Local Polynomial Regression for Symmetric Positive Definite Matrices (Supplementary Report)

Ying Yuan, Hongtu Zhu, Weili Lin, J. S. Marron

University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.

1. Cross validation bandwidth selection

In this section, we will derive the first-order approximation to the cross validation score $\mathrm{CV}_T(h)$ in Equation (27). We need some notation as follows:

$$
\begin{aligned}
&M(S,D) = \mathrm{tr}\{\log(S^{-1/2}DS^{-T/2})^2\},\\
&G_{n,T}^{(-i)}(\alpha_T(x)) = \sum_{j\neq i} g_T(S_j, D_T(x_j,\alpha_T(x),k_0))^2,\\
&G_{n,T}(\alpha_T(x)) = \sum g_T(S_j, D_T(x_j,\alpha_T(x),k_0))^2,\\
&\hat\alpha_{IT}^{(-i)}(x,h) = \mathrm{argmin}_{\alpha_T(x)} G_{n,T}^{(-i)}(\alpha_T(x)),\\
&\hat\alpha_{IT}(x,h) = \mathrm{argmin}_{\alpha_T(x)} G_{n,T}(\alpha_T(x)), \quad \Delta\hat\alpha_{IT}^{(-i)}(x,h) = \hat\alpha_{IT}^{(-i)}(x,h) - \hat\alpha_{IT}(x,h),\\
&\partial_D M(S,D) = \frac{\partial \mathrm{tr}\{\log(S^{-1/2}DS^{-T/2})^2\}}{\partial \mathrm{vecs}(D)},\\
&\partial_G M(S,GG^T) = \frac{\partial \mathrm{tr}\{\log(S^{-1/2}GG^TS^{-T/2})^2\}}{\partial \mathrm{vecs}(G)},\\
&\partial_D^2 M(S,D) = \frac{\partial^2 \mathrm{tr}\{\log(S^{-1/2}DS^{-T/2})^2\}}{\partial \mathrm{vecs}(D)\partial \mathrm{vecs}(D)^T},\\
&E(S,D,x,x_0) = D_T(x,\alpha_T(x_0),k_0)^{-1/2} S D_T(x,\alpha_T(x_0),k_0)^{-T/2},\\
&M(S,x-x_0,\alpha_T(x_0)) = \mathrm{tr}\{\log(S^{-1/2}D(x,\alpha_T(x_0),k_0)S^{-T/2})^2\},\\
&\partial_{\alpha_T(x_0)} M(S,x-x_0,\alpha_T(x_0)) = \frac{\partial \mathrm{tr}\{\log(S^{-1/2}D(x,\alpha_T(x_0),k_0)S^{-T/2})^2\}}{\partial \alpha_T(x_0)},\\
&\partial_{\alpha_T(x_0)}^2 M(S,x-x_0,\alpha_T(x_0)) = \frac{\partial^2 \mathrm{tr}\{\log(S^{-1/2}D(x,\alpha_T(x_0),k_0)S^{-T/2})^2\}}{\partial \alpha_T(x_0)\partial \alpha_T(x_0)^T},\\
&H^{(-i)}(x,\alpha_T(x),h) = \sum_{j\neq i} K_h(x_j-x)\partial_{\alpha_T(x)}^2 M(S_j,x_j-x,\alpha_T(x)).
\end{aligned}
$$

Let $\hat G_{IT}^{(-i)}(x_i,h)$ and $\hat G_{IT}(x_i,h)$ be, respectively, the subcomponents of $\hat\alpha_{IT}^{(-i)}(x,h)$ and $\hat\alpha_{IT}(x,h)$ corresponding to $G(x)$. Then, by using the Taylor's series expansion, we can approximate the cross validation at bandwidth h by

$$\begin{aligned}\mathrm{CV}_T(h) &= n^{-1}\sum_{i=1}^n g_T(S_i, \hat{D}_{IT}^{(-i)}(x_i,h))^2 = n^{-1}\sum_{i=1}^n \mathrm{tr}\{\log(S_i^{-1/2}\hat{D}_{IT}^{(-i)}(x_i,h)S_i^{-T/2})^2\} \\ &\approx n^{-1}\sum_{i=1}^n g_T(S_i,\hat{D}_{IT}(x_i,h))^2 + 2p_n(h),\end{aligned}$$

where $p_n(h)$ can be regarded as the degree of freedom for ILPR and is given by

$$p_n(h) = (2n)^{-1}\sum_{i=1}^n \{\partial_G M(S_i,\hat{D}_{IT}(x_i,h))^T \mathrm{vecs}(\hat{G}_{IT}^{(-i)}(x_i,h) - \hat{G}_{IT}(x_i,h))\}. \tag{1}$$

Since $\hat{\alpha}_{IT}^{(-i)}(x_i,h)$ and $\hat{\alpha}_{IT}(x_i,h)$ minimize $G_{n,T}^{(-i)}(\alpha_T(x_i))$ and $G_{n,T}(\alpha_T(x_i))$, respectively, we have

$$\begin{aligned} 0 &= \sum_{j\neq i}\{K_h(x_j-x_i)\partial_{\alpha(x_i)}M(S_j,x_j-x_i,\hat{\alpha}_{IT}^{(-i)}(x_i,h))\} \qquad (2)\\ &\approx \sum_{j\neq i}\{K_h(x_j-x_i)\partial_{\alpha(x_i)}M(S_j,x_j-x_i,\hat{\alpha}_{IT}(x_i,h))\} \\ &\quad + \sum_{j\neq i}K_h(x_j-x_i)\partial^2_{\alpha(x_i)}M(S_j,x_j-x_i,\hat{\alpha}_{IT}(x_i,h))\Delta\hat{\alpha}_{IT}^{(-i)}(x_i,h) \\ &= -K_h(0)\partial_{\alpha(x_i)}M(S_i,0,\hat{\alpha}_{IT}(x_i,h)) \\ &\quad + \sum_{j\neq i}K_h(x_j-x_i)\partial^2_{\alpha(x_i)}M(S_j,x_j-x_i,\hat{\alpha}_{IT}(x_i,h))\Delta\hat{\alpha}_{IT}^{(-i)}(x_i,h).\end{aligned}$$

This yields that

$$\Delta\hat{\alpha}_{IT}^{(-i)}(x_i,h) = K_h(0)H^{(-i)}(x_i,\hat{\alpha}_{IT}(x_i,h),h)^{-1}\partial_{\alpha(x_i)}M(S_i,0,\hat{\alpha}_{IT}(x_i,h)). \tag{3}$$

Furthermore, at a given x_i, we consider $\alpha_T(x_i) = (\alpha_{(1)}(x_i)^T, \alpha_{(2)}(x_i)^T)^T$ and $\alpha_{IT}^{(-i)}(x_i,h) = (\alpha_{(1),IT}^{(-i)}(x_i,h)^T, \alpha_{(2),IT}^{(-i)}(x_i,h)^T)^T$, in which $\alpha_{(1)}(x_i)$ and $\alpha_{(1),IT}^{(-i)}(x_i,h)$, respectively, correspond to the unknown parameters in $G(x_i)$. Suppose that $H^{(-i)}(x_i,\hat{\alpha}_{IT}(x_i,h),h)$ can be decomposed according to the decomposition $\alpha(x_i) = (\alpha_{(1)}(x_i)^T, \alpha_{(2)}(x_i)^T)^T$ as follows:

$$H^{(-i)}(x_i,\hat{\alpha}_{IT}(x_i,h),h) = \begin{pmatrix} H_{11,IT}^{(-i)}(x_i,h) & H_{12,IT}^{(-i)}(x_i,h) \\ H_{21,IT}^{(-i)}(x_i,h) & H_{22,IT}^{(-i)}(x_i,h)\end{pmatrix}.$$

It can be shown that all elements in $\partial_{\alpha_{(2)}(x_i)}M(S_i,0,\hat{\alpha}_{IT}(x_i,h))$ equal zero. Thus, by using the nullity theorem, we have

$$\Delta\hat{\alpha}_{(1),IT}^{(-i)}(x_i,h) = K_h(0)H_{11\cdot 2,IT}^{(-i)}(x_i,h)^{-1}\partial_{\alpha_{(1)}(x_i)}M(S_i,0,\hat{\alpha}_{IT}(x_i,h)), \tag{4}$$

where $H^{(-i)}_{11\cdot 2,IT}(x_i,h) = H^{(-i)}_{11,IT}(x_i,h) - H^{(-i)}_{12,IT}(x_i,h)H^{(-i)}_{22,IT}(x_i,h)^{-1}H^{(-i)}_{21,IT}(x_i,h)$. By substituting (4) into (1), we have

$$p_n(h) = (2n)^{-1}K_h(0)\sum_{i=1}^{n}\{\partial_G M(S_i,\hat{D}_{IT}(x_i,h))^T H^{(-i)}_{11\cdot 2,IT}(x_i,h)^{-1}\partial_G M(S_i,\hat{D}_{IT}(x_i,h))\}, \quad (5)$$

which finishes the proof of Equation (27).

In order to calculate $\text{CV}_T(h)$, we need the first order derivatives of matrix logarithm and exponential, $\partial_G M(S,GG^T)$, and the first-order and second-order derivatives of $M(S,x-x_0,\alpha_T(x_0))$ with respect to $\alpha_T(x_0)$ as follows.

Lemma 1. *(i) The first-order derivative of $M(S,GG^T)$ with respect to vecs(G) is given by*

$$\frac{\partial M(S,GG^T)}{\partial G_j} = 2tr\{\log(G^T S^{-1}G)G^{-1}\frac{\partial(GG^T)}{\partial G_j}G^{-T}\}, \quad (6)$$

where G_j is the $j-th$ unknown element in vecs(G).

(ii) Suppose that $D(\alpha)\in Sym^+(m)$ and $M(\alpha)\in Sym(m)$ are differentiable functions of α, the first order derivatives of $\log(D(\alpha))$ and $\exp(M(\alpha))$ with respect to the $j-th$ component of α_j are, respectively, given by

$$\frac{\partial\exp(M(\alpha))}{\partial\alpha_j} = \int_0^1 \exp((1-s)M(\alpha))\frac{\partial M(\alpha)}{\partial\alpha_j}\exp(sM(\alpha))ds, \quad (7)$$

$$\frac{\partial\log(D(\alpha))}{\partial\alpha_j} = \int_0^1 [\{D(\alpha)-I_m\}s+I_m]^{-1}\frac{\partial D(\alpha)}{\partial\alpha_j}[\{D(\alpha)-I_m\}s+I_m]^{-1}ds. \quad (8)$$

Proof of Lemma 1. Since $\text{tr}\{\log(S^{-1/2}GG^TS^{-T/2})^2\} = \text{tr}\{\log(G^{-1}SG^{-T})^2\}$, It follows from Proposition 2.1 in Maher (2005) that

$$\frac{\partial M(S,GG^T)}{\partial G_j} = 2\text{tr}\{\log(G^{-1}SG^{-T})G^TS^{-1}G\partial_{G_j}(G^{-1}SG^{-T})\}.$$

Because $\log(G^{-1}SG^{-T})G^TS^{-1}G = G^TS^{-1}G\log(G^{-1}SG^{-T})$ and

$$\partial_{G_j}(G^{-1}SG^{-T}) = -G^{-1}(\partial_{G_j}G)G^{-1}SG^{-T} - G^{-1}SG^{-T}(\partial_{G_j}G)G^{-T},$$

we have

$$\frac{\partial M(S,GG^T)}{\partial G_j} = -2\sum_{i=1}^{n}\text{tr}\{\log(G^{-1}SG^{-T})(G^{-1}\frac{\partial G}{\partial G_j} + \frac{\partial G^T}{\partial G_j}G^{-T})\},$$

which yields (6). The proof of (7) and (8) can be found in Higham (2008).

