} e(\mathbf{P} - x),$$

with equality if and only if A intersects every maximal chain.

We have already noted in (8.1) that we have equality if A is the antichain of minimal (or of maximal) elements.

Proof To follow this proof, it may help to consider the example in Figure **??**.

We form a directed network from $\mathbf{P} = (X, P)$, with one node for each element of X, together with a source s and sink t. Put an arc from s to each minimal element of $\mathbf{P}$, an arc from each maximal element to t, and an arc from x to y whenever y covers x in $\mathbf{P}$.

Now, for each linear extension $x_1 \prec x_2 \prec \cdots \prec x_n$ of $\mathbf{P}$, send one unit of flow along the path $s \to x_1 \to x_2 \to \cdots \to x_n \to t$. This flow is not restricted to using arcs of the network: it may also leap across incomparable pairs.

We now combine these unit flows, producing a total flow of value $e(\mathbf{P})$ from s to t.

Suppose x and y are incomparable elements of $\mathbf{P}$. Linear extensions of $\mathbf{P}$ where y is the successor of x are in 1-1 correspondence with linear extensions where x is the successor of y: the correspondence is obtained by swapping x and y whenever they are adjacent in the linear extension. What this means is that the net flow from x to y in our network is 0. Therefore, we genuinely have a flow of value $e(\mathbf{P})$ through the directed network.

For any node x, the value of the flow into x is the number of linear extensions in which x is the successor of one of its lower covers; for minimal x, this should be interpreted as the number in which x is the bottom element of the linear extension. In other words, the value of the flow into x is the number of linear extensions in which x is inserted into a linear extension of $\mathbf{P} - x$ in the lowest possible position. So the value of the flow into x is exactly $e(\mathbf{P} - x)$.

For any antichain A in the network, there is no directed path between elements of A, so the sum of the flow values through the nodes in A, which is equal to $\sum_{x \in A} e(\mathbf{P} - x)$, is at most the total flow value $e(\mathbf{P})$.

This completes the proof of the inequality.

We have equality as long as A separates s from t in the network, i.e., there is no covering edge from $D(A)$ to $U(A)$—conversely, if there is such an edge, then it carries some flow, which does not pass through A, so the inequality is strict. Any such edge can be extended to a maximal chain bypassing A, and conversely any such maximal chain must involve a covering edge from $D(A)$ to $U(A)$. $\square$

As a non-trivial example where we have equality, let $\mathbf{P}$ be a ranked poset with three ranks A_1, A_2, A_3. All maximal chains intersect A_2, so we have $\sum_{x \in A_2} e(\mathbf{P} - x) = e(\mathbf{P})$. There does not seem to be a simple direct proof of this fact via a partition of $E(\mathbf{P})$.

8.2 Polytopes: Preliminaries

For much of this chapter, we shall be dealing with convex polytopes in $\mathbb{R}^n$, where n is the number of elements of the poset. We start by briefly reviewing the basic terminology. For a fuller treatment, see ??? [99].

A *convex polyhedron* is the non-empty intersection of finitely many half-spaces in $\mathbb{R}^n$; so it is defined as the set of vectors in $\mathbb{R}^n$ satisfying finitely many linear inequalities. If a convex polyhedron is bounded, it is called a *convex polytope*. We will be considering convex polytopes contained within $[0,1]^n$.

Suppose that the convex polyhedron $\mathcal{K}$ has full dimension, i.e., is not contained in any hyperplane. Then there is a unique minimal collection of half-spaces whose intersection is $\mathcal{K}$. The boundaries of these half-spaces are called *defining hyperplanes*, and the intersection of a defining hyperplane with $\mathcal{K}$ is a *facet*. (This is the n-dimensional generalization of a "side" of a convex polygon in $\mathbb{R}^2$, or a "face" of a convex polyhedron in $\mathbb{R}^3$.)

An *extreme point* of a convex body $\mathcal{K}$ is a point in $\mathcal{K}$ that cannot be expressed as a convex combination of two different points of $\mathcal{K}$. A bounded convex body $\mathcal{K}$ in $\mathbb{R}^n$ is always the convex hull of its extreme points; indeed, every point in $\mathcal{K}$ can be written as a convex combination of at most $n+1$ of its extreme points (Carathéodory's Theorem). A convex polytope $\mathcal{K}$ has finitely many extreme points, called *vertices*, each of which is the intersection of some n of the defining hyperplanes of $\mathcal{K}$.

We will be identifying each element of a poset with one coordinate of our space $\mathbb{R}^n$. We achieve this by denoting the space $\mathbb{R}^X$, which is formally the set of functions from the n-element set X to $\mathbb{R}$. We use $\mathbf{a}$ to denote an element of $\mathbb{R}^X$, and a_x to denote the coordinate corresponding to element x of the ground-set X.

If Y is a subset of X, we define the *indicator vector* $\mathbf{e}^Y$ to be the vector with $e^Y_x = 1$ if $x \in Y$, and $e^Y_x = 0$ if $x \notin Y$.

The n-dimensional Euclidean volume of a body $\mathcal{K}$ in $\mathbb{R}^n$ is denoted $\mathrm{Vol}(\mathcal{K})$.

8.3 The Order Polytope

Let $\mathbf{P} = (X, P)$ be a poset. The most natural way of encoding $\mathbf{P}$ as an n-dimensional polytope is to consider the *order polytope*

$$\mathcal{O}(\mathbf{P}) = \{\mathbf{a} \in [0,1]^X : a_x \le a_y \text{ whenever } x < y \text{ in } \mathbf{P}\}.$$

A vector $\mathbf{a}$ in $\mathcal{O}(\mathbf{P})$, with no two entries equal, induces a linear order $\prec$ on X: $x \prec y$ if and only if $a_x < a_y$. The definition of the order polytope ensures that this order $\prec$ is a linear extension of P. Moreover, if we set

$$\mathcal{O}(\prec) = \{\mathbf{a} \in [0,1]^X : a_x \le a_y \text{ whenever } x \prec y\}$$

for each linear order $\prec$ on X, then $\mathcal{O}(\mathbf{P})$ can be written as the union $\bigcup_{\prec \in E(\mathbf{P})} \mathcal{O}(\prec)$. This union is disjoint, up to a set of measure zero where two coordinates are equal.

As the set of linear extensions of $\mathbf{P}$ can be recovered from $\mathcal{O}(\mathbf{P})$, so can

$\mathbf{P}$ itself. Not surprisingly, a good deal of information about $\mathbf{P}$ can be read off from $\mathcal{O}(\mathbf{P})$.

Proposition 8.3.1 *Let $\mathbf{P} = (X, P)$ be an n-element poset.*

(i) *The* facets *of $\mathcal{O}(\mathbf{P})$ are given by $\{\mathbf{a} \in \mathcal{O}(\mathbf{P}) : a_x = a_y\}$, where (x, y) is a covering pair in P, as well as $\{\mathbf{a} \in \mathcal{O}(\mathbf{P}) : a_x = 0\}$, for x minimal in P, and $\{\mathbf{a} \in \mathcal{O}(\mathbf{P}) : a_x = 1\}$, for x maximal in P.*

(ii) *The* vertices *of $\mathcal{O}(\mathbf{P})$ are the indicator vectors $\mathbf{e}^U$, where U is an up-set of P.*

(iii) *The* volume $\mathrm{Vol}(\mathcal{O}(\mathbf{P}))$ *is equal to $e(\mathbf{P})/n!$.*

Proof We'll prove just the last of these statements; the other two are left as exercises.

Notice that $[0, 1]^X$ is, up to a set of measure zero, partitioned into $n!$ pieces $\mathcal{O}(\prec)$, each of which—by symmetry—has the same volume, which is therefore $1/n!$. Now, again up to a set of measure zero, $\mathcal{O}(\mathbf{P})$ is the union of $e(\mathbf{P})$ of these sets. □

The simple observations in this proposition have some surprisingly far-reaching consequences, which we'll come back to at the end of this chapter.

8.4 The Chain Polytope

Interesting though the order polytope is, it is an awkward object to study geometrically. To get a feel for this polytope, observe that the entire diagonal (the set of points in $[0, 1]^X$ where all coordinates are equal) is contained in $\mathcal{O}(\mathbf{P})$, so the polytope is (typically) a fairly "thin" set "strung out" along the diagonal.

In this section, we look at a different polytope, the *chain polytope*, associated with $\mathbf{P}$. Geometrically, this is a much more pleasant beast than the order polytope. What makes it particularly interesting is the fact, far from immediately obvious, that it has the same volume $e(\mathbf{P})/n!$ as the order polytope.

Given a partial order $\mathbf{P} = (X, P)$, we define the *chain polytope* to be

$$\mathcal{C}(\mathbf{P}) = \{\mathbf{a} \in [0, 1]^X : \sum_{x \in C} a_x \le 1 \text{ for every chain } C \text{ of } \mathbf{P}\}.$$

So every chain C of $\mathbf{P}$ imposes a "constraint" on the vectors of the chain polytope: of course, it is only the *maximal* chains that matter (i.e., generate facets) in this context, as larger chains impose stronger constraints.

For a simple example, let $\mathbf{Q}$ be any linear order on an n-element set X. Then $\mathcal{C}(\mathbf{Q}) = \{\mathbf{a} \in [0,1]^X : \sum_{x \in X} a_x \leq 1\}$, the standard n-dimensional simplex, of volume $1/n!$. (Even slightly more complex examples will test the reader's powers of geometric visualization considerably.)

What can we say immediately about the chain polytope $\mathcal{C}(\mathbf{P})$ of a poset $\mathbf{P}$? For one thing, as it is defined purely in terms of the sets forming chains, i.e., the cliques in the comparability graph of $\mathbf{P}$, it is a comparability invariant: another partial order with the same comparability graph will have the same chain polytope.

The chain polytope is a down-set in $\mathbb{R}_+^X$: if $\mathbf{a}$ satisfies the constraints, and $0 \leq b_x \leq a_x$ for each x, then the vector $\mathbf{b}$ also satisfies the constraints. It also has full dimension, and is convex and compact. A body satisfying these conditions is called a *convex corner* in $\mathbb{R}_+^X$.

