



Estimation of conditional distribution functions from data with additional errors applied to shape optimization

Matthias Hansmann¹ · Benjamin M. Horn² · Michael Kohler¹ · Stefan Ulbrich²

Received: 25 January 2020 / Accepted: 5 July 2021 / Published online: 17 July 2021
© The Author(s) 2021

Abstract

We study the problem of estimating conditional distribution functions from data containing additional errors. The only assumption on these errors is that a weighted sum of the absolute errors tends to zero with probability one for sample size tending to infinity. We prove sufficient conditions on the weights (e.g. fulfilled by kernel weights) of a local averaging estimate of the codf, based on data with errors, which ensure strong pointwise consistency. We show that two of the three sufficient conditions on the weights and a weaker version of the third one are also necessary for the spc. We also give sufficient conditions on the weights, which ensure a certain rate of convergence. As an application we estimate the codf of the number of cycles until failure based on data from experimental fatigue tests and use it as objective function in a shape optimization of a component.

Keywords Conditional distribution function estimation · Consistency · Experimental fatigue tests · Local averaging estimate · Shape optimization · Isogeometric analysis

Mathematics Subject Classification 62G05 · 62G20

✉ Matthias Hansmann
hansmann@mathematik.tu-darmstadt.de

Benjamin M. Horn
bhorn@mathematik.tu-darmstadt.de

Michael Kohler
kohler@mathematik.tu-darmstadt.de

Stefan Ulbrich
ulbrich@mathematik.tu-darmstadt.de

¹ Department of Mathematics, TU Darmstadt, 64289 Darmstadt, Germany

² Department of Mathematics, TU Darmstadt, 64293 Darmstadt, Germany

1 Introduction

Let (X, Y) be a random vector, such that X is \mathbb{R}^d -valued and Y is real-valued, with conditional distribution function (codf) F , i.e.,

$$F(y, x) = \mathbf{P}\{Y \leq y \mid X = x\} = \mathbf{E} \left\{ I_{\{Y \leq y\}} \mid X = x \right\}.$$

One idea to construct estimates, approaching the codf asymptotically for some fixed $y \in \mathbb{R}$ and \mathbf{P}_X -almost all $x \in \mathbb{R}^d$ (where \mathbf{P}_X is the of X induced measure on $(\mathbb{R}^d, \mathcal{B}_d)$, i.e., $\mathbf{P}_X(B) = \mathbf{P}(X \in B)$ for every $B \in \mathcal{B}_d$), is to use an independent and identically distributed (i.i.d.) sample $(X_1, Y_1), \dots, (X_n, Y_n)$ of (X, Y) to compute a local averaging estimate

$$F_n(y, x) = \sum_{i=1}^n W_{n,i}(x) I_{\{Y_i \leq y\}} \tag{1}$$

of the codf. Here $W_{n,i}(x)$ for $i = 1, \dots, n$ are nonnegative weights, which can depend on the samples X_1, \dots, X_n .

A commonly used example for those weights are the weights of the so-called kernel estimate, which are defined by

$$W_{n,i}(x) = \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}, \tag{2}$$

where $0/0 = 0$ by definition (cf., e.g., Nadaraya (1964) and Watson (1964)). Here $h_n > 0$ is the so-called bandwidth and $K : \mathbb{R}^d \rightarrow \mathbb{R}$ is a so-called kernel function, e.g., the so-called naive kernel defined by

$$K(x) = I_{\{\|x\| \leq 1\}} \text{ for all } x \in \mathbb{R}^d.$$

For a fixed $y \in \mathbb{R}$, the estimate introduced in (1) is a special case of an estimate of a regression function $m(x) = \mathbf{E}\{Y \mid X = x\}$ (with the choice of $I_{\{Y \leq y\}}$ as dependent variable). Thus, all known results on estimates of the regression function do also apply for the corresponding estimates of the codf.

The regression estimate corresponding to the one of the codf in (1), i.e.,

$$m_n(x) = \sum_{i=1}^n W_{n,i}(x) \cdot Y_i$$

has been seminaly considered by Stone (1977). In particular, in Theorem 1 Stone gave necessary and sufficient conditions on the weights, such that the regression estimate m_n is weakly consistent in L_r for every real $r \geq 1$, i.e.,

$$\mathbf{E} \left\{ \int |m_n(x) - m(x)|^r \mathbf{P}_X(dx) \right\} \rightarrow 0 \quad (n \rightarrow \infty).$$

These conditions are for example fulfilled by the special choices of the weights of the partitioning, kernel and nearest-neighbor estimate, for details we refer to Chapters 4, 5 and 6 in Györfi et al. (2002).

Since the work of Stone (1977), several authors also dealt with strongly pointwise consistency of special regression function estimates for dependent variable Y , which are almost surely bounded by some constant. In case of the estimation of the codf we say that the estimate F_n of the codf F is *strongly pointwise consistent* for some fixed $y \in \mathbb{R}$, if

$$F_n(y, x) \rightarrow F(y, x) \quad a.s. \quad \text{for } \mathbf{P}_X\text{-almost every } x. \tag{3}$$

In the context of nonparametric regression, Devroye (1981) showed the strongly pointwise consistency of the kernel regression estimate, presuming that K is a so-called window kernel and that the bandwidth h_n fulfills some mild asymptotic conditions. Greblicki et al. (1984) generalized this consistency result to some broader class of kernels with possibly unbounded support. Stute (1986) also showed a result concerning the uniform pointwise consistency of the kernel estimate of the conditional distribution function. A proof of the strongly pointwise consistency of the partitioning regression estimate can be found in (Györfi et al. 2002, Theorem 25.6.). Györfi (1981a) and Devroye (1981) independently showed results concerning the strongly pointwise consistency of the nearest neighbor regression estimates. Devroye (1982) also gave necessary and sufficient conditions for the strongly pointwise consistency of the nearest neighbor regression estimates.

In order to obtain consistency results for all distributions (X, Y) with $\mathbf{E}|Y| < \infty$ (so-called universal consistency), some authors considered modified versions of the above mentioned estimates [cf., e.g., Walk (2001), Algoet and Györfi (1999)]. See also (Györfi et al. 2002, Chapter 25) and the literature cited therein for other estimates and further results on the strongly universal pointwise consistency.

Rates of convergence in probability for the kernel regression estimate have been obtained in Krzyzak and Pawlak (1987) and in Györfi (1981b) for the nearest neighbor regression estimate. Uniform almost sure rates of convergence for regression estimates have been shown in Härdle et al. (1988) by considering a more general setting of kernel-type estimators of conditional functionals. Optimal global rates of convergence for nonparametric regression estimates have been shown by Stone (1982).

Other estimates of the codf have been proposed by Hall et al. (1999), who studied the rate of convergence of a weighted kernel estimator. Cai (2002) showed asymptotic normality of this estimate. Furthermore, Hall and Yao (2005) used a dimension reduction technique to approximate the codf and study the asymptotic properties. Preadjusted local averaging estimates of the codf were proposed by Veraverbeke et al. (2014), who proved results concerning the uniform rate of convergence.

