Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 18, No. 2, pp. 349–366, 2022

MULTIVARIATE LINEAR REGRESSION WITH LOW-RANK AND ROW-SPARSITY*

Jun Sun†, Pan Shang, Qiuyun Xu and Bingzhen Chen

Abstract: In the era of big data, multivariate linear regression (MLR) model emerges from many modern science and technology fields, such as gene expression analysis, brain neural network, finance, economics, medical imaging diagnosis, risk management and so on. In these high-dimensional data, the data often have some low-rank structure in order to catch the most material information. Meanwhile, some data sets show the block-character in predictors. Combining these two aspects, we propose a new matrix regression model in this paper. The proposed model can induce an estimator which is low-rank and sparse in the sense of row-group with the help of nuclear norm and $\|\cdot\|_{2,1}$ norm. In order to obtain an estimator, we develop a linearized alternating direction method of multipliers and prove its global convergency. Moreover, we adopt an efficient method for the tuning parameter selection. Finally, some numerical experiments are carried out to demonstrate the properties of the new proposed model and the accuracy of the proposed algorithm.

Key words: *multivariate linear regression, low rank, row sparsity, linearized alternating direction method of multipliers, polyethylene data*

Mathematics Subject Classification: *90C46, 90C06, 90C25*

1 Introduction

Multivariate linear regression (MLR) model is commonly used in bioinformatics, chemometrics, econometrics, and other quantitative fields where one is interested in predicting several responses simultaneously. The MLR model is formulated as

$$Y = XB + W, \tag{1.1}$$

where $Y = (y_1, \cdots, y_n)^{\mathrm{T}} \in \mathbb{R}^{n\times m}$ is the response matrix, $X = (x_1, \cdots, x_n)^{\mathrm{T}} \in \mathbb{R}^{n\times p}$ is the predictor matrix, $B \in \mathbb{R}^{p\times m}$ is the unknown coefficient matrix and $W = (\epsilon_1, \cdots, \epsilon_n)^{\mathrm{T}} \in \mathbb{R}^{n\times m}$ is a stochastic error matrix. Our research concentrates on the case $m \geq 2$ and focuses on dealing with the high dimensional data that the dimension of response variables and predictors are both much larger than the number of observations, i.e., $n << \min(p, m)$. For example, the reaction process of low-density polyethylene (LDPE) is controlled by the reactor temperature measurements, the wall temperature and the feed rate. The number avg.molecular weight, weight avg.molecular weight, long chain branching, short chain branching, content of vinyl group, and content of vinyledene group are used to evaluate

*This work was supported by the National Natural Science Foundation of China (No. 12071022).
†Corresponding author

polyethylene quality. In addition, the reactor temperatures measured in a certain sequence. Thus, some adjacent temperatures may exhibit common properties, in other words, they have a block structure. To solve the unknown coefficient matrix B, we propose the MLR model with low-rank and row-sparsity (MLRLRRS) below

$$\min_{B\in\mathbb{R}^{p\times m}} \quad \frac{1}{2}\|Y - XB\|_{\mathrm{F}}^2 + \lambda_1\|B\|_* + \lambda_2\|B\|_{2,1}, \tag{1.2}$$

where $\lambda_1, \lambda_2 \geq 0$ are tuning parameters. This model takes advantage of the nuclear norm and $\|\cdot\|_{2,1}$ norm, which achieves low-rank and the sparse-block simultaneously. $\|B\|_* = \sum_{i=1}^{s}\sigma_i(B)$ makes the estimator to catch the material information, where $\sigma_1(B) \geq \sigma_2(B) \geq \cdots \geq \sigma_s(B) \geq 0$ are singular values of matrix B. $\|B\|_{2,1} = \sum_{j=1}^{J}\|B_j\|_{\mathrm{F}}$ induces the model to contain important block-predictors, where $\{B_j \in \mathbb{R}^{p\times m_j}, j = 1, 2, \cdots, J\}$ is a partition of the p rows in matrix B. In fact, the combination of the nuclear norm and the $\|\cdot\|_{2,1}$ norm has been previously employed by Tang et al. [16] and Guyonin et al. [5]. These two papers focused on robust principal component analysis (RPCA) based on low-rank and block-sparse matrix decomposition. And they didn't propose the model (1.2). In particular, when $\lambda_2 = 0$, this model can be viewed as the nuclear norm regularized matrix regression which is studied by Lu et al. [8], Yuan et al. [22] and so on. When $\lambda_1 = 0$ and considering every predictor as a group, there are some relevant algorithm works which considered the row-sparsity. Similä [14] designed two algorithms to solve this model with $\|B\|_{\ell_1/\ell_2}$ as constraint. The first algorithm gives a pointwise solution, while the second one computes the entire path of solutions as a function of the constraint parameter. Peng et al. [12] established the row-wise and element-wise sparsity multivariate regression model. They proposed a method called remMap regularized multivariate regression for identifying master predictors for fitting multivariate response regression models under the high-dimension-low-sample-size setting. Under high-dimensional scaling, Obozinski et al. [11] showed that ℓ_1/ℓ_2-regularized model exhibits a threshold for the recovery of the exact row pattern with high probability over the random design and noise. When $m = 1$, model (1.1) degrades into the linear regression model. Therefore, the model (1.2) is sparse group Lasso model proposed by Friedman et al. [3]. This model achieves the within-group and among-group sparsity constraints simultaneously, choosing important groups we are interested in and important features within the selected groups. To solve this advantageous model, Li et al. [7] set up the linearized alternating direction method of multipliers (LADMM). Thus, this paper will focus on designing LADMM to solve the proposed model (1.2).

The main contributions of our paper are threefold.

1. To analyse data with low-rank property and group-wise predictors, we propose a new MLR model MLRLRRS, which achieves the within-block and among-block low rank simultaneously.

2. For the new proposed model, we design a new LADMM for solving the new model and establish its global convergency by virtue of its variational inequality problem. In addition, we use the screening rule to select the penalty parameters.

3. We make some numerical experiments to show the accuracy and efficiency of the new proposed LADMM on solving the new model.

The remaining parts of the paper are organized as follows. Some notations and known results are summarized in Section 2. In Section 3, we give a new LADMM algorithm for solving the new model. The convergence theorem of the new algorithm is established in

Section 4. In Section 5, some numerical results are reported to show the accuracy of the proposed algorithm. In the end, some conclusions and future works are made in Section 6.

2 Notations and Useful Results

For readability, we collect some notations and known results used in this paper.

Suppose $B \in \mathbb{R}^{p\times m}$, $Z \in \mathbb{R}^{p\times m}$, then the following notations are used throughout this paper. The inner product of matrix: $\langle B, Z\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} b_{ij}z_{ij}$. Then we give some notations of matrix norm. The Frobenius norm $\|B\|_{\mathrm{F}} = \sqrt{\langle B, B\rangle} = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{m} b_{ij}^2}$. The ℓ_1/ℓ_2 norm $\|B\|_{\ell_1/\ell_2} = \sum_{i=1}^{p}\sqrt{\sum_{j=1}^{m} b_{ij}^2}$. The $\|\ \|_{2,1}$ norm $\|B\|_{2,1} = \sum_{j=1}^{J}\|B_j\|_{\mathrm{F}}$, where $\{B_j \in \mathbb{R}^{p\times m_j}, j = 1, 2, \cdots, J\}$ is a partition of the p rows in matrix B. The nuclear norm $\|B\|_* = \sum_{i=1}^{s}\sigma_i(B)$, where $\sigma_1(B) \geq \sigma_2(B) \geq \cdots \geq \sigma_s(B) \geq 0$ are singular values of matrix B. Note that $\|B\|_{2,1}$ can be seen as the generalization of $\|B\|_{\ell_1/\ell_2}$. When every B_j only has one row of B, $\|B\|_{2,1}$ becomes $\|B\|_{\ell_1/\ell_2}$. As we all know $\|B\|_{\ell_1/\ell_2}$ ensures the sparsity of predictors. Thus, $\|B\|_{2,1}$ makes the model possess group sparsity. The nuclear norm $\|B\|_*$ is a convex relaxation of non-convex Rank(B). Thus, using $\|B\|_*$ is easy to design optimization methods and can obtain an estimator with low rank.

