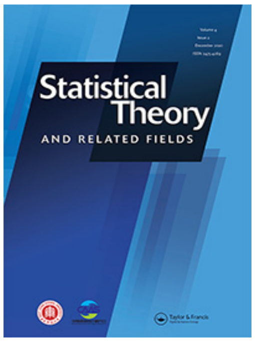

Statistical Theory and Related Fields

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Comment on ‘Review of sparse sufficient dimension reduction’

Michael Declan Power & Yuexiao Dong

To cite this article: Michael Declan Power & Yuexiao Dong (2020) Comment on ‘Review of sparse sufficient dimension reduction’, Statistical Theory and Related Fields, 4:2, 149-150, DOI: 10.1080/24754269.2020.1829394

To link to this article: https://doi.org/10.1080/24754269.2020.1829394

Published online: 13 Oct 2020.

STATISTICAL THEORY AND RELATED FIELDS
2020, VOL. 4, NO. 2, 149–150
https://doi.org/10.1080/24754269.2020.1829394

Comment on 'Review of sparse sufficient dimension reduction'

Michael Declan Power and Yuexiao Dong

Department of Statistical Science, Temple University, Philadelphia, PA, USA

ARTICLE HISTORY Received 18 September 2020; Accepted 23 September 2020

We congratulate the authors on a very interesting overview of sparse sufficient dimension reduction (SDR). Sparse SDR methods are discussed in both the classical $n > p$ setting as well as the high-dimensional $p > n$ setting. Related topics such as model-free variable selection and variable screening are also discussed in a most logical fashion. Last but not least, two new methodological contributions are made in this review paper. Namely, new variable screening methods are proposed as extensions of Yu et al. (2016), and novel sparse SDR methods are discussed following the sparse SIR in Tan et al. (2020).

While all the methods discussed in this review are in the frequentist domain, our comment will focus on sparse SDR through Bayesian methods. Reich et al. (2011) proposed an SDR approach via Bayesian mixture modelling. Take single-index model as an example. Let $\{(\boldsymbol{x}_i, Y_i), i = 1, \ldots, n\}$ be an i.i.d. sample from $(\boldsymbol{x}, Y)$. Let $\boldsymbol{\beta} \in \mathbb{R}^p$ be a basis for the central subspace and let $\lambda_i = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_i$ be the sufficient predictor. Then the conditional distribution of Y_i given $\boldsymbol{x}_i$ can be modelled as

$$p(Y_i \mid \lambda_i) = \sum_{k=1}^{K} p_k(\lambda_i)\mathcal{N}(\mu_k, \sigma_y^2), \qquad (1)$$

where there are K normal mixture components and $p_k(\lambda_i)$ denotes the weight of the kth component. By choosing the weights carefully, model (1) can be expressed as

$$\begin{cases} Y_i \sim \mathcal{N}(\mu_{g_i}, \sigma_y^2) & \\ g_i = k & \text{if } \psi_k < Z_i < \psi_{k+1} \\ Z_i \sim \mathcal{N}(\lambda_i, \sigma_z^2). & \end{cases}$$

Here Z_i is a latent continuous variable, $-\infty = \psi_1 < \psi_2 < \cdots < \psi_{K+1} = \infty$ are cutpoints. By placing a prior on $\boldsymbol{\beta}$ and the cutpoints, one can compute the conditional distributions and carry out the full Bayesian analysis. For SDR without sparsity, the prior for $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is set as $\beta_j \sim \mathcal{N}(0, 1)$, $j = 1, \ldots, p$. To introduce sparsity, a two-component mixture prior is assumed as

$$\begin{cases} \beta_j \sim \mathcal{N}(0, \pi_j + c^2(1 - \pi_j)) \\ \pi_j \sim \mathrm{Ber}(\bar{\pi}), \end{cases}$$

where $0 < c < 1$ is a fixed constant and $\bar{\pi}$ is the prior inclusion probability. If $\pi_j = 1$, then the jth predictor is included in the model. Otherwise the jth predictor is removed from the model. Reich et al. (2011) also discussed a similar Bayesian mixture model for the multiple-index model, and the details are omitted.