Let A be an antichain in $\mathbf{P}$. Then the indicator vector $\mathbf{e}^A$ is an element of $\mathcal{C}(\mathbf{P})$: each chain in $\mathbf{P}$ contains at most one element of A, so each constraint is satisfied by $\mathbf{e}^A$. These indicator vectors are certainly vertices of $\mathcal{C}(\mathbf{P})$, as they are vertices of $[0,1]^n$. What is not so obvious is that these are the *only* vertices of $\mathcal{C}(\mathbf{P})$.

To understand this issue, we need to consider a more general notion. What we have done is define a polytope associated with the comparability graph of a partial order, but there is no reason to restrict the definition to comparability graphs.

For any graph $\mathbf{G} = (V, E)$, we define a polytope

$$\mathcal{FS}(\mathbf{G}) = \{\mathbf{a} \in [0,1]^V : \sum_{v \in C} a_v \leq 1 \text{ for every clique } C \text{ of } \mathbf{G}\}.$$

If $\mathbf{G}$ is the comparability graph $\mathrm{Comp}(\mathbf{P})$ of a partial order $\mathbf{P}$, then this is just the chain polytope $\mathcal{C}(\mathbf{P})$. As before, we see that, for a stable set S of $\mathbf{G}$, $\mathbf{e}^S$ is a vertex of $\mathcal{FS}(\mathbf{G})$.

Now let the graph $\mathbf{H}$ be a 5-cycle. The vector $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ is in $\mathcal{FS}(\mathbf{H})$, as maximal cliques only have size 2. However, each vector $\mathbf{e}^S$, for S a stable set in $\mathbf{H}$, has coordinates summing to at most 2, so $(\frac{1}{2}, \dots, \frac{1}{2})$ is not in the convex hull of the vectors $\mathbf{e}^S$. In fact, $(\frac{1}{2}, \dots, \frac{1}{2})$ is a vertex of $\mathcal{FS}(\mathbf{H})$: it is the only other one besides the $\mathbf{e}^S$.

In this more general context, one is led to define two different polytopes. The *stable set polytope* $\mathcal{S}(\mathbf{G})$ of a graph $\mathbf{G}$ is the convex hull of the indicator vectors of its stable sets. This is always contained within the *fractional stable set polytope* $\mathcal{FS}(\mathbf{G})$ of $\mathbf{G}$, which is the polytope defined above. For $\mathbf{G}$ a 5-cycle, this containment is strict.

It is a theorem of Lovász [99] that $\mathcal{FS}(\mathbf{G}) = \mathcal{S}(\mathbf{G})$ if and only if $\mathbf{G}$ is perfect.

Now, the chain polytope of a poset $\mathbf{P}$ is just the fractional stable set polytope of the comparability graph Comp($\mathbf{P}$): comparability graphs are perfect, so this is also the stable set polytope of Comp($\mathbf{P}$). This means that the indicator vectors $\mathbf{e}^A$ of antichains are indeed the only vertices of the chain polytope.

Specialized to the chain polytope, Lovász's result can be restated as saying that every vector in $\mathcal{C}(\mathbf{P})$ can be written as a convex combination of indicator vectors of antichains. In the next section, we shall prove this and more.

8.5 The Laminar Decomposition

In this section, we introduce the idea of the laminar decomposition of an element of $\mathbb{R}_+^X$. This notion was first considered by Kahn and Kim [99].

Recall from Chapter 9.99 that the set of antichains in a partial order $\mathbf{P} = (X, P)$ forms a lattice, identifying the family of antichains with the up-set lattice, ordered by reverse inclusion, so that antichain A is below antichain A' if and only if $U[A] \supset U[A']$. A *laminar family* of antichains is a chain in this lattice.

It is clear that each maximal laminar family consists of exactly $n+1$ antichains $A_0 < \cdots < A_n$: the up-set $U[A_j]$ contains exactly $n-j$ elements, so A_0 is the antichain of minimal elements and A_n is the empty antichain. Moreover, the single element x_j in $U[A_{j-1}] \setminus U[A_j]$ is minimal in $U[A_{j-1}]$, so $x_1 \prec x_2 \prec \cdots \prec x_n$ is a linear extension of P. We can reverse this process: given a linear extension $x_1 \prec x_2 \prec \cdots \prec x_n$, the corresponding laminar family is found by setting $A_j = \mathrm{Min}(\{a_{j+1}, \ldots, a_n\})$, for $j = 0, \ldots, n$. Thus maximal laminar families are in 1-1 correspondence with linear extensions.

Proposition 8.5.1 *For any laminar family $(A_1, \ldots, A_m)$ of non-empty antichains in a poset $\mathbf{P}$, there is a chain C in $\mathbf{P}$ meeting each antichain in the family.*

Proof It's enough to prove the result for a maximal laminar family $(A_0, \ldots, A_{n-1})$, with the empty antichain A_n dropped, corresponding to a linear extension $x_1 \prec x_2 \prec \cdots \prec x_n$.

We construct the chain C by working greedily down the linear extension: we take x_n as the top element of C, then read down, taking x_j into C if it is below the current bottom element of the chain.

Now the antichain A_j is the set of minimal elements among $\{x_j, \ldots, x_n\}$,

Fig. 8.2. A vector in $\mathcal{O}(\mathbf{P})$ and its laminar decomposition.

so it includes the last element from this set taken into the chain C. Hence C meets each A_j, as required. □

For a *maximal* laminar family, inspection of the proof above reveals that the chain C is unique.

A *laminar decomposition* of a vector $\mathbf{a}$ in $\mathbb{R}_+^X$ is a representation

$$\mathbf{a} = \sum_{j=0}^{m} \lambda_j \mathbf{e}^{A_j},$$

where the λ_j are positive real numbers, and $A_0 < A_1 < \cdots < A_m$ is a laminar family of non-empty antichains in P.

Theorem 8.5.2 *Let* $\mathbf{P}$ *be any poset. Every vector* $\mathbf{a}$ *in* $\mathbb{R}_+^X$ *has a unique laminar decomposition*

$$\mathbf{a} = \sum_{j=0}^{m} \lambda_j \mathbf{e}^{A_j}.$$

Furthermore, $\mathbf{a} \in \mathcal{C}(\mathbf{P})$ *if and only if* $\sum_{j=0}^{m} \lambda_j \leq 1$.

Before we give a formal proof, let us examine a simple example. In Figure 8.2, we see a small poset and an element $\mathbf{a}$ of its chain polytope. The laminar decomposition is obtained by successively giving "weight" $\min_{x \in \mathrm{Min}(\mathbf{P})} a_x$ to the antichain of minimal elements, and removing at each stage any elements whose coefficients have become zero. So we take $A_0 = \{x, y\}$ with weight $\lambda_0 = 0.2$, delete y, and decrement a_x by 0.2. Then we give weight $\lambda_1 = 0.1$ to $A_1 = \{x, w\}$, and delete both x and w. Finally we give weight $\lambda_2 = 0.6$ to $A_2 = \{z\}$, and delete z. In this example, we used three antichains rather than the normal four because of the coincidence that two elements were removed simultaneously. If we were to perturb a_x or a_w, we would also need to include either $\{x\}$ or $\{z, w\}$ in our laminar family with positive weight.

The proof below amounts to saying that this process always works, and that we never have any choice.

Proof To establish the existence and uniqueness of the laminar decomposition, we proceed by induction on $n = |X|$. The result is trivial for the one-element poset.

Suppose the result is true for all posets with fewer than n elements. Let $\mathbf{P}$ be a poset on a ground-set X of size n, and take any vector $\mathbf{a} \in \mathbb{R}^X_+$.

If there is some element x of X with $a_x = 0$, then we can delete x from the poset, keeping all other a_y as before, and apply the induction hypothesis to $\mathbf{P} - x$.

Otherwise note that, in any laminar family, there is some minimal element of P that only appears in the antichain $A_0 = \mathrm{Min}(\mathbf{P})$. (This is the bottom element of the corresponding linear extension.) This tells us that any laminar decomposition of $\mathbf{a}$ must have $A_0 = \mathrm{Min}(\mathbf{P})$ and $\lambda_0 = \min_{x \in \mathrm{Min}(\mathbf{P})} a_x$.

Now we set x_1 to be some minimal element of $\mathbf{P}$ with $a_{x_1} = \lambda_0$, delete x_1 from $\mathbf{P}$, and set

$$a'_x = a_x - \lambda_0 e^{A_0}_x,$$

for each $x \neq x_1$. We apply the induction hypothesis to $\mathbf{a}'$, obtaining that this vector has a unique laminar decomposition.

It's clear that a laminar decomposition $\sum_{j=1}^m \lambda_j \mathbf{e}^{A_j}$ of $\mathbf{a}'$ yields a laminar decomposition

$$\lambda_0 \mathbf{e}^{A_0} + \sum_{j=1}^m \lambda_j \mathbf{e}^{A_j}$$

of $\mathbf{a}$, and vice versa. So there is a unique laminar decomposition of $\mathbf{a}$, and we are done by induction.

For the final statement, suppose that $\mathbf{a} = \sum_{j=0}^m \lambda_j \mathbf{e}^{A_j}$ is the laminar decomposition of $\mathbf{a} \in \mathbb{R}^X_+$. We see that, for any chain C, $\sum_{x \in C} a_x \leq \sum_{j=0}^m \lambda_j$, with equality if C hits all of the A_j. Therefore, by Proposition 8.5.1, $\sum_{x \in C} a_x \leq 1$ for all chains C if and only if $\sum_{j=0}^m \lambda_j \leq 1$, as claimed. □

The following is a restatement of the situation when $\mathbf{a} \in \mathcal{C}(\mathbf{P})$.

Corollary 8.5.3 *Every element* $\mathbf{a}$ *of* $\mathcal{C}(\mathbf{P})$ *has a unique representation as a convex combination*

$$\mathbf{a} = \sum_{j=0}^m \lambda_j \mathbf{e}^{A_j},$$

where $\sum_{j=0}^m \lambda_j = 1$*, each* λ_j *is positive, and* $(A_0, \ldots, A_m)$ *is a laminar family of antichains in* $\mathbf{P}$.

Proof The difference between the above convex representation and the laminar decomposition is that we are now required to have the λ_j sum to 1, but we are now permitted to introduce a term $\mathbf{0} = \lambda \mathbf{e}^{\emptyset}$ $(\lambda > 0)$, which clearly enables us to accomplish this. (Recall that the empty set is the maximum element in the antichain lattice.) □

For example, the vector illustrated in Figure 8.2 above can be written as the convex combination

$$0.2\mathbf{e}^{A_0} + 0.1\mathbf{e}^{A_1} + 0.6\mathbf{e}^{A_2} + 0.1\mathbf{e}^{\emptyset}.$$

As promised in the previous section, this result implies that $\mathcal{C}(\mathbf{P})$ is the convex hull of the indicator vectors of antichains in $\mathbf{P}$. But it has much wider implications than that, as we are about to see.

