Local Smoothing Methods in Image Processing

Vera Friederichs

Institut für Angewandte Mathematik, Universität Heidelberg

Im Neuenheimer Feld 294, D-69120 Heidelberg, Germany

E-mail address: vera@statlab.uni-heidelberg.de

January 15, 2002

Abstract

In this article a new data-adaptive method for smoothing of bivariate functions is developed. The smoothing is done by kernel regression with rotational invariant bivariate kernels. Two or three local bandwidth parameters are chosen automatically by a two-step plug-in approach. The algorithm starts with small global bandwidth parameters, which adapt during a few iterations to the noisy image. In the next step local bandwidths are estimated.

Some general asymptotic results about Gasser-Müller-estimators and optimal bandwidth selection are given. The derived local bandwidth estimators converge and are asymptotically normal.

Key words: 2-d kernel regression, data-adaptive bandwidth choice, iterative bandwidth choice, local bandwidths

1 Introduction

In this paper smoothing of bivariate functions is investigated. It can be applied e. g. for image processing or spatial data like geographical data. There exist bivariate smoothing methods like wavelet denoising, diffusion filtering, Gauss filtering with a Nadaraya-Watson estimator, nonlinear Gauss filtering by Godtliebsen et al. (1997) and modal regression by Scott (1992), but they are either not completely data-adaptive and the user needs a lot of experience to specify suitable parameters or they are not flexible enough to adapt to inhomogeneities. Polzehl and Spokoiny (1998) developed a completely data-adaptive procedure, but it works only for piecewise constant functions.

Another important class of nonparametric estimators are kernel estimators with global or local bandwidths. If the underlying true function is inhomogenous, then local bandwidths allow a better fit. However, an automatic data-adaptive procedure for the bandwidth selection is necessary, e. g. as in Fan and Gijbels (1995) for local polynomials or in Brockmann (1993); Brockmann et al. (1993) for Gasser-Müller-kernel-estimators (both for the one-dimensional case). Mammen and Gijbels (1995) have shown, that plug-in estimators do not achieve optimal rates of convergence for certain classes of functions, but, nevertheless, for finite samples they work well. Here, data-adaptive algorithms for two or three bandwidth parameters based on plug-in rules for Gasser-Müller-estimators are developed.

In section 2 general asymptotic results for Gasser-Müller-estimators are obtained and the pros and cons of local or global parameters and the number of bandwidth parameters are discussed. The algorithms are introduced in Chapter 3. In Chapter 4 asymptotic properties of the bandwidth estimators are considered.

2 Theory of Bivariate Kernel Regression

2.1 Definitions and Assumptions

A bivariate regression model of the form

$$Z_k = r(x_k, y_k) + \epsilon_k,$$

$k = 1, ..., n$, is considered, where Z_k are the data, r the unknown real-valued regression function on a compact subset $A \subset \mathbb{R}^2$ (w.l.o.g. $\lambda(A) = 1$, λ the Lebesgue-measure on $\mathbb{R}^2$) and ϵ_k the errors. If (x, y) is deterministic on a compact set $A \subset \mathbb{R}^2$, the design is called fixed. It is assumed, that the design is generated by a density f, that means that there is a partition $A_1, ..., A_n$ of A with $(x_k, y_k) \in A_k$, $\sup_k |\lambda(A_k) - \frac{1}{nf(x_k,y_k)}| = o(\frac{1}{n})$ and $\sup_k \sup_{x,y \in A_k} ||x - y|| = O(\frac{1}{\sqrt{n}})$. The design density f is called regular, if f is two times continuously differentiable on the interior of A and bounded away from 0. The errors ϵ_k are independent with mean 0, finite variance $\sigma^2(x_k, y_k)$ and all moments exist. The variance function $\sigma^2(x, y)$ is positive and two times continuously differentiable. Many regression estimators like kernel estimators, smoothing splines, orthogonal series estimators, local polynomials are at least asymptotically localized weighted averages of the data

$$\hat{r}(x, y) = \sum_{k=1}^{n} W_k(x, y) Z_k,$$

and they differ only in the weights W_k. For kernel estimators the weights W_k depend on a dilation and rescaling of a kernel.

Definition 2.1 *A function $K_{\nu_1,\nu_2} : \mathbb{R}^2 \longrightarrow \mathbb{R}$ with*

$$\int K_{\nu_1,\nu_2}(x, y) x^{\alpha_1} y^{\alpha_2} \, dx \, dy = \begin{cases} 0 & 0 \leq \alpha_1 + \alpha_2 < k \text{ and } (\alpha_1, \alpha_2) \neq (\nu_1, \nu_2) \\ & \\ \mu_{\nu_1,\nu_2} & (\alpha_1, \alpha_2) = (\nu_1, \nu_2) \end{cases},$$

where $\mu_{\nu_1,\nu_2} = (-1)^{\nu_1+\nu_2} \nu_1! \nu_2!$, is called kernel (for estimating the (ν_1, ν_2)-th derivative) of order k, $k \geq \nu_1 + \nu_2 + 1$, $\alpha_1, \alpha_2 \in \mathbb{N} \cup \{0\}$, $\nu_1, \nu_2 \in \{0, 1, 2, ...\}$.

In the following a kernel K_{ν_1,ν_2} is always assumed to be Lipschitz continuous and has compact support T, T the unit ball in $\mathbb{R}^2$. Here, only K (short for $K_{0,0}$), $K_{2,0}$, $K_{0,2}$ and $K_{1,1}$ are considered. $K_{0,0}$, $K_{2,0}$ and $K_{0,2}$ are assumed to be symmetric in both directions, i.e. $K_{\nu_1,\nu_2}(x, y) = K_{\nu_1,\nu_2}(-x, y)$ and $K_{\nu_1,\nu_2}(x, y) = K_{\nu_1,\nu_2}(x, -y)$. Moreover, $K(x, y) = K(y, x)$. It follows for $K_{0,0}$, $K_{2,0}$ and $K_{0,2}$, that $\int_T z_1^{\alpha_1} z_2^{\alpha_2} K_{\nu_1,\nu_2}(z_1, z_2) \, dz_1 \, dz_2 = 0$, if at least one integer α_i is odd. Therefore, K is of order $k \geq 2$ and $K_{2,0}$ and $K_{0,2}$ are of order $k \geq 4$. $K_{1,1}$ is assumed to be antisymmetric in both directions, i.e. $K_{1,1}(x,y) = -K_{1,1}(-x,y)$ and $K_{1,1}(x,y) = -K_{1,1}(x,-y)$. It follows, that $\int_T z_1^{\alpha_1} z_2^{\alpha_2} K_{1,1}(z_1,z_2)\, dz_1\, dz_2 = 0$, if at least one integer α_i is even. Later, it is required, that $K_{1,1}$ is of order $k \geq 4$, so it has automatically order $k \geq 5$. For more details on kernels see Müller (1988).

In the bivariate case the dilation is a vector and the rescaling is done by a so called bandwidth matrix B.

The bandwidth matrix can consist of up to three parameters. The sets

$$D_1 = \left\{ \begin{pmatrix} b_1 & 0 \\ 0 & b_1 \end{pmatrix} : b_1 > 0 \right\},$$

$$D = \left\{ \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix} : b_1, b_2 > 0 \right\}$$

and

$$P = \left\{ \begin{pmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{pmatrix} : b_{11}, b_{22} > 0, b_{11}b_{22} > b_{12}^2 \right\}$$

are considered. In D are diagonal matrices and in P symmetric positive definite matrices. If $B \in D$, then $B^{-1}\binom{x}{y}$ means, that the support of $K\left(B^{-1}\binom{x}{y}\right)$ is not the unit ball any more, but an ellipse with axes parallel to the coordinate axes and lengths b_1 and b_2. If the matrix $B \in P$, then the axes of the ellipse are not parallel to the coordinate axes any more, but rotated with an angle α. This geometric interpretation can be formalized by a diagonalization step. Since B is symmetric and positive definite, B can be diagonalized by

$$B = C^t B_d C,$$

where B_d is a diagonal matrix with entries b_1 and b_2 and C is an orthogonal transformation matrix which can be written in the Form

$$C = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

with some $\alpha \in [0, 2\pi]$.

$$b_{11} \quad = \quad b_1 \cos^2\alpha + b_2 \sin^2\alpha$$

$$
\begin{aligned}
b_{12} &= (b_2 - b_1)\sin\alpha\cos\alpha \qquad (1)\\
b_{22} &= b_1 \sin^2\alpha + b_2\cos^2\alpha\\
\det B &= b_1 b_2
\end{aligned}
$$

There are three bandwidth parameters: two axes of the ellipse and a rotation angle. In the following this parametrisation is used.

B may depend on (x, y) (local bandwidth) or may be independent of (x, y) (global bandwidth).

Definition 2.2 *If K_{ν_1,ν_2} is a kernel and B a bandwidth matrix, then kernel estimators of convolution type (to estimate the (ν_1, ν_2)-th derivative) are defined as*

- *for $B \in D$*

$$\hat{r}^{(\nu_1,\nu_2)}(x, y; B) = \frac{1}{b_1^{\nu_1+1} b_2^{\nu_2+1}} \sum_{k=1}^{n} \int_{A_k} K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv\ Z_k,$$

- *for $B \in P$*

$$\hat{r}(x, y; B) = \frac{1}{\det B} \sum_{k=1}^{n} \int_{A_k} K\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv\ Z_k,$$

and if $K_{2,0}$, $K_{0,2}$ and $K_{1,1}$ are the derivatives of the same kernel $\tilde{K}$ (e.g. $K_{2,0}(x, y) = \frac{\partial^2}{\partial x^2}\tilde{K}(x, y)$),

$$
\begin{aligned}
\hat{r}^{(2,0)}(x, y; B) &= \frac{1}{(\det B)^3} \sum_{k=1}^{n} \int_{A_k} b_{22}^2 K_{2,0}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) + b_{12}^2 K_{0,2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right)\\
&\qquad -2b_{12}b_{22}K_{1,1}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv\ Z_k,\\
\hat{r}^{(0,2)}(x, y; B) &= \frac{1}{(\det B)^3} \sum_{k=1}^{n} \int_{A_k} b_{12}^2 K_{2,0}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) + b_{11}^2 K_{0,2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right)\\
&\qquad -2b_{11}b_{12}K_{1,1}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv\ Z_k,\\
\hat{r}^{(1,1)}(x, y; B) &= \frac{1}{(\det B)^3} \sum_{k=1}^{n} \int_{A_k} -b_{12}b_{22} K_{2,0}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) - b_{11}b_{12} K_{0,2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right)\\
&\qquad +(b_{11}b_{22} + b_{12}^2)K_{1,1}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv\ Z_k.
\end{aligned}
$$

Kernel estimators of convolution type have better asymptotical properties concerning the bias and the variance than e.g. the Nadaraya-Watson estimator or the Priestley-Chao estimator. For fixed design they achieve the same rates as local polynomials (see Simonoff (1998)). For finite sample sizes they look smoother then most other estimators.

Multivariate kernel estimators were considered in Müller (1988) and the references therein.

2.2 Convergence of the Kernel Estimator

As criterion for optimality only the mean squared error (MSE) and a weighted version of the integrated mean squared error (MISE) (or the sample MSE and MISE) are considered. But the results of the simulation are also comparable in other distances, especially the visual impression.

$$MSE_{\hat{r}}(x,y;B) := \mathbb{E}(\hat{r}(x,y;B) - r(x,y))^2$$

$$\begin{aligned} MISE_{\hat{r}}(B) &:= \int v(x,y) MSE_{\hat{r}}(x,y;B)\,dx\,dy \\ &= \int v(x,y)\mathbb{E}(\hat{r}(x,y;B) - r(x,y))^2\,dx\,dy. \end{aligned}$$

The function $v(x,y)$ is a nonnegative, continuous weight function with support in A (to weaken boundary effects) and v integrates to one. If A is rectangular (w.l.o.g. $[0,1]^2$), then v can be chosen to have support $[\delta_1, 1-\delta_1] \times [\delta_2, 1-\delta_2] \in [0,1]^2$, $\delta_1, \delta_2 \geq 0$.

The mean squared error consists of two parts, the sqared bias and the variance of the estimator

$$MSE_{\hat{r}}(x,y;B) = [\mathbb{E}\hat{r}(x,y;B) - r(x,y)]^2 + \operatorname{var}\hat{r}(x,y;B).$$

Proposition 2.3 *a) Let r be twice continuously differentiable,* $b_1 + b_2 \longrightarrow 0$, $\sqrt{n}b_1 \longrightarrow \infty$, $\sqrt{n}b_2 \longrightarrow \infty$, *if* $n \longrightarrow \infty$, *and the general assumptions made above, then for any B in D or P (with the proposed parametrisation)*

- *for equidistant design* ($f \equiv 1$) *and constant* σ^2

$$MSE_{\hat{r}}(x,y;B) = AMSE_{\hat{r}}(x,y;B) + o(b_1^4 + b_2^4) + O\left(\frac{1}{n}\right) + O\left(\frac{b_1 + b_2}{n^{\frac{3}{2}} b_1^2 b_2^2}\right),$$

- *for equidistant design* ($f \equiv 1$) *and* σ^2 *non-constant*

$$MSE_{\hat{r}}(x,y;B) = AMSE_{\hat{r}}(x,y;B) + o(b_1^4 + b_2^4) + O\left(\frac{1}{n}\right) + O\left(\frac{b_1 + b_2}{n^{\frac{3}{2}} b_1^2 b_2^2}\right) + O\left(\frac{b_1^2 + b_2^2}{n b_1 b_2}\right),$$

- *for arbitrary fixed design*

$$MSE_{\hat{r}}(x,y;B) \quad = \quad AMSE_{\hat{r}}(x,y;B) + o(b_1^4 + b_2^4) + O\left(\frac{1}{n}\right) + o\left(\frac{1}{nb_1b_2}\right),$$

with asymptotic MSE

$$AMSE_{\hat{r}}(x,y;B) = \frac{\mu^2}{4}(\text{trace}(BHB))^2 + \frac{M}{nb_1b_2}\frac{\sigma^2(x,y)}{f(x,y)},$$

where

$$H = \begin{pmatrix} r^{(2,0)}(x,y) & r^{(1,1)}(x,y) \\ r^{(1,1)}(x,y) & r^{(0,2)}(x,y) \end{pmatrix}$$

is the Hessian of r, $M = \int_T K^2(z_1,z_2)\, dz_1\, dz_2$, *and* $\mu = \int_T K(z_1,z_2) z_1^2\, dz_1\, dz_2 = \int_T K(z_1,z_2) z_2^2\, dz_1\, dz_2$ *constants depending on K only.*

b) Equivalently

$$AMSE_{\hat{r}}(x,y;B) = \frac{\mu^2}{4}(b_1^2 r^{(2,0)}(x,y) + b_2^2 r^{(0,2)}(x,y))^2 + \frac{M}{nb_1b_2}\frac{\sigma^2(x,y)}{f(x,y)}$$

if B is diagonal and

$$\begin{aligned} AMSE_{\hat{r}}(x,y;B) \quad = \quad & \frac{\mu^2}{4}\Big((b_1^2c^2 + b_2^2s^2)r^{(2,0)}(x,y) + (b_1^2s^2 + b_2^2c^2)r^{(0,2)}(x,y) \\ & +2(b_2^2 - b_1^2)sc\, r^{(1,1)}(x,y)\Big)^2 + \frac{M}{nb_1b_2}\frac{\sigma^2(x,y)}{f(x,y)}, \end{aligned}$$

for arbitrary symmetric and positive definite B, where $c = \cos\alpha$ *and* $s = \sin\alpha$.

A lemma before the proof:

Lemma 2.4 *If B is arbitrarily symmetric and positive definite, then*

a)

$$\begin{aligned} & \sum_{k=1}^{n} \int_{A_k} K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right)\, du\, dv\; r(x_k,y_k) \\ & \qquad = \int_A K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right)\, r(u,v)\, du\, dv\; + O\left(\frac{b_1b_2}{\sqrt{n}}\right) \qquad (2) \end{aligned}$$

b) if $\sqrt{n}b_1 \longrightarrow \infty$ and $\sqrt{n}b_2 \longrightarrow \infty$ for $n \longrightarrow \infty$,

for arbitrary design

$$\begin{aligned}
&\sum_{k=1}^{n} \int_{A_k} K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv \int_{A_k} K_{\nu_3,\nu_4}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv\; \sigma^2(x_k,y_k) \\
&= \int_A \frac{\sigma^2(u,v)}{nf(u,v)} K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) K_{\nu_3,\nu_4}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv \\
&\qquad + O\left(\frac{b_1+b_2}{n^{\frac{3}{2}}}\right) + o\left(\frac{b_1 b_2}{n}\right) \qquad (3)
\end{aligned}$$

for equidistant design

$$\begin{aligned}
&\sum_{k=1}^{n} \int_{A_k} K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv \int_{A_k} K_{\nu_3,\nu_4}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv\; \sigma^2(x_k,y_k) \\
&= \int_A \frac{\sigma^2(u,v)}{n} K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) K_{\nu_3,\nu_4}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv + O\left(\frac{b_1+b_2}{n^{\frac{3}{2}}}\right). \quad (4)
\end{aligned}$$

Proof: First the proof of a)

$$\begin{aligned}
&\sum_{k=1}^{n} \int_{A_k} K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv\, r(x_k,y_k) \\
&= \sum_{k=1}^{n} \int_{A_k} K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) (r(x_k,y_k) + r(u,v) - r(u,v))\, du\, dv \\
&= \sum_{k=1}^{n} \int_{A_k} K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) r(u,v)\, du\, dv \\
&\quad + \sum_{k=1}^{n} \int_{A_k} K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) (r(x_k,y_k) - r(u,v))\, du\, dv \\
&= \int_A K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) r(u,v)\, du\, dv + O\left(\frac{b_1 b_2}{\sqrt{n}}\right),
\end{aligned}$$

because

$$\begin{aligned}
&\left|\sum_{k=1}^{n} \int_{A_k} K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) (r(x_k,y_k) - r(u,v))\, du\, dv\right| \\
&\leq \max_{(x,y)\in T} |K_{\nu_1,\nu_2}(x,y)| \sum_{k=1}^{n} \int_{\tilde{A}_k} |r(x_k,y_k) - r(u,v)|\, du\, dv \\
&= \max_{(x,y)\in T} |K_{\nu_1,\nu_2}(x,y)| \sum_{k=1}^{n} \lambda(\tilde{A}_k) |r(x_k,y_k) - r(\zeta_k,\xi_k)| \\
&= O(b_1 b_2)\, O\left(\frac{1}{\sqrt{n}}\right) \\
&= O\left(\frac{b_1 b_2}{\sqrt{n}}\right),
\end{aligned}$$

with $\tilde{A}_k = A_k \cap \{(x,y) \in A : K_{\nu_1,\nu_2}\left(B^{-1}\binom{x-u}{y-v}\right) \neq 0, (u,v) \in A_k\}$, $\sum_{k=1}^n \lambda(\tilde{A}_k) = \lambda(A \cap \{(x,y) \in A : K_{\nu_1,\nu_2}\left(B^{-1}\binom{x-u}{y-v}\right) \neq 0, (u,v) \in A\}) = O(b_1 b_2)$ and according to the mean value theorem, there exist such $(\zeta_k, \xi_k) \in \tilde{A}_k$. r is Lipschitz continuous and by the definition of A_k

$$\begin{aligned} |r(x_k,y_k) - r(\zeta_k,\xi_k)| &\leq L||(x_k,y_k) - (\zeta_k,\xi_k)|| \\ &= O\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

The proof of b)

According to the mean value theorem there exist $(u_k, v_k) \in A_k$ such that

$$\begin{aligned} &\sum_{k=1}^n \int_{A_k} K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv \int_{A_k} K_{\nu_3,\nu_4}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv\, \sigma^2(x_k,y_k) \\ &= \sum_{k=1}^n \sigma^2(x_k,y_k)\lambda(A_k) K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u_k \\ y-v_k \end{pmatrix}\right) \int_{A_k} K_{\nu_3,\nu_4}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv \\ &= \sum_{k=1}^n \sigma^2(x_k,y_k)\lambda(A_k) \int_{A_k} K_{\nu_3,\nu_4}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) \\ &\quad \cdot\left[K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) + K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u_k \\ y-v_k \end{pmatrix}\right) - K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right)\right] du\, dv \end{aligned}$$

K_{ν_1,ν_2} is Lipschitz continuous, $B^{-1} = C^t B_d^{-1} C$ (see equation (1)), the euclidian norm is not changed by orthonormal transformations and all norms are eqivalent in $\mathbb{R}^2$, therefore for all $(u,v) \in A_k$

$$\begin{aligned} &\left| K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u_k \\ y-v_k \end{pmatrix}\right) - K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) \right| \\ &\leq L\left\| B^{-1}\begin{pmatrix} x-u_k-(x-u) \\ y-v_k-(y-v) \end{pmatrix} \right\| \\ &= L\left\| \begin{pmatrix} \frac{u-u_k}{b_1} \\ \frac{v-v_k}{b_2} \end{pmatrix} \right\| \\ &= O\left(\frac{1}{\sqrt{n}}\left(\frac{1}{b_1} + \frac{1}{b_2}\right)\right) \end{aligned}$$

$\lambda(A_k) = \frac{1}{nf(x_k,y_k)} + o(\frac{1}{n}) = \frac{1}{nf(x_k,y_k)}(1+o(1))$ and for equidistant design $\lambda(A_k) = \frac{1}{n}$ and $f \equiv 1$.

$$\begin{aligned} &\sum_{k=1}^n \frac{\sigma^2(x_k,y_k)}{nf(x_k,y_k)}(1+o(1)) \int_{A_k} K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) \\ &\quad \cdot K_{\nu_3,\nu_4}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv \cdot \left(1 + O\left(\frac{1}{\sqrt{n}}\left(\frac{1}{b_1} + \frac{1}{b_2}\right)\right)\right) \\ &= \sum_{k=1}^n \int_{A_k} K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) K_{\nu_3,\nu_4}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv \end{aligned}$$

$$
\begin{aligned}
&\qquad \cdot\frac{\sigma^2(x_k,y_k)}{nf(x_k,y_k)}\left(1+O\left(\frac{1}{\sqrt{n}}\left(\frac{1}{b_1}+\frac{1}{b_2}\right)\right)+o(1)\right)\\
&= \frac{1}{n}\sum_{k=1}^{n}\int_{A_k}\left[\frac{\sigma^2(u,v)}{f(u,v)}-\frac{\sigma^2(u,v)}{f(u,v)}+\frac{\sigma^2(x_k,y_k)}{f(x_k,y_k)}\right]\\
&\qquad \cdot K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix}x-u\\y-v\end{pmatrix}\right)K_{\nu_3,\nu_4}\left(B^{-1}\begin{pmatrix}x-u\\y-v\end{pmatrix}\right)du\,dv\\
&\quad \cdot\left(1+O\left(\frac{1}{\sqrt{n}}\left(\frac{1}{b_1}+\frac{1}{b_2}\right)\right)+o(1)\right)
\end{aligned}
$$

For equidistant design the term $o(1)$ is not necessary.

