Statistica Sinica

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 1

.

These are conditions needed to establish the asymptotic properties of $\widehat{\boldsymbol\beta}$ in Theorem 1.

(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, $K_h(\mathbf{u}) = K(\mathbf{u}/h)/h^d = \Pi_{j=1}^d K_h(u_j) = \Pi_{j=1}^d K(u_j/h)/h^d$ for $\mathbf{u} = (u_1, \ldots, u_d)^{\mathrm{T}}$. Without causing misunderstanding, we use the same K regardless of the dimension of its argument.

(C2) The probability density function of $\boldsymbol\beta^{\mathrm{T}}\mathbf{x}$, denoted by $f\left(\boldsymbol\beta^{\mathrm{T}}\mathbf{x}\right)$, is bounded away from zero and infinity.

(C3) Let $\mathbf{r}(\boldsymbol\beta^{\mathrm{T}}\mathbf{x}) = E\{\mathbf{a}(\mathbf{x}) \mid \boldsymbol\beta^{\mathrm{T}}\mathbf{x}\}f(\boldsymbol\beta^{\mathrm{T}}\mathbf{x})$. The $(m-1)$-th derivatives of $\mathbf{r}(\boldsymbol\beta^{\mathrm{T}}\mathbf{x})$ and $f(\boldsymbol\beta^{\mathrm{T}}\mathbf{x})$ are locally Lipschitz-continuous as functions of $\boldsymbol\beta^{\mathrm{T}}\mathbf{x}$.

(C4) The bandwidth $h = O(n^{-\kappa})$ for $(2m)^{-1} < \kappa < (2d)^{-1}$.

S.2 Proof of the result regarding $\breve{\boldsymbol\beta}$ in Theorem 1

.

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

$$\begin{aligned}
\mathbf{0} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{A}(\boldsymbol\beta^{\mathrm T}\mathbf{X}_i)\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\widehat{\boldsymbol\alpha}(\boldsymbol\beta)\}\right]\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{A}(\boldsymbol\beta^{\mathrm T}\mathbf{X}_i)\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]\\
&\quad -\frac{1}{n}\sum_{i=1}^n \mathbf{A}(\boldsymbol\beta^{\mathrm T}\mathbf{X}_i)\frac{\partial\mathbf{m}(\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha)}{\partial\boldsymbol\alpha^{\mathrm T}}\Big|_{\boldsymbol\alpha=\boldsymbol\alpha_0(\boldsymbol\beta)}\sqrt{n}\{\widehat{\boldsymbol\alpha}(\boldsymbol\beta)-\boldsymbol\alpha_0(\boldsymbol\beta)\}+o_p(1).
\end{aligned}$$

This leads to

$$\sqrt{n}\{\widehat{\boldsymbol\alpha}(\boldsymbol\beta)-\boldsymbol\alpha_0(\boldsymbol\beta)\} = \frac{1}{\sqrt{n}}\mathbf{B}_1^{-1}\sum_{i=1}^n \mathbf{A}(\boldsymbol\beta^{\mathrm T}\mathbf{X}_i)\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]+o_p(1).$$

Let $\mathbf{m}_{\boldsymbol\beta}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\} = \partial\mathbf{m}(\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha)/\partial\mathrm{vecl}(\boldsymbol\beta)^{\mathrm T}|_{\boldsymbol\alpha=\boldsymbol\alpha_0(\boldsymbol\beta)}$, $\widehat{\boldsymbol\alpha}_{\boldsymbol\beta}(\boldsymbol\beta) = \partial\widehat{\boldsymbol\alpha}(\boldsymbol\beta)/\partial\mathrm{vecl}(\boldsymbol\beta)^{\mathrm T}$ and $\boldsymbol\alpha_{0,\boldsymbol\beta}(\boldsymbol\beta) = \partial\boldsymbol\alpha_0(\boldsymbol\beta)/\partial\mathrm{vecl}(\boldsymbol\beta)^{\mathrm T}$. Since $\breve{\boldsymbol\beta}$ solves (33), plugging the expression of $\widehat{\boldsymbol\alpha}(\boldsymbol\beta)-\boldsymbol\alpha_0(\boldsymbol\beta)$, we further have

