

Article

Kibria–Lukman-Type Estimator for Regularization and Variable Selection with Application to Cancer Data

Adewale Folaranmi Lukman ¹, Jeza Allohibi ^{2,*}, Segun Light Jegede ³, Emmanuel Taiwo Adewuyi ⁴,
Segun Oke ⁵ and Abdulmajeed Atiah Alharbi ²

¹ Department of Mathematics, University of North Dakota, Grand Forks, ND 58202, USA; adewale.lukman@und.edu

² Department of Mathematics, Faculty of Science, Taibah University, Al-Madinah Al-Munawara 42353, Saudi Arabia; aahharbi@taibahu.edu.sa

³ Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA; sjegede@kent.edu

⁴ Department of Statistics, Ladoke Akintola University of Technology, Ogbomoso 212102, Nigeria; etadewuyi46@lautech.edu.ng

⁵ Department of Physics, Chemistry and Mathematics, Alabama A&M University, Huntsville, AL 35762, USA; segunoke2016@gmail.com

* Correspondence: jlohibi@taibahu.edu.sa

Abstract: Following the idea presented with regard to the elastic-net and Liu-LASSO estimators, we proposed a new penalized estimator based on the Kibria–Lukman estimator with L1-norms to perform both regularization and variable selection. We defined the coordinate descent algorithm for the new estimator and compared its performance with those of some existing machine learning techniques, such as the least absolute shrinkage and selection operator (LASSO), the elastic-net, Liu-LASSO, the GO estimator and the ridge estimator, through simulation studies and real-life applications in terms of test mean squared error (TMSE), coefficient mean squared error (β MSE), false-positive (FP) coefficients and false-negative (FN) coefficients. Our results revealed that the new penalized estimator performs well for both the simulated low- and high-dimensional data in simulations. Also, the two real-life results show that the new method predicts the target variable better than the existing ones using the test RMSE metric.

Keywords: regularization; variable selection; elastic-net; LASSO; ridge estimator; Liu-LASSO

MSC: 62J05; 65F35



Citation: Lukman, A.F.; Allohibi, J.; Jegede, S.L.; Adewuyi, E.T.; Oke, S.; Alharbi, A.A. Kibria–Lukman-Type Estimator for Regularization and Variable Selection with Application to Cancer Data. *Mathematics* **2023**, *11*, 4795. <https://doi.org/10.3390/math11234795>

Academic Editor: Simeon Reich

Received: 16 October 2023

Revised: 10 November 2023

Accepted: 21 November 2023

Published: 28 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The linear regression model describes the response variable as a linear combination of one or more predictors, and it is widely applicable in fields such as medicine, biological science, economics, etc. The model is defined as follows:

$$y_i = x_i^T \beta + \varepsilon_i \quad (1)$$

where $y = (y_1, \dots, y_n)^T$ is a $n \times 1$ vector of the dependent (response) variable; $X = (x_1, \dots, x_n)^T$, $x_i \in \mathbb{R}^{n \times p}$ and $i = 1, \dots, n$ are the known $n \times p$ matrix of predictors; $\beta = (\beta_1, \dots, \beta_p)^T$ is a $p \times 1$ vector of regression coefficients; and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ is the random error vector such that $E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \sigma^2 \in \mathbb{R}^+$, $i = 1, \dots, n$.

The model parameters are usually estimated using the least squares (LS) estimator. The LS estimator minimized the residual sum of squares (RSS) of model (1) and produced the following expression:

$$\hat{\beta}_n^{LS} = \arg \min_{\beta \in \mathbb{R}^p} \frac{1}{2n} \|y - X\beta\|^2 \quad (2)$$

where $\|\cdot\|_2$ denotes the Euclidean norm. The high-dimensional quantitative structure–activity relationship (QSAR) modeling method is a popular example of the linear regression models. A high-dimensional problem arises when there are more features (variables) than the sample size (n) in a dataset. The popular OLSE does not give a unique solution via high-dimensional QSAR modeling since the number of molecular descriptors exceeds the compound number [1,2]. Likewise, the performance of the OLSE suffers a setback when there is linear dependency among the features (predictors), a condition called multicollinearity [3–6]. The setback includes a high standard error, insignificant t -tests, wrong conclusions, etc. [6–8].

