On the Efficiency of Online Approach to Nonparametric Smoothing of Big Data

[1] *University of Electronic Science and Technology, China*

[2] *National University of Singapore, Singapore*

Supplementary Material

Proof of Lemma 2.2 We prove the case where $p = 1$ to illustrate. Multivariate analogy follows through exactly the same arguments with notations replaced by their multivariate counterparts. First note that

$$E_i[K_{\tilde{h}_i}(\mathbf{X}_{ix})] = \int K(\mathbf{t}) f(\mathbf{x} + \tilde{h}_i t) dt \equiv f(\mathbf{x}) + b_i(\mathbf{x}),$$

uniformly in $i = 1, 2, \cdots$, where $b_i(\mathbf{x}) = \frac{1}{2} f^{(2)}(\mathbf{x}) \tilde{h}_i^2 + O(h_i^3)$. Write $v_i(\mathbf{x}) = K_{\tilde{h}_i}(\mathbf{X}_{ix}) - E[K_{\tilde{h}_i}(\mathbf{X}_{ix})]$ and we have $E\{v_i(\mathbf{x})\}^2 = \tilde{h}_i^{-1} f(\mathbf{x}) R_2(K) + O(\tilde{h}_i)$ again uniformly in $i = 1, 2, \cdots$. Therefore,

$$\begin{aligned} E\{|\tilde{f}_n(\mathbf{x}|\tilde{h}_n) - f(\mathbf{x})|^2\} &= \frac{1}{n^2}\Big(\sum_{i=1}^n b_i(\mathbf{x})\Big)^2 + \frac{1}{n^2}\sum_{i=1}^n E\{v_i^2(\mathbf{x})\} \\ &= \frac{25c^4}{36}\{f^{(2)}(\mathbf{x})\}^2 n^{-4\alpha} + \frac{5}{6c} f(\mathbf{x}) R_2(K) n^{\alpha-1} + o(n^{-4\alpha} + n^{\alpha-1}) \end{aligned}$$

where the last equality followed from the following facts

$$\sum_{i=1}^{n} b_i(\mathbf{x}) = \frac{c^2}{2} f^{(2)}(\mathbf{x}) \sum_{i=1}^{n} i^{-2\alpha}(1+o(1))$$
$$= \frac{c^2}{2} f^{(2)}(\mathbf{x}) n^{1-2\alpha} \int_0^1 t^{-2\alpha} dt(1+o(1)) = \frac{c^2 f^{(2)}(\mathbf{x})}{2(1-2\alpha)} n^{1-2\alpha}(1+o(1)),$$
$$\sum_{i=1}^{n} E\{v_i^2(\mathbf{x})\} \propto c^{-1} \sum_{i=1}^{n} i^{\alpha} = c^{-1} n^{1+\alpha} \int_0^1 t^{\alpha} dt(1+o(1)) = \frac{1}{c(1+\alpha)} n^{1+\alpha}(1+o(1)).$$

Therefore, to minimize the asymptotic MSE, we must have

$$\alpha = 1/5, \quad c = (0.3)^{1/5}[f(\mathbf{x}) R_2(K)/\{f^{(2)}(\mathbf{x})\}^2]^{1/5}.$$

The proof is complete. □

Proof of Lemma 2.3 Write $c_b = \frac{1}{2} c^2 f^{(2)}(\mathbf{x})$ and $c_v = c^{-p} R_2(K) f(\mathbf{x})$.

Therefore,

$$\sum_{n=1}^{N} E\{|\tilde{f}_n(\mathbf{x})|\tilde{h}_n) - f(\mathbf{x})|^2\} = \sum_{n=1}^{N} \frac{1}{n^2} \Big\{ \sum_{i=1}^{n} b_i \Big\}^2 + \sum_{n=1}^{N} n^{-2} \sum_{i=1}^{n} E\{v_i\}^2$$

where

$$\sum_{i=1}^{n} b_i = c_b \sum_{i=1}^{n} i^{-2\alpha}, \ \sum_{i=1}^{n} E\{v_i\}^2 = c_v \sum_{i=1}^{n} i^{p\alpha}.$$

We again only prove the case where $p = 1$. First note that

$$\sum_{i=1}^{n} \frac{n^{2\alpha-1}}{i^{2\alpha}} \in \Big[\int_{2/n}^{1+1/n} x^{-2\alpha} dx, \int_{1/n}^{1} x^{-2\alpha} dx \Big] = \frac{1}{1-2\alpha} \Big[\Big(1+\frac{1}{n}\Big)^{1-2\alpha} - \Big(\frac{2}{n}\Big)^{1-2\alpha}, 1 - \Big(\frac{1}{n}\Big)^{1-2\alpha} \Big],$$
$$\sum_{n=1}^{N} \frac{1}{n^{1+2\alpha}} \Big(1+\frac{1}{n}\Big)^{1-2\alpha} \in \Big[\int_{2}^{N+1} x^{-2}(1+x)^{1-2\alpha} dx, \int_{1}^{N} x^{-2}(1+x)^{1-2\alpha} dx \Big],$$
$$\sum_{n=1}^{N} \frac{1}{n^{1+2\alpha}} \Big(\frac{2}{n}\Big)^{1-2\alpha} \in 2^{2-2\alpha} \times [3/4, 1], \quad \sum_{n=1}^{N} \frac{1}{n^{1+2\alpha}} \Big(\frac{1}{n}\Big)^{1-2\alpha} \in [1/2, 1],$$
$$\sum_{n=1}^{N} \frac{1}{n^{1+2\alpha}} \in \Big[\int_{2}^{N+1} x^{-1-2\alpha} dx, \int_{1}^{N} x^{-1-2\alpha} dx \Big] \in \frac{1}{2\alpha} \Big[2^{-2\alpha} - (N+1)^{-2\alpha}, 1 - N^{-2\alpha} \Big].$$