So far, only Liero (1989) and Hansmann et al. (2019) studied a local averaging regression estimate with generalized weights and formulated conditions for consistency results. Liero (1989) assumed the weights to have the special form

$$W_{n,i}(x) = \frac{\phi_{n,i}(x, X_i)}{\sum_{i=1}^n \phi_{n,i}(x, X_i)},$$

where $\phi_{i,n}$ is a Borel-measurable function on $\mathbb{R}^d \times \mathbb{R}^d$ and does therefore only depend on X_i and formulated conditions that ensure a certain uniformly strong rate of convergence. Hansmann et al. (2019) gave conditions on the above introduced general weights $W_{n,i}$ of a local averaging regression estimate which imply the *strongly universally consistency*, i.e.

$$\int |m_n(x) - m(x)|^2 \mathbf{P}_X(dx) \rightarrow 0 \text{ a.s.}$$

for all distributions (X, Y) with $\mathbf{E}Y^2 < \infty$. To the authors knowledge there is no result so far, which characterizes necessary and sufficient conditions on the above mentioned general weights $W_{n,i}$, that ensure the strongly pointwise consistency or a certain rate of convergence in probability of the corresponding local averaging estimate of the conditional distribution function.

One of the main goals of this paper is to present these two results. A further aspect investigated in this paper is the consideration of additional errors in the data, which is motivated by an application in the context of shape optimization with respect to the fatigue life of a component. A short overview on the method used to asses the fatigue behaviour will be described in the following section.

1.1 Application in the context of experimental fatigue tests

In order to predict the fatigue life of a certain material, we use data from so-called strain-controlled fatigue tests, in which a material sample gets repeatedly elongated by a fixed strain amplitude ε . The repetitions, the so-called number of cycles N , until the material fails are counted and the corresponding stress amplitude τ is measured. Repeating this experiment yields data

$$\left\{ \left(\varepsilon_1^{(m)}, \left(N_1^{(m)}, \tau_1^{(m)} \right) \right), \dots, \left(\varepsilon_{l_m}^{(m)}, \left(N_{l_m}^{(m)}, \tau_{l_m}^{(m)} \right) \right) \right\}$$

for each material m . Since the mentioned strain-controlled fatigue tests are very time consuming, we only have 12 data points for the material of interest, which is not enough for a nonparametric estimation of the conditional distribution function of the number of cycles given a certain strain amplitude ε . Thus, we assume the model

$$N^{(m)}(\varepsilon) = \mu^{(m)}(\varepsilon) + \sigma^{(m)}(\varepsilon) \cdot \delta^{(m)} \tag{4}$$

to hold. In Sect. 3.1. we will describe a suitable method to estimate μ and σ by $\hat{\mu}$ and $\hat{\sigma}$, respectively, such that we can finally obtain data

$$\hat{\delta}_i^{(m)} = \frac{N_i^{(m)} - \hat{\mu}_i^{(m)}}{\hat{\sigma}_i^{(m)}} \text{ for } i = 1, \dots, l_m, \tag{5}$$

Due to the assumption in (4) the conditional distribution function of the number of cycles given a strain amplitude ε can be determined by a simple linear transformation

from the distribution function of $\delta^{(m)}$. Since we only have available 4 to 35 of the above data samples per material to estimate the distribution function of $\delta^{(m)}$, we will use data samples from other materials, that have similar static material properties. To this end we use an estimate of the conditional distribution function, with the vector of five static material properties (Young’s modulus, the yield limit for 0.2% residual elongation, the tensile strength, the static strength coefficient and the static strain hardening exponent) as covariate $X^{(m)}$ and the samples $\delta_i^{(m)}$ as dependent variable. More precisely we apply a nonparametric estimate of the codf to the data

$$\left\{ \left(X^{(m)}, \delta_1^{(m)} \right), \dots, \left(X^{(m)}, \delta_{l_m}^{(m)} \right) : m \text{ is a material in our database} \right\}$$

Furthermore the above data points contain errors in the dependent variable since we only estimated $\mu^{(m)}$ and $\sigma^{(m)}$, which leads to the topic of this paper, where we want to investigate how additional errors in the dependent variable influence an estimate of codf and show theoretical results concerning strongly pointwise consistency and rate of convergence in probability.

1.2 Data with errors

Motivated by the application described in the previous subsection, we generalize our mathematical setting and assume that we only have available data $(X_1, \bar{Y}_{1,n}), \dots, (X_n, \bar{Y}_{n,n})$ with errors in the samples of the dependent variable instead of the i.i.d. data $(X_1, Y_1), \dots, (X_n, Y_n)$.

In our above mentioned application we do not know anything explicitly on the errors $\bar{Y}_{i,n} - Y_i$ ($i = 1, \dots, n$). Thus, we are not able to impose a structure on those errors. In particular, we can not assume that those errors have to be random and in case that they are random they do not need to be independent or identically distributed and they do not need to have expectation zero, so estimates for convolution problems (see, e.g., Meister (2009) and the literature cited therein) are not applicable in the context of this paper. But we can assume that with increasing number n of total samples we also get more samples from the strain controlled fatigue tests for each of the materials. Thus, our estimates $\hat{\mu}^{(m)}$ and $\hat{\sigma}^{(m)}$ and therefore our data $\hat{\delta}_i^{(m)}$ of (5) get more reliable for all materials m . Consequently, with increasing n , our errors $\hat{\delta}_i^{(m)} - \delta_i^{(m)}$ get small for all materials m . Since the $\delta_i^{(m)}$ are the samples of our dependent variable it seems to be a natural idea in our application to assume that the absolute errors between Y_i and $\bar{Y}_{i,n}$ uniformly converge to zero almost surely, i.e., to assume that

$$\max_{i=1, \dots, n} |Y_i - \bar{Y}_{i,n}| \rightarrow 0 \text{ a.s.}$$

In our theoretical results in Section 2 it will turn out that we only have to assume the weaker condition

$$\eta_n(x) = \sum_{i=1}^n W_{n,i}(x) \cdot |Y_i - \bar{Y}_{i,n}| \rightarrow 0 \text{ a.s. for } \mathbf{P}_X\text{-almost every } x. \quad (E1)$$

where $W_{n,i}$ are the weights of the local averaging estimate above.

Note also that our set-up is triangular, which is the necessary in our application since the estimates $\hat{\mu}^{(m)}$ and $\hat{\sigma}^{(m)}$ can change with the number of data points n and can therefore lead in (5) to completely new samples with errors of the random variable $\delta^{(m)}$.

Errors, for which an (average) sum of the (squared) absolute errors tends to zero (as in (E1)), have been recently considered in the context of nonparametric regression with random design (cf., Kohler (2006), Fromkorth and Kohler (2011)), nonparametric regression with fixed design (cf., Furer et al. (2013), Furer and Kohler (2015)), (conditional) quantile estimation (cf., Hansmann and Kohler (2017) and Hansmann and Kohler (2019)), density estimation (cf., Felber et al. (2015)) and distribution estimation (cf., Bott et al. (2013)).

Since we do not assume anything on the nature of the errors besides that they are pointwise asymptotically negligible in the sense that (E1) holds, it seems to be a natural idea to ignore them completely and to try to use the same estimates as in the case that an independent and identically distributed sample is given.

1.3 Main Results

In Theorem 2.1 we present sufficient conditions on the weights and prove that these conditions ensure that the estimate $\bar{F}_n(y, x)$ applied to data $(X_1, \bar{Y}_{1,n}), \dots, (X_n, \bar{Y}_{n,n})$ with errors in the samples of the dependent variable is pointwise consistent in the sense that it approaches the interval

$$[F(y^-, x), F(y, x)] = \left[\lim_{\substack{\varepsilon \rightarrow 0, \\ \varepsilon > 0}} F(y - \varepsilon, x), F(y, x) \right]$$

for \mathbf{P}_X -almost all x asymptotically, presumed that the errors fulfill (E1). As we will show in Corollary 2.1, these assumptions on the weights are for example fulfilled by the weights of the kernel estimate.

We also show that two of the three sufficient conditions and a weaker version of the third condition are also necessary for the above pointwise consistency (see Theorem 2.2).