8.6 Stanley's Theorem

We are now ready to state and prove a theorem first proved by Richard Stanley in 1986 [99].

Theorem 8.6.1 *For any n-element poset $\mathbf{P}$,*

$$\mathrm{Vol}(\mathcal{C}(\mathbf{P})) = \frac{e(\mathbf{P})}{n!}.$$

Proof For each linear extension $\prec$, take the corresponding maximal laminar family $(A_0, \ldots, A_n)$, and let $\mathcal{T}(\prec)$ be the convex hull of $\mathbf{e}^{A_0}, \ldots, \mathbf{e}^{A_n}$.

As we saw in Corollary 8.5.3, every vector in $\mathcal{C}(\mathbf{P})$ has a unique representation as a convex combination of indicator vectors of antichains in a laminar family.

Convex hulls of sets of fewer than $n+1$ vectors have measure 0, so, for almost all vectors $\mathbf{a}$ in $\mathcal{C}(\mathbf{P})$, the laminar family used to represent $\mathbf{a}$ is maximal. In other words, apart from a set of measure zero, each vector $\mathbf{a} \in \mathcal{C}(\mathbf{P})$ lies in exactly one of the simplices $\mathcal{T}(\prec)$.

Of course, each $\mathcal{T}(\prec)$ is contained in $\mathcal{C}(\mathbf{P})$. To complete the proof, it suffices to check that each simplex $\mathcal{T}(\prec)$ has volume $1/n!$.

Given a linear extension $\prec$: $x_1 \prec x_2 \prec \cdots \prec x_n$, recall that antichain A_j consists of the minimal elements among $\{x_{j+1}, \ldots, x_n\}$. We start with $\mathbf{e}^{A_n}$ and introduce the other vectors in reverse order. The vectors $\mathbf{e}^{A_j}, \ldots, \mathbf{e}^{A_n}$ all lie in the flat $\{\mathbf{a} \in \mathbb{R}^X : a_{x_1} = \cdots = a_{x_j} = 0\}$ of $\mathbb{R}^X$. The next antichain A_{j-1} includes x_j but none of $x_1, \ldots, x_{j-1}$, so the vector $\mathbf{e}^{A_{j-1}}$ sits at height 1 above the flat containing the convex hull of $\{\mathbf{e}^{A_j}, \ldots, \mathbf{e}^{A_n}\}$. So the volume of the convex hull of $\{\mathbf{e}^{A_{j-1}}, \mathbf{e}^{A_j}, \ldots, \mathbf{e}^{A_n}\}$ is $1/(n-j+1)$ times

Fig. 8.3. A vector in $\mathcal{O}(\mathbf{P})$ and the corresponding vector in $\mathcal{C}(\mathbf{P})$.

the volume of the convex hull of $\{\mathbf{e}^{A_j}, \ldots, \mathbf{e}^{A_n}\}$. Therefore the volume of $\mathcal{T}(\prec)$ is $\prod_{j=n}^{1} 1/(n-j+1) = 1/n!$. □

Another way to carry out this proof is to give a measure-preserving bijection between $\mathcal{O}(\mathbf{P})$ and $\mathcal{C}(\mathbf{P})$. Inevitably, this is done by mapping $\mathcal{O}(\prec)$ to $\mathcal{T}(\prec)$ for each linear extension $\prec$ of P. If $x_1 \prec \cdots \prec x_n$, then $\mathcal{O}(\prec)$ is the convex hull of the indicator vectors of the up-sets $U_j = \{x_{j+1}, \ldots, x_n\}$, for $j = 0, \ldots, n$. So it is natural to construct our map from $\mathcal{O}(\mathbf{P})$ to $\mathcal{C}(\mathbf{P})$ by mapping $\mathbf{e}^U$ to $\mathbf{e}^{\mathrm{Min}(U)}$ for each up-set U, and interpolating linearly inside each simplex $\mathcal{O}(\prec)$.

This map can be described explicitly as follows: given a vector $\mathbf{a}$ in $\mathcal{O}(\mathbf{P})$, set $b_x = a_x$ if x is minimal, and $b_x = a_x - \max_{y<x} a_y$ for all other x. One can easily verify directly that this defines a bijection from $\mathcal{O}(\mathbf{P})$ to $\mathcal{C}(\mathbf{P})$: the inverse map is defined by $a_x = \max_C \sum_{y \in C} b_y$, where C runs over all chains in $\mathbf{P}$ with top element x. One can see immediately that the map is measure-preserving by noting that the matrix representing its action on any $\mathcal{O}(\prec)$ (coordinates in the order given by $\prec$) is upper-triangular with 1s on the diagonal, so has determinant 1.

The simple example in Figure 8.3 shows this map in operation.

8.7 Antiblockers and the Antichain Polytope

We have now related the number of linear extensions of a poset to the volume of the chain polytope, which is a convex corner (convex compact down-set of full dimension) in $\mathbb{R}_+^X$.

The *antiblocker* of a convex corner $\mathcal{K}$ in $\mathbb{R}_+^X$ is the body

$$\mathcal{K}^* = \{\mathbf{b} \in \mathbb{R}_+^X : \mathbf{b} \cdot \mathbf{a} \le 1 \text{ for every } \mathbf{a} \in \mathcal{K}\}.$$

It's easy to check that $\mathcal{K}^*$ is also a convex corner.

It's immediate that $\mathcal{K} \subseteq \mathcal{K}^{**} = (\mathcal{K}^*)^*$. To see the converse, take any $\mathbf{a}$ in $\mathbb{R}_+^X \setminus \mathcal{K}$. As $\mathcal{K}$ is compact and convex, there is some hyperplane separating $\mathbf{a}$ from $\mathcal{K}$; as $\mathcal{K}$ is a down-set, its normal can be taken to have non-negative coefficients. This means that there is some vector $\mathbf{u} \in \mathbb{R}_+^X$ such that $\mathbf{u} \cdot \mathbf{c} \le 1$ for all $\mathbf{c} \in \mathcal{K}$, while $\mathbf{u} \cdot \mathbf{a} > 1$. But then $\mathbf{u} \in \mathcal{K}^*$, so $\mathbf{a} \notin \mathcal{K}^{**}$. Therefore $\mathcal{K}^{**} = \mathcal{K}$.

The notion of an antiblocker is valuable in combinatorial optimization, and it also arises in the study of finite-dimensional normed spaces. To see the connection, define, for $\mathcal{K}$ a convex corner in $\mathbb{R}^n_+$, the *enlargement* $E(\mathcal{K})$ of $\mathcal{K}$ to be the body $\{\mathbf{a} \in \mathbb{R}^n : (|a_1|, |a_2|, \dots, |a_n|) \in \mathcal{K}\}$: $E(\mathcal{K})$ is the set of points that can be obtained by repeatedly reflecting a point of $\mathcal{K}$ in the coordinate planes. The fact that $\mathcal{K}$ is a convex corner is exactly what is required to ensure that $E(\mathcal{K})$ is the unit ball of some norm on $\mathbb{R}^n$. The enlargement $E(\mathcal{K}^*)$ is then the unit ball of the dual norm.

What is the antiblocker of the chain polytope? Let's ask the more general question: what is the antiblocker of the stable set polytope of a graph $\mathbf{G}$? Recall that the stable set polytope $\mathcal{S}(\mathbf{G})$ of $\mathbf{G}$ is the convex hull of indicator vectors $\mathbf{e}^S$ of stable sets S. For a vector $\mathbf{a}$ to be in $\mathcal{S}(\mathbf{G})^*$, it is necessary and sufficient that $\mathbf{a} \cdot \mathbf{e}^S \le 1$ for all the vertices $\mathbf{e}^S$, i.e., that $\sum_{x \in S} a_x \le 1$ for all stable sets S. Translating, this means that $\mathcal{S}(\mathbf{G})^* = \mathcal{FS}(\mathbf{G}^c)$, the fractional stable set polytope of the complement of $\mathbf{G}$. We deduce also that $\mathcal{FS}(\mathbf{G})^* = \mathcal{S}(\mathbf{G}^c)$.

We know that the chain polytope of a poset $\mathbf{P}$ is equal to both $\mathcal{S}(\mathrm{Comp}(\mathbf{P}))$ and $\mathcal{FS}(\mathrm{Comp}(\mathbf{P}))$. So its antiblocker is equal to both $\mathcal{FS}(\mathrm{Incomp}(\mathbf{P}))$ and $\mathcal{S}(\mathrm{Incomp}(\mathbf{P}))$. Explicitly, the antiblocker of the chain polytope is the convex hull of the indicator vectors of chains, and it is also given by

$$\{\mathbf{b} \in [0,1]^X : \sum_{x \in A} b_x \le 1 \text{ for all antichains } A \text{ of } \mathbf{P}\}.$$

This polytope is called the *antichain polytope* $\mathcal{A}(\mathbf{P})$ of $\mathbf{P}$.

Sidorenko's Inequality, Theorem 8.1.2, says that the vector $\mathbf{b}$ defined by $b_x = e(\mathbf{P} - x)/e(\mathbf{P})$ is in $\mathcal{A}(\mathbf{P})$. Showing the existence of the flow in that proof was tantamount to showing that $\mathbf{b}$ is a convex combination of indicator functions of (maximal) chains.

Let's take this a little further, and prove another result from Sidorenko's gem of a paper [99].

Theorem 8.7.1 *Suppose* $\mathbf{P}_1 = (X, P_1)$, $\mathbf{P}_2 = (X, P_2)$, $\mathbf{P}_3 = (X, P_3)$ *are three posets on the same ground-set* X *such that* $\mathrm{Comp}(\mathbf{P}_1) \cap \mathrm{Comp}(\mathbf{P}_2) \subseteq \mathrm{Comp}(\mathbf{P}_3)$. *Then* $e(\mathbf{P}_1)e(\mathbf{P}_2) \ge e(\mathbf{P}_3)$.

Proof We work by induction on n, the result being trivial for $n = 1$.

