# Nonparametric estimation of non-stationary velocity fields from 3D particle tracking velocimetry data

Michael Kohler[1] and Adam Krzyżak[2,*]

[1] *Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany, email: kohler@mathematik.tu-darmstadt.de*

[2] *Department of Computer Science and Software Engineering, Concordia University, 1455 De Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8, email: krzyzak@cs.concordia.ca*

December 3, 2010

**Abstract**

Nonparametric estimation of nonstationary velocity fields from 3D particle tracking velocimetry data is considered. The velocities of tracer particles are computed from their positions measured experimentally with random errors by high-speed cameras observing turbulent flows in fluids. Thus captured discrete data is plugged into a smoothing spline estimate which is used to estimate the velocity field at arbitrary points. The estimate is further smoothed over several time frames using fixed design kernel regression estimate. Consistency of the resulting estimate is investigated. Its performance is validated on the real data obtained by measuring a fluid flow of a liquid in a (rotating) squared tank agitated by an oscillating grid.

*AMS classification:* Primary 62G08; secondary 62P30.

*Key words and phrases:* Consistency, nonparametric regression, particle tracking velocimetry.

---

*Corresponding author. Tel: +1-514-848-2424 ext. 3007, Fax: +1-514-848-2830

Running title: *Estimation of velocity fields*

# 1 Introduction

In the recent years there has been tremendous progress in accurate and fast measurement techniques in fluid mechanics. This resulted in large amount of data which presently require development of new statistical tools to process and interpret the data.

In this article we analyze data produced by the 3D Particle Tracking Velocimetry (3D-PTV), see, e.g., Raffel et al. (1998). This technique allows visualization of a flow by recording the laser light scattered by naturally buoyant tracer particles in a fluid and subsequently using it to determine positions of the particles in consecutive frames. To do this 3D-PTV fits short trajectories of the particles to the observed pictures. These trajectories consist of approximately 20 time steps and are modeled by cubic splines. From these trajectories the estimates of the position and the velocity of the tracer particles are derived (cf., e.g., Lüthi, Tsinober and Kinzelbach (2005)). Thus data is produced which contains positions of particles and corresponding values of the fluid velocity field at these positions. This data contains two kinds of errors: firstly errors due to measurement errors for the locations of the tracer particles, and secondly errors due to fitting of the trajectories to these locations of the tracer particles. In this paper we want to use this data to estimate the velocity field at arbitrary locations and times.

Experimental studies on estimation of velocity fields in turbulent flows have been carried out, among others, by Guala et al. (2008), Kunnen, Geurts and Clercx (2010), Lüthi, Tsinober and Kinzelbach (2005), Messio et al. (2008) and Speetjens, Clercx and Van Heijst (2004). These researchers used kernel regression and local linear kernel regression estimates to smooth and interpolate the observed data. No theoretical analysis of the estimates was provided.

In this paper we pose the problem of recovering velocity fields at arbitrary locations and times as a non-stationary regression estimation problem with regression functions changing in time. The regression functions are estimated by smoothing spline estimates which are subsequently smoothed in time domain using the fixed design kernel regression estimate.

We prove consistency of the estimates and apply them to real data obtained from the 3D-PTV measuring a time-dependent velocity field in a (rotating) water tank agitated by an oscillating grid.

## 2 Definition of the estimates

Let $(X_t, Y_t)$ $(t \in [0,1])$ be $\mathbb{R}^d \times \mathbb{R}^d$-valued random vectors defined on the same probability space. Let the corresponding time dependent $d$-dimensional velocity field

$$m : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$$

be given by

$$m(t,x) = \mathbf{E}\{Y_t | X_t = x\},$$

and denote the distribution of $X_t$ by $\mu_t$. For $N \in \mathbb{N}$ we consider equidistant time points

$$t_k = t_k(N) = \frac{k}{N} \quad (k = 0, \dots, N)$$

and we assume that for each time point $t_k$ we are given a velocity field sample

$$\mathcal{D}_{n_{t_k}} = \left\{ (X_1^{(t_k)}, Y_1^{(t_k)}), \dots, (X_{n_{t_k}}^{(t_k)}, Y_{n_{t_k}}^{(t_k)}) \right\}.$$

Let $k \in \mathbb{N}$ with $2k > d$ and denote by $W^k(\mathbb{R}^d)$ the Sobolev space containing all functions $f : \mathbb{R}^d \to \mathbb{R}^d$ where all derivatives of total order $k$ of all components are in $L^2(\mathbb{R}^d)$. The condition $2k > d$ implies that the functions in $W^k(\mathbb{R}^d)$ are continuous and hence the evaluation of a function at a point is well defined. Let $m_{n_{t_k}}^{(t_k)}(\cdot) = m_{n_{t_k}}^{(t_k)}(\cdot, \mathcal{D}_{n_{t_k}})$ be the smoothing spline estimate of $m(t_k, \cdot)$ defined by

$$\tilde{m}_{n_{t_k}}^{(t_k)}(\cdot) = \arg \min_{f \in W^k(\mathbb{R}^d)} \left[ \frac{1}{n_{t_k}} \sum_{i=1}^{n_{t_k}} \left\| Y_i^{(t_k)} - f(X_i^{(t_k)}) \right\|_2^2 + \lambda_{t_k} \cdot J_k^2(f) \right] \tag{1}$$

where

$$J_k^2(f) = \sum_{\alpha_1, \dots, \alpha_d \in \mathbb{N},\, \alpha_1 + \dots + \alpha_d = k} \frac{k!}{\alpha_1! \cdot \ldots \cdot \alpha_d!} \int_{\mathbb{R}^d} \left\| \frac{\partial^k f}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}(x) \right\|_2^2 dx, \tag{2}$$

and by

$$m_{n_{t_k}}^{(t_k)}(x) = T_{\beta_N} \tilde{m}_{n_{t_k}}^{(t_k)}(x) := \max\left( \min\left( \tilde{m}_{n_{t_k}}^{(t_k)}(x), \beta_N \right), -\beta_N \right) \quad (x \in \mathbb{R}^d). \tag{3}$$

Here $\|\cdot\|_2$ denotes the Euclidean norm in $\mathbb{R}^d$ and the truncation level $\beta_N > 0$ is a parameter of the estimate which we will choose later such that $\beta_N \to \infty$ $(N \to \infty)$.

Let $l = \binom{d+k-1}{d}$ and let $\phi_1, \dots, \phi_l$ be all monomials $x_1^{\alpha_1} \cdot \ldots \cdot x_d^{\alpha_d}$ of total degree $\alpha_1 + \ldots + \alpha_d$ less than $k$. Define $R : \mathbb{R}_+ \to \mathbb{R}$ by

$$R(u) = \begin{cases} u^{2k-d} \cdot \log(u) & \text{if} \quad 2k-d \text{ is even}, \\ u^{2k-d} & \text{if} \quad 2k-d \text{ is odd}, \end{cases}$$

where $\log(z)$ is the natural logarithm of $z > 0$. It follows from Section V in Duchon (1976) that there exists a function of the form

$$\tilde{m}_{n_{t_k}}^{(t_k)}(x) = \sum_{i=1}^{n} a_i R(\|x - X_i^{(t_k)}\|_2) + \sum_{j=1}^{l} b_j \phi_j(x) \tag{4}$$

which achieves the minimum in (1), and that the coefficients $a_1, \dots, a_n, b_1, \dots, b_l \in \mathbb{R}^d$ of this function can be computed by solving $d$ linear systems of equations. Under some additional assumptions on the $X_1^{(t_k)}, \dots, X_n^{(t_k)}$ this is also shown in Section 2.4 of Wahba (1990).

