Entropy and Sorting

Jeff Kahn*† Jeong Han Kim*‡

Abstract

We reconsider the old problem of sorting under partial information, and give polynomial time algorithms for the following tasks.

(1) Given a partial order P, find (adaptively) a sequence of comparisons (questions of the form, "is $x < y$?") which *sorts* (i.e. finds an unknown linear extension of) P using $O(\log(e(P)))$ comparisons in worst case (where $e(P)$ is the number of linear extensions of P).

(2) Compute (on line) answers to any comparison algorithm for sorting a partial order P which force the algorithm to use $\Omega(\log(e(P)))$ comparisons.

(3) Given a partial order P of size n, estimate $e(P)$ to within a factor exponential in n. (We give upper and lower bounds which differ by the factor $n^n/n!$.)

Our approach, based on entropy of the comparability graph of P and convex minimization via the ellipsoid method, is completely different from earlier attempts to deal with these questions.

1 Background and results

The problem of *sorting under partial information* is:

*Department of Mathematics, Rutgers University, New Brunswick NJ 08903.

†Also, in Center for OR, Rutgers. Partially supported by grants from NSF(DMS-9003376) and AFOSR(89-0512 and 90-0008).

‡Partially supported by a DIMACS Graduate Assistantship.

given an unknown total order on a set $X = \{x_1, \ldots, x_n\}$, together with some of the relations $x_i < x_j$, determine the full order via questions of the form

$$\text{“is } x_i < x_j \text{ ?”} \tag{1}$$

In other words, we are given a partial order, say P, on X and want to determine some unknown *linear extension* of P (i.e. total ordering of X compatible with P) by means of questions as in (1). We will also call such a procedure *sorting P by comparisons*. (For more detailed discussions see e.g. [7, 20, 12]).

Our new results are summarized in the abstract, and will be elaborated upon below. We begin here with some

Background.

It is obvious that any comparison algorithm requires $\log(e(P))$ comparisons in worst case, where $e(P)$ is the number of linear extensions of P. This is the *“information theory lower bound”* (ITLB). (Our logarithms are base 2.)

How close the ITLB is to the truth was first considered by S. S. Kislitsyn [15] and then M. Fredman [7], who showed that sorting can be achieved with $\log(e(P)) + 2n$ comparisons, where $n = |P|$. (So in particular, the ITLB is nearly sharp unless $e(P)$ is quite small.)

In fact, Fredman showed much more generally that one can choose from among *any* collection Γ of permutations of X using no more than $\log|\Gamma| + 2n$ comparisons. As he remarks, construction of his algorithm “requires considerable enumerative information about the set Γ”: it is “practical” only if we ignore all costs other than comparisons.

At about the same time, Fredman raised the by now well-known conjecture [1] that if P is not a linear order, then it contains elements x, y with

$$1/3 \leq p(x < y) \leq 2/3, \tag{2}$$

where $p(x < y)$ denotes the fraction of extensions in which x precedes y. (This conjecture was later independently proposed by Linial [20]. That the value $1/3$ is best possible is shown by the poset with three elements and one relation.)

[1] This conjecture had actually been considered already by Kislitsyn in 1968, but people in the west seem to have been mostly unaware of his work until recently.

The point of this, of course, is that it shows the ITLB is sharp up to a constant factor, since an algorithm which always compares x, y satisfying (2) sorts with no more than $\log_{3/2} e(P) = (\log(3/2))^{-1} \log(e(P))$ comparisons. Furthermore (also of course), the exact bounds of (2) are not needed for this: any result which replaces (2) by

$$\delta < p(x < y) < 1 - \delta \tag{3}$$

with $\delta > 0$ constant gives the same application (with $(\log(3/2))^{-1}$ replaced by $(\log(1/(1-\delta)))^{-1}$).

Such a result, with $\delta = 3/11$, was proved in [12], and simpler proofs, with somewhat smaller δ's, were given in [14], [11]. (See also [16, 8] for more on this problem.)

All of the arguments [12, 14, 11] are geometric: that of [12] depends on the Aleksandrov-Fenchel inequalities, while the later ones use the simpler (but less precise) Brunn-Minkowski Theorem (see e.g. [2]). The opening to geometry is given by the observation (seemingly due to Linial [20]) that the volume of the *order polytope*

$$\mathcal{O}(P) := \{(v_1, \ldots, v_n) \in [0,1]^n : x_i <_P x_j \Rightarrow v_i \leq v_j\}$$

is just $e(P)/n!$, and that a comparison which splits extensions (in the sense of (3)) is the same thing as a hyperplane $v_i = v_j$ with at least δ of the volume of $\mathcal{O}(\mathcal{P})$ on each side.

Though it does resolve the question of the ITLB's accuracy, the $O(\log(e(P)))$ comparison algorithm whose existence is established by [12] suffers from the same drawback as Fredman's original approach: *finding* the comparisons has to date remained computationally infeasible.

On the other hand, the recent probabilistic algorithms for volume computation beginning with [6] do provide a means of finding the comparisons if one allows randomization. (For example, one can identify a good comparison by simply estimating the volumes of $\mathcal{O}(\mathcal{P})$ and its intersections with the "positive" sides of the various hyperplanes $v_i = v_j$; but see [13] for a more efficient procedure. Note that the validity of these algorithms still depends on the theorems mentioned above.)

Results

In this paper we show that it is indeed possible to sort with only $O(\log(e(P)))$ comparisons and to find the comparisons in (deterministic) polynomial time.

We also give polynomial algorithms for

(a) computing (on line) answers to any sorting algorithm which force it to use $\Omega(\log(e(P)))$ comparisons, and

(b) estimating the number of extensions of a given P to within an exponential (in $|P|$) factor.

Whether (b) was possible has also been an open question for a number of years. As far as we know, it was first proposed by Gary Miller in the early '80's (see e.g. [23]).