Suppose that the result is true whenever the common ground-set has fewer than n elements. Take any posets $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ as above, with a common ground-set X of n elements.

We start by noting that

$$e(\mathbf{P}_3) = \sum_{x \in \mathrm{Min}(\mathbf{P}_3)} e(\mathbf{P}_3 - x) \le \sum_{x \in \mathrm{Min}(\mathbf{P}_3)} e(\mathbf{P}_1 - x) e(\mathbf{P}_2 - x),$$

by the induction hypothesis, as $\mathrm{Comp}(\mathbf{P}_1 - x) \cap \mathrm{Comp}(\mathbf{P}_2 - x) \subseteq \mathrm{Comp}(\mathbf{P}_3 - x)$.

Now we define a vector $\mathbf{a} \in [0,1]^X$ by setting $a_x = e(\mathbf{P}_1 - x)/e(\mathbf{P}_1)$ for $x \in \mathrm{Min}(\mathbf{P}_3)$, and $a_x = 0$ otherwise. We claim that $\mathbf{a} \in \mathcal{C}(\mathbf{P}_2)$. To show this, let C be any chain in P_2, and consider

$$\sum_{x \in C} a_x = \sum_{x \in C \cap \mathrm{Min}(\mathbf{P}_3)} \frac{e(\mathbf{P}_1 - x)}{e(\mathbf{P}_1)}.$$

The set $A = C \cap \mathrm{Min}(\mathbf{P}_3)$ is an independent set in $\mathrm{Comp}(\mathbf{P}_3)$ and a clique in $\mathrm{Comp}(\mathbf{P}_2)$. By assumption, A is therefore an independent set in $\mathrm{Comp}(\mathbf{P}_1)$, i.e., an antichain in $\mathbf{P}_1$. By Theorem 8.1.2, $\sum_{x \in A} e(\mathbf{P}_1 - x) \le e(\mathbf{P}_1)$, and so we indeed have that $\mathbf{a} \in \mathcal{C}(\mathbf{P}_2)$.

Also by Theorem 8.1.2, this time applied to $\mathbf{P}_2$, we have that the vector $\mathbf{b}$ defined by $b_x = e(\mathbf{P}_2 - x)/e(\mathbf{P}_2)$ is in $\mathcal{A}(\mathbf{P}_2)$, and therefore $\mathbf{a} \cdot \mathbf{b} \le 1$. Translating, we have:

$$e(\mathbf{P}_3) \le \sum_{x \in \mathrm{Min}(\mathbf{P}_3)} e(\mathbf{P}_1 - x) e(\mathbf{P}_2 - x) = \sum_{x \in X} a_x e(\mathbf{P}_1) b_x e(\mathbf{P}_2) \le e(\mathbf{P}_1) e(\mathbf{P}_2).$$

This completes the proof. □

This theorem (and proof) can be generalized in several different ways, but we will settle for the form above, and the following simple application.

Corollary 8.7.2 *Let $\mathbf{L}_1$, $\mathbf{L}_2$ and $\mathbf{L}_3$ be three linear orders on the set X. Then $e(\mathbf{L}_1 \cap \mathbf{L}_2) e(\mathbf{L}_2 \cap \mathbf{L}_3) \ge e(\mathbf{L}_1 \cap \mathbf{L}_3)$.*

Hence the function $\rho(\mathbf{L}_1, \mathbf{L}_2) = \log e(\mathbf{L}_1 \cap \mathbf{L}_2)$ is a metric on the set of linear orders of X.

Proof All we need to check is that, if two elements are comparable in both $\mathbf{L}_1 \cap \mathbf{L}_2$ and $\mathbf{L}_2 \cap \mathbf{L}_3$, then they are comparable in $\mathbf{L}_1 \cap \mathbf{L}_3$. □

8.8 Bounds on $e(\mathbf{P})$ from the antichain polytope

Theorem 8.8.1 *Let $\mathbf{P} = (X, P)$ be any poset, and let $\mathbf{b}$ be any vector in $\mathcal{A}(\mathbf{P})$. Then $e(\mathbf{P}) \le \prod_{x \in X} 1/b_x$.*

Proof We know that the chain polytope $\mathcal{C}(\mathbf{P})$ is contained in $\{\mathbf{a} \in \mathbb{R}_+^X : \mathbf{a}\cdot\mathbf{b} \le 1\}$. This is a standard simplex, with a vertex on the x-axis at distance $1/b_x$ from the origin, so it has volume

$$\frac{1}{n!}\prod_{x\in X}\frac{1}{b_x}.$$

So, by Stanley's Theorem, $e(\mathbf{P}) = n!\,\mathrm{Vol}(\mathcal{C}(\mathbf{P})) \le \prod_{x\in X} 1/b_x$, as claimed. □

There is a second amusing proof of this result, using absolutely none of the theory we have developed.

Proof Consider the following random procedure for building a linear extension of $\mathbf{P}$ from the bottom up. At each stage, from the set of available elements, choose the next element to be x with probability proportional to b_x.

Now consider any single linear extension $\prec$ of $\mathbf{P}$, say $x_1 \prec x_2 \prec \cdots \prec x_n$. The probability that our random procedure results in $\prec$ is exactly

$$\prod_{j=1}^{n}\frac{b_{x_j}}{\sum_{y\in A_j} b_y},$$

where A_j is the set of elements minimal among $\{x_j,\ldots,x_n\}$. Since the A_j are antichains, and $\mathbf{b} \in \mathcal{A}(\mathbf{P})$, each sum $\sum_{y\in A_j} b_y$ is at most 1, and so the probability of $\prec$ is at least $\prod_{j=1}^n b_{x_j} = \prod_{x\in X} b_x$, a quantity independent of $\prec$.

We conclude that the number of linear extensions is indeed at most $\prod_{x\in X} 1/b_x$. □

Theorem 8.8.1 gives us a family of upper bounds on the number of linear extensions of a poset $\mathbf{P}$, one for each vector in the antichain polytope. Identifying the best of these bounds amounts to maximizing $\prod_{x\in X} b_x$, for $\mathbf{b} \in \mathcal{A}(\mathbf{P})$.

For a convex corner $\mathcal{K}$ in $\mathbb{R}_+^X$, set

$$V(\mathcal{K}) = \max_{\mathbf{a}\in\mathcal{K}}\prod_{x\in X} a_x,$$

and call any vector $\mathbf{a}$ achieving this maximum an *optimal point* of $\mathcal{K}$. This parameter $V(\mathcal{K})$ is closely related to the *entropy* $H(\mathcal{K})$ of $\mathcal{K}$; specifically $V(\mathcal{K}) = 2^{-nH(\mathcal{K})}$. What we saw in Theorem 8.8.1 is that $\mathrm{Vol}(\mathcal{K}) \le 1/(n!V(\mathcal{K}^*))$.

We'll discuss the practical side of the problem of finding $V(\mathcal{K})$ in Section 8.12. From a theoretical standpoint, the following result of Csiszar, Körner, Lovász, Marton and Simonyi [99] gives us a very clean method to verify that we have a solution.

Proposition 8.8.2 *For any antiblocking pair $(\mathcal{K}, \mathcal{K}^*)$ of convex corners in $\mathbb{R}^X_+$, the pair $\mathbf{a} \in \mathcal{K}$, $\mathbf{b} \in \mathcal{K}^*$ are optimal points of their respective polytopes if and only if $a_x b_x = 1/n$ for each $x \in X$.*

Therefore $V(\mathcal{K})V(\mathcal{K}^*) = 1/n^n$.

In particular, every convex corner has exactly one optimal point.

Proof Suppose $\mathbf{a} \in \mathcal{K}$ and $\mathbf{b} \in \mathcal{K}^*$. Then

$$\left(\prod_{x \in X} a_x b_x\right)^{1/n} \leq \frac{1}{n} \sum_{x \in X} a_x b_x \leq \frac{1}{n}.$$

The first inequality above is the inequality of the arithmetic and the geometric mean; equality holds if and only if all the terms $a_x b_x$ are equal. This shows that $V(\mathcal{K})V(\mathcal{K}^*) \leq 1/n^n$; if equality is achieved, then the optimum points $\mathbf{a}$ and $\mathbf{b}$ satisfy $a_x b_x = 1/n$ for each $x \in X$.

All that remains to be shown is that there are vectors $\mathbf{a}$, $\mathbf{b}$ in $\mathcal{K}$ and $\mathcal{K}^*$ respectively with $a_x b_x = 1/n$ for every $x \in X$.

Suppose $\mathbf{a}$ is an optimal point of $\mathcal{K}$. The normal vector to the surface $\{\mathbf{c} : \prod_{x \in X} c_x = V(\mathcal{K})\}$ at the point $\mathbf{c} = \mathbf{a}$ is in the direction of $\mathbf{b}$, where $b_x = 1/(na_x)$. So the convex body $\mathcal{K}$ lies under the tangent hyperplane $\{\mathbf{c} : \mathbf{c} \cdot \mathbf{b} = 1\}$. This means that $\mathbf{b}$ is in $\mathcal{K}^*$, as required. □

A consequence is the following result, due to Kahn and Kim [99].

Corollary 8.8.3 *For any poset $\mathbf{P}$, $n! V(\mathcal{C}(\mathbf{P})) \leq e(\mathbf{P}) \leq n^n V(\mathcal{C}(\mathbf{P}))$.*

Proof If $\mathbf{a}$ is the optimal point of $\mathcal{C}(\mathbf{P})$, then certainly $\mathcal{C}(\mathbf{P})$ contains the box with top point $\mathbf{a}$, of volume $V(\mathcal{C}(\mathbf{P}))$, so the lower bound follows from Theorem 8.6.1, Stanley's Theorem.

The upper bound is obtained by combining Theorem 8.8.1 and Proposition 8.8.2: $e(\mathbf{P}) \leq 1/V(\mathcal{A}(\mathbf{P})) = n^n V(\mathcal{C}(\mathbf{P}))$. □

8.9 2-Dimensional Posets

Normally, there is no particular interpretation of the volume of the antichain polytope. However, in the special case where the poset $\mathbf{P}$ has dimension 2, the antichain polytope of $\mathbf{P}$ is just the chain polytope of any complement $\overline{\mathbf{P}}$ of $\mathbf{P}$.

