Hindawi Publishing Corporation
Advances in Operations Research
Volume 2011, Article ID 263762, 18 pages
doi:10.1155/2011/263762

Research Article

Outlier-Resistant L_1 Orthogonal Regression via the Reformulation-Linearization Technique

J. Paul Brooks and Edward L. Boone

Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, P.O. Box 843083, 1015 Floyd Avenue, Richmond, VA 23284, USA

Correspondence should be addressed to J. Paul Brooks, jpbrooks@vcu.edu

Received 9 September 2010; Revised 7 January 2011; Accepted 14 January 2011

Academic Editor: I. L. Averbakh

Assessing the linear relationship between a set of continuous predictors and a continuous response is a well-studied problem in statistics and data mining. L_2-based methods such as ordinary least squares and orthogonal regression can be used to determine this relationship. However, both of these methods become impaired when influential values are present. This problem becomes compounded when outliers confound standard diagnostics. This work proposes an L_1-norm orthogonal regression method (L_1OR) formulated as a nonconvex optimization problem. Solution strategies for finding globally optimal solutions are presented. Simulation studies are conducted to assess the resistance of the method to outliers and the consistency of the method. The method is also applied to real-world data arising from an environmental science application.

1. Introduction and Background

Data analysts are often posed with the problem of determining the relationship between several variables and a response variable. The standard technique when all variables are defined on a continuous domain is ordinary least squares regression (OLS). When *outliers*, or unusual observations, are present in data, traditional regression techniques become impaired. Methods such as M-regression (M-R) use M estimates to reduce the impact of outliers. These methods are not designed for developing *errors-in-variables* models in which both the predictors and the response have measurement error or are considered random components. An example of such a situation is studying the relationship between pH and alkalinity in freshwater habitats, where both measurements are subject to error.

Orthogonal regression (L_2OR) is used when uncertainty is known to be present in both independent and dependent variables. This assumption is in contrast to OLS, where the predictors are assumed to be known with no measurement error. Furthermore, orthogonal regression measures the distances orthogonal to the fitted hyperplane whereas in OLS residuals are measured as the vertical distance of observations to the fitted surface.

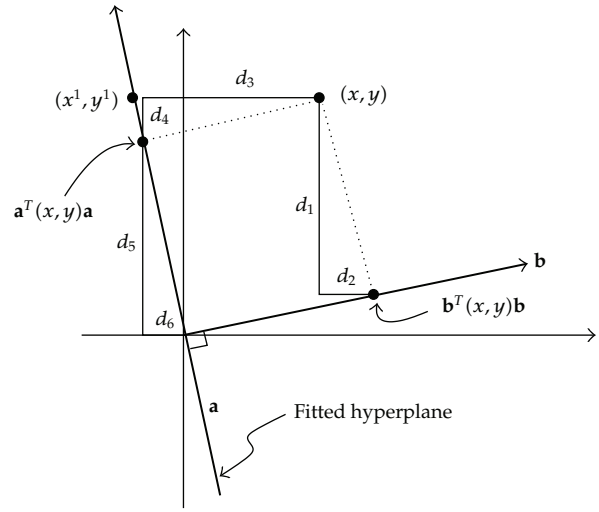

Figure 1: Two-dimensional illustration of different methods for incorporating the L_1 norm into orthogonal regression. In traditional orthogonal regression, the sum of distances of points (x, y) to $\mathbf{b}^T(x, y)\mathbf{b}$ is maximized, the sum of distances of points (x, y) to $\mathbf{a}^T(x, y)\mathbf{a}$ is minimized, and the sum of the magnitudes of $\mathbf{a}^T(x, y)\mathbf{a}$ is maximized. As noted in the text, each of these distance measures can be modified to incorporate the L_1 norm to derive different results. In this paper, the approach is to maximize the sum of L_1 distances of points (x, y) to $\mathbf{b}^T(x, y)\mathbf{b}$ which is illustrated by $d_1 + d_2$.

1.1. Previous Work on Robust Orthogonal Regression

The sensitivity of L_2OR to outliers has been noted, and other investigators have worked to develop robust methods [1–3]. The work of Zamar [3] includes the use of S and M estimates for orthogonal regression. Späth and Watson [4] introduce a method for incorporating the L_1 norm for measuring distances in orthogonal regression.

L_2OR can be formulated as equivalent to finding the last principal component, or the direction of minimum variation, in principal component analysis (PCA). Hence, any robust PCA method can be used for robust orthogonal regression. Two main approaches for robust PCA are (1) to find robust estimates of the covariance matrix (in traditional PCA, the principal components are eigenvectors of the covariance matrix) and (2) to use a robust measure of dispersion. Research in the former area includes [5–11]. Robust estimates of dispersion in PCA have been investigated in [12–16]; each of these works is based on a *projection pursuit* approach.

Our approach is closely related to that developed by Späth and Watson [4] and Kwak [16] by the manner in which we incorporate the L_1 norm into an orthogonal regression procedure. Späth and Watson [4] measure the error of an observation as the L_1 distance to its orthogonal projection onto a fitted hyperplane. Kwak [16] finds successive directions of maximum variation by maximizing the L_1 distance to the L_2 projection of points onto a line. In contrast to these methods, our approach is to directly find the direction of least variation by maximizing the L_1 distance between points and their L_2 projections onto a vector (see Figure 1). Also, the methods presented in [4, 16] guarantee only local minima to the respective optimization problems, while we present a method for deriving globally optimal solutions.

These three methods can be viewed as approximating a maximum likelihood estimator (MLE) for the linear errors-in-variables model with independent errors with a Laplace distribution (see [23, 24]). The MLE for such a model corresponds to a hyperplane that

minimizes the sum of L_1 projections. Zwanzig [25] considers an L_1 estimator for a nonlinear generalization of the error-in-variables model and shows that under certain assumptions on the error distribution, the estimator is consistent. When applied to the setting of L_1 orthogonal linear regression, the estimator is similar to the approach of Späth and Watson [4].

1.2. Traditional Orthogonal Regression

Suppose we are given observations with continuous predictors and responses $(\mathbf{x}_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$, $i = 1, \ldots, n$. L_2OR seeks to find an orthogonal projection of the data onto a hyperplane such that the sum of the orthogonal distances of the points $(\mathbf{x}_i, y_i)$ to the hyperplane is minimized. We assume throughout this work that the medians have been subtracted from samples and that the fitted hyperplane passes through the origin. We note that for large values of d, the coordinate-wise median may not be a good estimate of the center of a data cloud (see [26]).

In L_2OR, the sum of squared orthogonal distances of $(\mathbf{x}_i, y_i)$ to the hyperplane defined by $\mathbf{b}^T(\mathbf{x}, y) = 0$ is minimized. The vector $\mathbf{b}$ is normal to the best-fitting hyperplane and is the direction of least variation of the data. Because $\mathbf{b}$ is the direction of least variation, the sum of squared distances of observations to their projections along $\mathbf{b}$ is maximized. Therefore, we can find $\mathbf{b}$ by solving the following optimization problem:

$$[L_2\text{OR}] \max_{\mathbf{b}} \sum_{i=1}^{n} \left\| (\mathbf{x}_i, y_i) - \mathbf{b}^T(\mathbf{x}_i, y_i)\mathbf{b} \right\|_2^2, \tag{1.1}$$

subject to

$$\mathbf{b}^T\mathbf{b} = 1. \tag{1.2}$$

The variables are in the vector $\mathbf{b} \in \mathbb{R}^{d+1}$. The term $\mathbf{b}^T(\mathbf{x}_i, y_i)\mathbf{b}$ represents the orthogonal projection of observation i along $\mathbf{b}$ in terms of the original coordinates of the data.

