

# The Total Linear Discrepancy of an Ordered Set

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

Randy Shull  
Department of Computer Science  
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

April 21, 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  $|h_L(x) - h_L(y)|$ . 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  $h_L(x)$ , is the greatest cardinality of a chain whose maximum point is  $x$ . The *downset* of  $x \in X$  is  $D(x) = \{v \in X : v \prec x\}$  and likewise the *upset* of  $x \in X$  is  $U(x) = \{w \in X : x \prec w\}$ . For all terminology and notation not defined here, we refer the reader to [11].

The linear discrepancy of a poset was introduced by Tanenbaum, Trenk and Fishburn [10] as a measure of how far a poset is from being a linear order. It was studied further in [1], [4], [5] and [6]. Formally,

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

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 it is desirable to choose one that minimizes the difference in height of incomparable points. For example, a poset could represent a set of projects on a professor's desk, ordered by urgency (or perhaps importance). The professor must tackle these projects one at a time and should order them so that more urgent projects are done before less urgent ones (i.e., a linear extension). In addition, if two projects are incomparable, the professor would not want to complete them at widely different times, for this could be viewed as unfair by the person awaiting the completion of the second project. Thus the professor seeks a linear extension that achieves the linear discrepancy for this poset. Other examples appear in [10].

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  $P$  is

$$tld(P) = \min_L \sum_{x \parallel y} |h_L(x) - h_L(y)|.$$

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 1 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  $L$* , denoted by  $td_L(P)$ , is

$$td_L(P) = \sum_{x \parallel y} |h_L(x) - h_L(y)|.$$

For example, the linear extension  $L : a \prec b \prec c \prec d \prec f \prec e$  of the fish poset  $F$  from Figure 1 has  $td_L(F) = 6$ . 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  $td_{L'}(F) = 9$  and thus is not optimal.

There are no known efficient algorithms for computing the linear discrepancy of a poset. Indeed, the decision problem of determining whether  $ld(P) \leq k$  is NP-complete [2]. 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 7 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  $P = (V, \prec)$  be a poset. The *degree difference* of  $x \in V$ , denoted by  $dd(x)$ , is  $dd(x) = |D(x)| - |U(x)|$ .

**Definition 4** A linear extension  $L$  of poset  $P$  is *ordered by degree differences* if  $h_L(x) < h_L(y)$  whenever  $dd(x) \leq dd(y)$ .

Figure 1 shows three posets with the degree difference  $dd(v)$  listed next to each point  $v$ . Observe that two points with equal degree differences are incomparable and that comparable pairs of points have degree differences that differ by at least two. We record this in the following remark.

**Remark 5** If  $x \prec y$  in  $P$  then  $dd(x) + 2 \leq dd(y)$ .

**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  $x \in D(x) \setminus D(y)$  and  $y \in U(x) \setminus U(y)$ . The result follows from the definition of degree difference.  $\square$

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  $td_{L'}(F) = 9 = 6 + 0 - (-3) = td_L(F) + dd(c) - dd(b)$ .

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

$$td_{L'}(P) = td_L(P) + dd(y) - dd(x).$$

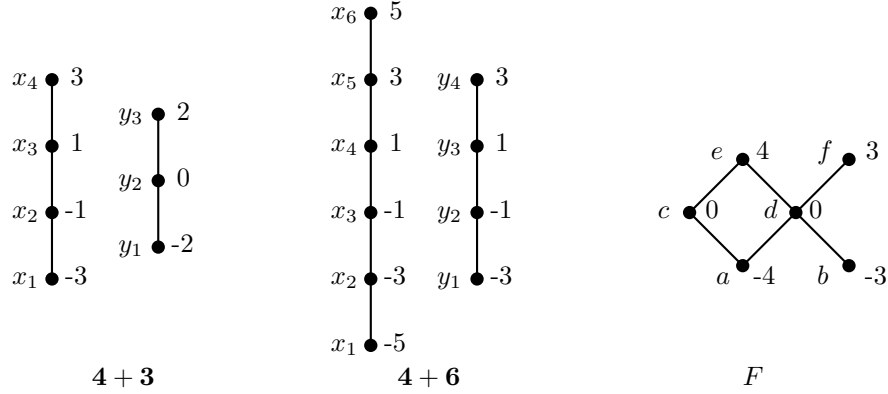


Figure 1: Three posets labeled with degree differences.

**Proof.** For incomparable pairs  $u, v$  with  $u, v \notin \{x, y\}$ , the terms  $|h_L(u) - h_L(v)|$  and  $|h_{L'}(u) - h_{L'}(v)|$  are identical. Similarly, they are identical for the incomparable pair  $x, y$ . Thus in computing  $\Delta = td_{L'}(P) - td_L(P)$ , 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  $|h_L(t) - h_L(x)| + |h_L(t) - h_L(y)|$  is equal to the sum  $|h_{L'}(t) - h_{L'}(y)| + |h_{L'}(t) - h_{L'}(x)|$ . 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  $\Delta$ .

