SUPPLEMENTAL MATERIAL FOR: THEORETICAL ANALYSIS OF NONPARAMETRIC FILAMENT ESTIMATION

By Wanli Qiao and Wolfgang Polonik

Due to page constraints on the main article this supplement presents the proofs of some technical results from Qiao and Polonik (2015) (Appendix A) as well as some miscellaneous results (Appendix B) that are used in the proofs. Appendix B also contains the derivan of the function G and some if its properties that play in important role in the paper.

APPENDIX A: TECHNICAL PROOFS

THEOREM 3.2 AND ITS PROOF

For convenience of the reader we first restate Theorem 3.2.

Theorem 3.2 *Under assumptions (**F**1)–(**F**2), (**K**1)–(**K**2) and (**H**1), for any* $x_0 \in \mathcal{G}$, $0 < \gamma < \infty$ *and* $0 \le T_{min}, T_{max} < \infty$, $T_{min} + T_{max} \neq 0$ *with* $\{\mathfrak{X}_{x_0}(t),\, t \in [-T_{min}, T_{max}]\} \subset \mathcal{H}$ *and*

$$\inf_{-T_{min} \le s < u \le T_{max}} \left\| \frac{1}{u-s} \int_s^u V(\mathfrak{X}_{x_0}(\lambda)) d\lambda \right\| \ge \gamma, \tag{A.1}$$

the sequence of stochastic process $\sqrt{nh^5}(\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t))$, $-T_{min} \le t \le T_{max}$, *converges weakly in the space* $C[-T_{min}, T_{max}] := C([-T_{min}, T_{max}], \mathbb{R}^2)$ *of* $\mathbb{R}^2$*-valued continuous functions on* $[-T_{min}, T_{max}]$ *to the Gaussian process* $\omega(t), -T_{min} \le t \le T_{max}$*, satisfying the SDE*

$$\begin{aligned} d\omega(t) =& \frac{\sqrt{\beta}}{2} \tilde{G}(\mathfrak{X}_{x_0}(t)) v(\mathfrak{X}_{x_0}(t)) dt + \nabla V(\mathfrak{X}_{x_0}(t)) \omega(t) dt \\ &+ \left\{ \tilde{G}(\mathfrak{X}_{x_0}(t)) \left[\int\int \mathbb{K}(\mathfrak{X}_{x_0}(t), \tau, z) f(\mathfrak{X}_{x_0}(t)) dz d\tau \right] \tilde{G}(\mathfrak{X}_{x_0}(t))^T \right\}^{1/2} dW(t) \end{aligned} \tag{A.2}$$

with initial condition $\omega(0) = 0$*, where* $W(t), t \ge 0$ *is a two-sided standard Brownian motion in* $\mathbb{R}^2$*,*

$$v(x) = \begin{pmatrix} \int K(z) z^T \nabla^2 f^{(2,0)}(x) z dz \\ \int K(z) z^T \nabla^2 f^{(1,1)}(x) z dz \\ \int K(z) z^T \nabla^2 f^{(0,2)}(x) z dz \end{pmatrix} \in \mathbb{R}^3, \tag{A.3}$$

and

$$\mathbb{K}(x, \tau, z) := d^2 K(z) \left[d^2 K(\tau V(x) + z) \right]^T \in \mathbb{R}^{3 \times 3}. \tag{A.4}$$

PROOF. The structure of the proof follows Koltchinskii et al. (2007). Let $\hat{\mathfrak{Y}}_{x_0}(t) = \hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)$. We will find sequences of stochastic processes $\hat{\mathfrak{Z}}_{x_0}(t) \equiv \hat{\mathfrak{Z}}_{x_0,n}(t)$ and $\hat{\mathfrak{D}}_{x_0}(t) \equiv \hat{\mathfrak{D}}_{x_0,n}(t)$ such that

$$\hat{\mathfrak{Y}}_{x_0}(t) = \hat{\mathfrak{Z}}_{x_0}(t) + \hat{\mathfrak{D}}_{x_0}(t), \quad t \in [-T_{min}, T_{max}],$$

where $\sqrt{nh^5}\hat{\mathfrak{Z}}_{x_0,n}(t)$, $-T_{min} \le t \le T_{max}$, converges weakly to the Gaussian process $\omega(t)$ defined in (A.2) and

$$\sup_{-T_{min} \le t \le T_{max}} |\hat{\mathfrak{D}}_{x_0}(t)| = o_p\Big(\frac{1}{\sqrt{nh^5}}\Big). \tag{A.5}$$

This immediately implies the assertion of the theorem. For ease of notation we drop from now on the index x_0 in this proof. Write

$$\hat{\mathfrak{Y}}(t) = \int_0^t (\hat{V} - V)(\mathfrak{X}(s))ds + \int_0^t \nabla V(\mathfrak{X}(s))\hat{\mathfrak{Y}}(s)ds + \hat{\mathfrak{R}}(t), \tag{A.6}$$

where

$$\hat{\mathfrak{R}}(t) := \int_0^t [\hat{V}(\hat{\mathfrak{X}}(s)) - \hat{V}(\mathfrak{X}(s)) - \nabla V(\mathfrak{X}(s))\hat{\mathfrak{Y}}(s)]ds.$$

It is not difficult to see (by following the proof on page 1585 of Koltchinskii et al. 2007) that

$$\sup_{-T_{min} \le t \le T_{max}} \|\hat{\mathfrak{R}}(t)\| = o_p\Big(\sup_{-T_{min} \le t \le T_{max}} \|\hat{\mathfrak{Y}}(t)\|\Big). \tag{A.7}$$

Suppose $\hat{\mathfrak{Z}}$ satisfies the differential equation

$$\frac{d\hat{\mathfrak{Z}}(t)}{dt} = \hat{V}(\mathfrak{X}(t)) - V(\mathfrak{X}(t)) + \nabla V(\mathfrak{X}(t))\hat{\mathfrak{Z}}(t), \quad \hat{\mathfrak{Z}}(0) = 0. \tag{A.8}$$

which means that

$$\hat{\mathfrak{Z}}(t) = \int_0^t [\hat{V}(\mathfrak{X}(s))) - V(\mathfrak{X}(s))]ds + \int_0^t \nabla V(\mathfrak{X}(s))\hat{\mathfrak{Z}}(s)ds. \tag{A.9}$$

Denote $\hat{\mathfrak{D}}(t) := \hat{\mathfrak{Y}}(t) - \hat{\mathfrak{Z}}(t)$. Then from (A.6) and (A.9) we have

$$\hat{\mathfrak{D}}(t) = \int_0^t \nabla V(\mathfrak{X}(s))\hat{\mathfrak{D}}(s)ds + \hat{\mathfrak{R}}(t).$$

Following the proof on page 1586 of Koltchinskii et al. (2007) we can show that

$$\sup_{-T_{min} \le t \le T_{max}} \|\hat{\mathfrak{D}}(t)\| = o_p\Big(\sup_{-T_{min} \le t \le T_{max}} \|\hat{\mathfrak{Z}}(t)\|\Big) \quad \text{as} \quad n \to \infty.$$

As we will show below, the sequence $\sqrt{nh^5}\hat{\mathfrak{Z}}(t), -T_{min} \le t \le T_{max}$, converges in distribution to the Gaussian process $\omega(t)$, it follows that

$$\sup_{-T_{min} \le t \le T_{max}} \|\hat{\mathfrak{Z}}(t)\| = O_P\left(\frac{1}{\sqrt{nh^5}}\right).$$

Immediately we get (A.5).

We now show the asserted weak convergence of $\sqrt{nh^5}\hat{\mathfrak{Z}}(t), -T_{min} \le t \le T_{max}$. Denote by $C_0^{(1)}[-T_{min}, T_{max}]$ the set of all $\mathbb{R}^2$-valued continuously differentiable functions on $[-T_{min}, T_{max}]$ with value zero at the point 0. We define a mapping $\mathcal{U} : C_0^{(1)}[-T_{min}, T_{max}] \mapsto C[-T_{min}, T_{max}]$ such that for any $S \in C_0^{(1)}[-T_{min}, T_{max}]$, $\mathcal{U}S(t)$ is the solution of the following differential equation in $\mathbb{R}^2$:

$$\frac{du(t)}{dt} = \frac{dS(t)}{dt} + \nabla V(\mathfrak{X}(t))u(t), \quad u(0) = 0. \tag{A.10}$$

As indicated on page 1587 of Koltchinskii et al. (2007), $\mathcal{U}$ is a Lipschitz mapping with respect to the uniform distance.

Define a stochastic process χ satisfying the SDE

$$d\chi(t) = \frac{\sqrt{\beta}}{2}\tilde{G}(t)v(\mathfrak{X}(t))dt + \left\{\tilde{G}(t)\left[\int\int \Psi(\mathfrak{X}(t), \tau, z)f(\mathfrak{X}(t))dzd\tau\right]\tilde{G}(t)^T\right\}^{1/2} dW(t) \tag{A.11}$$

with initial condition $\chi(0) = 0$. Then $\mathcal{U}\chi$ satisfies (A.2) with value zero at the point 0, i.e., $\mathcal{U}\chi = \omega$.

Denote two sequences of processes

$$\chi_n(t) := \sqrt{nh^5}\int_0^t [\hat{V}(\mathfrak{X}(s)) - V(\mathfrak{X}(s))]ds,$$
$$\omega_n(t) := \sqrt{nh^5}\hat{\mathfrak{Z}}(t).$$

Then by (A.8) we have $\omega_n = \mathcal{U}\chi_n$. As we will show below, the sequence χ_n converges weakly in the space $C[-T_{min}, T_{max}]$ to χ. Since $\mathcal{U}$ is Lipschitz, ω_n converges weakly to ω, which is the assertion.