We also investigate the rate of convergence of the estimate \bar{F}_n and present conditions on the weights, which ensure for \bar{F}_n a pointwise rate of convergence in probability of

$$r_n(x) + \sqrt{\eta_n(x)}$$

(see Theorem 2.3), where $r_n(x)$ is some deterministic rate fulfilling for \mathbf{P}_X -almost every $x \in \mathbb{R}^d$ $r_n(x) \rightarrow 0$ as $n \rightarrow \infty$, and where $\eta_n(x)$ is defined in (E1).

We also present an application to simulated and real data (see Section 3). In the real data application we use the considered method to estimate the distribution function of the numbers of cycles until failure in the context of fatigue behavior of steel under cyclic loading. This estimate is utilized as the objective in a shape optimization procedure,

which is embedded in an algorithm-based product development approach to determine an optimal profile geometry with respect to the fatigue behavior.

1.4 Notation

Throughout this paper the following notation is used: We write $U_n = O_{\mathbf{P}}(V_n)$ if the nonnegative random variables U_n and V_n satisfy

$$\lim_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbf{P}\{U_n > c \cdot V_n\} = 0.$$

The sets of natural positive, natural nonnegative and real numbers are denoted by \mathbb{N} , \mathbb{N}_0 and \mathbb{R} , respectively. We write $\rightarrow^{\mathbf{P}}$ as an abbreviation for convergence in probability and I_A for the indicator function of the set A . We denote the Euclidean Norm on \mathbb{R}^d by $\|\cdot\|$. For $z \in \mathbb{R}$ and a set $A \subseteq \mathbb{R}$, we define the distance from z to A as

$$dist(z, A) := \inf_{a \in A} |z - a|.$$

Furthermore, we write for the left-sided limit of a function G

$$G(y^-) = \lim_{\substack{\varepsilon \rightarrow 0, \\ \varepsilon > 0}} G(y - \varepsilon)$$

1.5 Outline

The outline of the paper is as follows: The main results are formulated in Section 2 and proven in the supplemental material. In Section 3, we present an application to simulated and real data.

2 Main results

Let

$$\bar{F}_n(y, x) = \sum_{i=1}^n W_{n,i}(x) I_{\{\bar{Y}_{i,n} \leq y\}} \tag{6}$$

be a local averaging estimate of the codf $F(y, x)$ corresponding to the data with errors $(X_1, \bar{Y}_{1,n}), \dots, (X_n, \bar{Y}_{n,n})$.

2.1 Consistency

First of all, we give sufficient conditions on the sequence of weights $W_{n,i}$, such that the estimate \bar{F}_n is pointwise consistent for all distributions of (X, Y) and all $y \in \mathbb{R}$.

The following result holds, which will be proven in Section S2 in the supplemental material.

Theorem 2.1 *Let $(X, Y), (X_1, Y_1), (X_2, Y_2) \dots$ be i.i.d. $\mathbb{R}^d \times \mathbb{R}$ -valued random vectors and let $W_{n,i}(x) := W_{n,i}(x, X_1, \dots, X_n)$ ($x \in \mathbb{R}^d$) be nonnegative weights, which fulfill*

- (A1) $\sum_{i=1}^n W_{n,i}(x) \rightarrow 1$ a.s. for \mathbf{P}_X -almost every $x \in \mathbb{R}^d$,
- (A2) for every Borel-measurable set $B \in \mathcal{B}_d$

$$\sum_{i=1}^n W_{n,i}(x) [I_{\{X_i \in B\}} - I_{\{x \in B\}}] \rightarrow 0 \text{ a.s. for } \mathbf{P}_X\text{-almost every } x \in \mathbb{R}^d,$$

- (A3) $\log(n) \cdot \sum_{i=1}^n W_{n,i}(x)^2 \rightarrow 0$ a.s. for \mathbf{P}_X -almost every $x \in \mathbb{R}^d$.

Furthermore let $\bar{Y}_{1,n}, \dots, \bar{Y}_{n,n}$ be random variables, which fulfill (E1) and let \bar{F}_n be the local averaging estimate defined in (6) with weights $W_{n,i}$. Then \bar{F}_n is pointwise consistent in the sense that for all $y \in \mathbb{R}$

$$\text{dist}(\bar{F}_n(y, x), [F(y^-, x), F(y, x)]) \rightarrow 0 \text{ a.s. for } \mathbf{P}_X\text{-almost every } x. \tag{7}$$

In the following corollary, which will be proven in Section S2 in the supplemental material, we formulate sufficient conditions for the pointwise strongly consistency of the kernel estimate of the codf, defined by the weights in (2).

Corollary 2.1 *Let $(X, Y), (X_1, Y_1), (X_2, Y_2) \dots$ be i.i.d. $\mathbb{R}^d \times \mathbb{R}$ -valued random vectors. Assume that K is the naive kernel and that the bandwidth $h_n > 0$ fulfills*

$$h_n \rightarrow 0 \text{ and } n \cdot h_n^d / \log(n) \rightarrow \infty \text{ (} n \rightarrow \infty \text{)}. \tag{8}$$

Furthermore, let $\bar{Y}_{1,n}, \dots, \bar{Y}_{n,n}$ be random variables, which fulfill

$$\sum_{i=1}^n \frac{|Y_i - \bar{Y}_{i,n}| \cdot K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} \rightarrow 0 \text{ a.s. for } \mathbf{P}_X\text{-almost every } x. \tag{9}$$

Let $W_{n,i}$ be the weights of the kernel estimate with kernel K and bandwidth h_n . Then the kernel estimate \bar{F}_n of the codf as defined in (6) is pointwise consistent in the sense that for all $y \in \mathbb{R}$

$$\text{dist}(\bar{F}_n(y, x), [F(y^-, x), F(y, x)]) \rightarrow 0 \text{ a.s. for } \mathbf{P}_X\text{-almost every } x.$$

Remark 2.1 Analogous results can be shown for estimates of the codf corresponding to the partitioning and nearest neighbor weights, assuming the conditions from Theorems 25.6. and 25.17., respectively, in Györfi et al. (2002). Furthermore, Corollary 2.1 can be extended to a more general class of kernels, which has been considered by Greblicki et al. (1984).

In Theorem 2.1, we formulated sufficient conditions on the sequence of weights that imply the pointwise consistency in the sense of (7). In the following theorem, we show that at least two of these three conditions and a weaker version of (A3) are also necessary, if F_n is pointwise consistent in the sense that for all distributions of (X, Y) for all $y \in \mathbb{R}$ (7) holds. As we will see in the following theorem, it is sufficient to consider an i.i.d. sample $(X_1, Y_1), (X_2, Y_2), \dots$ without errors.

Theorem 2.2 *Assume that $W_{n,i}$ is a sequence of nonnegative weights such that the corresponding estimate F_n from (1) is strongly pointwise consistent for all distributions of (X, Y) and all independent and identically as (X, Y) distributed random vectors $(X_1, Y_1), (X_2, Y_2), \dots$ in the sense that for all $y \in \mathbb{R}$ (7) holds. Then (A1), (A2) and*

$$(A3^*) \quad \sum_{i=1}^n W_{n,i}(x)^2 \xrightarrow{P} 0 \text{ for } P_X\text{-almost every } x \in \mathbb{R}^d,$$

which is a weaker version of (A3), are fulfilled.