In order to estimate the velocity field continuously in time, we make local averaging of the above regression estimates. Let $K : \mathbb{R} \to \mathbb{R}$ be a function (so-called kernel) satisfying

$$c_1 \cdot I_{[-a_1,a_1]}(x) \le K(x) \le c_2 \cdot I_{[-a_2,a_2]}(x) \quad (x \in \mathbb{R}) \tag{5}$$

for some $a_2 \ge a_1 > 0, c_2 \ge c_1 > 0$, where $I_A$ denotes the indicator function of the set $A$, and let $h_N > 0$ be the so-called bandwidth of the kernel estimate. We define the kernel estimate of the velocity field by

$$\hat{m}_N(t,x) = \frac{\sum_{k=0}^{N} m_{n_{t_k}}^{(t_k)}(x) \cdot K\left(\frac{t-t_k}{h_N}\right)}{\sum_{l=0}^{N} K\left(\frac{t-t_l}{h_N}\right)}. \tag{6}$$

# 3 Main theoretical result

In our main theoretical result we impose the following conditions on the underlying distribution:

(A1) $(X_t, Y_t) \in [0,1]^d \times \mathbb{R}^d$ $a.s.$ for all $t \in [0,1]$.

(A2) The random variables

$$(X_{t_k}, Y_{t_k}), (X_1^{(t_k)}, Y_1^{(t_k)}), \dots, (X_{n_k}^{(t_k)}, Y_{n_k}^{(t_k)})$$

are independent and identically distributed ($k \in \{0, \dots, N\}$).

(A3) $\sup_{t\in[0,1]} \mathbf{E}\{|Y_t|^8|\} \leq c_3$ for some constant $c_3 \in \mathbb{R}_+$.

(A4) $m$ (defined by $m(t,x) = \mathbf{E}\{Y_t|X_t = x\}$) is continuous on $[0,1] \times [0,1]^d$.

(A5)

$$\beta_N^2 \cdot \sup_{k:|t-t_k|\leq h_N} \sup_{A\in\mathcal{B}_d} |\mu_t(A) - \mu_{t_k}(A)| \to 0 \quad (N \to \infty)$$

for all $t \in [0,1]$.

**Remark 1.** Assumptions (*A*4) and (*A*5) are plausible in the application described in Section 4 due to the viscosity of the water.

**Remark 2.** In case that a density $f(t,\cdot)$ of $\mu_t$ exists (with respect to the Lebesgue-Borel measure), which is Holder continuous with exponent $\alpha$ with respect to $t$, i.e., which satisfies

$$|f(t,x) - f(s,x)| \leq C \cdot |t-s|^\alpha \quad (s,t \in [0,1], x \in [0,1]^d),$$

we have for any $A \in \mathcal{B}_d$, $A \subseteq [0,1]^d$

$$|\mu_t(A) - \mu_{t_k}(A)| \leq \int_A |f(t,x) - f(t_k,x)|\, dx \leq C \cdot |t - t_k|^\alpha.$$

Hence in this case condition (A6) is implied by

$$\beta_N^2 \cdot h_N^\alpha \to 0 \quad (N \to \infty).$$

**Theorem 1** *Let the estimate $\hat{m}_n$ be defined as in Section 2, where the kernel $K$ satisfies (5), assume that*

$$n_{t_k} \in \{n_{min}(N), n_{min}(N)+1, \ldots, n_{max}(N)\}\,, \tag{7}$$

*and assume that the parameters of the estimate satisfy*

$$\lambda_{t_k} \in [\lambda_{min}(N), \lambda_{max}(N)]\,, \tag{8}$$

$$\beta_N^4 / n_{min}(N) \to 0 \quad (N \to \infty), \tag{9}$$

$$\beta_N \to \infty \quad (N \to \infty), \tag{10}$$

$$\beta_N^2 \cdot \frac{n_{max}(N) - n_{min}(N)}{n_{min}(N)^2} \to 0 \quad (N \to \infty), \tag{11}$$

$$\frac{n_{min}(N) \cdot \lambda_{min}(N)^{d/(2k)}}{\beta_N^{4+d/k} \cdot \log(n_{max}(N)^2 \cdot \beta_N^2)} \to \infty \quad (N \to \infty), \tag{12}$$

$$\lambda_{max}(N) \to 0 \quad (N \to \infty), \tag{13}$$

*and*

$$h_N \to 0 \quad (N \to \infty). \tag{14}$$

*Assume furthermore that the underlying dsitribution satisfies* $(A1)$-$(A6)$. *Then*

$$\mathbf{E} \int \|\hat{m}_N(t,x) - m(t,x)\|_2^2 \, \mu_t(dx) \to 0 \quad (N \to \infty).$$

*for all* $t \in [0,1]$.

**Remark 3.** Assume that for some $r > 0$

$$n_{min}(N) \geq \log(N)^{5+2d/k}, \quad n_{max}(N) \leq N^r \quad \text{and} \quad \log(n)^2 \cdot \frac{n_{max}(N)}{n_{min}(N)^2} \to 0 \quad (N \to \infty).$$

Then the above conditions on the parameters are e.g. satisfied if we set $\beta_N = \log(N)$ and choose $\lambda_{max}(N)$ and $h_N$ such that (13) and (14) are satisfied and choose $\lambda_{min}(N) \leq \lambda_{max}(N)$ such that

$$\lambda_{min}(N) \geq \frac{\log(N)^{3+10k/d}}{n_{min}(N)^{2k/d}} \quad (N \in \mathbb{N}).$$

# 4 Application to real data

In this section we apply our estimation procedure to real data collected by Kinzel (2010). The aim of the experiment was to measure turbulent fluid flows in a water tank agitated in the first experiment by an oscillating grid and in a second experiment by a rotation of the water tank and by an oscillating grid in order to validate a theoretical model. The measurements were performed by a 3D-PTV method. Two kinds of Polystrene particles were seeded into the fluid and illuminated by a laser light. The scattered light was collected by two high-speed cameras with $1024 \times 1024$ pixels resolution in combination with image splitters that mimick a four-camera setup for each of the high-speed cameras. The cameras collected images at a rate of 125 $Hz$ over the period of 40 seconds yielding 5000 images for each camera. The first camera collected data from $60\mu m$ Polystyrene tracer particles within a large observation volume of size $50 \times 50 \times 40$ $mm^3$, while the second one was focused on $80\mu m$ Polysterene Rhodamin tracer particles within a small observation volume of size $15 \times 15 \times 15$ $mm^3$ inside the larger observation volume. With this setup the seeding density of 15 particles per 10 $mm^3$ was achieved for the large observation volume and 200 particles for the small observation volume.

In the sequel our estimates are computed from the data collected within the large observation volume. We use data from the small volume only for evaluation of the estimates. Trajectories of tracer particles of approximately 20 time steps modeled by cubic splines have been fitted to the observed pictures in order to estimate the positions and the velocities of the tracer particles. Here not all particles were sucessfully matched, in particular due to lack of light close to the boundaries of the observation volume. Hence it is impossible to determine complete trajectories of the tracer particles. The corresponding data consisting of pairs $(x, y)$, where $x \in \mathbb{R}^3$ is the location of a particle and $y \in \mathbb{R}^3$ is its velocity, for time $t = 20s$ is illustrated in Figure 1 (large tracer particles for the experiment without rotation), Figure 2 (small tracer particles for the experiment without rotation), Figure 3 (large tracer particles for the experiment with rotation) and Figure 4 (small tracer particles for the experiment with rotation). Here each data point $(x, y)$ is represented by an arrow located at the particles position with direction and length given by $y$.

In Figures 5 and 6 we show the kernel smoothing spline estimate with naive kernel applied to the data corresponding to the larger tracer particles for the experiment without rotation and for the experiment with rotation, resp. Here the smoothing parameter of the smoothing spline is chosen via generalized cross-validation, and the bandwidth of the kernel estimate is chosen from the set $\{1, 2, 3, 4, 6, 8, 12\}$ via 20-fold cross-validation. ¿From Figures 5 and 6 we can clearly see the different shapes of the currents in the flow in the two situations, and this is much better visible than in the originally measured data (cf. Figures 1 and 3).

If we compute the empirical $L_2$ risk for this estimate on the data corresponding to the smaller tracer particles (which is not used for computation of the estimate) we get for the experiment without rotation 0.000833. Here the same value for the empirical $L_2$ risk is also achieved for the smoothing spline estimate using only the data from time $t = 20s$.