The shrinkage estimators were mainly adopted to account for the linear dependency among the predictors. The shrinkage estimators include the ridge estimator [9], the Liu estimator [10], the Liu-type estimator [11], the two-parameter estimator [12], the modified ridge-type estimator [6], the modified Liu estimator [13], the Kibria–Lukman (KL) estimator [14], the modified KL estimator [15], and the Dawoud–Kibria (DK) estimator [16]. These shrinkage estimators account for more multicollinearity than the OLSE. The ridge estimator minimizes the residual sum of the squares (RSS) subject to an L2-norm of the coefficients, and it is defined as follows:

$$\hat{\beta}_n^\lambda = \arg \min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \frac{\lambda}{2} \|\beta\|_2^2 \right\}, \quad (3)$$

where $\lambda > 0$ represents the penalized parameter. The estimator produces a more stable estimate and a better prediction than the LS estimator. The ridge estimator will only perform regularization in high-dimensional settings and will not perform variable selection. It fails to reduce any of its coefficients to zero. Hence, it becomes difficult to interpret the regression models in high-dimensional settings.

Variable selection forms a crucial aspect of modeling high-dimensional problems. It is worthwhile to state that removing the important features (predictors) produces more biased regression estimates and predictions. Likewise, including irrelevant predictors degrades the estimation efficiency and leads to imprecise predictions [17]. Thus, variable selection seeks to select the most relevant variables from a large dataset and remove irrelevant and redundant variables. Moreover, it improves the performance of the model prediction and reduces the computational cost. The problem of variable selection has been studied extensively in the literature [17–22].

The least absolute penalty and selection operator (LASSO) [18] is a well-known variable selection technique obtained by subjecting the RSS to an L1-norm penalty, and it is defined as follows:

$$\hat{\beta}_n^L = \arg \min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\}, \quad (4)$$

where $\lambda > 0$ is the penalized parameter. LASSO retains the regularization property of the ridge and performs variable selection due to the introduction of the L1-norm penalty. It also improves the model prediction by eliminating irrelevant variables. However, the estimator often chooses only one variable from the group of pair-wise correlated predictors, and it is dominated by the ridge estimator when the multicollinearity level is high [1,8,18,22].

The elastic-net (E-net) approach [1] is another popular variable selection technique obtained by subjecting the RSS to the L1 and L2 norms penalty simultaneously. This technique retains the features of the ridge and the LASSO estimators. It is defined as follows:

$$\hat{\beta}_n^E = \arg \min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \frac{\lambda_2}{2} \|\beta\|_2^2 + \lambda_1 \|\beta\|_1 \right\}, \quad (5)$$

where $\lambda_1, \lambda_2 > 0$, $\|\cdot\|_2$ is the L2-norm, and $\|\cdot\|_1$ is L1-norm. The drawback of this method is that the L2-penalty term introduced extra bias that increases its variance [23].

Recently, the GO estimator [23] was developed as an improvement on the elastic net model. The estimator combines the features of the two-parameter estimator [24] and the LASSO. It is defined as follows:

$$\hat{\beta}_n^{GO} = \arg \min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \frac{\lambda_2}{2} \|\beta - d\hat{\beta}_n^{LS}\|_2^2 + \lambda_1 \|\beta\|_1 \right\}, \quad (6)$$

where $\lambda_1, \lambda_2, d > 0$.

The Liu-LASSO [21] is another type of E-net that is defined as follows:

$$\hat{\beta}_n^{LL} = \arg \min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \frac{1}{2} \|\beta - d\hat{\beta}_n^{LS}\|_2^2 + \lambda_1 \|\beta\|_1 \right\}, \quad (7)$$

where $\lambda_1, d > 0$. Both the GO estimator and the Liu-LASSO adopted two other shrinkage estimators (the two-parameter and the Liu estimators) other than the ridge estimator. Research shown that these shrinkage estimators compete well with the ridge estimator. Recently, the Kibria–Lukman (KL) estimator [14] shows competitive advantage over the ridge and the Liu estimator. Thus, the KL estimator will be adopted in a manner similar to the Liu-LASSO to form a new variable selection technique. The goal is to produce a new estimator in the class of the previously mentioned ones and assess its ability to deal with multicollinearity and high-dimensional problems.

The outline of the article is as follows: We define the new estimator in Section 2 and derive its properties. We analyze low-dimensional data and high-dimensional QSAR data to compare the proposed estimator to the existing ones, as shown in Section 4. The simulation study and conclusion are given in Sections 3 and 5, respectively.

2. Proposed Estimator (KL1)

The Kibria–Lukman estimator is an optimization problem given as follows:

$$\hat{\beta}_n^{KL} = \arg \min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \frac{\lambda_2}{2} \|\beta + \hat{\beta}_n^{LS}\|_2^2 \right\}, \quad (8)$$

where $\lambda_2 > 0$. The KL estimator possesses smaller bias compared to the ridge and the Liu estimator. Therefore, in accordance with the Liu-LASSO estimator, we proposed the KL1 estimator, and it is defined as follows:

$$\hat{\beta}_n^{KL1} = \arg \min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \frac{1}{2} \|\beta + \hat{\beta}_n^{LS}\|_2^2 + \lambda_1 \|\beta\|_1 \right\}, \quad (9)$$

where $\lambda_1, \lambda_2 > 0$. The estimator automatically retains the regularization and the variable selection property due to the L1-norm penalty.

