# Sliced Inverse Regression (SIR) via group lasso

**Bahr Kadhim Mohammed[1], Saif Hosam Raheem[2], Tahir R. Dikheel[3]**
[1, 2, 3]Statistics dept., Collage of Administration and Economics, University of Al-Qadisiyah, Iraq.

## Abstract

In this paper the authors propose a group-lasso for sliced inverse regression (group lasso-SIR). This proposed method can deal with the problem of correlation existence between predictor variables. Simulation is used to investigate the performance of proposed method comparing with ridge and lasso in sliced inverse regression (lasso-SIR). The results show that the group lasso-SIR method is performs well comparing with other methods depending on Mean Square Errors (MSE) criterion.

**Keywords:** sliced inverse regression, ridge regression, lasso, group lasso, lasso-SIR , group lasso-SIR.

## The corresponding author:

**Tahir R. Dikheel,** Statistics dept., Collage of Administration and Economics, University of Al-Qadisiyah, Iraq.

tahir.dikheel@qu.edu.iq

# الانحدار العكسي المقطوع (SIR) عبر مجموعة اللاسو

**بحر كاظم محمد[1] سيف حسام رحيم[2] طاهر ريسان دخيل[3]**

**[1،2،3]جامعة القادسية- كلية الادارة والاقتصاد – قسم الاحصاء**

**المستخلص:** في هذه البحث ، اقترح المؤلفون مجموعة لاسو للانحدار العكسي المقطوع (المجموعة اللاسو - SIR). يمكن أن تتعامل هذه الطريقة المقترحة مع مشكلة وجود الارتباط بين متغيرات التوقع. تم استخدام المحاكاة للتحقق من أداء الطريقة المقترحة بالمقارنة مع التلال واللاسو في الانحدار العكسي المقطوع-lasso) (SIR. أظهرت النتائج أن طريقة lasso-SIR للمجموعة تؤدي أداءً جيدًا مقارنة بالطرق الأخرى اعتمادًا على معيار متوسط الأخطاء المربعة.(MSE)
**الكلمات المفتاحية: الانحدار العكسي المقطوع ، الانحدار التلال ، اللاسو الجماعي ، اللاسو الجماعي ، اللاسو- SIR ، المجموعة اللاسو- SIR.**

 ISSN (1681- 6870)

## 1. Introduction

Regression analysis models are among the most common statistical tools in social, economic and medical ... etc. applications. Therefore, it was necessary for researchers to find and develop several methods to estimate the parameters of these models. As well as the selection of important and influential factors in applied studies, especially for studies that contain large data and variables. Therefore, there is an increasing need to reduce the number of these variables effectively, taking into account the retention of all information within these variables, whether the variables clarification or responses. To achieve this goal Li introduced a Sliced inverse Regression (SIR) method (Li, 1991) to reduce dimensions and reduce the dispersion of data, in this model, $E(X|Y)$ is estimated by the values of $Y$, i.e. the opposite of the usual regression. Which is divides the model into multiple slices by $Y$ values. In later years, researchers have developed and improved several methods in order to obtain accurate data analysis and better in SIR models. Li and Christopher proposed a method that combines the shrinkage of the lasso with SIR (Li and Christopher, 2006). Lixing et al. discussed the asymptotic behavior of the estimate of the central dimension-reduction space with high-dimensional data (Lixing et al., 2006). Li and Xiangrong introduced the SIR model based on the (OLS) method of SIR (Li and Xiangrong, 2008), the $L_2$ regularization is introduced, and an alternating least-squares algorithm is improvement. The robustness of sliced inverse regression were studied by Dikheel by proposing two robust methods (Dikheel, 2014). Alkenani and Dikheel suggested sliced inverse quintile regression model (Alkenani and Dikheel, 2016). Li and Christopher used the Lasso-SIR method, which did not take into account the problem of correlations between explanatory variables (Li and Christopher, 2006). The previous studies did not take in account the grouped correlation that may appear between explanatory variables. To overcome this problem we propose a group lasso with SIR to select the important variables within groups.