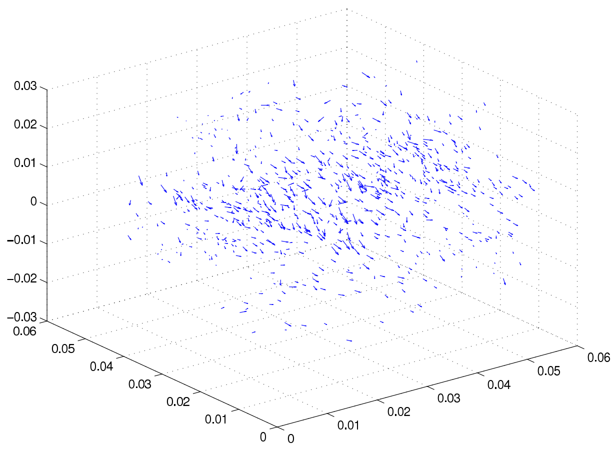


Figure 1: Measured data for the large tracer particles in the experiment without rotation for time $t = 20s$.

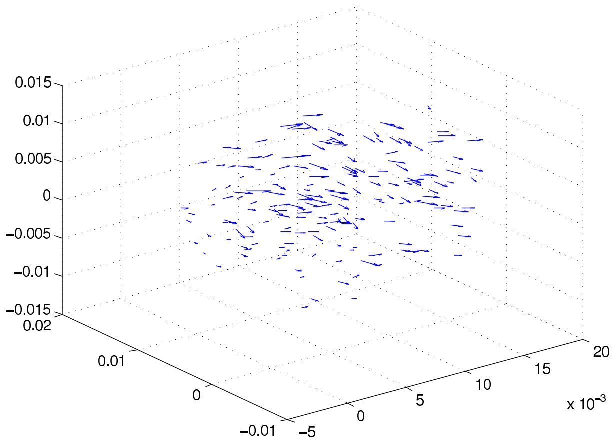


Figure 2: Measured data for the smaller tracer particles in the experiment without rotation for time $t = 20s$.

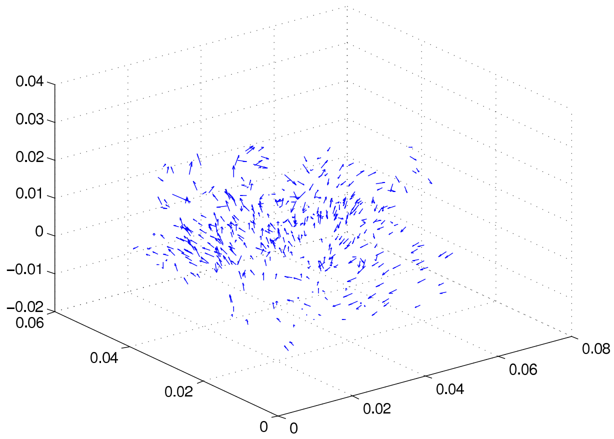


Figure 3: Measured data for the larger tracer particles in the experiment with rotation for time $t = 20s$.

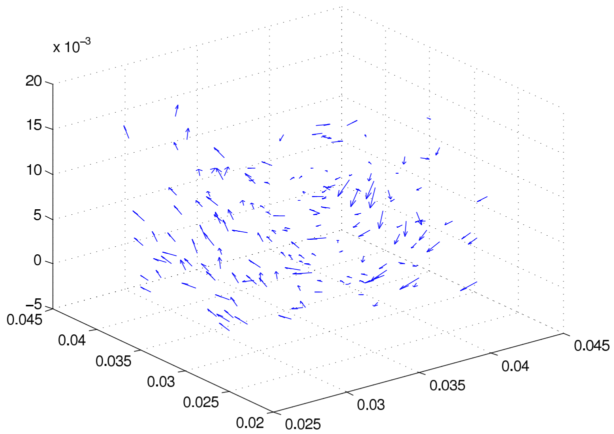


Figure 4: Measured data for the smaller tracer particles in the experiment with rotation for time $t = 20s$.

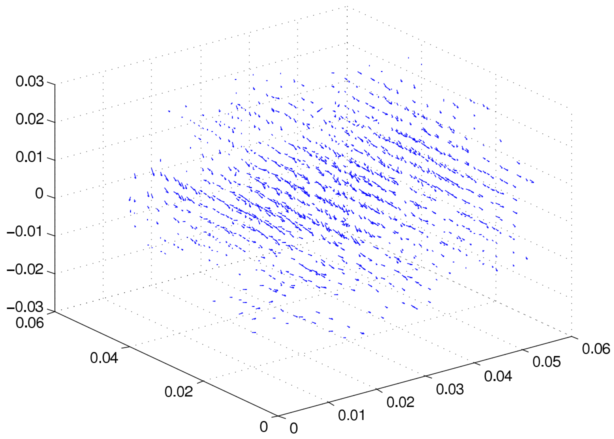


Figure 5: Kernel smoothing spline estimate for time $t = 20s$ with naive kernel for the experiment without rotation.

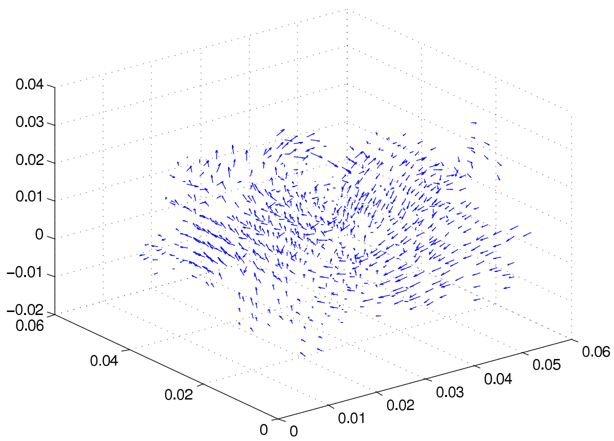


Figure 6: Kernel smoothing spline estimate for time $t = 20s$ with naive kernel for the experiment with rotation.

But if we compute the corresponding values for the experiments with rotation the empirical $L_2$ risks are $6.92 \cdot 10^{-5}$ in case of our newly proposed estimate and $7.18 \cdot 10^{-5}$ in case of the smoothing spline estimate using only data at time $t = 20s$.

# 5 Proofs

We begin with two auxiliary lemmas needed in the proof of the main result. As we will see in the proof of the main result, it suffices to formulate and prove these auxiliary results for scalar response variables.

For our first lemma we introduce the following notation: For a random function $f : \mathbb{R}^d \to \mathbb{R}$ (i.e., $f(x)$ is a random variable for each $x \in \mathbb{R}^d$) let $\mathbf{E}(f), \mathbf{Var}(f) : \mathbb{R}^d \to \mathbb{R}$ be deterministic functions defined by

$$
\begin{aligned}
&\mathbf{E}(f)(x) = \mathbf{E}\{f(x)\}, \\
&\mathbf{Var}(f)(x) = \mathbf{Var}\{f(x\} = \mathbf{E}\{|f(x) - \mathbf{E}(f(x))|^2\}.
\end{aligned}
$$

Set $t_k = k/N (k = 0, \dots, N)$, let $m : [0,1] \times \mathbb{R}^d \to \mathbb{R}$ and assume

$$
Y_k = m(t_k, \cdot) + \epsilon_k \quad (k = 0, \dots, N)
$$

where $\epsilon_0, \dots, \epsilon_N : \mathbb{R}^d \to \mathbb{R}$ are random functions satisfying

$$
\mathbf{E}(\epsilon_l \cdot \epsilon_k) = 0 \quad (l, k \in \{0, \dots, N\}, l \neq k).
$$

Let $\bar{Y}_0, \dots, \bar{Y}_N$ be arbitrary random functions $\mathbb{R}^d \to \mathbb{R}$ and define

$$
\hat{m}_N(t, \cdot) = \frac{\sum_{k=0}^N \bar{Y}_k K\left(\frac{t - t_k}{h_N}\right)}{\sum_{l=0}^N K\left(\frac{t - t_l}{h_N}\right)}
$$

for some $h_N > 0$ and some kernel $K : \mathbb{R} \to \mathbb{R}$ satisfying (5). For $t \in [0,1]$ let $\mu_t$ be a probability measure on $\mathbb{R}^d$. For $f : \mathbb{R}^d \to \mathbb{R}$ and $p > 0$ set

$$
\|f\|_{L_p(\mu_t)} = \left( \int_{\mathbb{R}^d} |f(x)|^p \, \mu_t(dx) \right)^{1/p}.
$$

Then the following result holds.

