Journal of Multivariate Analysis 109 (2012) 236–253

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Testing for symmetries in multivariate inverse problems

Melanie Birke [a,*], Nicolai Bissantz [b]

[a] Universität Bayreuth, Mathematisches Institut, 95440 Bayreuth, Germany

[b] Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany

ARTICLE INFO

Article history:
Received 19 April 2011
Available online 29 March 2012

AMS subject classifications:
primary 62F10
62F12
62G08
62H15
secondary 62F40
62H35

Keywords:
Deconvolution
Goodness-of-fit
Inverse problems
Semi-parametric regression
Symmetry

ABSTRACT

We propose a test for shape constraints which can be expressed by transformations of the coordinates of multivariate regression functions. The method is motivated by the constraint of symmetry with respect to some unknown hyperplane but can easily be generalized to other shape constraints of this type or other semi-parametric settings. In a first step, the unknown parameters are estimated and in a second step, this estimator is used in the L_2-type test statistic for the shape constraint. We consider the asymptotic behavior of the estimated parameter and show that it converges with parametric rate if the shape constraint is true. Moreover, we derive the asymptotic distribution of the test statistic under the null hypothesis and furthermore propose a bootstrap test based on the residual bootstrap. In a simulation study, we investigate the finite sample performance of the estimator as well as the bootstrap test.

1. Introduction

Several kinds of shape constraints play an important role in many areas of research. Some of them can be characterized by a linear transformation of the variables. Symmetry is one example for such a shape constraint. For example, many objects or parts of objects are symmetric with respect to reflection or rotation. Symmetry can be used in image compression and also in image analysis to detect certain objects. If symmetry of a certain object is violated one can sometimes deduce some results from it. Usually, parts of the human body are (nearly) symmetric, e.g. the left hand is symmetric to the right hand, the left part of the face to the right part and so on. This is usually also true for the thermographic distribution of those parts. If in a thermographic image of both hands this symmetry is severely violated, this can be a hint to some inflammation in this part. Problems of this and similar type make testing for symmetry to a problem of considerable interest. The method of testing described below is not only restricted to symmetry but can be generalized to other shape constraints characterized by a linear transformation of the variables. Technically, modeling the object of interest as a multivariate function and using linear transformations, we end up with the problem of testing for shape constraints of a multivariate function.

Whereas several results exist which discuss the symmetry of density functions (see e.g. Ahmad and Li [1], Cabaña and Cabaña [6] and Dette et al. [9] among many others) only few authors have considered testing for symmetry of a regression function so far. Recent results have been presented in [5,2], where both are for the case of bivariate functions in direct regression models and for symmetry with respect to some known axis.

* Corresponding author.
E-mail addresses: melanie.birke@uni-bayreuth.de (M. Birke), nicolai.bissantz@rub.de (N. Bissantz).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2012.03.008

In some cases, it is not possible to observe the object of interest directly. This leads to an inverse problem. Testing for symmetry in inverse regression problems can be of even higher interest than testing for symmetry in direct regression models. The reason is as follows. Whereas, at least in bivariate settings, symmetry in direct regression models can approximately be recognized by simply looking at the data, symmetrical structures in the true object can lack any symmetry in the observed (indirect) data. Consider, for example, the well known convolution problem which commonly appears in image analysis where the true object is distorted by a so called point-spread function we can easily find situations (e.g. for asymmetric point-spread functions or if the point-spread function has a different axis of symmetry than the true object) where the symmetry is not visible in the image. To the best of our knowledge there are no methods for testing for symmetry in inverse regression problems so far. An equivalent statement is more than ever true for other shape constraints of this type.

In the following we will develop a testing procedure for equality of a d-variate function under two different linear transformations. One example will be the reflection symmetry with respect to a $(d-1)$-dimensional hyperplane. Although we motivate the problem by the case of a symmetry constraint, the theoretical results and their proofs will be formulated as general as possible. Since the parameter of the linear transformations is unknown we estimate it in a first step by minimizing an L_2-criterion function. If the true function really fulfills the shape constraint, we derive, under some regularity conditions, consistency with parametric rate of the estimator and show that it is asymptotically normally distributed. In a second step, we use the minimized criterion function as test statistic for the shape constraint and show that it is asymptotically normal. Since the problem under consideration is closely related to certain semi-parametric problems we will use similar techniques as Härdle and Marron [11]. However note the important differences, that our problem is inverse and our regression function is multivariate. In nonparametric regression tests based on such asymptotic distributions usually do not perform satisfactorily in finite samples because the convergence is very slow and there is the problem of dealing with a bias term. To avoid this problem we propose a bootstrap test based on residual bootstrap and investigate the finite sample performance of this test in a simulation study.

The rest of the paper is organized as follows. In Section 2 we describe the model and define the estimator for the hyperplane as well as the test statistic. The asymptotic behavior of both is considered in Section 3 while we show the finite sample performance in Section 4. Finally all proofs are deferred to the Appendix.

2. The model and test statistic

We consider the nonparametric inverse regression model

$$Y_{\mathbf{r}} = \Psi m(\mathbf{x}_{\mathbf{r}}) + \sigma\varepsilon_{\mathbf{r}} \tag{1}$$

with $\mathbf{x}_{\mathbf{r}} = (r_1/(n_1 a_{n_1}), \dots, r_d/(n_d a_{n_d}))^T$, $r_j = -n_j, \dots, n_j$ and $a_{n_j} \to 0, j = 1, \dots, d$ such that with increasing sample size we have observations on the whole $\mathbb{R}^d$. For the sake of simplicity we assume in the following that $n_j = n$ and $a_{n_j} = a_n$ such that $\mathbf{x}_{\mathbf{r}} = (r_1, \dots, r_d)^T/(na_n)$ and for fixed n we have observations on the compact set $I_n = [-1/a_n, 1/a_n]^d$. In (1) m is a two times continuously differentiable regression function, and Ψ is an operator which maps m to the convolution $m * \psi$ with a known convolution function ψ. Finally, with $\mathbf{r} = (r_1, \dots, r_d)$, $\{\varepsilon_{\mathbf{r}}\}_{\mathbf{r}\in\{-n,\dots,n\}^d}$ are independent identically distributed errors with $\mathrm{E}[\varepsilon_{\mathbf{r}}] = 0$, $\mathrm{E}[\varepsilon_{\mathbf{r}}^2] = 1$ and $\mathrm{E}[\varepsilon_{\mathbf{r}}^4] < \infty$. If $\mathbf{j} = (j_1, \dots, j_d)$, $j = j_1 + \cdots + j_d$ and m is j times continuously differentiable according to Bissantz and Birke [3]

$$\hat{m}^{(\mathbf{j})}(\mathbf{x}) = \sum_{\mathbf{r}\in\{-n,\dots,n\}^d} w_{\mathbf{r},\mathbf{j}}(\mathbf{x}) Y_{\mathbf{r}} \tag{2}$$

with

$$w_{\mathbf{r},\mathbf{j}}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}(nh^j a_n)^d} \int_{[-1,1]^d} \frac{(-i\omega)^j e^{-i\omega^T(\mathbf{x}-\mathbf{x}_{\mathbf{r}})/h}}{\Phi_\psi(\omega/h)} d\omega \tag{3}$$

is an appropriate estimate for $\frac{\partial^{j_1+\cdots+j_d}}{\partial x_1^{j_1}\dots\partial x_d^{j_d}} m$. If $j = 0$ we write $\hat{m}^{(\mathbf{0})}(\mathbf{x}) = \hat{m}(\mathbf{x})$ and $w_{\mathbf{r},\mathbf{0}}(\mathbf{x}) = w_{\mathbf{r}}(\mathbf{x})$. As an abbreviation we write in the following $\Psi m = g$. In (3) Φ_f denotes the Fourier transform of a function f.

We consider linear transformations $T_\theta, S_\theta : \mathbb{R}^d \to \mathbb{R}^d$ parameterized by $\theta \in B \subset \mathbb{R}^d$ with B compact which are two times continuously differentiable with respect to θ and for which T_θ^{-1} and S_θ^{-1} exist. The testing problem we will consider is if for some set A_θ

$$m(\mathbf{z}) = m(T_\theta S_\theta^{-1}\mathbf{z}) \quad \text{for all } \mathbf{z} \in A_\theta \tag{4}$$

or equivalently

$$m(T_\theta\mathbf{x}) = m(S_\theta\mathbf{x}) \quad \text{for all } \mathbf{x} \in A = T_\theta^{-1}A_\theta. \tag{5}$$

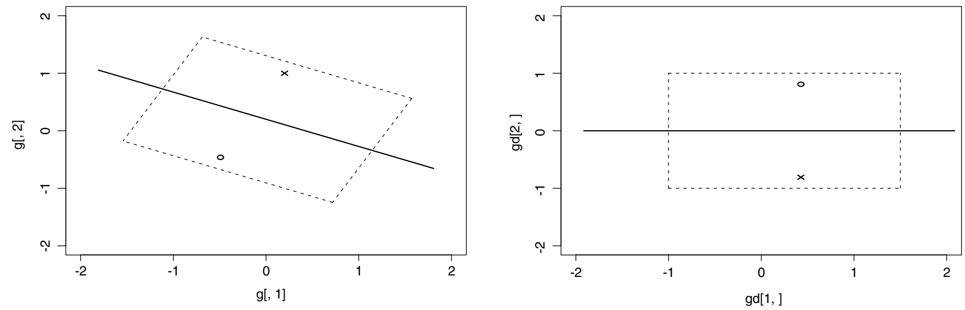

Fig. 1. Solid line: Symmetry axis $\mathfrak{g}_\theta$ (left) and transformed symmetry axis $T_\theta^{-1}\mathfrak{g}_\theta = \mathfrak{g}_0$ (right), rectangle with dashed line: A_θ (left) and A (right), $\circ$ respectively $\times$: z respectively $T_\theta S_\theta^{-1} z$, $z \in A_\theta$ (left) and $x = T_\theta^{-1} z$ respectively $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} x = S_\theta^{-1} z$, $x \in A$ (right).

To this end we will use

$$L(\theta) = \int_A (m(T_\theta \mathbf{x}) - m(S_\theta \mathbf{x}))^2 \, d\mathbf{x} \tag{6}$$

to check whether m exhibits such a structure on A_θ. The parameter ϑ of the true linear transformation minimizes this criterion function. In the following we will assume without loss of generality that $A = T_\theta^{-1} A_\theta$ is independent of θ and that $A_\theta \subset D$ for all $\theta \in B$ with D independent of θ.

Example. To illustrate the above definitions with our example of reflection symmetry, for every fixed $\theta \in \mathbb{R}^d$ mirroring m at the corresponding hyperplane can be realized by some linear functional $T_\theta S_\theta^{-1}$ where T_θ contains the shift of the hyperplane and the rotation and S_θ^{-1} is mainly the inverse of T_θ concatenated with the mirroring at the $(x_2, \ldots, x_d)$-hyperplane. To be even more precise we discuss reflection symmetry for the cases $d = 2$ and $d = 3$.

$\boldsymbol{d} = \mathbf{2}$: Here, the hyperplane reduces to a straight line parameterized by

$$\mathfrak{g}_\theta = \left\{ (\cos\theta_1, \sin\theta_1)^T \lambda + \theta_2 \left(-\sin(\theta_1), \cos(\theta_1)\right)^T | \lambda \in \mathbb{R} \right\}$$

with $\theta = (\theta_1, \theta_2)^T \in \mathbb{R}^2$ unknown such that mirroring $\mathbf{z} \in \mathbb{R}^2$ at that straight line can be obtained by transforming $\mathbf{z}$ to

$$T_\theta^{-1}\mathbf{z} = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix} \mathbf{z} - \begin{pmatrix} 0 \\ \theta_2 \end{pmatrix}, \tag{7}$$

mirroring at $\mathfrak{g}_0 = \left\{ (1, 0)^T \lambda | \lambda \in \mathbb{R} \right\}$ which gives

$$S_\theta^{-1}\mathbf{z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} T_\theta^{-1}\mathbf{z} \tag{8}$$

and transforming back, which finally yields

$$T_\theta S_\theta^{-1}\mathbf{z}.$$

The set A can e.g. be some rectangle $[a_1, b_1] \times [a_2, b_2]$, $a_1 < b_1$, $a_2 < b_2$ around the x_1-axis while $A_\theta = T_\theta(A)$ is the rotated and shifted rectangle. As a consequence, A and A_θ have the same Lebesgue measure. For an illustration also see Fig. 1.

$\boldsymbol{d} = \mathbf{3}$: Here, the hyperplane is given by the plane

$$\mathfrak{p}_\theta = \{(\cos\theta_1, 0, \sin\theta_1)^T \lambda + (-\sin\theta_1 \sin\theta_2, \cos\theta_2, \cos\theta_1 \sin\theta_2)^T \mu + \theta_3 (\sin\theta_1 \cos\theta_2, -\sin\theta_2, \cos\theta_1 \cos\theta_2)\}$$

with $\theta = (\theta_1, \theta_2, \theta_3)^T$ unknown. We define T_θ and S_θ as follows. For $\mathbf{z} \in \mathbb{R}^3$ let

$$T_\theta^{-1}\mathbf{z} = \begin{pmatrix} \cos\theta_1 & 0 & -\sin\theta_1 \\ \sin\theta_1 \sin\theta_2 & \cos\theta_2 & \cos\theta_1 \sin\theta_2 \\ \sin\theta_1 \cos\theta_2 & -\sin\theta_2 & \cos\theta_1 \cos\theta_2 \end{pmatrix} \mathbf{z} - \begin{pmatrix} 0 \\ 0 \\ \theta_3 \end{pmatrix}$$

and

$$S_\theta^{-1}\mathbf{z} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} T_\theta^{-1}\mathbf{z}.$$

Then, mirroring $\mathbf{z}$ at $\mathfrak{p}_0$ can be realized by the concatenation $T_\theta S_\theta^{-1}\mathbf{z}$.

