Estimation of the optimal design of a nonlinear parametric regression problem via Monte Carlo experiments

Ida Hertel* and Michael Kohler

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

March 27, 2012

Abstract

A Monte Carlo method for estimation of the optimal design of a nonlinear parametric regression problem is presented. The basic idea is to produce via Monte Carlo values of the error of a parametric regression estimate for randomly chosen designs and randomly chosen parameters and to use nonparametric regression to estimate from this data the design for which the maximal expected error with respect to all possible parameter values is minimal. A theoretical result concerning consistency of this estimate of the optimal design is presented, and the method is used to find an optimal design for an experimental fatigue test.

AMS classification: Primary 62K05; secondary 62P30.

Key words and phrases: Optimal design, Monte Carlo, nonparametric regression, consistency.

1 Introduction

Fatigue behaviour of materials can be described, e.g., by curves relating strain amplitudes and number of cycles till failure to each other. In corresponding experiments, for a given

*Corresponding author. Tel: +49-6151-16-2294, Fax: +49-6151-16-6822

Running title: *Estimation of the optimal design* strain amplitude the number of cycles till failure is observed which is rather time consuming since usually strain amplitudes are used in such a region that the corresponding number of cycles achieves values up to 10^7. For an efficient estimation of such curves it is necessary to choose the used strain amplitudes (usually between 7 and 15 for one material) carefully.

Mathematically, this can be considered as a problem of determing the optimal design of a fixed design regression problem. A parametric model in this context is given by the Manson-Coffin-Basquin relation (cf., e.g., Manson (1965))

$$\epsilon = \frac{\sigma_f'}{E} \cdot (2N_f)^b + \epsilon_f' \cdot (2N_f)^c, \tag{1}$$

which describes the dependency of the strain amplitude ϵ on the number N_f of cycles till failure. Here σ_f', b, ϵ_f' and c are cyclic material properties which characterize the fatigue behaviour of the material and E is a usually known parameter of the material. Since b and c are less than zero, the monotone function (1) has a well-defined inverse function, and it is this nonlinear model for the inverse function which describes the experiment where N is observed for given ϵ.

The purpose of this article is to develop a methodology which can determine values to be used in a sequence of experiments for the strain amplitudes such that by using the observed numbers till failure the above model can be estimated efficiently. Since this model is nonlinear and is given only implicitly, it seems to be difficult to determine these quantities theoretically. Instead we propose a simulation approach, which we fit into a more general framework introduced next.

We assume that we are interested in estimation of a function r_p, where $p \in \mathcal{P}$ is some unknown parameter from a given set of parameters $\mathcal{P}$. To do this, we have to choose for fixed $N \in \mathbb{N}$ a design

$$z = (z^{(1)}, \ldots, z^{(N)}) \in D^N$$

consisting of points $z^{(i)}$ from some given set D of possible design points. For this design we generate a data set

$$\mathcal{D}_N(z; p),$$

which we use to calculate an estimate

$$\hat{r}_N(\cdot, \mathcal{D}_N(z; p))$$

of r_p. Its error is denoted by

$$Err\left(\hat{r}_N\left(\cdot, \mathcal{D}_N\left(z;p\right)\right), r_p\right)$$

with expected value

$$\mathbf{E}\left\{Err\left(\hat{r}_N\left(\cdot, \mathcal{D}_N\left(z;p\right)\right), r_p\right)\right\}, \tag{2}$$

where the expectation is computed with respect to the data set $\mathcal{D}_N\left(z;p\right)$.

Our aim is to choose the design $z = (z^{(1)}, \ldots, z^{(N)}) \in D^N$ such that the maximal expected error

$$\max_{p\in\mathcal{P}} \mathbf{E}\left\{Err\left(\hat{r}_N\left(\cdot, \mathcal{D}_N\left(z;p\right)\right), r_p\right)\right\}$$

is as small as possible, i.e., we want to find a design $z = (z^{(1)}, \ldots, z^{(N)}) \in D^N$ such that

$$\max_{p\in\mathcal{P}} \mathbf{E}\left\{Err\left(\hat{r}_N\left(\cdot, \mathcal{D}_N\left(z;p\right)\right), r_p\right)\right\} \approx \inf_{u\in D^N} \max_{p\in\mathcal{P}} \mathbf{E}\left\{Err\left(\hat{r}_N\left(\cdot, \mathcal{D}_N\left(u;p\right)\right), r_p\right)\right\}.$$

To do this, we use a simulation approach. Here we assume that we know how to construct the data set for given design points and given parameter. The basic idea is to determine data points for a random design regression problem (where the independent variable consists of a randomly chosen design from D^N and a randomly chosen parameter from $\mathcal{P}$) in such a way that the corresponding regression function is the expected error in (2). Then we estimate this function by applying a nonparametric regression estimate to this data, and choose our estimated design such that for this design the maximal value of the estimate with respect to the parameter is as small as possible.

Our main result is that under some regularity conditions our estimated design is consistent in the sense that for this design the expected error indeed converges to the minimal possible value provided the sample size of the data used in estimation of this design converges to infinity. Furthermore we illustrate our methodology by applying it in the context of an experimental fatigue test.

1.1 Discussion of related results

Various criteria of classical theory of optimal experimental design can be found in Kiefer (1961), Atkinson (1982) and Pukelsheim (1993), where several methods and tools in analytical computation of optimal designs for a given linear model are introduced. In case of nonlinearity basically three ways to deal with this can be pointed out. Atkinson (1982) proposes a linearisation by expanding the model in a Taylor's series about a preliminary estimate of the parameter p_0 and additionally seeks a D-optimum design. As described by Imhof (2001) in case of a nonlinear model most tools of experimental design cannot be applied since the information matrix depends on unknown parameters. The two strategies to deal with this according to Imhof (2001) is to use a Bayesian or a minimax approach. The latter is based on the idea of choosing the best design for the worst parameter value p, cf., e.g., in Silvey (1980), where the uncertainty of p is treated by seeking the best design for the worst possible parameter value. For the same reason Imhof (2001) uses a standardized maximin D-optimal design in context of a exponential growth model first described in Dette (1997). Since the problem considered in this article is given implicit by the nonlinear inverse function of the underlying model we propose a minimax strategy in context of a simulation approach via Monte Carlo experiments. In this context Müller and Parmigiani (1996) pursue the idea of implementing stochastic optimization by curve fitting of Monte Carlo samples in connection with Bayesian design problems. The target is to find an optimal design by maximation of an expected utility function whose evaluation requires integration. Instead of this it reveals a curve fitting problem of a Monte Carlo sample.

Nonparametric regression estimation has been studied over many years. Two main approaches were developed: random design approach and fixed design approach. The most popular estimates for random design approach include kernel regression estimate (cf., e.g., Nadaraya (1964, 1970), Watson (1964), Devroye and Wagner (1980), Stone (1977) or Devroye and Krzyżak (1989)), partitioning regression estimate (cf., e.g., Györfi (1981) or Beirlant and Györfi (1998)), nearest neighbor regression estimate (cf., e.g., Devroye (1982), Devroye, Györfi, Krzyżak and Lugosi (1994), Mack (1981) or Zhao (1987)), least squares estimates (cf., e.g., Lugosi and Zeger (1995)) or smoothing spline estimates (cf., e.g., Kohler and Krzyżak (2001)). The main theoretical results are summarized in the monograph Györfi et al. (2002). For a survey of fixed design regression estimates we refer to the monograph Eubank (1999).

The proof of the consistency of our methodology is based on a consistency result in supremum norm for our nonparametric regression estimate. Various such results can be found, e.g., in Devroye (1978a), Devroye (1978b) and Härdle and Luckhaus (1984). In our theoretical result we use techniques from Kohler, Krzyżak and Walk (2011) in order to be able to weaken the conditions on the design distribution slightly more than in Devroye (1978a). Here the trick is that we increase the bandwidth of the kernel estimate if there are not enough data points in a ball around the point where we want to evaluate the estimate. As a consequence, we do not have to worry about the form of the parameter space we sample from, e.g., it is allowed that this space contains some sharp edges etc.