- (i)  $w : h_L(w) < h_L(x)$ ,  $w \parallel x$ , and  $w \prec y$ .
- (ii)  $z : h_L(z) > h_L(y)$ ,  $x \prec z$ , and  $z \parallel y$ .
- (iii)  $w' : h_L(w') < h_L(x)$ ,  $w' \prec x$ , and  $w' \parallel y$ .
- (iv)  $z' : h_L(z') > h_L(y)$ ,  $z' \parallel x$ , and  $y \prec z'$

Each point  $w$  in (i) and  $z$  in (ii) contributes  $+1$  to  $\Delta$ , and each point  $w'$  in (iii) and  $z'$  in (iv) contributes  $-1$  to  $\Delta$ . The number of points  $w$  in (i) is  $|D(y)| - |D(x) \cap D(y)|$  since each  $w$  with  $h_L(w) < h_L(x)$  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  $\Delta = |D(y)| + |U(x)| - |D(x)| - |U(y)| = dd(y) - dd(x)$  as desired.  $\square$

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 ordered by degree differences.*

**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 ordered by degree differences. Let  $x, y$  be a pair of points so that  $h_L(x) < h_L(y)$  and  $dd(x) > dd(y)$  and for which  $h_L(y) - h_L(x)$  is as small as possible. Suppose there exists a point  $z$  with  $h_L(x) < h_L(z) < h_L(y)$ . Then if  $dd(x) > dd(z)$ , the pair  $x, z$  violates the minimality of  $h_L(y) - h_L(x)$ , and otherwise,  $dd(z) \geq dd(x) > dd(y)$ , in which case the pair  $z, y$  violates this minimality condition. Thus no such  $z$  exists and in fact  $h_L(y) - h_L(x) = 1$ .

Because  $dd(x) > dd(y)$ , Remark 5 implies that  $x \not\prec y$ . Furthermore, since  $L$  is a linear extension of  $P$  and  $h_L(x) < h_L(y)$ , we know  $y \not\prec x$ . Thus  $x \parallel y$ . Swap  $x$  and  $y$  to obtain another linear extension  $L'$  of  $P$ . By Lemma 6 we have,  $td_P(L') = td_P(L) + dd(y) - dd(x) < td_P(L)$ . This contradicts the optimality of  $L$ .

Next we prove the converse. Let  $\hat{L}$  be a linear extension of  $P$  that is ordered by degree differences 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 ordered by degree differences. Therefore,  $\hat{L}$  and  $L$  differ only in the order of points with the same degree difference and we can transform  $\hat{L}$  to  $L$  by a sequence of swaps of consecutive points with equal degree difference. By the contrapositive of Remark 5, each such swap involves an *incomparable pair*  $x, y$  with  $dd(x) = dd(y)$ . By Lemma 6, each swap leaves the total linear discrepancy unchanged thus  $td(\hat{L}) = td(L)$  and  $\hat{L}$  is also an optimal linear extension.  $\square$

**Example 8** We use Theorem 7 to enumerate the optimal linear extensions for the posets in Figure 1.

- For the poset  $4 + 3$ , the degree differences are distinct and thus there is a unique optimal linear extension:  $x_1 \prec y_1 \prec x_2 \prec y_2 \prec x_3 \prec y_3 \prec x_4$ .
- For the poset  $4 + 6$ , there are 16 optimal linear extensions each having the following form, where the points in a set together may appear in either order:  $x_1 \prec \{y_1, x_2\} \prec \{y_2, x_3\} \prec \{y_3, x_4\} \prec \{y_4, x_5\} \prec x_6$ .
- The fish poset  $F$  has two optimal linear extensions as follows, where points  $c$  and  $d$  may appear in either order:  
 $a \prec b \prec \{c, d\} \prec f \prec e$ .

Theorem 7 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  $a_1, a_2, \dots, a_r$  be the set of distinct degree differences that occur among points of  $P$ . Furthermore, let  $b_i$  be the number of points of  $P$  that have degree difference 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 7 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  $tld(A_n) = \binom{n+1}{3}$ .*

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

$$td_L(A_k) = (1 + 2 + 3 + \dots + k - 1) + tld(A_{k-1}) = \binom{k}{2} + \binom{k}{3} = \binom{k+1}{3}$$

as desired.  $\square$

The poset  $S_n = (X, \prec)$  is called the *standard example* of a poset of dimension  $n$ . It has as its ground set  $X = \{x_1, x_2, \dots, x_n\} \cup \{y_1, y_2, \dots, y_n\}$ , 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  $tld(S_n) = 2\binom{n+1}{3} + n^2$ .*