Lemma 2. *Let $\alpha_j(x_0)$ be the jth element of $\alpha_T(x_0)$. The jth element of the vector $\partial_{\alpha_T(x_0)}M(S,x-x_0,\alpha_T(x_0))$ is given by*

$$\partial_{\alpha_j(x_0)}M(S,x-x_0,\alpha_T(x_0)) = -2tr[\log(E(S,D,x,x_0))D(x,\alpha_T(x_0),k_0)^{-1/2}\times \{\partial_{\alpha_j(x_0)}D(x,\alpha_T(x_0),k_0)\}D(x,\alpha_T(x_0),k_0)^{-T/2}].$$

The $(j,k)-$*th element of* $\partial^2_{\alpha_T(x_0)} M(S, x - x_0, \alpha_T(x_0))$ *is given by*

$$\frac{\partial^2 M(S, x - x_0, \alpha_T(x_0))}{\partial \alpha_j(x_0) \partial \alpha_k(x_0)} = -2tr[\frac{\partial \log(E(S, D, x, x_0))}{\partial \alpha_k(x_0)} D(x, \alpha_T(x_0), k_0)^{-1/2} \times$$
$$\{\partial_{\alpha_j(x_0)} D(x, \alpha_T(x_0), k_0)\} D(x, \alpha_T(x_0), k_0)^{-T/2}] - 2tr(\log(E(S, D, x, x_0))$$
$$\frac{\partial [D(x, \alpha_T(x_0), k_0)^{-1/2} \{\partial_{\alpha_j(x_0)} D(x, \alpha_T(x_0), k_0)\} D(x, \alpha_T(x_0), k_0)^{-T/2}]}{\partial \alpha_k(x_0)}).$$

Proof of Lemma 2. By using Lemma 1 and matrix differentiation, we can easily prove Lemma 2.

2. Annealing evolutionary stochastic approximation Monte Carlo

We now develop an annealing evolutionary stochastic approximation Monte Carlo algorithm for computing $\hat{\alpha}_{IT}(x_0; h)$. Quite recently, the stochastic approximation Monte Carlo algorithm Liang et al. (2007) has been proposed in the literature as a general simulation technique, which possesses a nice feature in that the moves are self-adjustable and thus not likely to get trapped by local energy minima. The annealing evolutionary SAMC algorithm (Liang, 2010) represents a further improvement of stochastic approximation Monte Carlo for optimization problems by incorporating some features of simulated annealing (Kirkpatrick et al., 1983) and the genetic algorithm (Goldberg, 1989) into its search process.

Like the genetic algorithm, annealing evolutionary stochastic approximation Monte Carlo works on a population of samples. Let $\alpha^l = (\alpha_{R,[1]}, \ldots, \alpha_{R,[l]})$ denote the population, where l is the population size, and $\alpha_{R,[i]} = (\alpha_{i1}, \ldots, \alpha_{iq(k_0+1)})$ is a $q(k_0 + 1)$-dimensional vector called an individual or chromosome in terms of genetic algorithms. Thus, the minimum of the objective function $G_n(\alpha_T(x_0))$, $\alpha_T(x_0) \in \mathcal{B}$, can be obtained by minimizing the function $U(\alpha^l) = \sum_{i=1}^{l} G_n(\alpha_{R,[i]})$. An unnormalized Boltzmann density can be defined for the population as follows,

$$\psi(\alpha^l) = \exp\left\{-U(\alpha^l)/\tau\right\}, \quad \alpha^l \in \mathcal{B}^l, \tag{9}$$

where $\tau = 1$ is called the temperature, and $\mathcal{B}^l = \mathcal{B} \times \cdots \times \mathcal{B}$ is a product sample space. The sample space can be partitioned according to the function $U(\alpha^l)$ into b subregions: $\mathbb{E}_1 = \{\alpha^l : U(\alpha^l) \le \delta_1\}$, $\mathbb{E}_2 = \{\alpha^l : \delta_1 < U(\alpha^l) \le \delta_2\}$, $\cdots$, $\mathbb{E}_{b-1} = \{\alpha^l : \delta_{b-2} < U(\alpha^l) \le \delta_{b-1}\}$, and $\mathbb{E}_b = \{\alpha^l : U(\alpha^l) > \delta_{b-1}\}$, where $\delta_1 < \delta_2 < \ldots < \delta_{b-1}$ are $b - 1$ known real numbers. We note that here the sample space is not necessarily partitioned according to the function $U(\alpha^l)$, for example, the function $\lambda(\alpha^l) = \min\{G_n(\alpha_{R,[1]}), \ldots, G_n(\alpha_{R,[l]})\}$ also works.

Let $\varpi(\delta)$ denote the index of the subregion that a sample with energy u belongs to. For example, if $\alpha^l \in E_j$, then $\varpi(U(\alpha^l)) = j$. Let $\mathcal{B}^{(t)}$ denote the sample space at iteration t. The algorithm initiates its search in the entire sample space $\mathcal{B}_0 = \bigcup_{i=1}^{b} E_i$, and then iteratively searches in the set

$$\mathcal{B}_t = \bigcup_{i=1}^{\varpi(U_{\min}^{(t)}+\aleph)} E_i, \quad t = 1, 2, \ldots, \tag{10}$$

where $U_{\min}^{(t)}$ is the best function value obtained until iteration t, and $\aleph > 0$ is a user specified parameter which determines the broadness of the sample space at each iteration. Note that in this method, the sample space shrinks iteration by iteration. To ensure the convergence of the algorithm to the set of global minima, the moves at each iteration are required to admit the following distribution as the invariant distribution,

$$f_{w^{(t)}}(\alpha^l) \propto \sum_{i=1}^{\varpi(U_{\min}^{(t)}+\aleph)} \frac{\psi(\alpha^l)}{e^{w_i^{(t)}}} I(\alpha^l \in \mathbb{E}_i), \quad x \in \mathcal{B}_t^l, \tag{11}$$

where $w_i^{(t)}$ are the working parameters which will be updated from iteration to iteration as described in the algorithm below.

The annealing evolutionary stochastic approximation Monte Carlo includes five types of moves, the MH-Gibbs mutation, K-point mutation, K-point crossover, snooker crossover, and linear crossover operators. See Liang (2010) for the details of the moves. Let $\rho_1, \ldots, \rho_5$, $0 < \rho_i < 1$ and $\sum_{i=1}^{5} \rho_i = 1$, denote the respective working probabilities of the five types of moves. The algorithm can be summarized as follows.

The algorithm:

(a) (Initialization) Partition the sample space $\mathcal{B}^l$ into b disjoint subregions $\mathbb{E}_1, \ldots, \mathbb{E}_b$; choose the threshold value $\aleph$ and the working probabilities $\rho_1, \ldots, \rho_5$; initialize a population $\alpha^{l(0)}$ at random; and set $w^{(0)} = (w_1^{(0)}, \ldots, w_b^{(0)}) = (0, 0, \ldots, 0)$, $\mathcal{B}_0^l = \bigcup_{i=1}^{b} \mathbb{E}_i$, $U_{\min}^{(0)} = U(\alpha^{l(0)})$ and $t = 0$. Let Θ be a compact set in R^m.

(b) (Sampling) Update the current population $\alpha^{l(t)}$ using the MH-Gibbs mutation, K-point mutation, K-point crossover, snooker crossover, and linear crossover operators according to the respective working probabilities.

(c) (Working weight updating) Update the working weight $w^{(t)}$ by setting

$$w_i^* = w_i^{(t)} + \gamma_{t+1} H_i(w^{(t)}, \alpha^{l(t+1)}), \quad i = 1, \ldots, \varpi(U_{\min}^{(t)} + \aleph),$$

where $H_i(w^{(t)}, \alpha^{l(t+1)}) = I(\alpha^{l(t+1)} \in \mathbb{E}_i)$ for the crossover operators, $H_i(w^{(t)}, \alpha^{l(t+1)}) = \sum_{j=1}^{l} I(\alpha^{l(t+1,j)} \in \mathbb{E}_i)/l$ for the mutation operators, and γ_{t+1} is called the gain factor. If

$w^* \in \Theta$, set $w^{(t+1)} = w^*$; otherwise, set $w^{(t+1)} = w^* + c^*$, where $c^* = (c^*, \ldots, c^*)$ and c^* is chosen such that $w^* + c^* \in \Theta$.

(d) (Termination Checking) Check the termination condition, e.g., whether a fixed number of iterations has been reached. Otherwise, set $t \to t+1$ and go to step (b).

In this article, we follow Liang (2010) to set $\rho_1 = \rho_2 = 0.05$, $\rho_3 = \rho_4 = \rho_5 = 0.3$, and the gain factor sequence

$$\gamma_t = \frac{t_0}{\max(t_0, t)}, \quad t = 0, 1, 2, \ldots, \tag{12}$$

with $t_0 = 500000$. In general, a large value of t_0 will allow the sampler to reach all the subregions very quickly even for a large system. As shown in Liang (2010), it can converge weakly toward a neighboring set of global minima of $U(\alpha^l)$ in the space of energy. More precisely, the sample $\alpha^{l(t)}$ converges in distribution to a random population with the density function

$$f_w(\alpha^l) \propto \sum_{i=1}^{\varpi(U_{\min}+\aleph)} \frac{\psi(\alpha^l)}{\int_{\mathbb{E}_i} \psi(\alpha^l) d\alpha^l} I(x \in \mathbb{E}_i), \tag{13}$$

where $U_{\min}$ is the global minimum value of $U(\alpha)$,

Regarding the setting of other parameters, we have the following suggestions. In the algorithm, the moves are reduced to the Metropolis-Hastings moves (Metropolis et al., 1953; Hastings, 1970) within the same subregions. Hence, the sample space should be partitioned such that the MH moves within the same subregion have a reasonable acceptance rate. In this article, we set $\delta_{i+1} - \delta_i \equiv 0.2$ for $i = 1, \ldots, b-1$.

