Discussion on 'Review of sparse sufficient dimension reduction'

Xin Zhang

Published online: 13 Oct 2020.

Submit your article to this journal ⬀

Article views: 51

View related articles ⬀

Taylor & Francis
Taylor & Francis Group

Check for updates

SHORT COMMUNICATION

Discussion on 'Review of sparse sufficient dimension reduction'

Xin Zhang

Department of Statistics, Florida State University, Tallahassee, FL, USA

I congratulate the authors on an excellent overview of an important research area. Sufficient dimension reduction methods are based on the model-free driving condition that $Y \perp\!\!\!\perp \mathbf{X} \mid \mathbf{P}_{\mathcal{S}}\mathbf{X}$, where $\mathbf{X} \in \mathbb{R}^p$ is multivariate and potentially high-dimensional, $\mathbf{P}_{\mathcal{S}}$ is the projection onto the dimension reduction subspace $\mathcal{S} \subseteq \mathbb{R}^p$. Equipped with variable selection and variable screening techniques, many modern sparse sufficient dimension reduction methods have been developed in the past few years, and they can work really well in the model development stage of high-dimensional data analysis. This review paper is very timely and provides a thorough overview of sparse sufficient dimension reduction methods and sheds lights on future research directions.

Specific contributions of this paper include the following. First of all, it recasts many moment-based sufficient dimension reduction methods as a generalised eigen-decomposition problem and also as a constrained trace optimisation (Equation (2.1) in the paper). These two formulations are crucial for studying sparse sufficient dimension reduction in high-dimensional settings where $p > n$ or $p \gg n$. This paper also provides a comprehensive overview of the key developments in sparse sufficient dimension reduction literature, from the first paper by Ni et al. (2005), to the newest theoretical and computational breakthrough in Tan et al. (2020). Various techniques for sparse sufficient dimension reduction are discussed, including the shrinkage and regularised estimation, variable screening and the trace pursuit, and the sequential procedures. Advantages and possible pitfalls of different approaches are well explained by the authors. Finally, the paper provides the theoretical foundations for establishing the minimax rates of convergence for estimating a sparse dimension reduction subspace. Further calculations are also included to provide insights about second-order dimension reduction methods such as SAVE and DR.

This paper is inspiring for substantive research in high-dimensional multivariate statistics, and encouraging for current and future studies in dimension reduction. In what follows, I draw two connections which may be suggestive of future directions.

1. Least squares formulations

In the seminal work of Li and Duan (1989), a well-known connection between the ordinary least square estimator in the regression of Y on $\mathbf{X} \in \mathbb{R}^p$ and the one-dimensional dimension reduction subspace can be derived. Specifically, consider the following bivariate loss function,

$$L(a + \mathbf{b}^T\mathbf{X}, Y), \quad L(u, v) \text{ is convex in } u,$$
$$\text{e.g. } L(u, v) = (u - v)^2; \quad (1)$$

and define the unique minimiser as follows,

$$(\alpha, \boldsymbol{\beta}) \equiv \arg \min_{a, \mathbf{b} \in \mathbb{R}^p} \mathrm{E}\{L(a + \mathbf{b}^T\mathbf{X}, Y)\}. \quad (2)$$

Then under the linearity condition of dimension reduction subspace, it was shown that $\boldsymbol{\beta}$ is contained in the central subspace (or any dimension reduction subspace).

A direct consequence of this somewhat surprising result is that one can simply use OLS to extract the direction in nonlinear models of the following form,

$$Y = g(\boldsymbol{\gamma}^T\mathbf{X}, \varepsilon), \quad (3)$$

where g is some unknown function, ε is the error term, and $\boldsymbol{\gamma} \in \mathbb{R}^{p \times d}$ spans the central subspace under this model. The model (3) is known as the single-index model when $d = 1$, and the multiple-index model when $d > 1$. Then Li–Duan Theorem (Li & Duan, 1989, Theorem 2.1) implies that the solution $\boldsymbol{\beta}$ from ordinary least squares estimation, or more generally from (2), lies within the central subspace: $\boldsymbol{\beta} \in \mathrm{span}(\boldsymbol{\gamma}) = \mathcal{S}_{Y|\mathbf{X}}$.