1.2 Notation

Throughout this paper we use the following notations: $\|x\|$ denotes the Euclidean norm of $x \in \mathbb{R}^d$, μ denotes the distribution of X, $supp(\mu)$ denotes the support of μ, $m(x) = \mathbf{E}\{Y|X=x\}$ is the regression function of (X,Y) and 1_B is the indicator of the set B. For $f : D \to \mathbb{R}$ we write

$$x = \arg\min_{z \in D} f(z)$$

in case that

$$x \in D \quad \text{and} \quad f(x) = \min_{z \in D} f(z).$$

1.3 Outline

This paper is organized as follows. The precise definition of the estimate is given in Section 2 and the main result is formulated in Section 3. Section 4 contains an application of our methodology to an experimental fatigue test. The proofs are given in Section 5.

2 Definition of the estimate of the optimal design

In the sequel we assume that we have given the number $N \in \mathbb{N}$ of design points, a compact set $D \subseteq \mathbb{R}^{d_z}$ from which we have to choose the design points, and a compact set $\mathcal{P} \subseteq \mathbb{R}^{d_p}$ of possible parameters, where for each $p \in \mathcal{P}$ a function r_p is given, which has to be estimated. For a given design $z \in D^N$ and a given parameter $p \in \mathcal{P}$ we asume that we can compute a data set $\mathcal{D}_N(z;p)$, an estimate

$$\hat{r}_N(\cdot, \mathcal{D}_N(z;p))$$

of r_p and its error

$$Err\left(\hat{r}_N(\cdot, \mathcal{D}_N(z;p)), r_p\right) \geq 0.$$

Throughout this we make the following assumption:

(A1) The nonnegative function

$$(z,p) \mapsto Err\left(\hat{r}_N(\cdot, \mathcal{D}_N(z;p)), r_p\right)$$

defined on $D^N \times \mathcal{P} \subseteq \mathbb{R}^{d_z \cdot N} \times \mathbb{R}^{d_p}$ is measurable with respect to the Borel sigma-algebra.

In order to find the optimal design, for which the maximal (in view of the parameter) expected error is minimal, we choose in a first step $n \in \mathbb{N}$ and random design points $Z_1, \ldots, Z_n \in D^N$ and random parameters $P_1, \ldots, P_n \in \mathcal{P}$. Here we assume that

(A2) $Z_1, \ldots, Z_n$ are uniformly distributed on $\mathcal{D}^N$,

(A3) $P_1, \ldots, P_n$ are uniformly distributed on $\mathcal{P}$,

(A4) $Z_1, P_1, \ldots, Z_n, P_n$ are independent,

(A5) $Z_1, P_1, \ldots, Z_n, P_n$ are independent from the data $\mathcal{D}_N(z;p)$ for all $z \in \mathcal{D}^N$ and $p \in \mathcal{P}$.

In a second step we construct for each $X_i = (Z_i, P_i)$ data sets

$$\mathcal{D}_N(X_i) = \mathcal{D}_N(Z_i; P_i),$$

where for values $z_i = Z_i(\omega)$ and $p_i = P_i(\omega)$ of Z_i and P_i the data set $\mathcal{D}_N(Z_i; P_i)$ is given by $\mathcal{D}_N(z;p)$. Then we use this data set to compute $\hat{r}_N(\cdot, \mathcal{D}_N(X_i))$ and denote its error by

$$Y_i = Err\left(\hat{r}_N(\cdot, \mathcal{D}_N(X_i)), r_{P_i}\right) = Err\left(\hat{r}_N(\cdot, \mathcal{D}_N((Z_i; P_i))), r_{P_i}\right).$$

In a third step we use nonparametric regression to estimate for $x = (z,p) \in \mathcal{D}^N \times \mathcal{P}$

$$m(x) := m(z,p) := \mathbf{E}\left\{Err\left(\hat{r}_N(\cdot, \mathcal{D}_N((z;p))), r_p\right)\right\}. \quad (3)$$

Because of $(A1)$ and $(A5)$, which imply

$$\mathbf{E}\left\{Err\left(\hat{r}_N(\cdot, \mathcal{D}_N((Z_1; P_1))), r_{P_1}\right) \middle| (Z_1, P_1) = (z,p)\right\}$$

$$= \mathbf{E}\left\{Err\left(\hat{r}_N(\cdot, \mathcal{D}_N((z;p))), r_p\right)\right\},$$

the above term can be written as conditional expectation via

$$m(x) = m(z,p) = \mathbf{E}\{Y_1|(Z_1,P_1) = (z,p)\} = \mathbf{E}\{Y_1|X_1 = x\}.$$

To estimate m, we use the data

$$(X_1,Y_1),\dots,(X_n,Y_n)$$

to compute the so-called Nadaraya-Watson kernel regression estimate (cf., e.g., Nadaraya (1964) and Watson (1964))

$$m_n(x) = \frac{\sum_{i=1}^n Y_i \cdot K\left(\frac{x-X_i}{\hat{h}_x}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{\hat{h}_x}\right)}.$$

Here $K : \mathbb{R}^{N \cdot d_z + d_p} \to \mathbb{R}$ is a so-called kernel function (e.g., $K(u) = 1_{S_1(0)}$, where $S_r(z)$ denotes the (closed) ball of radius r around z in a Euclidean space) and $\hat{h}_x$ is the bandwidth of the kernel. We define the latter one depending on x and the data in such a way that the ball around x with the radius given as the bandwidth contains at least a special number of data points. More precisely, we choose $r, h_n > 0$ and set

$$\hat{h}_x = \min\left\{h \geq h_n \;:\; \mu_n(S_{r \cdot h}) \geq \frac{\log n}{n^{1/4}}\right\},$$

where

$$\mu_n(A) = \frac{1}{n}\sum_{i=1}^n 1_A(X_i)$$

is the empirical measure of $A \subseteq \mathbb{R}^{N \cdot d_z + d_p}$ corresponding to $X_1,\dots,X_n$.

With the notation introduced above we can reformulate the aim of our procedure in the following way: Our goal is to find a design $(\hat{z}_1,\dots,\hat{z}_N) \in D^N$ such that

$$\max_{p \in \mathcal{P}} m\left((\hat{z}_1,\dots,\hat{z}_N), p\right) \approx \inf_{(z_1,\dots,z_N) \in D^N} \max_{p \in \mathcal{P}} m\left((z_1,\dots,z_N), p\right).$$

In the fourth and last step we define our estimate of the optimal design by

$$\left(\hat{z}^{(1)},\dots,\hat{z}^{(N)}\right) = \arg \min_{\left(\hat{u}^{(1)},\dots,\hat{u}^{(N)}\right) \in D^N} \max_{p \in \mathcal{P}} m_n\left(\left(\left(\hat{u}^{(1)},\dots,\hat{u}^{(N)}\right), p\right)\right). \qquad (4)$$

3 Main result

Our main result is the following theorem.

Theorem 1. *Assume that $D \subseteq \mathbb{R}^{d_z}$ and $\mathcal{P} \subseteq \mathbb{R}^{d_p}$ are compact sets, that the data is generated as in Section 2 and that the estimate is defined as in Section 2. Assume furthermore that (A1), ..., (A5) hold, that m defined by (3) is continuous, and that for some $L > 0$ we have with probability one*

$$0 \leq Err(\hat{r}_N(\cdot, \mathcal{D}_N(z;p)), r_p) \leq L \tag{5}$$

for all $z \in D^N$ and $p \in \mathcal{P}$. Let $\tilde{K} : \mathbb{R}_+ \to \mathbb{R}_+$ be a monotonically decreasing and left continuous function satisfying for some $R > 0$

$$\tilde{K}(+0) > 0 \quad and \quad \tilde{K}(t) = 0 \quad for\ t > R,$$

and define the kernel $K : \mathbb{R}^d \to \mathbb{R}_+$ by

$$K(u) = \tilde{K}(\|u\|) \quad (u \in \mathbb{R}^d).$$

Let $r > 0$ be such that $\tilde{K}(r) > 0$ and let the bandwidth $\hat{h}_x$ be defined as in Section 2 for some $h_n > 0$, $n \in \mathbb{N}$ satisfying

$$h_n \to 0 \quad (n \to \infty).$$

Then with probability one

$$\max_{p \in \mathcal{P}} m\left((\hat{z}_1, \ldots, \hat{z}_N), p\right) \to \inf_{(z_1,\ldots,z_N) \in D^N} \max_{p \in \mathcal{P}} m\left((z_1, \ldots, z_N), p\right)$$

as n tends to infinity.