The reminder of this paper is organized as follows. In section 2 we present SIR and their advantages. In section 3 we briefly introduce the concept of the methods lasso and group lasso. In section 4 we illustrate the SIR with group lasso. The results of the simulation study are discussed in section 5. A brief summarized conclusion is included in section 6.

## 2. Sliced Inverse Regression (SIR)

Li (1991) proposed the Sliced Inverse Regression model (SIR), which is one of the most common models for Sufficient Dimension Reduction (SDR) estimator. Finding a smooth regression function is the basis idea of SIR that operates on a variable set of projections. If we have a random vector of explanatory variables $X \epsilon R^p$ and a response variable Y, the SIR model is the opposite of the classical regression. To study the relationship between the explanatory variables and the response variable $E(Y|X)$, the model will be written as $E(X|Y)$, this means that the response variable is $X$ and independent variable is Y. The SIR is based on the following model:

$$Y = f\left(\beta_1^T x, \beta_2^T x, \dots \beta_p^T x, \epsilon\right), \qquad (1)$$

 ISSN (1681- 6870)

where $\beta_j, j = 1,2, \dots, p$ are unknown projection vector, $x = (x_1, x_2 \dots, x_p)^T$ is a $p \times 1$ predictor vector, $\epsilon$ is the random error with $E(X|\epsilon) = 0$.

Equation 1 shows the model when the response variable $Y$ depend on the p- dimension. The SIR method divides the model into multiple slices according to $Y$ values, then combines the information of all slides. Also, the SIR is based on finding estimates of effective trends to serve as parameters, which is considered a method to processing the problem of dimensions. The SIR is calculated through some arithmetic methods and conversions (Härdle and Simar, 2003). This can be explained in the following stages:

**Stage1**: $x$ is standardized by

$$\tilde{x} = \Sigma_{xx}^{1/2}(x_i - \bar{x}), \; i = 1,2, \dots, n, \qquad (2)$$

where $\bar{x}$ is the mean and $\Sigma_{xx}^{\frac{1}{2}}$ is the covariance matrix of $x$.

**Stage2**: The range of $y_i$ is divided into H slices, $Q_1, Q_2, \dots, Q_H$. Then calculate $\hat{p}_h = (\frac{1}{2}) \sum_{i=1}^{n} I_{Q_h}(y_i)$, which refers to the proportion of $y_i'$ that fall in the slice $Q_h, h = 1,2, \dots, H$, and $I_{Qh}$ is the indicator function.

**Stage3**: We compute the sample mean vector $\hat{M}_h$ of the $\tilde{x}_i's$ for each slice $Q_h$ as:

$$\hat{M}_h = \left(\frac{1}{n\hat{p}_h}\right) \sum \tilde{x}_i \qquad (3)$$

**Stage4**: Use a principle component analysis for $\hat{M}_h$ by the following formula:

$$\hat{V} = \sum_{h=1}^{H} \hat{p}_h \hat{M}_h \hat{M}'_h, \qquad (4)$$

then find the eigenvalues and eigenvectors of $\hat{V}$.

**Stage5**: Let $\hat{G}_j, \; j = 1,2, \dots, K$ represents the eigenvectors associated with the largest $k$ eigenvalues, $\hat{\beta}_j$ is transformed back to the original by the following:

$$\hat{\beta}_j = \hat{G}_j \hat{\Sigma}_{XX}^{\frac{-1}{2}} \qquad (5)$$

$\hat{\beta}_j$ represents the estimated Effective Dimension Reduction (E.D.R) directions.