$$\begin{aligned}
\mathbf{0} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{g}(Y_i)\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\breve{\boldsymbol\beta}^{\mathrm T}\mathbf{X}_i,\widehat{\boldsymbol\alpha}(\breve{\boldsymbol\beta})\}\right]^{\mathrm T}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{g}(Y_i)\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\widehat{\boldsymbol\alpha}(\boldsymbol\beta)\}\right]^{\mathrm T}\\
&\quad -\frac{1}{n}\sum_{i=1}^n \mathbf{g}(\mathbf{Y}_i)\sqrt{n}\{\mathrm{vecl}(\breve{\boldsymbol\beta}-\boldsymbol\beta)\}^{\mathrm T}\left[\mathbf{m}_{\boldsymbol\beta}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}+\mathbf{m}_{\boldsymbol\alpha}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\widehat{\boldsymbol\alpha}(\boldsymbol\beta)\}\widehat{\boldsymbol\alpha}_{\boldsymbol\beta}(\boldsymbol\beta)\right]^{\mathrm T}+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{g}(Y_i)\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]^{\mathrm T}-\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{g}(Y_i)\left[\mathbf{m}_{\boldsymbol\alpha}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\{\widehat{\boldsymbol\alpha}(\boldsymbol\beta)-\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]^{\mathrm T}\\
&\quad -\frac{1}{n}\sum_{i=1}^n \mathbf{g}(\mathbf{Y}_i)\sqrt{n}\{\mathrm{vecl}(\breve{\boldsymbol\beta}-\boldsymbol\beta)\}^{\mathrm T}\left[\mathbf{m}_{\boldsymbol\beta}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}+\mathbf{m}_{\boldsymbol\alpha}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\boldsymbol\alpha_{0,\boldsymbol\beta}(\boldsymbol\beta)\right]^{\mathrm T}+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{g}(Y_i)\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]^{\mathrm T}\\
&\quad -\frac{1}{n^{3/2}}\sum_{i=1}^n\sum_{j=1}^n \mathbf{g}(Y_i)\left(\mathbf{B}_1^{-1}\mathbf{A}(\boldsymbol\beta^{\mathrm T}\mathbf{X}_j)\left[\mathbf{a}(\mathbf{X}_j)-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_j,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]\right)^{\mathrm T}\mathbf{m}_{\boldsymbol\alpha}^{\mathrm T}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\\
&\quad -\frac{1}{n}\sum_{i=1}^n \mathbf{g}(\mathbf{Y}_i)\sqrt{n}\{\mathrm{vecl}(\breve{\boldsymbol\beta}-\boldsymbol\beta)\}^{\mathrm T}\left[\mathbf{m}_{\boldsymbol\beta}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}+\mathbf{m}_{\boldsymbol\alpha}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\boldsymbol\alpha_{0,\boldsymbol\beta}(\boldsymbol\beta)\right]^{\mathrm T}+o_p(1).
\end{aligned}$$

We vectorize the above display and write the relation equivalently as

$$\begin{aligned}
\mathbf{0} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{vec}\left(\mathbf{g}(Y_i)\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]^{\mathrm T}\right)\\
&\quad -\frac{1}{\sqrt{n}}\sum_{j=1}^n \frac{1}{n}\sum_{i=1}^n \mathbf{m}_{\boldsymbol\alpha}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\otimes\mathbf{g}(\mathbf{Y}_i)\left(\mathbf{B}_1^{-1}\mathbf{A}(\boldsymbol\beta^{\mathrm T}\mathbf{X}_j)\left[\mathbf{a}(\mathbf{X}_j)-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_j,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]\right)\\
&\quad -\frac{1}{n}\sum_{i=1}^n\left[\mathbf{m}_{\boldsymbol\beta}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}+\mathbf{m}_{\boldsymbol\alpha}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\boldsymbol\alpha_{0,\boldsymbol\beta}(\boldsymbol\beta)\right]\otimes\mathbf{g}(Y_i)\sqrt{n}\mathrm{vecl}(\breve{\boldsymbol\beta}-\boldsymbol\beta)+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{I}_{p_a}\otimes\mathbf{g}(Y_i)\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]\\
&\quad -\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{B}_2\left(\mathbf{B}_1^{-1}\mathbf{A}(\boldsymbol\beta^{\mathrm T}\mathbf{X}_i)\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]\right)\\
&\quad -E\left(\left[\mathbf{m}_{\boldsymbol\beta}\{\boldsymbol\beta^{\mathrm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}+\mathbf{m}_{\boldsymbol\alpha}\{\boldsymbol\beta^{\mathrm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\boldsymbol\alpha_{0,\boldsymbol\beta}(\boldsymbol\beta)\right]\otimes\mathbf{g}(Y)\right)\sqrt{n}\mathrm{vecl}(\breve{\boldsymbol\beta}-\boldsymbol\beta)+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n\{\mathbf{I}_{p_a}\otimes\mathbf{g}(Y_i)-\mathbf{B}_2\mathbf{B}_1^{-1}\mathbf{A}(\boldsymbol\beta^{\mathrm T}\mathbf{X}_i)\}\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]\\
&\quad -E\left(\left[\mathbf{m}_{\boldsymbol\beta}\{\boldsymbol\beta^{\mathrm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}+\mathbf{m}_{\boldsymbol\alpha}\{\boldsymbol\beta^{\mathrm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\boldsymbol\alpha_{0,\boldsymbol\beta}(\boldsymbol\beta)\right]\otimes E\{\mathbf{g}(Y)\mid\boldsymbol\beta^{\mathrm T}\mathbf{X}\}\right)\sqrt{n}\mathrm{vecl}(\breve{\boldsymbol\beta}-\boldsymbol\beta)\\
&\quad +o_p(1). \qquad\qquad \text{(S.1)}
\end{aligned}$$