σ^2 and f are Lipschitz continuous, and therefore $\frac{\sigma^2}{f}$ is also Lipschitz continuous, and therefore (similar to above calculations) for all $(u,v)\in A_k$

$$
\left|\frac{\sigma^2(x_k,y_k)}{f(x_k,y_k)}-\frac{\sigma^2(u,v)}{f(u,v)}\right| = O\left(\frac{1}{\sqrt{n}}\right)
$$

$O(\frac{1}{\sqrt{n}})$ converges faster than $O\left(\frac{1}{\sqrt{n}}\left(\frac{1}{b_1}+\frac{1}{b_2}\right)\right)$ and $o(1)$, therefore

$$
\begin{aligned}
&\sum_{k=1}^{n}\int_{A_k}K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix}x-u\\y-v\end{pmatrix}\right)K_{\nu_3,\nu_4}\left(B^{-1}\begin{pmatrix}x-u\\y-v\end{pmatrix}\right)du\,dv\\
&\qquad \cdot\frac{\sigma^2(x_k,y_k)}{nf(x_k,y_k)}\left(1+O\left(\frac{1}{\sqrt{n}}\left(\frac{1}{b_1}+\frac{1}{b_2}\right)\right)+o(1)\right)\\
&= \frac{1}{n}\int_{A}K_{\nu_1,\nu_2}\left(B^{-1}\begin{pmatrix}x-u\\y-v\end{pmatrix}\right)K_{\nu_3,\nu_4}\left(B^{-1}\begin{pmatrix}x-u\\y-v\end{pmatrix}\right)\\
&\qquad \cdot\frac{\sigma^2(u,v)}{f(u,v)}du\,dv+O\left(\frac{1}{\sqrt{n}}\left(\frac{1}{b_1}+\frac{1}{b_2}\right)\cdot\frac{b_1b_2}{n}\right)+o\left(\frac{b_1b_2}{n}\right)
\end{aligned}
$$

■

The proof of the Proposition:

Proof: According to Lemma 2.4 (2), because r is Lipschitz continuous, there is

$$
\begin{aligned}
\mathbb{E}\hat{r}(x,y;B) &= \frac{1}{b_1b_2}\sum_{k=1}^{n}\int_{A_k}K\left(B^{-1}\begin{pmatrix}x-u\\y-v\end{pmatrix}\right)du\,dv\,r(x_k,y_k)\\
&= \frac{1}{b_1b_2}\int_{A}K\left(B^{-1}\begin{pmatrix}x-u\\y-v\end{pmatrix}\right)r(u,v)\,du\,dv+O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

$b_1,b_2\geq\frac{1}{2\sqrt{n}}$, because for bandwidths that small the regression is almost interpolation and there is almost no smoothing (exact interpolation for equidistant, quadratic design). To avoid boundary effects, b_1 and b_2 are chosen small enough, that supp $K\left(B^{-1}\binom{x-u}{y-v}\right)\subset A$. After transforming the coordinates the region for integration is T.

$$\mathbb{E}\hat{r}(x,y;B) \;=\; \int_T K(z_1,z_2) r\left(\begin{pmatrix} x \\ y \end{pmatrix} - B\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}\right) dz_1\, dz_2 + O\left(\frac{1}{\sqrt{n}}\right)$$

After a Taylor expansion

$$\begin{aligned}
& r\left(\begin{pmatrix} x \\ y \end{pmatrix} - B\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}\right) \\
= \;& r(x,y) + \left\langle -B\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \begin{pmatrix} r^{(1,0)}(x,y) \\ r^{(0,1)}(x,y) \end{pmatrix} \right\rangle + \frac{1}{2}\left\langle B\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, HB\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right\rangle + o(b_{11}^2 + b_{12}^2 + b_{22}^2) \\
= \;& r(x,y) - r^{(1,0)}(x,y)(b_{11}z_1 + b_{12}z_2) - r^{(0,1)}(x,y)(b_{12}z_1 + b_{22}z_2) \\
& +\frac{1}{2}\Big[r^{(2,0)}(x,y)(b_{11}z_1 + b_{12}z_2)^2 + r^{(0,2)}(x,y)(b_{12}z_1 + b_{22}z_2)^2 \\
& \quad +2r^{(1,1)}(x,y)(b_{11}z_1 + b_{12}z_2)(b_{12}z_1 + b_{22}z_2)\Big] + o(b_{11}^2 + b_{12}^2 + b_{22}^2),
\end{aligned}$$

$$\begin{aligned}
& \mathbb{E}\hat{r}(x,y;B) \\
= \;& r(x,y)\int_T K(z_1,z_2)\, dz_1\, dz_2 && (5) \\
& -r^{(1,0)}(x,y)\Big(b_{11}\int_T z_1 K(z_1,z_2)\, dz_1\, dz_2 + b_{12}\int_T z_2 K(z_1,z_2)\, dz_1\, dz_2\Big) && (6) \\
& -r^{(0,1)}(x,y)\Big(b_{12}\int_T z_1 K(z_1,z_2)\, dz_1\, dz_2 + b_{22}\int_T z_2 K(z_1,z_2)\, dz_1\, dz_2\Big) && (7) \\
& +\frac{1}{2}r^{(2,0)}(x,y)\Big(b_{11}^2\int_T z_1^2 K(z_1,z_2)\, dz_1\, dz_2 + 2b_{11}b_{12}\int_T z_1 z_2 K(z_1,z_2)\, dz_1\, dz_2 \\
& \qquad +b_{12}^2\int_T z_2^2 K(z_1,z_2)\, dz_1\, dz_2\Big) && (8) \\
& +\frac{1}{2}r^{(0,2)}(x,y)\Big(b_{12}^2\int_T z_1^2 K(z_1,z_2)\, dz_1\, dz_2 + 2b_{12}b_{22}\int_T z_1 z_2 K(z_1,z_2)\, dz_1\, dz_2 \\
& \qquad +b_{22}^2\int_T z_2^2 K(z_1,z_2)\, dz_1\, dz_2\Big) && (9) \\
& +r^{(1,1)}(x,y)\Big(b_{11}b_{12}\int_T z_1^2 K(z_1,z_2)\, dz_1\, dz_2 + b_{12}b_{22}\int_T z_2^2 K(z_1,z_2)\, dz_1\, dz_2 \\
& \qquad +(b_{11}b_{22} + b_{12}^2)\int_T z_1 z_2 K(z_1,z_2)\, dz_1\, dz_2\Big) && (10) \\
& +o(b_{11}^2 + b_{12}^2 + b_{22}^2) && (11) \\
& +O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Because of the assumptions on K, (5) is $r(x,y)$, (6) and (7) are 0, (8) is $\frac{\mu}{2}(b_{11}^2 + b_{12}^2)r^{(2,0)}(x,y) = \frac{\mu}{2}(b_1^2c^2 + b_2^2s^2)r^{(2,0)}(x,y)$, (9) is $\frac{\mu}{2}(b_{12}^2 + b_{22}^2)r^{(0,2)}(x,y) = \frac{\mu}{2}(b_1^2s^2 + b_2^2c^2)r^{(0,2)}(x,y)$, (10) is $\mu(b_{11}b_{12} + b_{12}b_{22})r^{(1,1)}(x,y) = \mu(b_2^2 - b_1^2)sc\, r^{(1,1)}(x,y)$ and (11) is $o(b_1^2 + b_2^2)$.

If B is a diagonal matrix, then

$$\mathbb{E}\hat{r}(x,y;B) - r(x,y) = \frac{\mu}{2}\left(b_1^2 r^{(2,0)}(x,y) + b_2^2 r^{(0,2)}(x,y)\right) + o(b_1^2 + b_2^2) + O\left(\frac{1}{\sqrt{n}}\right),$$

$$\left[\mathbb{E}\hat{r}(x,y;B) - r(x,y)\right]^2 = \frac{\mu^2}{4}\left(b_1^2 r^{(2,0)}(x,y) + b_2^2 r^{(0,2)}(x,y)\right)^2 + o(b_1^4 + b_2^4) + O\left(\frac{1}{n}\right).$$

The term $O\left(\frac{b_1^2+b_2^2}{\sqrt{n}}\right)$ is not necessary, because one of the other terms is always dominating.

If B is arbitrarily symmetric and positive definite, then

$$\begin{aligned}&\left[\mathbb{E}\hat{r}(x,y;B) - r(x,y)\right]^2\\ &= \frac{\mu^2}{4}\left((b_1^2c^2 + b_2^2s^2)r^{(2,0)}(x,y) + (b_1^2s^2 + b_2^2c^2)r^{(0,2)}(x,y) + 2(b_2^2 - b_1^2)sc\,r^{(1,1)}(x,y)\right)^2\\ &\quad + o(b_1^4 + b_2^4) + O\left(\frac{1}{n}\right).\end{aligned}$$

Together,

$$\left[\mathbb{E}\hat{r}(x,y;B) - r(x,y)\right]^2 = \frac{\mu^2}{4}(\text{trace}(BHB))^2 + o(b_1^4 + b_2^4) + O\left(\frac{1}{n}\right)$$

According to Lemma 2.4 (3)

$$\begin{aligned}&\operatorname{var}\hat{r}(x,y;B)\\ &= \frac{1}{b_1^2b_2^2}\sum_{k=1}^{n}\left[\int_{A_k} K\left(B^{-1}\begin{pmatrix}x-u\\ y-v\end{pmatrix}\right)du\,dv\right]^2\sigma^2(x_k,y_k)\\ &= \frac{1}{nb_1^2b_2^2}\int_A \frac{\sigma^2(u,v)}{f(u,v)}K^2\left(B^{-1}\begin{pmatrix}x-u\\ y-v\end{pmatrix}\right)du\,dv + O\left(\frac{b_1+b_2}{n^{\frac{3}{2}}b_1^2b_2^2}\right) + o\left(\frac{b_1b_2}{nb_1^2b_2^2}\right)\\ &= \frac{1}{nb_1b_2}\int_T \frac{\sigma^2\left(\binom{x}{y} - B\binom{z_1}{z_2}\right)}{f\left(\binom{x}{y} - B\binom{z_1}{z_2}\right)}K^2(z_1,z_2)\,dz_1\,dz_2 + O\left(\frac{b_1+b_2}{n^{\frac{3}{2}}b_1^2b_2^2}\right) + o\left(\frac{1}{nb_1b_2}\right).\end{aligned}$$

For equidistant design the term $o(\frac{1}{nb_1b_2})$ is not necessary (Lemma 2.4 (4)), otherwise $o\left(\frac{1}{nb_1b_2}\right)$ is dominating.

If the design is equidistant and σ^2 constant, then

$$\operatorname{var}\hat{r}(x,y;B) = \frac{M\sigma^2}{nb_1b_2} + O\left(\frac{b_1+b_2}{n^{\frac{3}{2}}b_1^2b_2^2}\right).$$

If not, then, after a Taylor expansion,

$$\begin{aligned}
&\frac{\sigma^2\left(\binom{x}{y}-B\binom{z_1}{z_2}\right)}{f\left(\binom{x}{y}-B\binom{z_1}{z_2}\right)}\\
&= \frac{\sigma^2(x,y)}{f(x,y)} - (b_{11}z_1+b_{12}z_2)\frac{f\cdot(\sigma^2)^{(1,0)}-\sigma^2 f^{(1,0)}}{f^2}(x,y)\\
&\quad -(b_{12}z_1+b_{22}z_2)\frac{f\cdot(\sigma^2)^{(0,1)}-\sigma^2 f^{(0,1)}}{f^2}(x,y)\\
&\quad +\frac{1}{2}[(b_{11}z_1+b_{12}z_2)^2V_1(\zeta,\xi)+(b_{12}z_1+b_{22}z_2)^2V_2(\zeta,\xi)\\
&\quad +2(b_{11}z_1+b_{12}z_2)(b_{12}z_1+b_{22}z_2)V_3(\zeta,\xi)],
\end{aligned}$$

with V_1, V_2, V_3 bounded functions of f, σ^2 and their derivatives, (ζ,ξ) according to the mean value theorem (for a simpler notation the same symbol is used for possibly different values), then

$$\begin{aligned}
&\operatorname{var}\hat{r}(x,y;B)\\
&= \frac{1}{nb_1b_2}\frac{\sigma^2(x,y)}{f(x,y)}\int_T K^2(z_1,z_2)\,dz_1\,dz_2 && (12)\\
&\quad -\frac{f\cdot(\sigma^2)^{(1,0)}-\sigma^2 f^{(1,0)}}{nb_1b_2f^2}(x,y)\Big(b_{11}\int_T z_1K^2(z_1,z_2)\,dz_1\,dz_2\\
&\qquad\qquad +b_{12}\int_T z_2K^2(z_1,z_2)\,dz_1\,dz_2\Big) && (13)\\
&\quad -\frac{f\cdot(\sigma^2)^{(0,1)}-\sigma^2 f^{(0,1)}}{nb_1b_2f^2}(x,y)\Big(b_{12}\int_T z_1K^2(z_1,z_2)\,dz_1\,dz_2\\
&\qquad\qquad +b_{22}\int_T z_2K^2(z_1,z_2)\,dz_1\,dz_2\Big) && (14)\\
&\quad +\frac{1}{2nb_1b_2}\Big(b_{11}^2V_1(\zeta,\xi)+b_{12}^2V_2(\zeta,\xi)+2b_{11}b_{12}V_3(\zeta,\xi)\Big)\int_T z_1^2K^2(z_1,z_2)\,dz_1\,dz_2 && (15)\\
&\quad +\frac{1}{2nb_1b_2}\Big(b_{12}^2V_1(\zeta,\xi)+b_{22}^2V_2(\zeta,\xi)+2b_{12}b_{22}V_3(\zeta,\xi)\Big)\int_T z_2^2K^2(z_1,z_2)\,dz_1\,dz_2 && (16)\\
&\quad +\frac{1}{nb_1b_2}\Big(b_{11}b_{12}V_1(\zeta,\xi)+b_{12}b_{22}V_2(\zeta,\xi)+(b_{11}b_{22}+b_{12}^2)V_3(\zeta,\xi)\Big)\\
&\qquad\qquad \cdot\int_T z_1z_2K^2(z_1,z_2)\,dz_1\,dz_2 && (17)\\
&\quad +O\left(\frac{b_1+b_2}{n^{\frac{3}{2}}b_1^2b_2^2}\right)+o\left(\frac{1}{nb_1b_2}\right)
\end{aligned}$$

(12) is $\frac{M}{nb_1b_2}\frac{\sigma^2(x,y)}{f(x,y)}$, (13) and (14) are 0, because K^2 is symmetric, (15) is $O(\frac{b_{11}^2+b_{12}^2}{nb_1b_2})$, (16) is $O(\frac{b_{12}^2+b_{22}^2}{nb_1b_2})$, and (17) is 0, because K^2 is symmetric. $O(\frac{b_{11}^2+b_{12}^2+b_{22}^2}{nb_1b_2}) = O(\frac{b_1^2+b_2^2}{nb_1b_2})$ is smaller than $o(\frac{1}{nb_1b_2})$.

For equidistant design with non-constant σ^2 there is no term $o(\frac{1}{nb_1b_2})$. Therefore, there is the additional term $O(\frac{b_1^2+b_2^2}{nb_1b_2})$.

The result for arbitrary fixed design is

$$\operatorname{var}\hat{r}(x,y;B) = \frac{M}{nb_1b_2}\frac{\sigma^2(x,y)}{f(x,y)} + o\left(\frac{1}{nb_1b_2}\right).$$

If the design is equidistant, then

$$\operatorname{var}\hat{r}(x,y;B) = \frac{M\sigma^2(x,y)}{nb_1b_2} + O\left(\frac{b_1+b_2}{n^{\frac{3}{2}}b_1^2b_2^2}\right) + O\left(\frac{b_1^2+b_2^2}{nb_1b_2}\right),$$

$$MSE_{\hat{r}}(x,y;B) = [\mathbb{E}\hat{r}(x,y;B) - r(x,y)]^2 + \operatorname{var}\hat{r}(x,y;B).$$

■

Notation 2.5

$$I_{\sigma,f} = \int v(x,y)\frac{\sigma^2(x,y)}{f(x,y)}\,dx\,dy,$$

$$I_{i,j}^{k,l} = \int v(x,y)r^{(i,j)}(x,y)r^{(k,l)}(x,y)\,dx\,dy,$$

for $i,j,k,l \in \{0,1,2\}$.

Proposition 2.6 *Let r be twice continuously differentiable, $b_1+b_2 \longrightarrow 0$, $\sqrt{n}b_1 \longrightarrow \infty$, $\sqrt{n}b_2 \longrightarrow \infty$, if $n \longrightarrow \infty$, and the general assumptions made in section 2.1, then for any B in D or P (with the proposed parametrisation)*

- *for equidistant design ($f \equiv 1$) and constant σ^2*

$$MISE_{\hat{r}}(B) = AMISE_{\hat{r}}(B) + o(b_1^4+b_2^4) + O\left(\frac{1}{n}\right) + O\left(\frac{b_1+b_2}{n^{\frac{3}{2}}b_1^2b_2^2}\right),$$

- *for equidistant design ($f \equiv 1$) and σ^2 non-constant*

$$MISE_{\hat{r}}(B) = AMISE_{\hat{r}}(B) + o(b_1^4+b_2^4) + O\left(\frac{1}{n}\right) + O\left(\frac{b_1+b_2}{n^{\frac{3}{2}}b_1^2b_2^2}\right) + O\left(\frac{b_1^2+b_2^2}{nb_1b_2}\right),$$

- *for arbitrary fixed design*

$$MISE_{\hat{r}}(B) = AMISE_{\hat{r}}(B) + o(b_1^4+b_2^4) + O\left(\frac{1}{n}\right) + o\left(\frac{1}{nb_1b_2}\right),$$

with

$$AMISE_{\hat{r}}(B) = \frac{\mu^2}{4}\left[b_1^4 I_{2,0}^{2,0} + b_2^4 I_{0,2}^{0,2} + 2b_1^2 b_2^2 I_{2,0}^{0,2}\right] + \frac{M}{nb_1 b_2} I_{\sigma,f}$$

if B is diagonal and with

$$\begin{aligned} AMISE_{\hat{r}}(B) &= \frac{\mu^2}{4}\Big[(b_1^2c^2 + b_2^2s^2)^2 I_{2,0}^{2,0} + (b_1^2s^2 + b_2^2c^2)^2 I_{0,2}^{0,2} + 2(b_1^2c^2 + b_2^2s^2)(b_1^2s^2 + b_2^2c^2) I_{2,0}^{0,2} \\ &\quad +4(b_2^2 - b_1^2)^2 s^2c^2 I_{1,1}^{1,1} + 4(b_1^2c^2 + b_2^2s^2)(b_2^2 - b_1^2)sc I_{2,0}^{1,1} \\ &\quad +4(b_1^2s^2 + b_2^2c^2)(b_2^2 - b_1^2)sc I_{0,2}^{1,1}\Big] + \frac{M}{nb_1 b_2} I_{\sigma,f} \end{aligned}$$

for arbitrary symmetric and positive definite B, with $c = \cos\alpha$, $s = \sin\alpha$, $\mu = \int_T K(z_1,z_2) z_1^2\, dz_1\, dz_2$ *and* $M = \int_T K^2(z_1,z_2)\, dz_1\, dz_2$.

Proof: B is global, $\int v(x,y)\, dx\, dy = 1$ and the results of Proposition 2.3 are uniformly in (x,y), therefore integration changes only $AMSE_{\hat{r}}(x,y;B)$ in Proposition 2.3.

For B diagonal

$$\begin{aligned} AMISE_{\hat{r}}(B) &:= \int v(x,y) AMSE_{\hat{r}}(x,y;B)\, dx\, dy \\ &= \frac{\mu^2}{4}\int v(x,y)(b_1^2 r^{(2,0)}(x,y) + b_2^2 r^{(0,2)}(x,y))^2\, dx\, dy + \frac{M}{nb_1b_2} I_{\sigma,f} \\ &= \frac{\mu^2}{4}\left[b_1^4 I_{2,0}^{2,0} + b_2^4 I_{0,2}^{0,2} + 2b_1^2 b_2^2 I_{2,0}^{0,2}\right] + \frac{M}{nb_1 b_2} I_{\sigma,f} \end{aligned}$$

For arbitrary B

$$\begin{aligned} AMISE_{\hat{r}}(B) &:= \int v(x,y) AMSE_{\hat{r}}(x,y;B)\, dx\, dy \\ &= \frac{\mu^2}{4}\int v(x,y)\Big((b_1^2c^2 + b_2^2s^2) r^{(2,0)}(x,y) + (b_1^2s^2 + b_2^2c^2) r^{(0,2)}(x,y) \\ &\quad +2(b_2^2 - b_1^2)sc\, r^{(1,1)}(x,y)\Big)^2\, dx\, dy + \frac{M}{nb_1b_2} I_{\sigma,f} \\ &= \frac{\mu^2}{4}\Big[(b_1^2c^2 + b_2^2s^2)^2 I_{2,0}^{2,0} + 2(b_1^2c^2 + b_2^2s^2)(b_1^2s^2 + b_2^2c^2) I_{2,0}^{0,2} \\ &\quad +4(b_2^2 - b_1^2)^2 s^2c^2 I_{1,1}^{1,1} + (b_1^2s^2 + b_2^2c^2)^2 I_{0,2}^{0,2} + 4(b_1^2c^2 + b_2^2s^2)(b_2^2 - b_1^2)sc I_{2,0}^{1,1} \\ &\quad +4(b_1^2s^2 + b_2^2c^2)(b_2^2 - b_1^2)sc I_{0,2}^{1,1}\Big] + \frac{M}{nb_1b_2} I_{\sigma,f} \end{aligned}$$

■

Proposition 2.7 *If r is four times continuously differentiable, the rate of convergence in the bias part of the MSE and MISE in Proposition 2.3 and 2.6 is $O(b_1^6+b_2^6)+O\left(\frac{b_1^2+b_2^2}{\sqrt{n}}\right)$ instead of $o(b_1^4+b_2^4)$.*

Proof: The Taylor expansion in the proof of Proposition 2.3 can be written like

$$\begin{aligned}
r\left(\begin{pmatrix}x\\y\end{pmatrix}-B\begin{pmatrix}z_1\\z_2\end{pmatrix}\right) &= r(x,y)-r^{(1,0)}(x,y)(b_{11}z_1+b_{12}z_2)-r^{(0,1)}(x,y)(b_{12}z_1+b_{22}z_2)\\
&+\frac{1}{2}\Big[r^{(2,0)}(x,y)(b_{11}z_1+b_{12}z_2)^2+r^{(0,2)}(x,y)(b_{12}z_1+b_{22}z_2)^2\\
&\quad +2r^{(1,1)}(x,y)(b_{11}z_1+b_{12}z_2)(b_{12}z_1+b_{22}z_2)\Big]\\
&-\frac{1}{6}\left[\left((b_{11}z_1+b_{12}z_2)\frac{\partial}{\partial x}+(b_{12}z_1+b_{22}z_2)\frac{\partial}{\partial y}\right)^3 r(x,y)\right]\\
&+\frac{1}{24}\left[\left((b_{11}z_1+b_{12}z_2)\frac{\partial}{\partial x}+(b_{12}z_1+b_{22}z_2)\frac{\partial}{\partial y}\right)^4 r(\zeta,\xi)\right],
\end{aligned}$$

with $(\zeta,\xi)\in A$,

$$\begin{aligned}
&\mathbb{E}\hat{r}(x,y;B)\\
&= r(x,y)\int_T K(z_1,z_2)\,dz_1\,dz_2 &(18)\\
&\quad -r^{(1,0)}(x,y)\Big(b_{11}\int_T z_1K(z_1,z_2)\,dz_1\,dz_2+b_{12}\int_T z_2K(z_1,z_2)\,dz_1\,dz_2\Big) &(19)\\
&\quad -r^{(0,1)}(x,y)\Big(b_{12}\int_T z_1K(z_1,z_2)\,dz_1\,dz_2+b_{22}\int_T z_2K(z_1,z_2)\,dz_1\,dz_2\Big) &(20)\\
&\quad +\frac{1}{2}r^{(2,0)}(x,y)\Big(b_{11}^2\int_T z_1^2K(z_1,z_2)\,dz_1\,dz_2+2b_{11}b_{12}\int_T z_1z_2K(z_1,z_2)\,dz_1\,dz_2\\
&\qquad +b_{12}^2\int_T z_2^2K(z_1,z_2)\,dz_1\,dz_2\Big) &(21)\\
&\quad +\frac{1}{2}r^{(0,2)}(x,y)\Big(b_{12}^2\int_T z_1^2K(z_1,z_2)\,dz_1\,dz_2+2b_{12}b_{22}\int_T z_1z_2K(z_1,z_2)\,dz_1\,dz_2\\
&\qquad +b_{22}^2\int_T z_2^2K(z_1,z_2)\,dz_1\,dz_2\Big) &(22)\\
&\quad +r^{(1,1)}(x,y)\Big(b_{11}b_{12}\int_T z_1^2K(z_1,z_2)\,dz_1\,dz_2+b_{12}b_{22}\int_T z_2^2K(z_1,z_2)\,dz_1\,dz_2\\
&\qquad +(b_{11}b_{22}+b_{12}^2)\int_T z_1z_2K(z_1,z_2)\,dz_1\,dz_2\Big) &(23)\\
&\quad -\frac{1}{6}\int_T K(z_1,z_2)\Big((b_{11}z_1+b_{12}z_2)\frac{\partial}{\partial x}+(b_{12}z_1+b_{22}z_2)\frac{\partial}{\partial y}\Big)^3 r(x,y)\,dz_1\,dz_2 &(24)\\
&\quad +\frac{1}{24}\int_T K(z_1,z_2)\Big((b_{11}z_1+b_{12}z_2)\frac{\partial}{\partial x}+(b_{12}z_1+b_{22}z_2)\frac{\partial}{\partial y}\Big)^4 r(\zeta,\xi)\,dz_1\,dz_2 &(25)\\
&\quad +O(\frac{1}{\sqrt{n}}).
\end{aligned}$$

Because of the assumptions on K, (18) is $r(x,y)$, (19) and (20) are 0, (21) is $\frac{\mu}{2}(b_{11}^2+b_{12}^2)r^{(2,0)}(x,y)=$

$\frac{\mu}{2}(b_1^2c^2 + b_2^2s^2)r^{(2,0)}(x,y)$, (22) is $\frac{\mu}{2}(b_{12}^2 + b_{22}^2)r^{(0,2)}(x,y) = \frac{\mu}{2}(b_1^2s^2 + b_2^2c^2)r^{(0,2)}(x,y)$, (23) is $\mu(b_{11}b_{12} + b_{12}b_{22})r^{(1,1)}(x,y) = \mu(b_2^2 - b_1^2)sc\,r^{(1,1)}(x,y)$, (24) is 0 and (25) is $O(b_1^4 + b_2^4)$.

$$\left[\mathbb{E}\hat{r}(x,y;B) - r(x,y)\right]^2 = AMSE_{\hat{r}}(x,y;B) + O(b_1^6 + b_2^6) + O\left(\frac{b_1^2 + b_2^2}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right).$$

■

2.3 Optimal Bandwidth Parameters

The leading terms in $AMSE$ and $AMISE$ consist of the squared bias and the variance. The squared bias is increasing and the variance is decreasing in b_1 and b_2. The minimum of AMSE is the balance of the squared bias and the variance.