## 3. Lasso and group lasso

There are several new methods have emerged in recent years. Through which models and important variables can be selected. The most important characteristics of these methods are the identification and selection of important factors during the estimation process. One of these ways is Lasso (Least Absolute Shrinkage and Selection Operator) method proposed by Tibshirani (1996). It is an $\ell_1$ penalized least squares regression, by it some of the coefficients are shrunk while the rest of them are exactly set to zero. This feature makes lasso enjoys good properties of best subset regression and ridge regression

 **ISSN (1681- 6870)**

(Hastie et al., 2001). Lasso is sead to be better than ordinary least squares (OLS) regression for two reasons: Firstly, an over specified OLS model often has little bias but large variance, adversely affecting its prediction accuracy. This can be improved by shrinking or setting to zero some coefficients. Secondly, a large number of insignificant coefficients may be included in OLS models, a little value is added to the model and complicating the interpretation of the effects. The lasso estimate of $\beta$,

$$\hat{\beta}^* = argmin\left\{\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right|\right\}, \qquad (6)$$

where $\sum_{j=1}^{p}\lambda\left|\beta_j\right|$ is called lasso penalty.

Zou (2006) suggested a new version of the lasso, called adaptive lasso, based on the adaptive weights, which in turn lead different penalization to different coefficients in the $\ell_1$ penalty. The adaptive lasso can be defined as:

$$argmin\left\{\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda\sum_{j=1}^{p}w_j\left|\beta_j\right|\right\}, \qquad (7)$$

where $(w_1, w_2, \dots, w_p)$ are the adaptive weights.

Yuan and Lin (2006), suggested group lasso method. The main objective for it is to identify the common factors of groups by developing the lasso, the formula for the group lasso is:

$$\hat{\beta} = argmin_{\beta\in R^D}\|y - X\beta\|_2^2 + \sum_{g=1}^{G}\lambda_g\left\|\beta_{I_g}\right\|_2, \qquad (8)$$

where $X$ is the design matrix, $G$ represented the group sets and $g$ is the individual group, $I_g$ is the index set belonging to the $g^{th}$ group of variables.

## 4. Group lasso-SIR

Many researchers developed methods have emerged in recent to solve the problem of high-dimensional data as well the selection of important variables such as, lasso, adaptive lasso, group lasso…etc.. These methods have been employed with many classical regression analysis models. Li and Christopher (2006), introduced the lasso with SIR. They proposed replacing, the OLS estimator in SIR algorithm with a lasso estimator, and $\sum_{l=1}^{p}\left|\beta_{jl}\right| \leq \tau$ is the lasso constraint, $\beta_{jl}$ is the $lth$ coordinate of $\beta_j$, and $\tau$ is a given shrinkage factor. The SIR algorithm for their method for the $jth$ sparse SIR direction $\hat{\eta}_j$, $j = 1,2,\dots,d$,

1. Set a value for $\tau$, then initialize $\beta_j = \ddot{V}_j$, the $jth$ SIR estimate.

2. Find $t(y) = (\ddot{E}(X|y_1), \ddot{E}(X|y_2) \dots \ddot{E}(X|y_n))^T\beta_j$, where $\ddot{E}(X|y_i)$ refers to the sample estimate of $E(X|y_i)$ evaluated at $y_i$, $i = 1,2,\dots,n$.

 **ISSN (1681- 6870)**

3. $\beta_j$ is updated as the lasso solution with $t(y)$ as the response, $X$ as the predictors, and the shrinkage factor $\tau$.

4. $\beta_j$ is normalized as $\beta_j = \beta_j/|\beta_j|$. If $j > 1$, $\beta_j$ orthonormalized such that $\beta_j^T\beta_j = 1$ and $\hat{\eta}_k^T\hat{\Sigma}_x\beta_j = 0$, where $\hat{\eta}_k$, $k = 1,2,\dots,j-1$, are the sparse SIR estimates of the first $j-1$ directions.

5. steps 2-4 are repeated until $\beta_j$ converges. Set $\hat{\eta}_j = \beta_j$.

Here, we proposed to replace the lasso penalty with group lasso penalty in step (3) to get group lasso-SIR.