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

$$E\left([\mathbf{a}(\mathbf{X})-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}]\otimes E\{\mathbf{g}(Y)\mid\boldsymbol\beta^{\mathrm T}\mathbf{X}\}\right)=\mathbf{0}$$

for any $\boldsymbol\beta$, hence

$$\begin{aligned}
&E\left(\left[-\mathbf{m}_{\boldsymbol\beta}\{\boldsymbol\beta^{\mathrm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}-\mathbf{m}_{\boldsymbol\alpha}\{\boldsymbol\beta^{\mathrm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\boldsymbol\alpha_{0,\boldsymbol\beta}(\boldsymbol\beta)\right]\otimes E\{\mathbf{g}(Y)\mid\boldsymbol\beta^{\mathrm T}\mathbf{X}\}\right)\\
&+E\left(\left[\mathbf{a}(\mathbf{X})-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]\otimes\frac{\partial E\{\mathbf{g}(Y)\mid\boldsymbol\beta^{\mathrm T}\mathbf{X}\}}{\partial\mathrm{vecl}(\boldsymbol\beta)^{\mathrm T}}\right)\\
=\ &E\left(\left[-\mathbf{m}_{\boldsymbol\beta}\{\boldsymbol\beta^{\mathrm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}-\mathbf{m}_{\boldsymbol\alpha}\{\boldsymbol\beta^{\mathrm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\boldsymbol\alpha_{0,\boldsymbol\beta}(\boldsymbol\beta)\right]\otimes E\{\mathbf{g}(Y)\mid\boldsymbol\beta^{\mathrm T}\mathbf{X}\}\right)+\boldsymbol\Sigma_A\\
=\ &\mathbf{0}.
\end{aligned}$$

We thus can rewrite (S.1) as

$$\begin{aligned}
\mathbf{0} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\left\{\mathbf{I}_{p_a}\otimes\mathbf{g}(Y_i)-\mathbf{B}_2\mathbf{B}_1^{-1}\mathbf{A}(\boldsymbol\beta^{\mathrm T}\mathbf{X}_i)\right\}\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\mathrm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]\\
&\quad -\boldsymbol\Sigma_A\sqrt{n}\mathrm{vecl}(\breve{\boldsymbol\beta}-\boldsymbol\beta)+o_p(1),
\end{aligned}$$

which leads to the result in Theorem 1. □

S.3 Proof of Theorem 2

Because $\widehat{\boldsymbol\alpha}(\boldsymbol\beta)$ solves (45), we obtain the Taylor expansion

$$\begin{aligned}
\mathbf{0} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X}_i,\widehat{\boldsymbol\alpha}(\boldsymbol\beta)\}\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\rm T}\mathbf{X}_i,\widehat{\boldsymbol\alpha}(\boldsymbol\beta)\}\right]\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\rm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]\\
&\quad+\sum_{j=1}^{p_\alpha}\frac{1}{n}\sum_{i=1}^n \frac{\partial \mathbf{m}_{\boldsymbol\alpha}^{\rm T}(\boldsymbol\beta^{\rm T}\mathbf{X}_i,\boldsymbol\alpha)}{\partial \boldsymbol\alpha_j^{\rm T}}\Big|_{\boldsymbol\alpha=\boldsymbol\alpha_0(\boldsymbol\beta)}\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\rm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]\sqrt{n}\{\widehat{\boldsymbol\alpha}_j(\boldsymbol\beta)-\boldsymbol\alpha_{0j}(\boldsymbol\beta)\}\\
&\quad-\frac{1}{n}\sum_{i=1}^n \mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\frac{\partial \mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X}_i,\boldsymbol\alpha)}{\partial\boldsymbol\alpha^{\rm T}}\Big|_{\boldsymbol\alpha=\boldsymbol\alpha_0(\boldsymbol\beta)}\sqrt{n}\{\widehat{\boldsymbol\alpha}(\boldsymbol\beta)-\boldsymbol\alpha_0(\boldsymbol\beta)\}+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\rm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]\\
&\quad-\frac{1}{n}\sum_{i=1}^n \mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\frac{\partial \mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X}_i,\boldsymbol\alpha)}{\partial\boldsymbol\alpha^{\rm T}}\Big|_{\boldsymbol\alpha=\boldsymbol\alpha_0(\boldsymbol\beta)}\sqrt{n}\{\widehat{\boldsymbol\alpha}(\boldsymbol\beta)-\boldsymbol\alpha_0(\boldsymbol\beta)\}+o_p(1).
\end{aligned}$$

This leads to

$$\sqrt{n}\{\widehat{\boldsymbol\alpha}(\boldsymbol\beta)-\boldsymbol\alpha_0(\boldsymbol\beta)\}=\frac{1}{\sqrt{n}}\mathbf{B}_3^{-1}\sum_{i=1}^n \mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X}_i,\widehat{\boldsymbol\alpha}(\boldsymbol\beta)\}\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\left[\mathbf{a}(\mathbf{X}_i)-\mathbf{m}\{\boldsymbol\beta^{\rm T}\mathbf{X}_i,\boldsymbol\alpha_0(\boldsymbol\beta)\}\right]+o_p(1).$$

Following the same derivation as that in the proof of Theorem 1, we then obtain the expansion of $\breve{\boldsymbol\beta}$.