Since m is not known, we estimate the criterion function (6) as

$$\hat{L}_n(\theta) = \int_A \left(\hat{m}(T_\theta \mathbf{x}) - \hat{m}(S_\theta \mathbf{x})\right)^2 d\mathbf{x} \tag{9}$$

and find the estimator of ϑ by minimizing $\hat{L}_n(\theta)$

$$\hat{\vartheta} = \arg \min_{\theta \in B_0 \times B_1} \hat{L}_n(\theta),$$

where $B_0 \subset \mathbb{R}^{d-1}$ is the compact set of all possible rotation angles and $B_1 \subset \mathbb{R}$ the compact set of all possible shifts. If $\hat{m}$ is continuously differentiable, we can equivalently solve

$$\hat{l}_n(\theta) = \operatorname{grad} \hat{L}_n(\theta) = 0 \tag{10}$$

to find $\hat{\vartheta}$.

3. Asymptotic inference

Throughout the rest of the article let us denote $|\mathbf{y}|^{\mathbf{k}} = \prod_{j=1}^{d} |y_j|^{k_j}$ for $\mathbf{y} \in \mathbb{R}^d$ and $\mathbf{k} \in \mathbb{R}^d$. The first important assumption for all results in this section is

Assumption 1. Let $L(\theta)$ be locally convex near the true parameter ϑ in (5), that is for every $\delta > 0$ exists a constant $\kappa(\delta) > 0$ such that $L(\theta) - L(\vartheta) > \kappa(\delta)$ for all $\|\theta - \vartheta\| > \delta$

To consider asymptotic theory, we further assume that Ψ is ordinary smooth, i.e. we consider mildly ill-posed problems in model (1). This can be summarized in the following assumption.

Assumption 2. The Fourier transform Φ_ψ satisfies

$$\int_{[-1,1]^d} \frac{1}{|\Phi_\psi(\omega)|} d\omega \leq \kappa_1 h^{-\beta}$$
$$\int_{[-1,1]^d} \frac{|\omega|^{2\mathbf{j}}}{|\Phi_\psi(\omega)|^2} d\omega \sim \kappa h^{-2\beta} \int_{[-1,1]^d} |\omega|^{2\mathbf{j}+2\beta}$$

for some $\beta > 0$ and $\kappa_1, \kappa \in \mathbb{R} \setminus \{0\}$, $\mathbf{j}$ such that $j_1 + \cdots + j_d \leq 2$.

Assumption 3. The Fourier transform Φ_m of m satisfies $\int_{\mathbb{R}^d} |\Phi_m(\omega)|\, |\omega|^{\mathbf{k}} d\omega < \infty$ for any multiindex $\mathbf{k}$ with $k_1 + \cdots + k_d \leq s$ for some $s > \max\{\beta + 1, 3\}$ and m is at least two times continuously differentiable.

Assumption 4. The convolution $g = Km$ of m with ψ satisfies $\int_{\mathbb{R}^d} |g(\mathbf{y})|\, |\mathbf{y}|^{\mathbf{k}} d\omega < \infty$ for any multiindex $\mathbf{k}$ with $k_1 + \cdots + k_d \leq r$ for some $r > 0$.

Assumption 5. The bandwidth h fulfills $h \to 0$, $(\log n)^{1/4}/n^d h^{2d} a_n^d = o(1)$, $n^d a_n^{3d/2} h^{2\beta+4+d/2} \to \infty$, $n^d a_n^{3d/2} h^{2\beta+2s+d/2-2} = O(1)$ and $n^d a_n^{r+d/2} = o(h^{\beta+s+d-1})$.

Assumption 1 is e.g. fulfilled for functions $m \neq 0$ for which ϑ is unique and (5) holds. In contrast to this, Assumption 1 does not hold if $m \equiv 0$. Note, that in general, $m \equiv c$ for some constant c would also be a counter example for Assumption 1 but is completely excluded in this setting because we need L_2-integrability for m on $\mathbb{R}^d$. The conditions on m for a unique ϑ of course strongly depend on the particular form of the linear transformations T_θ and S_θ and we therefore restrict our discussion to the case of reflection symmetry for $d = 2$. In this case we have T_θ and S_θ like in (7) and (8) and local convexity is e.g. given for the function

$$\exp\left(-12(\cos(0.3)x + \sin(0.3)y - 0.1)^2 - 3(\cos(0.3)y - \sin(0.3)x + 0.1)^2\right)$$
$$+ 0.5 \cdot \exp\left(-3(\cos(0.3)x + \sin(0.3)y - 0.1)^2 - 9(\cos(0.3)y - \sin(0.3)x - 0.4)^2\right).$$

This function has only one axis of symmetry and therefore $L(\theta) = 0$ has as unique solution ϑ which is the parameter of this symmetry axis. For all other possible values $\theta \in \mathbb{R}^2$ we have $L(\theta) > 0$. Because we only consider θ from some compact set B and continuity we have $L(\theta) - L(\vartheta) > \kappa(\delta)$ for all $\|\theta - \vartheta\| > \delta$. Local convexity of $L(\theta)$ is not given for the function

$$\exp(-12(\cos(0.3)x + \sin(0.3)y - 0.1)^2 - 3(\cos(0.3)y - \sin(0.3)x + 0.1)^2)$$

which has two axes of symmetry and although $\|\vartheta_1 - \vartheta_2\| > \delta$ for the parameters $\vartheta_1, \vartheta_2 \in \mathbb{R}^2$ of the two symmetry axes and some $\delta > 0$ we have $L(\vartheta_1) - L(\vartheta_2) = 0 < \kappa(\delta)$ for any choice of κ.

Assumption 3 is, for example fulfilled, if for $\operatorname{grad}(m)$ (and hence also for the products and sums in the integral) the $\mathbf{k}$-th derivative exists for all $k_1 + \cdots + k_d \leq \beta$. Note also, that in Assumption 5 a_n cannot be seen as regularization parameter since it is determined by the underlying design. Therefore, all conditions have to be read as conditions on h_n, s, β and r dependent on the rate of a_n.

Under the above conditions we can now discuss the asymptotic properties. We first consider the consistency and the asymptotic distribution of the estimator $\hat{\vartheta}$.

Theorem 1. *Under Assumptions 1–5 there is $\hat{\vartheta}_n \xrightarrow{P} \vartheta$ for $n \to \infty$.*

Theorem 2. *Under Assumptions 1–5, if $\hat{m}$ is continuously differentiable, $\hat{\vartheta}$ is defined by (10) and ϑ is the true parameter in (5), we have*

$$\sqrt{n^d a_n^d}\left(\hat{\vartheta} - \vartheta\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2 h^{-1}(\vartheta)\Sigma(\vartheta)(h^{-1}(\vartheta))^T)$$

with

$$\begin{aligned}
\Sigma(\theta) &= \frac{\sigma^2}{(2\pi^2\kappa)^d}\left|\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\|\omega\|^{\beta} I_{[-1,1]^d}(\omega)e^{-i\omega^T\mathbf{y}}d\mathbf{y}d\omega\right|^2\int_{\mathbb{R}^d}\sigma_\theta(\mathbf{u})\sigma_\theta(\mathbf{u})^T d\mathbf{u}\\
\sigma_\theta(\mathbf{u}) &= \left(\left(\frac{\partial}{\partial\theta}T_\theta\right)(T_\theta^{-1}(\mathbf{u})) - M_\theta N_\theta^{-1}\left(\frac{\partial}{\partial\theta}S_\theta\right)(T_\theta^{-1}(\mathbf{u})) - N_\theta M_\theta^{-1}\left(\frac{\partial}{\partial\theta}T_\theta\right)(S_\theta^{-1}(\mathbf{u})) - \left(\frac{\partial}{\partial\theta}S_\theta\right)(S_\theta^{-1}(\mathbf{u}))\right)^T\\
&\quad\times(\text{grad}\, m(\mathbf{u}))^T\\
h(\theta) &= 2\int_A\left(\text{grad}\, m(T_\theta\mathbf{x})\frac{\partial}{\partial\theta}T_\theta\mathbf{x} - \text{grad}\, m(S_\theta\mathbf{x})\frac{\partial}{\partial\theta}S_\theta\mathbf{x}\right)\left(\text{grad}\, m(T_\theta\mathbf{x})\frac{\partial}{\partial\theta}T_\theta\mathbf{x} - \text{grad}\, m(S_\theta\mathbf{x})\frac{\partial}{\partial\theta}S_\theta\mathbf{x}\right)^T d\mathbf{x}.
\end{aligned}$$

The second point of interest is to test whether the image obeys a structure like in (5). We use the test statistic

$$\hat{L}_n(\hat{\vartheta}) = \int_A\left(\hat{m}(T_{\hat{\vartheta}}\mathbf{x}) - \hat{m}(S_{\hat{\vartheta}}\mathbf{x})\right)^2 d\mathbf{x} \tag{11}$$

which has the following asymptotic distribution.

Theorem 3. *Under Assumptions 1–5, if ϑ is the true parameter in (5), we have*

$$\sigma_n^{-1/2}\left(\hat{L}_n(\hat{\vartheta}) - \frac{2\sigma^2}{(2\pi)^d n^d h^{2\beta+d} a_n^d}\int_A\int_{[-1,1]^d}|\omega|^{2\beta}\left|\sin\left(\frac{\omega^T S_\vartheta\mathbf{x}}{h}\right)\right|^2 d\omega d\mathbf{x}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

with

$$\sigma_n = \frac{32\sigma^4}{\kappa^4(2\pi)^{2d}n^{2d}h^{2d+4\beta}a_n^{2d}}\int_{\mathbb{R}^{2d}}|\omega|^{2\beta}|\eta|^{2\beta}\left|\int_A\sin\left(\frac{\omega^T S_\vartheta\mathbf{x}}{h}\right)\sin\left(\frac{\eta^T S_\vartheta\mathbf{x}}{h}\right)d\mathbf{x}\right|^2 d(\omega,\eta).$$

It can be shown similarly as in the proof of Theorem 4 in the Appendix, that the effective rate of convergence is $n^d h^{2\beta+d/2} a_n^{3d/2}$.

4. Simulations

4.1. Simulation framework

In this section we present the results of a simulation study. To this end we generate observations according to model (1), i.e.

$$Y_{(r,s)} = \Psi m(\mathbf{x}_{(r,s)}) + \sigma\varepsilon_{(r,s)}.$$

In our simulations, the noise terms are i.i.d. normally distributed with variance 1 and $\mathbf{x}_{(r,s)} = \left(\frac{r}{n}, \frac{s}{n}\right)$, $(r,s) \in \{-n, -n+1, \ldots, n-1, n\}^2$ are the coordinates of a grid with equidistant stepsize in both coordinates and with $a_n = 1$. In the following we use the parameter values $n = 50$ and σ (in dependence of the underlying function m) such that σ makes up for 1/10-th and 1/25-th of the maximum of the signal Ψm, which amounts to signal-to-noise ratios – defined as the mean signal of the image divided by σ – of $\approx$10 and $\approx$4, respectively. These values amount to rather poor signal-to-noise ratios, and in a practical application, S/N will frequently be larger and our simulations be expected to be conservative with respect to the performance of our method.

We consider two different "true" images m_1 and m_2 from which the data is generated. These images represent the cases of having a unique axis of symmetry (image m_1) and of not having any axis of symmetry at all (image m_2). The images are generated from the following bivariate functions (with $(x_t, y_t) \in \mathbb{R}^2$).

$$m_1(x, y) = \exp(-3 \cdot (4 \cdot x_t^2 + (y_t + 0.1)^2)) + 0.5 \cdot \exp(-3 \cdot (x_t^2 + 3 \cdot (y_t - 0.4)^2))$$

$$\begin{aligned} m_2(x, y) \; = \; & 0.5 \cdot \exp(-5 \cdot ((x_t - 0.3)^2 + 5 \cdot (y_t + 0.3)^2 +)) + 0.5 \cdot \exp(-5 \cdot ((x_t + 0.2)^2 + 5 \cdot (y_t - 0.3)^2)) \\ & + 0.5 \cdot \exp(-5 \cdot ((x_t + 0.5)^2 + 5 \cdot (y_t + 0.6)^2)), \end{aligned}$$

where

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -\delta \\ 0 \end{pmatrix}$$

are the coordinates of a coordinate system which is rotated by an angle $\alpha = -0.3$ with respect to the original coordinate system of y in counterclockwise direction and shifted (along the transformed y_t-axis) by $\delta = 0.1$. Hence, image m_1 is symmetric with respect to an axis of symmetry which passes the x-axis at $x = 0.1$ and is tilted away to the right from the y-axis by an angle of -0.3 rad., that is $\vartheta = (\alpha, \delta)^T = (-0.3, 0.1)^T$.