In this paper, we present a new outlier-resistant method for orthogonal regression called L_1OR. The direction of least variation in data is found by maximizing the L_1 distances of observations to their projection points along a vector. The fitted hyperplane is orthogonal to the direction of least variation. The problem is formulated as a nonconvex optimization problem. We describe how globally optimal solutions can be derived based on a reformulation-linearization technique (RLT) developed by Sherali and Tuncbilek [27]. We present results of applying L_1OR to simulated data that is contaminated with outliers and compare the results to robust methods for orthogonal regression. The consistency of L_1OR is assessed using simulated data. L_1OR is applied to data collected for the evaluation of marine habitats, where uncertainty resides in both the dependent and independent variables.

2. Finding the Best-Fit Hyperplane

Suppose that instead of maximizing the sum of the squared perpendicular distances of observations to their projection along the direction of least variation, we maximize the sum of the L_1 distances. Using the L_1 metric reduces the impact of outlier observations.

In Figure 1, we illustrate different methods of incorporating the L_1 norm into an orthogonal regression procedure for a two-dimensional example. The fitted hyperplane is

defined by its normal vector $\mathbf{b}$, representing an approximation of the direction of least variation in data. The vector $\mathbf{a}$ spans the space defined by the fitted hyperplane. Our approach is to maximize the sum of L_1 distances of points onto their projections on $\mathbf{b}$. The L_1 distance of (x, y) to its L_2 projection on $\mathbf{b}$ is given by $d_1 + d_2$ in the figure. The procedure proposed by Späth and Watson [4] minimizes the sum of L_1 distances of points to their L_2 projections in a fitted hyperplane. The distance of (x, y) to its projection in the fitted subspace is indicated by $d_3 + d_4$. The procedure introduced by Kwak [16] maximizes the sum of L_1 magnitudes of the projections of points onto the fitted hyperplane. In Figure 1, this magnitude is given by $d_5 + d_6$. When these three distances are measured using the L_2 norm, the same regression plane is optimal [28]; however, because the distances in each case are measured using the L_1 norm, the resulting regression planes will not always coincide. The L_1 projection of (x, y) on to the fitted hyperplane is given by (x^1, y^1); an MLE approach would minimize the sum of L_1 distances of points to their L_1 projections.

Maximizing the sum of the L_1 distances of points to a line passing through the origin is written as

$$\begin{aligned}\max_{\mathbf{b}} \sum_{i=1}^{n} \left\| (\mathbf{x}_i, y_i) - \mathbf{b}^T (\mathbf{x}_i, y_i)\mathbf{b} \right\|_1 = \max \sum_{i=1}^{n} \sum_{j=1}^{d} \left| x_{ij} - b_j \left(\sum_{k=1}^{d} x_{ik} b_k + y_i b_{d+1} \right) \right| \\ + \sum_{i=1}^{n} \left| y_i - b_{d+1} \left(\sum_{k=1}^{d} x_{ik} b_k + y_i b_{d+1} \right) \right|.\end{aligned} \tag{2.1}$$

The objective function is nonlinear and nonconvex. As with [L_2OR], the optimal hyperplane is defined by $\mathbf{b}^T(\mathbf{x}, y) = 0$. Let r_{ij} be the L_1 residual for component j of observation i. Also, let $\mathbf{a} = \mathbf{b} + \mathbf{1}$, where $\mathbf{1}$ is a vector of 1's, so that all a_j variables are nonnegative. This substitution is necessary for our solution method which is explained below. The math program can then be formulated as

$$[L_1\text{OR}] \max \sum_{i=1}^{n} \sum_{j=1}^{d+1} r_{ij}, \tag{2.2}$$

subject to

$$\begin{gathered} r_{ij} = \begin{cases} \left[(\mathbf{x}_i, y_i) - (\mathbf{a} - \mathbf{1})^T (\mathbf{x}_i, y_i)(\mathbf{a} - \mathbf{1}) \right]_j & \text{if } z_{ij} = 0, \ \forall i, j, \\ \left[-(\mathbf{x}_i, y_i) + (\mathbf{a} - \mathbf{1})^T (\mathbf{x}_i, y_i)(\mathbf{a} - \mathbf{1}) \right]_j & \text{if } z_{ij} = 1, \ \forall i, j, \end{cases} \\ (\mathbf{a} - \mathbf{1})^T (\mathbf{a} - \mathbf{1}) = 1, \\ \mathbf{a} \geq \mathbf{0}, \\ \mathbf{a} \leq \mathbf{2}, \end{gathered} \tag{2.3}$$

$$z_{ij} \in \{0, 1\}, \quad i = 1, \ldots, n; \ j = 1, \ldots, d + 1. \tag{2.4}$$

The quantities $\mathbf{0}$, $\mathbf{1}$, and $\mathbf{2}$ are vectors with each coordinate having the value 0, 1, and 2, respectively. The objective function is now linear, and the first three sets of constraints are defined by nonconvex functions.

To derive globally optimal solutions for $[L_1\text{OR}]$, we combine the use of branch-and-bound for integer programming with branch-and-bound for the reformulation-linearization technique (RLT) as described in [27]. *Subproblem* will refer to a linear mixed-integer program (MIP) that corresponds to a node in a branch-and-bound tree for the RLT. Each subproblem can be converted to a linear MIP by expressing the conditional constraints as

$$\begin{aligned} r_{ij} &\le \left[(\mathbf{x}_i, y_i) - (\mathbf{a}-\mathbf{1})^T(\mathbf{x}_i, y_i)(\mathbf{a}-\mathbf{1})\right]_j + Mz_{ij}, \\ r_{ij} &\le \left[-(\mathbf{x}_i, y_i) + (\mathbf{a}-\mathbf{1})^T(\mathbf{x}_i, y_i)(\mathbf{a}-\mathbf{1})\right]_j + Mz_{ij}, \end{aligned} \tag{2.5}$$

for a sufficiently large constant M.

The following is a summary of RLT applied to $[L_1\text{OR}]$.

(i) *Subproblem optimization.* Select a subproblem to solve. Each subproblem is a linear MIP that relaxes the nonconvex constraints. If all subproblems are solved, then the incumbent solution is optimal.

(ii) *Check for new bound.* If the solution satisfies the original nonconvex constraints, the current solution is feasible. Update the incumbent solution and objective value if appropriate.

(iii) *Fathom.* Fathom if (1) the solution satisfies the original constraints, (2) the subproblem is infeasible, or (3) the objective value for the subproblem is less than the incumbent objective value.

(iv) *Branch.* Select a variable for branching, creating two subproblems.

A flow-chart detailing the steps in the RLT branch-and-bound process is included in Figure 2.

We now describe the construction of the root subproblem for RLT. For each occurrence of $a_j a_k$ in the constraints, substitute a new variable A_{jk} into the formulation. Also, add constraints of the form

$$\begin{aligned} &(2-a_j)(2-a_k) \ge 0, \quad j,k = 1,\ldots,d+1, \\ &(a_k-0)(2-a_j) \ge 0, \quad j,k = 1,\ldots,d+1, \\ &(a_k-0)(a_j-0) \ge 0, \quad j,k = 1,\ldots,d+1, \end{aligned} \tag{2.6}$$

but again replace occurrences of $a_j a_k$ with A_{jk}. The presence of 0 in the constraints is to reflect the lower bounds on the a_j variables; these lower bounds will be changed during the optimization algorithm as described below. The result is a linear MIP that is a relaxation of $[L_1\text{OR}]$ [27].

