

# Approximating L1-Norm Best-Fit Lines

J. Paul Brooks

Virginia Commonwealth University, [jpbrooks@vcu.edu](mailto:jpbrooks@vcu.edu)

José H. Dulá

Univeristy of Alabama, [jhdula@cba.ua.edu](mailto:jhdula@cba.ua.edu)

January 2, 2019

## Abstract

Sufficient conditions are provided for a deterministic algorithm for estimating an L1-norm best-fit one-dimensional subspace. To prove the conditions are sufficient, fundamental properties of the L1-norm projection of a point onto a one-dimensional subspace are derived. Also, an equivalence is established between the algorithm, which involves the calculation of several weighted medians, and independently-derived algorithms based on finding L1-norm solutions to overdetermined system of linear equations, each of which may be calculated via the solution of a linear program. The equivalence between the algorithms implies that each is a 2-factor approximation algorithm, which is the best-known factor among deterministic algorithms, and that the method based on weighted medians has the smallest worst-case computational requirements.

**Keywords.** L1-norm line fitting; L1-norm location; L1-norm subspace estimation; weighted median; L1-norm principal component analysis

## 1 Introduction

Given points  $x_i \in \mathbb{R}^m$ ,  $i = 1, \dots, n$ , an *L1-norm best-fit one-dimensional subspace*  $\{v\alpha : \alpha \in \mathbb{R}\}$  is defined by the vector  $v \in \mathbb{R}^m$  and is an optimal solution to

$$\min_{\substack{v \in \mathbb{R}^m, \\ \alpha_i \in \mathbb{R}, i=1, \dots, n}} \sum_{i=1}^n \|x_i - v\alpha_i\|_1. \quad (1)$$

Fitting subspaces and other affine sets to data is a fundamental task in data analysis techniques such as linear regression, logistic regression, and

principal component analysis (PCA) and for applications including ranking, recommendation engines, text mining, and video analysis. There are also applications in location analysis. Replacing the traditional sum-of-squared error criterion with the L1 norm can decrease the sensitivity of solutions to outlier observations in data.

The problem (1) is nonlinear, non-convex, and non-differentiable with a potential for many local optima. Efficient algorithms for the case when  $m = 2$  and for a related problem when fitting an  $m - 1$ -dimensional subspace (a hyperplane) are known. But as Schöbel [2015] notes, locating a line in more than two dimensions “turns out to be a difficult problem since all of the structure of line and hyperplane location problems gets lost.” The problem expressed in (1) has been shown to be NP-hard Gillis and Vavasis [2018].

In this work, we establish an equivalence between two algorithms for efficiently estimating an L1-norm best-fit one-dimensional subspace and we derive sufficient conditions for these algorithms to provide an optimal solution to (1). One of the algorithms is based on the calculation of several weighted medians [Brooks et al., 2017] and the other is based on finding the L1-norm solution to several overdetermined systems of linear equations, which may be cast as linear programs (LPs) [Tsagkarakis et al., 2016, Chierichetti et al., 2017a,b]. We demonstrate that these algorithms produce estimates with identical error and therefore have the same approximation factor (Section 3). The sufficient conditions that guarantee a globally optimal solution emerge from this equivalence and from fundamental properties of L1-norm projection of a point onto a one-dimensional subspace (Section 4).

The equivalence between the algorithms in [Brooks et al., 2017, Tsagkarakis et al., 2016, Chierichetti et al., 2017a,b] establishes them each as 2-factor approximation algorithms for (1), which is the best-known to date among deterministic algorithms [Chierichetti et al., 2017a,b] and may be the best-possible for a deterministic algorithm [Ban et al., 2018]. The equivalence also reveals the algorithm that has the smallest worst-case computational requirements.

## 1.1 Notation and Definitions

Unless otherwise specified, we suppose we are given points  $x_i \in \mathbb{R}^m$ ,  $i = 1, \dots, n$  that comprise the rows of a data matrix  $X \in \mathbb{R}^{n \times m}$ . The coordinate values of point  $x_i$  are indicated by  $x_{ij}$ ,  $j = 1, \dots, m$ . When  $m = 1$ , we drop the index  $j$  and the values are given by  $x_i$ ,  $i = 1, \dots, n$ . When we consider only one point ( $n = 1$ ), we denote the point as  $x$  and drop the index  $i$  so that the values are given by  $x_j$ ,  $j = 1, \dots, m$ . The notation  $x \neq 0$  will be

used to indicate that at least one coordinate value is nonzero. The notation  $\text{sgn}(x_j)$  indicates the sign of  $x_j$ .

The L1 norm of  $x \in \mathbb{R}^m$  is  $\|x\|_1 = \sum_{j=1}^m |x_j|$ . The L2 norm of  $x$  is  $\|x\|_2 = \left(\sum_{j=1}^m x_j^2\right)^{1/2}$ .

For a point/vector  $x \in \mathbb{R}^m$ ,  $x_{j'}$  is a *dominating coordinate* value if  $|x_{j'}| > \sum_{j \neq j'} |x_j|$ .

Because we are considering fitting one-dimensional subspaces, we will assume throughout the paper that all lines are *anchored* in that they contain the origin. A vector  $v \in \mathbb{R}^m$  *defines* a line if the points on the line are given by the set  $\{v\alpha : \alpha \in \mathbb{R}\}$ .

Given a one-dimensional subspace defined by  $v$ , one can solve (1) for  $\alpha_i$  to find projections  $v\alpha_i$  of the points onto the line. The *error* for point  $x_i$  is  $\|x_i - v\alpha_i\|_1 = \sum_{j=1}^m |x_{ij} - v_j\alpha_i|$ . If  $|x_{ij} - v_j\alpha_i| > 0$  for an optimal  $\alpha_i$  and coordinate  $j$ , then we say that the point *uses* unit direction  $e^j$ , given by the  $j^{\text{th}}$  unit vector, to project onto the line defined by  $v$ . We let  $\text{sgn}(j)$  denote the sign of the unit direction used for projection: if  $(x_{ij} - v_j\alpha_i) > 0$ , then  $\text{sgn}(j) = +$  and the point uses direction  $e^j$  to project; if  $(x_{ij} - v_j\alpha_i) < 0$ , then  $\text{sgn}(j) = -$  and the point uses direction  $-e^j$  to project. If  $|x_{ij} - v_j\alpha_i| = 0$ , we say that the  $j^{\text{th}}$  coordinate of  $x_i$  is *preserved* or *fixed*.

The following definition of the weighted median features prominently in our results.

**Definition 1.** A *weighted median* of numbers  $x_i \in \mathbb{R}$ ,  $i = 1, \dots, n$  with weights  $w_i \in \mathbb{R}_+$ ,  $i = 1, \dots, n$  is  $x_{\bar{i}}$ , where

$$\sum_{i: x_i > x_{\bar{i}}} w_i \leq \frac{1}{2} \sum_{i=1}^n w_i,$$

and

$$\sum_{i: x_i < x_{\bar{i}}} w_i \leq \frac{1}{2} \sum_{i=1}^n w_i.$$

Note that the weighted median is not necessarily unique.

The following proposition relating the weighted median to an optimization problem is well known and dates back to the 1700s (see [Bloomfield and Steiger, 1984, Farebrother, 1990]).

**Proposition 1.** *A weighted median of  $x_i \in \mathbb{R}$ ,  $i = 1, \dots, n$  with weights  $w_i \in \mathbb{R}_+$ ,  $i = 1, \dots, n$  is an optimal solution to the problem*

$$\min_{z \in \mathbb{R}^n} \sum_{i=1}^n w_i |x_i - z_i|. \quad (2)$$

The proposition may be proved by reformulating (2) as an LP and showing that a weighted median is an optimal solution to the primal and deriving a dual feasible solution with the same objective function value.