The crossover operator has been modified to serve as a proposal for the moves, and it is no longer as critical as to the genetic algorithm. Hence, the population size l is usually set to a moderate number, ranging from 10 to 100. Since $\aleph$ determines the size of the neighboring set toward which the method converges, $\aleph$ should be chosen carefully for efficiency of the algorithm. If $\aleph$ is too small, it may take a long time for the algorithm to locate the global minima. In this case, the sample space may contain a lot of separated regions, and most of the proposed transitions will be rejected if the proposal distribution is not spread out enough. If $\aleph$ is too large, it may also take a long time for the algorithm to locate the global energy minimum due to the broadness of the sample space. In practice, the values of l and $\aleph$ can be determined through a trial and error process based on the diagnosis for the convergence of the algorithm. If it fails to converge, the parameters should be tuned to larger values. As suggested by Liang (2010), the convergence of the method can be diagnosed by examining the difference of the patterns of the working weights obtained in multiple runs. In this article, we set $l = 50$ and $\aleph = 50$.

3. Proof of Theorem 2 and 3

First, we present the asymptotic properties of the intrinsic estimators under the Log-Euclidean metric at boundary points below.

Theorem 4. Suppose that $x_0 = dh$ *is a left boundary point of* $f_X(.)$. *Let* $d_{k_0,d} = (u_{k_0+1,d}, \cdots, u_{2k_0+1,d})^T$.

(i) Under conditions (C1)-(C4) in Appendix 2, we have $\mathcal{H}\{\hat{\alpha}_{IL}(x_0; h) - \alpha_L(x_0)\}$ *converges to* $\mathbf{0}$ *in probability as* $n \to \infty$.

(ii) For $k_0 = 0$, *under conditions (C1)-(C4) and (C10) in Appendix 2, conditioning on* $\mathbf{x} = \{x_1, \cdots, x_n\}$, *we have*

$$\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IL}(0+; h) - \alpha_L(0+)\} - hu_{0,d}^{-1}u_{1,d}vecs((\log(D(0+)))^{(1)})] \to^L N(\mathbf{0}, \Sigma_{0,d}(0+)), \quad (14)$$

where $\Sigma_{0,d}(0+) = f_X^{-1}(0+)u_{0,d}^{-2}v_{0,d}\Sigma_{\mathcal{E}_D}(0+)$.

(iii) For $k_0 > 0$, *under the conditions of Theorem 4 (ii), conditioning on* $\mathbf{x} = \{x_1, \cdots, x_n\}$, *we have*

$$\sqrt{nh}[\mathcal{H}\{\hat{\alpha}_{IL}(0+; h) - \alpha_L(0+)\} - \frac{h^{k_0+1}}{(k_0+1)!}(\mathcal{U}_{0,d}^{-1} \otimes I_q)(d_{k_0,d} \otimes vecs((\log(D(0+)))^{(k_0+1)}))] \to^L N(\mathbf{0}, \Sigma_d(0+)), \quad (15)$$

where $\Sigma_d(0+) = f_X^{-1}(0+)(\mathcal{U}_{0,d}^{-1}\mathcal{V}_{0,d}\mathcal{U}_{0,d}^{-1}) \otimes \Sigma_{\mathcal{E}_D}(0+)$.

Since the proofs of Theorems 1 and 4 can be easily followed from the same lines of arguments as those of Theorems 2 and 3, we omit them for simplicity. Since the proof of Theorem 3 involves more details, we only prove Theorem 3 and the proof of Theorem 2 follows the same lines of arguments. In addition, for notational simplicity, we only prove these theorems for the local linear regression for the $k_0 > 0$ case. We also prove theorems for the local constant regression because the proof requires some different treatments. The following lemmas are needed for our technical proofs.

Lemma 3. *Let* $\psi_{GG}(S, G, Y)$, $\psi_{GY}(S, G, Y)$, $\psi_{YG}(S, G, Y)$ *and* $\psi_{YY}(S, G, Y)$ *be the second order derivatives of* $\psi(S, G, Y)$ *with respect to* α_G *and* α_Y. *Let* $R(X) = Y(X) - Y^{(1)}(x_0)(X - x_0)$ *and assume that conditions (C1)-(C5), (C7) and (C8) hold. For any random* $m \times m$ *lower triangle matrix sequence* η_{i0} *and any random symmetric matrix sequence* $\eta_{i1} \in Sym(m)$, *for*

$i = 1, \cdots, n$, *if* $\max_{1\le i\le n} \|\eta_{i0}\| = o_p(1)$ *and* $\max_{1\le i\le n} \|\eta_{i1}\| = o_p(1)$, *then we have the following results:*

$$\begin{aligned}
&\sum_{i=1}^{n} hK_h(x_i - x_0)\psi_{GG}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1}) \\
= \; & nhf_X(0+)u_{0,d}\Psi_1(0+)\{1 + o_p(1)\}, && (16) \\
&\sum_{i=1}^{n} hK_h(x_i - x_0)\psi_{GY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1})(x_i - x_0)^l \\
= \; & nh^{l+1}f_X(0+)u_{l,d}\Psi_2(0+)\{1 + o_p(1)\}, && (17) \\
&\sum_{i=1}^{n} hK_h(x_i - x_0)\psi_{YY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1})(x_i - x_0)^l \\
= \; & nh^{l+1}f_X(0+)u_{l,d}\Psi_3(0+)\{1 + o_p(1)\}, && (18) \\
&\sum_{i=1}^{n} hK(x_i - x_0)\psi_{GY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1})^T vecs(R(x_i))(x_i - x_0)^l \\
= \; & \frac{1}{2}nh^{l+3}f_X(0+)u_{l+2,d}\Psi_2(0+)^T vecs(Y^{(2)}(0+))\{1 + o_p(1)\}, && (19) \\
&\sum_{i=1}^{n} hK(x_i - x_0)\psi_{YY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1})^T vecs(R(x_i))(x_i - x_0)^l \\
= \; & \frac{1}{2}nh^{l+3}f_X(0+)u_{l+2,d}\Psi_3(0+)^T vecs(Y^{(2)}(0+))\{1 + o_p(1)\}. && (20)
\end{aligned}$$

Proof of Lemma 3. We only prove (17), while the remainings can be shown using the same arguments. It is easy to see that

$$\begin{aligned}
&\sum_{i=1}^{n} hK_h(x_i - x_0)\psi_{GY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1})(x_i - x_0)^l \\
= \; & \sum_{i=1}^{n} hK_h(x_i - x_0)\psi_{GY}(S_i, G(x_0), Y(x_i))(x_i - x_0)^l + \\
& \sum_{i=1}^{n} hK_h(x_i - x_0)\{\psi_{GY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1}) - \psi_{GY}(S_i, G(x_0), Y(x_i))\}(x_i - x_0)^l \\
= \; & T_{n1} + T_{n2}.
\end{aligned}$$

Let $Z_{j,k} = hK_h(X - x_0)(\psi_{GY})_{j,k}(X - x_0)^l$, where $(\psi_{GY})_{j,k}$ is the (j,k)-th element of the matrix ψ_{GY}. For the (j,k)-th element $(T_{n1})_{j,k}$ in the matrix T_{n1}, we have

$$(T_{n1})_{j,k} \;=\; nE(Z_{j,k}) + O_p\left(\sqrt{nE(Z_{j,k}^2)}\right). \qquad (21)$$

We calculate the first two moments of $Z_{j,k}$ below. Note that

$$\begin{aligned} E(Z_{j,k}) &= E\{hK_h(X-x_0)(\psi_{GY})_{j,k}(X-x_0)^l\} \\ &= \int_0^1 hK_h(y-x_0)(y-x_0)^l(\Psi_2(y))_{j,k}f_X(y)dy \\ &= h^{l+1}\int_{-\min\{d,1\}}^1 K(z)z^l(\Psi_2(zh+x_0))_{j,k}f_X(zh+x_0)dz, \end{aligned} \tag{22}$$

which can be approximated by $h^{l+1}\mathcal{A}$ with $\mathcal{A} = \int_{-\min\{d,1\}}^1 K(z)z^l(\Psi_2(0+))_{j,k}f_X(x_0)dz$. Specifically, we consider the difference given by

$$\mathcal{I}_{n,1} \equiv \left|\int_{-\min\{d,1\}}^1 K(z)z^l(\Psi_2(zh+x_0))_{j,k}f_X(zh+x_0)dz - \mathcal{A}\right|.$$

Applying the dominated convergence theorem together with the boundedness and continuity assumptions on $f_X(.)$ and $\Psi_2(.)$, we get $\lim_{n\to\infty}\mathcal{I}_{n,1} = 0$. Thus,

$$E(Z_{j,k}) = h^{l+1}f_X(0+)(\Psi_2(0+))_{j,k}u_{l,d}\{1+o(1)\}. \tag{23}$$

Since $f_X(x)$ and $E\{(\psi_{GY})^2_{j,k}|X=x\}$ are bounded, there exists a $d_3 > 0$ such that $|f_X(x)E\{(\psi_{GY})^2_{j,}$ $x\}| < d_3$ for all x. So we have

$$\begin{aligned} E(Z^2_{j,k}) &= E\{h^2K^2_h(X-x_0)(\psi_{GY})^2_{j,k}(X-x_0)^{2l}\} \\ &= \int_0^1 h^2K^2_h(y-x_0)(y-x_0)^{2l}f_X(y)E\{(\psi_{GY})^2_{j,k}|X=y\}dy \\ &= h^{2l+1}\int_{-\min\{d,1\}}^1 K^2(z)z^{2l}f_X(zh+x_0)E\{(\psi_{GY})^2_{j,k}|X=zh+x_0\}dz \\ &\le d_3h^{2l+1}\int_{-\min\{d,1\}}^1 K^2(z)z^{2l}dz. \end{aligned}$$

By the continuity of $K(.)$, we have $\int_{-\min\{d,1\}}^1 K^2(z)z^{2l}dz < d_4$ for a given $d_4 > 0$ and

$$E(Z^2_{j,k}) \le d_3d_4h^{2l+1}. \tag{24}$$

Combining with (21), (23) and (24), we have

$$\begin{aligned} (T_{n1})_{j,k} &= nh^{l+1}(f_X(x_0)\Psi_2(x_0)u_{l,d}\{1+o(1)\} + O_p(1/\sqrt{nh}))_{j,k} \\ &= nh^{l+1}f_X(0+)(\Psi_2(0+))_{j,k}u_{l,d}\{1+o_p(1)\}. \end{aligned}$$

That is, $T_{n1} = nh^{l+1}f_X(0+)u_{l,d}\Psi_2(0+)\{1+o_p(1)\}$.