We now describe the branching procedure. The optimal solution to the relaxation is feasible for $[L_1\text{OR}]$ if $A_{jk} = a_j a_k$ for all j,k. If this condition is not satisfied, then choose a

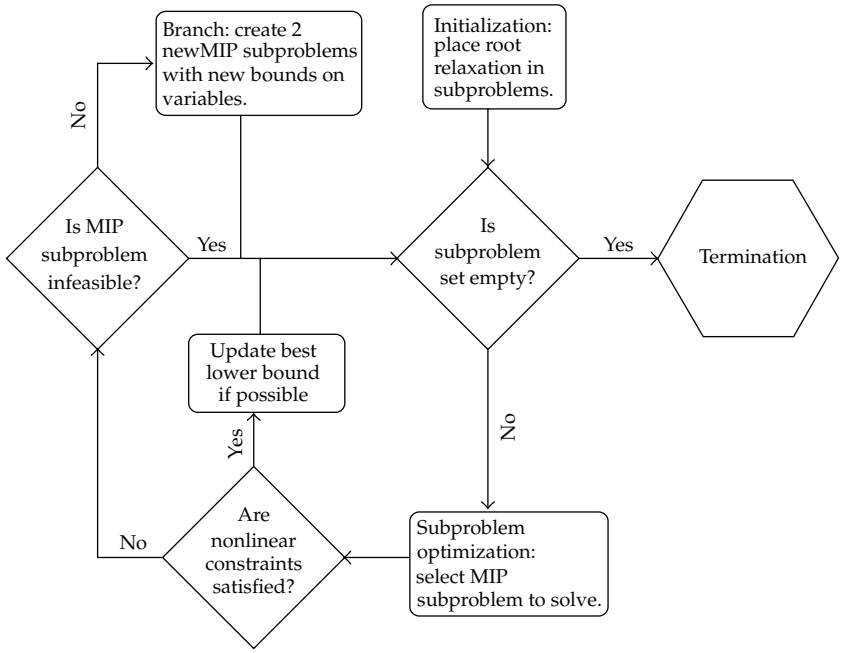

Figure 2: Flowchart of the steps involved in RLT branch and bound procedure applied to a nonlinear mixed-integer program.

variable a_j with $A_{jk} \neq a_j a_k$ for some k with current value $\overline{a}_j$ and create two new subproblems. One of the new subproblems will have constraints of the form

$$\begin{aligned} (\overline{a}_j - a_j)(2 - a_k) &\geq 0, \quad k = 1, \ldots, d+1, \\ (a_k - 0)(\overline{a}_j - a_j) &\geq 0, \quad k = 1, \ldots, d+1, \\ (a_k - 0)(a_j - 0) &\geq 0, \quad k = 1, \ldots, d+1, \\ a_j &\leq \overline{a}_j. \end{aligned} \tag{2.7}$$

Again, replace all occurrences of $a_j a_k$ with A_{jk} to create linear constraints. The other new subproblem will have the linearized form of the constraints

$$\begin{aligned} (2 - a_j)(2 - a_k) &\geq 0, \quad k = 1, \ldots, d+1, \\ (a_j - \overline{a}_j)(2 - a_k) &\geq 0, \quad k = 1, \ldots, d+1, \\ (a_k - 0)(a_j - \overline{a}_j) &\geq 0, \quad k = 1, \ldots, d+1, \\ a_j &\geq \overline{a}_j. \end{aligned} \tag{2.8}$$

As nodes in the branch-and-bound tree are traversed, the bounds for the a_j variables are successively tightened. Sherali and Tuncbilek [27] prove that either the search for optimal solutions terminates with a globally optimal solution in finite steps or else any accumulation point of solutions along an infinite branch of the branch-and-bound tree is a globally optimal solution.

3. Simulation Studies

In this section, the ability of L_1OR to resist the effects of two types of outliers is assessed using simulation studies. The approach is compared to L_2OR and several robust procedures. The consistency of L_1OR is also assessed using a simulation study.

[L_1OR] MIP subproblems are solved using CPLEX 12.1. If provable optimality is not achieved for MIP subproblems after 2 minutes, the best-known integer feasible solution is used. We implemented our own branch-and-bound algorithm for applying RLT in a C program, with a time limit of 7200 CPU seconds for each instance. Problems are solved on machines with 2 × 2.6 GHz Opteron processors and 2 GB RAM.

L_1OR is compared to a robust approach based on projection pursuit [12], a τ scale-based orthogonalized Gnanadesikan-Kettenring estimate [29] (hereafter τ-OGK), and a method based on PCA-L_1 [16]. The projection pursuit approach is applied by using the method for principal component analysis described in [15]. The method is modified for orthogonal regression by taking the last robust principal component as the coefficients of the orthogonal regression hyperplane. We denote this method by ppOR-mad or ppOR-qn, with the suffix indicating the scale function used. The other methods are denoted by τ-OGK and PCA-L_1. For PCA-L_1, the initial vector is set to $\mathbf{w}_0 = \arg\max_{\mathbf{x}_i} \|\mathbf{x}_i\|_2$ (see [16]).

L_2OR and ppOR models are derived using *prcomp()* and *PCAgrid()* functions, respectively, called in the R environment for statistical computing [30]. The function *PCAgrid()* is in the *pcaPP* [31] library. R code for the τ-OGK estimator was provided by an anonymous referee. We implemented the PCA-L_1 method [16] in a C program.

3.1. Vertical Outliers

A simulation study is conducted to assess the ability of L_1OR to detect linear relationships in bivariate data in the presence of *vertical outliers*. Vertical outliers have significant variation only in their response-variable values. A simulation design is utilized by varying the number of contaminated observations (C) and contamination magnitude (m). Each method is run on 30 datasets with 100 observations under each treatment condition. For this study, C is varied in the following manner: no contamination, $C = 0$, moderate contamination, $C = 10$, and high contamination, $C = 25$. The magnitude of contamination m is varied as $m = 1$: low contamination, $m = 10$: moderate magnitude, $m = 50$: large magnitude.

The data are sampled in the following manner.

(i) Generate the uncontaminated data: $x_i \sim U[-1, 1]$ and $y_i = x_i + \epsilon_i$, where $\epsilon_i \sim N(0, 0.1)$, for $i = 1, \ldots, 100 - C$.

(ii) Generate the contaminated data: $x_i \sim U[0.5, 1]$ and $y_i \sim |N(0, m \times 0.1)|$, for $i = 101 - C, \ldots, 100$.

An example dataset with fitted models generated using $m = 10$ and $C = 25$ is given in Figure 3(a).