It is easy to verify that

$$\begin{aligned}
&E\Bigg\{\Big(\Big[\mathbf{I}_{p_a}\otimes E\{\mathbf{g}(Y)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}-\mathbf{B}_2\mathbf{B}_3^{-1}\mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\Big]\Big[\mathbf{a}(\mathbf{X})-\mathbf{m}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\Big]\Big)^{\rm T}\\
&\mathbf{B}_2\mathbf{B}_3^{-1}\mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\Big[\mathbf{a}(\mathbf{X})-\mathbf{m}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\Big]\Bigg\}\\
=\;&\text{trace}E\Bigg\{\Big(\Big[\mathbf{I}_{p_a}\otimes E\{\mathbf{g}(Y)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}-\mathbf{B}_2\mathbf{B}_3^{-1}\mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\Big]\\
&\Big[\mathbf{a}(\mathbf{X})-\mathbf{m}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\Big]\Big)^{\rm T}\mathbf{B}_2\mathbf{B}_3^{-1}\mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\Big[\mathbf{a}(\mathbf{X})-\mathbf{m}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\Big]\Bigg\}\\
=\;&\text{trace}E\Big(\mathbf{B}_2\mathbf{B}_3^{-1}\mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\Big[\mathbf{a}(\mathbf{X})-\mathbf{m}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\Big]\\
&\Big[\mathbf{a}(\mathbf{X})-\mathbf{m}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\Big]^{\rm T}\Big[\mathbf{I}_{p_a}\otimes E\{\mathbf{g}(Y)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}-\mathbf{B}_2\mathbf{B}_3^{-1}\mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\Big]^{\rm T}\Big)\\
=\;&\text{trace}E\Big(\mathbf{B}_2\mathbf{B}_3^{-1}\mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\Big[\mathbf{I}_{p_a}\otimes E\{\mathbf{g}(Y)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}-\mathbf{B}_2\mathbf{B}_3^{-1}\mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\Big]^{\rm T}\Big)\\
=\;&\text{trace}E\Big(\mathbf{B}_2\mathbf{B}_3^{-1}\mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\Big[\mathbf{I}_{p_a}\otimes E\{\mathbf{g}^{\rm T}(Y)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}-\mathbf{Q}^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\mathbf{m}_{\boldsymbol\alpha}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\mathbf{B}_3^{-1}\mathbf{B}_2^{\rm T}\Big]\Big)\\
=\;&\text{trace}\Big(\mathbf{B}_2\mathbf{B}_3^{-1}E[\mathbf{m}_{\boldsymbol\alpha}^{\rm T}\{\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\alpha_0(\boldsymbol\beta)\}\otimes E\{\mathbf{g}^{\rm T}(Y)\mid\boldsymbol\beta^{\rm T}\mathbf{X}\}]-\mathbf{B}_2\mathbf{B}_3^{-1}\mathbf{B}_2^{\rm T}\Big)\\
=\;&0.
\end{aligned}$$

Thus the orthogonality result is verified. □

S.4 Proof of Theorem 3

Because no constraints are imposed on $f_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$ other than it is a valid pdf, hence its nuisance tangent space contains all mean zero functions of $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$. In addition to being a valid conditional pdf, $f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\epsilon})$ 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 Λ_2. The results of Λ_3 can be similarly derived as Λ_1, by treating Y as the random variable. We omit the details of the derivation of Λ_1, Λ_2 and Λ_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 Λ.

It is also not hard to see that $\Lambda_1^{\perp}$ contains all the functions $\mathbf{g}(\mathbf{X}, Y)$ such that $E\{\mathbf{g}(\mathbf{X}, Y) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\} = \mathbf{0}$, $\Lambda_2^{\perp}$ contains all the functions $\mathbf{g}(\mathbf{X}, Y)$ such that $E\{\mathbf{g}(\mathbf{X}, Y) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\epsilon}\}$ has the form $\mathbf{a}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}) + \mathbf{A}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\boldsymbol{\epsilon}$, and $\Lambda_3^{\perp}$ contains all the functions $\mathbf{g}(\mathbf{X}, Y)$ such that $E\{\mathbf{g}(\mathbf{X}, Y) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, Y\}$ has the form $\mathbf{a}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$. 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 $\boldsymbol{\beta}$, i.e. $\boldsymbol{\beta}_2$. The score function is

$$\begin{aligned}\mathbf{S}_{\boldsymbol{\beta}_2} &= \mathrm{vec}\left[\mathbf{X}_2\left\{\frac{\partial \log f_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} + \frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\epsilon})}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} - \frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\epsilon})}{\partial \boldsymbol{\epsilon}^{\mathrm{T}}}\frac{\partial \mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right.\right.\\ &\quad \left.\left. + \frac{\partial \log f_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, Y)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right\}\right] - \frac{\partial \mathbf{m}^{\mathrm{T}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)}{\partial \mathrm{vec}(\boldsymbol{\beta}_2)}\frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\epsilon})}{\partial \boldsymbol{\epsilon}}.\end{aligned}$$