To prove (17), it suffices to show that $T_{n2} = o_p(nh^{l+1})$. Let $\Delta_{n0} = \{\eta_{10},\cdots,\eta_{n0}\}$ and $\Delta_{n1} = \{\eta_{11},\cdots,\eta_{n1}\}$, where η_{i0} is lower triangle matrix and $\eta_{i1} \in \mathrm{Sym}(m)$ for $i = 1,\cdots,n$.

For any given $\delta > 0$, denote $D_\delta = \{\Delta_n = (\Delta_{n0}, \Delta_{n1}) : \|\eta_{i0}\|^2 + \|\eta_{i1}\|^2 \le \delta^2, \forall i \le n\}$. Define

$$\begin{aligned} V(\Delta_n) &= \frac{1}{nh^{l+1}} \sum_{i=1}^{n} hK_h(x_i - x_0)(x_i - x_0)^l \{\psi_{GY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1}) \\ &\quad - \psi_{GY}(S_i, G(x_0), Y(x_i))\}. \end{aligned}$$

Then, we have

$$\begin{aligned} \sup_{D_\delta} \|V(\Delta_n)\| &\le \frac{1}{nh^{l+1}} \sum_{i=1}^{n} hK_h(x_i - x_0)(x_i - x_0)^l \\ &\quad \sup_{D_\delta} \|\psi_{GY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1}) - \psi_{GY}(S_i, G(x_0), Y(x_i))\|. \end{aligned}$$

By using condition (C8), as $\delta \to 0$, we have

$$\epsilon_\delta = E\{\sup_{D_\delta} \|\psi_{GY}(S_i, G(x_0) + \eta_{i0}, Y(x_i) + \eta_{i1}) - \psi_{GY}(S_i, G(x_0), Y(x_i))\| | x_i = x\} = o(1).$$

Therefore, as $\delta \to 0$, we have

$$E\left\{\sup_{D_\delta} \|V(\Delta_n)\|\right\} \le \epsilon_\delta \frac{1}{nh^{l+1}} E\left\{\sum_{i=1}^{n} hK_h(x_i - x_0)(x_i - x_0)^l\right\} \to 0.$$

Since $\max_{1 \le i \le n} \|\eta_{i0}\| = o_p(1)$ and $\max_{1 \le i \le n} \|\eta_{i1}\| = o_p(1)$, we have $V(\hat{\Delta}_n) = o_p(1)$ for $\hat{\Delta}_n = (\hat{\Delta}_{n0}, \hat{\Delta}_{n1})$ with $\hat{\Delta}_{n0} = (\eta_{10}, \cdots, \eta_{n0})^T$ and $\hat{\Delta}_{n1} = (\eta_{11}, \cdots, \eta_{n1})^T$. Thus, $T_{n2} = nh^{l+1} V(\hat{\Delta}_n) = o_p(nh^{l+1})$, which finishes the proof of (17).

Lemma 4. *Let $\psi_G(S, G, Y)$ and $\psi_Y(S, G, Y)$ be the first order derivatives of $\psi(S, G, Y)$ with respect to α_G and α_Y, respectively. Assume that conditions (C1)-(C8) hold. Then we have*

$$\begin{aligned} & \sum_{i=1}^{n} hK_h(x_i - x_0)\psi_G(S_i, G(x_0), Y^{(1)}(x_0)(x_i - x_0)) \\ = & \frac{1}{2} nh^3 f_X(0+) u_{2,d} \Psi_2(0+)^T vecs(Y^{(2)}(0+))\{1 + o_p(1)\} + \\ & \sum_{i=1}^{n} hK_h(x_i - x_0)\psi_G(S_i, G(x_0), Y(x_i)), \end{aligned} \tag{25}$$

$$\begin{aligned} & \sum_{i=1}^{n} hK_h(x_i - x_0)\psi_Y(S_i, G(x_0), Y^{(1)}(x_0)(x_i - x_0))(x_i - x_0)^l \\ = & \frac{1}{2} nh^{l+3} f_X(0+) u_{l+2,d} \Psi_3(0+)^T vecs(Y^{(2)}(0+))\{1 + o_p(1)\} + \\ & \sum_{i=1}^{n} hK_h(x_i - x_0)\psi_Y(S_i, G(x_0), Y(x_i))(x_i - x_0)^l. \end{aligned} \tag{26}$$

Proof of Lemma 4. We just prove (25), while the second one can be similarly shown. We consider

$$\begin{aligned} J_n &\equiv \sum_{i=1}^n hK_h(x_i-x_0)\psi_G(S_i, G(x_0), Y(x_i)-R(x_i)) \\ &= \sum_{i=1}^n hK_h(x_i-x_0)[\psi_G(S_i, G(x_0), Y(x_i)) + \psi_{GY}(S_i, G(x_0), Y(x_i))^T \text{vecs}(-R(x_i)) \\ &\quad +\{\psi_G(S_i, G(x_0), Y(x_i)-R(x_i)) - \psi_G(S_i, G(x_0), Y(x_i)) \\ &\quad -\psi_{GY}(S_i, G(x_0), Y(x_i))^T \text{vecs}(-R(x_i))\}] \\ &= J_{n1}+J_{n2}+J_{n3}. \qquad (27) \end{aligned}$$

We need to consider J_{n1}, J_{n2}, and J_{n3} as follows. By using condition (C2) and the Taylor's series expansion, we have

$$\max_{1\le i\le n}\{\|R(x_i)\|\mathbf{1}(|x_i-x_0|\le h)\} \le \frac{1}{2}\sup_{|\xi-x_0|\le h}\|Y^{(2)}(\xi)\|h^2 = O_p(h^2).$$

Let $D_\delta = \{\Delta_n = (\eta_1, \cdots, \eta_n) : \|\eta_j\| \le \delta, \eta_j \in \text{Sym}(m), \forall j \le n\}$ for any $\delta > 0$. Define

$$\begin{aligned} V(\Delta_n) &= \frac{1}{nh}\sum_{i=1}^n hK_h(x_i-x_0)\{\psi_G(S_i, G(x_0), Y(x_i)+\eta_i) - \psi_G(S_i, G(x_0), Y(x_i)) \\ &\quad -\psi_{GY}(S_i, G(x_0), Y(x_i))^T \text{vecs}(\eta_i)\}, \end{aligned}$$

By using condition (C8), as $\delta \to 0$, we have $\epsilon_\delta = E\{\sup_{D_\delta}\|\psi_G(S_i, G(x_0), Y(x_i)+\eta_i) - \psi_G(S_i, G(x_0), Y(x_i)) - \psi_{GY}(S_i, G(x_0), Y(x_i))^T\text{vecs}(\eta_i)\||x_i = x\} = o(\delta)$ uniformly in a neighborhood of x_0. Therefore, for all $|x_i - x_0| \le h$, we have, as $\delta \to 0$,

$$E\{\sup_{D_\delta}\|V(\Delta_n)\|\} \le \epsilon_\delta \frac{1}{nh}E\left\{\sum_{i=1}^n hK_h(x_i-x_0)\right\} = o(1).$$

Since $\max_{1\le i\le n}\{\|R(x_i)\|\mathbf{1}(|x_i-x_0| \le h)\} = O_p(h^2) = o_p(1)$, we have $V(\hat{\Delta}_n) = o_p(h^2)$, where $\hat{\Delta}_n = (R(x_1), \cdots, R(x_n))$. This leads to $J_{n3} = nhV(\hat{\Delta}_n) = o_p(nh^3)$. Applying equation (19) in Lemma 3 to the second term J_{n2} in (27), we get

$$J_{n2} = \frac{1}{2}nh^3 f_X(0+)u_{2,d}\Psi_2(0+)^T\text{vecs}(Y^{(2)}(0+))\{1+o_p(1)\},$$

which yields (25).

Lemma 5. *Assume that conditions (C1)-(C9) hold. Let*

$$\mathcal{T}_n \equiv \begin{pmatrix} \sum_{i=1}^n hK_h(x_i-x_0)\psi_G(S_i, G(x_0), Y(x_i)) \\ \sum_{i=1}^n hK_h(x_i-x_0)(x_i-x_0)\psi_Y(S_i, G(x_0), Y(x_i))/h \end{pmatrix}. \qquad (28)$$

Then $\mathcal{T}_n/\sqrt{nh}$ is asymptotically normal with mean zero and covariance matrix

$$\Sigma_{\mathcal{T}} = f_X(0+) \begin{pmatrix} v_{0,d}\Psi_{11}(0+) & v_{1,d}\Psi_{12}(0+) \\ v_{1,d}\Psi_{12}(0+)^T & v_{2,d}\Psi_{22}(0+) \end{pmatrix} \{1+o(1)\}, \tag{29}$$

where $v_{k,d}$ for $k=0,1,2$ and $\Psi_{11}(x)$, $\Psi_{12}(x)$ and $\Psi_{22}(x)$ are defined in Section 4.2.

Proof of Lemma 5. Let $\mathcal{T}_G^{(ij)}$, and $\mathcal{T}_Y^{(ij)}$ denote the jth elements of $\psi_G(S_i, G(x_0), Y(x_i))$ and $\psi_Y(S_i, G(x_0), Y(x_i))$, respectively. Let

$$\mathcal{T}_{ni} = \begin{pmatrix} hK_h(x_i-x_0)\psi_G(S_i, G(x_0), Y(x_i)) \\ hK_h(x_i-x_0)(x_i-x_0)\psi_Y(S_i, G(x_0), Y(x_i))/h \end{pmatrix}.$$

Note that $\mathcal{T}_{ni}$ are independent and $\mathcal{T}_n = \sum_{i=1}^n \mathcal{T}_{ni}$.