Remark 1. In the above definition of the estimate we assume for notational simplicity that the minimum and the maximum in (4) exist. In case that they do not exist, it suffices to choose values which have distance less than $\delta_n > 0$ from the corresponding infimum and supremum, resp., where $\delta_n \to 0$ $(n \to \infty)$. It follows from the proof of Theorem 1 that in this case the assertion of Theorem 1 is still valid.

Remark 2. It follows from the proof of Theorem 1 and the consistency result in Devroye (1978a) that in case that for every ball in $\mathbb{R}^{d_z \cdot N} \times \mathbb{R}^{d_p}$ with center in $D^N \times \mathcal{P}$ the volume of this ball intersected with $D^N \times \mathcal{P}$ divided by the volume of the ball is greater than or equal to some constant $c > 0$, the above consistency result is still valid if we choose h_n instead of $\hat{h}_x$ as bandwidth of our kernel regression estimate, provided h_n satisfies

$$\frac{n \cdot h_n^d}{\log n} \to \infty \quad (n \to \infty).$$

4 Application to experimental fatigue tests

In this section we will apply the method described in Section 2 in the context of fatigue analysis, where the fatigue behaviour under cyclic loading is studied. Table 1 contains observed data from a strain controlled fatigue test, where a constant strain amplitude ϵ is adjusted and the corresponding stress amplitude σ and the corresponding number N_f of cycles till failure are observed. The aim of such an experiment is to estimate the so-called

ϵ	0.003	0.0035	0.004	0.004	0.0045	0.005	0.005
N_f	28572	8077	7878	2919	2950	1865	4015
σ	402.9	437.2	426.1	434.3	456.6	475.3	447.1

Table 1: Observed values in experimental fatigue tests.

strain life curve and the so-called stress strain curve describing the relation between the total strain amplitude ϵ and the number N_f of cycles till failure and between the total strain amplitude ϵ and the stress amplitude σ, respectively.

Based on Morrows proposal (cf., e.g., Manson (1965)) a parametric model for the strain life curve is given by

$$\epsilon = \frac{\sigma'_f}{E} \cdot (2N_f)^b + \epsilon'_f \cdot (2N_f)^c. \tag{6}$$

Here parameters of the strain life curve are given by the fatigue strength coefficient σ'_f, the fatigue strength exponent b, the fatigue ductility coefficient ϵ'_f and the fatigue ductility exponent c. E denotes the modulus of elasticity which can be assumed to be known from a previously performed static tensile material test.

A parametric model for the stress strain curve is given by the so-called Ramberg-Osgood relation (cf., e.g., Ramberg and Osgood (1943))

$$\epsilon = \frac{\sigma}{E} + (\frac{\sigma}{K'})^{\frac{1}{n'}}, \tag{7}$$

where K' and n' are further cyclic material parameters. Here both relations can be interpreted as the sum of an elastic and a plastic strain part, more precisely

$$\epsilon_{el} + \epsilon_{pl} = r_p(N_f) := \frac{\sigma'_f}{E} \cdot (2N_f)^b + \epsilon'_f \cdot (2N_f)^c$$

and

$$\epsilon_{el} + \epsilon_{pl} := f^{ss}_{(K',n')}(\sigma) = \frac{\sigma}{E} + (\frac{\sigma}{K'})^{\frac{1}{n'}}.$$

Williams, Lee and Rilly (2002) propose to find the cyclic parameter $p = (\sigma'_f, \epsilon'_f, b, c)$ by two separate linear regressions, where data consisting of triples (ϵ, N_f, σ) are used. This very common standard estimation method makes use of the condition of compatibility, which means that the elastic part and respectively the plastic part of the strain life curve equals the elastic part and respectively the plastic part of the stress strain curve. Now the idea is to determine for given data $(\epsilon_1, N_{f1}, \sigma_1), \cdots, (\epsilon_n, N_{fn}, \sigma_n)$ the elastic and plastic part of the strain amplitudes using relation (7)

$$\epsilon_{eli} = \frac{\sigma_i}{E} \quad \text{and} \quad \epsilon_{pl_i} = \epsilon_i - \epsilon_{eli} \quad (i = 1, \cdots, n)$$

and additionally to find the cyclic parameter p by separately fitting two linear functions based on the data $(\epsilon_{el1}, N_{f1}), \cdots, (\epsilon_{eln}, N_{fn})$ and $(\epsilon_{pl_1}, N_{f1}), \cdots, (\epsilon_{pl_n}, N_{fn})$, respectively. For example in case of the elastic part a classical linear least squares estimate is fitted to the data via

$$({\sigma'_f}^*, b^*) = \arg \min_{\sigma'_f, b \in (0,\infty)\times(-\infty,0)} \frac{1}{n} \sum_{i=1}^{n} \left| \log(2N_{fi}) - \frac{1}{b} \cdot (\log(\epsilon_{eli}) - \log(\frac{\sigma'_f}{E})) \right|^2.$$

In case of the plastic part one proceeds similarly, but here just data which satisfy $\epsilon_{pl} > 1 \cdot 10^{-6}$ are used in the regression.

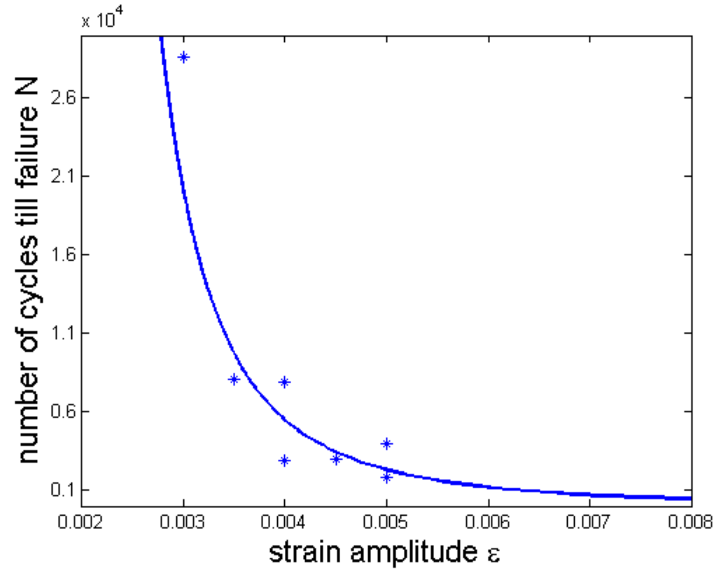
Application of this estimation method to the data of Table 1 yields the estimate

$$p_{ref} = [788.305, 0.118, -0.0631, -0.4427]$$

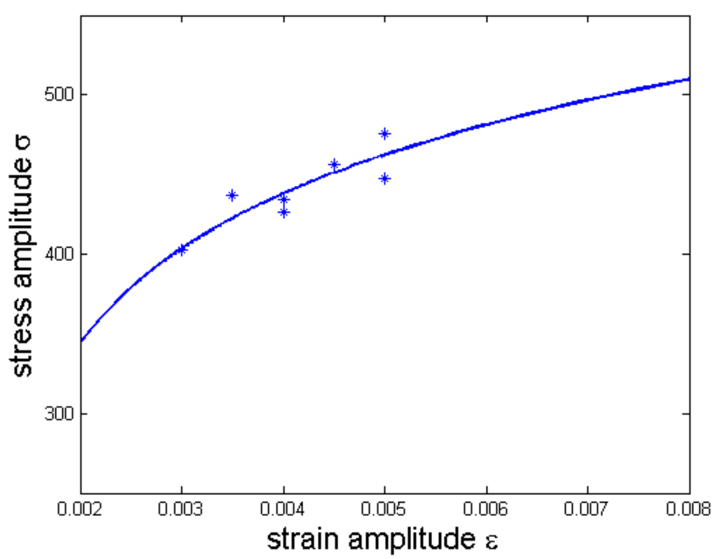
for the cyclic parameters of the above parametric model. Here the modulus of elasticity $E = 210000$ (MPa) was already known by previously performed static loading measurements.