We now decompose the score function into $\mathbf{S}_{\boldsymbol{\beta}_2} = \mathbf{S}_{\mathrm{eff}} + \mathbf{R}$, where

$$\begin{aligned}\mathbf{S}_{\mathrm{eff}} &= \mathrm{vec}\left(\boldsymbol{\epsilon}_2\frac{\partial \log f_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} + \frac{\partial \mathbf{Q}_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\left[\mathbf{I}_d \otimes \left\{\mathbf{Q}_2^{-1}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\boldsymbol{\epsilon}_2\right\}\right] + \mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\boldsymbol{\epsilon}_2^{\mathrm{T}}\right.\\ &\quad \left. \times \mathbf{Q}_2^{-1}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\frac{\partial \mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} + \boldsymbol{\epsilon}_2\frac{\partial \log f_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, Y)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right) + \frac{\partial \mathbf{m}^{\mathrm{T}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)}{\partial \mathrm{vec}(\boldsymbol{\beta}_2)}\mathbf{Q}_2^{-1}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\boldsymbol{\epsilon}_2,\end{aligned}$$

and

$$\begin{aligned}
\mathbf{R} &= \text{vec}\Bigg(\mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)\frac{\partial \log f_1(\boldsymbol\beta^{\rm T}\mathbf{X})}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta} + \mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta} + \boldsymbol\epsilon_2\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta} \\
&\quad -\frac{\partial \mathbf{Q}_2(\boldsymbol\beta^{\rm T}\mathbf{X})}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta}\left[\mathbf{I}_d\otimes\left\{\mathbf{Q}_2^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\boldsymbol\epsilon_2\right\}\right] - \mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)\left\{\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \boldsymbol\epsilon_2^{\rm T}} + \boldsymbol\epsilon_2^{\rm T}\mathbf{Q}_2^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\right\} \\
&\quad \times\frac{\partial \mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta} - \left\{\boldsymbol\epsilon_2\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \boldsymbol\epsilon_2^{\rm T}} + \mathbf{I}_{p-d}\right\}\frac{\partial \mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta} + \mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2) \\
&\quad \times\frac{\partial \log f_3(\boldsymbol\beta^{\rm T}\mathbf{X},Y)}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta} + \frac{\partial \mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta}\Bigg) - \frac{\partial \mathbf{m}^{\rm T}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)}{\partial \text{vec}(\boldsymbol\beta_2)}\left\{\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \boldsymbol\epsilon_2} + \mathbf{Q}_2^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\boldsymbol\epsilon_2\right\}.
\end{aligned}$$

Here, when taking derivative of a matrix with respect to a row vector, we obtain a block row matrix, with the jth block element is the derivative of the matrix with respect to the jth element of the vector. We can easily check that indeed $\mathbf{S}_{\boldsymbol\beta_2} = \mathbf{S}_{\rm eff} + \mathbf{R}$. It is also straightforward to verify that $\mathbf{S}_{\rm eff} \in \Lambda^{\perp}$. In addition, we easily obtain

$$\begin{aligned}
\mathbf{R}_1 &\equiv \text{vec}\left\{\mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)\frac{\partial \log f_1(\boldsymbol\beta^{\rm T}\mathbf{X})}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta} + \frac{\partial \mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta}\right\} \in \Lambda_1 \\
\mathbf{R}_3 &\equiv \mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)\frac{\partial \log f_3(\boldsymbol\beta^{\rm T}\mathbf{X},Y)}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta} \in \Lambda_3.
\end{aligned}$$

Finally, using the relation

$$\begin{aligned}
&E\left\{\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta} \mid \boldsymbol\beta^{\rm T}\mathbf{X}\right\} = \mathbf{0} \\
&E\left\{\boldsymbol\epsilon_2\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta} \mid \boldsymbol\beta^{\rm T}\mathbf{X}\right\} = \mathbf{0} \\
&E\left\{\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \boldsymbol\epsilon_2^{\rm T}} \mid \boldsymbol\beta^{\rm T}\mathbf{X}\right\} = \mathbf{0} \\
&E\left\{\boldsymbol\epsilon_2\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \boldsymbol\epsilon_2^{\rm T}} \mid \boldsymbol\beta^{\rm T}\mathbf{X}\right\} = -\mathbf{I}_{p-d} \\
&E\left\{\epsilon_{2j}\epsilon_{2j}\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \boldsymbol\epsilon_2^{\rm T}} \mid \boldsymbol\beta^{\rm T}\mathbf{X}\right\} = \mathbf{0},
\end{aligned}$$