**Proof.** Each minimal element  $x_i$  has degree difference  $dd(x_i) = -(n-1)$  and each maximal element  $y_i$  has degree difference  $dd(y_i) = (n-1)$ . By Theorem 7, 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  $L : x_1 \prec x_2 \prec \dots \prec x_n \prec y_1 \prec y_2 \prec \dots \prec y_n$ . 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  $|h_L(x_i) - h_L(y_i)| = n$ . Thus, using Lemma 11, we have

$$tld(S_n) = td_L(S_n) = 2 tld(A_n) + n^2 = 2\binom{n+1}{3} + n^2. \quad \square$$

Finally, we apply our results to the special case of posets which are the sum of chains, and consider our original goal of achieving linear extensions that are fair. The poset  $\mathbf{r} + \mathbf{s}$  consists of the disjoint union of a chain of  $r$  points with a chain of  $s$  points, and Figure 1 shows the posets  $\mathbf{4} + \mathbf{3}$  and  $\mathbf{4} + \mathbf{6}$ . In the poset  $\mathbf{r} + \mathbf{s}$ , the degree differences of points in the  $r$ -chain increase by twos from

$-(r-1)$  to  $r-1$ , and analogously for the  $s$ -chain. Thus if  $r$  and  $s$  have opposite parity, the degree differences are all distinct and by Theorem 7, there is a unique optimal linear extension with respect to total linear discrepancy. If  $r$  and  $s$  have the same parity, at most two points share the same degree difference, and in optimal linear extensions the points in these pairs can appear in either order, as in Example 8.

We can think of a poset  $P = \mathbf{r} + \mathbf{s}$  as representing a line of  $r$  people (or projects) and a separate line of  $s$  people (or projects) waiting to be served (or completed). A linear extension gives a way of merging these into one line while preserving the order of each line. If  $r = s$  or  $r = s - 1$ , a natural way to accomplish the merge is to have points from each line alternate. Indeed Theorem 7, shows that these are optimal solutions with respect to total linear discrepancy and the other optimal solutions (in the case  $r = s$ ) are not too different. More generally, Theorem 7 shows that for any values of  $r$  and  $s$ , in any optimal solution to the total discrepancy problem for  $\mathbf{r} + \mathbf{s}$ , two formerly adjacent members are never separated by more than three positions and there exists an optimal solution for which the separation is never more than two.

In contrast, the optimal solutions with respect to linear discrepancy insert one line into the middle of the other [10]. In this case, two formerly adjacent members from one line will find themselves separated by the entire length of the second, a situation that might appear unfair to the member of the first line occupying the lower position.

Analogously, if  $P$  is the sum of  $t$  chains, by Theorem 7, any optimal solution to the total discrepancy problem never separates two formerly adjacent members by more than  $2(t-1) + 1$  positions and there exists an optimal solution for which the separation is never more than  $t$  positions. In contrast, the optimal solutions with respect to linear discrepancy insert all the remaining points into the middle of a largest chain [10]. In this case, two formerly adjacent members from a longest line will find themselves separated by the all the member of the other lines. We conclude that at least in the case of posets which are the sums of chains, total linear discrepancy appears to provide a better measure of fairness than linear discrepancy.

**Acknowledgements:** This article is an extension of the ideas in [9] by the second and fourth authors. We acknowledge Wellesley College undergraduate Alicia West for her work on Lemma 11 and Proposition 12. We also thank Alan Shuchat for conversations that helped us to improve the presentation of this paper.

## References

- [1] 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.
- [2] P.C. Fishburn, P.J. Tanenbaum, and A.N. Trenk. Linear discrepancy and bandwidth. *ORDER*, 18:237–245, 2001.

- [3] J.G. Gimbel and A.N. Trenk. On the weakness of an ordered set. *SIAM J. Discrete Math.*, 11:655–663, 1998.
- [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] A. Shuchat, R. Shull, and A. Trenk. Range of the fractional weak discrepancy function. *ORDER*, 23:51–63, 2006.
- [8] A. Shuchat, R. Shull, and A. Trenk. Fractional weak discrepancy of posets and certain forbidden configurations. *The Mathematics of Preference, Choice and Order: Essays in Honor of Peter C. Fishburn (Studies in Choice and Welfare)* Edited by Steven Brams, William V. Gehrlein, and Fred S. Roberts, Springer-Verlag, New York (2009).
- [9] 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.
- [10] P.J. Tanenbaum, A.N. Trenk, and P.C. Fishburn. Linear discrepancy and weak discrepancy of partially ordered sets. *ORDER*, 18:201–225, 2001.
- [11] W. Trotter, *Combinatorics and partially ordered sets: Dimension theory* Johns Hopkins Univ. Press, Baltimore, 1992.