An Iterative Plug-in Algorithm for Optimal Bandwidth Selection in Kernel Intensity Estimation for Spatial Data

Vom Fachbereich Mathematik der Technischen Universität Kaiserslautern zur

Verleihung des akademischen Grades Doktor der Naturwissenschaften

(Doktor rerum naturalium, Dr. rer. nat.) genehmigte Dissertation

von

Pak Hang Lo

Gutachten:

Prof. Dr. Jürgen Franke, Technische Universität Kaiserslautern

Prof. Dr. Marie-Colette van Lieshout, Universiteit Twente

Datum der Disputation: 19. Januar 2018

D 386

Acknowledgements

I would like to express the most sincere gratitude to my supervisor, Prof. Dr. Jürgen Franke, for considering my application and recommending me for the research training group *Graduiertenkolleg* (GrK) 1932 in year 2014. If this had not happened, I would not have the opportunity to explore another research field with which I was not familiar. I am extremely grateful to his continuous support, patience and encouragement throughout my study. Without him the completion of this thesis is impossible.

I thank the board of Principal Investigators for accepting me into GrK 1932. Special thanks to *Deutsche Forschungsgemeinschaft* (DFG) for their funding which supports my study. I thank also Dr. habil. Christoph Lossen for his very helpful administrative support, Dr. Falk Triebsch and Ms. Jessica Borsche for supporting us as Ph.D. students.

Credits must go to the members of Statistics Group of Department of Mathematics, University of Kaiserslautern, especially Prof. Dr. Claudia Redenbach for her insightful advices and professional knowledge in point processes, and Mrs. Beate Siegler for her sincere help. I will never forget the friendship with Ms. Yuli Afrianti, Dr. Euna Nyarige, Dr. Danilo Pezo, Ms. Yolanda Rodríguez, Dr. Rumana Rois and Mr. Diego Roldán. Their care and support are invaluable. I am also deeply grateful to Ms. Lihua Chen from Financial Mathematics Group and her husband, Mr. Shihua Chen, who selflessly offered me help with the issues in my daily life.

I thank my family members including my parents and brother for their invaluable care, understanding and support throughout my life. For many years, my parents are very responsible and work very hard to support the family. My brother is a very good listener and gives helpful advices. I also thank my girlfriend, Miss. Pik Shan Hui, for her tremendous love and care.

Abstract

A popular model for the locations of fibres or grains in composite materials is the inhomogeneous Poisson process in dimension 3. Its local intensity function may be estimated non-parametrically by local smoothing, e.g. by kernel estimates. They crucially depend on the choice of bandwidths as tuning parameters controlling the smoothness of the resulting function estimate. In this thesis, we propose a fast algorithm for learning suitable global and local bandwidths from the data. It is well-known, that intensity estimation is closely related to probability density estimation. As a by-product of our study, we show that the difference is asymptotically negligible regarding the choice of good bandwidths, and, hence, we focus on density estimation.

There are quite a number of data-driven bandwidth selection methods for kernel density estimates. cross-validation is a popular one and frequently proposed to estimate the optimal bandwidth. However, if the sample size is very large, it becomes computational expensive. In material science, in particular, it is very common to have several thousand up to several million points. Another type of bandwidth selection is a solve-the-equation plug-in approach which involves replacing the unknown quantities in the asymptotically optimal bandwidth formula by their estimates.

In this thesis, we develop such an iterative fast plug-in algorithm for estimating the optimal global and local bandwidth for density and intensity estimation with a focus on 2- and 3-dimensional data. It is based on a detailed asymptotics of the estimators of the intensity function and of its second derivatives and integrals of second derivatives which appear in the formulae for asymptotically optimal bandwidths. These asymptotics are utilised to determine the exact number of iteration steps and some tuning parameters. For both global and local case, fewer than 10 iterations suffice. Simulation studies show that the estimated intensity by local bandwidth can better indicate the variation of local intensity than that by global bandwidth. Finally, the algorithm is applied to two real data sets from test bodies of fibre-reinforced high-performance concrete, clearly showing some inhomogeneity of the fibre intensity.

Zusammenfassung

Ein populäres Modell für die Lokation von Fasern oder Körnern in Verbundmaterialien ist der inhomogene Poisson-Prozess in Dimension 3. Seine lokale Intensitätsfunktion kann durch lokales Glätten, z.B. durch Kernschätzer, nichtparametrisch geschätzt werden. Diese Schätzer hängen wesentlich von der Wahl der Bandbreiten als Kontrollparameter für die lokale Glattheit der resultierenden Funktionsschätzer ab. In dieser Arbeit schlagen wir einen schnellen Algorithmus vor, mit dem geeignete globale und lokale Bandbreiten aus den Daten gelernt werden können. Bekanntlich hängen Intensitätsschätzer eng mit Schätzern für Wahrscheinlichkeitsdichten zusammen. Als ein Nebenprodukt unserer Untersuchungen zeigen wir, dass der Unterschied im Hinblick auf die Wahl guter Bandbreiten asymptotisch vernachlässigbar ist, und daher betrachten wir im größten Teil der Arbeit Dichteschätzer.

Für Kerndichteschätzer existieren bereits eine Reihe von Verfahren zur datengetriebenen Bandbreitenselektion. Kreuzvalidierung ist ein populärer Ansatz, der oft zur Bandbreitenwahl eingesetzt wird. Wenn der Stichprobenumfang sehr groß ist, wird dieses Verfahren allerdings sehr rechenaufwendig. In den Materialwissenschaften hat man es üblicherweise mit sehr großen Stichproben (Tausende bis Millionen von Punkten) zu tun. Ein anderer Ansatz zur Bandbreitenwahl nutzt asymptotische Approximationen für den Fehler, optimiert sie bzgl. der Bandbreite und ersetzt in den resultierenden Ausdrücken die unbekannten Größen durch Schätzer (plug-in).

In dieser Arbeit entwickeln wir einen iterativen, schnellen Plug-in Algorithmus zur Schätzung der optimalen globalen und lokalen Bandbreiten für Dichte- und Intensitätsschätzer, insbesondere in den Dimensionen 2 und 3. Er basiert auf einer detaillierten Asymptotik für die Funktionsschätzer, für ihre zweiten Ableitungen und für Integrale der zweiten Ableitungen, die in den Formeln für die asymptotisch optimalen Bandbreiten auftauchen. Aus dieser Asymptotik ergibt sich für den Algorithmus die exakte Anzahl der Iterationsschritte sowie die Wahl von gewissen Tuningparametern. Sowohl für globale wie auch für lokale Bandbreitenwahl reichen weniger als 10 Iterationen aus.

Simulationsstudien zeigen, dass der Intensitätsschätzer mit datengetriebener lokaler Bandbreite lokale Variationen in der zugrundeliegenden wahren Intensitätsfunktion deutlich besser zeigt als der entsprechende Schätzer mit datengetriebener globaler Bandbreite. Schließlich wenden wir den Algorithmus auf zwei reale Datensätze an, die von Testkörpern aus faserverstärktem Hochleistungsbeton stammen, und finden eine deutliche Inhomogenität in der Faserintensität.

Contents

Acknowledgements **i**

Abstract **iii**

Zusammenfassung **v**

Contents **vii**

List of Figures **x**

List of Tables **xi**

1 Introduction **1**

1.1 Non-parametric intensity estimation for inhomogeneous Poisson processes . . . 1

1.2 A leisurely introduction to kernel intensity estimates . . . 5

1.3 Local mean-squared error expansions . . . 7

1.4 Optimal bandwidth conditional on N . . . 13

1.5 Outline of the thesis . . . 15

2 Kernel intensity estimates and optimal bandwidths **17**

2.1 Mean squared error and mean integrated squared error . . . 17

2.2 Optimal asymptotic bandwidth . . . 24

2.3 The case of smoother intensity functions 30

3 Data adaptive bandwidth selection by the plug-in approach 33

3.1 An algorithm for automatic bandwidth selection 34

3.2 Asymptotics of kernel estimates with data-adaptive bandwidth . 36

4 Asymptotics for integrated mean-squared error estimates with random bandwidths 45

4.1 The one-dimensional case 46

4.2 The two-dimensional case 69

5 Asymptotics for local mean-squared error estimates with random bandwidths 93

6 Tuning parameters of the plug-in algorithm 103

6.1 Data-adaptive choice of global bandwidths 105

6.2 Data-adaptive choice of local bandwidths 109

6.3 The inflation factor 111

7 Bandwidth selection for spatial data ($d = 3$) 116

7.1 Approximation results in dimension $d = 3$ 117

7.2 Asymptotics for amise and amse estimates involving random bandwidths in dimension $d = 3$ 124

7.3 Tuning parameters of the plug-in algorithm in dimension $d = 3$. 126

8 Application to simulated and real data 133

8.1 Simulation results for the mixture bivariate normal distribution 134

8.2 Simulation results for a bivariate normal distribution with large correlation 139

8.3 Application to concrete fibres projected onto 2 dimensional plane 140

9 Conclusion **149**

References **151**

List of Figures

8.1 Plots of the iterated global bandwidths $\hat{h}_1^{(t)}$ and local bandwidth $\hat{h}_1^{(t)}(0.75, 0.75)$ for observations generated from a 2-dimensional normal distribution as specified in Section 8.1 with $\sigma_0 = \sigma_1 = \frac{1}{12}$ and $w = 0.75$. 135

8.2 True intensity, estimates by local bandwidth and global bandwidth for a mixture bivariate normal distribution, $\sigma_0 = \sigma_1 = \frac{1}{12}$, $w = 0.75$. 139

8.3 Scatter plots, kernel estimates by global and local bandwidths for the fibre locations projected onto $x - y$ plane, test body 1 143

8.4 Scatter plots, kernel estimates by global and local bandwidths for the fibre locations projected onto $x - z$ plane, test body 1 144

8.5 Scatter plots, kernel estimates by global and local bandwidths for the fibre locations projected onto $y - z$ plane, test body 1 145

8.6 Scatter plots, kernel estimates by global and local bandwidths for the fibre locations projected onto $x - y$ plane, test body 2 146

8.7 Scatter plots, kernel estimates by global and local bandwidths for the fibre locations projected onto $x - z$ plane, test body 2 147

8.8 Scatter plots, kernel estimates by global and local bandwidths for the fibre locations projected onto $y - z$ plane, test body 2 148

List of Tables

8.1 Iterated global bandwidths for observations generated from a 2-dimensional normal distribution as specified in Section 8.1 with $\sigma_0 = \sigma_1 = \frac{1}{12}$ and $w = 0.75$. 135

8.2 Iterated global and local bandwidths at $(x, y) = (0.75, 0.75)$ for observations generated from a 2-dimensional normal distribution as in Figure 8.1. 136

8.3 The difference between the true and estimated intensity by global bandwidth ($\bar{E}_G$) and local bandwidth ($\bar{E}_L$). Those $\bar{E}_L$ marked with * indicated that at least 1 sample has local bandwidth equal to 0, NaN or infinity. Those $\bar{E}_L$'s are computed by ignoring those samples. Please refer to Table 8.4 for the number of samples with local bandwidths = NaN, zero or infinity out of those 200 samples. 137

8.4 Number of samples with local bandwidths = NaN, zero or infinity out of 200 samples. 137

8.5 The difference between the true and estimated intensity by global bandwidth ($\bar{E}_G$) and local bandwidth ($\bar{E}_L$). Those $\bar{E}_L$ marked with * indicated that at least 1 sample has local bandwidth equal to 0, NaN or infinity. Those $\bar{E}_L$'s are computed by ignoring those samples. Please refer to Table 8.6 for the number of samples with local bandwidths = NaN, zero or infinity out of those 200 samples. 138

8.6 Number of samples with local bandwidths = NaN, zero or infinity out of 200 samples. 138

8.7 Difference between the estimated intensity and true intensity. True intensity is a bivariate normal distribution with mean $\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$ and covariance matrix $\begin{pmatrix} \frac{1}{144} & \frac{0.8}{144} \\ \frac{0.8}{144} & \frac{1}{144} \end{pmatrix}$. 140

8.8 True values of the intensity function and the estimates for a bivariate normal distribution with mean $\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$ and covariance matrix $\begin{pmatrix} \frac{1}{144} & \frac{0.8}{144} \\ \frac{0.8}{144} & \frac{1}{144} \end{pmatrix}$. 140

Chapter 1

Introduction

1.1 Non-parametric intensity estimation for inhomogeneous Poisson processes

Point processes are random processes for which a set of points are randomly distributed in time or geographic space. A very important point process is the Poisson point process for which the numbers of points in any two non-overlapping regions are independent. An intensity function which may depend on time or the location governs the distribution of the points. Non-parametric statistical tools such as kernel estimation have been used to estimate the feature of the intensity. Kernel estimation has been applied to regression and probability density over many years. A very crucial parameter for getting a good estimate which is also the only one is the smoothing parameter or bandwidth which controls the smoothness of the estimated intensity. When the bandwidth is too large, it smooths away the important features of the underlying structure. When it is too small, the resulting estimate is too rough and may contain features which are only noise of the underlying process hence can be a hindrance to data analysis.

Kernel estimation of probability density functions and regression functions are well studied. For choosing the crucial bandwidth parameter a subjective approach was adopted at the beginning of kernel estimation literature. Later, various data driven methods were developed for choosing the bandwidth for a more effective and objective way. Examples include least squares cross-validation and solve-the-equation plug-in approach. An overview of bandwidth selection can be found in Jones et al. (1996) or in Heidenreich et al. (2013). It has been shown that the variability of bandwidth obtained from cross-validation is much larger than that of plug-in, see Park and Marron (1990). For the particular problem of kernel intensity estimation, Diggle (1985) and Diggle (2014) propose a data-adaptive bandwidth selection based on the assumption of a Cox process, which we discuss in more detail in Section 1.2.

The mean integrated squared error (mise) criterion is a common way for measuring the error in the estimation. Asymptotically, mise is approximated by asymptotic mean squared error (amise). Through minimising the amise with respect to the bandwidth h, we can have an optimal bandwidth formula. However, the formula is expressed in terms of some to-be-estimated quantities such as the second derivative of the density function in case of density estimation. The idea of plug-in is to replace the unknown terms by their own estimates. In the optimal bandwidth formula for kernel density (regression) estimation, the unknown term is the integral of the second derivative of the density (regression function). For estimating the integral, Hall and Marron (1987) and Jones and Sheather (1991) consider using a kernel estimator with a bandwidth which is different from that for estimating the density. Gasser et al. (1991) consider an iterative algorithm for estimating the bandwidth for kernel regression, where the iterated bandwidth converges to a value close to the one which minimises the mean (integrated) squared error after several iteration steps. Engel et al. (1994) discuss how to choose suitable tuning parameters in the iterative algorithm for desired properties of the estimator for dimension 1.

In this thesis, plug-in approach is applied to intensity estimation of point process. Following Gasser et al. (1991), an iteration algorithm for estimating h is considered, and the asymptotic behaviours of the kernel estimates and those of the iterative bandwidth $\hat{h}$ for global and $\hat{h}(x)$ for local are derived.

The main competitors of our method discussed in the literature are Diggle's approach, which is tuned to Cox processes, and general resampling methods, in particular cross-validation and bootstrap. They provide estimates of mise(h) as a function of h and then optimise w.r.t. to h. This necessitates to calculate kernel estimates for many different values of h. For resampling methods like the popular bandwidth selection by cross-validation (compare, e.g., Brockmann and Marron (1991)), already the estimation of the mise(h) at a given value h requires the calculation of many kernel estimates. For smaller data sets this is computationally feasible. However, for many applications in material science, in particular for fibre directions in a μCT image of fibres in concrete, we have to deal with several thousands up to several million points. Therefore, resampling methods may be computationally quite expensive.

By contrast, for the iterative algorithm proposed in this thesis we only have to calculate a small (depending on dimension, a single digit or a bit larger) number of kernel estimates. We develop some theories, which show that the algorithm arrives at the best possible approximation already after a known finite number of iterations.

In a recent paper, Cronie and van Lieshout (2016) propose a bandwidth selection algorithm which in spirit follows Diggle's approach but in contrast to all other methods including ours does not try to choose h by minimising an approximation of mise(h). They observe that expectation of the sum of the inverse intensity evaluated at the points of the process lying in the observation window coincides with the window size and therefore is known. Then, they choose h such that the estimate of the expectation, which we get by replacing the intensity with its kernel estimate, is as close as possible to the window size.

Diggle's approach might still be feasible for larger samples as he also focuses on estimating the asymptotic mean integrated squared error. However, his method is based on the assumption of a Cox process and only allows to choose a global bandwidth parameter h. The same remark applies to the algorithm of Cronie and van Lieshout (2016) due to the summation over the whole observation window. However, we know from the theory of non-parametric kernel estimates, that optimal bandwidths for estimating a function $\mu(x)$ at a given location x depend on the local characteristics of the function, in particular on its curvature, i.e. on the square of the second derivative, as we shall see below. Where the curvature is high, small bandwidths are appropriate whereas large bandwidths are better suited for rather flat parts of the function. Therefore, we are also interested in a method allowing for data-adaptive selection of good local bandwidths $h(x)$ which can be achieved by the same kind of approach as the global bandwidths.

Note that the algorithm of Cronie and van Lieshout (2016) in principle could be extended to local bandwidth selection at a location x by considering a local window centred at x instead of the whole observation window. However, choosing the size of such a local window is a problem which has to be investigated in future research.

1.2 A leisurely introduction to kernel intensity estimates

In this section, we introduce some basic notions on point processes and the kind of estimates which are the focus of interest of this thesis. We strongly rely on the exposition of Diggle (2014), Chapters 5 and 6. To keep notation simple, we only consider dimension $d = 1$ here, but the concepts are straightforwardly generalised to $d > 1$.

We recall the definition of inhomogeneous Poisson processes with intensity function $\mu(x), x \in \mathbb{R}$. It is characterised by the following three properties, where $N(I)$ denotes the number of points in an interval $I \subseteq \mathbb{R}$:

P1: If $I_1, \ldots, I_m$ are disjoint intervals, then $N(I_1), \ldots, N(I_m)$ are independent.

P2: $N(I)$ is Poisson distributed with parameter $\int_I \mu(x)\,\mathrm{d}x$.

P3: Given $N(I) = n$, the n points of the process lying in I are i.i.d. with density

$$\lambda(x) = \frac{1}{\int_I \mu(y)\,\mathrm{d}y}\mu(x).$$

Note that these properties imply

$$\mu(x) = \lim_{\mathrm{d}x \to 0} \frac{\mathbb{E}N([x, x + \mathrm{d}x])}{\mathrm{d}x}.$$

Analogously the second-order intensity function is given by

$$\mu_2(x, y) = \lim_{\mathrm{d}x, \mathrm{d}y \to 0} \frac{\mathbb{E}N([x, x + \mathrm{d}x])\, N([y, y + \mathrm{d}y])}{\mathrm{d}x\mathrm{d}y}.$$

As the Poisson parameters of $N(I)$ in P2 depend on the location of I, such a process is not stationary. To achieve stationarity for modelling purpose, it is frequently assumed that $\mu(x)$ itself is random. This results in a so-called Cox process characterised by

CP1: $M(x), x \in \mathbb{R}$, is a stochastic process satisfying $M(x) \geq 0$.

CP2: Given $M(x) = \mu(x), x \in \mathbb{R}$, the random points form an inhomogeneous Poisson process with intensity function $\mu(x)$.

If the process $M(x)$ is stationary, then the Cox process is stationary too, i.e. we have

$$\lim_{\mathrm{d}x \to 0} \frac{\mathbb{E}N([x, x+\mathrm{d}x])}{\mathrm{d}x} = \mu, \lim_{\mathrm{d}x, \mathrm{d}y \to 0} \frac{\mathbb{E}N([x, x+\mathrm{d}x])\, N([y, y+\mathrm{d}y])}{\mathrm{d}x\mathrm{d}y} = \mu_2(x-y)$$

for all $x, y \in \mathbb{R}$, i.e. the inhomogeneity averages out if we take expectations w.r.t. the random intensity functions $M(x)$. In that case, the so-called K-function is given as

$$k(t) = \frac{2\pi}{\mu^2} \int_0^t \mu_2(s)\, \mathrm{d}s.$$

Diggle (2014), Section 5.3, considers estimates of the realised value $\mu(x)$ of $M(x)$ of the form

$$\hat{\mu}(x, h) = \frac{N([x-h, x+h])}{2h}, \quad h > 0,$$

i.e. he considers the point density in a small neighbourhood of x. The performance of the estimate mainly depends on the tuning parameter, the so-called bandwidth h. For choosing h, Diggle calculated the mean-squared error w.r.t. randomness of the point process as well as the randomness of the intensity function $M(x)$ as a function of h:

$$\mathrm{mse}(h) = \mathbb{E}\left(\hat{\mu}(x, h) - M(x)\right)^2.$$

Note that, due to the stationarity of the Cox process, this does not depend on x. Diggle derived a formula for $\mathrm{mse}(h)$ or, more precisely, for that part of $\mathrm{mse}(h)$ depending on h, which only depends on μ, $k(h)$ and an integral of $\mu_2(s)$. Those quantities all can be estimated such that an estimate of $\mathrm{mse}(h)$ can be derived and minimised w.r.t. h. Diggle recommends to plot this estimate and choose h visually in a range where this function is small and (usually) flat - compare Diggle's (2014) discussion of his figure 5.1.

Diggle also points out the relationship of $\hat{\mu}(x,h)$ to general probability density estimates of the Rosenblatt-Parzen type (compare, e.g., Silverman (1986)). To illustrate this fact, let us assume now that we observe the point process in a finite interval, say $[0,1]$. Let $N = N([0,1])$ denote the number of points, and $X_1, \ldots, X_N \in [0,1]$ their locations. Then,

$$\hat{\mu}(x,h) = \sum_{j=1}^{N} \frac{1}{2h} \mathbf{1}_{[x-h,x+h]}(X_j) = \sum_{j=1}^{N} K_h(x - X_j),$$

where $K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$ and $K(u) = \frac{1}{2}\mathbf{1}_{[-1,+1]}(u)$. The latter is called the rectangular kernel in kernel density estimation. It is well-known (compare Silverman (1986)) that the performance of the estimate is improved for smoother kernel functions provided they satisfy $K(u) \geq 0$ and $\int K(u)\,\mathrm{d}u = 1$.

Note that from P3, $X_1, \ldots, X_N$ are i.i.d. with probability density

$$\lambda(x) = \frac{1}{\int_0^1 \mu(y)\,\mathrm{d}y} \mu(x)$$

and

$$\hat{\lambda}(x,h) = \frac{1}{N}\hat{\mu}(x,h) = \frac{1}{N}\sum_{j=1}^{N} K_h(x - X_j)$$

is the usual kernel estimate for the density of $X_1, \ldots, X_N$ given the value of N. Note also from P2, that N is Poisson distributed with parameter $\int_0^1 \mu(y)\,\mathrm{d}y = \bar{\mu}$ and, hence, N is the maximum likelihood estimate of $\bar{\mu}$.

1.3 Local mean-squared error expansions

One of our main goals is quality inspection for given specimen of composite materials like fibre-reinforced concrete. Part of that problem is investigating the distribution of fibre locations for a given test volume. We model those locations as a realisation of an inhomogeneous Poisson process observed in the given volume. Whether the corresponding intensity function $\mu(x)$ is fixed or - more likely, due to the manufacturing process - random is not so much of interest if we want to judge the reliability of a given test body under stress. For that purpose, we want to estimate the given realisation well, using a kernel estimate $\hat{\mu}(x,h)$, which depends on the choice of bandwidth parameters.

Let $X_1,\dots,X_N$ denote the points of an inhomogeneous Poisson process with intensity $\mu(x)$ which lie in the unit interval $[0,1]$. Let

$$\bar{\mu} = \int_0^1 \mu\,(x)\,\mathrm{d}x = \mathbb{E}N, \qquad \text{and} \quad \lambda\,(x) = \frac{\mu\,(x)}{\bar{\mu}}.$$

Then, $\lambda(x), 0 \le x \le 1$, is a probability density on $[0,1]$, and given N, the X_j are i.i.d. with density λ. We write again

$$\hat{\mu}(x,h) = \sum_{j=1}^{N} K_h(x - X_j).$$

Following Diggle (2014), we first consider the rectangular kernel. Note that for asymptotic expansions of function estimates we need that the sample size increases. As here N is random, the corresponding assumption is $\mathbb{E}N = \bar{\mu} \to \infty$. Hence, as μ is, then, increasing too, we consider the invariant standardised mean-squared error for the asymptotic expansion

$$\frac{1}{\bar{\mu}^2}\ \mathrm{mse}\ \hat{\mu}(x,h) = \frac{1}{\bar{\mu}^2}\ \mathbb{E}\big(\hat{\mu}(x,h) - \mu(x)\big)^2.$$

Proposition 1. *Assume that $\lambda(x)$ is twice continuously differentiable, and the second derivative $\lambda''(x)$ is Hölder continuous with exponent $\beta > 0$, i.e. for all x, z and some $c_H > 0$*

$$|\lambda''(x) - \lambda''(z)| \le c_H |x - z|^{\beta}.$$

Then, we have for $\bar{\mu} \to \infty, h \to 0$ such that $\bar{\mu}h \to \infty$

$$\frac{1}{\bar{\mu}^2}\ \mathrm{mse}\ \hat{\mu}(x,h) = \frac{\lambda(x)}{2\bar{\mu}h} + \left(\frac{\lambda''(x)}{6}\right)^2 h^4 + O\left(\frac{h}{\bar{\mu}}\right) + O(h^{4+\beta}), \qquad 0 < x < 1.$$

Proof. To avoid discussion of boundary effects, we assume throughout the proof that h is already small enough such that $[x-h, x+h] \subset (0,1)$. As usual, the mse decomposes into variance and squared bias. We first consider the bias where we use

$$\mathbb{E}\hat{\mu}(x,h) = \frac{1}{2h}\mathbb{E}N([x-h, x+h]) = \frac{1}{2h}\int_{x-h}^{x+h} \mu(s)\mathrm{d}s.$$

Note also that from the Hölder condition on λ we get

$$|\mu''(z) - \mu''(x)| \leq c_H \ \bar{\mu}|z-x|^{\beta} \leq c_H \ \bar{\mu}h^{\beta}$$

for all $x - h \leq z \leq x + h$. Using a Taylor expansion of μ and $\int_{-h}^{h} t \ dt = 0$, we get

$$\begin{aligned}
\text{bias } \hat{\mu}(x,h) &= \mathbb{E} \ \hat{\mu}(x,h) - \mu(x) = \frac{1}{2h}\int_{x-h}^{x+h} (\mu(s) - \mu(x))\mathrm{d}s \\
&= \frac{1}{2h}\int_{x-h}^{x+h} \left(\mu'(x)(s-x) + \frac{\mu''(x) + O(\mu h^{\beta})}{2} \ (s-x)^2\right) \mathrm{d}s \\
&= \frac{1}{2h}\int_{-h}^{h} \left(\mu'(x)t + \frac{\mu''(x) + O(\bar{\mu}h^{\beta})}{2} \ t^2\right) \mathrm{d}t \\
&= \frac{h^2}{6}\mu''(x) + O(\bar{\mu} \ h^{2+\beta})
\end{aligned}$$

Now, looking at the variance, we have from P2

$$\begin{aligned}
\text{var } \hat{\mu}(x,h) &= \frac{1}{4h^2}\text{var } N([x-h, x+h]) \\
&= \frac{1}{4h^2}\mathbb{E} \ N([x-h, x+h]) = \frac{1}{2h}\mathbb{E}\hat{\mu}(x,h) \\
&= \frac{1}{2h}(\mu(x) + \text{bias } \hat{\mu}(x,h)) \\
&= \frac{\bar{\mu}}{2h}[\lambda(x) + O(h^2)] = \frac{1}{2h}\mu(x) + O(\bar{\mu}h)
\end{aligned}$$

using the bias expansion and $\mu''(x) = \bar{\mu} \ \lambda''(x) = O(\bar{\mu})$.

Combining both terms, we have

$$\text{mse}\hat{\mu}(x,h) = \text{var } \hat{\mu}(x,h) + \text{bias}^2\hat{\mu}(x,h) = \frac{1}{2h}\mu(x) + \left(\frac{h^2}{6}\mu''(x)\right)^2 + O(\bar{\mu}h) + O(\bar{\mu}h^{4+\beta}),$$

which concludes the proof. □

Corollary 1. *Under the assumptions of Proposition 1, the mse-optimal bandwidth is given asymptotically by*

$$\bar{h}_a^5(x) \sim \frac{9}{2} \frac{\lambda(x)}{(\lambda''(x))^2} \frac{1}{\bar{\mu}}.$$

Proof. The result follows from setting the derivative of

$$g(h) = \frac{\lambda(x)}{2\mu h} + \left(\frac{\lambda''(x)}{6}\right)^2 h^4$$

to 0. □

Now we replace the rectangular kernel with a general kernel function $K(u) \geq 0$ satisfying

$$\mathbf{K} : K(u) = 0 \text{ for } |u| > 1, \qquad \int_{-1}^{1} K(u)du = 1, \qquad \int_{-1}^{1} uK(u)du = 0.$$

The last assumption is, e.g., satisfied if K is symmetric around 0.

The key to the mse expansion now is Campbell's formula (compare Theorem 3.1.2 and the following remarks in Schneider and Weil (2008)). As the point processes which we are considering are simple by Lemma 3.2.1 of Schneider and Weil (2008), i.e. they do not allow for multiple points at the same location, and as the set of points of our Poisson process in the whole space is countable, we have

Lemma 1. *Let $g : \mathbb{R} \to \mathbb{R}$ be an integrable function, and let $X_1, X_2, \ldots$ denote all the points of our inhomogeneous Poisson process. Then,*

$$\mathbb{E} \sum_{j=1}^{\infty} g(X_j) = \int_{-\infty}^{\infty} g(z)\mu(z)\mathrm{d}z.$$

If $X_1, \ldots, X_N$ denote the points which lie in the unit interval $[0, 1]$ then we conclude from the lemma

$$\mathbb{E} \sum_{j=1}^{N} g(X_j) = \mathbb{E} \sum_{j=1}^{\infty} g(X_j)\mathbf{1}_{[0,1]}(X_j) = \int_{-\infty}^{\infty} g(z)\mathbf{1}_{[0,1]}(z)\mu(z)\,\mathrm{d}z = \int_{0}^{1} g(z)\mu(z)\mathrm{d}z. \tag{1.1}$$

We also need a corresponding result for calculating second-order moments. For that, we use that the second factorial moment measure of a Poisson process is just the product of the intensity measure with itself by Corollary 3.2.4 of Schneider and Weil (2008). Hence, we get from Theorem 3.1.3 of Schneider and Weil (2008)

Lemma 2. *Let $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be an integrable function, and let $X_1, X_2, \ldots$ denote all the points of our inhomogeneous Poisson process. Then,*

$$\mathbb{E} \sum_{j=1}^{\infty} \sum_{i=1, i \neq j}^{\infty} g(X_i, X_j) = \int_{-\infty}^{\infty} g(y, z) \mu(y) \mu(z) \mathrm{d}y \mathrm{d}z.$$

If we restrict our attention to the points $X_1, \ldots, X_N$ lying in $[0, 1]$, we get from the lemma

$$\begin{aligned} \mathbb{E} \sum_{j=1}^{N} \sum_{i=1, i \neq j}^{N} g(X_i, X_j) &= \mathbb{E} \sum_{j=1}^{\infty} \sum_{i=1, i \neq j}^{\infty} g(X_i, X_j) \mathbf{1}_{[0,1]}(X_i) \mathbf{1}_{[0,1]}(X_j) \\ &= \int_{-\infty}^{\infty} g(y, z) \mathbf{1}_{[0,1]}(y) \mathbf{1}_{[0,1]}(z) \mu(y) \mu(z) \mathrm{d}y \mathrm{d}z \\ &= \int_{0}^{1} g(y, z) \mu(y) \mu(z) \mathrm{d}y \mathrm{d}z. \qquad (1.2) \end{aligned}$$

Theorem 1. *Under the assumptions of Proposition 1 and if (**K**) holds for the kernel, we have for $\bar{\mu} \to \infty, h \to 0$ such that $\bar{\mu} h \to \infty$*

$$\frac{1}{\bar{\mu}^2} \operatorname{mse} \hat{\mu}(x, h) = \frac{\lambda(x)}{\bar{\mu} h} Q_K + \left(\frac{\lambda''(x)}{2} \right)^2 h^4 V_K^2 + O(\frac{1}{\bar{\mu}}) + O(h^{4+\beta}), \qquad 0 < x < 1,$$

with the known constants, depending on K only,

$$Q_K = \int_{-1}^{1} K^2(u) du, \qquad V_K = \int_{-1}^{1} u^2 K(u) du.$$

Proof. As $h \to 0$, we may again assume that it is already small enough such that $h \leq x \leq 1 - h$ such that we do not have to worry about boundary effects in the following calculations. As in the proof of Proposition 1, we investigate bias and variance separately. For the bias, we have

$$\begin{aligned}
\text{bias } \hat{\mu}(x,h) &= \mathbb{E}\ \hat{\mu}(x,h) - \mu(x) = \int_0^1 K_h(x-s)(\mu(s)-\mu(x))\mathrm{d}s \\
&= \int_0^1 K_h(x-s)\left(\mu'(x)(s-x) + \frac{\mu''(x)+O(\bar{\mu}h^\beta)}{2}\ (s-x)^2\right)\mathrm{d}s \\
&= \int_{-1}^1 K(t)\left(\mu'(x)th + \frac{\mu''(x)+O(\bar{\mu}h^\beta)}{2}\ t^2h^2\right)\mathrm{d}t \\
&= \frac{h^2}{2}\mu''(x)V_K + O(\bar{\mu}\ h^{2+\beta}) = \frac{\bar{\mu}h^2}{2}\lambda''(x)V_K + O(\bar{\mu}\ h^{2+\beta}),
\end{aligned}$$

where, for the first line, we use Campbell's formula (1.1) with $g(z) = K_h(x-z)$ and the fact that K and, hence, K_h integrate to 1. For the second line, we use Taylor expansion and Hölder continuity as in the proof of Proposition 1. For the third line, we substitute $t = \frac{x-s}{h}$, and for the last line, we use assumption **K** such that the first term vanishes.

For the variance, we first consider the second moment

$$\begin{aligned}
\mathbb{E}\ \hat{\mu}^2(x,h) &= \mathbb{E}\sum_{i,j=1}^N K_h(x-X_i)K_h(x-X_j) \\
&= \mathbb{E}\sum_{j=1}^N K_h^2(x-X_j) + \mathbb{E}\sum_{j=1}^N\sum_{i=1,i\neq j}^N K_h(x-X_i)K_h(x-X_j).
\end{aligned}$$

For the first term, we get by Campbell's formula (1.1) and substituting $t = \frac{x-s}{h}$

$$\begin{aligned}
\mathbb{E}\sum_{j=1}^N K_h^2(x-X_j) &= \int_0^1 K_h^2(x-s)\mu(s)\mathrm{d}s = \frac{1}{h}\int_{-1}^1 K^2(t)\mu(x-ht)\mathrm{d}t \\
&= \frac{1}{h}\int_{-1}^1 K^2(t)(\mu(x)+O(\bar{\mu}h))\mathrm{d}t = \frac{\mu(x)}{h}Q_K + O(\bar{\mu}),
\end{aligned}$$

where the second line follows from the mean-value theorem, $\mu'(x) = \mu\lambda'(x)$, the boundedness of λ' and $|t| \leq 1$. From (1.2), we have

$$\begin{aligned}
\mathbb{E}\sum_{j=1}^N\sum_{i=1,i\neq j}^N K_h(x-X_i)K_h(x-X_j) &= \int_0^1\int_0^1 K_h(x-s)K_h(x-z)\mu(s)\mu(z)\mathrm{d}s\mathrm{d}z \\
&= \left(\int_0^1 K_h(x-s)\mu(s)ds\right)^2 = \big(\mathbb{E}\ \hat{\mu}(x,h)\big)^2.
\end{aligned}$$

Therefore, we get

$$\text{var } \hat{\mu}(x,h) = \mathbb{E}\hat{\mu}^2(x,h) - \left(\mathbb{E}\ \hat{\mu}(x,h)\right)^2 = \frac{\mu(x)}{h}Q_K + O(\bar{\mu}).$$

Combining the bias and variance expansions, we get

$$\begin{aligned} \text{mse } \hat{\mu}(x,h) &= \text{var } \hat{\mu}(x,h) + \text{bias}^2\hat{\mu}(x,h) \\ &= \frac{1}{h}\mu(x)Q_K + O(\bar{\mu}) + \left(\frac{h^2}{2}\mu''(x)V_K + O(\bar{\mu}h^{2+\beta})\right)^2, \end{aligned}$$

which implies the asymptotic expansion of the mean-squared error of $\hat{\mu}(x,h)$.

□

In the same manner as Corollary 1, we get

Corollary 2. *Under the assumptions of Theorem 1, the mse-optimal bandwidth is*

$$\bar{h}_a^5(x) \sim \frac{Q_K}{V_K^2}\ \frac{\lambda(x)}{(\lambda''(x))^2}\ \frac{1}{\bar{\mu}}.$$

Note that the mse expansion of Theorem 1 has been stated already without proof in Cowling et al. (1996). Unfortunately, we were not able to get a copy of the long version of the paper, mentioned in the publication, from the authors. Therefore, we gave here our own proof.

1.4 Optimal bandwidth conditional on N

The asymptotically optimal bandwidth $\bar{h}_a(x)$ of Corollary 2 depends, among other quantities, on the unknown $\bar{\mu} = \mathbb{E}N$, where $N = N([0,1])$. It is well-known that for a sample of i.i.d. data with density $\lambda(x)$ and given size N, the asymptotically optimal bandwidth has exactly the same form with N replacing $\bar{\mu}$:

$$h_a^5(x) \sim \frac{Q_K}{V_K^2}\frac{\lambda(x)}{(\lambda''(x))^2}\frac{1}{N}$$

(compare, e.g., the analogous derivation in Section 2.2 for the case of dimension $d = 2$). Hence, we have for random N

$$\frac{\bar{h}_a(x)}{h_a(x)} \sim \left(\frac{N}{\bar{\mu}}\right)^{\frac{1}{5}} = 1 + O_p\left(\frac{1}{\sqrt{\bar{\mu}}}\right)$$

as, from P2, N is Poisson distributed with parameter $\bar{\mu}$, and, therefore,

$$\mathbb{E}\frac{N}{\bar{\mu}} = 1, \quad \mathrm{var}\frac{N}{\bar{\mu}} = \frac{1}{\bar{\mu}^2}\mathrm{var}N = \frac{1}{\bar{\mu}^2}\mathbb{E}N = \frac{1}{\bar{\mu}}.$$

Alternatively, as

$$\frac{1}{\sqrt{\bar{\mu}}} = \sqrt{\frac{N}{\bar{\mu}}}\frac{1}{\sqrt{N}} = \frac{1}{\sqrt{N}}\sqrt{1 + O_p\left(\frac{1}{\sqrt{\bar{\mu}}}\right)} = \frac{1}{\sqrt{N}}\left(1 + O_p\left(\frac{1}{\sqrt{\bar{\mu}}}\right)\right),$$

we could also write the asymptotic equivalence of $\bar{h}_a(x)$ and $h_a(x)$ in the form

$$\frac{\bar{h}_a(x)}{h_a(x)} \sim 1 + O_p\left(\frac{1}{\sqrt{N}}\right)$$

or

$$\begin{aligned}\bar{h}_a(x) &= h_a(x) + O_p\left(N^{-\frac{1}{2}}N^{-\frac{1}{5}}\right) \\ &= h_a(x) + O_p\left(N^{-\frac{7}{10}}\right).\end{aligned}$$

For dimension $d = 2$, we analogously get, compare (2.3), (2.4),

$$\bar{h}_{ai}(x) = h_{ai}(x) + O_p\left(N^{-\frac{2}{3}}\right), i = 1, 2,$$

for the two bandwidths involved in the two-dimensional kernel estimates.

As we shall see later in the thesis, the approximation error of, e.g., $N^{-\frac{2}{3}}$ for $d = 2$, is negligible compared to the difference between asymptotically optimal and optimal bandwidth (compare, e.g., Corollary 6) as well as to the difference between the asymptotically optimal bandwidth and its plug-in estimate (compare the derivation in Chapter 6). Therefore, for the purpose of bandwidth selection, it does not matter if we consider $h_a(x)$ or $\bar{h}_a(x)$. As the latter contains the unknown parameter $\bar{\mu}$, which anyhow would be replaced by its maximum likelihood estimate N, we prefer to immediately condition on N, treat it as given and work with $h_a(x)$. One might also argue that we know N anyhow and, therefore, should use this information in finding an optimal bandwidth for the particular sample at hand.

1.5 Outline of the thesis

In Chapter 2, the kernel estimates for the intensity function of an inhomogeneous spatial Poisson process for general dimension d will be introduced. The asymptotic approximations of the mean-squared error and the integrated mean-squared error are derived under assumptions of smoothness of the intensity and kernel function. The formulae for asymptotically optimal local and global bandwidths for $d = 2$ will be derived. All subsequent analysis will be done for $d = 2$ from Chapters 2 to 5. The case of $d = 3$ will be discussed in Chapter 7. We show that the amse (amise) and mse (mise) are asymptotically close when the sample size N goes to infinity. Based on this, we also show how close optimal (for finite sample size) and asymptotically optimal bandwidths are.

In Chapter 3, following Engel et al. (1994), we present an iterative algorithm for selecting the bandwidths automatically from the data based on the formulae of the asymptotic optimal bandwidths and plug-in estimates for the unknown quantities. Then, we present the asymptotics of the kernel estimates for the intensity itself and the second derivatives of the intensity with random bandwidths. However, the number of iteration steps and the choice of various tuning parameters will be left unspecified and discussed in Chapter 6. In later chapter, a more careful analysis of the asymptotics of estimates of the second derivatives of the intensity function will be done to facilitate choosing the tuning parameters.

In Chapter 4, we show the asymptotics for the integrated mean-squared error estimates with random bandwidths for $d = 1$. The proofs follow those in Engel et al. (1994). Such asymptotics and proofs are generalised to $d = 2$. Those asymptotics are necessary for choosing the tuning parameters and number of iterations steps for the global iteration in Chapter 6.

In Chapter 5, we show the asymptotics for the local mean-squared error estimates with random bandwidths. Those asymptotics are necessary for choosing the tuning parameters and number of iteration steps for the local iteration in Chapter 6.

In Chapter 7, those asymptotics presented in previous chapters are generalised to $d = 3$. Another set of tuning parameters for $d = 3$ will be chosen based on the asymptotics.

In Chapter 8, the algorithm presented in Chapter 6 is applied to some simulated 2 dimensional data sets. The fibre locations projected onto a plane obtained from concrete test bodies are analysed.

Chapter 2

Kernel intensity estimates and optimal bandwidths

In this chapter, we introduce kernel estimates for the intensity function of an inhomogeneous spatial Poisson process. We derive asymptotic approximations of the mean-squared error and the integrated mean-squared error which result in formulae for asymptotically optimal local and global bandwidths. We also investigate how close optimal (for finite sample size) and asymptotically optimal bandwidths are.

2.1 Mean squared error and mean integrated squared error

We consider an inhomogeneous Poisson process with intensity function $\mu(x)$ on $\mathbb{R}^d$. Let $X_1, \ldots, X_N$ be the points of the process lying in the unit cube $[0,1]^d$. We condition on N and treat it as a given number due to the discussion in Section 1.4.

Given $N \geq 1$, $X_1, \ldots, X_N$ are i.i.d. on $[0,1]^d$ with density

$$\lambda(x) = \frac{1}{\mu_0}\mu(x),$$

where

$$\mu_0 = \int_0^1 \cdots \int_0^1 \mu(x)\, \mathrm{d}x_1 \cdots \mathrm{d}x_d.$$

As $\lambda(x)$ is a probability density on $[0,1]^d$, we may estimate it by the common Rosenblatt-Parzen kernel estimate in dimension d. For the moment, we allow for different bandwidths in the coordinate directions, but not for adaptive smoothing into other directions. Let $h_1, \ldots, h_d > 0$ denote the bandwidths and $H = \operatorname{diag}(h_1, \ldots, h_d)$ the corresponding d-dimensional bandwidth matrix and that in particular $\Pi_{k=1}^d h_k = \det H$. Let $K : \mathbb{R}^d \to \mathbb{R}$ be a kernel function satisfying

Assumption 1. $K(u) \geq 0$, $\int \cdots \int K(u)\, \mathrm{d}u_1 \cdots \mathrm{d}u_d = 1$.

Assumption 2. *K has a compact support, say* $[-1,+1]^d$.

The latter assumption is for convenience only to simplify notation in the proofs. It may be relaxed to requiring that $K(u) \to 0$ for $\|u\| \to \infty$ fast enough. In particular, we could use the Gaussian kernel, i.e. the probability density of the d-variate standard normal distribution.

Using the rescaled kernel

$$K_H(u) = \frac{1}{\det H} K\left(H^{-1}u\right) = \frac{1}{\Pi_{k=1}^d h_k} K\left(\frac{u_1}{h_1}, \ldots, \frac{u_d}{h_d}\right),$$

we define the estimate $\hat{\lambda}(x,H)$ of $\lambda(x)$ as

$$\hat{\lambda}(x,H) = \frac{1}{N}\sum_{j=1}^{N} K_H(x - X_j).$$

We first derive an asymptotic expansion for the mean-squared error

$$\begin{aligned} \operatorname{mse}\hat{\lambda}(x,H) &= \mathbb{E}\left(\hat{\lambda}(x,H) - \lambda(x)\right)^2 \\ &= \operatorname{var}\hat{\lambda}(x,H) + \operatorname{bias}^2\hat{\lambda}(x,H). \end{aligned}$$

This result is well-known for $d = 1$ and common knowledge for $d > 1$, but we could not find a version of the latter in the literature which suits our particular needs. We make the following regularity assumption on $\lambda(x)$:

Assumption 3. *$\lambda(x)$ is a twice continuously differentiable probability density with support $[0,1]^d$, and the second partial derivatives*

$$\lambda_{ij}(x) = \frac{\partial^2}{\partial x_i \partial x_j} \lambda(x)$$

are Hölder continuous with exponent $\beta > 0$, i.e. for some $C > 0$

$$|\lambda_{ij}(x) - \lambda_{ij}(y)| \leq C \|x - y\|^\beta$$

for all $x, y \in [0,1]^d$.

Moreover, we make the following symmetry and standardisation assumption on the kernel $K(u)$:

Assumption 4. *$K(u)$ is symmetric around 0 in the following sense*

$$\int u_i K(u) \, \mathrm{d}u_i = 0$$

for all $-1 \leq u_j \leq 1, j \neq i$, and all $i = 1, \ldots, d$, and appropriately scaled and that for some $V_K > 0$

$$\int \cdots \int u_i^2 K(u) \, \mathrm{d}u_1 \cdots \mathrm{d}u_d = V_K$$

for $i = 1, \ldots, d$.

Assumption 4 is, e.g., satisfied if $K(u)$ is a product kernel, i.e. for a kernel $K_1 : \mathbb{R} \to \mathbb{R}$,

$$K(u) = \prod_{i=1}^{d} K_1(u_i),$$

where $\int t K_1(t) \, \mathrm{d}t = 0$, $\int t^2 K_1(t) \, \mathrm{d}t = V_K$. Moreover, we use the notation

$$Q_K = \int \cdots \int K^2(u) \, \mathrm{d}u_1 \cdots \mathrm{d}u_d.$$

For consistency of the kernel estimate $\hat{\lambda}(x, H)$, we need that $h_i \to 0, i = 1, \ldots, d$, for $N \to \infty$ with appropriate rates. We assume that the speed of convergence of the bandwidths is the same for all i, i.e. for some sequence $b_N \to 0$ for $N \to \infty$, we have

Assumption 5. $\frac{h_i}{b_N} \to \beta_i$, *for some constants* $0 < \beta_i < \infty, i = 1, \ldots, d$.

Theorem 2. *Let the Assumptions 1, 2, 3, 4, 5 be satisfied. Then*

a) $\text{bias}\hat{\lambda}(x, H) = \frac{1}{2} V_K \sum_{i=1}^{d} h_i^2 \lambda_{ii}(x) + O\left(b_N^{2+\beta}\right)$, *and*

b) $\text{var}\hat{\lambda}(x, H) = \frac{1}{N \det H}\left(Q_K \lambda(x) + O(b_N)\right) = O\left(\frac{1}{N b_N^{d-1}}\right)$ *for all* $x \in (0,1)^d$.