The approach we take, though still geometric, is completely different from those mentioned above. In particular, our sorting algorithm will *not* always be making comparisons which substantially decrease the number of extensions (though clearly it will necessarily be doing this much of the time).

The central notion is that of the entropy, $H(P)$, of the chain polytope of P (definitions in Section 2). The main developments are as follows.

We first establish a fairly close relationship between the ITLB and the quantity $n(\log n - H(P))$ (note $H(P) \leq \log n$, with equality exactly when P is a chain):

Theorem 1.1 *For any P,*

$$\begin{aligned} n(\log n - H(P)) &\geq \log(e(P)) \\ &\geq \max\{\log(n!) - nH(P), Cn(\log n - H(P))\}, \end{aligned} \tag{4}$$

where $C = (1 + 7\log e)^{-1} \approx .09$.

We then show, first, that there is always a comparison which forces an $\Omega(1/n)$ increase in entropy (Theorem 1.2), and second, that any comparison may be answered so as to limit the increase in entropy to $O(1/n)$ (Theorem 1.4).

For precise statements we write $P(x < y)$ for the poset generated by (i.e. the transitive closure of) P and the relation $x < y$.

Theorem 1.2 *In any P not a chain there exist x, y incomparable such that*

$$\min\{H(P(x < y), H(P(y < x)\} \geq H(P) + c/n, \tag{5}$$

where $c = \log(1 + 17/112) \approx 0.2$.

Corollary 1.3 *There is an algorithm which sorts any poset P in $O(\log(e(P)))$ comparisons and finds the comparisons in time polynomial in $|P|$.*

Theorem 1.4 *For any P and any incomparable $x, y \in P$,*

$$\min\{H(P(x < y), H(P(y < x)\} \leq H(P) + 2/n. \tag{6}$$

(The $2/n$ in Theorem 1.4 is sharp; we don't have a guess as to the best c in Theorem 1.2.)

Corollary 1.5 *There is an algorithm which computes answers to queries (1) for any poset P, runs in time polynomial in $|P|$, and forces any comparison algorithm for sorting $|P|$ to use at least $\Omega(\log(e(P)))$ comparisons.*

To derive our algorithmic results from Theorems 1.1-1.4 it is only necessary to point out that $H(P)$ is computationally manageable:

Theorem 1.6 *There is a polynomial time algorithm for approximating $H(P)$ to within (say) n^{-2}.*

This is so because $H(P)$ is the minimum of the convex function

$$f(\underline{a}) = -\frac{1}{n}\sum \log a_i$$

on a polytope (a slight truncation of the chain polytope) for which separations are easy to compute and on which the variation of f is not excessive. Estimating $H(P)$ is thus a problem which can be handled by the methods of [10].

Theorem 1.6 allows us, for example, to simply estimate $H(P(x < y))$, $H(P(y < x))$ for all incomparable x, y. By Theorem 1.2, this will produce at least one pair x, y, for which (5) holds (with c slightly decreased, say). On the other hand, the lower bound in Theorem 1.1 guarantees that choosing such a pair at each stage sorts in $O(\log(e(P)))$ steps, so we have Corollary 1.3.

Corollary 1.5 follows in similar fashion from Theorems 1.6 and 1.4, and the upper bound in Theorem 1.1.

In section 2 we review what we need in the way of graph entropy, and sections 3-5 are devoted to the proofs of Theorems 1.1-1.3.

2 Entropy

From now on, P is a partial order on the set $V = \{x_1, \ldots, x_n\}$. (We replace X by V because it will often be the vertex set of a graph.)

Graph entropy was introduced by Körner [17]. There are at least three (equivalent) definitions – two from [17], and the third due to Csiszár et. al. [4] – of which the last is most convenient for our purposes.

Recall that the *vertex packing polytope,* $VP(G)$, of a graph G on V is

$$\mathrm{conv}\{1_A : A \subseteq V \quad \text{independent}\},$$

where

$$(1_A)_i = \begin{cases} 1 & \text{if } x_i \in A \\ 0 & \text{otherwise.} \end{cases}$$

The *chain polytope* (the name is from [22]) of P is just $VP(P) := VP(G(P))$, where $G(P)$ is the comparability graph of P (i.e. the graph on V with $x_i \sim x_j$ iff x_i and x_j are comparable in P). (Equivalently (because $G(P)$ is perfect; see e.g. [9, 3]) $VP(P)$ consists of those $\underline{v} \in [0,1]^n$ for which $\sum_{i \in I} v_i \leq 1$ whenever $\{x_i : i \in I\}$ is a chain in P. This is the definition of chain polytope given in [22].) We also write $VP(\overline{P})$ for $VP(\overline{G(P)})$.

For probability distribution p on V, the *graph entropy of G with respect to p* is

$$H(G,p) = \min_{a \in VP(G)} -\sum_{i=1}^{n} p_i \log a_i. \tag{7}$$

When p is the constant vector $1/n$ we abbreviate this to $H(G)$.

As noted in [4], it is not hard to see that the minimum in (7) is always achieved and finite; the minimizing vector a is a convex combination of indicator vectors of maximal independent sets; and the ith coordinate of a is uniquely determined provided $p_i > 0$.

Actually the discussion of [4] is more general, concerning the entropy of a *convex corner*, i.e. a convex $K \subseteq (\Re^+)^n$ satisfying

$$\underline{0} \leq \underline{v}' \leq \underline{v} \in K \Rightarrow \underline{v}' \in K.$$

For such K the definition (7) extends to

$$H(K,p) = \min_{a \in K} -\sum_{i=1}^{n} p_i \log a_i,$$

and again we write $H(K)$ if p is uniform distribution.

We write $G(P)$ (as above) for the comparability graph of P, and set $H(G(P)) = H(P)$, $H(\overline{G(P)}) = H(\overline{P})$. We also write $a(P)$ and $b(P) = a(\overline{P})$ for the vectors achieving $H(P)$ and $H(\overline{P})$.