**Lemma 1** *Assume $N \cdot h_N > 1/a_1$. For any $t \in [0,1]$ we have*

$$\mathbf{E}\left(\|\hat{m}_N(t,\cdot) - m(t,\cdot)\|^2_{L_2(\mu_t)}\right)$$
$$\leq 3 \cdot \max_{k:|t-t_k|\leq a_2 \cdot h_N} \mathbf{E}\|\bar{Y}_k - Y_k\|^2_{L_2(\mu_t)} + 3 \cdot \frac{c_2}{c_1 \cdot a_1} \cdot \max_{k:|t-t_k|\leq a_2 \cdot h_N} \|\mathbf{E}(\epsilon_k^2)\|_{L_1(\mu_t)} \cdot \frac{1}{N \cdot h_N}$$
$$+3 \cdot \frac{c_2}{c_1 \cdot a_1} \cdot \max_{k:|t-t_k|\leq a_2 \cdot h_N} \|m(t_k,\cdot) - m(t,\cdot)\|^2_{L_2(\mu_t)}.$$

**Proof.**

$$\begin{aligned}
& \mathbf{E}\left(\|\hat{m}_N(t,\cdot) - m(t,\cdot)\|^2_{L_2(\mu_t)}\right) \\
= \; & \mathbf{E}\left(\left\| \frac{\sum_{k=0}^N (\bar{Y}_k - Y_k) K\left(\frac{t-t_k}{h_N}\right)}{\sum_{l=0}^N K\left(\frac{t-t_l}{h_N}\right)} + \frac{\sum_{k=0}^N \epsilon_k K\left(\frac{t-t_k}{h_N}\right)}{\sum_{l=0}^N K\left(\frac{t-t_l}{h_N}\right)} \right.\right. \\
& \qquad \left.\left. + \frac{\sum_{k=0}^N (m(t_k,\cdot) - m(t,\cdot)) K\left(\frac{t-t_k}{h_N}\right)}{\sum_{l=0}^N K\left(\frac{t-t_l}{h_N}\right)} \right\|^2_{L_2(\mu_t)}\right) \\
\leq \; & 3 \cdot \mathbf{E}\left(\left\| \frac{\sum_{k=0}^N (\bar{Y}_k - Y_k) K\left(\frac{t-t_k}{h_N}\right)}{\sum_{l=0}^N K\left(\frac{t-t_l}{h_N}\right)} \right\|^2_{L_2(\mu_t)}\right) \\
& +3 \cdot \mathbf{E}\left(\left\| \frac{\sum_{k=0}^N \epsilon_k K\left(\frac{t-t_k}{h_N}\right)}{\sum_{l=0}^N K\left(\frac{t-t_l}{h_N}\right)} \right\|^2_{L_2(\mu_t)}\right) \\
& +3 \cdot \mathbf{E}\left(\left\| \frac{\sum_{k=0}^N (m(t_k,\cdot) - m(t,\cdot)) K\left(\frac{t-t_k}{h_N}\right)}{\sum_{l=0}^N K\left(\frac{t-t_l}{h_N}\right)} \right\|^2_{L_2(\mu_t)}\right) \\
\stackrel{def}{=} \; & 3 \cdot T_{1,N} + 3 \cdot T_{2,N} + 3 \cdot T_{3,N}.
\end{aligned}$$

By pointwise application of Jensen's inequality and by the triangle inequality we get

$$\begin{aligned}
T_{1,N} &= \mathbf{E}\left(\left\| \left(\frac{\sum_{k=0}^N (\bar{Y}_k - Y_k) K\left(\frac{t-t_k}{h_N}\right)}{\sum_{l=0}^N K\left(\frac{t-t_l}{h_N}\right)}\right)^2 \right\|_{L_1(\mu_t)}\right) \\
&\leq \mathbf{E}\left(\left\| \frac{\sum_{k=0}^N (\bar{Y}_k - Y_k)^2 K\left(\frac{t-t_k}{h_N}\right)}{\sum_{l=0}^N K\left(\frac{t-t_l}{h_N}\right)} \right\|_{L_1(\mu_t)}\right) \\
&\leq \frac{\sum_{k=0}^N \mathbf{E}(\|(\bar{Y}_k - Y_k)^2\|_{L_1(\mu_t)}) \cdot K\left(\frac{t-t_k}{h_N}\right)}{\sum_{l=0}^N K\left(\frac{t-t_l}{h_N}\right)} \\
&\leq \max_{k:|t_k-t|\leq a_2 \cdot h_N} \mathbf{E}(\|\bar{Y}_k - Y_k\|^2_{L_2(\mu_t)}.
\end{aligned}$$

An application of Fubini's theorem, and use of $\mathbf{E}(\epsilon_l \cdot \epsilon_k) = 0$ for $l \neq k$, and application of the triangle inequality yields

$$
\begin{aligned}
T_{2,N} \quad &= \quad \left\| \mathbf{E} \left\{ \left( \frac{\sum_{k=0}^N \epsilon_k K\left(\frac{t-t_k}{h_N}\right)}{\sum_{l=0}^N K\left(\frac{t-t_l}{h_N}\right)} \right)^2 \right\} \right\|_{L_1(\mu_t)} \\
&= \quad \left\| \frac{\sum_{k=0}^N K^2\left(\frac{t-t_k}{h_N}\right) \mathbf{E}\{\epsilon_k^2\}}{\left(\sum_{l=0}^N K\left(\frac{t-t_l}{h_N}\right)\right)^2} \right\|_{L_1(\mu_t)} \\
&\leq \quad \frac{1}{\left(\sum_{l=0}^N K\left(\frac{t-t_l}{h_N}\right)\right)^2} \sum_{k=0}^N K^2\left(\frac{t-t_k}{h_N}\right) \|\mathbf{E}\{\epsilon_k^2\}\|_{L_1(\mu_t)} \\
&\overset{K \leq c_2 \cdot I_{[-a_2,a_2]}}{\leq} \quad c_2 \cdot \frac{\sum_{k=0}^N K\left(\frac{t-t_k}{h_N}\right)}{\left(\sum_{l=0}^N K\left(\frac{t-t_l}{h_N}\right)\right)^2} \cdot \max_{k:|t_k-t| \leq h_N} \|\mathbf{E}\{\epsilon_k^2\}\|_{L_1(\mu_t)} \\
&\overset{K \geq c_1 I_{[-a_1,a_1]}}{\leq} \quad \frac{c_2}{c_1 \cdot a_1} \cdot \frac{1}{N \cdot h_N} \max_{k:|t_k-t| \leq h_N} \|\mathbf{E}\{\epsilon_k^2\}\|_{L_1(\mu_t)}
\end{aligned}
$$

where the last inequality follows from equidistant distribution of $t_l$ in $[0,1]$ and the assumption $N \cdot h_N > 1/a_1$. Finally, by pointwise application of Jensen's inequality and by the triangle inequality we get

$$
\begin{aligned}
T_{3,N} \quad &\leq \quad \frac{1}{\sum_{l=0}^N K\left(\frac{t-t_l}{h_N}\right)} \sum_{k=0}^N K\left(\frac{t-t_k}{h_N}\right) \|(m(t_k,\cdot) - m(t,\cdot))^2\|_{L_1(\mu_t)} \\
&\overset{K \leq c_2 I_{[-a_2,a_2]}}{\leq} \quad \max_{k:|t_k-t| \leq a_2 \cdot h_N} \|m(t_k,\cdot) - m(t,\cdot)\|_{L_2(\mu_t)}^2.
\end{aligned}
$$

The proof is complete. □

In our next auxiliary lemma we prove a uniform consistency result for multivariate smoothing spline regression estimates.