## 1.2 Previous Related Work

L1 regression is a special case of optimal subspace fitting to data. The history of L1 regression involves weighted medians and LP formulations to solve overdetermined systems (see [Bloomfield and Steiger, 1984]). Efficient algorithms exist for L1 regression based on the calculation of several weighted medians or the solution of a single LP. Weighted medians and linear programming are closely intertwined themselves and play new an interesting roles in our presentation. The distance from a point to its projection on the regression hyperplane is limited to the single unit direction corresponding to the response variable. There is no such restriction for best-fit subspaces. L1 regression, and optimal hyperplane fitting more generally, are different from subspace fitting when the subspaces have fewer than  $m - 1$  dimensions (a line is a hyperplane only when  $m = 2$ ) in another important way; namely, the latter is NP-hard for general  $n$  and  $m$  and for lines in particular [Gillis and Vavasis, 2018].

The L1-norm best-fit line problem has been specifically treated in two and three dimensions. Recall that when  $m = 2$ , a line is also a hyperplane. Aspects of algorithms previously proposed for fitting general lines when  $m = 2$  and  $m = 3$  are evident in the approximation algorithms studied here. Morris and Norback [1983] note that for  $m = 2$ , optimal solutions can be derived by solving two LPs. Imai et al. [1989] provide an  $O(n)$  algorithm. When  $m = 2$ , a line is a hyperplane, so methods for finding best-fit hyperplanes may be employed [Zemel, 1984, Martini and Schöbel, 1998, Brooks and Dulá, 2013]. Zemel [1984] describes an optimal algorithm based on an  $O(n)$  method for solving the linear multiple choice knapsack problem. Brooks and Dulá [2013] derive a relationship between finding a best-fit hyperplane and L1-norm linear regression; Schöbel [1998] notes the relationship for the particular case of a line when  $m = 2$ . An L1-norm regression hyperplane can be derived by solving an LP [Charnes et al., 1955]

and exhibits several interesting properties [Appa and Smith, 1973]. For the anchored case, fitting an L1-norm regression line involves the calculation of a single weighted median (see [Bloomfield and Steiger, 1984]). We will show that for the case when  $m = 2$  and the line is anchored at the origin, the method introduced by Brooks et al. [2017] finds a best-fit line by calculating two L1-norm regression lines via the calculation of two weighted medians.

There is also work in the literature specifically for the case when the data is in  $\mathbb{R}^3$  where the problem begins to reveal its intricacies. For data in  $\mathbb{R}^3$ , Blanquero et al. [2011] develop a global optimization procedure using geometric branch-and-bound for fitting (unanchored) lines under the L1 norm. Brimberg et al. [2002] describe a heuristic for estimating L1-norm best-fit lines for  $m = 3$  by restricting the projections along horizontal paths. They also describe a heuristic for finding local optimal solutions by alternating optimization [Brimberg et al., 2003]. An algorithm specifically for solving (1) has not been proposed for  $m > 2$ .

Fitting lines in two and three dimensions is an example of fitting geometrical objects to points under different distance measures and assumptions which has a long history in the context of location theory. See [Hamacher and Nickel, 1998, Schöbel, 1999, Boltyanski et al., 1999, Díaz-Báñez et al., 2004, Blanquero et al., 2009, Schöbel, 2015] and references therein for more information. The two classes of points and lines established in Section 2 are reminiscent of properties established for the location of points and other objects.

Markopoulos et al. [2014] proved that a related problem, that of maximizing the sum of L1-norm lengths of the (L2-norm) projections of points onto a line, is NP-hard and provided an  $O(n^m)$  exact algorithm for  $n$  points. Maximizing the sum-of-squared lengths of the L2-norm of projections of points onto a line is equivalent to minimizing the sum-of-squared distances of points to their projections onto a line [Reris and Brooks, 2015, Jolliffe, 2002], but this equivalence fails to hold for the L1 norm. Several efficient heuristics for this surrogate objective have been proposed [Kwak, 2008, Nie et al., 2011, Kundu et al., 2014, Markopoulos et al., 2016].

There is a vast literature addressing more generic problems including fitting  $q$ -dimensional subspaces for  $q \geq 1$  with optimization criteria based on  $L_p$  norms with  $p \geq 1$ . In the remainder of this section, we summarize methods proposed for  $L_p$ -norm subspace fitting and their implications for deriving solutions to (1) which corresponds to the case when  $q = 1$  and  $p = 1$ . Also, see Lerman and Maunu [Lerman and Maunu, 2018] for a recent review of robust subspace estimation.

A stream of research under the name “robust principal component anal-

ysis” addresses the following L1-norm subspace estimation problem:

$$\min_{Y \in \mathbb{R}^{n \times m}} \sum_{i=1}^n \|x_i - y_i\|_1, \quad (3)$$

$$\text{s.t. } \text{rank}(Y) \leq q. \quad (4)$$

The points  $y_i \in \mathbb{R}^m$  that form the rows of  $Y$  are the projections of the points  $x_i$  in terms of the original coordinates. Basis vectors defining the best-fit subspace are not computed directly, but can be obtained by the singular value decomposition of  $Y$ . Because the rank function is difficult to compute, a convex approximation such as the nuclear norm is often employed. (Note that the difficulty with the rank function may also be avoided by explicitly decomposing the projections of points into a matrix of basis vectors and multipliers as in (1).) In robust principal component analysis, a convex approximation of the error plus a tradeoff with the constraint violation is minimized (see [Candès et al., 2011, Goldfarb et al., 2013, Vaswani et al., 2018, Ma and Aybat, 2018] and the references therein). Song et al. [2016, 2017] provide examples demonstrating that the method described by Candès et al. [2011] is not guaranteed to provide a rank  $q$  solution for a desired  $q$  and that the approximation factor may be arbitrarily poor for  $q$ -dimensional subspaces.

There have been a number of stochastic approximation algorithms suggested and analyzed for fitting L1-norm subspaces [Song et al., 2016, 2017, Chierichetti et al., 2017a,b, Kyrillidis, 2018, Dahiya et al., 2018, Ban et al., 2018]. In this paper, we analyze deterministic algorithms only.

Ke and Kanade [2003, 2005] propose a heuristic called L1-PCA which leverages LP in an alternating optimization framework where estimates for the subspace’s spanning vectors forming the columns of  $V$  are derived for fixed  $\alpha_i$ ,  $i = 1, \dots, n$  and then estimates for  $\alpha_i$ ,  $i = 1, \dots, n$  are derived for fixed  $V$ . The method is the extension of the method for  $m = 3$  proposed by Brimberg et al. [2003]. Song et al. [2016, 2017] show that the method has an approximation factor for (1) of at least  $n$  when using the sum-of-squared error solution as an initial starting point.

Tsagkarakis et al. [2016] develop a heuristic requiring the points to use the same  $m - q$  unit directions to project onto the fitted subspace. Such methods where a specified subset of the components of the points are preserved in their projection are referred to as *uniform feature preservation* methods. Their method can be seen as a generalization of the method for  $m = 3$  introduced in [Brimberg et al., 2002] where projections are restricted along horizontal paths, and also generalizes the LP-based methods

for fitting hyperplanes to the case of general  $q$ . Algorithm 2 described by Chierichetti et al. [2017a,b] is a  $(q + 1)$ -factor approximation algorithm for  $L_p$ -norm subspaces for  $p \geq 1$ ; for  $p = 1$ , the algorithm is the same as uniform feature preservation. For  $q = 1$  and  $p = 1$ , their algorithm provides an  $O(m \text{poly}(nm))$ -time 2-factor approximation algorithm. We will connect uniform feature preservation and the algorithm of Brooks et al. [2017] in Section 3, establishing the latter as the most efficient 2-factor deterministic approximation algorithm known to date.

## 2 Properties of L1-Norm Projection

In this section we present fundamental results which will provide preliminary background required to derive sufficient conditions on the data for the algorithm proposed by Brooks et al. [2017] to generate globally optimal solutions. We begin with properties of L1-norm projections onto a known line. Then we establish conditions for projections given certain characteristics of the line (Section 2.1) and the data (Section 2.2).