In accordance with model (1) for the observations, we do not assume to be able to observe m_i directly, but that at our disposal are only observations of the convolution of m_i, $i = 1, 2$ with a convolution function ψ given by

$$\psi(x, y) = \frac{\lambda}{2} \cdot \exp\left(-\lambda \cdot \sqrt{x^2 + 0.25 \cdot y^2}\right)$$

(with $\lambda = 5$). Fig. 2 shows the images of m_1 and m_2, their convolutions with Ψ and typical examples for estimates $\hat{m}_1$ and $\hat{m}_2$.

The convolution function ψ is symmetric with respect to the x- and y-axis of the (original) coordinate system, that is symmetric with respect to axes which are different to the axes of symmetry of m_1. In consequence, the convolved (observed) image Ψm_1 does not have any axis of axial symmetry. Note that this implies that testing for symmetry of m can in general not be substituted by testing for symmetry of Ψm, except under specific, strong assumptions on the symmetry properties of m and ψ. Instead, it is required that the observed image is deconvolved in a first step, with the symmetry test being performed in a subsequent second step.

In our simulations we use the spectral cut-off estimator (2) with equal bandwidths in both coordinate axes. From a visual inspection of 5 randomly selected noisy images and the associated estimates $\hat{m}$ we chose $h \approx 0.05$. This bandwidth was kept fixed in all subsequent simulations.

4.2. Critical functions and the distribution of estimated parameters and test statistics

In this section we describe the performance of the estimators for the symmetry axis parameters δ and α, and the properties of the underlying criterion function (9), which can, as already pointed out in Section 3, be used as test statistic for symmetry of the regression function, for the two different images considered here.

Fig. 3 shows the critical function $L_n(\delta, \alpha)$ both for the case of univariate estimation of the shift δ resp. the angle α (where the other parameter is assumed to be known) and for bivariate estimation of the pair (δ, α). For m_2 the criterion function for the selection of the shift only (top right panel) does not come close to the minimal value it attains for the symmetric function m_1 at all, but the situation is different for the estimation of the rotation angle, where the minimal values differ less strongly. Now consider the bivariate estimation of shift and rotation angle. For m_2, a complicated pattern appears without a distinct minimum.

Next, Fig. 4 shows the simulated distribution of the estimated parameters for rotation and shift for the various simulation setups. For m_2, which does not have an axis of symmetry at all, the critical function still shows clear minima of the criterion function if only one of the parameters was estimated. This is reflected in the right column of Fig. 4 for the estimated parameter, that is the value where the minimum is attained.

Finally, consider Fig. 5, which compares the simulated distributions of the test statistic for the case of one parameter estimated under H_0 (i.e. for m_1) with the results under H_1 (i.e. for m_2). In the latter case the distributions are shifted to significantly larger mean values, which reflects the fact that there exists no axis of symmetry. Moreover, their shape appears more symmetric than under H_0, where it is (much) more skewed to the right, similar to other L_2-based test statistics (e.g. [8,4,2]).

4.3. Testing for symmetry

In the final part of our simulations let us now turn to a more precise analysis of the performance of our proposed test for symmetry. Since the convergence of L_2-tests is known to be slow and the asymptotic distribution apparently depends on unknown parameters we use bootstrap quantiles as critical values for the test.

Hence, our testing procedure consists of two main parts. In the first *bootstrap* part we determine a bootstrap approximation to the distribution of the test statistics. In more detail, this consists of three steps: (1) to estimate the distribution of residuals, (2) to determine a "true image" $\hat{m}_B$ from which the bootstrap data are generated, and (3) to perform

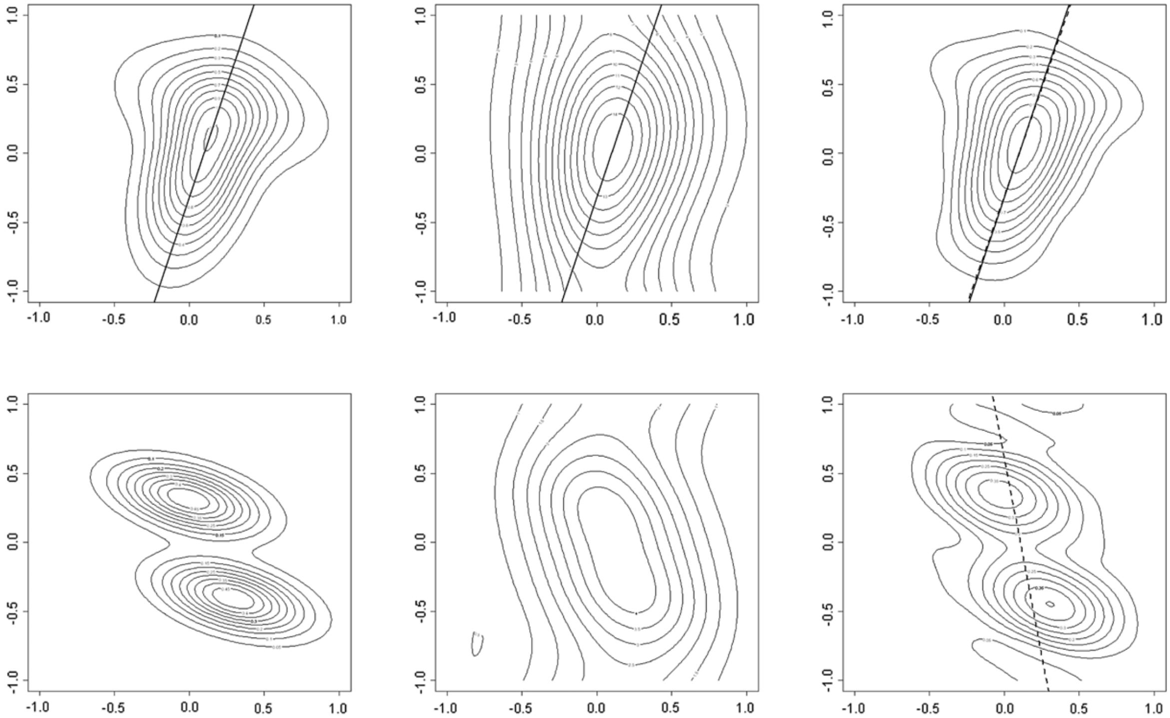

Fig. 2. True images and typical examples for the observed image and associated selected axis for m_1 (top panels) and m_2 (bottom panels). Left column: true functions, middle column: true function convolved with Ψ, right column: reconstructions from data with $n = 50$, $S/N = 25$. The full line indicates the true axis of symmetry and the dashed line the estimated symmetry axis. Note that m_2 is not symmetric to any axis, hence the full line is missing.

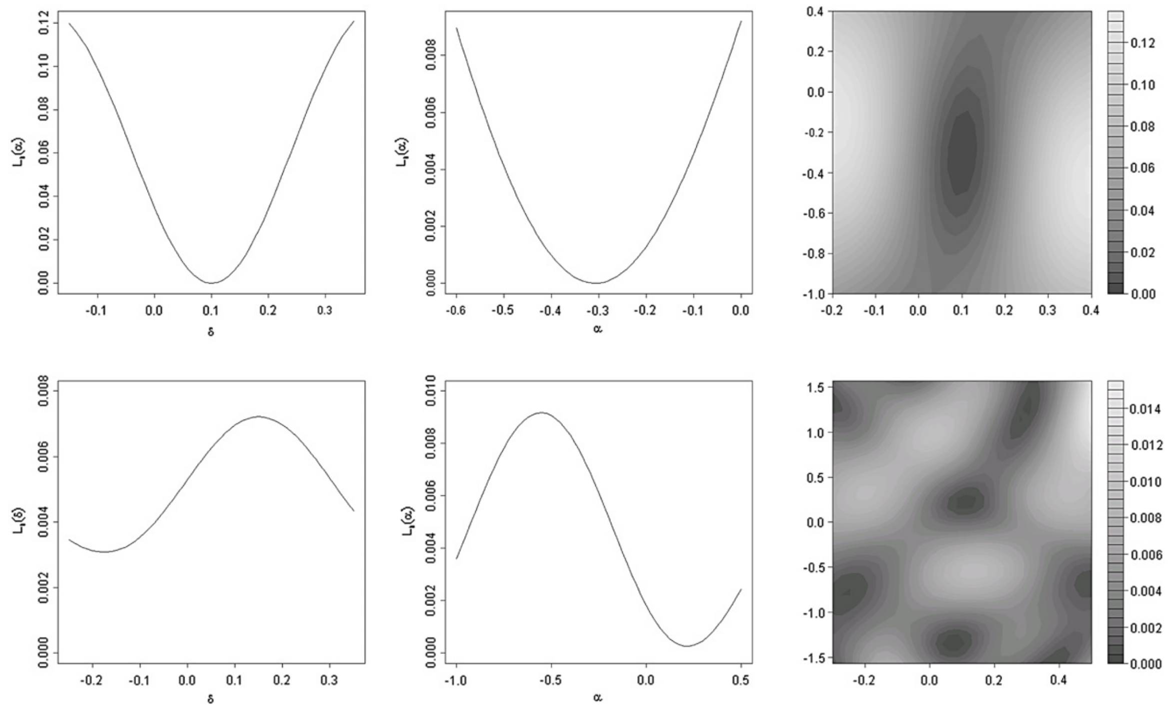

Fig. 3. True (noiseless) criterion function L_n for the translation axis for m_1 (top panels) and m_2 (bottom panels) for $n = 50$ and signal-to-noise ratio $S/N = 25$. Left column: $L_n(\delta)$ for $\alpha = -0.3$ assumed to be known, middle column: $L_n(\alpha)$ for $\delta = 0.1$ assumed to be known, right column: $L_n(\delta, \alpha)$.

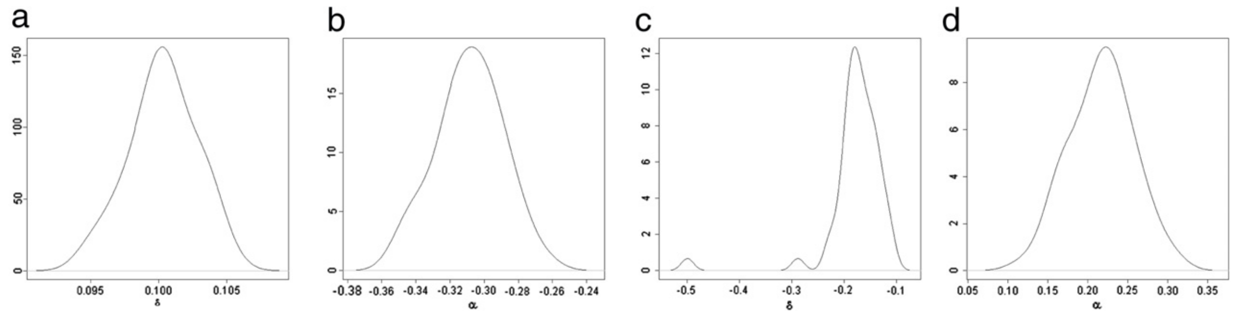

Fig. 4. Distribution of the estimated symmetry parameters for m_1 ((a) and (c)) and m_2 ((b) and (d)). (a) and (b): only shift estimated, (c) and (d): only rotation angle estimated for sample size parameter $n = 50$, and signal-to-noise ratio $S/N = 25$.

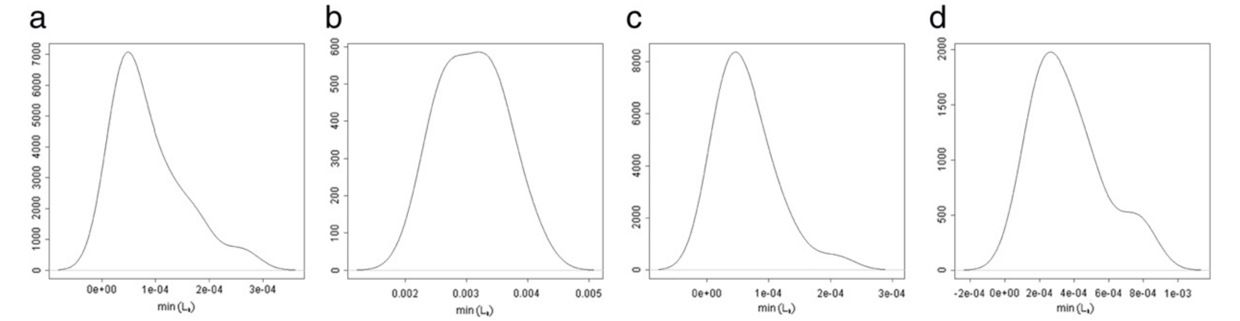

Fig. 5. Distribution of the test statistics under $H_0 : m = m_1$ ((a) and (c)) resp. $H_1 : m = m_2$ ((b) and (d)). (a) and (b): only shift estimated, (c) and (d): only rotation angle estimated for sample size parameter $n = 50$, and signal-to-noise ratio $S/N = 25$.

the bootstrap replications of the test statistic. The subsequent, second *test decision* part of the procedure is performed by computation of the test statistic for the original (observed) data and a decision based on this test statistic and the bootstrap approximation to its distribution. We now describe all steps in detail.