Figure 1 shows the observed data and the corresponding strain life curve and respectively stress strain curve as a result of the regression method described above. Since such a

Figure 1: Observed data points and estimated strain life and stress strain curve.



(a) strain life curve



(b) stress strain curve

series of experiments is rather time consuming and expensive, it is advisable to choose the experimental design, here the strain amplitudes ϵ, very carefully. So the question is if we can choose a better design than the reference design

$$z_{ref} = (0.003, 0.0035, 0.004, 0.004, 0.0045, 0.005, 0.005) \tag{8}$$

used in the experiment described above.

Next we use our methodology from Section 2 in order to choose a design consisting of $N = 7$ points from the set $D = [0.002, 0.006]$ which should lead to a rather small expected error if we generate data corresponding to this design and use the estimation method described above. To apply our methodology from Section 2, we need a model for the generation of the (random) values of N_f and σ. The basic idea is to assume that the above parametric relations hold and to add normally distributed random errors to functional values $r_P^{-1}(\epsilon)$ and $\sigma_{comp} = E \cdot \epsilon_{el}$, i.e., for a given strain amplitude ϵ we choose

$$N_f = N_f(\epsilon) = r_P^{-1}(\epsilon) + \delta_{N_f}$$

and

$$\sigma = \sigma(\epsilon) = \sigma_{comp} + \delta_\sigma,$$

where δ_{N_f} and δ_σ are independent truncated normally distributed random variables with expectation 0 and standard deviation $sd_N = 3674.60$ and $sd_\sigma = 11.56$, respectively. Here the property of compatibility is used to determine the value of the stress amplitude corresponding to r_P. The parameters sd_N and sd_σ are chosen using data from the real experiment for a steel presented in Figure 1. Since the variable N_f describes the number of cycles till failure for an adjusted strain amplitude ϵ just $N_f \geq 1$ are useful. Similarly it behaves with the variable σ since for this variable only positive values make sense. For this reason we use in both cases truncated normally distributed random variables in such a way that only random errors satisfying $|\delta_{N_f}| < r_p^{-1}(z) - 1$ and respectively $|\delta_\sigma| < \sigma_{comp} - 9.7 \cdot 10^{-7}$ will be permitted.

Another problem coming up concerns the value of the parameter P which is not explicitly given. We help ourselves by a construction of a neighborhood of the reference parameter $p_{ref} = [788.305, 0.118, -0.0631, -0.4427]$. We construct P according to another set $P_{data} = \{(p_1, E_1), \cdots, (p_{228}, E_{228})\}$, an excerpt of "‘Materials Database For Cyclic

Loading"' (cf., Boller, Seeger and Vormwald (2008)). More precisely we consider the 0.1-neighborhood of p_{ref}, with $c = (|p_{ref}^{(1)}|, |p_{ref}^{(2)}|, |p_{ref}^{(3)}|, |p_{ref}^{(4)}|) \cdot 0.1$ and randomly choose P_i from the cuboid $[p_{ref} - c, p_{ref} + c]$. We accept this chosen value as P_i if there exists a $p_i \in [P_i - c, P_i + c]$ from the database, and set the modulus of elasticity to $E_i = E_l$, where

$$p_l = \arg\min \{\|P_i - p_l\|_1, \ l = 1, \cdots, 228\}.$$

Finally we need to estimate the expected error of our estimate of the cyclic parameters for a given design Z_i and a given value P_i of the parameters. We do this by generating 100 independent synthetic data sets corresponding to this parameter value, each data set $D_N(Z_i, P_i)$ is used to compute the estimate $\hat{r}_N(\cdot, D_N(Z_i, P_i))$, and in addition approximately L_1-distances

$$Err(\hat{r}_N(\cdot, D_N(Z_i; P_i)), r_{P_i}) = \frac{1}{l} \sum_{j=1}^{l} \left| \hat{r}_N(\bar{N}_{f,j}, D_N(Z_i; P_i)) - r_{P_i}(\bar{N}_{f,j}) \right|$$

are computed with $\bar{N}_{f,j} = (100 + (j-1) \cdot 100)$ and $l = 10000$. We define the dependent variable Y_i as the median of these errors:

$$Y_i = Median \left(Err(\hat{r}_N(\cdot, D_N(Z_i; P_i)), r_{P_i})_1^{100} \right). \tag{9}$$

In a last step, we apply a nonparametric regression estimate to the generated data set. Here we use the so-called Nadaraya-Watson kernel regression estimate (cf., e.g., Nadaraya (1964) and Watson (1964)), where the parameter of the bandwidth h of the estimate is chosen by splitting of the sample (cf., e.g., Chapter 7 in Györfi et al. (2002)).

Finally, we estimate the optimal design by the design $\hat{z} \in D_N$ for which the maximal value $m_n(\hat{z}, p)$ corresponding to the parameter p is minimal.

We constructed a Monte-Carlo set consisting of 100000 values and compute the corresponding regression estimate described as above. Then this estimate was evaluated on a grid. Here 8 fixed and equidistant chosen values in D where combined in increasing order, where repetitions in ϵ-values were admitted, resulting in 3433 considered designs. The grid corresponding to the cyclic parameters consists of 4096 points by covering the cuboid $[p_{ref} - c, p_{ref} + c]$ with an equidistant grid. To get the estimate of the optimal design we used a Nadaraya-Watson kernel regression estimate with a Gaussian kernel and bandwidth $h = 0.0002199$.

$\hat{z}^{(1)}$	$\hat{z}^{(2)}$	$\hat{z}^{(3)}$	$\hat{z}^{(4)}$	$\hat{z}^{(5)}$	$\hat{z}^{(6)}$	$\hat{z}^{(7)}$
0.002038	0.002038	0.002038	0.004841	0.005962	0.005962	0.005962

Table 2. Estimated optimal design.

Table 2 describes our estimation of the optimal design. Here one design point is chosen at 0.004841, where the other 6 points are symmetrically located at the edge of the design interval. This design is better than our reference design (8) with respect to our minimax criteria, since we have:

$$\begin{aligned}
&\max_{p \in PGrid} Median\left(Err(\hat{r}_N(\cdot, D_N(\hat{z};p)), r_p)_1^{1000}\right) &&= 6.4369 \cdot 10^{-5} \\
\leq &\max_{p \in PGrid} Median\left(Err(\hat{r}_N(\cdot, D_N(z_{ref};p)), r_p)_1^{1000}\right) &&= 3.0754 \cdot 10^{-4}.
\end{aligned}$$

So we can conlude that our method yields an estimated design which is four times better than the considered reference design.

In order to validate the statistical significance of this result, ten of these values were computed by separately started simulations. Here in all ten cases our estimated design is with respect to our minimax criteria better than the reference design, since the errors of the estimated design and the errors of the reference design are located in the following range, respectively:

$$\max_{p \in PGrid} Median\left(Err(\hat{r}_N(\cdot, D_N(\hat{z};p)), r_p)_1^{1000}\right) \in \left[6.2781070 \cdot 10^{-5}, 6.7610200 \cdot 10^{-5}\right]$$

and

$$\max_{p \in PGrid} Median\left(Err(\hat{r}_N(\cdot, D_N(z_{ref};p)), r_p)_1^{1000}\right) \in [0.000307092, 0.000330605].$$

5 Proofs

5.1 Auxiliary results

In this subsection we prove two auxiliary results, which we will use later to prove Theorem 1. Our first auxiliary result concerns uniform convergence of the kernel regression estimate.