Coordinate Descent for KL1

We assume that the predictor variables are standardized and the response variable is centered. We attempt to find $\hat{\beta}_j$, which minimizes the following objective function:

$$\begin{aligned} (\hat{\beta}_j) &= \frac{1}{2n} \sum_{i=1}^n (y_i - x_{ij}\beta_j)^2 + \frac{1}{2} \sum_{j=1}^p (\beta_j + \hat{\beta}_j^{LS})^2 + \lambda_1 \sum_j |\beta_j| \\ &= \frac{1}{2n} \sum_{i=1}^n [y_i - x_{ij}\beta_j - \sum_{k \neq j} x_{ik}\beta_k]^2 + \frac{1}{2} \sum_{j=1}^p (\beta_j + \hat{\beta}_j^{LS})^2 + \lambda_1 |\beta_j| + \lambda_1 \sum_{k \neq j} |\beta_k| \\ &= \frac{1}{2n} \sum_{i=1}^n [y_i - \bar{y}_i^{(j)} - x_{ij}\beta_j]^2 + \frac{1}{2} \sum_{j=1}^p (\beta_j + \hat{\beta}_j^{LS})^2 + \lambda_1 |\beta_j| + \lambda_1 \sum_{k \neq j} |\beta_k| \end{aligned} \quad (10)$$

where $\bar{y}_i^{(j)} = \sum_{k \neq j} x_{ik} \beta_k$. The objective function (10) is differentiable everywhere except for $|\beta_j|$. The first derivative for $f(\hat{\beta}_j)$ when $\beta_j \neq 0$ is given as follows:

$$\frac{\partial f(\hat{\beta}_j)}{\partial \beta_j} = \frac{1}{n} \sum_{i=1}^n x_{ij} (y_i - \bar{y}_i^{(j)}) + \hat{\beta}_j^{LS} + \frac{1}{n} \sum_{i=1}^n x_{ij}^2 \beta_j + \lambda_1 \text{sgn}(\beta_j) \quad (11)$$

where $\text{sgn}(\beta_j) = \begin{cases} 1 & \beta_j > 0 \\ 0 & \beta_j = 0 \\ -1 & \beta_j < 0 \end{cases}$. The soft threshold estimate of β_j was obtained by setting Equation (11) to zero. Hence,

$$\beta_j \leftarrow \frac{S\left(\frac{1}{n} \sum_{i=1}^n x_{ij} (y_i - \bar{y}_i^{(j)}) - \hat{\beta}_j^{LS}, \lambda_1\right)}{\frac{1}{n} \sum_{i=1}^n x_{ij}^2 + 1} \quad (12)$$

where $y_i - \bar{y}_i^{(j)}$ is the partial residual for fitting β_j , and $S(z, \lambda_1)$ is the soft-thresholding operator defined by [22] as follows:

$$S(z, \lambda_1) = \begin{cases} z + \lambda_1, & z < -\lambda_1 \\ 0, & |z| \leq \lambda_1 \\ z - \lambda_1, & z > \lambda_1 \end{cases} \quad (13)$$

3. Simulation Studies

In this section, we designed three different cases to explore the performances of the proposed method (KL1) and other estimators. The simulation model is defined as follows:

$$y = X\beta + \sigma\epsilon \quad (14)$$

where the random error is $\epsilon \sim N(0, I)$. The predictor X is generated such that $X \sim N_p(0, \Sigma)$ where $\Sigma_{kj} = \rho^{|k-j|}$. ρ is the correlation between the predictors. Each independent simulated dataset is divided into three parts: a training set, a validation set and a test set. The division is performed at a ratio of 60%:20%:20%. The training data were used to train the model, while the validation data were used to predict the response while tuning the model's hyperparameters. The tuning parameters used in the coordinate descent algorithm (Algorithm 1) specify a grid in the interval $[0, 1]$. The test data are used to provide an unbiased evaluation of a final model fit on the training dataset. The simulation is conducted in two senses of dimensionality—high-dimensional data and low-dimensional data.