It thus remains to show the weak convergence of the sequence χ_n. In what follows we will write the explicit form of $\hat{V}(x)$ and $V(x)$, i.e., $\hat{V}(x) = G(d^2\hat{f}(x))$ and $V(x) = G(d^2 f(x))$. Recall that $G = (G_1, G_2)^T$ and $\tilde{G}(x) := \nabla G(d^2 f(x))$. Denote $\mathcal{J}_n(t) := \int_0^t \tilde{G}(\mathfrak{X}(s))d^2\hat{f}(\mathfrak{X}(s))ds$ and $\mathcal{J}(t) := \int_0^t \tilde{G}(\mathfrak{X}(s))d^2 f(\mathfrak{X}(s))ds$. A first order Taylor expansion with respect to the variables in G gives

$$\int_0^t [\hat{V}(\mathfrak{X}(s)) - V(\mathfrak{X}(s))]ds = \mathcal{J}_n(t) - \mathcal{J}(t) + \mathcal{R}_n(t), \tag{A.12}$$

where the remainder term

$$\mathcal{R}_n(t) := \begin{pmatrix} \int_0^t d^2(\hat{f}-f)(\mathfrak{X}(s))^T M_1(s) d^2(\hat{f}-f)(\mathfrak{X}(s))ds \\ \int_0^t d^2(\hat{f}-f)(\mathfrak{X}(s))^T M_2(s) d^2(\hat{f}-f)(\mathfrak{X}(s))ds \end{pmatrix}$$

with

$$M_i(s) := \int_0^1 \nabla^2 G_i(d^2 f(\mathfrak{X}(s)) + \tau d^2(\hat{f}-f)(\mathfrak{X}(s)))d\tau, \quad i=1,2. \tag{A.13}$$

For this remainder term we have

$$\begin{aligned}\sup_{t\in[-T_{min},T_{max}]} \|\mathcal{R}_n(t)\| &\le 2 \sup_{\substack{t\in[-T_{min},T_{max}]\\ i=1,2}} \Big| \int_0^t d^2(\hat{f}-f)(\mathfrak{X}(s))^T M_i(s) d^2(\hat{f}-f)(\mathfrak{X}(s))ds \Big| \\ &\le 2T\Big(\sup_{x\in\mathbb{R}^2} \|d^2(\hat{f}-f)(x)\|\Big)^2 \sup_{\substack{i=1,2\\ w\in\mathcal{H}^\epsilon}} \|\nabla^2 G_i(w)\| = O_P\Big(\frac{\log n}{nh^6}\Big) = o_p\Big(\frac{1}{\sqrt{nh^5}}\Big),\end{aligned}$$

because $\sup_{x\in\mathcal{H}^\epsilon} \|d^2(\hat{f}-f)(x)\| = O_p(\sqrt{\log n/nh^6})$ and $nh^8/\log n \to \infty$ by assumption (**H**1).

The linear approximation $\mathcal{J}_n(t) - \mathcal{J}(t)$ in (A.12) is a sum of iid random variables for each fixed n, indicating asymptotic normality. In fact, weak convergence of the process $\sqrt{nh^5}(\mathcal{J}_n(t) - \mathcal{J}(t))$ to $\chi(t)$ can be shown by proving convergence of finite dimensional distributions along with asymptotic stochastic equicontinuity by following the proof on pages 1591-1595 of Koltchinskii et al. (2007). Then we conclude the proof of this theorem. Further details are omitted. □

THE PROOF OF (5.11)

By our assumptions, the maps $K_\ell(z)$ are Lipschitz continuous. Let c_ℓ be the corresponding Lipschitz constants. Fix $\tau > 0$, and let $x_0, x_0^* \in \mathcal{G}$ be such that $\|x_0 - x_0^*\| \le \tau$, and assume that $T_{x_0} \le T_{x_0^*}$. We first show that there exists a constant $C > 0$ with

$$\sup_{x\in\mathbb{R}^2} \Big| \int_0^t K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s) - x}{h}\Big) - K_\ell\Big(\frac{\mathfrak{X}_{x_0^*}(s) - x}{h}\Big)ds\Big| \le C\tau \qquad \ell = 1,2,3, \;\; 0 < t < T_{x_0}. \tag{A.14}$$

Recall that the support of $K(z)$ is contained in a unit ball $\mathcal{B}(0,1)$. Thus we have with $A_{x,x_0}(h) = \{s : \|\mathfrak{X}_{x_0}(s) - x\| \le h\}$ that

$$\begin{aligned}&\int_0^t \Big|K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s) - x}{h}\Big) - K_\ell\Big(\frac{\mathfrak{X}_{x_0^*}(s) - x}{h}\Big)\Big|ds \\ &= \int_0^t \Big|K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s) - x}{h}\Big) - K_\ell\Big(\frac{\mathfrak{X}_{x_0^*}(s) - x}{h}\Big)\Big| \mathbf{1}_{A_{x,x_0}(h)\cup A_{x,x_0^*}(h)}(s)\, ds \\ &\le \int_0^t c_\ell \Big\|\frac{\mathfrak{X}_{x_0}(s) - \mathfrak{X}_{x_0^*}(s)}{h}\Big\| \mathbf{1}_{A_{x,x_0}(h)\cup A_{x,x_0^*}(h)}(s)\, ds.\end{aligned} \tag{A.15}$$

The Lebesgue measure of the set $A_{x,x_0}(h) \cup A_{x,x_0^*}(h)$ is of the order $O(h)$. To see that observe that for $s, s' \in A_{x,x_0}(h)$ we have $\|\mathfrak{X}_{x_0}(s) - \mathfrak{X}_{x_0}(s')\| \le 2h$, so that with $\gamma_{\mathcal{G}} > 0$ from (3.11):

$$2h \ge \|\mathfrak{X}_{x_0}(s) - \mathfrak{X}_{x_0}(s')\| = \Big\| \int_s^{s'} V(\mathfrak{X}_{x_0}(t))\, dt \Big\| = \Big\| \frac{1}{s-s'} \int_s^{s'} V(\mathfrak{X}_{x_0}(t)\, dt \Big\|\, |s - s'| \ge \gamma_{\mathcal{G}}\, |s - s'|.$$

It follows that $\mathrm{Leb}(A_{x,x_0}(h)) \le 2h/\gamma_{\mathcal{G}}$ and the same holds for $A_{x,x_0^*}(h)$, so that

$$\mathrm{Leb}(A_{x,x_0}(h) \cup A_{x,x_0^*}(h)) \le \frac{4h}{\gamma_{\mathcal{G}}}. \tag{A.16}$$

To continue the argument we will use the fact that $\mathfrak{X}_{x_0}(s)$ is Lipschitz continuous in x_0 under the sup-norm. To see this note that for any $x_0, x_0' \in \mathcal{G}$ and $s \in [0, \min(T_{x_0}, T_{x_0'})]$,

$$\begin{aligned} \|\mathfrak{X}_{x_0}(s) - \mathfrak{X}_{x_0'}(s)\| &= \Big\| x_0 - x_0' + \int_0^s [V(\mathfrak{X}_{x_0}(t)) - V(\mathfrak{X}_{x_0'}(t))]\, dt \Big\| \\ &\le \|x_0 - x_0'\| + \sup_{x\in\mathcal{H}} \|\nabla V(x)\|_F \int_0^s \|\mathfrak{X}_{x_0}(t) - \mathfrak{X}_{x_0'}(t)\| dt. \end{aligned}$$

Applying Gronwall's inequality, we have for all $s \in [0, \min(T_{x_0}, T_{x_0'})]$

$$\|\mathfrak{X}_{x_0}(s) - \mathfrak{X}_{x_0'}(s)\| \le \|x_0 - x_0'\| \exp\Big\{ \min(T_{x_0}, T_{x_0'}) \sup_{x\in\mathcal{H}} \|\nabla V(x)\|_F \Big\}. \tag{A.17}$$

By using (A.16) and (A.17), the integral in (A.15) can now be bounded by

$$\begin{aligned} c_\ell \int_0^t \Big\| \frac{\mathfrak{X}_{x_0}(s) - \mathfrak{X}_{x_0^*}(s)}{h} \Big\| \mathbf{1}_{A_{x,x_0} \cup A_{x,x_0^*}}(s)\, ds &\le\ c\, \frac{\tau}{h} c_\ell \int_0^t \mathbf{1}_{A_{x,x_0} \cup A_{x,x_0^*}}(s)\, ds \\ &\le\ c\, \frac{\tau}{h}\, c_\ell\, \frac{4h}{\gamma_{\mathcal{G}}} = \frac{4c\, c_\ell}{\gamma_{\mathcal{G}}}\, \tau, \end{aligned}$$

where $c = \exp\Big\{ T_{\mathcal{G}} \sup_{x\in\mathcal{H}} \|\nabla V(x)\|_F \Big\}$. We have verified (A.14). Next we show (5.11). Fix $0 < \tau \le 1$. Since $x_0 \in \mathcal{G}$ and $\mathcal{G}$ is compact there exist points $x_{0,1}, \ldots, x_{0,N_1}$ such that for all $x_0 \in \mathcal{G}$ we have $\min_{i=1,\ldots,N_1} \|x_0 - x_{0,i}\| \le \tau$ and $N_1 = N_1(\tau) \le A_1 \frac{1}{\tau}$ for some constant $A_1 > 0$. Further, let $t_1, \ldots, t_{N_2}$ be such that for all $t \in [0, \max_{x_0\in\mathcal{G}} T_{x_0}]$ we have $\min_{i=1,\ldots,N_2} |t - t_i| \le \tau$ and $N_2 = N_2(\tau) \le A_2 \frac{1}{\tau}$ for $A_2 > 0$. With these definitions let

$$\widetilde{\mathcal{F}}_{j,\ell}(\tau) = \Big\{ \omega_{j,\ell}(\cdot; x_{0,i}, t_k) : i = 1, \ldots, N_1(\tau),\, t_k \le T_{x_{0,i}}, k \in \{1, \ldots, N_2(\tau)\} \Big\}.$$

It is clear that the number of functions in $\widetilde{\mathcal{F}}_{j,\ell}(\tau)$ is bounded by $C\left(\frac{1}{\tau}\right)^2$. Without loss of generality we can assume that $T_{x_{0,i}} = \max\{T_{x_0} : \|x_0 - x_{0,i}\| \le \tau\}$, $i = 1, \ldots, N_1(\tau)$. Otherwise define $x_{0,i}^* = \mathrm{argmax}\{T_{x_0} : \|x_0 - x_{0,i}\| \le \tau\}$, replace $x_{0,i}$ by $x_{0,i}^*$ and change τ to 2τ.