It follows from Lemma 2 that $E(\mathcal{T}_{ni}) = E(\mathcal{T}_n) = \mathbf{0}$ and the covariance matrix of $\mathcal{T}_n/\sqrt{nh}$ is

$$\frac{n}{nh} E\left\{ h^2 K_h^2(X-x_0) \begin{pmatrix} \psi_G\psi_G^T, & \psi_G\psi_Y^T(X-x_0)/h \\ \psi_Y\psi_G^T(X-x_0)/h, & \psi_Y\psi_Y^T(X-x_0)^2/h^2 \end{pmatrix} \right\}.$$

Using the same arguments as in Lemma 3, we can obtain the asymptotic expression for the covariance matrix of $\mathcal{T}_n$, which equals Σ in (29). Finally, we will show that the sequence $\mathcal{T}_{ni}/\sqrt{nh}$ satisfies the Linderberg-Feller condition:

$$\sum_{i=1}^n E[\|\frac{\mathcal{T}_{ni}}{\sqrt{nh}}\|^2 1\left\{\|\frac{\mathcal{T}_{ni}}{\sqrt{nh}}\| > \epsilon\right\}] \to 0 \quad \text{for any} \quad \epsilon > 0. \tag{30}$$

Let $b>0$ and $\|\mathcal{T}_{ni}'\|^2 = (\mathcal{T}_G^{(i1)})^2 + \cdots + (\mathcal{T}_G^{(iq)})^2 + h^{-2}\{(\mathcal{T}_Y^{(i1)})^2 + \cdots + (\mathcal{T}_Y^{(iq)})^2\}(x_i-x_0)^2$.

$$\begin{aligned} & E\left[\left\|\frac{\mathcal{T}_{ni}}{\sqrt{nh}}\right\|^2 1\left\{\left\|\frac{\mathcal{T}_{ni}}{\sqrt{nh}}\right\| > \epsilon\right\}\right] \le \mathcal{K}_i \equiv E\left\{\frac{\left(K(h^{-1}(x_i-x_0))\right)^{b+2}\|\mathcal{T}_{ni}'\|^{(b+2)}}{(\sqrt{nh})^{b+2}(\epsilon)^b}\right\} \\ &= \int_{-\min\{d,1\}}^1 \frac{K(z)^{b+2}E\{\|\mathcal{T}_{ni}'\|^{(b+2)}|x_i = zh+x_0\}}{(\sqrt{nh})^{b+2}h^{-1}\epsilon^b} f_X(zh+x_0)dz \\ &\le \int_{-\min\{d,1\}}^1 \frac{E\{\|\psi_\alpha(S_i, G(x_0), Y(X))\|^{b+2}|X = zh+x_0\}}{\{K(z)\}^{-b-2}(\sqrt{nh})^{b+2}f_X(zh+x_0)^{-1}h^{-1}\epsilon^b} dz. \end{aligned}$$

Combining conditions (C4) and (C9) yields that there is a constant $d_5>0$ such that $\sum_{i=1}^n \mathcal{K}_i \le d_5 nh/(\sqrt{nh})^{b+2} \to 0$. Thus, it follows from the Linderberg-Feller theorem that $\mathcal{T}_n/\sqrt{nh}$ is asymptotically normal with mean 0 and covariance Σ.

Proof of Theorem 3 (i). Let $\tilde{Y}^{(1)}(x_0) = hY^{(1)}(x_0)$, $\gamma = (\text{vecs}(G(x))^T, \text{vecs}(\tilde{Y}^{(1)}(x))^T)^T$ and $\gamma_0 = (\text{vecs}(G(x_0))^T, \text{vecs}(hY^{(1)}(x_0))^T)^T$. Thus, $G_n(\gamma)$ can be written as

$$\sum_{i=1}^n hK_h(x_i-x_0)\psi(S_i, G(x_0), \tilde{Y}^{(1)}(x_0)(x_i-x_0)/h).$$

Let $U_\delta(\gamma_0) = \{\gamma : ||\gamma - \gamma_0|| < \delta\}$ for $\delta > 0$. We will show that for any small $\delta > 0$,

$$\lim_{n\to\infty} P\{\inf_{\gamma\in U_\delta(\gamma_0)} G_n(\gamma) > G_n(\gamma_0)\} = 1. \tag{31}$$

By a Taylor's series expansion, we have

$$G_n(\gamma) - G_n(\gamma_0) = G_n^{(1)}(\gamma_0)^T(\gamma-\gamma_0) + \frac{1}{2}(\gamma-\gamma_0)^T G_n^{(2)}(\gamma^*)(\gamma-\gamma_0), \tag{32}$$

where $\gamma \in U_\delta(\gamma_0)$ and $\gamma^* = (\text{vecs}(G(x_*))^T, \text{vecs}(\tilde{Y}^{(1)}(x_*))^T)^T$ lies in between $\gamma(x_0)$ and γ.

It follows from Lemmas 4 and 5 that

$$\begin{aligned}
&(nh)^{-1}G_n^{(1)}(\gamma_0)\\
=\;& \frac{1}{nh}\sum_{i=1}^n hK_h(x_i-x_0)\begin{pmatrix}\psi_G(S, G(x_0), Y^{(1)}(x_0)(x_i-x_0))\\ \psi_Y(S, G(x_0), Y^{(1)}(x_0)(x_i-x_0))(x_i-x_0)/h\end{pmatrix}\\
=\;& \frac{1}{2}h^2 f_X(0+)\text{vecs}(Y^{(2)}(0+))\begin{pmatrix}u_{2,d}\Psi_2(0+)\\ u_{3,d}\Psi_3(0+)\end{pmatrix}\{1+o_p(1)\}\\
&+\frac{1}{nh}\sum_{i=1}^n hK_h(x_i-x_0)\begin{pmatrix}\psi_G(S_i, G(x_0), Y(x_i))^T\\ h^{-1}(x_i-x_0)\psi_Y(S_i, G(x_0), Y(x_i))^T\end{pmatrix}\\
=\;& \frac{1}{2}h^2 f_X(0+)(u_{2,d}\Psi_2(0+)^T, u_{3,d}\Psi_3(0+)^T)^T\text{vecs}(Y^{(2)}(0+))\{1+o_p(1)\} + \frac{1}{nh}\mathcal{T}_n\\
=\;& o_p(1).
\end{aligned}$$

We define $\Delta(S, G, Y, X)$ as

$$\begin{pmatrix}\psi_{GG}(S,G,Y) & (X-x_0)h^{-1}\psi_{YG}(S,G,Y)\\ (X-x_0)h^{-1}\psi_{GY}(S,G,Y) & (X-x_0)^2h^{-2}\psi_{YY}(S,G,Y)\end{pmatrix}. \tag{33}$$

With some calculation, we have

$$\begin{aligned}
&\frac{1}{nh}G_n^{(2)}(\gamma^*) = \frac{1}{nh}\sum_{i=1}^n hK_h(x_i-x_0)[\{\Delta(S_i, G(x_*), \tilde{Y}^{(1)}(x_*)(x_i-x_0)h^{-1}, x_i)\\
&-\Delta(S_i, G(x_0), Y(x_i), x_i)\} + \Delta(S_i, G(x_0), Y(x_i), x_i)]\\
&\equiv M_{n1} + M_{n2}.
\end{aligned}$$

It follows from (16), (17) and (18) in Lemma 3 that

$$M_{n2} = f_X(x_0)\begin{pmatrix}u_{0,d}\Psi_1(0+) & u_{1,d}\Psi_2(0+)^T\\ u_{1,d}\Psi_2(0+) & u_{2,d}\Psi_3(0+)\end{pmatrix}\{1+o_p(1)\},$$

which is positive definite in probability. Note that $\|G(x_*) - G(x_0)\| \le \delta$ and for all $|x_i - x_0| < h$, we have $\max_i\|\tilde{Y}^{(1)}(x_*)(x_i-x_0)/h - Y(x_i)\| \le \max_i\|R(x_i)\| + \delta$. Then it follows from Lemma 3 that $\limsup_{\delta\to 0}\limsup_{n\to\infty}\|M_{n1}\| = 0$ in probability. Thus, for any $\gamma \in U_\delta(\gamma_0)$, we have for sufficiently small $\delta > 0$

$$\lim_{n\to\infty} P[\inf_{\gamma\in U_\delta(\gamma_0)} \frac{1}{nh}(\gamma-\gamma_0)^T G_n^{(2)}(\gamma^*)(\gamma-\gamma_0) > 0.5af_X(x_0)\delta^2] = 1,$$

which yields that $G_n(\gamma)$ has a local minimum $\hat{\gamma} = (\text{vecs}(\hat{G})^T, \text{vecs}(h\hat{Y}^{(1)})^T)^T$ in the interior of $U_\delta(\gamma_0)$. Then, we have $\lim_{n\to\infty} P\{\|\hat{\gamma}_n - \gamma_0\| \le \delta\} = 1$. This implies Theorem 3 (i).

Proof of Theorem 3(iii). Let $\hat{\eta}_0 = \hat{G} - G(x_0)$ and $\hat{\eta}_{1i} = \{\hat{Y}^{(1)} - Y^{(1)}(x_0)\}(x_i - x_0) - R(x_i)$. The estimator $\hat{\gamma}$ satisfies the following local estimating equations:

$$\sum_{i=1}^n K_h(x_i - x_0)\partial_\gamma \psi(S_i, \hat{G}, \hat{Y}^{(1)}(x_i - x_0)) = \mathbf{0}, \tag{34}$$

where $\partial_\gamma\psi(S,G,Y) \equiv \left(\psi_G(S,G,Y)^T, (X-x_0)\psi_Y(S,G,Y)^T\right)^T$. It follows from (34) that

$$\begin{aligned}\sum_{i=1}^n hK_h(x_i - x_0)[&\partial_\gamma\psi(S_i, G(x_0), Y(x_i)) + \partial^2_{\alpha\gamma}\psi(S, G(x_0), Y(x_i))^T(\text{vecs}(\hat{\eta}_0)^T, \text{vecs}(\hat{\eta}_{1i})^T)^T \\ &+ \{\partial_\gamma\psi(S, G(x_0) + \hat{\eta}_0, Y(x_i) + \hat{\eta}_{1i}) - \partial_\gamma\psi(S, G(x_0), Y(x_i)) \\ &\quad - \partial^2_{\alpha\gamma}\psi(S, G(x_0), Y(x_i))^T(\text{vecs}(\hat{\eta}_0)^T, \text{vecss}(\hat{\eta}_{1i})^T)^T\}] = 0.\end{aligned} \tag{35}$$

Note that the second term on the left hand side of (35) equals the sum of L_{n1} and L_{n2}, which can, respectively, be approximated by

$$\begin{aligned} L_{n1} &= \sum_{i=1}^n hK_h(x_i - x_0)\Delta(S_i, G(x_0), Y(x_i), x_i)\begin{pmatrix} \text{vecs}(\hat{\eta}_0) \\ h\cdot\text{vecs}(\hat{Y}^{(1)} - Y^{(1)}(x_0))\end{pmatrix} \\ &= f_X(0+)nh\begin{pmatrix} u_{0,d}\Psi_1(0+) & u_{1,d}\Psi_2(0+) \\ u_{1,d}\Psi_2(0+)^T & u_{2,d}\Psi_3(0+)\end{pmatrix}\begin{pmatrix} \text{vecs}(\hat{\eta}_0) \\ h\cdot\text{vecs}(\hat{Y}^{(1)} - Y^{(1)}(x_0))\end{pmatrix}\times \\ &\quad \{1 + o_p(1)\}, \\ L_{n2} &= \sum_{i=1}^n hK_h(x_i - x_0)\Delta(S_i, G(x_0), Y(x_i), x_i)\begin{pmatrix} \text{vecs}(O_m) \\ h\cdot\text{vecs}(R(x_i))\end{pmatrix} \\ &= \frac{1}{2}f_X(0+)nh^3\begin{pmatrix} u_{2,d}\Psi_2(0+) \\ u_{3,d}\Psi_3(0+)\end{pmatrix}\text{vecs}(Y^{(2)}(0+))\{1 + o_p(1)\}.\end{aligned}$$