The low-dimensional data simulation was carried out for sample sizes of $n = 100, 200$, and 400. The simulation was also computed to determine the various correlations between the predictors and variances such that $\rho = 0.5, 0.8$, and 0.99 and $\sigma = 5, 10$. We simulated 8 true β s with 50 datasets for different cases as follows:

Case 1: $\beta = (3, 1.5, 0, 0, 2, 0, 0, 0)'$

Case 2: $\beta = (3, 0, 0, 0, 0, 0, 0, 0)'$

Case 3: $\beta = (3, 1.5, 2, 1.6, 2, 1.5, 1.5, 1.5)'$

The high-dimensional data simulation was carried out for sample sizes of $n = 150$ and 200. The simulation was also computed to determine the various correlations between the predictors and variances such that $\rho = 0.5, 0.8$, and 0.99 and $\sigma = 5, 10$. We simulated 100 true β s with 50 datasets for different cases as follows:

$$\begin{aligned} \text{Case 1: } \beta &= \left(\underbrace{0.5, \dots, 0.5}_{15}, \underbrace{0, \dots, 0}_{25}, \underbrace{0.7, \dots, 0.7}_{15}, \underbrace{0, \dots, 0}_{30}, \underbrace{0.7, \dots, 0.7}_{15} \right)' \\ \text{Case 2: } \beta &= \left(\underbrace{0.5, \dots, 0.5}_{15}, \underbrace{0, \dots, 0}_{25}, \underbrace{0, \dots, 0}_{15}, \underbrace{0, \dots, 0}_{30}, \underbrace{0, \dots, 0}_{15} \right)' \end{aligned}$$

$$\text{Case 3: } \beta = \left(\underbrace{0.5, \dots, 0.5}_{15}, \underbrace{0.2, \dots, 0.2}_{25}, \underbrace{0.7, \dots, 0.7}_{15}, \underbrace{0.4, \dots, 0.4}_{30}, \underbrace{0.9, \dots, 0.9}_{15} \right)'$$

Algorithm 1 Pseudo-code for the coordinate descent algorithm

Input: maximum iteration (max_iter),
tolerance (tol), hyperparameters (λ_1),
predictors (matrix X), response (vector Y)

Initialize: $\tilde{\beta}_j, j = 1, \dots, p$

While iter < max_iter do:

$\beta_j \leftarrow \tilde{\beta}_j$

While $j < p$ do:

$u_j \leftarrow \frac{1}{n} \sum_{i=1}^n x_{ij} (y_i - \bar{y}_i^{(j)}) - \hat{\beta}_j^{LS}$

$v_j \leftarrow \frac{1}{n} \sum_{i=1}^n x_{ij}^2 + 1$

$\beta_j \leftarrow S\left(\frac{u_j}{v_j}, \frac{\lambda_1}{v_j}\right)$

End While

If $(\tilde{\beta} - \beta)'(\tilde{\beta} - \beta) < tol$:

Break

Else:

$\tilde{\beta}_j \leftarrow \beta_j$

End While

Output: $\beta_j, j = 1, \dots, p$

The model performance was measured based on the mean square error. The mean square errors of the predictions from the test data were estimated as given in (15). The mean square error for estimating true coefficients β -MSE through the training model was also observed in (16). The mean and standard errors of these MSEs were recorded and reported. We also reported the median of the false positives (FP) and false negatives (FN) involved in estimating the true coefficients of the model. By false positive, we mean the coefficients that were initially zero from the simulation model specification but returned a positive real number as the estimated coefficient. False-negative coefficients return negative estimated coefficients, though the true coefficients are positive. This allows us to see how well the estimator reports the significant coefficients when it performs dimension reduction.

$$TMSE = \frac{1}{n_{test}} (\hat{y}_{test} - y_{test})' (\hat{y}_{test} - y_{test}) \quad (15)$$

$$\beta MSE = \frac{1}{R} (\beta - \hat{\beta})' (\beta - \hat{\beta}) \quad (16)$$

The simulation results of the low-dimensional and high-dimensional simulations are reported in Supplementary Materials (Sections S1 and S2). We also report the spread of the TMSEs for the test prediction and β MSE for true coefficients for the high-dimensional cases in Supplementary Materials (Section S3).

A general observation of the results of the simulated data (Supplementary Materials, Sections S1 and S2) indicates that the values of the MSEs of the estimators increase as the standard deviation (σ) and the degree of multicollinearity (ρ) increase. There are also decreases in the values of the MSEs of the estimators as the sample size increases. Although our main focus is feature reduction in high-dimensional data, we observe a prediction for our proposed estimator better than those of other estimators.