In order to make our paper easier to be understood, some conclusions on the gradient and subdifferential of matrix functions are reviewed here.

Suppose $f(B) : \mathbb{R}^{p\times m} \mapsto \mathbb{R}$ is a matrix function, then the differential of $f(B)$ is

$$\frac{\partial f}{\partial B} = \begin{pmatrix} \frac{\partial f}{\partial b_{11}} & \frac{\partial f}{\partial b_{12}} & \cdots & \frac{\partial f}{\partial b_{1m}} \\ \frac{\partial f}{\partial b_{21}} & \frac{\partial f}{\partial b_{22}} & \cdots & \frac{\partial f}{\partial b_{2m}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial b_{p1}} & \frac{\partial f}{\partial b_{p2}} & \cdots & \frac{\partial f}{\partial b_{pm}} \end{pmatrix}.$$

If $Y \in \mathbb{R}^{n\times m}$, $X \in \mathbb{R}^{n\times p}$, A, $B \in \mathbb{R}^{p\times m}$, we have

(1) $\frac{\partial(\|Y-XB\|_{\mathrm{F}}^2)}{\partial B} = -2X^{\mathrm{T}}(Y - XB)$;

(2) $\frac{\partial\langle A,B\rangle}{\partial B} = A$.

Definition 2.1. Suppose $f(B) : \mathbb{R}^{p\times m} \mapsto \mathbb{R}$ is a matrix function, then the subdifferential of $f(B)$ is

$$\partial f(B) = \left\{M \in \mathbb{R}^{p\times m} | f(Z) \geq f(B) + \langle M, Z - B\rangle,\ \forall\ Z \in \mathbb{R}^{p\times m}\right\}.$$

There are two commonly used examples.

Example 2.2. If $B \in \mathbb{R}^{n\times m}$, then

$$\frac{\partial\|B\|_{\mathrm{F}}}{\partial B} = \begin{cases} \frac{B}{\|B\|_{\mathrm{F}}}, & \text{if } B \neq 0; \\ \left\{M \in \mathbb{R}^{p\times m} | \|M\|_{\mathrm{F}} \leq 1\right\}, & \text{if } B = 0. \end{cases}$$

Example 2.3. If the singular value decomposition (SVD) of B is $B = U\Sigma V^{\mathrm{T}}$, then the subdifferential of $\|B\|_*$ (see [18] for details) is

$$\frac{\partial\|B\|_*}{\partial B} = U \cdot \mathrm{Sgn}(\Sigma) \cdot V^{\mathrm{T}},$$

where

$$\mathrm{Sgn}(\Sigma) = \begin{pmatrix} I_r & O_{r\times(m-r)} \\ O_{(p-r)\times r} & T_{(p-r)\times(m-r)} \end{pmatrix},$$

$T \in \Gamma := \{T \in \mathbb{R}^{(p-r)\times(m-r)} | \sigma_1(T) \leq 1\}$, $\sigma_1(\cdot)$ is the largest singular value.

By virtue of Example 2.2 and Example 2.3, we can easily get the solution of the following two optimization problems.

Proposition 2.4. *Given $\lambda > 0$, $r > 0$, and $C \in \mathbb{R}^{p\times m}$. If C has the singular value decomposition $C = U\Lambda V^{\mathrm{T}}$, then the minimizer*

$$\hat{B} = \underset{B\in\mathbb{R}^{p\times m}}{\operatorname{argmin}} \quad \lambda\|B\|_* + \frac{r}{2}\|B - C\|_{\mathrm{F}}^2$$

has a closed-form and

$$\hat{B} = U\left(\Lambda - \frac{\lambda}{r} I_{p\times m}\right)_+ V^{\mathrm{T}}, \tag{2.1}$$

where the entries of $I_{p\times m}$ are 0s except 1 in the main diagonal, $z_+ = \max\{0, z\}$.

To see the detailed proof of Proposition 2.4, one can refer to be found Theorem 3 in [9]. This proposition gives an explicit solution of nuclear norm regularized least squares model. In view of the expression in (2.1), the rank of $\hat{B}$ depends on λ/r. Furthermore, $\mathrm{Rank}(B)$ gets small as the tuning parameter λ gets large. Thus we can obtain an low-rank estimator. In the following, we will use Proposition 2.4 to solve the B-subproblem in iteration scheme (3.3). We can select a large parameter λ_1 to obtain a low-rank estimator.

Now, we will come into an optimization problem containing Frobenius norm.

Proposition 2.5. *For the minimization problem*

$$\min_{B\in\mathbb{R}^{p\times m}} \quad \lambda\|B\|_{\mathrm{F}} + \frac{\mu}{2}\|B - C\|_{\mathrm{F}}^2, \tag{2.2}$$

where $\lambda > 0$, $\mu > 0$, and $C \in \mathbb{R}^{p\times m}$, the optimal solution has the following closed-form

$$\bar{B} = \left(1 - \lambda/(\mu\|C\|_{\mathrm{F}})\right)_+ C. \tag{2.3}$$

Proof. Considering Example 2.2, the KKT condition of (2.2) is

$$-\mu(B - C) \in \lambda\frac{\partial\|B\|_{\mathrm{F}}}{\partial B}.$$

If $B \neq 0$, then we have $\mu(B - C) = \lambda\frac{B}{\|B\|_{\mathrm{F}}}$, i.e.,

$$\left(\lambda/\|B\|_{\mathrm{F}} + \mu\right) B = \mu C. \tag{2.4}$$

Taking Frobenius norm on both sides of the above equality, it is easy to obtain

$$\|B\|_{\mathrm{F}} = \|C\|_{\mathrm{F}} - \lambda/\mu.$$

Inserting this expression into (2.4), the expression of B is

$$B = C - \lambda/(\mu\|C\|_{\mathrm{F}})\ C = \left(1 - \lambda/(\mu\|C\|_{\mathrm{F}})\right)\ C.$$

From (2.4), we can see every element of matrix B has the same sign as the corresponding element of C. Thus, $1 - \lambda/(\mu\|C\|_{\mathrm{F}})$ must be positive if $B \neq 0$. The solution of (2.2) has the form

$$\left(1 - \lambda/(\mu\|C\|_{\mathrm{F}})\right)_+ C.$$

□

Similar to Proposition 2.4, $\bar{B}$ approaches to zero if λ grows large. In the following, Proposition 2.5 will be used to obtain the estimator of the row-blocks in iteration scheme (3.3). As shown here, we can choose large λ to make some blocks to be zero. Thus, the predictors in model (1.2) could obtain group sparsity as discussed in linear regression, such as [23].

3 Construction of Optimization Method

In this section, we apply the LADMM to solve MLRLRRS (1.2). For the subproblems whose closed-form solutions are not easy to obtain, we illustrate how to linearize them and get corresponding results. Considering these works, we can obtain an ADMM-based algorithm for solving our model.

