Universität
Stuttgart

Fachbereich
Mathematik

On Convergence of Local Averaging Regression Function Estimates for the Regularization of Inverse Problems

Barbara Kaltenbacher, Harro Walk

Preprint 2010/005

Universität Stuttgart
Fachbereich Mathematik

On Convergence of Local Averaging Regression Function Estimates for the Regularization of Inverse Problems

Barbara Kaltenbacher, Harro Walk

Preprint 2010/005

Fachbereich Mathematik
Fakultät Mathematik und Physik
Universität Stuttgart
Pfaffenwaldring 57
D-70 569 Stuttgart

E-Mail: preprints@mathematik.uni-stuttgart.de
WWW: http://www.mathematik/uni-stuttgart.de/preprints

ISSN **1613-8309**

Abstract

In this paper we combine the ideas of regression function estimation on one hand and regularization by discretization on the other hand to a regularization method for linear ill-posed problems with additive stochastic noise. A general convergence result is provided and its assumptions are verified for the partitioning estimators and to some extent for kernel estimators. As an example of an inverse problem we consider Volterra integral equations of the first kind.

1 Introduction

Consider the inverse problem

$$\text{Find } f^\dagger \text{ such that } Y_i = (Kf^\dagger)(X_i) + \epsilon_i \tag{1}$$

with $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ independent identically distributed $(d+1)$-dimensional random vectors, the real random variables Y_i and ϵ_i being square integrable, $E\{\epsilon_i \,|\, X_i) = 0$, and

$$K : H_1 \to H_2$$

a linear bounded operator on appropriate function spaces, which is not boundedly invertible, so that the inverse problem is ill-posed. This is equivalent to

$$\text{Find } f^\dagger \text{ such that } Kf^\dagger = y, \text{ where } y(x) = E\{Y \,|\, X = x)$$

Local averaging nonparametric regression function estimates for y such as partitioning, nearest neighbour or kernel estimates have the form

$$y_n(x) = \sum_{i=1}^{n} W_{n,i}^h(x; X_1, \ldots, X_n) Y_i \,, \tag{2}$$

with appropriate weights $W_{n,i}^h(x; X_1, \ldots, X_n)$. On the other hand, regularization by discretization in image space (or self-regularization) of the deterministic inverse problem

$$\text{Find } f^\dagger \text{ such that } Kf^\dagger = y$$

is based on solving the discretized operator equation

$$Q^h K f = Q^h y \tag{3}$$

in a best approximate sense, i.e.,

$$f^h = (Q^h K)^\dagger Q^h y$$

where a superscript † denotes the generalized inverse. The idea of this paper is to use the operator $Q_n^h = Q_n^h(X_1, \ldots, X_n)$ defined by local averaging regression (2)

$$(Q_n^h g)(x) = \sum_{i=1}^{n} W_{n,i}^h(x; X_1, \ldots, X_n) g(X_i) \,, \tag{4}$$

as a discretization operator in (3), i.e., to apply regularization by discretization in a stochastic setting. Note that as opposed to, e.g., [4], the discretization operator itself is of stochastic nature, due to its dependence on the random variables $X_1, \ldots, X_n$ appearing in the weights. Using a regression estimate $\hat{y}_n^h \in Y_n^h$, e.g., $\hat{y}_n^h = y_n$ according to (2), we arrive at an inversion method of the form

$$\hat{f}_n^h = (Q_n^h K)^\dagger \hat{y}_n^h \,.$$

The aim of this paper is to show that under certain assumptions and with an appropriate choice $h = h(n)$, this is indeed a regularization method in the sense that

$$E\{\left\|\hat{f}_n^{h(n)} - f^\dagger\right\|_{H_1}^2\} \to 0 \text{ as } n \to \infty\,.$$

Here h is an abstract discretization parameter, which, e.g., can be interpreted as the bandwidth of a kernel estimator or is used to parametrize the number l_h of partition elements for a partitioning estimator. In analogy to the deterministic case, where the regularization parameter has to be chosen in dependence of the noise level, we will here select h as a function of n, where, sloppily speaking, large n corresponds to small noise level. Note however, that even without an ill-posed inverse problem in the background, a dependence $h = h(n)$ is always required in regression function estimation, as opposed to the deterministic setting. We will here see that for regression estimation in inverse problem, this choice additionally has to take into account the degree of ill-posedness.

For a convergence analysis of classical methods such as Tikhonov regularization, spectral cutoff and more general spectral theory based regularization methods for linear inverse problems with random noise, we refer to the existing literature (see, e.g. [**?**, **?**, **?**, 4, **?**, **?**] and the references therein). The aim of this paper is to consider possible regularizing effects for a class of methods for nonparametric regression estimation – so originally for well-posed problems – and to contribute to the analysis of such methods when applied in an appropriate manner to linear ill-posed operator equations.

2 Convergence analysis

2.1 A general convergence result

Theorem 1. *Let $K : H_1 \to H_2$ be a bounded linear operator between the Hilbert spaces H_1 and H_2, let $f^\dagger$ solve $Kf = y$ and let for all $n \in \mathbb{N}, h > 0$*

- *$\hat{y}_n^h \in Y_n^h \subseteq H_3$ be a regression estimate;*
- *$Q_n^h : H_2 \to H_3$ be a bounded linear operator;*
- *$\tilde{H}_1$, H_2 be possibly stochastically defined, (i.e., $\tilde{H}_1 = \tilde{H}_{n,1}^h$ $H_2 = H_{n,2}^h$) Hilbert spaces;*

and let the following assumptions hold:

$$\forall n \in \mathbb{N}, h > 0\ \exists \beta_n^h > 0 :\ E\{\left\|\hat{y}_n^h - Q_n^h y\right\|_{H_3}^2\} \le \beta_n^h \tag{5}$$

$$\forall n \in \mathbb{N}, h > 0\ \exists C_n^h > 0 :\ P\{\left\|(Q_n^h K)^\dagger|_{\tilde{Y}_n^h}\right\|_{H_3 \to \tilde{H}_1} \le C_n^h\} = 1 \tag{6}$$

$$\begin{aligned}&\forall (f_n^h)_{n\in\mathbb{N}, h>0}\,,\ f_n^h \in (Q_n^h K)^\dagger \tilde{Y}_n^h\ :\\ &\qquad \limsup_{\substack{h\to 0\\ n\to\infty}} E\{\left\|f_n^h\right\|_{\tilde{H}_1}^2\} \overset{\substack{h\to 0\\ n\to\infty}}{\longrightarrow} 0\ \Rightarrow\ \limsup_{\substack{h\to 0\\ n\to\infty}} E\{\left\|f_n^h\right\|_{H_1}^2\} \overset{\substack{h\to 0\\ n\to\infty}}{\longrightarrow} 0\end{aligned} \tag{7}$$

where $\tilde{Y}_n^h = Y_n^h + \mathcal{R}(Q_n^h)$. Moreover, define

$$\hat{f}_n^h = (Q_n^h K)^\dagger \hat{y}_n^h\,,$$

and let $h(n)$ be chosen such that

$$C_n^{h(n)} \beta_n^{h(n)} \to 0\,, \tag{8}$$

and

$$E\{\left\|(\mathrm{id} - Proj^{H_1}_{\mathcal{N}(Q_n^{h(n)}K)^\perp}) f^\dagger\right\|_{H_1}^2\} \overset{n\to\infty}{\longrightarrow} 0\,. \tag{9}$$

Then

$$E\{\left\|\hat{f}_n^{h(n)} - f^\dagger\right\|_{H_1}^2\} \to 0 \textit{ as } n \to \infty\,.$$

Proof. The assertion follows immediately from the error decomposition

$$\begin{aligned}\hat{f}_n^{h(n)} - f^\dagger &= (Q_n^{h(n)}K)^\dagger Q_n^{h(n)}Kf^\dagger - f^\dagger + (Q_n^{h(n)}K)^\dagger(\hat{y}_n^{h(n)} - Q_n^{h(n)}y) \\ &= -(\mathrm{id} - \mathrm{Proj}^{H_1}_{\mathcal{N}(Q_n^{h(n)}K)^\perp})f^\dagger + (Q_n^{h(n)}K)^\dagger(\hat{y}_n^{h(n)} - Q_n^{h(n)}y)\end{aligned}$$

by applying (9) to the first term on the right hand side and (5), (6), (8) as well as (7) to the second term on the right hand side. □

2.2 The approximation error

We first of all derive a sufficient condition for

$$E\{\left\|(\mathrm{id} - \mathrm{Proj}^{H_1}_{\mathcal{N}(Q_n^hK)^\perp})f^\dagger\right\|^2_{H_1}\} \overset{h\to 0, n\to\infty}{\longrightarrow} 0 \tag{10}$$

via pointwise convergence of the adjoint of Q_n^h, without any assumption on the distribution of X.