For the low-dimensional data result discussed in case 1 (Supplementary Materials, Section S1, Tables S1–S3), we observed that the KL1 and GO estimators performed better than the other estimators. The KL1 estimator performed better at predicting the test sample based on the MSE criterion, while the GO estimator had the minimum MSE for estimating

the beta coefficients. In case 2 (Supplementary Materials, Section S1, Tables S4–S6), the KL1 estimator performed better both at model prediction and beta coefficient estimation. The performance of the KL1 estimator in case 3 (Supplementary Materials, Section S1, Tables S7–S9) was the same as that in case 1. In the high-dimensional simulation result, the KL1 estimator performed better in model prediction, beta coefficient estimation and sensitivity with regard to detecting the right coefficients using FN and FP (Supplementary Materials, Section S2, Tables S10–S15).

Furthermore, though the GO estimator may compete with the KL1 estimator in the MSEs in terms of estimating the beta coefficients for the low-dimensional data, it is more appropriate to look at the FP and the FN. The FP and the FN proves that the KL1 estimator is preferable to the GO estimator where feature reduction is needed, that is, in cases 1 and 2.

4. Real-Life Application

In this section, we analyzed two datasets to evaluate the performances of the following estimators: the ridge, LASSO (L), E-net (E), GO, Liu-LASSO (LL) and the proposed (KL1) estimators. We used two datasets: the prostate cancer dataset and the asphalt binder dataset. Each dataset was analyzed by splitting them into two—training (80%) and test (20%) data. Model fitting and parameter tuning were carried out via k -fold cross classification, with $k = 10$. The tuning parameters were selected from a sequence 0 to 1. We computed the average errors over the folds for each of the tuning parameters. We chose the tuning parameters that minimized the mean square error. The coefficients from the better model were used to predict the response in the test sample, and the model performance was judged via the test mean squared error (TMSE), which is defined as follows:

$$\text{TMSE} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \left(y_{\text{test}} - x_i^T \hat{\beta} \right)^2 \quad (17)$$

Figure 1 below gives the general framework of the analytical process employed in the analysis of the real-life dataset. This process is followed through for implementing all the estimation techniques outlined in Section 2.

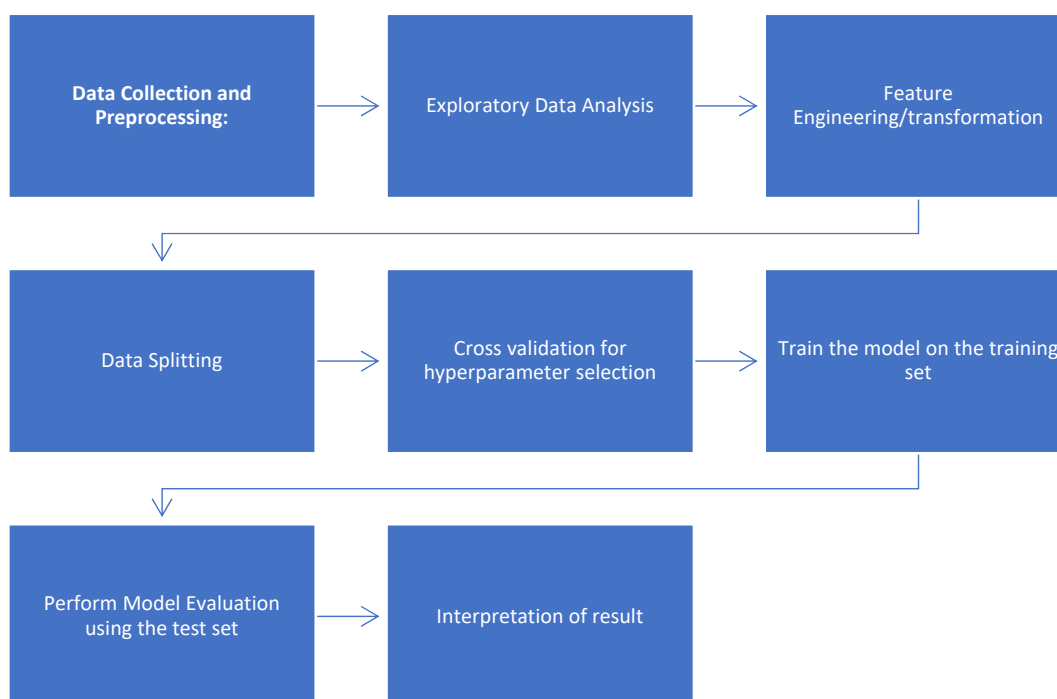


Figure 1. Roadmap for real-life data analysis.

Table 2. Regression coefficients and TMSE.