To evaluate each method's ability to accurately fit the known underlying model, the following model discrepancy, D, is used:

$$D\left(\hat{f}, f\right) = \int_{-1}^{1} \left|\hat{f}(x) - f(x)\right| dx, \tag{3.1}$$

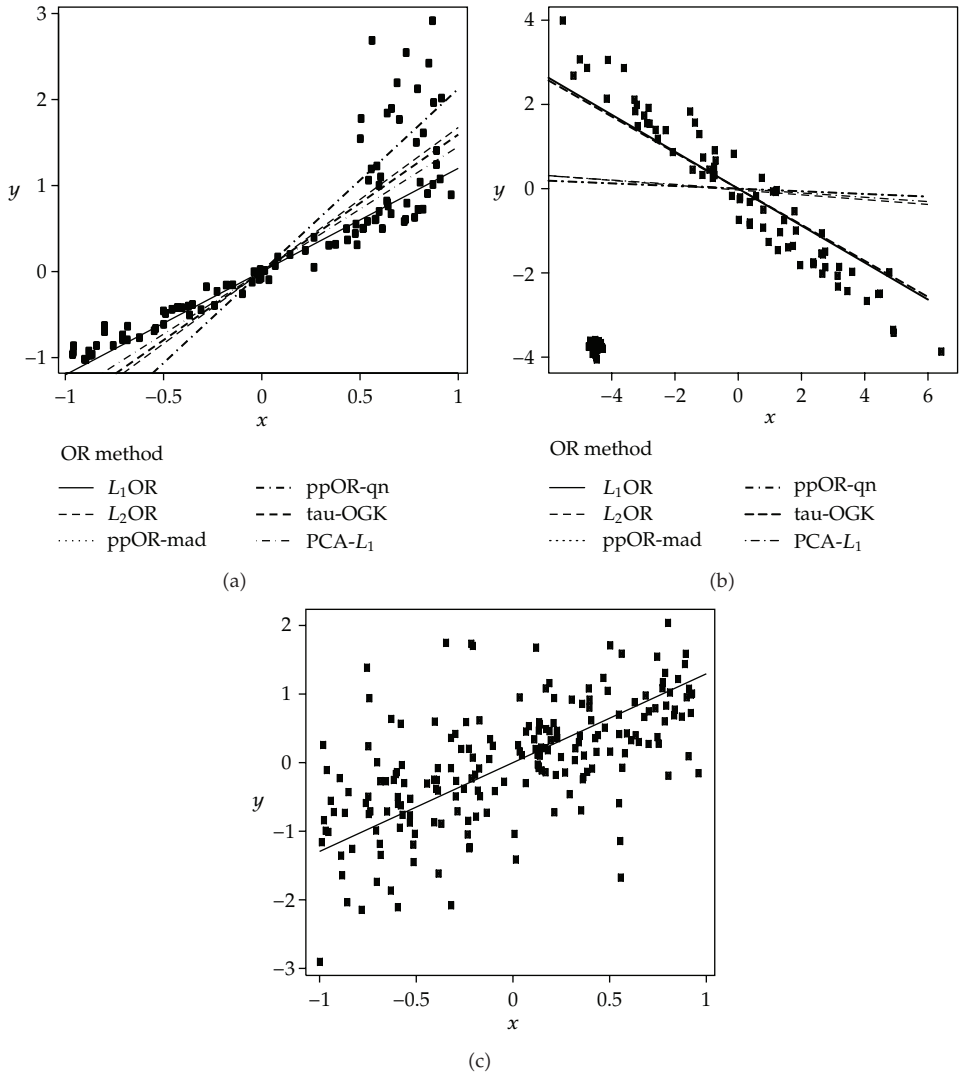

Figure 3: Examples of data sets used in simulation experiments and fitted models: (a) a data set with vertical outliers generated using parameters $m = 10$ and $C = 25$, (b) a data set with clustered leverage outliers with $n = 100$ and generated using $\epsilon = 0.25$, and (c) a data set with errors in both variables sampled from a Laplace distribution with $n = 200$.

where f is the known model and $\widehat{f}$ is the estimated model. Note that D corresponds to the area between f and $\widehat{f}$. If the estimated model is close to the true model, then D will be small. For each of the simulations D is computed and recorded. Using these results the average model discrepancy, $\overline{D}$, and standard error are computed.

To analyze the simulation, the means and standard deviations of D are computed for each setting of m and C and can be found in Table 1. For all configurations with $m \leq 10$, L_1OR has lower means and standard deviations than all other methods tested, indicating superior performance in resisting outlier contamination for such conditions. For $m = 50$, L_1OR performs worse than the robust methods with the exception of PCA-L_1 but better than

Table 1: Mean (standard deviation) of D for L_1OR, L_2OR, ppOR-mad, ppOR-qn, τ-OGK, and PCA-L_1 with contamination magnitudes $m = 1$, 1010, and 5050 and contamination levels $C = 0$, 10, and 25.

	Method	$m = 1$	$m = 10$	$m = 50$
$C = 0$	L_1OR	0.00997 (0.00540)		
	L_2OR	0.01818 (0.01459)		
	ppOR-mad	0.13624 (0.09616)		
	ppOR-qn	0.08398 (0.07724)		
	τ-OGK	0.01870 (0.01486)		
	PCA-L_1	0.02081 (0.01388)		
$C = 10$	L_1OR	0.00934 (0.00583)	0.08496 (0.01798)	0.31339 (0.05527)
	L_2OR	0.03070 (0.01578)	0.32365 (0.10149)	3.54666 (0.99552)
	ppOR-mad	0.13714 (0.12535)	0.11584 (0.10239)	0.08906 (0.07094)
	ppOR-qn	0.07475 (0.06369)	0.14938 (0.08210)	0.05840 (0.04831)
	τ-OGK	0.03018 (0.01696)	0.18032 (0.03857)	0.20396 (0.03736)
	PCA-L_1	0.02608 (0.01667)	0.17335 (0.04240)	0.76836 (0.16126)
$C = 25$	L_1OR	0.01190 (0.00573)	0.16172 (0.02743)	0.58962 (0.06106)
	L_2OR	0.04505 (0.01420)	0.62263 (0.12630)	6.26558 (1.35709)
	ppOR-mad	0.12443 (0.10311)	0.25518 (0.24805)	0.31136 (0.28315)
	ppOR-qn	0.08947 (0.08796)	0.59031 (0.18792)	0.24970 (0.12092)
	τ-OGK	0.03865 (0.01879)	0.45040 (0.09105)	0.54522 (0.08887)
	PCA-L_1	0.03940 (0.01382)	0.35664 (0.06198)	1.87768 (0.31515)

the outlier-sensitive L_2OR. In the case of extreme contamination ($C = 25, m = 50$), L_2OR and PCA-L_1 are extremely sensitive to outliers as indicated by large values for $\overline{D}$. The best-performing method for this configuration is ppOR-qn. L_1OR has mean discrepancy that is only 0.34 more than that of ppOR-qn but is at least 1.28 less than the outlier-sensitive methods. Overall, this suggests that L_1OR performs well when no contamination is present and in the presence of larger levels of contamination, but performance degrades relative to some of the robust methods when the contamination magnitude is very large.

3.2. Clustered Leverage Outliers

The ability of L_1OR to detect linear relationships in bivariate data with outliers is further analyzed with a simulation using datasets with *clustered leverage outliers*. Clustered leverage outliers in a dataset have very similar values but are far from the rest of the data set. The simulation design varies the number of observations (n) and the contamination level (ϵ). For each treatment condition and replication, a dataset is generated without contamination and a companion dataset is generated replacing the first $\lceil \epsilon n \rceil$ observations with contaminated data. There are 50 replications for each treatment condition. For this experiment, ϵ is varied in the following manner: low contamination: $\epsilon = 0.05$, moderate contamination: $\epsilon = 0.10$, and high contamination: $\epsilon = 0.25$.