2.2 Rate of convergence

Next, we investigate the rate of convergence. Therefore, we assume that for a fixed $y_0 \in \mathbb{R}$ and P_X -almost all x the codf $F(y, x)$ is locally Hölder continuous in x with exponent $0 < p \leq 1$, locally uniform in y . More precisely, we assume that for P_X -almost every x there exist finite constants $C(x), \kappa_1(x), \kappa_2(x) > 0$ such that

$$\sup_{y \in \mathbb{R}: |y - y_0| \leq \kappa_1(x)} |F(y, x) - F(y, z)| \leq C(x) \cdot \|x - z\|^p \tag{10}$$

for all $z \in \mathbb{R}^d$ with $\|z - x\| \leq \kappa_2(x)$. In the following result we present conditions on the weights, which ensure a certain pointwise rate of convergence in probability for \bar{F}_n .

Theorem 2.3 *Let $(X, Y), (X_1, Y_1), (X_2, Y_2) \dots$ be i.i.d. $\mathbb{R}^d \times \mathbb{R}$ -valued random vectors and $y_0 \in \mathbb{R}$ be fixed. Furthermore, let the conditional distribution function F fulfill the Hölder-assumption of (10) in $y_0 \in \mathbb{R}$ for some $0 < p \leq 1$ and let $F(\cdot, x)$ be continuous and differentiable at y_0 for P_X -almost every x . Furthermore, let $a_n(x), b_n(x)$ and $c_n(x)$ for every $x \in \mathbb{R}^d$ be real and positive sequences, which tend to zero as $n \rightarrow \infty$ for P_X -almost every $x \in \mathbb{R}^d$. Let $W_{n,i}(x) := W_{n,i}(x, X_1, \dots, X_n)$ be nonnegative weights, which fulfill for P_X -almost every $x \in \mathbb{R}^d$*

- (C1) $|\sum_{i=1}^n W_{n,i}(x) - 1| = O_P(a_n(x)),$
- (C2) $\sum_{i=1}^n W_{n,i}(x) I_{\{\|X_i - x\| > b_n(x)^{1/p}\}} = O_P(b_n(x)),$
- (C3) $\sum_{i=1}^n W_{n,i}(x)^2 = O_P(c_n(x)^2)$

Furthermore let $\bar{Y}_{1,n}, \dots, \bar{Y}_{n,n}$ be random variables, which fulfill

$$\eta_n(x) := \sum_{i=1}^n W_{n,i}(x) \cdot |Y_i - \bar{Y}_{i,n}| \xrightarrow{P} 0 \text{ for } P_X\text{-almost every } x \tag{11}$$

and let \bar{F}_n be the local averaging estimate defined in (6) with weights $W_{n,i}$. Then for \mathbf{P}_X -almost every $x \in \mathbb{R}^d$

$$|\bar{F}_n(y_0, x) - F(y_0, x)| = O_{\mathbf{P}}\left(a_n(x) + b_n(x) + c_n(x) + \sqrt{\eta_n(x)}\right). \tag{12}$$

It can be shown that the conditions from 2.3 are fulfilled by the kernel estimate for some appropriate sequences a_n, b_n and c_n , which leads to the following corollary, which will be proven in Section S2 in the supplemental material.

Corollary 2.2 *Let $(X, Y), (X_1, Y_1), (X_2, Y_2) \dots$ be i.i.d. $\mathbb{R}^d \times \mathbb{R}$ -valued random vectors and let $y_0 \in \mathbb{R}$ be fixed. Assume that the conditional distribution function F fulfills the assumptions of Theorem 3. Assume furthermore that K is the naive kernel and that the bandwidth $h_n > 0$ fulfills*

$$h_n \rightarrow 0 \text{ and } n \cdot h_n^d \rightarrow \infty \text{ (} n \rightarrow \infty \text{)}. \tag{13}$$

Furthermore, let $\bar{Y}_{1,n}, \dots, \bar{Y}_{n,n}$ be random variables, which fulfill

$$\eta_n(x) := \frac{\sum_{i=1}^n |Y_i - \bar{Y}_{i,n}| \cdot K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} \xrightarrow{\mathbf{P}} 0 \text{ for } \mathbf{P}_X\text{-almost every } x. \tag{14}$$

Let \bar{F}_n be the kernel estimate of the codf with kernel K and bandwidth h_n as defined in (6). Then for \mathbf{P}_X -almost every x

$$|\bar{F}_n(y_0, x) - F(y_0, x)| = O_{\mathbf{P}}\left(\sqrt{\frac{1}{n \cdot h_n^d}} + h_n^p + \sqrt{\eta_n(x)}\right). \tag{15}$$

In particular, the choice of $h_n = \tilde{c} \cdot \left(\frac{1}{n}\right)^{\frac{1}{2p+d}}$ leads to

$$|\bar{F}_n(y_0, x) - F(y_0, x)| = O_{\mathbf{P}}\left(\left(\frac{1}{n}\right)^{\frac{p}{2p+d}} + \sqrt{\eta_n(x)}\right) \tag{16}$$

for \mathbf{P}_X -almost every x .

Remark 2.2 The rate $\left(\frac{1}{n}\right)^{\frac{p}{2p+d}} + \sqrt{\eta_n(x)}$ can also be achieved by choosing a sufficient number of the nearest neighbors and special cubic partitions for the weights of the nearest neighbor and partitioning estimate of the conditional distribution function.

3 Application to simulated and real data

In this section we apply the above described methods to simulated and real data and estimate the codfs. Therefore we choose $W_{n,i}$ as kernel weights with naive kernel. The

bandwidth h_n is chosen data-dependent from the set $\{0.05, 0.1, 0.2, 0.3\}$ by cross-validation w.r.t. the corresponding regression estimate

$$\hat{m}(x) = \sum_{i=1}^n W_{n,i}(x) \bar{Y}_{i,n}$$

[cf. Section 8 in Györfi et al. (2002)]. More precisely, we try to find a $\hat{h} \in \{0.05, 0.1, 0.2, 0.3\}$ that minimizes

$$\frac{1}{n} \sum_{j=1}^n \left| \hat{m}_{\hat{h},-j}(X_j) - \bar{Y}_{j,n} \right|^2,$$

where $\hat{m}_{\hat{h},-j}$ is the above mentioned regression estimate with kernel weights and bandwidth \hat{h} corresponding to all n data points with additional errors in the dependent variable omitting $(X_j, \bar{Y}_{j,n})$. In order to get an impression regarding the convergence of our estimates, we firstly consider distributions with known codfs, afterwards we will apply our estimator to real data in the context of experimental fatigue tests. The latter estimate is then utilized as the objective in a shape optimization procedure, which is embedded in an algorithm-based product development approach to determine an optimal profile geometry with respect to the fatigue behavior.

3.1 Application to simulated data

Motivated by the application in the context of experimental fatigue tests, where we have 1222 data points (see Sect. 3.2), we will consider sample sizes of $n = 500, 1000$ and 2000 in order to classify our estimate of the codf. The goodness of our estimate of the codf will be assessed by the maximum absolute error

$$err_{max} := \max_{i=1,\dots,I} \max_{j=1,\dots,J} \left| \hat{F}_n(y_i, x_j) - F(y_i, x_j) \right|$$

on a grid that is determined by equidistant y_1, \dots, y_I and x_1, \dots, x_J for some fixed numbers $I, J \in N$. Due to the random number generation in our simulated data, our estimates of the codf contain randomness, therefore we repeat the codf estimation 100 times with new random numbers and subscript our maximum absolute errors by an upper index i . We will compare our estimates by considering the average value $\frac{1}{100} \sum_{i=1}^{100} err_{max}^i$ of the maximum absolute error.