Proposition 2.8 *$AMSE_{\hat{r}}(x,y;B)$ is minimized by the following bandwidth parameters*

- *if B is diagonal and $r^{(2,0)}(x,y)r^{(0,2)}(x,y) > 0$*

$$b_{1_{ASY}} = \left(\frac{M\sigma^2(x,y)}{2n\mu^2 f(x,y)}\right)^{\frac{1}{6}} \frac{|r^{(0,2)}(x,y)|^{\frac{1}{12}}}{|r^{(2,0)}(x,y)|^{\frac{5}{12}}} \qquad (26)$$

$$b_{2_{ASY}} = \left(\frac{M\sigma^2(x,y)}{2n\mu^2 f(x,y)}\right)^{\frac{1}{6}} \frac{|r^{(2,0)}(x,y)|^{\frac{1}{12}}}{|r^{(0,2)}(x,y)|^{\frac{5}{12}}} \qquad (27)$$

- *if B is arbitrary and $uz > 0$*

$$\alpha_{ASY} = \begin{cases} 0 & \text{if } r^{(2,0)}(x,y) - r^{(0,2)}(x,y) = 0 \\ \frac{1}{2}\arctan\frac{2r^{(1,1)}(x,y)}{r^{(0,2)}(x,y) - r^{(2,0)}(x,y)} & \text{otherwise} \end{cases}$$

$$b_{1_{ASY}} = \left(\frac{M\sigma^2(x,y)}{2n\mu^2 f(x,y)}\right)^{\frac{1}{6}} \frac{|z|^{\frac{1}{12}}}{|u|^{\frac{5}{12}}}$$

$$b_{2_{ASY}} = \left(\frac{M\sigma^2(x,y)}{2n\mu^2 f(x,y)}\right)^{\frac{1}{6}} \frac{|u|^{\frac{1}{12}}}{|z|^{\frac{5}{12}}},$$

with

$$u = r^{(2,0)}(x,y)c^2 + r^{(0,2)}(x,y)s^2 - 2r^{(1,1)}(x,y)sc,$$

$$z = r^{(2,0)}(x,y)s^2 + r^{(0,2)}(x,y)c^2 + 2r^{(1,1)}(x,y)sc,$$

where $s = \sin\alpha$ *and* $c = \cos\alpha$.

Let $I_{2,0}^{2,0} \neq 0$, $I_{0,2}^{0,2} \neq 0$ *and* $\sqrt{I_{2,0}^{2,0} I_{0,2}^{0,2}} + I_{2,0}^{0,2} \neq 0$. *The global bandwidth parameters which minimizes* $AMISE_{\hat{r}}(B)$ *are, if B is diagonal,*

$$b_{1_{IASY}} = \left(\frac{M}{n\mu^2} I_{\sigma,f}\right)^{\frac{1}{6}} \left(\frac{I_{0,2}^{0,2}}{I_{2,0}^{2,0}}\right)^{\frac{1}{8}} \left(\frac{1}{\sqrt{I_{2,0}^{2,0} I_{0,2}^{0,2}} + I_{2,0}^{0,2}}\right)^{\frac{1}{6}} \tag{28}$$

$$b_{2_{IASY}} = \left(\frac{M}{n\mu^2} I_{\sigma,f}\right)^{\frac{1}{6}} \left(\frac{I_{2,0}^{2,0}}{I_{0,2}^{0,2}}\right)^{\frac{1}{8}} \left(\frac{1}{\sqrt{I_{2,0}^{2,0} I_{0,2}^{0,2}} + I_{2,0}^{0,2}}\right)^{\frac{1}{6}} \tag{29}$$

Proof: If B is diagonal, then in the minimum

$$\begin{aligned} &\frac{\partial}{\partial b_1} AMSE_{\hat{r}}(x, y; B) = 0 \\ &= \mu^2 (b_1^2 r^{(2,0)}(x,y) + b_2^2 r^{(0,2)}(x,y)) b_1 r^{(2,0)}(x,y) - \frac{M\sigma^2(x,y)}{b_1^2 b_2 n f(x,y)} \end{aligned} \tag{30}$$

and

$$\begin{aligned} &\frac{\partial}{\partial b_2} AMSE_{\hat{r}}(x, y; B) = 0 \\ &= \mu^2 (b_1^2 r^{(2,0)}(x,y) + b_2^2 r^{(0,2)}(x,y)) b_2 r^{(0,2)}(x,y) - \frac{M\sigma^2(x,y)}{b_1 b_2^2 n f(x,y)} \end{aligned} \tag{31}$$

have to be fulfilled.

Equating (30) and (31)

$$b_1^2 |r^{(2,0)}(x,y)| = b_2^2 |r^{(0,2)}(x,y)|,$$

and taking square-root

$$b_1 = b_2 \sqrt{\frac{|r^{(0,2)}(x,y)|}{|r^{(2,0)}(x,y)|}}.$$

Plugging in (30)

$$\frac{M\sigma^2(x,y)}{\mu^2 n f(x,y)} = (b_2^2 |r^{(0,2)}(x,y)| \operatorname{sgn}(r^{(2,0)}(x,y)) + b_2^2 r^{(0,2)}(x,y)) \cdot b_2^4 r^{(2,0)}(x,y) \left|\frac{r^{(0,2)}(x,y)}{r^{(2,0)}(x,y)}\right|^{\frac{3}{2}},$$

rearrange the formula

$$\frac{M\sigma^2(x,y)}{\mu^2 n f(x,y)} = b_2^6 (\operatorname{sgn}(r^{(2,0)}(x,y)) + \operatorname{sgn}(r^{(0,2)}(x,y))) \cdot \operatorname{sgn}(r^{(2,0)}(x,y)) \frac{|r^{(0,2)}(x,y)|^{\frac{5}{2}}}{|r^{(2,0)}(x,y)|^{\frac{1}{2}}}$$

yielding

$$b_2 = \left(\frac{M\sigma^2(x,y)}{\mu^2 n f(x,y)}\right)^{\frac{1}{6}} \frac{|r^{(2,0)}(x,y)|^{\frac{1}{12}}}{|r^{(0,2)}(x,y)|^{\frac{5}{12}} a^{\frac{1}{6}}},$$

where $a = 1 + \text{sgn}(r^{(2,0)}(x,y))\text{sgn}(r^{(0,2)}(x,y))$.

If B is arbitrary, then in the minimum

$$\begin{aligned}
&\frac{\partial}{\partial\alpha} AMSE_{\hat{r}}(x,y;B) = 0 \\
&= \mu^2\Big(csr^{(2,0)}(x,y) - scr^{(0,2)}(x,y) + (c^2-s^2)r^{(1,1)}(x,y)\Big) \\
&\cdot(b_2^2-b_1^2)\Big((b_1^2c^2+b_2^2s^2)r^{(2,0)}(x,y) + (b_1^2s^2+b_2^2c^2)r^{(0,2)}(x,y) + 2(b_2^2-b_1^2)sc\,r^{(1,1)}(x,y)\Big). \quad (32)
\end{aligned}$$

It is sufficient to consider only the first term.

$$\begin{aligned}
0 &= scr^{(2,0)}(x,y) - scr^{(0,2)}(x,y) + (c^2-s^2)r^{(1,1)}(x,y) \\
&= \frac{1}{2}(r^{(2,0)}(x,y) - r^{(0,2)}(x,y))\sin 2\alpha + r^{(1,1)}(x,y)\cos 2\alpha
\end{aligned}$$

This is equivalent to

$$\frac{\sin 2\alpha}{\cos 2\alpha} = -2\frac{r^{(1,1)}(x,y)}{r^{(2,0)}(x,y) - r^{(0,2)}(x,y)}$$

and

$$\alpha = \frac{1}{2}\arctan\frac{2r^{(1,1)}(x,y)}{r^{(0,2)}(x,y) - r^{(2,0)}(x,y)},$$

if $r^{(2,0)}(x,y) - r^{(0,2)}(x,y) \neq 0$, otherwise $\alpha = 0$ solves (32).

$$\begin{aligned}
&\frac{\partial}{\partial b_1} AMSE_{\hat{r}}(x,y;B) = 0 \\
&= \mu^2\Big((b_1^2c^2+b_2^2s^2)r^{(2,0)}(x,y) + (b_1^2s^2+b_2^2c^2)r^{(0,2)}(x,y) + 2(b_2^2-b_1^2)sc\,r^{(1,1)}(x,y)\Big) \\
&\quad b_1\Big(c^2\,r^{(2,0)}(x,y) + s^2\,r^{(0,2)}(x,y) - 2sc\,r^{(1,1)}(x,y)\Big) - \frac{M}{nb_1^2b_2}\frac{\sigma^2(x,y)}{f(x,y)}, \quad (33)
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial}{\partial b_2} AMSE_{\hat{r}}(x,y;B) = 0 \\
&= \mu^2\Big((b_1^2c^2+b_2^2s^2)r^{(2,0)}(x,y) + (b_1^2s^2+b_2^2c^2)r^{(0,2)}(x,y) + 2(b_2^2-b_1^2)sc\,r^{(1,1)}(x,y)\Big) \\
&\quad b_2\Big(s^2\,r^{(2,0)}(x,y) + c^2\,r^{(0,2)}(x,y) + 2sc\,r^{(1,1)}(x,y)\Big) - \frac{M}{nb_1b_2^2}\frac{\sigma^2(x,y)}{f(x,y)}, \quad (34)
\end{aligned}$$

The last term in (32) occurs also in (33) and (34). If it is zero, it is not possible to solve (33) and (34).

Equating (33) and (34)

$$b_1^2(c^2\, r^{(2,0)}(x,y) + s^2\, r^{(0,2)}(x,y) - 2sc\, r^{(1,1)}(x,y))$$

$$= b_2^2(s^2\, r^{(2,0)}(x,y) + c^2\, r^{(0,2)}(x,y) + 2sc\, r^{(1,1)}(x,y)),$$

and taking square-root

$$b_1 = b_2\sqrt{\left|\frac{u}{z}\right|}.$$

Plugging in (33)

$$\begin{aligned}\frac{M\sigma^2(x,y)}{n\mu^2 f(x,y)} \quad = \quad & b_2^4\left|\frac{u}{z}\right|^{\frac{3}{2}}\Big((b_2^2\left|\frac{u}{z}\right|c^2 + b_2^2 s^2)r^{(2,0)}(x,y) + (b_2^2\left|\frac{u}{z}\right|s^2 + b_2^2 c^2)r^{(0,2)}(x,y)\\ & +2(b_2^2 - b_2^2\left|\frac{u}{z}\right|)sc\, r^{(1,1)}(x,y)\Big)\\ & \cdot\Big(c^2\, r^{(2,0)}(x,y) + s^2\, r^{(0,2)}(x,y) - 2sc\, r^{(1,1)}(x,y)\Big),\end{aligned}$$

rearranging

$$\frac{M\sigma^2(x,y)}{n\mu^2 f(x,y)} = b_2^6\left|\frac{u}{z}\right|^{\frac{3}{2}}(\left|\frac{u}{z}\right|u + z)u$$

yielding

$$b_2^6 = \frac{M\sigma^2(x,y)}{n\mu^2 f(x,y)(\mathrm{sgn}(u)\mathrm{sgn}(z)+1)}\frac{|u|^{\frac{1}{2}}}{|z|^{\frac{5}{2}}}$$

$$b_2 = \left(\frac{M\sigma^2(x,y)}{2n\mu^2 f(x,y)}\right)^{\frac{1}{6}}\frac{|u|^{\frac{1}{12}}}{|z|^{\frac{5}{12}}}$$

This ends the part for AMSE.

$$\frac{\partial}{\partial b_1}AMISE_{\hat{r}}(B) = \mu^2\left(b_1^3 I_{2,0}^{2,0} + b_1 b_2^2 I_{2,0}^{0,2}\right) - \frac{M}{n b_1^2 b_2}I_{\sigma,f} = 0 \tag{35}$$

$$\frac{\partial}{\partial b_2}AMISE_{\hat{r}}(B) = \mu^2\left(b_2^3 I_{0,2}^{0,2} + b_1^2 b_2 I_{2,0}^{0,2}\right) - \frac{M}{n b_1 b_2^2}I_{\sigma,f} = 0 \tag{36}$$

Equating (35) and (36)

$$b_1^4 = b_2^4\frac{I_{0,2}^{0,2}}{I_{2,0}^{2,0}}.$$

Taking fourth root

$$b_1 = b_2\left(\frac{I_{0,2}^{0,2}}{I_{2,0}^{2,0}}\right)^{\frac{1}{4}}.$$

Plugging in (35)

$$
\begin{aligned}
& b_2^4 \left(\frac{I_{0,2}^{0,2}}{I_{2,0}^{2,0}} \right)^{\frac{3}{4}} \left(b_2^2 I_{2,0}^{2,0} \left(\frac{I_{0,2}^{0,2}}{I_{2,0}^{2,0}} \right)^{\frac{2}{4}} + b_2^2 I_{2,0}^{0,2} \right) = \frac{M}{n\mu^2} I_{\sigma,f} \\
\Leftrightarrow\ & b_2^6 = \frac{M}{n\mu^2} I_{\sigma,f} \left(\frac{I_{2,0}^{2,0}}{I_{0,2}^{0,2}} \right)^{\frac{3}{4}} \frac{1}{\sqrt{I_{2,0}^{2,0} I_{0,2}^{0,2}} + I_{2,0}^{0,2}} \\
\Leftrightarrow\ & b_2 = \left(\frac{M}{n\mu^2} I_{\sigma,f} \right)^{\frac{1}{6}} \left(\frac{I_{2,0}^{2,0}}{I_{0,2}^{0,2}} \right)^{\frac{1}{8}} \left(\frac{1}{\sqrt{I_{2,0}^{2,0} I_{0,2}^{0,2}} + I_{2,0}^{0,2}} \right)^{\frac{1}{6}}
\end{aligned}
$$

■

The rate of the optimal local and global bandwidths b_1 and b_2 is always $n^{-\frac{1}{6}}$ under the assumptions of Proposition 2.8.

Remark 2.9 *If $b_1 = O(n^{-\frac{1}{6}})$, $b_2 = O(n^{-\frac{1}{6}})$ and the assumptions of Proposition 2.8 hold, then the error terms*

$$e_{MSE} = |MSE_{\hat{r}}(x, y; B) - AMSE_{\hat{r}}(x, y; B)|$$

and

$$e_{MISE} = |MISE_{\hat{r}}(x, y; B) - AMISE_{\hat{r}}(x, y; B)|$$

are of the same order e and

- *for r two times continuously partial differentiable*
  - *for equidistant design and constant σ^2*

    $$e = o(n^{-\frac{4}{6}}) + O(n^{-\frac{6}{6}}) + O(n^{-\frac{6}{6}}) = o(n^{-\frac{4}{6}})$$

  - *for equidistant design and non constant σ^2*

    $$e = o(n^{-\frac{4}{6}}) + O(n^{-\frac{6}{6}}) + O(n^{-\frac{6}{6}}) + O(n^{-\frac{6}{6}}) = o(n^{-\frac{4}{6}})$$

  - *for arbitrary fixed design*

    $$e = o(n^{-\frac{4}{6}}) + O(n^{-\frac{6}{6}}) + o(n^{-\frac{4}{6}}) = o(n^{-\frac{4}{6}})$$

- *for r four times continuously differentiable*
  - *for equidistant design and constant* σ^2

$$e = O(n^{-\frac{6}{6}}) + O(n^{-\frac{5}{6}}) + O(n^{-\frac{6}{6}}) = O(n^{-\frac{5}{6}})$$

  - *for equidistant design and non constant* σ^2

$$e = O(n^{-\frac{6}{6}}) + O(n^{-\frac{5}{6}}) + O(n^{-\frac{6}{6}}) + O(n^{-\frac{6}{6}}) = O(n^{-\frac{5}{6}})$$

  - *for arbitrary fixed design*

$$e = O(n^{-\frac{6}{6}}) + O(n^{-\frac{5}{6}}) + O(n^{-\frac{6}{6}}) + o(n^{-\frac{4}{6}}) = o(n^{-\frac{4}{6}})$$

Remark 2.10 *If r is twice continuously differentiable or four times continuously differentiable with arbitrary fixed design and the assumptions of Proposition 2.8 hold, then*

$$b_{1_{MISE}} = b_{1_{IASY}} + o(n^{-\frac{1}{6}}),$$

$$b_{2_{MISE}} = b_{2_{IASY}} + o(n^{-\frac{1}{6}}),$$

$$b_{1_{MSE}} = b_{1_{ASY}} + o(n^{-\frac{1}{6}}),$$

$$b_{2_{MSE}} = b_{2_{ASY}} + o(n^{-\frac{1}{6}})$$

and if B is not diagonal

$$\alpha_{MSE} = \alpha_{ASY} + o(1),$$

where $B_{MSE} := \min_B MSE_{\hat{r}}(x, y; B)$ *and* $B_{MISE} := \min_B MISE_{\hat{r}}(B)$. *If r is four times partial differentiable with equidistant design, then*

$$b_{1_{MISE}} = b_{1_{IASY}} + O(n^{-\frac{1}{4}}),$$

$$b_{2_{MISE}} = b_{2_{IASY}} + O(n^{-\frac{1}{4}}),$$

$$b_{1_{MSE}} = b_{1_{ASY}} + O(n^{-\frac{1}{4}}),$$

$$b_{2_{MSE}} = b_{2_{ASY}} + O(n^{-\frac{1}{4}}),$$

and if B is not diagonal

$$\alpha_{MSE} = \alpha_{ASY} + O(n^{-\frac{1}{12}}).$$

Proof: If $b_1 = O(n^{-\frac{1}{6}})$, $b_2 = O(n^{-\frac{1}{6}})$ and $\alpha \in (-\frac{\pi}{2}, \frac{\pi}{2})$, then, according to Remark (2.9),

$$MSE_{\hat{r}}(x,y;B) - AMSE_{\hat{r}}(x,y;B) = \begin{cases} o(n^{-\frac{4}{6}}) & \text{or} \\ O(n^{-\frac{5}{6}}) & \end{cases}$$

According to Proposition (2.8)

$$b_{1_{ASY}} = c_{1_{ASY}} n^{-\frac{1}{6}}$$

and

$$b_{2_{ASY}} = c_{2_{ASY}} n^{-\frac{1}{6}},$$

$c_{1_{ASY}}, c_{2_{ASY}} > 0$ and independent of n. Therefore

$$b_{1_{MSE}} = c_{1_{MSE}} n^{-\frac{1}{6}} + o(n^{-\frac{1}{6}})$$

and

$$b_{2_{MSE}} = c_{2_{MSE}} n^{-\frac{1}{6}} + o(n^{-\frac{1}{6}}),$$

$c_{1_{MSE}}, c_{2_{MSE}} > 0$ and independent of n, because otherwise the rates of MSE are getting worse. Therefore

$$\begin{aligned} & AMSE_{\hat{r}}(x,y;B_{MSE}) - AMSE_{\hat{r}}(x,y;B_{ASY}) \\ = \; & MSE_{\hat{r}}(x,y;B_{MSE}) - MSE_{\hat{r}}(x,y;B_{ASY}) + \begin{cases} o(n^{-\frac{4}{6}}) \\ O(n^{-\frac{5}{6}}) \end{cases} \end{aligned}$$

The lefthandside is positive, because $AMSE$ is minimized by B_{ASY}, the righthandside is negative, therefore

$$AMSE_{\hat{r}}(x,y;B_{MSE}) - AMSE_{\hat{r}}(x,y;B_{ASY}) = \begin{cases} o(n^{-\frac{4}{6}}) \\ O(n^{-\frac{5}{6}}) \end{cases}$$

A Taylor expansion of $AMSE_{\hat{r}}(x,y;B_{MSE})$, differentiated with respect to $b_{1_{MSE}}$, $b_{2_{MSE}}$ and α_{MSE} gives

$$AMSE_{\hat{r}}(x,y;B_{MSE}) \;=\; AMSE_{\hat{r}}(x,y;B_{ASY})$$

$$
+\sum_{|\nu|=1} D^\nu AMSE_{\hat r}(x,y;B_{ASY}) \begin{pmatrix} b_{1_{MSE}} - b_{1_{ASY}} \\ b_{2_{MSE}} - b_{2_{ASY}} \\ \alpha_{MSE} - \alpha_{ASY} \end{pmatrix}^\nu
$$

$$
+\frac{1}{2}\sum_{|\nu|=2} D^\nu AMSE_{\hat r}(x,y;\tilde B) \begin{pmatrix} b_{1_{MSE}} - b_{1_{ASY}} \\ b_{2_{MSE}} - b_{2_{ASY}} \\ \alpha_{MSE} - \alpha_{ASY} \end{pmatrix}^\nu , \qquad (37)
$$

where $|\nu| := \nu_1 + \nu_2 + \nu_3$, $D^\nu := \frac{\partial^{|\nu|}}{\partial b_1^{\nu_1} \partial b_2^{\nu_2} \partial \alpha^{\nu_3}}$ and there exist $\theta_1, \theta_2, \theta_\alpha \in [0,1]$ such that

$$
\begin{pmatrix} \tilde b_1 \\ \tilde b_2 \\ \tilde\alpha \end{pmatrix} = \begin{pmatrix} b_{1_{ASY}} \\ b_{2_{ASY}} \\ \alpha_{ASY} \end{pmatrix} + \begin{pmatrix} \theta_1(b_{1_{MSE}} - b_{1_{ASY}}) \\ \theta_2(b_{2_{MSE}} - b_{2_{ASY}}) \\ \theta_\alpha(\alpha_{MSE} - \alpha_{ASY}) \end{pmatrix}
$$

and therefore $\tilde b_1 = \tilde c_1 n^{-\frac{1}{6}} + o(n^{-\frac{1}{6}})$, $\tilde b_2 = \tilde c_2 n^{-\frac{1}{6}} + o(n^{-\frac{1}{6}})$ and $\tilde\alpha \in (-\frac{\pi}{2}, \frac{\pi}{2})$.

