Supplement to “Understanding and Utilizing
the Linearity Condition in Dimension Reduction”

Yanyuan Ma, Fei Jiang, Masayuki Henmi

Penn State University, University of Hong Kong, Institute of Statistical Mathematics

S.1 Conditions in Theorem

These are conditions needed to establish the asymptotic properties of $\hat{\beta}$ in Theorem.

(C1) The univariate kernel function $K(\cdot)$ is symmetric, has compact support and is Lipschitz continuous on its support. It satisfies

$$\int K(u)du = 1, \int u^i K(u)du = 0(i = 1, \ldots, m - 1), 0 \neq \int |u|^m K(u)du < \infty.$$ 

Thus $K$ is a $m$-th order kernel. The $d$-dimensional kernel function is a product of $d$ univariate kernel functions, that is, $K_h(u) = K(u/h)h^d = \Pi_{j=1}^d K_h(u_j) = \Pi_{j=1}^d K(u_j/h)/h^d$ for $u = (u_1, \ldots, u_d)^T$. Without causing misunderstanding, we use the same $K$ regardless of the dimension of its argument.

(C2) The probability density function of $\beta^T x$, denoted by $f(\beta^T x)$, is bounded away from zero and infinity.

(C3) Let $r(\beta^T x) = E[a(x) | \beta^T x]f(\beta^T x)$. The $(m-1)$-th derivatives of $r(\beta^T x)$ and $f(\beta^T x)$ are locally Lipschitz-continuous as functions of $\beta^T x$.

(C4) The bandwidth $h = O(n^{-\kappa})$ for $(2m)^{-1} < \kappa < (2d)^{-1}$. 
S.2 Proof of the result regarding $\tilde{\beta}$ in Theorem 1

Since $\hat{\alpha}(\beta)$ solves (34), we obtain the Taylor expansion

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A(\beta^T X_i) \left[ a(X_i) - m_\alpha(\beta^T X_i, \hat{\alpha}(\beta)) \right]$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A(\beta^T X_i) \left[ a(X_i) - m_\alpha(\beta^T X_i, \alpha_0(\beta)) \right]$$

$$- \frac{1}{n} \sum_{i=1}^{n} A(\beta^T X_i) \left[ \frac{\partial m(\beta^T X_i, \alpha)}{\partial \alpha} \right] \left|_{\alpha=\alpha_0(\beta)} \right. \sqrt{n}(\hat{\alpha}(\beta) - \alpha_0(\beta)) + o_p(1).$$

This leads to

$$\sqrt{n}(\hat{\alpha}(\beta) - \alpha_0(\beta)) = \frac{1}{\sqrt{n} B_1^{-1}} \sum_{i=1}^{n} A(\beta^T X_i) \left[ a(X_i) - m_\alpha(\beta^T X_i, \alpha_0(\beta)) \right] + o_p(1).$$

Let $m_\beta(\beta^T X_i, \alpha_0(\beta)) = \partial m(\beta^T X_i, \alpha)/\partial \text{vecl}(\beta)^T |_{\alpha=\alpha_0(\beta)}$, $\hat{\alpha}_\beta(\beta) = \partial \hat{\alpha}(\beta)/\partial \text{vecl}(\beta)^T$ and $\alpha_{0, \beta}(\beta) = \partial \alpha_0(\beta)/\partial \text{vecl}(\beta)^T$. Since $\tilde{\beta}$ solves (33), plugging the expression of $\hat{\alpha}(\beta) - \alpha_0(\beta)$, we further have

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(Y_i) \left[ a(X_i) - m(\hat{\beta}^T X_i, \hat{\alpha}(\beta)) \right]^T$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(Y_i) \left[ a(X_i) - m(\beta^T X_i, \hat{\alpha}(\beta)) \right]^T$$

$$- \frac{1}{n} \sum_{i=1}^{n} g(Y_i) \sqrt{n}(\text{vecl}(\tilde{\beta} - \beta))^T \left[ m_\beta(\beta^T X_i, \alpha_0(\beta)) + m_\alpha(\beta^T X_i, \hat{\alpha}(\beta)) \hat{\alpha}_\beta(\beta) \right]^T + o_p(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(Y_i) \left[ a(X_i) - m(\beta^T X_i, \alpha_0(\beta)) \right]^T - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(Y_i) \left[ m_\alpha(\beta^T X_i, \alpha_0(\beta)) \right] \left[ \hat{\alpha}(\beta) - \alpha_0(\beta) \right]^T$$