There is a trade-off: the larger $\mathcal{C}(\mathbf{P})$ is, the smaller $\mathcal{A}(\mathbf{P}) = \mathcal{C}(\overline{\mathbf{P}})$. Alternatively, the larger $e(\mathbf{P})$ is, the smaller $e(\overline{\mathbf{P}})$. What can we say about $e(\mathbf{P})e(\overline{\mathbf{P}})$? Bounds on this product can be obtained by combining Proposition 8.8.2 and Corollary 8.8.3, but we can do better.

For a start, we can apply Theorem 8.7.1 with $\mathbf{P}_1 = \mathbf{P}$, $\mathbf{P}_2 = \overline{\mathbf{P}}$, and $\mathbf{P}_3$ an antichain: we get the following result, due again to Sidorenko [99].

Theorem 8.9.1 *For $\mathbf{P} = (X, P)$ a 2-dimensional poset with $|X| = n$,*

$$e(\mathbf{P})e(\overline{\mathbf{P}}) \geq n!.$$

This is best possible, as can be seen by taking P and $\overline{P}$ to be a chain and an antichain. See Exercise 9 for a much wider class of extremal examples, namely all *series-parallel posets*.

In fact, the lower bound above holds in a much more general context.

Theorem 8.9.2 (Saint-Raymond's Theorem) *For $\mathcal{K}$ a convex corner in $\mathbb{R}^n_+$,*

$$\mathrm{Vol}(\mathcal{K})\,\mathrm{Vol}(\mathcal{K}^*) \geq 1/n!.$$

This result implies Theorem 8.9.1 via Stanley's Theorem, taking $\mathcal{K}$ to be the chain polytope of $\mathbf{P}$ and therefore $\mathcal{K}^*$ to be the chain polytope of $\overline{\mathbf{P}}$. The extreme cases of Saint-Raymond's Theorem are known, and turn out to be, up to scaling, exactly the chain polytopes of series-parallel orders.

The best known upper bound on the product $e(\mathbf{P})e(\overline{\mathbf{P}})$ is again a consequence of an extremal theorem from the geometry of $\mathbb{R}^n$.

Theorem 8.9.3 (Santaló's Theorem) *For $\mathcal{B}$ the unit ball of a norm on $\mathbb{R}^n$, and $\mathcal{B}^\circ$ the unit ball of the dual norm,*

$$\mathrm{Vol}(\mathcal{B})\,\mathrm{Vol}(\mathcal{B}^\circ) \leq \mathrm{Vol}(\mathcal{B}_2)^2,$$

where $\mathcal{B}_2$ is the unit ball of the ℓ_2 (Euclidean) norm on $\mathbb{R}^n$.

Corollary 8.9.4 *For $\mathbf{P}$ a 2-dimensional poset with n elements, and $\overline{\mathbf{P}}$ any complement,*

$$e(\mathbf{P})e(\overline{\mathbf{P}}) \leq (1 + o(1))n!\left(\frac{\pi}{2}\right)^n \sqrt{\frac{2}{\pi n}}.$$

Proof We apply Santaló's Inequality to the enlargement $E(\mathcal{K})$: the right-hand quantity is an estimate for the volume of the intersection of the Euclidean unit ball with the positive quadrant. $\square$

This bound seems unlikely to be close to best possible. Bollobás, Brightwell and Sidorenko [99] show that the supremum

$$\sup_{\mathbf{P} \text{ a dimension-2 poset}} \left(\frac{e(\mathbf{P})e(\overline{\mathbf{P}})}{n!} \right)^{1/n}$$

is at least some constant $\psi \simeq 1.123$, which is some way short of $\pi/2$.

8.10 LYM Posets

In Chapter 9.99, we studied ranked posets, and picked out the class of Sperner posets and LYM posets as being of special interest. Here we see how the LYM condition manifests itself in the context of this chapter.

Let $\mathbf{P} = (X, P)$ be a ranked poset with ranks $A_1, A_1, \ldots, A_h$ of sizes (*rank numbers*) $n_1, n_2, \ldots, n_h$ respectively. So each maximal chain is of the form $x_1 < x_2 < \cdots < x_h$, where $x_i \in A_i$ for $i = 1, \ldots, h$.

We define the *weight* w_x of an element $x \in A_i$ to be $1/n_i$, and the weight $w(Y)$ of a set $Y \subseteq X$ to be the sum $\sum_{x \in Y} w_x$ of the weights of its members.

The *LYM condition* is that the weight $w(A)$ of any antichain A of $\mathbf{P}$ is at most 1. (Note that equality is achieved for each of the ranks A_i.) A *LYM poset* is one satisfying the LYM condition.

We saw in Chapter 9.99 that the subset lattice $\mathbf{2^t}$ is a LYM poset. More generally, it is known [99] that Cartesian products of chains are LYM posets.

The LYM condition should look familiar: it says exactly that the weight vector $\mathbf{w}$ is in the antichain polytope!

In fact, $\mathbf{w}$ is always the optimal point of the antichain polytope. To see this, define $\mathbf{u}$ by setting $u_x = n_i/n$ for each $x \in A_i$; for any maximal chain C, $\sum_{x \in C} u_x = \sum_{i=1}^{n} (n_i/n) = 1$, so $\mathbf{u}$ is in the chain polytope and $u_x w_x = 1/n$ for each $x \in X$. By Proposition 8.8.2, this means that $\mathbf{w}$ is the optimal point of $\mathcal{A}(\mathbf{P})$ (and $\mathbf{u}$ is the optimal point of $\mathcal{C}(\mathbf{P})$). Thus we have $V(\mathcal{A}(\mathbf{P})) = \prod_{x \in X} w_x = \prod_{i=1}^{h} (1/n_i)^{n_i}$.

In 1974, Kleitman [99] proved that the LYM condition was equivalent to various others, one of which is the existence of a *regular covering* of $\mathbf{P}$ by chains, i.e., a non-empty collection of maximal chains such that, for each i, every element of rank i occurs in the same number of chains. A moment's thought reveals that this says that the vector $\mathbf{y}$ given by $y_j = 1/n_i$ for $j \in A_i$ can be written as a convex combination of indicator vectors of chains—the polyhedral theory assures us that this is equivalent to $\mathbf{y}$ being in $\mathcal{A}(\mathbf{P})$, which is exactly the LYM condition.

Let's turn to the problem of estimating the number of linear extensions of a LYM poset. Before applying the theory developed in this chapter, let us see what is trivial.

So, let $\mathbf{P}$ be a ranked poset with ranks $A_1, \dots, A_h$ of sizes $n_1, \dots, n_h$. To begin with, a LYM poset is Sperner, so we have the upper bound $e(\mathbf{P}) \le \mathrm{width}(\mathbf{P})^n = (\max_{i=1}^h n_i)^n$.

We also have the lower bound

$$e(\mathbf{P}) \ge \prod_{i=1}^{h} n_i!, \tag{8.2}$$

as this is the number of linear extensions in which all of rank A_i comes below all of rank A_{i+1}, for each i.

In many cases, these trivial bounds are not too far apart; for instance, if $\mathbf{P} = \mathbf{2^t}$, the two bounds turn out to differ by a factor of roughly $e^{3n/2}$. For many purposes, this is actually quite a narrow range.

The following improvement on the trivial upper bound is instant from Theorem 8.8.1, using the fact that the weight vector $\mathbf{w}$ is in the antichain polytope. (As $\mathbf{w}$ is the optimal point, this is always the best bound available via this method.)

Theorem 8.10.1 *Let $\mathbf{P}$ be a LYM poset, with rank numbers $n_1, \dots, n_h$. Then*

$$e(\mathbf{P}) \le \prod_{i=1}^{h} n_i^{n_i}.$$

This upper bound is now within a factor of e^n of the trivial lower bound for *all* LYM posets.

For general LYM posets, one can't hope to do much better than the upper bound in Theorem 8.10.1. The disjoint union $\mathbf{Q}(m,h)$ of m chains of height h is a LYM poset with $n = mh$ elements, and $e(\mathbf{Q}(m,h)) = n!/h!^m \ge m^n(2\pi h)^{-m/2}$, while our bound is simply $e(\mathbf{Q}(m,h)) \le m^n$.

However, for specific LYM posets of interest, the upper bound in Theorem 8.10.1 turns out not to be particularly tight. To have some perspective, we write, for a LYM poset $\mathbf{P}$ with rank-sizes $n_1, \dots, n_h$ summing to n, $e(\mathbf{P}) = q^n \prod_{i=1}^h n_i!$. So we are interested in determining, for specific families of LYM posets $\mathbf{P}$, where in the range $[1, e]$ the parameter $q = q(\mathbf{P})$ lies.

The most obvious special case is that of the subset lattice $\mathbf{2^t}$: Brightwell and Tetali [99] prove the following result.

Fig. 8.4. A Young tableau of shape (4,2,2,1), and the hook H_{21} of length 4.

Theorem 8.10.2

$$e(\mathbf{2^t}) \le \prod_{i=0}^{t} \binom{t}{i} \exp\left(2^t \frac{6 \log t}{t}\right).$$

Translating, this means that $q(\mathbf{2^t}) \le 1 + O(\log t/t)$, as $t \to \infty$. In other words, it is the trivial lower bound that is close to the truth. Brightwell and Tetali actually prove a more general result, showing (very roughly) that a height-h LYM poset $\mathbf{P}$ with $q(\mathbf{P})$ much above 1 has to be extremely sparse, with each element above at most about $\log h$ elements in the rank below.

8.11 Linear Extensions of the Grid: Young Tableaux and the Hook Formula

One interesting LYM poset that is sparse is the 2-dimensional grid, i.e., the product of two chains. Counting the linear extensions of the grid turns out to be a well-known problem, with a remarkable solution, in disguise.

For a positive integer n, let $\lambda = (\lambda_1, \ldots, \lambda_m)$ be a partition of n, i.e., a non-increasing sequence of positive integers summing to n. The *Young diagram* or *Ferrers diagram* $X(\lambda)$ of shape λ is an array of cells $x_{i,j}$, for $i = 1, \ldots, m$, $j = 1, \ldots, \lambda_i$. The diagram is traditional drawn as in Figure 8.4, so that there are m rows, the largest at the top.

A *Young tableau* of shape λ is an assignment of the integers $1, \ldots, n$ to the cells in the diagram of shape λ so that all rows and columns form increasing sequences; see Figure 8.4.