$D^\nu AMSE_{\hat r}(x,y;B_{ASY}) = 0$ for $|\nu| = 1$, because B_{ASY} minimizes $AMSE_{\hat r}(x,y;B)$.

Now $AMSE_{\hat r}(x,y;\tilde B)$ is differentiated.

$$
D^{2,0,0} AMSE_{\hat r}(x,y;\tilde B) = cn^{-\frac{2}{6}}(1+o(1))
$$

$$
D^{1,1,0} AMSE_{\hat r}(x,y;\tilde B) = cn^{-\frac{2}{6}}(1+o(1))
$$

$$
D^{1,0,1} AMSE_{\hat r}(x,y;\tilde B) = cn^{-\frac{3}{6}}(1+o(1))
$$

$$
D^{0,2,0} AMSE_{\hat r}(x,y;\tilde B) = cn^{-\frac{2}{6}}(1+o(1))
$$

$$
D^{0,1,1} AMSE_{\hat r}(x,y;\tilde B) = cn^{-\frac{3}{6}}(1+o(1))
$$

$$
D^{0,0,2} AMSE_{\hat r}(x,y;\tilde B) = cn^{-\frac{4}{6}}(1+o(1))
$$

For a simpler notation c stands for possibly different constants not equal to zero.

Equation (37) gives

$$
\begin{aligned}
& AMSE_{\hat r}(x,y;B_{MSE}) - AMSE_{\hat r}(x,y;B_{ASY}) \\
= \;\; & cn^{-\frac{2}{6}}(1+o(1))(b_{1_{MSE}} - b_{1_{ASY}})^2 + cn^{-\frac{2}{6}}(1+o(1))(b_{1_{MSE}} - b_{1_{ASY}})(b_{2_{MSE}} - b_{2_{ASY}})
\end{aligned}
$$

$$+cn^{-\frac{3}{6}}(1+o(1))(b_{1_{MSE}}-b_{1_{ASY}})(\alpha_{MSE}-\alpha_{ASY})+cn^{-\frac{2}{6}}(1+o(1))(b_{2_{MSE}}-b_{2_{ASY}})^2$$

$$+cn^{-\frac{3}{6}}(1+o(1))(b_{2_{MSE}}-b_{2_{ASY}})(\alpha_{MSE}-\alpha_{ASY})+cn^{-\frac{4}{6}}(1+o(1))(\alpha_{MSE}-\alpha_{ASY})^2$$

$$= \begin{cases} o(n^{-\frac{4}{6}}) \\ O(n^{-\frac{5}{6}}) \end{cases}. \tag{38}$$

Equation (38) is equivalent to

$$(b_{1_{MSE}}-b_{1_{ASY}})^2+2(b_{1_{MSE}}-b_{1_{ASY}})(b_{2_{MSE}}-b_{2_{ASY}})$$

$$+n^{-\frac{1}{6}}(b_{1_{MSE}}-b_{1_{ASY}})(\alpha_{MSE}-\alpha_{ASY})+(b_{2_{MSE}}-b_{2_{ASY}})^2$$

$$+n^{-\frac{1}{6}}(b_{2_{MSE}}-b_{2_{ASY}})(\alpha_{MSE}-\alpha_{ASY})+n^{-\frac{2}{6}}(\alpha_{MSE}-\alpha_{ASY})^2$$

$$= \begin{cases} o(n^{-\frac{2}{6}}) \\ O(n^{-\frac{3}{6}}) \end{cases}. \tag{39}$$

Every term has to be $o(n^{-\frac{1}{6}})$ or $O(n^{-\frac{3}{12}})$.

Therefore

$$b_{1_{MSE}}-b_{1_{ASY}} = \begin{cases} o(n^{-\frac{1}{6}}) \\ O(n^{-\frac{1}{4}}) \end{cases},$$

$$b_{2_{MSE}}-b_{2_{ASY}} = \begin{cases} o(n^{-\frac{1}{6}}) \\ O(n^{-\frac{1}{4}}) \end{cases}$$

and

$$\alpha_{MSE}-\alpha_{ASY} = \begin{cases} o(1) \\ O(n^{-\frac{1}{12}}) \end{cases}.$$

The proof for the global bandwidths follows the same lines, but without the parts dealing with α. ■

2.3.1 Global versus Local Bandwidth Parameters

The most crucial part in kernel estimation is the bandwidth selection. There arise two main questions: local or global bandwidth parameters and how many parameters.

Global parameters are easier to select and the estimation is more stable. And for a first impression they could be chosen by hand. But they are not able to adapt to spatial inhomogeneities. For the selection of local parameters an automatical procedure is necessary. Within certain ranges the exact size of the parameter does not matter (see Hall et al. (1995)for related ideas in the onedimensional case). So, for some inhomogenous functions the MISE can be reduced, if an automatical, stable procedure selects the parameter(s).

2.3.2 Number of Bandwidth Parameters

The next question is the number of parameters. With more parameters the estimator is more flexible to adapt to the underlying function and to reduce the variance of the noise. For three parameters

$$
\begin{aligned}
&AMSE_{\hat{r}}(x,y;B_{ASY}) \\
&= \frac{\mu^2}{4}\Bigg[\left(\frac{M\sigma^2(x,y)}{2n\mu^2 f(x,y)}\right)^{\frac{1}{3}}\left(\frac{|z|^{\frac{1}{6}}}{|u|^{\frac{5}{6}}}c^2+\frac{|u|^{\frac{1}{6}}}{|z|^{\frac{5}{6}}}s^2\right)r^{(2,0)}(x,y) \\
&\qquad +\left(\frac{M\sigma^2(x,y)}{2n\mu^2 f(x,y)}\right)^{\frac{1}{3}}\left(\frac{|z|^{\frac{1}{6}}}{|u|^{\frac{5}{6}}}s^2+\frac{|u|^{\frac{1}{6}}}{|z|^{\frac{5}{6}}}c^2\right)r^{(0,2)}(x,y) \\
&\qquad +2\left(\frac{M\sigma^2(x,y)}{2n\mu^2 f(x,y)}\right)^{\frac{1}{3}}\left(\frac{|u|^{\frac{1}{6}}}{|z|^{\frac{5}{6}}}-\frac{|z|^{\frac{1}{6}}}{|u|^{\frac{5}{6}}}\right)sc\,r^{(1,1)}(x,y)\Bigg]^2 \\
&\quad +\frac{M\sigma^2(x,y)}{nf(x,y)}\left(\frac{M\sigma^2(x,y)}{2n\mu^2 f(x,y)}\right)^{-\frac{1}{3}}(|z||u|)^{\frac{1}{3}} \\
&= \frac{\mu^2}{4}\left(\frac{M\sigma^2(x,y)}{2n\mu^2 f(x,y)}\right)^{\frac{2}{3}}\Bigg[\frac{|z|^{\frac{1}{6}}}{|u|^{\frac{5}{6}}}(c^2r^{(2,0)}(x,y)+s^2r^{(0,2)}(x,y)-2scr^{(1,1)}(x,y)) \\
&\qquad +\frac{|u|^{\frac{1}{6}}}{|z|^{\frac{5}{6}}}(s^2r^{(2,0)}(x,y)+c^2r^{(0,2)}(x,y)+2scr^{(1,1)}(x,y))\Bigg]^2 \\
&\quad +\left(\frac{M\sigma^2(x,y)}{nf(x,y)}\right)^{\frac{2}{3}}\left(2\mu^2\right)^{\frac{1}{3}}(|z||u|)^{\frac{1}{3}} \\
&= \frac{\mu^2}{4}\left(\frac{M\sigma^2(x,y)}{2n\mu^2 f(x,y)}\right)^{\frac{2}{3}}|uz|^{\frac{1}{3}}[\mathrm{sgn}(u)+\mathrm{sgn}(z)]^2+\left(\frac{M\sigma^2(x,y)}{nf(x,y)}\right)^{\frac{2}{3}}\left(2\mu^2\right)^{\frac{1}{3}}|uz|^{\frac{1}{3}} \\
&= \left(\frac{M\sigma^2(x,y)}{nf(x,y)}\right)^{\frac{2}{3}}|uz|^{\frac{1}{3}}\mu^{\frac{2}{3}}\left(\left(\frac{1}{2}\right)^{\frac{2}{3}}+2^{\frac{1}{3}}\right)
\end{aligned}
$$

under the assumptions of Proposition 2.8. For two parameters it is almost the same except that uz is $r^{(2,0)}(x,y)r^{(0,2)}(x,y)$. The order of the terms of higher order in the MSE does not depend on the number of parameters. Plugging in the definition for u, z and α yields

$$|uz| = |r^{(2,0)}(x,y)r^{(0,2)}(x,y)-(r^{(1,1)}(x,y))^2| \leq |r^{(2,0)}(x,y)r^{(0,2)}(x,y)|$$

for $r^{(2,0)}(x,y)r^{(0,2)}(x,y) > 0$, so $AMSE_{\hat{r}}(x,y;B_{ASY})$ with three parameters is always less or equal to the case for two parameters.

There is another nice interpretation of the parameter α : The diagonalization of a 2×2-matrix is a rotation and can be expressed by a matrix $\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$ and its transposed. Applying this to the Hessian of $r(x,y)$ yields as rotation angle the optimal bandwith parameter α. This can be seen by calculating the eigenvalues of the Hessian and using similar equations as for the bandwidth parametrization (see equations (1)). That means a good rotation angle (in terms of MSE) is a angle, which rotates the ellipse in such a way that one axis points in the direction of the smallest curvature and the other one in the direction of the largest curvature. Therefore, the support of the kernel can be increased (to reduce the variance), but the bias is not getting much larger. The ellipse is "long and thin".

3 Iterative Local Plug-in Bandwidth Selection

3.1 Two Bandwidth Parameters

This algorithm is a generalization for two dimensions of the algorithm of Brockmann et al. (1993); Brockmann (1993).

The bandwidth selection is based on the formulas for AMSE- and AMISE-optimal bandwidths ((26), (27), (28) and (29)). All unknown quantities are replaced by estimators.

The estimation of the design density and the variance of the noise has to be calculated in advance. The estimators of $I_{\sigma,f}$ and $\frac{\sigma^2(x,y)}{f(x,y)}$ have to be $O(n^{-\frac{1}{6}}) + O_p(n^{-\frac{1}{2}})$ and $O(n^{-\frac{1}{6}}) + O_p(n^{-\frac{1}{4}})$ resp. In the simulations the design ist equidistant and the variance constant.

The estimation of the second derivatives of $r(x,y)$ is a little bit more complicated, because kernel estimators can be used, but again the right bandwidths are not known. The idea is to start with small bandwidths, plug the derivatives obtained with the starting bandwidths inflated by a factor cn^{β} (because for the estimation of derivatives bandwidths need to be larger) in the formulas for the

AMSE-optimal bandwidths, get a new bandwidth estimation, inflate it again, obtain derivatives and so on.

It is not possible to start directly with the formula for the AMSE-optimal bandwidths, because for small bandwidths the variances of the estimators for the second derivatives are dominating (see proof of Theorem 4.1), so the estimators try to fit the noise rather than the underlying function. To stabilize the procedure it is better to start with global bandwidths. But once the bandwidth estimators have reached the right order $n^{-\frac{1}{6}}$ during the iteration steps, it is possible to switch to local bandwidths.

Step 1 estimate $I_{\sigma,f} = \int v(x,y)\frac{\sigma^2(x,y)}{f(x,y)}\,d(x,y)$ and $\frac{\sigma^2(x,y)}{f(x,y)}$

Initial values for b_1, b_2: $\hat{b}_1^{(0)} = n^{-\frac{1}{2}}$, $\hat{b}_2^{(0)} = n^{-\frac{1}{2}}$,

Step 2 for $i = 1, ..., i^*$ repeat

$$\hat{b}_1^{(i)} = \left(\frac{M}{n\mu^2}\hat{I}_{\sigma,f}\right)^{\frac{1}{6}} \left(\frac{\hat{I}_{0,2}^{0,2}}{\hat{I}_{2,0}^{2,0}}\right)^{\frac{1}{8}} \left(\frac{1}{\sqrt{\hat{I}_{2,0}^{2,0}\hat{I}_{0,2}^{0,2}} + \hat{I}_{2,0}^{0,2}}\right)^{\frac{1}{6}}$$

$$\hat{b}_2^{(i)} = \left(\frac{\hat{I}_{2,0}^{2,0}}{\hat{I}_{0,2}^{0,2}}\right)^{\frac{1}{4}} \hat{b}_1^{(i)}$$

where $I_{m,p}^{k,l}$ is approximated by $\hat{I}_{m,p}^{k,l}$ by the evaluation and weighted summation of $\hat{r}^{(m,p)}(x,y;c_1 n^\beta \hat{b}_1^{(i-1)}, c_2 n^\beta \hat{b}_2^{(i-1)}) \cdot \hat{r}^{(k,l)}(x,y;c_1 n^\beta \hat{b}_1^{(i-1)}, c_2 n^\beta \hat{b}_2^{(i-1)})v(x,y)$ on a fine grid

Step 3 for $j = 1, .., j^*$ repeat

$$\hat{b}_1^{(i^*+j)}(x,y) = \left(\frac{M\hat{\sigma}^2(x,y)}{n\mu^2 \hat{f}(x,y)}\frac{1}{S_1}\right)^{\frac{1}{6}}$$

$$\hat{b}_2^{(i^*+j)}(x,y) = \left(\frac{M\hat{\sigma}^2(x,y)}{n\mu^2 \hat{f}(x,y)}\frac{1}{S_2}\right)^{\frac{1}{6}},$$

where S_1 is the maximum of $2\frac{|r^{(2,0)}(x,y)|^{\frac{5}{2}}}{|r^{(0,2)}(x,y)|^{\frac{1}{2}}}$ itself or its smoothed version (with the bandwidths of the previous iteration step) and S_2 is the maximum of $2\frac{|r^{(0,2)}(x,y)|^{\frac{5}{2}}}{|r^{(2,0)}(x,y)|^{\frac{1}{2}}}$ itself or its smoothed version

In step 2 and 3 bandwidths are restricted to the interval $[\frac{1}{2\sqrt{n}}, \frac{1}{2}]$. If they are outside, then their values are set to $\frac{1}{2\sqrt{n}}$ or $\frac{1}{2}$ resp. Regression with bandwidths of size $\frac{1}{2\sqrt{n}}$ is almost interpolation, so there is no need for smaller bandwidths. An upper bound of $\frac{1}{2}$ is somewhat arbitrary, but it prevents the rescaled kernel to be out of the range of the image on opposite sides.

Starting with initial bandwidths $n^{-\frac{1}{2}}$ is motivated by simulations in the one-dimensional case, where sometimes bandwidth estimators starting with large values are trapped in local minima and do not get close to the optimal bandwidth (see Herrmann (1997)).

The number i^* of global steps to drive the bandwidths to the optimal order $n^{-\frac{1}{6}}$ is fixed and depends on the inflation factor n^β (see Chapter 4.2).

In step 3 local bandwidth estimators are calculated. But pure local bandwidths have drawbacks. If the signs of $r^{(2,0)}$ and $r^{(0,2)}$ are not the same, then the bandwidth estimators are infinite. If $r^{(2,0)}$ or $r^{(0,2)}$ are close to zero, then the bandwidth estimators are very large. The estimation of the bandwidths are based on a certain neighbourhood of (x, y) depending on the previous bandwidths and the inflation factor, so it is not sensible to let it be much larger. Simulations for one dimension can be found in Brockmann (1993); Brockmann et al. (1993). So, for some points (x, y) the estimated bandwidth function has small, high peaks, which produce strange peaks in the estimated function $\hat{r}(x, y)$. The peaks in the bandwidth functions have to be smoothed a little bit. There are many possibilities how to do this and the results are quite similar. Here, it is done by smoothing $2\frac{|r^{(2,0)}(x,y)|^{\frac{5}{2}}}{|r^{(0,2)}(x,y)|^{\frac{1}{2}}}$ for b_1 and $2\frac{|r^{(0,2)}(x,y)|^{\frac{5}{2}}}{|r^{(2,0)}(x,y)|^{\frac{1}{2}}}$ for b_2 with the previous bandwidths, but the non-smoothed value is kept if it is larger. Another possibility is to smooth the resulting bandwidth functions, or to mix the local and the global formulas

$$\begin{aligned}
\hat{b}_1^{(i^*+j)}(x,y) &= \left(\frac{M\hat{\sigma}^2(x,y)}{2n\mu^2\hat{f}(x,y)}\right)^{\frac{1}{6}} \left(\frac{\hat{r}_{j,(x,y)}^{(0,2)(0,2)}}{\hat{r}_{j,(x,y)}^{(2,0)(2,0)}}\right)^{\frac{1}{8}} \\
&\quad \cdot \left(\frac{1}{\sqrt{\hat{r}_{j,(x,y)}^{(2,0)(2,0)}\hat{r}_{j,(x,y)}^{(0,2)(0,2)}} + \hat{r}_{j,(x,y)}^{(0,2)(2,0)}}\right)^{\frac{1}{6}} \\
\hat{b}_2^{(i^*+j)}(x,y) &= \left(\frac{\hat{r}_{j,(x,y)}^{(2,0)(2,0)}}{\hat{r}_{j,(x,y)}^{(0,2)(0,2)}}\right)^{\frac{1}{4}} \hat{b}_1^{(i^*+j)}(x,y)
\end{aligned}$$

where $\hat{r}_{j,(x,y)}^{(i,j)(k,l)}$ is the smoothed estimator for $r^{(i,j)}(x,y)r^{(k,l)}(x,y)$ with bandwidths $c_1 n^\beta \hat{b}_1^{(i^*+j-1)}$ and $c_2 n^\beta \hat{b}_2^{(i^*+j-1)}$.

The constants c_1 and c_2 in the inflation factor are chosen to be one, but they are an additional possibility to control the behaviour of the bandwidth estimators during the iteration steps.

This algorithm involves no numerical minimization procedure with all its problems.

3.2 Three Bandwidth Parameters

For three bandwidth parameters the algorithm remains unchanged for the first and second step, because they are only necessary to obtain pilot estimators of the right order.

In step 3 estimators for α, b_1 and b_2 are obtained using the AMSE-optimal formulas. Again, smoothing is necessary.

4 Asymptotic Properties of the Proposed Bandwidth Estimators

4.1 Convergence of the Bandwidth Estimators

Theorem 4.1 *Under the assumptions of Chapter 2.1, if r is four times continuously differentiable and the inflation factor is $n^{\frac{1}{12}}$, then the following rates are obtained for the estimators of Chapter 3.1*

- *if* $I_{2,0}^{2,0} \neq 0$, $I_{0,2}^{0,2} \neq 0$ *and* $\sqrt{I_{2,0}^{2,0} I_{0,2}^{0,2}} + I_{2,0}^{0,2} \neq 0$

$$\begin{aligned} \hat{b}_1^{(7)} &= b_{1_{MISE}} \left(1 + O\left(n^{-\frac{1}{6}}\right) + O_P\left(n^{-\frac{1}{2}}\right)\right) \\ \hat{b}_2^{(7)} &= b_{2_{MISE}} \left(1 + O\left(n^{-\frac{1}{6}}\right) + O_P\left(n^{-\frac{1}{2}}\right)\right), \end{aligned}$$

- *if* $r^{(2,0)}(x,y) r^{(0,2)}(x,y) > 0$,

$$\begin{aligned} \hat{b}_1^{(6)}(x,y) &= b_{1_{MSE}}(x,y) \left(1 + O\left(n^{-\frac{1}{6}}\right) + O_P\left(n^{-\frac{1}{4}}\right)\right) \\ \hat{b}_2^{(6)}(x,y) &= b_{2_{MSE}}(x,y) \left(1 + O\left(n^{-\frac{1}{6}}\right) + O_P\left(n^{-\frac{1}{4}}\right)\right). \end{aligned}$$

A sketch of the proof for the global bandwidth estimator can be found in Herrmann et al. (1995).

Proof: The proof goes along the following lines: First, expansions are shown for the square of the second derivatives, which are used in step 2 and 3 of the algorithm. Then the effect of the iteration steps is considered giving the desired results.

In the proof, bandwidths b_1 and b_2 are considered which are random variables and have values in the interval $[\frac{1}{2\sqrt{n}}, \delta]$, such that there are no boundary effects to be considered. In some places it is assumed that nonrandom bandwidths $\tilde{b}_1$ and $\tilde{b}_2$ exist which approximate the random b_1 and b_2 in the following way

$$b_1 = \tilde{b}_1(1 + o_P(n^{-\gamma})) \tag{40}$$

$$b_2 = \tilde{b}_2(1 + o_P(n^{-\gamma})) \tag{41}$$

for some $\gamma > 0$. Further, it is assumed that $\tilde{b}_1$ and $\tilde{b}_2$ are of the same order and converge to zero. The existence of such $\tilde{b}_1$ and $\tilde{b}_2$ and all other assumptions on b_1 and b_2 will be ensured during the iteration steps.