$$- \frac{1}{n} \sum_{i=1}^{n} g(Y_i) \sqrt{n}(\text{vecl}(\tilde{\beta} - \beta))^T \left[ m_\beta(\beta^T X_i, \alpha_0(\beta)) + m_\alpha(\beta^T X_i, \alpha_0(\beta)) \alpha_{0, \beta}(\beta) \right]^T + o_p(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(Y_i) \left[ a(X_i) - m(\beta^T X_i, \alpha_0(\beta)) \right]^T$$

$$- \frac{1}{n^{1/2}} \sum_{i=1}^{n} g(Y_i) \left( B_1^{-1} A(\beta^T X_i) \left[ a(X_i) - m(\beta^T X_i, \alpha_0(\beta)) \right] \right)^T \left[ m_\beta(\beta^T X_i, \alpha_0(\beta)) + m_\alpha(\beta^T X_i, \alpha_0(\beta)) \alpha_{0, \beta}(\beta) \right]^T + o_p(1).$$
We vectorize the above display and write the relation equivalently as

\[
0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \text{vec} \left( g(Y_i) \left[ a(X_i) - m(\beta^T X_i, \alpha_0(\beta)) \right] \right)
\]

\[
- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{n} \sum_{i=1}^{n} m_{\alpha}(\beta^T X_i, \alpha_0(\beta)) \otimes g(Y_i) \left( B_1^{-1} A(\beta^T X_i) \right) \left[ a(X_i) - m(\beta^T X_i, \alpha_0(\beta)) \right]
\]

\[
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_{\beta}(\beta^T X_i, \alpha_0(\beta)) + m_{\alpha}(\beta^T X_i, \alpha_0(\beta)) \alpha_{0, \beta}(\beta) \otimes g(Y_i) \sqrt{n} \text{vec}(\tilde{\beta} - \beta) + o_p(1)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_{p_n} \otimes g(Y_i) \left[ a(X_i) - m(\beta^T X_i, \alpha_0(\beta)) \right]
\]

\[
- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} B_2 \left( B_1^{-1} A(\beta^T X_i) \right) \left[ a(X_i) - m(\beta^T X_i, \alpha_0(\beta)) \right]
\]

\[
- E \left( m_{\beta}(\beta^T X, \alpha_0(\beta)) + m_{\alpha}(\beta^T X, \alpha_0(\beta)) \alpha_{0, \beta}(\beta) \otimes g(Y) \right) \sqrt{n} \text{vec}(\tilde{\beta} - \beta) + o_p(1)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ I_{p_n} \otimes g(Y_i) - B_2 B_1^{-1} A(\beta^T X_i) \right] \left[ a(X_i) - m(\beta^T X_i, \alpha_0(\beta)) \right]
\]

\[
- E \left( m_{\beta}(\beta^T X, \alpha_0(\beta)) + m_{\alpha}(\beta^T X, \alpha_0(\beta)) \alpha_{0, \beta}(\beta) \right) \otimes E\left[ g(Y) \mid \beta^T X \right] \sqrt{n} \text{vec}(\tilde{\beta} - \beta)
\]

+ o_p(1).

(S.1)

Because \( \hat{\alpha}(\beta) \) solves \([34]\) and since we assume the model is correct, \( \alpha_0(\beta) = \lim_{n \to \infty} \hat{\alpha}(\beta) \) and \( E\{a(X) \mid \beta^T X\} = m(\beta^T X, \alpha_0(\beta)) \) for any \( \beta \). This leads to

\[
E \left( [a(X) - m(\beta^T X, \alpha_0(\beta))] \otimes E\left[ g(Y) \mid \beta^T X \right] \right) = 0
\]

for any \( \beta \), hence

\[
E \left( \left[ -m_{\beta}(\beta^T X, \alpha_0(\beta)) - m_{\alpha}(\beta^T X, \alpha_0(\beta)) \alpha_{0, \beta}(\beta) \right] \otimes E\left[ g(Y) \mid \beta^T X \right] \right)
\]

\[
+ E \left( \left[ a(X) - m(\beta^T X, \alpha_0(\beta)) \right] \otimes \frac{\partial E\left[ g(Y) \mid \beta^T X \right]}{\partial \text{vec}(\beta^T)} \right)
\]

\[
= E \left( \left[ -m_{\beta}(\beta^T X, \alpha_0(\beta)) - m_{\alpha}(\beta^T X, \alpha_0(\beta)) \alpha_{0, \beta}(\beta) \right] \otimes E\left[ g(Y) \mid \beta^T X \right] \right) + \Sigma_A
\]

\[
= 0.
\]