We show that $\widetilde{\mathcal{F}}_{j,\ell}(\tau)$ serves as an approximating class of functions assuring (5.11). To see that let $\omega_{j\ell}(\cdot; x_0, t) \in \mathcal{F}_{j,\ell}$ and let $\omega_{j,\ell}(\cdot; x_0^*, t^*) \in \widetilde{\mathcal{F}}_{j,\ell}$, where $\|x_0 - x_0^*\| \le \tau$ and $|t - t^*| \le \tau$. To verify (5.11) we show that for some constant $C > 0$ the following two bounds hold for $t \in [0, T_{x_0}]$:

$$d_\infty(\omega_{j,\ell}(\cdot; x_0, t), \omega_{j,\ell}(\cdot; x_0^*, t)) \le C\tau, \tag{A.18}$$

$$d_\infty(\omega_{j,\ell}(\cdot; x_0^*, t), \omega_{j,\ell}(\cdot; x_0^*, t^*)) \le C\tau, \tag{A.19}$$

where $d_\infty(f, g)$ denotes the supremum distance of functions f and g. First we show (A.18). Recall that $\widetilde{G}_j(u) = \nabla G_j(d^2 f(u))$, and denote $\widetilde{G}_{j,\ell}(u) = \frac{\partial}{\partial u_\ell}\widetilde{G}_j(u)$, $\ell = 1, 2, 3$. We obtain

$$\begin{aligned}
& d_\infty(\omega_{j,\ell}(\cdot; x_0, t), \omega_{j,\ell}(\cdot; x_0^*, t)) \\
= & \sup_{x\in\mathbb{R}^2} \Big| \int_0^t \Big[\widetilde{G}_{j,\ell}(\mathfrak{X}_{x_0}(s)) K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s) - x}{h}\Big) - \widetilde{G}_j(\mathfrak{X}_{x_0^*}(s)) K_\ell\Big(\frac{\mathfrak{X}_{x_0^*}(s) - x}{h}\Big)\Big] ds \Big| \\
\le & \sup_{x\in\mathbb{R}^2} \int_0^t \Big|\widetilde{G}_{j,\ell}(\mathfrak{X}_{x_0}(s)) - \widetilde{G}_{j,\ell}(\mathfrak{X}_{x_0^*}(s))\Big| K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s) - x}{h}\Big) ds & \text{(A.20)} \\
& \quad + \sup_{x\in\mathbb{R}^2} \int_0^t \Big| K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s) - x}{h}\Big) - K_\ell\Big(\frac{\mathfrak{X}_{x_0^*}(s) - x}{h}\Big)\Big| \, |\widetilde{G}_{j,\ell}(\mathfrak{X}_{x_0^*}(s))| ds. & \text{(A.21)}
\end{aligned}$$

Now, with $M_\ell = \sup_u K_\ell(u)$, the term in (A.20) can further be bounded by

$$M_\ell \int_0^t |\widetilde{G}_{j,\ell}(\mathfrak{X}_{x_0}(s)) - \widetilde{G}_{j,\ell}(\mathfrak{X}_{x_0^*}(s))| ds \le c' \, M_\ell T_{\mathcal{G}} \, \tau,$$

for some $c' > 0$, where we are using Lipschitz continuity of $\widetilde{G}_{j,\ell}$ along with (A.17). To bound (A.21) we first use the fact that the functions $\widetilde{G}_{j,\ell}$ are bounded, so that the integral in (A.21) is less than or equal to

$$\sup_u |\widetilde{G}_{j,\ell}(u)| \quad \sup_{x\in\mathbb{R}^2} \int_0^t \Big| K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s) - x}{h}\Big) - K_\ell\Big(\frac{\mathfrak{X}_{x_0^*}(s) - x}{h}\Big)\Big| ds \le C \sup_u |\widetilde{G}_{j,\ell}(u)| \, \tau,$$

by using (A.14). This shows (A.18). The bound (A.19) follows by using the boundedness of the integrant in the definition of the functions $\omega_{j,\ell}$ (see (5.8)). This completes the proof of (5.11). □

PROPOSITION 5.1 AND ITS PROOF

First we show uniform consistency of $\hat{\theta}_{x_0}$ which is needed for the proof of Proposition 5.1.

Proposition A.1 *Under assumptions (**F**1)–(**F**6), (**K**1)–(**K**2) and (**H**1), we have*

$$\sup_{x_0\in\mathcal{G}} |\hat{\theta}_{x_0} - \theta_{x_0}| = o_p(1), \tag{A.22}$$

PROOF. Fix $\epsilon > 0$ arbitrary (and small enough). We want to show that $P(\sup_{x_0\in\mathcal{G}} |\hat{\theta}_{x_0} - \theta_{x_0}| > \epsilon) \to 0$ as $n \to \infty$. Recall that by definition,

$$\theta_{x_0} = \underset{t}{\operatorname{argmin}}\{|t| : \langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t))\rangle = 0, \lambda_2(\mathfrak{X}_{x_0}(t)) < 0\},$$

and a similar definition holds for $\hat{\theta}_{x_0}$ (with $f, \mathfrak{X}, V$ and λ_2 replaced by our estimates). Without loss of generality, consider $\theta_{x_0} > 0$. Let ρ_{x_0} be the first time traveling backwards that the trajectory hits the boundary of $\mathcal{H}$, i.e., $\rho_{x_0} = \inf\{s : s \le \theta_{x_0}, \ \{\mathfrak{X}_{x_0}(t) : t \in [s, \theta_{x_0}]\} \subset \mathcal{H}\}$. Note that $\mathfrak{X}_{x_0}(\rho_{x_0}) \in \mathcal{H}$ since $\mathcal{H}$ is compact. Under assumption (**F**3), θ_{x_0} defined in (2.6) is unique, therefore for ϵ small enough, $\mathfrak{X}_{x_0}(\theta_{x_0})$ is the only filament point on $[(-\theta_{x_0} - \epsilon) \vee \rho_{x_0}, \theta_{x_0}]$. For any $x_0 \in \mathcal{G}$ let $\mathcal{C}_{x_0,\epsilon} = \{t \in [(-\theta_{x_0} - \epsilon) \vee \rho_{x_0}, \theta_{x_0} - \epsilon] : \langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t))\rangle = 0\}$. Assume for now that $\mathcal{C}_{x_0,\epsilon} \neq \emptyset$. Note that $\mathcal{C}_{x_0,\epsilon}$ is a compact set. For $\eta > 0$ let $\mathcal{C}^\eta_{x_0,\epsilon}$ denote the η-neighborhood of $\mathcal{C}_{x_0,\epsilon}$ intersected with $[(-\theta_{x_0} - \epsilon) \vee \rho_{x_0}, \theta_{x_0} - \epsilon]$. It suffices to show that

(i) $P(\forall x_0 \in \mathcal{G}, \exists\, t_{x_0} \in [\theta_{x_0} - \epsilon, \theta_{x_0} + \epsilon]$ s.t. $\langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(t_{x_0})), \hat{V}(\hat{\mathfrak{X}}_{x_0}(t_{x_0}))\rangle = 0) \to 1$,

(ii) $P(\sup_{x_0\in\mathcal{G}, t\in[\theta_{x_0}-\epsilon, \theta_{x_0}+\epsilon]} \hat{\lambda}_2(\hat{\mathfrak{X}}_{x_0}(t)) < 0) \to 1$,

(iii) There exists an $\eta > 0$ such that $P(\inf_{x_0\in\mathcal{G}, t\in\mathcal{C}^\eta_{x_0,\epsilon}} \hat{\lambda}_2(\hat{\mathfrak{X}}_{x_0}(t)) > 0) \to 1$ and

(iv) $P(\inf_{x_0\in\mathcal{G}, t\in[(-\theta_{x_0}-\epsilon)\vee\rho_{x_0}, \theta_{x_0}-\epsilon)\setminus\mathcal{C}^\eta_{x_0,\epsilon}} |\langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(t)), \hat{V}(\hat{\mathfrak{X}}_{x_0}(t))\rangle| > 0) \to 1$.

By our regularity assumptions $\{a_{x_0}(t) = \langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t))\rangle, \, x_0 \in \mathcal{G}\}$ and $\{\lambda_2(\mathfrak{X}_{x_0}(t)), x_0 \in \mathcal{G}\}$ are classes of equi-continuous functions on $t \in [\rho_{x_0}, \theta_{x_0} + a^*]$. Further, under assumption (**F**5) $a_{x_0}(t)$ is strictly monotonic at θ_{x_0}. Also the derivatives $a'_{x_0}(t), x_0 \in \mathcal{G}$ form an equi-continuous class of functions, and thus for any $\epsilon > 0$ sufficiently small there exists a $\delta > 0$ such that $\inf_{x_0\in\mathcal{G}} \inf_{t\in[\theta_{x_0}-\epsilon, \theta_{x_0}+\epsilon]} |a'_{x_0}(t)| > \delta$.

Moreover, since θ_{x_0} corresponds to the first filament point, we have that for any $x_0 \in \mathcal{G}$ and $t \in \mathcal{C}_{x_0,\epsilon}$, $\lambda_2(\mathfrak{X}_{x_0}(t)) > 0$. Note that here we have used assumption (**F**6). Since both $\mathcal{G}$ and $\mathcal{C}_{x_0,\epsilon}$ are compact, there exist $\eta, \zeta > 0$ with $\inf\limits_{x_0\in\mathcal{G}, t\in[(-\theta_{x_0}-\epsilon)\vee\rho_{x_0}, \theta_{x_0}-\epsilon)\setminus\mathcal{C}^\eta_{x_0,\epsilon}} |\langle \nabla f(\mathfrak{X}_{x_0}(t_{x_0})), V(\mathfrak{X}_{x_0}(t_{x_0}))\rangle| > \zeta$ and $\inf_{x_0\in\mathcal{G}, t\in\mathcal{C}^\eta_{x_0,\epsilon}} \lambda_2(\mathfrak{X}_{x_0}(t)) > \zeta$.