It is apparent that a Young tableau of shape λ is just a linear extension of a particular poset $\mathbf{Y}(\lambda)$ on the set $X(\lambda)$ of cells, namely that obtained by putting $x_{ij} \le x_{k\ell}$ if $i \le k$ and $j \le \ell$. (Warning: the minimum of $\mathbf{Y}(\lambda)$ is in the top left corner of the diagram.) Such a poset is a down-set in the *grid* $[m] \times [\lambda_1]$. The grid is a LYM poset, though $\mathbf{Y}(\lambda)$ need not be.

Remarkably, $e(\mathbf{Y}(\lambda))$, the number of Young tableaux with shape λ, has an exact formula. For a cell (i,j), the *hook* H_{ij} consists of the cells that are either below (i,j) in column j or to the right of (i,j) in row i, along with (i,j) itself. The *hook length* h_{ij} is the number of cells in the hook H_{ij}. A typical hook, of length 4, is shown in Figure 8.4.

We are now able to state the famous result of Frame, Robinson and Thrall [99].

Theorem 8.11.1 (The Hook Formula) *For any partition λ of n,*

$$e(\mathbf{Y}(\lambda)) = \frac{n!}{\prod_{(i,j)\in X(\lambda)} h_{ij}}.$$

It is natural to think that there should be a proof of Theorem 8.11.1 proceeding by defining independent events $A(i,j)$ in the space of all linear orders $\prec$ on $[n]$, with $\Pr(A(i,j)) = 1/h_{ij}$, and so that $\prec$ is a linear extension of $\mathbf{Y}(\lambda)$ if and only if all the $A(i,j)$ occur. But no such proof is known, and indeed there are no truly simple and explanatory proofs.

For the particular case of the $m \times m$ grid $\mathbf{G}_m$, the hook formula yields

$$e(\mathbf{G}_m) = \frac{m^2!}{\prod_{i=1}^{m} i^i \prod_{i=1}^{m-1} (2m-i)^i},$$

and the right hand side can be approximated as

$$m^{m^2} \left(\frac{\sqrt{e}}{4} + o(1) \right)^{m^2},$$

whereas the trivial lower bound works out to $m^{m^2} e^{-\frac{3}{2}m^2 + o(m^2)}$, so $q(\mathbf{G}_m) \to e^2/4$ as $m \to \infty$. It is interesting to note that this number is below e even for such a sparse LYM poset.

8.12 #P-completeness

For the rest of this chapter, we'll turn our attention to the computational question associated with $e(\mathbf{P})$: given a poset $\mathbf{P}$, how easy is it to count, or estimate, the number $e(\mathbf{P})$ of linear extensions? It is easy to see that the algorithm suggested by (8.1) takes exponential time in general, but what can be done in polynomial time?

We start with a negative result, due to Brightwell and Winkler [**?**]:

Theorem 8.12.1 *Computing $e(\mathbf{P})$ is a #P-complete problem.*

The complexity class #P is the "counting" equivalent of NP. The counting versions of all the most familiar NP-complete problems are complete for #P, so that a polynomial time algorithm for any one of them would entail a polynomial time algorithm for every problem in #P. This is regarded as

Fig. 8.5. The poset $\mathbf{Q}$. The ovals represent antichains of size $p-1$. Only the clause elements for the clause $c = x\overline{y}$ are shown.

highly implausible: a polynomial time algorithm for a #P-complete problem would imply not just P=NP, but the collapse of the entire polynomial hierarchy.

Proof (Sketch) We'll give just the main ideas of the proof, sliding over all the details and technicalities. The reader is invited to fill in all the gaps in Exercise 15.

The plan is to show that an oracle for counting linear extensions of posets can be used to calculate the number $N(F)$ of satisfying assignments of a 2-SAT formula F: this latter problem is known to be #P-complete.

Let us suppose then that we are given a 2-SAT formula F, i.e., a set of m variables and a set of n clauses, each containing exactly two literals (either a variable x or its negation $\overline{x}$). We wish to count the number $N(F)$ of satisfying assignments, i.e., the assignments of True/False to each variable so that each clause contains a True literal (a negated literal is True if and only if the corresponding variable is False). The first observation is that it is enough to count $N(F)$ modulo p, for many different primes p.

Given F and a large prime p, we form the poset $\mathbf{Q} = \mathbf{Q}_F(p)$ as in Figure 8.5. In this Figure, the ovals represent antichains of size $p-1$; there is one antichain U_x for each variable x, one antichain V_c for each clause c, and one extra antichain U_0, V_0 of each type. There are two special elements a and b, and elements x and $\overline{x}$ for each variable x, which we identify with the literals. There are four *clause elements* $c_1, \ldots, c_4$ of $\mathbf{Q}$ for each clause c: if x and y are the two variables involved in c, then one of these four elements is placed above each of the four pairs $(x, y), (x, \overline{y}), (\overline{x}, y), (\overline{x}, \overline{y})$, as shown. Only the clause element corresponding to the two literals in c is placed above the special element b.

For any partition (A, B, C) of the elements other than a and b, let $E(A, B, C)$ be the set of linear extensions $\prec$ of $\mathbf{Q}$ in which $A \prec a \prec B \prec b \prec C$. Note that $|E(A, B, C)| = e(\mathbf{Q}_A)e(\mathbf{Q}_B)e(\mathbf{Q}_C)$.

Suppose A does not contain either x or $\overline{x}$, for some variable x. Then the elements of $U_x \cup U_0$ are all isolated in $\mathbf{Q}_A$, so $e(\mathbf{Q}_A)$ is divisible by $(2p–2)!$, and thus by p. On the other hand, suppose that A contains at least one of each pair of literals, and k complete pairs ($0 \le k \le n$). Then (see Exercise 14):

$$e(\mathbf{Q}_A) = \frac{(np+p-1+k)!}{p^n((p+1)/2)^k},$$

which is divisible by p if $k > 0$. If $k = 0$ and $p > n$, then $e(\mathbf{Q}_A)$ is not divisible by p.

So, if p does not divide $e(\mathbf{Q}_A)$, then A contains exactly one of each pair of literals. The reader familiar with reductions from SAT will recognize this as one of the main levers: a choice of one of each pair of literals corresponds to a truth assignment—here it turns out that we need to regard the literals in A as being set to False—and we can restrict attention to partitions (A, B, C) in which A does encode a truth assignment, as all other choices do not contribute modulo p.

In a similar way, one can show that, if p does not divide $e(\mathbf{Q}_B)$, then B contains no literal, and exactly one clause element for each clause. Given A, when is there such a B? For the clause $c = x\overline{y}$ illustrated in Figure 8.5, if one of $(x, \overline{x})$ and one of $(y, \overline{y})$ are pushed up into C, then three of the four clause elements c_i are pushed up with them. Thus there is just one c_i that could be in B, provided that c_2 is one of the three elements pushed up, in other words, provided that the set of literals in C includes one appearing in the clause c, which in turn means that the encoded truth assignment satisfies the clause c.

Generally, let $\mathcal{S}$ be the set of partitions (A, B, C) such that: $E(A, B, C)$ is non-empty, each pair of literals has one element in A and the other in C, and each clause has one of its elements in B and the other three in C. Arguing as above, we see that $\mathcal{S}$ is in 1-1 correspondence with the set of satisfying assignments of F.

Finally, for each partition $(A, B, C) \in \mathcal{S}$, $e(\mathbf{Q}_C)$ is a fixed poset $\mathbf{Q}_0$, independent of p. Provided p is large enough and does not divide $e(\mathbf{Q}_0)$, we then have that $|E(A, B, C)|$ is equal to some fixed and readily calculable k, for all $(A, B, C) \in \mathcal{S}$, and that $e(\mathbf{Q}) = k|\mathcal{S}| = kN(F)$, modulo p. So, if we have a method of finding $e(\mathbf{Q})$, we can also find $N(F)$, modulo p, for all but a few primes p, as required. □

A consequence of this result, together with the fact that $e(\mathbf{P}) = \mathrm{Vol}(\mathcal{O}(\mathbf{P}))$, is that computing the exact volume of an n-dimensional polytope is #P-hard, even if the polytope is specified by at most quadratically many inequalities of the form $a_j - a_i \ge 0$. It is also known that counting the number of up-sets in a partial order is #P-hard, and again this gives as a consequence that counting the vertices of an n-dimensional polytope is hard, even in these very special and apparently benign circumstances.

Theorem 8.12.1 shows that it is hard to compute $e(\mathbf{P})$ in general, but one can still ask about special classes of posets. For instance, if $\mathbf{P}$ happens to be an up-set in a grid poset, then the hook formula of Section 8.11 enables us to calculate $e(\mathbf{P})$ efficiently. If we have a bound on the width of $\mathbf{P}$, then again there is a polynomial time algorithm to compute $e(\mathbf{P})$ (see Exercise 13). One final class for which there is a polynomial time algorithm is the class of posets whose covering graph is a tree; see Atkinson [99]. Brightwell and Winkler [99] indicate how to adapt their proof of Theorem 8.12.1 to show that it is #P-complete to compute $e(\mathbf{P})$ for posets $\mathbf{P}$ of height 3, but it is an open (and intriguing) problem whether the problem is still hard for posets of height 2.

8.13 Randomized Approximation Algorithms

In the previous section, we showed that computing $e(\mathbf{P})$ exactly is hard, and deduced that computing volumes of polytopes exactly is hard. Going the other way, there are celebrated positive results about *approximating* the volume of a convex polytope, starting with the seminal paper of Dyer, Frieze and Kannan [99], and of course these immediately yield positive results about approximating the number of linear extensions of a partial order.

Instead of discussing these results, we shall focus on a variant tailored specifically to the problem of counting linear extensions. Our account follows that of Karzanov and Khachiyan [99] from 1991.

We'll start by considering a slightly different problem. Given $\mathbf{P}$, how can we select a linear extension (nearly) uniformly at random from $\mathbf{P}$? It's not hard to see that, if we can solve the (approximate) counting problem in polynomial time, then we can also solve this problem in polynomial time. As we shall see later, the converse is also true.

There is a now well-established method for constructing an algorithm to generate an object approximately uniformly at random from a large set S whose size we do not know. The idea is to define a Markov chain whose state space is S, and whose stationary distribution is the uniform distribution. If the chain is ergodic (connected and aperiodic), then the distribution of the chain after N steps converges to the uniform distribution as $N \to \infty$; the algorithm is simply to start at an arbitrary point of the state space, and run the chain for sufficiently many steps.

