ARTICLE IN PRESS

Journal of Computational and Applied Mathematics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

Piecewise linear approximation methods with stochastic sampling sites

Junbin Li [a,*], Renhong Wang [a], Min Xu [a], Qin Fang [b]

[a] School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China

[b] College of Information, Dalian University, Dalian 116622, China

ARTICLE INFO

Article history:
Received 23 September 2016
Received in revised form 26 December 2016

MSC:
41A15

Keywords:
Piecewise linear approximation
Sampling sites
Normal distribution

ABSTRACT

We study a generalization of the classical piecewise linear approximation methods with equally spaced breaks by considering the sampling sites as random variables. The new methods are motivated by the facts that real-world data collected from what are perceived to be equally spaced sites suffer from random errors due to measurement inaccuracies and other known or unknown factors. We establish error estimates and convergence results under practical assumptions about the distribution of the sampling sites.

1. Introduction

Consider the knots sequence $\triangle_n = \{x_k\}_{k=0}^n$ with $x_k = k/n$. Let $x_{-1} := x_0,\ x_{n+1} := x_n$, and set

$$H_k(x) := \begin{cases} (x - x_{k-1})/(x_k - x_{k-1}), & x_{k-1} < x \le x_k, \\ (x_{k+1} - x)/(x_{k+1} - x_k), & x_k \le x < x_{k+1}, \\ 0, & \text{otherwise.} \end{cases}$$

These functions are called B-spline of order 2 for the knots sequence $\triangle_n$. Let $f(x)$ be a function defined on $[0, 1]$. Its piecewise linear approximant $I_n(f; x)$ with respect to $\triangle_n$ is given by

$$I_n(f; x) = \sum_{k=0}^{n} f(x_k)H_k(x). \tag{1}$$

One enjoys the piecewise linear approximation method for the following properties [1]:

1. *It reproduces linear functions.* This means that if $f(x)$ is linear, then $I_n(f; x) = f(x)$.
2. *It is variation diminishing.* This means that if $v(g)$ denotes the number of changes of sign in $[0, 1]$ of a function $g(x)$, then $v(I_n(f)) \le v(f)$.
3. *It is nearly optimal.* This means that

$$\operatorname{dist}(f, S_2(\triangle_n)) \le \|f - I_n(f)\| \le 2\operatorname{dist}(f, S_2(\triangle_n)),$$

* Corresponding author.
E-mail address: junbin@mail.dlut.edu.cn (J. Li).

http://dx.doi.org/10.1016/j.cam.2017.02.041

where $\|f - I_n(f)\|$, $S_2(\triangle_n)$, and $\text{dist}(f, S_2(\triangle_n))$ denote respectively the maximum norm of $f - I_n(f)$, the space of all piecewise linear functions on [0, 1] with breaks at $x_1, \ldots, x_{n-1}$, and the distance between f and $S_2(\triangle_n)$ with respect to maximum norm.

Moreover, properties 1 and 2 are the source of the shape preserving properties of the piecewise linear approximation method which is commonly employed in CAGD [2–4].

To form the approximant $I_n(f; x)$, one needs to sample f at sites $\{x_k\}_{k=0}^n$, and assembles them with "kits" $H_k(x)$. The sampling sites are $(n + 1)$ equally spaced points in the interval [0, 1]. In many real world problems, data collected from what are perceived to be equally spaced sites suffer from random errors due to signal delays, measurement inaccuracies, and other known or unknown factors. It motivates us to introduce a generalization of the piecewise linear approximant for which the sampling action takes place at scattered sites:

$$I_n^A(f; x) = \sum_{k=0}^{n} f(x_{n,k})H_k(x), \tag{2}$$

where $A = \langle x_{n,k}\rangle$ is a lower triangular array. If we take $A = \langle k/n\rangle$ in (2), the approximant $I_n^A(f; x)$ is exactly the classical piecewise linear approximant $I_n(f; x)$. In this paper we contemplate from the probabilistic perspective to study this generalized version of the piecewise linear approximation methods. Specifically, we focus on the following problem: Given $f \in C([0, 1])$ and $\varepsilon > 0$, draw $(n+1)$ points $x_{n,k} (k = 0, 1, \ldots, n)$ independently according to the normal distributions F_k with mean k/n and variance σ^2, respectively, and estimate the probability

$$P\left\{(x_{n,0}, x_{n,1}, \ldots, x_{n,n}) : \left\|I_n^A(f; x) - f(x)\right\| > \varepsilon\right\}. \tag{3}$$

This type of probabilistic estimate is similar to those in learning theory within the framework advocated by Cucker & Smale [5,6], with the difference that the estimate in learning theory is obtained without a priori knowledge of the probabilistic distributions the random variables $x_{n,0}, x_{n,1}, \ldots, x_{n,n}$ obey. Moreover, under some assumptions about the smoothness of approximated functions and the variance $\varepsilon > 0$, we provide the convergence order of $I_n^A(f; x)$ according to the probability estimated.

2. Error estimates

Let $x_{n,k} (k = 0, 1, \ldots, n)$ be random variables that obey the normal distributions F_k with mean $\text{E}(x_{n,k}) = k/n$ and variance $\text{Var}(x_{n,k}) = \sigma^2$. The following normal probability inequality is a standard result in probability theory. It can be found in many probability related textbooks (e.g. [7,8]).

Lemma 1. *If $x_{n,k}$ are random variables that obey the normal distribution with mean k/n and variance σ^2, then*

$$P\left(\left|x_{n,k} - \frac{k}{n}\right| \geq t\right) \leq \sqrt{\frac{2}{\pi}}\frac{\sigma}{t}e^{-\frac{t^2}{2\sigma^2}}, \quad t > 0. \tag{4}$$

Let $\omega(f, \delta) = \sup_{|u-v|\leq\delta} |f(u) - f(v)|$ denote the modulus of continuity of the function f. Then it is readily seen that

$$\left|I_n(f; x) - f(x)\right| \leq \omega\left(f, \frac{1}{n}\right). \tag{5}$$

In view of this and by using Lemma 1 we are ready to establish the main result of the paper.

Theorem 2. *Let $\varepsilon > 0$ and $f \in C([0, 1])$ be given. Suppose that $\omega\left(f, \frac{1}{n}\right) < \varepsilon/4$ and that $x_{n,k} (k = 0, 1, \ldots, n)$ are independently drawn according to the normal distributions F_k with mean k/n and variance σ^2. Then with $\alpha_n = \frac{1}{2}\frac{1}{n\omega\left(f, \frac{1}{n}\right)}$, we have the following probability estimate*

$$P\left\{\left\|I_n^A(f; x) - f(x)\right\| > \varepsilon\right\} \leq n\frac{\sqrt{2}\sigma}{\varepsilon\alpha_n}e^{-\left(\frac{\varepsilon\alpha_n}{\sqrt{2}\sigma}\right)^2}. \tag{6}$$

Please cite this article in press as: J. Li, et al., Piecewise linear approximation methods with stochastic sampling sites, Journal of Computational and Applied Mathematics (2017), http://dx.doi.org/10.1016/j.cam.2017.02.041