Coef.	LASSO	Liu-LASSO	Goest	E-Net	KL1	Ridge
lcavol	0.0139	0.0139	0.0156	0.0141	0.0139	0.0157
lweight	0.9721	0.9685	0.9737	0.9687	0.9641	0.9699
age	−0.0024	−0.0011	−0.0049	−0.0012	0.0000	−0.0037
lbph	0.0006	0.0006	0.0007	0.0006	0.0006	0.0007
svi	0.0000	0.0000	−0.0015	0.0000	0.0000	−0.0021
lcp	0.0000	0.0000	−0.0021	0.0000	0.0000	−0.0036
gleason	0.0046	0.0066	0.0096	0.0071	0.0091	0.0127
pgg45	−0.0035	−0.0035	−0.0036	−0.0034	−0.0034	−0.0035
TMSE	2.6167	2.6109	2.6329	2.6125	2.6038	2.6284

The results in Table 1 show that the LASSO estimator, Liu-LASSO estimator and E-Net selected six active sets. The GO estimator failed to shrink any of the variables to zero for this particular data, and the ridge estimator, as expected, only performed regularization. The proposed estimator, namely KL1, selected five active sets and possessed the lowest test mean squared error. This makes the KL1 estimator a useful fit for this dataset.

4.2. Dataset 2 (Asphalt Binder Data)

The data were employed to model the effects of some chemical compositions on surface free energy [8]. The model was defined as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{12} x_{i12} + \epsilon_i \quad i = 1, \dots, 23, \quad (19)$$

where y_i denotes surface free energy, x_{i1} denotes saturates, x_{i2} denotes aromatics, x_{i3} denotes resins, x_{i4} denotes asphaltenes, x_{i5} denotes wax, x_{i6} denotes carbon, x_{i7} denotes hydrogen, x_{i8} denotes oxygen, x_{i9} denotes nitrogen, x_{i10} denotes sulfur, x_{i11} denotes nickel and x_{i12} denotes vanadium. The summary statistics are presented in Table 3 and the correlation plot in Figure 3. The response variable is centered, and the predictors are standardized to have zero mean and unit standard deviation before model fitting. The following estimators were compared: the ridge estimator, the LASSO estimator, the Liu-LASSO estimator, the GO estimator, the elastic-net (E-net), and the KL1 estimator. The estimated TMSE and the regression coefficients are provided in Table 4. The ridge estimator did not shrink any of its coefficients, as expected. LASSO estimator, the Liu-LASSO estimator, the E-net estimator and the KL1 estimator selected ten active sets for the regression coefficients. Of all the estimators, the KL1 estimator possessed the smallest test mean squared error.

Table 3. Summary statistics for asphalt dataset.

Feature	Mean	SD	Kurtosis	Skewness
y	18.4213	4.104044	−0.11539	1.157349
x1	8.817391	2.863509	0.354511	−0.56121
x2	34.62174	5.639951	−0.58327	0.298716
x3	40.40435	7.272831	0.424932	0.025278
x4	14.67391	5.876549	−0.59483	−0.44598
x5	2.766087	1.417686	−1.54606	0.032829
x6	84.48261	1.745293	−1.21593	−0.1494

Table 3. Cont.

Feature	Mean	SD	Kurtosis	Skewness
x7	10.36478	0.332181	0.463533	0.679561
x8	0.966522	0.297911	4.368568	1.163691
x9	0.699565	0.252991	−0.24003	0.514575
x10	4.371739	2.15229	−1.17692	−0.07012
x11	71.34783	41.79562	−1.14499	0.434363
x12	243.1739	354.6355	8.152615	2.915543