Proof. As we consider x in the interior of the unit cube, and as $h_1, \ldots, h_d \to 0$ for $N \to \infty$, we may assume that N is large enough and that $h_i \leq x_i \leq 1 - h_i$, $i = 1, \ldots, d$. As K has support $[-1, +1]^d$ and, hence, K_H has support $[-h_1, h_1] \times \cdots \times [-h_d, h_d]$, we do not have to worry about boundary effects.

a) As K_H is a probability density, we have

$$\begin{aligned}
\text{bias}\hat{\lambda}(x, H) &= \mathbb{E}\hat{\lambda}(x, H) - \lambda(x) = \mathbb{E} K_H(x - X_1) - \lambda(x) \\
&= \int_0^1 \cdots \int_0^1 K_H(x - z)\left(\lambda(z) - \lambda(x)\right) \mathrm{d}z_1 \cdots \mathrm{d}z_d \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} K(u)\left(\lambda(x - Hu) - \lambda(x)\right) \mathrm{d}u_1 \cdots \mathrm{d}u_d
\end{aligned}$$

substituting $u = H^{-1}(x - z)$. Using a Taylor expansion up to order 2, we get with $0 \leq \theta \leq 1$

$$\begin{aligned}
&\text{bias}\hat{\lambda}(x, H) \\
= &\int \cdots \int K(u)\left\{-(Hu)^\top \nabla \lambda(x) + \frac{1}{2}(Hu)^\top \nabla^2 \lambda(x - \theta H u) H u\right\} \mathrm{d}u_1 \cdots \mathrm{d}u_d,
\end{aligned}$$

where $\nabla \lambda(x) = \left(\frac{\partial}{\partial x_1}\lambda(x), \ldots, \frac{\partial}{\partial x_d}\lambda(x)\right)^\top$ denotes the gradient, and

$$\nabla^2 \lambda(x) = \left(\frac{\partial^2}{\partial x_i \partial x_j}\lambda(x)\right)_{1 \leq i,j \leq d} = \left(\lambda_{ij}(x)\right)_{1 \leq i,j \leq d}$$

denotes the Hessian of $\lambda(x)$. The first term in the bias expansion vanishes due to Assumption 4 which in vector form reads

$$\int \cdots \int K(u)\, u \mathrm{d}u_1 \cdots \mathrm{d}u_d = 0.$$

For the second term we have, using Hölder continuity of $\lambda_{ij}(x)$

$$\begin{aligned}
& \left| u^\top H^\top \left(\nabla^2 \lambda (x - \theta H u) - \nabla^2 \lambda (x) \right) H u \right| \\
= & \left| \sum_{i,j=1}^{d} h_i h_j u_i u_j \left(\lambda_{ij} (x - \theta H u) - \lambda_{ij} (x) \right) \right| \\
\leq & \sum_{i,j=1}^{d} h_i h_j \left| u_i u_j \right| C \left\| \theta H u \right\|^\beta \\
\leq & \; C \sum_{i,j=1}^{d} h_i h_j \left\| H u \right\|^\beta = O\left(b_N^{2+\beta} \right)
\end{aligned}$$

as $|u_i| \leq 1, i = 1, \ldots, d$, for u in the support of K, and as $0 \leq \theta \leq 1$, using $h_i = O(b_N), i = 1, \ldots, d$, from Assumption 5. Therefore, we have

$$\begin{aligned}
\text{bias}\hat{\lambda}(x, H) &= \frac{1}{2} \int \cdots \int u^\top H^\top \nabla^2 \lambda(x) H u K(u) \mathrm{d}u_1 \cdots \mathrm{d}u_d + O\left(b_N^{2+\beta} \right) \\
&= \frac{1}{2} V_K \sum_{i=1}^{d} h_i^2 \lambda_{ii}(x) + O\left(b_N^{2+\beta} \right)
\end{aligned}$$

using Assumption 4.

b) As $X_1, \ldots, X_N$ are i.i.d. given N, we have

$$\begin{aligned}
\text{var}\hat{\lambda}(x, H) &= \frac{1}{N} \text{var} K_H (x - X_1) \\
&= \frac{1}{N} \mathbb{E} K_H^2 (x - X_1) - \frac{1}{N} \left(\mathbb{E} K_H (x - X_1) \right)^2 .
\end{aligned}$$

The second term is of order $\frac{1}{N}$ and, as we shall see, negligible compared to the first one, as, from a) and due to $b_N \to 0$

$$\mathbb{E} K_H (x - X_1) = \lambda(x) + \text{bias}\hat{\lambda}(x, H) = \lambda(x) + O\left(b_N^2 \right) = O(1).$$

So, we have to investigate

$$\begin{aligned}\mathbb{E}K_H^2(x-X_1) &= \int\cdots\int K_H^2(x-z)\,\lambda(z)\,\mathrm{d}z_1\cdots\mathrm{d}z_d\\ &= \frac{1}{\det H}\int\cdots\int K^2(u)\,\lambda(x-Hu)\,\mathrm{d}u_1\cdots\mathrm{d}u_d\end{aligned}$$

again substituting $u = H^{-1}(x-z)$. By a Taylor expansion of order 1, we have with $0<\theta<1$

$$\begin{aligned}\mathbb{E}K_H^2(x-X_1) &= \frac{1}{\det H}\int\cdots\int K^2(u)\,\mathrm{d}u_1\cdots\mathrm{d}u_d\lambda(x)\\ &\quad -\frac{1}{\det H}\int\cdots\int K^2(u)\,u^\top H^\top\nabla\lambda(x-\theta Hu)\,\mathrm{d}u_1\cdots\mathrm{d}u_d\\ &= \frac{Q_K\lambda(x)}{\det H}+\frac{1}{\det H}O\left(\sum_{k=1}^d h_k\right)\end{aligned}$$

as $\nabla\lambda(x)$ is continuous on $[0,1]^d$ and, hence, bounded, and the range of integration is the support of K, i.e. $[-1,+1]^d$. Note that from Assumption 5

$$\begin{aligned}\frac{O\left(\sum_{k=1}^d h_k\right)}{\det H} &= \frac{O(b_N)}{\det H}\\ &= \frac{O(b_N)}{\prod_{i=1}^d h_i}\\ &= O\left(\frac{1}{\prod_{i=1}^d \beta_i}\right)O\left(\frac{1}{b_N^{d-1}}\right)\\ &= O\left(\frac{1}{b_N^{d-1}}\right)\end{aligned}$$

such that finally we get b).

□

Corollary 3. *Under the assumptions of Theorem 2,*

a)

$$\mathrm{mse}\hat{\lambda}(x,H)=\mathrm{amse}(x,H)+O\left(\frac{1}{Nb_N^{d-1}}\right)+O\left(b_N^{4+\beta}\right)$$

with asymptotic mean-squared error

$$\text{amse}(x,H) = \frac{Q_K}{N\det H}\lambda(x) + \frac{1}{4}V_K^2\left(\sum_{i=1}^d h_i^2\lambda_{ii}(x)\right)^2.$$

b) Let $w_H(x) = \prod_{i=1}^d \mathbf{1}_{[h_i,1-h_i]}(x_i)$ be the indicator function of the interior hyper-rectangle $[h_1, 1-h_1]\times\cdots\times[h_d,1-h_d]$ of $[0,1]^d$. Then, the mean-integrated squared error over this region is

$$\begin{aligned}\text{mise}\hat{\lambda}(\cdot,H) &= \mathbb{E}\int\cdots\int\left(\hat{\lambda}(x,H)-\lambda(x)\right)^2 w_H(x)\,\mathrm{d}x_1\cdots\mathrm{d}x_d \\ &= \text{amise}(H) + O\left(\frac{1}{Nb_N^{d-1}}\right) + O\left(b_N^{4+\beta}\right)\end{aligned}$$

with

$$\text{amise}(H) = \frac{Q_K}{N\det H} + \frac{1}{4}V_K^2\int\cdots\int\left(\sum_{i=1}^d h_i^2\lambda_{ii}(x)\right)^2\mathrm{d}x_1\cdots\mathrm{d}x_d.$$

Proof. Part a) follows immediately from Theorem 2 and the bias-variance decomposition of $\text{mse}\hat{\lambda}(x,H)$. Note that from the proof of Theorem 2, the remainder terms can be chosen uniform with respect to $h_i \le x_i \le 1-h_i, i = 1,\ldots,d$. Therefore, integrating the relation a) results in

$$\text{mise}\hat{\lambda}(\cdot,H) = \int\cdots\int \text{amse}(x,H)\,w_H(x)\,\mathrm{d}x_1\cdots\mathrm{d}x_d + O\left(\frac{1}{Nb_N^{d-1}}\right) + O\left(b_N^{4+\beta}\right).$$

As $\lambda(x)$ and $\lambda_{ii}(x), i=1,\ldots,d$, are bounded on $[0,1]^d$ and, hence, $\left(\sum_{i=1}^d h_i^2\lambda_{ii}(x)\right)^2 = O\left(b_N^4\right)$ uniformly in x, we get, using

$$\int_0^1\cdots\int_0^1(1-w_H(x))\,\mathrm{d}x_1\cdots\mathrm{d}x_d = 1-\prod_{i=1}^d(1-2h_i) = O(b_N),$$

that

$$\int_0^1\cdots\int_0^1\lambda(x)\,w_H(x)\,\mathrm{d}x_1\cdots\mathrm{d}x_d = \int_0^1\cdots\int_0^1\lambda(x)\,\mathrm{d}x_1\cdots\mathrm{d}x_d + O(b_N) = 1+O(b_N),$$

$$\begin{aligned}&\int_0^1\cdots\int_0^1\left(\sum_{i=1}^d h_i^2\lambda_{ii}(x)\right)^2 w_H(x)\,\mathrm{d}x_1\cdots\mathrm{d}x_d \\ =\ &\int_0^1\cdots\int_0^1\left(\sum_{i=1}^d h_i^2\lambda_{ii}(x)\right)^2\mathrm{d}x_1\cdots\mathrm{d}x_d + O\left(b_N^5\right).\end{aligned}$$

Hence, we have

$$\int_0^1 \cdots \int_0^1 \operatorname{amse}(x, H)\, w_H(x)\, \mathrm{d}x_1 \cdots \mathrm{d}x_d = \operatorname{amise}(H) + O\left(\frac{1}{N b_N^{d-1}}\right) + O\left(b_N^{4+\beta}\right)$$

which implies the assertion as $\beta \leq 1$. □

Note that our definition of $\operatorname{mise}\hat{\lambda}(\cdot, H)$ neglects the intensity estimates close to the boundary where boundary effects may lead to different convergence rates. However, for $N \to \infty$, $h_i \to 0, i = 1, \ldots, d$, this modification becomes negligible and has no effect on the asymptotic mean integrated squared error $\operatorname{amise}(H)$.

2.2 Optimal asymptotic bandwidth

We now focus on the case $d = 2$ to keep notation as simple as possible. For the function arguments, we now write $(x_1, x_2)^\top \in \mathbb{R}^2$, e.g.

$$\operatorname{amse}(x_1, x_2, H) = \frac{Q_K}{N h_1 h_2} \lambda(x_1, x_2) + \frac{1}{4} V_K^2 \left(h_1^2 \lambda_{11}(x_1, x_2) + h_2^2 \lambda_{22}(x_1, x_2)\right)^2.$$

This is of the form

$$\frac{A}{N h_1 h_2} + \frac{1}{4}\left(B_1 h_1^2 + B_2 h_2^2\right)^2.$$

To minimise it, we set the partial derivatives with respect to h_1 and h_2 to 0:

$$\begin{aligned} -\frac{A}{N h_1^2 h_2} + \left(B_1 h_1^2 + B_2 h_2^2\right) B_1 h_1 &= 0 \qquad (2.1) \\ -\frac{A}{N h_1 h_2^2} + \left(B_1 h_1^2 + B_2 h_2^2\right) B_2 h_2 &= 0. \end{aligned}$$

Multiplying these equations by h_1 respectively h_2 and subtracting the second one from the first one results in

$$0 = \left(B_1 h_1^2 + B_2 h_2^2\right)\left(B_1 h_1^2 - B_2 h_2^2\right) = B_1^2 h_1^4 - B_2^2 h_2^4$$

and, hence,

$$h_2 = h_1 \sqrt{\frac{|B_1|}{|B_2|}}. \tag{2.2}$$

Plugging this relation into (2.1) results in

$$\left(B_1 + B_2 \frac{|B_1|}{|B_2|}\right) B_1 \sqrt{\frac{|B_1|}{|B_2|}} h_1^6 = \frac{A}{N}.$$

Writing $B_i = s_i |B_i|$, $s_i = \text{sgn} B_i = \text{sgn}\lambda_{ii}(x_1, x_2), i = 1, 2$, and $\rho = \sqrt{\frac{|B_1|^5}{|B_2|}}$, we get

$$(s_1 + s_2) s_1 \rho h_1^6 = (1 + s_1 s_2) \rho h_1^6 = \frac{A}{N},$$

i.e. for $s_1 s_2 \geq 0$

$$h_1 = \frac{A^{\frac{1}{6}}}{N^{\frac{1}{6}}} \frac{1}{\rho^{\frac{1}{6}}} \frac{1}{(1 + s_1 s_2)^{\frac{1}{6}}}.$$

Plugging in the expressions for A, B_i, ρ, we then have for the asymptotically optimal local bandwidths, using also (2.2),

$$h_{a1}(x_1, x_2) = \frac{Q_K^{\frac{1}{6}}}{N^{\frac{1}{6}}} \lambda^{\frac{1}{6}}(x_1, x_2) \frac{1}{V_K^{\frac{1}{3}}} \frac{|\lambda_{22}(x_1, x_2)|^{\frac{1}{12}}}{|\lambda_{11}(x_1, x_2)|^{\frac{5}{12}}} \frac{1}{(1 + s_1 s_2)^{\frac{1}{6}}}, \tag{2.3}$$

$$h_{a2}(x_1, x_2) = \frac{Q_K^{\frac{1}{6}}}{N^{\frac{1}{6}}} \lambda^{\frac{1}{6}}(x_1, x_2) \frac{1}{V_K^{\frac{1}{3}}} \frac{|\lambda_{11}(x_1, x_2)|^{\frac{1}{12}}}{|\lambda_{22}(x_1, x_2)|^{\frac{5}{12}}} \frac{1}{(1 + s_1 s_2)^{\frac{1}{6}}}. \tag{2.4}$$

Note that our argument only works if $\lambda_{11}(x_1, x_2) \lambda_{22}(x_1, x_2) \geq 0$, which, e.g., holds if $\lambda(x_1, x_2)$ is locally convex or concave around (x_1, x_2), i.e. the Hessian of $\lambda(x_1, x_2)$ is non-negative or non-positive definite respectively.

If $s_1 s_2 < 0$, i.e. if $\lambda_{11}(x_1, x_2)$, $\lambda_{22}(x_1, x_2)$ have opposite signs, $\lambda(x_1, x_2)$ has a saddlepoint-like behaviour around (x_1, x_2) which is rather the exception than the rule. Therefore, we consider only locations (x_1, x_2) where $s_1 s_2 \geq 0$ in the following. If $s_1 s_2 < 0$, i.e. $s_1 s_2 = -1$, then the dominant term in the bias expansion of Theorem 2 vanishes for suitable h_1, h_2. For getting asymptotically optimal bandwidths for this situation, a more detailed investigation of the remainder term of order $O\left(b_N^{2+\beta}\right)$ in that expansion would be necessary.

Turning now to the choice of global bandwidths, the asymptotic mean integrated squared error is in the case $d = 2$

$$\text{amise}\,(H) = \frac{Q_K}{Nh_1h_2} + \frac{1}{4}\left(h_1^4 I_{11} + 2h_1^2h_2^2 I_{12} + h_2^4 I_{22}\right)$$

with

$$I_{k\ell} = V_K^2 \int_0^1 \int_0^1 \lambda_{kk}\,(x_1, x_2)\,\lambda_{\ell\ell}\,(x_1, x_2)\,\mathrm{d}x_1\mathrm{d}x_2, k, \ell = 1, 2.$$

Setting the partial derivatives with respect to h_1,h_2 to 0, we get

$$\begin{aligned} -\frac{Q_K}{Nh_1^2h_2} + h_1^3 I_{11} + h_1h_2^2 I_{12} &= 0 \\ -\frac{Q_K}{Nh_1h_2^2} + h_2^3 I_{22} + h_1^2h_2 I_{12} &= 0. \end{aligned} \qquad (2.5)$$

Multiplying these equations by h_1 respectively h_2 and subtracting the second from the first one results in

$$h_1^4 I_{11} - h_2^4 I_{22} = 0 \implies h_2 = h_1 \left(\frac{I_{11}}{I_{22}}\right)^{\frac{1}{4}}. \qquad (2.6)$$

Plugging these relation into (2.5) results in

$$\frac{I_{11}^{\frac{5}{4}}}{I_{22}^{\frac{1}{4}}}\left(1 + \frac{I_{12}}{(I_{11}I_{22})^{\frac{1}{2}}}\right) h_1^6 = \frac{Q_K}{N}.$$

Note that by the Cauchy-Schwarz inequality, $|I_{12}| \leq (I_{11}I_{22})^{\frac{1}{2}}$ with equality only if $\lambda_{22}\,(x_1, x_2) = c\lambda_{11}\,(x_1, x_2)$ a.e. for some $c \in \mathbb{R}$. Therefore, the term in brackets only vanishes for $\lambda_{22}\,(x_1, x_2) = c\lambda_{11}\,(x_1, x_2)$ with some $c < 0$, which is a very special case which we exclude as part of the following assumption which also requires that $\lambda_{11}\,(x_1, x_2)$ and $\lambda_{22}\,(x_1, x_2)$ do not vanish a.e.:

$$I_{11}, I_{22}, \sqrt{I_{11}I_{22}} + I_{12} \neq 0.$$

Then, we get

$$h_1 = \frac{Q_K^{\frac{1}{6}}}{N^{\frac{1}{6}}}\left(\frac{I_{22}}{I_{11}}\right)^{\frac{1}{8}} \frac{1}{\left(\sqrt{I_{11}I_{22}} + I_{12}\right)^{\frac{1}{6}}}.$$

Using (2.6) and defining

$$\Lambda_{k\ell} = \int_0^1 \int_0^1 \lambda_{kk}(x_1, x_2)\, \lambda_{\ell\ell}(x_1, x_2)\, \mathrm{d}x_1 \mathrm{d}x_2, k, \ell = 1, 2,$$

such that $I_{k\ell} = V_K^2 \Lambda_{k\ell}$, we get for the asymptotically optimal global bandwidths

$$h_{a1} = \left(\frac{Q_K}{V_K^2}\right)^{\frac{1}{6}} \left(\frac{\Lambda_{22}}{\Lambda_{11}}\right)^{\frac{1}{8}} \frac{1}{\left(\sqrt{\Lambda_{11}\Lambda_{22}} + \Lambda_{12}\right)^{\frac{1}{6}}} \left(\frac{1}{N}\right)^{\frac{1}{6}} \tag{2.7}$$

$$h_{a2} = \left(\frac{Q_K}{V_K^2}\right)^{\frac{1}{6}} \left(\frac{\Lambda_{11}}{\Lambda_{22}}\right)^{\frac{1}{8}} \frac{1}{\left(\sqrt{\Lambda_{11}\Lambda_{22}} + \Lambda_{12}\right)^{\frac{1}{6}}} \left(\frac{1}{N}\right)^{\frac{1}{6}}. \tag{2.8}$$

We summarise the above derivations in the following theorem.

Theorem 3. *Let the assumptions of Theorem 2 be satisfied, and let* $d = 2$*.*

a) The bandwidths h_{a1}, h_{a2} *minimising* $\mathrm{amise}(H)$ *of Corollary 3 are given by (2.7), (2.8) if* $\Lambda_{11}, \Lambda_{22}, \sqrt{\Lambda_{11}\Lambda_{22}} + \Lambda_{12} \neq 0$.

b) The bandwidths $h_{a1}(x_1, x_2)$, $h_{a2}(x_1, x_2)$ *minimising* $\mathrm{amse}(x_1, x_2, H)$ *of Corollary 3 are given by (2.3), (2.4) if* $\lambda_{11}(x_1, x_2)\, \lambda_{22}(x_1, x_2) > 0$.

The asymptotically optimal bandwidths, hence, are locally and globally of order $N^{-\frac{1}{6}}$. The following corollary gives the rates of approximation of mse and mise by their asymptotic equivalents for this case. It follows immediately from Corollary 3.

Corollary 4. *Let the assumptions of Theorem 2 be fulfilled for* $d = 2$*, and let* $h_i N^{\frac{1}{6}} \to c_i > 0$ *for* $N \to \infty$*,* $i = 1, 2$*. Then for* $N \to \infty$*,*

a) $\mathrm{mse}\hat{\lambda}(x_1, x_2, H) - \mathrm{amse}(x_1, x_2, H) = o\left(N^{-\frac{2}{3}}\right)$,

b) $\mathrm{mise}\hat{\lambda}(\cdot, H) - \mathrm{amise}(H) = o\left(N^{-\frac{2}{3}}\right)$.

The next result states that the asymptotically optimal bandwidths of Theorem 3 approximate the finite-sample size optimal bandwidths for $N \to \infty$. Let $h_{0i}(x_1, x_2)$, h_{0i} denote for $i = 1, 2$ the bandwidth minimising $\text{mse}\hat{\lambda}(x, H)$ and $\text{mise}\hat{\lambda}(\cdot, H)$ respectively.

Corollary 5. *Let the assumptions of Theorem 3 be fulfilled. Then,*

a) $h_{0i}(x_1, x_2) = h_{ai}(x_1, x_2) + o\left(N^{-\frac{1}{6}}\right)$, $i = 1, 2$, *for all* $x_1, x_2 \in (0, 1)$*;*

b) $h_{0i} = h_{ai} + o\left(N^{-\frac{1}{6}}\right)$, $i = 1, 2$.

Proof. We only prove a), as b) can be shown analogously with a bit more notation. First we remark that $N^{\frac{1}{6}} h_{0i}(x_1, x_2)$ cannot converge to 0 or ∞. Otherwise, by Corollary 3, $N^{\frac{2}{3}} \text{mse}\hat{\lambda}(x_1, x_2, H_0) \to \infty$ with H_0 being the diagonal matrix with entries $h_{01}(x_1, x_2)$, $h_{02}(x_1, x_2)$. Let H_a denote the diagonal matrix with entries $h_{a1}(x_1, x_2)$, $h_{a2}(x_1, x_2)$ correspondingly. Then by Corollary 4

$$N^{\frac{2}{3}} \text{mse}\hat{\lambda}(x_1, x_2, H_a) = N^{\frac{2}{3}} \text{amse}(x_1, x_2, H_a) + o(1) \to C > 0$$

for $N \to \infty$, using that by Theorem 3, $N^{\frac{1}{6}} h_{ai}(x_1, x_2) \to c_i > 0$, $i = 1, 2$, and the expression for $\text{amse}(x_1, x_2, H)$ from Corollary 3. Hence, if $N^{\frac{2}{3}} \text{mse}\hat{\lambda}(x_1, x_2, H_0) \to \infty$ for large enough N, we would have

$$\text{mse}\hat{\lambda}(x_1, x_2, H_0) > \text{mse}\hat{\lambda}(x_1, x_2, H_a)$$

in contradiction to the definition of H_0. So, we have $N^{\frac{1}{6}} h_0(x_1, x_2) \to c_i^0$ for some $c_i^0 > 0$, $i = 1, 2$.

Hence, we have from Corollary 4

$$\begin{aligned} \text{mse}\hat{\lambda}(x_1, x_2, H_0) &= \text{amse}(x_1, x_2, H_0) + o\left(N^{-\frac{2}{3}}\right), \\ \text{mse}\hat{\lambda}(x_1, x_2, H_a) &= \text{amse}(x_1, x_2, H_a) + o\left(N^{-\frac{2}{3}}\right). \end{aligned}$$

Subtracting the first from the second relationship, we get

$$\text{mse}\hat{\lambda}(x_1, x_2, H_a) - \text{mse}\hat{\lambda}(x_1, x_2, H_0) = \text{amse}\hat{\lambda}(x_1, x_2, H_a) - \text{amse}\hat{\lambda}(x_1, x_2, H_0) + o\left(N^{-\frac{2}{3}}\right).$$

The left-hand side is non-negative as H_0 is mse-optimal, and the difference on the right-hand side is non-positive as H_a is amse-optimal. We conclude that both differences have to be 0 up to terms of order $o\left(N^{-\frac{2}{3}}\right)$, in particular

$$\text{amse}\left(x_1, x_2, H_a\right) - \text{amse}\left(x_1, x_2, H_0\right) = o\left(N^{-\frac{2}{3}}\right). \tag{2.9}$$

From a Taylor expansion of $\text{amse}\left(x_1, x_2, H\right)$ around H_a, we get with $\delta = \left(\delta_1, \delta_2\right)^\top$, $\delta_i = h_{0i}\left(x_1, x_2\right) - h_{ai}\left(x_1, x_2\right),\ i = 1, 2$

$$\text{amse}\left(x_1, x_2, H_0\right) - \text{amse}\left(x_1, x_2, H_a\right) = \nabla_h^\top \text{amse}\left(x_1, x_2, H_a\right)\delta + \frac{1}{2}\delta^\top \nabla_h^2 \text{amse}\left(x_1, x_2, \tilde{H}\right)\delta, \tag{2.10}$$

where ∇_h, ∇_h^2 denote gradient and Hessian with respect to $h = \left(h_1, h_2\right)^\top$, and $\tilde{H} = (1-\theta)\, H_a + \theta H_0$ for some $0 \leq \theta \leq 1$. As H_a minimises amse, the gradient in the first term on the right-hand side is 0. $\tilde{H}$ is a diagonal matrix with entries, say, $\tilde{h}_1, \tilde{h}_2$ which also satisfy $N^{\frac{1}{6}}\tilde{h}_i \to \tilde{c}_i > 0, i = 1, 2$. From Corollary 3, we get for the elements of the Hessian

$$\begin{aligned}
\frac{\partial^2}{\partial h_1^2}\text{amse}\left(x_1, x_2, H\right) &= \frac{2Q_K \lambda\left(x_1, x_2\right)}{N h_1^3 h_2} + V_K^2\left(3h_1^2 \lambda_{11}^2\left(x_1, x_2\right) + h_2^2 \lambda_{11}\left(x_1, x_2\right)\lambda_{22}\left(x_1, x_2\right)\right),\\
\frac{\partial^2}{\partial h_2^2}\text{amse}\left(x_1, x_2, H\right) &= \frac{2Q_K \lambda\left(x_1, x_2\right)}{N h_1 h_2^3} + V_K^2\left(h_1^2 \lambda_{11}\left(x_1, x_2\right)\lambda_{22}\left(x_1, x_2\right) + 3h_2^2 \lambda_{22}^2\left(x_1, x_2\right)\right),\\
\frac{\partial^2}{\partial h_1 \partial h_2}\text{amse}\left(x_1, x_2, H\right) &= \frac{2Q_K \lambda\left(x_1, x_2\right)}{N h_1^2 h_2^2} + 2V_K^2 h_1 h_2 \lambda_{11}\left(x_1, x_2\right)\lambda_{22}\left(x_1, x_2\right).
\end{aligned}$$

By the assumptions of Theorem 3, these terms are all greater than 0, and for $h_i = \tilde{h}_i, i = 1, 2$, they are of order $N^{-\frac{1}{3}}$. Hence we get from (2.9) and (2.10) for some $c > 0$

$$o\left(N^{-\frac{2}{3}}\right) = c\left\|\delta\right\|^2 N^{-\frac{1}{3}},$$

i.e.

$$\left\|h_0\left(x_1, x_2\right) - h_a\left(x_1, x_2\right)\right\|^2 = o\left(N^{-\frac{1}{3}}\right)$$

which implies the assertion. □

2.3 The case of smoother intensity functions

For later use, we briefly discuss in this section how the previous results change if we assume more smoothness about $\lambda(x)$. It is well-known from Stone (1984) seminal work that optimal bandwidths mainly depend on sample size, dimension and smoothness, i.e. assumed degree of differentiability, of the function to be estimated. In particular, we replace Assumption 3 now by

Assumption 6. $\lambda(x)$ *is a four times continuously differentiable probability density with support* $[0,1]^d$.

Moreover, we augment the symmetry Assumption 4 on K by

Assumption 7. $\int \cdots \int u_i^3 K(u)\, \mathrm{d}u_1 \cdots \mathrm{d}u_d = 0$ *for* $i = 1, \ldots, d$.

Then, we get an improved rate for the bias expansion in Theorem 2.

Proposition 2. *Let the assumptions of Theorem 2 and Assumptions 6 and 7 be satisfied. Then,*

$$\text{bias}\hat{\lambda}(x,H) = \frac{1}{2} V_K \sum_{i=1}^{d} h_i^2 \lambda_{ii}(x) + O\left(b_N^4\right).$$

Proof. We write $\lambda_i(x) = \frac{\partial}{\partial x_i}\lambda(x)$, $\lambda_{ijk}(x) = \frac{\partial^3}{\partial x_i \partial x_j \partial x_k}\lambda(x)$ and $\lambda_{ijk\ell}(x) = \frac{\partial^4}{\partial x_i \partial x_j \partial x_k \partial x_\ell}\lambda(x)$. We now can extend the Taylor expansion of $\lambda(x - Hu)$ in part a) of Theorem 2 to order 4:

$$\begin{aligned}
\lambda(x - Hu) - \lambda(x) &= -\sum_{i=1}^{d} h_i u_i \lambda_i(x) + \frac{1}{2} \sum_{i,j=1}^{d} h_i u_i h_j u_j \lambda_{ij}(x) \\
&\quad - \frac{1}{6} \sum_{i,j,k=1}^{d} h_i u_i h_j u_j h_k u_k \lambda_{ijk}(x) \\
&\quad + \frac{1}{24} \sum_{i,j,k,\ell=1}^{d} h_i u_i h_j u_j h_k u_k h_\ell u_\ell \lambda_{ijk\ell}(x - \theta H u)
\end{aligned}$$

for some $0 \le \theta \le 1$. If we multiply this by $K(u)$ and integrate with respect to u, then the first term on the right-hand side vanishes due to the symmetry

Assumption 4 on K. Analogously, the third term vanishes by Assumptions 4 and 7, and the components with $i \neq j$ in the second term also vanish. As, by Assumption 6, λ_{ijkl} is bounded, as K has compact support and as $|h_i h_j h_k h_\ell| \leq \sum_{i=1}^d h_i^4$, the fourth term is $O\left(\sum_{i=1}^d h_i^4\right) = O\left(b_N^4\right)$ from Assumption 5. □

As the dominant term in the bias expansion does not change by assuming more smoothness of λ, the asymptotically optimal bandwidths do not change. However, the approximation rate of the mse by the asymptotic mse improves, and we have the following analogue of Corollary 4. Recall that now we consider $d = 2$ only.

Corollary 6. *Let the assumptions of Corollary 4 and Assumptions 6 and 7 be satisfied. Then, for $h_i N^{\frac{1}{6}} \to c_i > 0, i = 1, 2,$ we have for $N \to \infty$*

a) $\operatorname{mse}\hat{\lambda}\left(x_1, x_2, H\right) - \operatorname{amse}\left(x_1, x_2, H\right) = O\left(N^{-\frac{5}{6}}\right)$,

b) $\operatorname{mise}\hat{\lambda}\left(\cdot, H\right) - \operatorname{amise}\left(H\right) = O\left(N^{-\frac{5}{6}}\right)$.

Proof. From Proposition 2, we have as in Corollary 3 the mse expansion

$$\operatorname{mse}\hat{\lambda}\left(x, H\right) = \operatorname{amse}\left(x, H\right) + O\left(\frac{1}{N b_N}\right) + O\left(b_N^6\right).$$

From the assumption on h_i, we have $b_N N^{\frac{1}{6}} \to c > 0$ such that

$$O\left(\frac{1}{N b_N}\right) + O\left(b_N^6\right) = O\left(N^{-\frac{5}{6}}\right).$$

b) follows as in the proof of Corollary 3 from a). □

Note that we could get better rates of convergence by using higher-order kernels if Assumption 6 is satisfied. However, for those kernels we would have $\int\int u_i^2 K\left(u\right) \mathrm{d}u_1 \mathrm{d}u_2 = 0$, i.e. K has to assume also negative values. This might lead to negative estimates $\hat{\lambda}\left(x_1, x_2\right)$ of the positive function $\lambda\left(x_1, x_2\right)$. Therefore, we do not investigate this direction further. We have to impose

Assumption 6 due to different reasons below: we need kernel estimates of the second derivatives $\lambda_{11}(x_1, x_2), \lambda_{22}(x_1, x_2)$, and for this we assume that they are twice continuously differentiable, i.e. Assumption 6. Finally, we also get better approximation rates of the optimal bandwidths by the asymptotically optimal ones, i.e. the following analogue of Corollary 5.

Corollary 7. *Let the assumptions of Corollary 5 and Assumptions 6 and 7 be satisfied. Then,*

a) $h_{0i}(x_1, x_2) = h_{ai}(x_1, x_2) + O\left(N^{-\frac{1}{4}}\right), i = 1, 2,$ *for all* $x_1, x_2 \in (0, 1)$*;*

b) $h_{0i} = h_{ai} + O\left(N^{-\frac{1}{4}}\right), i = 1, 2.$

Proof. The proof proceeds exactly as the proof of Corollary 5, except that we use the rate $O\left(N^{-\frac{5}{6}}\right)$ instead of $o\left(N^{-\frac{2}{3}}\right)$ for the approximation of mse by amse. In particular, we get

$$\|h_0(x_1, x_2) - h_a(x_1, x_2)\|^2 = N^{\frac{1}{3}} O\left(N^{-\frac{5}{6}}\right) = O\left(N^{-\frac{1}{2}}\right)$$

and, hence, $\|h_0(x_1, x_2) - h_a(x_1, x_2)\| = O\left(N^{-\frac{1}{4}}\right)$. □

Chapter 3

Data adaptive bandwidth selection by the plug-in approach

In this chapter, we use the formulae for the asymptotic optimal bandwidths and plug in estimates for the unknown quantities to get an iterative algorithm for selecting the bandwidths automatically from the data. Subsequently, we investigate the asymptotic behaviour of the kernel estimates with random bandwidths. Note that the algorithm depends on the choice of various tuning parameters which we leave unspecified for the moment. A rule how to choose them requires a more careful analysis of the asymptotics of estimates of the second derivatives of the intensity function which will be done in subsequent chapters. To keep notation simple, we only consider $d = 2$ here. The necessary adaptation to higher dimensions will be discussed later.

3.1 An algorithm for automatic bandwidth selection

In this section, we extend the ideas of Engel et al. (1994) to higher dimensions and formulate iterative procedures resulting in estimates $\hat{h}$ respectively $\hat{h}\,(x_1, x_2)$ for the globally respectively locally optimal bandwidths $h_0, h_0\,(x_1, x_2)$. For this purpose, we use estimates of $h_a, h_a\,(x_1, x_2)$ by (2.7), (2.8) respectively (2.3), (2.4). As those depend on unknown quantities involving, in particular, the second derivatives of $\lambda\,(x_1, x_2)$ respectively the integrals $\Lambda_{k\ell}, k, \ell = 1, 2$, and, in the local case, also $\lambda\,(x_1, x_2)$ itself, we need an iterative procedure which we formulate below.

First, we have to briefly discuss how to estimate $\lambda_{k\ell}\,(x_1, x_2)$ respectively $\Lambda_{k\ell}$, $k, \ell = 1, 2$. For the second derivatives, we use as usual the second derivatives of the kernel estimate $\hat{\lambda}\,(x_1, x_2, H)$, i.e.

$$
\begin{aligned}
\hat{\lambda}_{11}\,(x_1, x_2, H) &= \frac{\partial^2}{\partial x_1^2}\hat{\lambda}\,(x_1, x_2, H) = \frac{1}{Nh_1h_2}\sum_{j=1}^{N}\frac{\partial^2}{\partial x_1^2}K\left(\frac{x_1 - X_{j1}}{h_1}, \frac{x_2 - X_{j2}}{h_2}\right) \\
&= \frac{1}{Nh_1^3h_2}\sum_{j=1}^{N}K_{11}\left(\frac{x_1 - X_{j1}}{h_1}, \frac{x_2 - X_{j2}}{h_2}\right), \\
\hat{\lambda}_{22}\,(x_1, x_2, H) &= \frac{1}{Nh_1h_2^3}\sum_{j=1}^{N}K_{22}\left(\frac{x_1 - X_{j1}}{h_1}, \frac{x_2 - X_{j2}}{h_2}\right), \\
\hat{\lambda}_{12}\,(x_1, x_2, H) &= \frac{1}{Nh_1^2h_2^2}\sum_{j=1}^{N}K_{12}\left(\frac{x_1 - X_{j1}}{h_1}, \frac{x_2 - X_{j2}}{h_2}\right),
\end{aligned}
$$

where $K_{k\ell}\,(u_1, u_2) = \frac{\partial^2}{\partial u_k \partial u_\ell}K\,(u_1, u_2)$ denote the second derivatives of the kernel function, $k, \ell = 1, 2$.

Following Gasser et al. (1991), in the iteration we use larger bandwidths for the estimates of $\lambda_{k\ell}\,(x_1, x_2)$ than for the estimates of $\lambda\,(x_1, x_2)$. More precisely, for some inflation factor $N^\rho, \rho > 0$, if we use h_1, h_2 for $\hat{\lambda}\,(x_1, x_2, H)$, then we use $N^\rho h_1, N^\rho h_2$ for estimating $\lambda_{k\ell}\,(x_1, x_2)$, i.e. we consider the estimate

$\hat{\lambda}_{k\ell}\left(x_1, x_2, N^\rho H\right)$. We discuss appropriate choices for ρ later.

If we are interested in the local bandwidths, we nevertheless have to start with global bandwidths first and switch to local ones later in the iteration. The reason is rather large variability of $\hat{\lambda}_{k\ell}\left(x_1, x_2, N^\rho H\right)$ for the initially small bandwidths which would lead to instability of the algorithm. For the integrals $\Lambda_{k\ell}$ appearing in amise, this effect is not important as the variability averages out by integration. For $\Lambda_{k\ell}$, we use the estimates

$$\hat{\Lambda}_{k\ell}\left(N^\rho H\right) = \int_0^1 \int_0^1 \hat{\lambda}_{kk}\left(x_1, x_2, N^\rho H\right) \hat{\lambda}_{\ell\ell}\left(x_1, x_2, N^\rho H\right) v\left(x_1, x_2\right) \mathrm{d}x_1 \mathrm{d}x_2, k, \ell = 1, 2,$$

where $v\left(x_1, x_2\right)$ is a weight function integrating to 1 which we introduce to avoid boundary effects.

In the following $\hat{h}_k^{(i)}$, $k = 1, 2$, $\hat{H}^{(i)}$ denote the global bandwidths and the corresponding diagonal bandwidth matrix in the ith step of the iteration, and $\hat{h}_k^{(i)}\left(x_1, x_2\right), k = 1, 2$, $\hat{H}^{(i)}\left(x_1, x_2\right)$ denote the corresponding local quantities in later steps of the iteration. We also use the abbreviations

$$\begin{aligned}
\hat{\Lambda}_{k\ell}^{(i)} &= \hat{\Lambda}_{k\ell}\left(N^\rho \hat{H}^{(i)}\right), k, \ell = 1, 2, \\
\hat{\lambda}_{k\ell}^{(i)}\left(x_1, x_2\right) &= \hat{\lambda}_{k\ell}\left(x_1, x_2, N^\rho \hat{H}^{(i)}\left(x_1, x_2\right)\right), k, \ell = 1, 2, \\
\hat{\lambda}^{(i)}\left(x_1, x_2\right) &= \hat{\lambda}\left(x_1, x_2, \hat{H}^{(i)}\left(x_1, x_2\right)\right).
\end{aligned}$$

Step 0: We initialise the algorithm by choosing $\hat{h}_k^{(0)} = \frac{1}{\sqrt{N}}, k = 1, 2$.

Step 1: For $i = 1, \ldots, i^*$, iterate

$$\begin{aligned}
\hat{h}_1^{(i)} &= \left(\frac{Q_K}{NV_K^2}\right)^{\frac{1}{6}} \left(\frac{\hat{\Lambda}_{22}^{(i-1)}}{\hat{\Lambda}_{11}^{(i-1)}}\right)^{\frac{1}{8}} \frac{1}{\left(\sqrt{\hat{\Lambda}_{11}^{(i-1)}\hat{\Lambda}_{22}^{(i-1)}} + \hat{\Lambda}_{12}^{(i-1)}\right)^{\frac{1}{6}}}, \\
\hat{h}_2^{(i)} &= \hat{h}_1^{(i)} \left(\frac{\hat{\Lambda}_{11}^{(i-1)}}{\hat{\Lambda}_{22}^{(i-1)}}\right)^{\frac{1}{4}}, \\
\hat{h}_k^{(i)} &= \min\left(\hat{h}_k^{(i)}, \frac{1}{2\sqrt{N}}\right), k = 1, 2, \\
\hat{h}_k^{(i)} &= \max\left(\hat{h}_k^{(i)}, \frac{1}{2}\right), k = 1, 2.
\end{aligned}$$

Step 2: Set $\hat{H}^{(i^*)}(x_1, x_2) = \hat{H}^{(i^*)}$. For $i = i^*{+}1, \ldots, j^*$, assuming $\lambda_{11}(x_1, x_2)\, \lambda_{22}(x_1, x_2) > 0$:

$$\begin{aligned}
\hat{h}_1^{(i)}(x_1, x_2) &= \left(\frac{Q_K \hat{\lambda}^{(i-1)}(x_1, x_2)}{2NV_K^2}\right)^{\frac{1}{6}} \frac{\left|\hat{\lambda}_{22}^{(i-1)}(x_1, x_2)\right|^{\frac{1}{12}}}{\left|\hat{\lambda}_{11}^{(i-1)}(x_1, x_2)\right|^{\frac{5}{12}}}, \\
\hat{h}_2^{(i)}(x_1, x_2) &= \hat{h}_1^{(i)}(x_1, x_2) \left|\frac{\hat{\lambda}_{11}^{(i-1)}(x_1, x_2)}{\hat{\lambda}_{22}^{(i-1)}(x_1, x_2)}\right|^{\frac{1}{2}}.
\end{aligned}$$

3.2 Asymptotics of kernel estimates with data-adaptive bandwidth

The global and local bandwidths, derived from the data in the last section, depend on estimates of the density and of its second derivatives $\lambda_{ii}(x_1, x_2), i = 1, 2$. As a preliminary result, we briefly investigate the asymptotic mean-squared error of the kernel estimates $\lambda_{ii}(x_1, x_2, H)$ for deterministic bandwidths. In the following, we write

$$\lambda_i(x_1, x_2) = \frac{\partial}{\partial x_i}\lambda(x_1, x_2), K_i(u_1, u_2) = \frac{\partial}{\partial u_i}K(u_1, u_2), i = 1, 2,$$

for the first-order partial derivatives. We assume about the kernel that:

Assumption 8. *K is twice differentiable on $[-1,+1]^2$ with bounded second derivatives, and satisfies the boundary conditions $K(\pm 1, u_2) = K(u_1, \pm 1) = 0$, $K_1(\pm 1, u_2) = K_1(u_1, \pm 1) = 0$ for all $-1 \le u_1, u_2 \le 1$.*

Proposition 3. *Let the assumptions of Theorem 2, Assumptions 6 and 8 be satisfied. Assume, moreover, that the fourth-order partial derivative of $\lambda(x_1, x_2)$ are Hölder continuous with exponent $\beta > 0$. Then, for $i = 1, 2$*

a) $\mathbb{E}\hat{\lambda}_{ii}(x_1, x_2, H) = \lambda_{ii}(x_1, x_2) + \frac{1}{2} V_K \sum_{\ell=1}^{2} h_\ell^2 \frac{\partial^2}{\partial x_\ell^2} \lambda_{ii}(x_1, x_2) + O\left(b_N^{2+\beta}\right)$.

b) $\operatorname{var}\hat{\lambda}_{ii}(x_1, x_2, H) = \frac{1}{N h_1 h_2 h_i^4}\left(Q_K^{ii} \lambda(x_1, x_2) + O(b_N)\right)$ *with* $Q_K^{ii} = \int\int K_{ii}^2(u_1, u_2)\, \mathrm{d}u_1 \mathrm{d}u_2$.

Proof. We consider only $\hat{\lambda}_{11}(x_1, x_2, H)$, as the arguments for $\hat{\lambda}_{22}$ are the same.

a) As in the proof of Theorem 2, a) we have

$$\begin{aligned}
\mathbb{E}\hat{\lambda}_{11}(x_1, x_2, H) &= \frac{1}{h_1^3 h_2} \mathbb{E} K_{11}\left(\frac{x_1 - X_{11}}{h_1}, \frac{x_2 - X_{12}}{h_2}\right) \\
&= \frac{1}{h_1^2} \int_{-1}^{1} \int_{-1}^{1} K_{11}(u_1, u_2)\, \lambda(x_1 - h_1 u_1, x_2 - h_2 u_2)\, \mathrm{d}u_1 \mathrm{d}u_2.
\end{aligned}$$

Using integration by parts twice and, in particular, e.g.,

$$\frac{\partial}{\partial u_1} \lambda(x_1 - h_1 u_1, x_2 - h_2 u_2) = -h_1 \lambda_1(x_1 - h_1 u_1, x_2 - h_2 u_2),$$

we get

$$\begin{aligned}
\mathbb{E}\hat{\lambda}_{11}(x_1, x_2, H) &= \frac{1}{h_1} \int_{-1}^{1} \int_{-1}^{1} K_1(u_1, u_2)\, \lambda_1(x_1 - h_1 u_1, x_2 - h_2 u_2)\, \mathrm{d}u_1 \mathrm{d}u_2 \\
&= \int_{-1}^{1} \int_{-1}^{1} K(u_1, u_2)\, \lambda_{11}(x_1 - h_1 u_1, x_2 - h_2 u_2)\, \mathrm{d}u_1 \mathrm{d}u_2
\end{aligned}$$

as the constant terms in the integration-by-parts relations vanish due to Assumption 8. Using the same Taylor expansion arguments as in the proof of Theorem 2 with λ_{11} instead of λ, we get

$$\mathbb{E}\hat{\lambda}_{11}(x_1, x_2, H) = \lambda_{11}(x_1, x_2) + \frac{1}{2} V_K \sum_{i=1}^{2} h_i^2 \frac{\partial^2}{\partial x_i^2} \lambda_{11}(x_1, x_2) + O\left(b_N^{2+\beta}\right).$$

b) As $X_1, \ldots, X_N$ are i.i.d.,

$$\begin{aligned}
\operatorname{var}\hat{\lambda}_{11}(x_1, x_2, H) &= \frac{1}{N}\operatorname{var}\left(\frac{1}{h_1^3 h_2}K_{11}\left(\frac{x_1 - X_{11}}{h_1}, \frac{x_2 - X_{12}}{h_2}\right)\right) \\
&= \frac{1}{N h_1^6 h_2^2}\mathbb{E}K_{11}^2\left(\frac{x_1 - X_{11}}{h_1}, \frac{x_2 - X_{12}}{h_2}\right) \\
&\quad -\frac{1}{N}\left\{\mathbb{E}\frac{1}{h_1^3 h_2}K_{11}\left(\frac{x_1 - X_{11}}{h_1}, \frac{x_2 - X_{12}}{h_2}\right)\right\}^2 .
\end{aligned}$$

As in the proof of Theorem 2, b), the second term is of order $O\left(\frac{1}{N}\right)$ and therefore negligible as from part a)

$$\mathbb{E}\frac{1}{h_1^3 h_2}K_{11}\left(\frac{x_1 - X_{11}}{h_1}, \frac{x_2 - X_{12}}{h_2}\right) = \lambda_{11}(x_1, x_2) + O\left(b_N^2\right) = O(1) .$$

For the first term, we use the same Taylor expansion argument as in the proof of Theorem 2 b) to get

$$\begin{aligned}
&\frac{1}{h_1^6 h_2^2}\mathbb{E}K_{11}^2\left(\frac{x_1 - X_{11}}{h_1}, \frac{x_2 - X_{12}}{h_2}\right) \\
&= \frac{1}{h_1^6 h_2^2}\int\int K_{11}^2\left(\frac{x_1 - u_1}{h_1}, \frac{x_2 - u_2}{h_2}\right)\lambda(u_1, u_2)\,\mathrm{d}u_1\mathrm{d}u_2 \\
&= \frac{1}{h_1^5 h_2}\int\int K_{11}^2(u_1, u_2)\,\lambda(x_1 - h_1 u_1, x_2 - h_2 u_2)\,\mathrm{d}u_1\mathrm{d}u_2 \\
&= \frac{1}{h_1^5 h_2}\left(\int\int K_{11}^2(u_1, u_2)\,\mathrm{d}u_1\mathrm{d}u_2\,\lambda(x_1, x_2) + O(h_1 + h_2)\right) .
\end{aligned}$$

□

In the following, we impose a Lipschitz condition on K_{ii}.

Assumption 9. *K_{ii} is Lipschitz continuous with constant $L_{ii}, i = 1, 2$,*

$$|K_{ii}(u) - K_{ii}(v)| \leq L_{ii}|u - v|$$

for all $-1, u, v \leq 1$.

This condition could be relaxed to Hölder continuity with exponent $0 < \beta \leq 1$, but for ease of notation we choose $\beta = 1$.

We now consider random bandwidths h_1, h_2 which may depend on the data $(X_{j1}, X_{j2}), j = 1, \ldots, N$. We assume the following conditions which we later on guarantee to hold by construction:

Condition 1. $h_1, h_2 \in \left[\frac{1}{2\sqrt{N}}, \delta\right]$ for some fixed $\delta \leq \frac{1}{2}$.

Condition 2. For some $\gamma \geq 0$ and deterministic $\beta_1, \beta_2 > 0$, $b_N \to 0$, $\frac{1}{b_N\sqrt{N}} = O(1)$, $h_i = \beta_i b_N (1 + o_p(N^{-\gamma})), i = 1, 2$.

In particular, from Condition 2 we have $h_i \to 0$, $i = 1, 2$, $\frac{h_1}{h_2} \to \frac{\beta_1}{\beta_2} > 0$. H again denotes the diagonal matrix with entries h_1, h_2.

Proposition 4. *Let the assumptions of Proposition 3 and Assumption 9 be satisfied and additionally Conditions 1 and 2. Then, for* $i = 1, 2$,

$$\hat{\lambda}_{ii}(x_1, x_2, H) = \lambda_{ii}(x_1, x_2) + O\left(b_N^2\right) + o_p\left(N^{-\gamma} b_N^2\right) + o_p\left(\frac{1}{\alpha_N b_N^4}\sqrt{\frac{\log N}{N}}\right)$$

for any sequence $\alpha_N > 0$ *with* $\alpha_N \to 0$ *and* $\sqrt{\frac{\log N}{N}} = O(\alpha_N)$.