The proofs of (ii) - (iv) are straightforward by using uniform consistency of $\hat{\lambda}_2(\hat{\mathfrak{X}}_{x_0}(t))$ and $\langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(t)), \hat{V}(\hat{\mathfrak{X}}_{x_0}(t))\rangle$ along with the fact that the corresponding theoretical quantities satisfy the inequalities corresponding to the three probability statements from (ii) - (iv). Uniform consistency of $\langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(t)), \hat{V}(\hat{\mathfrak{X}}_{x_0}(t))\rangle$ can be shown by using Theorem 3.3 and uniform consistency results for the kernel estimators of ∇f and V under our assumptions. Uniform consistency of $\hat{\lambda}_2(\hat{\mathfrak{X}}_{x_0}(t))$ is inherited from uniform consistency of the second derivatives of the kernel estimator and uniform consistency of $\hat{\mathfrak{X}}_{x_0}(t)$ by observing that $\hat{\lambda}_2(\hat{\mathfrak{X}}_{x_0}(t)) = J(d^2(\hat{f}(\hat{\mathfrak{X}}_{x_0}(t)))$ with $J(\cdot)$ being Lipschitz-continuous. Further details are omitted. To see (i) first observe that since $\inf_{x_0\in\mathcal{G}} \inf_{t\in[\theta_{x_0}-\epsilon, \theta_{x_0}+\epsilon]} |a'_{x_0}(t)| > \delta$ there exists an $\eta > 0$ and $t_1, t_2 \in [\theta_{x_0} - \epsilon, \theta_{x_0} + \epsilon]$ with $a_{x_0}(t_1) \ge \eta$ and $a_{x_0}(t_2) \le -\eta$ for all $x_0 \in \mathcal{G}$. Uniform consistency of

$$\widehat{a}_{x_0}(t) = \langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(t)), \hat{V}(\hat{\mathfrak{X}}_{x_0}(t))\rangle \tag{A.23}$$

as an estimator of $a_{x_0}(t)$ implies that the probability of the event $B_n := \{$for all $x_0 \in \mathcal{G}$: $\hat{a}_{x_0}(t_1) \ge \eta/2$ and $\hat{a}_{x_0}(t_2) \le -\eta/2\}$ tends to one as $n \to \infty$. Since $\hat{a}_{x_0}(t)$ is continuous, we have that on B_n that for each $x_0 \in \mathcal{G}$ there exists a $t \in [\theta_{x_0} - \epsilon, \theta_{x_0} + \epsilon]$ with $\hat{a}_{x_0}(t) = 0$. This completes the proof of (A.22) in case $\mathcal{C}_{x_0,\epsilon} \neq \emptyset$. If $\mathcal{C}_{x_0,\epsilon} = \emptyset$, then we can ignore (iii) and the result follows from (i),(ii) and (iv). □

The above proof also shows that the probability of $\widehat{\Theta}_{x_0} = \emptyset$ or $\hat{\theta}_{x_0}$ is not unique for all $x_0 \in \mathcal{G}$ tends to zero as $n \to \infty$. We will thus only consider the case of $\widehat{\Theta}_{x_0} \neq \emptyset$ and unique $\hat{\theta}_{x_0}$ in what follows. Next we derive the convergence rate of $\sup_{x_0 \in \mathcal{G}} |\hat{\theta}_{x_0} - \theta_{x_0}|$.

Now we prove Proposition 5.1. For the convenience of the reader we restate this propositon here.

Proposition 5.1 *Under assumptions (**F1**)–(**F6**), (**K1**)–(**K2**) and (**H1**), we have*

$$\sup_{x_0 \in \mathcal{G}} |\hat{\theta}_{x_0} - \theta_{x_0}| = O_p(\alpha_n),$$

where $\alpha_n = \sqrt{\frac{\log n}{nh^6}}$, *and if in addition* $\sup_{x_0 \in \mathcal{G}} \|\nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\| = 0$, *then* $\alpha_n = \sqrt{\frac{\log n}{nh^5}}$.

PROOF. Note that $\widehat{a}_{x_0}(t)$ is the directional derivative of $\hat{f}$ at t when traversing the support of $\hat{f}$ along the curve $\hat{\mathfrak{X}}_{x_0}(t)$. By definition of a filament point we have $\widehat{a}_{x_0}(\hat{\theta}_{x_0}) = 0$ for all $x_0 \in \mathcal{G}$. A similar interpretation holds for the population quantity $a_{x_0}(t)$. We will use the behavior of $\widehat{a}_{x_0}(t) - a_{x_0}(t)$ around $t = \theta_{x_0}$ to determine the behavior of $\hat{\theta}_{x_0} - \theta_{x_0}$.

By using chain rule and noting that $\nabla\langle\nabla f(x), V(x)\rangle = \nabla^2 f(x)V(x) + \nabla V(x)\nabla f(x)$ we have

$$\begin{aligned} a'_{x_0}(t) &= V(\mathfrak{X}_{x_0}(t))^T \nabla^2 f(\mathfrak{X}_{x_0}(t)) V(\mathfrak{X}_{x_0}(t)) + \langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t))\rangle_{\nabla V(\mathfrak{X}_{x_0}(t))} \\ &= \lambda_2(\mathfrak{X}_{x_0}(t)) \|V(\mathfrak{X}_{x_0}(t))\|^2 + \langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t))\rangle_{\nabla V(\mathfrak{X}_{x_0}(t))} \end{aligned} \tag{A.24}$$

and similarly

$$\widehat{a}'_{x_0}(t) = \hat{\lambda}_2(\hat{\mathfrak{X}}_{x_0}(t)) \|\hat{V}(\hat{\mathfrak{X}}_{x_0}(t))\|^2 + \langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(t)), \hat{V}(\hat{\mathfrak{X}}_{x_0}(t))\rangle_{\nabla \hat{V}(\hat{\mathfrak{X}}_{x_0}(t))}. \tag{A.25}$$

Notice that $a'_{x_0}(t) = \tilde{a}'(\mathfrak{X}_{x_0}(t))$ with $\tilde{a}'(x)$ from (2.10). We can write

$$0 = \widehat{a}_{x_0}(\hat{\theta}_{x_0}) = \langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0})), \hat{V}(\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}))\rangle = \widehat{a}_{x_0}(\theta_{x_0}) + \widehat{a}'_{x_0}(\hat{\xi}_{x_0})(\hat{\theta}_{x_0} - \theta_{x_0}) \tag{A.26}$$

with some $\hat{\xi}_{x_0}$ between $\hat{\theta}_{x_0}$ and θ_{x_0}. We next show that for some $\eta > 0$

$$P(\inf_{x_0 \in \mathcal{G}} |\widehat{a}'_{x_0}(\hat{\xi}_{x_0})| \geq \eta) \to 1. \tag{A.27}$$

To this end we prove that

$$\sup_{x_0 \in \mathcal{G}} \sup_{t \in (\theta_{x_0} - \epsilon, \theta_{x_0} + \epsilon)} |\widehat{a}'_{x_0}(t) - a'_{x_0}(t)| = o_p(1) \quad \text{for } \epsilon > 0 \text{ sufficiently small,} \tag{A.28}$$

and

$$\tilde{a}'(x) = \lambda_2(x)\|V(x)\|^2 + \langle \nabla f(x), V(x)\rangle_{\nabla V(x)} \quad \text{is uniformly continuous in } x \in \mathcal{H}. \tag{A.29}$$

This then implies (A.27) by using standard arguments. Assertion (A.29) is a direct consequence of our regularity assumptions that assure continuity of $\tilde{a}'(x)$ by using compactness of $\mathcal{H}$. The consistency property (A.28) follows from uniform consistency of $\hat{\mathfrak{X}}_{x_0}(t)$ as an estimator for $\mathfrak{X}_{x_0}(t)$ (Lemma B.3) and uniform consistency of $\hat{V}(x)$, $\nabla\hat{V}(x)$, $\nabla\hat{f}(x)$ and $\nabla^2\hat{f}(x)$ (Lemmas B.1 and B.2) by using a continuous mapping argument or arguments similar to the ones presented in the following proof of (A.30).

To prove the assertion of the proposition it remains to show that

$$\sup_{x_0\in\mathcal{G}} |\hat{a}_{x_0}(\theta_{x_0})| = O_p(\alpha_n). \tag{A.30}$$

To see this write

$$\begin{aligned}\hat{a}_{x_0}(\theta_{x_0}) = \hat{a}_{x_0}(\theta_{x_0}) - a_{x_0}(\theta_{x_0}) &= \big\langle \nabla\hat{f}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})), \hat{V}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}))\big\rangle - \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle \\ &= \big\langle \big[\nabla\hat{f}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})) - \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\big], \hat{V}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}))\big\rangle \\ &\quad + \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \big[\hat{V}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})) - V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big]\big\rangle.\end{aligned}$$

The rate (A.30) now follows from the following facts:

$$\sup_{x_0\in\mathcal{G}} \big\|\nabla\hat{f}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})) - \mathbb{E}\nabla\hat{f}(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\| = O_p\Big(\sqrt{\frac{\log n}{nh^5}}\Big), \tag{A.31}$$

$$\sup_{x_0\in\mathcal{G}} \big\|\mathbb{E}\nabla\hat{f}(\mathfrak{X}_{x_0}(\theta_{x_0})) - \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\| = O(h^2), \tag{A.32}$$

$$\sup_{x_0\in\mathcal{G}} \big\|\hat{V}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})) - V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\| = O_p\Big(\sqrt{\frac{\log n}{nh^6}}\Big), \tag{A.33}$$

$$\text{both } V(\mathfrak{X}_{x_0}(\theta_{x_0})) \text{ and } \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})) \text{ are bounded uniformly in } x_0\in\mathcal{G}. \tag{A.34}$$

Properties (A.31) - (A.33) follow from Theorem 3.3, well-known properties of kernel estimators (see Lemma B.1) and Lemma B.2. Property (A.34) follows immediately from our regularity assumptions. The proof of Proposition 5.1 is complete. □

PROOFS OF THEOREM 3.4 AND LEMMA 3.1

Proof of Theorem 3.4 First write with $\tilde{\theta}_{x_0}$ between θ_{x_0} and $\hat{\theta}_{x_0}$,

$$\begin{aligned}\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}) &= \hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) + \hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}) \\ &= \hat{V}(\hat{\mathfrak{X}}_{x_0}(\tilde{\theta}_{x_0}))[\hat{\theta}_{x_0} - \theta_{x_0}] + \hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}) \\ &= V(\mathfrak{X}_{x_0}(\theta_{x_0}))[\hat{\theta}_{x_0} - \theta_{x_0}] + \big[\hat{V}(\hat{\mathfrak{X}}_{x_0}(\tilde{\theta}_{x_0})) - V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big]\,[\hat{\theta}_{x_0} - \theta_{x_0}] \\ &\quad + \hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}).\end{aligned} \tag{A.35}$$