If $H_2 \subseteq H_3$, then for arbitrary $\varepsilon > 0$ we have by $f^\dagger \in \mathcal{N}(K)^\perp = \overline{\mathcal{R}(K^*)}$ existence of a $v^\varepsilon \in H_2 = H_2 \cap H_3$ such that

$$\|f^\dagger - K^*v^\varepsilon\|_{H_1} < \varepsilon\,. \tag{11}$$

Therewith, due to $\|I - \mathrm{Proj}^{H_1}_{\mathcal{N}(Q_n^hK)^\perp}\|_{H_1\to H_1} = 1$ we get

$$\begin{aligned}&\|(I - \mathrm{Proj}^{H_1}_{\mathcal{N}(Q_n^hK)^\perp})f^\dagger\|_{H_1} \\ &\quad\leq \varepsilon + \|(I - \mathrm{Proj}^{H_1}_{\mathcal{N}(Q_n^hK)^\perp})K^*v^\varepsilon\|_{H_1} \\ &\quad= \varepsilon + \inf\{\|K^*v^\varepsilon - z\|_{H_1} \mid z \in \mathcal{N}(Q_n^hK)^\perp = \overline{\mathcal{R}(K^*Q_n^{h*})}\} \\ &\quad\leq \varepsilon + \inf\{\|K^*v^\varepsilon - z\|_{H_1} \mid z \in \mathcal{R}(K^*Q_n^{h*})\} \\ &\quad= \varepsilon + \inf\{\|K^*v^\varepsilon - K^*w\|_{H_1} \mid w \in \mathcal{R}(Q_n^{h*})\} \\ &\quad\leq \varepsilon + \|K^*\|_{H_2\to H_1}\inf\{\|v^\varepsilon - w\|_{H_2} \mid w \in \mathcal{R}(Q_n^{h*})\} \\ &\quad\leq \varepsilon + \|K\|_{H_1\to H_2}\|v^\varepsilon - Q_n^{h*}v^\varepsilon\|_{H_2}\,,\end{aligned}$$

which by the presumed boundedness of K as an operator from H_1 to H_2, yields convergence of the approximation error (10) provided that

$$\forall v \in H_2 \cap H_3: \quad E\{\|v - Q_n^{h*}v\|^2_{H_2}\} \to 0 \text{ as } h \to 0\ \ n \to \infty\,. \tag{12}$$

More generally, if there exists a Banach space $B \subseteq H_2 \cap H_3$ with B dense in H_2, then a similar argument shows that (12) is sufficient for (10). Namely, then for arbitrary $\varepsilon > 0$ we have by $f^\dagger \in \mathcal{N}(K)^\perp = \overline{\mathcal{R}(K^*)}$ and boundedness of $K^* : H_2 \to H_1$ existence of $\tilde{v}^\varepsilon \in H_2$ $v^\varepsilon \in B \subseteq H_2 \cap H_3$, such that $\|f^\dagger - K^*\tilde{v}^\varepsilon\|_{H_1} < \frac{\varepsilon}{2}$ and $\|\tilde{v}^\varepsilon - v^\varepsilon\|_{H_2} < \frac{\varepsilon}{2\|K^*\|_{H_2\to H_1}}$, hence (11).

We therefore state an auxiliary result for the adjoint of Q_n^h:

Lemma 1. *Let the spaces H_2, H_3 be defined by*

$$H_2 = L^2_{\mu''}\,, \quad H_3 = L^2_\mu \tag{13}$$

with μ the distribution of X_i and μ'' the empirical measure corresponding to $X_1, \ldots, X_n$, namely such that

$$\frac{1}{n}\sum_{i=1}^n f(X_i) = \int f(s)\mu''(ds)\,.$$

Moreover, let Q_n^h be given by (4) with the weights satisfying

$$W^h_{n,i}(x; x_1, \ldots, x_n) = F_n^h(x, x_i, \vec{N}_n^h(x, x_1, \ldots x_n)) \tag{14}$$

with

$$F_n^h : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{r^h} \to \mathbb{R}\,, \quad \vec{N}_n^h : (\mathbb{R}^d)^{n+1} \to \mathbb{R}^{r^h}\,.$$

Then

$$Q_n^{h\,*} : H_3 \to H_2\,,$$

$$(Q_n^{h\,*} z)(s) = n \int F_n^h(x, s, \vec{N}_n^h(x, X_1, \dots X_n)) z(x)\, \mu(dx)\,. \tag{15}$$

Proof. For any $w \in H_2$, $z \in H_3$, we have

$$\begin{aligned}
\langle Q_n^h w, z\rangle_{H_3} &= \int \sum_{i=1}^n W_{n,i}^h(x; X_1, \dots, X_n) w(X_i) z(x)\, \mu(dx) \\
&= \int \sum_{i=1}^n F_n^h(x, X_i, \vec{N}_n^h(x, X_1, \dots, X_n)) w(X_i) z(x)\, \mu(dx) \\
&= \int n \int F_n^h(x, s, \vec{N}_n^h(x, X_1, \dots, X_n)) w(s) z(x)\, \mu''(ds)\, \mu(dx) \\
&= \langle w, Q_n^{h\,*} z\rangle_{H_2}\,.
\end{aligned}$$

□

Remark 1. *It is readily checked that the assumptions of Lemma 1 hold for the following local averaging regression function estimators:*

- *partitioning estimator:*

$$W_{n,i}^h(X; X_1, \dots, X_n) = \sum_{j=1}^{l^h} \frac{I_{\{X_i \in A_j^h\}}}{\sum_{l=1}^n I_{\{X_l \in A_j^h\}}} I_{\{X \in A_j^h\}} \tag{16}$$

where $\frac{0}{0} := 0$ *and* $(\mathcal{P}^h)_{h>0}$ *is a family of partitions of* $\mathbb{R}^d$, $\mathcal{P}^h = \{A_1^h, A_2^h, \dots\}$;

$$r^h = l^h \le \infty\,, \quad \vec{N}_{n,j}^h(x, x_1, \dots, x_n) = \sum_{m=1}^n I_{\{x_m \in A_j^h\}}\,,\ j = 1, \dots l^h\,,$$

$$F_n^h(x, s, \vec{z}) = \sum_{j=1}^{l^h} \frac{I_{\{s \in A_j^h\}}}{z_j} I_{\{x \in A_j^h\}}\,,$$

with l^h *the number of partitioning elements* A_j^h. *In the special case of a cubic partition (i.e., of the* A_j^h *being d-dimensional hypercubes) the quantity* $h = h(n)$ *is just the side length of a cube.*

- *kernel estimator*

$$W_{n,i}^h(X; X_1, \dots, X_n) = \frac{\mathcal{K}(\frac{X - X_i}{h})}{\sum_{l=1}^n \mathcal{K}(\frac{X - X_l}{h})}$$

where $\mathcal{K} : \mathbb{R}^d \to \mathbb{R}$ *is a kernel function, and* $h = h(n)$ *is the bandwidth;*

$$r^h = 1\,, \quad N_n^h(x, x_1, \dots, x_n) = \sum_{l=1}^n \mathcal{K}(\tfrac{x - x_l}{h})\,,$$

$$F_n^h(x, s, z) = \frac{\mathcal{K}(\frac{x - s}{h})}{z}\,.$$

- k^h*-nearest-neighbor estimator:*

$$W_{n,i}^h(X, X_1, \dots, X_n) = \frac{1}{k^h} I_{\{X_i \in B_{N_n^h(X,X_1,\dots,X_n)}(X)\}}$$

where $B_R(x)$ *denotes the ball of radius* R *around the point* x*, and* $h = h(n) = \frac{k^h}{n}$*;*

$$r^h = 1\,, \quad N_n^h(x, x_1, \dots, x_n) = \min_{\{m_1,\dots,m_{k^h}\}\subseteq\{1,\dots,n\}} \max_{j\in\{1,\dots k^h\}} |x_{m_j} - x|\,,$$

$$F_n^h(x, s, z) = \frac{1}{k^h} I_{\{s\in B_z(x)\}}\,,$$

usually under the assumption that equal distances, so-called ties, appear with probability zero (cf. [2], pp 86,87).