Lemma 1. *Let $(X,Y),(X_1,Y_1),\cdots$ be independent and identically distributed random variables with values in $\mathbb{R}^d \times \mathbb{R}$ and denote the distribution of X by μ. Assume that*

$$\mathbf{E}\left(|Y|^2\right) < \infty, \tag{10}$$

and assume that the regression function $m : \mathbb{R}^d \to \mathbb{R}$ defined by $m(x) = \mathbf{E}\{Y|X = x\}$ is continuous. Let $\tilde{K} : \mathbb{R}_+ \to \mathbb{R}_+$ be a monotonically decreasing and left continuous function satisfying for some $R > 0$

$$\tilde{K}(+0) > 0 \quad \textit{and} \quad \tilde{K}(t) = 0 \quad \textit{for } t > R.$$

Define the kernel function $K : \mathbb{R}^d \to \mathbb{R}_+$ by

$$K(u) = \tilde{K}(\|u\|) \quad (u \in \mathbb{R}^d).$$

Let $r > 0$ be such that $\tilde{K}(r) > 0$, let $h_n > 0$ $(n \in \mathbb{N})$ and assume

$$h_n \to 0 \quad (n \to \infty) \tag{11}$$

and let m_n be the kernel estimate defined by

$$m_n(x) = \frac{\sum_{i=1}^n Y_i \cdot K\left(\frac{x - X_i}{\hat{h}_x}\right)}{\sum_{j=1}^n K\left(\frac{x - X_j}{\hat{h}_x}\right)}$$

and

$$\hat{h}_x = \min\left\{h \geq h_n \ : \ \mu_n(S_{r\cdot h}) \geq \frac{\log n}{n^{1/4}}\right\}.$$

Then for an arbitrary compact subset A of $\mathbb{R}^d$ the following assertion holds:

$$\sup_{x \in supp(\mu)\cap A} |m_n(x) - m(x)| \to 0 \quad a.s. \tag{12}$$

Proof of Lemma 1. Let $s \in (0,1)$ be arbitrary and $r > 0$. Using results from VC-theory (cf., e.g., Theorem 12.5, Corollary 13.2 in Devroye, Györfi and Lugosi (1996), Theorem 9.2 in Györfi et al. (2002), Theorem 9.1 in Györfi et al. (2002) and the proof of Lemma 3.2 in Kohler, Krzyżak and Walk (2003)) together with the Borel-Cantelli lemma it is possible to show

$$\frac{\sup_{x\in\mathbb{R}^d} |\mu_n(S_{r\cdot h_n}(x)) - \mu(S_{r\cdot h_n}(x))|}{\log n/\sqrt{n}} \to 0 \quad a.s., \tag{13}$$

$$\frac{\sup_{x\in\mathbb{R}^d}\left|\frac{1}{n}\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)-\mathbf{E}\left\{K\left(\frac{x-X}{h_n}\right)\right\}\right|}{\log n/\sqrt{n}} \to 0 \quad a.s. \tag{14}$$

and

$$\frac{\sup_{x\in\mathbb{R}^d}\left|\frac{1}{n}\sum_{i=1}^n Y_i\cdot 1_{\{|Y_i|\le n^{\frac{1}{4}}\}}\cdot K\left(\frac{x-X_i}{h_n}\right)-\mathbf{E}\left\{Y\cdot 1_{\{|Y|\le n^{\frac{1}{4}}\}}K\left(\frac{x-X}{h_n}\right)\right\}\right|}{\log n/n^{\frac{1}{4}}} \to 0 \quad a.s. \tag{15}$$

(cf., equations (11), (12) and (13) in Kohler, Krzyżak and Walk (2011)).

Since

$$\begin{aligned}\sup_{x\in supp(\mu)\cap A}|m_n(x)-m(x)| &\le \sup_{x\in supp(\mu)\cap A}\left|m_n(x)-\frac{\mathbf{E}\{Y\cdot K\left(\frac{x-X}{\hat{h}_x}\right)\}}{\mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\}}\right| \\ &\quad+\sup_{x\in supp(\mu)\cap A}\left|\frac{\mathbf{E}\{Y\cdot K\left(\frac{x-X}{\hat{h}_x}\right)\}}{\mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\}}-m(x)\right| \\ &:= T_{1,n}+T_{2,n}\end{aligned}$$

it suffices to show

$$T_{i,n}\to 0 \quad a.s. \quad (i\in\{1,2\}) \tag{16}$$

for all distributions of (X,Y) such that the regression function is continuous and such that $\mathbf{E}(|Y|^2)<\infty$. *In the first step of the proof* we show (16) for $i=1$. Because of $\mu_n(S_{r\cdot\hat{h}_x}(x))\ge\log(n)/n^{1/4}$ and because of $c_1\cdot 1_{S_r(x)}\le K(x)\le c_2$ (where $c_1=\tilde{K}(r)$ and $c_2=\tilde{K}(0)$) we have

$$\mathbf{E}\{K\left(\frac{x-X}{\hat{h}_x}\right)\}\ge\mathbf{E}\{c_1\cdot 1_{S_{r\cdot\hat{h}_x}(x)}(X)\}=c_1\cdot\mu\left(S_{r\cdot\hat{h}_x}\right) \tag{17}$$

and

$$\frac{1}{n}\sum_{i=1}^n K\left(\frac{x-X_i}{\hat{h}_x}\right)\ge c_1\cdot\frac{1}{n}\sum_{i=1}^n 1_{S_{r\cdot\hat{h}_x}(x)}(X_i)\ge c_1\cdot\frac{\log(n)}{n^{1/4}}.$$

It follows

$$\begin{aligned}&\left|\frac{\frac{1}{n}\sum_{i=1}^n Y_i\cdot K\left(\frac{x-X_i}{\hat{h}_x}\right)}{\frac{1}{n}\sum_{i=1}^n K\left(\frac{x-X_i}{\hat{h}_x}\right)}-\frac{\mathbf{E}\{Y\cdot K\left(\frac{x-X}{\hat{h}_x}\right)\}}{\mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\}}\right| \\ &=\left|\frac{\frac{1}{n}\sum_{i=1}^n Y_i\cdot 1_{\{|Y_i|>n^{1/4}\}}\cdot K\left(\frac{x-X_i}{\hat{h}_x}\right)}{\frac{1}{n}\sum_{i=1}^n K\left(\frac{x-X_i}{\hat{h}_x}\right)}+\frac{\sum_{i=1}^n Y_i\cdot 1_{\{|Y_i|\le n^{1/4}\}}\cdot K\left(\frac{x-X_i}{\hat{h}_x}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{\hat{h}_x}\right)}\right.\end{aligned}$$

$$
-\frac{\mathbf{E}\{Y\cdot 1_{\{|Y|\le n^{1/4}\}}\cdot K\left(\frac{x-X}{\hat{h}_x}\right)\}}{\mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\}}-\frac{\mathbf{E}\{Y\cdot 1_{\{|Y|>n^{1/4}\}}\cdot K\left(\frac{x-X}{\hat{h}_x}\right)\}}{\mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\}}\Bigg|
$$

$$
\begin{aligned}
&\le c_2\cdot\frac{\frac{1}{n}\sum_{i=1}^n |Y_i|\cdot 1_{\{|Y_i|>n^{1/4}\}}}{c_1\cdot\frac{\log(n)}{n^{1/4}}}+\frac{c_2\cdot\mathbf{E}\{|Y|\cdot 1_{\{|Y|>n^{1/4}\}}\}}{c_1\cdot\mu\left(S_{r\cdot\hat{h}_x}\right)}\\
&\qquad+\left|\frac{\sum_{i=1}^n Y_i\cdot 1_{\{|Y_i|\le n^{1/4}\}}\cdot K\left(\frac{x-X_i}{\hat{h}_x}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{\hat{h}_x}\right)}-\frac{\mathbf{E}\{Y\cdot 1_{\{|Y|\le n^{1/4}\}}\cdot K\left(\frac{x-X}{\hat{h}_x}\right)\}}{\mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\}}\right|\\
&=: T_{3,n}+T_{4,n}+T_{5,n}.
\end{aligned}
$$