However, just finding a chain with the right limiting distribution is not enough: we need to know that the chain is *rapidly mixing*: its convergence to near-stationarity takes place in a small enough number of steps. It's important that "small enough" refers to a number of steps polynomial not in the size of S, but (typically) in $\log |S|$. When $S = E(\mathbf{P})$, for instance, we wish to run the chain only for a number of steps that is polynomial in the number n of elements of $\mathbf{P}$, even though $E(\mathbf{P})$ is normally at least exponential in n.

For any poset $\mathbf{P}$, we define a graph $GE(\mathbf{P})$ on $E(\mathbf{P})$ by declaring two linear extensions $\prec$ and $\prec'$ to be adjacent if one can be obtained from the other by swapping a pair of consecutive elements, i.e., for some $k \in \{1, \ldots, n-1\}$, we have $x_1 \prec \cdots \prec x_k \prec x_{k+1} \prec \cdots \prec x_n$ and $x_1 \prec' \cdots \prec' x_{k+1} \prec' x_k \prec' \cdots \prec' x_n$. Let $d(\prec)$ be the degree of $\prec$ in $GE(\prec)$.

We define the *Karzanov-Khachiyan Markov chain* with state space $E(\mathbf{P})$ by specifying the transition matrix M:

$$m_{\prec,\prec'} = \begin{cases} 1/(2n-2) & \text{if } \prec \text{ and } \prec' \text{ are adjacent} \\ 1 - d(\prec)/(2n-2) & \text{if } \prec = \prec' \\ 0 & \text{otherwise.} \end{cases}$$

This has a very natural interpretation. If we are in the state $\prec$: $x_1 \prec x_2 \prec \cdots \prec x_n$, we choose, uniformly at random, an index j in $\{1, \ldots, n-1\}$, and consider the pair (x_i, x_{i+1}). If these two elements are comparable, necessarily with $x_i < x_{i+1}$, then we stay at the current state. If the two elements are incomparable, then we flip a fair coin; if it comes up Heads, we again stay at the current state, while if it comes up Tails then we move to the state obtained by swapping x_i and x_{i+1}.

It is fairly easy to check that the Karzanov-Khachiyan chain is connected and aperiodic. The fact that M is symmetric means that the chain is time-reversible and its stationary distribution is uniform. See the Exercises. If the probability vector $\mathbf{a}^t$ represents the distribution of the state after t steps, then we have $\mathbf{a}^t = M^t \mathbf{a}^0$. The initial distribution $\mathbf{a}^0$ is usually taken to be the indicator vector of some arbitrary state.

We say that a Markov chain with symmetric transition matrix M has *mixing time* at most $\tau(\varepsilon)$ if, for any $\varepsilon > 0$, any $t > \tau(\varepsilon)$, any initial distribution $\mathbf{a}^0$ and any subset A of the state space S,

$$\left| \Pr(\mathbf{a}^t \in A) - \frac{|A|}{|S|} \right| < \varepsilon.$$

Theorem 8.13.1 *For any n-element poset $\mathbf{P}$, the Karzanov-Khachiyan Markov chain has mixing time at most $n^6 \log n \log(1/\varepsilon)$.*

To prove this theorem, we shall invoke two results without proof, one giving a sufficient condition for rapid mixing and the other a geometric lemma.

What stops a Markov chain being rapidly mixing is some sort of "bottleneck": a partition of the state space S into A and A^c with only a small probability of transitions between A and A^c. Jerrum and Sinclair [99] made this idea precise by introducing the notion of the *conductance* of a Markov chain. When the transition probabilities are $m_{i,j}$ and the stationary distribution is uniform, the conductance of a subset A of S is given by

$$\Phi(A) = \frac{|S|}{|A||A^c|} \sum_{i \in A, j \in A^c} m_{i,j},$$

and the *conductance* Φ of the chain is the minimum conductance of any non-trivial subset of S.

Theorem 8.13.2 *For any Markov chain with state space S, uniform stationary distribution and conductance Φ, the mixing time is at most*

$$\frac{1}{\Phi^2} \log |S| \log(1/\varepsilon).$$

The following geometric result is due to Dyer and Frieze [99]. It is a strong form of the statement that a convex body does not have a bottleneck.

Lemma 8.13.3 *Let $\mathcal{K}$ be a convex body in $\mathbb{R}^n$, with diameter D, separated into two pieces $\mathcal{K}_1$ and $\mathcal{K}_2$ by a hypersurface $\mathcal{H}$. Then*

$$\mathrm{Vol}_{n-1}(\mathcal{H}) \geq \frac{4}{D} \frac{\mathrm{Vol}(\mathcal{K}_1)\,\mathrm{Vol}(\mathcal{K}_2)}{\mathrm{Vol}(\mathcal{K})}.$$

We are now ready to prove Theorem 8.13.1.

Proof Of course, we will proceed by showing that the Karzanov-Khachiyan Markov chain has high conductance. The underlying idea is to think of the state space as made up of the simplices $\mathcal{O}(\prec)$ making up the order polytope $\mathcal{O}(\mathbf{P})$, and the chain as stepping across facets from one simplex to a neighbor. Now, essentially, the fact that the order polytope is convex means that there cannot be a bottleneck.

Accordingly, let (A, B) be any non-trivial partition of $E(\mathbf{P})$. The conductance of A is

$$\Phi(A) = \frac{e(\mathbf{P})}{|A||B|}\frac{M}{2n-2}, \tag{8.3}$$

where M is the number of edges of $GE(\mathbf{P})$ going between A and B.

We consider the corresponding subsets $\mathcal{A} = \bigcup_{\prec \in A} \mathcal{O}(\prec)$ and $\mathcal{B} = \bigcup_{\prec \in B} \mathcal{O}(\prec)$ of the order polytope. Let $\mathcal{H}$ be the hypersurface separating $\mathcal{A}$ from $\mathcal{B}$. We now apply Lemma 8.13.3 with $\mathcal{K} = \mathcal{O}(\mathbf{P})$, $\mathcal{K}_1 = \mathcal{A}$ and $\mathcal{K}_2 = \mathcal{B}$, obtaining:

$$\mathrm{Vol}_{n-1}(\mathcal{H}) \geq \frac{4}{\sqrt{n}}\frac{|A||B|}{e(\mathbf{P})n!}. \tag{8.4}$$

The boundary hypersurface $\mathcal{H}$ is the essentially disjoint union of sets of the form $\mathcal{O}(\prec) \cap \mathcal{O}(\prec')$, where $\prec \in A$, $\prec' \in B$, and $\prec$ and $\prec'$ are adjacent. The $(n{-}1)$-dimensional volume of any such intersection is equal to $\sqrt{2}/(n{-}1)!$, as this is just a simplex in the hyperplane where the two swapped coordinates are equal. So

$$\mathrm{Vol}_{n-1}(\mathcal{H}) = \frac{M\sqrt{2}}{(n{-}1)!}. \tag{8.5}$$

Combining (8.3), (8.4) and (8.5) yields

$$\Phi(A) \geq \frac{4(n-1)!}{\sqrt{2}(2n{-}2)n!\sqrt{n}} \geq \frac{1}{n^{5/2}},$$

for any A, so also $\Phi \geq n^{-5/2}$. The theorem now follows from Theorem 8.13.2, using the bound $\log e(\mathbf{P}) \leq n \log n$. □

This is probably the most straightforward and attractive approach to proving polynomial mixing time, but the power n^6 is not particularly good. Bubley and Dyer [99], and Wilson [99] improved Theorem 8.13.1 by using a technique called *path coupling*, invented by Bubley and Dyer. They prove the following result.

Theorem 8.13.4 *For any n-element poset $\mathbf{P}$, the Karzanov-Khachiyan Markov chain has mixing time $O(n^3 \log n \log(1/\varepsilon))$.*

We'll indicate briefly how to use the results above to create a polynomial-time randomized algorithm to approximate $e(\mathbf{P})$. Again, we'll just give the basic idea, omitting the details in the calculation and not trying to get the best known exponent in the running time.

Set $\mathbf{P}_0 = \mathbf{P}$. Given $\mathbf{P}_k$, we let (x_k, y_k) be any pair of incomparable elements, and form $\mathbf{P}_{k+1}$ by adding to P_k the comparability $x < y$ and taking the transitive closure. We continue until we reach a linear order $\mathbf{P}_m$, and certainly $m \le \binom{n}{2}$. Then

$$e(\mathbf{P}) = \frac{e(\mathbf{P}_0)}{e(\mathbf{P}_1)}\frac{e(\mathbf{P}_1)}{e(\mathbf{P}_2)}\cdots\frac{e(\mathbf{P}_{m-1})}{e(\mathbf{P}_m)} = \prod_{k=0}^{m-1}\frac{1}{\Pr_k(x_k \prec y_k)},$$

where $\Pr_k(x \prec y)$ is the proportion of linear extensions of $\mathbf{P}_k$ with x_k below y_k. To estimate $e(\mathbf{P})$ within a factor $(1+\varepsilon)$, with probability at least $1-\delta$, it is thus sufficient—for some sequence of incomparable pairs (x_k, y_k)—to estimate each $\Pr_k(x_k \prec y_k)$ to within a factor $(1+\varepsilon n^{-2})$, with probability at least $1-\delta n^{-2}$.

Given $\mathbf{P}$, and any pair (x, y) of incomparable elements, we estimate $\Pr(x \prec y)$ by repeatedly generating nearly uniformly independent linear extensions, and simply counting how many of them have x below y. If we take $\varepsilon^{-2}n^5$ nearly independent, nearly uniformly random, samples, then with very high probability the proportion $Q(x \prec y)$ of samples with x below y will be within an additive error of $\frac{1}{3}\varepsilon n^{-2}$ of the true probability $\Pr(x \prec y)$. If $Q(x \prec y) \ge 1/2$, then with very high probability it is within a multiplicative constant $(1+\varepsilon n^{-2})$ of the true probability, as we need: if not, then we exchange the roles of x and y.

The best known result along these lines is the following, due to Bubley and Dyer [99].

Theorem 8.13.5 *There is a randomized algorithm, with running time $O(n^5\varepsilon^{-2}\log^2 n \log(n/\varepsilon))$, that takes as input an n-element poset $\mathbf{P}$ and a positive constant ε, and outputs a number Q such that*

$$\Pr\left((1-\varepsilon)e(\mathbf{P}) \le Q \le (1+\varepsilon e(\mathbf{P}))\right) > \frac{3}{4}.$$

An algorithm with these properties is called an *fpras* for $e(\mathbf{P})$. The constant $3/4$ is somewhat arbitrary: a better constant can be obtained by running the algorithm several times and taking the median of the outputs.

