Journal of Multivariate Analysis 132 (2014) 197–208

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Optimal global rates of convergence for noiseless regression estimation problems with adaptively chosen design

Michael Kohler

Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany

ARTICLE INFO

Article history:
Received 20 September 2013
Available online 6 September 2014

AMS subject classifications:
primary 62G08
secondary 62G05

Keywords:
L_1-error
Minimax rate of convergence
Noiseless regression
Scattered data approximation

ABSTRACT

Given the values of a measurable function $m : \mathbb{R}^d \to \mathbb{R}$ at n arbitrarily chosen points in $\mathbb{R}^d$ the problem of estimating m on whole $\mathbb{R}^d$ is considered. Here the estimate has to be defined such that the L_1 error of the estimate (with integration with respect to a fixed but unknown probability measure) is small. Under the assumption that m is (p, C)-smooth (i.e., roughly speaking, m is p-times continuously differentiable) it is shown that the optimal minimax rate of convergence of the L_1 error is $n^{-p/d}$, where the upper bound is valid even if the support of the design measure is unbounded but the design measure satisfies some moment condition. Furthermore it is shown that this rate of convergence cannot be improved even if the function is not allowed to change with the size of the data.

1. Introduction

In this article the problem of estimating a measurable function $m : \mathbb{R}^d \to \mathbb{R}$ from n observations of the value of the function m at points $z_1, \dots, z_n \in \mathbb{R}^d$, which might be arbitrarily chosen, is considered. Any estimate of m uses in a first step a strategy to choose the points

$$z_1, z_2 = z_2((z_1, m(z_1))), \dots, z_n = z_n((z_1, m(z_1)), \dots, (z_{n-1}, m(z_{n-1}))) \tag{1}$$

and then uses the data

$$\mathcal{D}_n = \{(z_1, m(z_1)), \dots, (z_n, m(z_n))\} \tag{2}$$

to estimate m by $m_n(\cdot) = m_n(\cdot, \mathcal{D}_n) : \mathbb{R}^d \to \mathbb{R}$. In numerical analysis this problem is known under the name scattered data approximation (usually with non-adaptively chosen points $z_1, \dots, z_n \in \mathbb{R}^d$), see, e.g., [27] and the literature cited therein. In this paper it is studied from a statistical point of view.

Motivated by a problem in density estimation, where m_n is used to generate additional data for the density estimate and where the error of the method crucially depends on the L_1 error of m_n (cf., [5,9]), the error of m_n is measured in this paper by the L_1 error computed with respect to a fixed but unknown probability measure μ, i.e., by

$$\int |m_n(x) - m(x)| \, \mu(dx). \tag{3}$$

In order to derive nontrivial rate of convergence results it is assumed in the sequel that the regression function is (p, C)-smooth according to the following definition.

E-mail address: kohler@mathematik.tu-darmstadt.de.

http://dx.doi.org/10.1016/j.jmva.2014.08.008
0047-259X/

Definition 1. Let $p = k + \beta$ for some $k \in \mathbb{N}_0$ and $0 < \beta \leq 1$, and let $C > 0$. A function $m : \mathbb{R}^d \to \mathbb{R}$ is called (p, C)-smooth, if for every $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ with $\sum_{j=1}^d \alpha_j = k$ the partial derivative $\frac{\partial^k m}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}$ exists and satisfies

$$\left| \frac{\partial^k m}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}(x) - \frac{\partial^k m}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}(z) \right| \leq C \cdot \|x - z\|^\beta$$

for all $x, z \in \mathbb{R}^d$, where $\mathbb{N}_0$ is the set of non-negative integers.