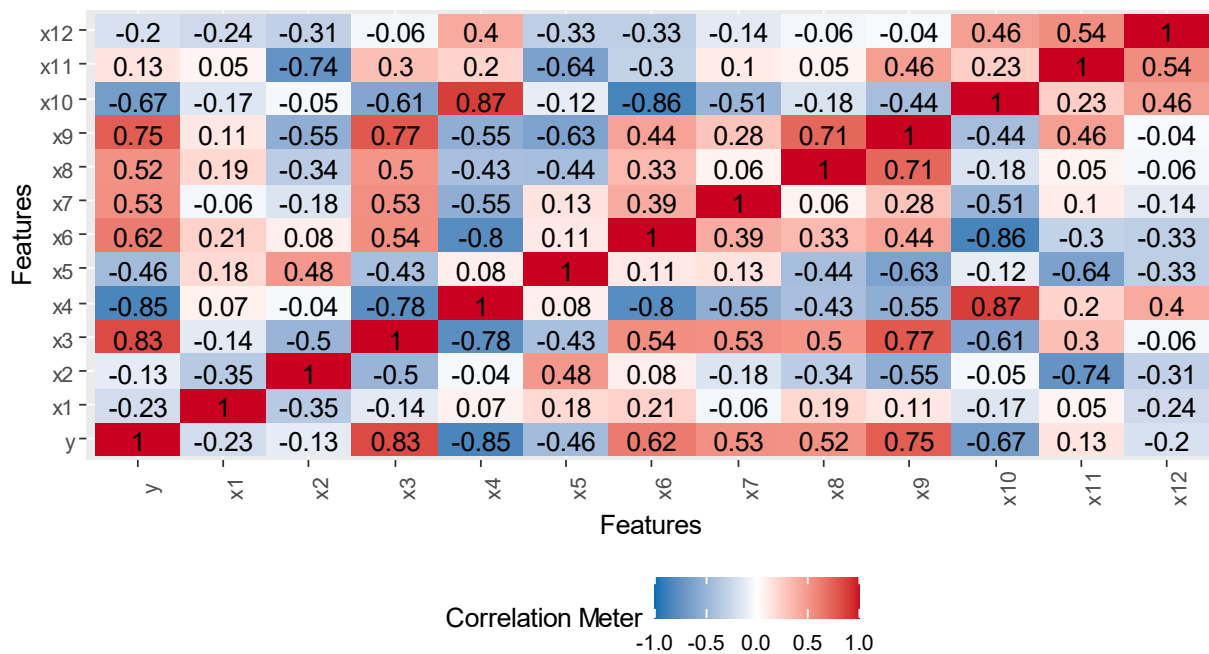


Figure 3. Correlation heatmap for asphalt dataset.

Table 4. Regression coefficients and MSE_y .

Coef.	LASSO	Liu-LASSO	Goest	E-Net	KL1	Ridge
saturates	0.0010	0.0010	0.0012	0.0010	0.0010	0.0012
aromatics	0.9323	0.9317	0.9326	0.9317	0.9310	0.9320
resins	−0.0524	−0.0523	−0.0526	−0.0523	−0.0521	−0.0525
asphaltenes	−0.0600	−0.0600	−0.0600	−0.0600	−0.0600	−0.0600
wax	−0.0337	−0.0337	−0.0338	−0.0337	−0.0336	−0.0338
carbon	−0.0070	−0.0070	−0.0109	−0.0066	−0.0071	−0.0113
hydrogen	0.0621	0.0621	0.0623	0.0621	0.0621	0.0623
oxygen	0.0134	0.0133	0.0138	0.0134	0.0133	0.0138
nitrogen	0.0000	0.0000	0.0000	0.0000	0.0000	0.0041
sulfur	0.0000	0.0000	0.0098	0.0000	0.0000	0.0123
nickel	−0.0027	−0.0027	−0.0070	−0.0025	−0.0026	−0.0084
vanadium	−0.0004	−0.0004	−0.0003	−0.0004	−0.0004	−0.0003
TMSE	9.1905	9.1849	9.2127	9.1862	9.1781	9.2165

5. Conclusions

In this study, we proposed a new penalized operator (KL1). It is a shrinkage and feature selection method that adopts the coordinate descent algorithm. The KL1 is an equivalent form of the Elastic net and the Liu-LASSO, which produces a sparse model with high prediction accuracy. The new estimator inherits the selection features of estimators such as Lasso and Elastic-net and equally possesses the stability of the ridge estimator. We applied the new estimator and the existing ones to two dataset (health data and chemical data). The application and simulation results supported the performance of KL1 and its dominance over the ridge estimator, elastic-net, GO, Liu-LASSO and LASSO. Although, the existing methods compete favorably.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/math11234795/s1>, Section S1. Quality of Estimators for Low Dimensional Simulated Data (Tables S1–S9); Section S2. Quality of Estimators for High Dimensional Simulated Data (Tables S10–S15); Section S3. Mean Square Error of the Estimated Coefficients for High Dimensional Simulated Data (Figures S1–S6).

Author Contributions: Conceptualization, A.F.L.; methodology, S.L.J. and E.T.A.; software, S.L.J. and E.T.A.; validation, J.A. and A.A.A.; investigation, J.A. and A.A.A.; writing—review & editing, J.A. and A.A.A.; supervision, S.O.; funding acquisition, J.A. and A.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Taibah University, grant number 445-9-469.

Data Availability Statement: All data generated or analyzed during this study are included in Supplementary Materials attached to this article: DATA 1_Prostate Cancer data.csv and DATA 2_asphalt Data.csv. And the code for both the simulation and the real-life applications is accessible via this link: <https://github.com/Teniola17/Teniola17/tree/main>, accessed on 25 September 2023.