Next we show

$$
\sup_{x\in A\cap supp(\mu)} T_{i,n}\to 0 \quad a.s. \tag{18}
$$

for $i\in\{3,4,5\}$. For $i=3$ we have for any $L>1$ and n sufficiently large

$$
\begin{aligned}
\frac{\frac{1}{n}\sum_{i=1}^n |Y_i|\cdot 1_{\{|Y_i|>n^{1/4}\}}}{\frac{\log(n)}{n^{1/4}}} &\le \frac{\frac{1}{n}\sum_{i=1}^n |Y_i|\cdot\frac{|Y_i|}{n^{1/4}}\cdot 1_{\{|Y_i|>n^{1/4}\}}}{\frac{\log(n)}{n^{1/4}}}\\
&= \frac{1}{\log n}\cdot\frac{1}{n}\sum_{i=1}^n |Y_i|^2\cdot 1_{\{|Y_i|>n^{1/4}\}}\\
&\le \frac{1}{n}\sum_{i=1}^n |Y_i|^2\cdot 1_{\{|Y_i|>L\}}\\
&\to \mathbf{E}\left\{|Y|^2\cdot 1_{\{|Y|>L\}}\right\} \quad a.s.
\end{aligned}
$$

by the strong law of large number and by (10). And because of (10) we get

$$
\mathbf{E}\left\{|Y|^2\cdot 1_{\{|Y|>L\}}\right\}\to 0
$$

for $L\to\infty$, from which (18) follows for $i=3$. For $i=4$ we observe for n sufficient large

$$
\begin{aligned}
&\sup_{x\in A\cap supp(\mu)}\frac{\mathbf{E}\{|Y|\cdot 1_{\{|Y|>n^{1/4}\}}\}}{\mu\left(S_{r\cdot\hat{h}_x}(x)\right)}\\
&\le \sup_{x\in A\cap supp(\mu)}\frac{\mathbf{E}\{|Y|\cdot\frac{|Y|}{n^{1/4}}\cdot 1_{\{|Y|>n^{1/4}\}}\}}{\mu_n(S_{r\cdot\hat{h}_x}(x))-(\mu_n(S_{r\cdot\hat{h}_x}(x))-\mu\left(S_{r\cdot\hat{h}_x}(x)\right))}\\
&= \sup_{x\in A\cap supp(\mu)}\frac{\mathbf{E}\{|Y|^2\cdot 1_{\{|Y|>n^{1/4}\}}\}}{n^{1/4}\cdot\left(\mu_n(S_{r\cdot\hat{h}_x}(x))-(\mu_n(S_{r\cdot\hat{h}_x}(x))-\mu\left(S_{r\cdot\hat{h}_x}(x)\right))\right)}\\
&\le \frac{\mathbf{E}\{|Y|^2\cdot 1_{\{|Y|>n^{1/4}\}}\}}{n^{1/4}\cdot\left(\log(n)/n^{1/4}-\sup_{x\in\mathbb{R}^d}|\mu_n(S_{r\cdot\hat{h}_x}(x))-\mu(S_{r\cdot\hat{h}_x}(x)))|\right)}
\end{aligned}
$$

$$= \frac{\mathbf{E}\{|Y|^2 \cdot 1_{\{|Y|>n^{1/4}\}}\}}{\log(n) - n^{1/4} \cdot \sup_{x\in\mathbb{R}^d} |\mu_n(S_{r\cdot\hat{h}_x}(x)) - \mu(S_{r\cdot\hat{h}_x}(x)))|}.$$

Because of (10) we have

$$\mathbf{E}\{|Y|^2 \cdot 1_{\{|Y|>n^{1/4}\}}\} \to 0 \quad (n \to \infty),$$

and together with (13) this implies (18) for $i = 4$. In order to show (18) for $i = 5$ we observe for n sufficiently large

$$\begin{aligned}
&\left| \frac{\sum_{i=1}^n Y_i \cdot 1_{\{|Y_i|\le n^{1/4}\}} \cdot K\left(\frac{x-X_i}{\hat{h}_x}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{\hat{h}_x}\right)} - \frac{\mathbf{E}\{Y \cdot 1_{\{|Y|\le n^{1/4}\}} \cdot K\left(\frac{x-X}{\hat{h}_x}\right)\}}{\mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\}} \right| \\
&= \Bigg| \frac{\mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\}\left(\frac{1}{n}\sum_{i=1}^n Y_i \cdot 1_{\{|Y_i|\le n^{1/4}\}} \cdot K\left(\frac{x-X_i}{\hat{h}_x}\right) - \mathbf{E}\{Y \cdot 1_{\{|Y|\le n^{1/4}\}} \cdot K\left(\frac{x-X}{\hat{h}_x}\right)\}\right)}{\frac{1}{n}\sum_{i=1}^n K\left(\frac{x-X_i}{\hat{h}_x}\right) \cdot \mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\}} \\
&\qquad + \frac{\mathbf{E}\left\{Y \cdot 1_{\{|Y|\le n^{1/4}\}} \cdot K\left(\frac{x-X}{\hat{h}_x}\right)\right\}\left(\mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\} - \frac{1}{n}\sum_{i=1}^n K\left(\frac{x-X_i}{\hat{h}_x}\right)\right)}{\frac{1}{n}\sum_{i=1}^n K\left(\frac{x-X_i}{\hat{h}_x}\right) \cdot \mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\}} \Bigg| \\
&\le \frac{\left|\frac{1}{n}\sum_{i=1}^n Y_i \cdot 1_{\{|Y_i|\le n^{1/4}\}} \cdot K\left(\frac{x-X_i}{\hat{h}_x}\right) - \mathbf{E}\{Y \cdot 1_{\{|Y|\le n^{1/4}\}} \cdot K\left(\frac{x-X}{\hat{h}_x}\right)\}\right|}{c_1 \cdot \mu_n(S_{r\cdot\hat{h}_x}(x))} \\
&\qquad + \frac{\mathbf{E}\{|Y| \cdot 1_{\{|Y|\le n^{1/4}\}} \cdot K\left(\frac{x-X}{\hat{h}_x}\right)\}}{\mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\}} \cdot \frac{|\mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\} - \frac{1}{n}\sum_{i=1}^n K\left(\frac{x-X_i}{\hat{h}_x}\right)|}{\frac{1}{n}\sum_{i=1}^n K\left(\frac{x-X_i}{\hat{h}_x}\right)} \\
&\le \frac{\left|\frac{1}{n}\sum_{i=1}^n Y_i \cdot 1_{\{|Y_i|\le n^{1/4}\}} \cdot K\left(\frac{x-X_i}{\hat{h}_x}\right) - \mathbf{E}\{Y \cdot 1_{\{|Y|\le n^{1/4}\}} \cdot K\left(\frac{x-X}{\hat{h}_x}\right)\}\right|}{c_1 \cdot \frac{\log(n)}{n^{1/4}}} \\
&\qquad + \frac{\mathbf{E}\{|Y| \cdot 1_{\{|Y|\le n^{1/4}\}} \cdot K\left(\frac{x-X}{\hat{h}_x}\right)\}}{c_1 \cdot \mu\left(S_{r\cdot\hat{h}_x}(x)\right)} \cdot \frac{\left|\mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\} - \frac{1}{n}\sum_{i=1}^n K\left(\frac{x-X_i}{\hat{h}_x}\right)\right|}{c_1 \cdot \mu_n\left(S_{r\cdot\hat{h}_x}(x)\right)} \\
&\le \frac{\left|\frac{1}{n}\sum_{i=1}^n Y_i \cdot 1_{\{|Y_i|\le n^{1/4}\}} \cdot K\left(\frac{x-X_i}{\hat{h}_x}\right) - \mathbf{E}\{Y \cdot 1_{\{|Y|\le n^{1/4}\}} \cdot K\left(\frac{x-X}{\hat{h}_x}\right)\}\right|}{c_1 \cdot \frac{\log(n)}{n^{1/4}}} \\
&\quad + \frac{c_2 \frac{\log(n)}{n^{1/4}} \cdot \mathbf{E}\{|Y| \cdot 1_{\{|Y|\le n^{1/4}\}}\}}{c_1 \frac{\log(n)}{n^{1/4}} \cdot \left(\frac{n^{1/4}}{\log(n)} - \left(\mu_n\left(S_{r\cdot\hat{h}_x}(x)\right) - \mu\left(S_{r\cdot\hat{h}_x}(x)\right)\right)\right)} \cdot \frac{\left|\mathbf{E}\left\{K\left(\frac{x-X}{\hat{h}_x}\right)\right\} - \frac{1}{n}\sum_{i=1}^n K\left(\frac{x-X_i}{\hat{h}_x}\right)\right|}{c_1 \frac{\log(n)}{n^{1/4}}}.
\end{aligned}$$