Properties of the L1-norm projection of a point onto a line can be derived using goal variables to reformulate (1) to a problem with a linear objective function and constraints having a single bilinear term in each:

$$\begin{aligned} \min_{\substack{v \in \mathbb{R}^m, \alpha \in \mathbb{R}^n, \\ \lambda^+, \lambda^- \in \mathbb{R}^{n \times m}}} & \sum_{i=1}^n \sum_{j=1}^m (\lambda_{ij}^+ + \lambda_{ij}^-), \\ \text{s.t.} & v_j \alpha_i + \lambda_{ij}^+ - \lambda_{ij}^- = x_{ij}, i = 1, \dots, n, j = 1, \dots, m, \\ & \lambda_{ij}^+, \lambda_{ij}^- \geq 0, i = 1, \dots, n, j = 1, \dots, m. \end{aligned} \tag{5}$$

When  $v$  is given, the projection of a point  $x$  onto the line defined by  $v$  is found by solving the following LP based on (5) for one point  $x \in \mathbb{R}^m$ .

$$\begin{aligned} \min_{\substack{\alpha \in \mathbb{R}, \\ \lambda^+, \lambda^- \in \mathbb{R}^m}} & \sum_{j=1}^m (\lambda_j^+ + \lambda_j^-), \\ \text{s.t.} & v_j \alpha + \lambda_j^+ - \lambda_j^- = x_j, j = 1, \dots, m, \\ & \lambda_j^+, \lambda_j^- \geq 0, j = 1, \dots, m. \end{aligned} \tag{6}$$

Examination of this LP reveals several interesting properties about L1-norm projection that will translate to useful properties for L1-norm line fitting. The following result is a special case of a property of the L1-norm solution

to an overdetermined set of linear equations, first noted for general  $m$  by Gauss (see Sheynin [1973], Sabo et al. [2011]). We provide a proof using arguments from linear programming.

**Proposition 2.** *Let  $v \neq 0$  be a given vector in  $\mathbb{R}^m$ . There is an L1-norm projection of the point  $x \in \mathbb{R}^m$  on the line defined by  $v$  that can be reached by using at most  $m-1$  unit directions. Moreover, if the preserved coordinate is  $\hat{j}$  and  $x_{\hat{j}} \neq 0$ , then  $v_{\hat{j}} \neq 0$ .*

*Proof.* Proof of Proposition 2. Any basic feasible solution for (6) will have  $\alpha$  as basic because it is unrestricted in sign. The remaining  $m-1$  basic variables will be one from each pair  $(\lambda_j^+, \lambda_j^-)$  for  $j \neq \hat{j}$  for some  $\hat{j}$ . The determinant of the basis is  $\pm v_{\hat{j}}$  and therefore this basis is non-singular if and only if  $v_{\hat{j}} \neq 0$ .  $\square$

Therefore, for data  $x_i \in \mathbb{R}^m$ ,  $i = 1, \dots, n$ , there is an L1-norm best-fit one-dimensional subspace  $\{v\alpha : \alpha \in \mathbb{R}\}$  solving (1) with the property that for each point  $i$  and for at least one of the coordinates  $j$ , there is no error. The algorithms that we analyze in Section 3 are optimal when there is no error for the same coordinate for all points; in Section 4 we characterize sets of points where this is always the case.

A basic feasible solution to (6) is given by  $\alpha = x_{\hat{j}}/v_{\hat{j}}$ , and, for  $j \neq \hat{j}$ :

$$\lambda_j^{\text{sgn}(j)} = \text{sgn}(j)(x_j - x_{\hat{j}}v_j/v_{\hat{j}}), \quad (7)$$

where  $\text{sgn}(j)$  is  $+$  if  $\lambda_j^+$  is basic, and  $\text{sgn}(j)$  is  $-$  if  $\lambda_j^-$  is basic. Feasibility requires that the  $m-1$  terms in (7) be nonnegative, or

$$\text{sgn}(j)(x_j - x_{\hat{j}}v_j/v_{\hat{j}}) \geq 0, \quad j \neq \hat{j}. \quad (8)$$

For a given  $v$ , this is always possible with the right choice of  $\text{sgn}(j)$ . Optimality requires that the reduced costs are nonnegative. The reduced cost for  $\alpha$  is zero because it is basic. The reduced costs for non-basic  $\lambda_j^+$  and  $\lambda_j^-$  are 2 for  $j \neq \hat{j}$ . The reduced costs associated with the pair  $(\lambda_j^+, \lambda_j^-)$  are

$$\begin{cases} 1 + \frac{1}{v_j} \sum_{j \neq \hat{j}} \text{sgn}(j)v_j, & \text{for } \lambda_j^+, \\ 1 - \frac{1}{v_j} \sum_{j \neq \hat{j}} \text{sgn}(j)v_j, & \text{for } \lambda_j^-. \end{cases} \quad (9)$$

Optimality reduces to having the pair in (9) be nonnegative. Therefore, an equivalent condition is:

$$\left| \sum_{j \neq \hat{j}} \text{sgn}(j)v_j \right| \leq |v_{\hat{j}}|. \quad (10)$$

The following proposition establishes that the L1-norm projection onto a line amounts to finding the weighted median of certain ratios.

**Proposition 3.** *For a point  $x \in \mathbb{R}^m$  and a vector  $v$  defining a one-dimensional subspace, there is an optimal multiplier  $\alpha$  for (6) that is a weighted median of  $\left\{\frac{x_j}{v_j} : v_j \neq 0\right\}$  with weights  $\{|v_j| : v_j \neq 0\}$ .*

*Proof.* Proof of Proposition 3. Given  $v$ , a vector defining a one-dimensional subspace, the objective function for (6) is

$$\sum_{j=1}^m |x_j - \alpha v_j| = \sum_{j:v_j \neq 0} |v_j| \left| \frac{x_j}{v_j} - \alpha \right| + \sum_{j:v_j=0} |x_j|.$$

The second sum is constant for all values of  $\alpha$ , so only the first sum is optimized. According to Proposition 1, an optimal value of  $\alpha$  is a weighted median of  $\left\{\frac{x_j}{v_j} : v_j \neq 0\right\}$  with weights  $\{|v_j| : v_j \neq 0\}$ .  $\square$

Proposition 3 implies that for data  $x_i \in \mathbb{R}^m$ ,  $i = 1, \dots, n$ , and for a vector  $v$  that is optimal for (1), there are optimal multipliers  $\alpha_i$  where each is a weighted median of  $\left\{\frac{x_{ij}}{v_j} : v_j \neq 0\right\}$  with weights  $\{|v_j| : v_j \neq 0\}$ . This fact was noted by Ke and Kanade [2003, 2005] in their development of an alternating optimization method for estimation solutions to (1). We will use this result to establish sufficient conditions for optimality for the algorithm proposed by Brooks et al. [2017] in Section 4.

## 2.1 There Are Two Kinds of Lines with Respect to L1-Norm Projection

In this section, we leverage the feasibility and optimality criteria for the projection LP in (6) to characterize lines for which all points project preserving a single coordinate. Lines in  $\mathbb{R}^m$  may therefore be partitioned into two categories: those where there is a dominating coordinate which then causes all points to project preserving this coordinate, and the rest in which case there are sets of points preserving different coordinates. The result has implications regarding the shortest path under the L1 norm from a customer to a linear facility. It will be instrumental in establishing the sufficient conditions in Section 4.

**Theorem 1.** *Consider a line  $\{v\alpha : \alpha \in \mathbb{R}\}$ . For a coordinate  $j'$ ,*

$$|v_{j'}| > \sum_{j \neq j'} |v_j| \tag{11}$$

if and only if all points project preserving coordinate  $j'$ .

*Proof.* Proof of Theorem 1 Suppose that

$$|v_{j'}| > \sum_{j \neq j'} |v_j|.$$

Then condition (10) is satisfied and points may project preserving the  $j'^{th}$  coordinate. Suppose  $j' \neq 1$ . There can be no points that project preserving the first coordinate because the condition expressed in (10) is violated for the first coordinate:

$$\begin{aligned} \left| \sum_{j \neq 1} \text{sgn}(j)v_j \right| &= \left| \text{sgn}(j')v_{j'} + \sum_{j \notin \{1, j'\}} \text{sgn}(j)v_j \right|, \\ &\geq \left| |v_{j'}| - \sum_{j \notin \{1, j'\}} |v_j| \right|, \\ &> |v_1|. \end{aligned}$$

Similar reasoning shows that no points project preserving any coordinate  $j \neq j'$ . Therefore, all points project preserving coordinate  $j'$ .

