Available online at www.sciencedirect.com
ScienceDirect
Journal of Approximation Theory 194 (2015) 35–61
Journal of Approximation Theory
www.elsevier.com/locate/jat

Full length article

On some aspects of approximation of ridge functions

Anton Kolleck[a], Jan Vybíral[b,*]

[a] Mathematical Institute, Technical University Berlin, Strasse des 17. Juni 136, D-10623 Berlin, Germany
[b] Department of Mathematical Analysis, Charles University, Sokolovská 83, 186 00, Prague 8, Czech Republic

Received 27 May 2014; received in revised form 2 January 2015; accepted 22 January 2015
Available online 7 February 2015

Communicated by Paul Nevai

Abstract

We present effective algorithms for uniform approximation of multivariate functions satisfying some prescribed inner structure. We extend, in several directions, the analysis of recovery of ridge functions $f(x) = g(\langle a, x\rangle)$ as performed earlier by one of the authors and his coauthors. We consider ridge functions defined on the unit cube $[-1, 1]^d$ as well as recovery of ridge functions defined on the unit ball from noisy measurements. We conclude with the study of functions of the type $f(x) = g(\|a - x\|^2_{l_2^d})$.

MSC: 65D15; 41A25

Keywords: Ridge functions; High-dimensional function approximation; Noisy measurements; Compressed sensing; Dantzig selector

1. Introduction

Functions depending on a large number of variables play nowadays a crucial role in many areas, including parametric and stochastic PDEs, bioinformatics, financial mathematics, data analysis and learning theory. Together with an extensive computational power being used in

* Corresponding author.
E-mail addresses: kolleck@math.tu-berlin.de (A. Kolleck), vybiral@karlin.mff.cuni.cz (J. Vybíral).

http://dx.doi.org/10.1016/j.jat.2015.01.003

these applications, results on basic numerical aspects of these functions become crucial. Unfortunately, multivariate problems often suffer from the *curse of dimension*, i.e. the minimal number of operations necessary to achieve (an approximation of) a solution grows exponentially with the underlying dimension of the problem. Although this effect was observed many times in the literature, we refer to [27] for probably the most impressive result of this kind—namely that even the uniform approximation of infinitely-differentiable functions is intractable in high dimensions.

In the area of *Information-Based Complexity* it was possible to achieve a number of positive results on tractability of multivariate problems by imposing an additional (structural) assumption on the functions under study. The best studied concepts in this area include tensor product constructions and different concepts of anisotropy and weights. We refer to the series of monographs [26,28,29] for an extensive treatment of these and related problems. We pursue the direction initiated by Cohen, Daubechies, DeVore, Kerkyacharian and Picard in [11] and further developed in a series of recent papers [18,20,25]. This line of study is devoted to *ridge functions*, which are multivariate functions f taking the form $f(x) = g(\langle a, x\rangle)$ for some univariate function g and a non-zero vector $a \in \mathbb{R}^d$. We refer also to [15,32,33] for a related approach.

Functions of this type are by no means new in mathematics. They appear for example very often in statistics in the frame of the so-called *single index models*. They play also an important role in approximation theory, where their simple structure motivated the question if a general function could be well approximated by sums of ridge functions. The pioneering work in this field is [24], where the term “ridge function” was first introduced, and also [22], where the fundamentality of ridge functions was investigated. Ridge functions appeared also in mathematical analysis of neural networks [4,31] and as the basic building blocks of *ridgelets* of Candès and Donoho [6]. A survey on approximation by (sums of) ridge functions was given in [30].

The biggest difference between our setting and the usual approach of statistical learning and data analysis is that we suppose that the sampling points of f can be freely chosen, and are not given in advance. This happens, for instance, if sampling of the unknown function at a point is realized by a (costly) PDE solver.

Most of the techniques applied so far in recovery of ridge functions are based on the simple formula

$$\nabla f(x) = g'(\langle a, x\rangle) \cdot a. \tag{1.1}$$

One way how to use (1.1) is to approximate the gradient of f at a point with non-vanishing $g'(\langle a, x\rangle)$. By (1.1), it is then collinear with a. Once a is recovered, one can use any one-dimensional sampling method to approximate g.

Another way to approximate a is inspired by the technique of *compressed sensing* [8,16]. Taking directional derivatives of f at x results into

$$\frac{\partial f(x)}{\partial \varphi} = \langle \nabla f(x), \varphi\rangle = g'(\langle a, x\rangle)\langle a, \varphi\rangle,$$

i.e. it gives access to the scalar product of a with a chosen vector φ. If we assume that most of the coordinates of a are zero (or at least very small) and choose the directions $\varphi_1, \dots, \varphi_m$ at random, one can recover a effectively by the algorithms of sparse recovery.

Our aim is to fill some gaps left so far in the analysis done in [18]. Although the possibility of extending the analysis also to functions defined on other domains than the unit ball was mentioned already in [18], no steps in this direction were done there. We study in detail ridge functions defined on the unit cube $[-1, 1]^d$. The crucial component of our analysis is the use of the sign of a vector $\mathrm{sign}(x)$, which is defined componentwise. Although the mapping $x \to \mathrm{sign}(x)$