First, let's rewrite the MLRLRRS (1.2). Suppose that all the p predictors are divided into J groups, with the number p_j in the jth group. So we use a notation $X_j \in \mathbb{R}^{n\times p_j}$ to represent the jth block of design matrix X, and the corresponding block matrix of regression coefficients is $B_j \in \mathbb{R}^{p_j\times m}$. Then the MLR model with J blocks can be defined as

$$Y = XB + W = \sum_{j=1}^{J} X_j B_j + W$$

and the optimization problem (1.2) can be rewritten as

$$\min_{B\in\mathbb{R}^{p\times m}} \quad \frac{1}{2}\|Y - XB\|_{\mathrm{F}}^2 + \lambda_1\|B\|_* + \lambda_2 \sum_{j=1}^{J} \|B_j\|_{\mathrm{F}}. \tag{3.1}$$

In order to use ADMM-based algorithm, we firstly introduce an auxiliary variable $C = (C_1, C_2, ..., C_J) \in \mathbb{R}^{p\times m}$ to the block term, where C_j $(j = 1, 2, ..., J)$ has the same dimension as B_j. Then optimization problem (3.1) can be rewritten as

$$\begin{aligned} \min_{\substack{B\in\mathbb{R}^{p\times m}\\ C\in\mathbb{R}^{p\times m}}} \quad & \frac{1}{2}\|Y - XB\|_{\mathrm{F}}^2 + \lambda_1\|B\|_* + \lambda_2 \sum_{j=1}^{J} \|C_j\|_{\mathrm{F}} \\ \text{s.t.} \quad & B = C. \end{aligned} \tag{3.2}$$

The augmented Lagrangian function of (3.2) is

$$L_\mu(B, C, A) := \frac{1}{2}\|Y - XB\|_{\mathrm{F}}^2 + \lambda_1\|B\|_* + \lambda_2 \sum_{j=1}^{J} \|C_j\|_{\mathrm{F}} - \langle A, B - C\rangle + \frac{\mu}{2}\|B - C\|_{\mathrm{F}}^2,$$

and the iterative scheme of ADMM is

$$\begin{cases} B^k = \underset{B\in\mathbb{R}^{p\times m}}{\operatorname{argmin}} \ L_\mu(B, C^{k-1}, A^{k-1}), \\ C^k = \underset{C\in\mathbb{R}^{p\times m}}{\operatorname{argmin}} \ L_\mu(B^k, C, A^{k-1}), \\ A^k = A^{k-1} - \tau\mu(B^k - C^k). \end{cases} \tag{3.3}$$

Now, we discuss the solutions of the subproblems in (3.3). The B-subproblem can be reformed as

$$\begin{aligned} B^k &= \underset{B\in\mathbb{R}^{p\times m}}{\operatorname{argmin}} \quad \frac{1}{2}\|Y - XB\|_{\mathrm{F}}^2 + \lambda_1\|B\|_* - \langle A^{k-1}, B - C^{k-1}\rangle + \frac{\mu}{2}\|B - C^{k-1}\|_{\mathrm{F}}^2 \\ &= \underset{B\in\mathbb{R}^{p\times m}}{\operatorname{argmin}} \quad \frac{1}{2}\|Y - XB\|_{\mathrm{F}}^2 + \lambda_1\|B\|_* + \frac{\mu}{2}\|B - C^{k-1} - A^{k-1}/\mu\|_{\mathrm{F}}^2 \\ &= \underset{B\in\mathbb{R}^{p\times m}}{\operatorname{argmin}} \quad \lambda_1\|B\|_* + \frac{1}{2}\|\widehat{X}B - \widehat{Y}_k\|_{\mathrm{F}}^2, \end{aligned} \tag{3.4}$$

where $\widehat{X} = (X^{\mathrm{T}}, \sqrt{\mu}I_{p\times p})^{\mathrm{T}}$, $I_{p\times p} \in \mathbb{R}^{p\times p}$ is an identity matrix and $\widehat{Y}_k = (Y^{\mathrm{T}}, \sqrt{\mu}(C^{k-1} + A^{k-1}/\mu)^{\mathrm{T}})^{\mathrm{T}}$. If $\widehat{X}^{\mathrm{T}}\widehat{X}$ is not a diagonal matrix, this subproblem does not have a closed-form solution. To deal with this problem, we can linearize the quadratic term $\frac{1}{2}\|\widehat{X}B - \widehat{Y}_k\|_{\mathrm{F}}^2$ in (3.4) and replace it by

$$\frac{1}{2}\|\widehat{X}B^{k-1} - \widehat{Y}_k\|_{\mathrm{F}}^2 + \langle \widehat{X}^{\mathrm{T}}(\widehat{X}B^{k-1} - \widehat{Y}_k), B - B^{k-1}\rangle + \frac{\nu}{2}\|B - B^{k-1}\|_{\mathrm{F}}^2, \tag{3.5}$$

where ν is a positive parameter and its value defined in Theorem 4.5. Omitting the constant term, we only need to consider the optimization problem

$$\begin{aligned} B^k &= \underset{B\in\mathbb{R}^{p\times m}}{\operatorname{argmin}} \quad \lambda_1\|B\|_* + \langle \widehat{X}^{\mathrm{T}}(\widehat{X}B^{k-1} - \widehat{Y}_k), B - B^{k-1}\rangle + \frac{\nu}{2}\|B - B^{k-1}\|_{\mathrm{F}}^2 \\ &= \underset{B\in\mathbb{R}^{p\times m}}{\operatorname{argmin}} \quad \lambda_1\|B\|_* + \frac{\nu}{2}\|B - B^{k-1} + \widehat{X}^{\mathrm{T}}(\widehat{X}B^{k-1} - \widehat{Y}_k)/\nu\|_{\mathrm{F}}^2. \end{aligned} \tag{3.6}$$

According to Proposition 2.4, if $B^{k-1} - \widehat{X}^{\mathrm{T}}(\widehat{X}B^{k-1} - \widehat{Y}_k)/\nu$ has the following SVD

$$B^{k-1} - \widehat{X}^{\mathrm{T}}(\widehat{X}B^{k-1} - \widehat{Y}_k)/\nu = U\Lambda V^{\mathrm{T}},$$

then the minimizer of (3.6) is

$$B^k = U\left(\Lambda - \frac{\lambda_1}{\nu}I_{p\times m}\right)_+ V^{\mathrm{T}}. \tag{3.7}$$

For the C-subproblem in (3.3), it can be written as

$$\begin{aligned} C^k &= \underset{C\in\mathbb{R}^{p\times m}}{\operatorname{argmin}} \quad \lambda_2\sum_{j=1}^{J}\|C_j\|_{\mathrm{F}} - \langle A^{k-1}, B^k - C\rangle + \frac{\mu}{2}\|B^k - C\|_{\mathrm{F}}^2. \\ &= \underset{C\in\mathbb{R}^{p\times m}}{\operatorname{argmin}} \quad \lambda_2\sum_{j=1}^{J}\|C_j\|_{\mathrm{F}} + \frac{\mu}{2}\|B^k - C - A^{k-1}/\mu\|_{\mathrm{F}}^2 \end{aligned}$$

This optimization problem can be separated to J subproblems and the jth subproblem has the following solution

$$\begin{aligned} C_j^k &= \underset{C_j\in\mathbb{R}^{p_j\times m}}{\operatorname{argmin}} \quad \lambda_2\|C_j\|_{\mathrm{F}} + \frac{\mu}{2}\|B_j^k - C_j - A_j^{k-1}/\mu\|_{\mathrm{F}}^2 \\ &= \left(1 - \frac{\lambda_2}{\mu\|B_j^k - A_j^{k-1}/\mu\|_{\mathrm{F}}}\right)_+ \left(B_j^k - A_j^{k-1}/\mu\right), \end{aligned} \tag{3.8}$$

which is from the result of Proposition 2.5.

Based on the above discussion, the iterative algorithm for MLRLRRS is described in Algorithm 1.

Algorithm 1: LADMM for MLRLRRS (3.1).

Input: X,Y and tol. Choose $\lambda_1 > 0$, $\lambda_2 > 0$, $\mu > 0$ and $\nu > \mu + \rho(X^{\mathrm{T}}X)$, where $\rho(\cdot)$ is the spectral radius.

Choose the start point (B^0, C^0, A^0).