Corollary 8. *Under the conditions of Proposition 4, for* $i = 1, 2$,

$$\hat{\lambda}_{ii}(x_1, x_2, H) = \lambda_{ii}(x_1, x_2) + O\left(b_N^2\right) + o_p\left(N^{-\gamma} b_N^2\right) + o_p\left(\frac{\log N}{\sqrt{N} b_N^4}\right).$$

Proof. The conditions of Proposition 4 are satisfied for $\alpha_N = \frac{1}{\sqrt{\log N}}$, and $\frac{1}{\alpha_N}\sqrt{\frac{\log N}{N}} = \frac{\log N}{\sqrt{N}}$. □

Proof. Proof of Proposition 4:

a) Let

$$\mu_{11}(H) = \int\int K_{11}\left(\frac{x_1 - u_1}{h_1}, \frac{x_2 - u_2}{h_2}\right) \lambda(u_1, u_2)\, \mathrm{d}u_1 \mathrm{d}u_2.$$

From the proof of Proposition 3 a), which works for random h_1, h_2 too, we have

$$\begin{aligned}
& \frac{1}{h_1^3 h_2} \int\int K_{11}\left(\frac{x_1-u_1}{h_1}, \frac{x_2-u_2}{h_2}\right) \lambda\left(u_1, u_2\right) \mathrm{d}u_1 \mathrm{d}u_2 \\
= & \frac{1}{h_1^2} \int\int K_{11}\left(u_1, u_2\right) \lambda\left(x_1-h_1 u_1, x_2-h_2 u_2\right) \mathrm{d}u_1 \mathrm{d}u_2 \\
= & \int\int K\left(u_1, u_2\right) \lambda_{11}\left(x_1-h_1 u_1, x_2-h_2 u_2\right) \mathrm{d}u_1 \mathrm{d}u_2 \\
= & \lambda_{11}\left(x_1, x_2\right)+\frac{1}{2} V_K \sum_{i=1}^2 h_i^2 \frac{\partial^2}{\partial x_i^2} \lambda_{11}\left(x_1, x_2\right)+O_p\left(h_1^{2+\beta}+h_2^{2+\beta}\right)
\end{aligned}$$

using substitution, integration by parts and a Taylor expansion. From Condition 2,

$$\frac{1}{h_1^3 h_2} \mu_{11}(H)=\lambda_{11}\left(x_1, x_2\right)+O\left(b_N^2\right)+o_p\left(N^{-\gamma} b_N^2\right).$$

b) We now consider

$$\hat{\lambda}_{11}\left(x_1, x_2, H\right)-\frac{1}{h_1^3 h_2} \mu_{11}(H)=\frac{1}{N h_1^3 h_2} \sum_{j=1}^N\left\{K_{11}\left(\frac{x_1-X_{j1}}{h_1}, \frac{x_2-X_{j2}}{h_2}\right)-\mu_{11}(H)\right\}.$$

To get rid of the technical problems caused by the randomness of h_1, h_2, we approximate them by deterministic bandwidths from an equidistant grid. For some $\tau>\frac{1}{2}$ to be chosen later, let $B_{N,\tau}$ be an equidistant grid in $\left[0, \frac{1}{2}\right]$ of width $N^{-\tau}$. Then, for any h_i satisfying Condition 1, there is a $\bar{h}_i \in B_{N,\tau}$ with $\left|h_i-\bar{h}_i\right| \leq N^{-\tau}$, $i=1,2$. Note that $\bar{h}_i$ is still random. By Assumption 8,

$$\frac{1}{N} \sum_{j=1}^N\left\{K_{11}\left(\frac{x_1-X_{j1}}{h_1}, \frac{x_2-X_{j2}}{h_2}\right)-\mu_{11}(H)\right\}=S_N\left(h_1, h_2\right)$$

is uniformly bounded. Moreover, from Condition 2,

$$\frac{1}{h_1^3 h_2}=\frac{1}{\beta_1^3 \beta_2 b_N^4\left(1+o_p\left(N^{-\gamma}\right)\right)^4}=\frac{1}{\beta_1^3 \beta_2 b_N^4}\left(1+o_p\left(N^{-\gamma}\right)\right)$$

using that the dominating term in $\left(1+o_p\left(N^{-\gamma}\right)\right)^4$ is $1+o_p\left(N^{-\gamma}\right)$ and a Taylor expansion for $\frac{1}{1+o_p\left(N^{-\gamma}\right)}=1+o_p\left(N^{-\gamma}\right)$. Hence,

$$\hat{\lambda}_{11}\left(x_1, x_2, H\right)=\frac{1}{h_1^3 h_2} \mu_{11}(H)+\frac{1+o_p\left(N^{-\gamma}\right)}{\beta_1^3 \beta_2 b_N^4} S_N\left(h_1, h_2\right).$$

So, we have to investigate $S_N(h_1, h_2)$. First, note that with $\bar{h}_i$, $i = 1, 2$, as above

$$|S_N(h_1, h_2)| \le \left|S_N\left(\bar{h}_1, \bar{h}_2\right)\right| + \left|S_N(h_1, h_2) - S_N\left(\bar{h}_1, \bar{h}_2\right)\right|.$$

The first term is bounded from above by

$$s_{N,1} = \sup\left\{|S_N(b_1, b_2)|; \frac{1}{2\sqrt{N}} - N^{-\tau} \le b_i \le \delta + N^{-\tau}, b_i \in B_{N,\tau}, i = 1, 2\right\}.$$

Note that $S_N(b_1, b_2)$ is a mean of, by Assumption 8, bounded, random variables as here $b_1, b_2 \in B_{N,\tau}$ are deterministic. Moreover, $\mathbb{E}S_N(b_1, b_2) = 0$ from the proof of Proposition 3 a). Let C denote an upper bound on $|K_{11}(u_1, u_2)|$. Then , the summands of $S_N(b_1, b_2)$ are bounded by $2C$, and their variance is bounded by $4C^2$. Let $\alpha_N > 0$, $\alpha_N \to 0$ for $N \to \infty$ such that $\sqrt{\frac{\log N}{N}} = O(\alpha_N)$. Then, from Bernstein's inequality, we have for $0 < \epsilon < 1$

$$\begin{aligned}
\text{pr}\left(\alpha_N\sqrt{\frac{N}{\log N}}|S_N(b_1, b_2)| > \epsilon\right) &= \text{pr}\left(|S_N(b_1, b_2)| > \frac{1}{\alpha_N}\sqrt{\frac{\log N}{N}}\epsilon\right) \\
&\le \exp\left\{\frac{-N\epsilon^2\frac{1}{\alpha_N^2}\frac{\log N}{N}}{\frac{4C}{3}\epsilon\frac{1}{\alpha_N}\sqrt{\frac{\log N}{N}} + 4C^2}\right\} \\
&\le \exp\left\{\frac{-\epsilon^2\frac{\log N}{\alpha_N^2}}{A}\right\}
\end{aligned}$$

for some suitable constant $A > 0$ and all large enough N, as, for $N \to \infty$, $\frac{1}{\alpha_N}\sqrt{\frac{\log N}{N}} = O(1)$. As $B_{N,\tau}$ is a finite set with less than $N^{2\tau}$ elements, we have with $B^*_{N,\tau} = B_{N,\tau} \cap \left[\frac{1}{2\sqrt{N}} - N^{-\tau}, \delta + N^{-\tau}\right] \subseteq B_{N,\tau}$

$$\begin{aligned}
\text{pr}\left(\alpha_N\sqrt{\frac{N}{\log N}}s_{N,1} > \epsilon\right) &\le \sum_{b_1, b_2 \in B^*_{N,\tau}} \text{pr}\left(\alpha_N\sqrt{\frac{N}{\log N}}|S_N(b_1, b_2)| > \epsilon\right) \\
&\le N^{2\tau}\exp\left\{-\frac{\epsilon^2}{A}\frac{\log N}{\alpha_N^2}\right\} \\
&= \exp\left\{\log N\left(2\tau - \frac{\epsilon^2}{A\alpha_N^2}\right)\right\} \to 0
\end{aligned}$$

for $N \to \infty$ as $\alpha_N \to 0$ and, hence, the factor of $\log N$ becomes negative for large enough N. Therefore, we have independent of the choice of τ

$$s_{N,1} = o_p\left(\frac{1}{\alpha_N}\sqrt{\frac{\log N}{N}}\right).$$

For the second term, we use from Assumption 9 with $|z_1|, |z_2| \leq 1$

$$\begin{aligned}\left|K_{11}\left(\frac{z_1}{b_1}, \frac{z_2}{b_2}\right) - K_{11}\left(\frac{z_1}{b_1'}, \frac{z_2}{b_2'}\right)\right| &\leq L_{11}\left\{\left|\frac{z_1}{b_1} - \frac{z_1}{b_1'}\right| + \left|\frac{z_2}{b_2} - \frac{z_2}{b_2'}\right|\right\} \\ &\leq L_{11}\left\{\frac{\left|b_1 - b_1'\right|}{b_1 b_1'} + \frac{\left|b_2 - b_2'\right|}{b_2 b_2'}\right\}.\end{aligned}$$

As $|z_i| \leq 1$ holds for $z_1 = x_1 - X_{j1}, z_2 = x_2 - X_{j2}$ and, in the integral defining $\mu_{11}, z_1 = x_1 - u_1, z_2 = x_2 - u_2$, and as $\lambda(u_1, u_2)$ integrates to 1, we get as $\left|h_i - \bar{h}_i\right| \leq N^{-\tau}$

$$\begin{aligned}&\left|S_N(h_1, h_2) - S_N\left(\bar{h}_1, \bar{h}_2\right)\right| \\ \leq\ & 2L_{11}\left\{\frac{1}{h_1\bar{h}_1} + \frac{1}{h_2\bar{h}_2}\right\} N^{-\tau} \\ =\ & 2L_{11}\left\{\frac{1}{h_1^2\left(1 + O\left(N^{-\tau}\right)\right)} + \frac{1}{h_2^2\left(1 + O\left(N^{-\tau}\right)\right)}\right\} N^{-\tau} \\ =\ & 2L_{11}\left\{\frac{1}{\beta_1^2} + \frac{1}{\beta_2^2}\right\} \frac{N^{-\tau}}{b_N^2\left(1 + o_p\left(N^{-\gamma}\right)\right)^2\left(1 + O\left(N^{-\tau}\right)\right)} \\ \leq\ & c\frac{1}{N^\tau b_N^2}\left(1 + o_p\left(N^{-\gamma}\right) + O\left(N^{-\tau}\right)\right)\end{aligned}$$

for some suitable constant $c > 0$, using Condition 2. Combining this with the bound on $\left|S_N\left(\bar{h}_1, \bar{h}_2\right)\right|$, we get

$$\begin{aligned}\left|S_N(h_1, h_2)\right| &\leq s_{N,1} + \left|S_N(h_1, h_2) - S_N\left(\bar{h}_1, \bar{h}_2\right)\right| \\ &= o_p\left(\frac{1}{\alpha_N}\sqrt{\frac{\log N}{N}}\right) + O\left(\frac{1}{N^\tau b_N^2}\right) + o_p\left(\frac{N^{-\gamma}}{N^\tau b_N^2}\right).\end{aligned}$$

From Condition 2, we then have

$$\begin{aligned}\frac{1}{h_1^3 h_2} S_N(h_1, h_2) &= \frac{S_N(h_1, h_2)}{\beta_1^3\beta_2\left(1 + o_p\left(N^{-\gamma}\right)\right)^4 b_N^4} = \frac{S_N(h_1, h_2)}{\beta_1^3\beta_2 b_N^4}\left(1 + o_p\left(N^{-\gamma}\right)\right) \\ &= O\left(\frac{1}{N^\tau b_N^6}\right) + o_p\left(\frac{N^{-\gamma}}{N^\tau b_N^6}\right) + o_p\left(\frac{1}{\alpha_N b_N^4}\sqrt{\frac{\log N}{N}}\right).\end{aligned}$$

As τ was arbitrary, we now choose $\tau = 4$. Then, as $\frac{1}{b_N\sqrt{N}}$ is bounded, we have

$$\frac{1}{N^{\tau} b_N^8} = \frac{1}{\left(b_N\sqrt{N}\right)^8} = O(1)$$

and

$$\frac{1}{N^{\tau} b_N^6} = O\left(b_N^2\right)$$

such that

$$\frac{1}{h_1^3 h_2} S_N(h_1, h_2) = o_p\left(\frac{1}{\alpha_N b_N^4}\sqrt{\frac{\log N}{N}}\right) + O\left(b_N^2\right) + o_p\left(N^{-\gamma} b_N^2\right).$$

c) Combining a) and b), we get finally

$$\begin{aligned}
\hat{\lambda}_{11}(x_1, x_2, H) &= \frac{1}{h_1^3 h_2}\left(\mu_{11}(H) + S_N(h_1, h_2)\right) \\
&= \lambda_{11}(x_1, x_2) + O\left(b_N^2\right) + o_p\left(N^{-\gamma} b_N^2\right) + o_p\left(\frac{1}{\alpha_N b_N^4}\sqrt{\frac{\log N}{N}}\right)
\end{aligned}$$

The same arguments analogously hold for $\hat{\lambda}_{22}(x_1, x_2, H)$.

□

The amse also depends on $\lambda(x_1, x_2)$ which for a plug-in method we also have to replace by an estimate with data-adaptive and, hence, random bandwidth. We have analogously to the last two results:

Proposition 5. *Let the assumptions of Proposition 4 be satisfied. Then*

$$\hat{\lambda}(x_1, x_2, H) = \lambda(x_1, x_2) + O\left(b_N^2\right) + o_p\left(N^{-\gamma} b_N^2\right) + o_p\left(\frac{1}{\alpha_N b_N^2}\sqrt{\frac{\log N}{N}}\right)$$

for any sequence $\alpha_N > 0$ with $\alpha_N \to 0$ and $\sqrt{\frac{\log N}{N}} = O(\alpha_N)$.

Corollary 9. *Under the conditions of Proposition 4,*

$$\hat{\lambda}(x_1, x_2, H) = \lambda(x_1, x_2) + O\left(b_N^2\right) + o_p\left(N^{-\gamma} b_N^2\right) + o_p\left(\frac{\log N}{\sqrt{N} b_N^2}\right).$$

Proof. Proof of Proposition 5: The proof is more or less identical to the proof of Proposition 4, and we only briefly discuss the differences.

a) Instead of $\mu_{11}(H)$, we consider

$$\mu(H) = \int\int K\left(\frac{x_1 - u_1}{h_1}, \frac{x_2 - u_2}{h_2}\right)\lambda(u_1, u_2)\,\mathrm{d}u_1\mathrm{d}u_2.$$

Instead of Proposition 3, we have to refer to Theorem 2 for the bias and variance expansion for deterministic bandwidth, and we do not need the integration-by-parts argument to finally get

$$\frac{1}{h_1 h_2}\mu(H) = \lambda(x_1, x_2) + O\left(b_N^2\right) + o_p\left(N^{-\gamma} b_N^2\right).$$

b) We decompose $\hat{\lambda}(x_1, x_2)$ into

$$\begin{aligned}\hat{\lambda}(x_1, x_2) &= \frac{1}{h_1 h_2}\mu(H) + \frac{1}{N h_1 h_2}\sum_{j=1}^{N}\left\{K\left(\frac{x_1 - X_{j1}}{h_1}, \frac{x_2 - X_{j2}}{h_2}\right) - \mu(H)\right\} \\ &= \frac{1}{h_1 h_2}\mu(H) + \frac{1}{h_1 h_2} S_N(h_1, h_2).\end{aligned}$$

K is uniformly bounded and Lipschitz continuous as K_{11}, such that $S_N(b_1, b_2)$ is a mean of bounded random variables again. We get by exactly the same arguments as in the proof of Proposition 4 b)

$$\frac{1}{h_1 h_2} S_N(h_1, h_2) = o_p\left(\frac{1}{\alpha_N b_N^2}\sqrt{\frac{\log N}{N}}\right) + O\left(b_N^2\right) + o_p\left(N^{-\gamma} b_N^2\right),$$

where, in the last step, we may choose $\tau = 3$ instead of $\tau = 4$.

c) The final result follows again from combining a) and b).

□

Chapter 4

Asymptotics for integrated mean-squared error estimates with random bandwidths

For estimating the integrated mean-squared error as a function of the bandwidths, let us recall from Corollary 3 that the asymptotic approximation is

$$\text{amise}\,(H) = \frac{Q_K}{N \det H} + \frac{V_K^2}{4} \int\int \left(\sum_{i=1}^{2} h_i^2 \lambda_{ii}\,(x_1, x_2) \right)^2 \mathrm{d}x_1 \mathrm{d}x_2.$$

Hence, we have to investigate

$$h_i^4 \int\int \hat{\lambda}_{ii}^2\,(x_1, x_2, H)\,\mathrm{d}x_1 \mathrm{d}x_2$$

and

$$h_1^2 h_2^2 \int\int \hat{\lambda}_{11}\,(x_1, x_2, H)\,\hat{\lambda}_{22}\,(x_1, x_2, H)\,\mathrm{d}x_1 \mathrm{d}x_2.$$

First note that in calculating $\int\int \hat{\lambda}_{ii}^2 (x_1, x_2, H)\, \mathrm{d}x_1 \mathrm{d}x_2$, we can avoid integration by using, e.g.,

$$\begin{aligned}
&\int\int \hat{\lambda}_{11}^2 (x_1, x_2, H)\, \mathrm{d}x_1 \mathrm{d}x_2 \\
=\ & \frac{1}{N^2 h_1^6 h_2^2} \sum_{i,j=1}^{N} \int\int K_{11}\left(\frac{x_1 - X_{j1}}{h_1}, \frac{x_2 - X_{j2}}{h_2}\right) K_{11}\left(\frac{x_1 - X_{i1}}{h_1}, \frac{x_2 - X_{i2}}{h_2}\right) \mathrm{d}x_1 \mathrm{d}x_2 \\
=\ & \frac{1}{N^2 h_1^5 h_2} \sum_{i,j=1}^{N} \int\int K_{11}(u, v)\, K_{11}\left(u + \frac{X_{j1} - X_{i1}}{h_1}, v + \frac{X_{j2} - X_{i2}}{h_2}\right) \mathrm{d}u \mathrm{d}v \\
=\ & \frac{1}{N^2 h_1^5 h_2} \sum_{i,j=1}^{N} L_{11}\left(\frac{X_{j1} - X_{i1}}{h_1}, \frac{X_{j2} - X_{i2}}{h_2}\right)
\end{aligned}$$

by substitution and denoting by $L_{11}(x, y) = K_{11} * K_{11}(x, y)$ the convolution of K_{11} with itself which may be calculated in advance:

$$L_{11}(x, y) = \int\int K_{11}(u, v)\, K_{11}(x - u, y - v)\, \mathrm{d}u \mathrm{d}v.$$

The same argument holds for the integral of $\hat{\lambda}_{22}^2$ with $L_{22} = K_{22} * K_{22}$ and for the integral of $\hat{\lambda}_{11}\hat{\lambda}_{22}$ with $L_{12} = K_{11} * K_{22}$.

4.1 The one-dimensional case

We start with investigating one-dimensional kernel density estimates as this is more suitable for a first exposition of the ideas. In higher dimensions the notation becomes more involved.

We prove an analogous result to Proposition 1 of Engel et al. (1994). As the proof of Engel et al. (1994) is very sketchy, we give our own proof using sometimes different arguments, in particular using Hoeffding's exponential inequality for means of bounded random variables and some properties of U-statistics. We could not follow every detail of the proof of Engel et al. (1994), e.g. they refer to Lemma 3.1 of Hall and Marron (1987) to get bound on a doubly indexed sum with full range for the two indices, but Hall and Marron (1987) only consider a summation over $i \neq j$, i.e. omitting the diagonal, and this is crucial for the derivation of their results (compare our Lemma 5, where we prove a version of that part of Hall and Marron's lemma which we need). Our modified proof leads to slightly different rates of the remainder term which, however, is not relevant for the application of the proposition.

First let us introduce notation and assumptions. $X_1, \ldots, X_N$ are i.i.d. with values in $[0,1]$, having density $\lambda(x)$. The kernel density estimate is

$$\hat{\lambda}(x,h) = \frac{1}{Nh} \sum_{j=1}^{N} K\left(\frac{x - X_j}{h}\right)$$

and the corresponding estimate of the second derivative $\lambda''(x)$ is

$$\hat{\lambda}''(x,h) = \frac{1}{Nh^3} \sum_{j=1}^{N} K_2\left(\frac{x - X_j}{h}\right)$$

with $K_2(u) = K''(u)$. We want to derive an asymptotic expansion of

$$\int \left(\hat{\lambda}''(x,h)\right)^2 \mathrm{d}x$$

for random, in particular for data-dependent, bandwidth h. We make the following assumptions:

Assumption 10. *λ is 4 times continuously differentiable on $[0,1]$, and the 4th derivative $\lambda^{(4)}$ is Hölder continuous with some exponent $\beta > 0$.*

Assumption 11. *$K(u)$ is a kernel function with support $[-1,+1]$ which is non-negative, twice continuously differentiable with Lipschitz continuous second derivative K'' and satisfies*

$$K(\pm 1) = K'(\pm 1) = 0.$$

As notation, we use $K_2 = K''$ and $V_K = \int u^2 K(u)\,\mathrm{d}u$.

Proposition 6. *Let h be a sequence of random bandwidths which can be approximated by a sequence b_N of deterministic bandwidths such that*

$$h = b_N \left(1 + o_p \left(N^{-\gamma}\right)\right)$$

for some $\gamma \geq 0$. Then,

$$\begin{aligned} & \int \left(\hat{\lambda}'' (x, h)\right)^2 \mathrm{d}x \\ = & \int \left(\lambda'' (x)\right)^2 \mathrm{d}x + V_K \int \lambda'' (x) \lambda^{(4)} (x) \mathrm{d}x b_N^2 + \frac{1}{N b_N^5} \int K_2^2 (u) \mathrm{d}u + R_N \end{aligned}$$

with remainder term

$$R_N = o\left(b_N^2\right) + o_p\left(N^{-\gamma} b_N^2 + \frac{\log N}{\sqrt{N}}\right) + O_p\left(\frac{1}{N b_N^4}\right) + o_p\left(\frac{\log N}{N^{\frac{3}{2}} b_N^5}\right) + o_p\left(\frac{N^{-\frac{\gamma}{2}}}{N b_N^5}\right).$$

Before starting with the proof let us remark that this expansion coincides with some slight differences in the o- and o_p- terms, due to the techniques of proof, with Proposition 1 of Engel et al. (1994). Note that in their formulation the last term of R_N is missing, but it can be inferred from (1.3) of their proof and is of order $o_p\left(\frac{1}{N b_N^5}\right)$.

Proof. We introduce

$$\nu (x, h) = \frac{1}{h^2} \int K_2 (u) \lambda (x - hu) \mathrm{d}u = \frac{1}{h^2} \mu (x, h).$$

Note that for deterministic b, $\nu(x, b) = \mathbb{E}\hat{\lambda}''(x, b)$ which immediately follows from substitution. $\mu(x, h)$ corresponds to $\mu_{11}(H)$ in the proof of Proposition 4. Now, we decompose

$$\int \left(\hat{\lambda}'' (x, h)\right)^2 \mathrm{d}x = E(h) + 2M(h) + V(h)$$

with

$$\begin{aligned} E(h) &= \int \nu^2 (x, h) \mathrm{d}x \\ M(h) &= \int \nu (x, h) \left\{\hat{\lambda}'' (x, h) - \nu (x, h)\right\} \mathrm{d}x \\ V(h) &= \int \left(\hat{\lambda}'' (x, h) - \nu (x, h)\right)^2 \mathrm{d}x \end{aligned}$$

a) We first derive an asymptotic expansion for $\nu(x,h)$. Integration by parts implies, using Assumption 10,

$$\begin{aligned}\nu(x,h) &= \frac{1}{h}\int K'(u)\,\lambda'(x-hu)\,\mathrm{d}u = \int K(u)\,\lambda''(x-hu)\,\mathrm{d}u \\ &= \int K(u)\left\{\lambda''(x) - hu\lambda'''(x) + \frac{1}{2}u^2h^2\lambda^{(4)}(x) + O\left(h^{2+\beta}\right)\right\}\mathrm{d}u \\ &= \lambda''(x) + \frac{1}{2}V_K\lambda^{(4)}(x)\,h^2 + O\left(h^{2+\beta}\right)\end{aligned}$$

by Taylor expansion and using Assumption 11 on K. Note that from the first line and boundedness of K, λ'' we have that $\nu(x,h)$ is uniformly bounded in x and h. If we approximate the random h by the deterministic b_N from our assumptions, we have

$$\begin{aligned}\nu(x,h) - \nu(x,b_N) &= \frac{1}{2}V_K\lambda^{(4)}(x)\left(h^2 - b_N^2\right) + O\left(\left|h^2 - b_N^2\right|\cdot\max(h,b_N)^\beta\right) \\ &= \frac{1}{2}V_K\lambda^{(4)}(x)\cdot(h+b_N)(h-b_N) + O\left(\left|h^2 - b_N^2\right|\cdot\max(h,b_N)^\beta\right).\end{aligned}$$

As $\lambda^{(4)}(x)$ is uniformly bounded, $\max(h,b_N) = \max(b_N(1+o_p(N^{-\gamma})), b_N) = b_N(1+o_p(N^{-\gamma}))$, and $h + b_N \le 2\max(h,b_N)$, $|h-b_N| = b_N\cdot o_p(N^{-\gamma})$, we get

$$|\nu(x,h) - \nu(x,b_N)| = o_p\left(b_N^2N^{-\gamma}\right).$$

b) As the next step, we look at $E(h)$. Using boundedness of $\nu(x,b)$ and a)

$$\begin{aligned}E(h) &= \int(\nu(x,h) - \nu(x,b_N) + \nu(x,b_N))^2\,\mathrm{d}x \\ &= \int\nu^2(x,b_N)\,\mathrm{d}x + o_p\left(b_N^2N^{-\gamma}\right) \\ &= \int\left(\lambda''(x) + \frac{1}{2}V_K\lambda^{(4)}(x)\,b_N^2 + o\left(b_N^2\right)\right)^2\mathrm{d}x + o_p\left(b_N^2N^{-\gamma}\right) \\ &= \int(\lambda''(x))^2\,\mathrm{d}x + V_K\int\lambda''(x)\,\lambda^{(4)}(x)\,\mathrm{d}x\cdot b_N^2 + o\left(b_N^2\right) + o_p\left(b_N^2N^{-\gamma}\right),\end{aligned}$$

i.e. $E(h)$ is the dominant term of $\int\left(\hat{\lambda}''(x,h)\right)^2\mathrm{d}x$ as an estimate of $\int(\lambda''(x))^2\,\mathrm{d}x$.

c) For any deterministic b, we have, using $\nu(x,b) = \mathbb{E}\hat{\lambda}''(x,b)$

$$\mathbb{E}V(b) = \int \mathrm{var}\hat{\lambda}''(x,b)\,\mathrm{d}x = \frac{1}{Nb^6}\int \mathrm{var}K_2\left(\frac{x-X_1}{b}\right)\mathrm{d}x$$

as $\hat{\lambda}''(x,b)$ is a sum of i.i.d. random variables. Hence using that $\mathbb{E}K_2\left(\frac{x-X_1}{b}\right) = b^3\mathbb{E}\hat{\lambda}''(x,b) = b^3\nu(x,b)$ and that ν is bounded

$$\begin{aligned}
\mathbb{E}V(b) &= \frac{1}{Nb^6}\int \mathbb{E}K_2^2\left(\frac{x-X_1}{b}\right)\mathrm{d}x - \frac{1}{N}\int \nu^2(x,b)\,\mathrm{d}x \\
&= \frac{1}{Nb^6}\int\int K_2^2\left(\frac{x-u}{b}\right)\lambda(u)\,\mathrm{d}u\mathrm{d}x + O\left(\frac{1}{N}\right) \\
&= \frac{1}{Nb^5}\int\int K_2^2(v)\,\lambda(x-bv)\,\mathrm{d}x\mathrm{d}v + O\left(\frac{1}{N}\right) \\
&= \frac{1}{Nb^5}\int\int K_2^2(v)\left(\lambda(x) - bv\lambda'(x) + o(b)\right)\mathrm{d}x\mathrm{d}v + O\left(\frac{1}{N}\right) \\
&= \frac{1}{Nb^5}\int K_2^2(v)\,\mathrm{d}v + O\left(\frac{1}{Nb^4}\right)
\end{aligned}$$

by a Taylor expansion, using the differentiability conditions on λ and that it integrates to 1.

d) Now, we consider the remainder term $V(b) - \mathbb{E}V(b)$. Note that

$$\begin{aligned}
V(b) &= \frac{1}{N^2b^6}\int\sum_{i,j=1}^{N}\left(K_2\left(\frac{x-X_i}{b}\right) - b^3\nu(x,b)\right)\left(K_2\left(\frac{x-X_j}{b}\right) - b^3\nu(x,b)\right)\mathrm{d}x \\
&= \frac{1}{Nb^5}W_N(b) + \frac{N-1}{2Nb^6}U_N(b),
\end{aligned}$$

where

$$W_N(b) = \frac{1}{Nb}\sum_{j=1}^{N}\int\left(K_2\left(\frac{x-X_j}{b}\right) - b^3\nu(x,b)\right)^2\mathrm{d}x = \frac{1}{N}\sum_{j=1}^{N}Q_b(X_j)$$

is a mean of i.i.d. random variables, and

$$\begin{aligned}
&U_N(b) \\
&= \frac{2}{N(N-1)}\sum_{i\neq j}\int\left(K_2\left(\frac{x-X_i}{b}\right) - b^3\nu(x,b)\right)\left(K_2\left(\frac{x-X_j}{b}\right) - b^3\nu(x,b)\right)\mathrm{d}x \\
&= \frac{2}{N(N-1)}\sum_{i\neq j}R_b(X_i,X_j)
\end{aligned}$$

is a U-statistic with symmetric kernel $R_b(x,z)$ (compare Chapter 5 of Serfling (1980)). First, we consider $W_N(b)$. Substituting $u = \frac{x-X_j}{b}$, we get

$$Q_b(X_j) = \int \left(K_2(u) - b^3\nu(X_j + bu, b)\right)^2 \mathrm{d}u \leq C, 1 \leq j \leq N,$$

due to boundedness of K_2 and ν. Applying Hoeffding's inequality (compare the lemma in Section 2.3.2 of Serfling (1980) for the one-sided version), we get for any sequence $a_N > 0$

$$\mathrm{pr}\left(a_N \left|W_N(b) - \mathbb{E}W_N(b)\right| \geq \epsilon\right) \leq 2 \cdot \exp\left(-\frac{2N\epsilon^2}{a_N^2 4C^2}\right) \to 0$$

if $\frac{N}{a_N^2} \to \infty$, and, then, $W_N(b) - \mathbb{E}W_N(b) = o_p\left(\frac{1}{a_N}\right)$. In particular, for $a_N = \frac{\sqrt{N}}{\log N}$, we get

$$\frac{1}{Nb^5}\left(W_N(b) - \mathbb{E}W_N(b)\right) = o_p\left(\frac{\log N}{N^{\frac{3}{2}}b^5}\right).$$

For the second component $U_N(b)$, we decompose as in the proof of Lemma 5, a),

$$R_b(X_1, X_2) = b\left\{L_2\left(\frac{X_1 - X_2}{b}\right) - \mathbb{E}L_2\left(\frac{X_1 - X_2}{b}\right)\right\} - R_b(X_1) - R_b(X_2)$$

with $L_2(z) = K_2 * K_2(z)$ and

$$\begin{aligned} R_b(u) &= \int b^3\nu(x,b)\left\{K_2\left(\frac{x-u}{b}\right) - b^3\nu(x,b)\right\}\mathrm{d}x \\ &= \int b^4\nu(u + bz, b)\, K_2(z)\,\mathrm{d}z + O\left(b^6\right) \\ &= \int b^4\lambda''(u + bz)\, K_2(z)\,\mathrm{d}z + O\left(b^6\right) \end{aligned}$$

substituting $z = \frac{x-u}{b}$, using boundedness of $\nu(x,b)$ and the expansion of $\nu(x,b)$ from a), where the latter implies $\nu(x,b) = \lambda''(x) + O(b^2)$ uniformly in x. Using a Taylor expansion of λ'' and uniform boundedness of $\lambda^{(4)}$

$$R_b(u) = b^4 \int K_2(z)\left\{\lambda''(u) + bz\lambda'''(u) + O\left(b^2\right)\right\}\mathrm{d}z + O\left(b^6\right) = O\left(b^6\right)$$

as, from Assumption 11, $\int K_2(z)\,\mathrm{d}z = 0 = \int zK_2(z)\,\mathrm{d}z$. Hence, $\frac{1}{b^6}R_b(u)$ is bounded, and we also have $\mathbb{E}R_b(X_1) = 0$. Using again Hoeffding's inequality, we have with $\left|\frac{1}{b^6}R_b(u)\right| \leq c$ and $a_N > 0$

$$\mathrm{pr}\left(a_N\left|\frac{1}{Nb^6}\sum_{j=1}^{N}R_b(X_j)\right| > \epsilon\right) \leq 2\cdot\exp\left\{-\frac{2N\epsilon^2}{a_N^2 4c^2}\right\}$$

and with $a_N = \frac{\sqrt{N}}{\log N}$

$$\frac{1}{Nb^6}\sum_{j=1}^{N}R_b(X_j) = o_p\left(\frac{\log N}{\sqrt{N}}\right).$$

Hence, we have

$$\frac{1}{b^6}U_N(b) = \frac{1}{b^6}\tilde{U}_N(b) + o_p\left(\frac{\log N}{\sqrt{N}}\right),$$

where $\tilde{U}_N(b)$ is also a U-statistic with kernel

$$\Lambda_b(x-z) = b\left\{L_2\left(\frac{x-z}{b}\right) - \mathbb{E}L_2\left(\frac{X_1-X_2}{b}\right)\right\}.$$

We set $\ell_b(x) = \mathbb{E}\Lambda_b(x-X_2)$, $\varsigma_1 = \mathrm{var}\ell_b(X_1)$, $\varsigma_2 = \mathrm{var}\Lambda_b(X_1-X_2)$. From Lemma A in Section 5.2.1 of Serfling (1980), we have

$$\mathrm{var}\tilde{U}_N(b) = \frac{4(N-2)}{N(N-1)}\varsigma_1 + \frac{2}{N(N-1)}\varsigma_2.$$

By the same argument as in the proof of Lemma 3 b),

$$b\int L_2\left(\frac{x-z}{b}\right)\lambda(z)\,\mathrm{d}z = b^2\int L_2(u)\,\lambda(x+bu)\,\mathrm{d}u = O\left(b^6\right)$$

uniformly in x, i.e. $|\ell_b(X_1)|$ is a random variable bounded by $c\cdot b^6$ for some $c > 0$, which implies $\varsigma_1 = O(b^{12})$. From Lemma 3, b) and c), we have

$$\varsigma_2 = b^2\mathrm{var}L_2\left(\frac{X_1-X_2}{b}\right) \leq b^2\mathbb{E}L_2^2\left(\frac{X_1-X_2}{b}\right) = O\left(b^4\right)$$

such that we conclude

$$\mathrm{var}\tilde{U}_N(b) = O\left(\frac{b^{12}}{N}\right) + O\left(\frac{b^4}{N^2}\right),$$

and, therefore, by Chebyshev's inequality

$$\frac{1}{b^6}\tilde{U}_N(b) = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{Nb^4}\right)$$

and, correspondingly,

$$\frac{1}{b^6}U_N(b) = O_p\left(\frac{1}{Nb^4}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right).$$

Together with the bound on $W_N(b) - \mathbb{E}W_N(b)$, we finally conclude

$$V(b) - \mathbb{E}V(b) = O_p\left(\frac{1}{Nb^4}\right) + o_p\left(\frac{\log N}{N^{\frac{3}{2}}b^5}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right).$$

e) For general random bandwidth h, we decompose $V(h)$ into $V(h) = \bar{V}(h) + \left(V(h) - \bar{V}(h)\right)$, where

$$\begin{aligned}\bar{V}(h) &= \frac{1}{Nh^6}\int\int K_2^2\left(\frac{x-u}{h}\right)\lambda(u)\,\mathrm{d}u\mathrm{d}x - \frac{1}{N}\int \nu^2(x,h)\,\mathrm{d}x \\ &= \frac{1}{Nh^5}\int\int K_2^2(v)\,\lambda(x-hv)\,\mathrm{d}x\mathrm{d}v - \frac{1}{N}\int \nu^2(x,h)\,\mathrm{d}x.\end{aligned}$$

Note, that for deterministic $h = b$, we have $\bar{V}(b) = \mathbb{E}V(b)$. Using the same expansion as in c), we also have for random h

$$\bar{V}(h) = \frac{1}{Nh^5}\int K_2^2(v)\,\mathrm{d}v + O_p\left(\frac{1}{Nh^4}\right).$$

As h satisfies $h = b_N(1 + o_p(N^{-\gamma}))$, we have for $m \geq 1$

$$\begin{aligned}h^m &= b_N^m\left(1 + o_p\left(N^{-\gamma}\right)\right)^m = b_N^m\left(1 + o_p\left(N^{-\gamma}\right)\right) \\ \frac{1}{h^m} &= \frac{1}{b_N^m\left(1 + o_p\left(N^{-\gamma}\right)\right)} = \frac{1 + o_p\left(N^{-\gamma}\right)}{b_N^m}\end{aligned}$$

using a Taylor expansion of $\frac{1}{1+z}$ for the last argument. We conclude

$$\begin{aligned}\bar{V}(h) &= \frac{1}{Nb_N^5}\int K_2^2(v)\,\mathrm{d}v\left(1 + o_p\left(N^{-\gamma}\right)\right) + O_p\left(\frac{1}{Nb_N^4}\right) + o_p\left(\frac{N^{-\gamma}}{Nb_N^4}\right) \\ &= \frac{1}{Nb_N^5}\int K_2^2(v)\,\mathrm{d}v + o_p\left(\frac{N^{-\gamma}}{Nb_N^5}\right) + O_p\left(\frac{1}{Nb_N^4}\right).\end{aligned}$$

It remains to study $V_0(h) = V(h) - \bar{V}(h)$. We decompose it into

$$V_0(h) = V_0(b_N) + (V_0(h) - V_0(b_N)),$$

where, from d),

$$V_0(b_N) = V(b_N) - \mathbb{E}V(b_N) = O_p\left(\frac{1}{Nb_N^4}\right) + o_p\left(\frac{\log N}{N^{\frac{3}{2}}b_N^5}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right).$$

We split

$$V_0(h) - V_0(b_N) = (V_0(h) - V_0(b_N))\left(\mathbf{1}_{\{|h-b_N| \geq b_N N^{-\gamma}\}} + \mathbf{1}_{\{|h-b_N| < b_N N^{-\gamma}\}}\right).$$

As $|h - b_N| = b_N \triangle_N$ with $\triangle_N = o_p(N^{-\gamma})$, we have for all $\epsilon > 0$, $a_N > 0$

$$\begin{aligned} \text{pr}\left(a_N |V_0(h) - V_0(b_N)| \mathbf{1}_{\{|h-b_N| \geq b_N N^{-\gamma}\}} > \epsilon\right) &\leq \text{pr}\left(\mathbf{1}_{\{|h-b_N| \geq b_N N^{-\gamma}\}} = 1\right) \\ &= \text{pr}\left(b_N \triangle_N \geq b_N N^{-\gamma}\right) \\ &= \text{pr}\left(N^{\gamma} \triangle_N \geq 1\right) \to 0, \end{aligned}$$

as $N^{\gamma} \triangle_N = o_p(1)$. Hence,

$$(V_0(h) - V_0(b_N)) \mathbf{1}_{\{|h-b_N| \geq b_N N^{-\gamma}\}} = o_p\left(\frac{1}{a_N}\right).$$

As in the proof of Proposition 4, b), we approximate h by $\bar{h} \in B_{N,\tau}$, where $B_{N,\tau}$ is a grid of finitely many points which are $N^{-\tau}$ apart for some suitably large $\tau > 0$. Let

$$B_{N,\tau}^{\gamma} = B_{N,\tau} \cap \left(b_N - b_N N^{-\gamma} - N^{-\tau}, b_N + b_N N^{-\gamma} + N^{-\tau}\right).$$

Note that $B_{N,\tau}^{\gamma}$ is part of an interval of length bounded by $c_B b_N$ for some constant $c_B > 0$ and large enough N, such that the number of points in $B_{N,\tau}^{\gamma}$ satisfies $\left|B_{N,\tau}^{\gamma}\right| \leq c_B N^{\tau} b_N$.

If $|h-b_N| < b_N N^{-\gamma}$ and as $\left|h-\bar{h}\right| \leq N^{-\tau}$, we have $\bar{h} \in B^{\gamma}_{N,\tau}$, and

$$\begin{aligned}
& \left|V_0(h) - V_0(b_N)\right| \mathbf{1}_{\{|h-b_N|<b_N N^{-\gamma}\}} \\
\leq\ & \left|V_0(h) - V_0\left(\bar{h}\right)\right| \mathbf{1}_{\{|h-b_N|<b_N N^{-\gamma}\}} + \sup_{b \in B^{\gamma}_{N,\tau}} \left|V_0(b) - V_0(b_N)\right| \\
\leq\ & \sup\left\{\left|V_0(b) - V_0\left(\bar{b}\right)\right|;\right. \\
& \left. b, \bar{b} \in \left[b_N - b_N N^{-\gamma} - N^{-\tau}, b_N + b_N N^{-\gamma} + N^{-\tau}\right], \left|b - \bar{b}\right| \leq N^{-\tau}\right\} \\
& + \sup_{b \in B^{\gamma}_{N,\tau}} \left|V_0(b) - V_0(b_N)\right| \\
=\ & s_1 + s_2.
\end{aligned}$$

For τ, N large enough,

$$\begin{aligned}
s_1 &\leq \sup\left\{\left|V_0(b) - V_0\left(\bar{b}\right)\right|; b, \bar{b} \geq b_N\left(1 - 2N^{-\gamma}\right), \left|b - \bar{b}\right| \leq N^{-\tau}\right\} \\
&\leq 2L \cdot \frac{N^{-\tau}}{b_N^8\left(1 - 2N^{-\gamma}\right)^8} = \frac{2L}{N^{\tau} b_N^8}\left(1 + O\left(N^{-\gamma}\right)\right)
\end{aligned}$$

using $\left|V_0(b) - V_0\left(\bar{b}\right)\right| \leq \left|V(b) - V\left(\bar{b}\right)\right| + \mathbb{E}\left|V(b) - V\left(\bar{b}\right)\right|$ and the Lipschitz property of $V(b)$ from Lemma 4 below.

For getting a bound on s_2, we decompose as in d)

$$V(b) = \frac{1}{Nb^5} W_N(b) + \frac{N-1}{2Nb^6} U_N(b)$$

and, as $\mathbb{E}U_N(b) = 0$,

$$\begin{aligned}
V_0(b) &= \frac{1}{Nb^5}\left(W_N(b) - \mathbb{E}W_N(b)\right) + \frac{N-1}{2Nb^6} U_N(b) \\
&= \frac{1}{Nb^5} W_{N,0}(b) + \frac{N-1}{2Nb^6} U_N(b).
\end{aligned}$$

Using $\left|B^{\gamma}_{N,\tau}\right| \leq c_B N^{\tau} b_N$ and again Hoeffding's inequality as in d), we get with $a_N = \frac{\sqrt{N}}{\log N}$

$$\begin{aligned}
\mathrm{pr}\left(a_N \sup_{b \in B^{\gamma}_{N,\tau}} \left|W_{N,0}(b)\right| \geq \epsilon\right) &\leq c_B N^{\tau} b_N \sup_{b \in B^{\gamma}_{N,\tau}} \mathrm{pr}\left(a_N \left|W_{N,0}(b)\right| \geq \epsilon\right) \\
&\leq 2c_B N^{\tau} b_N \exp\left(-\frac{2N\epsilon^2}{a_N^2 4c^2}\right) \\
&= 2c_B b_N \exp\left(\tau \log N - \frac{\epsilon^2}{2c^2}\left(\log N\right)^2\right) \to 0
\end{aligned}$$

for $N \to \infty$ such that

$$\sup_{b \in B_{N,\tau}^{\gamma}} |W_{N,0}(b)| = o_p\left(\frac{\log N}{\sqrt{N}}\right)$$

and the first component of s_2 is

$$\begin{aligned} & \sup_{b \in B_{N,\tau}^{\gamma}} \left| \frac{1}{Nb^5} W_{N,0}(b) - \frac{1}{Nb_N^5} W_{N,0}(b_N) \right| \\ \leq & \sup_{b \in B_{N,\tau}^{\gamma}} \frac{1}{Nb^5} |W_{N,0}(b)| + \frac{1}{Nb_N^5} |W_{N,0}(b_N)| \\ \leq & \; o_p\left(\frac{\log N}{N^{\frac{3}{2}} b_N^5}\right) \end{aligned}$$

as, for $b \in B_{N,\tau}^{\gamma}$, $b \geq b_N (1 - N^{-\gamma}) - N^{-\tau}$ and, hence, for large enough τ,

$$\begin{aligned} \frac{1}{b^5} & \leq \frac{1}{\left(b_N (1 - N^{-\gamma}) - N^{-\tau}\right)^5} \\ & = \frac{1}{b_N^5} (1 + o(1)) . \end{aligned}$$

Finally, we have to study

$$\begin{aligned} & \frac{N-1}{2N} \sup_{b \in B_{N,\tau}^{\gamma}} \left| \frac{1}{b^6} U_N(b) - \frac{1}{b_N^6} U_N(b_N) \right| \\ \leq & \frac{N-1}{2N} \sup_{b \in B_{N,\tau}^{\gamma}} \frac{1}{b^6} |U_N(b) - U_N(b_N)| + \frac{N-1}{2N} \sup_{b \in B_{N,\tau}^{\gamma}} \left| \frac{1}{b^6} - \frac{1}{b_N^6} \right| |U_N(b_N)| \\ = & \frac{(N-1)}{2N} \frac{(1 + O(N^{-\gamma}))}{b_N^6} \sup_{b \in B_{N,\tau}^{\gamma}} |U_N(b) - U_N(b_N)| \\ & + \frac{(N-1)}{2N} \frac{(1 + O(N^{-\gamma}))}{b_N^6} |U_N(b_N)| \end{aligned}$$

as, for τ, N large enough, $b_N (1 - 2N^{-\gamma}) \leq b \leq b_N (1 + 2N^{-\gamma})$ for all $b \in B_{N,\tau}^{\gamma}$.

From d), the second term satisfies

$$\frac{(N-1)}{2N} \frac{(1 + O(N^{-\gamma}))}{b_N^6} |U_N(b_N)| = O_p\left(\frac{1}{Nb_N^4}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right).$$

For the first term, we use from Lemma 5 with $q \geq 1$, $a_N = b_N\left(1+2N^{-\gamma}\right)$, $\alpha_N = b_N\left(1-2N^{-\gamma}\right) = a_N\left(1-\frac{4N^{-\gamma}}{1+2N^{-\gamma}}\right) = a_N\left(1-O\left(N^{-\gamma}\right)\right)$

$$\begin{aligned}\mathbb{E}\left(U_N(b)-U_N\left(\bar{b}\right)\right)^{2q} &\leq c\left(\frac{b_N}{N}\right)^{2q}\left(1+2N^{-\gamma}\right)^{2q}N^{-2q\gamma}\\ &\leq \bar{c}\left(\frac{b_N}{N}\right)^{2q}N^{-2q\gamma}\end{aligned}$$

for some suitable constant $\bar{c}$, not depending on N, and all $b, \bar{b} \in [\alpha_N, a_N]$.

Following Engel et al. (1994), we choose an arbitrary $\rho > 0$, to get

$$\begin{aligned}&\mathrm{pr}\left(N^{1+\gamma}b_N^5\frac{1}{b_N^6}\sup_{b\in B_{N,\tau}^{\gamma}}\left|U_N(b)-U_N(b_N)\right| \geq \epsilon N^{\rho}\right)\\ \leq\ & c_B N^{\tau}b_N\sup_{b\in B_{N,\tau}^{\gamma}}\mathrm{pr}\left(\left(\frac{N^{1+\gamma}}{b_N}\right)^{2q}\left|U_N(b)-U_N(b_N)\right|^{2q}\geq \epsilon^{2q}N^{2q\rho}\right)\\ \leq\ & c_B N^{\tau}b_N N^{-2q\rho}\frac{1}{\epsilon^{2q}}\sup_{b\in B_{N,\tau}^{\gamma}}\mathbb{E}\left(\frac{N^{1+\gamma}}{b_N}\left|U_N(b)-U_N(b_N)\right|\right)^{2q}\\ \leq\ & \frac{c_B\bar{c}}{\epsilon^{2q}}N^{\tau-2q\rho}b_N \to 0\end{aligned}$$

if q is chosen such that $2q\rho \geq \tau$. For the second inequality, we have used Markov's inequality, and for the third one the bound derived from Lemma 5. We conclude for arbitrarily small $\rho > 0$

$$\frac{1}{b_N^6}\sup_{b\in B_{N,\tau}^{\gamma}}\left|U_N(b)-U_N(b_N)\right| = o_p\left(\frac{N^{\rho}}{N^{1+\gamma}b_N^5}\right).$$

In particular, this term is of order $o_p\left(\frac{N^{-\frac{\gamma}{2}}}{Nb_N^5}\right)$ for $\rho \leq \frac{\gamma}{2}$.

Together with bound on $\frac{1}{Nb_N^5}\left|W_{N,0}(b)-W_{N,0}(b_N)\right|$ and on $\frac{1}{b_N^6}\left|U_N(b_N)\right|$, we finally conclude

$$\begin{aligned}s_2 &= \sup_{b\in B_{N,\tau}^{\gamma}}\left|V_0(b)-V_0(b_N)\right|\\ &= \frac{1}{Nb_N^5}\left\{o_p\left(\frac{\log N}{\sqrt{N}}\right)+o_p\left(N^{-\frac{\gamma}{2}}\right)+O_p(b_N)\right\}+o_p\left(\frac{\log N}{\sqrt{N}}\right)\end{aligned}$$

and we get the same asymptotic rate for $V_0(h) - V_0(b_N)$ as, for suitably large τ, s_1 is asymptotically negligible compared to s_2, and

$$(V_0(h) - V_0(b_N))\,\mathbf{1}_{\{|h-b_N|\geq b_N N^{-\gamma}\}}$$

is asymptotically negligible too.

Together with the expansions of $\bar{V}(h)$ and $V_0(b_N)$, we conclude

$$\begin{aligned} V(h) &= \bar{V}(h) + V_0(b_N) + (V_0(h) - V_0(b_N)) \\ &= \frac{1}{Nb_N^5}\int K_2^2(v)\,\mathrm{d}v + O_p\left(\frac{1}{Nb_N^4}\right) + o_p\left(\frac{N^{-\frac{\gamma}{2}}}{Nb_N^5}\right) \\ &\quad + o_p\left(\frac{\log N}{N^{\frac{3}{2}}b_N^5}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right). \end{aligned}$$

f) It remains to discuss $M(h)$ which follows the same line of arguments as for $V(h)$ with some simplification. First note that for deterministic b

$$\mathbb{E}M(b) = 0$$

as $\nu(x,b) = \mathbb{E}\hat{\lambda}''(x,b)$. For studying the remainder term $M(b) - \mathbb{E}M(b) = M(b)$ as in d), we write

$$\begin{aligned} M(b) &= \frac{1}{Nb^3}\int \sum_{j=1}^{N}\left\{K_2\left(\frac{x-X_j}{b}\right) - b^3\nu(x,b)\right\}\nu(x,b)\,\mathrm{d}x \\ &= \frac{1}{N}\sum_{j=1}^{N} R_b(X_j) \end{aligned}$$

as a mean of i.i.d. random variables with mean 0 by the definition of $\nu(x,b)$. Substituting $\frac{x-z}{b} = u$, we get

$$\begin{aligned} & R_b(z) \\ = \; & \frac{1}{b^2}\int \left\{K_2(u) - b^3\nu(z+bu,b)\right\}\nu(z+bu,b)\,\mathrm{d}u \\ = \; & \frac{1}{b^2}\int \left\{K_2(u) - b^3\lambda''(z+bu) - O(b^5)\right\}\left(\lambda''(z+bu) + O(b^2)\right)\mathrm{d}u \\ = \; & \frac{1}{b^2}\int \left\{K_2(u) - b^3\lambda''(z) - b^4u\lambda'''(z) + O(b^5)\right\}\left(\lambda''(z) + bu\lambda'''(z) + O(b^2)\right)\mathrm{d}u \\ = \; & \frac{1}{b^2}\int \left\{K_2(u) + O(b^3)\right\}\left\{\lambda''(z) + bu\lambda'''(z) + O(b^2)\right\}\mathrm{d}u \\ = \; & O(1) \end{aligned}$$

using the expansion of $\nu(x,b)$ from a), a Taylor expansion of $\lambda(x)$ and for the last step $\int K_2(u)\,\mathrm{d}u = 0 = \int uK_2(u)\,\mathrm{d}u$ from Assumption 11.