First consider

$$\begin{aligned}
&[\hat{r}^{(2,0)}(x,y;b_1,b_2)]^2 \\
&= \left[\frac{1}{b_1^3 b_2}\sum_{k=1}^{n}\int_{A_k} K_{2,0}\left(B^{-1}\begin{pmatrix} x-u \\ y-v\end{pmatrix}\right) du\, dv\, (r(x_k,y_k)+\epsilon_k)\right]^2 \\
&= \left[\frac{1}{b_1^3 b_2}\sum_{k=1}^{n}\int_{A_k} K_{2,0}\left(B^{-1}\begin{pmatrix} x-u \\ y-v\end{pmatrix}\right) du\, dv\, r(x_k,y_k)\right]^2 \qquad (42) \\
&\quad + \frac{2}{b_1^6 b_2^2}\sum_{k=1}^{n}\sum_{i=1}^{n}\int_{A_k}\int_{A_i} K_{2,0}\left(B^{-1}\begin{pmatrix} x-u \\ y-v\end{pmatrix}\right) \\
&\qquad\qquad \cdot K_{2,0}\left(B^{-1}\begin{pmatrix} x-\bar{u} \\ y-\bar{v}\end{pmatrix}\right) du\, dv\, d\bar{u}\, d\bar{v}\, r(x_k,y_k)\epsilon_i \qquad (43) \\
&\quad + \frac{1}{b_1^6 b_2^2}\sum_{k=1}^{n}\sum_{i=1}^{n}\int_{A_k}\int_{A_i} K_{2,0}\left(B^{-1}\begin{pmatrix} x-u \\ y-v\end{pmatrix}\right) K_{2,0}\left(B^{-1}\begin{pmatrix} x-\bar{u} \\ y-\bar{v}\end{pmatrix}\right) du\, dv\, d\bar{u}\, d\bar{v}\, \epsilon_k\epsilon_i \qquad (44)
\end{aligned}$$

Denote (42) by $\mathcal{B}^2(x,y;b_1,b_2)$, (43) by $\mathcal{M}(x,y;b_1,b_2)$ and decompose it in two parts $\mathcal{M}_1(x,y;b_1,b_2)$ and $\mathcal{M}_2(x,y;b_1,b_2)$ with

$$\mathcal{M}_1(x,y;b_1,b_2) = \frac{2}{b_1^6 b_2^2}\sum_{k=1}^{n}\sum_{i=1}^{n}\int_{A_k}\int_{A_i} K_{2,0}\left(B^{-1}\begin{pmatrix} x-u \\ y-v\end{pmatrix}\right)$$

$$\cdot K_{2,0}\left(B^{-1}\binom{x-\bar{u}}{y-\bar{v}}\right)(r(x_k,y_k)-r(u,v))\,du\,dv\,d\bar{u}\,d\bar{v}\,\epsilon_i$$

and

$$\begin{aligned}&\mathcal{M}_2(x,y;b_1,b_2)\\ &= \frac{2}{b_1^6 b_2^2}\sum_{k=1}^{n}\sum_{i=1}^{n}\int_{A_k}\int_{A_i} K_{2,0}\left(B^{-1}\binom{x-u}{y-v}\right)K_{2,0}\left(B^{-1}\binom{x-\bar{u}}{y-\bar{v}}\right) r(u,v)\,du\,dv\,d\bar{u}\,d\bar{v}\,\epsilon_i.\end{aligned}$$

Denote (44) by $\mathcal{V}(x,y;b_1,b_2)$ and decompose it in two parts $\mathcal{V}_1$ and $\mathcal{V}_2$ with

$$\begin{aligned}\mathcal{V}_1(x,y;b_1,b_2) &= \frac{1}{b_1^6 b_2^2}\sum_{k=1}^{n}\int_{A_k}\int_{A_k} K_{2,0}\left(B^{-1}\binom{x-u}{y-v}\right)\\ &\quad\cdot K_{2,0}\left(B^{-1}\binom{x-\bar{u}}{y-\bar{v}}\right)\,du\,dv\,d\bar{u}\,d\bar{v}\,\sigma^2(x_k,y_k)\end{aligned}$$

and

$$\begin{aligned}\mathcal{V}_2(x,y;b_1,b_2) &= \frac{1}{b_1^6 b_2^2}\sum_{k=1}^{n}\sum_{i=1}^{n}\int_{A_k}\int_{A_i} K_{2,0}\left(B^{-1}\binom{x-u}{y-v}\right)\\ &\quad\cdot K_{2,0}\left(B^{-1}\binom{x-\bar{u}}{y-\bar{v}}\right)\,du\,dv\,d\bar{u}\,d\bar{v}\,\left(\epsilon_k\epsilon_i-\delta_{ik}\sigma^2(x_k,y_k)\right).\end{aligned}$$

Now the parts $\mathcal{B}^2$, $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{V}_1$ and $\mathcal{V}_2$ are treated separately.

First, according to the last lines of the proof of Lemma 2.4b (3)

$$\begin{aligned}&\mathcal{V}_1(x,y;b_1,b_2)\\ &= \frac{1}{b_1^6 b_2^2}\sum_{k=1}^{n}\int_{A_k}\int_{A_k} K_{2,0}\left(B^{-1}\binom{x-u}{y-v}\right)K_{2,0}\left(B^{-1}\binom{x-\bar{u}}{y-\bar{v}}\right)\,du\,dv\,d\bar{u}\,d\bar{v}\,\sigma^2(x_k,y_k)\\ &= \frac{1}{b_1^6 b_2^2}\int_A \frac{\sigma^2(u,v)}{nf(u,v)}K_{2,0}^2\left(B^{-1}\binom{x-u}{y-v}\right)\,du\,dv\left(1+O\left(\frac{1}{\sqrt{n}}\left(\frac{1}{b_1}+\frac{1}{b_2}\right)\right)+o(1)\right).\end{aligned}$$

$\frac{\sigma^2}{f}$ can be expanded in a Taylor series as in the proof of Proposition 2.3

$$\begin{aligned}&\mathcal{V}_1(x,y;b_1,b_2)\\ &= \frac{1}{nb_1^5 b_2}\int_T \frac{\sigma^2\left(\binom{x}{y}-B\binom{z_1}{z_2}\right)}{f\left(\binom{x}{y}-B\binom{z_1}{z_2}\right)}K_{2,0}^2(z_1,z_2)\,dz_1\,dz_2\left(1+O\left(\frac{1}{\sqrt{n}}\left(\frac{1}{b_1}+\frac{1}{b_2}\right)\right)+o(1)\right)\\ &= \frac{1}{nb_1^5 b_2}\frac{\sigma^2(x,y)}{f(x,y)}M_{2,0}\left(1+O(b_1^2+b_2^2)\right)\left(1+O\left(\frac{1}{\sqrt{n}}\left(\frac{1}{b_1}+\frac{1}{b_2}\right)\right)+o(1)\right),\end{aligned}$$

where $M_{2,0}=\int_T K_{2,0}^2(z_1,z_2)\,dz_1\,dz_2$.

If $\frac{1}{b_1} = o_P(\sqrt{n})$, $\frac{1}{b_2} = o_P(\sqrt{n})$ and b_1, b_2 are $o_P(1)$, then

$$\mathcal{V}_1(x,y;b_1,b_2) = \frac{M_{2,0}}{nb_1^5 b_2}\frac{\sigma^2(x,y)}{f(x,y)}\left(1+o_P(1)\right).$$

The treatment of $\mathcal{V}_2$ is a little bit more complicated, because there is a mixture of two random structures: the errors ϵ_k and the bandwidths b_1, b_2.

$$\begin{aligned}\mathcal{V}_2(x,y;b_1,b_2) &= \frac{1}{b_1^6 b_2^2}\sum_{k=1}^n\sum_{i=1}^n \int_{A_k}\int_{A_i} K_{2,0}\left(B^{-1}\begin{pmatrix}x-u\\ y-v\end{pmatrix}\right) \\ &\qquad \cdot K_{2,0}\left(B^{-1}\begin{pmatrix}x-\bar{u}\\ y-\bar{v}\end{pmatrix}\right) du\, dv\, d\bar{u}\, d\bar{v}\ \left(\epsilon_k\epsilon_i - \delta_{ik}\sigma^2(x_k,y_k)\right) \\ &= \sum_{k=1}^n\sum_{i=1}^n a_{ik}(x,y;b_1,b_2)\left(\epsilon_k\epsilon_i - \delta_{ik}\sigma^2(x_k,y_k)\right),\end{aligned}$$

where

$$a_{ik}(x,y;b_1,b_2) = \frac{1}{b_1^6 b_2^2}\int_{A_k}\int_{A_i} K_{2,0}\left(B^{-1}\begin{pmatrix}x-u\\ y-v\end{pmatrix}\right) K_{2,0}\left(B^{-1}\begin{pmatrix}x-\bar{u}\\ y-\bar{v}\end{pmatrix}\right) du\, dv\, d\bar{u}\, d\bar{v}.$$

Therefore we need a discrete approximation of the bandwidths b_1, b_2 :

$B_{j,n}^{\rho_j}$, $j = 1,2$, is a set of bandwidths with the following properties:

$$\# B_{j,n}^{\rho_j} = n^{\rho_j}, \quad \max_{b_j \in [\frac{1}{2\sqrt{n}},\frac{1}{2}]}\ \min_{\bar{b}_j \in B_{j,n}^{\rho_j}} |b_j - \bar{b}_j| \le n^{-\rho_j}.$$

During the iteration steps it is ensured that $b_j \in [\frac{1}{2\sqrt{n}},\frac{1}{2}]$.

$$\begin{aligned}&\left| nb_1^5 b_2 \sum_{k=1}^n\sum_{i=1}^n a_{ik}(x,y;b_1,b_2)\left(\epsilon_k\epsilon_i - \delta_{ik}\sigma^2(x_k,y_k)\right)\right| \\ &\le \sup_{\bar{b}_j \in B_{j,n}^{\rho_j}} \left| n\bar{b}_1^5\bar{b}_2 \sum_{k=1}^n\sum_{i=1}^n a_{ik}(x,y;\bar{b}_1,\bar{b}_2)\left(\epsilon_k\epsilon_i - \mathbb{E}\epsilon_i\epsilon_k\right)\right| \qquad (45) \\ &\quad + \inf_{\bar{b}_j \in B_{j,n}^{\rho_j}} \left| nb_1^5 b_2 \sum_{k=1}^n\sum_{i=1}^n a_{ik}(x,y;b_1,b_2)\left(\epsilon_k\epsilon_i - \mathbb{E}\epsilon_i\epsilon_k\right) \right. \\ &\qquad \left. - n\bar{b}_1^5\bar{b}_2 \sum_{k=1}^n\sum_{i=1}^n a_{ik}(x,y;\bar{b}_1,\bar{b}_2)\left(\epsilon_k\epsilon_i - \mathbb{E}\epsilon_i\epsilon_k\right)\right| \qquad (46)\end{aligned}$$

First consider (46).

$K_{2,0}$ is Lipschitz continuous with respect to b_1 and b_2. Therefore a_{ik} is continuously differentiable and also Lipschitz continuous with respect to b_1 and b_2 on $[\frac{1}{2}n^{-\frac{1}{2}}, \frac{1}{2}]$.

The number of nonzero a_{ik} is $O(b_1^2 b_2^2 n^2)$ (similar to the considerations in the proof of Lemma 2.4). If ρ_1 and ρ_2 are chosen large enough.

$$\lim_{n\to\infty} P\Bigg(\inf_{\bar b_j \in B_{j,n}^{\rho_j}} \left| n \sum_{k=1}^n \sum_{i=1}^n \left(b_1^5 b_2 \, a_{ik}(x,y;b_1,b_2) - \bar b_1^5 \bar b_2 \, a_{ik}(x,y;\bar b_1,\bar b_2) \right) (\epsilon_k \epsilon_i - \mathbb{E}\epsilon_i\epsilon_k) \right| \geq \epsilon \Bigg) = 0,$$

because

$$\begin{aligned} & n \sum_{k=1}^n \sum_{i=1}^n \left(b_1^5 b_2 \, a_{ik}(x,y;b_1,b_2) - \bar b_1^5 \bar b_2 \, a_{ik}(x,y;\bar b_1,\bar b_2) \right) \\ = \;& O\left(n(b_1^2 b_2^2 n^2 + \bar b_1^2 \bar b_2^2 n^2) \left(n^{-\rho_1} + n^{-\rho_2} \right) \right) = o(1). \end{aligned}$$

Therefore (46) is $o_P(1)$.

Now consider (45). For all positive, real numbers ϵ and η and for all positive integers α

$$\begin{aligned} & P\Bigg(\sup_{\bar b_j \in B_{j,n}^{\rho_j}} \left| n \bar b_1^5 \bar b_2 \sum_{k=1}^n \sum_{i=1}^n a_{ik}(x,y;\bar b_1,\bar b_2) (\epsilon_k \epsilon_i - \mathbb{E}\epsilon_i\epsilon_k) \right| \geq \epsilon n^\eta \Bigg) \\ = \;& P\Bigg(\sup_{\bar b_j \in B_{j,n}^{\rho_j}} \left| \frac{n \bar b_1^5 \bar b_2}{\epsilon n^\eta} \sum_{k=1}^n \sum_{i=1}^n a_{ik}(x,y;\bar b_1,\bar b_2) (\epsilon_k \epsilon_i - \mathbb{E}\epsilon_i\epsilon_k) \right|^{2\alpha} \geq 1 \Bigg) \\ \leq \;& \mathbb{E} \sum_{\bar b_j \in B_{j,n}^{\rho_j}} \left| \frac{1}{\epsilon n^\eta} n \bar b_1^5 \bar b_2 \sum_{k=1}^n \sum_{i=1}^n a_{ik}(x,y;\bar b_1,\bar b_2) (\epsilon_k \epsilon_i - \mathbb{E}\epsilon_i\epsilon_k) \right|^{2\alpha}, \end{aligned}$$

because instead of taking the supremum the sum is taken and Markov's inequality is applied.

$$\begin{aligned} & \mathbb{E} \sum_{\bar b_j \in B_{j,n}^{\rho_j}} \left| \frac{1}{\epsilon n^\eta} n \bar b_1^5 \bar b_2 \sum_{k=1}^n \sum_{i=1}^n a_{ik}(x,y;\bar b_1,\bar b_2) (\epsilon_k \epsilon_i - \mathbb{E}\epsilon_i\epsilon_k) \right|^{2\alpha} \\ = \;& \sum_{\bar b_j \in B_{j,n}^{\rho_j}} \epsilon^{-2\alpha} n^{-2\alpha\eta} \mathbb{E} \left| \sum_{k=1}^n \sum_{i=1}^n n \bar b_1^5 \bar b_2 a_{ik}(x,y;\bar b_1,\bar b_2) (\epsilon_k \epsilon_i - \mathbb{E}\epsilon_i\epsilon_k) \right|^{2\alpha} \end{aligned}$$

Now an inequality of Whittle (1960) is required which can be stated as follows (a simplified version):

$$\mathbb{E} \left(\sum_{k=1}^n \sum_{i=1}^n c_{ik} \left(X_i X_j - \mathbb{E}\left(X_i X_j \right) \right) \right)^{2\alpha} \leq C_\alpha \left(\sum_{k=1}^n \sum_{i=1}^n c_{ik}^2 \right)^\alpha,$$

where α is a positive integer, c_{ik} real numbers and X_i, $i = 1, ..., n$, independent random variables with expectation zero. The constant C_α is positive and depends on the moments of X_i. The moments have to exist up to order 2α.

$$\begin{aligned} & \sum_{\bar b_j \in B_{j,n}^{\rho_j}} \epsilon^{-2\alpha} n^{-2\alpha\eta} \mathbb{E} \left| \sum_{k=1}^n \sum_{i=1}^n n \bar b_1^5 \bar b_2 a_{ik}(x,y;\bar b_1,\bar b_2) (\epsilon_k \epsilon_i - \mathbb{E}\epsilon_i\epsilon_k) \right|^{2\alpha} \\ \leq \;& \sum_{\bar b_j \in B_{j,n}^{\rho_j}} \epsilon^{-2\alpha} n^{-2\alpha\eta} C_\alpha \left(\sum_{k=1}^n \sum_{i=1}^n n^2 \bar b_1^{10} \bar b_2^2 a_{ik}^2(x,y;\bar b_1,\bar b_2) \right)^\alpha \end{aligned}$$

$$
\begin{aligned}
|a_{ik}(x,y;b_1,b_2)| &\leq \frac{1}{b_1^6 b_2^2}\left(\max_{(z_1,z_2)\in T} K_{2,0}(z_1,z_2)\right)^2 \lambda(A_k)\lambda(A_i) \\
&= O\left(\frac{1}{n^2 b_1^6 b_2^2}\right)
\end{aligned}
$$

The number of nonzero a_{ik} is $O(b_1^2 b_2^2 n^2)$.

$$
\sum_{k=1}^{n}\sum_{i=1}^{n} n^2 \bar{b}_1^{10}\bar{b}_2^2 a_{ik}^2(x,y;\bar{b}_1,\bar{b}_2) = O\left(n^2 \bar{b}_1^2 \bar{b}_2^2 n^2 \bar{b}_1^{10}\bar{b}_2^2 \frac{1}{n^4 \bar{b}_1^{12}\bar{b}_2^4}\right) = O(1)
$$

$$
\sum_{\bar{b}_j \in B_{j,n}^{\rho_j}} \epsilon^{-2\alpha} n^{-2\alpha\eta} C_\alpha \left(\sum_{k=1}^{n}\sum_{i=1}^{n} n^2 \bar{b}_1^{10}\bar{b}_2^2 a_{ik}^2(x,y;b_1,b_2)\right)^\alpha = O\left(n^{\rho_1+\rho_2} n^{-2\alpha\eta}\right) \longrightarrow 0
$$

for $n \to \infty$ and if α is chosen large enough. Therefore (45) is $o_P(n^\eta)$ and hence

$$
\mathcal{V}_2(x,y;b_1,b_2) = o_P\left(n^{-1+\eta} b_1^{-5} b_2^{-1}\right) = o_P\left(n^{-1+\eta} b_1^{-6}\right).
$$

The treatment of $\mathcal{M}_1(x,y;b_1,b_2)$ is analogously to $\mathcal{V}_2(x,y;b_1,b_2)$.

$$
\begin{aligned}
\mathcal{M}_1(x,y;b_1,b_2) &= \frac{2}{b_1^6 b_2^2}\sum_{k=1}^{n}\sum_{i=1}^{n}\int_{A_k}\int_{A_i} K_{2,0}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) \\
&\quad \cdot K_{2,0}\left(B^{-1}\begin{pmatrix} x-\bar{u} \\ y-\bar{v} \end{pmatrix}\right)(r(x_k,y_k)-r(u,v))\, du\, dv\, d\bar{u}\, d\bar{v}\, \epsilon_i \\
&= \sum_{i=1}^{n} d_i(x,y;b_1,b_2)\epsilon_i,
\end{aligned}
$$

where

$$
\begin{aligned}
&d_i(x,y;b_1,b_2) \\
&= \frac{2}{b_1^6 b_2^2}\int_{A_i}\sum_{k=1}^{n}\int_{A_k} K_{2,0}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) K_{2,0}\left(B^{-1}\begin{pmatrix} x-\bar{u} \\ y-\bar{v} \end{pmatrix}\right)(r(x_k,y_k)-r(u,v))\, du\, dv\, d\bar{u}\, d\bar{v}.
\end{aligned}
$$

Here another inequality of Whittle (1960) is used:

$$
\mathbb{E}\left(\sum_{i=1}^{n}\beta_i X_i\right)^{2\alpha} \leq C_\alpha \left(\sum_{i=1}^{n}\beta_i^2\right)^\alpha,
$$

where α is a positive integer, β_i real numbers and X_i, $i = 1,...,n$, independent random variables with expectation zero. The constant C_α is positive and depends on the moments of X_i. The moments have to exist up to order 2α.

$|d_i(x,y;b_1,b_2)|$

$$
\begin{aligned}
&\leq \frac{2}{b_1^6 b_2^2} \int_{A_i} \sum_{k=1}^{n} \int_{A_k} \left| K_{2,0}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) K_{2,0}\left(B^{-1}\begin{pmatrix} x-\bar{u} \\ y-\bar{v} \end{pmatrix}\right)\right| \cdot |r(x_k,y_k)-r(u,v)| \, du\, dv\, d\bar{u}\, d\bar{v} \\
&\leq \frac{2}{b_1^6 b_2^2}\lambda(A_i)\left(\sum_{k=1}^{n}\lambda(\tilde{A}_k)\right)\max_{z\in T} K_{2,0}^2(z)\, L \sup_k \sup_{w,z\in A_k} \| w-z \| \\
&= O\left(\frac{1}{b_1^6 b_2^2}\frac{b_1 b_2}{n}\frac{1}{\sqrt{n}}\right) = O\left(\frac{1}{b_1^5 b_2 n^{\frac{3}{2}}}\right),
\end{aligned}
$$

$\tilde{A}_k$ as in the proof of Lemma 2.4a.

The number of nonzero d_i is $O(b_1 b_2 n)$.

Therefore

$$
\mathcal{M}_1(x,y;b_1,b_2) = o_P\left(n^{-1+\eta} b_1^{-\frac{9}{2}} b_2^{-\frac{1}{2}}\right) = o_P\left(n^{-1+\eta} b_1^{-5}\right).
$$

This is a better rate than for $\mathcal{V}_2$.

The terms $\mathcal{B}^2$ and $\mathcal{M}_2$ are dominating $\mathcal{V}_1$, $\mathcal{V}_2$ and $\mathcal{M}_1$, so a more careful investigation is necessary.

First, $\mathcal{B}(x,y;\tilde{b}_1,\tilde{b}_2)$ is considered.

According to Lemma 2.4 and the Taylor expansion as in the proof of Proposition 2.7,

$$
\begin{aligned}
\mathcal{B}(x,y;b_1,b_2) &= \frac{1}{b_1^3 b_2}\sum_{k=1}^{n}\int_{A_k} K_{2,0}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) du\, dv\, r(x_k,y_k) \\
&= \frac{1}{b_1^3 b_2}\left[\int_A K_{2,0}\left(B^{-1}\begin{pmatrix} x-u \\ y-v \end{pmatrix}\right) r(u,v)\, du\, dv + O(b_1 b_2 n^{-\frac{1}{2}})\right] \\
&= \frac{1}{b_1^2}\int_T K_{2,0}(z_1,z_2) r\left(\begin{pmatrix} x \\ y \end{pmatrix} - B\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}\right) dz_1\, dz_2 + O(b_1^{-2} n^{-\frac{1}{2}}) \\
&= \frac{1}{b_1^2}\frac{1}{2} r^{(2,0)}(x,y) 2 b_1^2 + O(b_1^2 + \frac{b_2^4}{b_1^2} + b_1^{-2} n^{-\frac{1}{2}}) \\
&= r^{(2,0)}(x,y) + O(b_1^2 + b_1^{-2} n^{-\frac{1}{2}}) \\
&= r^{(2,0)}(x,y) + O(\tilde{b}_1^2 + \tilde{b}_1^{-2} n^{-\frac{1}{2}}) + o_P\left(\tilde{b}_1^2 n^{-\gamma} + \tilde{b}_1^{-2} n^{-\frac{1}{2}-\gamma}\right),
\end{aligned}
$$

with $\tilde{b}_1$ as in (40).