For the converse, suppose that all points project only preserving coordinate  $j'$ , and  $j' \neq 1$ . Then no point projects preserving coordinate 1. For any choice of  $\text{sgn}(j)$ ,  $j = 2, \dots, m$ , we can find points  $x$  where the feasibility condition (8) is satisfied. Therefore, condition (10) is not satisfied for coordinate 1 for every choice of directions; equivalently, for every choice of directions,

$$\left| \sum_{j=2}^m \text{sgn}(j)v_j \right| > |v_1|.$$

In particular,

$$- \sum_{j \notin \{1, j'\}} |v_j| + |v_{j'}| > |v_1|,$$

and so

$$|v_{j'}| > \sum_{j \neq j'} |v_j|.$$

□

We now characterize the sets of points in  $\mathbb{R}^m$  that use a particular set of  $m - 1$  unit directions to project onto a given line.

The dual of (6) is

$$\begin{aligned} \max_{\pi \in \mathbb{R}^m} \quad & \sum_{j=1}^m x_j \pi_j, \\ \text{s.t.} \quad & \sum_{j=1}^m v_j \pi_j = 0, \\ & -1 \leq \pi_j \leq 1, j = 1, \dots, m. \end{aligned} \tag{12}$$

From the dual, we can see that for a line and a point  $x$  not on the line, the points  $\{tx: -\infty \leq t \leq \infty\}$  project preserving the same coordinate. To see this, let  $(\alpha, \lambda^+, \lambda^-)$  and  $\pi$  solve the primal and dual LPs for  $x$  with objective function values  $z, w$  where  $z = w$  by strong duality. Replace the right-hand side in the primal and the objective function coefficients in the dual with  $tx$ . This maintains the primal-dual relation between the modified LPs. Note that the solution  $t(\alpha, \lambda^+, \lambda^-)$  is feasible to the primal when  $t \geq 0$  and  $t(\alpha, -\lambda^-, -\lambda^+)$  is feasible when  $t < 0$  with objective function value  $tz$ . The original optimal solution to the dual,  $\pi$ , remains feasible with new objective function value  $tw$ . Therefore  $t(\alpha, \lambda^+, \lambda^-)$  is an optimal solution when  $t \geq 0$  and  $t(\alpha, -\lambda^-, -\lambda^+)$  is optimal when  $t < 0$ . When basic, the projection unit directions from  $tx$  remain the same.

From this result for lines we see that regions where sets of points have projections that preserve the same coordinate are cones in  $\mathbb{R}^m$ . Expressions (8) and (10) are closely coupled since  $x$  determines in (8) the specific set of signs  $\text{sgn}(j)$  used in expression (10).

If condition (11) is not satisfied, then it is possible that for multiple coordinates, there are points projecting onto the line preserving them. However, it is not necessarily the case that for each coordinate, there are points that project preserving it. As in the case when condition (11) holds, for each set of signs satisfying (10) for a given set of  $m - 1$  unit directions, the feasibility conditions (8) define  $m - 1$  halfspaces containing the origin. The same set with the signs negated also generates the reflected cone. If there are no choices of signs for which  $\sum_{j=1}^m \text{sgn}(j)v_j = 0$ , then the cones have non-intersecting interiors, and there are an even number of them.

## 2.2 There Are Two Kinds of Points with Respect to L1-Norm Projection

In this section, we again leverage the feasibility and optimality criteria for the projection LP in (6) to classify points into two types: those with a dominating coordinate and those without. The distinction determines two cases for the sets of unit directions with signs that a point may use to project onto any line. In one case, nearly half of all possible combinations of projection directions will never be used for any line and in another case, we may exclude a much smaller number of combinations of projection directions. The sufficiency conditions in Section 4 are for datasets comprised of points with the same dominating coordinate, and the proof of the conditions relies on the result for the first case established here.

For a given point, there are certain directions for which there are no lines where that point projects using those directions. The following result characterizes the unit directions with signs that will never be used to project from a point to a line.

**Theorem 2.** *Suppose a point  $x \in \mathbb{R}^m$  is given. If for some coordinate  $j'$ ,*

$$|x_{j'}| > \sum_{j \neq j'} |x_j|, \quad (13)$$

*then the point will never project onto a line using the direction  $-\text{sgn}(x_{j'})e^{j'}$ .*

*Proof.* Proof of Theorem 2. Suppose that for a point  $x$  and coordinate  $j'$ , condition (13) holds. If the point preserves coordinate  $j'$ , then it does not use direction  $-\text{sgn}(x_{j'})e^{j'}$ .

Suppose that a point preserves coordinate  $j^* \neq j'$  when projecting onto a line defined by  $v$ . If  $x_{j^*} = 0$ , then the error is  $\sum_{j \neq j^*} |x_j| \geq |x_{j'}|$ , and the error would be smaller preserving coordinate  $j'$ . Therefore,  $x_{j^*} \neq 0$ . Recall that  $v_{j^*} \neq 0$ . Suppose for now that  $x_{j^*} > 0$ . Then from the feasibility condition expressed in (8),

$$\text{sgn}(j) \frac{v_j}{v_{j^*}} \leq \text{sgn}(j) \frac{x_j}{x_{j^*}}, \quad j \neq j^*.$$

If the point projects using  $-\text{sgn}(x_{j'})e^{j'}$ , then,

$$\begin{aligned}
1 + \frac{1}{v_{j^*}} \sum_{j \neq j^*} \text{sgn}(j)v_j &\leq 1 + \sum_{j \neq j^*} \text{sgn}(j) \frac{x_j}{x_{j^*}}, \\
&= \frac{1}{x_{j^*}} \left( \text{sgn}(j')x_{j'} + \sum_{j \notin \{j', j^*\}} \text{sgn}(j)x_j \right), \\
&= \frac{1}{x_{j^*}} \left( -\text{sgn}(x_{j'})x_{j'} + \sum_{j \notin \{j', j^*\}} \text{sgn}(j)x_j \right), \\
&< 1 - \frac{x_{j^*}}{x_{j^*}}, \\
&= 0,
\end{aligned}$$

violating the optimality condition (10). A symmetric argument treats the case when  $x_{j^*} < 0$ .  $\square$

For each  $j \neq j'$ , there are  $2^{m-2}$  possible remaining ways to choose the signs of projection, so there are  $2^{m-2}(m-1)$  total sets of directions with signs that cannot be used to project onto a line.

The number of  $(m-2)$ -dimensional faces of a hyperoctahedron in  $\mathbb{R}^m$  is  $2^{m-1}m$ . The L1 unit ball in  $\mathbb{R}^m$  is a regular hyperoctahedron and each set of  $m-1$  unit directions with signs can be associated with a unique  $(m-2)$ -dimensional face.

**Corollary 1.** *If for a point  $x \in \mathbb{R}^m$ , there exists an index  $j'$  such that condition (13) holds, then  $\frac{1}{2} \binom{m-1}{m}$  of the sets of unit directions with signs will not be used to project onto any line.*

*Proof.* Proof of Corollary 1. The result follows by taking the ratio of disallowed sets of directions to the total choices.  $\square$

**Theorem 3.** *If for a point  $x \in \mathbb{R}^m$  there is no dominating coordinate as in (13), and further*

$$|x_{j'}| < \sum_{j \neq j'} |x_j|,$$

*for  $j' = 1, \dots, m$ , then there are no lines for which  $x$  will preserve coordinate  $j^*$  and use unit directions  $-\text{sgn}(x_j)e^j$  for all  $j \neq j^*$  and for  $j^* = 1, \dots, m$ .*

*Proof.* Proof of Theorem 3. Suppose that a point  $x$  satisfies the conditions of the theorem and that  $x$  projects onto a line  $v$  by preserving  $j^*$ . If  $x_{j^*} = 0$  then  $\alpha = 0$  and  $\text{sgn}(j)x_j \geq 0$  for all  $j$  by (8) so that the point will not use the unit directions  $-\text{sgn}(x_j)e^j$ .