The data are sampled as follows.

(i) Generate the uncontaminated data: $(x_i, y_i) \sim N(\mathbf{0}, \mathbf{\Sigma})$, for $i = 1, \ldots, n$.

(ii) Generate the contaminated data: $(x_i, y_i) \sim N(\mathbf{m}, 10^{-2}\mathbf{I})$, for $i = 1, \ldots, \lceil \epsilon n \rceil$.

The covariance matrix ($\mathbf{\Sigma}$) is varied across replications. First, a 2×2 matrix $\mathbf{A}$ is generated such that each entry is sampled from a $N(0, 1)$ distribution. The QR decomposition $\mathbf{A} = \mathbf{QR}$ is calculated. Let $\mathbf{B} = \mathbf{Q}\mathbf{I}\,\mathrm{sgn}(\langle \mathbf{R} \rangle)$, where $\langle \cdot \rangle$ indicates taking the diagonal elements as a vector and $\mathrm{sgn}(\cdot)$ is the vector with the signs of the corresponding elements of a vector. Then $\mathbf{\Sigma}$ is sampled from a Wishart($\mathbf{B}$, 5). The means ($\mathbf{m}$) for the contaminated data are generated such that

(1) the Mahalanobis distance of $\mathbf{m}$ from the distribution $N(\mathbf{0}, \mathbf{\Sigma})$ is at least $\sqrt{2\chi^2_{0.99,2'}}$

(2) $\min\{x_i : i = 1, \ldots, n\} \leq m_1 \leq \max\{\mathrm{x}_i : i = 1, \ldots, n\}$, and

(3) $\min\{y_i : i = 1, \ldots, n\} \leq m_2 \leq \max\{y_i : i = 1, \ldots, n\}$.

An example dataset with 100 observations and fitted models generated using $\epsilon = 0.25$ is given in Figure 3(b).

Each method is evaluated based on the similarity of the models fit on the companion uncontaminated and contaminated datasets. The similarity measure S is defined as the absolute value of the inner product

$$S\left(\mathbf{b}^1, \mathbf{b}^2\right) = \left|\mathbf{b}^1 \cdot \mathbf{b}^2\right|, \tag{3.2}$$

where $\mathbf{b}^1$ and $\mathbf{b}^2$ are the vectors of coefficients derived for the uncontaminated and contaminated datasets. The values of S can be between 0 and 1, with larger values indicating that the models are in agreement and that outliers do not affect the estimate.

The means across replications and the percentage of instances with $S \geq 0.90$ for each value of n and ϵ are contained in Table 2. For $\epsilon = 0.05,\ 0.10$, the performance of L_1OR is nearly constant as n is increased. There is a slight degradation in performance for larger values of n, which is likely due to the increased computational complexity of instances (see Section 5). For $\epsilon = 0.05$, all methods have high mean values for S and high percentages of instances with $S \geq 0.9$, including the outlier-sensitive L_2OR. For $\epsilon = 0.10$, L_1OR and all of the robust methods have larger mean values of S than L_2OR. The ppOR-qn estimator has the most consistent performance across different values of n for $\epsilon = 0.10$, with mean values of S above 0.94 for each. The L_1OR estimator has mean values above 0.93 for $n \leq 100$, but performance degrades for $n = 200$. The τ-OGK estimator has the highest or second-highest mean values of S for $n \leq 100$. For $\epsilon = 0.25$, the performance of L_1OR lags the robust methods. For $n \leq 50$, the performance is similar to that of L_2OR. For $n \geq 100$, the mean values of S are less than those for L_2OR. For $\epsilon = 0.25$, the preferred estimator appears to be ppOR-mad, as it has the highest or second-highest value of S for each n.

3.3. Consistency

The consistency of L_1OR is assessed by performing tests on instances with various sample sizes. Bivariate data (x_i, y_i), $i = 1, \ldots, n$ are generated such that $x_i = \nu_i + \epsilon_i$, where $\nu_i \sim U[-1, 1]$ and $\epsilon_i \sim Laplace(0, 0.5)$, and $y_i = x_i + \xi_i$, where $\xi_i \sim Laplace(0, 0.5)$. The sample sizes tested are

Table 2: Mean of S/percentage of instances with $S \geq 0.9$ for L_1OR, L_2OR, ppOR-mad, ppOR-qn, τ-OGK, and PCA-L_1 with sample sizes $n = 25$, 50, 100, and 200 and contamination levels $\epsilon = 0.05$, 0.10, and 0.25.

	Method	$\epsilon = 0.05$	$\epsilon = 0.10$	$\epsilon = 0.25$
$n = 25$	L_1OR	0.996/1.000	0.993/1.000	0.680/0.520
	L_2OR	0.981/0.980	0.963/0.920	0.648/0.240
	ppOR-mad	0.967/0.900	0.933/0.740	0.859/0.500
	ppOR-qn	0.963/0.880	0.944/0.800	0.869/0.460
	τ-OGK	0.994/1.000	0.985/0.980	0.842/0.660
	PCA-L_1	0.962/0.940	0.969/0.960	0.794/0.380
$n = 50$	L_1OR	0.998/1.000	0.932/0.920	0.602/0.360
	L_2OR	0.988/1.000	0.912/0.860	0.609/0.260
	ppOR-mad	0.974/0.900	0.943/0.860	0.903/0.660
	ppOR-qn	0.989/1.000	0.962/0.900	0.858/0.400
	τ-OGK	0.997/1.000	0.974/0.980	0.818/0.640
	PCA-L_1	0.986/0.960	0.932/0.880	0.779/0.380
$n = 100$	L_1OR	0.973/0.960	0.931/0.900	0.519/0.180
	L_2OR	0.981/0.960	0.884/0.700	0.623/0.200
	ppOR-mad	0.979/0.960	0.956/0.900	0.923/0.700
	ppOR-qn	0.989/1.000	0.958/0.900	0.878/0.480
	τ-OGK	0.998/1.000	0.977/0.940	0.828/0.540
	PCA-L_1	0.979/0.960	0.940/0.880	0.810/0.340
$n = 200$	L_1OR	0.932/0.800	0.857/0.760	0.509/0.140
	L_2OR	0.917/0.820	0.805/0.580	0.608/0.160
	ppOR-mad	0.975/0.960	0.970/0.920	0.942/0.780
	ppOR-qn	0.978/0.980	0.959/0.860	0.893/0.560
	τ-OGK	0.997/1.000	0.954/0.920	0.834/0.600
	PCA-L_1	0.926/0.860	0.922/0.820	0.785/0.340

$n = 10, 25, 50, 100, 200$; data are generated for 100 datasets for each value of n. The *rlaplace()* function in the R package *rmutil* [32] is used to sample from the Laplace distribution. An example dataset with 200 observations and the fitted L_1OR model is given in Figure 3(c).

Figure 4 depicts the standard error of the absolute value of the slope as a function of sample size. As sample size increases, the standard error rapidly approaches zero, indicating that the procedure is consistent. For large sample sizes, L_1OR should provide good estimates.

4. An Environmental Example

The pH and alkalinity of the water in which the fish live are known to impact their overall health. Alkalinity is a measure of the ability of a solution to neutralize acids. Researchers expect pH and alkalinity be highly correlated. However, the relationship of the two variables is difficult to estimate in many datasets due to low variation in pH across streams and due to the presence of outliers. The dataset for this example is a subset of values collected across the state of Ohio resulting in 312 observations. Various subsets of this dataset have been considered previously by Norton [17], Lipkovich et al. [18], Noble et al. [19], and Boone et al.