Therefore

$$
\begin{aligned}
\mathcal{B}^2(x,y;b_1,b_2) &= \left[r^{(2,0)}(x,y)\right]^2 + O\left(\tilde{b}_1^2 + \tilde{b}_1^{-2} n^{-\frac{1}{2}} + \tilde{b}_1^{-4} n^{-1}\right) \\
&\quad + o_P\left(n^{-\gamma}(\tilde{b}_1^2 + \tilde{b}_1^{-2} n^{-\frac{1}{2}} + \tilde{b}_1^{-4} n^{-1} + \tilde{b}_1^{-4} n^{-1-\gamma})\right)
\end{aligned}
$$

The last part is the investigation of $\mathcal{M}_2(x,y;b_1,b_2)$. It is split in

$$
\mathcal{M}_2(x,y;\tilde{b}_1,\tilde{b}_2) + (\mathcal{M}_2(x,y;b_1,b_2) - \mathcal{M}_2(x,y;\tilde{b}_1,\tilde{b}_2)),
$$

with $\tilde b_1, \tilde b_2$ as in (40) and (41).

$$\begin{aligned}
&\mathcal{M}_2(x,y;\tilde b_1,\tilde b_2)\\
&= \frac{2}{\tilde b_1^6 \tilde b_2^2}\sum_{k=1}^{n}\sum_{i=1}^{n}\int_{A_k}\int_{A_i} K_{2,0}\left(\tilde B^{-1}\begin{pmatrix}x-u\\ y-v\end{pmatrix}\right) K_{2,0}\left(\tilde B^{-1}\begin{pmatrix}x-\bar u\\ y-\bar v\end{pmatrix}\right) r(u,v)\,du\,dv\,d\bar u\,d\bar v\,\epsilon_i\\
&= \sum_{i=1}^{n} c_i(x,y;\tilde b_1,\tilde b_2)\epsilon_i,
\end{aligned}$$

where

$$\begin{aligned}
&c_i(x,y;\tilde b_1,\tilde b_2)\\
&= \frac{2}{\tilde b_1^6 \tilde b_2^2}\int_{A_i}\sum_{k=1}^{n}\int_{A_k} K_{2,0}\left(\tilde B^{-1}\begin{pmatrix}x-u\\ y-v\end{pmatrix}\right) K_{2,0}\left(\tilde B^{-1}\begin{pmatrix}x-\bar u\\ y-\bar v\end{pmatrix}\right) r(\bar u,\bar v)\,d\bar u\,d\bar v\,du\,dv\\
&= \frac{2}{\tilde b_1^3 \tilde b_2}\int_{A_i} K_{2,0}\left(\tilde B^{-1}\begin{pmatrix}x-u\\ y-v\end{pmatrix}\right) du\,dv\,S(x,y;\tilde b_1,\tilde b_2), \qquad (47)
\end{aligned}$$

where

$$\begin{aligned}
S(x,y;\tilde b_1,\tilde b_2) &= \frac{1}{\tilde b_1^3 \tilde b_2}\int_A K_{2,0}\left(\tilde B^{-1}\begin{pmatrix}x-u\\ y-v\end{pmatrix}\right) r(u,v)\,du\,dv\\
&= \frac{1}{\tilde b_1^2}\int_T K_{2,0}(z_1,z_2) r\left(\begin{pmatrix}x\\ y\end{pmatrix} - \tilde B\begin{pmatrix}z_1\\ z_2\end{pmatrix}\right) dz_1\,dz_2 \qquad (48)\\
&= r^{(2,0)}(x,y) + O(\tilde b_1^2 + \tilde b_2^2)
\end{aligned}$$

analogously to the calculations of $\mathcal{B}(x,y;\tilde b_1,\tilde b_2)$.

$$|c_i(x,y;\tilde b_1,\tilde b_2)| = O\left(\frac{1}{n\tilde b_1^3 \tilde b_2}\right).$$

The number of nonzero c_i is $O(\tilde b_1 \tilde b_2 n)$.

Therefore, with similar calculations as for $\mathcal{V}_2(x,y;b_1,b_2)$

$$\mathcal{M}_2(x,y;\tilde b_1,\tilde b_2) = o_P\left(n^{-\frac{1}{2}+\eta}\tilde b_1^{-\frac{5}{2}}\tilde b_2^{-\frac{1}{2}}\right) = o_P\left(n^{-\frac{1}{2}+\eta}\tilde b_1^{-3}\right).$$

$$\begin{aligned}
\mathcal{M}_2(x,y;b_1,b_2) - \mathcal{M}_2(x,y;\tilde b_1,\tilde b_2) &= \sum_{i=1}^{n}(c_i(x,y;b_1,b_2) - c_i(x,y;\tilde b_1,\tilde b_2))\epsilon_i\\
&= \sum_{i=1}^{n}(c_i(x,y;b_1,b_2) - c_i(x,y;\tilde b_1,b_2))\epsilon_i \qquad (49)\\
&\quad + \sum_{i=1}^{n}(c_i(x,y;\tilde b_1,b_2) - c_i(x,y;\tilde b_1,\tilde b_2))\epsilon_i. \qquad (50)
\end{aligned}$$

Combining (47) and (48), then, for two different pairs of bandwidths b_1, b_2 and $\hat{b}_1, \hat{b}_2$

$$
\begin{aligned}
&|c_i(x,y;b_1,b_2) - c_i(x,y;\hat{b}_1,\hat{b}_2)| \\
= &\ \Bigg| \frac{2}{b_1^3 b_2} \int_{A_i} K_{2,0}(c)\, du\, dv && (51)\\
&\qquad \cdot \Bigg\{ \frac{1}{b_1^2} \int_T K_{2,0}(z_1,z_2)\, [r(a) - r(\hat{a})]\, dz_1\, dz_2 && (52)\\
&\qquad\quad + \int_T \left[\frac{1}{b_1^2} - \frac{1}{\hat{b}_1^2} \right] K_{2,0}(z_1,z_2) r(\hat{a})\, d(z_1,z_2) \Bigg\} && (53)\\
&\quad + \int_T \frac{1}{\hat{b}_1^2} K_{2,0}(z_1,z_2) r(\hat{a})\, d(z_1,z_2) && (54)\\
&\qquad \cdot \Bigg\{ \frac{2}{b_1^3 b_2} \int_{A_i} K_{2,0}(c) - K_{2,0}(\hat{c})\, du\, dv && (55)\\
&\qquad\quad + \left(\frac{2}{b_1^3 b_2} - \frac{2}{\hat{b}_1^3 \hat{b}_2} \right) \int_{A_i} K_{2,0}(\hat{c})\, du\, dv \Bigg\} \Bigg|, && (56)
\end{aligned}
$$

where $a = \binom{x}{y} - B\binom{z_1}{z_2}$, $\hat{a} = \binom{x}{y} - \hat{B}\binom{z_1}{z_2}$, $c = B^{-1}\binom{x-u}{y-v}$ and $\hat{c} = \hat{B}^{-1}\binom{x-u}{y-v}$.

First, consider (56).

$$
\left| \left(\frac{2}{b_1^3 b_2} - \frac{2}{\hat{b}_1^3 \hat{b}_2} \right) \int_{A_i} K_{2,0}\left(\hat{B}^{-1} \begin{pmatrix} x-u \\ y-v \end{pmatrix} \right) du\, dv \right| \quad \leq \quad C \frac{|\hat{b}_1^3 \hat{b}_2 - b_1^3 b_2|}{n b_1^3 b_2 \hat{b}_1^3 \hat{b}_2} \qquad (57)
$$

for an appropriate constant C (for short notation C is used for different constants).

For (55) consider first

$$
\int_{A_i} K_{2,0}\left(B^{-1} \begin{pmatrix} x-u \\ y-v \end{pmatrix} \right) - K_{2,0}\left(\hat{B}^{-1} \begin{pmatrix} x-u \\ y-v \end{pmatrix} \right) du\, dv.
$$

This expression is always zero, if $B^{-1}\binom{x-u}{y-v}$ and $\hat{B}^{-1}\binom{x-u}{y-v}$ are outside T (the unit ball) for all $(u,v) \in A_i$, e.g. if

$$
\frac{|x-u|}{b_1} > 1, \quad \frac{|x-u|}{\hat{b}_1} > 1, \quad \frac{|y-v|}{b_2} > 1 \text{ and } \frac{|y-v|}{\hat{b}_2} > 1, \text{ i.e.}
$$

$$
|x-u| > \max(b_1, \hat{b}_1) \text{ and } |y-v| > \max(b_2, \hat{b}_2) \text{ for all } (u,v) \in A_i.
$$

$K_{2,0}\left(B^{-1}\binom{x-u}{y-v}\right)$ is Lipschitz continuous in b_1 and b_2, therefore

$$
\int_{A_i} K_{2,0}\left(B^{-1} \begin{pmatrix} x-u \\ y-v \end{pmatrix} \right) - K_{2,0}\left(\hat{B}^{-1} \begin{pmatrix} x-u \\ y-v \end{pmatrix} \right) du\, dv
$$

$$
\begin{aligned}
&\leq L\int_{A_i}\left|\frac{x-u}{b_1}-\frac{x-u}{\hat{b}_1}\right|+\left|\frac{y-v}{b_2}-\frac{y-v}{\hat{b}_2}\right|\,du\,dv\\
&= L\int_{A_i}|x-u|\left|\frac{1}{b_1}-\frac{1}{\hat{b}_1}\right|+|y-v|\left|\frac{1}{b_2}-\frac{1}{\hat{b}_2}\right|\,du\,dv\\
&\leq \frac{L}{n}\left(\frac{\max(b_1,\hat{b}_1)}{b_1\hat{b}_1}|\hat{b}_1-b_1|+\frac{\max(b_2,\hat{b}_2)}{b_2\hat{b}_2}|\hat{b}_2-b_2|\right)\\
&= L\left(\frac{|\hat{b}_1-b_1|}{n\min(b_1,\hat{b}_1)}+\frac{|\hat{b}_2-b_2|}{n\min(b_2,\hat{b}_2)}\right).
\end{aligned}
$$

Later, these results are only used in (49) and (50). In these equations either $b_1=\hat{b}_1$ or $b_2=\hat{b}_2$. Under these conditions both terms in (58) can be bounded by (57), therefore (55) can be bounded by (57).

(54) is $r^{(2,0)}(x,y)+O(\hat{b}_1^2+\hat{b}_2^2)$ (see (48)), so the combination of (54), (55) and (56) is also bounded by (57).

For (52) a Taylor expansion of $r\left(\binom{x}{y}-B\binom{z_1}{z_2}\right)$ around $\binom{x}{y}-\hat{B}\binom{z_1}{z_2}=:\hat{a}$ is used:

$$
\begin{aligned}
& r\left(\begin{pmatrix}x\\y\end{pmatrix}-B\begin{pmatrix}z_1\\z_2\end{pmatrix}\right)\\
&\quad\leq r(\hat{a})+r^{(1,0)}(\hat{a})(\hat{b}_1-b_1)z_1+r^{(0,1)}(\hat{a})(\hat{b}_2-b_2)z_2\\
&\quad\quad+\frac{1}{2}\left[r^{(2,0)}(\hat{a})(\hat{b}_1-b_1)^2z_1^2+r^{(0,2)}(\hat{a})(\hat{b}_2-b_2)^2z_2^2+2r^{(1,1)}(\hat{a})(\hat{b}_1-b_1)(\hat{b}_2-b_2)z_1z_2\right]\\
&\quad\quad-\frac{1}{6}\left[(\hat{b}_1-b_1)z_1\frac{\partial}{\partial x}+(\hat{b}_2-b_2)z_2)\frac{\partial}{\partial y}\right]^3r(\hat{a})\\
&\quad\quad+C((\hat{b}_1-b_1)^4+(\hat{b}_2-b_2)^4+(\hat{b}_1-b_1)^2(\hat{b}_2-b_2)^2)
\end{aligned}
$$

For $r^{(\nu_1,\nu_2)}(\hat{a})$ expansions around (x,y) are plugged in. The kernel properties delete all odd terms.

$$
\begin{aligned}
&\int_T K_{2,0}(z_1,z_2)\,[r(a)-r(\hat{a})]\,dz_1\,dz_2\\
&\leq \int_T K_{2,0}(z_1,z_2)\Bigg[z_1(\hat{b}_1-b_1)\left(-z_1\hat{b}_1r^{(2,0)}\begin{pmatrix}x\\y\end{pmatrix}-z_1z_2^2\hat{b}_1\hat{b}_2^2r^{(2,2)}\begin{pmatrix}\xi\\\eta\end{pmatrix}-z_1^3\hat{b}_1^3r^{(4,0)}\begin{pmatrix}\xi\\\eta\end{pmatrix}\right)\\
&\quad+z_2(\hat{b}_2-b_2)\left(-z_1^2z_2\hat{b}_1^2\hat{b}_2r^{(2,2)}\begin{pmatrix}\xi\\\eta\end{pmatrix}-z_2^3\hat{b}_2^3r^{(0,4)}\begin{pmatrix}\xi\\\eta\end{pmatrix}\right)\\
&\quad+\frac{1}{2}z_1^2(\hat{b}_1-b_1)^2\left(r^{(2,0)}\begin{pmatrix}x\\y\end{pmatrix}+\frac{1}{2}z_2^2\hat{b}_2^2r^{(2,2)}\begin{pmatrix}\xi\\\eta\end{pmatrix}+\frac{1}{2}z_1^2\hat{b}_1^2r^{(4,0)}\begin{pmatrix}\xi\\\eta\end{pmatrix}\right)\\
&\quad+\frac{1}{2}z_2^2(\hat{b}_2-b_2)^2\left(\frac{1}{2}z_1^2\hat{b}_1^2r^{(2,2)}\begin{pmatrix}\xi\\\eta\end{pmatrix}+\frac{1}{2}z_2^2\hat{b}_2^2r^{(4,0)}\begin{pmatrix}\xi\\\eta\end{pmatrix}\right)\\
&\quad+(\hat{b}_1-b_1)(\hat{b}_2-b_2)z_1z_2\left(\frac{1}{2}z_1z_2\hat{b}_1\hat{b}_2r^{(2,2)}\begin{pmatrix}\xi\\\eta\end{pmatrix}\right)
\end{aligned}
$$

$$
\begin{aligned}
&+C(\hat{b}_1-b_1)^3 z_1^3\left(z_1\hat{b}_1 r^{(4,0)}\begin{pmatrix}\xi\\ \eta\end{pmatrix}\right)+C(\hat{b}_2-b_2)^3 z_2^3\left(z_2\hat{b}_2 r^{(0,4)}\begin{pmatrix}\xi\\ \eta\end{pmatrix}\right)\\
&+C(\hat{b}_1-b_1)^2(\hat{b}_2-b_2)z_1^2 z_2\left(z_2\hat{b}_2 r^{(2,2)}\begin{pmatrix}\xi\\ \eta\end{pmatrix}\right)\\
&+C(\hat{b}_1-b_1)(\hat{b}_2-b_2)^2 z_1 z_2^2\left(z_1\hat{b}_1 r^{(2,2)}\begin{pmatrix}\xi\\ \eta\end{pmatrix}\right)\Bigg]dz_1\,dz_2\\
&+C((\hat{b}_1-b_1)^4+(\hat{b}_2-b_2)^4+(\hat{b}_1-b_1)^2(\hat{b}_2-b_2)^2)\\
\leq\ & -2(\hat{b}_1-b_1)\hat{b}_1 r^{(2,0)}\begin{pmatrix}x\\ y\end{pmatrix}+\frac{1}{2}(\hat{b}_1-b_1)^2 2r^{(2,0)}\begin{pmatrix}x\\ y\end{pmatrix}\\
&+C((\hat{b}_1-b_1)(\hat{b}_1\hat{b}_2^2+\hat{b}_1^3)+(\hat{b}_2-b_2)(\hat{b}_1^2\hat{b}_2+\hat{b}_2^3)+(\hat{b}_1-b_1)^2(\hat{b}_1^2+\hat{b}_2^2)+(\hat{b}_2-b_2)^2(\hat{b}_1^2+\hat{b}_2^2)\\
&\quad+(\hat{b}_1-b_1)(\hat{b}_2-b_2)\hat{b}_1\hat{b}_2+(\hat{b}_1-b_1)^3\hat{b}_1+(\hat{b}_2-b_2)^3\hat{b}_2+(\hat{b}_1-b_1)^2(\hat{b}_2-b_2)\hat{b}_2\\
&\quad+(\hat{b}_1-b_1)(\hat{b}_2-b_2)^2\hat{b}_1+(\hat{b}_1-b_1)^4+(\hat{b}_2-b_2)^4+(\hat{b}_1-b_1)^2(\hat{b}_2-b_2)^2)\\
\leq\ & (b_1^2-\hat{b}_1^2)r^{(2,0)}\begin{pmatrix}x\\ y\end{pmatrix}+C(|\hat{b}_1-b_1|+|\hat{b}_2-b_2|)(|\hat{b}_1-b_1|^3+|\hat{b}_2-b_2|^3+\hat{b}_1^3+\hat{b}_2^3)\\
\leq\ & (b_1^2-\hat{b}_1^2)r^{(2,0)}\begin{pmatrix}x\\ y\end{pmatrix}+C(|\hat{b}_1-b_1|+|\hat{b}_2-b_2|)(\hat{b}_1^3+\hat{b}_2^3)
\end{aligned}
$$

The mean values (ξ,η) and the constants C are not necessarily the same.

For (53) a Taylor expansion of $r\left(\binom{x}{y}-\hat{B}\binom{z_1}{z_2}\right)$ around $\binom{x}{y}$ is used:

$$
\begin{aligned}
&\int_T K_{2,0}(z_1,z_2)r(\hat{a})dz_1\,dz_2\\
&\quad= \int_T K_{2,0}(z_1,z_2)\left(\frac{1}{2}z_1^2\hat{b}_1^2 r^{(2,0)}\begin{pmatrix}x\\ y\end{pmatrix}+Cz_1^4\hat{b}_1^4 r^{(4,0)}\begin{pmatrix}\xi\\ \eta\end{pmatrix}\right.\\
&\qquad\qquad\left.+Cz_1^2z_2^2\hat{b}_1^2\hat{b}_2^2 r^{(2,2)}\begin{pmatrix}\xi\\ \eta\end{pmatrix}+Cz_2^4\hat{b}_2^4 r^{(0,4)}\begin{pmatrix}\xi\\ \eta\end{pmatrix}\right)dz_1\,dz_2\\
&\quad\leq \hat{b}_1^2 r^{(2,0)}\begin{pmatrix}x\\ y\end{pmatrix}+C(\hat{b}_1^4+\hat{b}_2^4)
\end{aligned}
$$

The combination of (51), (52) and (53)

$$
\begin{aligned}
&\frac{2}{b_1^3b_2}\int_{A_i}K_{2,0}(c)\,du\,dv\cdot\left\{\frac{1}{b_1^2}\left((b_1^2-\hat{b}_1^2)r^{(2,0)}\begin{pmatrix}x\\ y\end{pmatrix}+C(|\hat{b}_1-b_1|+|\hat{b}_2-b_2|)(\hat{b}_1^3+\hat{b}_2^3)\right)\right.\\
&\qquad\qquad\left.+\left[\frac{1}{b_1^2}-\frac{1}{\hat{b}_1^2}\right]\left(\hat{b}_1^2 r^{(2,0)}\begin{pmatrix}x\\ y\end{pmatrix}+C(\hat{b}_1^4+\hat{b}_2^4)\right)\right\}\\
\leq\ & \frac{C}{nb_1^3b_2}\left\{\frac{|\hat{b}_1^2-b_1^2|}{b_1^2\hat{b}_1^2}(\hat{b}_1^4+\hat{b}_2^4)+\frac{1}{b_1^2}(|\hat{b}_1-b_1|+|\hat{b}_2-b_2|)(\hat{b}_1^3+\hat{b}_2^3)\right\}\\
\leq\ & \frac{1}{nb_1^3b_2}\frac{C}{b_1^2}\left\{(|\hat{b}_1-b_1|(\hat{b}_1^2+\hat{b}_2^2)(\hat{b}_1+b_1))+(|\hat{b}_1-b_1|+|\hat{b}_2-b_2|)(\hat{b}_1^3+\hat{b}_2^3)\right\}\\
\leq\ & \frac{1}{nb_1^3b_2}\frac{C}{b_1^2}\left\{|\hat{b}_1-b_1|(\hat{b}_1^3+\hat{b}_2^3+(\hat{b}_1^2+\hat{b}_2^2)b_1)+|\hat{b}_2-b_2|(\hat{b}_1^3+\hat{b}_2^3)\right\}.
\end{aligned}
$$

For bandwidth parameters as in (49)

$$\begin{aligned}\left|c_i(x,y;b_1,b_2)-c_i(x,y;\tilde{b}_1,b_2)\right| &\leq C\frac{|\tilde{b}_1^3-b_1^3|b_2}{nb_1^3b_2^2\tilde{b}_1^3}+C\frac{|\tilde{b}_1-b_1|}{nb_1^3b_2b_1^2}\cdot\left\{\tilde{b}_1^3+b_2^3+(\tilde{b}_1^2+b_2^2)b_1\right\}\\ &\leq C|\tilde{b}_1-b_1|T_1(b_1),\end{aligned}$$

where

$$T_1(b_1):=\frac{\tilde{b}_1^2+b_1^2}{nb_1^3b_2\tilde{b}_1^3}+\frac{1}{nb_1^3b_2b_1^2}\cdot\left\{\tilde{b}_1^3+b_2^3+(\tilde{b}_1^2+b_2^2)b_1\right\},$$

and for bandwidth parameters as in (50)

$$\begin{aligned}\left|c_i(x,y;\tilde{b}_1,b_2)-c_i(x,y;\tilde{b}_1,\tilde{b}_2)\right| &\leq C\frac{|\tilde{b}_2-b_2|\tilde{b}_1^3}{n\tilde{b}_1^3\tilde{b}_2\tilde{b}_1^3b_2}+C\frac{|\tilde{b}_2-b_2|}{n\tilde{b}_1^3b_2\tilde{b}_1^2}\cdot\left\{\tilde{b}_1^3+\tilde{b}_2^3\right\}\\ &\leq C|\tilde{b}_2-b_2|T_2(b_2),\end{aligned}$$

where

$$T_2(b_2):=\frac{1}{n\tilde{b}_1^3\tilde{b}_2b_2}+\frac{1}{n\tilde{b}_1^3b_2\tilde{b}_1^2}\cdot\left\{\tilde{b}_1^3+\tilde{b}_2^3\right\}.$$

Consider (49).

$$\begin{aligned}&\left|\frac{1}{T}\sum_{i=1}^n(c_i(x,y;b_1,b_2)-c_i(x,y;\tilde{b}_1,b_2))\epsilon_i\right|\\ &\leq \sup_{\bar{b}_1\in B_{1,n}^{\rho_1}}\left|\frac{1}{\bar{T}}\sum_{i=1}^n(c_i(x,y;\bar{b}_1,b_2)-c_i(x,y;\tilde{b}_1,b_2))\epsilon_i\right| \qquad (58)\\ &\quad+\inf_{\bar{b}_1\in B_{1,n}^{\rho_1}}\left|\frac{1}{T}\sum_{i=1}^n(c_i(x,y;b_1,b_2)-c_i(x,y;\tilde{b}_1,b_2))\epsilon_i\right.\\ &\qquad\left.-\frac{1}{\bar{T}}\sum_{i=1}^n(c_i(x,y;\bar{b}_1,b_2)-c_i(x,y;\tilde{b}_1,b_2))\epsilon_i\right|, \qquad (59)\end{aligned}$$

where

$$T:=\sqrt{nb_2(\tilde{b}_1+b_1)}|\tilde{b}_1-b_1|T_1(b_1)$$

and

$$\bar{T}:=\sqrt{nb_2(\tilde{b}_1+\bar{b}_1)}|\tilde{b}_1-\bar{b}_1|T_1(\bar{b}_1).$$

First consider (58).

The treatment of (58) is similar to $\mathcal{V}_2(x,y;b_1,b_2)$ and $\mathcal{M}_1(x,y;b_1,b_2)$.

The number of nonzero $c_i(x, y; b_1, b_2)$ and $c_i(x, y; \tilde{b}_1, b_2)$ is $O(nb_1b_2 + n\tilde{b}_1b_2)$,

$$|c_i(x, y; b_1, b_2) - c_i(x, y; \tilde{b}_1, b_2)| \le C|\tilde{b}_1 - b_1|T_1(b_1),$$

so (58) is $o_P(n^\eta)$.

For (59) it has to be shown that

$$\frac{c_i(x, y; b_1, b_2) - c_i(x, y; \tilde{b}_1, b_2)}{|\tilde{b}_1 - b_1|} \tag{60}$$

is Lipschitz continuous as a function of b_1. $\frac{1}{T}$ is also Lipschitz continuous, because the denominator is bounded away from zero, if $b_1, b_2, \tilde{b}_1$ are bounded away from zero, so (59) is $O_P((nb_1b_2+n\tilde{b}_1b_2)(n^{-\rho_1}+n^{-\rho_2})) = o_P(1)$, if $n^{-\rho_1}$ and $n^{-\rho_2}$ are chosen large enough and if $b_1, b_2 \le \frac{1}{2}$.