We thus can rewrite (S.1) as

\[
0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ I_{p_n} \otimes g(Y_i) - B_2 B_1^{-1} A(\beta^T X_i) \right] \left[ a(X_i) - m(\beta^T X_i, \alpha_0(\beta)) \right]
\]

\[
- \Sigma_A \sqrt{n} \text{vec}(\tilde{\beta} - \beta) + o_p(1),
\]

which leads to the result in Theorem 1 \( \square \)
S.3 Proof of Theorem 2

Because $\hat{\alpha}(\beta)$ solves $[15]$, we obtain the Taylor expansion

\[
0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_{\alpha}^T(\beta^T X_i, \hat{\alpha}(\beta)) Q^{-1}(\beta^T X) \left[ a(X_i) - m(\beta^T X_i, \hat{\alpha}(\beta)) \right]
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_{\alpha}^T(\beta^T X_i, \alpha_0(\beta)) Q^{-1}(\beta^T X) \left[ a(X_i) - m(\beta^T X_i, \alpha_0(\beta)) \right]
\]

\[
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\beta m_{\alpha}^T(\beta^T X_i, \alpha)}{\tilde{\beta} \alpha_j} |_{\alpha = \alpha_0(\beta)} Q^{-1}(\beta^T X) \left[ a(X_i) - m(\beta^T X_i, \alpha_0(\beta)) \right] \sqrt{n} \hat{\alpha}_j(\beta) - \alpha_0(\beta)) + o_p(1)
\]

\[
- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_{\alpha}^T(\beta^T X_i, \alpha_0(\beta)) Q^{-1}(\beta^T X) \frac{\hat{m}(\beta^T X_i, \alpha)}{\hat{\beta} \alpha_j} |_{\alpha = \alpha_0(\beta)} \sqrt{n} (\hat{\alpha}(\beta) - \alpha_0(\beta)) + o_p(1)
\]

This leads to

\[
\sqrt{n} (\hat{\alpha}(\beta) - \alpha_0(\beta)) = \frac{1}{\sqrt{n}} B_3^{-1} \sum_{i=1}^{n} m_{\alpha}^T(\beta^T X_i, \hat{\alpha}(\beta)) Q^{-1}(\beta^T X) \left[ a(X_i) - m(\beta^T X_i, \alpha_0(\beta)) \right] + o_p(1).
\]

Following the same derivation as that in the proof of Theorem [14] we then obtain the expansion of $\beta$.

It is easy to verify that

\[
E \left\{ \left[ I_{p_n} \otimes E(\beta^T X) - B_2 B_3^{-1} m_{\alpha}^T(\beta^T X, \alpha_0(\beta)) Q^{-1}(\beta^T X) \left[ a(X) - m(\beta^T X, \alpha_0(\beta)) \right] \right]^T \right\}
\]

\[
= \text{trace} E \left\{ \left[ I_{p_n} \otimes E(\beta^T X) - B_2 B_3^{-1} m_{\alpha}^T(\beta^T X, \alpha_0(\beta)) Q^{-1}(\beta^T X) \left[ a(X) - m(\beta^T X, \alpha_0(\beta)) \right] \right]^T \right\}
\]

\[
= \text{trace} E \left\{ \left[ B_2 B_3^{-1} m_{\alpha}^T(\beta^T X, \alpha_0(\beta)) Q^{-1}(\beta^T X) \left[ a(X) - m(\beta^T X, \alpha_0(\beta)) \right] \right]^T \left[ I_{p_n} \otimes E(\beta^T X) - B_2 B_3^{-1} m_{\alpha}^T(\beta^T X, \alpha_0(\beta)) Q^{-1}(\beta^T X) \right] \right\}
\]

