Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Complexity 21 (2005) 211–229
Journal of COMPLEXITY
www.elsevier.com/locate/jco

Information-based nonlinear approximation: an average case setting

Mark Kon[a,*], Leszek Plaskota[b]

[a] *Department of Mathematics and Statistics, Boston University, 111 Cummington Street, Boston, MA12215, USA*
[b] *Department of Mathematics, Informatics and Mechanics, Warsaw University, Banacha 2, 02-097 Warsaw, Poland*

Received 15 January 2004; accepted 15 November 2004

Abstract

Nonlinear approximation (NA) has usually been studied under deterministic assumptions and complete information about the underlying functions. In the present paper we assume only *partial information*, e.g., function values at some points, and we are interested in the *average case* error and complexity of NA. We show that the problem can be essentially decomposed in two independent problems related to average case nonlinear (restricted) approximation from complete information, and to average case unrestricted approximation from partial information. The results are then applied to average case piecewise polynomial approximation on $C([0,1])$ based on function values with respect to r-fold Wiener measure. In this case, to approximate with error ε it is necessary and sufficient to know the function values at $\Theta\left(\left(\varepsilon^{-1}\ln^{1/2}(1/\varepsilon)\right)^{1/(r+1/2)}\right)$ equidistant points and use $\Theta\left(\varepsilon^{-1/(r+1/2)}\right)$ adaptively chosen break points in piecewise polynomial approximation.

MSC: 41A46; 41A10

Keywords: Nonlinear approximation; Information-based complexity; Average case error; Wiener measure

* Corresponding author. Tel.: +1 617 353 2560; fax: +1 617 353 8100.
E-mail addresses: mkon@bu.edu (M. Kon), leszekp@mimuw.edu.pl (L. Plaskota).

0885-064X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jco.2004.11.001

1. Introduction

Nonlinear approximation (NA) relies on approximating a function f using a nonlinear manifold that consists of k-term linear combinations of functions from a dictionary. Such approximation is constructed based on full knowledge about f, and one is interested in the error as k goes to infinity, see, e.g., [1] for a review. Since in practice the underlying function is usually given via its (exact or noisy) values at finitely many points, the assumption that f is completely known can be sometimes questioned. On the other hand, there is *information-based complexity (IBC)* theory where partial information is essential; see, e.g., [7]. In this situation the study of NA based on partial and/or noisy information becomes quite natural.

In [2,3], *information-based nonlinear approximation (IBNA)* was studied in the context of neural networks. The authors considered the *worst case setting* and established, among other things, the following remarkable property. Suppose one wants to know how many function evaluations and how many terms in approximation he needs to approximate the function with error $\varepsilon > 0$. It turns out that this problem can be essentially solved by decomposing it in two problems corresponding to NA and IBC. Two crucial notions used are information complexity and neural complexity, which mean, respectively, the minimal number of function evaluations and the minimal number of terms in approximation sufficient to approximate with error ε. Since the two quantities have been studied (mostly independently) in NA and IBC, the results from both theories can be integrated to obtain results in IBNA.

In the present paper we study IBNA in the *average case setting*. That is, assuming the functions are distributed according to a probability measure, we ask for the average number of function evaluations and average number of terms in approximation necessary and sufficient to approximate within the average error ε.

Information complexity in the average case setting has been the topic of extensive study in IBC, but, unlike in the worst case, average case NA seems not to have been regularly studied yet. Therefore the purpose of this paper is twofold. First, we want to establish general results corresponding to those from the worst case. This is done in Sections 3 and 4 where we show that, again, to obtain the complexity results and construct best approximations in the average case IBNA, it suffices to combine the corresponding results and approximations in the average case IBC and NA. Second, we want to provide a first thorough analysis of complexity of average case NA for a nontrivial problem. This is done in Section 5, where piecewise polynomial uniform approximation with respect to r-fold Wiener measure is considered.

We now use the problem of Section 5 to illustrate results obtained in this paper. Suppose we want to uniformly approximate a random function $f : [0, 1] \to \mathbb{R}$ distributed according to the r-fold Wiener measure. The approximation is based on n evaluations, $f(x_1), \ldots, f(x_n)$, and given as a piecewise polynomial of degree at most s ($s \geqslant r$) with k break points. We stress that adaption is allowed, i.e., the choice of successive points at which the function is evaluated and the number n of them may depend on previously obtained values (adaptive information), and the break points of the approximation and the number k of them may depend on the gathered information about f (adaptive approximation). We ask how the points x_j in information and the break points in approximation should be chosen