(60) is differentiable with respect to b_1. Consider the absolute value of the derivative of (60) with respect to b_1 for $b_1 \neq \tilde{b}_1$:

$$\begin{aligned}
&\left|\frac{\partial}{\partial b_1}\frac{c_i(x, y; b_1, b_2) - c_i(x, y; \tilde{b}_1, b_2)}{|\tilde{b}_1 - b_1|}\right| \\
&= \left|\frac{\frac{\partial}{\partial b_1}c_i(x, y; b_1, b_2)(\tilde{b}_1 - b_1) - (c_i(x, y; b_1, b_2) - c_i(x, y; \tilde{b}_1, b_2))}{|\tilde{b}_1 - b_1|^2}\right| \\
&= \left|\frac{\frac{\partial}{\partial b_1}c_i(x, y; b_1, b_2)\mathrm{sgn}(\tilde{b}_1 - b_1) - \frac{\partial}{\partial b_1}c_i(x, y; \eta, b_2)\mathrm{sgn}(\tilde{b}_1 - b_1)}{|\tilde{b}_1 - b_1|}\right| \\
&= \frac{|\frac{\partial}{\partial b_1}c_i(x, y; b_1, b_2) - \frac{\partial}{\partial b_1}c_i(x, y; \eta, b_2)|}{|\tilde{b}_1 - b_1|} \\
&\le C\frac{|\tilde{b}_1 - \eta|}{|\tilde{b}_1 - b_1|} \le C
\end{aligned}$$

with a mean value η between $\tilde{b}_1$ and b_1. Here, it has been used that $\frac{\partial}{\partial b_1}c_i(x, y; b_1, b_2)$ is Lipschitz continuous with respect to b_1 due to the Lipschitz continuity of $K_{2,0}$.

So, (49) is $o_P(Tn^\eta) = o_P(n^{-\frac{1}{2}-\gamma+\eta}\tilde{b}_1^{-\frac{5}{2}}\tilde{b}_2^{-\frac{1}{2}})$.

The treatment of (50) is similar to (49).

$$\begin{aligned}
&\left|\frac{1}{U}\sum_{i=1}^{n}(c_i(x, y; \tilde{b}_1, b_2) - c_i(x, y; \tilde{b}_1, \tilde{b}_2))\epsilon_i\right| \\
&\le \sup_{\bar{b}_2 \in B_{2,n}^{\rho 2}} \left|\frac{1}{\bar{U}}\sum_{i=1}^{n}(c_i(x, y; \tilde{b}_1, \bar{b}_2) - c_i(x, y; \tilde{b}_1, \tilde{b}_2))\epsilon_i\right| \qquad (61)
\end{aligned}$$

$$+\inf_{\bar{b}_2\in B_{2,n}^{\rho 2}}\left|\frac{1}{U}\sum_{i=1}^{n}(c_i(x,y;\tilde{b}_1,b_2)-c_i(x,y;\tilde{b}_1,\tilde{b}_2))\epsilon_i\right.$$
$$\left.-\frac{1}{\bar{U}}\sum_{i=1}^{n}(c_i(x,y;\tilde{b}_1,\bar{b}_2)-c_i(x,y;\tilde{b}_1,\tilde{b}_2))\epsilon_i\right|, \qquad (62)$$

where

$$U := \sqrt{n\tilde{b}_1(\tilde{b}_2+b_2)}|\tilde{b}_2-b_2|T_2(b_2)$$

and

$$\bar{U} := \sqrt{n\tilde{b}_1(\tilde{b}_2+\bar{b}_2)}|\tilde{b}_2-\bar{b}_2|T_2(\bar{b}_2).$$

(61) is $o_P(n^\eta)$ and (62) is $o_P(1)$.

So, (50) is $o_P(Un^\eta)=o_P(n^{-\frac{1}{2}-\gamma+\eta}\tilde{b}_1^{-\frac{5}{2}}\tilde{b}_2^{-\frac{1}{2}})$.

Hence,

$$\mathcal{M}_2(x,y;b_1,b_2)=o_P(n^{-\frac{1}{2}+\eta}\tilde{b}_1^{-3}).$$

Together,

$$\begin{aligned}[\hat{r}^{(2,0)}(x,y;b_1,b_2)]^2 &= \left[r^{(2,0)}(x,y)\right]^2+O\left(\tilde{b}_1^2+\tilde{b}_1^{-2}n^{-\frac{1}{2}}+\tilde{b}_1^{-4}n^{-1}\right)\\ &+o_P(n^{-\gamma}(\tilde{b}_1^2+\tilde{b}_1^{-2}n^{-\frac{1}{2}}+\tilde{b}_1^{-4}n^{-1}+\tilde{b}_1^{-4}n^{-1-\gamma}))\\ &+o_P(n^{-\frac{1}{2}+\eta}\tilde{b}_1^{-3})+o_P(n^{-1+\eta}\tilde{b}_1^{-6})\end{aligned}$$

For $\hat{I}_{2,0}^{2,0}(b_1,b_2)=\int v(x,y)\left[\hat{r}^{(2,0)}(x,y)\right]^2\,dx\,dy$ some improvements are possible.

First, consider $\mathcal{V}_2$.

$$\begin{aligned}&\int_A v(x,y)a_{ik}(x,y;b_1,b_2)\,dx\,dy\\ &= \frac{1}{b_1^6b_2^2}\int_{A_k}\int_{A_i}\int_A v(x,y)K_{2,0}\left(B^{-1}\begin{pmatrix}x-u\\ y-v\end{pmatrix}\right)K_{2,0}\left(B^{-1}\begin{pmatrix}x-\bar{u}\\ y-\bar{v}\end{pmatrix}\right)\,dx\,dy\,du\,dv\,d\bar{u}\,d\bar{v}\\ &= \frac{1}{b_1^5b_2}\int_{A_k}\int_{A_i}\int_{A'} v\left(B\begin{pmatrix}z_1\\ z_2\end{pmatrix}+\begin{pmatrix}u\\ v\end{pmatrix}\right)K_{2,0}(z_1,z_2)\\ &\qquad\cdot K_{2,0}\left(\begin{pmatrix}z_1\\ z_2\end{pmatrix}+B^{-1}\begin{pmatrix}u-\bar{u}\\ v-\bar{v}\end{pmatrix}\right)\,dz_1\,dz_2\,du\,dv\,d\bar{u}\,d\bar{v}\\ &\le \frac{1}{b_1^5b_2}\lambda(A_k)\lambda(A_i)(\max_{(x,y)\in T}K_{2,0}(x,y))^2\\ &= O\left(\frac{1}{n^2b_1^5b_2}\right).\end{aligned}$$

The number of nonzero $\int_A v(x,y)a_{ik}(x,y;b_1,b_2)\,dx\,dy$ is $O(n^2 b_1 b_2)$.

Therefore

$$\begin{aligned}\int_A v(x,y)\mathcal{V}_2(x,y;b_1,b_2)\,dx\,dy &= o_P\left(n^{-1+\eta}b_1^{-\frac{9}{2}}b_2^{-\frac{1}{2}}\right)\\ &= o_P\left(n^{-1+\eta}b_1^{-5}\right).\end{aligned}$$

$\mathcal{M}_2(x,y;b_1,b_2)$ is again split in $\mathcal{M}_2(x,y;b_1,b_2)-\mathcal{M}_2(x,y;\tilde{b}_1,\tilde{b}_2)$ and $\mathcal{M}_2(x,y;\tilde{b}_1,\tilde{b}_2)$.

First, consider $\mathcal{M}_2(x,y;\tilde{b}_1,\tilde{b}_2)$.

$$\begin{aligned}&\left|\int_A v(x,y)c_i(x,y;\tilde{b}_1,\tilde{b}_2)\,dx\,dy\right|\\
&= \left|\frac{2}{\tilde{b}_1^5\tilde{b}_2}\int_A\int_{A_i}\int_T v(x,y)K_{2,0}\left(\tilde{B}^{-1}\begin{pmatrix}x-u\\y-v\end{pmatrix}\right)K_{2,0}(z_1,z_2)\right.\\
&\qquad\left.\cdot r\left(\begin{pmatrix}x\\y\end{pmatrix}-\tilde{B}\begin{pmatrix}z_1\\z_2\end{pmatrix}\right)dz_1\,dz_2\,du\,dv\,dx\,dy\right|\\
&= \left|\frac{2}{\tilde{b}_1^4}\int_{A_i}\int_T\int_T v\left(\tilde{B}\begin{pmatrix}t_1\\t_2\end{pmatrix}+\begin{pmatrix}u\\v\end{pmatrix}\right)K_{2,0}(t_1,t_2)K_{2,0}(z_1,z_2)\right.\\
&\qquad\left.\cdot r\left(\tilde{B}\begin{pmatrix}t_1\\t_2\end{pmatrix}+\begin{pmatrix}u\\v\end{pmatrix}-\tilde{B}\begin{pmatrix}z_1\\z_2\end{pmatrix}\right)dz_1\,dz_2\,dt_1\,dt_2\,du\,dv\right|\\
&\le \left|\frac{2}{\tilde{b}_1^4}\int_{A_i}\int_T v\left(\tilde{B}\begin{pmatrix}t_1\\t_2\end{pmatrix}+\begin{pmatrix}u\\v\end{pmatrix}\right)K_{2,0}(t_1,t_2)\right.\\
&\qquad\left.\cdot\left[r^{(2,0)}\left(\tilde{B}\begin{pmatrix}t_1\\t_2\end{pmatrix}+\begin{pmatrix}u\\v\end{pmatrix}\right)\tilde{b}_1^2+C(\tilde{b}_1^4+\tilde{b}_2^4)\right]dt_1\,dt_2\,du\,dv\right|\\
&\le \left|2\int_{A_i}[v(u,v)r^{(2,0)}(u,v)]^{(2,0)}\,du\,dv+\frac{C}{n}\left(1+\frac{\tilde{b}_2^4}{\tilde{b}_1^4}\right)\right|\\
&\le \frac{C}{n}\left(1+\frac{\tilde{b}_2^4}{\tilde{b}_1^4}\right)\\
&= O\left(\frac{1}{n}\right)\end{aligned}$$

A Taylor expansion of $r\left(\tilde{B}\binom{t_1}{t_2}+\binom{u}{v}-\tilde{B}\binom{z_1}{z_2}\right)$ around $\tilde{B}\binom{t_1}{t_2}+\binom{u}{v}$ and kernel properties are used.

The number of nonzero $\int_A v(x,y)c_i(x,y;\tilde{b}_1,\tilde{b}_2)\,dx\,dy$ is $O(n)$, therefore

$$\int_A v(x,y)\mathcal{M}_2(x,y;\tilde{b}_1,\tilde{b}_2)\,dx\,dy = o_P\left(n^{-\frac{1}{2}+\eta}\right).$$

The same holds for $\mathcal{M}_2(x,y;b_1,b_2)-\mathcal{M}_2(x,y;\tilde{b}_1,\tilde{b}_2)$.

$$\left|\int_A v(x,y)c_i(x,y;b_1,b_2)\,dx\,dy-\int_A v(x,y)c_i(x,y;\tilde{b}_1,b_2)\,dx\,dy\right|\le\frac{C}{n}$$

and

$$\left| \int_A v(x,y) c_i(x,y;\tilde{b}_1,b_2)\,dx\,dy - \int_A v(x,y) c_i(x,y;\tilde{b}_1,\tilde{b}_2)\,dx\,dy \right| \le \frac{C}{n}$$

because b_i and $\tilde{b}_i$ are of the same order for $i = 1,2$ (similar splitting as in the local case). So,

$$\mathcal{M}_2(x,y;b_1,b_2) = o_P\left(n^{-\frac{1}{2}+\eta}\right).$$

Together,

$$\begin{aligned} \hat{I}^{2,0}_{2,0}(b_1,b_2) &= \frac{1}{n b_1^5 b_2} M_{2,0} I_{\sigma,f}\,(1+o_P(1)) + I^{2,0}_{2,0} + O\left(\tilde{b}_1^2 + \tilde{b}_1^{-2} n^{-\frac{1}{2}} + \tilde{b}_1^{-4} n^{-1}\right) \\ &\quad + o_P\left(n^{-\gamma}(\tilde{b}_1^2 + \tilde{b}_1^{-2} n^{-\frac{1}{2}} + \tilde{b}_1^{-4} n^{-1} + \tilde{b}_1^{-4} n^{-1-\gamma})\right) + o_P\left(n^{-1+\eta} b_1^{-5}\right) + o_P\left(n^{-\frac{1}{2}+\eta}\right), \end{aligned}$$

Where $M_{2,0} = \int_T K^2_{2,0}(x,y)\,dx\,dy$, $I_{\sigma,f} = \int_T v(x,y) \frac{\sigma^2(x,y)}{f(x,y)}\,dx\,dy$ and $I^{2,0}_{2,0} = \int_T v(x,y) \left[r^{(2,0)}(x,y)\right]^2\,dx\,dy$.

The considerations for $[\hat{r}^{(0,2)}(x,y;b_1,b_2)]^2$ and $\hat{I}^{0,2}_{0,2}(b_1,b_2)$ are similar to above calculations, but with b_1 instead of b_2 and vice versa and $K_{0,2}$ and $r^{(0,2)}$ instead of $K_{2,0}$ and $r^{(2,0)}$. So,

$$\begin{aligned} [\hat{r}^{(0,2)}(x,y;b_1,b_2)]^2 &= \left[r^{(0,2)}(x,y)\right]^2 + O\left(\tilde{b}_1^2 + \tilde{b}_1^{-2} n^{-\frac{1}{2}} + \tilde{b}_1^{-4} n^{-1}\right) \\ &\quad + o_P(n^{-\gamma}(\tilde{b}_1^2 + \tilde{b}_1^{-2} n^{-\frac{1}{2}} + \tilde{b}_1^{-4} n^{-1} + \tilde{b}_1^{-4} n^{-1-\gamma})) \\ &\quad + o_P(n^{-\frac{1}{2}+\eta} \tilde{b}_1^{-3}) + o_P(n^{-1+\eta} \tilde{b}_1^{-6}) \end{aligned}$$

and

$$\begin{aligned} \hat{I}^{0,2}_{0,2}(b_1,b_2) &= \frac{1}{n b_1 b_2^5} M_{0,2} I_{\sigma,f}\,(1+o_P(1)) + I^{0,2}_{0,2} + O\left(\tilde{b}_1^2 + \tilde{b}_1^{-2} n^{-\frac{1}{2}} + \tilde{b}_1^{-4} n^{-1}\right) \\ &\quad + o_P\left(n^{-\gamma}(\tilde{b}_1^2 + \tilde{b}_1^{-2} n^{-\frac{1}{2}} + \tilde{b}_1^{-4} n^{-1} + \tilde{b}_1^{-4} n^{-1-\gamma})\right) + o_P\left(n^{-1+\eta} b_1^{-5}\right) + o_P\left(n^{-\frac{1}{2}+\eta}\right), \end{aligned}$$

where $M_{0,2} = \int_T K^2_{0,2}(x,y)\,dx\,dy$.

The lines of thoughts for $\hat{I}^{2,0}_{0,2}$ are similar to those for $\hat{I}^{2,0}_{2,0}$ and $\hat{I}^{0,2}_{0,2}$, so

$$\begin{aligned} \hat{I}^{2,0}_{0,2}(b_1,b_2) &= \frac{1}{n b_1^3 b_2^3} M^{2,0}_{0,2} I_{\sigma,f}\,(1+o_P(1)) + I^{2,0}_{0,2} + O\left(\tilde{b}_1^2 + \tilde{b}_1^{-2} n^{-\frac{1}{2}} + \tilde{b}_1^{-4} n^{-1}\right) \\ &\quad + o_P\left(n^{-\gamma}(\tilde{b}_1^2 + \tilde{b}_1^{-2} n^{-\frac{1}{2}} + \tilde{b}_1^{-4} n^{-1} + \tilde{b}_1^{-4} n^{-1-\gamma})\right) + o_P\left(n^{-1+\eta} b_1^{-5}\right) + o_P\left(n^{-\frac{1}{2}+\eta}\right), \end{aligned}$$

where $M^{2,0}_{0,2} = \int_T K_{2,0}(x,y) K_{0,2}(x,y)\,dx\,dy$.

Now the behaviour of the bandwidth estimators during the iteration steps can be investigated.

The initial bandwidths are $\hat{b}_1^{(0)} = \hat{b}_2^{(0)} = n^{-\frac{1}{2}}$, $\hat{I}_{\sigma,f}$ and $\frac{\hat{\sigma}^2(x,y)}{\hat{f}(x,y)}$ estimators with

$$\hat{I}_{\sigma,f} = I_{\sigma,f}\left(1 + O\left(n^{-\frac{1}{6}}\right) + O_P\left(n^{-\frac{1}{2}}\right)\right)$$

and

$$\frac{\hat{\sigma}^2(x,y)}{\hat{f}(x,y)} = \frac{\sigma^2(x,y)}{f(x,y)}\left(1 + O\left(n^{-\frac{1}{6}}\right) + O_P\left(n^{-\frac{1}{4}}\right)\right).$$

First, global steps are performed.

During the first iteration step

$$\begin{aligned}
\hat{I}_{2,0}^{2,0}(b_1^{(0)} n^{\frac{1}{12}}, b_2^{(0)} n^{\frac{1}{12}}) &= n^{\frac{3}{2}} M_{2,0} I_{\sigma,f}\left(1 + o_P(1)\right) \\
\hat{I}_{0,2}^{0,2}(b_1^{(0)} n^{\frac{1}{12}}, b_2^{(0)} n^{\frac{1}{12}}) &= n^{\frac{3}{2}} M_{0,2} I_{\sigma,f}\left(1 + o_P(1)\right) \\
\hat{I}_{0,2}^{2,0}(b_1^{(0)} n^{\frac{1}{12}}, b_2^{(0)} n^{\frac{1}{12}}) &= n^{\frac{3}{2}} M_{0,2}^{2,0} I_{\sigma,f}\left(1 + o_P(1)\right),
\end{aligned}$$

so

$$\begin{aligned}
\hat{b}_1^{(1)} &= C_K^{\frac{1}{6}}\left(\frac{M_{0,2}}{M_{2,0}}\right)^{\frac{1}{8}} n^{-\frac{5}{12}}(1 + o_P(1)) \\
\hat{b}_2^{(1)} &= C_K^{\frac{1}{6}}\left(\frac{M_{2,0}}{M_{0,2}}\right)^{\frac{1}{8}} n^{-\frac{5}{12}}(1 + o_P(1)),
\end{aligned}$$

where $C_K = \frac{M}{\mu^2(\sqrt{M_{2,0}M_{0,2}} + M_{0,2}^{2,0})}$.

All the assumptions made at the beginning of the proof are fulfilled.

For the second iteration step

$$\begin{aligned}
\hat{I}_{2,0}^{2,0}(\hat{b}_1^{(1)} n^{\frac{1}{12}}, \hat{b}_2^{(1)} n^{\frac{1}{12}}) &= n M_{2,0} I_{\sigma,f} C_K^{-1}\left(\frac{M_{2,0}}{M_{0,2}}\right)^{\frac{1}{2}}(1 + o_P(1)) \\
\hat{I}_{0,2}^{0,2}(\hat{b}_1^{(1)} n^{\frac{1}{12}}, \hat{b}_2^{(1)} n^{\frac{1}{12}}) &= n M_{0,2} I_{\sigma,f} C_K^{-1}\left(\frac{M_{0,2}}{M_{2,0}}\right)^{\frac{1}{2}}(1 + o_P(1)) \\
\hat{I}_{0,2}^{2,0}(\hat{b}_1^{(1)} n^{\frac{1}{12}}, \hat{b}_2^{(1)} n^{\frac{1}{12}}) &= n M_{0,2}^{2,0} I_{\sigma,f} C_K^{-1}\left(1 + o_P(1)\right),
\end{aligned}$$

so

$$\begin{aligned}
\hat{b}_1^{(2)} &= C_K^{\frac{2}{6}}\left(\frac{M_{0,2}}{M_{2,0}}\right)^{\frac{1}{8}} n^{-\frac{4}{12}}(1 + o_P(1)) \\
\hat{b}_2^{(2)} &= C_K^{\frac{2}{6}}\left(\frac{M_{2,0}}{M_{0,2}}\right)^{\frac{1}{8}} n^{-\frac{4}{12}}(1 + o_P(1))
\end{aligned}$$

and so on until

$$\begin{aligned}\hat{b}_1^{(3)} &= C_K^{\frac{3}{6}} \left(\frac{M_{0,2}}{M_{2,0}}\right)^{\frac{1}{8}} n^{-\frac{3}{12}}(1+o_P(1)) \\ \hat{b}_2^{(3)} &= C_K^{\frac{3}{6}} \left(\frac{M_{2,0}}{M_{0,2}}\right)^{\frac{1}{8}} n^{-\frac{3}{12}}(1+o_P(1)).\end{aligned}$$

Now, the constant terms are coming into account.

$$\begin{aligned}\hat{I}_{2,0}^{2,0}(\hat{b}_1^{(3)} n^{\frac{1}{12}}, \hat{b}_2^{(3)} n^{\frac{1}{12}}) &= M_{2,0} I_{\sigma,f} C_K^{-3} \left(\frac{M_{2,0}}{M_{0,2}}\right)^{\frac{1}{2}} (1+o_P(1)) + I_{2,0}^{2,0} \\ \hat{I}_{0,2}^{0,2}(\hat{b}_1^{(3)} n^{\frac{1}{12}}, \hat{b}_2^{(3)} n^{\frac{1}{12}}) &= M_{0,2} I_{\sigma,f} C_K^{-3} \left(\frac{M_{0,2}}{M_{2,0}}\right)^{\frac{1}{2}} (1+o_P(1)) + I_{0,2}^{0,2} \\ \hat{I}_{0,2}^{2,0}(\hat{b}_1^{(3)} n^{\frac{1}{12}}, \hat{b}_2^{(3)} n^{\frac{1}{12}}) &= M_{0,2}^{2,0} I_{\sigma,f} C_K^{-3} (1+o_P(1)) + I_{0,2}^{2,0},\end{aligned}$$

so

$$\begin{aligned}\hat{b}_1^{(4)} &= C_1 n^{-\frac{2}{12}}(1+o_P(1)) \\ \hat{b}_2^{(4)} &= C_2 n^{-\frac{2}{12}}(1+o_P(1)),\end{aligned}$$

where C_1 and C_2 are constants depending on kernel constants, second derivatives of $r(x,y)$, $\sigma^2(x,y)$ and $f(x,y)$.

After 4 global steps the bandwidth estimators have reached the right order $n^{-\frac{1}{6}}$. Now other terms in the expansion are dominating.

The procedure can be continued like the first 4 steps to obtain global bandwidth estimators or can use the formulas for the local case.