We need a convergence rate of $\sup_{x_0\in\mathcal{G}}\|\hat V(\hat{\mathfrak{X}}_{x_0}(\tilde\theta_{x_0}))-V(\mathfrak{X}_{x_0}(\theta_{x_0}))\|$. Let a^* be as in assumption (**F**3), and $\epsilon>0$ arbitrary. On the set $\{\sup_{x_0\in\mathcal{G}}|\widehat\theta_{x_0}-\theta_{x_0}|<a^*\}\cap\{\sup_{x_0\in\mathcal{G};t\in[\theta_{x_0}-a*,\theta_{x_0}+a*]}\|\hat{\mathfrak{X}}_{x_0}(t)-\mathfrak{X}_{x_0}(t)\|<\epsilon\}$ we have by recalling that $\mathcal{H}^\epsilon$ denotes the ϵ-enlargement of $\mathcal{H}$,

$$\begin{aligned}
&\sup_{x_0\in\mathcal{G}}\|\hat V(\hat{\mathfrak{X}}_{x_0}(\tilde\theta_{x_0}))-V(\mathfrak{X}_{x_0}(\theta_{x_0}))\|\\
&\quad\le \sup_{x_0\in\mathcal{G},t\in[\theta_{x_0}-a*,\theta_{x_0}+a*]}\|\hat V(\hat{\mathfrak{X}}_{x_0}(t))-V(\mathfrak{X}_{x_0}(t))\|+\sup_{x_0\in\mathcal{G}}\|V(\mathfrak{X}_{x_0}(\tilde\theta_{x_0}))-V(\mathfrak{X}_{x_0}(\theta_{x_0}))\|\\
&\quad\le \sup_{x_0\in\mathcal{G},t\in[\theta_{x_0}-a*,\theta_{x_0}+a*]}\|\hat V(\hat{\mathfrak{X}}_{x_0}(t))-V(\mathfrak{X}_{x_0}(t))\|+\sup_{x\in\mathcal{H}}\|\nabla V(x)V(x)\|\sup_{x_0\in\mathcal{G}}|\tilde\theta_{x_0}-\theta_{x_0}|\\
&\quad\le \sup_{x\in\mathcal{H}^\epsilon}\|\hat V(x)-V(x)\|+\sup_{\substack{x_0\in\mathcal{G}\\ t\in[\theta_{x_0}-a*,\theta_{x_0}+a*]}}|V(\hat{\mathfrak{X}}_{x_0}(t))-V(\mathfrak{X}_{x_0}(t))|\\
&\qquad\qquad+\sup_{x\in\mathcal{H}}\|\nabla V(x)V(x)\|\sup_{x_0\in\mathcal{G}}|\tilde\theta_{x_0}-\theta_{x_0}|=\ O_p\Big(\sqrt{\frac{\log n}{nh^6}}\Big)
\end{aligned}\tag{A.36}$$

by using Theorem 3.3, Proposition 5.1 and Lemma B.2. It follows from (A.35), (A.36), Theorem 3.3 and Proposition 5.1 that

$$\sup_{x_0\in\mathcal{G}}\big\|[\hat{\mathfrak{X}}_{x_0}(\hat\theta_{x_0})-\mathfrak{X}_{x_0}(\theta_{x_0})]-V(\mathfrak{X}_{x_0}(\theta_{x_0}))[\hat\theta_{x_0}-\theta_{x_0}]\big\|=O_p\Big(\frac{\log n}{nh^6}\Big)+O_p\Big(\sqrt{\frac{\log n}{nh^5}}\Big)=O_p\Big(\sqrt{\frac{\log n}{nh^5}}\Big).$$

□

Proof of Lemma 3.1 We continue using the notation introduced in (A.23) - (A.25) and (A.29). Since assumption (**F**5) implies that $0<\inf_{x_0\in\mathcal{G}}|a'_{x_0}(\theta_{x_0})|<\infty$ (see discussion given after the assumptions) we obtain from (A.26) and (A.28) that

$$\hat\theta_{x_0}-\theta_{x_0}=-\frac{\widehat a_{x_0}(\theta_{x_0})}{a'_{x_0}(\theta_{x_0})}+O_p(R_n),$$

where $|R_n|\le\sup_{x_0\in\mathcal{G}}|\widehat a'_{x_0}(\hat\xi_{x_0})-a'_{x_0}(x_0)|\,|\hat\theta_{x_0}-\theta_{x_0}|$. Using Proposition 5.1, the assertion of the lemma is now a consequence of assumption (**H**1) as well as

$$\sup_{x_0\in\mathcal{G}}|\widehat a'_{x_0}(\hat\xi_{x_0})-a'_{x_0}(\theta_{x_0})|=O_p\Big(\sqrt{\frac{\log n}{nh^8}}\Big),\tag{A.37}$$

$$\sup_{x_0\in\mathcal{G}}\Big|\,\widehat a_{x_0}(\theta_{x_0})-\langle\nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})),\hat V(\mathfrak{X}_{x_0}(\theta_{x_0}))\rangle\,\Big|=O_p\Big(\sqrt{\frac{\log n}{nh^5}}\Big)\tag{A.38}$$

and

$$\sup_{x_0\in\mathcal{G}}\Big|\,\hat\varphi_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0}))-\frac{\langle\nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})),\hat V(\mathfrak{X}_{x_0}(\theta_{x_0}))\rangle}{a'_{x_0}(\theta_{x_0})}\,\Big|=O_p\Big(\frac{1}{nh^7}\Big).\tag{A.39}$$

To see (A.37), observe that on $A_n(\epsilon) = \{|\hat{\theta}_{x_0} - \theta_{x_0}| \le \epsilon\},\ \epsilon > 0$, uniformly in $x_0 \in \mathcal{G}$

$$\begin{aligned}|\widehat{a}'_{x_0}(\hat{\xi}_{x_0}) - a'_{x_0}(\theta_{x_0})| &\le \sup_{t\in(\theta_{x_0}-\epsilon,\theta_{x_0}+\epsilon)} |\widehat{a}'_{x_0}(t) - a'_{x_0}(t)| + \sup_{\substack{|s-t|\le\epsilon,\\ s,t\in(\theta_{x_0}-\epsilon,\theta_{x_0}+\epsilon)}} |a'_{x_0}(s) - a'_{x_0}(t)| \\ &\le \sup_{t\in(\theta_{x_0}-\epsilon,\theta_{x_0}+\epsilon)} |\widehat{a}'_{x_0}(t) - a'_{x_0}(t)| + C\epsilon\end{aligned}$$

for some $C > 0$, where the last inequality follows from the fact that the derivatives of $a'_{x_0}(t)$ are continuous in both t and x_0 (note here that Lipschitz continuity of $\mathfrak{X}_{x_0}(s)$ in x_0 is shown in (A.17), and thus $a'_{x_0}(t)$ is uniformly bounded for $t \in (\theta_{x_0} - \epsilon, \theta_{x_0} + \epsilon)$ and $x_0 \in \mathcal{G}$). Proposition 5.1 implies that with α_n from Proposition 5.1 we have $P\big(A_n(\alpha_n)\big) \to 1$ as $n \to \infty$, and clearly $\alpha_n = o\Big(\sqrt{\frac{\log n}{nh^8}}\Big)$. It thus remains to show that

$$\sup_{x_0\in\mathcal{G}} \sup_{t\in(\theta_{x_0}-\epsilon,\theta_{x_0}+\epsilon)} |\widehat{a}'_{x_0}(t) - a'_{x_0}(t)| = O_P\bigg(\sqrt{\frac{\log n}{nh^8}}\bigg). \tag{A.40}$$

We have

$$\begin{aligned}&|\widehat{a}'_{x_0}(t) - a'_{x_0}(t)| \\ &\quad\le \Big| \|\hat{V}(\hat{\mathfrak{X}}_{x_0}(t))\|_{\nabla^2\hat{f}(\hat{\mathfrak{X}}_{x_0}(t))} - \|V(\mathfrak{X}_{x_0}(t))\|_{\nabla^2 f(\mathfrak{X}_{x_0}(t))}\Big| \\ &\qquad + \big|\langle\nabla\hat{f}(\hat{\mathfrak{X}}_{x_0}(t)),\, \hat{V}(\hat{\mathfrak{X}}_{x_0}(t))\rangle_{\nabla\hat{V}(\hat{\mathfrak{X}}_{x_0}(t))} - \langle\nabla f(\mathfrak{X}_{x_0}(t)),\, V(\mathfrak{X}_{x_0}(t))\rangle_{\nabla V(\mathfrak{X}_{x_0}(t))}\big|\end{aligned} \tag{A.41}$$

and we will derive the rate of convergence for each of the terms on the right-hand side. As for the first term, we have by using a telescoping argument:

$$\begin{aligned}&\Big| \|\hat{V}(\hat{\mathfrak{X}}_{x_0}(t))\|_{\nabla^2\hat{f}(\hat{\mathfrak{X}}_{x_0}(t))} - \|V(\mathfrak{X}_{x_0}(t))\|_{\nabla^2 f(\mathfrak{X}_{x_0}(t))}\Big| \\ &\qquad\le \big|\,\langle\hat{V}(\hat{\mathfrak{X}}_{x_0}(t)),\, \hat{V}(\hat{\mathfrak{X}}_{x_0}(t)) - V(\mathfrak{X}_{x_0}(t))\rangle_{\nabla^2\hat{f}(\hat{\mathfrak{X}}_{x_0}(t))}\,\big| \\ &\qquad\quad + \big|\,\langle\hat{V}(\hat{\mathfrak{X}}_{x_0}(t)),\, V(\mathfrak{X}_{x_0}(t))\rangle_{\nabla^2\hat{f}(\hat{\mathfrak{X}}_{x_0}(t))-\nabla^2 f(\mathfrak{X}_{x_0}(t))}\,\big| \\ &\qquad\quad\quad + \big|\,\langle\hat{V}(\hat{\mathfrak{X}}_{x_0}(t)) - V(\mathfrak{X}_{x_0}(t)),\, V(\mathfrak{X}_{x_0}(t))\rangle_{\nabla^2 f(\mathfrak{X}_{x_0}(t))}\big|.\end{aligned}$$

Now we can use similar arguments as in the proof of (A.30). The asserted rate then is the slowest of the rates of convergence of $\sup_{x_0\in\mathcal{G}}\sup_{t\in(\theta_{x_0}-\epsilon,\theta_{x_0}+\epsilon)} \|\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\|$ and the rates of $\sup_{x\in\mathcal{H}^\epsilon} \|\nabla^2\hat{f}(x) - \nabla^2 f(x)\|_F$ and $\sup_{x\in\mathcal{H}^\epsilon} \|\hat{V}(x) - V(x)\|$, respectively (see Theorem 3.3 and Lemmas B.1 and B.2), where $\epsilon > 0$ is arbitrary. The resulting rate for the last term on the right-hand side of the above inequality is $O_P\big(\sqrt{\frac{\log n}{nh^6}}\big)$. A similar argument applied to the second term on the right-hand side of (A.41) gives the (slower) asserted rate $O_P\big(\sqrt{\frac{\log n}{nh^8}}\big)$ in (A.40) inherited by the rate of $\sup_{x\in\mathcal{H}^\epsilon} \|\nabla\hat{V}(x) - \nabla V(x)\|_F$ (see Lemma B.2), which in turn is inherited from the rate of convergence of $\sup_{x\in\mathcal{H}^\epsilon} \|\nabla^2\hat{f}(x) - \nabla^2 f(x)\|_F$. This proves (A.37).