Now suppose  $x_{j^*} > 0$ . Then from the feasibility condition expressed in (8),

$$\operatorname{sgn}(j) \frac{v_j}{v_{j^*}} \leq \operatorname{sgn}(j) \frac{x_j}{x_{j^*}}, \quad j \neq j^*.$$

If the point projects using the unit directions  $-\operatorname{sgn}(x_j)e^j$ , then,

$$\begin{aligned} 1 + \frac{1}{v_{j^*}} \sum_{j \neq j^*} \operatorname{sgn}(j) v_j &\leq 1 + \sum_{j \neq j^*} \operatorname{sgn}(j) \frac{x_j}{x_{j^*}}, \\ &= \frac{1}{x_{j^*}} \left( x_{j^*} + \sum_{j \neq j^*} \operatorname{sgn}(j) x_j \right), \\ &= \frac{1}{x_{j^*}} \left( x_{j^*} - \sum_{j \neq j^*} \operatorname{sgn}(x_j) x_j \right), \\ &< 0, \end{aligned}$$

violating the optimality condition in (10). A symmetric argument treats the case when  $x_{j^*} < 0$ .  $\square$

Theorem 2 and Theorem 3 present two cases. The first case establishes that if there is a dominating coordinate of  $x$ , then by Corollary 1, we may ignore almost half of the possible unit directions for projection. If a point  $x$  satisfies the second case, we may omit consideration of directions associated with a facet of the unit ball. The number of facets of the L1 unit ball is  $2^m$ , and there are  $m - 1$  sets of directions with signs associated with each.

### 3 Approximation of L1-Norm Best-Fit One-Dimensional Subspaces via Weighted Medians and Linear Programming

In this section, we show that an L1-norm line-fitting algorithm based on weighted medians produces fitted subspaces with the same objective function value as an LP-based method, thereby providing the most efficient 2-factor approximation algorithm. We then characterize arrangements of points for which the algorithms are guaranteed to provide global optimal solutions to (1).

Consider Algorithm 1 proposed by Brooks et al. [2017] for estimating an L1-norm best-fit one-dimensional subspace. One coordinate  $\hat{j}$  is chosen for all points to preserve and sets  $v_j = 1$ . The data are projected into the two-dimensional space defined by the  $j$ - and  $\hat{j}$ -axes. The  $v_j$  coordinate is the tangent of the angle that one of the points makes with the  $\hat{j}$ -axis. The selected angle is the weighted median of the angles with weights  $|x_{i\hat{j}}|$

---

**Algorithm 1** Procedure for estimating an L1-norm best-fit one-dimensional subspace. [Brooks et al., 2017]

---

Given points  $x_i \in \mathbb{R}^m$  for  $i = 1, \dots, n$ .

```

1: Set  $z^* = \infty$ .
2: for ( $\hat{j} = 1; \hat{j} \leq m; \hat{j} = \hat{j} + 1$ )          /* Loop on fixed coordinate. */ do
3:   for ( $j = 1; j \leq m; j = j + 1$ )              /* Find each weight. */ do
4:     for ( $i = 1; i \leq n; i = i + 1$ )            /* Calculate angles with fixed
       coordinate axis. */ do
5:       Set  $\theta_i = \begin{cases} \arctan \frac{x_{ij}}{x_{i\hat{j}}} & \text{if } x_{i\hat{j}} \neq 0, \\ 0 & \text{o.w.} \end{cases}$ 
6:       Find the weighted median  $\tilde{\theta}$  of  $\{\theta_i : i = 1, \dots, n\}$  with weights
        $|x_{ij}|$ .
7:       Set  $v_j = \tan \tilde{\theta}$ .                      /* Set  $j^{\text{th}}$  weight. */
8:       Set  $z_j = \sum_{i=1}^n \sum_{j=1}^m |x_{ij} - x_{ij}v_j|$ .
9:       if  $z_j < z^*$ ,                               /* Update optimal solution. */ then
10:        Set  $z^* = z_j, v^* = v$ .

```

---

(Figure 3). The estimated line with the smallest error among all choices for  $\hat{j}$  is returned.

The angle that each point makes with respect to the  $\hat{j}$ -axis in the space defined by the  $j$ - and  $\hat{j}$ -axes is  $\theta_i = \arctan \frac{x_{ij}}{x_{i\hat{j}}}$ . Note that without loss of generality, all points may be projected into the non-negative halfspace  $\{x \in \mathbb{R}^m : x_{j^*} \geq 0\}$ . Also, because the arctangent function is monotonically increasing, sorting the ratios  $\frac{x_{ij}}{x_{i\hat{j}}}$ ,  $i = 1, \dots, n$  is equivalent to sorting the angles  $\theta_i$ ,  $i = 1, \dots, n$ .

Now consider an estimate of an L1-norm best-fit one-dimensional subspace that results from modifying the nonlinear program in (5). The modification is to impose the preservation of one of the coordinates,  $\hat{j}$ , in the projections of the  $n$  data points which means each point will use the same  $m - 1$  unit directions to project onto the line defined by  $v$ , as suggested independently by Brooks and Dulá [2016], Tsagkarakis et al. [2016], and Chierichetti et al. [2017a,b]. This condition transforms (5) into an LP. The basic step of the procedure requires fixing one of the  $m$  coordinates,  $\hat{j}$ , which will be preserved in the projections. By Proposition 2, if  $x_{i\hat{j}} \neq 0$  for any  $i$ , then  $v_{\hat{j}} \neq 0$ . Therefore, without restricting the line that will be defined by the vector  $v$ , we can set  $v_{\hat{j}} = 1$ . This amounts to a simple normalization of the vector  $v$ ; the rest of its components remain variable. From this we

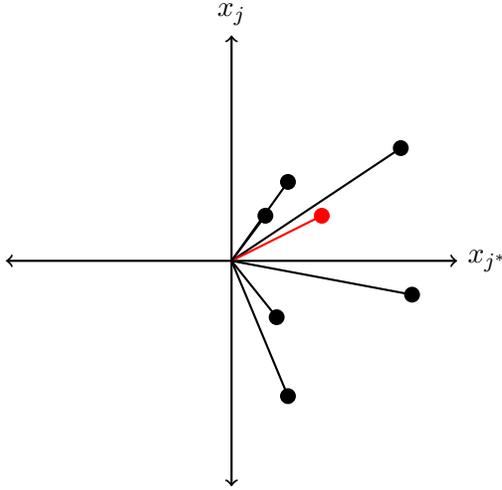


Illustration of Algorithm 1 for estimating an L1-norm best-fit line. Points are projected into the two-dimensional space defined by the  $j$ - and  $j^*$ -axes and the non-negative half-space  $\{x \in \mathbb{R}^m : x_{j^*} \geq 0\}$ . The red point is the one whose angle with the  $j^*$ -axis is the weighted median and is used to calculate the value of  $v_j$ .

get  $\alpha_i = x_{ij}$  for  $i = 1, \dots, n$  in the constraints in (5) and the formulation becomes:

$$\begin{aligned}
 z_j = & \min_{\substack{v \in \mathbb{R}^m, v_j = 1, \\ \lambda^+, \lambda^- \in \mathbb{R}^{m \times n}}} \sum_{i=1}^n \sum_{j=1}^m (\lambda_{ij}^+ + \lambda_{ij}^-), \\
 \text{s.t. } & v_j x_{i\hat{j}} + \lambda_{ij}^+ - \lambda_{ij}^- = x_{ij}, i = 1, \dots, n, j = 1, \dots, m, j \neq \hat{j}, \\
 & \lambda_{ij}^+, \lambda_{ij}^- \geq 0, i = 1, \dots, n, j = 1, \dots, m, j \neq \hat{j},
 \end{aligned} \tag{14}$$

which is an LP. Each of the  $n$  data points generates  $m - 1$  constraints in this LP. An optimal solution defines a vector  $v$  such that the sum of the L1-norm distances of paths from the points to points on the line using the  $m - 1$  directions that preserve the  $\hat{j}$ -th coordinate is minimized.