By the consistency of $\hat{\gamma} = (\text{vecs}(\hat{G})^T, \text{vecs}(h\hat{Y}^{(1)})^T)^T$, we have $\|\hat{\eta}_0\| = o_p(1) = O_p(\|\hat{G} - G(x_0)\|)$ and

$$\begin{aligned}\sup_{\{i:|x_i - x_0|\le h\}}\|\hat{\eta}_{i1}\| &\le \sup_{\{i:|x_i - x_0|\le h\}}\{\|R(x_i)\| + h\|\hat{Y}^{(1)} - Y^{(1)}(x_0)\|\} \\ &= O_p(h^2 + h\|\hat{Y}^{(1)} - Y^{(1)}(x_0)\|) = o_p(1).\end{aligned}$$

Thus, it follows from (C8) and Lemma 4 that the third term on the left hand side of (35) is at the order of $o_p(nh)\{h^2 + \|\hat{G} - G(x_0)\| + h\|\hat{Y}^{(1)} - Y^{(1)}(x_0)\|\}$. Let

$$B_n = \frac{h^2}{2}\begin{pmatrix} u_{0,d}\Psi_1(0+) & u_{1,d}\Psi_2(0+)^T \\ u_{1,d}\Psi_2(0+) & u_{2,d}\Psi_3(0+) \end{pmatrix}^{-T}\begin{pmatrix} u_{2,d}\Psi_2(0+)^T \\ u_{3,d}\Psi_3(0+)^T \end{pmatrix}\text{vecs}(Y^{(2)}(0+))\{1 + o_p(1)\}.$$

Then from (35), we obtain that

$$\hat{\gamma} - \gamma_0 = B_n + \{nhf_X(0+)\}^{-1}\begin{pmatrix} u_{0,d}\Psi_1(0+) & u_{1,d}\Psi_2(0+)^T \\ u_{1,d}\Psi_2(0+) & u_{2,d}\Psi_3(0+) \end{pmatrix}^{-T}\mathcal{T}_n\{1 + o_p(1)\}.$$

Finally, Theorem 3 (iii) follows from Lemma 5 and the Slutsky's theorem.

The above derivation holds for any $k_0 > 0$. When $k_0 = 0$ and $K(.)$ is symmetric, the following modifications need to be made.

Lemma 6. *Let $\eta_0(X) = G(X) - G(x_0)$. Assume that conditions (C1)-(C5) and (C7) hold. If x_0 is an interior point of $f_X(\cdot)$, then we have*

$$\begin{aligned}&\sum_{i=1}^n hK_h(x_i - x_0)\psi_{GG}(S_i, G(x_i))vecs(\eta_0(x_i)) \\ =\;& nh^3\Psi_1(x_0)\{G^{(1)}(x_0)f_X^{(1)}(x_0) + 0.5G^{(2)}(x_0)f_X(x_0)\}u_2\{1 + o_p(1)\}. \qquad (36)\end{aligned}$$

If x_0 is a boundary point of $f_X(\cdot)$, then we have

$$\begin{aligned}&\sum_{i=1}^n hK_h(x_i - x_0)\psi_{GG}(S_i, G(x_i))vecs(\eta_0(x_i)) \\ &\qquad = nh^2 f_X(0+)\Psi_1(0+)G^{(1)}(0+)u_{1,d}\{1 + o_p(1)\}. \qquad (37)\end{aligned}$$

Proof of Lemma 6. We only prove equation (36). Let

$$T_{n1} = \sum_{i=1}^n hK_h(x_i - x_0)\psi_{GG}(S_i, G(x_i))\text{vecs}(\eta_0(x_i)).$$

For the (j)-th element $(T_{n1})_j$ of the vector T_{n1}, we have

$$(T_{n1})_j = nE(Z_j) + O_p(\sqrt{nE(Z_j^2)}). \qquad (38)$$

We calculate the first two moments of Z_j below. Note that

$$\begin{aligned} E(Z_j) &= E[hK_h(X - x_0)(\psi_{GG}(S_i, G(X))\text{vecs}(\eta_0(X)))_j] \\ &= \int_{-\infty}^{\infty} hK_h(y - x_0)(E\{\psi_{GG}(S_i, G(X))|X = y\}\text{vecs}(\eta_0(y)))_j f_X(y)dy \\ &= h\int_{-1}^{1} K(z)(E\{\psi_{GG}(S_i, G(X))|X = zh + x_0\}\text{vecs}(\eta_0(zh + x_0)))_j f_X(zh + x_0)dz, \end{aligned}$$

By a Taylor's series expansion, we have $\eta_0(zh+x_0) = G^{(1)}(x_0)zh + G^{(2)}(x_0)(zh)^2/2$. Applying the dominated convergence theorem together with the continuity assumptions on $f_X^{(1)}(\cdot)$, $G(.)^{(2)}$ and $\Psi_1(.)$ yields

$$E(Z_j) = h^3 u_2(\Psi_1(x_0)\{G^{(1)}(x_0)f_X^{(1)}(x_0) + 0.5G^{(2)}(x_0)f_X(x_0)\})_j\{1+o(1)\}. \tag{39}$$

By the continuity $f_X^{(1)}(x)$ and $G(x)^{(2)}$ together with condition (C7), we have

$$\begin{aligned} E(Z_j^2) &= E\{h^2K_h^2(X-x_0)(\psi_{GG}(S_i,G(x_i))\text{vecs}(\eta_0(x_i)))_j^2\} \\ &= \int_{-\infty}^{\infty} h^2K_h^2(y-x_0)f_X(y)(E\{\psi_{GG}(S,G(X))|X=y\}\text{vecs}(\eta_0(y)))_j^2 dy \\ &= h\int_{-1}^{1} K^2(z)z^{2l}f_X(zh+x_0)(E\{\psi_{GG}(S,G(X))|X=zh+x_0\} \\ &\quad \text{vecs}(\eta_0(zh+x_0)))_j^2 dz \\ &= h^3\{1+O(h)\} \end{aligned} \tag{40}$$

Combining with (38),(39) and (40), we have

$$\begin{aligned} (T_{n1})_j &= nh^3[(\Psi_1(x_0)\{G^{(1)}(x_0)f_X^{(1)}(x_0) + 0.5G^{(2)}(x_0)f_X(x_0)\})_j u_2\{1+o(1)\} \\ &\quad + O_p(1/\sqrt{nh^3})] \\ &= nh^3(\Psi_1(x_0)\{G^{(1)}(x_0)f_X^{(1)}(x_0) + 0.5G^{(2)}(x_0)f_X(x_0)\})_j u_2\{1+o_p(1)\}. \end{aligned}$$

That is, $T_{n1} = nh^3\Psi_1(x_0)\{G^{(1)}(x_0)f_X^{(1)}(x_0) + 0.5G^{(2)}(x_0)f_X(x_0)\}u_2\{1+o_p(1)\}$.

Lemma 7. *Assume that conditions (C1)-(C7) hold. If x_0 is an interior point of $f_X(\cdot)$, then we have*

$$\begin{aligned} &\sum_{i=1}^n hK_h(x_i-x_0)\psi_G(S_i,G(x_i)) \\ = \; & nh^3\Psi_1(x_0)\{G^{(1)}(x_0)f_X^{(1)}(x_0) + 0.5G^{(2)}(x_0)f_X(x_0)\}u_2\{1+o_p(1)\} + \\ &\sum_{i=1}^n hK_h(x_i-x_0)\psi_G(S_i,G(x_0)). \end{aligned} \tag{41}$$

If x_0 is the left boundary point of $f_X(\cdot)$,

$$\begin{aligned} &\sum_{i=1}^n hK_h(x_i-x_0)\psi_G(S_i,G(x_i)) \\ = \; & nh^2 f_X(0+)\Psi_1(0+)G^{(1)}(0+)u_{1,d}\{1+o_p(1)\} + \sum_{i=1}^n hK_h(x_i-x_0)\psi_G(S_i,G(x_0)). \end{aligned} \tag{42}$$

Lemma 8. *Assume that conditions (C1)-(C9) hold. Let*

$$\mathcal{T}_n \equiv \sum_{i=1}^{n} hK_h(x_i - x_0)\psi_G(S_i, G(x_i)). \tag{43}$$

Then $\mathcal{T}_n/\sqrt{nh}$ is asymptotically normal with mean zero and covariance matrices

$$\Sigma = f_X(x_0)v_0\Psi_{11}(x_0)\{1+o(1)\} \text{ if } x_0 \text{ is an interior point of } f_X(x), \tag{44}$$

$$\Sigma = f_X(0+)v_0\Psi_{11}(0+)\{1+o(1)\} \text{ if } x_0 = dh \text{ is a boundary point of } f_X(x). \tag{45}$$

The proof is similar to the proof of Lemma 5.

Proof of Theorem 3(ii). We use Lemmas 6-8 and follow the same lines of Theorem 3 (iii) to prove Theorem 3 (ii).

4. Comparisons

We study the asymptotic relative efficiency of the intrinsic local constant and linear estimators in an interior point x_0 under the trace and Log-Euclidean metrics by comparing their AMSEs/AMISEs. We use AMSE_{opt} and AMISE_{opt} to denote the AMSE and AMISE evaluated at their optimal bandwidth. We first compare AMSE_{opt} for the intrinsic local constant estimators under the trace and Log-Euclidean metrics. Specifically, with some calculations, we see that as n approaches ∞, the ratio of $\text{AMSE}_{opt}(\log\{\hat{D}_{IT}(x_0;h,0)\})$ over $\text{AMSE}_{opt}(\log\{\hat{D}_{IL}(x_0;h,0)\})$ converges to

$$\text{rMSE}(T,L;0) = \left[\frac{\text{tr}\{G_D(x_0)^{\otimes 2}\Psi_1(x_0)^{-1}\Psi_{11}(x_0)\Psi_1(x_0)^{-1}\}}{\text{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}}\right]^{4/5}$$
$$\times\left\{\frac{\text{tr}([G_D(x_0)^T\text{vecs}\{f_X^{(1)}(x_0)f_X(x_0)^{-1}G^{(1)}(x_0)+0.5G^{(2)}(x_0)\}]^{\otimes 2})}{\text{tr}[\text{vecs}\{f_X^{(1)}(x_0)f_X(x_0)^{-1}\log(D(x_0))^{(1)}+0.5\log(D(x_0))^{(2)}\}^{\otimes 2}]}\right\}^{1/5}, \tag{46}$$

which is the product of two terms. The first term is associated with the ratio of the covariances of the intrinsic local constant estimators of $\log(D(x_0))$ under the two metrics, while the second term is associated with their biases. Consider the simplest scenario with $m = 1$ such that $D(x_0) = G(x_0)^2$ and $G(x_0) > 0$. By simple calculations, we can show that the first term equals one and the second term equals

$$\left[\frac{f_X^{(1)}(x_0)f_X(x_0)^{-1}G^{(1)}(x_0)+0.5G^{(2)}(x_0)}{f_X^{(1)}(x_0)f_X(x_0)^{-1}G^{(1)}(x_0)+0.5G^{(2)}(x_0)-0.5\{G^{(1)}(x_0)\}^2/G(x_0)}\right]^{2/5},$$

which yields that $\text{rMSE}(T,L;0) > 1$ if and only if

$$0.25\{G^{(1)}(x_0)\}^2/G(x_0) < f_X^{(1)}(x_0)f_X(x_0)^{-1}G^{(1)}(x_0)+0.5G^{(2)}(x_0).$$

Thus, whether $\text{rMSE}(T, L; 0)$ is greater than 1 depends on both the design density and $D(x)$ itself as a function of x.