A. *Bootstrap part of the testing procedure*:

1. *Estimation of the distribution of residuals*: In our simulations we use a residual bootstrap as follows. In the first step we determine the empirical distribution of the residuals as the centered distribution of differences between the observations and an estimate $\Psi\hat{m}$ of Ψm. Then, in each of the bootstrap replications, we draw residuals from this distribution and generate bootstrap data as the sum of a suitable "true bootstrap image" $\hat{m}_B$ and these residuals.
2. *Determination of a "true image"* $\hat{m}_B$: The "true bootstrap image" $\hat{m}_B$ is generated as follows such that it obeys a known axis of symmetry and closely resembles the true (unknown) function m, assuming H_0 to be true.

   Step 2.1—Estimating m: Determination of an estimate $\hat{m}$ of m as described above.

   Step 2.2—Estimation of symmetry axis parameter: Minimization of the criterion function yields estimates $\hat{\delta}$ and/or $\hat{\alpha}$ of the symmetry axis parameter(s) of $\hat{m}$.

   Step 2.3—Backshift and rotation of $\hat{m}$: We shift and rotate $\hat{m}$ back by the estimated parameters $\hat{\delta}$ and/or $\hat{\alpha}$ (and, if applicable, the known true values of the other parameter). Under H_0, and if no noise would be present in the observed data, the new image $\check{m}$ would now be symmetric with respect to the y-axis.

   Step 2.4—Symmetrization: To ensure symmetry, we average the image over both sides of the y-axis, that is according to the scheme $\tilde{m}(x, y) = \frac{1}{2}\left(\check{m}(x, y) + \check{m}(-x, y)\right)$ for all (x, y).

   Step 2.5—Backrotation and shifting of the image to the estimated symmetry axis: The image $\tilde{m}$ is rotated and shifted such that it is symmetric with respect to the axis with the estimated parameters $\hat{\delta}$ and/or $\hat{\alpha}$, or – if applicable – the known values of shift and rotation, respectively. We call the resulting image $\hat{m}_B$.
3. *Bootstrap replications*: In the final step of the bootstrap part of the testing procedure we generate bootstrap data from the model $Y^*_{\mathbf{r}} = \Psi\hat{m}_B(\mathbf{x_r}) + \varepsilon^*_{\mathbf{r}}$, where $\varepsilon^*_{\mathbf{r}}$ are drawn independently from the empirical distribution of the residuals $\hat{\varepsilon}_{\mathbf{r}} = Y_{\mathbf{r}} - \Psi\hat{m}(\mathbf{x_r})$. From each set of bootstrap data the image is estimated and the minimal value of the criterion function, that is the test statistics, determined. In our simulations we always use $B = 200$ bootstrap replications. The $\lfloor B(1-\alpha)\rfloor$-th order statistic of all those bootstrap test statistics gives the critical value for the test.

Test decision part of the testing procedure:

In the second part of the testing procedure we use once more the estimate $\hat{m}$ of m described above. From this estimate we determine the test statistics $\hat{L}_n(\hat{\alpha}, \hat{\delta})$, that is the minimal value of the criterion function (11). The test decision by itself is

Table 1
Estimated rejection probabilities of the test for axial symmetry from 200 simulations each in case of estimating the axis-shift δ (with α known) under $H_0 : m = m_1$ and under an alternative $m = \kappa \cdot m_2 + (1 - \kappa) \cdot m_1$, respectively.

Hypothesis/Nominal level	$S/N = 10$			$S/N = 25$		
	5%	10%	20%	5%	10%	20%
$H_0 : m = m_1$	5.5%	10.5%	21.5%	6.5%	11.0%	20.5%
$H_1, \kappa = 0.1$	8.0%	12.0%	23.5%	8.5%	17.0%	27.0%
$H_1, \kappa = 0.2$	10.5%	20.0%	33.0%	54.0%	70.5%	81.5%
$H_1, \kappa = 0.4$	57.0%	71.5%	82.0%	100%	100%	100%

Table 2
Estimated rejection probabilities of the test for axial symmetry from 100 simulations each in case of estimating both the axis-shift δ and the angle of rotation α, and under an alternative $m = \kappa \cdot m_2 + (1 - \kappa) \cdot m_1$, respectively.

Hypothesis/Nominal level	$S/N = 10$			$S/N = 25$		
	5%	10%	20%	5%	10%	20%
$H_0 : m = m_1$	0%	2%	7%	6%	12%	20%
$H_1, \kappa = 0.4$	3%	5%	15%	8%	19%	39%
$H_1, \kappa = 1.0$	9%	19%	50%	78%	87%	96%

then to reject the null hypothesis of m obeying an axial symmetry to level α, if the test statistics for the original set of data is larger than the $(1 - \alpha)$-quantile of the bootstrap distribution of the test statistics.

In the following, we consider the functions

$$m_\kappa(x, y) = \kappa m_2(x, y) + (1 - \kappa)m_1(x, y), \quad \kappa = 0, 0.1, 0.2, 0.4, 1$$

to analyze the sensitivity of our test to small deviations from symmetry. Tables 1 and 2 summarize the simulated levels and power of the test for axial symmetry for the case of an unknown shift parameter δ only (with α known), and for the case that both parameters are unknown. The results demonstrate the substantial additional difficulty of disproving the existence of *any* axis of symmetry if both δ and α are unknown. Slightly acceptable results for the moderate sample size of $n = 50$ only appear for a comparable large deviation from symmetry (i.e. $\kappa = 1$). This effect is to a large part due to the complicated shape of the critical function in this case (cf. Fig. 3) with several local minima. If only the shift parameter is unknown, the test already performs well for small deviations from symmetry (e.g. $\kappa = 0.2$ for a signal-to-noise ratio of $S/N = 25$ or $\kappa = 0.4$ for $S/N = 10$).

Acknowledgments

This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823 project C4: Reconstruction of time variable distributions in statistical inverse problems) of the German Research Foundation and the BMBF project INVERS.

Appendix. Proofs

Theorem 4.

$$n^d h^{2\mathbf{j}+2\beta+d/2} a_n^{3d/2} \left(\int_B \left[\hat{m}^{(\mathbf{j})}(x) - m^{(\mathbf{j})}(x)\right]^2 dx - \frac{2^d \sigma^2 \prod_{k=1}^{d} (2(j_k + \beta_k) + 1)^{-1}}{\kappa \pi^d n^d h^{2\mathbf{j}+2\beta+d} a_n^{2d}} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^{(\mathbf{j})})$$

for $\mathbf{j} = (j_1, \ldots, j_k)$ *with* $j_1 + \cdots + j_k \le 2$ *and*

$$s^{(\mathbf{j})} = \frac{2\sigma^4}{\kappa^2 (2\pi)^{2d}} \lim_{n\to\infty} \prod_{l=1}^{d} a_n h^{4\beta_l+4j_l+1} \iint I_{[-1,1]}(\omega_l) I_{[-1,1]}(\eta_l) |\omega_l \eta_l|^{2j_l+2\beta_l} \frac{\sin^2\left(\frac{\omega_l - \eta_l}{a_n}\right)}{(\omega_l - \eta_l)^2} \, d\omega_l \, d\eta_l.$$

Proof. In the following we write the L^2-distance as a quadratic form and some bias terms and apply a central limit theorem by de Jong [7]. There is

$$\begin{aligned}\int_B \left[\hat{m}^{(\mathbf{j})}(x) - m^{(\mathbf{j})}(x)\right]^2 dx &= \int_B \left(\sum_{\mathbf{r}} w_{\mathbf{r},\mathbf{j}}(x)\varepsilon_{\mathbf{r}}\right)^2 dx + 2\int_B \left(\sum_{\mathbf{r}} w_{\mathbf{r},\mathbf{j}}(x)\varepsilon_{\mathbf{r}}\right)(\mathrm{E}[\hat{m}^{(\mathbf{j})}(x)] - m^{(\mathbf{j})}(x))dx \\ &\quad + \int_B (\mathrm{E}[\hat{m}^{(\mathbf{j})}(x)] - m^{(\mathbf{j})}(x))^2 dx \\ &= I_1^{(\mathbf{j})} + I_2^{(\mathbf{j})} + I_3^{(\mathbf{j})}.\end{aligned}$$

Using the definition of $w_{r,\mathbf{j}}(x)$ and Parseval's equality we obtain

$$\begin{aligned}I_1^{(\mathbf{j})} &= \frac{1}{(2\pi)^d n^{2d} h^{2j+d} a_n^{2d}} \int_{\mathbb{R}^d} |\omega|^{2j} \frac{I_{[-1,1]^d}(\omega)}{|\Phi_\psi(\omega/h)|^2} \left|\sum_{\mathbf{r}} e^{i\omega^T x_{\mathbf{r}}/h}\varepsilon_{\mathbf{r}}\right|^2 d\omega \\ &\quad - \frac{1}{(2\pi)^d n^{2d} h^{2j+d} a_n^{2d}} \int_{(B/h)^c} \left|\int_{\mathbb{R}^d} e^{-i\omega^T x}(-i\omega)^j \frac{I_{[-1,1]^d}(\omega)}{\Phi_\psi(\omega/h)} \sum_{\mathbf{r}} e^{i\omega^T x_{\mathbf{r}}/h}\varepsilon_{\mathbf{r}} d\omega\right|^2 dx \\ &= I_{1.1}^{(\mathbf{j})} - I_{1.2}^{(\mathbf{j})}.\end{aligned}$$

We write

$$I_{1.1}^{(\mathbf{j})} = \sum_u a_{u,u}^{(\mathbf{j})} \tilde{\varepsilon}_u^2 + \tilde{\varepsilon}^T \tilde{A}^{(\mathbf{j})} \tilde{\varepsilon} = I_{1.1.1}^{(\mathbf{j})} + I_{1.1.2}^{(\mathbf{j})}$$

with

$$\begin{aligned}a_{u,v}^{(\mathbf{j})} &= \frac{1}{(2\pi)^d n^{2d} h^{2j+d} a_n^{2d}} \int_{\mathbb{R}^d} |\omega|^{2j} \frac{I_{[-1,1]^d}(\omega)}{|\Phi_\psi(\omega/h)|^2} e^{i\omega^T \tilde{x}_u/h} e^{-i\omega^T \tilde{x}_v/h} d\omega \\ \tilde{A}^{(\mathbf{j})} &= (\tilde{a}_{u,v}^{(\mathbf{j})})_{1\le u,v\le (2n+1)^d}, \quad \tilde{a}_{u,v}^{(\mathbf{j})} = a_{u,v}^{(\mathbf{j})} \text{ for } u \ne v, \qquad \tilde{a}_{u,u}^{(\mathbf{j})} = 0 \\ \tilde{x}_1 &= x_{(-n,\ldots,-n)}, \ldots, \tilde{x}_{(2n+1)^d} = x_{(n,\ldots,n)} \\ \tilde{\varepsilon}^T &= (\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_{(2n+1)^d}) = (\varepsilon_{(-n,\ldots,-n)}, \ldots, \varepsilon_{(n,\ldots,n)}) \in \mathbb{R}^{(2n+1)^d}.\end{aligned}$$

For $I_{1.1.1}^{(\mathbf{j})}$ we obtain

$$\begin{aligned}\mathrm{E}[I_{1.1.1}^{(\mathbf{j})}] &= \sigma^2 \sum_u a_{u,u}^{(\mathbf{j})} = \frac{\sigma^2}{(2\pi)^d n^{2d} h^{2j+d} a_n^{2d}} \sum_r \int_{\mathbb{R}^d} |\omega|^{2j} \frac{I_{[-1,1]^d}(\omega)}{|\Phi_\psi(\omega/h)|^2} d\omega \\ &= \frac{\sigma^2 (2n+1)^d}{(2\pi)^d n^{2d} h^{2j+d} a_n^{2d}} \int_{\mathbb{R}^d} |\omega|^{2j} \frac{I_{[-1,1]^d}(\omega)}{|\Phi_\psi(\omega/h)|^2} d\omega \\ &\sim \frac{\sigma^2 (2n+1)^d}{\kappa^2 (2\pi)^d n^{2d} h^{2j+2\beta+d} a_n^{2d}} \int_{\mathbb{R}^d} |\omega|^{2j+2\beta} I_{[-1,1]^d}(\omega) d\omega \\ &= \frac{\sigma^2 (2n+1)^d}{\kappa^2 \pi^d n^{2d} h^{2j+2\beta+d} a_n^{2d}} \prod_{k=1}^d \frac{1}{2(j_k+\beta_k)+1} = O\left(\frac{1}{n^d h^{2j+2\beta+d} a_n^{2d}}\right)\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(I_{1.1.1}) &= \sum_u \left(a_{u,u}^{(\mathbf{j})}\right)^2 \mu_4(\varepsilon) = \frac{\mu_4(\varepsilon)}{(2\pi)^{2d} n^{4d} h^{4j+2d} a_n^{4d}} \sum_r \left(\int_{\mathbb{R}^d} |\omega|^{2j} \frac{I_{[-1,1]^d}(\omega)}{|\Phi_\psi(\omega/h)|^2} d\omega\right)^2 \\ &= \frac{\mu_4(\varepsilon)(2n+1)^d}{(2\pi)^{2d} n^{4d} h^{4j+2d} a_n^{4d}} \left(\int_{\mathbb{R}^d} |\omega|^{2j} \frac{I_{[-1,1]^d}(\omega)}{|\Phi_\psi(\omega/h)|^2}\right)^2 d\omega \\ &\sim \frac{\mu_4(\varepsilon)(2n+1)^d}{\kappa^4 (2\pi)^{2d} n^{4d} h^{4j+4\beta+2d} a_n^{4d}} \left(\int_{\mathbb{R}^d} |\omega|^{2j+2\beta} |I_{[-1,1]^d}(\omega)|^2 d\omega\right)^2 = O\left(\frac{1}{n^{3d} h^{4j+4\beta+2d} a_n^{4d}}\right) \\ &= o\left(\frac{1}{n^{2d} h^{4j+4\beta+d} a_n^{3d}}\right).\end{aligned}$$