As $N \to \infty$,

$$\sum_{n=1}^{N} n^{-1-2\alpha} \Big(1 - 1/n\Big)^{1-2\alpha} \in \Big[\frac{1}{2\alpha} \Big(\frac{1}{2}\Big)^{2\alpha} - 1, \frac{1}{2\alpha} - \frac{1}{2} \Big],$$

whence

$$\sum_{n=1}^{N} \frac{1}{n^2} \sum_{i=1}^{n} b_i = \Big(\frac{1}{2\alpha} - 1 \Big)(1 + o(1)), \tag{S0.1}$$
$$\sum_{n=1}^{N} n^{-2} \Big\{ \sum_{i=1}^{n} b_i \Big\}^2 = \frac{c_b^2}{(1-2\alpha)^2 (1-4\alpha)} N^{1-4\alpha} (1 + o(1)). \tag{S0.2}$$

Similarly,

$$\sum_{i=1}^{n} i^{\alpha} / n^{1+\alpha} \in \Big[\int_{1/n}^{1} x^{\alpha} dx, \int_{2/n}^{(n+1)/n} x^{\alpha} dx \Big]$$
$$= \frac{1}{1+\alpha} \Big[1 - \Big(\frac{1}{n}\Big)^{1+\alpha}, \Big(1 + \frac{1}{n}\Big)^{1+\alpha} - \Big(\frac{2}{n}\Big)^{1+\alpha} \Big]$$

and

$$\sum_{n=1}^{N} n^{\alpha-1} \propto \alpha^{-1} N^{\alpha}\Big[1-\Big(\frac{2}{N}\Big)^{\alpha}\Big];\ \sum_{n=1}^{N} n^{\alpha-1}\Big(\frac{1}{n}\Big)^{1+\alpha} \in (1/2,1),$$
$$\sum_{n=1}^{N} n^{\alpha-1}\Big(1+\frac{1}{n}\Big)^{1+\alpha} = \sum_{n=1}^{N} n^{-2}(n+1)^{1+\alpha} \in \Big[\sum_{n=1}^{N} n^{\alpha-1}, \sum_{n=1}^{N}(n+1)^{\alpha-1}\Big],$$
$$\sum_{n=1}^{N} n^{\alpha-1}/N^{\alpha} \in \Big[\int_{1/N}^{1+1/N} x^{\alpha-1}dx, \int_{0}^{1} x^{\alpha-1}dx\Big] \to \alpha^{-1} \quad (N\to\infty).$$

Therefore,

$$\sum_{n=1}^{N} n^{-2}\sum_{i=1}^{n} E\{v_i\}^2 = \frac{c_v}{\alpha(\alpha+1)} N^{\alpha}(1+o(1)). \tag{S0.3}$$

Combining (S0.1)-(S0.3), we see that (S0.1) will be negligible compared to the other two, with $n_0/N \to 0$, whence the α which maximizes the sum of these three terms, also maximizes the sum of (S0.2) and (S0.3). Therefore, in general, the leading term of $MSE(N,\alpha,c)$ is given by

$$\frac{c_b^2}{(1-2\alpha)^2(1-4\alpha)} N^{-4\alpha} + \frac{c_v}{p\alpha(p\alpha+1)} N^{p\alpha-1}.$$

Although the coefficients also vary with α, this does not change the fact that provided that N is large enough, the α which minimizes the above quantity is $1/(p+4)$. As for the optimal value for c, note that the c minimizes the above quantity for any given value of α is given by

$$\Big(\frac{R_2(K)f(\mathbf{x})(1-2\alpha)^2(1-4\alpha)p}{p\alpha(p\alpha+1)[tr\{\mathcal{H}_f(\mathbf{x})\}]^2}\Big)^{1/(p+4)}.$$

The proof is thus complete by setting $\alpha = 1/(p+4)$. □

Proof of Lemma 2.4 Define

$$b_n = n^{2\alpha-1}\sum_{i=1}^{n} i^{-2\alpha}(1-\theta_i)^2, \quad v_n = n^{-1-p\alpha}\sum_{i=1}^{n} i^{p\alpha}(1-\theta_i)^{-1}.$$

The online estimator with bandwidth $\tilde{h}_i = c(1-\theta_i)i^{-\alpha}$ then has its AMSE given by

$$\frac{c^4}{4}\{f^{(2)}((\mathbf{x}))\}^2 n^{-4\alpha} b_n^2 + c^{-p} f(\mathbf{x}) R_2(K) n^{p\alpha-1} v_n,$$

which, with c chosen optimally, turns out to be

$$\frac{p+4}{4} p^{-p/(4+p)} [f(\mathbf{x})R_2(K)]^{4/(4+p)} \{f^{(2)}((\mathbf{x}))\}^{2p/(p+4)} v_n^{4/(4+p)} b_n^{2p/(p+4)}.$$

Therefore, its relative efficiency against the off-line estimator is given by

$$\lim_n v_n^{4/(4+p)} b_n^{2p/(p+4)}.$$

This together with the facts that

$$\sum_{i=1}^{n} i^{-2\alpha}(1-\theta_i(a))^2 = \sum_{i=1}^{n} i^{-2\alpha}(1+o(1)), \quad \sum_{i=1}^{n} i^{p\alpha}(1-\theta_i(a))^{-1} = \sum_{i=1}^{n} i^{p\alpha}(1+o(1)),$$

means that the relative efficiency is identical to that suggested by Lemma 2.2. □

Proof of Lemma 2.5 First note that by the definition of $w_{N,n}$, we easily see that

$$\tilde{f}_N(\mathbf{x}|\tilde{h}_N, \beta_N) = \sum_{n=1}^{N} w_{N,n} K_{\tilde{h}_n}(\mathbf{X}_n - \mathbf{x});$$

the AMSE of $\tilde{f}_N(\mathbf{x}|\tilde{h}_N\beta_N)$ is easily seen to be as required. The rest of the proof then follows from the following two corollaries: Corollary S0.1 and Corollary S0.2, and the fact that if $n\beta_n \to 0$, then for any $a > 0$,

$$\sum \beta_n \leq a \sum n^{-1} \propto a \log N,$$

so that $n^a \exp(-\sum_{n=1}^{N} \beta_n) \to \infty$, for any $a > 0$. □

Corollary S0.1. Depending on the speed at which β_n converges to 0, as $n \to \infty$,

(A) $n\beta_n \to b$ for some $b > 0$: $S_N \approx \frac{b}{(b-2\alpha)} N^{-2\alpha}$;

(B) $n\beta_n \to \infty : S_N \propto N^{-2\alpha}$;

(C) $n\beta_n \to 0 : S_N \propto \exp\Big[-\Big(\sum_{n=1}^{N} \beta_n\Big)(1 + o(1))\Big]$.