\[
= \text{trace} E \left\{ B_2 B_3^{-1} m_{\alpha}^T(\beta^T X, \alpha_0(\beta)) \left[ I_{p_n} \otimes E(\beta^T X) - B_2 B_3^{-1} m_{\alpha}^T(\beta^T X, \alpha_0(\beta)) Q^{-1}(\beta^T X) \right] \right\}
\]

\[
= 0.
\]
S.4 Proof of Theorem

Because no constraints are imposed on \( f_1(\beta^T X) \) other than it is a valid pdf, hence its nuisance tangent space contains all mean zero functions of \( \beta^T X \). In addition to being a valid conditional pdf, \( f_2(\beta^T X, \epsilon) \) is subject to the mean zero condition. This restricts the corresponding nuisance tangent space, and it is easy to verify that it has the form given in \( \Lambda_2 \). The results of \( \Lambda_3 \) can be similarly derived as \( \Lambda_1 \), by treating \( Y \) as the random variable. We omit the details of the derivation of \( \Lambda_1, \Lambda_2 \) and \( \Lambda_3 \) since they involve only standard practice. It is also easy to verify that the three spaces are orthogonal to each other, hence we obtain the results concerning \( \Lambda \).

It is also not hard to see that \( \Lambda_1^\perp \) contains all the functions \( g(X, Y) \) such that \( E[g(X, Y) \mid \beta^T X] = 0 \), \( \Lambda_2^\perp \) contains all the functions \( g(X, Y) \) such that \( E[g(X, Y) \mid \beta^T X, \epsilon] \) has the form \( a(\beta^T X) + A(\beta^T X) \epsilon \), and \( \Lambda_3^\perp \) contains all the functions \( g(X, Y) \) such that \( E[g(X, Y) \mid \beta^T X, Y] \) has the form \( a(\beta^T X) \). Thus, taking the intersection of \( \Lambda_1^\perp, \Lambda_2^\perp \) and \( \Lambda_3^\perp \), we obtain \( \Lambda^\perp \) as described in Theorem 3.

To obtain the efficient score, we first calculate the score function with respect to the parameter of interest contained in \( \beta \), i.e. \( \beta_2 \). The score function is

\[
S_{\beta_2} = \text{vec} \left[ X_2 \left\{ \frac{\partial \log f_1(\beta^T X)}{\partial \epsilon_{1} \beta} + \frac{\partial \log f_2(\beta^T X, \epsilon)}{\partial \epsilon_{1} \beta} - \frac{\partial \log f_2(\beta^T X, \epsilon) \cdot \partial m(\beta^T X, \beta)}{\partial \epsilon_{1} \beta} \frac{\partial m(\beta^T X, \beta)}{\partial \epsilon_{1} \beta} \right\} - \frac{\partial m^T(\beta^T X, \beta_2)}{\partial \epsilon_{1} \beta} \frac{\partial \log f_2(\beta^T X, \epsilon)}{\partial \epsilon_{1} \beta} \right].
\]

We now decompose the score function into \( S_{\beta_2} = S_{\text{eff}} + R \), where

\[
S_{\text{eff}} = \text{vec} \left( \epsilon_2 \frac{\partial \log f_1(\beta^T X)}{\partial \epsilon_{1} \beta} + \frac{\partial Q_2(\beta^T X)}{\partial \epsilon_{1} \beta} \left[ I_{d} \otimes \left\{ Q_2^{-1}(\beta^T X) \epsilon_2 \right\} \right] + m(\beta^T X, \beta_2) \epsilon_2^T \right) \times Q_2^{-1}(\beta^T X) \frac{\partial m(\beta^T X, \beta_2)}{\partial \epsilon_{1} \beta} + \epsilon_2 \frac{\partial \log f_2(\beta^T X, Y)}{\partial \epsilon_{1} \beta} \right) + \frac{\partial m^T(\beta^T X, \beta_2)}{\partial \epsilon_{1} \beta} Q_2^{-1}(\beta^T X) \epsilon_2,
\]
S.4 Proof of Theorem 3