As a first example we choose $(X, Y), (X_1, Y_1), (X_2, Y_2), \dots$ as independent and identically distributed random vectors such that X is F -distributed with 5 numerator and 2 denominator degrees of freedom and Y is normal-distributed with mean $X \cdot (X - 1)$ and variance 1. As data with errors we set $\bar{Y}_{i,n} = Y_i + \frac{100}{n}$. Observe that we get completely new samples, when n changes. As a comparison to that we also consider $\bar{Z}_{i,n} = Y_i + \frac{100}{i}$, where the samples with bigger errors are kept by. The values x_1, \dots, x_{20} and y_1, \dots, y_{20} for the grid for the maximum absolute error are

Table 1 Average value of err_{max} for $F_{Y,n}$, $F_{\bar{Y},n}$, $F_{\bar{Z},n}$ in the first example

Size of n	500	1000	2000
Average value of err_{max} for $F_{Y,n}$	0.1587	0.1180	0.0989
Average value of err_{max} for $F_{\bar{Y},n}$	0.2168	0.1456	0.1139
Average value of err_{max} for $F_{\bar{Z},n}$	0.3821	0.2446	0.1733

Table 2 Average value of err_{max} for $F_{Y,n}$, $F_{\bar{Y},n}$ in the second example

Size of n	500	1000	2000
Average value of err_{max} for $F_{Y,n}$	0.1741	0.1456	0.1328
Average value of err_{max} for $F_{\bar{Y},n}$	0.2293	0.1734	0.1471

chosen equidistantly on $[0, 5]$ and $[-1, 2]$, respectively. Corollary 2.1 implies for all $y \in \mathbb{R}$ that $F_{Y,n}(y, x)$, $F_{\bar{Y},n}(y, x)$ and $F_{\bar{Z},n}(y, x)$ are pointwise consistent estimates for $[F(y^-, x), F(y, x)]$ for \mathbf{P}_X -almost all $x \in \mathbb{R}$. Due to the considered setting this interval does in fact only consist of one point that is equal to $F(y, x)$. The mentioned result of Corollary 2.1 is confirmed by the average values of the maximum absolute error in Table 1. Due to the fact that the samples with bigger errors are kept by, the estimator $F_{\bar{Y},n}$ yields smaller average squared errors than the estimator $F_{\bar{Z},n}$, in particular for the small sample sizes n .

As a second example we choose $(X, Y), (X_1, Y_1), (X_2, Y_2), \dots$ as independent and identically distributed random vectors such that X is $t(5)$ -distributed and Y is exponentially distributed with mean $\sqrt{|X|}$. As data with errors we choose $\bar{Y}_{i,n} = Y_i + U_{i,n}$, where $U_{1,n}, \dots, U_{n,n}$ are independent and uniformly on $(0, 100/n)$ -distributed, which are also independent of $(X_1, Y_1), \dots, (X_n, Y_n)$. The values x_1, \dots, x_{20} and y_1, \dots, y_{20} for the grid for the maximum absolute error are chosen equidistantly on $[-1, 1]$ and $[0.5, 1]$. Again, we can conclude from Corollary 2.1 for all $y \in \mathbb{R}$ the pointwise consistency of $\hat{F}_{\bar{Y},n}$, which is confirmed by the average maximum errors in Table 2.

As a third example we choose $(X, Y), (X_1, Y_1), (X_2, Y_2), \dots$ as independent and identically distributed random vectors with a discrete covariate X which is uniformly distributed on $\{1, 2, 3, 4, 5\}$. The dependend variable Y is chosen as χ^2 -distributed with X degrees of freedom. As data with errors we set $\bar{Y}_{i,n} = Y_i + \epsilon_{i,n}$ where $\epsilon_{1,n}, \dots, \epsilon_{n,n}$ are independent and normal-distributed with mean and variance $100/n$ and also independent of $(X_1, Y_1), \dots, (X_n, Y_n)$. As the covariate is discrete in this setting and the distance between the discrete values is one, the set $\{0.5, 1, 2\}$ is used for the choice of the bandwidth. A choice of $\hat{h} = 0.5$ would for example mean that for $x = j$ only samples of Y will be used in the estimate for which the corresponding samples of the covariate are equal to j . In this example the x -values for the grid for the maximum absolute error can be chosen as $x_j = j$ for $j = 1, 2, 3, 4, 5$. The values for y_1, \dots, y_{20} will be chosen equidistantly on $[0.5, 5]$. Again, we can conclude from Corollary 2.1

Table 3 Average value of err_{max} for $F_{Y,n}$ $F_{\bar{Y},n}$ in the third example

Size of n	500	1000	2000
Average value of err_{max} for $F_{Y,n}$	0.1082	0.0771	0.0528
Average value of err_{max} for $F_{\bar{Y},n}$	0.1812	0.1060	0.0639

for all $y \in \mathbb{R}$ the pointwise consistency of $\hat{F}_{\bar{Y},n}$, which is confirmed by the average maximum errors in Table 3.

3.2 Application to real data

In the following, we provide an application of the methods above in the context of shape optimization of steel profiles with respect to the fatigue behavior under cyclic loading.

From a practical point of view the robustness against cyclic loading is one of the major aspects to guarantee a long product lifetime, which results in an increased sustainability. Usually, the design process of such a profile geometry takes a lot of time and (human) resources, with no guarantee of the resulting profile geometry to be optimal. The new method described in this paper automates this process, which reduces the development effort and leads to a provable optimal profile design. For further details on the algorithm-based development process we refer to Roos et al. (2016) and Groche et al. (2017).

Our focus in the subsequent steps lies on the optimization of an integral sheet metal profile (made of the material HC 480 LA) with respect to the fatigue behavior. In particular we study a three-chambered profile, which is continuously produced in an integral way and can for example be used to separate oil, water and power supply. The used manufacturing technology is developed within the Collaborative Research Center 666 (CRC 666) at the Technische Universität Darmstadt. One main aspect of this technology is to produce those integral structures out of one part by linear flow and bend splitting. On the one hand this production technique requires less joining operations involving, for example, stress concentration or the action of heat and on the other hand the linear flow splitting leads to a ultrafine-grained microstructure (UFG) at the upper side of the steel flanges (cf., e.g., Bohn et al. (2008)). Both points yield significant advantages concerning the material properties, which can be utilized to produce lightweight structures with an improved fatigue life. Due to the significantly changed material properties in comparison to the material in as-received state, we model the linear flow split and bent parts of the structure as a different material.

In order to assess the fatigue behavior, we use data of experimental fatigue tests, in which a material sample gets repeatedly elongated by a fixed strain amplitude ε . The repetitions, the so called number of cycles N , until the material fails are counted and the corresponding stress amplitude τ is measured. Based on this data we estimate the conditional distribution function of the number of cycles $N^{(m)}$ until failure given a fixed strain amplitude ε for both of two considered materials m (as-received and linear flow split state). This estimation will be described in Section 3.1 in detail. Finally,

both estimates of the codf are evaluated at $N_{min} = 50,000$ in order to determine the approximate probability of a failure before N_{min} number of cycles for a fixed strain amplitude ε , which will be used as the objective of the optimization. Details on the shape optimization can be found in Sect. 3.2.

