Journal of Multivariate Analysis 142 (2015) 167–182

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Convergence estimates in probability and in expectation for discrete least squares with noisy evaluations at random points

Giovanni Migliorati [a,*], Fabio Nobile [a], Raúl Tempone [b]

[a] MATHICSE-CSQI, École Polytechnique Fédérale de Lausanne, Lausanne CH-1015, Switzerland

[b] Applied Mathematics and Computational Sciences, and SRI Center for Uncertainty Quantification in Computational Science and Engineering, KAUST, Thuwal 23955-6900, Saudi Arabia

ARTICLE INFO

Article history:
Received 24 December 2014
Available online 28 August 2015

AMS subject classifications:
41A10
41A25
41A50
41A63
62G08
65M70

Keywords:
Approximation theory
Discrete least squares
Noisy evaluations
Error analysis
Convergence rates
Large deviations
Learning theory
Multivariate polynomial approximation

ABSTRACT

We study the accuracy of the discrete least-squares approximation on a finite-dimensional space of a real-valued target function from noisy pointwise evaluations at independent random points distributed according to a given sampling probability measure. The convergence estimates are given in mean-square sense with respect to the sampling measure. The noise may be correlated with the location of the evaluation and may have nonzero mean (offset). We consider both cases of bounded or square-integrable noise/offset. We prove conditions between the number of sampling points and the dimension of the underlying approximation space that ensure a stable and accurate approximation. Particular focus is on deriving estimates in probability within a given confidence level. We analyze how the best approximation error and the noise terms affect the convergence rate and the overall confidence level achieved by the convergence estimate. The proofs of our convergence estimates in probability use arguments from the theory of large deviations to bound the noise term. Finally we address the particular case of multivariate polynomial approximation spaces with any density in the beta family, including uniform and Chebyshev.

1. Introduction

The motivations of our analysis come from the development of discrete least-squares approximation methods for functions depending on a multivariate random variable distributed according to a known probability measure. This topic falls at the intersection of approximation theory and learning theory [6,7], and is related to nonparametric regression with random design [10] and statistical learning theory [21]. More specifically, our framework is an instance of the *projection learning problem* (or improper function learning problem) described in [6,18,19].

We focus on the discrete least-squares approximation of a target function on a given finite dimensional (linear) vector space using pointwise evaluations at independent and randomly selected points, identically distributed according to the underlying probability measure. In particular, we are interested in the case of discrete least-squares projection on

* Corresponding author.
E-mail addresses: giovanni.migliorati@epfl.ch (G. Migliorati), fabio.nobile@epfl.ch (F. Nobile), raul.tempone@kaust.edu.sa (R. Tempone).

http://dx.doi.org/10.1016/j.jmva.2015.08.009

not necessarily bounded multivariate approximation sets, *i.e.* the minimizer of the discrete least-squares problem is not constrained to be in a compact subset. Two situations might occur, depending on the context and on the origin of the evaluations of the target function: *noiseless* evaluations or *noisy* evaluations. The former situation arises for example in an abstract modeling context, where round-off or other discretization errors can be properly controlled. The latter situation typically arises when dealing with experimental data, which are polluted by measurement and/or systematic errors.

A vast literature is available for discrete least-squares approximations on compact sets or linear vector spaces in the noisy case. In the case of linear vector spaces, we mention the bound in [10, Theorem 11.3] or those in [3,2], which hold in expectation under the assumption that the target function itself is bounded. Often a truncation operator has to be used to obtain those bounds. Moreover, these results are nonoptimal in the noiseless case, as the best approximation error in the subspace is not recovered when the amount of noise tends to zero.

The stability and accuracy of discrete least squares on finite-dimensional vector spaces in the noiseless and noisy cases have been recently analyzed in several works [5,16,4,15,12]. It is shown that optimal convergence rates can be recovered in the noiseless case if a suitable relation between the number of evaluations and the dimension of the approximation space is enforced. Moreover, such relation guarantees stability of the discrete projection with high probability.

