The Total Linear Discrepancy of an Ordered Set

David Howard
School of Mathematics
Georgia Institute of Technology,
Atlanta, GA 33032 USA

Randy Shull
Computer Science Department
Wellesley College
Wellesley, MA 02481 USA

Noah Streib
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 33032 USA

Ann N. Trenk
Department of Mathematics
Wellesley College
Wellesley, MA 02481 USA

October 18, 2009

ABSTRACT

In this paper we introduce the notion of the *total linear discrepancy* of a poset as a way of measuring the fairness of linear extensions. If L is a linear extension of a poset P, and x, y is an incomparable pair in P, the height difference between x and y in L is $|L(x) - L(y)|$. The total linear discrepancy of P in L is the sum over all incomparable pairs of these height differences. The total linear discrepancy of P is the minimum of this sum taken over all linear extensions L of P. While the problem of computing the (ordinary) linear discrepancy of a poset is NP-complete, the total linear discrepancy can be computed in polynomial time. Indeed, in this paper, we characterize those linear extensions that are optimal for total linear discrepancy. The characterization provides an easy way to count the number of optimal linear extensions.

1 Introduction

In this paper we consider only finite posets. We begin with some definitions and notation. We denote the cardinality of set S by $|S|$. A poset $P = (X, \prec)$ consists of a ground set X together with an order relation $\prec$. If there are several posets under consideration, we write $\prec_P$. When points $x, y \in X$ are incomparable we write $x \parallel_P y$ or just $x \parallel y$. If there are no incomparabilities then P is a *linear order* or *chain*. A *linear extension* L of a poset P is a linear order that respects the relation of P, that is, $x \prec_L y$ whenever $x \prec_P y$. The *height* of a point x in a linear order L, denoted by $L(x)$, is the greatest cardinality of a chain whose maximum point is x. The *downset* of $x \in X$, denoted $D(x)$, is $\{v \in X : v \prec x\}$. Similarly, the *upset* of $x \in X$, denoted $U(x)$, is $\{w \in X : x \prec w\}$.

The linear discrepancy of a poset, written $\mathrm{ld}(P)$, was introduced by Tanenbaum, Trenk and Fishburn [**?**] as a measure of how far a poset is from being a linear order. It was studied further in [**?**], [**?**], [**?**], [**?**] and [**?**]. Formally,

$$\mathrm{ld}(P) = \min_L \max_{x \parallel y} |L(x) - L(y)|$$

where the minimum is taken over all linear extensions L of P.

The concept of linear discrepancy arises in many real world problems where a linear extension of a poset is required and in the interest of fairness it is desirable to choose one that minimizes the difference in height of incomparable points. Examples appear in [**?**].

In this paper we consider a different measure of fairness. Rather than seeking to minimize the *maximum* difference in height between incomparable elements, we now seek to minimize the *average* such difference. Equivalently, we seek to minimize the *sum* of such differences.

2 Total Linear Discrepancy

Definition 1 The *total linear discrepancy* of a poset P, written $\mathrm{tl}(P)$, is

$$\min_L \sum_{x \parallel y} |L(x) - L(y)|$$

where the minimum is taken over all linear extensions L of P. A linear extension for which this minimum value is achieved is called *optimal*.

It will sometimes be useful to refer to the sum in Definition **??** for a particular linear extension.

Definition 2 Let P be a poset and L be a linear extension of P. The *total discrepancy of P in L*, written $t_L(P)$, is

$$\sum_{x \parallel y} |L(x) - L(y)|.$$

For example, the linear extension $L : a \prec b \prec c \prec d \prec f \prec e$ of the fish poset F from Figure **??** has $t_L(F) = 6$. This linear extension is optimal since F contains five incomparable pairs, and point c participates in three of them. The linear extension $L' : a \prec c \prec b \prec d \prec f \prec e$ has $t_{L'}(F) = 9$ and thus is not optimal.

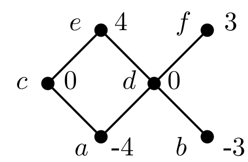

Figure 1: Poset F labeled with net heights.

There are no known efficient algorithms for computing the linear discrepancy of a poset. Indeed, the decision problem of determining whether $\mathrm{ld}(P) \leq k$ is NP-complete [**?**]. Surprisingly, the situation is quite different for total linear discrepancy. Not only can the total linear discrepancy of a poset be computed in polynomial time, but Theorem **??** characterizes those linear extensions that are optimal.

Intuitively, a point with large downset should appear higher in an optimal linear extension and one with large upset should appear lower. This motivates the following definitions which play a key role in characterizing optimal linear extensions.

Definition 3 Let $P = (X, \prec)$ be a poset. The *net height* of $x \in X$, written $\hat{h}(x)$, is $|D(x)| - |U(x)|$.

Definition 4 A linear extension L of poset P is *height ordered* if $L(x) < L(y)$ whenever $\hat{h}(x) < \hat{h}(y)$.

Figure **??** shows the net height $\hat{h}(x)$ listed next to each point x of the poset F. Observe that two points with equal net heights are incomparable and that comparable pairs of points have net heights that differ by at least two. We record this in the following remark.