Note that in each of these cases the value of $\vec{N}_n^h$ *is invariant under permutations of the last* n *arguments:*

$$\begin{aligned} &\forall (x, x_1, \dots, x_n) \in (\mathbb{R}^d)^{n+1}\, \forall \pi : \{1,\dots,n\} \to \{1,\dots,n\} \text{ bijective } : \\ &\vec{N}_n^h(x, x_{\pi(1)}, \dots, x_{\pi(n)}) = \vec{N}_n^h(x, x_1, \dots, x_n)\,. \end{aligned} \tag{17}$$

To obtain (12), we proceed similary to the proof of Stone's theorem, see Theorem 4.1 of [2].

Lemma 2. *Let* H_2*,* H_3 *be defined as in (13). Then for any* $v \in H_2$

$$E\{\|v - Q_n^{h^*} v\|_{H_2}^2\} \le 2J_n^* + 2L_n^* \tag{18}$$

with

$$\begin{aligned} J_n^* &= E\{\frac{1}{n}\sum_{i=1}^n \left(n \int W_{n,i}^h(x; X_1, \dots, X_n)\, [v(X_i) - v(x)]\, \mu(dx)\right)^2\} \\ &= n^2 E\{\left(\int W_{n,1}^h(x; X_1, \dots, X_n)\, [v(X_1) - v(x)]\, \mu(dx)\right)^2\}\,, \end{aligned} \tag{19}$$

$$\begin{aligned} L_n^* &= E\{\frac{1}{n}\sum_{i=1}^n \left(\left[1 - n\int W_{n,i}^h(x; X_1, \dots, X_n)\, \mu(dx)\right] v(X_i)\right)^2\} \\ &= E\{v(X_1)^2 \left[1 - n\int W_{n,1}^h(x; X_1, \dots, X_n)\, \mu(dx)\right]^2\}\,. \end{aligned} \tag{20}$$

Proof.

$$\begin{aligned} &E\{\|v - Q_n^{h^*} v\|_{H_2}^2\} \\ &= E\{\int (v(s) - (Q_n^{h^*} v)(s))^2\, \mu''(ds)\} \\ &= E\{\frac{1}{n}\sum_{i=1}^n (v(X_i) - (Q_n^{h^*} v)(X_i))^2\} \\ &= E\{\frac{1}{n}\sum_{i=1}^n \left(v(X_i) - n\int F_n^h(x, X_i, \vec{N}_n^h(x, X_1, \dots X_n)) v(x)\, \mu(dx)\right)^2\} \\ &= E\{\frac{1}{n}\sum_{i=1}^n \left(v(X_i) - n\int W_{n,i}^h(x; X_1, \dots, X_n) v(x)\, \mu(dx)\right)^2\}\,. \end{aligned}$$

□

We now prove convergence to zero of the terms J_n^*, L_n^* for the partitioning estimator (16) without any assumption on the distribution of X. For this purpose, we denote by $A^h(x)$ the partition element containing the point x and make use of the notation

$$W_{n,i}^h(X;X_1,\dots,X_n) = \frac{I_{\{X_i\in A^h(X)\}}}{1+\sum_{\substack{m=1\\ m\neq i}}^n I_{\{X_m\in A^h(X)\}}}\,. \tag{21}$$

Lemma 3. *Let $(\mathcal{P}^{h(n)})_{n\in\mathbb{N}}$ be a sequence of partitions of $\mathbb{R}^d$, i.e., $\mathcal{P}^{h(n)} = \{A_1^{h(n)}, A_2^{h(n)},\dots\}$ such that for each sphere S centered at the origin*

$$\max_{j\,:\,A_j^{h(n)}\cap S\neq\emptyset} \operatorname{diam}(A_j^{h(n)}) \to 0 \text{ as } n\to\infty \tag{22}$$

and

$$\frac{|\{j\,:\,A_j^{h(n)}\cap S\neq\emptyset\}|}{n} \to 0 \text{ as } n\to\infty\,. \tag{23}$$

Then

$$J_n^*\to 0 \text{ and } L_n^*\to 0 \text{ as } n\to\infty\,.$$

For the proof we need two auxiliary results.

Lemma 4. *Let the random variable Z be binomially distributed with parameters n and p. Then*

$$E\left\{\frac{1}{(1+Z)^2}\right\} \le \min\left\{\frac{2}{(n+1)^2p^2}\,,\ \frac{1}{(n+1)^2p^2}+\frac{3}{(n+1)^3p^3}\right\}\,.$$

Proof.

$$\begin{aligned}
E\left\{\frac{1}{(1+Z)^2}\right\} &= \sum_{k=0}^n \frac{1}{(1+k)^2}\binom{n}{k}p^k(1-p)^{n-k}\\
&\le 2\sum_{k=0}^n \frac{1}{(k+1)(k+2)}\binom{n}{k}p^k(1-p)^{n-k}\\
&= 2\frac{1}{(n+1)(n+2)}\sum_{k=0}^n \binom{n+2}{k+2}p^k(1-p)^{n-k}\\
&= 2\frac{1}{(n+1)(n+2)p^2}\sum_{l=2}^n \binom{n+2}{l}p^l(1-p)^{n+2-l}\\
&\le \frac{2}{(n+1)^2p^2}
\end{aligned}$$

and by the decomposition

$$\begin{aligned}
\frac{1}{(k+1)^2} &= \frac{1}{(k+1)(k+2)}+\frac{1}{(k+1)^2(k+2)}\\
&\le \frac{1}{(k+1)(k+2)}+\frac{3}{(k+1)(k+2)(k+3)}\,,
\end{aligned}$$

similarly

$$E\left\{\frac{1}{(1+Z)^2}\right\} \le \sum_{k=0}^n \frac{1}{(k+1)(k+2)}\binom{n}{k}p^k(1-p)^{n-k}$$

$$
\begin{aligned}
&+\sum_{k=0}^{n}\frac{3}{(k+1)(k+2)(k+3)}\binom{n}{k}p^k(1-p)^{n-k}\\
= \;& \frac{1}{(n+1)(n+2)}\sum_{k=0}^{n}\binom{n+2}{k+2}p^k(1-p)^{n-k}\\
&+\frac{3}{(n+1)(n+2)(n+3)}\sum_{k=0}^{n}\binom{n+3}{k+3}p^k(1-p)^{n-k}\\
\le \;& \frac{1}{(n+1)^2p^2}+\frac{3}{(n+1)^3p^3}\,.
\end{aligned}
$$

□

Lemma 5. *Let $(\mathcal{P}^h)_{h>0}$ be a family of partitions of $\mathbb{R}^d$, denote by $A^h(x)$ the partition element containing the point x and let $W_{n,i}^h(X;X_1,\dots,X_n)$ according to (21) be the weight of the corresponding partitioning estimate.*

Then for any measurable $w:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}_+$

$$
\begin{aligned}
&E\left\{\left(\int W_{n,i}^h(x;X_1,\dots,X_n)w(x,X_1)\mu(dx)\right)^2\right\}\\
&\le \int \min\left\{\frac{2}{n^2}\,,\;\frac{1}{n^2}+\frac{3}{n^3\mu(A_j^h)}\right\}\left(\frac{\int_{A^h(t)}w(x,t)\mu(dx)}{\mu(A^h(t))}\right)^2\mu(dt)\,.
\end{aligned}
$$

Proof.