In order to see (A.38), first observe that

$$\begin{aligned}&\widehat{a}_{x_0}(\theta_{x_0}) - \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle\\ &\qquad\qquad = \big\langle \nabla \hat f(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})), \hat V(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}))\big\rangle - \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle.\end{aligned}$$

Since $\sup_{x_0\in\mathcal{G}} \|\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})\| = O_p\Big(\sqrt{\frac{\log n}{nh^5}}\Big)$ (Theorem 3.3), and $\sup_{x_0\in\mathcal{G}} \|\nabla \hat f(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})) - \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\| = O_p\Big(\sqrt{\frac{\log n}{nh^5}}\Big)$ (see (A.31), (A.32) and (**H**1)), and since $d^2\hat f$ is uniformly consistent, it is straightforward to see that

$$\sup_{x_0\in\mathcal{G}} \Big|\widehat{a}_{x_0}(\theta_{x_0}) - \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle\Big| = O_p\Big(\sqrt{\frac{\log n}{nh^5}}\Big),$$

This is (A.38). To see (A.39) observe that with

$$\widehat{W}_n(x) = \langle \nabla f(x), d^2(\hat f - f)(x)\rangle_{\nabla G(d^2 f(x))},$$

we have $\hat\varphi_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0})) = \frac{\widehat{W}_n(\mathfrak{X}_{x_0}(\theta_{x_0})) - \mathbb{E}\widehat{W}_n(\mathfrak{X}_{x_0}(\theta_{x_0}))}{a'_{x_0}(\theta_{x_0})}$, so that the assertion follows from

$$\sup_{x_0\in\mathcal{G}} \Big|\,\mathbb{E}\widehat{W}_n(\mathfrak{X}_{x_0}(\theta_{x_0}))\,\Big| = O\Big(\frac{1}{nh^7}\Big) \tag{A.42}$$

and

$$\sup_{x_0\in\mathcal{G}} \big|\,\widehat{W}_n(\mathfrak{X}_{x_0}(\theta_{x_0})) - \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle\,\big| = O_p\Big(\frac{\log n}{nh^6}\Big). \tag{A.43}$$

To see (A.42) we use $\sup_{x\in\mathcal{H}^\epsilon} \|\mathbb{E}d^2\hat f(x) - d^2 f(x)\| = O(h^2)$, which follows by standard arguments. Since $\mathbb{E}\widehat{W}_n(\mathfrak{X}_{x_0}(\theta_{x_0}))$ is a linear combination of the components of bias vector it is of the same order. Assumptions (**H**1) assures that $h^2 = O(\frac{1}{nh^7})$.

As for (A.43), recall our notation $V(x) = G(d^2 f(x))$ and $\hat V(x) = G(d^2\hat f(x))$. We see that

$$\begin{aligned}&\big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle = \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))^T\, [\hat V(\mathfrak{X}_{x_0}(\theta_{x_0})) - V(\mathfrak{X}_{x_0}(\theta_{x_0}))]\\ &\quad = \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))^T \Big[\int_0^1 \nabla G\big[(d^2\hat f + \lambda d^2(\hat f - f))(\mathfrak{X}_{x_0}(\theta_{x_0}))\big]\, d\lambda\,\Big]\; d^2(\hat f - f)(\mathfrak{X}_{x_0}(\theta_{x_0})).\end{aligned}$$

Using standard arguments we obtain

$$\begin{aligned}&\sup_{x_0\in\mathcal{G}} \Big|\widehat{W}_n(\mathfrak{X}_{x_0}(\theta_{x_0})) - \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle\Big|\\ &\qquad\qquad = O_p\Big(\sup_{x_0\in\mathcal{G}} \big|d^2\hat f(\mathfrak{X}_{x_0}(\theta_{x_0})) - d^2 f(\mathfrak{X}_{x_0}(\theta_{x_0}))\big|^2\Big) = O_p\Big(\frac{\log n}{nh^6}\Big).\end{aligned}$$

This completes the proof of (3.14). To prove (3.15) where $\sup_{x_0\in\mathcal{G}} \|\nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\| = 0$, we approximate $\hat a_{x_0}(\theta_{x_0})$ by $\big\langle \nabla f(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})), V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle$ rather than by $\big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), V(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}))\big\rangle$

as we did above. We also will have to deal with the bias of $\nabla\hat{f}(\mathfrak{X}_{x_0}(\theta_{x_0}))$ because this is not negligible here. Let $\mu_n(x) = \mathbb{E}\nabla\hat{f}(x)$. Then a simple telescoping argument gives

$$\begin{aligned}
&\sup_{x_0\in\mathcal{G}}\Big|\hat{a}_{x_0}(\theta_{x_0}) - \langle\mu_n(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})), V(\mathfrak{X}_{x_0}(\theta_{x_0}))\rangle\Big| \\
&\qquad\le \sup_{x_0\in\mathcal{G}}\Big\|\nabla\hat{f}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}))\Big\| \sup_{x_0\in\mathcal{G}}\Big\|\hat{V}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})) - V(\mathfrak{X}_{x_0}(\theta_{x_0}))\Big\| \\
&\qquad\quad + \sup_{x_0\in\mathcal{G}}\Big\|\nabla\hat{f}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})) - \mu_n(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}))\Big\| \sup_{x_0\in\mathcal{G}}\Big\|V(\mathfrak{X}_{x_0}(\theta_{x_0}))\Big\| = O_p\bigg(\sqrt{\frac{\log n}{nh^4}}\bigg) \quad \text{(A.44)}
\end{aligned}$$

by using (A.31) and (A.33) and Lemma B.1. Further, a one-term Taylor expansion gives

$$\begin{aligned}
\langle\nabla f(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})), V(\mathfrak{X}_{x_0}(\theta_{x_0}))\rangle &= \langle\nabla f(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})), V(\mathfrak{X}_{x_0}(\theta_{x_0}))\rangle - \langle\nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), V(\mathfrak{X}_{x_0}(\theta_{x_0}))\rangle \\
&= \langle V(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})\rangle_{\nabla^2 f(\mathfrak{X}_{x_0}(\theta_{x_0}))} + r_n \qquad \text{(A.45)}
\end{aligned}$$

where $r_n = \langle V(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})\rangle_{\hat{A}_n}$ with $\hat{A}_n = \int_0^1 [\nabla^2 f(\mathfrak{X}_{x_0}(\theta_{x_0}) + \lambda(\mathfrak{X}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}))) - \nabla^2 f(\mathfrak{X}_{x_0}(\theta_{x_0}))]\, d\lambda$, and we have

$$r_n = O_p\Big(\frac{\log n}{nh^5}\Big). \qquad \text{(A.46)}$$

The rate is uniform in $x_0 \in \mathcal{G}$ and follows from Theorem 3.3 and our regularity assumptions. Using (A.44) and (A.46) the assertion now follows by using similar arguments as in the first part of the proof. □

PROOFS OF COROLLARIES 3.1 - 3.3

We first prove Corollary 3.2. It follows from Theorem 3.2 that as $n \to \infty$,

$$\sqrt{nh^5}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})) \to \mathcal{N}(m(\theta_{x_0}), \Sigma(\theta_{x_0})) \qquad \text{(A.47)}$$

where $m(\cdot) \in \mathbb{R}^2$ and $\Sigma(\cdot) \in \mathbb{R}^{2\times 2}$ satisfy the ODE

$$\begin{aligned}
\dot{m}(t) &= \frac{\sqrt{\beta}}{2}\tilde{G}(t)\mathcal{B}(\mathfrak{X}_{x_0}(t)) + \tilde{G}(t)m(t), \\
\dot{\Sigma}(t) &= \nabla V(\mathfrak{X}_{x_0}(t))\Sigma(t) + \Sigma(t)[\nabla V(\mathfrak{X}_{x_0}(t))]^T \\
&\qquad + \tilde{G}(t)\bigg[\int\int \Psi(\mathfrak{X}_{x_0}(t), \tau, z) f(\mathfrak{X}_{x_0}(t))\,dz\,d\tau\bigg]\tilde{G}(t)^T
\end{aligned}$$

with $m(0) = x_0$ and $\Sigma(0) = 0$. Here $\tilde{G}(t)$, $\mathcal{B}(\mathfrak{X}_{x_0}(t))$ and $\Psi(\mathfrak{X}_{x_0}(t), \tau, z)$ have been defined in Theorem 3.2.

Observe that by assumption (**H**1) we have $\log n/nh^{13/2} = o(1/\sqrt{nh^5})$ and $\sqrt{nh^5}h^2 \to \beta \geq 0$. Thus the assertion follows from (A.47) above and (3.17), by observing that standard arguments show that

$$h^{-2}\big((\mathbb{E}\nabla\hat{f} - \nabla f)(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}))\big) \to \tilde{b}(\mathfrak{X}_{x_0}(\theta_{x_0})), \tag{A.48}$$

where $\tilde{b}(x) = \frac{1}{2}\mu_2(K)\left(f^{(3,0)}(x) + f^{(1,2)}(x),\ f^{(0,3)} + f^{(2,1)}(x)\right)^T$.

Next we prove Corollary 3.1. By definition of $\hat{\varphi}_{1n}(x)$ (see (3.12)) the approximation (3.16) shows that in this case the convergence properties of $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$ are essentially determined by the deviation $d^2\hat{f}(\mathfrak{X}_{x_0}(\theta_{x_0})) - \mathbb{E}d^2\hat{f}(\mathfrak{X}_{x_0}(\theta_{x_0}))$. The behavior of this difference is well known. With $\mathbf{R} := \mathbf{R}(d^2K)$ where $\mathbf{R}(\cdot)$ defined in assumption (**K**2), we have that under our assumptions (e.g. see Duong et al. 2008)

$$\sqrt{nh^6}(d^2\hat{f}(x) - \mathbb{E}d^2\hat{f}(x)) \to \mathcal{N}(0, f(x)\mathbf{R}) \quad \text{as } n \to \infty, \quad \text{in distribution.} \tag{A.49}$$

It follows from (3.16) and assumption (**H**1) that $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$ has the same asymptotic distribution as $\hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0}))\,V(\mathfrak{X}_{x_0}(\theta_{x_0}))$. The assertion now follows by standard arguments.

Corollary 3.3 follows immediately from (3.16) by observing that

$$\begin{aligned} O_P\Big(\frac{\log n}{nh^7}\Big) &= \sup_{x_0\in\mathcal{G}} \Big|\Big(\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}) + \hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0}))V(\mathfrak{X}_{x_0}(\theta_{x_0}))\Big)^T V^{\perp}(\mathfrak{X}_{x_0}(\theta_{x_0}))\Big| \\ &= \sup_{x_0\in\mathcal{G}} \Big|\Big(\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})\Big)^T V^{\perp}(\mathfrak{X}_{x_0}(\theta_{x_0}))\Big|. \end{aligned}$$

□

PROOF OF CONTINUITY OF $\boldsymbol{\theta_{x_0}}$ AS A FUNCTION IN $\boldsymbol{x_0 \in \mathcal{G}}$.