Proof of Corollary S0.1 What follows from (2.3) and the definition of S_n is that

$$S_{N+1} = (1 - \beta_{N+1})S_N + (N+1)^{-2\alpha}\beta_{N+1}. \tag{S0.4}$$

Divide either side by a factor of $(N+1)^{-2\alpha}$:

$$(N+1)^{2\alpha}S_{N+1} = (1 - \beta_{N+1})N^{2\alpha}S_N \frac{(N+1)^{2\alpha}}{N^{2\alpha}} + \beta_{N+1}. \tag{S0.5}$$

Take the limits of either side and suppose $n^{2\alpha}S_n \to s$, where s could be 0, finite, or ∞:

$$s \approx s(1 - \beta_{N+1})(1 + N^{-1})^{2\alpha} + \beta_{N+1},$$

from which it can be inferred that $s(\beta_{N+1}-2\alpha N^{-1}) \approx \beta_{N+1}$. Three possible scenarios depending on the speed at which $\beta_n \to 0$:

(A) $n\beta_n \to \infty$: in this case $s = 1$, i.e. $S_n \approx n^{-2\alpha}$;

(B) $n\beta_n \to b$ for some $b > 0$ in this case $s = b/(b-2\alpha)$, i.e. $S_n \approx \frac{b}{(b-2\alpha)} n^{-2\alpha}$;

(C) $n\beta_n \to 0$: in this case $S_N \propto \exp\Big[-\Big(\sum_{n=1}^{N}\beta_n\Big)(1+o(1))\Big]$.

The proof of case (C) is as follows. First note that S_n is decreasing, i.e. $S_n > S_{n+1}$, which together with (S0.4) means that $S_n > n^{-2\alpha}$. In fact, $S_n n^{2\alpha} \uparrow \infty$, for if it is bounded, then it must have a limit, say $s \geq 1$ which together with (S0.5):

$$(N+1)^{2\alpha} S_{N+1} = N^{2\alpha} S_N (1 - \beta_{N+1} + \frac{2\alpha}{N+1}) + \beta_{N+1}.$$

where since $\beta_n = o(n^{-1})$, we approximately have

$$(N+1)^{2\alpha} S_{N+1} > N^{2\alpha} S_N (1 + \frac{\alpha}{N+1}) > N^{2\alpha} S_N + \frac{\alpha}{N+1},$$

which could only imply that $S_n n^{2\alpha} \uparrow \infty$. Now rewrite (S0.4) as

$$\frac{S_{N+1} - S_N}{S_N} = -\beta_{N+1} + \beta_{N+1} \frac{1}{(N+1)^{2\alpha} S_N} = -\beta_{N+1}(1 + o(1)). \quad \text{(S0.6)}$$

Expressed in the form of differential equations

$$d(log S_N) = -\beta_{N+1}(1 + o(1)) \Rightarrow S_N = C \exp\Big[-\Big(\sum_{n=1}^{N} \beta_n\Big)(1 + o(1))\Big];$$

for some $C > 0$; the proof of case (C) is thus complete. □

Corollary S0.2. Depending on the rate of β_n converging to 0,

(A) $n\beta_n \to b$: in this case what holds in general is that $(\beta_{n+1} - \beta_n)/\beta_n = n^{-1}$, so that

$$\tilde{S}_N = \frac{b^2}{(2b + p\alpha - 1)} N^{p\alpha - 1}(1 + o(1));$$

(B) $n\beta_n \to \infty,\ (\beta_N - \beta_{N+1})/\beta_{N+1} = o(\beta_N)$: $\tilde{S}_n \propto n^{p\alpha}\beta_n$;

(C) $n\beta_n \to \infty,\ (\beta_N - \beta_{N+1})/(\beta_{N+1}\beta_N) = b$ for some $b > 0$: $\tilde{S}_n \propto n^{p\alpha}\beta_n$;

(D) $n\beta_n \to 0$: $\tilde{S}_N \propto \exp\Big[\Big(- 2\sum_{n=1}^{N} \beta_n\Big)(1 + o(1)\Big]$.

Proof of Corollary S0.2 What follows from (2.3) and the definition of $\tilde{S}_n$ is that

$$\tilde{S}_{N+1} = (1 - \beta_{N+1})^2 \tilde{S}_N + (N+1)^\alpha \beta_{N+1}^2.$$

Divide either side by a factor of $(N+1)^\alpha \beta_{N+1}$ and suppose $n^{-\alpha}\tilde{S}_n/\beta_n \to s$, where s again could be 0, finite, or ∞:

$$\begin{aligned} & s \approx (1 - \beta_{N+1})^2 s \frac{N^\alpha \beta_N}{(N+1)^\alpha \beta_{N+1}} + \beta_{N+1} \\ & \Rightarrow s(2\beta_{N+1} + \alpha N^{-1} - \frac{\beta_N - \beta_{N+1}}{\beta_{N+1}}) \approx \beta_{N+1}; \end{aligned}$$

Proof of cases (A), (B) and (C) thus follows.

For case (D), first note that $\tilde{S}_n \le n^\alpha$ and since $\beta_n = o(n^{-1})$, we have

$n^{1-\alpha}\tilde{S}_n \uparrow \infty$, which could be inferred from the following

$$\begin{aligned}\frac{\tilde{S}_{N+1}}{(N+1)^{\alpha-1}} &= \frac{\tilde{S}_N}{N^{\alpha-1}}(1-\beta_{N+1})^2\frac{N^{\alpha-1}}{(N+1)^{\alpha-1}} + (N+1)\beta_{N+1}^2 \\ &= \frac{\tilde{S}_N}{N^{\alpha-1}}(1+\frac{1-\alpha}{2N}) + o(N^{-1}).\end{aligned}$$