It is obvious that entropy is monotone, that is, if $F \subseteq G$ are graphs on V (containment is as sets of edges), then

$$H(F) \leq H(G). \tag{8}$$

That it is also subadditive was shown by Körner [18]:

Theorem 2.1 *For any two graphs F, G and probability distribution p on V,*

$$H(F \cup G, p) \leq H(F, p) + H(G, p). \tag{9}$$

In particular, since $H(G \cup \overline{G}, p) = H(p) := -\sum_i p_i \log p_i$, we have

$$H(G, p) + H(\overline{G}, p) \geq H(p). \tag{10}$$

On the other hand, it was conjectured in [19] and proved in [4] that

Theorem 2.2 *A graph G is perfect if and only if*

$$H(G, p) + H(\overline{G}, p) = H(p)$$

for every probability distribution p.

In particular, since comparability graphs are perfect, we have

$$H(P) + H(\overline{P}) = \log n \tag{11}$$

for every P. From this, using the fact that $\sum a_i b_i \leq 1$ whenever $a \in VP(G)$, $b \in VP(\overline{G})$, one easily deduces (see[4])

$$a_i(P) b_i(P) = 1/n \quad \forall i. \tag{12}$$

There is a useful representation of $a(P)$ based on Dilworth's ordering (from [5]) on the set $\mathcal{A}$ of maximal antichains of P. Recall that for X, Y maximal antichains of P, this ordering sets $X \prec Y$ if

$$\forall x \in X \quad \exists y \in Y, \quad y \geq_P x.$$

(Or equivalently: every $y \in Y$ is $\geq$ some $x \in X$.) It's easy to see that this is a partial order.

Lemma 2.3 *There is a unique representation*

$$a(P) = \sum_{j=1}^{r} w_j 1_{A_j} \tag{13}$$

with $w_j > 0$, $\sum w_j = 1$ *and* $A_1 \prec \cdots \prec A_r$ *distinct maximal antichains of* P.

Proof. We first prove existence. Note that in *any* representation

$$a = \sum w_A 1_A \tag{14}$$

of $a := a(P)$ as a convex combination of (indicator vectors of) antichains A, all A's in the support of w must be maximal, since expanding any of these antichains gives a strictly better a.

Given a representation as in (14), choose, if possible, A, A' incomparable under $\prec$ with $0 < w_A \leq w_{A'}$. (If no such choice exists, then (14) is the desired representation.) Let

$$\begin{aligned} B &= \min(A \cup A'), \\ B' &= \max(A \cup A') \end{aligned}$$

(where $\min X$ and $\max X$ are the sets of minimal and maximal elements of $X \subseteq P$). Then B, B' are antichains, and each element of P appears the same number of times in B, B' as in A, A'. Thus with w' given by

$$w'_C = \begin{cases} w_C - w_A & \text{if } C = A \text{ or } A' \\ w_C + w_A & \text{if } C = B \text{ or } B' \\ w_C & \text{otherwise,} \end{cases}$$

$\sum w'_C 1_C$ again represents a as a convex combination of (necessarily) maximal antichains. This will complete the proof of existence provided we can show this procedure doesn't cycle. One way to see this is to fix a linear extension $\propto$ of $\prec$, and order functions $u : \mathcal{A} \to \Re^+$ "lexicographically": $u' < u$ if $u'_C < u_C$ with C the least element (under $\propto$) of $\mathcal{A}$ for which $u'_C \neq u_C$. It's then clear that with w, w' as above, we have $w < w'$.

We now turn to the proof of uniqueness. This does not even require the assumption $\sum w_i = 1$. We show more generally that

Proposition 2.4 *Any $a : P \to \Re^+$ has a unique representation of the form*

$$a = \sum_{i=1}^{t} w_i 1_{A_i}$$

with $w_i > 0$ and $A_1 \prec \cdots \prec A_t$ distinct maximal antichains.

We call this representation the *laminar decomposition* of a.

Proof. Suppose we have a representation as in the Proposition. Let $P^+ = \{x \in P : a(x) > 0\}$, $A = \min(P^+) \supseteq A_1$ and $\alpha = \min\{a(x) : x \in A\}$. Then

$$A_1 = A, \quad w_1 = \alpha. \tag{15}$$

For suppose first that $A_1 \neq A$, and let $x \in A \backslash A_1$. Then $x \in A_i$ for some $i > 1$, contradicting the assumption $A_1 \prec A_i$. If instead, $A_1 = A$ but $w_1 = \beta < \alpha$, then $\sum_{i=2}^{t} w_i 1_{A_i}$ is a laminar decomposition of the function $a' : P \to \Re^+$ given by

$$a'(x) = \begin{cases} a(x) - \beta & \text{if } x \in A \\ a(x) & \text{otherwise,} \end{cases}$$

and the same argument again gives a contradiction.

The proof by induction (on $|P^+|$, say) is now completed by observing that for A_1, w_1 as in (15), $\sum_{i=1}^{t} w_i 1_{A_i}$ is a laminar decomposition of a if and only if $\sum_{i=2}^{t} w_i 1_{A_i}$ is a laminar decomposition of the function $a' : P \to \Re^+$ given by

$$a'(x) = \begin{cases} a(x) - \alpha & \text{if } x \in A \\ a(x) & \text{otherwise.} \end{cases}$$

□

Remark Notice that the proof of Proposition 2.4 gives an easy algorithm for finding the laminar decomposition of a.

Notice that, with the notation of the lemma, if we set

$$\alpha_i = \min\{j : x_i \in A_j\}, \quad \beta_i = \max\{j : x_i \in A_j\},$$

then

$$x_i \in A_j \Leftrightarrow \alpha_i \leq j \leq \beta_i. \tag{16}$$

For if not, then there is j strictly between α_i and β_i such that $x_i \notin A_j$. Since $A_{\alpha_i} \prec A_j \prec A_{\beta_i}$ there exist $x \in A_j$ and $y \in A_{\beta_i}$ so that $x_i < x \leq y$. Thus $\{x_i < y\}$ is a subset of the antichain A_{β_i}, a contradiction.