$$
\begin{aligned}
&E\left\{\left(\int W_{n,1}^h(x;X_1,\dots,X_n)w(x,X_1)\mu(dx)\right)^2\right\}\\
=\;& E\left\{\int\frac{I_{\{X_1\in A^h(x)\}}}{1+\sum_{m=2}^n I_{\{X_m\in A^h(x)\}}}w(x,X_1)\mu(dx)\right.\\
&\qquad\left.\int\frac{I_{\{X_1\in A^h(z)\}}}{1+\sum_{m=2}^n I_{\{X_m\in A^h(z)\}}}w(z,X_1)\mu(dz)\right\}\\
=\;& \int\left(\int\int I_{\{t\in A^h(x)\}}I_{\{t\in A^h(z)\}}w(x,t)w(z,t)\right.\\
&\qquad E\left\{\frac{1}{1+\sum_{m=2}^n I_{\{X_m\in A^h(x)\}}}\frac{1}{1+\sum_{m=2}^n I_{\{X_m\in A^h(z)\}}}\right\}\\
&\qquad\left.\mu(dx)\mu(dz)\right)\mu(dt)\\
\le\;& \int\left(\int\int I_{\{t\in A^h(x)\}}I_{\{t\in A^h(z)\}}w(x,t)w(z,t)\right.\\
&\qquad\sqrt{E\left\{\frac{1}{(1+\sum_{m=2}^n I_{\{X_m\in A^h(x)\}})^2}\right\}}\\
&\qquad\left.\sqrt{E\left\{\frac{1}{(1+\sum_{m=2}^n I_{\{X_m\in A^h(z)\}})^2}\right\}}\mu(dx)\mu(dz)\right)\mu(dt)\\
=\;& \int\left(\int I_{\{t\in A^h(x)\}}w(x,t)\sqrt{E\left\{\frac{1}{(1+\sum_{m=2}^n I_{\{X_m\in A^h(x)\}})^2}\right\}}\mu(dx)\right)^2\mu(dt)\\
\le\;& \int\left(\int I_{\{t\in A^h(x)\}}w(x,t)\right.
\end{aligned}
$$

$$\sqrt{\min\left\{\frac{2}{n^2\mu(A^h(x))^2}\,,\ \frac{1}{n^2\mu(A^h(x))^2}+\frac{3}{n^3\mu(A^h(x))^3}\right\}}\mu(dx)\Bigg)^2\mu(dt) \tag{24}$$

$$= \int \min\left\{\frac{2}{n^2}\,,\ \frac{1}{n^2}+\frac{3}{n^3\mu(A^h(t))}\right\}\left(\frac{\int_{A^h(t)} w(x,t)\mu(dx)}{\mu(A^h(t))}\right)^2\mu(dt)\,,$$

where we have used independence in the third line, the Cauchy-Schwarz inequality in the fourth line, and Lemma 4 together with the fact that $\sum_{m=2}^{n} I_{\{X_m\in A^h(x)\}}$ is $b(n-1,\mu(A^h(x)))$ distributed, in the sixth line. □

Proof. (Lemma 3) Let $\varepsilon>0$ be arbitrary. According to Theorem A.1 in [2] we choose a uniformly continuous compactly supported function $\tilde{v}$ such that $\int|v(x)-\tilde{v}(x)|^2\mu(dx)<\varepsilon$. Therewith,

$$\begin{aligned}
J_n^* &\le 3n^2E\{\left(\int W_{n,1}^h(x;X_1,\dots,X_n)\,[v(X_1)-\tilde{v}(X_1)]\ \mu(dx)\right)^2\}\\
&\quad +3n^2E\{\left(\int W_{n,1}^h(x;X_1,\dots,X_n)\,[\tilde{v}(X_1)-\tilde{v}(x)]\ \mu(dx)\right)^2\}\\
&\quad +3n^2E\{\left(\int W_{n,1}^h(x;X_1,\dots,X_n)\,[\tilde{v}(x)-v(x)]\ \mu(dx)\right)^2\}\\
&= 3J_{n,1}+3J_{n,2}+3J_{n,3}\,,
\end{aligned}$$

with $h=h(n)$, where due to Lemma 5

$$J_{n,1}\le 2\int|v(t)-\tilde{v}(t)|^2\left(\frac{\mu(A^h(t))}{\mu(A^h(t))}\right)^2\le 2\varepsilon\,,$$

since $\frac{\mu(A^h(t))}{\mu(A^h(t))}=\begin{cases}0 & \text{if } \mu(A^h(t))=0\\ 1 & \text{else}\end{cases}$ and

$$J_{n,2}\le 2\int_{S^*}\left(\frac{\int_{A^h(t)}|\tilde{v}(x)-\tilde{v}(t)|\mu(dx)}{\mu(A^h(t))}\right)^2\mu(dt)\,,$$

with suitable compact set S^*, where the integrand is bounded by $4(\max_{t\in\mathbb{R}^d}|\tilde{v}(t)|)$ and goes to zero as $n\to\infty$, since $\tilde{v}$ is uniformly continuous and (22) is assumed. Hence by Lebesgue's Dominated Convergence Theorem, $J_{n,2}\to 0$ as $n\to\infty$. By Lemma 5 and the Cauchy-Schwarz inequality we get

$$\begin{aligned}
J_{n,3} &\le 2\int\left(\frac{\int_{A^h(t)}|\tilde{v}(x)-v(x)|\mu(dx)}{\mu(A^h(t))}\right)^2\mu(dt)\\
&\le 2\int\frac{\mu(A^h(t))\int_{A^h(t)}|\tilde{v}(x)-v(x)|^2\mu(dx)}{\mu(A^h(t))^2}\mu(dt)\\
&= 2\int\frac{\int I_{A^h(t)}(x)|\tilde{v}(x)-v(x)|^2\mu(dx)}{\mu(A^h(t))}\mu(dt)\\
&= 2\int|\tilde{v}(x)-v(x)|^2\left[\int\frac{I_{A^h(x)}(t)}{\mu(A^h(t))}\mu(dt)\right]\mu(dx)\\
&= 2\int|\tilde{v}(x)-v(x)|^2\frac{\int I_{A^h(x)}(t)\mu(dt)}{\mu(A^h(x))}\mu(dx)\\
&\le 2\int|\tilde{v}(x)-v(x)|^2\mu(dx)\\
&\le 2\varepsilon\,.
\end{aligned}$$

With $\tilde{v}$ as above we obtain

$$\begin{aligned}
L_n^* &\leq 2E\{(v(X_1)-\tilde{v}(X_1))^2\left[1-n\int W_{n,1}^h(x;X_1,X_2,\ldots,X_n)\,\mu(dx)\right]^2\} \\
&\quad +2E\{\tilde{v}(X_1)^2\left[1-n\int W_{n,1}^h(x;X_1,X_2,\ldots,X_n)\,\mu(dx)\right]^2\} \\
&= 2L_n'+2L_n''
\end{aligned}$$

with $h=h(n)$, where by Lemma 5

$$\begin{aligned}
L_n' &\leq 2E\{(v(X_1)-\tilde{v}(X_1))^2\}+2n^2E\{(v(X_1)-\tilde{v}(X_1))^2 \\
&\qquad \left[\int W_{n,1}^h(x;X_1,X_2,\ldots,X_n)\,\mu(dx)\right]^2\} \\
&\leq 2\varepsilon+4\int(v(t)-\tilde{v}(t))^2\left(\frac{\int_{A^h(t)}\mu(dx)}{\mu(A^h(t))}\right)^2\mu(dt) \\
&\leq 6\varepsilon\,.
\end{aligned}$$