1 **for** $k = 1, 2, \cdots$ **do**
2 Compute B^k by (3.7);
3 Compute C^k by (3.8);
4 Updata A^k via $A^k = A^{k-1} - \tau\mu(B^k - C^k)$;
5 Check stopping criterion;
6 **end**

4 Convergency

In this section, we present the convergence result for the proposed LADMM in the last section. We mainly establish the global convergence theorem and study the convergence rate of LADMM. The procedure of the convergence analysis follows the idea in [7].

In order to make convergence analysis easily, we write the problem (3.2) as a variational form. In order to do this, we give the Lagrange function of (3.2)

$$\frac{1}{2}\|Y - XB\|_{\mathrm{F}}^2 + \lambda_1\|B\|_* + \lambda_2\sum_{j=1}^{J}\|C_j\|_{\mathrm{F}} - \langle A, B - C\rangle, \tag{4.1}$$

where $A \in \mathbb{R}^{p\times m}$ is the Lagrangian matrix multiplier. Because problem (3.2) is a convex problem with linear constraints and $(0, 0)$ is a relatively interior point of its feasible area, there is a Karush-Kuhn-Tucker (KKT) point which is comprised by solutions of problem (3.2) and its dual form. Therefore, minimizing problem (3.2) is the same as finding a point $(B^{*\mathrm{T}}, C^{*\mathrm{T}}, A^{*\mathrm{T}})^{\mathrm{T}} \in \Omega := \mathbb{R}^{p\times m} \times \mathbb{R}^{p\times m} \times \mathbb{R}^{p\times m}$ satisfying

$$\begin{cases} 0 = \lambda_1 f(B^*) + X^{\mathrm{T}}(XB^* - Y) - A^*, \\ 0 = \lambda_2 g(C^*) + A^*, \\ 0 = B^* - C^*. \end{cases} \tag{4.2}$$

where $f(B^*) \in \partial(\|B^*\|_*)$ and $g(C^*) \in \partial\left(\sum_{j=1}^{J}\|C_j^*\|_{\mathrm{F}}\right)$. Denote Ω^* as the collection of the points in Ω satisfying (4.2). Let $\omega^* = (B^{*\mathrm{T}}, C^{*\mathrm{T}}, A^{*\mathrm{T}})^{\mathrm{T}}$ be any point in Ω^*, then (3.2) can be transformed to the following variational inequality (VI)

$$VI(\Omega, F): \ \langle \omega - \omega^*, F(\omega^*)\rangle \geq 0, \ \forall\ \omega \in \Omega,$$

where

$$\omega = \begin{pmatrix} B \\ C \\ A \end{pmatrix}, \ F(\omega) = \begin{pmatrix} \lambda_1 f(B) + X^{\mathrm{T}}(XB - Y) - A \\ \lambda_2 g(C) + A \\ B - C \end{pmatrix}. \tag{4.3}$$

With the help of this transformation, the estimator in every iteration of the proposed algorithm can be expressed as the following VI problem.

Lemma 4.1. *Denote* $M = (-\mu I_p \ \mu I_p \ O_p)^{\mathrm{T}}$, $G = \begin{pmatrix} (\nu-\mu)I_p - X^{\mathrm{T}}X & O_p & O_p \\ O_p & I_p & O_p \\ O_p & O_p & \frac{1}{\mu}I_p \end{pmatrix}$. *Then the solution sequence* $\{\omega^{k-1}\}$ *satisfies*

$$\langle \omega' - \omega^k, F(\omega^k) + M(C^{k-1} - C^k) - G(\omega^{k-1} - \omega^k)\rangle \geq 0, \ \forall\ \omega' \in \Omega. \tag{4.4}$$

Proof. Consider the first order optimality conditions of the minimization problems (3.6) and (3.8), we can see that the iterative scheme in Algorithm 1 is equivalent to find $\omega^k = (B^{k^{\mathrm{T}}}, C^{k^{\mathrm{T}}}, A^{k^{\mathrm{T}}})^{\mathrm{T}} \in \Omega$, $f(B^k) \in \partial(\|B^k\|_*)$ and $g(C^k) \in \partial(\sum_{j=1}^J \|C_j^k\|_{\mathrm{F}})$ such that

$$\begin{cases} 0 = \lambda_1 f(B^k) + \widehat{X}^{\mathrm{T}}(\widehat{X}B^{k-1} - \widehat{Y}_k) + \nu(B^k - B^{k-1}), \\ 0 = \lambda_2 g(C^k) + \mu(C^k - B^k + A^{k-1}/\mu), \\ 0 = B^k - C^k - (A^{k-1} - A^k)/\mu, \end{cases} \tag{4.5}$$

where

$$\widehat{X}^{\mathrm{T}}(\widehat{X}B^{k-1} - \widehat{Y}_k) = (X^{\mathrm{T}}X + \mu I_{p\times p})B^{k-1} - (X^{\mathrm{T}}Y + A^{k-1} + \mu C^{k-1}). \tag{4.6}$$

Inserting (4.6) into (4.5) and considering notation M and G, problem (4.5) can be rewritten as the VI problem (4.4). □

In view of Lemma 4.1, we can easily obtain the following lemma.

Lemma 4.2. *For any* $\omega^* \in \Omega^*$, *the solution sequence* $\{\omega^{k-1}\}$ *satisfies*

$$\begin{aligned} &\langle \omega^{k-1} - \omega^*, G(\omega^{k-1} - \omega^k)\rangle \\ &\qquad \geq \langle \omega^{k-1} - \omega^k, G(\omega^{k-1} - \omega^k)\rangle - \langle A^{k-1} - A^k, C^{k-1} - C^k\rangle. \end{aligned} \tag{4.7}$$

Proof. Since (4.4) holds for any $\omega' \in \Omega$, let $\omega' = \omega^* \in \Omega^*$, then we have

$$\langle \omega^* - \omega^k, F(\omega^k) + M(C^{k-1} - C^k) - G(\omega^{k-1} - \omega^k)\rangle \geq 0. \tag{4.8}$$

Moreover, we have $B^* - C^* = 0$. So (4.8) becomes to

$$\begin{aligned} \langle \omega^k - \omega^*, G(\omega^{k-1} - \omega^k)\rangle &\geq \langle \omega^k - \omega^*, F(\omega^k)\rangle - \mu\langle B^k - C^k, C^{k-1} - C^k\rangle \\ &= \langle \omega^k - \omega^*, F(\omega^k)\rangle - \langle A^{k-1} - A^k, C^{k-1} - C^k\rangle, \end{aligned} \tag{4.9}$$

where $A^{k-1} - A^k = \mu(B^k - C^k)$.

Note that $F(\omega)$ defined in (4.3) is monotone. It follows that

$$\langle \omega^k - \omega^*, F(\omega^k) - F(\omega^*)\rangle \geq 0,$$

and

$$\langle \omega^k - \omega^*, F(\omega^k)\rangle \geq \langle \omega^k - \omega^*, F(\omega^*)\rangle \geq 0. \tag{4.10}$$

Substituting $(\omega^k - \omega^{k-1}) - (\omega^{k-1} - \omega^*)$ for $\omega^k - \omega^*$ in inequality (4.9) and considering (4.10), the desired result is proved. □

Now, combining Lemma 4.1 and Lemma 4.2, we can show that the sequence $\{\omega^{k-1}\}$ generated by LADMM is contractive with respect to the solution set Ω^*.