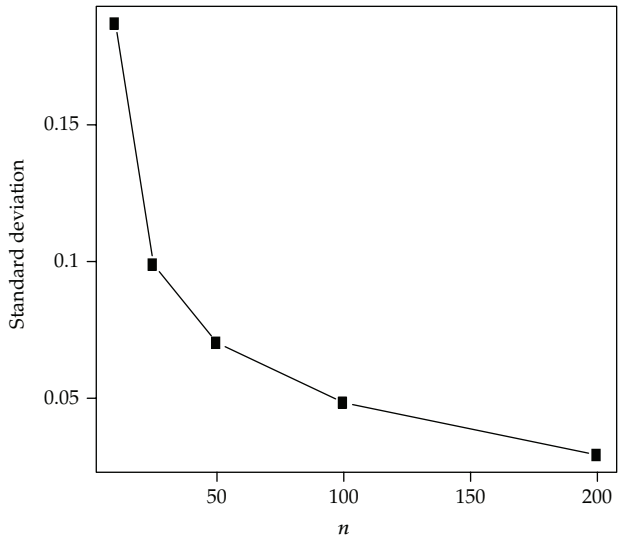

Figure 4: Plot of the standard error of the absolute value of the slope in a bivariate experiment as a function of sample size (n).

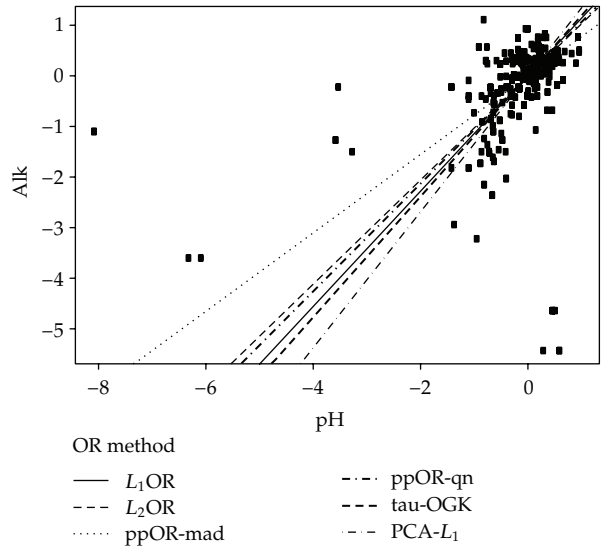

Figure 5: Scatter plot of pH versus alkalinity with models from L_1OR, L_2OR, and ppOR-mad; ppOR-qn, τ-OGK, and PCA-L_1.

[20] with varying degrees of success at estimating the relationship between pH and alkalinity. For the purposes of this work both pH and alkalinity have been normalized. Note that in this data both pH and alkalinity have measurement error, and hence an orthogonal regression method should be used. The same computational settings as in the simulation studies are used for this analysis.

Figure 5 shows the scatter plot of pH versus alkalinity. There appears to be a linear relationship between alkalinity and pH. Also notice the vertical and leverage outliers present

Table 3: Summary of regression models for alkalinity on pH using L_1OR, L_2OR, ppOR-mad, ppOR-qn, τ-OGK, and PCA-L_1.

Method	Estimate	Standard error*	T value	P value
L_1OR	−0.87760	0.05993	−14.64350	.00000
L_2OR	−0.97168	0.42857	−2.26728	.02555
ppOR-mad	−1.28919	0.74707	−1.72567	.08753
ppOR-qn	−0.93906	0.20355	−4.61344	.00001
τ-OGK	−0.83845	0.10632	−7.88578	.00000
PCA-L_1	−0.74421	0.11951	−6.22713	.00000

in the data. The Pearson correlation coefficient for the relationship between pH and alkalinity is $r = 0.3366$ which is biased down due to the outliers in the data. Furthermore, since the correlation is biased down, extracting a pH-alkalinity component and using that as a predictor would not be prudent. Hence, a regression method is needed that is insensitive to outliers/influential points. With the exception of ppOR-mad, the outlier-insensitive methods all demonstrate resistance to the outliers in the measurements of pH when compared to the outlier-sensitive L_2OR. The method based on PCA-L_1 produces a model that appears to be least affected by outliers, followed by the τ-OGK estimator and then L_1OR.

Table 3 shows the summary of regression models for each method. Here the standard errors are bootstrap standard errors based on 100 bootstrap samples. The bootstrap standard errors vary widely across the methods, with L_1OR having the most stable estimates, as evidenced by smaller standard errors, followed by τ-OGK and PCA-L_1. With the exception of ppOR-mad, the P values indicate that the relationship between pH and alkalinity is statistically significant. Notice that the P value for the relationship is statistically significant using L_1OR, τ-OGK, and PCA-L_1 with a P value of less than .00001. The L_1OR, τ-OGK, and PCA-L_1 estimators appear to be the best choice for this data, with PCA-L_1 producing the best estimate and L_1OR providing the most stable estimate.

An expansion of this problem is to consider how alkalinity, pH, and habitat measure qualitative habitat evaluation index (QHEI) impact the index of biotic integrity (IBI). QHEI measures the quality of the habitat in which the fish reside [21]. QHEI is determined from the following six measures: stream substrate, in stream cover, channel morphology, riparian and bank condition, pool and riffle quality, and gradient. Here higher values correspond to better habitat quality, and lower values correspond to poorer habitat quality. IBI measures the health of the fish community. Lower values of IBI correspond only to tolerant species present, low community organization, and high proportion of fish with physical anomalies. High values correspond to highly organized fish communities, many intolerant species present, and high diversity among species [22]. The data consist of 312 observations from the same sites.

Orthogonal regression models are fit to the data with IBI as the response and QHEI, pH, and alkalinity as predictors. The method presented by Croux and Haesbroeck [10] (hereafter CH) is used instead of τ-OGK because of the increased number of variables. The CH method is a robust PCA based on finding eigenvalues of a robust estimate of the covariance matrix.

Table 4 shows the coefficients, bootstrap standard errors, t value, and P values for the regressions using each method. Notice that the L_1OR estimates for the coefficients for pH and alkalinity are the most stable, as indicated by lower standard errors. The standard errors for the QHEI coefficient estimates are the smallest for each method. The estimate for the QHEI coefficient by CH has the lowest standard error and is the largest positive estimate

Table 4: Summary of regression models for IBI on QHEI, pH, and alkalinity (ALK) using L_1OR, L_2OR, ppOR-mad, ppOR-qn, CH, and PCA-L_1.