Therefore, similar to (S0.6) we have

$$\frac{\tilde{S}_{N+1}-\tilde{S}_N}{\tilde{S}_N} = -2\beta_{N+1} + \beta_{N+1}^2 + \frac{N\beta_{N+1}^2}{N^{1-\alpha}\tilde{S}_N} = -2\beta_{N+1}(1+o(1));$$

the proof of case (D) is thus complete. □

Proof of Lemma 2.6 The derivation of optimal β_n is as follows. First note that the asymptotically equivalent problem is as follows:

$$\min_{w(.),c} \frac{c^4}{4}\Big\{\int_0^1 w(x)x^{-2/(p+4)}dx\Big\}^2 + c^{-1}\int_0^1 w(x)^2 x^{1/(p+4)}dx \qquad \text{(S0.7)}$$

subject to $c > 0$, $w(.) \geq 0$ and $\int_0^1 w(x)dx = 1$. Yet we also need to ensure that $w(.)$ could be realized via the sequential updating procedure associated with the ONLINE estimator with weighting series $\{\beta_i,\ i \geq 1\}$, i.e.

$$\beta_i \prod_{k=i+1}^{n}(1-\beta_k) = \frac{1}{n}w(i/n), \quad i = 1, 2, \cdots, n$$

A sufficient and necessary condition is that $w(.)$ meets this requirement: for any a, b, $x \in R$, $w(ax)/w(bx)$ is a function of a and b only. Alternatively, we have $w(ax) = g(a)g(x)$ for all a, $x \in R$ and some function $g(.)$. Since this means $w(x) = [g(x^{1/n})]^n$ for any $x > 0$ and positive integer n, we know immediately that $g(1) = 1$ and thus $w(ax) = w(a)w(x)$ and $w(1) = 1$. Furthermore

$$\frac{\partial w(ax)}{\partial a} = xw'(ax) = w(x)w'(a);$$

with $a = 1$, this translates into $w(x) = aw'(x)x$ for some constant a, whence

$$\partial\Big(\log w(x)\Big) = ax^{-1} \Rightarrow w(x) = cx^a$$

for some c. In this case, the function to minimize in (S0.7) turns out to be

$$\min_{a,c}\Big[\frac{c^4}{4}\Big(\int_0^1 x^{a-2/(p+4)}dx\Big)^2 + c^{-1}\int_0^1 x^{2a+1/(p+4)}dx\Big](a+1)^2. \qquad \text{(S0.8)}$$

For any given $a > 0$, the optimal $c = \{p(a+(p+2)/(p+4))/2\}^{1/(p+4)}$; plug this into (S0.8) and it equates the following

$$\min_a \frac{(a+1)^{p+4}}{\{a+(p+2)/(p+4)\}^{p+2}},$$

which is again minimized when $a = 0$. □

Proof of Lemma 3.3 The following results in matrix theory will be used: suppose $a_n \to 0$, A and B are two fixed matrices of the same dimension and A^{-1} exists, then

$$(A + a_nB)^{-1} = A^{-1} - a_nA^{-1}BA^{-1} + O(a_n^2).$$

First, it is easy to establish that

$$\tilde{\mathcal{S}}_n(\mathbf{x}) = f(\mathbf{x})\mathbf{I}_{p+1} + \frac{cn^{-\alpha}}{1-\alpha}\begin{bmatrix} 0 & \nabla f(\mathbf{x}) \\ \nabla^\top f(\mathbf{x}) & \mathbf{0} \end{bmatrix} + O_p(n^{-2\alpha} + n^{(\alpha-1)/2})$$

$$[\tilde{\mathcal{S}}_n(\mathbf{x})]^{-1} = f^{-1}(\mathbf{x})\mathbf{I}_{p+1} - \frac{cn^{-\alpha}}{(1-\alpha)f^2(\mathbf{x})}\begin{bmatrix} 0 & \nabla f(\mathbf{x}) \\ \nabla^\top f(\mathbf{x}) & \mathbf{0} \end{bmatrix} + O_p(n^{-2\alpha} + n^{(\alpha-1)/2}).$$

Secondly, based on the local Taylor expansion of $m(.)$ around $\mathbf{x}$

$$Y_i = \varepsilon_i + \tilde{\mathbf{X}}_{in}^\top(\mathbf{x})m_i(\mathbf{x}) + \frac{1}{2}\mathbf{X}_{ix}^\top\mathcal{H}_m(\mathbf{x})\mathbf{X}_{ix} + O(h_i^3), \quad m_i(\mathbf{x}) = [m(\mathbf{x}), h_i\nabla^\top m(\mathbf{x})]^\top,$$

we have

$$\begin{aligned}
\tilde{\mathcal{S}}_n(\mathbf{x}), Y) = &\frac{1}{n}\sum_{i=1}^n K_{\tilde{h}_i}(\mathbf{X}_{ix})\tilde{\mathbf{X}}_{in}(\mathbf{x})\tilde{\mathbf{X}}_{in}^\top(\mathbf{x})m_i(\mathbf{x}) + \frac{1}{n}\sum_{i=1}^n K_{\tilde{h}_i}(\mathbf{X}_{ix})\tilde{\mathbf{X}}_{in}(\mathbf{x})\varepsilon_i \\
&+\frac{1}{2n}\sum_{i=1}^n K_{\tilde{h}_i}(\mathbf{X}_{ix})\tilde{\mathbf{X}}_{in}(\mathbf{x})\mathbf{X}_{ix}^\top\mathcal{H}_m(\mathbf{x})\mathbf{X}_{ix} + O(n^{-3\alpha}) \\
= &\frac{1}{n}\sum_{i=1}^n K_{\tilde{h}_i}(\mathbf{X}_{ix})\tilde{\mathbf{X}}_{in}(\mathbf{x})\tilde{\mathbf{X}}_{in}^\top(\mathbf{x})m_i(\mathbf{x}) + \frac{1}{n}\sum_{i=1}^n K_{\tilde{h}_i}(\mathbf{X}_{ix})\tilde{\mathbf{X}}_{in}(\mathbf{x})\varepsilon_i \\
&+\frac{c^2n^{-2\alpha}f(\mathbf{x})}{2(1-2\alpha)}\begin{bmatrix} tr\{\mathcal{H}_m(\mathbf{x})\} \\ \mathbf{0} \end{bmatrix} + o(n^{-2\alpha}) + O_p(n^{-1/2}).
\end{aligned}$$