Corollary 4.3. *For any $\omega^* \in \Omega^*$, we obtain*

$$\|\omega^k - \omega^*\|_G^2 \leq \|\omega^{k-1} - \omega^*\|_G^2 - \|\omega^{k-1} - \omega^k\|_G^2. \tag{4.11}$$

Proof. Considering (4.7), $\forall\ \omega^* \in \Omega^*$, it is easy to obtain

$$\begin{aligned}\|\omega^k - \omega^*\|_G^2 &= \|\omega^{k-1} - \omega^*\|_G^2 + \|\omega^{k-1} - \omega^k\|_G^2 - 2\langle \omega^{k-1} - \omega^*, G(\omega^{k-1} - \omega^k)\rangle \\ &\leq \|\omega^{k-1} - \omega^*\|_G^2 - \|\omega^{k-1} - \omega^k\|_G^2 + 2\langle A^{k-1} - A^k, C^{k-1} - C^k\rangle. \end{aligned}\tag{4.12}$$

According to (4.2), we have $\lambda_2 g(C^{k-1}) + A^{k-1} = 0$. So the following results are true

$$\begin{aligned}\langle C^{k-1} - C^k, \lambda_2 g(C^k) + A^k\rangle \geq 0,\\ \langle C^k - C^{k-1}, \lambda_2 g(C^{k-1}) + A^{k-1}\rangle \geq 0.\end{aligned}$$

Then, we have

$$\langle A^{k-1} - A^k, C^{k-1} - C^k\rangle \leq \lambda_2 \langle C^{k-1} - C^k, g(C^k) - g(C^{k-1})\rangle \leq 0. \tag{4.13}$$

Combining (4.13) and (4.12), we can obtain the desired conclusion (4.11). □

Furthermore, considering Lemma 4.1, Lemma 4.2 and Corollary 4.3, we can easily get the following three conclusions.

Corollary 4.4. (i) $\lim_{k\to\infty} \|\omega^{k-1} - \omega^k\|_G = 0$.

(ii) *The solution sequence $\{\omega^{k-1}\}$ is bounded.*

(iii) *The sequence $\{\|\omega^{k-1} - \omega^*\|_G\}$ decrease for any $\omega^* \in \Omega^*$.*

Now we can give the convergence theorem for Algorithm 1.

Theorem 4.5. *Given $\mu > 0$, $\nu > \mu + \rho(X^{\mathrm{T}}X)$ and $(B^{0\mathrm{T}}, C^{0\mathrm{T}}, A^{0\mathrm{T}})^{\mathrm{T}} \in \Omega$. Denote $\{\omega^{k-1}\}$ as the solution sequence generated by LADMM Algorithm 1. Then, $\{\omega^{k-1}\}$ converges to the solution of model* (3.2).

Proof. The conclusion (i) in Corollary 4.4 implies that

$$\lim_{k\to\infty} \|B^{k-1} - B^k\|_G = 0, \lim_{k\to\infty} \|C^{k-1} - C^k\|_G = 0, \lim_{k\to\infty} \|A^{k-1} - A^k\|_G = 0.$$

Conclusion (ii) in Corollary 4.4 shows that $\{\omega^{k-1}\}$ converge. Assume that this point of convergence is $\omega^\infty = (B^{\infty\mathrm{T}}, C^{\infty\mathrm{T}}, A^{\infty\mathrm{T}})^{\mathrm{T}}$. If $\{\omega^{k_j}\}$ is a subsequence of $\{\omega^{k-1}\}$ converging to ω^∞, then we have

$$B^{k_j} \to B^\infty,\quad C^{k_j} \to C^\infty,\quad A^{k_j} \to A^\infty,$$

and

$$\begin{cases} \lim_{j\to\infty} \|B^{k_j} - B^{k_j+1}\|_G = 0,\ \lim_{j\to\infty} \|C^{k_j} - C^{k_j+1}\|_G = 0, \\ \lim_{j\to\infty} \|A^{k_j} - A^{k_j+1}\|_G = 0. \end{cases} \tag{4.14}$$

Now, we only need to demonstrate that ω^∞ satisfies the KKT conditions (4.2). In view of (4.4) and (4.14), it is easy to obtain

$$\lim_{j\to\infty} \langle \omega' - \omega^{k_j}, F(\omega^{k_j})\rangle \geq 0,\ \forall\ \omega' \in \Omega.$$

It follows that

$$\langle \omega' - \omega^\infty, F(\omega^\infty)\rangle \geq 0,\ \forall\ \omega' \in \Omega.$$

So, the cluster point ω^∞ satisfies (4.2), i.e., $\omega^\infty \in \Omega^*$. Note that the conclusion (iii) in Corollary 4.4 means that for any $k \geq 0$,

$$\|\omega^k - \omega^\infty\|_G \leq \|\omega^{k-1} - \omega^\infty\|_G.$$

Thus, ω^∞ is the unique point of convergence. That is, $\{\omega^{k-1}\}$ converges to ω^∞. The desired result is obtained. □

Remark 4.6. As a direct result of Theorem 4.1 in [6], the iteration complexity for the proposed Algorithm 1 is $O(1/k)$. Here the iteration complexity is in the sense of average. The detailed procedure can be referred to [6].

Now, let's introduce the stopping criteria for Algorithm 1. Considering the kth iteration of Algorithm 1, we obtain

$$\begin{gathered}\partial(\lambda_1\|B^k\|_*) \ni -\nu(B^k - B^{k-1}) - \widehat{X}^{\mathrm{T}}(\widehat{X}B^{k-1} - \widehat{Y}_k),\\ \partial(\lambda_2\|C^k\|_{2,1}) \ni \mu(B^k - C^k) - A^{k-1}.\end{gathered}$$

It follows that

$$D_k \in \partial(\lambda_1\|B^k\|_* + \lambda_2\|C^k\|_{2,1} + \frac{1}{2}\|Y - XB^k\|_{\mathrm{F}}^2),$$

where

$$\begin{aligned}D_k &:= -\nu(B^k - B^{k-1}) - \widehat{X}^{\mathrm{T}}(\widehat{X}B^{k-1} - \widehat{Y}_k) + \partial(\lambda_2\|C^k\|_{2,1}) + X^{\mathrm{T}}(XB^k - Y)\\ &= (\mu - \nu)(B^k - B^{k-1}) - \mu(C^k - C^{k-1}).\end{aligned}$$

Note that, the KKT system for problem (3.2) is given by:

$$B - C = 0,\quad X^{\mathrm{T}}(XB - Y) + \partial(\lambda_1\|B\|_*) + \partial(\lambda_2\|C\|_{2,1}) = 0.$$

Thus the following stopping criterion is employed

$$\max\left\{\|B^k - C^k\|_{\mathrm{F}},\ \frac{\|D_k\|_{\mathrm{F}}}{\max\{1, \|Y\|_{\mathrm{F}}\}}\right\} < 10^{-4}.$$

5 Numerical Experiments

In this section, we report some numerical results of Algorithm 1 when solving problem (1.2). We have implemented the LADMM algorithm in MATLAB(Version R2015a). All runs are performed on a desktop with Intel(R) Core(TM) i5-8500 CPU (3.00 GHz) and 8 GB RAM. For the tuning parameters λ_1 and λ_2, we choose them with the help of tuning parameter selection works [2, 13]. As in classical ADMM algorithm, the parameters is set as $\mu = (1 + \sqrt{5})/2$. And as discussed in Lemma 4.1, we choose $\nu = 1.01(\mu + \rho(X^{\mathrm{T}}X))$.

5.1 Simulation studies

The random dat used in our experiments is generated as follows. We construct the predictor matrix X by generating its rows x_i as i.i.d. realizations from a multivariate normal distribution $\mathcal{N}(0,\Sigma)$, with $\Sigma = \left(\rho^{|j-k|}\right)_{p\times p}, 0 < \rho \le 1, 1 \le j,k \le p$. The true coefficient matrix $B^* = bB_1B_2^{\mathrm{T}}$, with $b > 0, B_1$ is a $p \times r$ matrix and B_2 is a $m \times r$ matrix. All entries in B_1 and B_2 are i.i.d. $\mathcal{N}(0,1)$. Each row in $Y = (y_1, y_2, \cdots, y_n)^{\mathrm{T}}$ is then generated as $y_i = B^{\mathrm{T}}x_i + \epsilon_i, 1 \le i \le n$, where ϵ_i denoting the i-th row of the noise matrix $W = (w_{ij})_{n\times m}$ and w_{ij}s are i.i.d. $\mathcal{N}(0,1)$.