through tedious but straightforward calculation, we can verify that

$$\begin{aligned}
\mathbf{R}_2 &\equiv \text{vec}\Bigg(\mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta} + \boldsymbol\epsilon_2\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta} \\
&\quad -\frac{\partial \mathbf{Q}_2(\boldsymbol\beta^{\rm T}\mathbf{X})}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta}\left[\mathbf{I}_d\otimes\left\{\mathbf{Q}_2^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\boldsymbol\epsilon_2\right\}\right] - \mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)\left\{\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \boldsymbol\epsilon_2^{\rm T}} + \boldsymbol\epsilon_2^{\rm T}\mathbf{Q}_2^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\right\} \\
&\quad \times\frac{\partial \mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta} - \left[\boldsymbol\epsilon_2\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \boldsymbol\epsilon_2^{\rm T}} + \mathbf{I}_{p-d}\right]\frac{\partial \mathbf{m}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)}{\partial \mathbf{X}^{\rm T}\boldsymbol\beta}\Bigg) \\
&\quad -\frac{\partial \mathbf{m}^{\rm T}(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\beta_2)}{\partial \text{vec}(\boldsymbol\beta_2)}\left\{\frac{\partial \log f_2(\boldsymbol\beta^{\rm T}\mathbf{X},\boldsymbol\epsilon_2)}{\partial \boldsymbol\epsilon_2} + \mathbf{Q}_2^{-1}(\boldsymbol\beta^{\rm T}\mathbf{X})\boldsymbol\epsilon_2\right\} \in \Lambda_2.
\end{aligned}$$

We can see that $\mathbf{R} = \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3$, hence $\mathbf{R} \in \Lambda$. This shows that $\mathbf{S}_{\rm eff}$ is indeed the efficient score. □

S.5 Proof of Theorem 4

Similar to the derivation in proving Theorem 3, the form of Λ_1, Λ_3 are unchanged. Regarding Λ_2, because in addition to being a valid conditional pdf, $f_2(\boldsymbol{\beta}^{\rm T}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)$ 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 Λ_2. The orthogonality of the three spaces Λ_1, Λ_2 and Λ_3 still holds, hence we obtain the results concerning Λ.

Obviously, $\Lambda_1^\perp$ and $\Lambda_3^\perp$ remain unchanged from in those in Theorem 3. $\Lambda_2^\perp$ contains all the functions $\mathbf{g}(\mathbf{X}, Y)$ such that $E\{\mathbf{g}(\mathbf{X}, Y) \mid \boldsymbol{\beta}^{\rm T}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2\}$ has the form $\mathbf{a}(\boldsymbol{\beta}^{\rm T}\mathbf{X}) + \mathbf{A}(\boldsymbol{\beta}^{\rm T}\mathbf{X})\widetilde{\boldsymbol{\epsilon}}_2 + \mathbf{B}(\boldsymbol{\beta}^{\rm T}\mathbf{X})\widetilde{\boldsymbol{\epsilon}}_2\widetilde{\boldsymbol{\epsilon}}_2^{\rm T}$. Thus, taking the intersection of $\Lambda_1^\perp, \Lambda_2^\perp$ and $\Lambda_3^\perp$, we obtain $\Lambda^\perp$ as described in Theorem 4. Note that our construction ensures that $E(\widetilde{\boldsymbol{\epsilon}}_2\mathbf{v}^{\rm T} \mid \boldsymbol{\beta}^{\rm T}\mathbf{X}) = \mathbf{0}$. We then can write

$$
\begin{aligned}
\Lambda_1 &= \Big[\mathbf{h}(\boldsymbol{\beta}^{\rm T}\mathbf{X}) : E\{\mathbf{h}(\boldsymbol{\beta}^{\rm T}\mathbf{X})\} = \mathbf{0}, E\{\mathbf{h}^{\rm T}(\boldsymbol{\beta}^{\rm T}\mathbf{X})\mathbf{h}(\boldsymbol{\beta}^{\rm T}\mathbf{X})\} < \infty, \mathbf{h}(\boldsymbol{\beta}^{\rm T}\mathbf{X}) \in \mathcal{R}^{(p-d)d}\Big] \\
\Lambda_2 &= \Big[\mathbf{h}(\boldsymbol{\beta}^{\rm T}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2) : E\{\mathbf{h}(\boldsymbol{\beta}^{\rm T}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2) \mid \boldsymbol{\beta}^{\rm T}\mathbf{X}\} = \mathbf{0}, E\{\widetilde{\boldsymbol{\epsilon}}_2\mathbf{h}^{\rm T}(\boldsymbol{\beta}^{\rm T}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2) \mid \boldsymbol{\beta}^{\rm T}\mathbf{X}\} = \mathbf{0}, \\
&\quad E\{\mathbf{v}\mathbf{h}^{\rm T}(\boldsymbol{\beta}^{\rm T}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2) \mid \boldsymbol{\beta}^{\rm T}\mathbf{X}\} = \mathbf{0}, E\{\mathbf{h}^{\rm T}(\boldsymbol{\beta}^{\rm T}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)\mathbf{h}(\boldsymbol{\beta}^{\rm T}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)\} < \infty, \\
&\quad \mathbf{h}(\boldsymbol{\beta}^{\rm T}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2) \in \mathcal{R}^{(p-d)d}\Big] \\
\Lambda_3 &= \Big[\mathbf{h}(\boldsymbol{\beta}^{\rm T}\mathbf{X}, Y) : E\{\mathbf{h}(\boldsymbol{\beta}^{\rm T}\mathbf{X}, Y) \mid \boldsymbol{\beta}^{\rm T}\mathbf{X}\} = \mathbf{0}, E\{\mathbf{h}^{\rm T}(\boldsymbol{\beta}^{\rm T}\mathbf{X}, Y)\mathbf{h}(\boldsymbol{\beta}^{\rm T}\mathbf{X}, Y)\} < \infty, \\
&\quad \mathbf{h}(\boldsymbol{\beta}^{\rm T}\mathbf{X}, Y) \in \mathcal{R}^{(p-d)d}\Big] \\
\Lambda^\perp &= \Big[\mathbf{g}(\mathbf{X}, Y) : E\{\mathbf{g}(\mathbf{X}, Y) \mid \boldsymbol{\beta}^{\rm T}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2\} = \mathbf{A}(\boldsymbol{\beta}^{\rm T}\mathbf{X})\widetilde{\boldsymbol{\epsilon}}_2 + \mathbf{B}(\boldsymbol{\beta}^{\rm T}\mathbf{X})\mathbf{v}, \\
&\quad E\{\mathbf{g}(\mathbf{X}, Y) \mid \boldsymbol{\beta}^{\rm T}\mathbf{X}, Y\} = \mathbf{0}, E\{\mathbf{g}^{\rm T}(\mathbf{X}, Y)\mathbf{g}(\mathbf{X}, Y)\} < \infty, \mathbf{g}(\mathbf{X}, Y) \in \mathcal{R}^{(p-d)d}\Big].
\end{aligned}
$$