3.2.1 Estimation of the conditional distribution function

For the estimation of the codf we use a database that contains for each material m data

$$\left\{ \left(\varepsilon_1^{(m)}, \left(N_1^{(m)}, \tau_1^{(m)} \right) \right), \dots, \left(\varepsilon_{l_m}^{(m)}, \left(N_{l_m}^{(m)}, \tau_{l_m}^{(m)} \right) \right) \right\},$$

which was obtained by the above mentioned experimental fatigue tests and consists of the strain amplitude $\varepsilon_i^{(m)}$ the corresponding number of cycles $N_i^{(m)}$ until failure and the stress amplitude $\tau_i^{(m)}$. For each material m we have available a number of experimental fatigue tests l_m in a range from 4 to 35. Aggregated over all 132 studied materials the database includes 1222 of the above data points in total, i.e., we have

$$\sum_{m=1}^{132} l_m = 1222.$$

Since the experimental fatigue tests for obtaining one of the above data points are very time consuming, there are only 12 and 8 data points available for the considered material HC 480 LA in as-received and linear flow split state, respectively, which is not enough for a nonparametric estimation of the conditional distribution function. In order to nevertheless estimate the codf of the number of cycles until failure, we assume the model

$$N^{(m)}(\varepsilon) = \mu^{(m)}(\varepsilon) + \sigma^{(m)}(\varepsilon) \cdot \delta^{(m)}$$

to hold, where $\mu^{(m)}(\varepsilon)$ is the expected number of cycles until failure, $\sigma^{(m)}(\varepsilon)$ is the standard deviation for each material m and strain amplitude ε ; $\delta^{(m)}$ is an error term that has expectation 0 and variance 1 for each material m . We estimate the conditional distribution function of $\delta^{(m)}$ as well as $\mu^{(m)}(\varepsilon)$ and $\sigma^{(m)}(\varepsilon)$, so that we can obtain an estimate of the codf of $N^{(m)}$ by a simple linear transformation. For this purpose we use a similar approach as in Bott and Kohler (2017):

In order to obtain an estimate $\hat{\mu}^{(m)}(\varepsilon)$ of the expected number of cycles $\mu^{(m)}(\varepsilon)$, we apply a standard-method from the literature[(cf. Williams et al. (2003))], which uses the measured data of material m to estimate the coefficients $p = (\sigma'_f, \varepsilon'_f, b, c)$ of the strain life curve according to Coffin-Morrow-Manson(cf. Manson (1965))by linear regression and estimate $\mu^{(m)}(\varepsilon)$ from the corresponding strain life curve.

The estimation of the standard deviation $\sigma^{(m)}(\varepsilon)$ is more complicated, since we need to apply a nonparametric estimator to the squared deviations

$$Y_i^{(m)} = \left(N_i^{(m)} - \hat{\mu}_i^{(m)} \right)^2 \quad (i = 1, \dots, l_m)$$

for each material m , which usually needs more samples. So we augmented our data points per material m by 100 artificial ones as in Furer and Kohler (2015):

At first, we interpolate the squared deviations $Y_i^{(k)}$ for each material $k \neq m$ on a grid of 100 equidistant strain amplitudes ε . In order to generate an artificial data point at a fixed grid point, we also use interpolated values from materials, that are similar to the material m , assuming that similar materials yield similar fatigue behavior. Observe that we use the whole database consisting of 132 materials in order to obtain more interpolated values and to improve the statistical power of our estimation. The similarity is measured using 5 static material properties, namely Young’s modulus, the yield limit for 0.2% residual elongation, the tensile strength, the static strength coefficient and the static strain hardening exponent. In order to ensure that we only use interpolated values from those materials that have similar static material properties, we apply the Nadaraya-Watson kernel regression estimates with the static material properties as covariate and the interpolated data as dependent variable. In this way we obtain 100 artificial data points (one at each grid point) per material m . Finally, the estimation of the standard deviation $\sigma^{(m)}(\varepsilon)$ is done by weighting the Nadaraya-Watson kernel regression estimates applied to the real and the artificial data of the squared deviations as dependent variable and the corresponding ε -values as covariate.

Thus, we finally determine the data samples

$$\hat{\delta}_i^{(m)} = \frac{N_i^{(m)} - \hat{\mu}_i^{(m)}}{\hat{\sigma}_i^{(m)}} \quad \text{for } i = 1, \dots, l_m$$

of the random variables $\delta^{(m)}$ for each material m . Notice that these samples contain errors because we only estimated $\mu^{(m)}(\varepsilon)$ and $\sigma^{(m)}(\varepsilon)$. Since only 12 and 8 of the above data samples for the two material states of HC 480 LA are available, we also use data samples from other materials of the database, that have similar static material properties (with the same justification as above), in order to estimate the codf of $\delta^{(m)}$.

This consideration of similar materials in the estimation of the conditional distribution function is done by using the kernel estimate of the codf with the static material properties as covariate X_i and the data samples of $\delta^{(m)}$ for all 122 materials m as the dependent variable. The bandwidth h of the kernel weights is determined by a crossvalidation of the corresponding regression estimate as described in the beginning of this section.

As described in Sect. 1.2 it can be assumed that (E1) holds for the errors $\hat{\delta}_i^{(m)} - \delta_i^{(m)}$. Thus, evaluating the mentioned estimate of the codf at the static material properties $x = X^{(m)}$ of some material m leads to an estimate $\hat{G}_{\delta^{(m)}}$ of the codf of $\delta^{(m)}$ (see Theorem 2.2 for a theoretical justification). However, this estimate $\hat{G}_{\delta^{(m)}}$ can be transformed to an estimate $\hat{F}^{(m)}$ of the codf of $N^{(m)}$ given a strain amplitude ε by

$$\hat{F}_{N^{(m)}}(y, \varepsilon) = \hat{G}_{\delta^{(m)}}\left(\frac{y - \hat{\mu}^{(m)}(\varepsilon)}{\hat{\sigma}^{(m)}(\varepsilon)}\right). \tag{17}$$

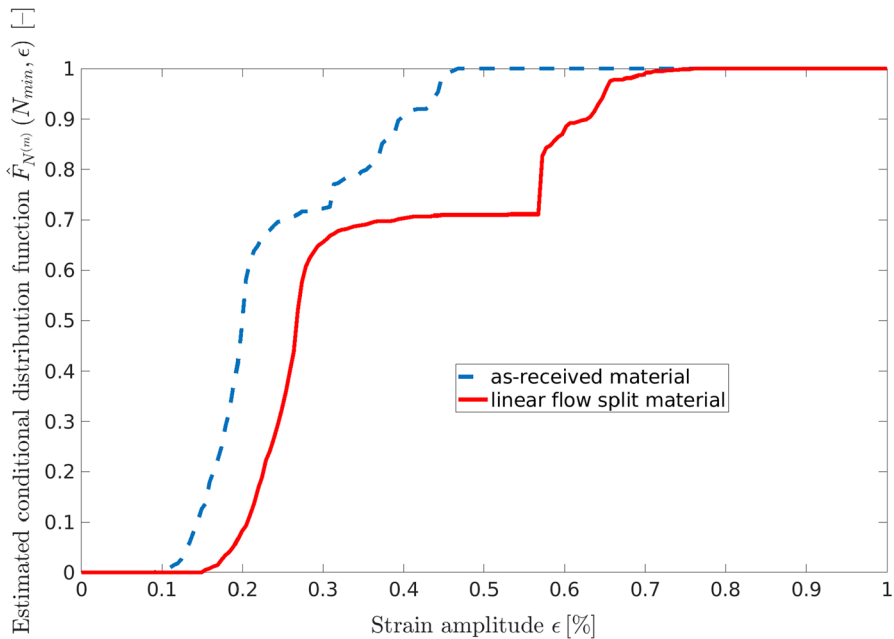


Fig. 1 Estimated conditional distribution function $\hat{F}_{N^{(m)}}(N_{min}, \epsilon)$ for $N_{min} = 50,000$ for the material HC 480 LA in as received and linear flow split state

This estimate of the conditional distribution function is evaluated at $y = N_{min} = 50,000$ numbers of cycles to obtain an approximate probability of a failure before N_{min} number of cycles for a fixed strain amplitude ϵ . In Fig. 1, we illustrated this estimated probability $\hat{F}_{N^{(m)}}(N_{min}, \epsilon)$ for the considered material HC 480 LA in as-received and linear flow split state and $\epsilon \in [0\%, 1\%]$. Here the strain amplitude ϵ is given proportional to the length of the material sample used in the experiments. As expected, $\hat{F}_{N^{(m)}}(N_{min}, \epsilon)$ is increasing in ϵ .