Each of the  $m$  choices for  $\hat{j}$  defines a different LP; one for each subset of  $m - 1$  unit directions that can be used exclusively to project the data. The uniform feature preservation method is based on solving these  $m$  LPs and selecting the vector  $v$  from the solution associated with the smallest of the  $m$  objective function values (Algorithm 2).

The uniform feature preservation procedure requires finding the solution to  $m$  LPs each with  $n(m - 1)$  constraints. Although solving LPs is efficient, the LPs can be large and solving them directly using an LP solver can be

---

**Algorithm 2** Uniform feature preservation for estimating an L1-norm best-fit one-dimensional subspace [Tsagkarakis et al., 2016, Chierichetti et al., 2017a,b].

---

Given points  $x_i \in \mathbb{R}^m$  for  $i = 1, \dots, n$ .

- 1: Set  $z^* = \infty$ .
  - 2: **for** ( $\hat{j} = 1$ ;  $\hat{j} \leq m$ ;  $\hat{j} = \hat{j} + 1$ ) **do**
  - 3:     Solve the LP in (14) to obtain  $v$ .
  - 4:     **if**  $z_j < z^*$ , **then**
  - 5:         Set  $z^* = z_j$ ,  $v^* = v$ .
- 

computationally demanding and time consuming. Remarkably, it is possible to avoid solving the LPs in (14) by applying Algorithm 1 because for each choice of  $\hat{j}$ , an optimal solution is provided by the calculation of  $m - 1$  weighted medians.

**Proposition 4.** *An optimal solution  $v$  to the LP in (14) can be constructed as follows. If  $x_{i\hat{j}} = 0$  for all  $i$ , then set  $v = 0$ . Otherwise, set  $v_j = 1$  and for each  $j \neq \hat{j}$ , set  $v_j = \frac{x_{i\hat{j}}}{x_{ij}}$ , where  $\frac{x_{i\hat{j}}}{x_{ij}}$  is the weighted median of ratios  $\frac{x_{ij}}{x_{i\hat{j}}}$ ,  $i = 1, \dots, n$  with weights  $|x_{i\hat{j}}|$ ,  $i = 1, \dots, n$ .*

*Proof.* Proof of Proposition 4. If  $x_{i\hat{j}} = 0$  for all points  $i$ , then the solution  $v = 0$  achieves an objective function value of  $\sum_{i=1}^n \sum_{j \neq \hat{j}} |x_{ij}|$  and is therefore optimal.

If  $x_{i\hat{j}} = 0$  for a point  $i$ , then the constraints for point  $i$  in (14) are of the form

$$\lambda_{ij}^+ - \lambda_{ij}^- = x_{ij}, j \neq \hat{j},$$

and the contribution to the objective function value for that point will be a constant value equal to  $\sum_{j \neq \hat{j}} |x_{ij}|$  for any  $v$ , so we may exclude them from consideration in deriving  $v^*$ .

We will assume without loss of generality that  $x_{i\hat{j}} \neq 0$  for all points  $i$ . The LP in (14) can be rewritten as

$$\begin{aligned} z_{\hat{j}} &= \min_{v \in \mathbb{R}^m, v_j=1} \sum_{i=1}^n \sum_{j=1}^m |x_{ij} - v_j x_{i\hat{j}}|, \\ &= \min_{v \in \mathbb{R}^m, v_j=1} \sum_{i=1}^n \sum_{j=1}^m |x_{i\hat{j}}| \left| \frac{x_{ij}}{x_{i\hat{j}}} - v_j \right|. \end{aligned} \tag{15}$$

By Proposition 1, an optimal solution to the problem is to set  $v_j$ , for each  $j$ , to the weighted median of  $\frac{x_{ij}}{x_j}$ ,  $i = 1, \dots, n$  with weights  $|x_{ij}|$ ,  $i = 1, \dots, n$ .  $\square$

Access to a solution to LP (14) provided by Proposition 4 means the instruction in Step 3 in the uniform feature preservation procedure requires only calculating  $m(m - 1)$  weighted medians. A weighted median may be calculated via sorting a list of  $n$  numbers rather than applying an LP solver. Sorting each list is independent and therefore both fixing coordinates and calculating components of  $v$  can be fully parallelized into subprocesses for increased computational efficiency. As has been noted in different contexts, efficiencies can be achieved in the calculation of weighted medians to achieve a linear-time worst-case performance for the weighted median subroutine [Bloomfield and Steiger, 1984, Zemel, 1984, Imai et al., 1989, Gurwitz, 1990].

Proposition 4 establishes an equivalence between the calculation of  $m(m - 1)$  weighted medians of  $n$  ratios in Algorithm 1 [Brooks et al., 2017] and solving the LP in the uniform feature preservation method [Tsagkarakis et al., 2016, Chierichetti et al., 2017a,b]. It therefore follows that the Algorithm 1 procedure is a 2-factor approximation algorithm, as established by Chierichetti et al. for the uniform feature preservation method [Chierichetti et al., 2017a,b].

Song et al. [2016] provide a family of examples for which Algorithms 1 and 2 have an approximation factor of  $\frac{2(m-1)}{m}$  that we will describe here. Consider the following  $m + 1$  points in  $\mathbb{R}^m$ :  $e^1, e^2, e^3, \dots, e^m$ , and  $\mathbb{1}$ , where  $\mathbb{1}$  is the vector of all ones. Algorithms 1 and 2 can return  $v = e^1$  as optimal by projecting on  $v$  preserving coordinate 1 with errors 0, 1, 1,  $\dots$ , 1, and  $m - 1$  for a total of  $2m - 2$ . A globally optimal solution is to set  $v = \mathbb{1}$  and project onto  $v$  preserving any coordinate except  $j$  for each of the first  $m - 1$  points for errors 1, 1, 1,  $\dots$ , 1, and 0 for a total of  $m$ . Note that for  $v = \mathbb{1}$  and projection on to  $v$  preserving coordinate 1 for all points yields errors of  $m - 1, 1, 1, \dots, 1$ , and 0 for a total of  $2m - 2$  so that Algorithms 1 and 2 are unable to distinguish between the suboptimal solution  $v = e^1$  and the optimal solution  $v = \mathbb{1}$ .

## 4 Sufficient Conditions for Finding an L1-Norm Best-Fit One-Dimensional Subspace

We first establish that Algorithm 1 produces optimal solutions for the case when  $m = 2$ . Then we leverage results established in Section 2 to charac-

terize sufficient conditions for the algorithm to provide a globally optimal L1-norm best-fit one-dimensional subspace for general  $m$ .

**Proposition 5.** *Algorithm 1 finds a globally optimal best-fit one-dimensional subspace when  $m = 2$ .*

*Proof.* Proof of Proposition 5. When  $m = 2$ , a best-fit one-dimensional subspace is a hyperplane. Brooks et al. [2013] establish that an L1-norm best-fit  $m - 1$ -dimensional subspace is an L1-norm regression hyperplane where one of the variables serves as the response and there is no intercept. Therefore, one need only check two L1-norm regression hyperplanes, one where the first variable serves as the response, and one where the second serves as the response, and select the hyperplane with the smallest error. An L1-norm regression line with one predictor, one response, and having an intercept of zero can be found by finding the weighted median of the ratios of the responses to the predictors with the absolute value of the predictors as the weights - this fact dates back hundreds of years to Boscovich, Simpson, and Laplace (see Bloomfield and Steiger [1984], Farebrother [1990]). Algorithm 1 calculates the two regression hyperplanes via weighted medians and selects the one with the lowest error, thereby providing an optimal solution.  $\square$

We now provide a characterization of sets of points in  $\mathbb{R}^m$  for which Algorithm 1 derives an optimal solution to (1). If all points in a data cloud have the same dominating coordinate, then there is an L1-norm best-fit one-dimensional subspace that has a dominating coordinate. By Theorem 1, all points project onto that line preserving the same coordinate. Because Algorithm 1 assumes at each iteration that points preserve the same coordinate, it will find an optimal line. The proof of the following theorem uses Theorem 1, Theorem 2, Proposition 3, and Proposition 4 directly.