## 5. **Simulation study**

To evaluate the performance of group lasso-SIR, we use the numerical study by a code written with R depending on *Lasso SIR* package. The results of this method are compared with ridge and lasso-SIR with several models depending on MSE. We suppose the number of predictors to be p=1000, while the correlation between predictor variables is taken to be (r=0.25, 0.50, 0.75), the sample sizes are (n=100, 150, 300), the non zero coefficients s are (5, 10, 15, 20). For the regression model $y = X\beta + e$, we suppose $\beta = (.2, 1.3, .2, 1.3, \dots)$ for the first s coefficients, zero otherwise. $e$ is independently and identically distributed (i.i.d) with standard normal. First, we generate $z_j$ as $z_j = (\delta_j + a_j\delta)/\sqrt{1+a_j^2}, j = 1{:}p$, where $\delta$ and $\delta_j$ is a standard normal distribution and $a_j$ is generated as $a_1 = \sqrt{\frac{r}{1-r}}$ , $a_j = a_1$ for $j \le 7$, $a_j = a_1^2$ for $7 < j \le 12$, $a_j = a_1^4$ for $12 < j \le 18$, $a_j = 0$ for $j > 18$, the predictor variables are generated as $x_j = \sin(z_j)$ for j=1:5, $x_j = \exp(z_j)$ for j=6:12, and $x_j = z_j$ for j>=12. The simulated samples are repeated 500 times to reach stable results.

**Case1**: 500 simulated samples are generated for s=20 and the results are included in Table1:

Table 1: MSE results for the used methods when s=20.

| n | r | ridge | lasso-SIR | group lasso-SIR |
|---|---|---|---|---|
| 100 | .25 | 0.9156187 | 0.6579719 | **0.5437485** |
| | .5 | 3.088036 | 2.667877 | **2.409792** |
| | .75 | 2.711442 | 2.310508 | **2.281719** |
| 150 | .25 | 0.5222085 | 0.3522492 | **0.3300406** |
| | .5 | 0.7642423 | 0.5362527 | **0.4245361** |
| | .75 | 1.541709 | 1.293854 | **1.082153** |
| 300 | .25 | 0.9799152 | 0.8655148 | **0.7718458** |
| | .5 | 0.1767782 | 0.1641021 | **0.1259184** |
| | .75 | 1.944356 | 1.613517 | **1.468753** |

**Case2**: The same as Case1 except s=15 and the results are included in Table2:

Table 2: MSE results for the used methods when s=15.

| n | R | ridge | lasso-SIR | group lasso-SIR |
|---|---|---|---|---|
| 100 | .25 | 1.084368 | 0.8718594 | **0.5176119** |
| | .5 | 1.481249 | 1.213465 | **1.191358** |
| | .75 | 2.16188 | 1.739144 | **1.434361** |
| 150 | .25 | 1.05884 | 0.8650053 | **0.7217975** |
| | .5 | 1.142425 | 0.9611598 | **0.9027045** |
| | .75 | 1.492051 | 1.169872 | **0.9427215** |
| 300 | .25 | 0.4947488 | 0.3839826 | **0.3704343** |
| | .5 | 0.8336184 | 0.7398179 | **0.5827363** |
| | .75 | 1.076618 | 1.003812 | **0.7756428** |

**Case 3**: The same as Case1 except s=10 and the results are included in Table3:

Table3: MSE results for the used methods when s=10.

| N | R | ridge | lasso-SIR | group lasso-SIR |
|---|---|---|---|---|
| 100 | .25 | 0.7036851 | 0.3935908 | **0.4061151** |
| | .5 | 0.6723328 | 0.4613948 | **0.4408378** |
| | .75 | 1.841008 | 1.182484 | **1.161038** |
| 150 | .25 | 0.5041425 | 0.4297877 | **0.371805** |
| | .5 | 0.4743354 | 0.2549559 | **0.2130783** |
| | .75 | 2.57582 | 2.152348 | **1.70304** |
| 300 | .25 | 0.2495151 | 0.1850454 | **0.1734728** |
| | .5 | 0.4158943 | 0.2916982 | **0.2165369** |
| | .75 | 0.6822518 | 0.5554065 | **0.370098** |

Table 1, Table 2, and Table 3 show that the group lasso-SIR gives good performance when s=20, 15, 10 in comparison with other methods for all used samples and correlations. Furthermore, the results show that the MSE decreases when the sample size increases, while the MSE values increases when the correlation values decrease.