Further,

$$\begin{aligned}
L_n'' &= \int\tilde{v}(t)^2\mu(dt) \\
&\quad -2n\int\tilde{v}(t)^2E\{\int W_{n,1}^h(x;t,X_2,\ldots,X_n)\,\mu(dx)\}\mu(dt) \\
&\quad +n^2\int\tilde{v}(t)^2E\{\left[\int W_{n,1}^h(x;t,X_2,\ldots,X_n)\,\mu(dx)\right]^2\}\mu(dt)\,,
\end{aligned}$$

where, by Jensen's inequality, for all $t\in\mathbb{R}^d$

$$\begin{aligned}
&nE\{\int W_{n,1}^h(x;t,X_2,\ldots,X_n)\,\mu(dx)\} \\
&= nE\{\int\frac{I_{A^h(t)}(x)}{1+\sum_{m=2}^n I_{A^h(x)}(X_m)}\,\mu(dx)\} \\
&\geq n\int\frac{I_{A^h(t)}(x)}{1+E\{\sum_{m=2}^n I_{A^h(x)}(X_m)\}}\,\mu(dx) \\
&= n\int\frac{I_{A^h(t)}(x)}{1+(n-1)\mu(A^h(x))}\,\mu(dx) \\
&= n\int\frac{I_{A^h(t)}(x)}{1+(n-1)\mu(A^h(t))}\,\mu(dx) \qquad (25) \\
&\geq \frac{n\mu(A^h(t))}{1+n\mu(A^h(t))} \\
&= 1-\frac{1}{1+n\mu(A^h(t))}
\end{aligned}$$

and, by Lemma 5,

$$\begin{aligned}
&n^2\int\tilde{v}(t)^2E\{\left[\int W_{n,1}^h(x;t,X_2,\ldots,X_n)\,\mu(dx)\right]^2\}\mu(dt) \\
&\leq \int\left(1+\frac{3}{n\mu(A^h(t))}\right)\left(\frac{\int_{A^h(t)}\mu(dx)}{\mu(A^h(t))}\right)^2\tilde{v}(t)^2\mu(dt) \\
&\leq \int\tilde{v}(t)^2\mu(dt)+\frac{3}{n}\int\frac{\tilde{v}(t)^2}{\mu(A^h(t))}\mu(dt)\,.
\end{aligned}$$

Thus, the terms of the form $\int \tilde{v}(t)^2\mu(dt)$ in L''_n cancel out and with suitable compact set S^*,

$$
\begin{aligned}
L''_n &\leq 2\int_{S^*} \frac{1}{1+n\mu(A^h(t))}\tilde{v}(t)^2\mu(dt) + 3\int_{S^*} \frac{1}{n\mu(A^h(t))}\tilde{v}(t)^2\mu(dt) \\
&\leq \frac{5\max_{s\in S^*}\tilde{v}(s)^2}{n}\int_{S^*}\frac{1}{\mu(A^h(t))}\mu(dt) \\
&\leq \frac{5\max_{s\in S^*}\tilde{v}(s)^2}{n}\sum_{j\,:\,A_j^{h(n)}\cap S^*\neq\emptyset}\frac{\mu(A_j^{h(n)})}{\mu(A_j^{h(n)})} \\
&\leq 5\max_{s\in S^*}\tilde{v}(s)^2\frac{|\{j\,:\,A_j^{h(n)}\cap S^*\neq\emptyset\}|}{n} \to 0 \text{ as } n\to\infty\,. \qquad (26)
\end{aligned}
$$

by (23).

Therefore $L^*_n \to 0$ as $n\to\infty$. □

Remark 2. *An analogous result to Lemma 3 can be proven for the kernel estimator with $\mathcal{K} = I_{S_{0,1}}$, replacing $A^h(x)$ by $x + hS_{0,1}$ as follows: Note that the analoga of (24), (25) need not be valid any more, however, the limit (26) can be carried over according to (5.1) on page 76 of [2] under the assumption $nh(n)^d \to \infty$ as $n\to\infty$. In the special case $\mu = const.\lambda$ (λ Lebesgue-Borel measure on $[0,T]^d$, see also Corollary 1 below) equations (24), (25) hold due to the translation invariance of λ.*

2.3 Convergence rates for $\hat{y}_n^h$

Lemma 6. *Assume that*

$$\forall x\in\mathbb{R}^d\,:\; Var\{Y|X=x\}\leq\sigma^2\,.$$

Then for $\hat{y}_n^h = y_n$ according to (2) with (16) and $l^h<\infty$

$$\forall n\in\mathbb{N}, h>0\;\exists\beta_n^h>0\,:\quad E\{\left\|\hat{y}_n^h - Q_n^h y\right\|_{H_3}^2\}\leq\beta_n^h \qquad (27)$$

holds with

$$\beta_n^h = 2\sigma^2\frac{l^h}{n}$$

Proof. The proof follows along the lines of part of the proof of Theorem 4.3 in [2], see the estimate of the term $E\{\int(m_n(x)-\hat{m}_n(x))^2\mu(dx)\}$ on pages 65, 66. □

Note that similar results can also be obtained for the kernel and the nearest neighbour estimate, see, Theorems 5.2, 6.2 in [2].

2.4 Stability estimates

Estimates of the form

$$\forall n\in\mathbb{N}, h>0\;\exists C_n^h>0\,:\quad P\{\left\|(Q_n^hK)^\dagger|_{Y_n^h+\mathcal{R}(Q_n^h)}\right\|_{H_3\to\tilde{H}_1}\leq C_n^h\} \;=\; 1 \qquad (28)$$

heavily depend on the degree of illposedness of the operator equation $Kf=y$ and its interplay with the discretization operator Q_n^h and are therefore problem-dependent. We here consider first kind Volterra integral equations

$$\int_0^t k(t,\tau)f(\tau)\,d\tau = y(t)\,,\quad t\in(0,T)\,.$$

since on one hand they appear in a wide range of applications, on the other hand their degree of ill-posedness can be specified in a precise and clear manner via the concept of k-smoothing Volterra operators. Namely, k-fold differentiation of a first kind integral equation with a k-smoothing Volterra operator leads to a well posed second kind Volterra equation, hence an integral equation with a k-smoothing Volterra operator is k times as ill-posed as numerical differentiation. We will here only describe the case $k = 1$ in detail, since the generalization to $k > 1$ is quite obvious. More precisely, let K be defined by

$$(Kf)(t) = \int_0^t \mathbf{k}(t,\tau) f(\tau)\, d\tau\,. \tag{29}$$

with

$$|\mathbf{k}(t,t)| \geq \underline{\gamma} > 0 \quad \forall t \in (0,T)\,, \tag{30}$$

and a sufficiently smooth kernel

$$\mathbf{k} \in L^\infty(0,T;L^2(0,T))\,. \tag{31}$$

Also, let

$$H_1 = L^2(0,T) \tag{32}$$

and let the spaces H_2, H_3 be defined as in (13). The norm $\|\cdot\|_{\tilde{H}_1}$ to be defined in Proposition 1 below will be an approximation of $\|\cdot\|_{H_1}$ in the sense of (7), see Lemma 8.

Moroever, consider again the partitioning estimator for defining Q_n^h, i.e., we use (4) with

$$W_{n,i}^h(X;X_1,\ldots,X_n) = \sum_{j=1}^{l^h} \frac{I_{\{X_i \in A_j^h\}}}{\sum_{m=1}^n I_{\{X_m \in A_j^h\}}} I_{\{X \in A_j^h\}}\,.$$

Boundedness of $K : H_1 \to H_2$ can be seen as follows:

$$\begin{aligned}
\|Kf\|_{H_2}^2 &= \frac{1}{n}\sum_{i=1}^n \left(\int_0^{X_i} \mathbf{k}(X_i,\tau) f(\tau)\, d\tau\right)^2 \\
&\leq \frac{1}{n}\sum_{i=1}^n \int_0^{X_i} (\mathbf{k}(X_i,\tau))^2\, d\tau \int_0^{X_i} (f(\tau))^2\, d\tau \\
&\leq \|\mathbf{k}\|_{L^\infty(0,T;L^2(0,T))} \|f\|_{L^2(0,T)}^2\,.
\end{aligned}$$

Due to the estimate

$$\begin{aligned}
\left\|(Q_n^h K)^\dagger|_{\tilde{Y}_n^h}\right\|_{H_3 \to \tilde{H}_1} &= \sup_{q \in \tilde{Y}_n^h\,,\ q \neq 0} \frac{\left\|(Q_n^h K)^\dagger q\right\|_{\tilde{H}_1}}{\|q\|_{H_3}} \\
&= \sup_{q \in \tilde{Y}_n^h\,,\ q \neq 0} \frac{\left\|(Q_n^h K)^\dagger q\right\|_{\tilde{H}_1}}{\sqrt{\left\|\mathrm{Proj}_{\mathcal{R}(Q_n^h K)}^{H_3} q\right\|_{H_3}^2 + \left\|\mathrm{Proj}_{\mathcal{R}(Q_n^h K)^\perp}^{H_3} q\right\|_{H_3}^2}} \\
&\leq \sup_{f \in (Q_n^h K)^\dagger \tilde{Y}_n^h\,,\ f \neq 0} \frac{\|f\|_{\tilde{H}_1}}{\|Q_n^h K f\|_{H_3}}\,,
\end{aligned}$$

where we have used $\mathrm{Proj}_{\mathcal{R}(Q_n^h K)}^{H_3} = (Q_n^h K)(Q_n^h K)^\dagger$, we can derive

$$\begin{aligned}
&\forall n \in \mathbb{N}, h > 0\ \exists C_n^h > 0\ \forall f \in (Q_n^h K)^\dagger \tilde{Y}_n^h\ : \\
&\qquad P\{C_n^h \left\|Q_n^h K f\right\|_{H_3} \geq \|f\|_{\tilde{H}_1}\} \ = \ 1
\end{aligned} \tag{33}$$

as a sufficient condition for (28).