and

\[
R = \text{vec} \left( m(\beta^T X, \beta_2) \frac{\partial \log f_1(\beta^T X)}{\partial X^\beta} + m(\beta^T X, \beta_2) \frac{\partial \log f_2(\beta^T X, \epsilon_2)}{\partial \epsilon_2} + \epsilon_2 \frac{\partial \log f_2(\beta^T X, \epsilon_2)}{\partial X^\beta} \right) \\
- \frac{\partial Q_2(\beta^T X)}{\partial X^\beta} \left[ I_d \otimes \left( Q_2^{-1}(\beta^T X) \epsilon_2 \right) \right] - m(\beta^T X, \beta_2) \left\{ \frac{\partial \log f_2(\beta^T X, \epsilon_2)}{\partial \epsilon_2} + \epsilon_2 \frac{\partial Q_2^{-1}(\beta^T X)}{\partial \epsilon_2} \right\} \\
\times \frac{\partial m(\beta^T X, \beta_2)}{\partial X^\beta} - \epsilon_2 \left\{ \frac{\partial \log f_2(\beta^T X, \epsilon_2)}{\partial \epsilon_2} + \epsilon_2 \frac{\partial m(\beta^T X, \beta_2)}{\partial \epsilon_2} + m(\beta^T X, \beta_2) \left\{ \frac{\partial m(\beta^T X, \beta_2)}{\partial \epsilon_2} + \epsilon_2 \frac{\partial Q_2^{-1}(\beta^T X)}{\partial \epsilon_2} \right\} \right\}.
\]

Here, when taking derivative of a matrix with respect to a row vector, we obtain a block row matrix, with the \(j\)th block element is the derivative of the matrix with respect to the \(j\)th element of the vector. We can easily check that indeed \(S_{\beta_2} = S_{\text{eff}} + R\). It is also straightforward to verify that \(S_{\text{eff}} \in \Lambda^+\). In addition, we easily obtain

\[
R_1 = \text{vec} \left( m(\beta^T X, \beta_2) \frac{\partial \log f_1(\beta^T X)}{\partial X^\beta} + \frac{\partial m(\beta^T X, \beta_2)}{\partial X^\beta} \right) \in \Lambda_1
\]

\[
R_3 = m(\beta^T X, \beta_2) \frac{\partial \log f_3(\beta^T X, Y)}{\partial X^\beta} \in \Lambda_3.
\]

Finally, using the relation

\[
E \left\{ \frac{\partial \log f_2(\beta^T X, \epsilon_2)}{\partial X^\beta} \mid \beta^T X \right\} = 0
\]

\[
E \left\{ \epsilon_2 \frac{\partial \log f_2(\beta^T X, \epsilon_2)}{\partial X^\beta} \mid \beta^T X \right\} = 0
\]

\[
E \left\{ \frac{\partial \log f_2(\beta^T X, \epsilon_2)}{\partial \epsilon_2} \mid \beta^T X \right\} = 0
\]

\[
E \left\{ \epsilon_2 \frac{\partial \log f_2(\beta^T X, \epsilon_2)}{\partial \epsilon_2} \mid \beta^T X \right\} = -I_{p-d}
\]

\[
E \left\{ \epsilon_2 \frac{\partial \log f_2(\beta^T X, \epsilon_2)}{\partial \epsilon_2} \mid \beta^T X \right\} = 0,
\]

through tedious but straightforward calculation, we can verify that

\[
R_2 = \text{vec} \left( m(\beta^T X, \beta_2) \frac{\partial \log f_2(\beta^T X, \epsilon_2)}{\partial X^\beta} + \epsilon_2 \frac{\partial \log f_2(\beta^T X, \epsilon_2)}{\partial X^\beta} \right) \\
- \frac{\partial Q_2(\beta^T X)}{\partial X^\beta} \left[ I_d \otimes \left( Q_2^{-1}(\beta^T X) \epsilon_2 \right) \right] - m(\beta^T X, \beta_2) \left\{ \frac{\partial \log f_2(\beta^T X, \epsilon_2)}{\partial \epsilon_2} + \epsilon_2 \frac{\partial Q_2^{-1}(\beta^T X)}{\partial \epsilon_2} \right\} \\
\times \frac{\partial m(\beta^T X, \beta_2)}{\partial X^\beta} - \epsilon_2 \left\{ \frac{\partial \log f_2(\beta^T X, \epsilon_2)}{\partial \epsilon_2} + \epsilon_2 \frac{\partial m(\beta^T X, \beta_2)}{\partial \epsilon_2} + m(\beta^T X, \beta_2) \left\{ \frac{\partial m(\beta^T X, \beta_2)}{\partial \epsilon_2} + \epsilon_2 \frac{\partial Q_2^{-1}(\beta^T X)}{\partial \epsilon_2} \right\} \right\}.
\]