**Theorem 4.** *For points  $x_i \in \mathbb{R}^m$ ,  $i = 1, \dots, n$ , if there is a single dominating coordinate for each point  $i$ :*

$$|x_{ij'}| > \sum_{j \neq j'} |x_{ij}|,$$

*then the Algorithm 1 provides an L1-norm best-fit one-dimensional subspace.*

*Proof.* Proof of Theorem 4. Taking the reflection of any point  $x_i$  about the origin does not affect the set of optimal solutions to (1): replace  $x_i$  with  $-x_i$  and replace  $\alpha_i$  with  $-\alpha_i$  and the objective function value remains unchanged. Take the reflections of points  $x_i$  about the origin so that  $x_{ij'} > 0$  for each point  $i = 1, \dots, n$ .

Consider the terms in the objective function for (1),  $|x_{ij'} - \alpha_i v_{j'}|$ ,  $i = 1, \dots, n$ . By Theorem 2, the two terms in the absolute value must have the same sign. Because  $x_{ij'} > 0$  for all  $i$ , the multipliers  $\alpha_i$  must have the same sign as  $v_{j'}$ . Having them all positive or all negative results in the same objective function value. We will assume they are all positive without loss of generality.

Suppose  $v$  defines an L1-norm best-fit one-dimensional subspace. Consider a new point  $x_\ell$  with a dominating coordinate  $x_{\ell j'} > 0$  and with  $\text{sgn}(x_{\ell j}) \neq \text{sgn}(v_j)$  for  $j \neq j'$ . By Proposition 3, the optimal multiplier  $\alpha_\ell$  for the projection of  $x_{\ell j}$  onto the line defined by  $v$  is a weighted median of non-positive numbers  $\frac{x_{\ell j}}{v_j}$ , for  $j \neq j'$  and  $v_j \neq 0$ , and one positive number  $\frac{x_{\ell j'}}{v_{j'}}$ , with weights  $\{|v_j| : v_j \neq 0\}$ .

By Theorem 2,  $\alpha_\ell > 0$ , and so  $\frac{x_{\ell j'}}{v_{j'}}$  is the unique weighted median. Therefore,  $\sum_{j \neq j'} |v_j| \leq 1/2 \sum_{j=1}^m |v_j|$ , so that  $|v_{j'}| \geq \sum_{j \neq j'} |v_j|$ . To see that the inequality is strict, consider the second-largest ratio  $\frac{x_{ij}}{v_j}$ . Because it is not a weighted median, it must be that  $|v_{j'}| > 1/2 \sum_{j=1}^m |v_j|$  which means that  $|v_{j'}| > \sum_{j \neq j'} |v_j|$ .

Therefore,  $v_{j'}$  is a dominating coordinate. By Theorem 1, all points project onto such an optimal  $v$  preserving coordinate  $j'$  and using the  $m - 1$  unit directions  $j \neq j'$ . By Proposition 4, Algorithm 1 will find such a  $v$  when  $\hat{j} = j'$ .  $\square$

## 5 Conclusions

Problems arising from outliers in data when using methods based on minimizing the sum-of-squared errors are acknowledged as a serious limitation to these methods. The response has been an increase in interest in outlier-insensitive methods based on the L1 norm. Using the L1 norm allows outliers to be present in the data without this unduly affecting the final results.

Many analytics methods, including linear regression, logistic regression, and traditional PCA, require fitting subspaces to data to extract information about properties such as location, dispersion, and orientation. Except in the case of a hyperplane or a point, finding an L1-norm best-fit subspace for a point set in  $m$ -dimensions is a nonlinear, nonconvex, nonsmooth optimization problem and has been shown to be NP-hard.

This work treats the L1-norm best-fit subspace when this is a line. It establishes the equivalence between a class of algorithms for estimating the line based on preserving the same dimension in all the points' projections

and another algorithm based on calculating weighted medians. The former approach referred to as “uniform feature preservation” has been proposed independently by several authors. It formulates and solves large LPs, one for each dimension, and the resultant estimate is guaranteed to generate a total error bounded to within a factor of two of the optimum. The equivalent algorithm requires calculating weighted medians. This reflects the close relation between LP and weighted medians. The equivalence makes these algorithms more accessible and practical especially when dealing with large data sets.

This paper also establishes that uniform preservation algorithms can generate optimal L1-norm best-fit lines. We present sufficient conditions on the data for this to occur. The conditions are determined by deriving fundamental properties of projections of points on to lines using the L1 norm.

The L1 norm plays a prominent role in the design and implementation of procedures based on minimizing errors. The case when the objective is to fit hyperplanes, as in regression, using the L1 norm has been studied for a long time and is well-understood: finding the L1-norm best-fit  $m - 1$ -dimensional subspace is tractable. Another interesting subspace for fitting data is a line; this problem is intractable under the L1 norm. Uniform feature preservation algorithms for estimating L1-norm best-fit lines with their bounds on worst performance and efficient solution using weighted medians places at our disposal a valuable tool to a multitude of applications that rely on subspace estimation as a subproblem.

## **Acknowledgments.**

The first author was supported in part by National Institute of Justice Award 2016-DN-BX-0151, National Institutes of Health Award 8U54HD080784, PPB Grant 15011 from the Global Alliance to Prevent Prematurity and Stillbirth, and the National Science Foundation Award DMS-1127914 to the Statistical and Applied Mathematical Sciences Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## References

- G. Appa and C. Smith. On  $L_1$  and Chebyshev Estimation. *Mathematical Programming*, 5(1):73–87, 1973. ISSN 00255610. doi: 10.1007/BF01580112. URL <http://www.springerlink.com/index/XT54Q26436688716.pdf>.
- F. Ban, V. Bhattiprolu, K. Bringmann, P. Koley, E. Lee, and D.P. Woodruff. A PTAS for  $\ell_p$ -Low Rank Approximation. 2018. URL <http://arxiv.org/abs/1807.06101>.
- R. Blanquero, E. Carrizosa, A. Schöbel, and D. Scholz. A Global Optimization Procedure for the Location of a Median Line in the Three-Dimensional Space. *European Journal of Operational Research*, 215(1): 14–20, 2011.
- Rafael Blanquero, Emilio Carrizosa, and Pierre Hansen. Locating Objects in the Plane Using Global Optimization Techniques. *Mathematics of Operations Research*, 34(4):837–858, 2009. ISSN 0364-765X. doi: 10.1287/moor.1090.0406. URL <http://dx.doi.org/10.1287/moor.1090.0406>.
- P. Bloomfield and W.L. Steiger. *Least Absolute Deviations*. Birkhäuser Boston, Boston, MA, 1984. ISBN 978-1-4684-8576-9. doi: 10.1007/978-1-4684-8574-5. URL <http://link.springer.com/10.1007/978-1-4684-8574-5>.
- V. Boltjanski, H. Martini, and V. Soltan. *Geometric Methods and Optimization Problems*. Springer Science and Business Media, Dordrecht, 1999.
- J. Brimberg, H. Juel, and A. Schöbel. Linear Facility Location in Three Dimensions-Models and Solution Methods. *Operations Research*, 50(6): 1050–1057, 2002.
- J. Brimberg, H. Juel, and A. Schöbel. Properties of Three-Dimensional Median Line Location Models. *Annals of Operations Research*, 122(1): 71–85, 2003.
- J.P. Brooks and J.H. Dulá. The  $L_1$ -Norm Best-Fit Hyperplane Problem. *Applied Mathematics Letters*, 26:51–55, 2013.
- J.P. Brooks and J.H. Dulá. Robust PCA via  $L_1$ -Norm Line Fitting. *Winner, 2016 INFORMS Data Mining Best Paper Award*, <http://connect.informs.org/data-mining/awards>, 2016.

