



S2IR: Standardized Sliced Inverse Regression

Li Kevin

► To cite this version:

| Li Kevin. S2IR: Standardized Sliced Inverse Regression. 2013. hal-00833686

HAL Id: hal-00833686

<https://hal.science/hal-00833686>

Preprint submitted on 13 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

S2IR: Standardized Sliced Inverse Regression

Kevin B. Li

Beijing Jiaotong University

Abstract

In this paper we consider a semiparametric regression model involving a p -dimensional explanatory variable \mathbf{x} and including a dimension reduction of \mathbf{x} via an index $B'\mathbf{x}$. In this model, the main goal is to estimate B and to predict the real response variable Y conditionally to \mathbf{x} . A standard approach is based on sliced inverse regression (SIR). We propose a new version of this method: the standardized sliced inverse regression.

The purpose of the regression of a univariate response y on a p -dimensional predictor vector \mathbf{x} is to make inference on the conditional distribution of $y|\mathbf{x}$. Following [17], \mathbf{x} can be replaced by its standardized version

$$\mathbf{z} = [\Sigma_{\mathbf{x}}]^{-1/2}(\mathbf{x} - \mu_{\mathbf{x}}) , \quad (1)$$

where $\mu_{\mathbf{x}}$ and $\Sigma_{\mathbf{x}}$ denote the mean and covariance matrix of \mathbf{x} respectively assuming non-singularity of $\Sigma_{\mathbf{x}}$.

The goal of dimension reduction in regression is to find out a $p \times d$ matrix B such that

$$y \perp\!\!\!\perp \mathbf{z} | B'\mathbf{z} , \quad (2)$$

where “ $\perp\!\!\!\perp$ ” indicates independence. Then the p -dimensional \mathbf{z} can be replaced by the d -dimensional vector $B'\mathbf{z}$ without specifying any parametric model and without losing any information on predicting y . The column space $\text{Span}\{B\}$ is called a dimension reduction subspace. The smallest applicable d is called the dimension of the regression.

Based on the inverse mean $E(\mathbf{z}|y)$, [35] proposed Sliced Inverse Regression (SIR) for dimension reduction in regression. It is realized that SIR can not recover the symmetric dependency [35, 13]. After SIR, many dimension reduction methods have been introduced. Sliced Average Variance Estimate (SAVE) proposed by [13] and Principle Hessian Directions (pHd) proposed by [36] are another two popular ones. Both pHd and SAVE refer to the second inverse moment, centered or non-centered. Compared with SAVE, pHd can not detect certain dependency hidden in the second moment [68, 67] and the linear dependency [36, 16]. Among those dimension reduction methods using only the first two inverse moments, SAVE seems to be the preferred one. Nevertheless, SAVE is not always the winner. For example, [67] implied that a linear combination of

SIR and pHd may perform better than SAVE in some cases. It is not surprising since [35] already suggested that a suitable combination of two different methods might sharpen the dimension reduction results. [67] further proposed that a bootstrap method could be used to pick up the “best” linear combination of two known methods, as well as the dimension of the regression, in the sense of the variability of the estimators, although lower variability under the bootstrap procedure does not necessarily lead to a better estimator. [40] pointed out that linear combinations of two known methods selected by the bootstrap criterion may not perform as well as a single new method, their Directional Regression method (DR), even though the bootstrap one is computationally intensive.

As pointed out in [45], Sliced Inverse Regression also suffers from limitations coming from the need to inverse the covariance matrix in (1). We thus propose a new method to overcome this problem. The idea is to assume that the predictors are standardized. The so-called Standardized Sliced Inverse Regression (S2IR) is based on the inverse mean $E(\mathbf{x}|y)$ instead of $E(\mathbf{z}|y)$ as in [35]. A basis of the central subspace is estimated by computing the eigenvectors of the covariance matrix of $E(\mathbf{x}|y)$. This new method is no more sensible to the curse of dimensionality since it does not require the inversion of the covariance matrix in (1).