Hence, $|R_b(z)|$ is uniformly bounded in z and b by some constant, say r. Therefore, we may use Hoeffding's inequality for means of i.i.d. bounded random variables (compare, e.g., the lemma in Section 2.3.2 of Serfling (1980) for the one-sided version) and get for $a_N > 0$

$$\begin{aligned} \mathrm{pr}(a_N |M(b)| > \epsilon) &= \mathrm{pr}\left(|M(b)| \geq \frac{\epsilon}{a_N}\right) \\ &\leq 2\exp\left\{-\frac{N\epsilon^2}{4r^2 a_N^2}\right\} \to 0 \end{aligned}$$

if $\frac{N}{a_N^2} \to \infty$, and in that case, we have $a_N M(b) = o_p(1)$ or $M(b) = o_p\left(\frac{1}{a_N}\right)$.

It remains to study $M(h)$ for random h, which can be traced back to the behaviour of $M(b_N)$ as for $V(h)$ in e). We decompose

$$M(h) = M(b_N) + (M(h) - M(b_N))$$

and then split the second term into

$$(M(h) - M(b_N))\left(\mathbf{1}_{\{|h-b_N| \geq b_N N^{-\gamma}\}} + \mathbf{1}_{\{|h-b_N| < b_N N^{-\gamma}\}}\right).$$

By the same argument as in c), the first term is asymptotically negligible. For the second term, we again approximate h by $\bar{h} \in B_{N,\tau}^{\gamma}$, and we get for $a_N = \frac{\sqrt{N}}{\log N}$, for which $\frac{N}{a_N^2} = (\log N)^2 \to \infty$ holds,

$$\begin{aligned} & a_N |M(h) - M(b_N)| \mathbf{1}_{\{|h-b_N| < b_N N^{-\gamma}\}} \\ \leq\ & a_N \sup\left\{\left|M(b) - M(\bar{b})\right|;\right. \\ & \left. b, \bar{b} \in \left[b_N - b_N N^{-\gamma} - N^{-\tau}, b_N + b_N N^{-\gamma} + N^{-\tau}\right], \left|b - \bar{b}\right| \leq N^{-\tau}\right\} \\ & + a_N \sup_{b \in B_{N,\tau}^{\gamma}} |M(b)| + a_N |M(b_N)|. \end{aligned}$$

From the discussion above

$$a_N M(b_N) = o_p(1), \text{i.e. } M(b_N) = o_p\left(\frac{\log N}{\sqrt{N}}\right).$$

Using again Hoeffding's inequality and the same argument as for the supremum of $|W_{N,0}(b)|$ as in e),

$$\operatorname{pr}\left(a_N \sup_{b\in B_{N,\tau}^{\gamma}} |M(b)| > \epsilon\right) \leq 2c_B b_N \exp\left\{\tau \log N - \frac{\epsilon^2}{2r^2}(\log N)^2\right\} \to 0$$

i.e.

$$\sup_{b\in B_{N,\tau}^{\gamma}} |M(b)| = o_p\left(\frac{1}{a_N}\right) = o_p\left(\frac{\log N}{\sqrt{N}}\right)$$

too. Finally, we conclude with the help of Corollary 10 below, that for $\left|b-\bar{b}\right| \leq N^{-\tau}$ with $b, \bar{b} \geq b_N\left(1-2N^{-\gamma}\right)$

$$a_N \left|M(b) - M\left(\bar{b}\right)\right| \leq L\frac{\sqrt{N}}{\log N}\frac{N^{-\tau}}{b_N^5}\left(1+O\left(N^{-\gamma}\right)\right) \to 0$$

for large enough τ. Combining all terms, we get for $N\to\infty$

$$M(h) = o_p\left(\frac{\log N}{\sqrt{N}}\right).$$

g) Combining the expansions for $E(h)$, $M(h)$, $V(h)$, we finally have

$$\begin{aligned}
&\int \left(\hat{\lambda}''(x,h)\right)^2 \mathrm{d}x \\
=\;& \int \left(\lambda''(x)\right)^2 \mathrm{d}x + V_K \int \lambda''(x)\,\lambda^{(4)}(x)\,\mathrm{d}x b_N^2 + o\left(b_N^2\right) + o_p\left(b_N^2 N^{-\gamma}\right) \\
&+ \frac{1}{Nb_N^5}\int K_2^2(u)\,\mathrm{d}u + O_p\left(\frac{1}{Nb_N^4}\right) + o_p\left(\frac{N^{-\frac{\gamma}{2}}}{Nb_N^5}\right) + o_p\left(\frac{\log N}{N^{\frac{3}{2}}b_N^5}\right) \\
&+ o_p\left(\frac{\log N}{\sqrt{N}}\right).
\end{aligned}$$

□

Lemma 3. *Let the assumptions of Proposition 6 be fulfilled, and set*

$$L_2(u) = K_2 * K_2(u) = \int K_2(v)\,K_2(u-v)\,\mathrm{d}v.$$

a) $\int u^k L_2(u)\,\mathrm{d}u = 0$ *for* $0 \leq k \leq 3$, $\int u^4 L_2(u)\,\mathrm{d}u = 6\left(\int u^2 K_2(u)\,\mathrm{d}u\right)^2 = 24.$

b) $\mathbb{E}L_2\left(\frac{X_1-X_2}{b}\right) = O\left(b^5\right)$.

c) $\mathbb{E}L_2^2\left(\frac{X_1-X_2}{b}\right) = O\left(b^2\right)$.

Proof. a) From Assumption 11, we have, using integration by parts for the last two terms, $\int K_2(u)\,\mathrm{d}u = K'(1) - K'(-1) = 0$, $\int uK_2(u)\,\mathrm{d}u = -\int K'(u)\,\mathrm{d}u = K(-1) - K(1) = 0$. Substituting $w = u - v$, we then have

$$\int L_2(u)\,\mathrm{d}u = \int\int K_2(v)\,K_2(u-v)\,\mathrm{d}v\mathrm{d}u = \int\int K_2(v)\,K_2(w)\,\mathrm{d}w\mathrm{d}v = 0$$

$$\begin{aligned}\int uL_2(u)\,\mathrm{d}u &= \int\int uK_2(v)\,K_2(u-v)\,\mathrm{d}v\mathrm{d}u \\ &= \int\int (v+w)\,K_2(v)\,K_2(w)\,\mathrm{d}w\mathrm{d}v = 0\end{aligned}$$

$$\begin{aligned}\int u^2L_2(u)\,\mathrm{d}u &= \int\int (v+w)^2\,K_2(v)\,K_2(w)\,\mathrm{d}w\mathrm{d}v \\ &= \int w^2K_2(w)\,\mathrm{d}w \int K_2(v)\,\mathrm{d}v + \int v^2K_2(v)\,\mathrm{d}v \int K_2(w)\,\mathrm{d}w \\ &\quad +2\int wK_2(w)\,\mathrm{d}w \int vK_2(v)\,\mathrm{d}v \\ &= 0\end{aligned}$$

and, analogously, as all factors of the form $\int K_2(v)\,\mathrm{d}v$ or $\int vK_2(v)\,\mathrm{d}v$ vanish,

$$\begin{aligned}\int u^3L_2(u)\,\mathrm{d}u &= \int\int (v+w)^3\,K_2(v)\,K_2(w)\,\mathrm{d}w\mathrm{d}v = 0 \\ \int u^4L_2(u)\,\mathrm{d}u &= \int\int (v+w)^4\,K_2(v)\,K_2(w)\,\mathrm{d}w\mathrm{d}v \\ &= 6\int v^2K_2(v)\,\mathrm{d}v \int w^2K_2(w)\,\mathrm{d}w\end{aligned}$$

and, again by integration by parts, using Assumption 11,

$$\int u^2K_2(u)\,\mathrm{d}u = -2\int uK'(u)\,\mathrm{d}u = 2\int K(u)\,\mathrm{d}u = 2.$$

b) By a Taylor expansion up to order 4, we have for $0 \leq \theta \leq b$,

$$\begin{aligned}\lambda(z+bu) &= \sum_{\ell=0}^{3} b^{\ell}\lambda^{(\ell)}(z)\frac{1}{\ell!}u^{\ell} + b^{4}\lambda^{(4)}(z+\theta u)\frac{1}{4!}u^{4} \\ &= \sum_{\ell=0}^{3} b^{\ell}\lambda^{(\ell)}(z)\frac{1}{\ell!}u^{\ell} + O\left(b^{4}\right)\end{aligned}$$

uniformly in $u \in [-2,2]$, $0 \leq z \leq 1$, as $\lambda^{(4)}$ is bounded. Substituting $u = \frac{x-z}{b}$, we have, using a) for the last equality,

$$\begin{aligned}\mathbb{E}L_2\left(\frac{X_1-X_2}{b}\right) &= \int\int L_2\left(\frac{x-z}{b}\right)\lambda(x)\lambda(z)\,\mathrm{d}x\mathrm{d}z \\ &= b\int\int L_2(u)\lambda(z+bu)\,\mathrm{d}u\lambda(z)\,\mathrm{d}z \\ &= b\int\lambda(z)\left\{\sum_{\ell=0}^{3} b^{\ell}\lambda^{(\ell)}(z)\frac{1}{\ell!}\int u^{\ell}L_2(u)\,\mathrm{d}u + O\left(b^{4}\right)\right\}\mathrm{d}z \\ &= O\left(b^{5}\right).\end{aligned}$$

c) Substituting $u = \frac{x}{b}$, $v = \frac{z}{b}$, we have for some $C > 0$,

$$\begin{aligned}\mathbb{E}L_2^2\left(\frac{X_1-X_2}{b}\right) &= \int\int L_2^2\left(\frac{x-z}{b}\right)\lambda(x)\lambda(z)\,\mathrm{d}x\mathrm{d}z \\ &= b^2\int\int L_2^2(u-v)\lambda(bu)\lambda(bv)\,\mathrm{d}u\mathrm{d}v \\ &\leq C\cdot b^2\max_w L_2^2(w)\left(\max_u \lambda(u)\right)^2 \\ &= O\left(b^2\right)\end{aligned}$$

as λ is bounded and L_2 has a bounded support.

□

Lemma 4. *Under the assumptions of Proposition 6,*

$$\left|V(b)-V\left(\bar{b}\right)\right| \leq L\cdot\frac{\left|b-\bar{b}\right|}{b_m^8}, 0 < b, \bar{b} < 1,$$

for some suitable constant $L > 0$ and $b_m = \min\left(b,\bar{b}\right)$.

Proof. Throughout the proof, C denotes a generic constant which may assume different values. We write

$$\begin{aligned} V(b) &= \int \left(\frac{1}{b^3} G(x,b)\right)^2 \mathrm{d}x, G(x,b) = \frac{1}{N}\sum_{j=1}^{N} g_j(x,b) \\ g_j(x,b) &= K_2\left(\frac{x-X_j}{b}\right) - b^3 \nu(x,b). \end{aligned}$$

Due to Lipschitz continuity of λ,

$$\begin{aligned} & \left|b^3\nu(x,b) - \bar{b}^3\nu\left(x,\bar{b}\right)\right| \\ = & \left|b\int K_2(u)\lambda(x-bu)\,\mathrm{d}u - \bar{b}\int K_2(u)\lambda\left(x-\bar{b}u\right)\mathrm{d}u\right| \\ \leq & \; C\cdot\left|b-\bar{b}\right| + \bar{b}\left|\int K_2(u)\left\{\lambda(x-bu) - \lambda\left(x-\bar{b}u\right)\right\}\mathrm{d}u\right| \\ \leq & \; C\cdot\left|b-\bar{b}\right| \end{aligned}$$

as K_2 is bounded and $\bar{b}\leq 1$. Due to Lipschitz continuity of K_2

$$\left|K_2\left(\frac{x-X_j}{b}\right) - K_2\left(\frac{x-X_j}{\bar{b}}\right)\right| \leq C\left|x-X_j\right|\left|\frac{\bar{b}-b}{b\bar{b}}\right| \leq C\cdot\frac{\left|b-\bar{b}\right|}{b_m^2}$$

as $x, X_j\in[0,1]$. Hence, uniformly in j and x

$$\left|g_j(x,b) - g_j\left(x,\bar{b}\right)\right| \leq C\cdot\frac{\left|b-\bar{b}\right|}{b_m^2}$$

and, then,

$$\left|G(x,b) - G\left(x,\bar{b}\right)\right| \leq C\cdot\frac{\left|b-\bar{b}\right|}{b_m^2}.$$

Note that K_2 and ν are bounded, such that $G(x,b)$ is bounded too, such that we also have

$$\begin{aligned} \left|G^2(x,b) - G^2\left(x,\bar{b}\right)\right| &= \left|G(x,b) - G\left(x,\bar{b}\right)\right|\left|G(x,b) + G\left(x,\bar{b}\right)\right| \\ &\leq C\cdot\frac{\left|b-\bar{b}\right|}{b_m^2}. \end{aligned}$$

Finally, we get

$$\begin{aligned}\left|V(b)-V(\bar{b})\right| &\leq \frac{1}{b^6}\int\left|G^2(x,b)-G^2(x,\bar{b})\right|\mathrm{d}x\\&+\left|\frac{1}{b^6}-\frac{1}{\bar{b}^6}\right|\int G^2(x,\bar{b})\,\mathrm{d}x\\&\leq \frac{1}{b^6}C\frac{\left|b-\bar{b}\right|}{b_m^2}+C\frac{\left|b-\bar{b}\right|}{b_m^7}\leq C\cdot\frac{\left|b-\bar{b}\right|}{b_m^8}\end{aligned}$$

using the mean-value theorem for the function $\frac{1}{b^6}$. □

Corollary 10. *Under the assumptions of Lemma 4*

$$\left|M(b)-M(\bar{b})\right|\leq L\cdot\frac{\left|b-\bar{b}\right|}{b_m^5}, 0<b,\bar{b}\leq 1$$

for some suitable $L>0$.

Proof. As $\nu(x,b)$ is uniformly bounded in x,b, which follows from the proof of Proposition 6, a), using boundedness of K and λ'', we get from

$$M(b)=\int\frac{1}{b^3}G(x,b)\,\nu(x,b)\,\mathrm{d}x$$

with G as in the proof Lemma 4

$$\begin{aligned}\left|M(b)-M(\bar{b})\right| &\leq \int\left|\frac{1}{b^3}G(x,b)-\frac{1}{\bar{b}^3}G(x,\bar{b})\right|\nu(x,b)\,\mathrm{d}x\\&+\int\frac{1}{\bar{b}^3}G(x,\bar{b})\left|\nu(x,b)-\nu(x,\bar{b})\right|\mathrm{d}x\\&\leq C\cdot\frac{\left|b-\bar{b}\right|}{b_m^5}+C\frac{1}{b_m^3}\leq L\cdot\frac{\left|b-\bar{b}\right|}{b_m^5},\end{aligned}$$

where, for the first term, we use the same arguments as in the proof of Lemma 4 and for the second term boundedness of $G(x,b)$. □

Lemma 5. *Let the assumptions of Proposition 6 be satisfied. Let for some $a_N>0$ with $a_N^4=O(N^{-\gamma})$ and $Na_N\to\infty$, the bandwidths $b,\bar{b}$ satisfying $\alpha_N=a_N(1-O(N^{-\gamma}))<b,\bar{b}<a_N$. Then, for $q\geq 1$ and some $c>0$*

$$\mathbb{E}\left(U_N(b)-U_N(\bar{b})\right)^{2q}\leq c\left(\frac{a_N}{N}\right)^{2q}N^{-2q\gamma}.$$

Proof. The proof follows the same line of argument as the treatment of S_{11} in the proof of Lemma 3.1 of Hall and Marron (1987). Recall that

$$U_N(b) = \frac{2}{N(N-1)} \sum_{i \neq j} R_b(X_i, X_j)$$

with

$$\begin{aligned} & R_b(X_1, X_2) \\ = & \int \left\{ K_2\left(\frac{x - X_1}{b}\right) - \mathbb{E}K_2\left(\frac{x - X_1}{b}\right) \right\} \left\{ K_2\left(\frac{x - X_2}{b}\right) - \mathbb{E}K_2\left(\frac{x - X_2}{b}\right) \right\} \mathrm{d}x. \end{aligned}$$

As X_1, X_2 are independent, $\mathbb{E}R_b(X_1, X_2) = 0$, and we even have

$$\mathbb{E}\{R_b(X_1, X_2) | X_2\} = 0 = \mathbb{E}\{R_b(X_1, X_2) | X_1\}.$$

a) Let $L_2(z) = \int K_2(u) K_2(z - u) \mathrm{d}u$ be the convolution of K_2 with itself. As K_2 has support $[-1, +1]$, L_2 has support $[-2, +2]$. Note that, using the same notation as in the proof of Proposition 6,

$$\mathbb{E}K_2\left(\frac{x - X_1}{b}\right) = b^3 \mathbb{E}\hat{\lambda}''(x, b) = b^3 \nu(x, b).$$

Moreover, substituting $z = \frac{x-v}{b}$ and using symmetry of K_2

$$\int K_2\left(\frac{x - u}{b}\right) K_2\left(\frac{x - v}{b}\right) \mathrm{d}x = b \int K_2\left(z - \frac{u - v}{b}\right) K_2(z) \mathrm{d}z = bL_2\left(\frac{u - v}{b}\right).$$

Due to independence of X_1, X_2,

$$b\mathbb{E}L_2\left(\frac{X_1 - X_2}{b}\right) = \int \mathbb{E}K_2\left(\frac{x - X_1}{b}\right) \mathbb{E}K_2\left(\frac{x - X_2}{b}\right) \mathrm{d}x = b^6 \int \nu^2(x, b) \mathrm{d}x.$$

Writing

$$R_b(u) = \int b^3 \nu(x, b) \left\{ K_2\left(\frac{x - u}{b}\right) - b^3 \nu(x, b) \right\} \mathrm{d}x$$

which satisfies $\mathbb{E}R_b(X_1) = 0$, we get the decomposition

$$\begin{aligned} R_b(X_1, X_2) &= b\left\{ L_2\left(\frac{X_1 - X_2}{b}\right) - \mathbb{E}L_2\left(\frac{X_1 - X_2}{b}\right) \right\} - R_b(X_1) - R_b(X_2) \\ &= \Lambda_b(X_1 - X_2) - R_b(X_1) - R_b(X_2). \end{aligned}$$

b) Using Lipschitz continuity of K_2 and $X_1, X_2 \in [0,1]$, we have for some constants $c, \tilde{c}$

$$\begin{aligned}
& \left| bL_2\left(\frac{X_1 - X_2}{b}\right) - \bar{b}L_2\left(\frac{X_1 - X_2}{\bar{b}}\right) \right| \\
\leq \; & \int \left| K_2\left(\frac{x - X_1}{b}\right) - K_2\left(\frac{x - X_1}{\bar{b}}\right) \right| \cdot \left| K_2\left(\frac{x - X_2}{b}\right) \right| \mathrm{d}x \\
& + \int \left| K_2\left(\frac{x - X_1}{\bar{b}}\right) \right| \cdot \left| K_2\left(\frac{x - X_2}{b}\right) - K_2\left(\frac{x - X_2}{\bar{b}}\right) \right| \mathrm{d}x \\
\leq \; & c \left| \frac{1}{b} - \frac{1}{\bar{b}} \right| \left\{ \int |x - X_1| \left| K_2\left(\frac{x - X_2}{b}\right) \right| \mathrm{d}x + \int |x - X_2| \left| K_2\left(\frac{x - X_1}{\bar{b}}\right) \right| \mathrm{d}x \right\} \\
\leq \; & \tilde{c} \left| \frac{1}{b} - \frac{1}{\bar{b}} \right| \left(b + \bar{b}\right) \leq 2\tilde{c}\frac{a_N}{\alpha_N^2} \left|b - \bar{b}\right|
\end{aligned}$$

substituting $z = \frac{x - X_2}{b}$ respectively $z = \frac{x - X_1}{\bar{b}}$. Noting that $L_2\left(\frac{X_1 - X_2}{b}\right) = 0$ for $|X_1 - X_2| \geq 2a_N$ if $b \leq a_N$, we also have

$$\left| bL_2\left(\frac{X_1 - X_2}{b}\right) - \bar{b}L_2\left(\frac{X_1 - X_2}{\bar{b}}\right) \right| \leq 2\tilde{c}\frac{a_N}{\alpha_N^2} \left|b - \bar{b}\right| \cdot \mathbf{1}_{[-2,+2]}\left(\frac{X_1 - X_2}{a_N}\right).$$

Note that from substituting $u = \frac{x}{a_N}$, $v = \frac{z}{a_N}$

$$\begin{aligned}
\mathbb{E}\mathbf{1}_{[-2,+2]}\left(\frac{X_1 - X_2}{a_N}\right) &= \int\int \mathbf{1}_{[-2,+2]}\left(\frac{x - z}{a_N}\right) \lambda(x)\, \lambda(z)\, \mathrm{d}x\mathrm{d}z \\
&= a_N^2 \int\int \mathbf{1}_{[-2,+2]}(u - v)\, \lambda(a_N u)\, \lambda(a_N v)\, \mathrm{d}u\mathrm{d}v = O\left(a_N^2\right)
\end{aligned}$$

as λ is bounded. Hence, we have

$$\left|\Lambda_b(X_1 - X_2) - \Lambda_{\bar{b}}(X_1 - X_2)\right| \leq 2\tilde{c}\frac{a_N}{\alpha_N^2} \left|b - \bar{b}\right| \left\{ \mathbf{1}_{[-2,+2]}\left(\frac{X_1 - X_2}{a_N}\right) + O\left(a_N^2\right) \right\}.$$

c) From the proof of Proposition 6, a), we have

$$\left|\nu(x, b) - \nu\left(x, \bar{b}\right)\right| = \frac{1}{2} V_K \left|\lambda^{(4)}(x)\right| \cdot \left|b^2 - \bar{b}^2\right| + o\left(a_N^2\right)$$

for $b, \bar{b} \leq a_N$, and therefore, with some constant $c > 0$, using $b^2 - \bar{b}^2 = \left(b - \bar{b}\right)\left(b + \bar{b}\right)$,

$$\left|\nu(x, b) - \nu\left(x, \bar{b}\right)\right| \leq c a_N \left|b - \bar{b}\right| + o\left(a_N^2\right).$$

For getting an upper bound on $|R_b(X_1) - R_{\bar{b}}(X_1)|$, we decompose

$$
\begin{aligned}
& \int \left| b^3 \nu(x,b) K_2\left(\frac{x-X_1}{b}\right) - \bar{b}^3 \nu\left(x,\bar{b}\right) K_2\left(\frac{x-X_1}{\bar{b}}\right)\right| \mathrm{d}x \\
\leq \; & a_N^3 \int \left|\nu(x,b) - \nu\left(x,\bar{b}\right)\right| \left|K_2\left(\frac{x-X_1}{b}\right)\right| \mathrm{d}x \\
& + a_N^3 \int \left|\nu\left(x,\bar{b}\right)\right| \left|K_2\left(\frac{x-X_1}{b}\right) - K_2\left(\frac{x-X_1}{\bar{b}}\right)\right| \mathrm{d}x \\
\leq \; & \left\{c a_N^4 \left|b-\bar{b}\right| + o\left(a_N^5\right)\right\} \int \left|K_2\left(\frac{x-X_1}{b}\right)\right| \mathrm{d}x \\
& + a_N^3 \int \bar{c}\,|x-X_1| \cdot \left|\frac{1}{b} - \frac{1}{\bar{b}}\right| \mathrm{d}x \\
\leq \; & \tilde{c} a_N^5 \left|b-\bar{b}\right| + o\left(a_N^6\right) + \bar{c}\frac{a_N^3}{\alpha_N^2}\left|b-\bar{b}\right| \\
= \; & \bar{c}\frac{a_N^3}{\alpha_N^2}\left|b-\bar{b}\right| + o\left(a_N^6\right)
\end{aligned}
$$

with suitable constants $c, \bar{c}, \tilde{c}$, where we have used boundedness of $\nu(x,b)$, x, X_1, Lipschitz continuity of K_2 and, for the last inequality, substituted $u = \frac{x-X_1}{b}$. Note, from $\alpha_N \leq b, \bar{b} \leq a_N$, $a_N - \alpha_N = a_N O\left(N^{-\gamma}\right)$, the first term in the second last line is $o\left(a_N^6\right)$ too. The same inequality also holds for the expectations, such that we finally get

$$
|R_b(X_1) - R_{\bar{b}}(X_1)| \leq \bar{c}\frac{a_N^3}{\alpha_N^2}\left|b-\bar{b}\right| + o\left(a_N^6\right).
$$

Combining this with the inequality for $\Lambda_b(X_1 - X_2)$ from b), we finally get for some $c > 0$

$$
\begin{aligned}
& |R_b(X_1,X_2) - R_{\bar{b}}(X_1,X_2)| \\
\leq \; & c\frac{a_N}{\alpha_N^2}\left|b-\bar{b}\right| \left\{\mathbf{1}_{[-2,+2]}\left(\frac{X_1-X_2}{a_N}\right) + O\left(a_N^2\right)\right\} + o\left(a_N^6\right) \\
= \; & O\left(N^{-\gamma}\right)\left\{\mathbf{1}_{[-2,+2]}\left(\frac{X_1-X_2}{a_N}\right) + O\left(a_N^2\right)\right\}
\end{aligned}
$$

as $\left|b-\bar{b}\right| \leq a_N - \alpha_N = a_N O\left(N^{-\gamma}\right)$, $\frac{1}{\alpha_N^2} = \frac{1}{a_N^2}\left(1 + O\left(N^{-\gamma}\right)\right)$ and $o\left(a_N^6\right) = a_N^4 o\left(a_N^2\right) = o\left(a_N^2 N^{-\gamma}\right)$.

d) Using the abbreviation $\rho_N = \frac{2}{N(N-1)}$, we write

$$U_N(b) - U_N(\bar{b}) = \rho_N \sum_{i \neq j} \left(R_b(X_i, X_j) - R_{\bar{b}}(X_i, X_j)\right) = \rho_N \sum_{i \neq j} D_{ij}.$$

Then, for any $q \geq 1$,

$$\mathbb{E}\left(U_N(b) - U_N(\bar{b})\right)^{2q} = \rho_N^{2q} \sum_{i_1 \neq j_1, \ldots, i_{2q} \neq j_{2q}} \mathbb{E} D_{i_1 j_1} \cdots D_{i_{2q} j_{2q}}.$$

Following Hall and Marron (1987), we rearrange the sum with regard to the number m of different indices in $\{i_1, j_1, \ldots, i_{2q}, j_{2q}\}$, $m = 2, \ldots, 4q$. In the m-th group, there are at most $c_q N^m$ terms for some constant c_q depending only on q. Note also, that in the groups $m = 2q+1, \ldots 4q$, all expectations are 0, as, in this case, at least one index, say w.l.o.g. i_1, only appears once, and then

$$\begin{aligned}
\mathbb{E} D_{i_1 j_1} \cdots D_{i_{2q} j_{2q}} &= \mathbb{E}\left(\mathbb{E}\left\{D_{i_1 j_1} | X_{j_1}, X_{i_2}, X_{j_2}, \ldots, X_{i_{2q}}, X_{j_{2q}}\right\} D_{i_2 j_2} \cdots D_{i_{2q} j_{2q}}\right) \\
&= \mathbb{E}\left(\mathbb{E}\left\{D_{i_1 j_1} | X_{j_1}\right\} D_{i_2 j_2} \cdots D_{i_{2q} j_{2q}}\right) = 0
\end{aligned}$$

as $D_{i_1 j_1}$ is independent of $X_k, k \neq i_1, j_1$ and $\mathbb{E}\left\{R_b(X_i, X_j) | X_j\right\} = 0$ for $i \neq j$.

Using the upper bound on $|D_{ij}|$ from the end of c), we conclude for some constant c

$$\mathbb{E}\left(U_N(b) - U_N(\bar{b})\right)^{2q} \leq c \cdot \rho_N^{2q} \sum_{m=2}^{2q} N^m O\left(N^{-2q\gamma}\right) \eta_m,$$

where η_m is a bound on

$$\mathbb{E} \prod_{\ell=1}^{2q} \left(\mathbf{1}_{[-2,+2]}\left(\frac{X_{i_\ell} - X_{j_\ell}}{a_N}\right) + O\left(a_N^2\right)\right)$$

for $i_1, j_1, \ldots, i_{2q}, j_{2q}$ from group m, i.e. containing exactly m different indices.

Write $\mathbf{1}_{ij} = \mathbf{1}_{[-2,+2]}\left(\frac{X_i - X_j}{a_N}\right)$, and use $\mathbf{1}_{ij}^k = \mathbf{1}_{ij}$ for all $k \geq 2$. For $m = 2$, as $i_\ell \neq j_\ell$, we only have the case where $i_1 = \cdots = i_{2q} \neq j_1 = \cdots = j_{2q}$, and the expectation is

$$\mathbb{E}\left(\mathbf{1}_{i_1j_1}+O\left(a_N^2\right)\right)^{2q}=\mathbb{E}\mathbf{1}_{i_1j_1}+O\left(a_N^2\right)\cdot\mathbb{E}\mathbf{1}_{i_1j_1}=O\left(a_N^2\right)$$

as, from b), $\mathbb{E}\mathbf{1}_{ij}=O\left(a_N^2\right)$ for $i\neq j$. We conclude $\eta_2=O\left(a_N^2\right)$. By the same argument, the dominant term in the expectation is always $\mathbb{E}\mathbf{1}_{i_1j_1}\cdots\mathbf{1}_{i_{2q}j_{2q}}$. For given m, this results in an m-fold integral of a product of indicator functions $\mathbf{1}_{[-2,+2]}\left(\frac{z_k-z_\ell}{a_N}\right)$ with respect to $\lambda\left(z_1\right),\ldots,\lambda\left(z_m\right)$. Substituting $u_k=\frac{z_k}{a_N}$, this becomes a_N^m times the m-fold integral of the product of indicator functions $\mathbf{1}_{[-2,+2]}\left(u_k-u_\ell\right)$ with respect to $\lambda\left(a_Nu_1\right),\ldots,\lambda\left(a_Nu_m\right)$. As λ is bounded and the indicator functions are bounded with bounded support, this integral is $O\left(1\right)$, and, therefore, $\eta_m=a_N^mO\left(1\right)$, and we finally have

$$\begin{aligned}\mathbb{E}\left(U_N\left(b\right)-U_N\left(\bar{b}\right)\right)^{2q} &= \rho_N^{2q}\sum_{m=2}^{2q}N^ma_N^mO\left(N^{-2q\gamma}\right)\\ &\leq \rho_N^{2q}\sum_{m=0}^{2q}\left(Na_N\right)^mO\left(N^{-2q\gamma}\right)\\ &= \rho_N^{2q}O\left(N^{-2q\gamma}\right)\frac{\left(Na_N\right)^{2q+1}-1}{Na_N-1}\end{aligned}$$

from the formula of the geometric sum. As $\rho_N=O\left(\frac{1}{N^2}\right)$ and $Na_N\to\infty$, the right-hand side is $\frac{a_N^{2q}}{N^{2q}}O\left(N^{-2q\gamma}\right)$.

□

4.2 The two-dimensional case

We now study the analogous two-dimensional problems where we have to derive asymptotic expansions for

$$\hat{\Lambda}_{i\ell}=\int\int\hat{\lambda}_{ii}\left(x_1,x_2,H\right)\hat{\lambda}_{\ell\ell}\left(x_1,x_2,H\right)\mathrm{d}x_1\mathrm{d}x_2$$

for $i, \ell = 1, 2$ with random diagonal bandwidth matrix H. We need the following assumptions:

Assumption 12. *λ is 4-times continuously differentiable on $[0,1]^2$, and the partial derivatives of order 4 are Hölder continuous with some exponent $\beta > 0$.*

The kernel K has to satisfy the previous assumptions 1, 2, 4, 7, 8, 9, which we state again as following:

Assumption 13. *We make the following assumption for the kernel:*

i) *$K(u_1, u_2)$ is a non-negative kernel function on $[-1,+1]^2$, integrating to 1.*

ii) *K is twice continuously differentiable, and the second-order derivatives $K_{ii}(u_1, u_2) = \frac{\partial^2}{\partial u_i^2} K(u_1, u_2)$, $i = 1, 2$, are Lipschitz continuous.*

iii) *K and its first-order derivatives $K_i(u) = \frac{\partial}{\partial u_i} K(u)$ satisfy the symmetry conditions*

a) $K(\pm 1, u_2) = K(u_1, \pm 1) = 0$,

$K_i(\pm 1, u_2) = K_i(u_1, \pm 1) = 0$, $i = 1, 2$, *for all* $-1 \leq u_1, u_2 \leq 1$.

b) $\int u_i K(u)\, \mathrm{d}u_i = 0$ *for all* u_j, $j \neq i$, $i = 1, 2$.

c) $\int \int u_i^2 K(u)\, \mathrm{d}u_1 \mathrm{d}u_2 = V_K$, $i = 1, 2$, $\int \int u_i^3 K(u)\, \mathrm{d}u_1 \mathrm{d}u_2 = 0$, $i = 1, 2$.

Note that from iii)a), we in particular have

$$\int \int K_{11}(u_1, u_2)\, \mathrm{d}u_1 \mathrm{d}u_2 = \int (K_1(+1, u_2) - K_1(-1, u_2))\, \mathrm{d}u_2 = 0$$

and, analogously, $\int \int K_{22}(u_1, u_2)\, \mathrm{d}u_1 \mathrm{d}u_2 = 0$ *too.*

Theorem 4. *Let Assumptions 12 and 13 be fulfilled. Let h_1 and h_2 be sequences of random bandwidths which can be approximated by sequences of deterministic bandwidths converging to 0 with the same rate such that for some $0 < b_N \to 0$ $(N \to \infty)$*

$$h_i = \beta_i b_N \left(1 + o_p\left(N^{-\gamma}\right)\right)$$

for some $\gamma \geq 0$.

Then, for $i = 1, 2$

$$\begin{aligned}
\hat{\Lambda}_{ii} &= \int\int \hat{\lambda}_{ii}^2 (x_1, x_2, H)\, \mathrm{d}x_1 \mathrm{d}x_2 \\
&= \int\int \lambda_{ii}^2 (x_1, x_2)\, \mathrm{d}x_1 \mathrm{d}x_2 + b_N^2 V_K \int\int \lambda_{ii}(x_1, x_2) \sum_{\ell=1}^{2} \beta_\ell^2 \frac{\partial^2}{\partial x_\ell^2} \lambda_{ii}(x_1, x_2)\, \mathrm{d}x_1 \mathrm{d}x_2 \\
&\quad + \frac{1}{N\beta_1\beta_2\beta_i^4 b_N^6} \int\int K_{ii}^2(u)\, \mathrm{d}u_1 \mathrm{d}u_2 + R_{N,ii}
\end{aligned}$$

and

$$\begin{aligned}
\hat{\Lambda}_{12} &= \int\int \hat{\lambda}_{11}(x_1, x_2, H)\, \hat{\lambda}_{22}(x_1, x_2, H)\, \mathrm{d}x_1 \mathrm{d}x_2 \\
&= \int\int \lambda_{11}(x_1, x_2)\, \lambda_{22}(x_1, x_2)\, \mathrm{d}x_1 \mathrm{d}x_2 \\
&\quad + b_N^2 \frac{V_K}{2} \int\int \sum_{\ell=1}^{2} \beta_\ell^2 \left\{ \lambda_{11}(x_1, x_2) \frac{\partial^2}{\partial x_\ell^2} \lambda_{22}(x_1, x_2) + \lambda_{22}(x_1, x_2) \frac{\partial^2}{\partial x_\ell^2} \lambda_{11}(x_1, x_2) \right\} \mathrm{d}x_1 \mathrm{d}x_2 \\
&\quad + \frac{1}{N\beta_1^3\beta_2^3 b_N^6} \int\int K_{11}(u)\, K_{22}(u)\, \mathrm{d}u_1 \mathrm{d}u_2 + R_{N,12},
\end{aligned}$$

where the remainder terms $R_{N,i\ell}$, $i, \ell = 1, 2$, are all of the order

$$R_{N,i\ell} = o\left(b_N^2\right) + o_p\left(b_N^2 N^{-\gamma}\right) + O_p\left(\frac{1}{N b_N^5}\right) + o_p\left(\frac{N^{-\gamma}}{N b_N^6}\right) + o_p\left(\frac{\log N}{N^{\frac{3}{2}} b_N^6}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right).$$

Note that the transition from dimension 1 to 2 changes the rate of the variance part from $\left(N b_N^5\right)^{-1}$ to $\left(N b_N^6\right)^{-1}$ whereas the rate of the bias part of the expansion remains b_N^2. This is in line with well-known other results on kernel estimates.

Proof. As the proof is completely analogous to that of Proposition 6, we only formulate the main steps. H, B denote diagonal bandwidths matrices with random entries h_1, h_2 respectively with deterministic entries b_1, b_2. From Condition 2, $\frac{h_i}{b_N} \underset{p}{\rightarrow} \beta_i > 0$ for $N \rightarrow \infty$. We also consider only deterministic sequences of bandwidths sharing this asymptotic behaviour, i.e. b_1, b_2 have the same rate which is given by b_N. In particular, we frequently use $O\left(b_1^k b_2^\ell\right) = O\left(b_N^{k+\ell}\right)$, $k, \ell \geq 0$.

We introduce

$$\nu_i\left(x_1, x_2, H\right) = \frac{1}{h_i^2} \int\int K_{ii}\left(u_1, u_2\right) \lambda\left(x_1 - h_1 u_1, x_2 - h_2 u_2\right) \mathrm{d}u_1 \mathrm{d}u_2, i = 1, 2$$

such that for deterministic b_1, b_2

$$\nu_i\left(x_1, x_2, B\right) = \mathbb{E}\hat{\lambda}_{ii}\left(x_1, x_2, B\right).$$

We first consider the integral of $\hat{\lambda}_{11}^2\left(x_1, x_2, H\right)$, and split it into

$$\int\int \hat{\lambda}_{11}^2\left(x_1, x_2, H\right) \mathrm{d}x_1 \mathrm{d}x_2 = E\left(H\right) + 2M\left(H\right) + V\left(H\right)$$

with

$$\begin{aligned} E\left(H\right) &= \int\int \nu_1^2\left(x_1, x_2, H\right) \mathrm{d}x_1 \mathrm{d}x_2, \\ M\left(H\right) &= \int\int \nu_1\left(x_1, x_2, H\right) \left\{\hat{\lambda}_{11}\left(x_1, x_2, H\right) - \nu_1\left(x_1, x_2, H\right)\right\} \mathrm{d}x_1 \mathrm{d}x_2 \\ V\left(H\right) &= \int\int \left(\hat{\lambda}_{11}\left(x_1, x_2, H\right) - \nu_1\left(x_1, x_2, H\right)\right)^2 \mathrm{d}x_1 \mathrm{d}x_2. \end{aligned}$$

The treatment of the integral of $\hat{\lambda}_{22}^2\left(x_1, x_2, H\right)$ is completely analogous. For the integral of $\hat{\lambda}_{11}\left(x_1, x_2, H\right) \hat{\lambda}_{22}\left(x_1, x_2, H\right)$, we discuss it at the end of the proof.

a) From the proof of Proposition 3, a), we have

$$\nu_1\left(x_1, x_2, H\right) = \lambda_{11}\left(x_1, x_2\right) + \frac{1}{2} V_K \sum_{i=1}^{2} h_i^2 \frac{\partial^2}{\partial x_i^2} \lambda_{11}\left(x_1, x_2\right) + O_p\left(b_N^{2+\beta}\left(1 + o_p\left(N^{-\gamma}\right)\right)\right)$$

and, then, writing B_N for the diagonal matrix with entries $\beta_1 b_N$ respectively $\beta_2 b_N$

$$\begin{aligned} \nu_1\left(x_1, x_2, H\right) - \nu_1\left(x_1, x_2, B_N\right) &= \frac{1}{2} V_K \sum_{i=1}^{2} \left(h_i^2 - \beta_i^2 b_N^2\right) \frac{\partial^2}{\partial x_i^2} \lambda_{11}\left(x_1, x_2\right) + o_p\left(b_N^{2+\beta} N^{-\gamma}\right) \\ &= o_p\left(b_N^2 N^{-\gamma}\right). \end{aligned}$$

b) From the boundedness of $\nu_1(x_1, x_2, B)$ and a) we get

$$\begin{aligned} E(H) &= \int\int \lambda_{11}^2(x_1,x_2)\,\mathrm{d}x_1\mathrm{d}x_2 + V_K \int\int \lambda_{11}(x_1,x_2) \sum_{i=1}^{2} \beta_i^2 \frac{\partial^2}{\partial x_i^2}\lambda_{11}(x_1,x_2)\,\mathrm{d}x_1\mathrm{d}x_2 \cdot b_N^2 \\ &\quad + o\left(b_N^2\right) + o_p\left(b_N^2 N^{-\gamma}\right). \end{aligned}$$

c) Using $\nu_1(x_1, x_2, B) = \mathbb{E}\hat{\lambda}_{11}(x_1, x_2, B)$, we have for deterministic b_1, b_2

$$\mathbb{E}V(B) = \frac{1}{Nb_1^6 b_2^2}\int\int \mathrm{var}K_{11}\left(\frac{x_1 - X_{11}}{b_1}, \frac{x_2 - X_{12}}{b_2}\right)\mathrm{d}x_1\mathrm{d}x_2.$$

Using again boundedness of $\nu_1(x_1, x_2, B)$ and

$$\mathbb{E}K_{11}\left(\frac{x_1 - X_{11}}{b_1}, \frac{x_2 - X_{12}}{b_2}\right) = b_1^3 b_2 \nu_1(x_1, x_2, B)$$

we get

$$\begin{aligned} &\mathbb{E}V(B) \\ =\ & \frac{1}{Nb_1^6 b_2^2}\int\int \mathbb{E}K_{11}^2\left(\frac{x_1 - X_{11}}{b_1}, \frac{x_2 - X_{12}}{b_2}\right)\mathrm{d}x_1\mathrm{d}x_2 - \frac{1}{N}\int\int \nu_1^2(x_1,x_2,B)\,\mathrm{d}x_1\mathrm{d}x_2 \\ =\ & \frac{1}{Nb_1^5 b_2}\int\int\int\int K_{11}^2(u_1,u_2)\,\lambda(x_1 - b_1u_1, x_2 - b_2u_2)\,\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}x_1\mathrm{d}x_2 + O\left(\frac{1}{N}\right) \\ =\ & \frac{1}{Nb_1^5 b_2}\int\int\int\int K_{11}^2(u_1,u_2)\,\lambda(x_1,x_2)\,\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}x_1\mathrm{d}x_2 \\ & + \frac{1}{Nb_1^5 b_2}\int\int\int\int K_{11}^2(u_1,u_2)\left\{-\sum_{i=1}^{2} b_i u_i \lambda_i(x_1,x_2) + o(b_1 + b_2)\right\}\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}x_1\mathrm{d}x_2 \\ & + O\left(\frac{1}{N}\right) \\ =\ & \frac{1}{Nb_1^5 b_2}\int\int K_{11}^2(u_1,u_2)\,\mathrm{d}u_1\mathrm{d}u_2 + O\left(\frac{1}{Nb_N^5}\right). \end{aligned}$$

d) To study the asymptotic behaviour of $V(B) - \mathbb{E}V(B)$, we decompose

$$V(B) = \frac{1}{Nb_1^5 b_2}W_N(B) + \frac{N-1}{2Nb_1^6 b_2^2}U_N(B)$$

with

$$\begin{aligned} W_N(B) &= \frac{1}{Nb_1 b_2}\sum_{j=1}^{N}\int\int\left(K_{11}\left(\frac{x_1 - X_{j1}}{b_1}, \frac{x_2 - X_{j2}}{b_2}\right) - b_1^3 b_2 \nu_1(x_1,x_2,B)\right)^2 \mathrm{d}x_1\mathrm{d}x_2 \\ &= \frac{1}{N}\sum_{j=1}^{N} Q_B(X_j) \end{aligned}$$

being a mean of i.i.d. random variables, and

$$\begin{aligned} U_N(B) &= \frac{2}{N(N-1)}\sum_{i\neq j}\int\int\left(K_{11}\left(\frac{x_1-X_{i1}}{b_1},\frac{x_2-X_{i2}}{b_2}\right)-b_1^3b_2\nu_1(x_1,x_2,B)\right) \\ &\quad \left(K_{11}\left(\frac{x_1-X_{j1}}{b_1},\frac{x_2-X_{j2}}{b_2}\right)-b_1^3b_2\nu_1(x_1,x_2,B)\right)\mathrm{d}x_1\mathrm{d}x_2 \\ &= \frac{2}{N(N-1)}\sum_{i\neq j}R_B(X_i,X_j) \end{aligned}$$

being a U-statistic with kernel R_B.

Substituting $u_1=\frac{x_1-X_{j1}}{b_1}$, $u_2=\frac{x_2-X_{j2}}{b_2}$, $u=(u_1,u_2)^\top$, $Bu=(b_1u_1,b_2u_2)^\top$

$$Q_B(X_j)=\int\int\left(K_{11}(u_1,u_2)-b_1^3b_2\nu_1(X_j+Bu,B)\right)^2\mathrm{d}u_1\mathrm{d}u_2\leq C, 1\leq j\leq N$$

and we can again apply Hoeffding's inequality, to show

$$\frac{1}{Nb_1^5b_2}\left(W_N(B)-\mathbb{E}W_N(B)\right)=o_p\left(\frac{\log N}{N^{\frac{3}{2}}b_1^5b_2}\right)=o_p\left(\frac{\log N}{N^{\frac{3}{2}}b_N^6}\right).$$

For the second component, we decompose

$$\begin{aligned} R_B(X_1,X_2) &= b_1b_2\left\{L_2\left(B^{-1}(X_2-X_1)\right)-\mathbb{E}L_2\left(B^{-1}(X_2-X_1)\right)\right\} \\ &\quad -R_B(X_1)-R_B(X_2), \end{aligned}$$

where L_2, R_B are defined as

$$L_2(u)=\int\int K_{11}(v)K_{11}(v-u)\mathrm{d}v_1\mathrm{d}v_2, u\in\mathbb{R}^2,$$

$$\begin{aligned} R_B(u) &= \int\int b_1^3b_2\nu_1(x_1,x_2,B)\left\{K_{11}\left(\frac{x_1-u_1}{b_1},\frac{x_2-u_2}{b_2}\right)-b_1^3b_2\nu_1(x_1,x_2,B)\right\}\mathrm{d}x_1\mathrm{d}x_2 \\ &= \int\int b_1^4b_2^2\nu_1(u+Bz,B)K_{11}(z)\mathrm{d}z_1\mathrm{d}z_2+O\left(b_1^6b_2^2\right) \\ &= \int\int b_1^4b_2^2\lambda_{11}(u+Bz)K_{11}(z)\mathrm{d}z_1\mathrm{d}z_2+O\left(b_1^6b_2^2\right)+O\left(b_1^4b_2^4\right) \\ &= b_1^4b_2^2\int\int K_{11}(z)\left\{\lambda_{11}(u)+\nabla^\top\lambda_{11}(u)Bz+O\left(b_1^2+b_2^2\right)\right\}\mathrm{d}z_1\mathrm{d}z_2+O\left(b_N^8\right) \\ &= O\left(b_N^8\right) \end{aligned}$$

as from Assumption 11, the integrals of $K_{11}(z)$ and of $zK_{11}(z)$ vanish. Hence, $\frac{1}{b_1^6 b_2^2} R_B(u)$ is bounded, and we also have $\mathbb{E} R_B(X_1) = 0$, such that we may again apply Hoeffding's inequality to get

$$\frac{1}{N b_1^6 b_2^2} \sum_{j=1}^{N} R_B(X_j) = o_p\left(\frac{\log N}{\sqrt{N}}\right)$$

and

$$\frac{1}{b_1^6 b_2^2} U_N(B) = \frac{1}{b_1^6 b_2^2} \tilde{U}_N(B) + o_p\left(\frac{\log N}{\sqrt{N}}\right),$$

where $\tilde{U}_N(B)$ is the U-statistic with kernel

$$\Lambda_B(x - z) = b_1 b_2 \left\{ L_2\left(B^{-1}(x - z)\right) - \mathbb{E} L_2\left(B^{-1}(X_1 - X_2)\right) \right\}.$$

We set $\ell_B(x) = \mathbb{E}\Lambda_B(x - X_2)$, $\varsigma_1 = \mathrm{var}\,\ell_B(X_1)$, $\varsigma_2 = \mathrm{var}\Lambda_B(X_1 - X_2)$. By the same argument as in the proof of Lemma 6, b),

$$\begin{aligned} b_1 b_2 \int\int L_2\left(B^{-1}(x - z)\right) \lambda(z)\, \mathrm{d}z_1 \mathrm{d}z_2 &= b_1^2 b_2^2 \int\int L_2(u)\, \lambda(x + Bu)\, \mathrm{d}u_1 \mathrm{d}u_2 \\ &= O\left(b_N^8\right) \end{aligned}$$

uniformly in x, which implies $\varsigma_1 = O(b_N^{16})$. From Lemma 6, b), c), we have

$$\varsigma_2 = b_1^2 b_2^2 \mathrm{var} L_2\left(B^{-1}(X_1 - X_2)\right) \leq b_1^2 b_2^2 \mathbb{E} L_2^2\left(B^{-1}(X_1 - X_2)\right) = O\left(b_N^8\right)$$

such that from Lemma A in Section 5.2.1 of Serfling (1980)

$$\mathrm{var}\tilde{U}_N(B) = O\left(\frac{b_N^{16}}{N}\right) + O\left(\frac{b_N^8}{N^2}\right)$$

and

$$\frac{1}{b_1^6 b_2^2} \tilde{U}_N(B) = O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{N b_N^4}\right)$$

and, correspondingly,

$$\frac{1}{b_1^6 b_2^2} U_N(B) = O_p\left(\frac{1}{N b_N^4}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right).$$

Together with the bound on $W_N(B) - \mathbb{E}W_N(B)$, we finally have

$$V(B) - \mathbb{E}V(B) = O_p\left(\frac{1}{Nb_N^4}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right) + o_p\left(\frac{\log N}{N^{\frac{3}{2}}b_N^6}\right).$$

e) For random bandwidths h_1, h_2, we decompose $V(H)$ into

$$V(H) = \bar{V}(H) + \left(V(H) - \bar{V}(H)\right),$$

where

$$\begin{aligned}
\bar{V}(H) &= \frac{1}{Nh_1^6h_2^2}\int\int\int\int K_{11}^2\left(H^{-1}(z-u)\right)\lambda(u)\,\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}z_1\mathrm{d}z_2 \\
&\quad -\frac{1}{N}\int\int \nu_1^2(x_1,x_2,H)\,\mathrm{d}x_1\mathrm{d}x_2 \\
&= \frac{1}{Nh_1^5h_2}\int\int\int\int K_{11}^2(v_1,v_2)\,\lambda(z-Hv)\,\mathrm{d}v_1\mathrm{d}v_2\mathrm{d}z_1\mathrm{d}z_2 \\
&\quad -\frac{1}{N}\int\int \nu_1^2(x_1,x_2,H)\,\mathrm{d}x_1\mathrm{d}x_2.
\end{aligned}$$

From c), $\bar{V}(B) = \mathbb{E}V(B)$ for deterministic B. Using the same expansion as in c), together with $h_i = \beta_i b_N\left(1 + o_p\left(N^{-\gamma}\right)\right)$, $i = 1, 2$,

$$\bar{V}(H) = \frac{1}{N\beta_1^5\beta_2 b_N^6}\int\int K_{11}^2(u_1,u_2)\,\mathrm{d}u_1\mathrm{d}u_2 + o_p\left(\frac{N^{-\gamma}}{Nb_N^6}\right) + O_p\left(\frac{1}{Nb_N^5}\right).$$

It remains to study $V_0(H) = V(H) - \bar{V}(H)$, which we decompose as

$$V_0(H) = V_0(B_N) + \left(V_0(H) - V_0(B_N)\right)$$

writing B_N for the diagonal matrix with entries $\beta_1 b_N$ and $\beta_2 b_N$. From d),

$$V_0(B_N) = V(B_N) - \mathbb{E}V(B_N) = O_p\left(\frac{1}{Nb_N^4}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right) + o_p\left(\frac{\log N}{N^{\frac{3}{2}}b_N^6}\right).$$

We split

$$V_0(H) - V_0(B_N) = \left(V_0(H) - V_0(B_N)\right)\left(\mathbf{1}_{A_N^C} + \mathbf{1}_{A_N}\right),$$

where $A_N = \{|h_1 - \beta_1 b_N| < \beta_1 b_N N^{-\gamma}\} \cap \{|h_2 - \beta_2 b_N| < \beta_2 b_N N^{-\gamma}\}$. As in the proof of Proposition 6, e), we have that, due to $h_i = \beta_i b_N (1 + o_p (N^{-\gamma}))$, $i = 1, 2$, the first term of the right-hand side is asymptotically negligible.