For every fixed (n,m,p,r,ρ,b), we generate the data 100 times. Then we report the estimation MSE $\|\hat{B} - B^*\|_{\mathrm{F}}^2/(pm)$ (Est-Err) and prediction mean-squared-errors MSE $\|Y - X\hat{B}\|_{\mathrm{F}}^2/(mn_{\mathrm{v}})$ (Pre-Err) of validation data. We also presented the median rank of the estimator and the average CPU times (in seconds).

The following two cases are considered.

• **Experiment 1**

In this simulation, we set $n = 200, m = 600, p = 1000$, and the true rank of coefficient matrix $r \in \{2,5\}$, the correlation coefficient $\rho = 0.1, 0.5, 0.9$ and signal strength $b = 0.1, 0.3$. All combinations of correlation and signal strength are covered in the simulations. The results of the numerical study are summarized in Table 1.

• **Experiment 2**

In this case, we generated $n = 100$ observations with $p = 250$ predictors and $m = 250$ responses. To indicate the group structure of a predictors, all predictors are designed with ten blocks, say $\{X_1, X_2, \cdots, X_{10}\}$. Each block obeys Gaussian distribution $\mathcal{N}(0,\Sigma)$. For the block $X_j, j = 1, \cdots, 10$, we select $\Sigma_j = j \cdot (\Sigma_{kl})_{p\times p}$ and Σ_{kl} satisfies $\Sigma_{kl} = \begin{cases} \rho, & k \ne l, \\ 1, & k = l, \end{cases}$ where $0 < \rho < 1$. Considering the sparsity of the blocks, we select the second 125 rows of B^* to be zero. To ensure the low-rank property of B^*, we generate it as $B^* = b\begin{pmatrix} B_1 \\ O \end{pmatrix} * B_2^{\mathrm{T}}$, with $b > 0, B_1 \in \mathbb{R}^{125\times r}$ and $B_2 \in \mathbb{R}^{250\times r}$. All entries in B_1 and B_2 are i.i.d. $\mathcal{N}(0,1)$. Finally, each row in $Y = (y_1, y_2, \cdots, y_n)^{\mathrm{T}}$ is then generated as $y_i = B^{*\mathrm{T}}x_i + \epsilon_i, 1 \le i \le n$, where ϵ_i denoting the i-th row of the noise matrix $W = (w_{ij})_{n\times m}$ and w_{ij}s are i.i.d. $\mathcal{N}(0,1)$. The values of (r,b,ρ) are set as in **Experiment 1**. The numerical results are demonstrated in Table 2 and Table 3. And the bold results are the best one among the compared methods.

In Table 2 and Table 3, APGL is the accelerated proximal gradient (APG) algorithm with line-search designed in [17]. LADM is the linearized alternating direction method proposed in [19]. LADMM represents the linearized alternating direction method of multipliers for solving sparse group Lasso [7]. Note that, APGL and LADM are used to solve nuclear norm penalized multivariate regression model, i.e., model (1.2) with $\lambda_2 == 0$. LADMM is designed for solving sparse group Lasso in the scenario of multiple linear regression. Ignoring the structural characteristics of MLR, MLR can be transformed to multiple linear regression. Thus it can be solved by LADMM proposed in [7].

From Table 2 and Table 3, we can see our method performs better than APGL, LADM and LADMM in terms of Est-Err and Pre-Err. The reason is that the experiments designed here possess a block structure. Our model is specifically designed for data with this structural characteristic. But APGL and LADM solve optimization problem with only nuclear norm regularizer and they don't consider block structure. When it comes to CPU time, our method seems to be the worse one. This can be explained as follows. In all the compared methods, the main work is SVD. Since different ways of computing SVD give different CPU times, here we aim to show how these methods perform. Thus, we used the efficient Matlab

Table 1: The performance of Algorithm 1 for **Experiment 1**.

b		$r=2$			$r=5$		
		$\rho=0.1$	$\rho=0.5$	$\rho=0.9$	$\rho=0.1$	$\rho=0.5$	$\rho=0.9$
0.1	Est–Err	1.17e-1	1.14e-1	1.23e-1	2.14e-1	2.07e-1	2.02e-1
	Pre–Err	1.60e-1	1.47e-1	1.56e-1	2.31e-1	2.22e-1	2.21e-1
	Rank	2	2	2	5	5	5
	CPU	7.50	9.30	9.50	7.88	8.33	9.29
0.3	Est–Err	2.82e-1	2.81e-1	2.82e-1	5.70e-1	6.13e-1	6.02e-1
	Pre–Err	4.70e-1	4.62e-1	5.05e-1	6.60e-1	7.07e-1	7.07e-1
	Rank	2	2	2	5	5	5
	CPU	8.43	7.57	6.10	8.54	6.07	8.63

Table 2: The performance of Algorithm 1 for **Experiment 2** with $r=2$.

b		$\rho=0.5$			
		APGL	LADM	LADMM	Algorithm 1
0.05	Est-Err	5.03e-2(4.91e-4)	4.75e-2(4.82e-4)	4.23e-2(4.63e-4)	**3.42e-2**(8.49e-4)
	Pre-Err	1.25(8.54e-3)	1.06(6.22e-3)	8.21e-1(4.61e-3)	**7.11e-1**(3.01e-3)
	Rank	2	2	–	2
	CPU	0.06	**0.04**	0.79	0.64
0.06	Est-Err	6.23e-2(4.96-4)	4.87e-2(4.81e-4)	4.47e-2(2.39e-4)	**3.31e-2**(6.94e-4)
	Pre-Err	1.39(6.18e-3)	1.09(6.31e-3)	8.65e-1(5.21e-3)	**7.35e-1**(1.03e-2)
	Rank	2	2	–	2
	CPU	0.07	**0.04**	0.81	0.65
0.07	Est-Err	6.62e-2(2.15-3)	5.06e-2(5.03e-4)	4.63e-2(2.67e-4)	**3.29e-2**(1.59e-4)
	Pre-Err	1.43(1.42e-2)	1.14(6.98e-3)	9.01e-1(5.53e-3)	**7.82e-1**(1.17e-3)
	Rank	2	2	–	2
	CPU	0.06	**0.04**	0.76	0.64

Mex interface mexsvd which can be found in the package APGL. In addition, some other accelerated techniques such as truncation technique, continuation technique are employed in APGL and LADM. Here, we mainly focus on designing an algorithm to solve optimization problem extracted from data having low-rank and row-block structure. This also shows that we must design particular algorithms to analysis the data sets with special properties. For the measurement Rank, we can see that our model gives the same median rank as APGL and LADM. However, the LADMM don't consider the low-rank property. Thus, it can't achieve an estimator with low rank. Combining the above discussion, the proposed model not only gives a more accurate estimator but also a low-rank estimator.