In our formal proofs we repeatedly apply Theorem 3.3 with $T_{x_0}^{min} = \theta_{x_0} - a^*$ and $T_{x_0}^{max} = \theta_{x_0} + a^*$, and so we need to know that $x_0 \to \theta_{x_0}$ is continuous. We will show this here. First we indicate that filament points are continuous in the starting points of paths, namely, $x_0 \to \mathfrak{X}_{x_0}(\theta_{x_0})$, $x_0 \in \mathcal{G}$ is continuous. To see this, recall that the filament $\mathcal{L} = \{\mathfrak{X}_{x_0}(\theta_{x_0}),\, x_0 \in \mathcal{G}\}$ by assumption is a smooth curve with bounded curvature. For a given $x_0^* \in \mathcal{G}$ and $\epsilon > 0$, consider the set $\mathcal{L}_{x_0^*}(\epsilon) = \{x \in \mathcal{L} : \|\mathfrak{X}_{x_0^*}(\theta_{x_0^*}) - x\| \leq \epsilon\}$. This curve (piece of the filament) has finite length. Now let $\mathcal{G}_{x_0^*} = \{x_0 \in \mathcal{G} : x_0 \rightsquigarrow \mathcal{L}_{x_0^*}(\epsilon)\}$ denote the set of all points on integral curves $\mathfrak{X}_{x_0}(t)$ passing through a filament point on $\mathcal{L}_{x_0^*}(\epsilon)$. By definition, $x_0^* \in \mathcal{G}_{x_0^*}$. Since integral curves are non-overlapping, the set $\mathcal{G}_{x_0^*}$ is delineated by two 'boundary curves' $\mathfrak{X}^L$ and $\mathfrak{X}^U$, say, corresponding to the filament points on the endpoints of the curve $\mathcal{L}_{x_0^*}(\epsilon)$. Let $\delta = \inf\{\|x_0^* - x\|, x \in \mathfrak{X}^L \cup \mathfrak{X}^U\}$. Then, $\delta > 0$. (Otherwise x_0^* would lie on one of the boundary curves.) With this δ we have, $\|x - x_0^*\| < \delta \Rightarrow \|\mathfrak{X}_{x_0^*}(\theta_{x_0^*}) - \mathfrak{X}_x(\theta_x)\| < \epsilon$ by construction.

Now let $x_0, x_0' \in \mathcal{G}$ and without loss of generality assume $\theta_{x_0} \geq \theta_{x_0'}$. Then we have

$$\Big(\mathfrak{X}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0'}(\theta_{x_0'})\Big) + \Big(\mathfrak{X}_{x_0'}(\theta_{x_0'}) - \mathfrak{X}_{x_0}(\theta_{x_0'})\Big) = \mathfrak{X}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0'})$$
$$= (\theta_{x_0} - \theta_{x_0'})\Big(\frac{1}{\theta_{x_0} - \theta_{x_0'}}\int_{\theta_{x_0'}}^{\theta_{x_0}} V(\mathfrak{X}_{x_0}(t))dt\Big).$$

Taking L^2 norm on both sides it follows that

$$|\theta_{x_0} - \theta_{x_0'}|\Big\|\frac{1}{\theta_{x_0} - \theta_{x_0'}}\int_{\theta_{x_0'}}^{\theta_{x_0}} V(\mathfrak{X}_{x_0}(t))dt\Big\| \leq \Big\|\mathfrak{X}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0'}(\theta_{x_0'})\Big\| + \Big\|\mathfrak{X}_{x_0'}(\theta_{x_0'}) - \mathfrak{X}_{x_0}(\theta_{x_0'})\Big\|. \tag{A.50}$$

Similar to (A.17), we can show that there exists a constant $C > 0$ such that $\|\mathfrak{X}_{x_0'}(\theta_{x_0'}) - \mathfrak{X}_{x_0}(\theta_{x_0'})\| \leq C\|x_0 - x_0'\|$. Considering the statement above that $\mathfrak{X}_{x_0}(\theta_{x_0})$ as a function of x_0 is continuous in $\mathcal{G}$, we have that there exists another constant $C' > 0$ such that R.H.S. of (A.50) $\leq C'\|x_0 - x_0'\|$. Using assumption (**F**4), we complete the proof. □

PROOF OF (5.24)

A Taylor expansion of $A_h(x+y)$ gives

$$A_h(x+y) = A_h(x) + \nabla A_h(x)y + \frac{1}{2}\nabla^{\otimes 2}A_h(x)y^{\otimes 2} + o(\|y^{\otimes 2}\|),$$

where the little-o term can be chosen to be independent of h due to assumptions (**F**1)–(**F**2) and the fact that h is bounded. Similarly, a Taylor expansion of $a_h(x+y)$ leads to

$$a_h(x+y) = a_h(x) - a_h(x)^3 A_h(x)^T\mathbf{R}\nabla A_h(x)y - \frac{1}{2}a_h(x)^3\|\nabla A_h(x)y\|_{\mathbf{R}}$$
$$- \frac{1}{2}a_h(x)^3 A_h(x)^T\mathbf{R}\nabla^{\otimes 2}A_h(x)y^{\otimes 2} + o(\|y^{\otimes 2}\|), \tag{A.51}$$

where again the little-o term is independent of h due to the same reason as above. A Taylor expansion of $\int d^2K(x+y-s)[d^2K(x-s)]^T ds$ about $y=0$ gives $\int d^2K(x+y-s)[d^2K(x-s)]^T ds = \mathbf{R} + \frac{1}{2}\int \nabla^{\otimes 2}d^2K(s)y^{\otimes 2}[d^2K(s)]^T ds + o(\|y^{\otimes 2}\|)$, so that

$$A_h(x+y)^T\int d^2K(x+y-s)[d^2K(x-s)]^T ds\; A_h(x)$$
$$= (a_h(x))^{(-2)} + (\nabla A_h(x)y)^T\mathbf{R}A_h(x) + \frac{1}{2}(\nabla^{\otimes 2}A_h(x)y^{\otimes 2})^T\mathbf{R}A_h(x)$$
$$+ \frac{1}{2}A_h(x)^T\int \nabla^{\otimes 2}d^2K(s)y^{\otimes 2}[d^2K(s)]^T ds A_h(x) + o(\|y^{\otimes 2}\|). \tag{A.52}$$

Plugging all these expansions into (5.23) leads to

$r_h(x+y, x)$

$$
\begin{aligned}
&= a_h(x+y)a_h(x)\ A_h(x+y)^T \int d^2K(x+y-s)[d^2K(x-s)]^T ds\ A_h(x) \\
&= \Big\{1 - (a_h(x))^2 A_h(x)^T \mathbf{R}\nabla A_h(x)y - \frac{1}{2}(a_h(x))^2\{(\nabla A_h(x)y)^T\mathbf{R}\nabla A_h(x)y \\
&\qquad + A_h(x)^T\mathbf{R}\nabla^{\otimes 2}A_h(x)y^{\otimes 2}\} + o(\|y^{\otimes 2}\|)\Big\}\Big\{1 + (a_h(x))^2(\nabla A_h(x)y)^T\mathbf{R}A_h(x) \\
&\qquad\qquad + \frac{1}{2}(a_h(x))^2\Big\{A_h(x)^T \int \nabla^{\otimes 2}d^2K(s)y^{\otimes 2}[d^2K(s)]^T ds A_h(x) \\
&\qquad\qquad\qquad + (\nabla^{\otimes 2}A_h(x)y^{\otimes 2})^T\mathbf{R}A_h(x))\Big\} + o(\|y^{\otimes 2}\|)\Big\} \\
&= 1 - \frac{1}{2}(a_h(x))^2(\nabla A_h(x)y)^T\mathbf{R}\nabla A_h(x)y - [(a_h(x))^2 A_h(x)^T\mathbf{R}\nabla A_h(x)y]^2 \\
&\qquad + \frac{1}{2}(a_h(x))^2 A_h(x)^T \int \nabla^{\otimes 2}d^2K(s)y^{\otimes 2}[d^2K(s)]^T ds A_h(x) + o(\|y^{\otimes 2}\|) \\
&= 1 - y^T\Lambda_1(h,x)y - y^T\Lambda_2(hx)y + o(\|y^{\otimes 2}\|) \\
&= 1 - y\Lambda(h,x)y^T + o(\|y^{\otimes 2}\|)
\end{aligned}
$$

This is (5.24). □

PROOF OF THE FACT THAT (5.22) IMPLIES (5.19)

In the proof of Theorem 3.1 as presented in the main article it is claimed that (5.19) follows from (5.22). Here we show in some details why this in fact is the case.

Write a two-dimensional standard Brownian bridge $B(x), x \in [0,1]^2$ as

$$B(x) = W(x) - x_1 x_2 W(1,1),$$

where W is a two-dimensional Wiener process. Let $X = (X_1, X_2)^T$ be a random vector in $\mathbb{R}^2$. Let $\mathfrak{T} : \mathbb{R}^2 \mapsto [0,1]^2$ denote the Rosenblatt transformation (Rosenblatt 1952) defined as

$$\mathfrak{T}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} P(X_1 \le x_1) \\ P(X_2 \le x_2 | X_1 = x_1) \end{pmatrix}.$$

Define

$$
\begin{aligned}
{}_0Y_n(x) &= h^{-1}\left\langle \frac{a(x)A(x)}{\sqrt{f(x)}}, \int_{\mathbb{R}^2} d^2K\Big(\frac{x-s}{h}\Big)dB(\mathfrak{T}s)\right\rangle, \\
{}_1Y_n(x) &= h^{-1}\left\langle \frac{a(x)A(x)}{\sqrt{f(x)}}, \int_{\mathbb{R}^2} d^2K\Big(\frac{x-s}{h}\Big)dW(\mathfrak{T}s)\right\rangle, \\
{}_2Y_n(x) &= h^{-1}\left\langle \frac{a(x)A(x)}{\sqrt{f(x)}}, \int_{\mathbb{R}^2} d^2K\Big(\frac{x-s}{h}\Big)\sqrt{f(s)}dW(s)\right\rangle, \\
{}_3Y_n(x) &= h^{-1}\left\langle a(x)A(x), \int_{\mathbb{R}^2} d^2K\Big(\frac{x-s}{h}\Big)dW(s)\right\rangle.
\end{aligned}
$$

