# 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:

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

□

## 2. Proof of Theorem 2

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

$$\begin{aligned}
g :& \mathbb{R}^{pH} \to \mathbb{R}^{p(H-1)} \\
(\boldsymbol{a}_1^{\intercal}, \ldots, \boldsymbol{a}_H^{\intercal})^{\intercal} \mapsto& ((\boldsymbol{a}_H + \ldots + \boldsymbol{a}_2 - \boldsymbol{a}_1)^{\intercal}, \ldots, (\boldsymbol{a}_H - (\boldsymbol{a}_1 + \ldots + \boldsymbol{a}_{H-1}))^{\intercal})^{\intercal}
\end{aligned}$$

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

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


*Corresponding Author

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

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

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

$$\boldsymbol{W} = \begin{bmatrix} -\boldsymbol{I} & \boldsymbol{I} & \boldsymbol{I} \cdots & \boldsymbol{I} & \boldsymbol{I} \\ -\boldsymbol{I} & -\boldsymbol{I} & \boldsymbol{I} \cdots & \boldsymbol{I} & \boldsymbol{I} \\ \vdots & \vdots & \vdots \ddots & \vdots & \vdots \\ -\boldsymbol{I} & -\boldsymbol{I} & -\boldsymbol{I} \cdots & -\boldsymbol{I} & \boldsymbol{I} \end{bmatrix}$$

□

## 3. Asymptotic results for OVA

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

The column vectors of $\boldsymbol{\Gamma}$ are $m^{Z}_{\mathrm{OVA}}(r, s), 1 \leq r < s \leq H$. We know that

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

where $A_i$ denotes the $i^{\text{th}}$ 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 $\boldsymbol{\Gamma}$. The proof is similar to the one for Theorem 2, therefore we omit it.

**Theorem 3.**

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

*where* $\boldsymbol{W}$ *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* $\boldsymbol{W}_{ij}$ *the element at the* $i^{\text{th}}$ *row and* $j^{\text{th}}$ *column of* $\boldsymbol{W}$, $\boldsymbol{W}_{ij} = \boldsymbol{I}$ *if* $j = s$ *at the corresponding row,* $\boldsymbol{W}_{ij} = -\boldsymbol{I}$ *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.