5.2 Real data analysis

In this subsection, we will apply our model to the polyethylene data set [15, 1]. This data set focuses on studying the reaction process of low-density polyethylene (LDPE). The reaction process is controlled by 20 reactor temperature measurements T_1—T_{20}, the wall temperature and the feed rate. The number avg.molecular weight, weight avg.molecular weight, long

Table 3: The performance of Algorithm 1 for **Experiment 2** with $r = 5$.

b		$\rho = 0.5$			
		APGL	LADM	LADMM	Algorithm 1
0.05	Est-Err	7.59e-2(9.81e-4)	6.51e-2(7.36e-4)	5.67e-2(4.85e-4)	**4.61e-2**(4.77e-4)
	Pre-Err	1.57(1.58e-2)	1.31(8.84e-3)	1.01(6.67e-3)	**8.78e-1**(6.93e-3)
	Rank	5	5	–	5
	CPU	0.07	**0.05**	0.17	0.68
0.06	Est-Err	9.29e-2(1.69e-4)	7.68e-2(5.87e-4)	6.25e-2(5.23e-4)	**5.31e-2**(4.69e-4)
	Pre-Err	1.79(8.23e-3)	1.53(6.57e-3)	1.18(7.47e-3)	**9.85e-1**(6.57e-3)
	Rank	5	5	–	5
	CPU	0.08	**0.05**	0.20	0.70
0.07	Est-Err	1.07e-1(3.35e-3)	8.71e-2(1.25e-3)	7.11e-2(3.22e-3)	**6.12e-2**(3.89e-3)
	Pre-Err	2.02(2.53e-2)	1.65(1.89e-2)	1.26(1.24e-2)	**1.03**(1.07e-2)
	Rank	5	5	–	5
	CPU	0.07	**0.05**	0.17	0.69

chain branching, short chain branching, content of vinyl group, and content of vinyledene group are used to evaluate polyethylene quality. Thus, there are $n = 56$ observations with $p = 22$ and $m = 6$ in this data set. As in [10], we applied log transformation to the responses, and then standardized them. In addition, the 20 reactor temperatures measured in a certain sequence. Thus, some adjacent temperatures may exhibit common properties, in other words they have a block structure. From the the correlation coefficient matrix of predictors displayed in Table 4 and Table 5, we can see that the two adjacent temperatures are in high correlation except T_7 and T_8, T_{11} and T_{12}, T_{12} and T_{13}, T_{14} and T_{15}. This implies that, they are in one group. Therefore, we should consider the block structure in predictors when analyzing data.

Consider that this data set has no test set and contains few samples. We split this data set equally into two parts: training set $(X_{\text{training}}, Y_{\text{training}})$ and test set $(X_{\text{test}}, Y_{\text{test}})$. Then, we use training data $(X_{\text{training}}, Y_{\text{training}})$ and regression the model to obtain an estimator $\hat{B}$. After that, we use $\hat{B}$ and the test data $(X_{\text{test}}, Y_{\text{test}})$ to measure the mean squared prediction error (MSPE)

$$\text{MSPE} = \mathbb{E}\left\{\|Y_{\text{test}} - X_{\text{test}}\hat{B}\|^2\right\}.$$

To avoid the contingency of division, we split the data set 100 times. Then we compute the mean value and standard deviation of MSPE. The numerical results are shown in Table 6.

To show the performance clearly, we also draw the boxplot of the 100 repeated results in Fig. 1.

In view of Table 6 and Fig. 1, the MLRLRRS performs an excellent prediction. The good performance of model (1.2) can also be explained by the fact that the predictors show some block/group structures. In view of the correlation coefficient matrix of predictors displayed in Table 4 and Table 5, some block structure are shown, such as $\{T_1, T_2, T_3, T_4, T_5\}$, $\{T_6, T_7\}$, $\{T_8, T_9\}$, $\{T_{10}, T_{11}\}$, $\{T_{13}, T_{14}\}$,$\{T_{15}, T_{16}\}$, $\{T_{17}, T_{18}, T_{19}, T_{20}\}$. Our proposed model is designed specially for the data set with block structure among predictors.

Form the above analysis, we can summarize how to apply the new model to real data. Given a data set in reality, the procedure of using our model to make analysis is: first we compute the correlation coefficient matrix of predictors. Then we pick up high correlated

Table 4: The correlation coefficient matrix of predictors.

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
T_1	1.00	1.00	0.98	0.95	0.94	0.84	0.79	0.65	0.53	0.38
T_2	**1.00**	1.00	0.99	0.96	0.95	0.85	0.80	0.66	0.55	0.40
T_3	**0.98**	**0.99**	1.00	0.98	0.95	0.83	0.77	0.61	0.51	0.38
T_4	**0.95**	**0.96**	**0.98**	1.00	0.96	0.82	0.74	0.55	0.46	0.35
T_5	**0.94**	**0.95**	**0.95**	**0.96**	1.00	0.90	0.81	0.59	0.49	0.37
T_6	0.84	0.85	0.83	0.82	**0.90**	1.00	0.90	0.62	0.51	0.38
T_7	0.79	0.80	0.77	0.74	0.81	**0.90**	1.00	0.77	0.66	0.51
T_8	0.65	0.66	0.61	0.55	0.59	0.62	*0.77*	1.00	0.92	0.74
T_9	0.53	0.55	0.51	0.46	0.49	0.51	0.66	**0.92**	1.00	0.90
T_{10}	0.38	0.40	0.38	0.35	0.37	0.38	0.51	0.74	**0.90**	1.00
T_{11}	0.18	0.21	0.22	0.21	0.21	0.20	0.31	0.51	0.69	**0.90**
T_{12}	-0.11	-0.09	-0.05	-0.04	-0.07	-0.10	-0.06	0.01	0.18	0.42
T_{13}	-0.27	-0.25	-0.19	-0.16	-0.21	-0.26	-0.27	-0.28	-0.12	0.12
T_{14}	-0.41	-0.39	-0.31	-0.25	-0.32	-0.39	-0.44	-0.53	-0.38	-0.14
T_{15}	-0.52	-0.50	-0.40	-0.33	-0.40	-0.48	-0.57	-0.74	-0.64	-0.45
T_{16}	-0.52	-0.50	-0.39	-0.32	-0.40	-0.49	-0.61	-0.84	-0.75	-0.59
T_{17}	-0.46	-0.45	-0.35	-0.28	-0.36	-0.46	-0.59	-0.84	-0.79	-0.68
T_{18}	-0.39	-0.37	-0.27	-0.20	-0.28	-0.38	-0.52	-0.80	-0.76	-0.64
T_{19}	-0.33	-0.31	-0.20	-0.12	-0.20	-0.31	-0.46	-0.76	-0.72	-0.61
T_{20}	-0.25	-0.22	-0.12	-0.03	-0.11	-0.22	-0.38	-0.70	-0.66	-0.55

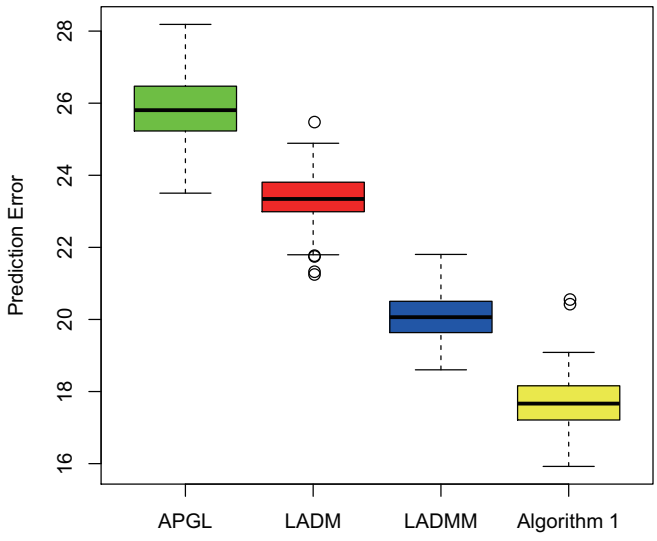

Figure 1: The boxplot of MSPE for analyzing polyethylene data.

Table 5: The correlation coefficient matrix of predictors.