To obtain the efficient score, we first calculate the score function with respect to the parameter of interest contained in $\boldsymbol{\beta}$, i.e. $\boldsymbol{\beta}_2$. The score function is different from that in

Theorem 3 and we obtain

$$
\begin{aligned}
\mathbf{S}_{\boldsymbol{\beta}_2} &= \text{vec}\left[\mathbf{X}_2\left\{\frac{\partial \log f_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} + \frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} - \frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2^{\mathrm{T}}}\mathbf{D}^{-1}(\boldsymbol{\beta}_2)\frac{\partial \mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right.\right.\\
&\quad \left.\left. + \frac{\partial \log f_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, Y)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right\}\right] + \left(\frac{\partial \mathbf{D}^{-1}(\boldsymbol{\beta}_2)}{\partial \text{vec}(\boldsymbol{\beta}_2)^{\mathrm{T}}}[\mathbf{I}_{(p-d)d}\otimes\{\mathbf{X}-\mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\}]\right.\\
&\quad \left. - \mathbf{D}^{-1}(\boldsymbol{\beta}_2)\frac{\partial \mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)}{\partial \text{vec}(\boldsymbol{\beta}_2)^{\mathrm{T}}}\right)^{\mathrm{T}}\frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2} - \frac{\partial \log\det\{\mathbf{D}(\boldsymbol{\beta}_2)\}}{\partial \text{vec}(\boldsymbol{\beta}_2)}\\
&= \text{vec}\left[\mathbf{X}_2\left\{\frac{\partial \log f_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} + \frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} - \frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2^{\mathrm{T}}}\mathbf{D}^{-1}(\boldsymbol{\beta}_2)\frac{\partial \mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right.\right.\\
&\quad \left.\left. + \frac{\partial \log f_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, Y)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right\}\right] + \left(\frac{\partial \mathbf{D}^{-1}(\boldsymbol{\beta}_2)}{\partial \text{vec}(\boldsymbol{\beta}_2)^{\mathrm{T}}}[\mathbf{I}_{(p-d)d}\otimes\{\mathbf{D}(\boldsymbol{\beta}_2)\widetilde{\boldsymbol{\epsilon}}_2\}] - \mathbf{D}^{-1}(\boldsymbol{\beta}_2)\frac{\partial \mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)}{\partial \text{vec}(\boldsymbol{\beta}_2)^{\mathrm{T}}}\right)^{\mathrm{T}}\\
&\quad \times\frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2} - \frac{\partial \log\det\{\mathbf{D}(\boldsymbol{\beta}_2)\}}{\partial \text{vec}(\boldsymbol{\beta}_2)}.
\end{aligned}
$$

The key difference is in how the score function should be decomposed, reflecting the change of the spaces Λ and $\Lambda^{\perp}$. We can rewrite