**Lemma 2** *Let $L : [0,\infty) \to \mathbb{R}_+$ be a monotonically decreasing function satisfying*

$$L(h) \to 0 \quad (h \to 0).$$

*For constants $c_4, c_5$ let $\mathcal{D}$ be the class of all random variables $(X,Y)$ satisfying*

*(A1′) $X$ $\mathbb{R}^d$-valued, $Y$ $\mathbb{R}$-valued,*

*(A2′)* $X \in [0,1]^d$ *a.s.*

*(A3′)* $\mathbf{E}\{|Y|^8\} \leq c_4$,

*(A4′)* $\sup_{x \in [0,1]^d} |m(x)| \leq c_5$,

*(A5′) For all* $x, z \in \mathbb{R}^d$*:* $|m(x) - m(z)| \leq L(\|x - z\|_2)$.

*For* $N \in \mathbb{N}$ *let* $\beta_N, \lambda_{min}(N), \lambda_{max}(N) \in \mathbb{R}_+$ *and let* $n_{min}(N), n_{max}(N) \in \mathbb{N}$*, such that* $\lambda_{min}(N) \leq \lambda_{max}(N)$ *and* $n_{min}(N) \leq n_{max}(N)$*. For* $(X,Y) \in \mathcal{D}$ *let* $(X_1,Y_1), (X_2,Y_2), \ldots$ *be i.i.d. copies of* $(X,Y)$*, which are independent of* $(X,Y)$*, and let* $\hat{\lambda} = \hat{\lambda}(N)$ *and* $\hat{n} = \hat{n}(N)$ *be random variables with values in*

$$[\lambda_{min}(N), \lambda_{max}(N)] \quad \textit{and} \quad \{n_{min}(N), n_{min}(N)+1, \ldots, n_{max}(N)\}, \; \textit{resp.}$$

*Define a smoothing spline estimate* $m_{\hat{n},\hat{\lambda}}$ *by*

$$\tilde{m}_{\hat{n},\hat{\lambda}}(\cdot) = \arg \min_{f \in W^k(\mathbb{R}^d)} \left[ \frac{1}{\hat{n}} \sum_{i=1}^{\hat{n}} |Y_i - f(X_i)|^2 + \hat{\lambda} \cdot J_k^2(f) \right]$$

*where*

$$J_k^2(f) = \sum_{\alpha_1,\ldots,\alpha_d \in \mathbb{N},\, \alpha_1+\ldots+\alpha_d = k} \frac{k!}{\alpha_1! \cdot \ldots \cdot \alpha_d!} \int_{\mathbb{R}^d} \left| \frac{\partial^k f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}(x) \right|^2 dx,$$

*and by*

$$m_{\hat{n},\hat{\lambda}}(x) = T_{\beta_N} \tilde{m}_{\hat{n},\hat{\lambda}}(x)$$

*where*

$$T_{\beta_N} z := \max\{\min\{z, \beta_N\}, -\beta_N\} \quad (z \in \mathbb{R}).$$

*Assume that (9)–(13) hold. Then*

$$\sup_{(X,Y) \in \mathcal{D}} \mathbf{E} \int |m_{\hat{n},\hat{\lambda}}(x) - m(x)|^2 \mathbf{P}_X(dx) \to 0 \quad (N \to \infty).$$

**Proof.** The proof is an extension of the proof of Theorem 1 in Kohler and Krzyżak (2001).

Let $\epsilon > 0$ be arbitrary. *In the first step of the proof* we construct for $(X,Y) \in \mathcal{D}$ a deterministic approximation $g_\epsilon$ of $m(x) = \mathbf{E}\{Y|X=x\}$ such that

$$\int |m(x) - g_\epsilon(x)|^2 \mathbf{P}_X(dx) < \epsilon \quad \text{and} \quad J_k^2(g_\epsilon) \leq c_6 \cdot c_5^2$$

where $c_6 \in \mathbb{R}_+$ does not depend on the distribution of $(X, Y)$. To do this, let $B$ be the (univariate) B-spline of order $M = k + 1$ with support $supp(B) = [0, M + 1]$ and knots $0, 1, \ldots, M + 1$ (cf., e.g., de Boor (1978)). Then $B$ is $k$-times continuously differentiable, and its $k$th derivative is bounded in absolute value by some constant. Choose $K \in \mathbb{N}$ such that

$$L(\sqrt{d} \cdot (M + 1)/K) < \sqrt{\epsilon}$$

and define $g_\epsilon$ by

$$\begin{aligned} & g_\epsilon(x^{(1)}, \ldots, x^{(d)}) \\ &= \sum_{j_1, \ldots, j_d \in \{-M, \ldots, K-1\}} m\left(\frac{j_1}{K}, \ldots, \frac{j_d}{K}\right) \cdot B\left(K \cdot (x^{(1)} - \frac{j_1}{K})\right) \cdot \ldots \cdot B\left(K \cdot (x^{(d)} - \frac{j_d}{K})\right). \end{aligned}$$

Using the fact that the basis functions from a B-spline basis sum up to one (cf., e.g., de Boor (1978)) we get for any $x = (x^{(1)}, \ldots, x^{(d)}) \in [0, 1]^d$

$$|m(x) - g_\epsilon(x)| \leq \sup_{z: \max_{i=1,\ldots,d} |x^{(i)} - z^{(i)}| \leq (M+1)/K} |m(x) - m(z)| \leq L(\sqrt{d} \cdot (M + 1)/K) < \sqrt{\epsilon},$$

hence

$$\int |m(x) - g_\epsilon(x)|^2 \mathbf{P}_X(dx) < \epsilon.$$

Furthermore, using that the derivatives of the B-splines are bounded and have compact support and that at each point only $(M + 1)^d$ of the products of the B-splines occuring in the definition of $g_\epsilon$ are unequal to zero we get for some constant $c_7 > 0$ depending on $M$, $k$ and $d$:

$$\begin{aligned} & J_k^2(g_\epsilon) \\ & \leq c_7 \cdot c_5^2 \cdot \max_{\substack{\alpha_1, \ldots, \alpha_d \in \mathbb{N}, \\ \alpha_1 + \ldots + \alpha_d = k}} \max_{x \in \mathbb{R}^d} \left| \frac{\partial^k}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}} B\left(K \cdot (x^{(1)} - \frac{j_1}{K})\right) \cdot \ldots \cdot B\left(K \cdot (x^{(d)} - \frac{j_d}{K})\right) \right|^2 \\ & \leq c_8 \cdot c_5^2 \cdot K^{2k}. \end{aligned}$$

*In the second step of the proof* we split the $L_2$ error as follows:

$$\begin{aligned} & \mathbf{E} \int |m_{\hat{n}, \hat{\lambda}}(x) - m(x)|^2 \mathbf{P}_X(dx) = \mathbf{E}\{|m_{\hat{n}, \hat{\lambda}}(X) - Y|^2\} - \mathbf{E}\{|m(X) - Y|^2\} \\ & = \mathbf{E}\{|m_{\hat{n}, \hat{\lambda}}(X) - Y|^2\} - (1 + \epsilon)\mathbf{E}\{|m_{\hat{n}, \hat{\lambda}}(X) - Y_{\beta_N}|^2\} \\ & \quad + (1 + \epsilon)\mathbf{E}\left(\mathbf{E}\{|m_{\hat{n}, \hat{\lambda}}(X) - Y_{\beta_N}|^2 | \mathcal{D}_n\} - \frac{1}{n}\sum_{i=1}^n |m_{\hat{n}, \hat{\lambda}}(X_i) - Y_{i, \beta_N}|^2\right) \end{aligned}$$

$$
\begin{aligned}
&+(1+\epsilon)\mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n}|m_{\hat{n},\hat{\lambda}}(X_i)-Y_{i,\beta_N}|^2-(1+\epsilon)\cdot\frac{1}{n}\sum_{i=1}^{n}|\tilde{m}_{\hat{n},\hat{\lambda}}(X_i)-Y_i|^2\right)\\
&+(1+\epsilon)^2\mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n}|\tilde{m}_{\hat{n},\hat{\lambda}}(X_i)-Y_i|^2-\frac{1}{n}\sum_{i=1}^{n}|g_\epsilon(X_i)-Y_i|^2\right)\\
&+(1+\epsilon)^2\left(\mathbf{E}\{|g_\epsilon(X)-Y|^2\}-\mathbf{E}\{|m(X)-Y|^2\}\right)\\
&+((1+\epsilon)^2-1)\mathbf{E}\{|m(X)-Y|^2\}\\
\stackrel{\text{def}}{=}&\sum_{j=1}^{6}T_{j,N},
\end{aligned}
$$

where $Y_{\beta_N}=T_{\beta_N}Y$ and $Y_{i,\beta_N}=T_{\beta_N}Y_i$.