We now check the assumptions of Theorem 5.2 in [7] for $I_{1.1.2}$. First of all we calculate the variance

$$\sigma(n)^2 = \mathrm{Var}(\tilde{\varepsilon}^T \tilde{A}^{(\mathbf{j})} \tilde{\varepsilon}) = 2\sigma^4 \mathrm{tr}(\tilde{A}^{(\mathbf{j})})^2 = 2\sigma^4 \sum_{u\ne v} (a_{u,v}^{(\mathbf{j})})^2$$

$$= \frac{2\sigma^4}{(2\pi)^{2d} n^{4d} h^{4j+2d} a_n^{4d}} \sum_{r \neq s} \left(\int_{\mathbb{R}^d} |\omega|^{2j} \frac{I_{[-1,1]^d}(\omega)}{|\Phi_\psi(\omega/h)|^2} e^{i\omega^T x_r/h} e^{-i\omega^T x_s/h} d\omega \right)^2$$

$$\sim \frac{2\sigma^4}{(2\pi)^{2d} n^{2d} h^{4j} a_n^{2d}} \int_{I_n/h} \int_{I_n/h} \left(\int_{\mathbb{R}^d} |\omega|^{2j} \frac{I_{[-1,1]^d}(\omega)}{|\Phi_\psi(\omega/h)|^2} e^{i\omega^T y/h} e^{-i\omega^T z/h} d\omega \right)^2 dydz$$

$$= \frac{2\sigma^4}{(2\pi)^{2d} n^{2d} h^{4j} a_n^{2d}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\omega|^{2j} |\eta|^{2j} \frac{I_{[-1,1]^d}(\omega) I_{[-1,1]^d}(\eta)}{|\Phi_\psi(\omega/h)|^2 |\Phi_\psi(\eta/h)|^2} \left| \int_{I_n/h} e^{i(\omega-\eta)^T u} du \right|^2 d\omega d\eta$$

$$= \frac{2\sigma^4}{(2\pi)^{2d} n^{2d} h^{4j} a_n^{2d}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\omega|^{2j} |\eta|^{2j} \frac{I_{[-1,1]^d}(\omega) I_{[-1,1]^d}(\eta)}{|\Phi_\psi(\omega/h)|^2 |\Phi_\psi(\eta/h)|^2} \prod_{l=1}^d \frac{|e^{i(\omega_l-\eta_l)/(ha_n)} - e^{-i(\omega_l-\eta_l)/(ha_n)}|^2}{|\omega_l - \eta_l|^2} d\omega d\eta$$

$$= \frac{2\sigma^4}{\kappa^4 (2\pi)^{2d} n^{2d} h^{4j+4\beta} a_n^{2d}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} I_{[-1,1]^d}(\omega) I_{[-1,1]^d}(\eta) \prod_{l=1}^d |\omega_l|^{2j_l+2\beta_l} |\eta_l|^{2j_l+2\beta_l} \frac{|\sin\left(\frac{\omega_l-\eta_l}{ha_n}\right)|^2}{|\omega_l-\eta_l|^2} d\omega d\eta$$

$$= \frac{2\sigma^4}{\kappa^4 (2\pi)^{2d} n^{2d} h^{4j+4\beta} a_n^{2d}} \prod_{l=1}^d \int_{\mathbb{R}} \int_{\mathbb{R}} I_{[-1,1]}(\omega_l) I_{[-1,1]}(\eta_l) |\omega_l \eta_l|^{2j_l+2\beta_l} \frac{\sin^2\left(\frac{\omega_l-\eta_l}{ha_n}\right)}{(\omega_l-\eta_l)^2} d\omega_l \, d\eta_l$$

$$= \frac{2\sigma^4 h^{\sum_{l=1}^d (4j_l+4\beta_l+2)}}{\kappa^4 (2\pi)^{2d} n^{2d} h^{4j+\beta+2d} a_n^{2d}} \prod_{l=1}^d \int_{-1/h}^{1/h} \int_{-1/h}^{1/h} |\omega_l \eta_l|^{2j_l+2\beta_l} \frac{\sin^2\left(\frac{\omega_l-\eta_l}{a_n}\right)}{(\omega_l-\eta_l)^2} d\omega_l \, d\eta_l$$

$$= \frac{2 \prod_{l=1}^d C_l \sigma^4}{\kappa^4 (2\pi)^{2d} n^{2d} h^{4j+4\beta+d} a_n^{3d}}$$

using that

$$\lim_{n\to\infty} a_n h^{4\beta_l+4j_l+1} \int_{-1/h}^{1/h} \int_{-1/h}^{1/h} |\omega_l \eta_l|^{2j_l+2\beta_l} \frac{\sin^2\left(\frac{\omega_l-\eta_l}{a_n}\right)}{(\omega_l-\eta_l)^2} d\omega_l \, d\eta_l = C_l,$$

following from the integrability of sinc^2 by some slightly tedious algebra. In the following, we check the assumptions (1)–(3) of Theorem 5.2 in [7] to show the asymptotic normality of $I_{1.1.2}^{(\mathbf{j})}$.

(1) We have uniformly over all $\mathbf{s} \in \{-n, \dots, n\}^d$

$$\begin{aligned} \sum_{\mathbf{r} \in \{-n,\dots,n\}^d} |a_{\mathbf{r},\mathbf{s}}^{(\mathbf{j})}|^2 &= \frac{1}{(2\pi)^{4d} n^{4d} h^{4j+2d} a_n^{4d}} \sum_{\mathbf{r} \in \{-n,\dots,n\}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\omega\eta|^{2j} \frac{I_{[-1,1]^d}(\omega) I_{[-1,1]^d}(\eta)}{|\Phi_\psi(\omega/h)|^2 |\Phi_\psi(\eta/h)|^2} \\ &\quad \times e^{i(\omega-\eta)^T x_{\mathbf{r}}/h} e^{-i(\omega-\eta)^T x_{\mathbf{s}}/h} d\omega d\eta \\ &\sim \frac{1}{(2\pi)^{4d} n^{3d} h^{4j+d} a_n^{3d}} \int_{A_n} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\omega\eta|^{2j} \frac{I_{[-1,1]^d}(\omega) I_{[-1,1]^d}(\eta)}{|\Phi_\psi(\omega/h)|^2 |\Phi_\psi(\eta/h)|^2} \\ &\quad \times e^{i(\omega-\eta)^T u} e^{-i(\omega-\eta)^T x_{\mathbf{s}}/h} d\omega d\eta du \\ &= \frac{1}{(2\pi)^{4d} n^{3d} h^{4j+d} a_n^{3d}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\omega\eta|^{2j} \frac{I_{[-1,1]^d}(\omega) I_{[-1,1]^d}(\eta)}{|\Phi_\psi(\omega/h)|^2 |\Phi_\psi(\eta/h)|^2} \\ &\quad \times \left(\prod_{\nu=1}^d \frac{\sin\left(\frac{\omega_\nu-\eta_\nu}{ha_n}\right)}{\omega_\nu - \eta_\nu} \right) e^{-i(\omega-\eta)^T x_{\mathbf{s}}/h} d\omega d\eta \\ &= \frac{1}{(2\pi)^{4d} n^{3d} h^{4j+4\beta+d} a_n^{3d}} \prod_{\nu=1}^d \int_{\mathbb{R}} \int_{\mathbb{R}} |\omega_\nu \eta_\nu|^{2j} I_{[-1,1]^d}(\omega_\nu) I_{[-1,1]^d}(\eta_\nu) \\ &\quad \times \left(\frac{\sin\left(\frac{\omega_\nu-\eta_\nu}{ha_n}\right)}{\omega_\nu - \eta_\nu} \right) e^{-i(\omega_\nu-\eta_\nu)^T x_{\mathbf{s},\nu}/h} d\omega_\nu d\eta_\nu. \end{aligned}$$

Since $|\sin((\omega_\nu - \eta_\nu)/(ha_n))/(\omega_\nu - \eta_\nu)| \leq (ha_n)^{-1}$ we obtain

$$\sum_{\mathbf{r}\in\{-n,\dots,n\}^d} |a^{(\mathbf{j})}_{\mathbf{r},\mathbf{s}}|^2 = O\left(\frac{1}{n^{3d}h^{4j+4\beta+2d}a_n^{4d}}\right)$$

and therefore with $\kappa(n) = (\log n)^{1/4}$

$$\frac{\kappa(n)}{\sigma(n)^2} \sum_{\mathbf{r}\in\{-n,\dots,n\}^d} |a_{\mathbf{r},\mathbf{s}}|^2 = O\left(\frac{(\log n)^{1/4}}{n^d h^d a_n^d}\right) = o(1)$$

(2) Since $\kappa(n) \to \infty$ and $\varepsilon_\mathbf{r}$ are independent identically distributed with $\mathrm{E}[\varepsilon_\mathbf{r}^2] = \sigma^2 < \infty$, it immediately follows that

$$\mathrm{E}[\varepsilon_\mathbf{r}^2 I\{|\varepsilon_\mathbf{r}| > \kappa(n)\}] = o(1).$$

(3) For estimating the eigenvalues $\mu_\mathbf{r}$ of $\tilde{A}^{(\mathbf{j})}$ we use Gerschgorin's Theorem and obtain uniformly over all $\mathbf{s} \in \{-n,\dots,n\}^d$

$$\begin{aligned}
\mu_\mathbf{s} &\leq \sum_{\mathbf{r}\in\{-n,\dots,n\}^d} |a^{(\mathbf{j})}_{\mathbf{r},\mathbf{s}}| \\
&\sim \frac{1}{(2\pi)^{2d} n^d h^{2j} a_n^d} \int_{A_n} \left| \int_{\mathbb{R}^d} |\omega|^{2j} \frac{I_{[-1,1]^d}(\omega)}{|\Phi_\psi(\omega/h)|^2} e^{i\omega^T u} e^{-i\omega^T x_\mathbf{s}/h} d\omega \right| du \\
&= \frac{1}{(2\pi)^{2d} n^d h^{2j+2\beta+d} a_n^{2d}} \prod_{\nu=1}^{d} \int_{-1/(ha_n)}^{1/(ha_n)} \int_{\mathbb{R}} |\omega_\nu|^{2j_\nu+2\beta_\nu} I_{[-1,1]^d}(\omega_\nu) e^{i\omega_\nu u_\nu} e^{-i\omega_\nu x_{\mathbf{s},\nu}/h} d\omega_\nu du_\nu.
\end{aligned}$$

It now follows by similar but tedious calculations as above, that this term is of order $O(\log n/n^d a_n^d h^{2j+2\beta})$ and

$$\frac{1}{\sigma(n)^2} \max_{\mathbf{s}\in\{-n,\dots,n\}^d} \mu_\mathbf{s}^2 = O(ha_n \log n) = o(1).$$

It now remains to discuss the remainder terms For $I_{1.2}$ we get

$$I_{1.2} = o_P(I_{1.1})$$

since it consists of the tails of the integral in $I_{1.1}$, before Parseval's equality was used, and the upper respective lower bound of the integral tails asymptotically diverge to $\pm\infty$. This means, that $I_{1.2}$ is asymptotically negligible.