Generalizations of the previous analyses to the noisy case have been presented as well in the aforementioned works. In the particular case of bounded noise (stochastic or deterministic) with zero mean, an estimate in expectation has been proposed in [5]. Estimates in expectation with the deterministic noise model have been proven in [4]. Estimates in probability have been proven in [4] but using the best approximation error in L^∞ rather than L^2 and focusing only on the deterministic noise model. In both the noiseless and noisy cases, the analyses in [5,4] rely on the Chernoff bounds for sums of random matrices proven in [1,20]. The analysis in [16] uses different techniques to derive a convergence estimate in probability, and covers only the noiseless case.

The purpose of the present work is to derive new convergence estimates in probability and in expectation, in the general case of noise of stochastic type with nonzero mean, that recover optimal convergence rates in the limit of zero noise. We split the noise into two parts: the conditional expectation of the noise w.r.t. the sampling measure, that we name in the following as the *offset* of the noise, and the part of the noise due to its intrinsic randomness, hereafter called fluctuations. According to this splitting, we consider three types of noise models: (i) square-integrable offset and uniformly bounded conditional variance of the fluctuations with respect to the sampling measure, (ii) square-integrable offset and bounded fluctuations, (iii) bounded offset and fluctuations. Using arguments coming from the theory of large deviations [8,22], we prove in Theorem 9 a probabilistic bound for the fluctuation term in the discrete least-square projection, *i.e.* taking out the effect of the offset. Afterwards, exploiting Theorem 9, for each one of the aforementioned noise models we prove convergence estimates in probability for the discrete least-square projection error when a specific condition is satisfied between the number of pointwise evaluations and the dimension of the underlying approximation space. The derived convergence estimates relate the L^2 approximation error of the discrete least-squares approximation with the best approximation error measured either in the L^2 norm or in the L^∞ norm. These probability estimates do not require the use of any truncation operator. Moreover, we prove a convergence estimate in expectation with the unbounded noise model, that generalizes a result previously given in [5] to the case of nonzero offset. Our convergence estimates, both in probability and in expectation, separate the contribution to the error due to the best approximation error on a given approximation space and the contribution due to the presence of noise, similarly to the so-called *bias–variance trade off*, see *e.g.* [6,17].

Finally we apply our results to the particular setting of multivariate polynomial approximation spaces, which is a provably effective choice in many situations where a smooth dependence on many parameters needs to be approximated. Examples of such a situation arise when approximating the parameter-to-solution map of many types of PDEs with stochastic data, see *e.g.* the monographs [9,11] or the works [4,12,15] focused on discrete least squares. In [5,16,4], discrete least squares on multivariate polynomial spaces with evaluations at random points have been analyzed with the uniform and arcsine density: in any dimension and with polynomial spaces associated with downward closed multi-index sets, stability and accuracy have been proven, provided a specific proportionality relation is satisfied between the number of evaluations and the dimension of the polynomial approximation space. Then the analysis has been extended to any density in the beta family, using the results proven in [13].

In [14] it has been proven that, in the case of uniform density and with anisotropic tensor product polynomial spaces in any dimension, the random point set can be replaced by suitable low-discrepancy point sets, leading to analogous results concerning stability and accuracy of discrete least squares in the noiseless case. These results can be combined with those of the present paper, to provide convergence estimates for discrete least squares with noisy evaluations at low-discrepancy point sets, rather than random point sets.

Another analysis of discrete least squares with deterministic points has been proposed in [23], with points that are asymptotically distributed according to the arcsine density.

The outline of the paper is the following. In Section 2 we introduce the discrete least-squares approximation, the observation models, the assumptions on the noise and the noise models. In Section 2.1 we briefly present the algebraic formulation of discrete least squares and in Section 2.2 we recall the results achieved in [5]. In Section 3 we present our estimates in expectation (Section 3.1) and in probability (Section 3.2). Several intermediate results used in the proofs of these estimates have been collected in Section 5 where, in particular, we derive an estimate for the noise term using arguments from the theory of large deviations. In Section 4 we apply our convergence estimates in the noisy case to the setting of polynomial approximation. Finally in Section 6 we draw some conclusions.