Acknowledgments: The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through project number 445-9-469. Also, the authors would like to extend their appreciation to Taibah University for its supervision support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [\[CrossRef\]](#)
2. Fu, W.J. Penalized Regressions: The Bridge versus the Lasso. *J. Comput. Graph. Stat.* **1998**, *7*, 397–416.
3. Dawoud, I.; Abonazel, M.R.; Awwad, F.A.; Tag Eldin, E. A New Tobit Ridge-Type Estimator of the Censored Regression Model with Multicollinearity Problem. *Front. Appl. Math. Stat.* **2022**, *8*, 952142. [\[CrossRef\]](#)
4. Ugwuowo, F.I.; Oranye, H.E.; Arum, K.C. On the jackknife Kibria-Lukman estimator for the linear regression model. *Commun. Stat. Simul. Comput.* **2021**, *1*, 1–13. [\[CrossRef\]](#)
5. Idowu, J.I.; Oladapo, O.J.; Owolabi, A.T.; Ayinde, K.; Akinmoju, O. Combating multicollinearity: A new two-parameter approach. *Nicel Bilim. Derg.* **2023**, *5*, 90–116. [\[CrossRef\]](#)
6. Lukman, A.F.; Ayinde, K.; Binuomote, S.; Clement, O.A. Modified Ridge-Type Estimator to Combat Multicollinearity: Application to Chemical Data. *J. Chemom.* **2019**, *33*, e3125. [\[CrossRef\]](#)
7. Gujarati, D.N. *Basic Econometrics*, 4th ed.; McGraw-Hill: New York, NY, USA, 2004.
8. Arashi, M.; Asar, Y.; Yüzbaşı, B. SLASSO: A scaled LASSO for multicollinear situations. *J. Stat. Comput. Simul.* **2021**, *91*, 3170–3183. [\[CrossRef\]](#)
9. Hoerl, A.E.; Kennard, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **1970**, *12*, 55–67. [\[CrossRef\]](#)
10. Liu, K. A new class of biased estimate in linear regression. *Commun. Stat.* **1993**, *22*, 393–402.
11. Liu, K. Using Liu-type estimator to combat collinearity. *Commun. Stat. Theory Methods* **2003**, *32*, 1009–1020. [\[CrossRef\]](#)
12. Şiray, G.Ü.; Toker, S.; Özbay, N. Defining a two-parameter estimator: A mathematical programming evidence. *J. Stat. Comput. Simul.* **2021**, *91*, 2133–2152. [\[CrossRef\]](#)
13. Dawoud, I.; Abonazel, M.R.; Awwad, F.A. Modified Liu estimator to address the multicollinearity problem in regression models: A new biased estimation class. *Sci. Afr.* **2022**, *17*, e01372. [\[CrossRef\]](#)

14. Kibria, B.M.G.; Lukman, A.F. A New Ridge-Type Estimator for the Linear Regression Model: Simulations and Applications. *Scientifica* **2020**, *2020*, 9758378. [[CrossRef](#)] [[PubMed](#)]
15. Aladeitan, B.B.; Adebimpe, O.; Lukman, A.F.; Oludoun, O.; Abiodun, O. Modified Kibria-Lukman (MKL) estimator for the Poisson Regression Model: Application and simulation. *F1000Research* **2021**, *10*, 548. [[CrossRef](#)] [[PubMed](#)]
16. Dawoud, I.; Kibria, B.M.G. A new biased estimator to combat the multicollinearity of the gaussian linear regression model. *Stat. J.* **2020**, *3*, 526–541. [[CrossRef](#)]
17. Wang, H.; Li, G.; Jiang, G. Robust regression shrinkage and consistent variable selection through the LAD-lasso. *J. Bus. Econ. Stat.* **2007**, *25*, 347–355. [[CrossRef](#)]
18. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]
19. Friedman, J.; Hastie, T.; Höfling, H.; Tibshirani, R. Pathwise coordinate optimization. *Ann. Appl. Stat.* **2007**, *1*, 302–332. [[CrossRef](#)]
20. Breheny, P.; Huang, J. Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann. Appl. Stat.* **2011**, *5*, 232. [[CrossRef](#)]
21. Genç, M. A new double-regularized regression using Liu and lasso Regularization. *Comput. Stat.* **2021**, *37*, 159–227. [[CrossRef](#)]
22. Ozkale, M.R.; Kaciranlar, S. The restricted and unrestricted two-parameter estimators. *Commun. Stat. Theory Methods* **2007**, *36*, 2707–2725. [[CrossRef](#)]
23. Genc, M.; Ozkale, M.R. Usage of the GO estimator in high dimensional linear models. *Comput. Stat.* **2021**, *36*, 217–239. [[CrossRef](#)]
24. Knight, K.; Fu, W. Asymptotics for lasso-type estimators. *Ann. Stat.* **2000**, *28*, 1356–1378.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.