In the sequel A_j, w_j, α_i and β_i are as given above.

3 Bounds

Here we prove Theorem 1.1. We make use of the fact, proved by Stanley [22], that $VP(P)$ and $\mathcal{O}(P)$ have the same volume, whence

$$vol(VP(P)) = e(P)/n!. \tag{17}$$

The upper bound and first half of the lower bound in (4) are thus equivalent to

$$\frac{n^n}{n!}2^{-nH(P)} \geq vol(VP(P)) \geq 2^{-nH(P)}. \tag{18}$$

Proof. Let $a = a(P)$. Then since $VP(P)$ is a convex corner,

$$vol(VP(P)) \geq \prod_i a_i = 2^{-nH(P)}.$$

On the other hand, according to (12), $b = b(P)$ is given by $b_i = 1/(na_i)$, so that

$$VP(P) \subset L := \{(s_1, ..., s_n) : s_i \geq 0, \sum_{i=1}^{n} s_i b_i \leq 1\}$$

implies

$$vol(VP(P)) \leq vol(L) = \frac{n^n}{n!}\prod_i a_i = \frac{n^n}{n!}2^{-nH(P)}.$$

□

Notice that these quite easy inequalities are already enough to settle the problem of estimating $e(P)$ mentioned in the introduction.

Notice also that the proof of (18) is valid for convex corners K, that is,

$$\frac{n^n}{n!}2^{-nH(K)} \geq vol(K) \geq 2^{-nH(K)}.$$

It would be interesting to see whether this simple observation could be extended to give either better volume estimates for convex corners, or estimates for more general bodies.

To derive an $O(\log(e(P)))$ sorting algorithm from Theorem 1.2, we must show that $nH(\overline{P}) = O(\log(e(P)))$ (note we replace the left hand side of (4) by $nH(\overline{P})$ using (11)). The lower bound in (18) gives such a bound unless $e(P)$ is quite small; namely, if $\log e(P) \geq cn$, with $c > 0$ constant, then

$$nH(\overline{P}) \leq ((c + \log e)/c) \log(e(P)). \tag{19}$$

Strangely (or perhaps not, since this is where Fredman's algorithm fails to be $O(\log(e(P)))$), instances with $e(P)$ small require more work. This is the other half of the lower bound in (4):

$$nH(\overline{P}) \leq (1 + 7 \log e) \log(e(P)). \tag{20}$$

For the proof of this we require a few preliminaries. We write $x \sim y$ if x, y are comparable in P, and denote $\{1, ..., q\}$ by $[q]$. Fix a maximum length chain C of P, say $C = \{x_1 < \cdots < x_l\}$, and let $T = P \setminus C = \{y_1, ..., y_t\}$.

We prove this by induction on t, roughly as follows. It's easy to see that t may be assumed fairly small relative to n. To induct we try to find i, j such that inserting y_j between x_i and x_{i+1} multiplies $e(P)$ by some factor $1/k$, while decreasing $H(\overline{P})$ by at most about $\log k$.

To help in keeping track of the effects of such insertions, we define

$$\begin{aligned}
&K(j) = \{i \in [l] : x_i \not\sim y_j\}, \quad |K(j)| = k_j \\
&f(j) = \min\{i \in [l] : x_i > y_j\}, \quad \min \emptyset = l + 1 \\
&g(j) = \max\{i \in [l] : x_i < y_j\}, \quad \max \emptyset = 0 \\
&U(i) = \{j \in [t] : f(j) = i\}, \quad |U(i)| = u_i \\
&V(i) = \{j \in [t] : g(j) = i\}, \quad |V(i)| = v_i
\end{aligned}$$

Note $k_j > 0$ for all $j = 1, ..., t$ since C is a maximal chain.

Lemma 3.1 *If $t < n/7$ and P has no cut point (element comparable to all others), then there is $j \in [t]$ such that $k_j \geq 3$ and*

$$\sum_{i \in K(j)} (u_i + v_i) \leq k_j \ . \tag{21}$$

Proof. Suppose this is false, and consider a minimal $T' \subset T$ for which

$$\bigcup_{j: y_j \in T'} K_j = [l] \ .$$

We may assume $T' = \{y_1, ..., y_r\}$. By our assumption we have

$$\sum_{i \in K(j)} (u_i + v_i) \geq k_j - 2 \quad for \ j = 1, ..., r \ ,$$

so

$$\sum_{j \in [r]} \sum_{i \in K(j)} (u_i + v_i) \geq \sum_{j \in [r]} k_j - 2r \ .$$

But the right hand side here is at least $l - 2t$ (since $\bigcup_{j \in [r]} K_j = [l]$), while the left hand side is at most $\sum_{i \in [l]} 2(u_i + v_i) \leq 4t$ (since the minimality of T' implies that no i is in more than two of $K(1), ..., K(r)$). This gives $6t \geq l$, contradicting the assumption $t < n/7$.

□

Lemma 3.2 *Suppose P has no cutpoint and (as a set of relations) is maximal with given entropy. Then if $t < n/7$ there exist $j \in [t]$ and $i \in [l]$ such that $P' := P(x_i < y_j < x_{i+1})$ satisfies*

$$e(P') \leq (k_j - 1)^{-1} e(P)$$

and

$$nH(\bar{P}) \leq nH(\bar{P}') + 2 \log(2k_j + 1) \ .$$

Proof. Let y_j be as in Lemma 3.1 and $K_j = \{x_h < ... < x_m\}$. Choose $i \in \{h, ..., m-1\}$ with

$$Pr(x_i < y_j < x_{i+1}) := \frac{e(P(x_i < y_j < x_{i+1}))}{e(P)}$$

minimum, and set $P' = P(x_i < y_j < x_{i+1})$. Then clearly

$$e(P') \leq (k_j - 1)^{-1} e(P).$$

On the other hand, the maximality of P implies that P' differs from P only in the at most $2k_j$ new relations involving y_j (c.f. (21)), that is, $z < y_j \Rightarrow z < x_{i+1}$ and $w > y_j \Rightarrow w > x_i$.