Remark 5 *If $x \prec y$ in P then $\hat{h}(x) + 2 \leq \hat{h}(y)$.*

Proof. Given that $x \prec y$, transitivity implies that $D(x) \subset D(y)$ and $U(y) \subset U(x)$. Indeed, $|D(x)| + 1 \leq |D(y)|$ and $|U(y)| + 1 \leq |U(x)|$ because $x \in D(x) \backslash D(y)$ and $y \in U(x) \backslash U(y)$. The result follows from the definition of net height. $\Box$

The next lemma calculates the effect on total discrepancy of swapping two consecutive points in a linear extension. We have seen an example of this lemma in the linear extensions L and L' of the fish poset F. In that instance (with $x = b$ and $y = c$) we have $t_{L'}(F) = 9 = 6 + 0 - (-3) = t_L(F) + \hat{h}(c) - \hat{h}(b)$.

Lemma 6 *Let L be a linear extension of poset P and let x, y be incomparable elements in P with $L(y) = L(x) + 1$. If L' is the linear extension of P formed by swapping x and y, then*

$$t_{L'}(P) = t_L(P) + \hat{h}(y) - \hat{h}(x).$$

Proof. Define Δ to be $t_{L'}(P) - t_L(P)$. For incomparable pairs u, v with $u, v \notin \{x, y\}$, the terms $|L(u) - L(v)|$ and $|L'(u) - L'(v)|$ are identical. Similarly, they are identical for the incomparable pair x, y. Thus in computing Δ we need only consider the contribution arising from incomparable pairs in which one point is in the set $\{x, y\}$ and the other point t is not. Furthermore, if t is incomparable to both x and y then the sum $|L(t) - L(x)| + |L(t) - L(y)|$ is equal to the sum $|L'(t) - L'(y)| + |L'(t) - L'(x)|$. Thus we need only consider the pairs in which t is incomparable to one of x, y and comparable to the other. There are four such cases to consider in computing Δ.

(i) $w : L(w) < L(x), \;\; w \parallel x, \text{ and } w \prec y.$

(ii) $z : L(z) > L(y), \;\; x \prec z, \text{ and } z \parallel y.$

(iii) $w' : L(w') < L(x), \;\; w' \prec x, \text{ and } w' \parallel y.$

(iv) $z' : L(z') > L(y), \;\; z' \parallel x, \text{ and } y \prec z'$

Each point w in (i) and z in (ii) contributes $+1$ to Δ, and each point w' in (ii) and z' in (iv) contributes -1 to Δ. The number of points w in (i) is $|D(y)| - |D(x) \cap D(y)|$ since each w with $L(w) < L(x)$ will have either $w \prec x$ or $w \parallel x$. Similarly, the number of points z in (ii) is $|U(x)| - |U(x) \cap U(y)|$, the number of points w' in (iii) is $|D(x)| - |D(x) \cap D(y)|$ and the number of points z' in (iv) is $|U(y)| - |U(x) \cap U(y)|$. Thus $\Delta = |D(y)| + |U(x)| - |D(x)| - |U(y)| = \hat{h}(y) - \hat{h}(x)$ as desired. □

We are now ready to characterize the linear extensions of P that are optimal with respect to total linear discrepancy.

Theorem 7 *A linear extension L is optimal with respect to total linear discrepancy if and only if L is height ordered.*

Proof. First we prove the forward direction. Assume, for a contradiction, that L is an optimal linear extension of P but that it is not height ordered. Let x, y be a pair of points so that $L(x) < L(y)$ and $\hat{h}(x) > \hat{h}(y)$ and for which $L(y) - L(x)$ is as small as possible. Suppose there exists a point z with $L(x) < L(z) < L(y)$. If $\hat{h}(x) > \hat{h}(z)$, then the pair x, z violates the minimality of $L(y) - L(x)$, and otherwise, $\hat{h}(z) \geq \hat{h}(x) > \hat{h}(y)$, in which case the pair z, y violates this minimality condition. Thus no such z exists and in fact $L(y) - L(x) = 1$.

Because $\hat{h}(x) > \hat{h}(y)$, Remark **??** implies that $x \not\prec y$. Furthermore, since L is a linear extension of P and $L(x) < L(y)$, we know $y \not\prec x$. Thus $x \parallel y$. Swap x and y to obtain another linear extension L' of P. By Lemma **??** we have, $t_P(L') = t_P(L) + \hat{h}(y) - \hat{h}(x) < t_P(L)$. This contradicts the optimality of L.

Next we prove the converse. Let $\hat{L}$ be a linear extension of P that is height ordered and let L be a linear extension of P that is optimal with respect to total linear discrepancy. By the first half of this proof, L is also height ordered. Therefore, $\hat{L}$ and L differ only in the order of points with the same net height and we can transform $\hat{L}$ to L by a sequence of swaps of consecutive points with equal net height. By the contrapositive of Remark **??**, each such swap involves an *incomparable pair* x, y with $\hat{h}(x) = \hat{h}(y)$. By Lemma **??**, each swap leaves the total discrepancy unchanged thus $t_{\hat{L}}(P) = t_L(P)$ and $\hat{L}$ is also an optimal linear extension. □

Example 8 It follows from Theorem **??** that the poset F in Figure **??** has exactly two optimal linear extensions, where points c and d may appear in either order: $a \prec b \prec \{c, d\} \prec f \prec e$.