	T_{11}	T_{12}	T_{13}	T_{14}	T_{15}	T_{16}	T_{17}	T_{18}	T_{19}	T_{20}
T_1	0.18	-0.11	-0.27	-0.41	-0.52	-0.52	-0.46	-0.39	-0.33	-0.25
T_2	0.21	-0.09	-0.25	-0.39	-0.50	-0.50	-0.45	-0.37	-0.31	-0.22
T_3	0.22	-0.05	-0.19	-0.31	-0.40	-0.39	-0.35	-0.27	-0.20	-0.12
T_4	0.21	-0.04	-0.16	-0.25	-0.33	-0.32	-0.28	-0.20	-0.12	-0.03
T_5	0.21	-0.07	-0.21	-0.32	-0.40	-0.40	-0.36	-0.28	-0.20	-0.11
T_6	0.20	-0.10	-0.26	-0.39	-0.48	-0.49	-0.46	-0.38	-0.31	-0.22
T_7	0.31	-0.06	-0.27	-0.44	-0.57	-0.61	-0.59	-0.52	-0.46	-0.38
T_8	0.51	0.01	-0.28	-0.53	-0.74	-0.84	-0.84	-0.80	-0.76	-0.70
T_9	0.69	0.18	-0.12	-0.38	-0.64	-0.75	-0.79	-0.76	-0.72	-0.66
T_{10}	0.90	0.42	0.12	-0.14	-0.45	-0.59	-0.68	-0.64	-0.61	-0.55
T_{11}	1.00	0.72	0.45	0.20	-0.18	-0.34	-0.51	-0.48	-0.45	-0.40
T_{12}	*0.72*	1.00	0.86	0.67	0.26	0.08	-0.16	-0.17	-0.17	-0.17
T_{13}	0.45	*0.86*	1.00	0.92	0.57	0.40	0.13	0.11	0.11	0.09
T_{14}	0.20	0.67	**0.92**	1.00	0.84	0.71	0.44	0.42	0.41	0.39
T_{15}	-0.18	0.26	0.57	*0.84*	1.00	0.94	0.73	0.71	0.69	0.66
T_{16}	-0.34	0.08	0.40	0.71	**0.94**	1.00	0.91	0.90	0.88	0.86
T_{17}	-0.51	-0.16	0.13	0.44	0.73	**0.91**	1.00	0.99	0.98	0.96
T_{18}	-0.48	-0.17	0.11	0.42	0.71	**0.90**	**0.99**	1.00	1.00	0.98
T_{19}	-0.45	-0.17	0.11	0.41	0.69	0.88	**0.98**	**1.00**	1.00	0.99
T_{20}	-0.40	-0.17	0.09	0.39	0.66	0.86	**0.96**	**0.98**	**0.99**	1.00

Table 6: The numerical results of Algorithm 1 for analysing polyethylene data set.

	APGL	LADM	LADMM	Algorithm 1
MSPE	25.57(9.58e-1)	23.31(8.84e-1)	20.22(6.67e-1)	**17.71**(7.93e-1)

predictors to form blocks. At last, using the new model and the designed algorithm to analyze the data under the obtained block structures. Note that, if some a predictor belongs to two or more block, we can train the model for each one.

6 Conclusion

In this paper, we propose a MLR model with low-rank and row-sparsity which can deal with the matrix data whose coefficient matrix possesses low-rank and sparse row-group properties simultaneously. To solve the new model, we design a LADMM algorithm and establish its global convergency. An advantage of the proposed LADMM is that it is easily implementable. Moreover, we carry out some numerical studies including real data analysis to show the accuracy and efficiency of the proposed method.

References

[1] L. Breiman and J.H. Friedman, Predicting multivariate responses in multiple linear regression, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 59 (1997) 3–54.

[2] Y. Dan, P. Shang and L. Kong, The row sparse Huber trace regression, working paper, 2021.

[3] J. Friedman, T. Hastie and R. Tibshirani, A note on the group Lasso and a sparse group Lasso, (2010), arXiv: 1001.0736.

[4] D. Goldfarb, S. Ma and K. Scheinberg, Fast alternating linearization methods for minimizing the sum of two convex function, *Math. Program. A* 141 (2013) 349–382.

[5] C. Guyon, T. Bouwmans and E. Zahzah, Foreground detection based on low-rank and block-sparse matrix decomposition, in: *IEEE ICIP,* Orlando, Florida, USA, 2012, pp. 1225–1228.

[6] B.S. He and X.M. Yuan, On the $O(1/n)$ convergence rate of Douglas-Rachford alternating direction method, *SIAM J. Numer. Anal.* 50 (2012) 700–709.

[7] X. Li, L. Mo, X. Yuan and J. Zhang, Linearized alternating direction direction method of multipliers for sparse group and fused Lasso models, *Comput. Stat. Data. Anal.* 79 (2014) 203–221.

[8] Z. Lu, R. Monteiro and M. Yuan, Convex optimization methods for dimension reduction and coefficient estimation in multivariate linear regression, *Math. Program. Ser. A* 131 (2010) 163–194.

[9] S. Ma, D. Goldfarb and L. Chen, Fixed point and Bregman iterative methods for matrix rank minimization, *Math. Program.* 128 (2011) 321–353.

[10] A. Mukherjee and J. Zhu, Reduced rank ridge regression and its kernel extensions, *Stat. Anal. Data Min.: ASA Data Sci. J.* 4 (2011) 612–622.

[11] G. Obozinski, M. Wainwright and M. Jordan, Support union recovery in high-dimensional multivariate regression, *Ann. Statist.* 39 (2011) 1–47.

[12] J. Peng, J. Zhu, A. Bergamaschi, W. Han, D. Noh, J. Pollack and P. Wang, Regularized multivariate regression for identifying master predictors with application to integrative genomics study of breast cancer, *Ann. Appl. Stat.* 4 (2010) 53–77.

[13] P. Shang and L. Kong, Singular value screening rules for the nuclear norm regularized multivariate linear regression, *Pac. J. Optim.* 17 (2021) 1–22.

[14] T. Similä and J. Tikka, Input selection and shrinkage in multiresponse linear regression, *Comput. Statist. Data Anal.* 52 (2007) 406–422.

[15] B. Skagerberg, J.F. MacGregor and C. Kiparissides, Multivariate data analysis applied to low-density polyethylene reactors, *Chemometr. Intell. Lab. Syst.* 14 (1992) 341–356.

[16] G. Tang and A. Nehorai, Robust principal component analysis based on low-rank and block-sparse matrix decomposition, in: *IEEE ICIP,* Baltimore, Maryland, USA, 2011, pp. 1–5.

[17] K. Toh and S. Yun, An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems, *Pacific J. Optim.* 6 (2010) 615–640.

[18] G. Watson, Characterization of the subdifferential of some matrix norms, *Linear Algebra Appl.* 170 (1992) 33–45.

[19] J. Yang and X. Yuan, Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization, *Math. Comput.* 82 (2013) 301–329.

[20] M. Yang, Smoothing technique and fast alternating direction method for robust PCA, in: *Proceeding of 33rd CCC,* Nanjing, China, 2014, pp. 4782–4785.

[21] M. Yang and Y. Wang, Fast alternating direction method of multipliers for robust PCA, *J. Nanjing Univ. Posts Telecommun. (Natural Science)* 34 (2014) 83–88.

[22] M. Yuan, A. Ekici, Z. Lu and R. Monteiro, Dimension reduction and coefficient estimation in multivariate linear regression, *J. R. Stat. Soc. Series B Stat. Methodol.* 69 (2007) 329–346.

[23] M. Yuan and Y. Lin, Model selection and estimation in regression with grouped variables, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 68 (2006) 49–76.

Manuscript received 24 June 2021
revised 15 August 2021
accepted for publication 30 September 2021

Jun Sun
School of Science, Beijing Jiaotong University
Beijing, 100044, P.R. China
E-mail address: 17118473@bjtu.edu.cn

Pan Shang
School of Science, Beijing Jiaotong University
Beijing, 100044, P.R. China
E-mail address: 18118019@bjtu.edu.cn

Qiuyun Xu
School of Science, Beijing Jiaotong University
Beijing, 100044, P.R. China
E-mail address: 15121549@bjtu.edu.cn

Bingzhen Chen
Institute of Statistics and Big Data
Renmin University of China
Beijing, 100872, P.R. China
E-mail address: chenbingzhen6026@163.com