Since we also needed the derivative of $\hat{F}_{N^{(m)}}(N_{min}, \epsilon)$ w.r.t. ϵ , we interpolated the function $\hat{F}_{N^{(m)}}(N_{min}, \epsilon)$ by a piece-wise cubic smoothing spline, using 200 equidistant data points of ϵ and corresponding function values.

3.2.2 Fatigue Strength Shape Optimization

The former presented estimate of the failure probability $\hat{F}_{N^{(m)}}(N_{min}, \cdot)$ is in the following applied to the shape optimization of a multichambered profile. Our aim is to find the optimal geometry for a specific load scenario and a given starting geometry under certain design constraints, to reach minimal failure probability, as defined above. In order to calculate the failure probability of every point of the profile, we model the physical behavior of the considered geometry under applied loads at each point. For this purpose we describe the mechanical system in terms of the linear elasticity equations, for further details on the elasticity equations we refer to the supplement material Section S1. For numerical treatment of the elasticity equations are discretized in the

sense of *isogeometric analysis*. Thus, the discretized linear elasticity equations are denoted by

$$A_h(u)\mathbf{y}_h = b_h(u).$$

The discretization by methods of the isogeometric approach is explained in detail in the supplement material.

3.2.3 Shape Optimization

In this section we briefly describe the shape optimization problem governed by the linear elasticity problem as defined in the supplement material Section S1. The *finite dimensional shape optimization problem* can be written as

$$\begin{aligned} \min \quad & J_h(u, \mathbf{y}_h) \\ \text{s.t.} \quad & u \in U_{\text{ad}}, \quad A_h(u)\mathbf{y}_h = b_h(u). \end{aligned}$$

The design variables are denoted by $u \in \mathbb{R}^n$, where $n \in \mathbb{N}$ is the dimension of the design space. By $\mathbf{y}_h \in \mathbb{R}^{\tilde{n}}$ the displacement is described, which is determined by the linear elasticity equations $A_h(u)\mathbf{y}_h = b_h(u)$. The number of the control points of the isogeometric mesh is denoted by $\tilde{n} \in \mathbb{N}$. An introduction to shape optimization is given in Haslinger and Mäkinen (2003). Since the elasticity problem has a unique solution, we define a Lipschitz continuous solution operator $u \mapsto \mathbf{y}_h(u)$ such that the reduced form of the objective function can be written as

$$j_h(u) := J_h(u, \mathbf{y}_h(u)).$$

The corresponding shape gradient $g_h(u)$ can be efficiently determined by the adjoint approach as described in Hinze et al. (2009). The reduced shape optimization is stated as

$$\begin{aligned} \min \quad & j_h(u) \\ \text{s.t.} \quad & u \in U_{\text{ad}}. \end{aligned}$$

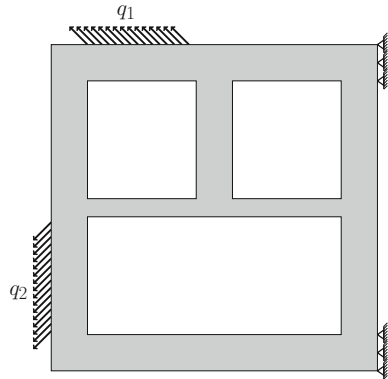
Here the set of admissible designs $U_{\text{ad}} \subset \mathbb{R}^n$ is defined by design constraints, for example angle or length restrictions.

In this work we use the *accumulated failure probability* as objective function

$$j_h(u) := \sum_{i \in \{1, \dots, \tilde{n}\}} \hat{F}_{N^{(m)}}(N_{\text{min}}, \bar{\varepsilon}([\mathbf{y}_h(u)]_i), u),$$

as defined above, see (17), with $N_{\text{min}} = 50,000$ fixed, where the sum is calculated over all control points (coefficients of the basis functions in \mathbf{y}_h above). If we define the failure of the whole profile by the failure of one of its parts, the accumulated failure probability over all parts is a (discretized) upper bound on the failure probability of

Fig. 2 Applied load scenario. For simplicity the bending radii are neglected in this draft. The right side of the profile is clamped at the bottom and the top. A uniform surface load $q_1 \in \mathbb{R}^3$, with $\|q_1\| = 70N$, is applied with 45° to the surface at the upper left part and a load $q_2 \in \mathbb{R}^3$, with $\|q_2\| = 70N$, is applied with 45° to the lower left side of the profile. The load scenario is constant in the third dimension



the whole profile and thus a reasonable objective function. It can be shown, that the objective function is nonconvex with respect to the design. The main principal strain $\bar{\varepsilon}$ can be determined by calculating the maximal eigenvalue with respect to the absolute value of the linearized strain tensor ε , by *Cardano’s formula*.

The optimization is done using a *sequential quadratic programming method (SQP)* [see, e.g., Nocedal and Wright (2006)].

3.2.4 Numerical Result

We apply the above described methods to perform a shape optimization of a three-chambered profile with respect to fatigue strength. Therefore, we assume a static load scenario, as shown in Fig. 2. The profile is clamped at the boundary on the right-hand side. Additionally, there are surface loads applied at the upper and lower left of the geometry. The loads act on the surface at an angle of 45° .

The geometry is modeled as a tricubic NURBS solid, with 25,920 degrees of freedom and 1350 elements. The outer dimensions are $50\text{ cm} \times 50\text{ cm} \times 2\text{ cm}$. To reduce the need of numerous additional constraints, we applied a parametrization with only twelve degrees of freedom. For this purpose, we subdivide the profile into four parts and determine the barycentric coordinates of each control point.

As constraints, we consider an upper bound on the total volume, and we fix the volume of the Neumann and Dirichlet boundaries. For technical reasons, we also add a minimal bound for the volume of each element to circumvent negative element volumes. After 58 iterations with 375 function evaluations, the SQP method found the solution depict in Fig. 3. The accumulated failure probability could be reduced about almost 53.58%. The used SQP method is the standard MATLAB R2018a implementation. Additionally, we compare the result to the optimization with respect to the *compliance*

$$j_h(u) := (\mathbf{y}_h(u))^\top (M_h^\Omega \mathbf{f}_h + M_h^\Gamma \mathbf{q}_h),$$

where \mathbf{f}_h and \mathbf{q}_h are the discretized volume force and surface load acting on the geometry, respectively, and M_h^Ω and M_h^Γ are the mass matrices of the interior Ω and

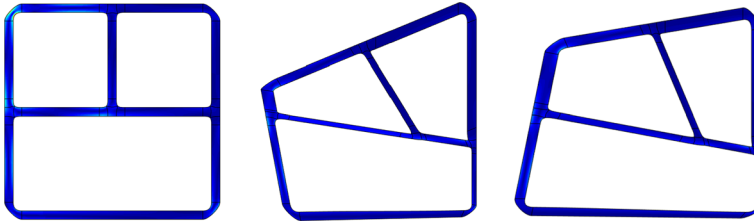


Fig. 3 Starting solution (left) compared to the optimal geometries with respect to the accumulated failure probability (middle) and the compliance (right). The color represents the von Mises stress in MPa. The displacement is neglected. The accumulated failure probabilities of the optimal solutions could be reduced about 53.58% (middle) or 37.82% (right), respectively

boundary Γ of the considered geometry, respectively. In this case the compliance could be reduced about 58.34% after 29 iterations and 106 function evaluations. The optimal solution is visualized in Fig. 3. The accumulated failure probability of this geometry is reduced about 37.82% compared to the starting solution. We see that in general the optimization of the accumulated failure probability can not be replaced by the classical compliance optimization. All the calculations are performed on an Intel Core i7-4790 CPU with 3.60 GHz and 16 GB RAM. The used software was Mathworks MATLAB R2018a running in single thread mode.