$$
\begin{aligned}
\mathbf{S}_{\boldsymbol{\beta}_2} &= \text{vec}\left[\mathbf{X}_2\left\{\frac{\partial \log f_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} + \frac{\partial \log f_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, Y)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} + \frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right\}\right]\\
&\quad - \text{vec}\left\{\mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2^{\mathrm{T}}}\mathbf{D}^{-1}(\boldsymbol{\beta}_2)\frac{\partial \mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right\}\\
&\quad - \frac{\partial \mathbf{m}^{\mathrm{T}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)}{\partial \text{vec}(\boldsymbol{\beta}_2)}\left\{\mathbf{D}^{-1}(\boldsymbol{\beta}_2)\right\}^{\mathrm{T}}\frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2}\\
&\quad - \text{vec}\left\{\mathbf{D}(\boldsymbol{\beta}_2)\widetilde{\boldsymbol{\epsilon}}_2\frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2^{\mathrm{T}}}\mathbf{D}^{-1}(\boldsymbol{\beta}_2)\frac{\partial \mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right\}\\
&\quad + \left(\frac{\partial \mathbf{D}^{-1}(\boldsymbol{\beta}_2)}{\partial \text{vec}(\boldsymbol{\beta}_2)^{\mathrm{T}}}[\mathbf{I}_{(p-d)d}\otimes\{\mathbf{D}(\boldsymbol{\beta}_2)\widetilde{\boldsymbol{\epsilon}}_2\}]\right)^{\mathrm{T}}\frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2} - \frac{\partial \log\det\{\mathbf{D}(\boldsymbol{\beta}_2)\}}{\partial \text{vec}(\boldsymbol{\beta}_2)}\\
&= \text{vec}\left[\mathbf{X}_2\left\{\frac{\partial \log f_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} + \frac{\partial \log f_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, Y)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} + \frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right\}\right]\\
&\quad - \left\{\frac{\partial \mathbf{m}^{\mathrm{T}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}\otimes\mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2) + \frac{\partial \mathbf{m}^{\mathrm{T}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)}{\partial \text{vec}(\boldsymbol{\beta}_2)}\right\}\left\{\mathbf{D}^{-1}(\boldsymbol{\beta}_2)\right\}^{\mathrm{T}}\frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2}\\
&\quad - \left(\left[\frac{\partial \mathbf{m}^{\mathrm{T}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}\left\{\mathbf{D}^{-1}(\boldsymbol{\beta}_2)\right\}^{\mathrm{T}}\right]\otimes\mathbf{D}(\boldsymbol{\beta}_2)\right)\text{vec}\left\{\widetilde{\boldsymbol{\epsilon}}_2\frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2^{\mathrm{T}}}\right\}\\
&\quad + \mathbf{C}_1(\boldsymbol{\beta}_2)\text{vec}\left\{\widetilde{\boldsymbol{\epsilon}}_2\frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2^{\mathrm{T}}}\right\} - \frac{\partial \log\det\{\mathbf{D}(\boldsymbol{\beta}_2)\}}{\partial \text{vec}(\boldsymbol{\beta}_2)}\\
&= \text{vec}\left[\mathbf{X}_2\left\{\frac{\partial \log f_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} + \frac{\partial \log f_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, Y)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} + \frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right\}\right]\\
&\quad + \mathbf{K}_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2} + \mathbf{K}_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\text{vec}\left\{\widetilde{\boldsymbol{\epsilon}}_2\frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2^{\mathrm{T}}}\right\}\\
&\quad - \frac{\partial \log\det\{\mathbf{D}(\boldsymbol{\beta}_2)\}}{\partial \text{vec}(\boldsymbol{\beta}_2)}
\end{aligned}
$$

We decompose the score function into $\mathbf{S}_{\boldsymbol{\beta}_2} = \mathbf{S}_{\text{eff}} + \mathbf{R}$, where $\mathbf{R} \in \Lambda$ and $\mathbf{S}_{\text{eff}} \in \Lambda^{\perp}$ and hence is the efficient score and. Here,

$$
\begin{aligned}
\mathbf{S}_{\text{eff}} \quad &= \quad \text{vec}\left(\mathbf{D}(\boldsymbol{\beta}_2)\widetilde{\boldsymbol{\epsilon}}_2 \frac{\partial \log f_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} + \mathbf{D}(\boldsymbol{\beta}_2)\widetilde{\boldsymbol{\epsilon}}_2 \frac{\partial \log f_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, Y)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right) \\
&\quad -\mathbf{K}_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\widetilde{\boldsymbol{\epsilon}}_2 + \mathbf{K}_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\mathbf{v} - \mathbf{K}_4(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\mathbf{v}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{R} \quad &= \quad \text{vec}\left\{\mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\frac{\partial \log f_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} + \mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\frac{\partial \log f_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, Y)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}} + \mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right\} \\
&\quad +\text{vec}\left\{\mathbf{D}(\boldsymbol{\beta}_2)\widetilde{\boldsymbol{\epsilon}}_2 \frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}}\right\} + \mathbf{K}_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2} \\
&\quad +\mathbf{K}_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\widetilde{\boldsymbol{\epsilon}}_2 - \mathbf{K}_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\mathbf{v} + \mathbf{K}_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\text{vec}\left\{\widetilde{\boldsymbol{\epsilon}}_2 \frac{\partial \log f_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \widetilde{\boldsymbol{\epsilon}}_2)}{\partial \widetilde{\boldsymbol{\epsilon}}_2^{\mathrm{T}}} + \mathbf{I}_{p-d}\right\} \\
&\quad +\mathbf{K}_4(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\mathbf{v} - \frac{\partial \log \det\{\mathbf{D}(\boldsymbol{\beta}_2)\}}{\partial \text{vec}(\boldsymbol{\beta}_2)} - \mathbf{K}_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\beta}_2)\text{vec}(\mathbf{I}_{p-d}).
\end{aligned}
$$

It is obvious that $\mathbf{S}_{\text{eff}} \in \Lambda^{\perp}$. Careful and tedious calculations, through grouping the terms in $\mathbf{R}$ as the second, the third, the fourth+fifth, the sixth+seventh+eighth, ninth+tenth, and first+eleventh+twelfth terms, verify that $\mathbf{R} \in \Lambda$. □