Since the bias of $\hat{m}^{(\mathbf{j})}$ is uniformly of order $o(h^{s-j-1})$ on B (see e.g. [3]) we have with condition 5

$$I_3 = o(h^{2s-2j-2}) = o\left(\frac{1}{n^d h^{2\beta+2j+d/2} a_n^{3d/2}}\right)$$

and by applying the Cauchy–Schwarz inequality also

$$I_2 = O\left(\frac{1}{n^{d/2} h^{\beta+j+d/4} a_n^{3d/4}}\right) o(h^{s-j-1}) = o\left(\frac{1}{n^d h^{2\beta+2j+d/2} a_n^{3d/2}}\right). \quad \square$$

A.1. Proof of Theorem 1

Since $L(\theta)$ is locally convex near ϑ, for every $\varepsilon > 0$ exists a constant $K_\varepsilon > 0$ with

$$\mathrm{P}(|\hat{\vartheta}_n - \vartheta_n| > \varepsilon) \leq \mathrm{P}(L(\hat{\vartheta}_n) - L(\vartheta) > K_\varepsilon) \leq \mathrm{P}(|\hat{L}(\hat{\vartheta}_n) - L(\hat{\vartheta}_n)| > K_\varepsilon/2) + \mathrm{P}(|\hat{L}(\vartheta) - L(\vartheta)| > K_\varepsilon/2)$$

since $\hat{\vartheta}_n$ minimizes $\hat{L}(\theta)$ and the assertion follows if we show that $\hat{L}(\theta) - L(\theta)$ stochastically converges to 0 uniformly in θ. To this end note that

$$\begin{aligned}
|\hat{L}(\theta) - L(\theta)| &= \left| \int_A (\hat{m}(T_\theta x) - \hat{m}(S_\theta x))^2 dx - \int_A (m(T_\theta x) - m(S_\theta x))^2 dx \right| \\
&\leq C\left(\int_A (\hat{m}(T_\theta x) - m(T_\theta x))^2 dx + \int_A (\hat{m}(S_\theta x) - m(S_\theta x))^2 dx \right) \\
&\leq 2C \int_{A_\theta} (\hat{m}(z) - m(z))^2 dz \leq 2C \int_D (\hat{m}(z) - m(z))^2 dz.
\end{aligned}$$

Therefore we have for any $\tilde{\delta} > 0$ and $\delta = \tilde{\delta}/(2C)$

$$\mathrm{P}(\sup_\theta |\hat{L}(\theta) - L(\theta)| > \tilde{\delta}) \leq \mathrm{P}\left(\int_D (\hat{m}(z) - m(z))^2 dz > \delta \right).$$

But the right probability converges to 0 because of Theorem 4. $\square$

A.2. Proof of Theorem 2

Note, that $\hat{l}_n(\hat{\vartheta}) = 0$. With this and a first order Taylor expansion of $\hat{l}_n$ in ϑ we write

$$-\hat{h}(\xi_n)(\hat{\vartheta} - \vartheta) = \hat{l}_n(\vartheta) \tag{12}$$

for some ξ_n with $\|\xi_n - \vartheta\| \le \|\hat{\vartheta} - \vartheta\|$. Theorem 2 now follows after we have shown the following two Lemmata

Lemma 1. *Under the assumptions of Theorem 2 we have*

$$\sqrt{n^d a_n^d}\hat{l}_n(\vartheta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma(\vartheta))$$

with $\Sigma(\theta)$ and $\sigma_\theta(u)$ defined as in Theorem 2.

Lemma 2. *Under the assumptions of Theorem 2 we have*

$$\hat{h}(\xi_n) \xrightarrow{P} h(\vartheta).$$

Proof of Lemma 1. We write

$$\Delta_{m,\theta}(\mathbf{x}) = \left(\operatorname{grad} m(T_\theta \mathbf{x}) \frac{\partial}{\partial\theta} T_\theta \mathbf{x} - \operatorname{grad} m(S_\theta \mathbf{x}) \frac{\partial}{\partial\theta} S_\theta \mathbf{x}\right)^T$$

and

$$\begin{aligned}
\hat{l}_n(\vartheta) &= 2\int_A \left[\hat{m}(T_\vartheta \mathbf{x}) - \hat{m}(S_\vartheta \mathbf{x})\right] \Delta_{m,\vartheta}(\mathbf{x}) d\mathbf{x} + R_{n,1} \\
&= \sum_{\mathbf{r}\in\{-n,\dots,n\}^d} \left(2\int_A (w_\mathbf{r}(T_\vartheta \mathbf{x}) - w_\mathbf{r}(S_\vartheta \mathbf{x}))\, \Delta_{m,\vartheta}(\mathbf{x}) d\mathbf{x}\right) Z_\mathbf{r} + R_{n,1} \\
&= \sum_{\mathbf{r}\in\{-\mathbf{n},\dots,\mathbf{n}\}^\mathbf{d}} v_\mathbf{r}(\vartheta)\varepsilon_\mathbf{r} + R_{n,1} + R_{n,2} = \tilde{l}_n(\vartheta) + 2R_{n,1} + 2R_{n,2}
\end{aligned}$$

with

$$\begin{aligned}
v_\mathbf{r}(\vartheta) &= 2\int_A (w_\mathbf{r}(T_\vartheta \mathbf{x}) - w_\mathbf{r}(S_\vartheta \mathbf{x})) \left(\operatorname{grad} m(T_\vartheta \mathbf{x}) \left.\frac{\partial}{\partial\theta} T_\theta\right|_{\theta=\vartheta} \mathbf{x} - \operatorname{grad} m(S_\vartheta \mathbf{x}) \left.\frac{\partial}{\partial\theta} S_\theta\right|_{\theta=\vartheta} \mathbf{x}\right)^T d\mathbf{x} \in \mathbb{R}^d \\
R_{n,1} &= \int_A \left[\hat{m}(T_\vartheta \mathbf{x}) - \hat{m}(S_\vartheta \mathbf{x})\right] \left(\operatorname{grad}(\hat{m} - m)(T_\vartheta \mathbf{x}) \left.\frac{\partial}{\partial\theta} T_\theta\right|_{\theta=\vartheta} \mathbf{x} - \operatorname{grad}(\hat{m} - m)(S_\vartheta \mathbf{x}) \left.\frac{\partial}{\partial\theta} S_\theta\right|_{\theta=\vartheta} \mathbf{x}\right) d\mathbf{x} \\
R_{n,2} &= \int_A \left(\mathrm{E}[\hat{m}(T_\vartheta \mathbf{x})] - \mathrm{E}[\hat{m}(S_\vartheta \mathbf{x})]\right) \Delta_{m,\vartheta}(\mathbf{x}) d\mathbf{x}.
\end{aligned}$$

This means, that $\hat{l}_n(\vartheta)$ consists of a sum of weighted independently distributed random variables for which we determine the asymptotic distribution by using Lemma 3.1 in [10] and remainders $R_{n,1}$ and $R_{n,2}$ for which we show that they are asymptotically negligible. We will first consider the asymptotic distribution of $\tilde{l}_n$. To this end we have to check the condition

$$\frac{\max\limits_{\mathbf{r}\in\{-\mathbf{n},\dots,\mathbf{n}\}^\mathbf{d}} |c^T v_\mathbf{r}(\vartheta)|}{\left(\sum\limits_{\mathbf{r}\in\{-\mathbf{n},\dots,\mathbf{n}\}^\mathbf{d}} c^T v_\mathbf{r}(\vartheta) v_\mathbf{r}^T c(\vartheta)\right)^{1/2}} = o(1) \tag{13}$$

for every $c \in \mathbb{R}^d$. Note that from (4) we have

$$\begin{aligned}
\operatorname{grad} m(S_\vartheta \mathbf{x}) &= \operatorname{grad} m(T_\vartheta \mathbf{x}) M_\vartheta N_\vartheta^{-1} \\
\operatorname{grad} m(T_\vartheta \mathbf{x}) &= \operatorname{grad} m(S_\vartheta \mathbf{x}) N_\vartheta M_\vartheta^{-1}.
\end{aligned}$$

Therefore we get

$$\begin{aligned}
|c^T v_\mathbf{r}(\vartheta)| &= \left| 2\int_A \frac{1}{(nha_n)^d} \int_{\mathbb{R}^d} (e^{-i\omega(T_\vartheta \mathbf{x} - \mathbf{x}_\mathbf{r})/h} - e^{-i\omega(S_\vartheta \mathbf{x} - \mathbf{x}_\mathbf{r})/h}) \frac{I_{[-1,1]^d}(\omega)}{\Phi_\psi(\omega/h)} d\omega\, c^T \Delta_{m,\vartheta}(\mathbf{x}) d\mathbf{x} \right| \\
&\le \frac{2}{(na_n)^d} \iint_{\mathbb{R}^2} |e^{-i\omega T_\vartheta \mathbf{u}} - e^{-i\omega S_\vartheta \mathbf{u}}|\, |e^{i\omega \mathbf{x}_\mathbf{r}/h}| \frac{I_{[-1,1]^d}(\omega)}{|\Phi_\psi(\omega/h)|} d\omega \left|c^T \Delta_{m,\vartheta}(h\mathbf{u})\right| d\mathbf{u}
\end{aligned}$$

$$\le \frac{4}{(na_n)^d}\int_{\mathbb{R}^2}\frac{I_{[-1,1]^d}(\omega)}{|\Phi_\psi(\omega/h)|}d\omega\int_A\left|c^T\Delta_{m,\vartheta}(h\mathbf{u})\right|d\mathbf{u}$$
$$= O\left(\frac{1}{n^d h^\beta a_n^d}\right)$$

and

$$\begin{aligned}\sum_{\mathbf{r}\in\{-\mathbf{n},\dots,\mathbf{n}\}^\mathbf{d}}(c^T v_\mathbf{r}(\vartheta))^2 &= \frac{4}{(2\pi nha_n)^{2d}}\sum_\mathbf{r}\left(\int_A\int_{\mathbb{R}^d}(e^{-i\omega^T(T_\vartheta\mathbf{x}-\mathbf{x_r})/h}-e^{-i\omega^T(S_\vartheta\mathbf{x}-\mathbf{x_r})/h})\frac{I_{[-1,1]^d}(\omega)}{\Phi_\psi(\omega/h)}d\omega c^T\Delta_{m,\vartheta}(\mathbf{x})d\mathbf{x}\right)^2\\
&= \frac{4}{(na_n)^d h^{2d}}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\frac{I_{[-1,1]^d}(\omega)}{\Phi_\psi(\omega/h)}c^T\left(\int_A e^{-i\omega^T(T_\vartheta\mathbf{x}-\mathbf{u})/h}\Delta_{m,\vartheta}(\mathbf{x})(\operatorname{grad} m(T_\vartheta\mathbf{x}))^T d\mathbf{x}\right.\right.\\
&\quad\left.\left.-\int_A e^{-i\omega^T(S_\vartheta\mathbf{x}-\mathbf{u})/h}\left(N_\vartheta M_\vartheta^{-1}\frac{\partial}{\partial\theta}T_\theta\bigg|_{\theta=\vartheta}\mathbf{x}-\frac{\partial}{\partial\theta}S_\theta\bigg|_{\theta=\vartheta}\mathbf{x}\right)^T(\operatorname{grad} m(S_\vartheta\mathbf{x}))^T d\mathbf{x}\right)d\omega\right)^2\\
&\quad\times d\mathbf{u}(1+o(1))\\
&= \frac{4}{(na_n)^d}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\frac{I_{[-1,1]^d}(h\omega)}{\Phi_\psi(\omega)}c^T\left(\int_A e^{-i\omega^T(T_\vartheta\mathbf{x}-\mathbf{u})}\Delta_{m,\vartheta}(\mathbf{x})(\operatorname{grad} m(T_\vartheta\mathbf{x}))^T d\mathbf{x}\right.\right.\\
&\quad\left.\left.-\int_A e^{-i\omega^T(S_\vartheta\mathbf{x}-\mathbf{u})}\left(N_\vartheta M_\vartheta^{-1}\frac{\partial}{\partial\theta}T_\theta\bigg|_{\theta=\vartheta}\mathbf{x}-\frac{\partial}{\partial\theta}S_\theta\bigg|_{\theta=\vartheta}\mathbf{x}\right)^T(\operatorname{grad} m(S_\vartheta\mathbf{x}))^T d\mathbf{x}\right)d\omega\right)^2\\
&\quad\times d\mathbf{u}(1+o(1)).\end{aligned}$$

With Assumption 3 the integral on the r.h.s. of the equation exists, and we have

$$\sum_{\mathbf{r}\in\{-\mathbf{n},\dots,\mathbf{n}\}^\mathbf{d}}(c^T v_\mathbf{r}(\vartheta))^2 = \frac{4C_a}{(na_n)^d}$$

with

$$\begin{aligned}C_a &= \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\frac{I_{[-1,1]^d}(h\omega)}{\Phi_\psi(\omega)}c^T\left(\int_A e^{-i\omega^T(T_\vartheta\mathbf{x}-\mathbf{u})}\Delta_{m,\vartheta}(\mathbf{x})(\operatorname{grad} m(T_\vartheta\mathbf{x}))^T d\mathbf{x}\right.\right.\\
&\quad\left.\left.-\int_A e^{-i\omega^T(S_\vartheta\mathbf{x}-\mathbf{u})}\left(N_\vartheta M_\vartheta^{-1}\frac{\partial}{\partial\theta}T_\theta\bigg|_{\theta=\vartheta}\mathbf{x}-\frac{\partial}{\partial\theta}S_\theta\bigg|_{\theta=\vartheta}\mathbf{x}\right)^T(\operatorname{grad} m(S_\vartheta\mathbf{x}))^T d\mathbf{x}\right)d\omega\right)^2 d\mathbf{u}.\end{aligned}$$

This yields by

$$\frac{\max_{\mathbf{r}\in\{-\mathbf{n},\dots,\mathbf{n}\}^\mathbf{d}}|c^T v_\mathbf{r}(\vartheta)|}{\left(\sum_\mathbf{r}^n c^T v_\mathbf{r}(\vartheta)v_\mathbf{r}^T c(\vartheta)\right)^{1/2}} = O\left(\frac{1}{(na_n)^{d/2}h^\beta}\right) = o(1)$$

and the Cramér–Wold device the asymptotic normality of $\tilde{l}_n(\vartheta)$. We will now discuss the remainder terms. Using the Cauchy–Schwarz inequality we get

$$\begin{aligned}R_{n,1} &\le \left(\int_A\left[\hat{m}(T_\vartheta\mathbf{x})-\hat{m}(S_\vartheta\mathbf{x})\right]^2 d\mathbf{x}\right)^{1/2}\\
&\quad\times\left(\int_A\left(\left(\frac{\partial}{\partial\theta}T_\theta\bigg|_{\theta=\vartheta}\mathbf{x}\right)^T\left(\operatorname{grad}(\hat{m}-m)(T_\vartheta\mathbf{x})\right)^T-\left(\frac{\partial}{\partial\theta}S_\theta\bigg|_{\theta=\vartheta}\mathbf{x}\right)^T\left(\operatorname{grad}(\hat{m}-m)(S_\vartheta\mathbf{x})\right)^T\right)^2 d\mathbf{x}\right)^{1/2}.\end{aligned}$$