References

- [1] Aragon, Y. (1997). A Gauss implementation of multivariate sliced inverse regression. *Computational Statistics*, 12, 355–372.
- [2] Aragon, Y. and Saracco, J. (1997). Sliced Inverse Regression (SIR): an appraisal of small sample alternatives to slicing. *Computational Statistics*, 12, 109–130.
- [3] Azaïs, R., Gégout-Petit, A., and Saracco, J. (2012). Optimal quantization applied to sliced inverse regression. *Journal of Statistical Planning and Inference*, 142, 481–492.
- [4] Barreda, L., Gannoun, A., and Saracco, J. (2007). Some extensions of multivariate SIR. *Journal of Statistical Computation and Simulation*, 77, 1–17.
- [5] Bentler, P.M., Xie, J., (2000). Corrections to test statistics in principal Hessian directions. *Statistics and Probability Letters*, 47, 381–389.
- [6] Bernard-Michel, C., Gardes, L. and Girard, S. (2008). A Note on Sliced Inverse Regression with Regularizations, *Biometrics*, 64, 982–986.
- [7] Bernard-Michel, C., Douté, S., Fauvel, M., Gardes, L. and Girard, S. (2009). Retrieval of Mars surface physical properties from OMEGA hyperspectral images using Regularized Sliced Inverse Regression. *Journal of Geophysical Research - Planets*, 114, E06005.

- [8] Bernard-Michel, C., Gardes, L. and Girard, S. (2009). Gaussian Regularized Sliced Inverse Regression, *Statistics and Computing*, 19, 85-98.
- [9] Bura, E. and Cook, R. D. (2001). Estimating the structural dimension of regressions via parametric inverse regression. *Journal of the Royal Statistical Society, Series B*, 63, 393–410.
- [10] Chavent, M., Kuentz, V., Liquet, B., and Saracco, J. (2011). A sliced inverse regression approach for a stratified population. *Communications in statistics - Theory and methods*, 40, 1–22.
- [11] Chavent, M., Girard, S., Kuentz, V., Liquet, B., Nguyen, T.M.N. and Saracco, J. (2013). A sliced inverse regression approach for data stream, <http://hal.inria.fr/hal-00688609>
- [12] Chen, C.-H. and Li, K.-C. (1998). Can SIR be as popular as multiple linear regression? *Statistica Sinica*, 8, 289–316.
- [13] Cook, R.D., Weisberg, S., (1991). Discussion of ‘sliced inverse regression for dimension reduction’. *Journal of the American Statistical Association*, 86, 328-332.
- [14] Cook, R.D., (1994). Using dimension-reduction subspaces to identify important inputs in models of physical systems. Proceedings of the Section on Physical and Engineering Sciences. Alexandria, VA: American Statistical Association. 18-25.
- [15] Cook, R.D., 1996. Graphics for regressions with a binary response. *Journal of the American Statistical Association*, 91, 983-992.
- [16] Cook, R. D.(1998). Principal hessian directions revisited (with discussion). *Journal of the American Statistical Association*, 93, 84–100.
- [17] Cook, R.D., (1998b). Regression Graphics, Ideas for Studying Regressions through Graphics. Wiley, New York.
- [18] Cook, R.D., Lee, H., (1999). Dimension-reduction in binary response regression. *Journal of the American Statistical Association*, 94, 1187-1200.
- [19] Cook, R. D. (2000). SAVE: a method for dimension reduction and graphics in regression. *Communications in statistics - Theory and methods*, 29, 2109–2121.
- [20] Cook, R.D., Critchley, F., (2000). Identifying regression outliers and mixtures graphically. *Journal of the American Statistical Association*, 95, 781-794.
- [21] Cook, R. D. and Li, B. (2002). Dimension reduction for conditional mean in regression. *The Annals of Statistics*, 30, 450–474.