Similarly, we define $\text{rMISE}(T, L; 0)$ as the ratio of $\text{AMISE}_{opt}(\log\{\hat{D}_{IT}(x_0; h, 0)\})$ over AMISE_{opt} $(\log\{\hat{D}_{IL}(x_0; h, 0)\})$. Following similar arguments to $\text{rMSE}(T, L; 0)$, when $m = 1$, we have that $\text{rMISE}(T, L; 0) > 1$ if and only if

$$0.25 \int \{G^{(1)}(x)\}^2 G(x)^{-1} w(x) dx < \int \{f_X^{(1)}(x) f_X(x)^{-1} G^{(1)}(x) + 0.5 G^{(2)}(x)\} w(x) dx.$$

Therefore, in terms of AMSE_{opt} and AMISE_{opt}, the trace metric is not uniformly superior to or worse than the Log-Euclidean metric for reconstructing all $D(x)$.

We compare AMSE_{opt} for the intrinsic local linear estimators under the trace and Log-Euclidean metrics. As n approaches ∞, the ratio of $\text{AMSE}_{opt}(\log\{\hat{D}_{IT}(x_0; h, 1)\})$ over AMSE_{opt} $(\log\{\hat{D}_{IL}(x_0; h, 1)\})$ converges to

$$\begin{aligned} \text{rMSE}(T, L; 1) &= \left[\frac{\text{tr}\{G_D(x_0)^{\otimes 2}\Omega_0(x_0)\}}{\text{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}}\right]^{4/5} \\ &\quad \times \left\{\frac{\text{tr}([G_D(x_0)^T \Psi_1(x_0)^{-1}\Psi_2^T(x_0)\text{vecs}\{Y^{(2)}(x_0)\}]^{\otimes 2})}{\text{tr}(\text{vecs}[\log\{D(x_0)\}^{(2)}]^{\otimes 2})}\right\}^{1/5}. \qquad (47) \end{aligned}$$

We also consider the simplest scenario with $m = 1$ such that $D(x) = G(x_0)^2 \exp(Y(x))$ and $G(x_0) > 0$. With some calculations, we can show that $\text{rMSE}(T, L; 1)$ equals one when $m = 1$. Thus, the two metrics are actually the same for one dimensional case. However, when $m > 1$, it is unclear whether $\text{rMSE}(T, L; 1)$ equals to 1 or not.

5. Expressions of optimal bandwidths

By Theorem 1 (iii), $\text{AMSE}(\log\{\hat{D}_{IL}(x_0; h, 0)\})$ equals

$$\begin{aligned} &h^4 u_2^2 \text{tr}\{(\text{vecs}[0.5\log\{D(x_0)\}^{(2)} + f_X^{(1)}(x_0) f_X(x_0)^{-1}\log\{D(x_0)\}^{(1)}])^{\otimes 2}\} \\ &\qquad + v_0\{nhf_X(x_0)\}^{-1}\text{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}. \qquad (48) \end{aligned}$$

Minimizing $\text{AMSE}(\log\{\hat{D}_{IL}(x_0; h, 0)\})$ and $\text{AMISE}(\log\{\hat{D}_{IL}(x_0; h, 0)\})$ leads to

$$h_{opt,L}(x_0; 0)^5 = \frac{v_0\{nf_X(x_0)\}^{-1}\text{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}}{4u_2^2 \text{tr}\{(\text{vecs}[0.5\log\{D(x_0)\}^{(2)} + f_X^{(1)}(x_0) f_X(x_0)^{-1}\log\{D(x_0)\}^{(1)}])^{\otimes 2}\}}, \qquad (49)$$

$$h_{opt,L}(0)^5 = \frac{v_0\{nf_X(x_0)\}^{-1}\int \text{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}w(x)dx}{4u_2^2 \int \text{tr}\{(\text{vecs}[0.5\log\{D(x_0)\}^{(2)} + f_X^{(1)}(x_0) f_X(x_0)^{-1}\log\{D(x_0)\}^{(1)}])^{\otimes 2}\}w(x)dx}.$$

For the intrinsic local linear estimator, $\text{AMSE}(\log\{\hat{D}_{IL}(x_0; h, 1)\})$ is given by

$$0.25 h^4 u_2^2 \text{tr}\{(\text{vecs}[\log\{D(x_0)\}^{(2)}])^{\otimes 2}\} + v_0\{nhf_X(x_0)\}^{-1}\text{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}. \qquad (50)$$

Intrinsic local constant and linear estimators have the same asymptotic covariance and their differences are concerned only with their biases. The local constant estimator has one more term $h^2 u_2 f_X^{(1)}(x_0) f_X(x_0)^{-1} \text{vecs}[\log\{D(x_0)\}^{(1)}]$, which depends on the marginal density $f_X(.)$. Subsequently, we can get the optimal bandwidths, which detailed expression can be found in supplement.

$$h_{opt,L}(x_0;1)^5 = \frac{\{nf_X(x_0)\}^{-1} v_0 \text{tr}\{\Sigma_{\mathcal{E}_D}(x_0)\}}{u_2^2 \text{tr}\{(\text{vecs}[\log\{D(x_0)\}^{(2)}])^{\otimes 2}\}}, \tag{51}$$

$$h_{opt,L}(1)^5 = \frac{n^{-1} v_0 \int \text{tr}\{\Sigma_{\mathcal{E}_D}(x)\}\{f_X(x)\}^{-1} w(x)dx}{u_2^2 \int \text{tr}\{(\text{vecs}[\log\{D(x_0)\}^{(2)}])^{\otimes 2}\} w(x)dx}. \tag{52}$$

It follows from the delta method that AMSE$(\log\{\hat{D}_{IT}(x_0;h,0)\})$ can be approximated as

$$\begin{aligned} & h^4 u_2^2 \text{tr}([G_D(x_0)^T \text{vecs}\{G^{(1)}(x_0) f_X^{(1)}(x_0) f_X(x_0)^{-1} + 0.5 G^{(2)}(x_0)\}]^{\otimes 2}) \\ & \quad + (nh)^{-1} \text{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x_0)\} + o(h^4 + (nh)^{-1}), \end{aligned} \tag{53}$$

where $G_D(x_0) = \{\partial \text{vec}(\log(G(x_0)^{\otimes 2}))/\partial \text{vecs}(G(x_0))^T\}^T$. The asymptotic bias and variance of $\hat{D}_{IT}(x_0;h,0)$ are similar to those of the Nadaraya-Watson estimator when response is in Euclidean space (Fan, 1992). By minimizing AMSE$(\log(\hat{D}_{IT}(x_0;h,0)))$ and AMISE$(\log(\hat{D}_{IT}(x_0;h,0)))$, we have

$$h_{opt,T}(x_0;0)^5 = \frac{n^{-1} \text{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x_0)\}}{4u_2^2 \text{tr}([G_D(x_0)^T \text{vecs}\{G^{(1)}(x_0) f_X^{(1)}(x_0) f_X(x_0)^{-1} + 0.5 G^{(2)}(x_0)\}]^{\otimes 2})},$$

$$h_{opt,T}(0)^5 = \frac{n^{-1} \int \text{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x_0)\} w(x)dx}{4u_2^2 \int \text{tr}([G_D(x_0)^T \text{vecs}\{G^{(1)}(x_0) f_X^{(1)}(x_0) f_X(x_0)^{-1} + 0.5 G^{(2)}(x_0)\}]^{\otimes 2}) w(x)dx}.$$

For the intrinsic local linear estimator, AMSE$(\log(\hat{D}_{IT}(x_0;h,1)))$ is given by

$$0.25 h^4 u_2^2 \text{tr}[\{G_D(x_0)^T \Psi_1(x_0)^{-1} \Psi_2^T(x_0) \text{vecs}(Y^{(2)}(x_0))\}^{\otimes 2}] + (nh)^{-1} \text{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x_0)\}. \tag{54}$$

Minimizing AMSE$(\log(\hat{D}_{IT}(x_0;h,1)))$ and AMISE$(\log(\hat{D}_{IT}(x_0;h,1)))$, respectively, leads to

$$h_{opt,T}(x_0;1)^5 = \frac{n^{-1} \text{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x_0)\}}{u_2^2 \text{tr}[\{G_D(x_0)^T \Psi_1(x_0)^{-1} \Psi_2^T(x_0) \text{vecs}(Y^{(2)}(x_0))\}^{\otimes 2}]}, \tag{55}$$

$$h_{opt,T}(1)^5 = \frac{n^{-1} \int \text{tr}\{G_D(x_0)^{\otimes 2} \Omega_0(x)\} w(x)dx}{u_2^2 \int \text{tr}[\{G_D(x_0)^T \Psi_1(x_0)^{-1} \Psi_2^T(x_0) \text{vecs}(Y^{(2)}(x_0))\}^{\otimes 2}] w(x)dx}.$$

6. Figures

This section displays AGD and LAGD for the trace and Log-Euclidean metrics. Figure 1 and Figure 2 are for the trace and Log-Euclidean metrics, respectively at the moderate noise level Σ_1. We observe that the comparison measurements based on these two metrics reveal results similar to that under the Euclidean metric. Under all metrics, the local linear estimator is superior to the local constant estimator. Also, our ILPREs outperform the corresponding estimators under the Euclidean metric and the tensor spline estimators under the noise models (a) and (b). For the Rician noise model, our ILPREs under the Log-Euclidean metric slightly outperform than those under the trace and Euclidean metrics. Moreover, the local constant and linear estimators outperform the tensor spline estimators under all noise distributions. The variations of AGDs for ILPREs under the trace metric are larger than those under the Log-Euclidean metric under all three noise distributions. The U shape of the LAGD curves indicates that interior points have smaller LAGDs than those near the boundaries and there are more design points in the center than those on the boundaries.