Method	Variable	Estimate	Standard error*	T value	P value
L_1OR	QHEI	−0.67248	5.58039	−0.12051	.90433
	pH	0.70695	9.31363	0.07591	.93965
	ALK	−1.26714	5.60148	−0.22622	.82150
L_2OR	QHEI	0.17841	5.79662	0.03078	.97551
	pH	−11.25396	86.54848	−0.13003	.89681
	ALK	3.97315	64.00076	0.06208	.95062
ppOR-mad	QHEI	0.07038	5.57955	0.01261	.98996
	pH	4.31182	68.31064	0.06312	.94980
	ALK	−4.36975	41.89404	−0.10430	.91714
ppOR-qn	QHEI	−1.88655	4.41714	−0.42710	.67024
	pH	21.51640	83.28626	0.25834	.79668
	ALK	−13.92729	60.63738	−0.22968	.81881
CH	QHEI	0.33704	3.75244	0.08982	.92861
	pH	21.25614	53.26711	0.39905	.69072
	ALK	−21.16159	67.57228	−0.31317	.75481
PCA-L_1	QHEI	−0.97810	9.09927	−0.10749	.91462
	pH	−3.61571	107.89751	−0.03351	.97333
	ALK	3.33549	79.46839	0.04197	.96661

*Standard errors are bootstrap standard errors based on 100 bootstrap samples.

which seems to agree best with biological expectations. The better the habitat the fish have to live in, the better the health of the fish community. For all methods except for L_2OR and PCA-L_1, the coefficients indicate a positive correlation between IBI and pH and a negative correlation between IBI and alkalinity. While none of the variables in any of the regressions are statistically significant, this dataset provides an example of how the regression coefficients from orthogonal regression with outliers may be suspect.

5. Computation Time

The solution method proposed for L_1OR is more computationally intensive than the other methods used for comparison in this paper. The alternative methods solve all of the instances used here in less than a few seconds. In this section, we evaluate the computational performance of our implementation of L_1OR.

Tables 5–8 contain data on the computational performance of L_1OR in each of the experiments conducted. In each table, the first column(s) indicates the configuration of the data: for Table 5 the contamination level C and contamination magnitude m; for Table 6 the sample size n, contamination level ϵ, and whether the data has outliers; for Table 7 the sample size n; for Table 8 the number of variables d. The second column *% Optimal* indicates the percentage of instances solved to optimality, meaning that all MIP subproblems solved to optimality and the RLT branch and bound tree are fully explored. The third column *Avg. MIPs Solved* contains the average number of MIPs solved for each configuration. The fourth column *Avg. MIPs Suboptimal* contains the average number of MIPs that were not solved to optimality within the 120 CPU second time limit. The fifth column *Avg. Time-to-Term. (s)*

Table 5: Computational performance of L_1OR implementation for simulation with vertical outliers.

C	m	% Optimal	Avg. MIPs solved	Avg. MIPs suboptimal	Avg. time to Term. (s)	Avg. time to Best Soln. (s)
0	0	7.0	193.2	1.2	289.4	235.2
10	1	6.0	217.9	1.0	282.2	242.7
10	10	6.0	187.0	0.9	279.1	255.5
10	50	14.0	117.9	0.5	199.0	186.4
25	1	9.0	236.4	0.9	287.1	226.2
25	10	1.0	105.7	1.2	231.3	223.0
25	50	24.0	154.1	0.2	163.1	149.4

Table 6: Computational performance of L_1OR implementation for simulation with clustered leverage outliers.

n	ϵ	Contamination	% Optimal	Avg. MIPs Solved	Avg. MIPs Suboptimal	Avg. time to Term. (s)	Avg. time to Best Soln. (s)
25	0.05	N	100.00	112.5	0.0	5.7	5.1
25	0.05	Y	100.00	116.4	0.0	7.3	6.5
25	0.1	N	100.00	110.8	0.0	6.1	5.3
25	0.1	Y	100.00	114.8	0.0	8.2	7.6
25	0.25	N	100.00	140.6	0.0	7.8	6.7
25	0.25	Y	100.00	104.1	0.0	10.3	9.9
50	0.05	N	90.00	115.8	0.1	55.4	52.2
50	0.05	Y	82.00	111.2	0.2	82.7	80.8
50	0.1	N	94.00	127.3	0.1	53.0	49.8
50	0.1	Y	76.00	113.5	0.3	102.4	100.1
50	0.25	N	86.00	115.1	0.2	62.8	60.7
50	0.25	Y	44.00	125.9	0.8	186.3	184.1
100	0.05	N	10.00	119.5	2.4	445.3	434.5
100	0.05	Y	6.00	124.5	3.1	548.3	541.4
100	0.1	N	16.00	106.1	2.1	389.0	378.0
100	0.1	Y	4.00	112.7	4.2	697.2	671.8
100	0.25	N	6.00	118.7	2.6	465.8	452.3
100	0.25	Y	0.00	114.4	5.9	911.2	886.6
200	0.05	N	0.00	96.9	7.2	1243.3	1154.3
200	0.05	Y	0.00	99.1	9.1	1459.7	1398.4
200	0.1	N	0.00	93.4	7.3	1201.8	1150.1
200	0.1	Y	0.00	102.2	10.8	1662.2	1617.2
200	0.25	N	0.00	108.4	7.3	1249.9	1206.4
200	0.25	Y	0.00	102.3	11.8	1743.9	1704.6

contains the average of the lesser of the CPU seconds before the RLT branch and bound tree is explored and 7200 seconds. The last column *Avg. Time to Best Soln. (s)* contains the average time to find the best feasible solution.

With the exception of the bootstrap samples for the environmental data with $d = 4$ (Table 8), the RLT branch-and-bound tree is explored in every instance. However, for many of

Table 7: Computational performance of L_1OR implementation for consistency experiment.

n	% Optimal	Avg. MIPs solved	Avg. MIPs suboptimal	Avg. time to term. (s)	Avg. time to best soln. (s)
10	100.00	134.1	0.0	1.7	1.3
25	100.00	131.9	0.0	9.2	7.9
50	89.00	127.8	0.1	54.8	51.6
100	1.00	115.7	1.6	358.2	346.8
200	0.00	105.1	7.7	1481.8	1418.5

Table 8: Computational performance of L_1OR implementation for bootstrap simulations.

d	% Optimal	Avg. MIPs solved	Avg. MIPs suboptimal	Avg. time to term. (s)	Avg. time to best soln. (s)
2	0.00	105.1	3.2	1099.3	1020.5
4	0.00	69.5	57.5	7257.9	5511.9

these instances, at least one of the MIP subproblems is not solved to optimality. The solution taken in these instances is therefore not "provably" optimal. All instances with $n \leq 25$ are solved to optimality. As n is increased to 50 and larger, fewer instances are solved to optimality. For $n \leq 100$, the number of MIP subproblems that are not solved to optimality is less than 5% of the subproblems solved in those instances on average. For $n = 200$ in the simulation for clustered leverage outliers and the consistency experiment, about 10% of the MIPs are not solved to optimality. For the bootstrap simulation with $d = 4$, more than half of the MIPs were not solved to optimality.

In the simulation with vertical outliers (Table 5), more instances are solved to optimality when the outlier contamination is larger. In contrast, in the simulation with clustered leverage outliers (Table 6), fewer instances with contamination are solved to optimality, than the companion datasets without contamination. Also, the number of instances solved to optimality seems to decrease as the contamination level increases.

In the simulation with vertical outliers (Table 5), at least one MIP is not solved to optimality in most instances. Except in the case of extreme outlier contamination, L_1OR performed competitively when compared to robust methods. Also, the standard error for the slope in the consistency experiment (Table 7), the percentage of instances solved to optimality decreases dramatically as n increases, but the standard error for the estimates continues to decrease. For these instances then, the time limit for the MIP subproblems does not appear to hamper the ability to find good solutions.

Only one experiment, the bootstrap simulation with $d = 4$, used data with more than 2 variables. The degradation in computational performance is more dramatic in the shift from $d = 2$ to $d = 4$ in the bootstrap simulation than the degradation observed when n is increased in the bivariate experiments. This phenomenon is likely due to the increase in nonlinear constraints needed to produce the RLT relaxation.