Observing that

$$m_i(\mathbf{x}) = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & h_i\mathbf{I}_p \end{bmatrix}\begin{bmatrix} m(\mathbf{x}) \\ \nabla m(\mathbf{x}) \end{bmatrix},$$

we further have

$$[\tilde{\mathcal{S}}_n(\mathbf{x})]^{-1}\frac{1}{n}\sum_{i=1}^{n}K_{\tilde{h}_i}(\mathbf{X}_{ix})\tilde{\mathbf{X}}_{in}(\mathbf{x})\tilde{\mathbf{X}}_{in}^{\top}(\mathbf{x})\begin{bmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}m(\mathbf{x})\\ \nabla m(\mathbf{x})\end{bmatrix}=\begin{bmatrix}m(\mathbf{x})\\ \mathbf{0}\end{bmatrix},$$

$$\frac{1}{n}\sum_{i=1}^{n}K_{\tilde{h}_i}(\mathbf{X}_{ix})\tilde{\mathbf{X}}_{in}(\mathbf{x})\tilde{\mathbf{X}}_{in}^{\top}(\mathbf{x})\begin{bmatrix}0 & \mathbf{0}\\ \mathbf{0} & h_i\mathbf{I}_p\end{bmatrix}$$

$$= f(\mathbf{x})\begin{bmatrix}0 & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_p\sum_{i=1}^{n}h_i/n\end{bmatrix}+\begin{bmatrix}0 & \nabla^{\top}f(\mathbf{x})\sum_{i=1}^{n}h_i^2/n\\ \mathbf{0} & \mathbf{0}\end{bmatrix}+O_p(n^{-3\alpha}+n^{-1/2}),$$

$$\frac{1}{n}\sum_{i=1}^{n}K_{\tilde{h}_i}(\mathbf{X}_{ix})\tilde{\mathbf{X}}_{in}(\mathbf{x})\tilde{\mathbf{X}}_{in}^{\top}(\mathbf{x})\begin{bmatrix}0 & \mathbf{0}\\ \mathbf{0} & h_i\mathbf{I}_p\end{bmatrix}\begin{bmatrix}m(\mathbf{x})\\ \nabla m(\mathbf{x})\end{bmatrix}$$

$$= f(\mathbf{x})\begin{bmatrix}0\\ \frac{cn^{-\alpha}}{1-\alpha}\nabla^{\top}m(\mathbf{x})\end{bmatrix}+\begin{bmatrix}\frac{c^2n^{-2\alpha}}{1-2\alpha}\nabla^{\top}m(\mathbf{x})\nabla f(\mathbf{x})\\ \mathbf{0}\end{bmatrix}+O_p(n^{-3\alpha}+n^{-1/2}),$$

$$[\tilde{\mathcal{S}}_n(\mathbf{x})]^{-1}\frac{1}{n}\sum_{i=1}^{n}K_{\tilde{h}_i}(\mathbf{X}_{ix})\tilde{\mathbf{X}}_{in}(\mathbf{x})\tilde{\mathbf{X}}_{in}^{\top}(\mathbf{x})\begin{bmatrix}0 & \mathbf{0}\\ \mathbf{0} & h_i\mathbf{I}_p\end{bmatrix}\begin{bmatrix}m(\mathbf{x})\\ \nabla m(\mathbf{x})\end{bmatrix}$$

$$=\begin{bmatrix}0\\ \frac{cn^{-\alpha}}{1-\alpha}\nabla^{\top}m(\mathbf{x})\end{bmatrix}+\frac{c^2\alpha^2n^{-2\alpha}}{(1-2\alpha)(1-\alpha)^2}\begin{bmatrix}\nabla^{\top}m(\mathbf{x})\nabla f(\mathbf{x})/f(\mathbf{x})\\ \mathbf{0}\end{bmatrix}+O_p(n^{-3\alpha}+n^{-1/2}).$$

Therefore,

$$\tilde{m}_n(\mathbf{x}) = [\tilde{\mathcal{S}}_n(\mathbf{x})]^{-1}\tilde{\mathcal{S}}_n(\mathbf{x},Y) = \begin{bmatrix} m(\mathbf{x}) \\ \frac{cn^{-\alpha}}{1-\alpha}\nabla^\top m(\mathbf{x}) \end{bmatrix} + \frac{c^2\alpha^2 n^{-2\alpha}}{(1-2\alpha)(1-\alpha)^2}\begin{bmatrix} \nabla^\top m(\mathbf{x})\nabla f(\mathbf{x})/f(\mathbf{x}) \\ \mathbf{0} \end{bmatrix}$$
$$+\frac{c^2 n^{-2\alpha}}{2(1-2\alpha)}\begin{bmatrix} tr\{\mathcal{H}_m(\mathbf{x})\} \\ \mathbf{0} \end{bmatrix} + \frac{1}{nf\mathbf{x}}\sum_{i=1}^n K_{\tilde{h}_i}(\mathbf{X}_{ix})\tilde{\mathbf{X}}_{in}(\mathbf{x})\varepsilon_i + O_p(n^{-3\alpha}+n^{-1/2}).$$

Proof of Lemma 4.1. If U_i is near u_0, we apply the following local Taylor expansion concerning function $g_k(.)$:

$$g_k(U_i) = g_k(u_0) + g_k^{(1)}(u_0)U_{i0} + \frac{1}{2}g_k^{(2)}(u_0)U_{i0}^2 + O(|U_{i0}|^3), \quad U_{i0} = U_i - u_0.$$

The proof of (4.18) is done similarly to Fan and Zhang (1999).

$$\frac{1}{n}\sum_{i=1}^n K_{h_n}(U_{i0})\mathbf{X}_i Y_i = \frac{1}{n}\sum_{i=1}^n K_{h_n}(U_{i0})\mathbf{X}_i\varepsilon_i + \frac{1}{n}\sum_{i=1}^n K_{h_n}(U_{i0})\mathbf{X}_i\mathbf{X}_i^\top\mathbf{g}(u_0)$$
$$+\frac{1}{n}\sum_{i=1}^n K_{h_n}(U_{i0})U_{i0}\mathbf{X}_i\mathbf{X}_i^\top\mathbf{g}^{(1)}(u_0) + \frac{1}{2n}\sum_{i=1}^n K_{h_n}(U_{i0})U_{i0}^2\mathbf{X}_i\mathbf{X}_i^\top\mathbf{g}^{(2)}(u_0),$$
$$\frac{1}{n}\sum_{i=1}^n K_{h_n}(U_{i0})\mathbf{X}_i\mathbf{X}_i = (\nu.f)(u_0) + \frac{1}{2}(\nu.f)^{(2)}(u)h_n^2 + O_p(h_n^4 + (nh_n)^{-1/2})$$
$$+O_p(h_n^3 + (nh_n)^{-1/2}).$$