It suffices to show

$$
\limsup_{N\to\infty}\sup_{(X,Y)\in\mathcal{D}}T_{j,N}\leq c_9\cdot\epsilon \tag{15}
$$

($j\in\{1,2,\ldots,6\}$) for some constant $c_9\in\mathbb{R}_+$ not depending on $\epsilon$.

*In the third step of the proof* we show (15) for $j=1$. Because of

$$
(a+b)^2\leq(1+\epsilon)a^2+(1+\frac{1}{\epsilon})b^2\quad(a,b>0),
$$

assumption (A3′) and condition (10) we have

$$
T_{1,N}\leq(1+\frac{1}{\epsilon})\cdot\mathbf{E}\{|Y_{\beta_N}-Y|^2\}\leq(1+\frac{1}{\epsilon})\cdot\frac{\mathbf{E}\{|Y_{\beta_N}-Y|^8\}}{\beta_N^6}\leq(1+\frac{1}{\epsilon})\cdot\frac{c_4}{\beta_N^6}\to 0
$$

$(N\to\infty)$.

*In the fourth step of the proof* we show (15) for $j=2$. Let $A_n$ be the event that

$$
\frac{1}{\hat{n}}\sum_{i=1}^{\hat{n}}|Y_i|^2\leq 2+c_4.
$$

Using $\mathbf{E}Y^2\leq 1+\mathbf{E}\{|Y|^8\}\leq 1+c_4$, the union bound and the inequality of Markov we get

$$
\begin{aligned}
\mathbf{P}(A_n^c)&=\mathbf{P}\left\{\frac{1}{\hat{n}}\sum_{i=1}^{\hat{n}}|Y_i|^2>2+c_4\right\}\\
&\leq\mathbf{P}\left\{\frac{1}{\hat{n}}\sum_{i=1}^{\hat{n}}|Y_i|^2-\mathbf{E}Y^2>1\right\}\\
&\leq(n_{max}(N)-n_{min}(N))\cdot\max_{n\in\{n_{min}(N),\ldots,n_{max}(N)\}}\mathbf{P}\left\{\frac{1}{n}\sum_{i=1}^{n}|Y_i|^2-\mathbf{E}Y^2>1\right\}\\
&\leq(n_{max}(N)-n_{min}(N))\cdot\max_{n\in\{n_{min}(N),\ldots,n_{max}(N)\}}\frac{\mathbf{E}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}(|Y_i|^2-\mathbf{E}Y^2)\right|^4\right\}}{1^4}
\end{aligned}
$$

$$
\begin{aligned}
&\leq (n_{max}(N) - n_{min}(N)) \cdot \\
&\quad \max_{n \in \{n_{min}(N),\dots,n_{max}(N)\}} \frac{n\mathbf{E}\left\{(|Y|^2 - \mathbf{E}Y^2)^4\right\} + n^2\left(\mathbf{E}\left\{(|Y|^2-\mathbf{E}Y^2)^2\right\}\right)^2}{n^4}.
\end{aligned}
$$

Using

$$
\begin{aligned}
\mathbf{E}\left\{(|Y|^2 - \mathbf{E}Y^2)^4\right\} &\leq \mathbf{E}\{|Y|^8\} + 6\cdot\mathbf{E}\{|Y|^4\}\cdot(\mathbf{E}\{|Y|^2\})^2 + (\mathbf{E}\{|Y|^2\})^4 \\
&\leq c_4 + 6\cdot(1+c_4)\cdot(1+c_4) + c_4 \leq 8\cdot(1+c_4)^2
\end{aligned}
$$

and

$$
\left(\mathbf{E}\left\{(|Y|^2 - \mathbf{E}Y^2)^2\right\}\right)^2 \leq (\mathbf{E}\{|Y|^4\})^2 \leq c_4
$$

we get

$$
\begin{aligned}
\mathbf{P}(A_n^c) &\leq (n_{max}(N) - n_{min}(N)) \cdot \max_{n \in \{n_{min}(N),\dots,n_{max}(N)\}} \frac{n\cdot 8\cdot(1+c_4)^2 + n^2\cdot c_4}{n^4} \\
&\leq (n_{max}(N) - n_{min}(N)) \cdot \frac{9\cdot(1+c_4)^2}{n_{min}(N)^2}.
\end{aligned}
$$

On $A_n$ we have

$$
\begin{aligned}
\hat{\lambda}\cdot J_k^2(\tilde{m}_{\hat{n},\hat{\lambda}}) &\leq \frac{1}{\hat{n}}\sum_{i=1}^{\hat{n}} |Y_i - \tilde{m}_{\hat{n},\hat{\lambda}}(X_i)|^2 + \hat{\lambda}\cdot J_k^2(\tilde{m}_{\hat{n},\hat{\lambda}}) \\
&\leq \frac{1}{\hat{n}}\sum_{i=1}^{\hat{n}} |Y_i - 0|^2 + \hat{\lambda}\cdot J_k^2(0) \\
&\leq 2 + c_4,
\end{aligned}
$$

hence

$$
J_k^2(\tilde{m}_{\hat{n},\hat{\lambda}}) \leq \frac{2+c_4}{\lambda_{min}(N)},
$$

which implies

$$
m_{\hat{n},\hat{\lambda}} \in \mathcal{F}_N = \left\{ T_{\beta_N} f \;:\; f \in W^k(\mathbb{R}^d) \text{ and } J_k^2(f) \leq \frac{2+c_4}{\lambda_{min}(N)} \right\}.
$$

Let $\mathcal{G}_N = \left\{ g : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R} \;:\; g(x,y) = |f(x) - T_{\beta_N} y|^2 \text{ for some } f \in \mathcal{F}_N \right\}$. Using

$$
\begin{aligned}
&\mathbf{E}\{|m_{\hat{n},\hat{\lambda}}(X) - Y_{\beta_n}|^2 | \mathcal{D}_n\} - \frac{1}{\hat{n}}\sum_{i=1}^{\hat{n}} |m_{\hat{n},\hat{\lambda}}(X_i) - Y_{i,\beta_n}|^2 \\
&\leq \left( \mathbf{E}\{|m_{\hat{n},\hat{\lambda}}(X) - Y_{\beta_n}|^2 | \mathcal{D}_n\} - \frac{1}{\hat{n}}\sum_{i=1}^{\hat{n}} |m_{\hat{n},\hat{\lambda}}(X_i) - Y_{i,\beta_n}|^2 \right) \cdot I_{A_n} + 4\beta_n^2 \cdot I_{A_n^c}
\end{aligned}
$$

we get

$$
\begin{aligned}
\frac{1}{1+\epsilon} T_{2,N} \quad &\leq \quad \mathbf{E}\left\{\sup_{g \in \mathcal{G}_N} \left| \mathbf{E}\{g(X,Y)\} - \frac{1}{\hat{n}} \sum_{i=1}^{\hat{n}} g(X_i, Y_i) \right| \right\} + 4\beta_n^2 \cdot \mathbf{P}\{A_n^c\} \\
&\leq \quad \mathbf{E}\left\{\sup_{g \in \mathcal{G}_N} \left| \mathbf{E}\{g(X,Y)\} - \frac{1}{\hat{n}} \sum_{i=1}^{\hat{n}} g(X_i, Y_i) \right| \right\} \\
&\qquad\qquad +4\beta_n^2 \cdot (n_{max}(N) - n_{min}(N)) \cdot \frac{9 \cdot (1+c_4)^2}{n_{min}(N)^2}.
\end{aligned}
$$