- [22] Cook, R.D., Ni, L., (2005). Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. *Journal of the American Statistical Association*, 100, 410-428.
- [23] Coudret, R., Girard, S. and Saracco, J. (2013). A new sliced inverse regression method for multivariate response regression, <http://hal.inria.fr/hal-00688609>
- [24] Duan, N. and Li, K.-C. (1991). Slicing regression: a link-free regression method. *The Annals of Statistics*, 19, 505–530.
- [25] Dunia, R. and Joe Qin, S. (1998). Subspace approach to multidimensional fault identification and reconstruction. *AIChE Journal*, 44, 1813–1831.
- [26] Ferré, L. (1998). Determining the dimension in sliced inverse regression and related methods. *Journal of the American Statistical Association*, 93, 132–140.
- [27] Gannoun, A., Girard, S., Guinot, C. and Saracco, J. (2002). Reference ranges based on nonparametric quantile regression, *Statistics in Medicine*, 21, 3119-3135.
- [28] Gannoun, A., Girard, S., Guinot, C. and Saracco, J. (2004). Sliced inverse regression in reference curves estimation. *Computational Statistics and Data Analysis*, 46, 103–122.
- [29] Gannoun, A. and Saracco, J. (2003). An asymptotic theory for SIR_α method. *Statistica Sinica*, 13, 297–310.
- [30] Hall, P. and Li, K. C. (1993). On almost linearity of low dimensional projections from high dimensional data. *The Annals of Statistics*, 21, 867-889.
- [31] Hsing, T. (1999). Nearest neighbor inverse regression. *The Annals of Statistics*, 27, 697–731.
- [32] Hsing, T. and Carroll, R. J. (1992). An asymptotic theory for sliced inverse regression. *The Annals of Statistics*, 20, 1040–1061.
- [33] Kuentz, V., Lique, B., and Saracco, J. (2010). Bagging versions of sliced inverse regression. *Communications in statistics - Theory and methods*, 39, 1985–1996.
- [34] Kuentz, V. and Saracco, J. (2010). Cluster-based sliced inverse regression. *Journal of the Korean Statistical Society*, 39, 251–267.
- [35] Li, K.C. (1991). Sliced inverse regression for dimension reduction (with discussion). *Journal of the American Statistical Association*, 86, 316-342.
- [36] Li, K.C. (1992). On principal Hessian directions for data visualization and dimension reduction: another application of Steins lemma. *Journal of the American Statistical Association*, 87, 1025–1039.

- [37] Li, K.-C., Aragon, Y., Shedden, K., and Agnan, C. T. (2003). Dimension reduction for multivariate response data. *Journal of the American Statistical Association*, 98, 99–109.
- [38] Li, L. and Nachtsheim, C. J. (2006). Sparse sliced inverse regression. *Technometrics*, 48, 503–510.
- [39] Li, L., Cook, R. D. and Tsai, C. L. (2007) Partial inverse regression. *Biometrika*, 94, 615–625.
- [40] Li, B., Wang, S., (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association*, 102, 997–1008.
- [41] Li, L. and Yin, X. (2008). Sliced inverse regression with regularizations, *Biometrics*, 64, 124–131.
- [42] Li, B. and Wang, S. (2007). On directional regression for dimension reduction, *Journal of the American Statistical Association*, 102, 997–1008.
- [43] Li, K.B. (2013). Invariance properties of Sliced Inverse Regression, <http://hal.archives-ouvertes.fr/hal-00805491>
- [44] Li, K.B. (2013). A review on Sliced Inverse Regression, <http://hal.archives-ouvertes.fr/hal-00803698>
- [45] Li, K.B. (2013b). Some limitations of Sliced Inverse Regression, <http://hal.archives-ouvertes.fr/hal-00803602>
- [46] Lique, B. and Saracco, J. (2008). Application of the bootstrap approach to the choice of dimension and the α parameter in the SIR_α method. *Communications in statistics - Simulation and Computation*, 37, 1198–1218.
- [47] Lique, B. and Saracco, J. (2012). A graphical tool for selecting the number of slices and the dimension of the model in SIR and SAVE approaches. *Computational Statistics*, 27, 103–125.
- [48] Lue, H.-H. (2009). Sliced inverse regression for multivariate response regression. *Journal of Statistical Planning and Inference*, 139, 2656–2664.
- [49] Naik, P. and Tsai, C. L. (2000) Partial least squares estimator for single-index models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62, 763–771.
- [50] Nkiet, G.-M. (2008). Consistent estimation of the dimensionality in sliced inverse regression. *Annals of the Institute of Statistical Mathematics*, 60, 257–271.
- [51] Prendergast, L. A. (2005). Influence functions for sliced inverse regression. *Scandinavian Journal of Statistics*, 32, 385–404.