We apply Theorem 4 and obtain $R_{n,1} = O_P(1/n^d a_n^{3d/2}h^{2\beta+1+d/2}) = o_p(1/n^{d/2}a_n^{3d/4}h^{\beta+d/4})$ since $n^{d/2}a_n^{3d/4}h^{\beta+1/2+d/4}\to\infty$ by Assumption 5. Now it remains to estimate

$$R_{n,2} = \frac{1}{(2\pi na_n h)^d}\sum_\mathbf{r}\int_A\int_{\mathbb{R}^d}(e^{-i\omega^T(T_\vartheta\mathbf{x}-\mathbf{x_r})/h}-e^{-i\omega^T(S_\vartheta\mathbf{x}-\mathbf{x_r})/h})\frac{I_{[-1,1]^d}(\omega)}{\Phi_\psi(\omega/h)}d\omega\Delta_{m,\vartheta}(\mathbf{x})d\mathbf{x}g(\mathbf{x_r})$$

$$
\begin{aligned}
&= \frac{1}{(2\pi h)^d}\int_{[-1/a_n,1/a_n]^d}\int_A\int_{\mathbb{R}^d}(e^{-i\omega^T T_\vartheta \mathbf{x}/h}-e^{-i\omega^T S_\vartheta \mathbf{x}/h})e^{i\omega^T\mathbf{u}/h}\frac{I_{[-1,1]^d}(\omega)}{\Phi_\psi(\omega/h)}d\omega\Delta_{m,\vartheta}(\mathbf{x})d\mathbf{x}g(\mathbf{u})d\mathbf{u}\\
&\quad + O\left(\frac{1}{n^d a_n^d}\right)\int_A\int_{\mathbb{R}^d}(e^{-i\omega^T T_\vartheta \mathbf{x}/h}-e^{-i\omega^T S_\vartheta \mathbf{x}/h})e^{i\omega^T\mathbf{u}/h}\frac{I_{[-1,1]^d}(\omega)}{\Phi_\psi(\omega/h)}d\omega\Delta_{m,\vartheta}(\mathbf{x})d\mathbf{x}\\
&= \frac{1}{(2\pi h)^2}\int_A\int_{\mathbb{R}^d}(e^{-i\omega^T T_\vartheta \mathbf{x}/h}-e^{-i\omega^T S_\vartheta \mathbf{x}/h})\Phi_m(\omega/h)I_{[-1,1]^d}(\omega)d\omega\Delta_{m,\vartheta}(\mathbf{x})d\mathbf{x}\\
&\quad - \frac{1}{(2\pi h)^2}\int_A\int_{\mathbb{R}^d}(e^{-i\omega^T T_\vartheta \mathbf{x}/h}-e^{-i\omega^T S_\vartheta \mathbf{x}/h})\left(\int_{([-1/a_n,1/a_n]^d)^c}e^{i\omega^T\mathbf{u}/h}g(\mathbf{u})d\mathbf{u}\right)\frac{I_{[-1,1]^d}(\omega)}{\Phi_\psi(\omega/h)}d\omega\Delta_{m,\vartheta}(\mathbf{x})d\mathbf{x}\\
&\quad + O\left(\frac{1}{n^d a_n^d}\right)\int_A\int_{\mathbb{R}^d}(e^{-i\omega^T T_\vartheta \mathbf{x}/h}-e^{-i\omega^T S_\vartheta \mathbf{x}/h})e^{i\omega^T\mathbf{u}/h}\frac{I_{[-1,1]^d}(\omega)}{\Phi_\psi(\omega/h)}d\omega\Delta_{m,\vartheta}(\mathbf{x})d\mathbf{x}\\
&= R_{n,2}^{[1]} + R_{n,2}^{[2]} + R_{n,2}^{[3]} O\left(\frac{1}{n^d a_n^d h^d}\right).
\end{aligned}
$$

There is

$$R_{n,2}^{[1]} = R_{n,2}^{[1.1]} - R_{n,2}^{[1.2]}$$

with

$$
\begin{aligned}
R_{n,2}^{[1.1]} &= \frac{1}{(2\pi h)^d}\int_{A_\vartheta}\int_{\mathbb{R}^d}e^{-i\omega^T\mathbf{y}/h}\Phi_m\left(\frac{\omega}{h}\right)I_{[-1,1]^d}(\omega)d\omega\Delta_{m,\vartheta}(T_\vartheta^{-1}\mathbf{y})d\mathbf{y}\\
R_{n,2}^{[1.2]} &= \frac{1}{(2\pi h)^d}\int_{A_\vartheta}\int_{\mathbb{R}^d}e^{-i\omega^T(S_\vartheta T_\vartheta^{-1}\mathbf{y})/h}\Phi_m\left(\frac{\omega}{h}\right)I_{[-1,1]^d}(\omega)d\omega\Delta_{m,\vartheta}(T_\vartheta^{-1}\mathbf{y})d\mathbf{y}.
\end{aligned}
$$

Since $m(\mathbf{z}) = m(T_\vartheta S_\vartheta^{-1}\mathbf{z})$ it is easy to show that $\Phi_m = \Phi_{m(T_\vartheta S_\vartheta^{-1}\cdot)}$ and

$$
\begin{aligned}
\Phi_{m(T_\vartheta S_\vartheta^{-1}\cdot)}(\omega/h) &= \int_{\mathbb{R}^2}e^{i\omega^T\mathbf{v}/h}m(T_\vartheta S_\vartheta^{-1}\mathbf{v})d\mathbf{v}\\
&= \int_{\mathbb{R}^d}e^{i\omega^T(S_\vartheta T_\vartheta^{-1}\mathbf{u})/h}m(\mathbf{u})d\mathbf{u} = e^{-i\omega^T b_\vartheta(I-N_\vartheta M_\vartheta^{-1})/h}\int_{\mathbb{R}^d}e^{i\omega^T N_\vartheta M_\vartheta^{-1}\mathbf{u}/h}m(\mathbf{u})d\mathbf{u}\\
&= e^{-i\omega^T b_\vartheta(I-N_\vartheta M_\vartheta^{-1})/h}\Phi_m\left((N_\vartheta M_\vartheta^{-1})^T\omega/h\right).
\end{aligned}
$$

Furthermore

$$e^{-i\omega^T(S_\vartheta T_\vartheta^{-1}\mathbf{y})/h} = e^{i\omega^T b_\vartheta(I-N_\vartheta M_\vartheta^{-1})/h}e^{-i\omega^T N_\vartheta M_\vartheta^{-1}\mathbf{y}/h}.$$

Substituting this in $R_{n,2}^{[1.2]}$ we obtain

$$
\begin{aligned}
R_{n,2}^{[1.2]} &= \frac{1}{(2\pi h)^d}\int_{A_\vartheta}\int_{\mathbb{R}^d}e^{-i\omega^T(S_\vartheta T_\vartheta^{-1}\mathbf{y})/h}\Phi_{m(T_\vartheta S_\vartheta^{-1}\cdot)}\left(\frac{\omega}{h}\right)I_{[-1,1]^d}(\omega)d\omega\Delta_{m,\vartheta}(T_\vartheta^{-1}\mathbf{y})d\mathbf{y}\\
&= \frac{1}{(2\pi h)^d}\int_{A_\vartheta}\int_{\mathbb{R}^d}e^{-i((N_\vartheta M_\vartheta^{-1})^T\omega)^T\mathbf{y}/h}\Phi_m\left(\frac{(N_\vartheta M_\vartheta^{-1})^T\omega}{h}\right)I_{[-1,1]^d}(\omega)d\omega\Delta_{m,\vartheta}(T_\vartheta^{-1}\mathbf{y})d\mathbf{y}\\
&= R_{n,2}^{[1.1]}
\end{aligned}
$$

with $(N_\vartheta M_\vartheta^{-1})^T\omega = \eta$. Therefore $R_{n,2}^{[1]} = 0$.

$$
\begin{aligned}
\|R_{n,2}^{[2]}\| &\le \frac{1}{2\pi^d h^{d+\beta}}\int_{([-1/a_n,1/a_n]^d)^c}\frac{1}{\|\mathbf{u}\|^r}\|\mathbf{u}\|^r|g(\mathbf{u})|d\mathbf{u}\int_{\mathbb{R}^d}\|\omega\|^\beta\frac{1}{\|\omega/h\|^\beta}\frac{|I_{[-1,1]^d}(\omega)|}{|\Phi_\psi(\omega/h)|}d\omega\\
&\quad\times\int_A\left\|\text{grad } m(T_\vartheta\mathbf{x})\frac{\partial}{\partial\theta}T_\theta\bigg|_{\theta=\vartheta}\mathbf{x} - \text{grad } m(S_\vartheta\mathbf{x})\frac{\partial}{\partial\theta}S_\theta\bigg|_{\theta=\vartheta}\mathbf{x}\right\|d\mathbf{x}\\
&\le O\left(\frac{a_n^r}{h^{d+\beta}}\right)\int_{\mathbb{R}^d}\|\mathbf{u}\|^r|g(\mathbf{u})|d\mathbf{u}\int_{\mathbb{R}^d}\|\omega\|^\beta I_{[-1,1]^d}(\omega)
\end{aligned}
$$

$$\times \int_A \left\| \text{grad } m(T_\vartheta \mathbf{x}) \frac{\partial}{\partial\theta} T_\theta \bigg|_{\theta=\vartheta} \mathbf{x} - \text{grad } m(S_\vartheta \mathbf{x}) \frac{\partial}{\partial\theta} S_\theta \bigg|_{\theta=\vartheta} \mathbf{x} \right\| d\mathbf{x}$$

$$= O\left(\frac{a_n^r}{h^{d+\beta}}\right)$$

and

$$|R_{n,2}^{[3]}| \le \frac{2}{h^\beta} C \int_A \left\| \text{grad } m(T_\vartheta \mathbf{x}) \frac{\partial}{\partial\theta} T_\theta \bigg|_{\theta=\vartheta} \mathbf{x} - \text{grad } m(S_\vartheta \mathbf{x}) \frac{\partial}{\partial\theta} S_\theta \bigg|_{\theta=\vartheta} \mathbf{x} \right\| d\mathbf{x}$$
$$= O\left(\frac{1}{h^\beta}\right).$$

Altogether this yields with the assumptions $n^{d/2} a_n^{r+d/2}/h^d \to 0$ and $n^{d/2} a_n^{d/2} h^{d+\beta} \to \infty$

$$|R_{n,2}| = 0 + O\left(\frac{a_n^r}{h^{d+\beta}}\right) + O\left(\frac{1}{h^\beta}\right) O\left(\frac{1}{n^d a_n^d h^d}\right) = o\left(\frac{1}{n^{d/2} a_n^{d/2}}\right). \quad \square$$

Proof of Lemma 2. First of all note that $\|\xi_n - \vartheta\| \le \|\hat{\vartheta}_n - \vartheta\|$ and therefore $\xi_n \xrightarrow{P} \vartheta$ for $n \to \infty$.

$$\hat{h}(\xi_n) - h(\vartheta) = (\hat{h}(\xi_n) - h(\xi_n)) + (h(\xi_n) - h(\vartheta)).$$

With the above remark and the continuity of h it is immediately clear that the second part stochastically converges to 0. For the first part it suffices to show that $\sup_\theta \|\hat{h}(\theta) - h(\theta)\|_M$ stochastically converges to 0 where $\|\cdot\|_M$ denotes the maximum norm of a matrix. We have

$$\begin{aligned}
\frac{1}{2}(\hat{h}(\theta) - h(\theta)) &= \frac{1}{2}\left(\frac{\partial}{\partial\theta}\hat{l}_n(\theta) - \frac{\partial}{\partial\theta} l(\theta)\right) \\
&= \int_A (\Delta_{\hat{m},\theta}(\mathbf{x}) - \Delta_{m,\theta}(\mathbf{x}))^T (\Delta_{\hat{m},\theta}(\mathbf{x}) - \Delta_{m,\theta}(\mathbf{x})) d\mathbf{x} + \int_A \Delta_{m,\theta}(\mathbf{x})^T (\Delta_{\hat{m},\theta}(\mathbf{x}) - \Delta_{m,\theta}(\mathbf{x})) d\mathbf{x} \\
&\quad + \int_A (\Delta_{\hat{m},\theta}(\mathbf{x}) - \Delta_{m,\theta}(\mathbf{x}))^T \Delta_{m,\theta}(\mathbf{x}) d\mathbf{x} + \int_A (\hat{m}(T_\theta\mathbf{x}) - m(T_\theta\mathbf{x}) - (\hat{m}(S_\theta\mathbf{x}) - m(S_\theta\mathbf{x}))) \\
&\quad \times \left(\frac{\partial}{\partial\theta}\Delta_{\hat{m},\theta}(\mathbf{x}) - \frac{\partial}{\partial\theta}\Delta_{m,\theta}(\mathbf{x})\right) d\mathbf{x} + \int_A (m(T_\theta\mathbf{x}) - m(S_\theta\mathbf{x})) \left(\frac{\partial}{\partial\theta}\Delta_{\hat{m},\theta}(\mathbf{x}) - \frac{\partial}{\partial\theta}\Delta_{m,\theta}(\mathbf{x})\right) d\mathbf{x} \\
&\quad + \int_A (\hat{m}(T_\theta\mathbf{x}) - m(T_\theta\mathbf{x}) - (\hat{m}(S_\theta\mathbf{x}) - m(S_\theta\mathbf{x}))) \frac{\partial}{\partial\theta}\Delta_{m,\theta}(\mathbf{x}) d\mathbf{x}.
\end{aligned}$$