By using Theorem 9.1 and Lemma 20.6 in Györfi et al. (2002) we get as in Kohler and Krzyżak (2001) for $\delta > 0$ and $t \geq \delta > 0$ arbitrary

$$
\begin{aligned}
&\mathbf{P}\left\{\sup_{g \in \mathcal{G}_N} \left| \mathbf{E}\{g(X,Y)\} - \frac{1}{\hat{n}} \sum_{i=1}^{\hat{n}} g(X_i, Y_i) \right| > t \right\} \\
&\leq n_{max}(N) \cdot \max_{n \geq n_{min}(N)} \mathbf{P}\left\{\sup_{g \in \mathcal{G}_N} \left| \mathbf{E}\{g(X,Y)\} - \frac{1}{n} \sum_{i=1}^{n} g(X_i, Y_i) \right| > t \right\} \\
&\leq n_{max}(N) \cdot 8 \left( \frac{c_{10}\beta_N \cdot n_{max}(N)}{\frac{t}{32\beta_n}} \right)^{c_{11}\left(\sqrt{\frac{2+c_4}{\lambda_{min}(N)}} \cdot \frac{32\beta_N}{t}\right)^{\frac{d}{k}} + c_{12}} \exp\left( - \frac{n_{min}(N) \cdot t^2}{128(4\beta_N^2)^2} \right) \\
&\leq \left( \frac{c_{13} \cdot n_{max}(N)^2 \cdot \beta_N^2}{t} \right)^{c_{14}\left(\frac{\beta_N}{\sqrt{\lambda_{min}(N) \cdot t}}\right)^{\frac{d}{k}}} \exp\left( -c_{15} \cdot \frac{n_{min}(N) \cdot t^2}{\beta_N^4} \right) \\
&\leq \exp\left( -\frac{c_{15}}{2} \cdot \frac{n_{min}(N) \cdot t^2}{\beta_N^4} \right)
\end{aligned}
$$

for $N$ sufficiently large, since

$$
\frac{\left(\frac{\beta_N}{\sqrt{\lambda_{min}(N) \cdot t}}\right)^{\frac{d}{k}} \cdot \log\left(\beta_N^2 \cdot n_{max}(N)^2 / t\right)}{n_{min}(N) \cdot t^2 / \beta_N^4} \leq \frac{\beta_N^{4+\frac{d}{k}} \cdot \log\left(\beta_N^2 \cdot n_{max}(N)^2 / \delta\right)}{\delta^{2+d/k} \cdot n_{min}(N) \cdot \lambda_{min}(N)^{d/(2k)}} \to 0 \quad (N \to \infty)
$$

by (12).

Hence, for any $\delta > 0$ we get for $N$ sufficiently large

$$
\begin{aligned}
&\mathbf{E}\left\{\sup_{g \in \mathcal{G}_N} \left| \mathbf{E}\{g(X,Y)\} - \frac{1}{\hat{n}} \sum_{i=1}^{\hat{n}} g(X_i, Y_i) \right| \right\} \\
&\leq \delta + \int_\delta^\infty \mathbf{P}\left\{\sup_{g \in \mathcal{G}_N} \left| \mathbf{E}\{g(X,Y)\} - \frac{1}{\hat{n}} \sum_{i=1}^{\hat{n}} g(X_i, Y_i) \right| > t \right\} dt \\
&\leq \delta + \int_\delta^\infty \exp\left( -\frac{c_{15}}{2} \cdot \frac{n_{min}(N) \cdot t \cdot \delta}{\beta_N^4} \right) dt \\
&\leq \delta + \frac{2\beta_N^4}{c_{15} \cdot \delta \cdot n_{min}(N)} \cdot \exp\left( -\frac{c_{15}}{2} \cdot \frac{n_{min}(N) \cdot \delta^2}{\beta_N^4} \right) \to \delta \quad (N \to \infty)
\end{aligned}
$$

since $\beta_N^4/n_{min}(N) \to 0$ $(N \to \infty)$, and with $\delta \to 0$ we get (15) for $j = 2$.

*In the fifth step of the proof* we show (15) for $j = 3$. Using

$$|T_{\beta_N} z - y| \le |z - y| \quad \text{for } |y| \le \beta_N, z \in \mathbb{R}$$

and proceeding otherwise as in the third step of the proof we see that

$$\begin{aligned}
& T_{3,n} \\
& \le (1+\epsilon)\mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n} |\tilde{m}_{\hat{n},\hat{\lambda}}(X_i) - Y_{i,\beta_N}|^2 - (1+\epsilon)\cdot\frac{1}{n}\sum_{i=1}^{n} |\tilde{m}_{\hat{n},\hat{\lambda}}(X_i) - Y_i|^2\right) \\
& \le (1+\epsilon)\cdot\left(1+\frac{1}{\epsilon}\right)\cdot\mathbf{E}\{|Y_{\beta_N} - Y|^2\} \\
& \le (1+\epsilon)\cdot\left(1+\frac{1}{\epsilon}\right)\cdot\frac{1+c_4}{\beta_N^6} \to 0 \quad (N \to \infty).
\end{aligned}$$

*In the sixth step of the proof* we show (15) for $j = 4$. By the definition of the estimate we get

$$\begin{aligned}
& \frac{1}{(1+\epsilon)^2}\cdot T_{4,n} \\
& \le \mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n} |\tilde{m}_{\hat{n},\hat{\lambda}}(X_i) - Y_i|^2 + \hat{\lambda}\cdot J_k^2(\tilde{m}_{\hat{n},\hat{\lambda}}) - \frac{1}{n}\sum_{i=1}^{n} |g_\epsilon(X_i) - Y_i|^2 - \hat{\lambda}\cdot J_k^2(g_\epsilon)\right) \\
& \quad + \hat{\lambda}\cdot J_k^2(g_\epsilon) \\
& \le \lambda_{max}(N)\cdot J_k^2(g_\epsilon) \le \lambda_{max}(N)\cdot c_6\cdot c_5^2 \to 0 \quad (N \to \infty).
\end{aligned}$$

*In the seventh step of the proof* we show (15) for $j = 5$. Here the assertion follows from

$$\begin{aligned}
\frac{1}{(1+\epsilon)^2} T_{5,n} &= \left(\mathbf{E}\{|g_\epsilon(X) - Y|^2\} - \mathbf{E}\{|m(X) - Y|^2\}\right) \\
&= \int |g_\epsilon(x) - m(x)|^2 \mathbf{P}_X(dx) < \epsilon.
\end{aligned}$$

*In the eighth (and final) step of the proof* we finish the proof by showing (15) for $j = 6$. We have

$$\sup_{(X,Y)\in\mathcal{D}} T_{6,n} \le \epsilon\cdot(2+\epsilon)\cdot \sup_{(X,Y)\in\mathcal{D}} 4\cdot\mathbf{E}\{|Y|^2\} \le \epsilon\cdot(2+\epsilon)\cdot 4\cdot(1+c_4).$$

□

**Proof of Theorem 1.** Since

$$\int \|\hat{m}_N(t,x) - m(t,x)\|_2^2 \, \mu_t(dx)$$

is the sum of the $L_2$ errors of the $d$ components of $\hat{m}_n(t,x)$, and since each of these components is a penalized least squares estimate applied to a scalar response variable, it suffices to prove Theorem 1 in case of a scalar response variable, which we will do in the sequel.