Motivated by a frequentist SDR method recently proposed by Fang and Yu (2020), Power and Dong (2020) proposed a new Bayesian approach for sparse sufficient dimension reduction. Let $\mathrm{Var}(\boldsymbol{x}) = \boldsymbol{\Sigma}$, $\mathrm{E}(\boldsymbol{x}) = \boldsymbol{\mu}$, and denote $\{J_1, \ldots, J_H\}$ as a partition for the support of Y. The classical sliced inverse regression (SIR) (Li, 1991) uses the kernel matrix $\boldsymbol{M}_{\mathrm{SIR}} = \sum_{h=1}^{H} \boldsymbol{\xi}_h \boldsymbol{\xi}_h^{\mathrm{T}}/p_h$, where $\boldsymbol{\xi}_h = \boldsymbol{\Sigma}^{-1}\mathrm{E}\{(\boldsymbol{x} - \boldsymbol{\mu})\delta_h\}$ with $\delta_h = I(Y \in J_h)$ and $p_h = E(\delta_h)$ for $h = 1, \ldots, H$. Note that $\boldsymbol{\xi}_h$ can be solved as an optimisation problem

$$\boldsymbol{\xi}_h = \underset{\boldsymbol{\gamma} \in \mathbb{R}^p}{\operatorname{argmin}}\ \mathrm{E}[\{\delta_h - p_h - \boldsymbol{\gamma}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{\mu})\}^2]. \qquad (2)$$

Fang and Yu (2020) then applied the Mallows model averaging (MMA) of Hansen (2007) to solve the least squares problem (2). Power and Dong (2020) utilised Bayesian model averaging (BMA) (Raftery et al., 1997) to solve (2) instead. Similar to MMA, BMA works well for sparse models and may also adapt to models with dense signals. Furthermore, instead of solving for $\boldsymbol{\xi}_h$, $h = 1, \ldots, H$, individually, we may solve for them jointly. Let $\boldsymbol{W} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_H)$ and $\boldsymbol{U} = (\delta_1 - p_1, \ldots, \delta_H - p_H)^{\mathrm{T}}$. Then we have

$$\boldsymbol{W} = \underset{\boldsymbol{\Theta} \in \mathbb{R}^{p \times H}}{\operatorname{argmin}}\ \mathrm{E}[\{\boldsymbol{U} - \boldsymbol{\Theta}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{\mu})\}^{\mathrm{T}}\{\boldsymbol{U} - \boldsymbol{\Theta}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{\mu})\}]. \qquad (3)$$

To the best of our knowledge, there is no frequentist model averaging approach to solve (3). On the

CONTACT Yuexiao Dong ydong@temple.edu Department of Statistical Science, Temple University, Philadelphia, PA 19122, USA

other hand, multi-response BMA (Brown et al., 1998) can be easily adapted to solve (3). As shown in Power and Dong (2020), the multi-response BMA outperforms the frequentist MMA for SDR.

We congratulate the authors again for providing a stimulating review of existing sparse SDR techniques, which should motivate further development of new SDR methods. It is our belief that more Bayesian approaches may be applied for this cause.

References

Brown, P. J., Vannucci, M., & Fearn, T. (1998). Multivariate Bayesian variable selection and prediction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *60*(3), 627–641. https://doi.org/10.1111/rssb.1998.60.issue-3

Fang, F., & Yu, Z. (2020). Model averaging assisted sufficient dimension reduction. *Computational Statistics and Data Analysis*, *152*. https://doi.org/10.1016/j.csda.2020.106993

Hansen, B. E. (2007). Least squares model averaging. *Econometrica*, *75*(4), 1175–1189. https://doi.org/10.1111/ecta.2007.75.issue-4

Li, K. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, *86*(414), 316–327. https://doi.org/10.1080/01621459.1991.10475035

Power, M. D., & Dong, Y. (2020). Bayesian model averaging sufficient dimension reduction. *Statistics and Probability Letters*. Submitted.

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, *92*(437), 179–191. https://doi.org/10.1080/01621459.1997.10473615

Reich, B. J., Bondell, H. D., & Li, L. (2011). Sufficient dimension reduction via Bayesian mixture modeling. *Biometrics*, *67*(3), 886–895. https://doi.org/10.1111/biom.2011.67.issue-3

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

Yu, Z., Dong, Y., & Shao, J. (2016). On marginal sliced inverse regression for ultrahigh dimensional model-free feature selection. *The Annals of Statistics*, *44*(6), 2594–2623. https://doi.org/10.1214/15-AOS1424