For the second term, we approximate h by $\bar{h}$ from a finite grid $B^2_{N,\tau}$, where $B_{N,\tau}$ is defined as in the proof of Proposition 6. In particular, we have $\left|h_i - \bar{h}_i\right| \le N^{-\tau}$, $i = 1, 2$. We set

$$\begin{aligned} B^{\gamma,2}_{N,\tau} &= B^2_{N,\tau} \cap \left(\beta_1 b_N \left(1 - N^{-\gamma}\right) - N^{-\tau}, \beta_1 b_N \left(1 + N^{-\gamma}\right) + N^{-\tau}\right) \\ &\quad \times \left(\beta_2 b_N \left(1 - N^{-\gamma}\right) - N^{-\tau}, \beta_2 b_N \left(1 + N^{-\gamma}\right) + N^{-\tau}\right). \end{aligned}$$

Note that $B^{\gamma,2}_{N,\tau}$ satisfies $\left|B^{\gamma,2}_{N,\tau}\right| \le c_B N^{2\tau} b_N^2$ for some constant $c_B > 0$. We decompose, with $\bar{H}$ denoting the diagonal matrix with entries $\bar{h}_1, \bar{h}_2$,

$$\begin{aligned} |V_0(H) - V_0(B_N)| \mathbf{1}_{A_N} &\le \left|V_0(H) - V_0(\bar{H})\right| \mathbf{1}_{A_N} + \sup_{b \in B^{\gamma,2}_{N,\tau}} |V_0(B) - V_0(B_N)| \\ &\le S_1 + S_2, \end{aligned}$$

where, with $\bar{B}$ denoting a diagonal matrix with entries $\bar{b}_1, \bar{b}_2$

$$\begin{aligned} & S_1 \\ = \ & \sup \left\{ \left|V_0(B) - V_0(\bar{B})\right|; \right. \\ & \left. \left|b_i - \bar{b}_i\right| \le N^{-\tau}, b_i, \bar{b}_i \in \left[\beta_i b_N \left(1 - N^{-\gamma}\right) - N^{-\tau}, \beta_i b_N \left(1 + N^{-\gamma}\right) + N^{-\tau}\right], i = 1, 2 \right\} \\ \le \ & \sup \left\{ \left|V_0(B) - V_0(\bar{B})\right|; b_i, \bar{b}_i \ge \beta_i b_N \left(1 - 2N^{-\gamma}\right), \left|b_i - \bar{b}_i\right| \le N^{-\tau}, i = 1, 2 \right\} \\ \le \ & \frac{4L}{N^\tau b_N^{10}} \left(1 + O\left(N^{-\gamma}\right)\right) \end{aligned}$$

from the Lipschitz property of $V(B)$ stated in Lemma 7.

For getting a bound on S_2, we decompose as in d)

$$V(B) = \frac{1}{N b_1^5 b_2} W_N(B) + \frac{N-1}{2N b_1^6 b_2^2} U_N(B)$$

and, with $W_{N,0}(B) = W_N(B) - \mathbb{E} W_N(B)$, using $\mathbb{E} U_N(B) = 0$,

$$V_0(B) = \frac{1}{N b_1^5 b_2} W_{N,0}(B) + \frac{N-1}{2N b_1^6 b_2^2} U_N(B).$$

Using $\left|B_{N,\tau}^{\gamma;2}\right| \leq c_B N^{2\tau} b_N^2$, we get from Hoeffding's inequality as in d) and using the same argument as in the proof of Proposition 6

$$\sup_{b\in B_{N,\tau}^{\gamma;2}} \left|W_{N,0}(B)\right| = o_p\left(\frac{\log N}{\sqrt{N}}\right)$$

and the first component of S_2 is

$$\sup_{b\in B_{N,\tau}^{\gamma;2}} \left|\frac{1}{Nb_1^5 b_2} W_{N,0}(B) - \frac{1}{N\beta_1^5\beta_2 b_N^6} W_{N,0}(B_N)\right| = o_p\left(\frac{\log N}{N^{\frac{3}{2}} b_N^6}\right).$$

Finally, we have to study

$$\begin{aligned}
&\frac{N-1}{2N} \sup_{b\in B_{N,\tau}^{\gamma;2}} \left|\frac{1}{b_1^6 b_2^2} U_N(B) - \frac{1}{\beta_1\beta_2 b_N^8} U_N(B_N)\right| \\
\leq\ & \frac{(N-1)}{2N}\frac{(1+O(N^{-\gamma}))}{b_N^8} \sup_{b\in B_{N,\tau}^{\gamma;2}} \left|U_N(B) - U_N(B_N)\right| \\
&+\frac{(N-1)}{2N}\frac{(1+O(N^{-\gamma}))}{b_N^8}\left|U_N(B_N)\right|
\end{aligned}$$

as, for large enough N, τ, $\beta_i b_N(1-2N^{-\gamma}) \leq b_i \leq \beta_i b_N(1+2N^{-\gamma})$, $i=1,2$, for all $b \in B_{N,\tau}^{\gamma;2}$. Again, from d), the second term satisfies

$$\frac{(N-1)}{2N}\frac{(1+O(N^{-\gamma}))}{b_N^8}\left|U_N(B_N)\right| = O_p\left(\frac{1}{Nb_N^4}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right).$$

For the first term, we use Lemma 8 with $q \geq 1$, $a_{iN} = \beta_i b_N(1+2N^{-\gamma})$, $\alpha_{iN} = \beta_i b_N(1-2N^{-\gamma})$, $i=1,2$,

$$\begin{aligned}
\mathbb{E}\left(U_N(B) - U_N(\bar{B})\right)^{2q} &\leq c\cdot\left(\frac{b_N^3}{N}\right)^{2q}\left(1+2N^{-\gamma}\right)^{6q} N^{-2q\gamma} \\
&\leq \bar{c}\left(\frac{b_N^3}{N}\right)^{2q} N^{-2q\gamma}
\end{aligned}$$

for some suitable constant $\bar{c}$, not depending on N, and all $b, \bar{b}$ with $b_i, \bar{b}_i \in [\alpha_{iN}, a_{iN}]$, $i = 1, 2$. Using the same argument as in Proposition 6, we conclude for arbitrarily small $\rho > 0$

$$\frac{1}{b_N^8}\sup_{b\in B_{N,\tau}^{\gamma;2}} \left|U_N(B) - U_N(B_N)\right| = o_p\left(\frac{N^\rho}{N^{1+\gamma} b_N^5}\right).$$

In particular, this term is of order $o_p\left(\frac{N^{-\frac{\gamma}{2}}}{Nb_N^5}\right)$ for $\rho \leq \frac{\gamma}{2}$. We finally conclude

$$S_2 = o_p\left(\frac{\log N}{N^{\frac{3}{2}}b_N^6}\right) + o_p\left(\frac{N^{-\frac{\gamma}{2}}}{Nb_N^5}\right) + O_p\left(\frac{1}{Nb_N^4}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right),$$

and we have the same rate for $V_0(H) - V_0(B_N)$, as S_1 is negligible for large enough τ, and $(V_0(H) - V_0(B_N))\,\mathbf{1}_{A_N^C}$ is negligible, too. Together with the expansion of $\bar{V}(H)$ and $V_0(B_N)$, we conclude

$$\begin{aligned} V(H) &= \bar{V}(H) + V_0(B_N) + (V_0(H) - V_0(B_N)) \\ &= \frac{1}{N\beta_1^5\beta_2 b_N^6}\int\int K_{11}^2(u)\,\mathrm{d}u_1\mathrm{d}u_2 + O_p\left(\frac{1}{Nb_N^5}\right) + o_p\left(\frac{N^{-\gamma}}{Nb_N^6}\right) \\ &\quad + o_p\left(\frac{\log N}{N^{\frac{3}{2}}b_N^6}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right). \end{aligned}$$

f) For deterministic B, we write

$$M(B) = \frac{1}{Nb_1^6 b_2^2}\sum_{j=1}^{N} R_B(X_j),$$

where $R_B(u)$ is defined in d). From d), we therefore have $M(B) = o_p\left(\frac{\log N}{\sqrt{N}}\right)$.

To study the behaviour of $M(H)$ for random H, we decompose it as

$$M(H) = M(B_N) + (M(H) - M(B_N)),$$

where B_N is defined in e). We split the second term into

$$M(H) - M(B_N) = (M(H) - M(B_N))\left(\mathbf{1}_{A_N^C} + \mathbf{1}_{A_N}\right)$$

with A_N as in e), and, as there, we conclude that the first term is negligible. The second term is treated as $(V_0(H) - V_0(B_N))\,\mathbf{1}_{A_N}$ in e) by approximating h_1, h_2 by the closest grid points $\bar{h}_1, \bar{h}_2$ in $B_{N,\tau}$

$$|M(H) - M(B_N)|\,\mathbf{1}_{A_N} \leq \left|M(H) - M\left(\bar{H}\right)\right|\mathbf{1}_{A_N} + \sup_{b\in B_{N,\tau}^{\gamma,2}} |M(B) - M(B_N)|.$$

The first term is again asymptotically negligible for large enough τ using the Lipschitz property of $M(B)$ from Corollary 11. For the second term, we use

$$\sup_{b \in B_{N,\tau}^{\gamma,2}} |M(B) - M(B_N)| \leq \sup_{b \in B_{N,\tau}^{\gamma,2}} |M(B)| + |M(B_N)|,$$

where $M(B_N) = o_p\left(\frac{\log N}{\sqrt{N}}\right)$ from the considerations above. Using Hoeffding's inequality, noting that $\mathbb{E}M(B) = 0$, and the same argument as in the proof of Proposition 6, we also get the same rate for the supremum

$$\sup_{b \in B_{N,\tau}^{\gamma,2}} |M(B)| = o_p\left(\frac{\log N}{\sqrt{N}}\right),$$

and, finally,

$$M(H) = o_p\left(\frac{\log N}{\sqrt{N}}\right).$$

g) Combining the expansions for $E(H)$, $M(H)$, $V(H)$, we finally have

$$\begin{aligned}
&\int\int \hat{\lambda}_{11}^2(x_1, x_2, H)\,\mathrm{d}x_1\mathrm{d}x_2 \\
=\; &\int\int \lambda_{11}^2(x_1, x_2)\,\mathrm{d}x_1\mathrm{d}x_2 + V_K \int\int \lambda_{11}(x_1, x_2) \sum_{i=1}^{2} \beta_i^2 \frac{\partial^2}{\partial x_i^2} \lambda_{11}(x_1, x_2)\,\mathrm{d}x_1\mathrm{d}x_2 b_N^2 \\
&+ o\left(b_N^2\right) + o_p\left(b_N^2 N^{-\gamma}\right) + \frac{1}{N\beta_1^5\beta_2 b_N^6} \int\int K_{11}^2(u)\,\mathrm{d}u_1\mathrm{d}u_2 \\
&+ O_p\left(\frac{1}{N b_N^5}\right) + o_p\left(\frac{N^{-\gamma}}{N b_N^6}\right) + o_p\left(\frac{\log N}{N^{\frac{3}{2}} b_N^6}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right).
\end{aligned}$$

h) The asymptotic expansion of the integral of $\hat{\lambda}_{22}^2(x_1, x_2, H)$ is obviously the same as in g) with 1 and 2 exchanged. The decomposition of the mixed term is

$$\int\int \hat{\lambda}_{11}(x_1, x_2, H)\, \hat{\lambda}_{22}(x_1, x_2, H)\,\mathrm{d}x_1\mathrm{d}x_2 = E'(H) - M'(H) + V'(H)$$

with

$$
\begin{aligned}
E^{'}(H) &= \int\int \nu_1(x_1,x_2,H)\,\nu_2(x_1,x_2,H)\,\mathrm{d}x_1\mathrm{d}x_2 \\
M^{'}(H) &= \int\int \nu_1(x_1,x_2,H)\left\{\hat{\lambda}_{22}(x_1,x_2,H)-\nu_2(x_1,x_2,H)\right\}\mathrm{d}x_1\mathrm{d}x_2 \\
&\quad +\int\int \nu_2(x_1,x_2,H)\left\{\hat{\lambda}_{11}(x_1,x_2,H)-\nu_1(x_1,x_2,H)\right\}\mathrm{d}x_1\mathrm{d}x_2 \\
V^{'}(H) &= \int\int \left\{\hat{\lambda}_{11}(x_1,x_2,H)-\nu_1(x_1,x_2,H)\right\}\left\{\hat{\lambda}_{22}(x_1,x_2,H)-\nu_2(x_1,x_2,H)\right\}\mathrm{d}x_1\mathrm{d}x_2.
\end{aligned}
$$

Using a), we get analogously to b)

$$
\begin{aligned}
&E^{'}(H) \\
=\;& \int\int \lambda_{11}(x_1,x_2)\,\lambda_{22}(x_1,x_2)\,\mathrm{d}x_1\mathrm{d}x_2 \\
&+\frac{b_N^2}{2}V_K\int\int\sum_{i=1}^{2}\beta_i^2\left\{\lambda_{11}(x_1,x_2)\frac{\partial^2}{\partial x_i^2}\lambda_{22}(x_1,x_2)+\lambda_{22}(x_1,x_2)\frac{\partial^2}{\partial x_i^2}\lambda_{11}(x_1,x_2)\right\}\mathrm{d}x_1\mathrm{d}x_2 \\
&+o\left(b_N^2\right)+o_p\left(b_N^2N^{-\gamma}\right).
\end{aligned}
$$

Using $\nu_i(x_1,x_2,B)=\mathbb{E}\hat{\lambda}_{ii}(x_1,x_2,B)$

$$
\begin{aligned}
&\mathbb{E}V^{'}(B) \\
=\;& \frac{1}{N^2b_1^4b_2^4}\int\int\sum_{i,j=1}^{N}\operatorname{cov}\left(K_{11}\left(\frac{x_1-X_{i1}}{b_1},\frac{x_2-X_{i2}}{b_2}\right),K_{22}\left(\frac{x_1-X_{j1}}{b_1},\frac{x_2-X_{j2}}{b_2}\right)\right)\mathrm{d}x_1\mathrm{d}x_2 \\
=\;& \frac{1}{Nb_1^4b_2^4}\int\int\operatorname{cov}\left(K_{11}\left(\frac{x_1-X_{11}}{b_1},\frac{x_2-X_{12}}{b_2}\right),K_{22}\left(\frac{x_1-X_{11}}{b_1},\frac{x_2-X_{12}}{b_2}\right)\right)\mathrm{d}x_1\mathrm{d}x_2
\end{aligned}
$$

as X_i, X_j are independent for $i \neq j$. Similar to c), we conclude

$$
\mathbb{E}V^{'}(B)=\frac{1}{Nb_1^3b_2^3}\int\int K_{11}(u_1,u_2)\,K_{22}(u_1,u_2)\,\mathrm{d}u_1\mathrm{d}u_2+O\left(\frac{1}{Nb_N^5}\right).
$$

We decompose as in d)

$$
V^{'}(B)=\frac{1}{Nb_1^3b_2^3}W_N^{'}(B)+\frac{N-1}{2Nb_1^4b_2^4}U_N^{'}(B)
$$

with

$$
\begin{aligned}
W_N^{'}(B) &= \frac{1}{N}\sum_{j=1}^{N}Q_B(X_j)\,,U_N^{'}(B)=\frac{2}{N(N-1)}\sum_{i\neq j}R_B^{'}(X_i,X_j) \\
Q_B^{'}(X_j) &= \frac{1}{b_1b_2}\int\int\left\{K_{11}\left(\frac{x_1-X_{j1}}{b_1},\frac{x_2-X_{j2}}{b_2}\right)-b_1^3b_2\nu_1(x_1,x_2,B)\right\} \\
&\quad \left\{K_{22}\left(\frac{x_1-X_{j1}}{b_1},\frac{x_2-X_{j2}}{b_2}\right)-b_2^3b_1\nu_2(x_1,x_2,B)\right\}\mathrm{d}x_1\mathrm{d}x_2 \\
R_B^{'}(X_i,X_j) &= \int\int\left\{K_{11}\left(\frac{x_1-X_{i1}}{b_1},\frac{x_2-X_{i2}}{b_2}\right)-b_1^3b_2\nu_1(x_1,x_2,B)\right\} \\
&\quad \left\{K_{22}\left(\frac{x_1-X_{j1}}{b_1},\frac{x_2-X_{j2}}{b_2}\right)-b_2^3b_1\nu_2(x_1,x_2,B)\right\}\mathrm{d}x_1\mathrm{d}x_2.
\end{aligned}
$$

As in d), we conclude that $Q_B^{'}(X_j)$ is bounded, and, then, from Hoeffding's inequality

$$\frac{1}{Nb_1^3 b_2^3}\left(W_N^{'}(B) - \mathbb{E}W_N^{'}(B)\right) = o_p\left(\frac{\log N}{N^{\frac{3}{2}} b_1^3 b_2^3}\right) = o_p\left(\frac{\log N}{N^{\frac{3}{2}} b_N^6}\right).$$

For the second component, we decompose

$$\begin{aligned} R_B^{'}(X_1, X_2) &= b_1 b_2 \left\{L_2^{'}\left(B^{-1}(X_1 - X_2)\right) - \mathbb{E}L_2^{'}\left(B^{-1}(X_1 - X_2)\right)\right\} \\ &\quad - R_B^{'}(X_1) - R_B^{''}(X_2) \end{aligned}$$

with

$$\begin{aligned} L_2^{'}(u) &= \int\int K_{11}(u) K_{22}(v-u)\, \mathrm{d}v_1 \mathrm{d}v_2, u \in \mathbb{R}^2, \\ R_B^{'}(u) &= \int\int b_2^3 b_1 \nu_2(x_1, x_2, B)\left\{K_{11}\left(\frac{x_1 - u_1}{b_1}, \frac{x_2 - u_2}{b_2}\right) - b_1^3 b_2 \nu_1(x_1, x_2, B)\right\} \mathrm{d}x_1 \mathrm{d}x_2 \\ &= \int\int b_2^4 b_1^2 \lambda_{22}(u + Bz) K_{11}(z)\, \mathrm{d}z_1 \mathrm{d}z_2 + O\left(b_N^8\right), \\ R_B^{''}(u) &= \int\int b_1^3 b_2 \nu_1(x_1, x_2, B)\left\{K_{22}\left(\frac{x_1 - u_1}{b_1}, \frac{x_2 - u_2}{b_2}\right) - b_2^3 b_1 \nu_2(x_1, x_2, B)\right\} \mathrm{d}x_1 \mathrm{d}x_2 \\ &= \int\int b_1^4 b_2^2 \lambda_{11}(u + Bz) K_{22}(z)\, \mathrm{d}z_1 \mathrm{d}z_2 + O\left(b_N^8\right). \end{aligned}$$

By a Taylor expansion of λ_{22} respectively λ_{11}, we conclude as in d), that $\frac{1}{b_1^4 b_2^4}\left(R_B^{'}(u) + R_B^{''}(u)\right)$ is bounded, and, using that its mean is 0 and Hoeffding's inequality,

$$\frac{1}{b_1^4 b_2^4} U_N^{'}(B) = \frac{1}{b_1^4 b_2^4} \tilde{U}_N^{'}(B) + o_p\left(\frac{\log N}{\sqrt{N}}\right),$$

where $\tilde{U}_N^{'}(B)$ is the U-statistic with kernel

$$\Lambda_B^{'}(x - z) = b_1 b_2 \left\{L_2^{'}\left(B^{-1}(x - z)\right) - \mathbb{E}L_2^{'}\left(B^{-1}(X_1 - X_2)\right)\right\}.$$

An analogous result to Lemma 6 with $L_2^{'}$ replacing L_2 follows by exactly the same arguments, such that we conclude as in d)

$$\frac{1}{b_1^4 b_2^4} U_N^{'}(B) = O_p\left(\frac{1}{N b_N^4}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right)$$

and, therefore,

$$V^{'}(B)-\mathbb{E}V^{'}(B)=O_p\left(\frac{1}{Nb_N^4}\right)+o_p\left(\frac{\log N}{\sqrt{N}}\right)+o_p\left(\frac{\log N}{N^{\frac{3}{2}}b_N^6}\right).$$

For random bandwidths h_1, h_2, we decompose $V^{'}(H)$ as $V^{'}(H)=\bar{V}^{'}(H)+\left(V^{'}(H)-\bar{V}^{'}(H)\right)$ with

$$\begin{aligned}
&\bar{V}^{'}(H)\\
=&\frac{1}{Nh_1^4h_2^4}\int\int\int\int K_{11}\left(H^{-1}(z-u)\right)K_{22}\left(H^{-1}(z-u)\right)\lambda(u)\,\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}z_1\mathrm{d}z_2\\
&-\frac{1}{N}\int\int\nu_1(x_1,x_2,H)\,\nu_2(x_1,x_2,H)\,\mathrm{d}x_1\mathrm{d}x_2\\
=&\frac{1}{Nh_1^3h_2^3}\int\int K_{11}(u_1,u_2)\,K_{22}(u_1,u_2)\,\mathrm{d}u_1\mathrm{d}u_2+o_p\left(\frac{N^{-\gamma}}{Nb_N^6}\right)+O_p\left(\frac{1}{Nb_N^5}\right)
\end{aligned}$$

by the same argument as in e). $V_0^{'}(H)=V^{'}(H)-\bar{V}^{'}(H)$ is decomposed as

$$V_0^{'}(H)=V_0^{'}(B_N)+\left(V_0^{'}(H)-V_0^{'}(B_N)\right)$$

and it can be shown exactly as in e) that the second term has the same order as S_2 in e). The order of the first term $V_0^{'}(B_N)=V^{'}(B_N)-\mathbb{E}V^{'}(B_N)$ has already been given above as B_N is deterministic. We finally conclude

$$\begin{aligned}
V^{'}(H)\;=\;&\frac{1}{N\beta_1^3\beta_2^3b_N^6}\int\int K_{11}(u)\,K_{22}(u)\,\mathrm{d}u_1\mathrm{d}u_2+O_p\left(\frac{1}{Nb_N^5}\right)+o_p\left(\frac{N^{-\gamma}}{Nb_N^6}\right)\\
&+o_p\left(\frac{\log N}{N^{\frac{3}{2}}b_N^6}\right)+o_p\left(\frac{\log N}{\sqrt{N}}\right).
\end{aligned}$$

Finally, $M^{'}(H)=o_p\left(\frac{\log N}{\sqrt{N}}\right)$ can be shown analogously to f). Combining the expansions of $E^{'}(H)$, $M^{'}(H)$, $V^{'}(H)$ we have

$$\begin{aligned}
&\int\int\hat{\lambda}_{11}(x_1,x_2,H)\,\hat{\lambda}_{22}(x_1,x_2,H)\,\mathrm{d}x_1\mathrm{d}x_2\\
=&\int\int\lambda_{11}(x_1,x_2)\,\lambda_{22}(x_1,x_2)\,\mathrm{d}x_1\mathrm{d}x_2+\frac{1}{N\beta_1^3\beta_2^3b_N^6}\int\int K_{11}(u)\,K_{22}(u)\,\mathrm{d}u_1\mathrm{d}u_2\\
&+\frac{b_N^2}{2}V_K\int\int\sum_{i=1}^{2}\beta_i^2\left\{\lambda_{11}(x_1,x_2)\frac{\partial^2}{\partial x_i^2}\lambda_{22}(x_1,x_2)+\lambda_{22}(x_1,x_2)\frac{\partial^2}{\partial x_i^2}\lambda_{11}(x_1,x_2)\right\}\mathrm{d}x_1\mathrm{d}x_2\\
&+o\left(b_N^2\right)+o_p\left(b_N^2N^{-\gamma}\right)+O_p\left(\frac{1}{Nb_N^5}\right)+o_p\left(\frac{N^{-\gamma}}{Nb_N^6}\right)\\
&+o_p\left(\frac{\log N}{N^{\frac{3}{2}}b_N^6}\right)+o_p\left(\frac{\log N}{\sqrt{N}}\right).
\end{aligned}$$

□

Lemma 6. *Let the assumptions of Theorem 4 be fulfilled, and set*

$$L_2(u) = \int\int K_{11}(v) K_{11}(v-u) \, \mathrm{d}v_1 \mathrm{d}v_2.$$

a) For $i, j, k = 1, 2$

$$\int\int L_2(u) \, \mathrm{d}u_1 \mathrm{d}u_2 = 0,$$

$$\int\int u_i L_2(u) \, \mathrm{d}u_1 \mathrm{d}u_2 = 0,$$

$$\int\int u_i u_j L_2(u) \, \mathrm{d}u_1 \mathrm{d}u_2 = 0,$$

$$\int\int u_i u_j u_k L_2(u) \, \mathrm{d}u_1 \mathrm{d}u_2 = 0.$$

b) $\mathbb{E} L_2 \left(B^{-1}(X_1 - X_2)\right) = O\left(b_M^6\right)$ *with* $b_M = \max(b_1, b_2)$.

c) $\mathbb{E} L_2^2 \left(B^{-1}(X_1 - X_2)\right) = O\left(b_1^2 b_2^2\right)$.

Proof. a) Recall that from our assumptions

$$\int\int K_{11}(u) \, \mathrm{d}u_1 \mathrm{d}u_2 = 0$$

and

$$\int\int u_i K(u) \, \mathrm{d}u_1 \mathrm{d}u_2 = 0, i = 1, 2.$$

Substituting $w = u - v$, we then have

$$\begin{aligned} \int\int L_2(u) \, \mathrm{d}u_1 \mathrm{d}u_2 &= \int\int\int\int K_{11}(v) K_{11}(v-u) \, \mathrm{d}v_1 \mathrm{d}v_2 \mathrm{d}u_1 \mathrm{d}u_2 \\ &= \int\int K_{11}(v) \, \mathrm{d}v_1 \mathrm{d}v_2 \int\int K_{11}(w) \, \mathrm{d}w_1 \mathrm{d}w_2 = 0 \end{aligned}$$

$$\int\int u_i L_2(u) \, \mathrm{d}u_1 \mathrm{d}u_2 = \int\int\int\int (v_i + w_i) K_{11}(v) K_{11}(w) \, \mathrm{d}v_1 \mathrm{d}v_2 \mathrm{d}w_1 \mathrm{d}w_2 = 0$$

$$\begin{aligned}\int\int u_i u_j L_2(u)\,\mathrm{d}u_1\mathrm{d}u_2 &= \int\int v_i v_j K_{11}(v)\,\mathrm{d}v_1\mathrm{d}v_2 \int\int K_{11}(w)\,\mathrm{d}w_1\mathrm{d}w_2 \\ &\quad + \int\int w_i w_j K_{11}(w)\,\mathrm{d}w_1\mathrm{d}w_2 \int\int K_{11}(v)\,\mathrm{d}v_1\mathrm{d}v_2 \\ &\quad + \int\int v_i K_{11}(v)\,\mathrm{d}v_1\mathrm{d}v_2 \int\int w_j K_{11}(w)\,\mathrm{d}w_1\mathrm{d}w_2 \\ &\quad + \int\int v_j K_{11}(v)\,\mathrm{d}v_1\mathrm{d}v_2 \int\int w_i K_{11}(w)\,\mathrm{d}w_1\mathrm{d}w_2 \\ &= 0.\end{aligned}$$

The last relationship is shown analogously.

b) Substituting $u_1 = \frac{x_1 - z_1}{b_1}$, $u_2 = \frac{x_2 - z_2}{b_2}$, $u = (u_1, u_2)^\top$, we have

$$\begin{aligned}\mathbb{E}L_2\left(B^{-1}(X_1 - X_2)\right) &= \int\int\int\int L_2\left(B^{-1}(x-z)\right)\lambda(x)\lambda(z)\,\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}z_1\mathrm{d}z_2 \\ &= b_1 b_2 \int\int\int\int L_2(u)\lambda(z+Bu)\,\mathrm{d}u_1\mathrm{d}u_2\lambda(z)\,\mathrm{d}z_1\mathrm{d}z_2.\end{aligned}$$

For a multi-index $\alpha = (\alpha_1, \alpha_2)$, $\alpha_1, \alpha_2 \geq 0$, we use the common notation

$$|\alpha| = \alpha_1 + \alpha_2, \alpha! = \alpha_1!\alpha_2!, u^\alpha = u_1^{\alpha_1} u_2^{\alpha_2} \quad \text{for } u \in \mathbb{R}^2,$$

$$D_g^\alpha = \frac{\partial^{|\alpha|}}{\partial u_1^{\alpha_1}\partial u_2^{\alpha_2}} g \quad \text{for a function } g : \mathbb{R}^2 \to \mathbb{R}.$$

Then, we have the Taylor expansion of $\lambda(z + Bu)$

$$\lambda(z+Bu) = \lambda(z) + \sum_{\ell=1}^{3}\sum_{|\alpha|=\ell}\frac{1}{\alpha!}D^\alpha\lambda(z)(Bu)^\alpha + O\left(b_M^4\right).$$

Using $(Bu)^\alpha = b_1^{\alpha_1}u_1^{\alpha_1}b_2^{\alpha_2}u_2^{\alpha_2} = b_1^{\alpha_1}b_2^{\ell-\alpha_1}u_1^{\alpha_1}u_2^{\ell-\alpha_1}$ for $|\alpha| = \ell$, the integrals of the first four terms in the Taylor expansion multiplied with $L_2(u)$ vanish by a). Hence

$$\mathbb{E}L_2\left(B^{-1}(X_1 - X_2)\right) = b_1 b_2 \int\int \lambda(z)\,\mathrm{d}z_1\mathrm{d}z_2 O\left(b_M^4\right) = O\left(b_M^6\right).$$

c) Substituting $u_1 = \frac{x_1}{b_1}$, $u_2 = \frac{x_2}{b_2}$, $v_1 = \frac{z_1}{b_1}$, $v_2 = \frac{z_2}{b_2}$, we have

$$\begin{aligned}\mathbb{E}L_2^2\left(B^{-1}(X_1 - X_2)\right) &= \int\int\int\int L_2^2\left(B^{-1}(x-z)\right)\lambda(x)\lambda(z)\,\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}z_1\mathrm{d}z_2 \\ &= b_1^2 b_2^2 \int\int\int\int L_2^2(u-v)\lambda(Bu)\lambda(Bv)\,\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}v_1\mathrm{d}v_2 \\ &= O\left(b_1^2 b_2^2\right)\end{aligned}$$

as λ is bounded and L_2 has bounded support.

□

Lemma 7. *Under the assumptions of Theorem 4,*

$$\left|V(B)-V\left(\bar{B}\right)\right| \leq L \cdot \frac{\left\|b-\bar{b}\right\|}{b_m^{10}}, b, \bar{b} \in (0,1]^2$$

for some suitable constant $L > 0$ *and* $b_m = \min\left(b_1, b_2, \bar{b}_1, \bar{b}_2\right)$.

Proof. For abbreviation, we write

$$\begin{aligned} V(B) &= \int\int\left(\frac{1}{b_1^3 b_2} G(x_1, x_2, B)\right)^2 \mathrm{d}x_1 \mathrm{d}x_2, \\ G(x_1, x_2, B) &= \frac{1}{N}\sum_{j=1}^{N} g_j(x_1, x_2, B), \\ g_j(x_1, x_2, B) &= K_{11}\left(\frac{x_1 - X_{j1}}{b_1}, \frac{x_2 - X_{j2}}{b_2}\right) - b_1^3 b_2 \nu_1(x_1, x_2, B). \end{aligned}$$

From Lipschitz continuity of λ, using C for generic constants with varying values and $b_M = \max\left(b_1, b_2, \bar{b}_1, \bar{b}_2\right)$,

$$\begin{aligned} & \left|b_1^3 b_2 \nu_1(x_1, x_2, B) - \bar{b}_1^3 \bar{b}_2 \nu_1\left(x_1, x_2, \bar{B}\right)\right| \\ = & \left| b_1 b_2 \int\int K_{11}(u_1, u_2)\, \lambda(x_1 - b_1 u_1, x_2 - b_2 u_2)\, \mathrm{d}u_1 \mathrm{d}u_2 \right. \\ & \left. - \bar{b}_1 \bar{b}_2 \int\int K_{11}(u_1, u_2)\, \lambda\left(x_1 - \bar{b}_1 u_1, x_2 - \bar{b}_2 u_2\right) \mathrm{d}u_1 \mathrm{d}u_2 \right| \\ \leq & \ C\left|b_1 b_2 - \bar{b}_1 \bar{b}_2\right| \\ & + \bar{b}_1 \bar{b}_2 \int\int \left|K_{11}(u_1, u_2)\right| \left|\lambda(x_1 - b_1 u_1, x_2 - b_2 u_2) - \lambda\left(x_1 - \bar{b}_1 u_1, x_2 - \bar{b}_2 u_2\right)\right| \mathrm{d}u_1 \mathrm{d}u_2 \\ \leq & \ C \cdot b_M\left(\left|b_1 - \bar{b}_1\right| + \left|b_2 - \bar{b}_2\right|\right) + C \cdot b_M^2 \left\|b - \bar{b}\right\| \leq C \cdot b_M \left\|b - \bar{b}\right\| \end{aligned}$$

as K_{11} has bounded support, and as $b_M \leq 1$. From Lipschitz continuity of

K_{11},

$$
\begin{aligned}
& \left| K_{11}\left(\frac{x_1 - X_{j1}}{b_1}, \frac{x_2 - X_{j2}}{b_2}\right) - K_{11}\left(\frac{x_1 - X_{j1}}{\bar{b}_1}, \frac{x_2 - X_{j2}}{\bar{b}_2}\right)\right|^2 \\
\leq \; & C^2\left(\left|\frac{x_1 - X_{j1}}{b_1} - \frac{x_1 - X_{j1}}{\bar{b}_1}\right|^2 + \left|\frac{x_2 - X_{j2}}{b_2} - \frac{x_2 - X_{j2}}{\bar{b}_2}\right|^2\right) \\
\leq \; & C^2 \cdot \frac{1}{b_m^4}\left(\left(b_1 - \bar{b}_1\right)^2 + \left(b_2 - \bar{b}_2\right)^2\right) \\
= \; & \frac{C^2}{b_m^4}\left\|b - \bar{b}\right\|^2
\end{aligned}
$$

as $x_1, x_2, X_{j1}, X_{j2} \in [0,1]$. Hence, uniformly in j, x_1, x_2, as $b_M \leq 1 \leq \frac{1}{b_m^2}$

$$
\begin{aligned}
\left|g_j\left(x_1, x_2, B\right) - g_j\left(x_1, x_2, \bar{B}\right)\right| & \leq C \cdot \left(b_M + \frac{1}{b_m^2}\right)\left\|b - \bar{b}\right\| \\
& \leq C \cdot \frac{1}{b_m^2}\left\|b - \bar{b}\right\|
\end{aligned}
$$

and, then,

$$
\left|G\left(x_1, x_2, B\right) - G\left(x_1, x_2, \bar{B}\right)\right| \leq C \cdot \frac{1}{b_m^2}\left\|b - \bar{b}\right\|.
$$

As K_{11}, ν_1 are bounded, $G\left(x_1, x_2, B\right)$ is bounded too, and we get

$$
\left|G^2\left(x_1, x_2, B\right) - G^2\left(x_1, x_2, \bar{B}\right)\right| \leq C \cdot \frac{1}{b_m^2}\left\|b - \bar{b}\right\|
$$

too. Finally, we have

$$
\begin{aligned}
\left|V(B) - V\left(\bar{B}\right)\right| & \leq \frac{1}{b_m^8}\int\int\left|G^2\left(x_1, x_2, B\right) - G^2\left(x_1, x_2, \bar{B}\right)\right| \mathrm{d}x_1 \mathrm{d}x_2 \\
& \quad + \left|\frac{1}{b_1^6 b_2^2} - \frac{1}{\bar{b}_1^6 \bar{b}_2^2}\right| \int\int G^2\left(x_1, x_2, \bar{B}\right) \mathrm{d}x_1 \mathrm{d}x_2 \\
& \leq \frac{C}{b_m^8} \cdot \frac{\left\|b - \bar{b}\right\|}{b_m^2} + C \cdot \frac{\left\|b - \bar{b}\right\|}{b_m^9} \\
& \leq C \cdot \frac{\left\|b - \bar{b}\right\|}{b_m^{10}},
\end{aligned}
$$

where we use for $i = 1, 2$, $\ell = 6$ or 2

$$
\begin{aligned}
\left|\frac{1}{b_i^\ell - \bar{b}_i^\ell}\right| & = \frac{\left|\bar{b}_i^\ell - b_i^\ell\right|}{b_i^\ell \bar{b}_i^\ell} = \frac{O\left(\max\left(b_i^{\ell-1}, \bar{b}_i^{\ell-1}\right)\right)\left|\bar{b}_i - b_i\right|}{b_i^\ell \bar{b}_i^\ell} \\
& \leq \frac{1}{b_m^{\ell+1}}\left|b_i - \bar{b}_i\right|,
\end{aligned}
$$

and the boundedness of G^2. □

Corollary 11. *Under the assumptions of Lemma 7,*

$$\left|M(B)-M\left(\bar{B}\right)\right| \leq L\cdot\frac{\left\|b-\bar{b}\right\|}{b_m^{10}} \quad \textit{with } b_m=\min\left(b_1,b_2\right)$$

for some suitable $L>0$.

Proof. From the proof of Proposition 3, a), $\nu_1\left(x_1,x_2,B\right)$ is uniformly bounded in x_1,x_2,B. We write with G as in the proof of Lemma 7

$$M(B)=\frac{1}{b_1^6b_2^2}\int\int G\left(x_1,x_2,B\right)b_1^3b_2\nu_1\left(x_1,x_2,B\right)\mathrm{d}x_1\mathrm{d}x_2$$

such that

$$\begin{aligned}
& \left|M(B)-M\left(\bar{B}\right)\right| \\
\leq\ & \frac{1}{b_m^8}\int\int\left|G\left(x_1,x_2,B\right)-G\left(x_1,x_2,\bar{B}\right)\right|b_1^3b_2\left|\nu_1\left(x_1,x_2,B\right)\right|\mathrm{d}x_1\mathrm{d}x_2 \\
& +\frac{1}{b_m^8}\int\int\left|G\left(x_1,x_2,\bar{B}\right)\right|\left|b_1^3b_2\nu_1\left(x_1,x_2,B\right)-\bar{b}_1^3\bar{b}_2\nu_1\left(x_1,x_2,\bar{B}\right)\right|\mathrm{d}x_1\mathrm{d}x_2 \\
\leq\ & C\cdot\frac{\left\|b-\bar{b}\right\|}{b_m^{10}}+C\cdot\frac{b_M\left\|b-\bar{b}\right\|}{b_m^8}\leq L\cdot\frac{\left\|b-\bar{b}\right\|}{b_m^{10}}
\end{aligned}$$

using the inequalities for G and $b_1^3b_2\nu_1$ from the proof of Lemma 7 and $b_m\leq 1$, $b_M=\max\left(b_1,b_2\right)\leq 1$. □

Lemma 8. *Let the assumptions of Theorem 4 be satisfied. Let the bandwidths $b_1,b_2,\bar{b}_1,\bar{b}_2$ satisfy $0<\alpha_N<b_i,\bar{b}_i<a_N$, $i=1,2$, for some $\alpha_N=a_N\left(1-O\left(N^{-\gamma}\right)\right)$ and a_N satisfying $a_N^4=O\left(N^{-\gamma}\right)$ and $Na_N\to\infty$. Then, for $q\geq 1$*

$$\mathbb{E}\left(U_N(B)-U_N\left(\bar{B}\right)\right)^{2q}=\frac{a_N^{6q}}{N^{2q}}O\left(N^{-2q\gamma}\right),$$

where $U_N(B)$ is defined in the proof of Theorem 4, d).

Proof. The proof is analogous to that of Lemma 5, such that we only discuss the differences. Note that again

$$U_N(B)=\frac{2}{N(N-1)}\sum_{i\neq j}R_B\left(X_i,X_j\right)$$

is a U-statistic with symmetric kernel $R_B(x,z)$, $x,z \in \mathbb{R}^2$, and $\mathbb{E}R_B(X_1,X_2) = 0$, and, more generally,

$$\mathbb{E}\{R_B(X_1,X_2)|X_2\} = 0 = \mathbb{E}\{R_B(X_1,X_2)|X_1\}$$

which is crucial for the argument.

a) In the two-dimensional case, we have

$$L_2(u) = \int\int K_{11}(v) K_{11}(v-u)\,\mathrm{d}v_1\mathrm{d}v_2,$$

and substituting $z_1 = \frac{x_1-v_1}{b_1}$, $z_2 = \frac{x_2-v_2}{b_2}$

$$\begin{aligned} & \int\int K_{11}\left(\frac{x_1-u_1}{b_1},\frac{x_2-u_2}{b_2}\right) K_{11}\left(\frac{x_1-v_1}{b_1},\frac{x_2-v_2}{b_2}\right)\mathrm{d}x_1\mathrm{d}x_2 \\ = & \; b_1b_2\int\int K_{11}(z_1,z_2) K_{11}\left(z_1-\frac{u_1-v_1}{b_1}, z_2-\frac{u_2-v_2}{b_2}\right)\mathrm{d}z_1\mathrm{d}z_2 \\ = & \; b_1b_2L_2\left(B^{-1}(u-v)\right). \end{aligned}$$

From the independence of X_1, X_2, we get

$$\begin{aligned} & b_1b_2\mathbb{E}L_2\left(B^{-1}(X_1-X_2)\right) \\ = & \int\int \mathbb{E}K_{11}\left(\frac{x_1-X_{11}}{b_1},\frac{x_2-X_{12}}{b_2}\right)\mathbb{E}K_{11}\left(\frac{x_1-X_{21}}{b_1},\frac{x_2-X_{22}}{b_2}\right)\mathrm{d}x_1\mathrm{d}x_2 \\ = & \; b_1^6b_2^2\int\int \nu_1^2(x_1,x_2,B)\,\mathrm{d}x_1\mathrm{d}x_2. \end{aligned}$$

Setting

$$\Lambda_B(u-v) = b_1b_2\left\{L_2\left(B^{-1}(u-v)\right) - \mathbb{E}L_2\left(B^{-1}(X_1-X_2)\right)\right\},$$

we decompose

$$R_B(X_1,X_2) = \Lambda_B(X_1-X_2) - R_B(X_1) - R_B(X_2),$$

where $R_B(u)$ is defined as in the proof of Theorem 4, d), again.

b) With $b, \bar{b}$ denoting the bandwidth vectors with coordinates b_i respectively $\bar{b}_i$, $i = 1, 2$,

$$\begin{aligned}
\left\|B^{-1}z - \bar{B}^{-1}z\right\|^2 &= z_1^2\left(\frac{1}{b_1} - \frac{1}{\bar{b}_1}\right)^2 + z_2^2\left(\frac{1}{b_2} - \frac{1}{\bar{b}_2}\right)^2 \\
&\leq \|z\|^2\left\{\left(\frac{\bar{b}_1 - b_1}{b_1\bar{b}_1}\right)^2 + \left(\frac{\bar{b}_2 - b_2}{b_2\bar{b}_2}\right)^2\right\} \\
&\leq \frac{\|z\|^2}{\alpha_N^4}\left\|b - \bar{b}\right\|^2
\end{aligned}$$

using $b_i, \bar{b}_i \geq \alpha_N$. Then, using Lipschitz continuity of K_{11}, we have as in the proof of Lemma 5, b) for some constant $\tilde{c} > 0$

$$\left|b_1b_2L_2\left(B^{-1}\left(X_1 - X_2\right)\right) - \bar{b}_1\bar{b}_2L_2\left(\bar{B}^{-1}\left(X_1 - X_2\right)\right)\right| \leq 2\tilde{c}\frac{a_N^2}{\alpha_N^2}\left\|b - \bar{b}\right\|.$$

As the support of L_2 is $[-2, +2]^2$, $L_2\left(B^{-1}\left(X_1 - X_2\right)\right) = 0$ if $B^{-1}\left(X_1 - X_2\right) \notin [-2, +2]^2$, and, as $b_i, \bar{b}_i \leq a_N, i = 1, 2$, we conclude

$$\begin{aligned}
&\left|b_1b_2L_2\left(B^{-1}\left(X_1 - X_2\right)\right) - \bar{b}_1\bar{b}_2L_2\left(\bar{B}^{-1}\left(X_1 - X_2\right)\right)\right| \\
\leq\; & 2\tilde{c}\frac{a_N^2}{\alpha_N^2}\left\|b - \bar{b}\right\|\mathbf{1}_{[-2,+2]^2}\left(\frac{X_1 - X_2}{a_N}\right).
\end{aligned}$$

As λ is bounded,

$$\begin{aligned}
&\mathbb{E}\mathbf{1}_{[-2,+2]^2}\left(\frac{X_1 - X_2}{a_N}\right) \\
=\; & \int\int\int\int \mathbf{1}_{[-2,+2]}\left(\frac{u_1 - v_1}{a_N}\right)\mathbf{1}_{[-2,+2]}\left(\frac{u_2 - v_2}{a_N}\right)\lambda(u)\lambda(v)\,\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}v_1\mathrm{d}v_2 \\
=\; & O\left(a_N^4\right)
\end{aligned}$$

by substituting $\frac{u_i}{a_N} = z_i$, $\frac{v_i}{a_N} = w_i$, $i = 1, 2$, and we have

$$\left|\Lambda_B\left(X_1 - X_2\right) - \Lambda_{\bar{B}}\left(X_1 - X_2\right)\right| \leq 2\tilde{c}\frac{a_N^2}{\alpha_N^2}\left\|b - \bar{b}\right\|\left\{\mathbf{1}_{[-2,+2]^2}\left(\frac{X_1 - X_2}{a_N}\right) + O\left(a_N^4\right)\right\}.$$

c) From the proof of Proposition 6, a), we have

$$\begin{aligned}
\left|\nu_1\left(x_1, x_2, B\right) - \nu_1\left(x_1, x_2, \bar{B}\right)\right| &\leq \frac{1}{2}V_K\sum_{i=1}^{2}\left|b_i^2 - \bar{b}_i^2\right|\left|\frac{\partial^2}{\partial x_i^2}\lambda_{11}\left(x_1, x_2\right)\right| + o\left(a_N^2\right) \\
&\leq ca_N\left\|b - \bar{b}\right\| + o\left(a_N^2\right)
\end{aligned}$$

for some $c > 0$, where we have used $\left|b_i + \bar{b}_i\right| = O(a_N), i = 1, 2,$ and $\left|b_i - \bar{b}_i\right|^2 \leq \left\|b - \bar{b}\right\|^2, i = 1, 2$. To get an upper bound for $|R_B(X_1) - R_{\bar{B}}(X_1)|$ we get as in the proof of Lemma 5, with $x = (x_1, x_2)^\top$,

$$\begin{aligned}
& \int \left|b_1^3 b_2 \nu_1(x_1, x_2, B) K_{11}\left(B^{-1}(x - X_1)\right) - \bar{b}_1^3 \bar{b}_2 \nu_1\left(x_1, x_2, \bar{B}\right) K_{11}\left(\bar{B}^{-1}(x - X_1)\right)\right| \mathrm{d}x \\
\leq \ & \tilde{c} a_N^7 \left\|b - \bar{b}\right\| + o\left(a_N^8\right) + \bar{c} \frac{a_N^4}{\alpha_N^2} \left\|b - \bar{b}\right\| \\
= \ & \bar{c} \frac{a_N^4}{\alpha_N^2} \left\|b - \bar{b}\right\| + o\left(a_N^8\right)
\end{aligned}$$

for suitable $\tilde{c}, c > 0$, and, then, also for some $c > 0$, using $a_N^4 = O\left(N^{-\gamma}\right)$,

$$\begin{aligned}
|R_B(X_1) - R_{\bar{B}}(X_1)| &\leq c \frac{a_N^4}{\alpha_N^2} \left\|b - \bar{b}\right\| + o\left(a_N^8\right) \\
&= c\left(1 + O\left(N^{-\gamma}\right)\right) a_N^2 \left\|b - \bar{b}\right\| + o\left(a_N^4 N^{-\gamma}\right).
\end{aligned}$$

Combining this with the inequality for $\Lambda_B(X_1 - X_2)$ from b), we get

$$|R_B(X_1, X_2) - R_{\bar{B}}(X_1, X_2)| = a_N O\left(N^{-\gamma}\right) \left\{\mathbf{1}_{[-2,+2]^2}\left(\frac{X_1 - X_2}{a_N}\right) + O\left(a_N^3\right)\right\}.$$

d) Writing as in the proof of Lemma 5 with $\rho_N = \frac{2}{N(N-1)}$,

$$U_N(B) - U_N\left(\bar{B}\right) = \rho_N \sum_{i \neq j} D_{ij},$$

then, we get by the same argument as in the proof of Lemma 5

$$\mathbb{E}\left(U_N(B) - U_N\left(\bar{B}\right)\right)^{2q} \leq \rho_N^{2q} \sum_{m=2}^{2q} N^m a_N^{2q} O\left(N^{-2q\gamma}\right) \eta_m,$$

where η_m is an upper bound on

$$\mathbb{E} \prod_{\ell=1}^{2q} \left\{\mathbf{1}_{[-2,+2]^2}\left(\frac{X_{i_\ell} - X_{j_\ell}}{a_N}\right) + O\left(a_N^3\right)\right\}.$$

As from b),

$$\mathbb{E}\mathbf{1}_{[-2,+2]^2}\left(\frac{X_{i_\ell} - X_{j_\ell}}{a_N}\right) = O\left(a_N^4\right)$$

for $i_\ell \neq j_\ell$, we can argue as in the proof of Lemma 5 that $\eta_m = a_N^{2m} O(1)$, substituting $u_{k_i} = \frac{z_{k_i}}{a_N}$, $k = 1, \ldots, m$, $i = 1, 2$, in the m-fold integral of indicator functions

$$\mathbf{1}_{[-2,+2]^2}\left(\frac{z_k - z_\ell}{a_N}\right) = \mathbf{1}_{[-2,+2]}\left(\frac{z_{k_1} - z_{\ell_1}}{a_N}\right) \mathbf{1}_{[-2,+2]}\left(\frac{z_{k_2} - z_{\ell_2}}{a_N}\right).$$

Finally, we have

$$\begin{aligned}\mathbb{E}\left(U_N\left(B\right)-U_N\left(\bar{B}\right)\right)^{2q} &= \rho_N^{2q}\sum_{m=2}^{2q}N^m a_N^{2q}a_N^{2m}O\left(N^{-2q\gamma}\right)\\ &= \frac{a_N^{6q}}{N^{2q}}O\left(N^{-2q\gamma}\right).\end{aligned}$$

□

Chapter 5

Asymptotics for local mean-squared error estimates with random bandwidths

We now consider a detailed asymptotic analysis of the local estimates of intensity and its second derivatives analogous to the integrated quantities in the previous chapter. These results are needed for local adaptive bandwidth selection.