Proposition 1. *Define*

$$\|f\|^2_{\tilde{H}_1} := \sum_{j=1}^{l^h}\left(\frac{1}{\hat{c}^h_{n,j}}\sum_{i=1}^{n} w^h_{n,i,j}\int_0^{X_i}\mathbf{k}(X_i,\tau)f(\tau)\,d\tau - \frac{1}{\hat{c}^h_{n,j-1}}\sum_{\nu=1}^{n} w^h_{n,\nu,j-1}\int_0^{X_\nu}\mathbf{k}(X_\nu,\tau)f(\tau)\,d\tau\right)^2 \tag{34}$$

with the abbreviation

$$w^h_{n,i,j} = \frac{I_{\{X_i\in A^h_j\}}}{\sum_{m=1}^n I_{\{X_m\in A^h_j\}}}\,,\quad \hat{c}^h_{n,j} = \sqrt{\mu(A^h_j)}\left(1-(1-\mu(A^h_j))^n\right)\,,\quad j\in\{1,\ldots,l^h\}$$

(note that $\sum_{i=1}^n w^h_{n,i,j} = 1$ *and* $\sum_{j=1}^{l^h} w^h_{n,i,j} I_{\{X\in A^h_j\}} = W^h_{n,i}(X;X_1,\ldots,X_n)$*).*

Then (33) and therewith (28) holds with

$$C^h_n = \frac{1}{\min_{k\in\{1,\ldots,l^h\}}\hat{c}_{n,k}\sqrt{\mu(A^h_k)}}\tilde{C}^h\,, \tag{35}$$

where

$$\tilde{C}^h = \sqrt{\sum_{j=1}^{l^h}\left(\prod_{k=j+1}^{l^h}(1+\sqrt{k-1})\right)^2}\,. \tag{36}$$

Here, $\hat{c}^h_{n,k} \geq \hat{c}^h_k := (1-e^{-\alpha})\sqrt{\mu(A^h_k)}$ *and*

$$C^h_n \leq C^h := \frac{1}{(1-e^{-\alpha})\min_{k\in\{1,\ldots,l^h\}}\mu(A^h_k)}\tilde{C}^h$$

if $h = h(n)$ *is chosen such that* $\min_{k\in\{1,\ldots,l^h\}}\mu(A^h_k) \geq \frac{\alpha}{n}$ *for some* $\alpha > 0$*, which is, e.g., the case under condition*

$$\forall x\in\mathrm{I\!R}^d\ :\ n\mu(A^{h(n)}(x)) \to \infty \text{ as } n\to\infty\,.$$

Proof. We get

$$\begin{aligned}\left\|Q^h_n K f\right\|^2_{H_3} &= \int\left(\sum_{i=1}^n\sum_{j=1}^{l^h}\frac{I_{\{X_i\in A^h_j\}}}{\sum_{m=1}^n I_{\{X_m\in A^h_j\}}} I_{\{x\in A^h_j\}}\int_0^{X_i}\mathbf{k}(X_i,\tau)f(\tau)\,d\tau\right)^2\mu(dx)\\ &= \sum_{k=1}^{l^h}\int_{A^h_k}\left(\sum_{i=1}^n\sum_{j=1}^{l^h} w^h_{n,i,j} I_{\{x\in A^h_k\}}\int_0^{X_i}\mathbf{k}(X_i,\tau)f(\tau)\,d\tau\right)^2\mu(dx)\\ &= \sum_{k=1}^{l^h}\mu(A^h_k)\left(\sum_{i=1}^n w^h_{n,i,k}\int_0^{X_i}\mathbf{k}(X_i,\tau)f(\tau)\,d\tau\right)^2\,.\end{aligned}$$

Denoting

$$\begin{aligned}\xi_j &= \frac{1}{\hat{c}^h_{n,j}}\sum_{i=1}^n w^h_{n,i,j}\int_0^{X_i}\mathbf{k}(X_i,\tau)f(\tau)\,d\tau - \frac{1}{\hat{c}^h_{n,j-1}}\sum_{\nu=1}^n w^h_{n,\nu,j-1}\int_0^{X_\nu}\mathbf{k}(X_\nu,\tau)f(\tau)\,d\tau\\ &\qquad j\in\{2,\ldots,l^h\} \qquad (37)\\ \xi_1 &= \frac{1}{\hat{c}_{n,1}}\sum_{i=1}^n w^h_{n,i,1}\int_0^{X_i}\mathbf{k}(X_i,\tau)f(\tau)\,d\tau\end{aligned}$$

we obviously have

$$\xi_j = -\sum_{k=1}^{j-1} \xi_k + \frac{1}{\hat{c}_{n,j}^h} \sum_{i=1}^n w_{n,i,j}^h \int_0^{X_i} \mathbf{k}(X_i,\tau) f(\tau)\, d\tau \,.$$

Applying the simple estimate (that can be shown by induction)

$$\left(\forall j \in \{1,\dots,l^h\} : \; \zeta_j \le \sum_{k=1}^{j-1} \zeta_k + B_k\right) \;\Rightarrow\; \left(\forall j \in \{1,\dots,l^h\} : \; |\vec{\zeta}|_{l^2}^2 \le (\tilde{C}^h)^2 |\vec{B}|_{l^2}^2\right)$$

to $\zeta_j = |\xi_j|$, $B_j = |\frac{1}{\hat{c}_{n,j}^h} \sum_{i=1}^n w_{n,i,j}^h \int_0^{X_i} \mathbf{k}(X_i,\tau) f(\tau)\, d\tau|$, we get

$$\|f\|_{\tilde{H}_1}^2 = |\vec{\xi}|_{l^2}^2 \le (C_n^h)^2 \left\|Q_n^h K f\right\|_{H_3}^2 \,.$$

□

Lemma 7.

$$E\{w_{n,i,j}^h g(X_i)\} = \frac{1}{n\mu(A_j^h)} \left(1 - (1-\mu(A_j^h))^n\right) \int_{A_j^h} g(t)\mu(dt)\,.$$

Proof. By the identity

$$\frac{I_{\{X_i \in A_j^h\}}}{\sum_{m=1}^n I_{\{X_m \in A_j^h\}}} = \frac{1}{1 + \sum_{m=1, m\ne i}^n I_{\{X_m \in A_j^h\}}} I_{\{X_i \in A_j^h\}}\,,$$

that can be obtained by distinction between the cases $X_i \in A_j^h$ and $X_i \notin A_j^h$, we get that

$$w_{n,i,j}^h g(X_i) = \underbrace{\frac{1}{1 + \sum_{m=1, m\ne i}^n I_{\{X_m \in A_j^h\}}}}_{=a(X_m\,:\,m\ne i)} \underbrace{I_{\{X_i \in A_j^h\}} g(X_i)}_{=b(X_i)}\,,$$

hence by independence

$$E\{w_{n,i,j}^h g(X_i)\} = E\{a(X_m \,:\, m \ne i)\} E\{b(X_i)\}$$

where

$$E\{a(X_m \,:\, m\ne i)\} = \frac{1}{n\mu(A_j^h)} \left(1 - (1-\mu(A_j^h))^n\right)$$

(see the proof of Lemma 4.1 in [2] with $p = \mu(A_j^h)$). □

Lemma 8. *Let $\mu = \lambda$ be the restricted Lebesgue measure times $\frac{1}{T}$, with $(\mathcal{P}^h)_{h>0}$ a family of equidistant partitions of $[0,T]$, $\mathcal{P}^h = \{A_1^h, A_2^h, \dots\}$, and let in addition to the assumptions made above,*

$$\partial_1 \mathbf{k} \in L^2((0,T)^2) \tag{38}$$

hold.