Because of

$$\frac{\log(n)}{n^{1/4}} \cdot \mathbf{E}\{|Y| \cdot 1_{\{|Y|\le n^{1/4}\}}\} \le \frac{\log(n)}{n^{1/4}} \cdot \mathbf{E}\{|Y|\} \to 0 \quad (n \to \infty)$$

and

$$\liminf_{n\to\infty} \frac{\log(n)}{n^{1/4}} \cdot \left(\frac{n^{1/4}}{\log(n)} - \sup_{x\in\mathbb{R}^d} \left| \mu_n(S_{r\cdot\hat{h}_x}(x)) - \mu(S_{r\cdot\hat{h}_x}(x)) \right| \right) > 0,$$

which follows from (13), we conclude from (14) and (15) that (18) also holds for $i = 5$. This completes the proof of (16) for $i = 1$.

In the second step of the proof we show (16) for $i = 2$. For $x \in supp(\mu)$ we know that $\mathbf{P}\{X \in S_\epsilon(x)\} > 0$ for all $\epsilon > 0$. By (17) we get

$$\mathbf{E}\left\{ K\left(\frac{x-X}{\hat{h}_x} \right) \right\} \geq c_1 \cdot \mathbf{P}\{X \in S_\epsilon(x)\} > 0 \quad \text{for } x \in supp(\mu),$$

hence

$$\begin{aligned}
\sup_{x\in supp(\mu)\cap A} \left| \frac{\mathbf{E}\left\{ Y \cdot K\left(\frac{x-X}{\hat{h}_x}\right)\right\}}{\mathbf{E}\left\{ K\left(\frac{x-X}{\hat{h}_x}\right)\right\}} - m(x) \right| &= \sup_{x\in supp(\mu)\cap A} \left| \frac{\mathbf{E}\left\{ K\left(\frac{x-X}{\hat{h}_x}\right)(Y - m(x))\right\}}{\mathbf{E}\left\{ K\left(\frac{x-X}{\hat{h}_x}\right)\right\}} \right| \\
&= \sup_{x\in supp(\mu)\cap A} \left| \frac{\mathbf{E}\left\{ \mathbf{E}\left\{ K\left(\frac{x-X}{\hat{h}_x}\right)(Y - m(x)) | X \right\}\right\}}{\mathbf{E}\left\{ K\left(\frac{x-X}{\hat{h}_x}\right)\right\}} \right| \\
&= \sup_{x\in supp(\mu)\cap A} \left| \frac{\mathbf{E}\left\{ K\left(\frac{x-X}{\hat{h}_x}\right)(m(X) - m(x))\right\}}{\mathbf{E}\left\{ K\left(\frac{x-X}{\hat{h}_x}\right)\right\}} \right| \\
&\leq \sup_{x\in supp(\mu)\cap A} \sup_{z\in\mathbb{R}^d : \|x-z\|\leq R\cdot\hat{h}_x} |m(x) - m(z)|
\end{aligned}$$

The last inequality follows from the fact that $K(u) = 0$ for $\|u\| > R$ for some $R > 0$.

Since $supp(\mu) \cap A$ is compact and m is continuous, we know that the term on the right hand side is less than

$$\sup_{x\in supp(\mu)\cap A} L(\hat{h}_x)$$

for some function $L : (0, \infty) \to \mathbb{R}$ satisfying

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

So it suffices to show

$$\limsup_{n\to\infty} \sup_{x\in supp(\mu)\cap A} \hat{h}_x = 0 \quad a.s.$$

This in turn follows from

$$\limsup_{n\to\infty} \sup_{x\in supp(\mu)\cap A} \hat{h}_x \leq h \quad a.s.$$

for all $h > 0$, which we show in the sequel.

Let $h > 0$ be arbitrary. Since $supp(\mu) \cap A$ is compact, there exists finitely many balls with centers in $supp(\mu) \cap A$ and radius $\frac{r \cdot h}{2}$ such that $supp(\mu) \cap A$ is contained in the union of these balls. Let $S_{\frac{r \cdot h}{2}}(\bar{x}_j)$, $\bar{x}_j \in supp(\mu) \cap A$, $j = 1, \cdots, \bar{k}$ be these balls. Since the centers of these balls are in $supp(\mu)$ we know that all balls have a μ–measure greater than zero, so there exists some constant $c > 0$ such that $\mu\left(S_{\frac{r \cdot h}{2}}(\bar{x}_j)\right) > c$ for $j = 1, \cdots, \bar{k}$. Since every ball $S_{r \cdot h}(x)$ with $x \in supp(\mu) \cap A$ contains at least one of these balls (e.g., the ball where x lies within) we get that $\mu(S_{r \cdot h}(x)) > c$ for all $x \in supp(\mu) \cap A$. From this and (13) we can conclude that all balls $S_{r \cdot h}(x)$ with $x \in supp(\mu) \cap A$ contain with probability one for n sufficiently large at least $n/\log(n)$ data points. But this together with (11) implies

$$\sup_{x \in supp(\mu) \cap A} \hat{h}_x \leq h$$

for n sufficiently large almost surely. The proof is complete. □

Next we use Lemma 1 to show

Lemma 2. *Let $((Z,P),Y), ((Z_1,P_1),Y_1), \cdots$ be independent and identically distributed random variables taking values in $\left(D^N \times \mathcal{P}\right) \times \mathbb{R}_+$, where $D \subset \mathbb{R}^{d_z}$ and $\mathcal{P} \subset \mathbb{R}^{d_p}$ are compact. Let $\mu_{Z,P}$ be the probability measure of (Z,P) and assume that $supp(\mu_{Z,P}) = D^N \times \mathcal{P}$. Define the kernel estimate by*

$$m_n(z,p) = \frac{\sum_{i=1}^n Y_i \cdot K\left(\frac{\binom{z}{p} - \binom{Z_i}{P_i}}{\hat{h}_{z,p}}\right)}{\sum_{i=1}^n K\left(\frac{\binom{z}{p} - \binom{Z_i}{P_i}}{\hat{h}_{z,p}}\right)}$$

with

$$\hat{h}_{z,p} = \min\left\{h \geq h_n \quad : \quad \mu_n(S_{r \cdot h}(z,p)) \geq \frac{\log(n)}{n^{1/4}}\right\}, \tag{19}$$

and set

$$\hat{Z}_n = \arg\min_{z \in D^N} \max_{p \in \mathcal{P}} m_n(z,p), \tag{20}$$

and assume for notational simplicity again that the maximum and minimum above exist. Furthermore assume that the conditions of Lemma 1 hold. Then

$$\max_{p \in \mathcal{P}} m\left(\hat{Z}_n, p\right) \to \min_{z \in D^N} \max_{p \in \mathcal{P}} m(z,p) \quad a.s. \tag{21}$$