(For suppose $z < y_j$. Note that by maximality there is an antichain A in the laminar decomposition of $a(P)$ with $x_i, y_j \in A$. Since $x_{i+1} > x_i$, all antichains in the laminar decomposition which contain x_{i+1} follow A, and similarly all thos containing z precede A. But then, again using maximality, we have $z < x_{i+1}$.)

Thus, according to Theorem 2.1,

$$\begin{aligned} nH(\bar{P}) &\leq nH(\bar{P'}) + (2k_j + 1)H(\frac{1}{2k_j + 1}) \\ &\leq nH(\bar{P'}) + 2\log(2k_j + 1) \end{aligned}$$

(where $H(z) := -z \log z - (1 - z) \log(1 - z)$ is the entropy function).

□

Proof of (20). We may, of course, assume that P is maximal with given entropy. We retain the notation introduced above and induct on n and t, the result being obvious if either $n = 1$ or $t = 0$. We assume therefore that $n > 1$ and $t > 0$. If P has a cutpoint x, then we finish by induction since $nH(\bar{P}) = (n - 1)H(\overline{P \setminus \{x\}})$ and $e(P) = e(P \setminus \{x\})$; so we assume this is not the case. We next observe that the easy inequality $e(P) \geq 2^t$ allows us to assume $t < n/7$, since otherwise (20) follows from (19).

We now have the hypotheses of Lemma 3.2, so also its conclusion. Since (inducting on t), (20) is true for P', we have

$$\begin{aligned} nH(\bar{P}) &\leq nH(\overline{P'}) + 2\log(2k_j + 1) \\ &\leq (1 + 7\log e)\log e(P') + 4\log(k_j + 1) \\ &\leq (1 + 7\log e)\log e(P) + (8 - (1 + 7\log e))\log(k_j - 1) \\ &\leq (1 + 7\log e)\log e(P)\ . \end{aligned}$$

completing the proof.

□

There are various possibilities for extending the lower bounds here, of which we mention just one:

Conjecture 3.3 *If $l = l(P)$ is the length of a longest chain in P, then*

$$vol(VP(P)) \geq (l^l/l!)2^{-nH(P)}.$$

This would improve the constant in (20) to $1 + \log e$. Notice it is tight for any union of a chain and an antichain.

4 Offense

Here we prove Theorem 1.2. To put the task of locating a good comparison in some perspective, let us first mention two curious examples:

Suppose P consists of two disjoint and unrelated chains of size $k = n/2$. Comparison of the minima of the two chains then turns out to be a good comparison in our sense, forcing an entropy increase of about $1/n$. But comparison of the *middle* elements is *not* good – it gives only an $O(n^{-2})$ increase – even though it splits the extensions perfectly.

On the other hand, suppose P is the poset on $\{x_1, \ldots, x_k, y_1, \ldots, y_k\}$ ($n = 2k$) with relations $x_i < y_j$ iff $i = j$ or $i = 1$. Then the comparison $x_1 : x_2$ is good in our sense, but does a poor job of splitting extensions.

For the proof of Theorem 1.2, it's more natural to work in the complement, showing that we can *decrease* $H(\overline{P})$ by some specified amount (say ε), since for this we only need to exhibit *some* $b' \in VP(\overline{P'})$ for which

$$-\frac{1}{n}\sum \log b'_k \leq H(\overline{P}) - \varepsilon$$

(with P' the new poset).

For example, suppose x_i, x_j are minimal in P,

$$b = b(P) = \sum_{m=1}^{s} z_m 1_{B_m}$$

(with each B_k a chain of P), and let $P' = P(x_i < x_j)$,

$$B'_k = \begin{cases} B_k \cup \{x_i\} & \text{if } x_j \in B_k \\ B_k & \text{otherwise.} \end{cases}$$

Then

$$b' = \sum_{m=1}^{s} z_m 1_{B'_m} \in VP(\overline{P'}),$$

$b'_i = b_i + b_j$, and $H(\overline{P'}) \geq H(\overline{P}) + \log(1 + b_j/b_i)$. This already gives Theorem 1.2 if there are minimal x_i, x_j with the ratios b_i/b_j, b_j/b_i bounded.

In general, if the new covering relation is $x_i < x_j$, we may modify the weight function z by transferring some fraction (say μ) of the weight on each chain B (of P) containing x_j to a chain (of P') obtained by replacing the portion of B below x_j by a chain with largest element x_i.

The effect of this procedure is quantified in

Lemma 4.1 *For any incomparable $x_i, x_j \in P$ and $\mu \in [0,1]$, and w_k's as in Proposition 2.4, the entropy of $P' = P(x_i < x_j)$ satisfies*

$$nH(P') \;\geq\; nH(P) + \log(1 + \mu \sum_{k=1}^{\beta_i} w_k/a_j) + \log(1 - \mu \sum_{k=1}^{\alpha_j - 1} w_k/a_j).$$

(assuming the right hand side is defined).