Then there exists a $c > 0$ such that

$$\forall f \in L^2(0,T) : \quad \liminf_{\substack{h\to 0\\ n\to\infty}} E\{\|f\|_{\tilde{H}_1}^2\} \ge c\,\|f\|_{H_1}^2 \,. \tag{39}$$

Consequently (7) holds.

Proof. Let $f \in L^2(0,T)$. Taking into account the fact that by the Cauchy-Schwarz inequality

$$E\{\|f\|_{\tilde{H}_1}^2\} = \sum_{j=1}^{l^h} E\{\xi_j^2\} \geq \sum_{j=1}^{l^h} (E\{\xi_j\})^2 \tag{40}$$

with ξ according to (37), we first of all show that

$$\liminf_{\substack{h\to 0\\ n\to\infty}} \sum_{j=1}^{l^h} (E\{\xi\})^2 \geq c\,\|f\|_{H_1}^2 \,. \tag{41}$$

Due to Lemma 7, we have

$$\begin{aligned} E\{\xi_j\} &= \frac{1}{\sqrt{\mu(A_j^h)}^3} \int_{A_j^h} \int_0^t \mathbf{k}(t,\tau) f(\tau)\, d\tau \mu(dt) \\ &\quad - \frac{1}{\sqrt{\mu(A_{j-1}^h)}^3} \int_{A_{j-1}^h} \int_0^t \mathbf{k}(t,\tau) f(\tau)\, d\tau \mu(dt)\,. \end{aligned}$$

Since we are in the 1-d case with μ being the restricted Lebesgue measure up to a constant factor, we can write $A_j^h = [t_j^h, t_{j+1}^h)$, $\mu(A_j) = h$, for $t_j^h = (j-1)h$, $h = \frac{T}{l^h}$. Therewith,

$$\sum_{j=1}^{l^h} (E\{\xi_j\})^2 = \frac{1}{h^3} \sum_{j=1}^{l^h} \left(\int_{t_j^h}^{t_{j+1}^h} \int_0^t \mathbf{k}(t,\tau) f(\tau)\, d\tau\, dt - \int_{t_{j-1}^h}^{t_j^h} \int_0^t \mathbf{k}(t,\tau) f(\tau)\, d\tau\, dt \right)^2 .$$

Integration by parts yields

$$\begin{aligned} &\int_{t_{j-1}^h}^{t_j^h} \int_0^t \mathbf{k}(t,\tau) f(\tau)\, d\tau\, dt \\ &= -\int_{t_{j-1}^h}^{t_j^h} (t - t_{j-1}^h) \left(\mathbf{k}(t,t) f(t) + \int_0^t \partial_1 \mathbf{k}(t,\tau) f(\tau)\, d\tau \right) dt + h \int_0^{t_j^h} \mathbf{k}(t,\tau) f(\tau)\, d\tau \end{aligned}$$

$$\begin{aligned} &\int_{t_j^h}^{t_{j+1}^h} \int_0^t \mathbf{k}(t,\tau) f(\tau)\, d\tau\, dt \\ &= -\int_{t_j^h}^{t_{j+1}^h} (t - t_{j+1}^h) \left(\mathbf{k}(t,t) f(t) + \int_0^t \partial_1 \mathbf{k}(t,\tau) f(\tau)\, d\tau \right) dt + h \int_0^{t_j^h} \mathbf{k}(t,\tau) f(\tau)\, d\tau\,, \end{aligned}$$

so

$$\sum_{j=1}^{l^h} (E\{\xi_j\})^2 = \frac{2}{3} \sum_{j=1}^{l^h} \langle \phi_j^n, \mathbf{k}(\cdot,\cdot) f + \int_0^{\cdot} \partial_1 \mathbf{k}(\cdot,\tau) f(\tau)\, d\tau \rangle^2 \,,$$

where $\langle \cdot,\cdot \rangle$ denotes the L^2 inner product and ϕ_j^n, $j \in \{1,\dots,l^h\}$ are the L^2 normalized hat functions on $(0,T)$:

$$\phi_j^n(t) = \sqrt{\frac{3}{2h^3}} \begin{cases} t - t_{j-1}^h & t \in [t_{j-1}^h, t_j^h] \\ t_{j+1}^h - t & t \in [t_j^h, t_{j+1}^h] \\ 0 & \text{else.} \end{cases}$$

Since the union of the spaces $\mathcal{H}^h := \operatorname{span}(\phi_j^n : j \in \{1,\dots,l^h\})$ is dense in $L^2(0,T)$, we get

$$\begin{aligned} \lim_{\substack{h\to 0\\ n\to\infty}} \sum_{j=1}^{l^h} (E\{\xi_j\})^2 &= \frac{2}{3} \left\| \mathbf{k}(\cdot,\cdot) f + \int_0^{\cdot} \partial_1 \mathbf{k}(\cdot,\tau) f(\tau)\, d\tau \right\|_{L^2(0,T)}^2 \\ &\geq \frac{2}{3C^2} \|f\|_{L^2(0,T)}^2 \end{aligned}$$

for some constant $C > 0$ depending only on T, $\underline{\gamma}$ and $\|\partial_1 \mathbf{k}\|_{L^2((0,T)^2)}$, by bounded invertibility of the second kind Volterra integral operator $L^2(0,T) \to L^2(0,T)$, $f \mapsto \mathbf{k}(\cdot,\cdot)f + \int_0^{\cdot} \partial_1 \mathbf{k}(\cdot,\tau) f(\tau)\, d\tau$, see, e.g., [1], [3].

The conclusion of (7) from (39) can be done by the following argument: Assume that for some family $(f_n^h)_{n\in\mathbb{N},h>0}$ of elements of $(Q_n^h K)^\dagger \tilde{Y}_n^h \subseteq L^2(0,T)$ there exists an $\varepsilon > 0$ and a subsequence $(h_k, n_k)_{k\in\mathbb{N}}$ with $h_k \to 0$, $n_k \to \infty$ as $k \to \infty$, such that for $f_k := f_{n_k}^{h_k}$, $\tilde{H}_{k,1} = \tilde{H}_{n_k,1}^{h_k}$

$$(a) \lim_{k\to\infty} E\{\|f_k\|^2_{\tilde{H}_{k,1}}\} \stackrel{k\to\infty}{\longrightarrow} 0 \text{ but } (b)\ \forall k \in \mathbb{N} :\ E\{\|f_k\|^2_{H_1}\} \geq \varepsilon\,.$$

Then due to (a), for $k \geq k_0$ with k_0 sufficiently large we have

$$E\{\|f_k\|^2_{\tilde{H}_{k,1}}\} < c\frac{\varepsilon}{2}\,.$$

On the other hand, due to (b) we can divide by $E\{\|f_k\|^2_{H_1}\}$ and consider the normalized sequence $\frac{\|f_k\|^2_{H_1}}{E\{\|f_k\|^2_{H_1}\}}$ whose expectation is constant 1 and which is therewith convergent to 1 in expectation. Hence there exists an almost surely convergent subsequence, i.e., k_m such that $\frac{\|f_{k_m}\|^2_{H_1}}{E\{\|f_{k_m}\|^2_{H_1}\}} \stackrel{l\to\infty}{\longrightarrow} 1$ almost surely, so there exists an m_0 such that for all $m \geq m_0$ there holds $\frac{\|f_{k_m}\|^2_{H_1}}{E\{\|f_{k_m}\|^2_{H_1}\}} \geq \frac{1}{2}$ almost surely and therewith by (b)

$$\|f_{k_m}\|^2_{H_1} \geq \frac{\varepsilon}{2} \text{ almost surely}$$

which implies

$$E\{\|f_{k_m}\|^2_{\tilde{H}_{k,1}}\} \geq cE\{\|f_{k_m}\|^2_{H_1}\} \geq c\frac{\varepsilon}{2}\,.$$

This gives a contradiction. □

Summarizing, we get the following convergence result:

Corollary 1. *Let K in (1) be defined by (29) with (30), (31), (38). Moreover, let $\mu = \frac{1}{T}\lambda$ (λ Lebesgue-Borel measure on $[0,T]$), with $(\mathcal{P}^h)_{h>0}$ a family of equidistant partitions of $[0,T]$. Define $\hat{f}_n^h = (Q_n^h K)^\dagger \hat{y}_n^h$ with $\hat{y}_n^h = y_n$ according to (2) and Q_n^h according to (4) with (16). and let $h = h(n) = \frac{T}{l_n}$ be chosen such that*

$$\left(l_n \to \infty \quad \textit{and} \quad \frac{l_n}{n} \to 0 \quad \textit{and} \quad \frac{(l_n)^2 \tilde{C}^{h(n)}}{n} \to 0\right) \quad \textit{as } n \to \infty \tag{42}$$

with $\tilde{C}^h$ according to (36).