**Case 4**: The same as Case1 except s=5 and the results are included in Table4:

Table 4: MSE results for the used methods when s=5.

| n | r | ridge | lasso-SIR | group lasso-SIR |
|---|---|---|---|---|
| 100 | .25 | 0.4651417 | 0.1696405 | **0.1897502** |
| | .5 | 0.5027688 | **0.2309072** | 0.2411894 |
| | .75 | 0.9814881 | **0.2718451** | 0.2853204 |
| 150 | .25 | 0.3199197 | **0.1559372** | 0.1918986 |
| | .5 | 0.1204836 | **0.07101108** | 0.09445693 |
| | .75 | 0.1401998 | **0.07905097** | 0.0809922 |
| 300 | .25 | 0.2283673 | **0.1579774** | 0.1644473 |
| | .5 | 0.1524528 | **0.1098665** | 0.1147017 |
| | .75 | 0.1271715 | **0.07357599** | 0.07467907 |

In Table 4 , when s=5, we note that the lasso-SIR has the smallest MSE for all sample sizes and correlations except when n=100 and r=.25 which the group lasso-SIR is the best. Also, in this case the MSE values decrease when the sample sizes increase. Whereas, the MSE values increases when the correlation values decrease. This different behavior belong to remove the groups case because the zero values of beta include only the first group.

## 6. Conclusion

The correlated predictor variables stand as a serious problem in variable selection methods. In current paper we proposed group lasso-SIR, this new method is compared with ridge and lasso-SIR by using several simulation cases. To judge the group lasso-SIR performance, we build numerical data which is simulating what we describe in previous sections. Based on results and MSE criterion we conclude that group lasso-SIR do well when the non zero coefficients, represented by s, increasing to include all groups of case of correlated predictor variables. Furthermore, we conclude that the lasso-SIR gives best results in case of small value of s because the groups are removed in this case. Consequently, the authors believe that group lasso-SIR can give a good results in practice when the predictors are correlated. Also, the graphical lasso-SIR, robust lasso-SIR, and robust group lasso-SIR can be a future works in this sense.

## References

1- Alkenani, A., & Dikheel, T. R. (2016). Sparse Sliced Inverse Quantile Regression.
2- Dikheel, T. R. (2014). Robust Sliced Inverse Regression. AL-Qadisiyah Journal For Administrative and Economic sciences, 16(1), 227-242.
3- Härdle, W. and Simar, L. (2003) Applied Multivariate Statistical Analysis, Springer Verlag. ISBN 3-540-03079-4
4- Hastie, T., Tibshirani, R., & Friedman, J. J. H. (2001). The elements of statistical learning (Vol. 1): Springer New York.
5- Li, K. C. (1991). Sliced inverse regression for dimension reduction. Journal of the American Statistical Association, 86(414), 316-327.
6- Li, L., & Nachtsheim, C. J. (2006). Sparse sliced inverse regression. Technometrics, 48(4), 503-510.
7- Li, L., & Yin, X. (2008). Sliced inverse regression with regularizations. Biometrics, 64(1), 124-131.
8- Lin, Q., Zhao, Z., & Liu, J. S. (2019). Sparse sliced inverse regression via lasso. Journal of the American Statistical Association, 1-33.
9- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Series B 58, 267–288.
10- Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), 49-67.
11- Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association 101, 1418–1429
12- Zhu, L., Miao, B., & Peng, H. (2006). On sliced inverse regression with high-dimensional covariates. Journal of the American Statistical Association, 101(474), 630-643.