In Proposition 4, we have derived an asymptotic expansion of $\hat{\lambda}_{ii}(x_1, x_2, H)$. For the purpose of bandwidth selection, the remainder terms have to be somewhat improved which we do in the following local version of Theorem 4.

Theorem 5. *Let h_1, h_2 be a sequence of bandwidths satisfying*

$$h_i = \beta_i b_N \left(1 + o_p\left(N^{-\gamma}\right)\right), i = 1, 2$$

for some fixed $\beta_1, \beta_2 > 0$, $\gamma \geq 0$ and a deterministic bandwidth rate $b_N \to 0$ with $b_N^4 = o_p(N^{-\gamma})$. We abbreviate $Q_{ii} = \int\int K_{ii}^2(u)\, \mathrm{d}u_1 \mathrm{d}u_2$, $i = 1, 2$. Under the assumptions of Theorem 4, we have for $i = 1, 2$ and fixed $x_1, x_2 \in [0, 1]$

with $\lambda(x_1,x_2)>0$

$$\hat{\lambda}_{ii}(x_1,x_2,H) = \lambda_{ii}(x_1,x_2)+\frac{1}{2}V_K\sum_{\ell=1}^{2}\beta_\ell^2\frac{\partial^2}{\partial x_\ell^2}\lambda_{ii}(x_1,x_2)\,b_N^2+o_p\left(b_N^2N^{-\gamma}\right)$$
$$+\sqrt{\lambda(x_1,x_2)\,Q_{ii}}\,O_p\left(\frac{1}{\sqrt{N}b_N^3}\right)+o_p\left(\frac{N^{-\frac{\gamma}{2}}}{\sqrt{N}b_N^3}\right)+O_p\left(\frac{N^{-\gamma}}{\sqrt{N}b_N^3}\right).$$

Proof. We decompose as before

$$\hat{\lambda}_{11}(x_1,x_2,H) = \nu_1(x_1,x_2,H)+\left(\hat{\lambda}_{11}(x_1,x_2,H)-\nu_1(x_1,x_2,H)\right)$$
$$= \nu_1(x_1,x_2,H)+D(H),$$

where as in the proof of Theorem 4

$$\nu_1(x_1,x_2,H)=\frac{1}{h_1^2}\int\int K_{11}(u_1,u_2)\,\lambda(x_1-h_1u_1,x_2-h_2u_2)\,\mathrm{d}u_1\mathrm{d}u_2$$

such that, for deterministic B, $\nu_1(x_1,x_2,B)=\mathbb{E}\hat{\lambda}_{11}(x_1,x_2,B)$. In the following, we write $z=(x_1,x_2)^\top$.

a) From the proof of Theorem 4, a), we have

$$\nu_1(z,H)=\lambda_{11}(z)+\frac{1}{2}V_K\sum_{i=1}^{2}\beta_i^2\frac{\partial^2}{\partial x_i^2}\lambda_{11}(z)\,b_N^2+o_p\left(b_N^2N^{-\gamma}\right).$$

b) With $L_B(u)=K_{11}\left(B^{-1}(z-u)\right)-b_1^3b_2\nu_1(z,B)$, we have for deterministic B

$$D(B)=\frac{1}{Nb_1^3b_2}\sum_{j=1}^{N}L_B(X_j)=\frac{1}{b_1^3b_2}S_N(B),$$

where $L_B(X_j)\,,j=1,\ldots,N$, are i.i.d. zero-mean random variables with

$$\mathrm{var}L_B(X_j)=\mathrm{var}K_{11}\left(B^{-1}(z-X_j)\right)=\lambda(z)\,Q_{11}b_1b_2+O\left(b_N^3\right)$$

as in the proof of Theorem 4, c). Here, we have assumed that $b_i=\beta_ib_N\left(1+O\left(N^{-\gamma}\right)\right)$, $i=1,2$. Therefore, we have

$$D(B)=O_p\left(\frac{1}{\sqrt{Nb_1^5b_2}}\right)\sqrt{\lambda(z)\,Q_{11}}.$$

c) Let B_N again denote the diagonal bandwidth matrix with entries $\beta_i b_N$. We decompose

$$\begin{aligned} D(H) &= D(B_N) + (D(H) - D(B_N)) \\ &= \sqrt{\lambda(z) Q_{11}} O_p\left(\frac{1}{\sqrt{N} b_N^3}\right) + (D(H) - D(B_N)) \end{aligned}$$

from b). Now, we follow the same line of arguments as in the investigation of $V_0(H) - V_0(B_N)$ in the proof of Theorem 4, e), to conclude that $D(H) - D(B_N)$ coincides up to asymptotically negligible terms with

$$\begin{aligned} \sup_{b \in B_{N,\tau}^{\gamma,2}} |D(B) - D(B_N)| \leq & \frac{O(N^{-\gamma})}{b_N^4} |S_N(B_N)| \\ & + \frac{O(1 + N^{-\gamma})}{b_N^4} \sup_{b \in B_{N,\tau}^{\gamma,2}} |S_N(B) - S_N(B_N)|, \end{aligned}$$

where we have used $|b_i - \beta_i b_N| = O(b_N N^{-\gamma})$, $i = 1, 2$, for $b \in B_{N,\tau}^{\gamma,2}$. From b), we have $S_N(B_N) = \beta_1^3 \beta_2 b_N^4 D(B_N) = O_p\left(\frac{b_N}{\sqrt{N}}\right)$, i.e. the first term is of order $O_p\left(\frac{N^{-\gamma}}{\sqrt{N} b_N^3}\right)$. For the second term, we conclude from Lemma 11 with $a_N = \max(\beta_1, \beta_2) b_N (1 + 2N^{-\gamma})$, $\alpha_N = \max(\beta_1, \beta_2) b_N (1 - 2N^{-\gamma})$

$$\mathbb{E}(S_N(B) - S_N(B_N))^{2q} = O\left(b_N^{2q} N^{-q-2q\gamma}\right).$$

As in the proof of Theorem 4, e), this implies

$$\frac{1}{b_N^4} \sup_{b \in B_{N,\tau}^{\gamma,2}} |S_N(B) - S_N(B_N)| = o_p\left(\frac{N^{-\frac{\gamma}{2}}}{\sqrt{N} b_N^3}\right).$$

Together, we have

$$D(H) = \sqrt{\lambda(z) Q_{11}} O_p\left(\frac{1}{\sqrt{N} b_N^3}\right) + O_p\left(\frac{N^{-\gamma}}{\sqrt{N} b_N^3}\right) + o_p\left(\frac{N^{-\frac{\gamma}{2}}}{\sqrt{N} b_N^3}\right)$$

which together with a) implies the result.

□

For $\hat{\lambda}(x_1, x_2, H)$, we have the following analogous expansion:

Theorem 6. *Let h_1, h_2 satisfy the conditions of Theorem 5. Then, under the assumptions of Theorem 4, we have for fixed $x_1, x_2 \in [0,1]$*

$$\begin{aligned}\hat{\lambda}(x_1, x_2, H) &= \lambda(x_1, x_2) + \frac{1}{2} V_K \sum_{\ell=1}^{2} \beta_\ell^2 \lambda_{\ell\ell}(x_1, x_2) b_N^2 + o_p\left(b_N^2 N^{-\gamma}\right) \\ &+ \sqrt{\lambda(x_1, x_2) Q_K} O_p\left(\frac{1}{\sqrt{N} b_N}\right) + o_p\left(\frac{N^{-\frac{\gamma}{2}}}{\sqrt{N} b_N}\right) + O_p\left(\frac{N^{-\gamma}}{\sqrt{N} b_N}\right).\end{aligned}$$

Proof. We decompose as in the proof of Theorem 5

$$\begin{aligned}\hat{\lambda}(x_1, x_2, H) &= \mu_1(x_1, x_2, H) + \left(\hat{\lambda}(x_1, x_2, H) - \mu_1(x_1, x_2, H)\right) \\ &= \mu_1(x_1, x_2, H) + D(H),\end{aligned}$$

where

$$\mu_1(x_1, x_2, H) = \int\int K(u_1, u_2) \lambda(x_1 - h_1 u_1, x_2 - h_2 u_2) \mathrm{d}u_1 \mathrm{d}u_2$$

such that, for deterministic B, $\mu_1(x_1, x_2, B) = \mathbb{E}\hat{\lambda}(x_1, x_2, B)$. In the following, we write $z = (x_1, x_2)^\top$. □

a) From the proof of Theorem 2, a), we have

$$\mu_1(z, H) = \lambda(z) + \frac{1}{2} V_K \sum_{i=1}^{2} \beta_i^2 \lambda_{ii}(z) b_N^2 + o_p\left(b_N^2 N^{-\gamma}\right).$$

b) With $M_B(u) = K\left(B^{-1}(z-u)\right) - b_1 b_2 \mu_1(z, B)$, we have for deterministic B

$$D(B) = \frac{1}{N b_1 b_2} \sum_{j=1}^{N} M_B(X_j) = \frac{1}{b_1 b_2} T_N(B),$$

where $M_B(X_j), j = 1, \ldots, N$ are i.i.d. with mean 0 and variance

$$\mathrm{var} M_B(X_j) = \mathrm{var} K\left(B^{-1}(z - X_j)\right) = \lambda(z) Q_K b_1 b_2 + O\left(b_N^3\right)$$

as in the proof of Theorem 2, b) with $b_i = \beta_i b_N (1 + o(N^{-\gamma}))$, $i = 1, 2$. Therefore, we have

$$D(B) = O_p\left(\frac{1}{\sqrt{N b_1 b_2}}\right) \sqrt{\lambda(z) Q_K}.$$

c) Now, we argue as in the proof of Theorem 5, c) to get

$$D(H)=\sqrt{\lambda(z)\,Q_K}O_p\left(\frac{1}{\sqrt{N}b_N}\right)+\left(D(H)-D(B_N)\right),$$

where the second term coincides up to asymptotically negligible terms with

$$\begin{aligned}\sup_{b\in B^{\gamma;2}_{N,\tau}}\left|D(B)-D(B_N)\right| &\leq \frac{O\left(N^{-\gamma}\right)}{b_N^2}\left|T_N(B_N)\right| \\ &\quad +\frac{O\left(1+N^{-\gamma}\right)}{b_N^2}\sup_{b\in B^{\gamma;2}_{N,\tau}}\left|T_N(B)-T_N(B_N)\right|.\end{aligned}$$

From b), the first term on the right-hand side is $O_p\left(\frac{N^{-\gamma}}{\sqrt{N}b_N}\right)$. For the second term we conclude from Lemma 10

$$\mathbb{E}\left(T_N(B)-T_N\left(\bar{B}\right)\right)^{2q}=O\left(b_N^{2q}N^{-q-2q\gamma}\right)$$

which implies as in the proof of Theorem 5, c)

$$\frac{1}{b_N^2}\sup_{b\in B^{\gamma;2}_{N,\tau}}\left|T_N(B)-T_N(B_N)\right|=o_p\left(\frac{N^{-\frac{\gamma}{2}}}{\sqrt{N}b_N}\right),$$

i.e. we have

$$D(H)=\sqrt{\lambda(z)\,Q_K}O_p\left(\frac{1}{\sqrt{N}b_N}\right)+O_p\left(\frac{N^{-\gamma}}{\sqrt{N}b_N}\right)+o_p\left(\frac{N^{-\frac{\gamma}{2}}}{\sqrt{N}b_N}\right)$$

which together with a) implies the result.

In the first lemma, we collect some Lipschitz properties needed for the two subsequent lemmas.

Lemma 9. *Let for $a_N>0$ the bandwidths $b_i,\bar{b}_i, i=1,2$, satisfying $\alpha_N=a_N\left(1-O\left(N^{-\gamma}\right)\right)<b_i,\bar{b}_i<a_N$ for some $\gamma\geq 0$. Then,*

a) Uniformly in $w\in\mathbb{R}^2$

$$\begin{aligned}\left|K(w)-K\left(\bar{B}^{-1}Bw\right)\right| &= O\left(N^{-\gamma}\right),\\ \left|K_{11}(w)-K_{11}\left(\bar{B}^{-1}Bw\right)\right| &= O\left(N^{-\gamma}\right).\end{aligned}$$

b) With μ_i, ν_i defined in the proofs of Theorem 5 respectively 6

$$\begin{aligned} \left|\mu_i\left(z, B\right) - \mu_i\left(z, \bar{B}\right)\right| &= O\left(a_N^2 N^{-2\gamma}\right) \\ \left|\nu_i\left(z, B\right) - \nu_i\left(z, \bar{B}\right)\right| &= O\left(a_N^2 N^{-2\gamma}\right) \end{aligned}$$

for any $z \in [0,1]^2$ and $i = 1, 2$.

Proof. a) Due to the Lipschitz continuity of K, we have for some $c^*, c > 0$ and large enough N

$$\begin{aligned} \left|K\left(w\right) - K\left(\bar{B}^{-1}Bw\right)\right| &\leq c\left\{\left|1 - \frac{b_1}{\bar{b}_1}\right| |w_1| + \left|1 - \frac{b_2}{\bar{b}_2}\right| |w_2|\right\} \cdot \mathbf{1}_{[-c^*, c^*]^2}\left(w\right) \\ &\leq cc^* \left\{\frac{\left|b_1 - \bar{b}_1\right|}{\bar{b}_1} + \frac{\left|b_2 - \bar{b}_2\right|}{\bar{b}_2}\right\} = O\left(N^{-\gamma}\right) \end{aligned}$$

as $K\left(w\right) - K\left(\bar{B}^{-1}Bw\right)$ has support contained in $[-c^*, c^*]^2$ following from $\frac{b_i}{\bar{b}_i} \leq \frac{a_N}{\alpha_N} = 1 + O\left(N^{-\gamma}\right) \leq c^*$ for large enough N and suitable c^*. For the last assertion, we have used $\left|b_i - \bar{b}_i\right| \leq a_N O\left(N^{-\gamma}\right)$ and $\frac{a_N}{\bar{b}_i} \leq \frac{a_N}{\alpha_N} = 1 + O\left(N^{-\gamma}\right)$. The same argument holds for K_{11} as it also is Lipschitz continuous and has a bounded support.

b) As $\mu_1\left(z, B\right) = \mathbb{E}\hat{\lambda}\left(z, B\right)$, we get as in the bias expansion in the proof of Theorem 2, a)

$$\begin{aligned} &\mu_1\left(z, B\right) - \mu_1\left(z, \bar{B}\right) \\ =\ & \int\int K\left(u\right)\left\{\lambda\left(z - Bu\right) - \lambda\left(z - \bar{B}u\right)\right\} \mathrm{d}u_1 \mathrm{d}u_2 \\ =\ & \frac{1}{2}\int\int K\left(u\right)\left\{\left(\bar{B} - B\right)u\right\}^\top \nabla^2 \lambda\left(z - \bar{B}u + \theta\left(\bar{B} - B\right)u\right)\left(\bar{B} - B\right)u \mathrm{d}u_1 \mathrm{d}u_2 \\ =\ & \frac{1}{2}\sum_{i,\ell=1}^{2}\left(\bar{b}_i - b_i\right)\left(\bar{b}_\ell - b_\ell\right)\int\int K\left(u\right) u_i u_\ell \lambda_{i\ell}\left(z - \bar{B}u + \theta\left(\bar{B} - B\right)u\right) \mathrm{d}u_1 \mathrm{d}u_2 \\ =\ & O\left(a_N^2 N^{-2\gamma}\right) \end{aligned}$$

as $\lambda_{i\ell}$ is bounded and has a bounded support. Analogously, $\nu_1\left(z, B\right) =$

$\mathbb{E}\hat{\lambda}_{11}(z,B)$ such that from the proof of Proposition 3, a)

$$\begin{aligned}\nu_1(z,B)-\nu_1\left(z,\bar{B}\right) &= \int\int K(u)\left\{\lambda_{11}(z-Bu)-\lambda_{11}\left(z-\bar{B}u\right)\right\}\mathrm{d}u_1\mathrm{d}u_2 \\ &= O\left(a_N^2 N^{-2\gamma}\right)\end{aligned}$$

by the same Taylor expansion argument.

□

Lemma 10. *Let the assumptions of Theorem 5 be satisfied. Let for some $a_N > 0$ with $\frac{1}{Na_N^2} = O(1)$, the bandwidths $b_i, \bar{b}_i$, $i = 1,2$, satisfying $\alpha_N = a_N\left(1-O\left(N^{-\gamma}\right)\right) < b_i, \bar{b}_i < a_N$ for some $\gamma \geq 0$. Let*

$$T_N(B) = \frac{1}{N}\sum_{j=1}^{N} M_B(X_j)$$

with

$$M_B(u) = K\left(B^{-1}(z-u)\right) - b_1 b_2 \mu_1(z,B).$$

Then, for $q \geq 1$

$$\mathbb{E}\left(T_N(B)-T_N\left(\bar{B}\right)\right)^{2q} = O\left(\frac{a_N^{2q}}{N^q N^{2q\gamma}}\right).$$

Proof. The proof is similar to those of Lemma 5 and 8 but requires a somewhat more careful argument, as a direct analogy of the arguments for the integrated estimates here would lead to suboptimal rates which are not good enough for the intended application. c denotes various constants in the following.

a) As in the proof of Lemma 8, d), we write

$$T_N(B) - T_N\left(\bar{B}\right) = \frac{1}{N}\sum_{j=1}^{N} D_j,$$

where $D_j = M_B(X_j) - M_{\bar{B}}(X_j)$, $j = 1,\ldots,N$, are i.i.d. zero-mean random variables. In particular, we have for any $j_1,\ldots,j_{2q} \in \{1,\ldots,N\}$

$$\mathbb{E}D_{j_1}\cdots D_{j_{2q}} = 0$$

if $j_1, \ldots, j_{2q}$ contain more than q different indices such that at least one index appears only once in the product.

As in the proof of Lemma 5, d), there are at most $c_q N^m$ terms $\mathbb{E}D_{j_1}\cdots D_{j_{2q}}$ with $j_1, \ldots, j_{2q}$ containing exactly m different values, and we have

$$\mathbb{E}\left(T_N(B) - T_N\left(\bar{B}\right)\right)^{2q} \leq \frac{c}{N^{2q}} \sum_{m=1}^{q} N^m \pi_m,$$

where π_m is an upper bound for $\left|\mathbb{E}D_{j_1}\cdots D_{j_{2q}}\right|$ in case of m different values in $\{j_1, \ldots, j_{2q}\}$.

b) To calculate π_m, we set $\{j_1, \ldots, j_{2q}\} = \{k_1, \ldots, k_m\}$, where $k_1 \neq \cdots \neq k_m$, and we write

$$D(u) = M_B(u) - M_{\bar{B}}(u)$$

such that $D_j = D(X_j)$. Due to independence of $X_1, \ldots, X_N$

$$\begin{aligned}
&\left|\mathbb{E}D_{j_1}\cdots D_{j_{2q}}\right| \\
= &\left|\int\cdots\int \prod_{\ell=1}^{2q} D\left(u_{j_\ell}\right) \prod_{i=1}^{m} \lambda\left(u_{k_i}\right) \mathrm{d}u_{k_1,1}\mathrm{d}u_{k_1,2}\cdots\mathrm{d}u_{k_m,1}\mathrm{d}u_{k_m,2}\right| \\
= &\left|b_1^m b_2^m \int\cdots\int \prod_{\ell=1}^{2q} d\left(v_{j_\ell}\right) \prod_{i=1}^{m} \lambda\left(z - Bv_{k_i}\right) \mathrm{d}v_{k_1,1}\mathrm{d}v_{k_1,2}\cdots\mathrm{d}v_{k_m,1}\mathrm{d}v_{k_m,2}\right| \\
\leq &\; c\left(b_1 b_2\right)^m \int\cdots\int \left|\prod_{\ell=1}^{2q} d\left(v_{j_\ell}\right)\right| \mathrm{d}v_{k_1,1}\mathrm{d}v_{k_1,2}\cdots\mathrm{d}v_{k_m,1}\mathrm{d}v_{k_m,2}
\end{aligned}$$

satisfying $v_{k_\ell} = B^{-1}\left(z - u_{k_\ell}\right),\ \ell = 1, \ldots, m$, and writing

$$d(v) = K(v) - K\left(\bar{B}^{-1}Bv\right) - b_1 b_2 \mu_1(z, B) + \bar{b}_1 \bar{b}_2 \mu_1(z, B).$$

For the last inequality, we have used boundedness of λ. From Lemma 9, we conclude $d(v) = O\left(N^{-\gamma}\right)\left(1 + O\left(a_N^2\right)\right)$ uniformly in $b, \bar{b}$ and v, as $\left|b_i - \bar{b}_i\right| \leq a_N - \alpha_N = a_N O\left(N^{-\gamma}\right)$, $i = 1, 2$. We conclude as $b_i \leq a_N$, $i = 1, 2$, that $\pi_m = a_N^{2m} O\left(N^{-2q\gamma}\right)$.

c) Combining a) and b), we have

$$\begin{aligned}
\mathbb{E}\left(T_N(B) - T_N\left(\bar{B}\right)\right)^{2q} &= O\left(\frac{N^{-2q\gamma}}{N^{2q}}\right)\sum_{m=1}^{q} N^m a_N^{2m} \\
&= O\left(\frac{N^{-2q\gamma}}{N^{2q}}\right)\sum_{m=0}^{q} \left(Na_N^2\right)^m \\
&= O\left(\frac{N^{-2q\gamma}}{N^{2q}}\right)\frac{\left(Na_N^2\right)^{q+1}-1}{Na_N^2-1} \\
&= O\left(\frac{a_N^{2q}N^{-2q\gamma}}{N^q}\right).
\end{aligned}$$

□

Lemma 11. *Let the assumptions of Lemma 10 be satisfied, then*

$$\mathbb{E}\left(S_N(B) - S_N\left(\bar{B}\right)\right)^{2q} = O\left(\frac{a_N^{2q}}{N^q N^{2q\gamma}}\right),$$

where

$$S_N(B) = \frac{1}{N}\sum_{j=1}^{N} L_B(X_j).$$

Proof. The proof uses the same ideas as the proof of Lemma 10.

a) Set $D_j = L_B(X_j) - L_{\bar{B}}(X_j)$ such that $D_1, \ldots, D_N$ are i.i.d. with $\mathbb{E}D_j = 0$ and

$$S_N(B) - S_N\left(\bar{B}\right) = \frac{1}{N}\sum_{j=1}^{N} D_j.$$

Note that for $j_1, \ldots, j_{2q}$ containing more than q different indices,

$$\mathbb{E}D_{j_1}\cdots D_{j_{2q}} = 0,$$

and we get as in the proof of Lemma 10, a)

$$\mathbb{E}\left(S_N(B) - S_N\left(\bar{B}\right)\right)^{2q} \leq \frac{c}{N^{2q}}\sum_{m=1}^{q} N^m \pi_m,$$

where now π_m is an upper bound for $\left|\mathbb{E}D_{j_1}\cdots D_{j_q}\right|$ in case of exactly m different values among $j_1, \ldots, j_{2q}$.

b) We have to calculate π_m and set $\{k_1, \ldots, k_m\} = \{j_1, \ldots, j_{2q}\}$ with different $k_1, \ldots, k_m$. We write

$$D(u) = L_B(u) - L_{\bar{B}}(u)$$

such that $D_j = D(X_j)$. Then, we have as in the proof of Lemma 10, b)

$$\begin{aligned}
&\left|\mathbb{E} D_{j_1} \cdots D_{j_{2q}}\right| \\
= &\left| \int \cdots \int \prod_{\ell=1}^{2q} D\left(u_{j_\ell}\right) \prod_{i=1}^{m} \lambda\left(u_{k_i}\right) \mathrm{d}u_{k_1,1} \mathrm{d}u_{k_1,2} \cdots \mathrm{d}u_{k_m,1} \mathrm{d}u_{k_m,2} \right| \\
= & b_1^m b_2^m \left| \int \cdots \int \prod_{\ell=1}^{2q} d\left(v_{j_\ell}\right) \prod_{i=1}^{m} \lambda\left(z - B v_{k_i}\right) \mathrm{d}v_{k_1,1} \mathrm{d}v_{k_1,2} \cdots \mathrm{d}v_{k_m,1} \mathrm{d}v_{k_m,2} \right| \\
\leq & c \cdot (b_1 b_2)^m \int \cdots \int \left| \prod_{\ell=1}^{2q} d\left(v_{j_\ell}\right) \right| \mathrm{d}v_{k_1,1} \mathrm{d}v_{k_1,2} \cdots \mathrm{d}v_{k_m,1} \mathrm{d}v_{k_m,2},
\end{aligned}$$

where

$$\begin{aligned}
d(v) &= K_{11}(v) - K_{11}\left(\bar{B}^{-1} B v\right) - b_1^3 b_2 \nu_1(z, B) + \bar{b}_1^3 \bar{b}_2 \nu_1\left(z, \bar{B}\right) \\
&= O\left(N^{-\gamma}\right)\left(1 + O\left(a_N^4\right)\right)
\end{aligned}$$

uniformly in $\bar{b}_1, \bar{b}_2$ and v from Lemma 9, and we again conclude $\pi_m = a_N^{2m} O\left(N^{-2q\gamma}\right)$ and, finally,

$$\begin{aligned}
\mathbb{E}\left(S_N(B) - S_N\left(\bar{B}\right)\right)^{2q} &\leq c \frac{1}{N^{2q(1+\gamma)}} \sum_{m=1}^{q} \left(N a_N^2\right)^m \\
&\leq \frac{c}{N^{2q(1+\gamma)}} \frac{\left(N a_N^2\right)^{q+1} - 1}{N a_N^2 - 1} \\
&= O\left(\frac{a_N^{2q}}{N^q N^{2q\gamma}}\right).
\end{aligned}$$

□

Chapter 6

Tuning parameters of the plug-in algorithm

In this chapter, we finally complete the description of the algorithm for automatic bandwidth selection. Using the asymptotics of Chapters 4 and 5, we recommend an inflation factor N^ρ with $\rho = \frac{1}{12}$, but also discuss an alternative. Then, we illustrate how ρ together with the asymptotic analysis determines the numbers of iteration steps in the global respectively local bandwidth selection algorithm. It turns out that a very small number of iterations suffices as they already lead to approximations of the optimal bandwidths with remaining approximation errors caused by unavoidable purely random effects.

The algorithm for data-adaptive selection of global and local bandwidth parameters described in Section 3.1 depends on the choice of some tuning parameters: the initial bandwidth values $\hat{h}_i^{(0)}$, $i = 1, 2$, the inflation factor N^ρ for the bandwidths of second derivative estimates and the number of iterations i^* respectively j^* of the global and local part of the algorithm. In particular, one may ask why we propose a small and fixed number of iterations. This is inspired by the one-dimensional case studied in Engel et al. (1994), but will be explained in the first section below. In the following section, we discuss the choice of the inflation factor.

Before we start with the derivation, let us briefly explain why the algorithm is split into global and local part. The asymptotically optimal local bandwidths from Theorem 3 are of order $N^{-\frac{1}{6}}$ globally and locally, and from Corollary 5, we know that they approximate the optimal bandwidths well. So, one might wonder why we do not immediately start with the local part of the algorithm with $\hat{h}_i^{(0)} = \frac{1}{\sqrt{N}}$, $i = 1, 2$? This, however, would lead to highly variable estimates of $\lambda_{ii}(x_1, x_2)$, $i = 1, 2$, which are needed in the iteration, and the algorithm would suffer from pronounced random effects.

We get a much more stable behaviour for the global iterations as we only have to estimate

$$\Lambda_{k\ell} = \int \int \lambda_{kk}(x_1, x_2) \lambda_{\ell\ell}(x_1, x_2) \, \mathrm{d}x_1 \mathrm{d}x_2, k, \ell = 1, 2,$$

which suffers less from randomness as integration acts as a smoothing operation cancelling the effect of the strong local variation of estimates $\hat{\lambda}_{ii}(x_1, x_2, H)$, $i = 1, 2$.

The algorithm, in particular through the choice of i^*, is constructed in a manner such that it uses global bandwidths until the right rate $N^{-\frac{1}{6}}$ is achieved. Then, it switches to local iterations to improve the constant of the selected bandwidths.

Let us briefly discuss the choice of initial values. We start with small bandwidths such that the first estimates involve only a little smoothing and are close to the data, i.e. the density estimate is not far away from the empirical measure (which would correspond roughly to a bandwidth choice of order N^{-1}).

6.1 Data-adaptive choice of global bandwidths

The numbers i^* and j^* of global and local iterations depend crucially on the inflation factor N^ρ. In this section, we illustrate this for $\rho = \frac{1}{12}$ which we shall recommend for reasons explained in the next section.

We use the following abbreviations

$$C_K = \frac{Q_K}{V_K^2}, M_{i\ell} = \int \int K_{ii}(u) K_{\ell\ell}(u) \, \mathrm{d}u_1 \mathrm{d}u_2, i, \ell = 1, 2,$$

which are known constants depending on the kernel K only. We also write

$$I_{i\ell}(k) = \int \int \lambda_{ii}(x_1, x_2) \frac{\partial^2}{\partial x_k^2} \lambda_{\ell\ell}(x_1, x_2) \, \mathrm{d}x_1 \mathrm{d}x_2 + \int \int \lambda_{\ell\ell}(x_1, x_2) \frac{\partial^2}{\partial x_k^2} \lambda_{ii}(x_1, x_2) \, \mathrm{d}x_1 \mathrm{d}x_2,$$

which are constants depending only on λ. From Theorem 4, we have, using $\beta_{i\ell} = \beta_1 \beta_2 \beta_i^2 \beta_\ell^2$ as abbreviation,

$$\hat{\Lambda}_{i\ell}(H) = \Lambda_{i\ell} + \frac{M_{i\ell}}{\beta_{i\ell}} \frac{1}{N b_N^6} + \frac{V_K}{2} \sum_{k=1}^{2} \beta_k^2 I_{i\ell}(k) \, b_N^2 + R_{N,i\ell},$$

where the components of the remainder term are of smaller order than the second or the third term respectively.

Step 1: Starting with $\hat{h}_1^{(0)} = \hat{h}_2^{(0)} = \frac{1}{\sqrt{N}}$, we can choose $\beta_1 = \beta_2 = 1$, $b_N = \frac{1}{\sqrt{N}}$. Then, for $\rho = \frac{1}{12}$,

$$\begin{aligned} \hat{\Lambda}_{i\ell}\left(N^\rho \hat{H}^{(0)}\right) &= \Lambda_{i\ell} + \frac{M_{i\ell}}{\beta_{i\ell}} \frac{1}{(N^\rho b_N)^6 N} + \frac{V_K}{2} \sum_{k=1}^{2} \beta_k^2 I_{i\ell}(k) (N^\rho b_N)^2 + R_{N,i\ell} \\ &= N^{\frac{3}{2}} \frac{M_{i\ell}}{\beta_{i\ell}} (1 + o_p(1)) \quad \text{for } i, \ell = 1, 2, \end{aligned}$$

as $(N^\rho b_N)^6 N = N^{\frac{3}{2}} b_N^6 = N^{\frac{3}{2}} N^{-\frac{6}{2}} = N^{-\frac{3}{2}}$ and $(N^\rho b_N)^2 = N^{\frac{1}{6}} \frac{1}{N} = N^{-\frac{5}{6}}$. Therefore, $\Lambda_{i\ell}$ is a constant, the dominant term in the expansion of $\hat{\Lambda}_{i\ell}\left(N^\rho \hat{H}^{(0)}\right)$ is the second one. Plugging this into the formula for

$\hat{h}_1^{(1)}, \hat{h}_2^{(1)}$, we get

$$\begin{aligned}
\hat{h}_1^{(1)} &= C_K^{\frac{1}{6}} \left(\frac{M_{22}}{M_{11}}\right)^{\frac{1}{8}} \frac{1}{\left(\sqrt{M_{11}M_{22}}N^{\frac{3}{2}} + N^{\frac{3}{2}}M_{12}\right)^{\frac{1}{6}}} \frac{1}{N^{\frac{1}{6}}} \left(1 + o_p\left(1\right)\right) \\
&= C_K^{\frac{1}{6}} \left(\frac{M_{22}}{M_{11}}\right)^{\frac{1}{8}} \frac{1}{\left(\sqrt{M_{11}M_{22}} + M_{12}\right)^{\frac{1}{6}}} N^{-\frac{5}{12}} \left(1 + o_p\left(1\right)\right), \\
\hat{h}_2^{(1)} &= C_K^{\frac{1}{6}} \left(\frac{M_{11}}{M_{22}}\right)^{\frac{1}{8}} \frac{1}{\left(\sqrt{M_{11}M_{22}} + M_{12}\right)^{\frac{1}{6}}} N^{-\frac{5}{12}} \left(1 + o_p\left(1\right)\right).
\end{aligned}$$

Step 2: Let us denote by $\beta_i^{(1)}$ the factors depending only on the kernel such that $\hat{h}_i^{(1)} = \beta_i^{(1)} N^{-\frac{5}{12}} \left(1 + o_p\left(1\right)\right)$, and $b_N = N^{-\frac{5}{12}}$. Then, the dominant term in the expansion of $\hat{\Lambda}_{i\ell}\left(N^\rho \hat{H}^{(1)}\right)$ is again the second one, and

$$\hat{\Lambda}_{i\ell}\left(N^\rho \hat{H}^{(1)}\right) = N \frac{M_{i\ell}}{\beta_{i\ell}^{(1)}} \left(1 + o_p\left(1\right)\right) \quad \text{for } i, \ell = 1, 2,$$

with $\beta_{i\ell}^{(1)}$ being defined as $\beta_{i\ell}$ with $\beta_i = \beta_i^{(1)}$, $i = 1, 2$. We get

$$\begin{aligned}
\hat{h}_1^{(2)} &= C_K^{\frac{1}{6}} \left(\frac{M_{22}\beta_{11}^{(1)}}{M_{11}\beta_{22}^{(1)}}\right)^{\frac{1}{8}} \frac{\sqrt{\beta_1^{(1)}\beta_2^{(1)}}}{\left(\sqrt{M_{11}M_{22}} + M_{12}\right)^{\frac{1}{6}}} N^{-\frac{1}{3}} \left(1 + o_p\left(1\right)\right), \\
\hat{h}_2^{(2)} &= C_K^{\frac{1}{6}} \left(\frac{M_{11}\beta_{22}^{(1)}}{M_{22}\beta_{11}^{(1)}}\right)^{\frac{1}{8}} \frac{\sqrt{\beta_1^{(1)}\beta_2^{(1)}}}{\left(\sqrt{M_{11}M_{22}} + M_{12}\right)^{\frac{1}{6}}} N^{-\frac{1}{3}} \left(1 + o_p\left(1\right)\right)
\end{aligned}$$

as $\beta_{11}^{(1)}\beta_{22}^{(1)} = \left(\beta_1^{(1)}\beta_2^{(1)}\right)^6 = \left(\beta_{12}^{(1)}\right)^2$.

Step 3: Write again $\hat{h}_i^{(2)} = \beta_i^{(2)} b_N \left(1 + o_p\left(1\right)\right)$ with $b_N = N^{-\frac{1}{3}}$. Again the second term is dominant, and, for $i, \ell = 1, 2$,

$$\hat{\Lambda}_{i\ell}\left(N^\rho \hat{H}^{(2)}\right) = \sqrt{N} \frac{M_{i\ell}}{\beta_{i\ell}^{(2)}} \left(1 + o_p\left(1\right)\right)$$

and

$$\begin{aligned}
\hat{h}_1^{(3)} &= C_K^{\frac{1}{6}} \left(\frac{M_{22}\beta_{11}^{(2)}}{M_{11}\beta_{22}^{(2)}}\right)^{\frac{1}{8}} \frac{\sqrt{\beta_1^{(2)}\beta_2^{(2)}}}{\left(\sqrt{M_{11}M_{22}} + M_{12}\right)^{\frac{1}{6}}} N^{-\frac{1}{4}} \left(1 + o_p\left(1\right)\right), \\
\hat{h}_2^{(3)} &= C_K^{\frac{1}{6}} \left(\frac{M_{11}\beta_{22}^{(2)}}{M_{22}\beta_{11}^{(2)}}\right)^{\frac{1}{8}} \frac{\sqrt{\beta_1^{(2)}\beta_2^{(2)}}}{\left(\sqrt{M_{11}M_{22}} + M_{12}\right)^{\frac{1}{6}}} N^{-\frac{1}{4}} \left(1 + o_p\left(1\right)\right).
\end{aligned}$$

Step 4: Write again $\hat{h}_i^{(3)} = \beta_i^{(3)} b_N \left(1 + o_p\left(1\right)\right)$ with $b_N = N^{-\frac{1}{4}}$. Now, the constant term in the expansion of $\hat{\Lambda}_{i\ell}\left(N^{\rho}\hat{H}^{(3)}\right)$ is no longer negligible compared to the second one. Only the third term is of smaller order as $\left(N^{\rho}b_N\right)^2 = N^{\frac{1}{6}}N^{-\frac{1}{2}} = N^{-\frac{1}{3}}$. Hence, we have now for $i, \ell = 1, 2$

$$\hat{\Lambda}_{i\ell}\left(N^{\rho}\hat{H}^{(3)}\right) = \Lambda_{i\ell} + \frac{M_{i\ell}}{\beta_{i\ell}^{(3)}}\left(1 + o_p\left(1\right)\right),$$

and we get

$$\begin{aligned} \hat{h}_1^{(4)} &= \beta_1^{(4)} N^{-\frac{1}{6}}\left(1 + o_p\left(1\right)\right), \\ \hat{h}_2^{(4)} &= \beta_2^{(4)} N^{-\frac{1}{6}}\left(1 + o_p\left(1\right)\right), \end{aligned}$$

where the constants $\beta_i^{(4)}$, $i = 1, 2$, depend on $C_K, M_{i\ell}, \beta_{i\ell}^{(3)}, \Lambda_{i\ell}$, $i = 1, 2$. Now, after the 4th step, we have reached the optimal bandwidth order $N^{-\frac{1}{6}}$.

For the remaining steps, the main term of the bandwidths will always be of optimal rate $N^{-\frac{1}{6}}$, and, hence, the inflated bandwidths for the second-derivative estimates will be of order $N^{\frac{1}{12}}N^{-\frac{1}{6}} = N^{-\frac{1}{12}}$. Now, the remainder terms in Proposition 4 and 5 for the local case and in Theorem 4 for the global case become relevant.

We first consider the situation where we are only interested in a global bandwidth. Then, the remainder terms of Theorem 4 for bandwidth rate $N^{\rho}b_N = N^{\frac{1}{12}}N^{-\frac{1}{6}} = N^{-\frac{1}{12}}$ are

$$\begin{aligned} R_{N,i\ell} &= o\left(N^{-\frac{1}{6}}\right) + o_p\left(N^{-\gamma}N^{-\frac{1}{6}}\right) + O_p\left(N^{-\frac{7}{12}}\right) + o_p\left(\frac{N^{-\gamma}}{\sqrt{N}}\right) + o_p\left(\frac{\log N}{N}\right) \\ &\quad + o_p\left(\frac{\log N}{\sqrt{N}}\right) \\ &= o\left(N^{-\frac{1}{6}}\right) + o_p\left(N^{-\gamma}N^{-\frac{1}{6}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right), \end{aligned}$$

$i, \ell = 1, 2$, and, correspondingly,

$$\begin{aligned}
\hat{\Lambda}_{i\ell}(H) &= \Lambda_{i\ell} + \frac{M_{i\ell}}{\beta_{i\ell}} O\left(\frac{1}{\sqrt{N}}\right) + \frac{V_K}{2} \sum_{k=1}^{2} \beta_k^2 I_{i\ell}(k) O\left(N^{-\frac{1}{6}}\right) + R_{N,i\ell} \\
&= \Lambda_{i\ell} + O\left(N^{-\frac{1}{6}}\right) + o_p\left(N^{-\gamma} N^{-\frac{1}{6}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right) \\
&= \Lambda_{i\ell} + r_N(\gamma).
\end{aligned}$$

Step 5: From the expansion of $\hat{h}_i^{(4)}$, they satisfy Condition 2 of Proposition 4 and the subsequent expansions with $b_N = N^{-\frac{1}{6}}$ and $\gamma = 0$. Hence, $r_N(\gamma) = r_N(0) = O\left(N^{-\frac{1}{6}}\right) + o_p\left(N^{-\frac{1}{6}}\right)$, and

$$\begin{aligned}
&\left(\frac{\hat{\Lambda}_{22}}{\hat{\Lambda}_{11}}\right)^{\frac{1}{8}} \frac{1}{\left(\sqrt{\hat{\Lambda}_{11}\hat{\Lambda}_{22}} + \hat{\Lambda}_{12}\right)^{\frac{1}{6}}} \\
= &\left(\frac{\Lambda_{22}}{\Lambda_{11}}\right)^{\frac{1}{8}} \frac{1}{\left(\sqrt{\Lambda_{11}\Lambda_{22}} + \Lambda_{12}\right)^{\frac{1}{6}}} \left(1 + O\left(N^{-\frac{1}{6}}\right) + o_p\left(N^{-\frac{1}{6}}\right)\right)
\end{aligned}$$

which implies

$$\begin{aligned}
\hat{h}_1^{(5)} &= h_{a1}\left(1 + O\left(N^{-\frac{1}{6}}\right) + o_p\left(N^{-\frac{1}{6}}\right)\right) \\
&= h_{a1}\left(1 + O\left(N^{-\frac{1}{6}}\right)\right)\left(1 + o_p\left(N^{-\frac{1}{6}}\right)\right)
\end{aligned}$$

and, analogously, we have the same relation for $\hat{h}_2^{(5)}$, h_{a2}.

Step 6: As $1 + O\left(N^{-\frac{1}{6}}\right) + o_p\left(N^{-\frac{1}{6}}\right) = \left(1 + O\left(N^{-\frac{1}{6}}\right)\right)\left(1 + o_p\left(N^{-\frac{1}{6}}\right)\right)$, $\hat{h}_i^{(5)}$ satisfies Condition 2 with $b_N = N^{-\frac{1}{6}}$, $\gamma = \frac{1}{6}$. Hence, $r_N(\gamma) = r_N\left(\frac{1}{6}\right) = O\left(N^{-\frac{1}{6}}\right) + o_p\left(N^{-\frac{1}{3}}\right)$, and, as in Step 5,

$$\hat{h}_i^{(6)} = h_{ai}\left(1 + O\left(N^{-\frac{1}{6}}\right) + o_p\left(N^{-\frac{1}{3}}\right)\right), i = 1, 2.$$

Step 7: $\hat{h}_i^{(6)}$ satisfies Condition 2 with $b_N = N^{-\frac{1}{6}}$, $\gamma = \frac{1}{3}$, such that

$$\begin{aligned}
r_N(\gamma) &= r_N\left(\frac{1}{3}\right) = O\left(N^{-\frac{1}{6}}\right) + o_p\left(N^{-\frac{1}{2}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right) \\
&= O\left(N^{-\frac{1}{6}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right)
\end{aligned}$$

and we have

$$\hat{h}_i^{(7)} = h_{ai}\left(1 + O\left(N^{-\frac{1}{6}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right)\right), \quad i = 1, 2.$$

Further iterations do not improve the rate with which $\hat{h}_i^{(7)}$ approximates h_{ai}, as the term $o_p\left(N^{-\gamma}N^{-\frac{1}{6}}\right)$ will stay negligible compared to $o_p\left(\frac{\log N}{\sqrt{N}}\right)$, i.e. the algorithm can stop here.

6.2 Data-adaptive choice of local bandwidths

Now, we consider the situation where we are interested in local bandwidth selection. We start with 4 steps of global iteration with $\rho = \frac{1}{12}$ such that the optimal bandwidth rate $b_N = N^{-\frac{1}{6}}$ is reached. From Theorems 5 and 6 we have for $i = 1, 2$

$$\begin{aligned}
\hat{\lambda}_{ii}^2\left(x_1, x_2, N^\rho H\right) &= \lambda_{ii}^2\left(x_1, x_2\right) + O\left(N^{2\rho}b_N^2\right) + O_p\left(\frac{1}{\sqrt{N}\left(N^\rho b_N\right)^3}\right) + r_N'\left(\gamma\right) \\
&= \lambda_{ii}^2\left(x_1, x_2\right) + O\left(N^{-\frac{1}{6}}\right) + O_p\left(N^{-\frac{1}{4}}\right) + r_N'\left(\gamma\right) \\
\hat{\lambda}\left(x_1, x_2, H\right) &= \lambda\left(x_1, x_2\right) + O\left(b_N^2\right) + O_p\left(\frac{1}{\sqrt{N}b_N}\right) + r_N\left(\gamma\right) \\
&= \lambda\left(x_1, x_2\right) + O\left(N^{-\frac{1}{3}}\right) + O_p\left(N^{-\frac{1}{3}}\right) + r_N\left(\gamma\right)
\end{aligned}$$

with

$$\begin{aligned}
r_N\left(\gamma\right) &= o_p\left(b_N^2 N^{-\gamma}\right) + o_p\left(\frac{N^{-\frac{\gamma}{2}}}{\sqrt{N}b_n}\right) + O_p\left(\frac{N^{-\gamma}}{\sqrt{N}b_N}\right) \\
&= o_p\left(N^{-\frac{1}{3}}N^{-\frac{\gamma}{2}}\right) + O_p\left(N^{-\frac{1}{3}}N^{-\gamma}\right) \\
r_N'\left(\gamma\right) &= o_p\left(\left(N^\rho b_N\right)^2 N^{-\gamma}\right) + o_p\left(\frac{N^{-\frac{\gamma}{2}}}{\sqrt{N}\left(N^\rho b_N\right)^3}\right) + O_p\left(\frac{N^{-\gamma}}{\sqrt{N}\left(N^\rho b_N\right)^3}\right) \\
&= o_p\left(N^{-\frac{1}{6}}N^{-\gamma}\right) + o_p\left(N^{-\frac{1}{4}}N^{-\frac{\gamma}{2}}\right) + O_p\left(N^{-\frac{1}{4}}N^{-\gamma}\right).
\end{aligned}$$

Therefore, we have

$$\hat{\lambda}^{\frac{1}{6}}\left(x_1, x_2, H\right)\left|\frac{\hat{\lambda}_{22}\left(x_1, x_2, N^\rho H\right)}{\hat{\lambda}_{11}^5\left(x_1, x_2, N^\rho H\right)}\right|^{\frac{1}{12}} = \lambda^{\frac{1}{6}}\left(x_1, x_2\right)\left|\frac{\lambda_{22}\left(x_1, x_2\right)}{\lambda_{11}^5\left(x_1, x_2\right)}\right|^{\frac{1}{12}} + R_N\left(\gamma\right)$$

with

$$R_N(\gamma) = O\left(N^{-\frac{1}{6}}\right)+O_p\left(N^{-\frac{1}{4}}\right)+o_p\left(N^{-\frac{1}{6}}N^{-\gamma}\right)+o_p\left(N^{-\frac{1}{4}}N^{-\frac{\gamma}{2}}\right)+O_p\left(N^{-\frac{1}{4}}N^{-\gamma}\right).$$

Step 5': As we use $\hat{h}_i^{(4)}$ for the kernel estimates here, we have as in Step 5 of the global iteration $\gamma = 0$ and $R_N(\gamma) = O\left(N^{-\frac{1}{6}}\right) + o_p\left(N^{-\frac{1}{6}}\right)$, which implies

$$\hat{h}_1^{(5)}(x_1, x_2) = h_{a1}(x_1, x_2)\left(1 + O\left(N^{-\frac{1}{6}}\right) + o_p\left(N^{-\frac{1}{6}}\right)\right)$$

and, analogously, for $\hat{h}_2^{(5)}(x_1, x_2)$ and $h_{a2}(x_1, x_2)$.