There is

$$\Delta_{m,\theta}(\mathbf{x})^T (\Delta_{\hat{m},\theta}(\mathbf{x}) - \Delta_{m,\theta}(\mathbf{x})) = (a_{i,j}(\mathbf{x}))_{1\le i,j\le k}$$
$$\frac{\partial}{\partial\theta}\Delta_{\hat{m},\theta}(\mathbf{x}) - \frac{\partial}{\partial\theta}\Delta_{m,\theta}(\mathbf{x}) = (h_{i,j}(\mathbf{x}))_{1\le i,j\le k}$$

with

$$\begin{aligned}
a_{i,j}(\mathbf{x}) &= \sum_{s=1}^d \sum_{t=1}^d \left(\frac{\partial}{\partial x_s} m(T_\theta\mathbf{x}) \frac{\partial}{\partial\theta_i}(T_\theta\mathbf{x})_s - \frac{\partial}{\partial x_s} m(S_\theta\mathbf{x}) \frac{\partial}{\partial\theta_i}(S_\theta\mathbf{x})_s\right) \frac{\partial}{\partial\theta_j}(T_\theta\mathbf{x})_t \left(\frac{\partial}{\partial x_s}\hat{m}(T_\theta\mathbf{x}) - \frac{\partial}{\partial x_s} m(T_\theta\mathbf{x})\right) \\
&\quad - \sum_{s=1}^d \sum_{t=1}^d \left(\frac{\partial}{\partial x_s} m(T_\theta\mathbf{x}) \frac{\partial}{\partial\theta_i}(T_\theta\mathbf{x})_s - \frac{\partial}{\partial x_s} m(S_\theta\mathbf{x}) \frac{\partial}{\partial\theta_i}(S_\theta\mathbf{x})_s\right) \frac{\partial}{\partial\theta_j}(S_\theta\mathbf{x})_t \\
&\quad \times \left(\frac{\partial}{\partial x_s}\hat{m}(S_\theta\mathbf{x}) - \frac{\partial}{\partial x_s} m(S_\theta\mathbf{x})\right)
\end{aligned}$$

$$\begin{aligned}h_{i,j}(\mathbf{x}) &= \sum_{s=1}^{d}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}(T_\theta\mathbf{x})_s\left(\frac{\partial}{\partial x_s}\hat{m}(T_\theta\mathbf{x})-\frac{\partial}{\partial x_s}m(T_\theta\mathbf{x})\right)-\frac{\partial^2}{\partial\theta_i\partial\theta_j}(S_\theta\mathbf{x})_s\left(\frac{\partial}{\partial x_s}\hat{m}(S_\theta\mathbf{x})-\frac{\partial}{\partial x_s}m(S_\theta\mathbf{x})\right)\right]\\&\quad+\sum_{s=1}^{d}\sum_{t=1}^{d}\left[\frac{\partial}{\partial\theta_i}(T_\theta\mathbf{x})_s\frac{\partial}{\partial\theta_j}(T_\theta\mathbf{x})_t\left(\frac{\partial^2}{\partial x_s\partial x_t}\hat{m}(T_\theta\mathbf{x})-\frac{\partial^2}{\partial x_s\partial x_t}m(T_\theta\mathbf{x})\right)\right.\\&\quad\left.-\frac{\partial}{\partial\theta_i}(S_\theta\mathbf{x})_s\frac{\partial}{\partial\theta_j}(S_\theta\mathbf{x})_t\left(\frac{\partial^2}{\partial x_s\partial x_t}\hat{m}(S_\theta\mathbf{x})-\frac{\partial^2}{\partial x_s\partial x_t}m(S_\theta\mathbf{x})\right)\right]\\&=\sum_{s=1}^{d}I_s^{[1]}(\mathbf{x},i,j)+\sum_{s=1}^{d}\sum_{t=1}^{d}I_{s,t}^{[2]}(\mathbf{x},i,j).\end{aligned}$$

From the definition of T_θ and S_θ it is immediately clear, that terms like $\|\partial/\partial\theta T_\theta\mathbf{x}\|$ are uniformly bounded over θ and $\mathbf{x}\in B$. By applying the Cauchy–Schwarz inequality several times it therefore suffices to show that

$$\int_A(\hat{m}(T_\theta\mathbf{x})-m(T_\theta\mathbf{x}))^2d\mathbf{x}=o_P(1),\qquad\int_A(\hat{m}(S_\theta\mathbf{x})-m(S_\theta\mathbf{x}))^2d\mathbf{x}=o_P(1),$$

$$\int_A\left(\frac{\partial}{\partial x_i}\hat{m}(T_\theta\mathbf{x})-\frac{\partial}{\partial x_i}m(T_\theta\mathbf{x})\right)^2d\mathbf{x}=o_P(1),$$

$$\int_A\left(\frac{\partial}{\partial x_i}\hat{m}(S_\theta\mathbf{x})-\frac{\partial}{\partial x_i}m(S_\theta\mathbf{x})\right)^2d\mathbf{x}=o_P(1),\quad 1\le i\le d$$

$$\int_A\left(\frac{\partial^2}{\partial x_i\partial x_j}\hat{m}(T_\theta\mathbf{x})-\frac{\partial^2}{\partial x_i\partial x_j}m(T_\theta\mathbf{x})\right)^2d\mathbf{x}=o_P(1),$$

$$\int_A\left(\frac{\partial^2}{\partial x_i\partial x_j}\hat{m}(S_\theta\mathbf{x})-\frac{\partial^2}{\partial x_i\partial x_j}m(S_\theta\mathbf{x})\right)^2d\mathbf{x}=o_P(1),\quad 1\le i,j\le d$$

uniformly over θ. We obtain for example, if $\max\{|\partial^2/\partial\theta_i\partial\theta_j(T_\theta\mathbf{x})_s|,|\partial^2/\partial\theta_i\partial\theta_j(S_\theta\mathbf{x})_s|\}\le C$ for some $C>0$

$$\begin{aligned}&\int_A|(\hat{m}(T_\theta\mathbf{x})-m(T_\theta\mathbf{x})-(\hat{m}(S_\theta\mathbf{x})-m(S_\theta\mathbf{x})))I_s^{[1]}(\mathbf{x},i,j)|d\mathbf{x}\\&\le C\int_A|\hat{m}(T_\theta\mathbf{x})-m(T_\theta\mathbf{x})-(\hat{m}(S_\theta\mathbf{x})-m(S_\theta\mathbf{x}))|\left|\frac{\partial}{\partial x_s}\hat{m}(T_\theta\mathbf{x})-\frac{\partial}{\partial x_s}m(T_\theta\mathbf{x})\right|d\mathbf{x}\\&\quad+C\int_A|\hat{m}(T_\theta\mathbf{x})-m(T_\theta\mathbf{x})-(\hat{m}(S_\theta\mathbf{x})-m(S_\theta\mathbf{x}))|\left|\frac{\partial}{\partial x_s}\hat{m}(S_\theta\mathbf{x})-\frac{\partial}{\partial x_s}m(S_\theta\mathbf{x})\right|d\mathbf{x}\\&=C_2\left(\int_{A_\theta}(\hat{m}(\mathbf{z})-m(\mathbf{z}))^2d\mathbf{z}\right)^{1/2}\left(\int_{A_\theta}\left(\frac{\partial}{\partial z_s}\hat{m}(\mathbf{z})-\frac{\partial}{\partial z_s}m(\mathbf{z})\right)^2d\mathbf{z}z\right)^{1/2}\\&\le C_2\left(\int_D(\hat{m}(\mathbf{z})-m(\mathbf{z}))^2d\mathbf{z}\right)^{1/2}\left(\int_D\left(\frac{\partial}{\partial z_s}\hat{m}(\mathbf{z})-\frac{\partial}{\partial z_s}m(\mathbf{z})\right)^2d\mathbf{z}\right)^{1/2}=o_P(1)\end{aligned}$$

by using Theorem 4. The other terms are estimated similarly. □

A.3. Proof of Theorem 3

We use the decomposition

$$\hat{L}_n(\hat{\vartheta})=\hat{L}_n(\vartheta)-(\vartheta-\hat{\vartheta})^T\hat{l}_n(\hat{\vartheta})-(\vartheta-\hat{\vartheta})^T\hat{h}(\xi_n)(\vartheta-\hat{\vartheta})$$

and immediately see from the previous proof that the second term on the right is 0 and the last term on the right is of order $O_P(n^{-d}a_n^{-d})=o_P((n^dh^{2\beta+d/2}a_n^{3d/2})^{-1})$. Therefore it suffices to show the weak convergence of the first term to the desired distribution. It is

$$\begin{aligned}L_n(\vartheta)&=\int_A\left(\sum_{\mathbf{r}}(w_{\mathbf{r}}(S_\vartheta\mathbf{x})-w_{\mathbf{r}}(T_\vartheta\mathbf{x}))\varepsilon_{\mathbf{r}}\right)^2d\mathbf{x}\\&\quad+2\int_A\left(\sum_{\mathbf{r}}(w_{\mathbf{r}}(S_\vartheta\mathbf{x})-w_{\mathbf{r}}(T_\vartheta\mathbf{x}))\varepsilon_{\mathbf{r}}\right)\left(\sum_{\mathbf{s}}(w_{\mathbf{s}}(S_\vartheta\mathbf{x})-w_{\mathbf{s}}(T_\vartheta\mathbf{x}))\Psi m(\mathbf{x_s})\right)d\mathbf{x}.\end{aligned}$$

As in the proof of Theorem 4 one easily sees that the last two terms on the right are of order $o_P((n^d h^{2\beta+d/2} a_n^{3d/2})^{-1})$. We get

$$L_n(\vartheta) = \sum_{\mathbf{r}} \int_A (w_{\mathbf{r}}(S_\vartheta \mathbf{x}) - w_{\mathbf{r}}(T_\vartheta \mathbf{x}))^2 d\mathbf{x}\varepsilon_{\mathbf{r}} + \sum_{\mathbf{r}\neq\mathbf{s}} \int_A (w_{\mathbf{r}}(S_\vartheta \mathbf{x}) - w_{\mathbf{r}}(T_\vartheta \mathbf{x}))(w_{\mathbf{s}}(S_\vartheta \mathbf{x}) - w_{\mathbf{s}}(T_\vartheta \mathbf{x})) d\mathbf{x}\varepsilon_{\mathbf{r}}\varepsilon_{\mathbf{s}}.$$

The rest of the proof now follows along the lines of the proof of Theorem 4 when considering $I_1^{(\mathbf{j})}$.

References

[1] I.A. Ahmad, Q. Li, Testing symmetry of an unknown density function by kernel method, Nonparametr. Stat. 7 (1997) 279–293.
[2] M. Birke, H. Dette, K. Stahljans, Testing symmetry of a nonparametric bivariate regression function, J. Nonparam. Statist. 23 (2) (2011) 547–565.
[3] N. Bissantz, M. Birke, Asymptotic normality and confidence intervals for inverse regression models with convolution-type operators, J. Multivariate Anal. 100 (2009) 2364–2375.
[4] N. Bissantz, H. Dette, K. Proksch, Model checks in inverse regression models with convolution-type operators, Technical Report SFB 823.
[5] Bissantz, Holzmann, Pawlak, Testing for image symmetries—with application to confocal microscopy, IEEE Trans. Inform. Theory 55 (2010) 1841–1855.
[6] A. Cabaña, M. Cabaña, Tests of symmetry based on transformed empirical processes, Canad. J. Statist. 28 (2000) 829–839.
[7] P. de Jong, A central limit theorem for generalized quadratic forms, Probab. Theory Related Fields 75 (1987) 261–277.
[8] H. Dette, A consistent test for the functional form of a regression based on a difference of variance estimators, Ann. Statist. 27 (1999) 1012–1040.
[9] H. Dette, S. Kusi-Appiah, N. Neumeyer, Testing symmetry in nonparametric regression models, Nonparametr. Stat. 14 (5) (2002) 477–494.
[10] R.L. Eubank, Nonparametric Regression and Spline Smoothing, second ed., in: Statistics: Textbooks and Monographs, vol. 157, Marcel Dekker, Inc., New York, 1999.
[11] W. Härdle, J.S. Marron, Semiparametric comparison of regression curves, Ann. Statist. 18 (1990) 63–89.