Acknowledgements The authors would like to thank the German Research Foundation (DFG) for funding this project within the Collaborative Research Centre 666. The authors would also like to thank an associate editor and a referee for their helpful comments.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Supplementary Materials The supplement contains additional information on the linear elasticity equations and the isogeometric approach complementing the information in Chapter 3.2 and furthermore the proofs of the Theorems and the Corollaries.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Algoet P, Györfi L (1999) Strong universal pointwise consistency of some regression function estimates. *J Multivar Anal* 71(1):125–144. <https://doi.org/10.1006/jmva.1999.1836>

- Bohn T, Bruder E, Müller C (2008) Formation of ultrafine-grained microstructure in hsla steel profiles by linear flow splitting. *J Mater Sci* 43:7307–7312
- Bott AK, Kohler M (2017) Nonparametric estimation of a conditional density. *Electr J Stat* 69:189–214. <https://doi.org/10.1214/13-EJS850>
- Bott AK, Devroye L, Kohler M (2013) Estimation of a distribution from data with small measurement errors. *Electr J Stat* 7:2457–2476. <https://doi.org/10.1214/13-EJS850>
- Cai Z (2002) Regression quantiles for time series. *Econ Theory* 18(1):169–192
- Devroye L (1981) On the almost everywhere convergence of nonparametric regression function estimates. *Ann Stat* 9(6):1310–1319. <https://doi.org/10.1214/aos/1176345647>
- Devroye L (1982) Necessary and sufficient conditions for the pointwise convergence of nearest neighbor regression function estimates. *Probab Theory Related Fields* 61(4):467–481
- Felber T, Kohler M, Krzyzak A (2015) Adaptive density estimation from data with small measurement errors. *IEEE Trans Inf Theory* 61(6):3446–3456. <https://doi.org/10.1109/TIT.2015.2421297>
- Fromkorth A, Kohler M (2011) Analysis of least squares regression estimates in case of additional errors in the variables. *J Stat Plan Inference* 141(1):172–188. <https://doi.org/10.1016/j.jspi.2010.05.031>
- Furer D, Kohler M (2015) Smoothing spline regression estimation based on real and artificial data. *Metrika* 78(6):711–746. <https://doi.org/10.1007/s00184-014-0524-6>
- Furer D, Kohler M, Krzyzak A (2013) Fixed-design regression estimation based on real and artificial data. *J Nonparametr Stat* 25(1):223–241. <https://doi.org/10.1080/10485252.2012.749257>
- Greblicki W, Krzyzak A, Pawlak M (1984) Distribution-free pointwise consistency of kernel regression estimate. *Ann Stat* 12(4):1570–1575. <https://doi.org/10.1214/aos/1176346815>
- Groche P, Bruder E, Gramlich S (2017) Manufacturing integrated design - sheet metal product and process innovation. Springer, Berlin
- Györfi L (1981a) Recent results on nonparametric regression estimate and multiple classification. *Probl Control Inf Theory* 10(1):43–52
- Györfi L (1981b) The rate of convergence of k_n -nn regression estimates and classification rules (corresp.). *IEEE Trans Inf Theory* 27(3):362–364. <https://doi.org/10.1109/TIT.1981.1056344>
- Györfi L, Kohler M, Krzyzak A, Walk H (2002) A distribution-free theory of nonparametric regression. Springer, New York
- Hall P, Yao Q (2005) Approximating conditional distribution functions using dimension reduction. *Ann Stat* 33(3):1404–1421. <https://doi.org/10.1214/009053604000001282>
- Hall P, Wolff RCL, Yao Q (1999) Methods for estimating a conditional distribution function. *J Am Stat Assoc* 94(445):154–163
- Hansmann M, Kohler M (2017) Estimation of quantiles from data with additional measurement errors. *Statistica Sinica* 27:1661–1673
- Hansmann M, Kohler M (2019) Estimation of conditional quantiles from data with additional measurement errors. *J Stat Plan Inference* 200:176–195. <https://doi.org/10.1016/j.jspi.2018.09.013>
- Hansmann M, Kohler M, Walk H (2019) On the strong universal consistency of local averaging regression estimates. *Ann Inst Stat Math* 71:1233–1263. <https://doi.org/10.1007/s10463-018-0674-9>
- Härdle W, Janssen P, Serfling R (1988) Strong uniform consistency rates for estimators of conditional functionals. *Ann Stat* 16(4):1428–1449. <https://doi.org/10.1214/aos/1176351047>
- Haslinger J, Mäkinen RAE (2003) Introduction to Shape Optimization. SIAM
- Hinze M, Pinnau R, Ulbrich M, Ulbrich S (2009) Optimization with PDE constraints. Springer, Berlin
- Kohler M (2006) Nonparametric regression with additional measurement errors in the dependent variable. *J Stat Plan Inference* 136(10):3339–3361. <https://doi.org/10.1016/j.jspi.2005.01.009>
- Krzyzak A, Pawlak M (1987) The pointwise rate of convergence of the kernel regression estimate. *J Stat Plan Inference* 16:159–166. [https://doi.org/10.1016/0378-3758\(87\)90065-6](https://doi.org/10.1016/0378-3758(87)90065-6)
- Liero H (1989) Strong uniform consistency of nonparametric regression function estimates. *Probab Theory Relat Fields* 82(4):587–614. <https://doi.org/10.1007/BF00341285>
- Manson SS (1965) Fatigue: a complex subject—some simple approximations. *Exp Mech* 5(7):193–226
- Meister A (2009) Deconvolution Problems in Nonparametric Statistics -, vol 193, 2009th edn. Lecture Notes in Statistics, vol. Springer, Berlin
- Nadaraya EA (1964) On estimating regression. *Theory Probab Appl* 9(1):141–142
- Nocedal J, Wright SJ (2006) Numerical optimization. Springer, Berlin
- Roos M, Horn BM, Gramlich S, Ulbrich S, Kloberdanz H (2016) Manufacturing integrated algorithm-based product design—case study of a snap-fit fastening. *Procedia CIRP* 50:123 – 128, 26th CIRP Design Conference

- Stone CJ (1977) Consistent nonparametric regression. *Ann Stat* 5(4):595–620. <https://doi.org/10.1214/aos/1176343886>
- Stone CJ (1982) Optimal global rates of convergence for nonparametric regression. *Ann Stat* 10(4):1040–1053. <https://doi.org/10.1214/aos/1176345969>
- Stute W (1986) On almost sure convergence of conditional empirical distribution functions. *Ann Probab* 14(3):891–901. <https://doi.org/10.1214/aop/1176992445>
- Veraverbeke N, Gijbels I, Omelka M (2014) Preadjusted non-parametric estimation of a conditional distribution function. *J R Stat Soc Ser B Stat Methodol* 76(2):399–438. <https://doi.org/10.1111/rssb.12041>
- Walk H (2001) Strong universal pointwise consistency of recursive regression estimates. *Ann Instit Stat Math* 53(4):691–707
- Watson GS (1964) Smooth regression analysis. *Sankhya: The Indian Journal of Statistics. Series A* 26:359–372
- Williams CR, Lee YL, Rilly JT (2003) A practical method for statistical analysis of strain-life fatigue data. *Intl J Fatigue* 25(5):427–436

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.