Following the arguments in the proof of Theorem 1 in Rosenblatt (1976), we similarly have

$$\sup_{x\in\mathcal{L}} |Y_n(x) - {}_0Y_n(x)| = O_p(h^{-1}n^{-1/6}(\log n)^{3/2}),$$
$$\sup_{x\in\mathcal{L}} |{}_0Y_n(x) - {}_1Y_n(x)| = O_p(h),$$
$$\sup_{x\in\mathcal{L}} |{}_2Y_n(x) - {}_3Y_n(x)| = O_p(h).$$

The two Gaussian fields ${}_1Y_n(x)$ and ${}_2Y_n(x)$ have the same probability structure. Note that the Gaussian fields ${}_3Y_n(hx)$ and $U_h(x)$ have the same probability structure on $\mathcal{H}_h$. Hence in order to prove (5.19) it suffices to prove (5.22). □

APPENDIX B: MISCELLANEOUS RESULTS

We will also need to estimate the derivative of $V(x)$ given by

$$\nabla V(x) = \nabla G(d^2 f(x))\nabla d^2 f(x), \quad x \in \mathcal{H}^\epsilon.$$

The corresponding plug-in kernel estimators of $\nabla V(x)$ then is

$$\nabla \hat{V}(x) = \nabla G(d^2 \hat{f}(x))\nabla d^2 \hat{f}(x), \quad x \in \mathcal{H}^\epsilon.$$

Lemma B.1 *Under assumptions (**F**1),(**K**1)–(**K**2) and (**H**1), we have for* $\epsilon > 0$ *that*

$$\sup_{x\in\mathcal{H}^\epsilon} \|\nabla \hat{f}(x) - \mathbb{E}[\nabla \hat{f}(x)]\| = O_p\Big(\sqrt{\frac{\log n}{nh^4}}\Big)$$
$$\sup_{x\in\mathcal{H}^\epsilon} \|\nabla^2 \hat{f}(x) - \nabla^2 f(x)\|_F = O_p\Big(\sqrt{\frac{\log n}{nh^6}}\Big),$$
$$\sup_{x\in\mathcal{H}^\epsilon} \|\nabla d^2 \hat{f}(x) - \nabla d^2 f(x)\|_F = O_p\Big(\sqrt{\frac{\log n}{nh^8}}\Big).$$

The same rate holds for $\nabla^2 f(x)$ replaced by $\mathbb{E}\nabla^2 \hat{f}(x)$.

Lemma B.2 *Under assumptions (**F**1)–(**F**2), (**K**1)–(**K**2) and (**H**1), we have for* $\epsilon > 0$ *small enough that*

$$\sup_{x\in\mathcal{H}^\epsilon} \|\hat{V}(x) - V(x)\| = O_p\Big(\sqrt{\frac{\log n}{nh^6}}\Big),$$
$$\sup_{x\in\mathcal{H}^\epsilon} \|\nabla^2 \hat{f}(x)\hat{V}(x) - \nabla^2 f(x)V(x)\| = O_p\Big(\sqrt{\frac{\log n}{nh^6}}\Big),$$
$$\sup_{x\in\mathcal{H}^\epsilon} \|\nabla G(d^2 \hat{f}(x)) - \nabla G(d^2 f(x))\|_F = O_p\Big(\sqrt{\frac{\log n}{nh^6}}\Big),$$

$$\sup_{x\in\mathcal{H}^\epsilon} \|\nabla\hat{V}(x) - \nabla V(x)\|_F = O_p\Big(\sqrt{\frac{\log n}{nh^8}}\Big),$$

$$\sup_{x\in\mathcal{H}^\epsilon} \|\nabla\hat{V}(x)\hat{V}(x) - \nabla V(x)V(x)\| = O_p\Big(\sqrt{\frac{\log n}{nh^8}}\Big)$$

The following result shows the uniform consistency of the estimator $\hat{\mathfrak{X}}_{x_0}(t)$.

Lemma B.3 *For $x_0 \in \mathcal{G}$ let $T_{x_0}^{min}, T_{x_0}^{max} \geq 0$ be such that $T_{x_0}^{min} + T_{x_0}^{max} > 0$ and $T_{\mathcal{G}} := \max\{\sup_{x_0\in\mathcal{G}} T_{x_0}^{min}, \sup_{x_0\in\mathcal{G}} T_{x_0}^{max}\} < \infty$ and $\{\mathfrak{X}_{x_0}(t), t \in [-T_{x_0}^{min}, T_{x_0}^{max}]\} \subset \mathcal{H}$. Under assumptions (**F**1)–(**F**2), (**K**1)–(**K**2) and (**H**1), we have that*

$$\sup_{x_0\in\mathcal{G}, t\in[-T_{x_0}^{min}, T_{x_0}^{max}]} \|\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\| = o_p(1).$$

PROOF. Following the proof on page 1584 of Koltchinskii et al. (2007), we obtain that for all $x_0 \in \mathcal{G}$ and $t \in [0, T_{x_0}^{max}]$ and for some constant $L > 0$

$$\|\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\| \leq T_{x_0}^{max} \sup_{x\in\mathbb{R}^2} \|\hat{V}(x) - V(x)\| e^{Lt}.$$

Therefore by Lemma B.1, and the fact that G is Lipschitz continuous (recall the definitions $\hat{V}(x) = G(d^2\hat{f}(x))$ and $V(x) = G(d^2f(x))$)

$$\sup_{x_0\in\mathcal{G}, t\in[0,T_{x_0}^{max}]} \|\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\| \leq T_{\mathcal{G}} \sup_{x\in\mathbb{R}^2} \|\hat{V}(x) - V(x)\| e^{LT_{\mathcal{G}}} = o_p(1).$$

Similarly we can prove $\sup_{x_0\in\mathcal{G}, t\in[-T_{x_0}^{min}, 0]} \|\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\| = o_p(1)$ and therefore the lemma is proved. □

The function G and some of its properties. In what follows we show that with

$$G(u, v, w) = \begin{pmatrix} 2u - 2w + 2v - 2\sqrt{(w-u)^2 + 4v^2} \\ w - u + 4v - \sqrt{(w-u)^2 + 4v^2} \end{pmatrix} \tag{B.1}$$

we have that

$$V(x) = G(d^2f(x)) \quad \text{is a second eigenvector of the Hessian } \nabla^2 f(x). \tag{B.2}$$

We also write $G = (G_1, G_2)^T$. Let $\lambda_2(x)$ be the second eigenvalue of $\nabla^2 f(x)$. Then $\lambda_2(x)$ is the smaller root of equation

$$\begin{vmatrix} f^{(2,0)}(x) - \lambda_2(x) & f^{(1,1)}(x) \\ f^{(1,1)}(x) & f^{(0,2)}(x) - \lambda_2(x) \end{vmatrix} = 0.$$

Calculation shows that the second eigenvalue of the Hessian is

$$\lambda_2(x) = J(d^2f(x))$$

where

$$J(u,v,w) = \frac{u + w - \sqrt{(u-w)^2 + 4v^2}}{2}. \tag{B.3}$$

We denote $V(x) = (u,v)^T$. Since $V(x)$ is the second eigenvector of the Hessian, we have $(\nabla^2 f(x) - \lambda_2(x)\mathbf{I})V(x) = 0$, i.e.,

$$\begin{pmatrix} f^{(2,0)}(x) - \lambda_2(x) & f^{(1,1)}(x) \\ f^{(1,1)}(x) & f^{(0,2)}(x) - \lambda_2(x) \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = 0. \tag{B.4}$$

Note that the two equations above are linearly dependent. Also notice that both $V_1(x) := \Big(\lambda_2(x) - f^{(0,2)}(x),\ f^{(1,1)}(x)\Big)^T$ and $V_2(x) := \Big(f^{(1,1)}(x),\ \lambda_2(x) - f^{(2,0)}(x)\Big)^T$ are solutions to one of (and therefore both of) the equations in (B.4). For any $c_1(x),\ c_2(x) \in \mathbb{R}$, if

$$V(x) = c_1(x)V_1(x) + c_2(x)V_2(x),$$

then $V(x)$ satisfies the equations in (B.4). We want to find $c_1(x)$ and $c_2(x)$ such that $V(x) \neq 0$ if the two eigenvalues of $\nabla^2 f(x)$ are not equal, i.e. $f^{(2,0)}(x) \neq f^{(0,2)}(x)$ and $f^{(1,1)}(x) \neq 0$. For this purpose we choose $c_1(x) \equiv 4$ and $c_2(x) \equiv 2$. As a result, $V(x) = G(d^2 f(x))$ with G defined in (B.1), which is (B.2)

There are other ways of choosing $c_1(x)$ and $c_2(x)$ but all that matters for our results is that $V(x)$ is smooth and that $\|V(x)\|$ is bounded away from zero (and infinity) as long as the two eigenvalues of Hessian $\nabla^2 f(x)$ are distinct.

Lipschitz continuity of G and ∇G. First observe that by assumption (**F**2), there exists a $\delta > 0$ such that $\{d^2 f(x) : x \in \mathcal{H}\} \subset \mathcal{Q}_\delta$ where

$$\mathcal{Q}_\delta = \{(u,v,w) \in \mathbb{R}^3 : |u - w| > \delta \ \text{ or } \ |v| > \delta\}, \tag{B.5}$$

since two eigenvalues of a 2×2 symmetric matrix are equal iff the matrix is a scaled identity matrix (see also the discussion given after the assumptions in the paper).

To verify Lipschitz continuity of $G(u,v,w)$ on $\mathbb{R}^3$, it suffices to notice that

$$\begin{aligned} &\left|\sqrt{(u_1 - w_1)^2 + 4v_1^2} - \sqrt{(u_2 - w_2)^2 + 4v_2^2}\right| \\ &= \frac{|(u_1 - u_2 - w_1 + w_2)(u_1 - w_1 + u_2 - w_2) + 4(v_1 - v_2)(v_1 + v_2)|}{\sqrt{(u_1 - w_1)^2 + 4v_1^2} + \sqrt{(u_2 - w_2)^2 + 4v_2^2}} \\ &\leq |u_1 - u_2| + |w_1 - w_2| + 2|v_1 - v_2|. \end{aligned}$$

As for Lipschitz continuity of $\nabla G(u,v,w)$ on $\mathcal{Q}_\delta$, it suffices to notice that the components of $\nabla^2 G_i(u,v,w)$ are all bounded on $\mathcal{Q}_\delta$, where $((u-w)^2 + 4v^2)^{3/2}$ in the denominator is bounded away from zero on $\mathcal{Q}_\delta$.