We can see that \(R = R_1 + R_2 + R_3\), hence \(R \in \Lambda\). This shows that \(S_{\text{eff}}\) is indeed the efficient score.
S.5 Proof of Theorem 4

Similar to the derivation in proving Theorem 3, the form of $\Lambda_1$, $\Lambda_3$ are unchanged. Regarding $\Lambda_2$, because in addition to being a valid conditional pdf, $f_2(\beta^T X, \epsilon_2)$ is subject to the mean zero and constant variance conditions, the corresponding nuisance tangent space is further restricted. It is also easy to verify that it has the form given in $\Lambda_2$. The orthogonality of the three spaces $\Lambda_1, \Lambda_2, \Lambda_3$ still holds, hence we obtain the results concerning $\Lambda$.

Obviously, $\Lambda_1^\perp$ and $\Lambda_3^\perp$ remain unchanged from those in Theorem 3. $\Lambda_2^\perp$ contains all the functions $g(X, Y)$ such that $E[g(X, Y) \mid \beta^T X, \epsilon_2]$ has the form $a(\beta^T X) + A(\beta^T X)\epsilon_2 + B(\beta^T X)\epsilon_2^T T$. Thus, taking the intersection of $\Lambda_1^\perp, \Lambda_2^\perp, \Lambda_3^\perp$, we obtain $\Lambda^\perp$ as described in Theorem 4. Note that our construction ensures that $E(\epsilon_2 \epsilon_2^T \mid \beta^T X) = 0$. We then can write

$$\Lambda_1 = \left[ h(\beta^T X) : E[h(\beta^T X)] = 0, E[h^T(\beta^T X)h(\beta^T X)] < \infty, h(\beta^T X) \in R^{(p-d)d} \right]$$
$$\Lambda_2 = \left[ h(\beta^T X, \epsilon_2) : E[h(\beta^T X, \epsilon_2) \mid \beta^T X] = 0, E[\epsilon_2 h^T(\beta^T X, \epsilon_2) \mid \beta^T X] = 0, \right.$$
$$E\{vh^T(\beta^T X, \epsilon_2) \mid \beta^T X\} = 0, E[h^T(\beta^T X, \epsilon_2)h(\beta^T X, \epsilon_2)] < \infty, \left. h(\beta^T X, \epsilon_2) \in R^{(p-d)d} \right]$$
$$\Lambda_3 = \left[ h(\beta^T X, Y) : E[h(\beta^T X, Y) \mid \beta^T X] = 0, E[h^T(\beta^T X, Y)h(\beta^T X, Y)] < \infty, \right.$$
$$h(\beta^T X, Y) \in R^{(p-d)d} \left. \right]$$
$$\Lambda^\perp = \left[ g(X, Y) : E[g(X, Y) \mid \beta^T X, \epsilon_2] = A(\beta^T X)\epsilon_2 + B(\beta^T X)\nu, \right.$$
$$E[g(X, Y) \mid \beta^T X, Y] = 0, E[g^T(X, Y)g(X, Y)] < \infty, g(X, Y) \in R^{(p-d)d} \left. \right].$$

To obtain the efficient score, we first calculate the score function with respect to the parameter of interest contained in $\beta$, i.e. $\beta_2$. The score function is different from that in
The key difference is in how the score function should be decomposed, reflecting the change of the spaces $\Lambda$ and $\Lambda^I$. We can rewrite

$$S_{\beta_2} = \text{vec} \left[ X_2 \left\{ \frac{\partial \log f_1 (\beta^T X, \hat{e}_2)}{\partial X^T \beta} + \frac{\partial \log f_2 (\beta^T X, \hat{e}_2)}{\partial X^T \beta} - \frac{\partial \log f_2 (\beta^T X, \hat{e}_2)}{\partial \hat{e}_2^T} D^{-1} (\beta_2) \frac{\partial \mathbf{m}(\beta^T X, \beta_2)}{\partial \hat{e}_2^T} \right\} \right]$$