Step 6': As in Step 6 of the global iteration, we now have $\gamma = \frac{1}{6}$, and $R_N(\gamma) = O\left(N^{-\frac{1}{6}}\right) + O_p\left(N^{-\frac{1}{4}}\right)$ which is the fastest possible rate as the second term in the definition of $R_N(\gamma)$ does not depend on γ. Further iterations would not improve the approximation, and we stop with

$$\hat{h}_i^{(6)}(x_1, x_2) = h_{ai}(x_1, x_2)\left(1 + O\left(N^{-\frac{1}{6}}\right) + O_p\left(N^{-\frac{1}{4}}\right)\right).$$

Let us close the discussion with a remark on situations where a mixed local-global bandwidth selection is advisable, even if we are interested in locally optimal smoothing. If $\lambda(x_1, x_2)$ is close to 0 and flat, i.e. $\lambda_{ii}(x_1, x_2) \approx 0$, $i = 1, 2$, too, then the local bandwidth $\hat{h}_i^{(\ell)}(x_1, x_2)$ of the algorithm become quite unstable. The optimal bandwidths should be quite large in these regions but due to random variations the crucial factor, for e.g. $i = 1$,

$$\left\{\hat{\lambda}^2\left(x_1, x_2, \hat{H}^{(\ell-1)}(x_1, x_2)\right)\frac{\hat{\lambda}_{22}\left(x_1, x_2, N^\rho \hat{H}^{(\ell-1)}(x_1, x_2)\right)}{\hat{\lambda}_{11}^5\left(x_1, x_2, N^\rho \hat{H}^{(\ell-1)}(x_1, x_2)\right)}\right\}^{\frac{1}{12}}$$

may assume small and large value alike. In this case, it is recommendable to stick with a good global bandwidth in those regions, i.e. to use $\hat{h}_i^{(7)}$, $i = 1, 2$, where $\hat{\lambda}\left(x_1, x_2, \hat{H}^{(4)}\right)$ and $\left|\hat{\lambda}_{ii}\left(x_1, x_2, N^\rho \hat{H}^{(4)}\right)\right|$ are below some constant, say, c_λ, in a neighbourhood

$$U(x_1, x_2) = \left\{(u_1, u_2)^\top ; |u_1 - x_1|, |u_2 - x_2| \leq \delta\right\}$$

of $(x_1, x_2)^\top$. E.g. we could check if

$$\min_{u \in U(x_1, x_2)} \left\{\hat{\lambda}\left(u, \hat{H}^{(4)}\right), \left|\hat{\lambda}_{ii}\left(u, N^\rho \hat{H}^{(4)}\right)\right|, i = 1, 2\right\} > c_\lambda$$

as a criterion if we proceed with the local bandwidth optimisation or not.

We combine the results of our discussion into the following theorem.

Theorem 7. *Under the assumptions of Theorem 4, we have that the algorithm of Section 3.1 obtains the following rates for approximating the asymptotically optimal bandwidths for inflation factor* $N^\rho = N^{\frac{1}{12}}$.

a) $\hat{h}_i^{(7)} = h_{ai}\left(1 + O\left(N^{-\frac{1}{6}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right)\right), i = 1, 2,$

b) $\hat{h}_i^{(6)}(x_1, x_2) = h_{ai}(x_1, x_2)\left(1 + O\left(N^{-\frac{1}{6}}\right) + O_p\left(N^{-\frac{1}{4}}\right)\right), i = 1, 2.$

Further iterations do not improve these rates. In case b), we have to switch from the global to the local part of the iteration after 4 steps.

6.3 The inflation factor

In this section, we argue why we have chosen $\rho = \frac{1}{12}$ and discuss some alternatives. Note that, from the derivation in the last section, the number of iterations of the plug-in algorithm crucially depends on the choice of ρ. We again first consider the case where we are only interested in suitable global bandwidths.

From Theorem 4 and its proof the two main terms of the approximation error $\hat{\Lambda}_{i\ell} - \Lambda_{i\ell}$ are the bias term, which is of order $O\left(N^{2\rho} b_N^2\right)$, and the variance term, which is of order $O\left(\frac{1}{N(N^\rho b_N)^6}\right)$ if we use the inflation factor N^ρ for the bandwidths of the estimates $\hat{\Lambda}_{i\ell}$. Let us assume that we already have reached the optimal rate $b_N = N^{-\frac{1}{6}}$. Then, the bias and variance term are $O\left(N^{2\rho-\frac{1}{3}}\right)$ respectively $O\left(\frac{1}{N^{6\rho}}\right)$, i.e. the first term increases with ρ, the second decreases with ρ. Balancing both terms requires $\frac{1}{3} - 2\rho = 6\rho$, i.e. $\rho = \frac{1}{24}$, which would result in both error terms being of order $O\left(\frac{1}{N^{6\rho}}\right) = O\left(N^{-\frac{1}{4}}\right)$.

If we start the plug-in algorithm again with $\hat{h}_i^{(0)} = \frac{1}{\sqrt{N}}$, $i = 1, 2$, then with inflation factor $N^{\frac{1}{24}}$ we need 8 steps for reaching the optimal bandwidth rate $N^{-\frac{1}{6}}$ by the same kind of derivation as for $\rho = \frac{1}{12}$, and we have $\hat{h}_i^{(8)} = \beta_1^{(8)} N^{-\frac{1}{6}}\left(1 + o_p(1)\right)$ for some constant $\beta_i^{(8)}$, $i = 1, 2$. In the next, now the 9th, step we have to take into account the remainder terms as in Step 5 of the previous section. Again, we have $b_N = N^{-\frac{1}{6}}$, and now, the bandwidth used for second derivative estimates is of order $N^\rho b_N = N^{\frac{1}{24}} N^{-\frac{1}{6}} = N^{-\frac{3}{24}} = N^{-\frac{1}{8}}$. In that case the remainder terms of Theorem 4 are

$$
\begin{aligned}
R_{N,i\ell} &= o\left(N^{-\frac{1}{4}}\right) + o_p\left(N^{-\gamma} N^{-\frac{1}{4}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + o_p\left(N^{-\frac{\gamma}{2}} N^{-\frac{1}{4}}\right) \\
&\quad + o_p\left(\frac{\log N}{\sqrt{N}}\right) + o_p\left(\frac{\log N}{N^{\frac{3}{4}}}\right) \\
&= o\left(N^{-\frac{1}{4}}\right) + o_p\left(N^{-\frac{\gamma}{2}} N^{-\frac{1}{4}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{\Lambda}_{i\ell}(H) &= \Lambda_{i\ell} + O\left(N^{-\frac{1}{4}}\right) + o_p\left(N^{-\frac{\gamma}{2}} N^{-\frac{1}{4}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right) \\
&= \Lambda_{i\ell} + r_N(\gamma).
\end{aligned}
$$

Step 9: As, now, $\hat{h}_i^{(8)}$ satisfies Condition 2 with $\gamma = 0$, we have $r_N(\gamma) = r_N(0) = O\left(N^{-\frac{1}{4}}\right) + o_p\left(N^{-\frac{1}{4}}\right)$ and

$$
\left(\frac{\hat{\Lambda}_{22}}{\hat{\Lambda}_{11}}\right)^{\frac{1}{8}} \left(\frac{1}{\sqrt{\hat{\Lambda}_{11}\hat{\Lambda}_{22}} + \hat{\Lambda}_{12}}\right)^{\frac{1}{6}} = \left(\frac{\Lambda_{22}}{\Lambda_{11}}\right)^{\frac{1}{8}} \left(\frac{1}{\sqrt{\Lambda_{11}\Lambda_{22}} + \Lambda_{12}}\right)^{\frac{1}{6}} \left(1 + r_N(0)\right)
$$

which implies

$$\hat{h}_1^{(9)} = h_{a1}\left(1 + O\left(N^{-\frac{1}{4}}\right) + o_p\left(N^{-\frac{1}{4}}\right)\right)$$

and analogously for $\hat{h}_2^{(9)}, h_{a2}$. Hence $\hat{h}_i^{(9)}$ are a better approximation to h_{ai}, $i = 1, 2$, than $\hat{h}_i^{(8)}$ as the latter provides an approximation only up to a factor of order $1 + o_p(1)$.

Step 10: As $\hat{h}_i^{(9)}$ satisfies Condition 2 with $\gamma = \frac{1}{4}$, we have $r_N(\gamma) = O\left(N^{-\frac{1}{4}}\right) + o_p\left(\frac{1}{\sqrt{N}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right) = O\left(N^{-\frac{1}{4}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right)$ and we have

$$\hat{h}_i^{(10)} = h_{ai}\left(1 + O\left(N^{-\frac{1}{4}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right)\right), i = 1, 2.$$

Further iterations do not improve the approximation rate of $\hat{h}_i^{(10)}$ further, so the algorithm stops here.

For the local bandwidth selection with inflation factor $N^{\frac{1}{24}}$, we have again

$$\hat{\lambda}^{\frac{1}{6}}(x_1, x_2, H)\left|\frac{\hat{\lambda}_{22}(x_1, x_2, N^\rho H)}{\hat{\lambda}_{11}(x_1, x_2, N^\rho H)}\right|^{\frac{1}{12}} = \lambda^{\frac{1}{6}}(x_1, x_2)\left|\frac{\lambda_{22}(x_1, x_2)}{\lambda_{22}(x_1, x_2)}\right|^{\frac{1}{12}} + r_N'(\gamma),$$

where now with $\rho = \frac{1}{24}$

$$r_N'(\gamma) = O\left(N^{-\frac{1}{4}}\right) + o_p\left(N^{-\gamma}N^{-\frac{1}{4}}\right) + o_p\left(N^{-\frac{1}{3}}\right).$$

Step 9': For the first local iteration step with $N^\rho = N^{\frac{1}{24}}$, we have $\gamma = 0$ and $r_N'(\gamma) = r_N'(0) = O\left(N^{-\frac{1}{4}}\right) + o_p\left(N^{-\frac{1}{4}}\right)$, which implies

$$\hat{h}_i^{(9)}(x_1, x_2) = h_{ai}(x_1, x_2)\left(1 + O\left(N^{-\frac{1}{4}}\right) + o_p\left(N^{-\frac{1}{4}}\right)\right), i = 1, 2.$$

Step 10': Similar to Step 10, $\gamma = \frac{1}{4}$, $r_N'(\gamma) = O\left(N^{-\frac{1}{4}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right)$, and we stop with

$$\hat{h}_i^{(10)}(x_1, x_2) = h_{ai}(x_1, x_2)\left(1 + O\left(N^{-\frac{1}{4}}\right) + o_p\left(N^{-\frac{1}{3}}\right)\right), i = 1, 2.$$

We summarise our findings to

Proposition 7. *Under the assumptions of Theorem 4, the algorithm of Section 3.1 achieves the following rates for inflation factor* $N^{\rho} = N^{\frac{1}{24}}$

a) $\hat{h}_i^{(10)} = h_{ai}\left(1 + O\left(N^{-\frac{1}{4}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right)\right), i = 1, 2.$

b) $\hat{h}_i^{(10)}(x_1, x_2) = h_{ai}(x_1, x_2)\left(1 + O\left(N^{-\frac{1}{4}}\right) + O_p\left(N^{-\frac{1}{4}}\right)\right), i = 1, 2.$

In case b), we have to switch from the global to the local part of the iteration after 8 steps.

The inflation factor $N^{\frac{1}{24}}$ provides a better approximation of the asymptotically optimal bandwidths compared to $N^{\frac{1}{12}}$, as the deterministic part of the approximation factor is $O\left(N^{-\frac{1}{4}}\right)$ instead of $O\left(N^{-\frac{1}{6}}\right)$. However, recall that we are really interested in the mise respectively mse-optimal bandwidth h_{0i} respectively $h_{0i}(x_1, x_2)$ from Corollary 7, which coincide with asymptotically optimal bandwidths h_{ai} respectively $h_{ai}(x_1, x_2)$ only up to an error term of order $O\left(N^{-\frac{1}{4}}\right)$. Therefore, it makes no sense to try to approximate the asymptotically optimal bandwidth better than that. Note that for $\rho = \frac{1}{24}$, we get from Proposition 7 and Corollary 7, e.g.,

$$\begin{aligned} \hat{h}_i^{(10)} &= \left(h_{0i} + O\left(N^{-\frac{1}{4}}\right)\right)\left(1 + O\left(N^{-\frac{1}{4}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right)\right) \\ &= h_{0i} + O\left(N^{-\frac{1}{4}}\right), i = 1, 2, \end{aligned}$$

and, using that h_{ai} is of order $N^{-\frac{1}{6}}$, for $\rho = \frac{1}{12}$, $\hat{h}_i^{(7)} = h_{0i} + O\left(N^{-\frac{1}{4}}\right)$, too. Therefore, for approximating, both choices of ρ finally lead to the same approximation error of the optimal bandwidth.

We prefer $\rho = \frac{1}{12}$ due to two reasons. On the one hand, the algorithm requires fewer iterations for achieving an approximation rate of h_{ai} which cannot be improved further. On the other hand, from the discussion at the beginning of this section it achieves that the random part of $\Lambda_{i\ell} - \hat{\Lambda}_{i\ell}$ is of minimal achievable order $O\left(\frac{1}{N^{6\rho}}\right) = O\left(\frac{1}{\sqrt{N}}\right)$, which leads to more stable estimates. We have to pay for it with a larger bias order $O\left(N^{2\rho-\frac{1}{3}}\right) = O\left(N^{-\frac{1}{6}}\right)$ instead of the balanced rate $O\left(N^{-\frac{1}{4}}\right)$. Note that by Corollary 5, $O\left(N^{-\frac{1}{6}}\right)$ almost is the rate of the difference $h_{ai} - h_{0i}$, if λ is only twice continuously differentiable, such that in this case, a further improvement of bias is not worthwhile. However, for four times continuously differentiable λ, which we have assumed in our derivation, $h_{ai} - h_{0i}$ is of order $O\left(N^{-\frac{1}{4}}\right)$ by Corollary 7. Nevertheless, we prefer a slightly more biased, but less variable and more stable estimate, i.e. $\rho = \frac{1}{12}$.

Analogous arguments hold for preferring $\rho = \frac{1}{12}$ for the local part of the algorithm.

Chapter 7

Bandwidth selection for spatial data ($d = 3$)

In this chapter, we give an overview over the plug-in method for bandwidth selection in dimension $d = 3$ which, together, with $d = 2$, is related to the kind of applications from material science which motivated this thesis. The results can be easily extended to arbitrary dimension d.

Note that for a diagonal bandwidth matrix H in dimension d with entries $h_1, \ldots, h_d$ satisfying (compare Theorem 4, e.g.)

$$h_i = \beta_i b_N \left(1 + o_p\left(N^{-\gamma}\right)\right), i = 1, \ldots, N,$$

the variance component of the asymptotic mse and mise expansion is of order $O\left(\frac{1}{N b_N^d}\right)$ whereas the squared bias component is of order $O\left(b_N^4\right)$ independent of d. This follows from Corollary 3. So, the asymptotic mse and mise increase with d. The approximation results, which are the justification for the plug-in algorithm in dimension $d = 2$, change accordingly, but can be used in exactly the same manner as in $d = 2$. Technically, the main reason is the observation that in substituting $u_i = \frac{z_i - x_i}{h_i}, i = 1, \ldots, d$, which is a main tool in the proofs, we get a factor $h_1 \cdot \ldots \cdot h_d = O\left(b_N^d\right)$ before the integral w.r.t. $u_1, \ldots, u_d$ which takes care of the dimension-dependent increase of the factor $\frac{1}{N\det H} = O\left(\frac{1}{Nb_N^d}\right)$ in the mse and mise-expansions.

7.1 Approximation results in dimension $d = 3$

Recall that for general dimension d, we have considered the kernel estimate

$$\hat{\lambda}(x, H) = \frac{1}{N}\sum_{j=1}^{N} K_H(x - X_j)$$

with

$$K_H(u) = \frac{1}{\det H} K\left(H^{-1}u\right)$$

with $x, u \in \mathbb{R}^d$ (compare Section 2.1). In Corollary 3, we have derived asymptotic approximations for the mean-squared error and integrated mean-squared error. They allow the calculation of asymptotically optimal bandwidths $h_{a1}, \ldots, h_{ad}$ and corresponding local bandwidths $h_{a1}(x), \ldots, h_{ad}(x), x \in (0,1)^d$ which are the basis of the plug-in algorithm. In the following, we discuss the main results which justify this algorithm for $d = 3$. Their derivation is largely identical to the case $d = 2$, and so, we only discuss the proofs as far as they differ from dimension $d = 2$ which has been extensively discussed in the previous chapters.

The only relevant difference between the cases $d = 3$ and $d = 2$ is the fact, that the formula for the asymptotically optimal global bandwidths becomes a bit more complicated as, in dimension 2, some terms are cancelling which is no longer true in higher dimensions. Recall from Corollary 3, that

$$\begin{aligned}
\text{amise}(H) &= \frac{Q_K}{N\det H} + \frac{1}{4}V_K^2 \int \left(\sum_{i=1}^{3} h_i^2 \lambda_{ii}(x)\right)^2 \mathrm{d}x \\
&= \frac{Q_K}{N\det H} + \frac{1}{4}\sum_{k,\ell=1}^{3} h_k^2 h_\ell^2 I_{k\ell}
\end{aligned}$$

with

$$I_{k\ell} = V_K^2 \int \lambda_{kk}(x)\,\lambda_{\ell\ell}(x)\,\mathrm{d}x = V_K^2 \Lambda_{k\ell}.$$

Here and in the following, we write $\int \cdots \mathrm{d}x$ as a shorthand notation for $\int\int\int \cdots \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}x_3$. To minimise amise$(H)$, we set the partial derivatives w.r.t. h_1, h_2, h_3 to 0 and get the following system of polynomial equations:

$$\sum_{k=1}^{3} h_\ell^2 I_{k\ell} h_k^2 \det H = \frac{Q_K}{N}, \ell = 1, 2, 3,$$

or, equivalently, using $I_{k\ell} = I_{\ell k}$,

$$h_1^4 I_{11} + h_1^2 h_2^2 I_{12} + h_1^2 h_3^2 I_{13} = \frac{Q_K}{N \det H}, \tag{7.1a}$$

$$h_2^4 I_{22} + h_1^2 h_2^2 I_{12} + h_2^2 h_3^2 I_{23} = \frac{Q_K}{N \det H}, \tag{7.1b}$$

$$h_3^4 I_{33} + h_1^2 h_3^2 I_{13} + h_2^2 h_3^2 I_{23} = \frac{Q_K}{N \det H}. \tag{7.1c}$$

Subtracting (7.1a) from (7.1b) respectively (7.1c), dividing by h_1^4 and setting $u = \frac{h_2^2}{h_1^2}, v = \frac{h_3^2}{h_1^2}$, we get

$$I_{22} u^2 + I_{23} uv - I_{13} v = I_{11}, \tag{7.2a}$$

$$I_{33} v^2 + I_{23} uv - I_{12} u = I_{11}, \tag{7.2b}$$

which implies

$$I_{22} u^2 + I_{12} u = I_{33} v^2 + I_{13} v. \tag{7.3}$$

Now, set $w = \frac{I_{33}}{I_{23}} v + u$, i.e. $u = w - \frac{I_{33}}{I_{23}} v$, and plug it into (7.2b) such that

$$I_{33} v^2 + I_{23} v \left(w - \frac{I_{33}}{I_{23}} v \right) - I_{12} \left(w - \frac{I_{33}}{I_{23}} v \right) = I_{11},$$

i.e. the quadratic term cancels, and

$$w = \frac{I_{11} - \frac{I_{12} I_{33}}{I_{23}} v}{I_{23} v - I_{12}}$$

such that

$$\begin{aligned} u &= w - \frac{I_{33}}{I_{23}} v = \frac{I_{11} - \frac{I_{12} I_{33}}{I_{23}} v - \frac{I_{33}}{I_{23}} I_{23} v^2 + \frac{I_{12} I_{33}}{I_{23}} v}{I_{23} v - I_{12}} \\ &= \frac{I_{11} - I_{33} v^2}{I_{23} v - I_{12}} = \frac{\Lambda_{11} - \Lambda_{33} v^2}{\Lambda_{23} v - \Lambda_{12}}. \end{aligned} \tag{7.4}$$

Plugging this into (7.3), we get

$$I_{22}\left(I_{11}-I_{33}v^2\right)^2+I_{12}\left(I_{23}v-I_{12}\right)\left(I_{11}-I_{33}v^2\right)=\left(I_{33}v^2+I_{13}v\right)\left(I_{23}v-I_{12}\right)^2,$$

i.e., taking into account that $I_{k\ell}=V_K^2\Lambda_{k\ell}, V_K>0$, v solves the polynomial equation of degree 4

$$\sum_{k=0}^{4}a_k v^k=0 \tag{7.5}$$

with coefficients

$$\begin{aligned}
a_0 &= \left(\Lambda_{11}\Lambda_{22}-\Lambda_{12}^2\right)\Lambda_{11},\\
a_1 &= \left(\Lambda_{11}\Lambda_{23}-\Lambda_{12}\Lambda_{13}\right)\Lambda_{12},\\
a_2 &= 2\left(\Lambda_{12}\Lambda_{13}\Lambda_{23}-\Lambda_{11}\Lambda_{22}\Lambda_{33}\right),\\
a_3 &= \left(\Lambda_{12}\Lambda_{33}-\Lambda_{13}\Lambda_{23}\right)\Lambda_{23},\\
a_4 &= \left(\Lambda_{22}\Lambda_{33}-\Lambda_{23}^2\right)\Lambda_{33},
\end{aligned}$$

where, in particular, $a_0, a_4 \geq 0$ by definition of $\Lambda_{k\ell}$ and the Cauchy-Schwarz inequality.

Finally, we use $h_2^2=uh_1^2, h_3^2=vh_1^2$ and plug it into (7.1a) to get

$$V_K^2\left(\Lambda_{11}+\Lambda_{12}u+\Lambda_{13}v\right)h_1^4=\frac{Q_K}{N\sqrt{uv}h_1^3}.$$

This implies the first part of the following result which is analogous to Theorem 3. The second part for the local bandwidths follows exactly as in dimension $d=2$. Note that the polynomial equation for v may be solved rather precisely by using solvers like the MATLAB routine “solve”.

Theorem 8. *Let the assumptions of Theorem 2 be satisfied, and let $d=3$.*

a) The bandwidths h_{a1}, h_{a2}, h_{a3} minimising amise(H) *of Corollary 3 are* *given by*

$$\begin{aligned} h_{a1}^7 &= \frac{Q_K}{V_K^2} \frac{1}{\sqrt{uv}\left(\Lambda_{11} + \Lambda_{12} u + \Lambda_{13} v\right)} \frac{1}{N}, \\ h_{a2} &= \sqrt{u} h_{a1}, \\ h_{a3} &= \sqrt{v} h_{a1}, \end{aligned}$$

where v solves (7.5) and u is given as a function of v by (7.4), provided (7.5) has a positive solution for which $u > 0$ and $\Lambda_{11} + \Lambda_{12}u + \Lambda_{13}v > 0$, too.

b) *The bandwidths $h_{a1}(x), h_{a2}(x), h_{a3}(x)$ minimising* $\operatorname{amse}(x, H)$ *of Corollary 3 are given by*

$$h_{ak}^7(x) = \frac{Q_K}{V_K^2} \lambda(x) \frac{\left|\prod_{i=1}^3 \lambda_{ii}(x)\right|^{\frac{1}{2}}}{\left|\lambda_{kk}(x)\right|^{\frac{7}{2}}} \frac{1}{3N},$$

where $k = 1, 2, 3$, provided $\operatorname{sgn}\lambda_{kk}(x) = 1, k = 1, 2, 3$, *or* $\operatorname{sgn}\lambda_{kk}(x) = -1, k = 1, 2, 3$.

Let us briefly discuss the condition on the signs of $\lambda_{kk}(x)$ in part b). It also appears in the analogous Theorem 3, but in higher dimensions it seems to be more and more restrictive. Let us briefly discuss why this condition appears. With $A = Q_K \lambda(x), B_i = V_K \lambda_{ii}(x), i = 1, 2, 3$, the amse is of the form

$$\operatorname{amse}(x, H) = \frac{A}{NP} + \frac{1}{4} S^2 \quad \text{with } P = \prod_{i=1}^3 h_i, S = \sum_{i=1}^3 B_i h_i^2.$$

Setting the partial derivatives w.r.t. h_1, h_2, h_3 to 0 results in

$$S B_k h_k^2 = \frac{A}{NP}, k = 1, 2, 3.$$

Subtracting the 2nd respectively 3rd equation from the first implies

$$S\left(B_1 h_1^2 - B_2 h_2^2\right) = 0 = S\left(B_2 h_2^2 - B_3 h_3^2\right).$$

If $S \neq 0$, then $B_1 h_1^2 = B_2 h_2^2 = B_3 h_3^2$ and $S = 3B_1 h_1^2$, such that

$$SB_1 h_1^2 = 3B_1^2 h_1^4 = \frac{A}{NP}.$$

Moreover, we can replace h_2, h_3 in P by $\sqrt{\frac{|B_1|}{|B_k|}} h_1$, $k = 2$ respectively $k = 3$, and we get the formula in b). If $S = 0$, on the other hand, which only may happen if not all B_i, i.e. all $\lambda_{ii}(x)$, have the same sign, we are in a degenerate situation where the bias term in $\operatorname{amse}(x, H)$ vanishes. Replacing h_i by ch_i for $c > 0$, we then have

$$\operatorname{amse}(x, cH) = \frac{A}{Nc^3 P} + \frac{1}{4} c^4 S^2 = \frac{A}{Nc^3 P} \to 0$$

for $c \to \infty$. This does not imply that very large bandwidths are recommendable, but it means that we would have to take higher-order terms in the bias expansion into account which are negligible in the usual amse formula (i.e. in the case where $S \neq 0$). A detailed examination of this case is beyond the scope of this chapter. Let us just mention that the assumption of b) covers situations where $\lambda(x)$ is locally convex respectively concave at x which will usually hold for most points.

For the further discussion, it is important to note that the positive solution of (7.5), if it exists, is an explicit, though complicated function of $a_0, \ldots, a_4$. This follows from the subsequent steps: Substituting $\bar{v} = v + \frac{1}{4}\frac{a_3}{a_4}$ and dividing by a_4, (7.5) becomes the reduced equation $\bar{v}^4 + p\bar{v}^2 + q\bar{v} + r = 0$, where p, q, r are polynomials in $\frac{a_i}{a_4}$, $i = 0, 1, 2, 3$. Euler's solution is representing the solution $\bar{v}$ as a linear combination of $\sqrt{y_i}, i = 1, 2, 3$, where y_1, y_2, y_3 are the solutions of the cubic equation

$$y_3 + py^2 + \left(p^2 - 4r\right) y - q^2 = 0.$$

This cubic equation can be brought into reduced form by substituting $\bar{y} = y + \frac{1}{3}p$, resulting in a cubic equation $\bar{y}^3 + 3\bar{p}\bar{y} + 2\bar{q} = 0$, where $\bar{p}, \bar{q}$ are polynomials in p, q, r. Finally, by e.g. the Cardano formula, the solutions $\bar{y}_1, \bar{y}_2, \bar{y}_3$ are functions of $\bar{p}, \bar{q}$ involving polynomials and cubic and square roots (compare, e.g., Arnold (1965), or any other textbook on calculus).

From Theorem 8, the optimal bandwidth rate for $d = 3$ is $N^{-\frac{1}{7}}$. Together with Corollary 3, we get the following analogue of Corollary 4:

Corollary 12. *Under the assumptions of Theorem 8, let* $h_i N^{\frac{1}{7}} \to c_i > 0$ *for* $N \to \infty$*,* $i = 1, 2, 3$*. Then, for* $N \to \infty$*,*

a) $\text{mse}\hat{\lambda}(x, H) - \text{amse}(x, H) = o\left(N^{-\frac{4}{7}}\right)$,

b) $\text{mise}\hat{\lambda}(\cdot, H) - \text{amise}(H) = o\left(N^{-\frac{4}{7}}\right)$.

Analogously to Corollary 5, we also have for the mse and mise optimal bandwidths $h_{0i}(x), h_{0i}, i = 1, 2, 3$:

Corollary 13. *Let the assumptions of Theorem 8 be satisfied. Then,*

a) $h_{0i}(x) = h_{ai}(x) + o\left(N^{-\frac{1}{7}}\right), i = 1, 2, 3,$ *for all* $x \in (0, 1)^3$,

b) $h_{0i} = h_{ai} + o\left(N^{-\frac{1}{7}}\right), i = 1, 2, 3.$

If we assume that λ is four times continuously differentiable, we have the analogue of Corollary 6:

Corollary 14. *Let the assumptions of Corollary 12 and Assumptions 6 and 7 be satisfied. Then, for* $h_i N^{\frac{1}{7}} \to c_i > 0$*,* $i = 1, 2, 3$*, we have for* $N \to \infty$

a) $\text{mse}\hat{\lambda}(x, H) - \text{amse}(x, H) = O\left(N^{-\frac{5}{7}}\right)$,

b) $\text{mise}\hat{\lambda}(\cdot, H) - \text{amise}(H) = O\left(N^{-\frac{5}{7}}\right)$.

Beyond differentiability conditions on $\lambda(x)$, we also need various regularity and in particular symmetry conditions on the kernel function $K(u)$ analogous to Assumption 13. We collect them as

Assumption 14. *i) $K(u)$ is a non-negative kernel function on $[-1,+1]^3$, integrating to 1.*

ii) K is twice continuously differentiable, and the second-order derivatives $K_{ii}(u) = \frac{\partial^2}{\partial u_i^2} K(u)$, $i = 1,2,3$, are Lipschitz continuous.

iii) K and its first-order derivatives $K_i(u) = \frac{\partial}{\partial u_i} K(u)$ satisfy the symmetry conditions

a) $K(\pm 1, u_2, u_3) = K(u_1, \pm 1, u_3) = K(u_1, u_2, \pm 1) = 0$,

$K_i(\pm 1, u_2, u_3) = K_i(u_1, \pm 1, u_3) = K_i(u_1, u_2, \pm 1) = 0$, $i = 1,2,3$, for all $-1 \le u_1, u_2, u_3 \le 1$.

b) $\int u_i K(u)\, \mathrm{d}u_i = 0$ for all u_j, $j \neq i$, $i = 1,2,3$.

c) $\int u_i^2 K(u)\, \mathrm{d}u = V_K$, $\int u_i^3 K(u)\, \mathrm{d}u = 0$, $i = 1,2,3$.

For reference, we also formulate the 3-dimensional version of Assumption 12:

Assumption 15. *λ is 4-times continuously differentiable on $[0,1]^3$, and the partial derivatives of order 4 are Hölder continuous with some exponent $\beta > 0$.*

We have the following analogue to Proposition 3:

Proposition 8. *Let the assumptions of Theorem 2, Assumptions 14 and 15 be satisfied. Then*

i) $\mathbb{E}\hat{\lambda}_{ii}(x,H) = \lambda_{ii}(x) + \frac{1}{2} V_K \sum_{\ell=1}^{3} h_\ell^2 \frac{\partial^2}{\partial x_\ell^2} \lambda_{ii}(x) + O\left(b_N^{2+\beta}\right)$,

ii) $\mathrm{var}\hat{\lambda}_{ii}(x,H) = \frac{1}{N h_i^4 \det H}\left(Q_K^{ii} \lambda(x) + O(b_N)\right)$ with Q_K^{ii} as in Proposition 3.

7.2 Asymptotics for amise and amse estimates involving random bandwidths in dimension $d = 3$

In this section, we formulate the 3-dimensional analogues of the results of Chapters 4 and 5. As those, we denote by

$$\hat{\Lambda}_{k\ell} = \int \hat{\lambda}_{kk}(x, H)\, \hat{\lambda}_{\ell\ell}(x, H)\, \mathrm{d}x$$

the estimates of $\Lambda_{k\ell}, 1 \leq k, \ell \leq 3$, which, by Theorem 8, determine the asymptotically optimal global bandwidths. The estimates of the second derivatives $\hat{\lambda}_{\ell\ell}(x, H)$ are given by

$$\begin{aligned}\hat{\lambda}_{\ell\ell}(x, H) &= \frac{\partial^2}{\partial x_\ell^2} \hat{\lambda}(x, H) \\ &= \frac{1}{N h_\ell^2 \det H} \sum_{j=1}^{N} K_{\ell\ell}\left(H^{-1}(x - X_j)\right), \ell = 1, 2, 3,\end{aligned}$$

where $K_{\ell\ell}(u)$ denotes the second derivative $\frac{\partial^2}{\partial u_\ell^2} K(u)$ of the kernel.

First note that as in Chapter 4 we may write, e.g.,

$$\begin{aligned}\hat{\Lambda}_{11} &= \frac{1}{N^2 h_1^6 h_2^2 h_3^2} \sum_{i,j=1}^{N} \int K_{11}\left(H^{-1}(x - X_i)\right) K_{11}\left(H^{-1}(x - X_j)\right) \mathrm{d}x \\ &= \frac{1}{N^2 h_1^5 h_2 h_3} \sum_{i,j=1}^{N} \int K_{11}(u)\, K_{11}\left(H^{-1}(X_i - X_j) + u\right) \mathrm{d}u \\ &= \frac{1}{N^2 h_1^5 h_2 h_3} \sum_{i,j=1}^{N} L_{11}\left(H^{-1}(X_j - X_i)\right),\end{aligned}$$

where $L_{11} = K_{11} * K_{11}$ again denotes the convolution of K_{11} with itself, i.e. we have the same kind of relation used in the proof of Theorem 4. We get the following analogue of Theorem 4.

Theorem 9. *Let the Assumptions 14 and 15 be fulfilled. Let h_1, h_2, h_3 be random bandwidths satisfying for some $\beta_i > 0$, $i = 1, 2, 3$, and a deterministic* *sequence* $0 < b_N \to 0$ *(*$N \to \infty$*)*

$$h_i = \beta_i b_N \left(1 + o_p\left(N^{-\gamma}\right)\right), i = 1, 2, 3$$

for some $\gamma \geq 0$*. Then, for* $i = 1, 2, 3$

$$\begin{aligned}\hat{\Lambda}_{ii} &= \int \lambda_{ii}^2(x)\,\mathrm{d}x + b_N^2 V_K \int \lambda_{ii}(x) \sum_{\ell=1}^{3} \beta_\ell^2 \frac{\partial^2}{\partial x_\ell^2} \lambda_{ii}(x)\,\mathrm{d}x \\ &\quad + \frac{1}{N\beta_1\beta_2\beta_3\beta_i^4 b_N^7} \int K_{ii}^2(u)\,\mathrm{d}u + R_{N,ii}, \qquad (7.6)\end{aligned}$$

and, for $i \neq \ell$

$$\begin{aligned}&\hat{\Lambda}_{i\ell} \\ = &\int \lambda_{ii}(x)\lambda_{\ell\ell}(x)\,\mathrm{d}x + b_N^2 \frac{V_K}{2} \int \sum_{k=1}^{3} \beta_k^2 \left\{\lambda_{ii}(x) \frac{\partial^2}{\partial x_k^2} \lambda_{\ell\ell}(x) + \lambda_{\ell\ell}(x) \frac{\partial^2}{\partial x_k^2} \lambda_{ii}(x)\right\} \mathrm{d}x \\ &+ \frac{1}{N\beta_1\beta_2\beta_3\beta_i^2\beta_\ell^2 b_N^7} \int K_{ii}(u) K_{\ell\ell}(u)\,\mathrm{d}u + R_{N,i\ell}, \qquad (7.7)\end{aligned}$$

where the remainder terms $R_{N,i\ell}, i, \ell = 1, 2, 3$*, are all of the order*

$$R_{N,i\ell} = o\left(b_N^2\right) + o_p\left(b_N^2 N^{-\gamma}\right) + O_p\left(\frac{1}{N b_N^6}\right) + o_p\left(\frac{N^{-\gamma}}{N b_N^7}\right) + o_p\left(\frac{\log N}{N^{\frac{3}{2}} b_N^7}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right).$$

Similarly, we have the following analogue of Theorem 5:

Theorem 10. *Under the assumptions of Theorem 9 with* $h_i = \beta_i b_N \left(1 + o_p\left(N^{-\gamma}\right)\right), i = 1, 2, 3$ *again, where in particular* $b_N^4 = o_p\left(N^{-\gamma}\right)$*, we have with* $Q_{ii} = \int K_{ii}^2(u)\,\mathrm{d}u, i = 1, 2, 3$*, for all* $x \in [0, 1]^3$ *satisfying* $\lambda(x) > 0$

$$\begin{aligned}\hat{\lambda}_{ii}(x, H) &= \lambda_{ii}(x) + \frac{1}{2} V_K \sum_{\ell=1}^{3} \beta_\ell^2 \frac{\partial^2}{\partial x_\ell^2} \lambda_{ii}(x)\, b_N^2 + o_p\left(b_N^2 N^{-\gamma}\right) \\ &\quad + \sqrt{\lambda(x) Q_{ii}}\, O_p\left(\frac{1}{\sqrt{N b_N^7}}\right) + o_p\left(\frac{N^{-\frac{\gamma}{2}}}{\sqrt{N b_N^7}}\right) + O_p\left(\frac{N^{-\gamma}}{\sqrt{N b_N^7}}\right).\end{aligned}$$

We also need the following analogue of Theorem 6.

Theorem 11. *Under the assumptions of Theorem 10, we have for fixed* $x \in (0,1)^3$

$$\begin{aligned}\hat{\lambda}(x,H) &= \lambda(x) + \frac{1}{2}V_K \sum_{\ell=1}^{3} \beta_\ell^2 \lambda_{\ell\ell}(x) b_N^2 + o_p\left(b_N^2 N^{-\gamma}\right) \\ &\quad + \sqrt{\lambda(x) Q_K} O_p\left(\frac{1}{\sqrt{N b_N^3}}\right) + o_p\left(\frac{N^{-\frac{\gamma}{2}}}{\sqrt{N b_N^3}}\right) + O_p\left(\frac{N^{-\gamma}}{\sqrt{N b_N^3}}\right).\end{aligned}$$

7.3 Tuning parameters of the plug-in algorithm in dimension $d = 3$

The plug-in algorithm in dimension 3 uses the same kind of iteration as in dimension 2 (compare Section 3.1). We start with initial guesses for the global bandwidths $\hat{h}_k^{(0)}, k = 1,2,3$, and choose an inflation factor N^ρ for the kernel estimates of second derivatives. We use the abbreviations

$$\begin{aligned}\hat{\Lambda}_{k\ell}^{(i)} &= \hat{\Lambda}_{k\ell}^{(i)}\left(N^\rho \hat{H}^{(i)}\right), \quad k,\ell = 1,2,3, \\ \hat{\lambda}_{k\ell}^{(i)}(x) &= \hat{\lambda}_{k\ell}\left(x, N^\rho \hat{H}^{(i)}(x)\right), \quad k,\ell = 1,2,3, \\ \hat{\lambda}^{(i)}(x) &= \hat{\lambda}\left(x, \hat{H}^{(i)}(x)\right),\end{aligned}$$

where $\hat{H}^{(i)}, \hat{H}^{(i)}(x)$ as usual denote the diagonal bandwidth matrices with entries $\hat{h}_k^{(i)}$ respectively $\hat{h}_k^{(i)}(x)$, $k = 1,2,3$. Then, the algorithm starts with iteratively improving the global bandwidths. If we are interested in local bandwidth optimisation, then we switch from a certain point (after i^* iterations) on to the iteration improving local bandwidths.

Step 0: Choose initial bandwidths $\hat{h}_k^{(0)} = \frac{1}{\sqrt{N}}, k = 1,2,3$.

Step 1: (global) For $i = 1, \ldots, i^*$, iterate:

Find solution $\hat{v}^{(i)}$ of $\sum_{\ell=0}^{4} \hat{a}_\ell^{(i-1)} v^\ell = 0$, $v > 0$, with $\hat{a}_\ell^{(i-1)}$ given by the equations following (7.5) with $\hat{\Lambda}_{k\ell}^{(i-1)}$ replacing $\Lambda_{k\ell}$. Calculate

$$\begin{aligned}
\hat{u}^{(i)} &= \frac{\hat{\Lambda}_{11}^{(i-1)} - \hat{\Lambda}_{33}^{(i-1)} \left(\hat{v}^{(i)}\right)^2}{\hat{\Lambda}_{23}^{(i-1)} \hat{v}^{(i)} - \hat{\Lambda}_{12}^{(i-1)}}, \\
\hat{h}_1^{(i)} &= \left\{ \frac{Q_K}{NV_K^2} \frac{1}{\sqrt{\hat{u}^{(i)} \hat{v}^{(i)}} \left(\hat{\Lambda}_{11}^{(i-1)} + \hat{\Lambda}_{12}^{(i-1)} \hat{u}^{(i)} + \hat{\Lambda}_{13}^{(i-1)} \hat{v}^{(i)} \right)} \right\}^{\frac{1}{7}}, \\
\hat{h}_2^{(i)} &= \sqrt{\hat{u}^{(i)}} \hat{h}_1^{(i)}, \\
\hat{h}_3^{(i)} &= \sqrt{\hat{v}^{(i)}} \hat{h}_1^{(i)}. \\
\hat{h}_k^{(i)} &= \max\left\{ \hat{h}_k^{(i)}, \frac{1}{2\sqrt{N}} \right\}, \quad \hat{h}_k^{(i)} = \min\left\{ \hat{h}_k^{(i)}, \frac{1}{2} \right\}, k = 1, 2, 3.
\end{aligned}$$

Step 2: (local) Set $\hat{H}^{(i^*)}(x) = \hat{H}^{(i^*)}$. For $i = i^* + 1, \ldots, j^*$,

$$\hat{h}_k^{(i)}(x) = \left(\frac{Q_K \hat{\lambda}^{(i-1)}(x)}{NV_K^2} \right)^{\frac{1}{7}} \frac{\left| \prod_{\ell=1}^{3} \hat{\lambda}_{\ell\ell}^{(i-1)}(x) \right|^{\frac{1}{14}}}{\left| \hat{\lambda}_{kk}^{(i-1)}(x) \right|^{\frac{1}{2}}} \frac{1}{\left\{ \hat{s}_k^{(i-1)} \left(\hat{s}_1^{(i-1)} + \hat{s}_2^{(i-1)} + \hat{s}_3^{(i-1)} \right) \right\}^{\frac{1}{7}}},$$

$k = 1, 2, 3$, where $\hat{s}_\ell^{(i-1)} = \mathrm{sgn} \hat{\lambda}_{\ell\ell}^{(i-1)}(x)$, $\ell = 1, 2, 3$.

Let us first discuss the choice of ρ. In dimension 3, the bias term of $\hat{\Lambda}_{k\ell} - \Lambda_{k\ell}$ is of order $O\left((N^\rho b_N)^2\right)$ by Theorem 9 if we choose the bandwidth $N^\rho b_N$ for calculating $\hat{\Lambda}_{k\ell}$. Correspondingly, the variance term, i.e. the 3rd term in the expansion of $\hat{\Lambda}_{k\ell}$, is of order $O\left(\frac{1}{N(N^\rho b_N)^7}\right)$. Assuming, that b_N is already of optimal order $N^{-\frac{1}{7}}$, the bias term becomes $O\left(N^{2\rho - \frac{2}{7}}\right)$ and the variance term $O\left(\frac{1}{N^{7\rho}}\right)$. Balancing both terms would lead to $\frac{2}{7} - 2\rho = 7\rho$, i.e. $\rho = \frac{2}{63}$, and both error terms would be of order $O\left(N^{-\frac{2}{9}}\right)$.

However, as discussed in Section 6.3, we prefer to concentrate on keeping the variance term small which leads to the condition $\frac{1}{N^{7\rho}} = \frac{1}{\sqrt{N}}$, i.e. $\rho = \frac{1}{14}$. In the following, we choose therefore the inflation factor $N^\rho = N^{\frac{1}{14}}$.

We now follow the arguments of Section 6.1 without describing all the details. We use $c_{k\ell}$ for some constants with changing values throughout the iteration, which are functions of $\beta_1, \beta_2, \beta_3$ and various known functionals of the kernel K.

Step 1: We start with $\hat{h}_k^{(0)} = \frac{1}{\sqrt{N}}, k = 1, 2, 3$, i.e. b_N from Theorem 9 is $\frac{1}{\sqrt{N}}$ too. For $\rho = \frac{1}{14}$, the dominant term in the expansions (7.6), (7.7) in the 3rd one, and as $N^\rho b_N = N^{-\frac{3}{7}}$, we get

$$\hat{\Lambda}_{k\ell}^{(0)} = c_{k\ell} N^2 \left(1 + o_p(1)\right).$$

Plugging this into the formulae for $\hat{h}_k^{(1)}$, we use that the solution of (7.5) is invariant w.r.t. rescaling of the coefficients $\hat{a}_k^{(0)}$, i.e. we may replace them by $\frac{1}{N^6}\hat{a}_k^{(0)} = \hat{\alpha}_k^{(0)} = O(1)\left(1 + o_p(1)\right)$, such that $\hat{v}^{(1)} = O_p(1)$. The same holds for $\hat{u}^{(1)}$, as from (7.4) and the asymptotic approximation of $\hat{\Lambda}_{k\ell}^{(0)}$, numerator and denominator are both of order N^2. Hence, $\hat{h}_k^{(1)}$ is of order $\left\{\frac{1}{N \cdot N^2}\right\}^{\frac{1}{7}} = N^{-\frac{3}{7}}$.

Step 2: For $i = 1$, we have $b_N = N^{-\frac{3}{7}}$ and $N^\rho b_N = N^{-\frac{5}{14}}$, such that

$$\hat{\Lambda}_{k\ell}^{(1)} = c_{k\ell} N^{\frac{3}{2}} \left(1 + o_p(1)\right)$$

and $\hat{h}_k^{(2)}$ is of order $\left\{\frac{1}{N \cdot N^{\frac{3}{2}}}\right\}^{\frac{1}{7}} = N^{-\frac{5}{14}}$.

Step 3: With $b_N = N^{-\frac{5}{14}}$ and $N^\rho b_N = N^{-\frac{2}{7}}$, we have

$$\hat{\Lambda}_{k\ell}^{(2)} = c_{k\ell} N \left(1 + o_p(1)\right)$$

and $\hat{h}_k^{(3)}$ is of order $\left\{\frac{1}{N \cdot N}\right\}^{\frac{1}{7}} = N^{-\frac{2}{7}}$.

Step 4: With $b_N = N^{-\frac{2}{7}}$ and $N^\rho b_N = N^{-\frac{3}{14}}$, we have

$$\hat{\Lambda}_{k\ell}^{(3)} = c_{k\ell} \sqrt{N} \left(1 + o_p(1)\right)$$

and $\hat{h}_k^{(4)}$ is of order $\left\{\frac{1}{N \cdot \sqrt{N}}\right\}^{\frac{1}{7}} = N^{-\frac{3}{14}}$.

Step 5: With $b_N = N^{-\frac{3}{14}}$ and $N^\rho b_N = N^{-\frac{1}{7}}$, we have

$$\hat{\Lambda}^{(4)}_{k\ell} = \Lambda_{k\ell} + c_{k\ell}\left(1 + o_p\left(1\right)\right)$$

as now the constant term is no longer negligible. The bandwidth approximation $\hat{h}^{(5)}_k$ has now reached the optimal rate $N^{-\frac{1}{7}}$.

For the further iterations, we have to take into account the remainder terms $R_{N,k\ell}$ of (7.6), (7.7), which with $b_N = N^{-\frac{1}{7}}$ and $N^\rho b_N = N^{-\frac{1}{14}}$ are

$$R_{N,k\ell} = o\left(N^{-\frac{1}{7}}\right) + o_p\left(N^{-\gamma}N^{-\frac{1}{7}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right),$$

and, correspondingly,

$$\begin{aligned}\hat{\Lambda}_{k\ell}\left(H\right) &= \Lambda_{k\ell} + O\left(N^{-\frac{1}{7}}\right) + o_p\left(N^{-\gamma}N^{-\frac{1}{7}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right) \\ &= \Lambda_{k\ell} + r_N\left(\gamma\right).\end{aligned}$$

Step 6: $\hat{h}^{(5)}_k$ satisfies the approximation assumption of Theorem 9 with $b_N = N^{-\frac{1}{7}}$ and $\gamma = 0$ such that $r_N\left(\gamma\right) = O\left(N^{-\frac{1}{7}}\right) + o_p\left(N^{-\frac{1}{7}}\right)$, and we get

$$\begin{aligned}\hat{h}^{(6)}_k &= h_{ak}\left(1 + O\left(N^{-\frac{1}{7}}\right) + o_p\left(N^{-\frac{1}{7}}\right)\right) \\ &= h_{ak}\left(1 + O\left(N^{-\frac{1}{7}}\right)\right)\left(1 + o_p\left(N^{-\frac{1}{7}}\right)\right).\end{aligned}$$

Step 7: We now have $b_N = N^{-\frac{1}{7}}, \gamma = \frac{1}{7}$ such that $r_N\left(\gamma\right) = O\left(N^{-\frac{1}{7}}\right) + o_p\left(N^{-\frac{2}{7}}\right)$, and

$$\hat{h}^{(7)}_k = h_{ak}\left(1 + O\left(N^{-\frac{1}{7}}\right) + o_p\left(N^{-\frac{2}{7}}\right)\right).$$

Step 8: Now $b_N = N^{-\frac{1}{7}}, \gamma = \frac{2}{7}$ such that $r_N\left(\gamma\right) = O\left(N^{-\frac{1}{7}}\right) + o_p\left(N^{-\frac{3}{7}}\right)$, and

$$\hat{h}^{(8)}_k = h_{ak}\left(1 + O\left(N^{-\frac{1}{7}}\right) + o_p\left(N^{-\frac{3}{7}}\right)\right).$$

Step 9: Now $b_N = N^{-\frac{1}{7}}, \gamma = \frac{3}{7}$ such that $r_N(\gamma) = O\left(N^{-\frac{1}{7}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right)$ as now $N^{-\gamma}N^{-\frac{1}{7}} = N^{-\frac{4}{7}}$ is of smaller order. Therefore, further iterations do not improve the approximation quality, and we stop with

$$\hat{h}_k^{(9)} = h_{ak}\left(1 + O\left(N^{-\frac{1}{7}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right)\right).$$

We now turn to the local bandwidth selection. After $i^* = 5$ global steps we have attained the optimal rate $b_N = N^{-\frac{1}{7}}$. From Theorem 10 and 11, we have with bandwidths of order $N^\rho b_N = N^{-\frac{1}{14}}$ respectively $b_N = N^{-\frac{1}{7}}$

$$\begin{aligned}
\hat{\lambda}^2_{kk}(x, N^\rho H) &= \lambda^2_{kk}(x) + O\left(N^{-\frac{1}{7}}\right) + O_p\left(N^{-\frac{1}{4}}\right) + r'_N(\gamma),\\
\hat{\lambda}^2(x, H) &= \lambda^2(x) + O\left(N^{-\frac{2}{7}}\right) + O_p\left(N^{-\frac{2}{7}}\right) + r_N(\gamma)
\end{aligned}$$

with

$$\begin{aligned}
r_N(\gamma) &= o_p\left(N^{-\gamma}b_N^2\right) + o_p\left(\frac{N^{-\frac{\gamma}{2}}}{\sqrt{Nb_N^3}}\right) + O_p\left(\frac{N^{-\gamma}}{\sqrt{Nb_N^3}}\right)\\
&= o_p\left(N^{-\frac{2}{7}}N^{-\frac{\gamma}{2}}\right) + O_p\left(N^{-\frac{2}{7}}N^{-\gamma}\right),\\
r'_N(\gamma) &= o_p\left(N^{-\frac{1}{7}}N^{-\gamma}\right) + o_p\left(N^{-\frac{1}{4}}N^{-\frac{\gamma}{2}}\right) + O_p\left(N^{-\frac{1}{4}}N^{-\gamma}\right)\\
&= o_p\left(N^{-\frac{1}{7}}N^{-\gamma}\right) + o_p\left(N^{-\frac{1}{4}}N^{-\frac{\gamma}{2}}\right)
\end{aligned}$$

and, therefore,

$$\hat{\lambda}^{\frac{1}{7}}(x, H)\frac{\left|\prod_{\ell=1}^3 \hat{\lambda}_{\ell\ell}(x)\right|^{\frac{1}{14}}}{\left|\hat{\lambda}_{kk}(x)\right|^{\frac{1}{2}}} = \lambda^{\frac{1}{7}}(x)\frac{\left|\prod_{\ell=1}^3 \lambda_{\ell\ell}(x)\right|^{\frac{1}{14}}}{\left|\lambda_{kk}(x)\right|^{\frac{1}{2}}} + R_N(\gamma)$$

with

$$R_N(\gamma) = O\left(N^{-\frac{1}{7}}\right) + O_p\left(N^{-\frac{1}{4}}\right) + o_p\left(N^{-\frac{1}{7}}N^{-\gamma}\right).$$

Step 6': Using $\hat{h}_k^{(5)}$, we have as in Step 6 $b_N = N^{-\frac{1}{7}}, \gamma = 0$, such that $R_N(\gamma) = O\left(N^{-\frac{1}{7}}\right) + o_p\left(N^{-\frac{1}{7}}\right)$, implying

$$\hat{h}_k^{(6)}(x) = h_{ak}(x)\left(1 + O\left(N^{-\frac{1}{7}}\right) + o_p\left(N^{-\frac{1}{7}}\right)\right).$$

Step 7': With $\hat{h}_k^{(6)}(x)$, we have $b_N = N^{-\frac{1}{7}}, \gamma = \frac{1}{7}$ such that $R_N(\gamma) = O\left(N^{-\frac{1}{7}}\right) + O_p\left(N^{-\frac{1}{4}}\right)$

$$\hat{h}_k^{(7)}(x) = h_{ak}(x)\left(1 + O\left(N^{-\frac{1}{7}}\right) + O_p\left(N^{-\frac{1}{4}}\right)\right).$$

Step 8': Here, the term $o_p\left(N^{-\frac{1}{7}}N^{-\gamma}\right) = o_p\left(N^{-\frac{2}{7}}\right)$ is negligible to the unavoidable term $O_p\left(N^{-\frac{1}{4}}\right)$. Therefore, further iterations do not improve the approximation quality.