For single-index models, this means that $\boldsymbol{\beta}$ from OLS fitting, $(\alpha, \boldsymbol{\beta}) = \arg \min_{a, \mathbf{b}} \mathrm{E}\{(Y - a - \mathbf{b}^T\mathbf{X})^2\}$, is the same as $\boldsymbol{\gamma}$ in population up to a scalar multiplication (i.e. $\boldsymbol{\beta} = c \cdot \boldsymbol{\gamma}$ for some constant $c \neq 0$). As such, we

CONTACT Xin Zhang ✉ henry@stat.fsu.edu

might still use OLS to study regression graphics even when there is a nonlinear relationship between Y and X (Cook, 1998). In high-dimensional setting, one may replace OLS with its penalised version such as LASSO regression and achieve consistent variable selection and directional estimation (Neykov, Lin, et al., 2016; Neykov, Liu, et al., 2016).

For multiple-index models and model-free sufficient dimension reduction, the (sparse) estimation of the central subspace is much more challenging. In high-dimensional sparse sufficient dimension reduction, it is thus desirable to have a penalised least squares formulation. Indeed, as this paper points out, the computationally tractable and rate optimal sparse SIR is eventually obtained by the adaptive estimation scheme based on a least squares formulation (Tan et al., 2020). Finally, the paper concludes that *'Since SAVE and DR can not be rewritten as a least-square formulation, we do not define refined sparse SAVE and DR estimator'*. Not surprisingly, developing theoretical solid and computationally feasible methods for high-dimensional SAVE and DR is very challenging and requires substantial efforts.

2. Sparse/constrained canonical correlation

As discussed in Section 3.1 of this paper, sparse sufficient dimension reduction subspace can be obtained by the C^3 method: constrained canonical correlation (Zhou and He, 2008). The idea is to estimate the constrained canonical variates between the B-spline basis functions of response transformation, $\pi(\mathbf{Y}) \in \mathbb{R}^{m+k_n}$, where m is the spline order and k_n is the number of interval knots in B-spline transformation, and the predictor $\mathbf{X} \in \mathbb{R}^p$. Then this procedure can be viewed as an estimation for sparse SIR directions. However, as the authors noted, this method may not be directly applicable to very high-dimensional settings, at least in theory. In the past few years, there are some advances in both the theoretical and computational aspects of high-dimensional canonical correlation analysis. In Mai and Zhang (2019), they solved the sparse canonical correlation analysis (CCA) problem using an iterative penalised least squares approach. This new iterative penalised least squares approach can be directly applied to estimate the sparse sufficient dimension reduction directions when combined with B-spline transformations of the response.

To illustrate the idea, we consider the estimation of the leading CCA directions. For a multivariate $\mathbf{X} \in \mathbb{R}^p$ and a multivariate $\mathbf{Y} \in \mathbb{R}^q$, the leading CCA directions are defined through a pair of linear combinations $\boldsymbol{\alpha}_1^T \mathbf{Y}$ and $\boldsymbol{\beta}_1^T \mathbf{X}$ such that the correlation between them are maximised. When $\max(p, q) \gg n$, the sparse CCA problem assumes that the population solution $\boldsymbol{\alpha}_1^\star \in \mathbb{R}^q$ and $\boldsymbol{\beta}_1^\star \in \mathbb{R}^p$ are both sparse so that we can estimate them with a limited sample size. Then the leading sparse CCA directions can be obtained by solving the

following constrained optimisation problem,

$$(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) = \arg \min_{\boldsymbol{\alpha} \in \mathbb{R}^q, \boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n (\boldsymbol{\alpha}^T \mathbf{Y}_i - \boldsymbol{\beta}^T \mathbf{X}_i)^2$$
$$+ \lambda_\alpha \|\boldsymbol{\alpha}\|_1 + \lambda_\alpha \|\boldsymbol{\beta}\|_1, \quad (4)$$

