

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/74725/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Artemiou, Andreas and Tian, Lipu 2015. Using sliced inverse mean difference for sufficient dimension reduction. *Statistics and Probability Letters* 106 , pp. 184-190. 10.1016/j.spl.2015.07.025

Publishers page: <https://doi.org/10.1016/j.spl.2015.07.025>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Supplementary to “Using Sliced Inverse Mean Difference for Sufficient Dimension Reduction”

Andreas Artemiou^{a,*}, Lipu Tian^b

^a*School of Mathematics, Cardiff University*

^b*Department of Mathematical Sciences, Michigan Technological University*

1. Proof of Theorem 1

Proof of Theorem 1: From the definition of m_d we have:

$$\begin{aligned}
 m_d &= E(\mathbf{X}I(\tilde{Y} = 1)) - E(\mathbf{X}I(\tilde{Y} = -1)) \\
 &= E(E(\mathbf{X}|Y)I(\tilde{Y} = 1)) - E(E(\mathbf{X}|Y)I(\tilde{Y} = -1)) \\
 &= E(E(E(\mathbf{X}|\beta^\top \mathbf{X})|Y)I(\tilde{Y} = 1)) - E(E(E(\mathbf{X}|\beta^\top \mathbf{X})|Y)I(\tilde{Y} = -1)) \\
 &= E(E(\mathbf{P}_\beta^\top(\Sigma)\mathbf{X}|Y)I(\tilde{Y} = 1)) - E(E(\mathbf{P}_\beta^\top(\Sigma)\mathbf{X}|Y)I(\tilde{Y} = -1)) \\
 &= \mathbf{P}_\beta^\top(\Sigma)E(E(\mathbf{X}|Y)I(\tilde{Y} = 1)) - \mathbf{P}_\beta^\top(\Sigma)E(E(\mathbf{X}|Y)I(\tilde{Y} = -1)) \\
 &= \mathbf{P}_\beta^\top(\Sigma)E(\mathbf{X}I(\tilde{Y} = 1)) - \mathbf{P}_\beta^\top(\Sigma)E(\mathbf{X}I(\tilde{Y} = -1)) \\
 &= \mathbf{P}_\beta^\top(\Sigma)(E(\mathbf{X}I(\tilde{Y} = 1)) - E(\mathbf{X}I(\tilde{Y} = -1))) = \mathbf{P}_\beta^\top(\Sigma)m_d
 \end{aligned}$$

□

2. Proof of Theorem 2

Proof of Theorem 2: First we define a function g as follows:

$$\begin{aligned}
 g : \mathbb{R}^{pH} &\rightarrow \mathbb{R}^{p(H-1)} \\
 (\mathbf{a}_1^\top, \dots, \mathbf{a}_H^\top)^\top &\mapsto ((\mathbf{a}_H + \dots + \mathbf{a}_2 - \mathbf{a}_1)^\top, \dots, (\mathbf{a}_H - (\mathbf{a}_1 + \dots + \mathbf{a}_{H-1}))^\top)^\top
 \end{aligned}$$

where $\mathbf{a}_i, i = 1, \dots, H$ are p dimensional vectors.

Now since each column vector of $\mathbf{\Gamma}$ can be created as a function of columns of \mathbf{B} based on (9) it is easy to see that $g(\text{vec}(\mathbf{B})) = \text{vec}(\mathbf{\Gamma})$. Therefore using

*Corresponding Author

the Delta method and applying function g on $\text{vec}(\tilde{\mathbf{Z}}_n)$ we get the desired result. To see how matrix \mathbf{W} is constructed one needs to carefully look at each entry in the range of function g , as $\mathbf{W} = \nabla(g)^\top$. For the first entry we have the p -dimensional vector $\mathbf{v}_1 = \mathbf{a}_H + \dots \mathbf{a}_2 - \mathbf{a}_1$. Then:

$$\begin{aligned} \frac{\partial \mathbf{v}_1}{\partial \mathbf{a}_1} &= -\mathbf{I} \\ \frac{\partial \mathbf{v}_1}{\partial \mathbf{a}_2} &= \mathbf{I} \\ &\vdots \\ \frac{\partial \mathbf{v}_1}{\partial \mathbf{a}_H} &= \mathbf{I} \end{aligned}$$

Similarly one can do this for the other $H-2$ vectors in the range of g . Therefore it is easy to see that \mathbf{W} takes the form:

$$\mathbf{W} = \begin{bmatrix} -\mathbf{I} & \mathbf{I} & \mathbf{I} \dots & \mathbf{I} & \mathbf{I} \\ -\mathbf{I} & -\mathbf{I} & \mathbf{I} \dots & \mathbf{I} & \mathbf{I} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\mathbf{I} & -\mathbf{I} & -\mathbf{I} \dots & -\mathbf{I} & \mathbf{I} \end{bmatrix}$$

□

3. Asymptotic results for OVA

In this section we derive the asymptotic distribution of $\mathbf{\Gamma}_{\text{OVA}}$.

The column vectors of $\mathbf{\Gamma}$ are $m_{\text{OVA}}^Z(r, s), 1 \leq r < s \leq H$. We know that

$$\begin{aligned} m_{\text{OVA}}^Z(q, s) &= E(\mathbf{Z}I(Y \in A_s)) - E(\mathbf{Z}I(Y \in A_r)) \\ &= p_s E(\mathbf{Z}|Y \in A_s) - p_r E(\mathbf{Z}|Y \in A_r) \end{aligned} \tag{1}$$

where A_i denotes the i^{th} slice and p_i the proportion of points in slice A_i .

Using the result of Lemma 1 together with the Delta method one can prove the following result which gives the asymptotic distribution of $\mathbf{\Gamma}$. The proof is similar to the one for Theorem 2, therefore we omit it.

Theorem 3.

$$\sqrt{n}\text{vec}(\hat{\Gamma} - \Gamma) \xrightarrow{\mathcal{D}} N_{p\binom{H}{2}}(0, \mathbf{W}\Delta\mathbf{W}^\top)$$

where \mathbf{W} is a $p\binom{H}{2} \times pH$ matrix which is an $\binom{H}{2} \times H$ array of $p \times p$ matrices that can take 3 possible values; positive or negative identity matrices or zero matrices. Each row of the array corresponds to a pair (r, s) where $1 \leq r < s \leq H$. Denoting by \mathbf{W}_{ij} the element at the i^{th} row and j^{th} column of \mathbf{W} , $\mathbf{W}_{ij} = \mathbf{I}$ if $j = s$ at the corresponding row, $\mathbf{W}_{ij} = -\mathbf{I}$ if $j = r$ at the corresponding row or $\mathbf{0}$ otherwise.

This result can be used to provide asymptotic sequential tests for the estimation of the dimension of the CS in the same manner as was discussed for LVR in Section 3.3 of the article.