Figure 3 and Figure 4 are for the trace and Log-Euclidean metrics, respectively at relatively high noise level $\Sigma = 4\Sigma_1$. We observe that the comparison measurements based on these two metrics reveal results similar to that under the Euclidean metric. When the noise level is high, the intrinsic local linear estimators under the trace metric slightly outperforms those under the Log-Euclidean metric under the noise models (a) and (b).

References

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning.* Reading, Massachusetts: Addison Wesley.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika 57*, 97–109.

Higham, N. J. (2008). *Functions of Matrices: Theory and Computation.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi (1983). Optimization by simulated annealing. *Science 220*, 671–680.

Liang, F. (2010). Evolutionary stochastic approximation monte carlo for global optimization. *Statistics and Computing*, In press.

Liang, F., C. Liu, and R. Carroll (2007). Stochastic approximation in monte carlo computation. *J. Am. Statist. Assoc. 102*, 305–320.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *J. Chemical Physics 21*, 1087–1091.

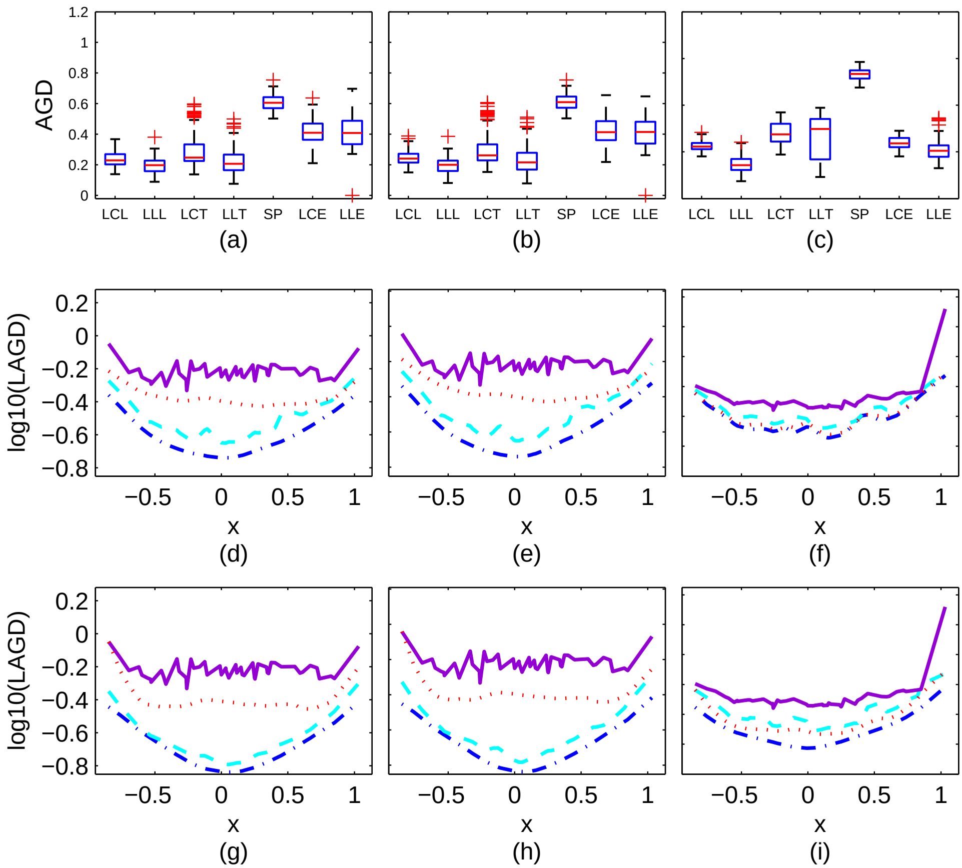

Fig. 1. Comparisons of the local constant and linear estimators under the three metrics and the tensor spline estimators under the three noise distributions. Panels (a)-(c) of the first row show the boxplots of $1000\times$AGDs obtained from seven different estimators, in which LCL, LCT, and LCE, respectively, represent the local constant estimators under the Log-Euclidean, trace and Euclidean metrics, LLL, LLT, and LLE , respectively, represent the corresponding local constant and linear estimators under the metrics, and SP represents the tensor spline estimator. Panels (d)-(f) of the second row show the $\log_{10}$(LAGD) curves based on LCL (dash-dotted line), LCT (dashed line), LCE (dotted line), and SP (solid line). Panels (g)-(i) of the third row show the $\log_{10}$(LAGD) curves based on LLL (dash-dotted line), LLT (dashed line), LLE (dotted line), and SP (solid line). The columns correspond to the three noise models: column 1: Riemannian log normal; column 2: log normal; and column 3: Rician.

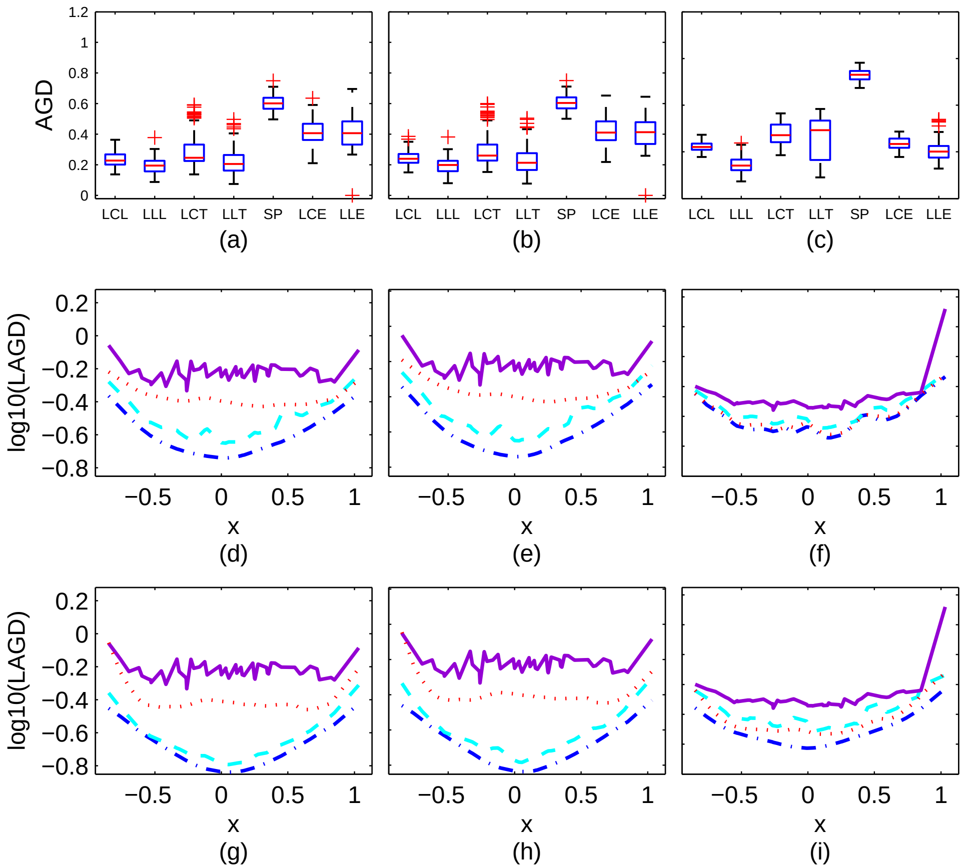

Fig. 2. Comparisons of the local constant and linear estimators under the three metrics and the tensor spline estimators under the three noise distributions. Panels (a)-(c) of the first row show the boxplots of $1000\times$AGDs obtained from seven different estimators, in which LCL, LCT, and LCE, respectively, represent the local constant estimators under the Log-Euclidean, trace and Euclidean metrics, LLL, LLT, and LLE , respectively, represent the corresponding local constant and linear estimators under the metrics, and SP represents the tensor spline estimator. Panels (d)-(f) of the second row show the $\log_{10}$(LAGD) curves based on LCL (dash-dotted line), LCT (dashed line), LCE (dotted line), and SP (solid line). Panels (g)-(i) of the third row show the $\log_{10}$(LAGD) curves based on LLL (dash-dotted line), LLT (dashed line), LLE (dotted line), and SP (solid line). The columns correspond to the three noise models: column 1: Riemannian log normal; column 2: log normal; and column 3: Rician.

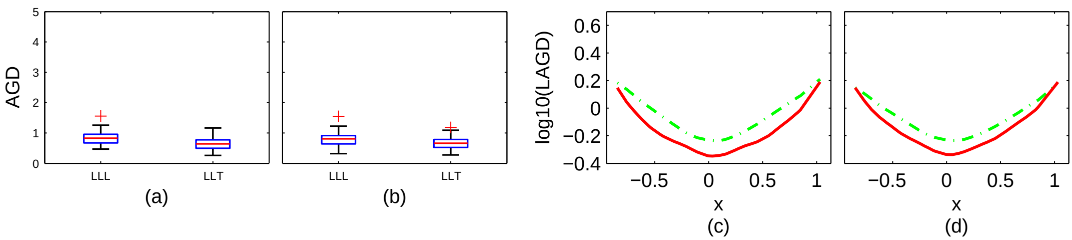

Fig. 3. Comparison of the intrinsic local linear estimators under the Log-Euclidean and trace metrics for the first two noise models at a relatively higher noise level: Panels (a) and (c): Riemannian log normal; Panels (b) and (d): log normal; Panels (a) and (b): the boxplots of AGDs for LLL and LLT; Panels (c) and (d): the $\log_{10}$(LAGD) curves of LLL and LLT.

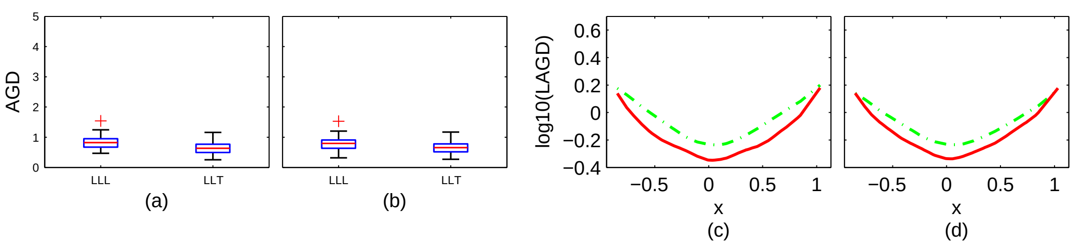

Fig. 4. Comparison of the intrinsic local linear estimators under the Log-Euclidean and trace metrics for the first two noise models at a relatively higher noise level: Panels (a) and (c): Riemannian log normal; Panels (b) and (d): log normal; Panels (a) and (b): the boxplots of AGDs for LLL and LLT; Panels (c) and (d): the $\log_{10}$(LAGD) curves of LLL and LLT.