Proof. Let

$$b = b(P) = \sum_{m=1}^{s} z_m 1_{B_m}$$

with $B_1, \ldots, B_s$ chains of P and $x_j \in B_m$ iff $m \in [t]$. Also, denote

$$b_{q,j} = \sum_{m : x_q, x_j \in B_m} z_m$$

and for $m \in [t]$

$$C_m = B_m \setminus \{x \in P : x < x_j\} .$$

Now fix a chain $C = \{x_{i_1} < ... < x_{i_h} = x_i\}$ such that

$$\sum_{p=1}^{h} a_{i_p} = \sum_{k=1}^{\beta_i} w_k \ , \tag{22}$$

and consider the chains B'_m and weights z'_m given by

$$\begin{array}{ll} B'_m = B_m & 1 \le m \le s, \\ B'_{m+s} = C \cup C_m & 1 \le m \le t, \\ z'_{m+s} = \mu z_m, \quad z'_m = (1-\mu) z_m & 1 \le m \le t, \\ z'_m = z_m & t+1 \le m \le s. \end{array}$$

(That is, we transfer a μ-fraction of the z-weight of each B_m containing x_j to the associated chain $C \cup C_m$.) Then

$$b' = \sum_{m=1}^{s+t} z'_m 1_{B'_m} \in VP(\overline{P'})$$

is easily seen to satisfy

$$b'_q = \begin{cases} b_q + \mu(b_j - b_{q,j}) & \text{if } x_q \in C \\ b'_q = b_q - \mu b_{q,j} & \text{if } x_q \notin C, x_q < x_j \\ b'_q = b_q & \text{otherwise.} \end{cases}$$

Thus by the definition of $H(\bar{P})$, we have

$$\begin{aligned} nH(P') - nH(P) &= nH(\bar{P}) - nH(\bar{P}') \\ &\ge -\sum_{q=1}^{n} (\log b_q - \log b'_q) \\ &\ge \sum_{q: x_q \in C} \log(1 + \mu b_j / b_q) + \sum_{q: x_q < x_j} \log(1 - \mu b_{q,j}/b_q) \\ &\ge \log(1 + \mu b_j \sum_{q: x_q \in C} 1/b_q) + \log(1 - \mu \sum_{q: x_q < x_j} b_{q,j}/b_q) \end{aligned}$$

where in the second inequality we use $\log(1+u-v) \ge \log(1+u) + \log(1-v)$ for nonnegative real numbers u, v, and in the third inequality we inductively use $\log(1+u) + \log(1+v) \ge \log(1+u+v)$ for all real numbers u, v with $uv \ge 0$.

On the other hand, using $a_i b_i = 1/n$ and (12),

$$\sum_{q:x_q \in C} 1/b_q = \sum_{q:x_q \in C} n a_q = n \sum_{k=1}^{\beta_i} w_k \ ,$$

and

$$\begin{aligned} \sum_{q:x_q < x_j} b_{q,j}/b_q &= n \sum_{q:x_q < x_j} a_q b_{q,j} \\ &= n \sum_{q:x_q < x_j} \sum_{k:x_q \in A_k} w_k b_{q,j} \\ &= n \sum_{k=1}^{\alpha_j - 1} w_k \sum_{x_q \in A_k} b_{q,j} \\ &\leq n \sum_{k=1}^{\alpha_j - 1} w_k b_j \end{aligned}$$

where the inequality holds since A_k is an antichain. Therefore,

$$\begin{aligned} nH(P') - nH(P) &\geq \log(1 + \mu b_j \sum_{q:x_q \in C} 1/b_q) + \log(1 - \mu \sum_{q:x_q < x_j} b_{q,j}/b_q) \\ &\geq \log(1 + \mu n \sum_{k=1}^{\beta_i} w_k b_j) + \log(1 - \mu n \sum_{k=1}^{\alpha_j - 1} w_k b_j) \\ &= \log(1 + \mu \sum_{k=1}^{\beta_i} w_k / a_j) + \log(1 - \mu \sum_{k=1}^{\alpha_j - 1} w_k / a_j) \ . \end{aligned}$$

□

Also, we need the following easy lemma.

Lemma 4.2 *Given $0 < \varepsilon_1, \varepsilon_2 < 1$, choose i with a_i as large as possible subject to*

$$\sum_{k=1}^{\alpha_i - 1} w_k \leq \varepsilon_1 a_i$$

and let t be the smallest number for which

$$\sum_{k=\alpha_i}^{t} w_k \geq \varepsilon_2 a_i \ .$$

Then for any $x_j \in A_t \setminus \{x_i\}$,

$$a_j < \frac{\varepsilon_1 + \varepsilon_2}{\varepsilon_1} a_i \ .$$

Proof. If $a_j \leq a_i$ then we are done. Suppose $a_j > a_i$. Then by the choices of a_i and $t \geq \alpha_j$

$$\begin{aligned} \varepsilon_1 a_j &< \sum_{k=1}^{\alpha_j - 1} w_k \\ &= \sum_{k=1}^{\alpha_i - 1} w_k + \sum_{k=\alpha_i}^{\alpha_j - 1} w_k \\ &< \varepsilon_1 a_i + \varepsilon_2 a_i \ . \end{aligned}$$

□

Proof of Theorem 1.2 Notice first of all that we may assume P has no cut point, since if it does then the Theorem follows by induction using the fact that (for any cut point x)

$$nH(\bar{P}) = (n-1)H(\overline{P \setminus \{x\}}) \ .$$

For $\varepsilon_1 = 1/4$, $\varepsilon_2 = 1/3$, take x_i, x_i as in Lemma 4.2. Also, let $\delta := \sum_{k=1}^{\alpha_i - 1} w_k / a_i \leq \varepsilon_1$. Then by Lemmas 4.1 and 4.2, for $P' = P(x_j > x_i)$ and

$$\mu := \frac{\varepsilon_1 a_j}{(\varepsilon_1 + \varepsilon_2) a_i} \leq 1 \ ,$$

$$\begin{aligned} nH(P') - nH(P) &\geq \log(1 + \mu \sum_{k=1}^{\beta_i} w_k / a_j) + \log(1 - \mu \sum_{k=1}^{\alpha_j - 1} w_k / a_j) \\ &\geq \log(1 + \mu(\delta + 1)\frac{a_i}{a_j}) + \log(1 - \mu(\delta + \varepsilon_2)\frac{a_i}{a_j}) \\ &\geq \log(1 + \frac{\varepsilon_1 - \varepsilon_1\varepsilon_2 - \varepsilon_1^2 - \varepsilon_1^3}{\varepsilon_1 + \varepsilon_2}) \\ &= \log(1 + \frac{17}{112}) \ . \end{aligned}$$

On the other hand, for $P'' = P(x_j > x_i)$ and $\mu = 1$, Lemma 4.1 and the choice of x_j imply

$$\begin{aligned} nH(P'') - nH(P) &\geq \log(1 + \sum_{k=1}^{\beta_j} w_k/a_i) + \log(1 - \sum_{k=1}^{\alpha_i - 1} w_k/a_i) \\ &\geq \log(1 + (\delta + \varepsilon_2)) + \log(1 - \delta) \\ &\geq \log(1 + \frac{3}{16}) \end{aligned}$$

completing the proof.