6. Discussion

This work introduces a new L_1 orthogonal regression technique that is designed to be resistant to outliers. We develop a method for deriving globally optimal solutions for problem instances. Via simulation, the method shows promise for being resistant to outliers. An

application to an environmental example further demonstrates that the method produces results which are more resistant to outliers than traditional orthogonal regression and competes with other robust methods. Hence, this method gives data analysts that deal with errors-in-variables data contaminated with outliers a resistant alternative to orthogonal regression.

The computational studies presented here indicate that different robust or outlier-resistant methods are suitable in different situations, and there is no clearly superior method. The pcaPP-mad method is among the best performers in the presence of vertical and clustered leverage outliers in simulated data but has perhaps the poorest estimate in the real-world example that contains both types of outliers. PCA-L_1 is among the poorest performers in the presence of vertical and clustered leverage outliers in simulated data but produces some of the best estimates in the real-world analysis. The inconsistency of the results for PCA-L_1 may be due to the dependence of the method on having a good starting point for finding a good local optimal solution. The L_1OR method presented here performs best with respect to the other methods in the presence of moderate contamination by vertical outliers but suffers in cases of extreme contamination.

Traditional orthogonal regression (L_2OR) can be formulated as a special case of PCA. The approach presented in this work for formulation and optimization can potentially be adapted to develop an outlier-resistant method for PCA. An outlier-resistant PCA algorithm would be useful for data analysts that work with contaminated data. Another possible extension is for an outlier-resistant factor analysis procedure for analyzing categorical data.

Acknowledgment

The authors would like to thank two anonymous referees for numerous suggestions for improving the content and presentation of this work.

References

[1] M. L. Brown, "Robust line estimation with errors in both variables," *Journal of the American Statistical Association*, vol. 77, pp. 71–79, 1982.

[2] R. J. Carroll and P. P. Gallo, "Aspects of robustness in the functional errors-in-variables regression model," *Communications in Statistics*, vol. 11, pp. 2573–2585, 1982.

[3] R. H. Zamar, "Robust estimation in the errors-in-variables model," *Biometrika*, vol. 76, pp. 149–160, 1989.

[4] H. Späth and G. A. Watson, "On orthogonal linear ℓ1 approximation," *Numerische Mathematik*, vol. 51, no. 5, pp. 531–543, 1987.

[5] N. A. Campbell, "Robust procedures in multivariate analysis—I: robust covariance estimation," *Applied Statistics*, vol. 29, pp. 231–237, 1980.

[6] S. J. Devlin, R. Gnandesikan, and J. R. Kettenring, "Robust estimation of dispersion matrices and principal components," *Journal of the American Statistical Association*, vol. 76, pp. 354–362, 1981.

[7] J. S. Galpin and D. M. Hawkins, "Methods of L_1 estimation of a covariance matrix," *Computational Statistics and Data Analysis*, vol. 5, no. 4, pp. 305–319, 1987.

[8] R. A. Naga and G. Antille, "Stability of robust and non-robust principal components analysis," *Computational Statistics and Data Analysis*, vol. 10, no. 2, pp. 169–174, 1990.

[9] J. I. Marden, "Some robust estimates of principal components," *Statistics and Probability Letters*, vol. 43, no. 4, pp. 349–359, 1999.

[10] C. Croux and G. Haesbroeck, "Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies," *Biometrika*, vol. 87, no. 3, pp. 603–618, 2000.

[11] H. Kamiya and S. Eguchi, "A class of robust principal component vectors," *Journal of Multivariate Analysis*, vol. 77, no. 2, pp. 239–269, 2001.

[12] G. Li and Z. Chen, "Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and Monte Carlo," *Journal of the American Statistical Association*, vol. 80, pp. 759–766, 1985.

[13] Y. Xie, J. Wang, Y. Liang, L. Sun, X. Song, and R. Yu, "Robust principal components analysis by projection pursuit," *Journal of Chemometrics*, vol. 7, pp. 527–541, 1993.

[14] R. Maronna, "Principal components and orthogonal regression based on robust scales," *Technometrics*, vol. 47, no. 3, pp. 264–273, 2005.

[15] C. Croux and A. Ruiz-Gazen, "High breakdown estimators for principal components: the projection-pursuit approach revisited," *Journal of Multivariate Analysis*, vol. 95, no. 1, pp. 206–226, 2005.

[16] N. Kwak, "Principal component analysis based on L1-norm maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 9, pp. 1672–1680, 2008.

[17] S. B. Norton, *Using Biological Monitoring Data to Distinguish Among Types of Stress in Streams of the Eastern Corn Belt Plains Ecoregion*, Ph.D. thesis, George Mason University, Fairfax, Va, USA, 1999.

[18] I. Lipkovich, E. P. Smith, and K. Ye, "Evaluating the impact environmental stressors on Benthic macroinvertebrate communities via Bayesian model averaging," in *Case Studies in Bayesian Statistics*, pp. 267–283, 2002.

[19] R. Noble, E. P. Smith, and K. Ye, "Model selection in canonical correlation analysis (CCA) using Bayesian model averaging," *Environmetrics*, vol. 15, no. 4, pp. 291–311, 2004.

[20] E. L. Boone, K. Ye, and E. P. Smith, "Evaluating the relationship between ecological and habitat conditions using hierarchical models," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 10, no. 2, pp. 131–147, 2005.

[21] Ohio Environmental Protection Agency, *The Qualitative Habitat Evaluation Index (QHEI): Rationale, Methods and Application*, State of Ohio Environmental Protection Agency, 1989.

[22] Ohio Environmental Protection Agency, *Biological Criteria for the Protection of Aquatic Life: Volume II: Users Manual for Biological Assessment of Ohio Surface Waters*, State of Ohio Environmental Protection Agency, 1988, WQMA-SWS-6.

[23] A. Baccini, P. Besse, and A. de Faguerolles, "A L1-norm PCA and heuristic approach," in *Proceedings of the International Conference on Ordinal and Symbolic Data Analysis*, pp. 359–368, 1987.

[24] S. Agarwal, M. K. Chandraker, F. Kahl, D. Kriegman, and S. Belongie, "Practical global optimization for multiview geometry," *Lecture Notes in Computer Science*, vol. 3951, pp. 592–605, 2006.

[25] S. Zwanzig, "On L_1-norm estimators in nonlinear regression and nonlinear errors-in-variables models," *IMS Lecture Notes—Monograph Series*, vol. 35, pp. 101–118, 1997.

[26] P. J. Rousseeuw and A. Struyf, "Computing location depth and regression depth in higher dimensions," *Statistics and Computing*, vol. 8, no. 3, pp. 193–203, 1998.

[27] H. D. Sherali and C. H. Tuncbilek, "A global optimization algorithm for polynomial programming problems using a Reformulation-Linearization Technique," *Journal of Global Optimization*, vol. 2, no. 1, pp. 101–112, 1992.

[28] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, USA, 2nd edition, 2002.

[29] R. A. Maronna and R. H. Zamar, "Robust estimates of location and dispersion for high-dimensional datasets," *Technometrics*, vol. 44, no. 4, pp. 307–317, 2002.

[30] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2009.

[31] P. Filzmozer, H. Fritz, and K. Kalcher, *pcaPP: Robust PCA by Projection Pursuit*, 2009.

[32] J. Lindsey, *rmutil: Utilities for Nonlinear Regression and Repeated Measurements*, 2009.