We summarise the discussion in the following result which is the analogue of Theorem 7.

Theorem 12. *Under the assumptions of Theorem 11, we have that the plug-in algorithm at the beginning of this section obtains the following rates for approximating the asymptotically optimal bandwidths with the inflation factor* $N^\rho = N^{\frac{1}{14}}$.

a) $\hat{h}_k^{(9)} = h_{ak}\left(1 + O\left(N^{-\frac{1}{7}}\right) + o_p\left(\frac{\log N}{\sqrt{N}}\right)\right), k = 1, 2, 3.$

b) $\hat{h}_k^{(7)}(x) = h_{ak}(x)\left(1 + O\left(N^{-\frac{1}{7}}\right) + O_p\left(N^{-\frac{1}{4}}\right)\right), k = 1, 2, 3.$

Further iterations do not improve these rates. In case b), we switch after 5 steps from the global to the local iteration, i.e. $i^* = 5, j^* = 2$.

It may happen during the iteration that there is more than one positive solution $\hat{v}^{(i)}$ of the polynomial equation of degree 4 for which also the corresponding $\hat{u}^{(i)} > 0$. If this happens prior to the last step of the iteration, i.e. in local bandwidth selection always, then it does not matter which solution is chosen. The ambiguity of solutions does influence only a constant factor of the intermediate bandwidths, but not the rates which are all that counts here. We prefer to choose the larger one of both solutions to make the algorithm automatic. If we use the algorithm for global bandwidth selection and if more than one solution appears during the last step 9, then we recommend to look at the different resulting vectors of bandwidths and choose that one for which $h_1 h_2 h_3$ is largest. This should lead to a smoother appearance of the final intensity estimate.

Chapter 8

Application to simulated and real data

In this chapter, we consider simulated and real data in 2 dimension ($d = 2$). The kernel function used is quartic kernel:

$$K(u,v) = \begin{cases} \left(\frac{15}{16}\right)^2 (1-u^2)^2 (1-v^2)^2, & |u| \le 1, |v| \le 1; \\ 0, & \text{otherwise,} \end{cases}$$

so that

$$\begin{aligned} V_K &= \int_{-1}^{1}\int_{-1}^{1} u^2 K(u,v)\,\mathrm{d}u\mathrm{d}v \\ &= \left(\frac{15}{16}\right)^2 \int_{-1}^{1}\int_{-1}^{1} u^2 \left(1-u^2\right)^2 \left(1-v^2\right)^2 \mathrm{d}u\mathrm{d}v \\ &= \frac{1}{7}, \end{aligned}$$

$$\begin{aligned} Q_K &= \int_{-1}^{1}\int_{-1}^{1} K^2(u,v)\,\mathrm{d}u\mathrm{d}v \\ &= \left(\frac{15}{16}\right)^4 \int_{-1}^{1}\int_{-1}^{1} \left(1-u^2\right)^4 \left(1-v^2\right)^4 \mathrm{d}u\mathrm{d}v \\ &= \frac{25}{49}. \end{aligned}$$

8.1 Simulation results for the mixture bivariate normal distribution

Samples were generated from the following distribution:

$$X = (1 - W)\, X_0 + W X_1,$$

where X_0 is a bivariate normal distribution with mean $\begin{pmatrix} 0.5 & 0.5 \end{pmatrix}$ and covariance matrix $\begin{pmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_0^2 \end{pmatrix}$. Similarly, X_1 is a bivariate normal distribution with mean $\begin{pmatrix} 0.75 & 0.75 \end{pmatrix}$ and covariance matrix $\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix}$. σ_0 and σ_1 will be specified later. W is a Bernoulli random variable such that $W = 1$ with probability w and $W = 0$ with probability $1 - w$. Observations outside of $[0, 1]^2$ were rejected. The total sample size was $N = 500$.

We chose $w = 0.75$ and $\sigma_0^2 = \sigma_1^2 = \frac{1}{144}$ to illustrate how the bandwidth changes as the iteration proceeds. The iterated global bandwidths shown in Tables 8.1 and 8.2 illustrate that the iterated bandwidth becomes steady after 7 steps for global case and 6 steps for local case. Note that in Figure 8.1, the bandwidth at the point $(0.75, 0.75)$ in the local iteration steps (from step 5 onwards) decreases first and becomes steady afterwards. Probably it is due to the fact that at the point $(0.75, 0.75)$ at where the peak is located, the second derivatives of the intensity function $\lambda_{11}\,(x, y)$ and $\lambda_{22}\,(x, y)$ attain a very high value. Therefore, the algorithm chooses a relatively smaller bandwidth to accommodate this feature.

Iteration step t	$h_{a1}^{(t)}$	$h_{a2}^{(t)}$
0	0.04472136	0.04472136
1	0.07110748	0.07343927
2	0.08994116	0.09747116
3	0.09833928	0.108044528
4	0.10254668	0.11285674
5	0.10489176	0.11490613
6	0.105874438	0.11606366
7	0.10670693	0.11649906
8	0.10722224	0.11666709
9	0.10734786	0.11681582
10	0.10743845	0.11682822
11	0.10748641	0.11680769
12	0.10752665	0.11682475
13	0.10755093	0.11679476
14	0.10754256	0.11681992
15	0.10755222	0.11680228

Table 8.1: Iterated global bandwidths for observations generated from a 2-dimensional normal distribution as specified in Section 8.1 with $\sigma_0 = \sigma_1 = \frac{1}{12}$ and $w = 0.75$.

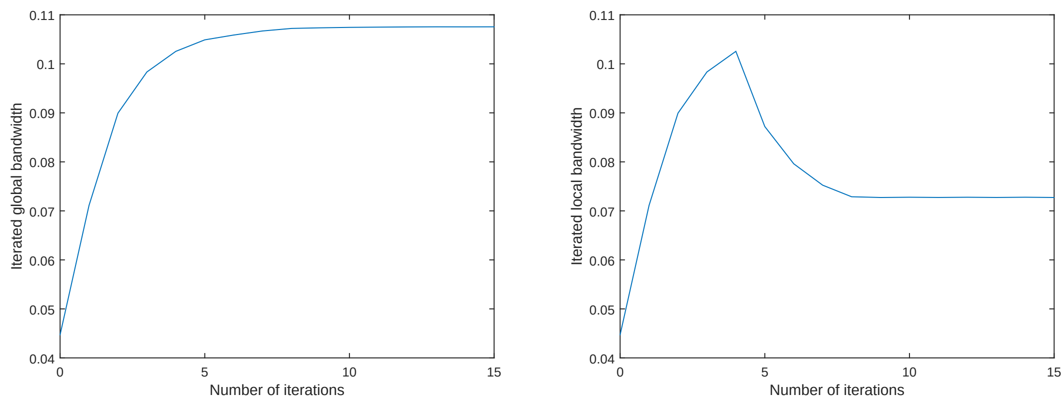

Figure 8.1: Plots of the iterated global bandwidths $\hat{h}_1^{(t)}$ and local bandwidth $\hat{h}_1^{(t)}(0.75, 0.75)$ for observations generated from a 2-dimensional normal distribution as specified in Section 8.1 with $\sigma_0 = \sigma_1 = \frac{1}{12}$ and $w = 0.75$.

Next, we conducted a Monte-Carlo study. We generated $M = 200$ samples with each sample having the sample size $N = 500$. For pure global bandwidth selection, there are 7 iteration steps. For the local case, the first 4 steps will be

Global iteration step t	$h_{a1}^{(t)}$	$h_{a2}^{(t)}$
0	0.04472136	0.04472136
1	0.07110748	0.07343927
2	0.08994116	0.09747116
3	0.09833928	0.10804453
4	0.10254668	0.11285674
Local iteration step t	$h_{a1}^{(t)}(0.75, 0.75)$	$h_{a2}^{(t)}(0.75, 0.75)$
5	0.08720507	0.09686396
6	0.07960701	0.09037397
7	0.07524026	0.08601521
8	0.07287823	0.08635201
9	0.07272479	0.08579517
10	0.07277950	0.08578627
11	0.07273217	0.08581165
12	0.07277762	0.08577490
13	0.07273078	0.08582061
14	0.07278107	0.08576696
15	0.07272592	0.08582843

Table 8.2: Iterated global and local bandwidths at $(x, y) = (0.75, 0.75)$ for observations generated from a 2-dimensional normal distribution as in Figure 8.1.

global iteration steps followed by 2 local iteration steps. For each sample, the estimated intensity using the global or local bandwidths at the points $(x, y) = (0.5, 0.5), \left(\frac{3}{4}, \frac{3}{4}\right), \left(\frac{1}{3}, \frac{1}{3}\right), \left(\frac{2}{3}, \frac{2}{3}\right)$, and the squared difference between the true intensity and estimated intensity $E_G = \left(\hat{\lambda}_G(x, y) - \lambda(x, y)\right)^2$ for pure global case and $E_L = \left(\hat{\lambda}_L(x, y) - \lambda(x, y)\right)^2$ for local case were computed. The whole process was repeated for $M = 200$ times and we computed $\bar{E}_G = \frac{1}{M}\sum_{k=1}^{M} E_G^{(k)}$ and $\bar{E}_L = \frac{1}{M}\sum_{k=1}^{M} E_L^{(k)}$ to compare the performance between global bandwidth and local bandwidth.

From Table 8.3, it seems that generally the intensity estimated by local bandwidth is better than the one by global bandwidth. In some cases, $\bar{E}_L$ is even much smaller than $\bar{E}_G$. On the other hand, there are some occasions in which $\bar{E}_L$ are a bit larger than $\bar{E}_G$, but most of which are the samples in which there are some invalid local bandwidths showing 0, NaN or infinity due

w	(x,y)	$(0.5, 0.5)$	$\left(\frac{3}{4}, \frac{3}{4}\right)$	$\left(\frac{1}{3}, \frac{1}{3}\right)$	$\left(\frac{2}{3}, \frac{2}{3}\right)$
$\frac{3}{4}$	$\sigma_0 = \frac{1}{12}$ $\sigma_1 = \frac{1}{12}$	$\bar{E}_G = 1.9688$ $\bar{E}_L = 2.1334$	$\bar{E}_G = 14.0260$ $\bar{E}_L = 7.3532$	$\bar{E}_G = 0.0215$ $\bar{E}_L = 0.0290*$	$\bar{E}_G = 0.5160$ $\bar{E}_L = 0.4704*$
	$\sigma_0 = \frac{1}{8}$ $\sigma_1 = \frac{1}{12}$	$\bar{E}_G = 0.2564$ $\bar{E}_L = 0.4874*$	$\bar{E}_G = 15.6275$ $\bar{E}_L = 7.3849$	$\bar{E}_G = 0.0400$ $\bar{E}_L = 0.0334*$	$\bar{E}_G = 0.4384$ $\bar{E}_L = 0.3769$
	$\sigma_0 = \frac{1}{8}$ $\sigma_1 = \frac{1}{16}$	$\bar{E}_G = 0.2956$ $\bar{E}_L = 0.3615*$	$\bar{E}_G = 48.7982$ $\bar{E}_L = 23.8801$	$\bar{E}_G = 0.0590$ $\bar{E}_L = 0.0275*$	$\bar{E}_G = 1.5306$ $\bar{E}_L = 2.3907*$

Table 8.3: The difference between the true and estimated intensity by global bandwidth ($\bar{E}_G$) and local bandwidth ($\bar{E}_L$). Those $\bar{E}_L$ marked with * indicated that at least 1 sample has local bandwidth equal to 0, NaN or infinity. Those $\bar{E}_L$'s are computed by ignoring those samples. Please refer to Table 8.4 for the number of samples with local bandwidths = NaN, zero or infinity out of those 200 samples.

w	(x,y)	$(0.5, 0.5)$	$\left(\frac{3}{4}, \frac{3}{4}\right)$	$\left(\frac{1}{3}, \frac{1}{3}\right)$	$\left(\frac{2}{3}, \frac{2}{3}\right)$
$\frac{3}{4}$	$\sigma_0 = \frac{1}{12} = \sigma_1$	0	0	70	4
	$\sigma_0 = \frac{1}{8}, \sigma_1 = \frac{1}{12}$	48	0	133	0
	$\sigma_0 = \frac{1}{8}, \sigma_1 = \frac{1}{16}$	54	0	131	89

Table 8.4: Number of samples with local bandwidths = NaN, zero or infinity out of 200 samples.

to numerical issues. The numbers of such samples are tabulated in Table 8.4. Tables 8.5 and 8.6 show the corresponding statistics with w changed to $\frac{1}{2}$.

Figure 8.2 shows the scatter plot of the points generated from a bivariate normal distribution with $\sigma_0 = \frac{1}{12} = \sigma_1$ and $w = 0.75$. Note that at the region where there are almost no points generated (e.g. $x > 0.7$ and $y < 0.5$), the local iteration cannot proceed as $\hat{\lambda}(x, y, H)$ is 0, leading to a sudden drop in the kernel estimate. At the peak where there is a high variation of intensity, e.g. $(x, y) = (0.75, 0.75)$, the estimate by local iteration is closer to the true intensity than the global iteration.

w	(x,y)	$(0.5, 0.5)$	$\left(\frac{3}{4}, \frac{3}{4}\right)$	$\left(\frac{1}{3}, \frac{1}{3}\right)$	$\left(\frac{2}{3}, \frac{2}{3}\right)$
$\frac{1}{2}$	$\sigma_0 = \frac{1}{12}$ $\sigma_1 = \frac{1}{12}$	$\bar{E}_G = 7.9731$ $\bar{E}_L = 4.4406$	$\bar{E}_G = 8.2182$ $\bar{E}_L = 4.4976$	$\bar{E}_G = 0.0685$ $\bar{E}_L = 0.0827*$	$\bar{E}_G = 0.2406$ $\bar{E}_L = 0.1615*$
	$\sigma_0 = \frac{1}{8}$ $\sigma_1 = \frac{1}{12}$	$\bar{E}_G = 0.9142$ $\bar{E}_L = 0.8799$	$\bar{E}_G = 13.5416$ $\bar{E}_L = 5.3731$	$\bar{E}_G = 0.0458$ $\bar{E}_L = 0.0564*$	$\bar{E}_G = 0.2257$ $\bar{E}_L = 0.2515$
	$\sigma_0 = \frac{1}{8}$ $\sigma_1 = \frac{1}{16}$	$\bar{E}_G = 0.7081$ $\bar{E}_L = 0.8110*$	$\bar{E}_G = 44.8710$ $\bar{E}_L = 16.1661$	$\bar{E}_G = 0.0832$ $\bar{E}_L = 0.0708*$	$\bar{E}_G = 0.9734$ $\bar{E}_L = 1.2096*$

Table 8.5: The difference between the true and estimated intensity by global bandwidth ($\bar{E}_G$) and local bandwidth ($\bar{E}_L$). Those $\bar{E}_L$ marked with * indicated that at least 1 sample has local bandwidth equal to 0, NaN or infinity. Those $\bar{E}_L$'s are computed by ignoring those samples. Please refer to Table 8.6 for the number of samples with local bandwidths = NaN, zero or infinity out of those 200 samples.

w	(x,y)	$(0.5, 0.5)$	$\left(\frac{3}{4}, \frac{3}{4}\right)$	$\left(\frac{1}{3}, \frac{1}{3}\right)$	$\left(\frac{2}{3}, \frac{2}{3}\right)$
$\frac{1}{2}$	$\sigma_0 = \frac{1}{12} = \sigma_1$	0	0	27	23
	$\sigma_0 = \frac{1}{8}, \sigma_1 = \frac{1}{12}$	0	0	136	0
	$\sigma_0 = \frac{1}{8}, \sigma_1 = \frac{1}{16}$	2	0	128	16

Table 8.6: Number of samples with local bandwidths = NaN, zero or infinity out of 200 samples.

We conclude that the bandwidth selection algorithm usually works quite well, where the local method is considerably better where the true underlying intensity function has a high curvature, i.e. at $\left(\frac{3}{4}, \frac{3}{4}\right)$, as it is to be expected considering the asymptotic mean-squared error approximation.

The local bandwidth selection causes sometimes numerical problems, in particular at points where the intensity function is close to 0 and where we have no or few observations in the neighbourhood. For stability, it would be advisable to modify the algorithm such that the local iterations are only applied to locations with not too few data in an appropriate neighbourhood, and, otherwise, the optimal global bandwidth will be used.

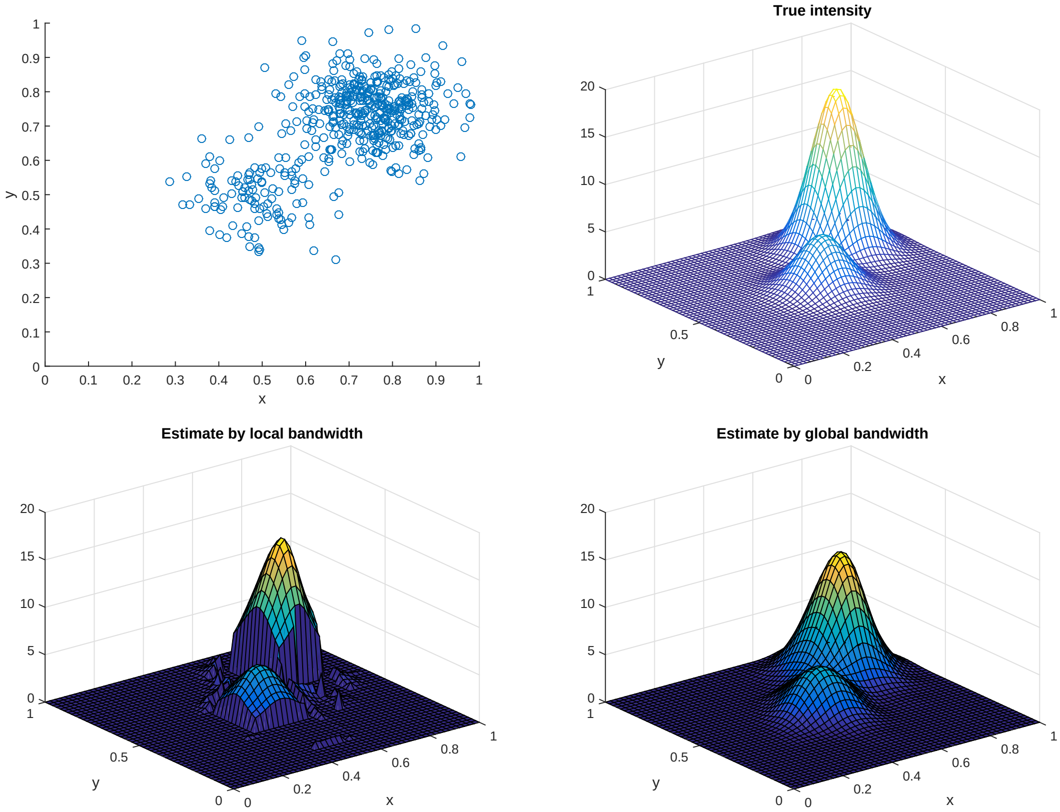

Figure 8.2: True intensity, estimates by local bandwidth and global bandwidth for a mixture bivariate normal distribution, $\sigma_0 = \sigma_1 = \frac{1}{12}$, $w = 0.75$.

8.2 Simulation results for a bivariate normal distribution with large correlation

500 observations were generated from a bivariate normal distribution with mean $\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$ and covariance matrix $\begin{pmatrix} \frac{1}{144} & \frac{0.8}{144} \\ \frac{0.8}{144} & \frac{1}{144} \end{pmatrix}$, i.e. the correlation coefficient between the first and second component is 0.8. 200 samples were generated. The differences between the estimated intensity and true intensity at the points $(0.4, 0.4)$, $(0.45, 0.45)$, $(0.55, 0.55)$and $(0.6, 0.6)$ are tabulated in Table 8.7. At the point $(0.5, 0.5)$ where the intensity function has the largest second order derivatives, the local bandwidth performs much better than the

(x,y)	$(0.4,0.4)$	$(0.45,0.45)$	$(0.5,0.5)$	$(0.55,0.55)$	$(0.6,0.6)$
$\bar{E}_G$	10.3390	43.0134	68.1889	41.7772	12.4429
$\bar{E}_L$	12.6652	31.9088	42.0751	30.7512	15.3291

Table 8.7: Difference between the estimated intensity and true intensity. True intensity is a bivariate normal distribution with mean $\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$ and covariance matrix $\begin{pmatrix} \frac{1}{144} & \frac{0.8}{144} \\ \frac{0.8}{144} & \frac{1}{144} \end{pmatrix}$.

(x,y)	True value	Estimate by global bandwidth	Estimate by local bandwidth
$(0.4,0.4)$	17.1631	15.7835	14.6840
$(0.45,0.45)$	31.2732	25.5565	26.4292
$(0.5,0.5)$	38.1972	32.2937	34.6547
$(0.55,0.55)$	31.2732	26.2438	27.6993
$(0.6,0.6)$	17.1631	12.8752	11.8492

Table 8.8: True values of the intensity function and the estimates for a bivariate normal distribution with mean $\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$ and covariance matrix $\begin{pmatrix} \frac{1}{144} & \frac{0.8}{144} \\ \frac{0.8}{144} & \frac{1}{144} \end{pmatrix}$.

global bandwidth does. However, at the points $(0.4, 0.4)$ and $(0.6, 0.6)$ which are a bit far away from the peak, the local bandwidth performs slightly worse than the global bandwidth does. The true values of the intensity function and the estimates from a typical sample are tabulated in Table 8.8.

8.3 Application to concrete fibres projected onto 2 dimensional plane

Data sets of fibre locations of several concrete test bodies were provided by the engineering partner, Daniele Casucci, of the research group GrK 1932. See also Casucci (2018). The image processing software MAVI developed by

Fraunhofer ITWM was used to extract the fibre locations. The plug-in algorithm was applied to estimate the intensity of fibre locations projected onto 2 dimensional plane. Both global and local bandwidth iteration algorithms were deployed. Figures 8.3, 8.4 and 8.5 show the scatter plots and kernel estimates for the fibre locations projected onto $x-y$, $x-z$ and $y-z$ planes of a concrete test body. For each scatter plot, there are three graphs showing the kernel estimates by global and local bandwidths. For local bandwidth iteration, recall that we need to start with 4 steps of global iteration and switch to local iteration on 5th step. Sometimes the global iteration will lead to a situation where $\hat{\lambda}_{11}\left(x, y, N^{\rho}H\right)$ and $\hat{\lambda}_{22}\left(x, y, N^{\rho}H\right)$ have opposite signs violating the assumptions of our theoretical results. We could not proceed the local iteration and continue the global iteration instead for those local steps. The graph with the title "Estimate by local and global bandwidth" shows the kernel estimate for which the local iteration cannot proceed and global iteration proceeds instead for those local steps. It is of interest to note that in Figure 8.4 which contains the graphs for $x-z$ plane, the kernel estimates by local bandwidth can better indicate the variation of the fibre intensity than the one by global bandwidth. Along the close-to-zero z-coordinate (approximately $z = 0.15$), the kernel estimate by local bandwidth shows that there is a particularly high concentration of fibres as illustrated in the scatter plot. Along the close-to-one z-coordinate (approximately $z = 0.85$) the kernel estimate by local bandwidth shows that there are two sharp peaks and a trough where the fibre intensity is particularly low as illustrated in the scatter plot. Clearly, the kernel estimate by global bandwidth cannot indicate that along $z = 0.85$, the fibre intensity is relatively low compared to the neighbourhood. In Figure 8.5, along the close-to-one z-coordinate (approximately $z = 0.9$), the fibre intensity is more or less constant. The estimate by local bandwidth shows a flat portion while the estimate by global bandwidth shows that the intensity changes continuously along $z = 0.9$ probably due to over-smoothing.

However, when the fibre intensity is too low, the local iteration fails to proceed. In fact, at the point $(x, z) = (0.4, 0.5)$, the estimate $\hat{\lambda}_{11}(x, z, N^{\rho}H)$ and $\hat{\lambda}_{22}(x, z, N^{\rho}H)$ are of opposite signs after 4 steps of global iterations.

Figures 8.6, 8.7 and 8.8 show the same graphs for another test body. As before, the local iteration can better indicate the variation of intensity than the global bandwidth can. Note that in all plots, the intensity estimates become small close to the boundary. That is purely due to the well-known boundary effects for function estimates based on local smoothing. The estimate pretends that there are no data outside $[0, 1]^2$ though they only are not observed. In its current form, our method does not give reliable results close to the boundaries. As a remedy, we could use special asymmetric boundary kernels (compare, e.g., Section 4.4 of Härdle (1990), for the regression problem). For purely global bandwidth selection, this already should suffice if the number of points close to the boundary is small compared to the whole observation region. For local bandwidth selection, however, we would have to modify the algorithm to include boundary kernels as the resulting kernel estimates have a different kind of asymptotics near the boundary. Of course, for $N \to \infty$ and $h \to 0$, the boundary effect will become smaller, but it is a problem for finite sample sizes except for situations like that in Sections 8.1 and 8.2 where all data are well in the interior.

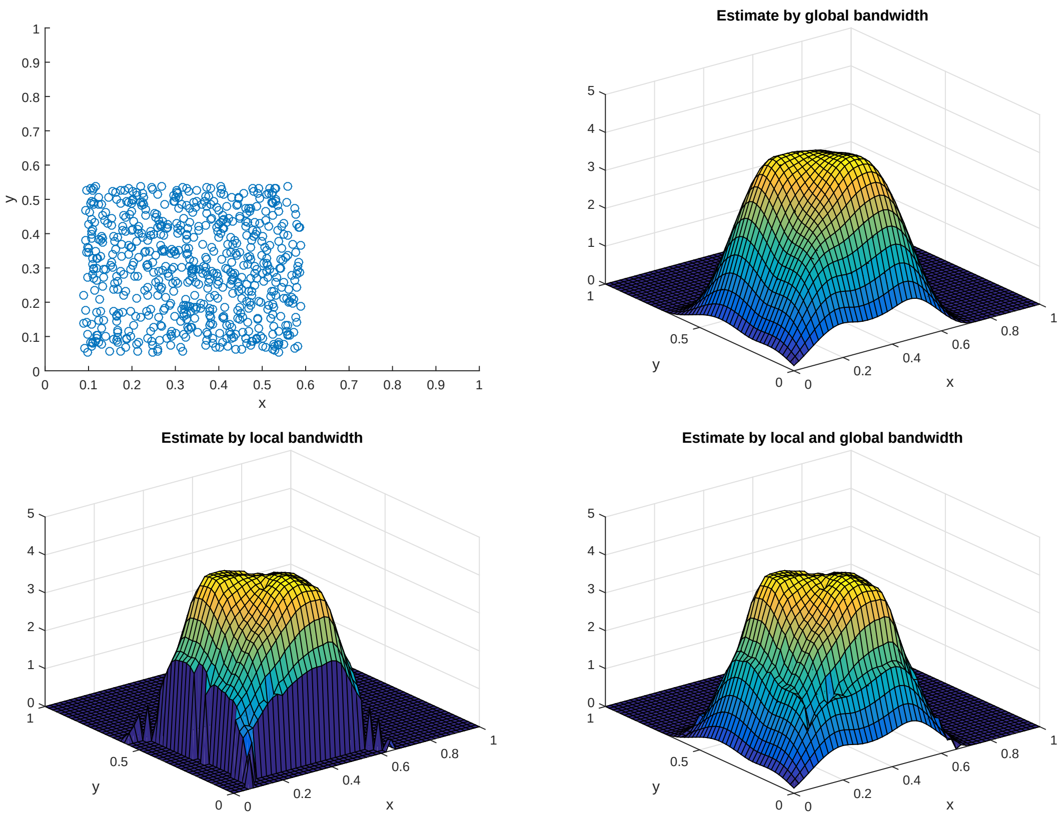

Figure 8.3: Scatter plots, kernel estimates by global and local bandwidths for the fibre locations projected onto $x - y$ plane, test body 1

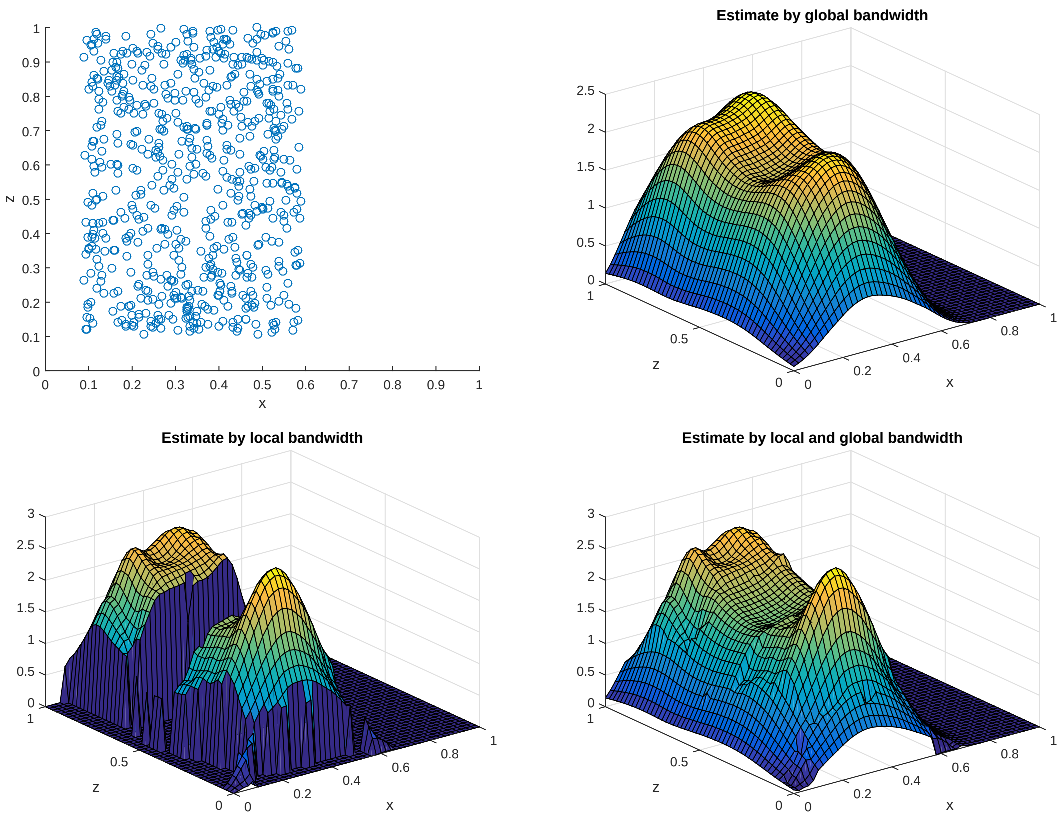

Figure 8.4: Scatter plots, kernel estimates by global and local bandwidths for the fibre locations projected onto $x - z$ plane, test body 1

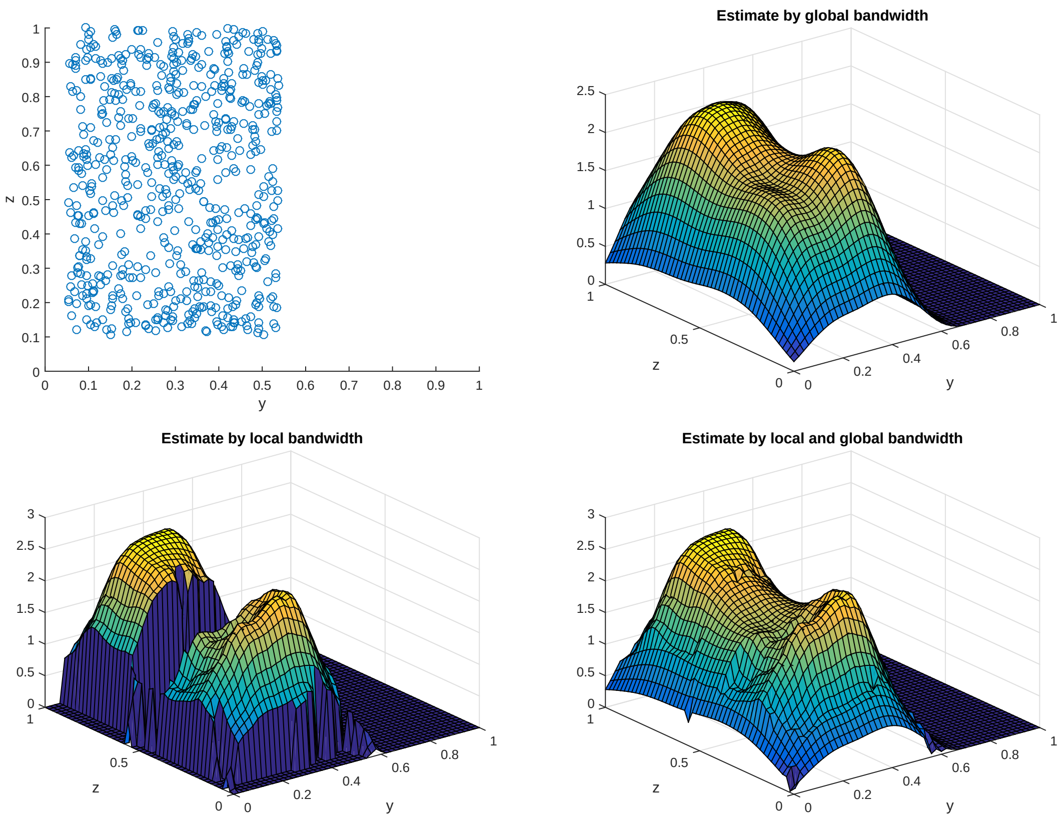

Figure 8.5: Scatter plots, kernel estimates by global and local bandwidths for the fibre locations projected onto $y - z$ plane, test body 1

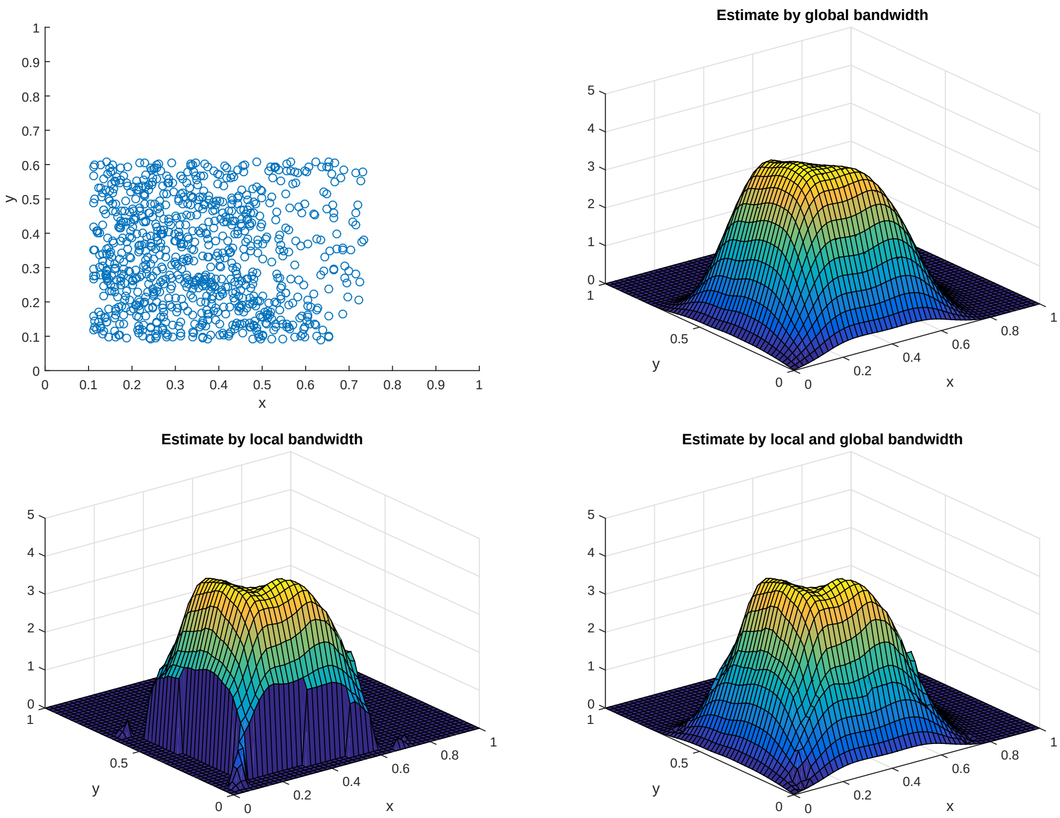

Figure 8.6: Scatter plots, kernel estimates by global and local bandwidths for the fibre locations projected onto $x - y$ plane, test body 2

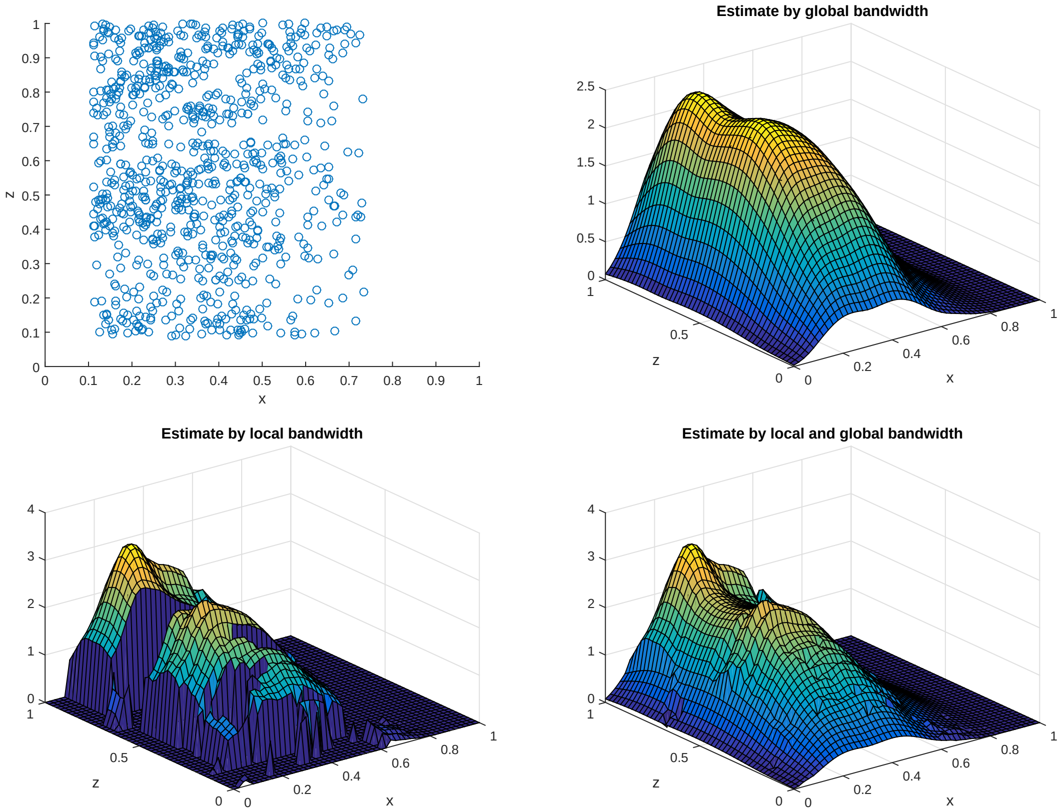

Figure 8.7: Scatter plots, kernel estimates by global and local bandwidths for the fibre locations projected onto $x - z$ plane, test body 2

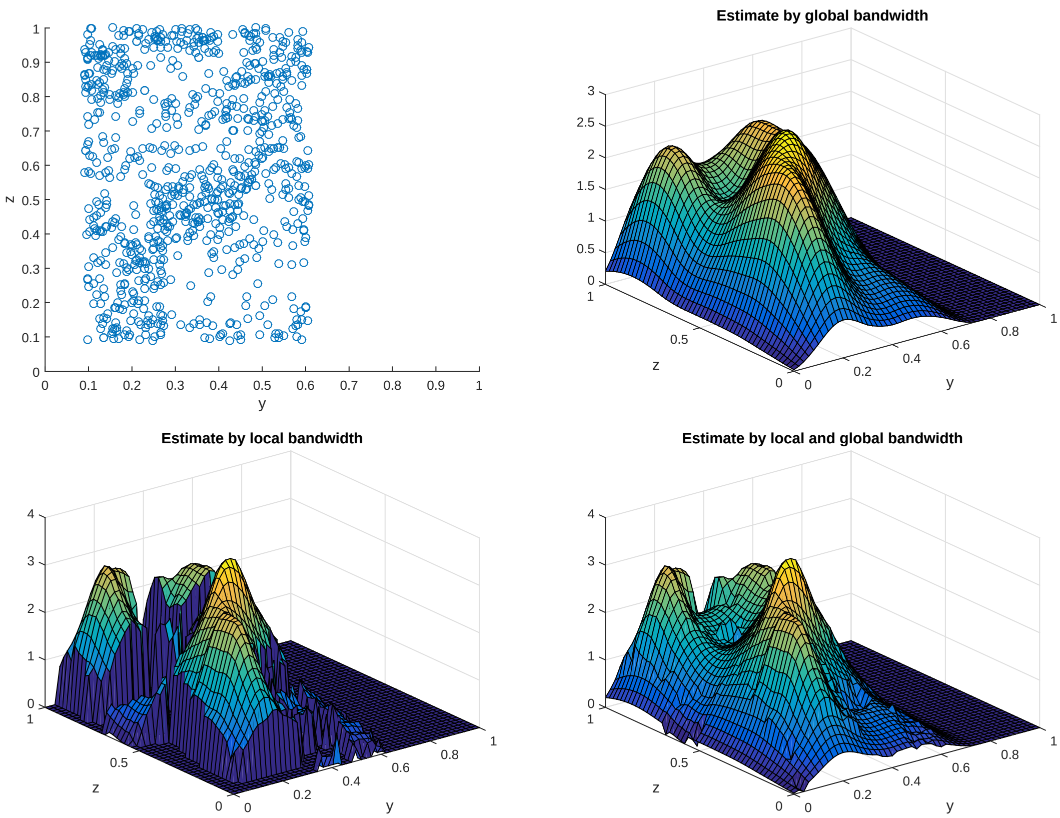

Figure 8.8: Scatter plots, kernel estimates by global and local bandwidths for the fibre locations projected onto $y - z$ plane, test body 2

Chapter 9

Conclusion

In this thesis, we have developed an iterative plug-in algorithm for bandwidth selection for kernel intensity estimation. In Chapter 2, we derived the asymptotic expressions for the mean squared error (mse) and mean integrated squared error (mise). From them, we derived, for $d = 2$, the asymptotically optimal bandwidth for local and global cases and proved that the difference between the bandwidth minimising the mse respectively mise and the one minimising the asymptotic mse respectively asymptotic mise is up to $o\left(N^{-\frac{1}{6}}\right)$. In case of smoother intensity function, the difference is even up to a smaller rate.

In Chapter 4, we derived the asymptotics for the integrated mean-squared error estimates with random bandwidths. In particular, we have derived the asymptotic expansions for $\hat{\Lambda}_{k\ell} = \int\int \hat{\lambda}_{kk}(x_1, x_2, H)\, \hat{\lambda}_{\ell\ell}(x_1, x_2, H)\, \mathrm{d}x_1 \mathrm{d}x_2, k, \ell = 1, 2$ for $d = 2$. Such expansions are utilised to choose the tuning parameters for the iterative algorithm. In Chapter 5, we derived the asymptotics for the local mean-squared error estimates with random bandwidths. In particular, we have derived the asymptotic expansions for $\hat{\lambda}_{ij}(x_1, x_2, H), i = 1, 2$ and $\hat{\lambda}(x_1, x_2, H)$ for $d = 2$. In Chapter 6, based on the asymptotic expansions derived in Chapters 4 and 5, we recommend an inflation factor N^{ρ} with $\rho = \frac{1}{12}$ because the algorithm requires fewer iteration steps and it leads to more stable estimates. Choosing $\rho = \frac{1}{24}$ leads to a better approximation of the asymptotically optimal bandwidths but it does not lead to a better approximation of the mse-optimal bandwidth. Moreover, choosing $\rho = \frac{1}{24}$ leads to more iteration steps. Therefore, we have chosen $\rho = \frac{1}{12}$. Based on them, we determined the number of iteration steps in the global and local bandwidth selection algorithm. Only a small number of iterations suffices. In Chapter 7, the asymptotic analysis and tuning parameter selection were repeated for $d = 3$.

In Chapter 8, the iterative bandwidth selection algorithm was applied to some simulated data sets. As expected, the local bandwidth can better indicate the variation of point intensity than the global one. However, special attention is required when only a few points are present. In this case, local bandwidth iteration cannot proceed and optimal global bandwidth will be used.

In the short future, the data analysis for 3 dimensional data sets will be done to complete the scope of this thesis.

References

Arnold, G. (1965). *NETZ-Formeln de Mathematatik*. Westermann, Braunschweig.

Brockmann, M. and Marron, J. (1991). Asymptotic optimality of the least-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stochastic processes and their Appl.* **38**, 157–165.

Casucci, D. (2018). Experimentelle Untersuchungen zum Zugkriechen in gerissenem ultrahochfestem stahlfaserverstärkten Beton. Ph.D thesis, University of Kaiserslautern. Submitted in Feb 2018 .

Cowling, A., Hall, P., and Phillips, M. J. (1996). Bootstrap confidence regions for the intensity of a poisson point process. *Journal of the American Statistical Association* **91**, 1516–1524.

Cronie, O. and van Lieshout, M. N. M. (2016). Bandwidth selection for kernel estimates of the spatial intensity function. **arXiv:1611.10221v1[stat.ME]**.

Diggle, P. (1985). A kernel method for smoothing point process data. *Applied Statistics* **34**, 138–147.

Diggle, P. (2014). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. Chapman and Hall.

Engel, J., Herrmann, E., and Gasser, T. (1994). An iterative bandwidth selector for kernel estimation of densities and their derivatives. *Nonparametric Statistics* **4**, 21–34.

Gasser, T., Kneip, A., and Köhler, W. (1991). A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association* **86**, 643–652.

Hall, P. and Marron, J. S. (1987). Estimation of integrated squared density derivatives. *Statistics and Probability Letters* **6**, 109–115.

Härdle, W. (1990). *Applied Nonparametric Regression.* Germany: Cambridge University Press.

Heidenreich, N.-B., Schindler, A., and Sperlich, S. (2013). Bandwidth selection for kernel density estimation: a review of fully automatic selectors. *Advances in Statistical Analysis* **97**, 403–433.

Jones, M. C., Marron, J. S., and Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association* **91**, 401–407.

Jones, M. C. and Sheather, S. J. (1991). Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statistics and Probability Letters* **11**, 511–514.

Park, B. U. and Marron, J. S. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association* **85**, 66–72.

Schneider, R. and Weil, W. (2008). *Stochastic and Integral Geometry.* Springer, Berlin-Heidelberg-New York.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley and Sons.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall/CRC.

Stone, C. (1984). An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics* **12**, 1285–1297.

Scientific Career

2007	Bachelor of Science, Major in Risk Management and Finance (held in 11/2010), The University of Hong Kong
2010	Master of Statistics (held in 11/2011), The University of Hong Kong
2011	Master of Philosophy (held in 01/2014), The University of Hong Kong
09/2010-05/2011	Teaching Assistant (part-time), Department of Statistics and Actuarial Science, The University of Hong Kong
06/2011-08/2011	Research Assistant, Department of Statistics and Actuarial Science, The University of Hong Kong
09/2014-2017	PhD in Mathematics, University of Kaiserslautern, Germany

Wissenschaftlicher Werdegang

2007	Bachelor of Science, Vertiefung in Risikomanagement und Finanzen (erhalten 11/2010), The University of Hong Kong
2010	Master in Statistik (erhalten 11/2011), The University of Hong Kong)
2011	Master in Philosophie (erhalten 01/2014), The University of Hong Kong
09/2010-05/2011	Lehrassistent (Teilzeit), Fachbereich Statistik und Versicherungswissenschaft, The University of Hong Kong
06/2011-08/2011	Wissenschaftlicher Mitarbeiter, Fachbereich Statistik und Versicherungswissenschaft, The University of Hong Kong
09/2014-2017	Doktorand im Fachbereich Mathematik, Technische Universität Kaiserslautern, Deutschland