subject to constraints that $\boldsymbol{\alpha}^T \widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}} \boldsymbol{\alpha} = 1$ and $\boldsymbol{\beta}^T \widehat{\boldsymbol{\Sigma}}_{\mathbf{X}} \boldsymbol{\beta} = 1$. Note that the data are centred so that $\sum_{i=1}^n \mathbf{X}_i = 0$ and $\sum_{i=1}^n \mathbf{Y}_i = 0$, also that the sample covariance $\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{X}}$ do not need to be positive definite. Then it can be shown that, the above optimisation can be solved by iteratively solving the following two LASSO regression problems. Specifically, the sparse CCA solutions can be obtained as follows,

$$\widetilde{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha} \in \mathbb{R}^q} \sum_{i=1}^n (\widehat{\boldsymbol{\beta}}_1^T \mathbf{X}_i - \boldsymbol{\alpha}^T \mathbf{Y}_i)^2 + \lambda_\alpha \|\boldsymbol{\alpha}\|_1,$$

$$\widehat{\boldsymbol{\alpha}}_1 = \frac{\widetilde{\boldsymbol{\alpha}}}{\sqrt{\widetilde{\boldsymbol{\alpha}}^T \widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}} \widetilde{\boldsymbol{\alpha}}}}; \quad (5)$$

$$\widetilde{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n (\widehat{\boldsymbol{\alpha}}_1^T \mathbf{Y}_i - \boldsymbol{\beta}^T \mathbf{X}_i)^2 + \lambda_\beta \|\boldsymbol{\beta}\|_1,$$

$$\widehat{\boldsymbol{\beta}}_1 = \frac{\widetilde{\boldsymbol{\beta}}}{\sqrt{\widetilde{\boldsymbol{\beta}}^T \widehat{\boldsymbol{\Sigma}}_{\mathbf{X}} \widetilde{\boldsymbol{\beta}}}}. \quad (6)$$

For sequential sparse CCA directions, $\widehat{\boldsymbol{\alpha}}_k \in \mathbb{R}^q$ and $\widehat{\boldsymbol{\beta}}_k \in \mathbb{R}^p$, $k = 1, 2, \ldots$, we can use the similar iterative penalised least squares formulation after deflation of the data from the previous $(k-1)$ estimated directions. Theoretical results show that this approach can consistently estimate the population directions (any fix number of pairs $k = 1, \ldots, K$) with an overwhelming probability in ultra-high dimensions $\log(p + q) = o(n)$. More importantly, due to its simplicity, the iterative penalised least squares approach can be extremely fast and scalable – even much faster than some existing convex formulations. This approach might be useful for sparse sufficient dimension reduction when the response is multivariate and even high-dimensional.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Xin Zhang is an Associate Professor in Statistics at the Florida State University.

References

Cook, R. D. (1998). *Regression graphics: Ideas for studying regressions through graphics* (Vol. 318). John Wiley & Sons.

Li, K.-C., & Duan, N. (1989). Regression analysis under link violation. *The Annals of Statistics*, 17(3), 1009–1052. https://doi.org/10.1214/aos/1176347254

Mai, Q., & Zhang, X. (2019). An iterative penalized least squares approach to sparse canonical correlation analysis.

Biometrics, *75*(3), 734–744. https://doi.org/10.1111/biom.v75.3

Neykov, M., Lin, Q., & Liu, J. S. (2016). Signed support recovery for single index models in high-dimensions. *Annals of Mathematical Sciences and Applications*, *1*(2), 379–426. https://doi.org/10.4310/AMSA.2016.v1.n2.a5

Neykov, M., Liu, J. S., & Cai, T. (2016). L1-regularized least squares for support recovery of high dimensional single index models with gaussian designs. *The Journal of Machine Learning Research*, *17*(1), 2976–3012.

Ni, L., Cook, R. D., & Tsai, C.-L. (2005). A note on shrinkage sliced inverse regression. *Biometrika*, *92*(1), 242–247. https://doi.org/10.1093/biomet/92.1.242

Tan, K., Shi, L., & Yu, Z. (2020). Sparse sir: Optimal rates and adaptive estimation. *The Annals of Statistics*, *48*(1), 64–85. https://doi.org/10.1214/18-AOS1791

Zhou, J., & He, X. (2008). Dimension reduction based on constrained canonical correlation and variable filtering. *The Annals of Statistics*, *36*(4), 1649–1668. https://doi.org/10.1214/07-AOS529