□

Remark As shown by the poset with three elements and one relation, the value of c in Theorem 1.2 cannot be increased beyond $3 \log 3 - 4 \approx .755$.

5 Defense

Here we prove Theorem 1.3. The reader may check that the Theorem is sharp whenever x, y are isolated elements of P. The proof of Theorem 1.4 is again based on the laminar decomposition of $a(P)$. The effect on this decomposition of adding a relation $x < y$ is that some of the A_l's may no longer be antichains (in the new partial order). However when this happens, because of the nature of the decomposition, at least one of $A_l \setminus \{x\}$, $A_l \setminus \{y\}$ *will* be an antichain. The proof consists of showing that for at least one of the answers to the comparison $x : y$ we may modify the decomposition by such deletions to produce an a' in the chain polytope of the new poset with $-(1/n) \sum \log a'_i \leq H(P) + 2/n$. (In most cases, the correct procedure is to replace A_l by the two antichains $A_l \setminus \{x\}$, $A_l \setminus \{y\}$, dividing the weight w_l between them.)

For the proof we use x_1 and x_2 in place of x and y, and retain the notations A_k, w_k, α_k and β_k used earlier.

Proof of Theorem 1.3. Without loss of generality we may assume $\alpha_1 \leq \alpha_2$. We consider three cases.

<u>**Case 1:**</u> $\alpha_2 > \beta_1$

Set $P' := P(x_2 > x_1)$. Then for all $x_k \leq x_1$ and $x_l \geq x_2$, we have

$$\alpha_l \geq \alpha_2 > \beta_1 \geq \beta_k \ .$$

Thus $A_1, ..., A_r$ are still antichains of P'. This implies

$$H(P') = H(P) \ .$$

Case 2: $\alpha_1 \leq \alpha_2 \leq \beta_1 \leq \beta_2$.
Set $P' := P(x_2 > x_1)$ and consider

$$\begin{array}{cc} A'_m = A_m \setminus \{x_1\}, A''_m = A_m \setminus \{x_2\} & \text{if } \alpha_2 \leq m \leq \beta_1 \\ A'_m = A_m & \text{otherwise.} \end{array}$$

Since

$$\alpha_l > \beta_2 \geq \beta_1, \ \beta_k < \alpha_1 \leq \alpha_2 \quad \text{if } x_k < x_1, \ x_l > x_2$$

the sets defined above are antichains of P'. Now define w' by

$$\begin{array}{ll} w'_m = w''_m = w_m/2 & \text{if } \alpha_2 \leq m \leq \beta_1 \\ w'_m = w_m & \text{otherwise.} \end{array}$$

Then

$$a' := \sum_m w'_m 1_{A'_m} + \sum_{\alpha_2 \leq m \leq \beta_1} w''_m 1_{A''_M}$$

belongs to $VP(P')$ and satisfies $a'_1 \geq a_1/2$, $a'_2 \geq a_2/2$ and $a'_k = a_k$ if $k \neq 1, 2$.
Thus

$$H(P') \leq -\frac{1}{n} \sum_i \log a'_i \leq -\frac{1}{n} \sum_i \log a_i + \frac{2}{n} \ .$$

Case 3: $\alpha_1 \leq \alpha_2 \leq \beta_2 \leq \beta_1$.
Without loss of generality, we may assume

$$\sum_{k=\alpha_1}^{\alpha_2 - 1} w_k \geq \sum_{k=\beta_2+1}^{\beta_1} w_k \ .$$

Again set $P' = P(x_2 > x_1)$ and

$$\begin{array}{cc} A'_m = A_m \setminus \{x_1\}, A''_m = A_m \setminus \{x_2\} & \text{if } \alpha_2 \leq m \leq \beta_2 \\ A'_m = A_m \setminus \{x_1\} & \text{if } \beta_2 < m \leq \beta_1 \\ A'_m = A_m & \text{otherwise.} \end{array}$$

Since for all $x_k < x_1$ and $x_l > x_2$, we have

$$\beta_k < \alpha_1 \le \alpha_2 < \alpha_l \ ,$$

the sets defined above are antichains of P'. Now define w' by

$$\begin{array}{ll} w'_m = w''_m = w_m/2 & \text{if } \alpha_2 \le m \le \beta_2 \\ w'_m = w_m & \text{otherwise.} \end{array}$$

Then the vector

$$a' := \sum_m w'_m 1_{A'_m} + \sum_{\alpha_2 \le m \le \beta_2} w''_m 1_{A''_M}$$

belongs to $VP(P')$ and satisfies $a'_1 \ge a_1/2$, $a'_2 = a_2/2$ and $a'_k = a_k$ if $k \ne 1, 2$. Thus

$$H(P') \le -\frac{1}{n}\sum_i \log a'_i \le -\frac{1}{n}\sum_i \log a_i + \frac{2}{n} \ .$$