Then (4.18) follows by considering the ratio of these two terms. Similarly, from its definition in (4.17), we have

$$\begin{aligned}\tilde{\mathbf{g}}_n(u_0) = \mathbf{g}(u_0) &+ \Big[\sum_{i=1}^n K_{\tilde{h}_i}(U_{i0})\mathbf{X}_i\mathbf{X}_i^\top\Big]^{-1}\sum_{i=1}^n K_{\tilde{h}_i}(U_{i0})U_{i0}\mathbf{X}_i\mathbf{X}_i^\top\mathbf{g}^{(1)}(u_0)\\ &+\Big[2\sum_{i=1}^n K_{\tilde{h}_i}(U_{i0})\mathbf{X}_i\mathbf{X}_i^\top\Big]^{-1}\sum_{i=1}^n K_{\tilde{h}_i}(U_{i0})U_{i0}^2\mathbf{X}_i\mathbf{X}_i^\top\mathbf{g}^{(2)}(u_0) \quad (S0.9)\\ &+\Big[\sum_{i=1}^n K_{\tilde{h}_i}(U_{i0})\mathbf{X}_i\mathbf{X}_i^\top\Big]^{-1}\sum_{i=1}^n K_{\tilde{h}_i}(U_{i0})\mathbf{X}_i\varepsilon_i + O(n^{-3/5}).\end{aligned}$$

For the 'inverse matrix',

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n K_{\tilde{h}_i}(U_{i0})\mathbf{X}_i\mathbf{X}_i^\top &= (\nu.f)(u_0) + (\nu.f)^{(2)}(u_0)\frac{1}{2n}\sum_{i=1}^n \tilde{h}_i^2\\ &+O(n^{-1}\sum_{i=1}^n \tilde{h}_i^3) + O_p\Big(n^{-1}(\sum_{i=1}^n \tilde{h}_i^{-1})^{1/2}\Big).\end{aligned}$$

The proof is thus complete when plugging this into (S0.9). □

Proof of Lemma 4.2 Again with $h_n = O(n^{-1/5})$, we have standard results such as

$$\begin{aligned}\Sigma_n &= (\nu.f)(u_0)\otimes\mathbf{I}_2 + h_n(\nu.f)^{(1)}(u_0)\otimes\tilde{\mathbf{I}}_2 + O(h_n^2)\\ \mathcal{S}_n &= \Sigma_n\mathbf{g}_n(u_0) + \frac{1}{2n}\sum_{i=1}^n K_{\tilde{h}_i}(U_{i0})U_{i0}^2\mathbf{X}_{n,i}(u_0)\mathbf{X}_i^\top\mathbf{g}^{(2)}(u_0)\\ &+\frac{1}{n}\sum_{i=1}^n K_{\tilde{h}_i}(U_{i0})\mathbf{X}_{n,i}(u_0)\varepsilon_i + O_p(n^{-1/2}),\end{aligned}$$

where $\mathbf{I}_2$ is the 2×2 identity matrix, $\tilde{\mathbf{I}}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Consequently,

$$\hat{\mathbf{g}}_n(u_0) = \mathbf{g}_n(u_0) + \Sigma_n^{-1}\frac{1}{n}\sum_{i=1}^{n} K_{\tilde{h}_i}(U_{i0})\mathbf{X}_{n,i}(u_0)\varepsilon_i$$
$$+\Sigma_n^{-1}\frac{1}{2n}\sum_{i=1}^{n} K_{\tilde{h}_i}(U_{i0})\mathbf{X}_{n,i}(u_0)U_{i0}^2\mathbf{X}_i^\top\mathbf{g}^{(2)}(u_0) + O_p(n^{-1/2}).$$

(4.22) thus follows from facts

$$\mathbf{I}_{q,2q}\Sigma_n^{-1} = [(\nu.f)(u_0)]^{-1} \otimes [1,0] + O(h_n);$$
$$\Big([(\nu.f)(u_0)]^{-1} \otimes [1,0]\Big)\mathbf{X}_{n,i}(u_0) = [(\nu.f)(u_0)]^{-1}\mathbf{X}_i.$$

The proof is complete. □

Proof of Lemma 4.3 We will repeatedly refer to he following properties of Kronecker product: when the order of matrices permit the indicated operations,

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD);\ (A \otimes B)^{-1} = A^{-1} \otimes B^{-1};\ (A \otimes B)^\top = A^\top \otimes B^\top.$$

Write $\tilde{g}_n(u_0) \equiv (n\tilde{\Sigma}_n)^{-1}(n\tilde{\mathcal{S}}_n)$, i.e.

$$\tilde{\Sigma}_n = \frac{1}{n}\sum_{i=1}^{n} K_{\tilde{h}_i}(U_{i0})\tilde{\mathbf{X}}_{n,i}(u_0)\tilde{\mathbf{X}}_{n,i}^\top(u_0);\ \tilde{\mathcal{S}}_n = \frac{1}{n}\sum_{i=1}^{n} K_{\tilde{h}_i}(U_{i0})\tilde{\mathbf{X}}_{n,i}(u_0)Y_i$$

We start with the inverse matrix.