Exercises

8.1 What is the maximum of $e(\mathbf{P})$ over all posets $\mathbf{P}$ with n elements and width at most w? Compare this to the bound in Theorem 8.1.1.

8.2 What is the minimum of $e(\mathbf{P})$ over all posets $\mathbf{P}$ with n elements and height at most h?

8.3 Prove the first part of Proposition 8.3.1, describing the facets of the order polytope.

Fig. 8.6. The element x is in no cutset.

8.4 Show that the vertices of the order polytope are exactly the indicator vectors of up-sets.

8.5 Describe all the facets of the chain polytope.

8.6 Let $\mathbf{P}$ be the three-element poset with ground-set $\{x_1, x_2, x_3\}$, with the only comparability being $x_1 < x_2$. Describe the order polytope and the chain polytope of $\mathbf{P}$, and demonstrate directly that each has volume $1/2$.

8.7 Let $\mathbf{P}$ be the three-element poset with ground-set $\{x_1, x_2, x_3\}$, and relations $x_1 < x_2$, $x_1 < x_3$. Describe the order polytope and the chain polytope of $\mathbf{P}$, and demonstrate directly that each has volume $1/3$.

8.8 A *cutset* in a poset $\mathbf{P}$ is an antichain meeting every chain. Show that the vector $\mathbf{b}$ defined by $b_x = e(\mathbf{P} - x)/e(\mathbf{P})$ is maximal in $\mathcal{A}(\mathbf{P})$ if and only if every element of $\mathbf{P}$ is in a cutset.

Suppose there is an induced subposet of $\mathbf{P}$ of any of the forms in Figure 8.6. Show that x is in no cutset of $\mathbf{P}$.

8.9 Recall that a series-parallel poset is one that can be obtained from single-element posets by repeated disjoint unions and linear sums. Show that all series-parallel posets have dimension 2.

Show that, if $\mathbf{P}$ is a series-parallel poset and $\overline{\mathbf{P}}$ any complement, then $e(\mathbf{P})e(\overline{\mathbf{P}}) = n!$.

8.10 Let $\mathbf{P}$ be a poset, let $\mathbf{u}$ be the optimal point in $\mathcal{C}(\mathbf{P})$, and let $\sum_{j=1}^{m} \lambda_j \mathbf{e}^{A_j}$ be the laminar decomposition of $\mathbf{u}$. Show that all the A_j are maximal antichains and that $\sum_{j=1}^{m} \lambda_j = 1$.

Now suppose $\mathbf{P}$ is a ranked poset. Show that $\mathbf{P}$ is LYM if and only if the A_j in the laminar decomposition of $\mathbf{u}$ are exactly the ranks.

8.11 Show that, in every poset $\mathbf{P} = (X, P)$, there is a chain C such that

$$\sum_{x \in C} \frac{1}{e(\mathbf{P} - x)} \geq \frac{|X|}{e(\mathbf{P})}.$$

8.12 Prove Theorem 8.11.1 for a partition $(\lambda_1, 1, \ldots, 1)$.

8.13 Let $\mathcal{W}_k$ be the class of posets of width at most k. Show that there is an algorithm using dynamic programming that calculates $e(\mathbf{P})$

exactly, and runs in time $O(n^{k+1})$ (or thereabouts), for all posets in $\mathcal{W}_k$.

8.14 For p an odd prime, let $\mathbf{P} = (X, P)$ be a partial order consisting of: r components $K_{1,p-1}$ with one element above p–1 others, s components $K_{2,p-1}$ with two elements above $p-1$ others, and t isolated elements. What is the probability that a uniformly randomly chosen linear order on X is a linear extension of $\mathbf{P}$? For which values of r, s and t does p divide $e(\mathbf{P})$?

8.15 Give a complete proof of Theorem 8.12.1.

8.16 For any poset $\mathbf{P}$, show that the graph $GE(\mathbf{P})$ is connected. (This means that the Karzanov-Khachiyan Markov chain is always connected.) Show also that the uniform distribution on $\mathrm{E}(P)$ is the stationary distribution for the Karzanov-Khachiyan Markov chain.

8.17 In the description of Karzanov-Khachiyan Markov chain on the set of linear extensions, why did we flip a fair coin in the case where x_i and x_{i+1} are incomparable? What would go wrong if we *always* swapped x_i and x_{i+1} in this case?

8.18 Suppose $\mathbf{P}$ is the poset $\mathbf{k} + \mathbf{1}$, consisting of a single element x and an incomparable chain $y_0 < y_1 < \cdots < y_{k-1}$ of k elements. Describe the graph defined in Exercise 16 explicitly in this case. Now consider running the Karzanov-Khachiyan Markov chain in this case, starting from any fixed linear extension $\prec_0$. Explain why the mixing time is $\Omega(n^3)$.

8.14 Notes and References

The theory of stable set and fractional stable set polytopes of graphs was developed by ... . The key breakthrough, showing that the two polytopes are equal if and only if the graph is perfect, is due to Lovász [99]. It is easy to see that this result implies Theorem 9.99, the Weak Perfect Graph Theorem, and indeed Lovász's proof establishes both results together. For much more on this topic, see Grötschel, Lovász and Schrijver [99].

Stanley's original proof of Theorem 8.6.1 in [99] uses the Ehrhart polynomial of a polytope, a polynomial whose leading coefficient is equal to the volume.

The first paper to exploit the potential of Stanley's Theorem was that of Kahn and Kim [99]. Theorem 8.8.1 is just one minor aspect of the work reported in that paper: their primary purpose was to give a deterministic algorithm for sorting, starting from P, using $O(\log(e(\mathbf{P}))$ comparisons. The existence and uniqueness of the laminar decomposition is also proved in that paper. The derivation we give here of Stanley's Theorem from that result does not seem to have appeared in print before.

The alternative, probabilistic, proof of Theorem 8.8.1 first appears in Brightwell and Tetali [99].

Theorem 8.9.2 was first proved by Saint-Raymond [99]; there are now several simpler proofs, including one by Meyer [**?**] that also gives the cases of equality. The lower bound on $e(\mathbf{P})e(\overline{\mathbf{P}})$ in Theorem 8.9.1 was first proved by Sidorenko [99]; the derivation given here, and the upper bound, are due to Bollobás, Brightwell and Sidorenko [99].

Bollobás and Brightwell [99] studied the *content* $\mu(\mathcal{K})$ of a convex corner $\mathcal{K}$, defined recursively by

$$\mu(\mathcal{K}) = \max_{\mathbf{a}\in\mathcal{K}} \sum_{i=1}^{n} a_i \mu(\mathcal{K}_i),$$

where $\mathcal{K}_i = \{\mathbf{b} \in \mathcal{K} : b_i = 0\}$. The content can be found in various guises elsewhere: in the paper of Sidorenko [99], and in Meyer's proof [99] of Saint-Raymond's Inequality.

If $\mathcal{K}$ is the chain polytope of a poset $\mathbf{P}$, then it follows from Sidorenko's Inequality that $\mu(\mathcal{K}) = e(\mathbf{P})$. Bollobás and Brightwell prove, among other things, that $\mu(\mathcal{K}) \leq n!\,\mathrm{Vol}(\mathcal{K})$, with equality if (and, for $\mathcal{K}$ a polytope all of whose vertices are 0-1 vectors, only if) $\mathcal{K}$ is the stable set polytope of a strongly perfect graph: this extends Stanley's Theorem. Exercise 11 is taken from [99].

The paper of Bollobás and Brightwell also contains various generalizations of Sidorenko's results, in particular of Theorem 8.7.1, as well as the inequalities of Santaló and Saint-Raymond.

Corollary 8.7.2 was conjectured to us by Tuuka Ilomäki in September 2005; he was completing a Doctorate in Music at the Sibelius Academy in Finland, and wished to compare similarities of musical objects. We haven't come across the inequality in the literature, and it does not seem to be at all easy to prove directly. The properties of the Ilomäki Metric are unexplored.

The problem of estimating the number of linear extensions of the subset lattice $\mathbf{2^t}$ was first raised by Richard Stanley. In 1987, Sha and Kleitman proved the result in Theorem 8.10.1 for the special case of $\mathbf{2^t}$, and this was extended by Shastri in 1998 to LYM posets satisfying certain conditions on the level sizes. The observation that Theorem 8.10.1 is a consequence of Theorem 8.8.1 is made by Brightwell and Tetali [99].

The original proof by Frame, Robinson and Thrall[99] of Theorem 8.11.1 in 1954 used the representation theory of the symmetric group. There are now some short elementary proofs, especially that of Greene, Nijenhuis and Wilf [99] in 1979, which also provides a straightforward mechanism for generating a random linear extension of $\mathbf{Y}(\lambda)$. There has also been much interest in the rough structure of a typical linear extension of the grid; see Pittel and Romik [99]b:Pittel-Romik) for some recent progress.

The first proof that there is an fpras, based on a rapidly mixing Markov chain, for estimating the volume of a suitably presented convex body in $\mathbb{R}^n$ is due to Dyer, Frieze and Kannan [99]: the authors mention in that paper the application to counting linear extensions of a poset. There have been several subsequent improvements to the methods, and in particular to the bounds on the running time.

Conductance was introduced and first used to prove rapid mixing by Jerrum and Sinclair [99]; the basic technique has been extended since, most recently by Kannan, Lovász and Montenegro [99].

Karzanov and Khachiyan [99] introduced their Markov chain, and gave a proof very similar to our proof of Theorem 8.13.1, as well as the application to counting $e(\mathbf{P})$. Dyer and Frieze [99] improved their analysis, and also gave Lemma 8.13.3: earlier work had used a weaker form. Bubley and Dyer [99] used path coupling to show that the Karzanov-Khachiyan chain has mixing time at most around n^4, and that a slight variant of the chain has mixing time $O(n^3 \log n)$. Wilson [99] extended their analysis to show that the Karzanov-Khachiyan chain has mixing time $\Theta(n^3 \log n)$.

Excellent recent accounts of the theory of rapidly mixing Markov chains are the book of Jerrum [99], and the survey article of Dyer and Greenhill [99].