- [52] Prendergast, L. A. (2007). Implications of influence function analysis for sliced inverse regression and sliced average variance estimation. *Biometrika*, 94, 585–601.
- [53] Saracco, J. (1997). An asymptotic theory for Sliced Inverse Regression. *Communications in statistics - Theory and methods*, 26, 2141–2171.
- [54] Saracco, J. (1999). Sliced inverse regression under linear constraints. *Communications in statistics - Theory and methods*, 28, 2367–2393.
- [55] Saracco, J. (2001). Pooled slicing methods versus slicing methods. *Communications in statistics - Simulation and Computation*, 30, 489–511.
- [56] Saracco, J. (2005). Asymptotics for pooled marginal slicing estimator based on SIR_α approach. *Journal of Multivariate Analysis*, 96, 117–135.
- [57] Schott, J. R. (1994). Determining the dimensionality in sliced inverse regression. *Journal of the American Statistical Association*, 89, 141–148.
- [58] Scrucca, L. (2007). Class prediction and gene selection for DNA microarrays using regularized sliced inverse regression. *Computational Statistics & Data Analysis*, 52, 438–451.
- [59] Setodji, C. M. and Cook, R. D. (2004). K-means inverse regression. *Technometrics*, 46, 421–429.
- [60] Shao, Y., Cook, R.D., Weisberg, S., (2007). Marginal tests with sliced average variance estimation. *Biometrika*, 94, 285–296.
- [61] Shao, Y., Cook, R. D., and Weisberg, S. (2009). Partial central subspace and sliced average variance estimation. *Journal of Statistical Planning and Inference*, 139, 952–961.
- [62] Szretter, M. E. and Yohai, V. J. (2009). The sliced inverse regression algorithm as a maximum likelihood procedure. *Journal of Statistical Planning and Inference*, 139, 3570–3578.
- [63] Weisberg, S., (2002). Dimension reduction regression in R. *Journal of Statistical Software*, Available from <http://www.jstatsoft.org>
- [64] Wen, X., Cook, R.D., (2007). Optimal sufficient dimension reduction in regressions with categorical predictors. *Journal of Statistical Inference and Plannin*, 137, 1961–1979.
- [65] Wu, H.-M. (2008). Kernel sliced inverse regression with applications to classification. *Journal of Computational and Graphical Statistics*, 17, 590–610.
- [66] Ye, Z. and Yang, J. (2010). Sliced inverse moment regression using weighted chi-squared tests for dimension reduction. *Journal of Statistical Planning and Inference*, 140, 3121–3131.

- [67] Ye, Z., Weiss, R.E., (2003). Using the bootstrap to select one of a new class of dimension reduction methods. *Journal of the American Statistical Association*, 98, 968-979.
- [68] Yin, X., Cook, R.D., (2002). Dimension reduction for the conditional k th moment in regression. *Journal of the Royal Statistical Society*, Ser. B. 64, 159-175.
- [69] Yin, X., Cook, R.D., (2003). Estimating central subspaces via inverse third moments. *Biometrika*, 90, 113-125.
- [70] Yin, X. and Bura, E. (2006). Moment-based dimension reduction for multivariate response regression. *Journal of Statistical Planning and Inference*, 136, 3675-3688.
- [71] Zhong, W., Zeng, P., Ma, P., Liu, J.S. and Zhu, Y. (2005). RSIR: Regularized Sliced Inverse Regression for motif discovery. *Bioinformatics*, 21, 4169-4175.
- [72] Zhu, L., Miao, B., and Peng, H. (2006). On sliced inverse regression with high-dimensional covariates. *Journal of the American Statistical Association*, 101, 630-643.
- [73] Zhu, L.-P. and Yu, Z. (2007). On spline approximation of sliced inverse regression. *Science in China Series A: Mathematics*, 50, 1289-1302.
- [74] Zhu, L. X. and Fang, K. T. (1996). Asymptotics for kernel estimate of sliced inverse regression. *The Annals of Statistics*, 24, 1053-1068.
- [75] Zhu, L. X. and Ng, K. W. (1995). Asymptotics of sliced inverse regression. *Statistica Sinica*, 5, 727-736.
- [76] Zhu, L. X., Ohtaki, M., and Li, Y. (2007). On hybrid methods of inverse regression-based algorithms. *Computational Statistics*, 51, 2621-2635.