$$\tilde{\mathbf{X}}_{n,i}(u_0)\tilde{\mathbf{X}}_{n,i}^\top(u_0) = (\mathbf{X}_i\mathbf{X}_i^\top) \otimes \left(\begin{bmatrix} 1 \\ U_{i0}/\tilde{h}_i \end{bmatrix} [1, U_{i0}/\tilde{h}_i] \right),$$

$$E(\tilde{\Sigma}_n) = (\nu.f)(u_0) \otimes \mathbf{I}_2 + \tilde{h}_i(\nu.f)^{(1)}(u_0) \otimes \tilde{\mathbf{I}}_2 + O(\tilde{h}_i^2), \quad \text{(S0.10)}$$

$$\mathrm{Var}[(\tilde{\Sigma}_n)_{ij}] = O(\tilde{h}_i^{-1}), \quad \text{(S0.11)}$$

where the $O(.)$ terms are all uniform in $i \geq 1$. Therefore,

$$\begin{aligned} \tilde{\Sigma}_n &= (\nu.f)(u_0) \otimes \mathbf{I}_2 + (\nu.f)^{(1)}(u_0) \otimes \tilde{\mathbf{I}}_2 \\ &\quad \times (\frac{1}{n}\sum_{i=1}^n \tilde{h}_i)\frac{5c}{4}n^{-1/5} + O(\frac{1}{n}\sum_{i=1}^n \tilde{h}_i^2 +)n^{-2/5} + O\Big(n^{-1}(\sum_{i=1}^n \tilde{h}_i^{-1})\Big)n^{-2/5}, \\ \tilde{\Sigma}_n^{-1} &= [(\nu.f)(u_0)]^{-1} \otimes \mathbf{I}_2 \quad \text{(S0.12)} \\ &\quad - \frac{5c}{4}n^{-1/5}\Big([(\nu.f)]^{-1}(u_0)[(\nu.f)^{(1)}(u_0)][(\nu.f)]^{-1}(u_0)\Big) \otimes \tilde{\mathbf{I}}_2 + O(n^{-2/5}). \end{aligned}$$

Next, based on expansion like

$$\begin{aligned} Y_i &= \sum_{k=1}^q \Big(g_k(u_0) + g_k^{(1)}(u_0)U_{i0} + \frac{1}{2}g_k^{(2)}(u_0)U_{i0}^2\Big)x_{ik} + O(|U_{i0}|^3) + \varepsilon_i \\ &= \tilde{\mathbf{X}}_{n,i}^\top(u_0)\tilde{\mathbf{g}}_i(u_0) + \frac{1}{2}\sum_{k=0}^q g_k^{(2)}(u_0)U_{i0}^2 x_{ik} + O(|U_{i0}|^3) + \varepsilon_i, \end{aligned}$$

where recall that $\mathbf{g}_i(u_0) = [g_1(u_0), h_i g_1^{(1)}(u_0), \cdots, g_q(u_0), h_i g_q^{(1)}(u_0)]^\top$, we have

$$\begin{aligned} \tilde{\mathcal{S}}_n &= \frac{1}{n}\sum_{i=1}^n [K_{\tilde{h}_i}(U_{i0})\tilde{\mathbf{X}}_{n,i}(u_0)\tilde{\mathbf{X}}_{n,i}^\top(u_0)\tilde{\mathbf{g}}_i(u_0)] + \frac{1}{n}\sum_{i=1}^n K_{\tilde{h}_i}(U_{i0})\tilde{\mathbf{X}}_{n,i}(u_0)\varepsilon_i \\ &\quad + \frac{1}{2n}\sum_{i=1}^n K_{\tilde{h}_i}(U_{i0})U_{i0}^2\tilde{\mathbf{X}}_{n,i}(u_0)\mathbf{X}_i^\top\mathbf{g}^{(2)}(u_0) + O(n^{-1}(\sum_{i=1}^n \tilde{h}_i^2)^{1/2}). \end{aligned}$$

Seeing that

$$\tilde{\mathbf{g}}_i(u_0) = \left(\mathbf{I}_{q+1} \otimes \begin{bmatrix} 1 & 0 \\ 0 & \tilde{h}_i \end{bmatrix} \right) \tilde{\mathbf{g}}(u_0),$$

we have

$$\begin{aligned} \tilde{\mathbf{X}}_{n,i}(u_0)\tilde{\mathbf{X}}_{n,i}^{\top}(u_0)\tilde{\mathbf{g}}_i(u_0) &= (\mathbf{X}_i\mathbf{X}_i^{\top}) \otimes \left(\begin{bmatrix} 1 \\ U_{i0}/\tilde{h}_i \end{bmatrix} [1, U_{i0}/\tilde{h}_i] \begin{bmatrix} 1 & 0 \\ 0 & \tilde{h}_i \end{bmatrix} \right) \tilde{\mathbf{g}}(u_0) \\ &= (\mathbf{X}_i\mathbf{X}_i^{\top}) \otimes \begin{bmatrix} 1 & U_{i0} \\ U_{i0}/\tilde{h}_i & U_{i0}^2/\tilde{h}_i \end{bmatrix} \tilde{\mathbf{g}}(u_0) \qquad \text{(S0.13)} \\ &= (\mathbf{X}_i\mathbf{X}_i^{\top}) \otimes \begin{bmatrix} 1 & 0 \\ U_{i0}/\tilde{h}_i & 0 \end{bmatrix} \tilde{\mathbf{g}}(u_0) + (\mathbf{X}_i\mathbf{X}_i^{\top}) \otimes \begin{bmatrix} 0 & U_{i0} \\ 0 & U_{i0}^2/\tilde{h}_i \end{bmatrix} \tilde{\mathbf{g}}(u_0), \end{aligned}$$

with the first matrix having vectors of zeros as its even-numbered columns, while the second having vectors of zeros as its odd-numbered columns. For the first matrix, it follows from the definition of Σ_n that

$$(\tilde{\Sigma}_n)^{-1}\frac{1}{n}\sum_{i=1}^{n} K_{\tilde{h}_i}(U_{i0})(\mathbf{X}_i\mathbf{X}_i^{\top}) \otimes \begin{bmatrix} 1 & 0 \\ U_{i0}/\tilde{h}_i & 0 \end{bmatrix} \tilde{\mathbf{g}}(u_0) = \mathbf{g}(u_0) \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \text{(S0.14)}$$

The dealing of the second matrix is more complicated. First we claim that its (non-zero) entries of the matrix are all of order $o_p(n^{-1/2})$: for any $i, j = 1, \cdots, q$,

$$\frac{1}{n}\sum_{i=1}^{n} K_{\tilde{h}_i}(U_{i0})x_{ik}x_{ij}U_{i0} = O_p(n^{-3/5}),\ \ \frac{1}{n}\sum_{i=1}^{n} K_{\tilde{h}_i}(U_{i0})x_{ik}x_{ij}U_{i0}^2/\tilde{h}_i = O_p(n^{-3/5}).$$