$$- \text{vec} \left\{ \mathbf{m}(\beta^T X, \beta_2) \frac{\partial \log f_2 (\beta^T X, \hat{e}_2)}{\partial \hat{e}_2^T} D^{-1} (\beta_2) \frac{\partial \mathbf{m}(\beta^T X, \beta_2)}{\partial \hat{e}_2^T} \right\}$$

$$- \frac{\partial \mathbf{m}(\beta^T X, \beta_2)}{\partial \hat{e}_2^T} \left\{ \frac{\partial D^{-1} (\beta_2)}{\partial \hat{e}_2^T} \right\} \text{vec} \left\{ \frac{\partial \log f_2 (\beta^T X, \hat{e}_2)}{\partial \hat{e}_2^T} \right\}$$

$$+ \frac{\partial \log f_2 (\beta^T X, \hat{e}_2)}{\partial \hat{e}_2^T} \left\{ \frac{\partial D^{-1} (\beta_2)}{\partial \hat{e}_2^T} \right\} \text{vec} \left\{ \frac{\partial \log f_2 (\beta^T X, \hat{e}_2)}{\partial \hat{e}_2^T} \right\}$$

$$- \left\{ \frac{\partial \mathbf{m}(\beta^T X, \beta_2)}{\partial \beta^T X} \right\} \left\{ \frac{\partial \mathbf{m}(\beta^T X, \beta_2)}{\partial \beta^T X} \right\}$$

$$+ C_1(\beta_2) \text{vec} \left\{ \frac{\partial \log f_2 (\beta^T X, \hat{e}_2)}{\partial \hat{e}_2^T} \right\}$$

$$- \frac{\partial \log \det |D(\beta_2)|}{\partial \hat{e}_2^T}$$

$$+ K_1(\beta^T X, \beta_2) \text{vec} \left\{ \frac{\partial \log f_2 (\beta^T X, \hat{e}_2)}{\partial \hat{e}_2^T} \right\}$$

$$- \frac{\partial \log \det |D(\beta_2)|}{\partial \hat{e}_2^T}$$
We decompose the score function into $S_{\beta_2} = S_{\text{eff}} + R$, where $R \in \Lambda$ and $S_{\text{eff}} \in \Lambda^*$ and hence is the efficient score and. Here,

$$S_{\text{eff}} = \text{vec} \left( D(\beta_2)e_{\beta_2} \frac{\partial \log f_1(\beta^T X)}{\partial X^T \beta} + D(\beta_2)e_{\beta_2} \frac{\partial \log f_2(\beta^T X, Y)}{\partial X^T \beta} \right)$$

$$-K_1(\beta^T X, \beta_2)e_{\beta_2} + K_2(\beta^T X, \beta_2)v - K_4(\beta^T X, \beta_2)v$$

and

$$R = \text{vec} \left\{ m(\beta^T X, \beta_2) \frac{\partial \log f_1(\beta^T X)}{\partial X^T \beta} + m(\beta^T X, \beta_2) \frac{\partial \log f_3(\beta^T X, Y)}{\partial X^T \beta} + m(\beta^T X, \beta_2) \frac{\partial \log f_2(\beta^T X, \bar{\beta}_2)}{\partial \bar{\beta}_2} \right\}$$

$$+ \text{vec} \left( D(\beta_2)e_{\beta_2} \frac{\partial \log f_2(\beta^T X, \bar{\beta}_2)}{\partial \bar{\beta}_2} \right) + K_1(\beta^T X, \beta_2)e_{\beta_2} + K_2(\beta^T X, \beta_2)v + K_3(\beta^T X, \beta_2)\text{vec} \left\{ \bar{\beta}_2 \frac{\partial \log f_2(\beta^T X, \bar{\beta}_2)}{\partial \bar{\beta}_2} + I_{p-d} \right\}$$

$$+ K_4(\beta^T X, \beta_2)v - \frac{\partial \log \det \{D(\beta_2)\}}{\partial \text{vec}(\beta_2)} - K_3(\beta^T X, \beta_2)v(\text{vec}(I_{p-d}).$$

It is obvious that $S_{\text{eff}} \in \Lambda^*$. Careful and tedious calculations, through grouping the terms in $R$ as the second, the third, the fourth+fifth, the sixth+seventh+eighth, nineth+tenth, and first+eleventh+twelfth terms, verify that $R \in \Lambda$. $\square$