First, the global branch is considered. The order of the terms in $\hat{I}_{2,0}^{2,0}$, $\hat{I}_{0,2}^{0,2}$ and $\hat{I}_{0,2}^{2,0}$ is such, that γ is becoming important. After 4 steps γ is now zero.

$$\begin{aligned}\hat{I}_{2,0}^{2,0}(\hat{b}_1^{(4)} n^{\frac{1}{12}}, \hat{b}_2^{(4)} n^{\frac{1}{12}}) &= I_{2,0}^{2,0} + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{6}} + n^{-\frac{1}{2}+\eta}\right) \\ \hat{I}_{0,2}^{0,2}(\hat{b}_1^{(4)} n^{\frac{1}{12}}, \hat{b}_2^{(4)} n^{\frac{1}{12}}) &= I_{0,2}^{0,2} + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{6}} + n^{-\frac{1}{2}+\eta}\right) \\ \hat{I}_{0,2}^{2,0}(\hat{b}_1^{(4)} n^{\frac{1}{12}}, \hat{b}_2^{(4)} n^{\frac{1}{12}}) &= I_{0,2}^{2,0} + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{6}} + n^{-\frac{1}{2}+\eta}\right),\end{aligned}$$

$$\begin{aligned}\hat{b}_1^{(5)} &= b_{1_{IASY}} \left(1 + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{6}} + n^{-\frac{1}{2}+\eta}\right)\right) \\ \hat{b}_2^{(5)} &= b_{2_{IASY}} \left(1 + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{6}} + n^{-\frac{1}{2}+\eta}\right)\right).\end{aligned}$$

This means, that now $\gamma = \frac{1}{6}$.

$$\begin{aligned}
\hat{I}_{2,0}^{2,0}(\hat{b}_1^{(5)} n^{\frac{1}{12}}, \hat{b}_2^{(5)} n^{\frac{1}{12}}) &= I_{2,0}^{2,0} + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{3}} + n^{-\frac{1}{2}+\eta}\right) \\
\hat{I}_{0,2}^{0,2}(\hat{b}_1^{(5)} n^{\frac{1}{12}}, \hat{b}_2^{(5)} n^{\frac{1}{12}}) &= I_{0,2}^{0,2} + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{3}} + n^{-\frac{1}{2}+\eta}\right) \\
\hat{I}_{0,2}^{2,0}(\hat{b}_1^{(5)} n^{\frac{1}{12}}, \hat{b}_2^{(5)} n^{\frac{1}{12}}) &= I_{0,2}^{2,0} + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{3}} + n^{-\frac{1}{2}+\eta}\right),
\end{aligned}$$

$$\begin{aligned}
\hat{b}_1^{(6)} &= b_{1_{IASY}}\left(1 + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{3}} + n^{-\frac{1}{2}+\eta}\right)\right) \\
\hat{b}_2^{(6)} &= b_{2_{IASY}}\left(1 + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{3}} + n^{-\frac{1}{2}+\eta}\right)\right).
\end{aligned}$$

Now, $\gamma = \frac{1}{3}$.

$$\begin{aligned}
\hat{I}_{2,0}^{2,0}(\hat{b}_1^{(6)} n^{\frac{1}{12}}, \hat{b}_2^{(6)} n^{\frac{1}{12}}) &= I_{2,0}^{2,0} + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{2}+\eta}\right) \\
\hat{I}_{0,2}^{0,2}(\hat{b}_1^{(6)} n^{\frac{1}{12}}, \hat{b}_2^{(6)} n^{\frac{1}{12}}) &= I_{0,2}^{0,2} + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{2}+\eta}\right) \\
\hat{I}_{0,2}^{2,0}(\hat{b}_1^{(6)} n^{\frac{1}{12}}, \hat{b}_2^{(6)} n^{\frac{1}{12}}) &= I_{0,2}^{2,0} + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{2}+\eta}\right),
\end{aligned}$$

$$\begin{aligned}
\hat{b}_1^{(7)} &= b_{1_{IASY}}\left(1 + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{2}+\eta}\right)\right) \\
\hat{b}_2^{(7)} &= b_{2_{IASY}}\left(1 + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{2}+\eta}\right)\right).
\end{aligned}$$

No further improvement is possible, so the iteration stops after 7 steps. $b_{i_{IASY}}$ can be replaced by $b_{i_{MISE}}$ for $i = 1, 2$ according to Remark 2.10.

Now, the local branch is considered.

$$\begin{aligned}
\left[\hat{r}^{(2,0)}(x, y; \hat{b}_1^{(4)} n^{\frac{1}{12}}, \hat{b}_2^{(4)} n^{\frac{1}{12}})\right]^2 &= \left[r^{(2,0)}(x, y)\right]^2 + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{6}} + n^{-\frac{1}{4}+\eta}\right) \\
\left[\hat{r}^{(0,2)}(x, y; \hat{b}_1^{(4)} n^{\frac{1}{12}}, \hat{b}_2^{(4)} n^{\frac{1}{12}})\right]^2 &= \left[r^{(0,2)}(x, y)\right]^2 + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{6}} + n^{-\frac{1}{4}+\eta}\right)
\end{aligned}$$

$$\begin{aligned}
\hat{b}_1^{(5)}(x, y) &= b_{1_{ASY}}(x, y)\left(1 + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{6}}\right)\right) \\
\hat{b}_2^{(5)}(x, y) &= b_{2_{ASY}}(x, y)\left(1 + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{6}}\right)\right).
\end{aligned}$$

Now, $\gamma = \frac{1}{6}$.

$$\left[\hat{r}^{(2,0)}(x, y; \hat{b}_1^{(5)} n^{\frac{1}{12}}, \hat{b}_2^{(5)} n^{\frac{1}{12}})\right]^2 = \left[r^{(2,0)}(x, y)\right]^2 + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{2}{6}} + n^{-\frac{1}{4}+\eta}\right)$$

$$\left[\hat{r}^{(0,2)}(x,y;\hat{b}_1^{(5)}n^{\frac{1}{12}},\hat{b}_2^{(5)}n^{\frac{1}{12}})\right]^2 = \left[r^{(0,2)}(x,y)\right]^2 + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{2}{6}} + n^{-\frac{1}{4}+\eta}\right)$$

$$\hat{b}_1^{(6)}(x,y) = b_{1_{ASY}}(x,y)\left(1 + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{4}+\eta}\right)\right)$$
$$\hat{b}_2^{(6)}(x,y) = b_{2_{ASY}}(x,y)\left(1 + O\left(n^{-\frac{1}{6}}\right) + o_P\left(n^{-\frac{1}{4}+\eta}\right)\right).$$

No further improvement is possible, so the iteration stops after 4 global and 2 local steps. $b_{i_{ASY}}(x,y)$ can be replaced by $b_{i_{MSE}}(x,y)$ for $i = 1,2$ according to Remark 2.10. ■

4.2 The Inflation Factor

The inflation factor n^β influences strongly the number of iteration steps and the rate of convergence of the bandwidth estimators in Theorem 4.1. Some arguments for the choice of the inflation factor $n^{\frac{1}{12}}$ are given and alternatives discussed.

It is distinguished between the global and the local case. For the global case the terms $O\left(n^{-\frac{1}{6}}\right)$ and $O_P\left(n^{-\frac{1}{2}}\right)$ in Theorem 4.1 come from the terms

$$O\left(\tilde{b}_1^2\right) = O\left(n^{-\frac{2}{6}+2\beta}\right)$$

(the bias term $\mathcal{B}$) and

$$O_P\left(n^{-1}b_1^{-6}\right) = O_P\left(n^{-6\beta}\right)$$

(the variance term $\mathcal{V}$) for $b_1 = n^{-\frac{1}{6}+\beta}$. The first is increasing in β and the second decreasing.

If both terms are balanced, then $\beta = \frac{1}{24}$, leading to $O\left(n^{-\frac{1}{4}}\right)$ and $O_P\left(n^{-\frac{1}{4}}\right)$. But with the replacement of $b_{i_{ASY}}(x,y)$ by $b_{i_{MSE}}(x,y)$ a term $o\left(n^{-\frac{1}{6}}\right)$ for arbitrary fixed design and $O\left(n^{-\frac{1}{4}}\right)$ for equidistant design is introduced. So, for arbitrary fixed design it is not necessary to balance both terms, but the variance term can be decreased to $O_P\left(n^{-\frac{1}{2}}\right)$ by choosing $\beta = \frac{1}{12}$. This choice is variance optimal.

In the local case the variance term is $O_P\left(n^{-\frac{1}{2}}b_1^{-3}\right) = O_P\left(n^{-3\beta}\right)$. So, in the balanced situation $\beta = \frac{1}{15}$. Again, $\beta = \frac{1}{12}$ is variance optimal, reducing the variance term to $O_P\left(n^{-\frac{1}{4}}\right)$.

The rate of $b_i^{(j)}$, $i = 1, 2$, is increased in each step by β, the rate of the inflation factor. Therefore,

$$\left[\frac{\frac{1}{2} - \frac{1}{6}}{\beta}\right]$$

steps are necessary to drive the bandwidth estimators from the initial order $n^{\frac{1}{2}}$ to the optimal order $n^{\frac{1}{6}}$. These are 4 steps for $\beta = \frac{1}{12}$ and 8 steps for $\beta = \frac{1}{24}$.

After these steps $o_P\left(n^{\gamma}\tilde{b}_1^2\right)$ (stemming from the bias part $\mathcal{B}$) is becoming important. Now $\gamma = 0$. In the balanced situation ($\beta = \frac{1}{24}$ for the global or $\beta = \frac{1}{15}$ for the local case) no further improvement is possible. So, after one additional step the optimal order is achieved. In the variance optimal situation the order of the o_P-term can be reduced by increasing γ until it reaches $O_P\left(n^{-\frac{1}{2}}\right)$. in the global case or $O_P\left(n^{-\frac{1}{4}}\right)$ in the local case. So, there are 2 additional steps in the local case and 3 additional steps in the global case increasing γ from 0 to $\frac{1}{6}$ and finally $\frac{1}{3}$ (see the proof of Theorem 4.1).

4.3 Generalizations

4.3.1 Three Bandwidth Parameters

In the algorithm in Chapter 3.1 two bandwidth parameters are estimated. But it is also possible to estimate the third parameter as described in Chapter 3.2.

There is an additional term $[\hat{r}^{(1,1)}(x, y; b_1, b_2)]^2$. Similar to above calculations, it is

$$\begin{aligned}
[\hat{r}^{(1,1)}(x, y; b_1, b_2)]^2 &= \left[r^{(1,1)}(x, y)\right]^2 + O\left(\tilde{b}_1^2 + \tilde{b}_1^{-2}n^{-\frac{1}{2}} + \tilde{b}_1^{-4}n^{-1}\right) \\
&\quad + o_P(n^{-\gamma}(\tilde{b}_1^2 + \tilde{b}_1^{-2}n^{-\frac{1}{2}} + \tilde{b}_1^{-4}n^{-1} + \tilde{b}_1^{-4}n^{-1-\gamma})) \\
&\quad + o_P(n^{-\frac{1}{2}+\eta}\tilde{b}_1^{-3}) + o_P(n^{-1+\eta}\tilde{b}_1^{-6}).
\end{aligned}$$

Actually, the rates can be improved, if $K_{1,1}$ is of order 5. But then other rates in u and z are dominating.

The quantities u and z are estimated by

$$\hat{u} = \left[r^{(2,0)}(x, y)c^2 + r^{(0,2)}(x, y)s^2 - 2r^{(1,1)}(x, y)sc\right] \cdot \left(1 + O\left(\tilde{b}_1^2 + \tilde{b}_1^{-2}n^{-\frac{1}{2}} + \tilde{b}_1^{-4}n^{-1}\right)\right.$$

$$+o_P\left(n^{-\gamma}(\tilde{b}_1^2+\tilde{b}_1^{-2}n^{-\frac{1}{2}}+\tilde{b}_1^{-4}n^{-1}+\tilde{b}_1^{-4}n^{-1-\gamma})\right)+o_P\left(n^{-\frac{1}{2}+\eta}\tilde{b}_1^{-3}\right)+o_P\left(n^{-1+\eta}\tilde{b}_1^{-6}\right)\Big)$$

and

$$\begin{aligned}\hat{z} &= \left[r^{(2,0)}(x,y)s^2+r^{(0,2)}(x,y)c^2+2r^{(1,1)}(x,y)sc\right]\cdot\Big(1+O\left(\tilde{b}_1^2+\tilde{b}_1^{-2}n^{-\frac{1}{2}}+\tilde{b}_1^{-4}n^{-1}\right)\\ &\quad +o_P\left(n^{-\gamma}(\tilde{b}_1^2+\tilde{b}_1^{-2}n^{-\frac{1}{2}}+\tilde{b}_1^{-4}n^{-1}+\tilde{b}_1^{-4}n^{-1-\gamma})\right)+o_P\left(n^{-\frac{1}{2}+\eta}\tilde{b}_1^{-3}\right)+o_P\left(n^{-1+\eta}\tilde{b}_1^{-6}\right)\Big).\end{aligned}$$

So, the estimators for b_1 and b_2 achieve the same rate as in the case with two parameters. The third parameter helps to get better results for finite sample sizes and reduces the leading term of the MSE.

4.3.2 Additional smoothing

In step 3 of the algorithms in Chapter 3.1 and 3.2 there is an additional smoothing of the pure local bandwidth estimators. It does not change the rates, because the smoothing is done with a smaller bandwidth, so only smaller additional terms occur.

4.4 Asymptotic Normality of the Estimator

It is not only possible to derive rates of convergence of the bandwidth estimator, but the distribution can also be calculated.

Theorem 4.2 *Under the assumptions of Theorem 4.1, equidistant design and constant variance of the noise the bandwidth estimators $\hat{b}_1^{(6)}(x,y)$ and $\hat{b}_2^{(6)}(x,y)$ are asymptotically normal with expectation $b_{1_{ASY}}$ and $b_{2_{ASY}}$ and bias terms of order $O\left(n^{-\frac{1}{3}}\right)$.*

Proof: Let $\widehat{AMSE}_{\hat{r}}(B)$ denote the estimator obtained by substituting the second derivatives and the variance by their estimators. The dependence on (x,y) is omitted for a simpler notation. Let denote the bandwidth estimator of Chapter 3.1 by $\hat{B}$. The estimators for the second derivatives are obtained by using $\hat{B}n^{\frac{1}{12}}$.

$$\begin{aligned}0 &= \frac{\partial}{\partial b_1}\widehat{AMSE}_{\hat{r}}(\hat{B})\\ &= \frac{\partial}{\partial b_1}AMSE_{\hat{r}}(\hat{B})-\frac{\partial}{\partial b_1}AMSE_{\hat{r}}(B_{ASY})+\frac{\partial}{\partial b_1}\widehat{AMSE}_{\hat{r}}(\hat{B})-\frac{\partial}{\partial b_1}AMSE_{\hat{r}}(\hat{B})\end{aligned}$$

A Taylor expansion of $\frac{\partial}{\partial b_1} AMSE_{\hat{r}}(\hat{B})$ around B_{ASY}, the fact that $D^\nu \frac{\partial}{\partial b_1} AMSE_{\hat{r}}(\hat{B})$ and $D^\nu \frac{\partial}{\partial b_1} AMSE_{\hat{r}}(B_{ASY})$ are of order $n^{-\frac{1}{6}}$ for $|\nu| = 2$ and (from Theorem 4.1 and Remark 2.10)

$$(\hat{b}_1 - b_{1_{ASY}})^2 = O\left(n^{-\frac{2}{6}\cdot 2}\right) + O_P\left(n^{-\frac{5}{12}\cdot 2}\right) \tag{63}$$

and

$$(\hat{b}_2 - b_{2_{ASY}})^2 = O\left(n^{-\frac{2}{6}\cdot 2}\right) + O_P\left(n^{-\frac{5}{12}\cdot 2}\right) \tag{64}$$

yield

$$\begin{aligned} 0 &= \sum_{|\nu|=1} D^\nu \frac{\partial}{\partial b_1} AMSE_{\hat{r}}(B_{ASY})(\hat{B} - B_{ASY})^\nu + \frac{\partial}{\partial b_1} \widehat{AMSE}_{\hat{r}}(\hat{B}) - \frac{\partial}{\partial b_1} AMSE_{\hat{r}}(\hat{B}) \\ &\quad + O\left(n^{-\frac{5}{6}}\right) + O_P\left(n^{-1}\right). \end{aligned}$$

With

$$\begin{aligned} &\frac{\partial}{\partial b_1} \widehat{AMSE}_{\hat{r}}(\hat{B}) - \frac{\partial}{\partial b_1} AMSE_{\hat{r}}(\hat{B}) \\ &= \mu^2 \hat{b}_1 \Big[\hat{b}_1^2 \Big((r^{(2,0)}(x,y))^2 - (\hat{r}^{(2,0)}(x,y;\hat{b}_1 n^{\frac{1}{12}}, \hat{b}_2 n^{\frac{1}{12}}))^2\Big) \\ &\qquad + \hat{b}_2^2 \Big(r^{(2,0)}(x,y) r^{(0,2)}(x,y) - \hat{r}^{(2,0)}(x,y;\hat{b}_1 n^{\frac{1}{12}}, \hat{b}_2 n^{\frac{1}{12}}) \hat{r}^{(0,2)}(x,y;\hat{b}_1 n^{\frac{1}{12}}, \hat{b}_2 n^{\frac{1}{12}})\Big)\Big] \\ &\quad - \frac{M}{n \hat{b}_1^2 \hat{b}_2} \left(\sigma^2 - \hat{\sigma}^2\right) \end{aligned}$$

and

$$\begin{aligned} &\left[\hat{r}^{(2,0)}(x,y;b_{1_{ASY}} n^{\frac{1}{12}}, b_{2_{ASY}} n^{\frac{1}{12}})\right]^2 \\ &\quad = \left[r^{(2,0)}(x,y)\right]^2 + \sum_{i=1}^n c_i(x,y;b_{1_{ASY}} n^{\frac{1}{12}}, b_{2_{ASY}} n^{\frac{1}{12}})\epsilon_i + O\left(n^{-\frac{1}{6}}\right) + o_p\left(n^{-\frac{1}{6}}\right) \end{aligned}$$

and

$$\begin{aligned} &\left[\hat{r}^{(0,2)}(x,y;b_{1_{ASY}} n^{\frac{1}{12}}, b_{2_{ASY}} n^{\frac{1}{12}})\right]^2 \\ &\quad = \left[r^{(0,2)}(x,y)\right]^2 + \sum_{i=1}^n c_i(x,y;b_{1_{ASY}} n^{\frac{1}{12}}, b_{2_{ASY}} n^{\frac{1}{12}})\epsilon_i + O\left(n^{-\frac{1}{6}}\right) + o_p\left(n^{-\frac{1}{6}}\right) \end{aligned}$$

(see proof of Theorem 4.1) and equation (63) and (64) it is

$$\begin{aligned} 0 &= \sum_{|\nu|=1} D^\nu \frac{\partial}{\partial b_1} AMSE_{\hat{r}}(B_{ASY})(\hat{B} - B_{ASY})^\nu \\ &\quad + \mu^2 b_{1_{ASY}} \left[b_{1_{ASY}}^2 + b_{2_{ASY}}^2\right] \sum_{i=1}^n c_i(x,y;b_{1_{ASY}} n^{\frac{1}{12}}, b_{2_{ASY}} n^{\frac{1}{12}})\epsilon_i + O\left(n^{-\frac{4}{6}}\right) + o_p\left(n^{-\frac{4}{6}}\right). \end{aligned}$$

The term with $\sigma^2 - \hat{\sigma}^2$ is of smaller order and has been neglected, since $\operatorname{var} \hat{\sigma}^2 = O(n^{-1})$.

With similar considerations for $\frac{\partial}{\partial b_2} \widehat{AMSE}_{\hat{r}}(\hat{B})$

$$
\begin{aligned}
0 &= \sum_{|\nu|=1} D^{\nu} \frac{\partial}{\partial b_2} AMSE_{\hat{r}}(B_{ASY})(\hat{B} - B_{ASY})^{\nu} \\
&\quad + \mu^2 b_{2_{ASY}} \left[b^2_{1_{ASY}} + b^2_{2_{ASY}} \right] \sum_{i=1}^{n} c_i(x, y; b_{1_{ASY}} n^{\frac{1}{12}}, b_{2_{ASY}} n^{\frac{1}{12}}) \epsilon_i + O\left(n^{-\frac{4}{6}}\right) + o_p\left(n^{-\frac{4}{6}}\right).
\end{aligned}
$$

$o_p\left(n^{-\frac{4}{6}}\right)$ is smaller then the middle term (with similar calculations as in the proof of Theorem 4.1), so it can be omitted.

Solving this system of equations leads to

$$
\begin{aligned}
\hat{b}_1 &= b_{1_{ASY}} + O\left(n^{-\frac{2}{6}}\right) \\
&\quad + \left(b_{1_{ASY}} \frac{\partial^2}{\partial b_1 \partial b_2} AMSE_{\hat{r}}(B_{ASY}) - b_{2_{ASY}} \frac{\partial^2}{\partial b_2^2} AMSE_{\hat{r}}(B_{ASY}) \right) \\
&\quad \frac{1}{N} \mu^2 \left[b^2_{1_{ASY}} + b^2_{2_{ASY}} \right] \sum_{i=1}^{n} c_i(x, y; b_{1_{ASY}} n^{\frac{1}{12}}, b_{2_{ASY}} n^{\frac{1}{12}}) \epsilon_i
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{b}_2 &= b_{2_{ASY}} + O\left(n^{-\frac{2}{6}}\right) \\
&\quad + \left(b_{2_{ASY}} \frac{\partial^2}{\partial b_1 \partial b_2} AMSE_{\hat{r}}(B_{ASY}) - b_{1_{ASY}} \frac{\partial^2}{\partial b_1^2} AMSE_{\hat{r}}(B_{ASY}) \right) \\
&\quad \frac{1}{N} \mu^2 \left[b^2_{1_{ASY}} + b^2_{2_{ASY}} \right] \sum_{i=1}^{n} c_i(x, y; b_{1_{ASY}} n^{\frac{1}{12}}, b_{2_{ASY}} n^{\frac{1}{12}}) \epsilon_i,
\end{aligned}
$$

where

$$
N = \frac{\partial^2}{\partial b_1^2} AMSE_{\hat{r}}(B_{ASY}) \frac{\partial^2}{\partial b_2^2} AMSE_{\hat{r}}(B_{ASY}) - \left(\frac{\partial^2}{\partial b_1 \partial b_2} AMSE_{\hat{r}}(B_{ASY}) \right)^2.
$$

So, $\hat{b}_1$ and $\hat{b}_2$ are asymptotic normal as weighted sums of independent random variables. ■

5 Acknowledgement

This article is part of the author's PhD thesis which was conducted under the supervision of Prof. Jürgen Franke at the Universität Kaiserslautern, Germany.

References

Michael Brockmann. *Local Bandwidth Selection in Nonparametric Kernel Regression.* Dissertation, Universität Heidelberg, 1993. Shaker, Aachen.

Michael Brockmann, Theo Gasser, and Eva Herrmann. Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association*, 88(424):1302–1309, 1993.

Jianqing Fan and Irène Gijbels. Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. R. Stat. Soc., Ser. B*, 57(2):371–394, 1995.

F. Godtliebsen, E. Spjøtvoll, and J. S. Marron. A nonlinear gaussian filter applied to images with discontinuities. *Journal of Nonparametric Statistics*, 8(1):21–43, 1997.

Peter Hall, J. Stephen Marron, and D. M. Titterington. On partial local smoothing rules for curve estimation. *Biometrika*, 82(3):575–587, 1995.

Eva Herrmann. Local bandwidth choice in kernel regression estimation. *Journal of Computational and Graphical Statistics*, 6(1):35–54, 1997.

Eva Herrmann, M. P. Wand, Joachim Engel, and Theo Gasser. A bandwidth selector for bivariate kernel regression. *J. R. Statist. Soc. B*, 57(1), 1995.

Enno Mammen and Irène Gijbels. On local adaptivity of kernel estimates with plug-in local bandwidth selectors. Preprint, 1995.

Hans-Georg Müller. *Nonparametric Regression Analysis of Longitudinal Data.* Lecture Notes in Statistics. Springer, Berlin Heidelberg, 1988.

Jörg Polzehl and Vladimir G. Spokoiny. Image denoising: Pointwise adaptive approach. Discussion Paper 38, SFB 373, Humboldt-Universität, Berlin, 1998.

David W. Scott. *Multivariate Density Estimation.* Wiley series in probability and mathematical statistics. John Wiley, New York, 1992.

Jeffrey S. Simonoff. *Smoothing Methods in Statistics.* Springer Series in Statistics. Springer, New York, 1998.

P. Whittle. Bounds for the moments of linear and quadratic forms in independent variables. *Theory of Probability and its Applications*, V(3):302–305, 1960.