As for their expectations,

$$\frac{1}{n}\sum_{i=1}^{n} E\Big(K_{\tilde h_i}(U_{i0})x_{ik}x_{ij}U_{i0}\Big) = (\nu.f)^{(1)}_{kj}(u_0)\frac{1}{n}\sum_{i=1}^{n}\tilde h_i^2(\frac{5c^2}{3}n^{-2/5}) + O(n^{-4/5}),$$
$$\frac{1}{n}\sum_{i=1}^{n} E\Big(K_{\tilde h_i}(U_{i0})x_{ik}x_{ij}U_{i0}^2/\tilde h_i\Big) = (\nu.f)_{kj}(u_0)\frac{1}{n}\sum_{i=1}^{n}\tilde h_i(\frac{5c}{4}n^{-1/5}) + O(n^{-3/5}).$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i\mathbf{X}_i^\top)\otimes\begin{bmatrix}0 & U_{i0}\\ 0 & U_{i0}^2/\tilde h_i\end{bmatrix} = \frac{5c}{4}n^{-1/5}(\nu.f)(u_0)\otimes\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}$$
$$+\frac{5c^2}{3}n^{-2/5}(\nu.f)^{(1)}(u_0)\otimes\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} + o_p(n^{-1/2}),$$

$$\tilde\Sigma_n^{-1}\frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i\mathbf{X}_i^\top)\otimes\begin{bmatrix}0 & U_{i0}\\ 0 & U_{i0}^2/\tilde h_i\end{bmatrix} = \frac{5c}{4}n^{-1/5}\mathbf{I}_{q+1}\otimes\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix} + o_p(n^{-1/2})$$
$$+\frac{5c^2}{3}n^{-2/5}\Big([(\nu.f)(u_0)]^{-1}(\nu.f)^{(1)}(u_0)\Big)\otimes\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$$
$$-\frac{25c^2}{16}n^{-2/5}\Big([(\nu.f)(u_0)]^{-1}[(\nu.f)^{(1)}(u_0)]\Big)\otimes\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$$
$$= \frac{5c}{4}n^{-1/5}\mathbf{I}_{q+1}\otimes\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix} + o_p(n^{-1/2})] \qquad \text{(S0.15)}$$
$$+\frac{5c^2}{48}n^{-2/5}\Big([(\nu.f)(u_0)]^{-1}[(\nu.f)^{(1)}(u_0)]\Big)\otimes\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}.$$

Since

$$\mathbf{I}_{q+1} \otimes \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tilde{\mathbf{g}}(u_0) = \mathbf{g}^{(1)}(u_0) \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{S0.16}$$

$$\Big([(\nu.f)(u_0)]^{-1}(\nu.f)^{(1)}(u_0)\Big) \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \tilde{\mathbf{g}}(u_0)$$

$$= \Big([(\nu.f)(u_0)]^{-1}(\nu.f)^{(1)}(u_0)\mathbf{g}^{(1)}(u_0)\Big) \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{S0.17}$$

combining (S0.13), (S0.13), (S0.14), (S0.15), (S0.16) and(S0.17), we have

$$(\tilde{\Sigma}_n)^{-1}\frac{1}{n}\sum_{i=1}^{n}[K_{\tilde{h}_i}(U_{i0})\tilde{\mathbf{X}}_{n,i}(u_0)\tilde{\mathbf{X}}_{n,i}^{\top}(u_0)\mathbf{g}_i(u_0)] \tag{S0.18}$$

$$= \mathbf{g}(u_0) \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{5c}{4}\mathbf{g}^{(1)}(u_0) \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} + o_p(n^{-1/2})$$

$$+\frac{5c^2}{48}n^{-2/5}\Big([(\nu.f)(u_0)]^{-1}[(\nu.f)^{(1)}(u_0)]\mathbf{g}^{(1)}(u_0)\Big) \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

For the second term in (S0.13), since it is easy to verify that its variance of order $o(n^{-8/5})$, thus we only need to consider its expectation.

$$E\Big(K_{\tilde{h}_i}(U_{i0})U_{i0}^2\tilde{\mathbf{X}}_{n,i}(u_0)\mathbf{X}_i^{\top}\mathbf{g}^{(2)}(u_0)\Big) = \tilde{h}_i^2[(\nu.\mathbf{g}^{(2)}.f)(u_0)] \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$+\tilde{h}_i^3[(\nu.f)^{(1)}.g^{(2)}](u_0) \otimes \begin{bmatrix} 0 \\ \mu_4(K) \end{bmatrix} + O(\tilde{h}_i^4).$$

Therefore,

$$\frac{1}{2n}\sum_{i=1}^{n} K_{\tilde{h}_i}(U_{i0})U_{i0}^2\tilde{\mathbf{X}}_{n,i}(u_0)\mathbf{X}_i^\top\mathbf{g}^{(2)}(u_0)$$
$$= \frac{5c^2}{6}n^{-2/5}[(\nu.\mathbf{g}^{(2)}.f)(u_0)] \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{5c^3}{4}n^{-3/5}[(\nu.f)^{(1)}.\mathbf{g}^{(2)}](u_0) \otimes \begin{bmatrix} 0 \\ \mu_4(K) \end{bmatrix} + O_p(n^{-4/5}),$$

and consequently

$$(\tilde{\Sigma}_n)^{-1}\frac{1}{2n}\sum_{i=1}^{n} K_{\tilde{h}_i}(U_{i0})U_{i0}^2\tilde{\mathbf{X}}_{n,i}(u_0)\mathbf{X}_i^\top\mathbf{g}^{(2)}(u_0) \tag{S0.19}$$
$$= \frac{5c^2}{6}n^{-2/5}\mathbf{g}^{(2)}(u_0) \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{5c^3}{4}n^{-3/5}\Big([(\nu.f)^{-1}(u_0)][(\nu.f)^{(1)}.\mathbf{g}^{(2)}](u_0)\Big) \otimes \begin{bmatrix} 0 \\ \mu_4(K) \end{bmatrix}$$
$$-\frac{25c^3}{24}n^{-3/5}\Big([(\nu.f)]^{-1}(u_0)[(\nu.f)^{(1)}(u_0)]\mathbf{g}^{(2)}(u_0)\Big) \otimes \begin{bmatrix} \mu_4(K) \\ 0 \end{bmatrix} + O_p(n^{-4/5});$$

(4.24) thus follows from (S0.13), (S0.18) and (S0.19). □