In the sequel minimax rate of convergence results for the L_1 error (3) are derived. More precisely, for function classes $\mathcal{F}^{(p,C)}$ of (p, C)-smooth functions $f : \mathbb{R}^d \to \mathbb{R}$ the behavior of

$$\inf_{\tilde{m}_n} \sup_{m \in \mathcal{F}^{(p,C)}} \int |\tilde{m}_n(x) - m(x)| \, \mu(dx)$$

is analyzed, and estimates m_n are constructed such that

$$\int |m_n(x) - m(x)| \, \mu(dx) \approx \inf_{\tilde{m}_n} \sup_{m \in \mathcal{F}^{(p,C)}} \int |\tilde{m}_n(x) - m(x)| \, \mu(dx).$$

A related problem is nonparametric regression estimation, where the x-values of the data $\mathcal{D}_n$ defined by (1) and (2) are given by an independent and identically distributed sample of μ and the corresponding function values are observed with additional errors with mean zero. This problem has been extensively studied in the literature. The most popular estimates include kernel regression estimate (cf., e.g., [18,19,26,8,20,22,7] or [16]), partitioning regression estimate (cf., e.g., [10,2] or [15]), nearest neighbor regression estimate (cf., e.g., [4] or [6]), least squares estimates (cf., e.g., [17] or [12]) and smoothing spline estimates (cf., e.g., [25] or [13]). Minimax rates of convergence in this context have been derived in [21–24,1,15,16].

Kohler and Krzyżak [14] have analyzed how the minimax rate of convergence results in [22] change in case that the function m can be observed without error. The main results are that firstly for estimating (p, C)-smooth functions no estimate can achieve a rate better than $n^{-p/d}$. Secondly, a nearest neighbor estimate achieves this rate if $p \leq 1$. Thirdly, a nearest neighbor polynomial interpolation estimate achieves this rate for arbitrary $p \in \mathbb{N}$ in case $d = 1$ provided the distribution μ satisfies regularity assumptions (which are satisfied, e.g., in case of the uniform distribution). And fourthly it is shown that without regularity assumption on μ no estimate can achieve a rate of convergence better than n^{-1}. Throughout this paper it is assumed that the support of μ is bounded.

In this article it is investigated how these rate of convergence results change in case that the estimate is allowed to choose the design points, i.e., the points where the function values of m are observed, in an adaptive way as described by (1). Surprisingly, the minimax rate of convergence for estimation of (p, C)-smooth functions still remains $n^{-p/d}$, but this time it is achievable even in case $p/d > 1$ without regularity conditions on the measure μ. In order to prove the corresponding lower bounds techniques from the standard minimax theory are applied. There it is allowed that the function to be estimated changes whenever the data size changes. It is shown furthermore that the above rate of convergence cannot be improved even if the function is not allowed to change with the size of the data, and that this rate of convergence can be achieved even if the support of the measure μ is unbounded but μ satisfies some moment condition.

Throughout the paper the following notation is used: The sets of natural numbers, integers and real numbers are denoted by $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{R}$, resp. For $z \in \mathbb{R}$ the smallest integer greater than or equal to z is denoted by $\lceil z \rceil$, and $\lfloor z \rfloor$ is the largest integer less than or equal to z. For $f : \mathbb{R}^d \to \mathbb{R}$

$$\|f\|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)|$$

is its supremum norm, and the supremum norm of f on a set $A \subseteq \mathbb{R}^d$ is denoted by

$$\|f\|_{\infty,A} = \sup_{x \in A} |f(x)|$$

$\|x\|$ is the Euclidean norm of a vector $x \in \mathbb{R}^d$. The components of $x \in \mathbb{R}^d$ are denoted by $x^{(1)}, \ldots, x^{(d)}$, i.e.,

$$x = (x^{(1)}, \ldots, x^{(d)})^T.$$

The support of a probability measure μ defined on the Borel sets in $\mathbb{R}^d$ is denoted by

$$supp(\mu) = \left\{ x \in \mathbb{R}^d : \mu(S_r(x)) > 0 \text{ for all } r > 0 \right\},$$

where $S_r(x)$ is the ball of radius r around x.

The outline of the paper is as follows: The main results are formulated in Section 2. The proofs are contained in Section 3.

2. Main results

In our first result we assume that the support of μ is bounded. In order to simplify the notation we assume w.l.o.g. that $supp(\mu) = [0, 1]^d$. We will use well-known results from spline theory to show that if we choose in this case the design points $z_1, \ldots, z_n$ equidistantly in $[0, 1]^d$, then a properly defined spline approximation of a (p, C)-smooth function achieves the rate of convergence $n^{-p/d}$.