Proof of Lemma 2. Since

$$\begin{aligned}\left|\max_{p\in\mathcal{P}} m\left(\hat{Z}_n,p\right)-\min_{z\in D^N}\max_{p\in\mathcal{P}} m\left(z,p\right)\right| &\le \left|\max_{p\in\mathcal{P}} m\left(\hat{Z}_n,p\right)-\min_{z\in D^N}\max_{p\in\mathcal{P}} m_n\left(z,p\right)\right| \\ &+\left|\min_{z\in D^N}\max_{p\in\mathcal{P}} m_n\left(z,p\right)-\min_{z\in D^N}\max_{p\in\mathcal{P}} m\left(z,p\right)\right| \\ &= T_{1,n}+T_{2,n}\end{aligned}$$

it suffices to show

$$T_{i,n}\to 0 \quad a.s. \quad for \quad i\in\{1,2\}. \tag{22}$$

To show (22) for $i=1$ we use the definition of $\hat{Z}_n$ and the relation

$$\left|\max_j a_j-\max_j b_j\right|\le\max_j\left|a_j-b_j\right|.$$

It follows

$$\begin{aligned}T_{1,n} &= \left|\max_{p\in\mathcal{P}} m\left(\hat{Z}_n,p\right)-\max_{p\in\mathcal{P}} m_n\left(\hat{Z}_n,p\right)\right| \\ &\le \max_{p\in\mathcal{P}}\left|m\left(\hat{Z}_n,p\right)-m_n\left(\hat{Z}_n,p\right)\right| \\ &\le \sup_{(z,p)\in D^N\times\mathcal{P}}\left|m\left(z,p\right)-m_n\left(z,p\right)\right|\to 0 \quad a.s.\end{aligned}$$

by Lemma 1.

To show (22) for $i=2$ we use

$$|\min_j a_j-\min_j b_j|=|\max_j(-b_j)-\max_j(-a_j)|\le\max_j\left|a_j-b_j\right|,$$

and can continue with

$$\begin{aligned}&\left|\min_{z\in D^N}\max_{p\in\mathcal{P}} m_n\left(z,p\right)-\min_{z\in D^N}\max_{p\in\mathcal{P}} m\left(z,p\right)\right| \\ &\le \max_{z\in D^N}\left|\max_{p\in\mathcal{P}} m_n\left(z,p\right)-\max_{p\in\mathcal{P}} m\left(z,p\right)\right| \\ &\le \sup_{(z,p)\in D^N\times\mathcal{P}}\left|m_n(z,p)-m(z,p)\right|\to 0 \quad a.s.,\end{aligned}$$

where the last step follows again from Lemma 1. □

5.2 Proof of Theorem 1

Set

$$Y_i=Err\left(\hat{r}_N(\cdot,\mathcal{D}_n(Z_i;P_i)),r_{P_i}\right) \quad (i\in\{1,\ldots,n\}).$$

Then m defined by (3) is the regression function to $((Z_1, P_1), Y_1)$. The assumptions of Theorem 1 imply the assumption of Lemma 2, in particular (5) implies that Y_1 satisfies the integrability condition of Lemma 2 and $(A2) - (A4)$ imply $supp(\mu_{Z,P}) = D^N \times \mathcal{P}$. Application of Lemma 2 yields the assertion. □

Acknowledgement

The authors would like to thank the German Research Foundation (DFG) for founding this project within the Collaborative Research Center 666.

References

[1] Atkinson, A. C. (1982). Development in the Design of Experiments. *International Statistical Review* **50**, pp. 161-177.

[2] Beirlant, J. and Györfi, L. (1998). On the asymptotic L_2-error in partitioning regression estimation. *Journal of Statistical Planning and Inference* **71**, pp. 93–107.

[3] Boller, Chr., Seeger, T. and Vormwald, M.(2008). Materials Database for Cyclic Loading. *Fachgebiet Werkstoffmechanik, TU Darmstadt.*

[4] Dette, H. (1997). Designing Experiments with Respect to 'Standardized' Optimality Criteria. *Journal of the Royal Statistical Society.Series B (Methodological)* **59**, pp. 97-110.

[5] Devroye, L. (1978a). The uniform convergence of the Nadaraya-Watson regression function estimate. *Canadian Journal of Statistics* **6**, pp. 179-191.

[6] Devroye, L. (1978b). The uniform convergence of nearest neighbor regression function estimators and their application in optimization. *IEEE Transactions on Information Theory* **24**, pp. 142-151.

[7] Devroye, L. (1982). Necessary and sufficient conditions for the almost everywhere convergence of nearest neighbor regression function estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **61**, pp. 467–481.

[8] Devroye, L., Györfi, L. and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition.* Springer-Verlag, New York.

[9] Devroye, L., Györfi, L., Krzyżak, A., and Lugosi, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *Annals of Statistics* **22**, pp. 1371–1385.

[10] Devroye, L. and Krzyżak, A. (1989). An equivalence theorem for L_1 convergence of the kernel regression estimate. *Journal of Statistical Planning and Inference* **23**, pp. 71–82.

[11] Devroye, L. and Wagner, T. J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Annals of Statistics* **8**, pp. 231–239.

[12] Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing.* 2nd edition, Marcel Dekker, New York.

[13] Györfi, L. (1981). Recent results on nonparametric regression estimate and multiple classification. *Problems of Control and Information Theory* **10**, pp. 43–52.

[14] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression.* Springer Series in Statistics, Springer-Verlag, New York.

[15] Härdle, W., and Luckhaus, S. (1984). Uniform consistency of a class of regression function estimates. *Annals of Statistics* **12**, pp. 612-623.

[16] Imhof, L. A. (2001). Maximin Designs for Exponential Growth Models and Heteroscedastic Polynomial Models. *The Annals of Statistics* **29**, pp. 561-576.

[17] Kiefer, J. (1961). Optimum Design in Regression Problems. *Annals of Mathematical Statistics* **32**, pp. 298-325.

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

[19] Kohler, M., Krzyżak, A. and Walk, H. (2003). Strong consistency of automatic kernel regression estimates. *Annals of the Institute of Statistical Mathematics* **55**, pp. 287-308.

[20] Kohler, M., Krzyżak, A. and Walk, H. (2011). Estimation of the essential supremum of a regression function. *Statistics and Probability Letters* **81**, pp. 685-693.

[21] Lugosi, G. and Zeger, K. (1995). Nonparametric estimation via empirical risk minimization. *IEEE Transactions on Information Theory* **41**, pp. 677-687.

[22] Mack, Y. P. (1981). Local properties of k–nearest neighbor regression estimates. *SIAM Journal on Algebraic and Discrete Methods* **2**, pp. 311–323.

[23] Manson, S. S. (1965). Fatigue: A complex subject - some simple approximation. *Experimental Mechanics* **5**, pp. 193-226.

[24] Müller, P. and Parmigiani, G. (1996). Optimal Design via curve fitting of Monte Carlo experiments. *J. Amer. Statist. Assoc.* **90**, pp. 1322-1330.

[25] Nadaraya, E. A. (1964) On estimating regression. *Theory of Probability and its Applications* **9**, pp. 141–142.

[26] Nadaraya, E. A. (1970). Remarks on nonparametric estimates for density functions and regression curves. *Theory of Probability and its Applications* **15**, pp. 134–137.

[27] Pukelsheim, F. (1993). Optimal Design of Experiments. *Wiley, New York.*

[28] Ramberg, W. and Osgood, W. R. (1943). Description of stress-strain curves by three parameters. *Technical Note No. 902, National Advisory Committee For Aeronautics, Washington DC.*

[29] Silvey, S. D. (1980). Optimal Design. *London: Chapman and Hall.*

[30] Stone, C. J. (1977). Consistent nonparametric regression. *Annals of Statistics* **5**, pp. 595–645.

[31] Watson, G. S. (1964). Smooth regression analysis. *Sankhya Series A* **26**, pp. 359–372.

[32] Williams, C.R., Lee, Y.-L. and Rilly, J.T. (2002). A practical method for statistical analysis of strain-life fatigue data. *International Journal of Fatigue* **25**, pp. 427-436.

[33] Zhao, L. C. (1987). Exponential bounds of mean error for the nearest neighbor estimates of regression functions. *Journal of Multivariate Analysis* **21**, pp. 168–178.