In general, Theorem **??** allows us to find an optimal linear extension efficiently and from there to calculate the total linear discrepancy. It also allows us to calculate the number of optimal linear extensions. We record these as corollaries.

Corollary 9 *Let P be a poset and $a_1, a_2, \ldots, a_r$ be the set of distinct net heights that occur among points of P. If b_i is the number of points of P that have net height equal to a_i, then the number of linear extensions of P that are optimal is $b_1!\ b_2!\ \cdots\ b_r!$.*

Corollary 10 *A linear extension of a poset P that is optimal with respect to total linear discrepancy can be constructed in polynomial time.*

3 Special Classes of Posets

In this section, we consider applying our results to several special classes of posets – antichains, the standard examples S_n of posets of dimension n, and the sum of chains.

While Theorem **??** allows us to determine precisely which linear extensions of a poset are optimal, it does not provide a closed form expression for the value of the total linear discrepancy. We do have formulas for the total linear discrepancy in two special cases.

Lemma 11 *If A_n is an antichain on n points then* $\mathrm{tl}(A_n) = \binom{n+1}{3}$.

Proof. We proceed by induction. For A_2 the result is clearly true. We assume $\mathrm{tl}(A_{k-1}) = \binom{k}{3}$ and show $\mathrm{tl}(A_k) = \binom{k+1}{3}$. Any linear extension $L : x_1 \prec x_2 \prec \cdots \prec x_k$ of A_k will be optimal, so we need only calculate $t_L(A_k) = \sum_{1 \leq i < j \leq k} |L(x_i) - L(x_j)|$. Separating out the terms involving x_k yields

$$t_L(A_k) = (1 + 2 + 3 + \cdots + k - 1) + \mathrm{tl}(A_{k-1}) = \binom{k}{2} + \binom{k}{3} = \binom{k+1}{3}. \ \square$$

The poset $S_n = (X, \prec)$ is called the *standard example* of a poset of dimension n. It has as its ground set $X = \{x_1, x_2, \ldots, x_n\} \cup \{y_1, y_2, \ldots, y_n\}$, and the only comparabilities are $x_i \prec y_j$ for $i \neq j$.

Proposition 12 *If S_n is the standard example poset on n points then* $\text{tl}(S_n) = 2\binom{n+1}{3} + n^2$.

Proof. Each minimal element x_i has net height $\hat{h}(x_i) = -(n-1)$ and each maximal element y_i has net height $\hat{h}(y_i) = (n-1)$. By Theorem **??**, any linear extension in which all the x's appear below all of the y's is optimal, so we will use the linear extension $L : x_1 \prec x_2 \cdots \prec x_n \prec y_1 \prec y_2 \cdots \prec y_n$. The minimal points form an antichain as do the maximal points. For each incomparable pair of the form $x_i \parallel y_i$, we have $|L(x_i) - L(y_i)| = n$. Thus, using Lemma **??**, we have

$$\text{tl}(S_n) = t_L(S_n) = 2\ \text{tl}(A_n) + n^2 = 2\binom{n+1}{3} + n^2. \ \Box$$

Theorem **??** was obtained almost simultaneously in [**?**] where the authors use the term *average relational distance.*

Acknowledgements: This article is an extension of the ideas in [**?**] by the second and fourth authors. We acknowledge Wellesley College undergraduate Alicia West for her work on Lemma **??** and Proposition **??**. We also thank Alan Shuchat and Douglas West for suggestions that helped improve the presentation of this paper.

References

[1] G. Brightwell and V. Patel. Average relational distance in linear extensions of posets. *Discrete Mathematics,* this issue.

[2] G.B. Chae, M. Cheong, and S.M. Kim. Irreducible posets of linear discrepancy 1 and 2. *Far East J. Math. Sci. (FJMS)*, 22(2):217–226, 2006.

[3] P.C. Fishburn, P.J. Tanenbaum, and A.N. Trenk. Linear discrepancy and bandwidth. *Order*, 18:237–245, 2001.

[4] D.M. Howard, M.T. Keller, and S.J. Young. A characterization of partially ordered sets with linear discrepancy equal to 2. *Order*, 24:139–153, 2007.

[5] D.M. Howard, G.B. Chae, M. Cheong, and S.M. Kim. Irreducible width 2 posets of linear discrepancy 3. *Order*, 25(2):105–119, 2008.

[6] M.T. Keller, and S.J. Young. A Brooks-type theorem for the bandwidth of interval graphs. Submitted for publication, 2008.

[7] R. Shull, and A.N. Trenk. Total linear discrepancy for the sum of two chains. Department of Computer Science Technical Report #CSD-TR-35-2009, Wellesley College, 2009.

[8] P.J. Tanenbaum, A.N. Trenk, and P.C. Fishburn. Linear discrepancy and weak discrepancy of partially ordered sets. *Order*, 18:201–225, 2001.