For $k \in \{0, \dots, N\}$ let

$$\mathcal{D}_X^{(k)} = \left\{ X_i^{(t_l)}, Y_i^{(t_l)} \,:\, i \in \{1, \dots, n_l\}, l \in \{0, \dots, k-1\} \right\} \cup \left\{ X_j^{(t_k)} \,:\, j \in \{1, \dots, n_k\} \right\}.$$

Set

$$Y_k = m(t_k, \cdot) + \epsilon_k$$

where

$$\epsilon_k = m_{n_{t_k}}^{(t_k)}(\cdot) - \mathbf{E}\left\{ m_{n_{t_k}}^{(t_k)}(\cdot) | \mathcal{D}_X^{(k)} \right\},$$

and

$$\bar{Y}_k = Y_k + \mathbf{E}\left\{ m_{n_{t_k}}^{(t_k)}(\cdot) | \mathcal{D}_X^{(k)} \right\} - m(t_k, \cdot) = m_{n_{t_k}}^{(t_k)}(\cdot)$$

$(k = 0, \dots, N)$. Then we have for $l, k \in \{0, \dots, N\}, l < k$

$$\mathbf{E}\{\epsilon_l \cdot \epsilon_k\} = \mathbf{E}\{\mathbf{E}\{\epsilon_l \cdot \epsilon_k | \mathcal{D}_X^{(k)}\}\} = \mathbf{E}\{\epsilon_l \cdot \mathbf{E}\{\epsilon_k | \mathcal{D}_X^{(k)}\}\} = \mathbf{E}\{\epsilon_l \cdot 0\} = 0.$$

Application of Lemma 1 yields

$$\begin{aligned}
&\mathbf{E} \int |\hat{m}_N(t,x) - m(t,x)|^2 \mu_t(dx) \\
&\leq 3 \cdot \max_{k:|t-t_k| \leq a_2 \cdot h_N} \mathbf{E}\{\|\bar{Y}_k - Y_k\|^2_{L_2(\mu_t)}\} + 3 \cdot \frac{c_2}{c_1 \cdot a_1} \cdot \max_{k:|t-t_k| \leq a_2 \cdot h_N} \|\mathbf{E}\{\epsilon_k^2\}\|_{L_1(\mu_t)} \cdot \frac{1}{N \cdot h_N} \\
&\quad + 3 \cdot \frac{c_2}{c_1 \cdot a_1} \cdot \max_{k:|t-t_k| \leq a_2 \cdot h_N} \|m(t_k, \cdot) - m(t, \cdot)\|^2_{L_2(\mu_t)} \\
&\leq 3 \cdot \max\{1, \frac{c_2}{c_1 \cdot a_1} \cdot \frac{1}{N \cdot h_N}\} \cdot \max_{k:|t-t_k| \leq a_2 \cdot h_N} \mathbf{E} \int \left| m_{n_{t_k}}^{(t_k)}(x) - m(t_k, x) \right|^2 \mu_t(dx) \\
&\quad + 3 \cdot \frac{c_2}{c_1 \cdot a_1} \cdot \max_{k:|t-t_k| \leq a_2 \cdot h_N} \|m(t_k, \cdot) - m(t, \cdot)\|^2_{L_2(\mu_t)},
\end{aligned}$$

where the last inequality follows from

$$\begin{aligned}
&\mathbf{E} \int \left| m_{n_{t_k}}^{(t_k)}(x) - m(t_k, x) \right|^2 \mu_t(dx) \\
&= \int \mathbf{E} \left\{ \left| m_{n_{t_k}}^{(t_k)}(x) - m(t_k, x) \right|^2 \right\} \mu_t(dx) \\
&= \int \left( \mathbf{E} \left\{ \left| m_{n_{t_k}}^{(t_k)}(x) - \mathbf{E}\left\{ m_{n_{t_k}}^{(t_k)}(x) | \mathcal{D}_X^{(k)} \right\} \right|^2 \right\} + \mathbf{E} \left\{ \left| \mathbf{E}\left\{ m_{n_{t_k}}^{(t_k)}(x) | \mathcal{D}_X^{(k)} \right\} - m(t_k, x) \right|^2 \right\} \right) \mu_t(dx) \\
&= \|\mathbf{E}\{\epsilon_k^2\}\|_{L_1(\mu_t)} + \|\mathbf{E}\{|\bar{Y}_k - Y_k|^2\}\|_{L_1(\mu_t)}.
\end{aligned}$$

It suffices to show

$$\max_{k:|t-t_k|\le a_2\cdot h_N} \mathbf{E}\int \left|m_{n_{t_k}}^{(t_k)}(x)-m(t_k,x)\right|^2 \mu_t(dx) \to 0 \quad (N\to\infty) \tag{16}$$

and

$$\max_{k:|t-t_k|\le a_2\cdot h_N} \|m(t_k,\cdot)-m(t,\cdot)\|_{L_2(\mu_t)}^2 \to 0 \quad (N\to\infty). \tag{17}$$

(17) follows from the (uniform) continuity of $m$ on the compact set $[0,1]\times[0,1]^d$ and condition (14).

If $\nu_1,\nu_2$ are measures on $\mathbb{R}^d$ and $f$ is a function on $\mathbb{R}^d$ bounded in absolute value by $4\cdot\beta_N^2$, then

$$\left|\int_{\mathbb{R}^d} f(x)\,\nu_1(dx)-\int_{\mathbb{R}^d} f(x)\,\nu_2(dx)\right| \le 4\cdot\beta_N^2\cdot \sup_{A\in\mathcal{B}_d}|\nu_1(A)-\nu_2(A)|.$$

Hence because of (A5) it suffices to show

$$\max_{k:|t-t_k|\le a_2\cdot h_N} \mathbf{E}\int \left|m_{n_{t_k}}^{(t_k)}(x)-m(t_k,x)\right|^2 \mu_{t_k}(dx) \quad (N\to\infty), \tag{18}$$

which we do in the sequel by using Lemma 2.

Let $\mathcal{D}$ be the class of all distributions of $(X_s,Y_s)$ for some

$$s\in\{l/N\,:\,l\in\{0,\dots,N\},N\in\mathbb{N}\}.$$

For $s=l/N$ set $\hat{n}(N)=\hat{n}=n_{t_l}=n_{l/N}$ and $\hat{\lambda}=\hat{\lambda}(N)=\lambda_{t_l}=\lambda_{l/N}$. By the assumptions of Theorem 1 (in particular by the uniform continuity of $m$ on $[0,1]\times[0,1]^d$) it is easy to see that the assumptions of Lemma 2 are satisfied, from which we can conclude

$$\begin{aligned}&\max_{k:|t-t_k|\le a_2\cdot h_N} \mathbf{E}\int \left|m_{n_{t_k}}^{(t_k)}(x)-m(t_k,x)\right|^2 \mu_{t_k}(dx)\\ &\le \sup_{(X,Y)\in\mathcal{D}} \mathbf{E}\int |m_{\hat{n},\hat{\lambda}}(x)-m(x)|^2\mathbf{P}_X(dx)\to 0 \quad (N\to\infty).\end{aligned}$$

The proof is complete. $\square$

## References


[1] De Boor, C. (1978). *A Practical Guide to Splines.* Springer, New York.

[2] Duchon, J. (1976). Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *R.A.I.R.O. Analyse Numérique* **10**, 5–12.

[3] Guala, M., Liberzon, A., Hoyer, K., Tsinober, A. and Kinzelbach, W. (2009). Experimental study of clustering of large particles in homogeneous turbulent flow. *Journal of Turbulence* **9**, 1-20.

[4] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A distribution-free theory of nonparametric regression.* Springer Series in Statistics, Springer, 2002.

[5] Kinzel, M. (2010). Experimental Investigation of Turbulence under the Influence of Confinement and Rotation. PhD thesis, Department of Mechanical Engineering, Technische Universität Darmstadt.

[6] Kohler, M. and Krzyżak, A. (2001). Nonparametric regression estimation using penalized least squares. *IEEE Transactions on Information Theory* **47**, 3054–3058.

[7] Kunnen, R. P. J., Geurts, B. J. and Clercx, H. J. H. (2010). Experminental and numerical investigation of turbulent convection in a rotating cylinder. *Journal of Fluid Mechanics* **642**, 445-476.

[8] Lüthi, B., Tsinober, A. and Kinzelbach, W. (2005). Lagrangian measurement of vorticity dynamics in turbulent flow. *Journal of Fluid Mechanics* **528**, 87-118.

[9] Messio, L., Morize, C., Rabaud, M. and Moisy, F. (2008) Experimental observation using particle image velocimetry of inertial waves in a rotating fluid. *Experiments in Fluids* **44**, 519-528.

[10] Raffel, M., Willert, C., and Kompenhans, J. (1998) *Particle Image Velocimetry: A Practical Guide.* Springer-Verlag, Heidelberg.

[11] Speetjens, M. F. M., Clercx, H. J. H. and Van Heijst, G. J. F. (2004). A numerical and experimental study on advection in three-dimensional Stokes flows. *Journal of Fluid Mechanics* **514**, 77-105.

[12] Wahba, G. (1990). *Spline Models for Observational Data.* SIAM, Philadelphia.