- J.P. Brooks, J.H. Dulá, and E.L. Boone. A Pure L1-norm Principal Component Analysis. *Computational Statistics & Data Analysis*, 61:83–98, 2013.
- J.P. Brooks, J.H. Dulá, A.L. Pakyz, and R.E. Polk. Identifying Hospital Antimicrobial Resistance Targets via Robust Ranking. *IISE Transactions on Healthcare Systems Engineering*, 5579(August), 2017. doi: 10.1080/24725579.2017.1339148.
- E.J. Candès, X. Li, Y. Ma, and J. Wright. Robust Principal Component Analysis? *Journal of the ACM*, 58(3):11:1–11:37, 2011.
- A. Charnes, W.W. Cooper, and R.O. Ferguson. Optimal Estimation of Executive Compensation. *Management Science*, 1(2):138–151, 1955.
- F. Chierichetti, S. Gollapudi, R. Kumar, S. Lattanzi, R. Panigrahy, and D.P. Woodruff. Algorithms for  $\ell_p$  Low Rank Approximation. *ArXiv*, 2017a. URL <http://arxiv.org/abs/1705.06730>.
- F. Chierichetti, S. Gollapudi, R. Kumar, S. Lattanzi, R. Panigrahy, and D.P. Woodruff. Algorithms for  $\ell_p$  Low-Rank Approximation. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 806–814, Sydney, Australia, 2017b. URL <http://proceedings.mlr.press/v70/chierichetti17a.html>.
- Y. Dahiya, D. Konomis, and D.P. Woodruff. An Empirical Evaluation of Sketching for Numerical Linear Algebra. In *2018 ACM SIGKDD International Conference on Knowledge Discovery in Databases*, London, 2018. ISBN 9781450355520.
- J.M. Díaz-Báñez, J. A. Mesa, and A. Schöbel. Continuous Location of Dimensional Structures. *European Journal of Operational Research*, 152(1):22–44, 2004. ISSN 03772217. doi: 10.1016/S0377-2217(02)00647-1.
- R.W. Farebrother. Studies in the History of Probability and Statistics XLII. Further Details of Contacts Between Boscovich and Simpson in June 1760. *Biometrika*, 77(2):397–400, 1990.
- N. Gillis and S.A. Vavasis. On the Complexity of Robust PCA and  $\ell_1$ -Norm Low-Rank Matrix Approximation. *Mathematics of Operations Research*, Articles i:1–13, 2018.

- D. Goldfarb, S. Ma, and K. Scheinberg. Fast Alternating Linearization Methods for Minimizing the Sum of Two Convex Functions. *Mathematical Programming*, 141(1):349–382, 2013.
- C. Gurwitz. Weighted Median Algorithms for  $L_1$  Approximation. *BIT*, 30(2):301–310, 1990.
- H.W. Hamacher and S. Nickel. Classification of Location Models. *Location Science*, 6(1-4):229–242, 1998. ISSN 09668349. doi: 10.1016/S0966-8349(98)00053-9.
- H. Imai, K. Kato, and P. Yamamoto. A Linear-Time Algorithm for Linear  $L_1$  Approximation of Points. *Algorithmica*, 4(1):77–96, 1989. ISSN 14320541. doi: 10.1007/BF01553880.
- I.T. Jolliffe. *Principal Component Analysis*. Springer, 2nd edition, 2002.
- Q. Ke and T. Kanade. Robust Subspace Computation using L1 Norm. Technical report, Carnegie Mellon University, 2003.
- Q. Ke and T. Kanade. Robust L1 Norm Factorization in the Presence of Outliers and Missing Data by Alternative Convex Programming. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- S. Kundu, P.P. Markopoulos, and D.A. Pados. Fast Computation of the  $L_1$ -Principal Component of Real-Valued Data. *IEEE International Conference on Acoustic, Speech, and Signal Processing*, pages 8028–8032, 2014.
- N. Kwak. Principal Component Analysis based on L1-Norm Maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(9):1672–1680, 2008.
- A. Kyrillidis. Simple and Practical Algorithms for  $\ell_p$ -Norm Low-Rank Approximation. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, page Article ID 167, Monterey, CA, 2018. Association for Uncertainty in Artificial Intelligence. URL <http://arxiv.org/abs/1805.09464>.
- G. Lerman and T. Maunu. An Overview of Robust Subspace Recovery. 2018.
- S. Ma and N.S. Aybat. Efficient Optimization Algorithms for Robust Principal Component Analysis and Its Variants. 2018. URL <http://arxiv.org/abs/1806.03430>.

- P.P. Markopoulos, K.N. Karystinos, and D.A. Pados. Optimal Algorithms for  $L_1$ -Subspace Signal Processing. *IEEE Transactions on Signal Processing*, 62(19):5046–5058, 2014.
- P.P. Markopoulos, S. Kundu, S. Chamadia, and D.A. Pados. L1-Norm Principal-Component Analysis via Bit Flipping. In *Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016*, pages 326–332, Anaheim, CA, 2016. ISBN 9781509061662. doi: 10.1109/ICMLA.2016.120.
- H. Martini and A. Schöbel. Median Hyperplanes in Normed Spaces - A Survey. *Discrete Applied Mathematics*, 89(1-3):181–195, 1998.
- J.G. Morris and J.P. Norback. Linear Facility Location - Solving Extensions to the Basic Problem. *European Journal of Operational Research*, 12(1): 1–22, 1983. ISSN  $\text{null}$ . doi: 10.1016/j.pcad.2015.11.006. URL <https://gwu.illiad.oclc.org/illiad/pdf/277299.pdf%5Cnpapers2://publication/uuid/FA3E868D-2CDC-4D5B-B2F8-DDC06AE176E7>.
- F. Nie, H. Huang, C. Ding, D. Luo, and H. Wang. Robust Principal Component Analysis with Non-Greedy  $\ell_1$ -Norm Maximization. *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, pages 1433–1438, 2011.
- R.A. Reris and J.P. Brooks. *Principal Component Analysis and Optimization: A Tutorial*, pages 212–225. INFORMS, 2015.
- K. Sabo, R. Scitovski, and I. Vazler. Searching for a Best LAD-Solution of an Overdetermined System of Linear Equations Motivated by Searching for a Best LAD-Hyperplane on the Basis of Given Data. *Journal of Optimization Theory and Applications*, 149(2):293–314, 2011.
- A. Schöbel. Locating Least-Distant Lines in the Plane. *European Journal of Operational Research*, 7(97):152–159, 1998. ISSN 03772217.
- A. Schöbel. *Locating Lines and Hyperplanes: Theory and Algorithms*. Kluwer Academic Publishers, Dordrecht, 1999.
- A. Schöbel. Location of Dimensional Facilities in a Continuous Space. In G Laporte, S Nickel, and F Saldanha da Gama, editors, *Location Science*, pages 135–175. Springer International Publishing, Cham, 2015. ISBN 978-3-319-13110-8. doi: 10.1007/978-3-319-13111-5\_7. URL [http://link.springer.com/10.1007/978-3-319-13111-5\\_7](http://link.springer.com/10.1007/978-3-319-13111-5_7).

- O.B. Sheynin. R.J. Boscovich's Work on Probability. *Archive for History of Exact Sciences*, 9(4):306–324, 1973.
- Z. Song, D.P. Woodruff, and P. Zhong. Low Rank Approximation with Entrywise  $\ell_1$ -Norm Error. *ArXiv*, 2016. URL <http://arxiv.org/abs/1611.00898>.
- Z. Song, D.P. Woodruff, and P. Zhong. Low Rank Approximation with Entrywise  $\ell_1$ -Norm Error. In *STOC 2017 Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 688–701, Montreal, 2017.
- N. Tsagkarakis, P.P. Markopoulos, and D.A. Pados. On the  $L_1$ -norm Approximation of a Matrix by Another of Lower Rank. *15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 768–773, 2016.
- N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy. Robust Subspace Learning: Robust PCA, Robust Subspace Tracking, and Robust Subspace Recovery. *IEEE Signal Processing Magazine*, 35(July):32–55, 2018. ISSN 10535888. doi: 10.1109/MSP.2018.2826566.
- E. Zemel. An  $O(n)$  Algorithm for the Linear Multiple Choice Knapsack Problem and Related Problems. *Information Processing Letters*, 18(3): 123–128, 1984.