□

Another Proof of Theorem 1.3. Set

$$\begin{array}{ll} U = \{x \in P : x < x_1\}, & V = \{x \in P : x > x_1\} \\ W = \{x \in P : x < x_2\}, & Z = \{x \in P : x > x_2\} \end{array}$$

and choose a chain $K \subseteq U$ of P with the weight

$$w(K) := \sum_{x_i \in K} a_i$$

as large as possible. Similarly, choose chains $L \subseteq V$, $M \subseteq W$ and $N \subseteq Z$ with maximum weights. Then since the chain polytope of P is $VP(P)$,

$$\begin{array}{l} w(K) + w(L) + a_1 \le 1 \\ w(M) + w(N) + a_2 \le 1 \ . \end{array}$$

Therefore,

$$w(K) + w(N) + (a_1 + a_2)/2 \le 1 \tag{23}$$

or

$$w(L) + w(M) + (a_1 + a_2)/2 \le 1 \ . \tag{24}$$

Without loss of generality we may assume that (23) is true. It is enough to show that the vector a' with

$$a'_i = \begin{cases} a_i/2 & \text{if } i = 1, 2 \\ a_i & \text{otherwise} \end{cases}$$

is in the chain polytope of $P' := P(x_1 < x_2)$. Suppose Q is a maximal chain of P'. If $\{x_1, x_2\} \not\subseteq Q$ then it is easy, by maximality of Q, to see that Q is a chain of P. Thus $a'_i \leq a_i$ for all i implies

$$w'(Q) := \sum_{i: x_i \in Q} a'_i \leq w(Q) \leq 1 \; .$$

If $\{x_1, x_2\} \subseteq Q$ then set

$$K' = \{x \in Q : x <_{P'} x_1\}, \quad N' = \{x \in Q : x >_{P'} x_2\} \; .$$

Note that $K' \subset U$, $N' \subset Z$ are chains of P and $Q = K' \cup N' \cup \{x_1, x_2\}$ since there is no element x such that $x_1 <_{P'} x <_{P'} x_2$. Therefore by our choices of K and N, we have

$$\begin{aligned} w'(Q) &= w'(K') + w'(N') + a_1/2 + a_2/2 \\ &= w(K') + w(N') + a_1/2 + a_2/2 \\ &\leq w(K) + w(N) + a_1/2 + a_2/2 \\ &\leq 1 \; . \end{aligned}$$

□

Let us also just mention that in the case of *ordinary* sorting (i.e. starting from an antichain) – or more generally, in any situation where we know $a(P)$ for the initial poset P – we get the algorithmic application (that is, forcing $\Omega(\log(e(P)))$ comparisons) without recourse to the ellipsoid algorithm, by maintaining some good, though not necessarily optimal, $a \in VP(P')$ (with P' the evolving poset).

Acknowledgments Thanks to Leo Khachiyan for acting as separation oracle oracle. The idea that entropy might have something to do with comparison algorithms was first suggested to us by Ravi Boppana's elegant lower bound argument in [1].

References

[1] R. Boppana, Optimal separations between concurrent-write parallel machines, *Proc. 21st Annual ACM Symposium on Theory of Computing* (1989), 320-326.

[2] H. Buseman,*Convex Surfaces*, Interscience, New York, 1985.

[3] V. Chvátal, On certain polytopes associated with graphs, *J. Combinatorial Th. B* **18** (1975), 138-154.

[4] I. Csiszár, J. Körner, L. Lovász, K. Marton and G. Simonyi, Entropy splitting for antiblocking corners and perfect graphs, *Combinatorica* **10** (1990), 27-40.

[5] R.P. Dilworth, Some combinatorial problems on partially ordered sets, *Proc. AMS Symposia in Appl. Math* **10** (1960), 85-90.

[6] M.E. Dyer, A.M. Frieze and R. Kannan, A random polynomial time algorithm for approximating the volume of convex bodies, *Proc. 21st Annual ACM Symposium on Theory of Computing* (1989), 375-381.

[7] M.L. Fredman, How good is the information theory bound in sorting?, *Theoretical Computer Science* **1** (1976), 355-361.

[8] J. Friedman, A note on poset geometries,

[9] D.R. Fulkerson, On the perfect graph theorem, pp. 69-77 in *Mathematical Programming* (T.C. Hu and S.M. Robinson, eds.), Academic Press, New York, 1973.

[10] M. Grötschel, L. Lovász and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization,* Springer-Verlag, 1988.

[11] J. Kahn and N. Linial, Balancing extensions via Brunn-Minkowski, *Combinatorica* **11** (1991), 363-368.

[12] J. Kahn and M. Saks, Balancing poset extensions, *Order* **1** (1984), 113-126.

[13] A. Karzanov and L. Khachiyan, On the conductance of order Markov chains, *Order* **8** (1991), 7-15.

[14] L. Khachiyan, Optimal algorithms in convex programming, decomposition and sorting, in *Computers and Decision Problems* (Ju. Jaravlev, ed.) Moscow, Nauka, 1989, pp. 161-205 (Russian).

[15] S. S. Kislitsyn, A finite partially ordered set and its corresponding set of permutations, *Mat. Zametki* **4** (1968), 511-518.

[16] J. Komlós, A strange pigeon-hole principle, *Order* **7** (1990), 107-113.

[17] J. Körner, Coding of an information source having ambiguous alphabet and the entropy of graphs, *Trans. 6th Prague Conf. information Th. etc.* (1973) 411-425

[18] J. Körner, Fredman-Komlós bounds and information theory, *SIAM J. Alg. Disc. Meth.* **7** (1986), 560-570.

[19] J. Körner and K. Marton, New bounds for perfect hashing via information theory, *Europ. J. Combinatorics*

[20] N. Linial, The information theoretic bound is good for merging, *SIAM J. Computing* **13** (1984), 795-801.

[21] L. Lovász and M. Simonovits, The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume, *Proc. 31st IEEE Symp. Found. Comp. Sci.* (1990), 346-355.

[22] R.P. Stanley, Two poset polytopes, *Discrete and Computational Geom.* **1** (1986), 9-23.

[23] W.T. Trotter, Problems and conjectures in the combinatorial theory of ordered sets, *Ann. Discrete. Math.* **41** (1989), 401-416.