Then

$$E\{\left\|\hat{f}_n^{h(n)} - f^\dagger\right\|^2_{L^2(0,T)}\} \to 0 \textit{ as } n \to \infty\,.$$

Proof. The result is a consequence of Theorem 1 together with inequality (18), Lemmas 2, 3, 6, 8, and the fact that $B = C(0,T)$ is contained in $H_2 \cap H_3$ according to (13) and dense in H_2. □

Remark 3. *An example for a choice of l_n such that (42) is satisfied is*

$$l_n = [(\log n)^{1-\delta}] \quad \textit{for some } \delta \in (0,1)\,,$$

(with the notation $[\alpha]$ for the largest integer $\leq \alpha \in \mathbb{R}$)

$$\frac{l_n^2 \tilde{C}^h}{n} \leq \frac{l_n^2}{n}\sqrt{\sum_{j=1}^{l_n}\left(\prod_{k=j+1}^{l_n}(\sqrt{e}\sqrt{k})\right)^2}$$

$$
\begin{aligned}
&= \frac{l_n^2}{n}\sqrt{\sum_{j=1}^{l_n} e^{l_n-j}\frac{l_n!}{j!}} \\
&\le \frac{l_n^2}{n}\sqrt{l_n e^{l_n} l_n!} \\
&\le \textit{const.}\frac{l_n^{11/4}}{n} l_n^{l_n/2} \\
&= \textit{const.}\exp(-\log n + \tfrac{1}{2} l_n \log(l_n) + \tfrac{11}{4}\log(l_n)) \\
&\le \textit{const.}\exp(-\log n + \tfrac{1}{2}(\log n)^{1-\delta}(1-\delta)\log\log n + \tfrac{11}{4}(1-\delta)\log\log n) \\
&\to 0 \;\; \textit{as } n \to \infty\,,
\end{aligned}
$$

where we have used Stirling's formula in the form $l! \le \textit{const.}\sqrt{l}(\frac{l}{e})^l$. *On the other hand, note that* $l_n = [\log n]$ *would not be a feasible choice since*

$$
\begin{aligned}
\frac{[\log n]^2 \tilde{C}^h}{n} &\ge \frac{1}{n}\sqrt{\sum_{j=1}^{[\log n]}\left(\prod_{k=j+1}^{[\log n]}(\sqrt{k-1})\right)^2} \\
&= \frac{1}{n}\sqrt{\sum_{j=1}^{[\log n]}\frac{([\log n]-1)!}{(j-1)!}} \\
&\ge \frac{1}{n}\sqrt{[\log n]!} \to \infty \;\; \textit{as } n \to \infty
\end{aligned}
$$

3 Conclusions and Remarks

Investigating the question of how regression function estimation can be used as regularization methods for linear ill-posed problems with additional measurement errors, we have shown convergence mainly for the partitioning estimators and to some extent also for kernel estimator.

Future research will be devoted to a posteriori regularization parameter choice (e.g. by cross-validation or splitting the sample) and to a generalization to nonlinear problems. Also use of least squares methods for nonparametric regression estimation instead of local averaging methods is of interest, further establishing almost sure convergence under sharpened integrability assumptions instead of convergence in squared mean, and rate of convergence under smoothness assumptions (on $f^\dagger$ in (1)).

References

[1] H.W. Engl, *Integralgleichungen*, Springer, Wien, 1997.

[2] L. Györfi, M. Kohler, A. Krzyzak, and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*, Springer Series in Statistics, Springer, 2002.

[3] R. Kress, *Linear Integral Equations*, Springer, New York, 1999. 2nd ed.

[4] P.Mathé and S.V. Pereverzev, Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods, SIAM J. Numer.Anal. 38 (2000), 1999–2021.

Barbara Kaltenbacher
Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, 8010 Graz, Austria.
E-Mail: barbara.kaltenbacher@uni-graz.at

Harro Walk
Fachbereich Mathematik, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany.
E-Mail: Harro.Walk@mathematik.uni-stuttgart.de

Erschienene Preprints ab Nummer 2007/001

Komplette Liste: `http://www.mathematik.uni-stuttgart.de/preprints`

2010/005 *Kaltenbacher, B.; Walk, H.:* On convergence of local averaging regression function estimates for the regularization of inverse problems

2010/004 *Kühnel, W.; Solanes, G.:* Tight surfaces with boundary

2010/003 *Kohler, M; Walk, H.:* On optimal exercising of American options in discrete time for stationary and ergodic data

2010/002 *Gulde, M.; Stroppel, M.:* Stabilizers of Subspaces under Similitudes of the Klein Quadric, and Automorphisms of Heisenberg Algebras

2010/001 *Leitner, F.:* Examples of almost Einstein structures on products and in cohomogeneity one

2009/008 *Griesemer, M.; Zenk, H.:* On the atomic photoeffect in non-relativistic QED

2009/007 *Griesemer, M.; Moeller, J.S.:* Bounds on the minimal energy of translation invariant n-polaron systems

2009/006 *Demirel, S.; Harrell II, E.M.:* On semiclassical and universal inequalities for eigenvalues of quantum graphs

2009/005 *Bächle, A, Kimmerle, W.:* Torsion subgroups in integral group rings of finite groups

2009/004 *Geisinger, L.; Weidl, T.:* Universal bounds for traces of the Dirichlet Laplace operator

2009/003 *Walk, H.:* Strong laws of large numbers and nonparametric estimation

2009/002 *Leitner, F.:* The collapsing sphere product of Poincaré-Einstein spaces

2009/001 *Brehm, U.; Kühnel, W.:* Lattice triangulations of E^3 and of the 3-torus

2008/006 *Kohler, M.; Krzyżak, A.; Walk, H.:* Upper bounds for Bermudan options on Markovian data using nonparametric regression and a reduced number of nested Monte Carlo steps

2008/005 *Kaltenbacher, B.; Schöpfer, F.; Schuster, T.:* Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems

2008/004 *Leitner, F.:* Conformally closed Poincaré-Einstein metrics with intersecting scale singularities

2008/003 *Effenberger, F.; Kühnel, W.:* Hamiltonian submanifolds of regular polytope

2008/002 *Hertweck, M.; Hofert, C.R.; Kimmerle, W.:* Finite groups of units and their composition factors in the integral group rings of the groups $PSL(2, q)$

2008/001 *Kovarik, H.; Vugalter, S.; Weidl, T.:* Two dimensional Berezin-Li-Yau inequalities with a correction term

2007/006 *Weidl, T.:* Improved Berezin-Li-Yau inequalities with a remainder term

2007/005 *Frank, R.L.; Loss, M.; Weidl, T.:* Polya's conjecture in the presence of a constant magnetic field

2007/004 *Ekholm, T.; Frank, R.L.; Kovarik, H.:* Eigenvalue estimates for Schrödinger operators on metric trees

2007/003 *Lesky, P.H.; Racke, R.:* Elastic and electro-magnetic waves in infinite waveguides

2007/002 *Teufel, E.:* Spherical transforms and Radon transforms in Moebius geometry

2007/001 *Meister, A.:* Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions