4-13-2012

Bounding the Errors in Constructive Function Approximation

Chaouki T. Abdallah

D. Docampo

BOUNDING THE ERRORS IN CONSTRUCTIVE FUNCTION APPROXIMATION

D. Docampo and C.T. Abdallah∗
ETSIT, Universidad de Vigo, Campus Universitario, Vigo 36207
∗EECE Department, University of New Mexico, USA
Tel: +3486 812134; fax: +3486 812116
e-mail: ddocampo@tsc.uvigo.es

ABSTRACT

In this paper we study the theoretical limits of finite constructive convex approximations of a given function in a Hilbert space using elements taken from a reduced subset. We also investigate the trade-off between the global error and the partial error during the iterations of the solution.
The results obtained constitute a refinement of well established convergence analysis for constructive iterative sequences in Hilbert spaces with applications in projection pursuit regression and neural network training.

1 Introduction

Continuous functions on compact subsets of $\mathcal{R}^d$ can be uniformly approximated by linear combinations of sigmoidal functions [1], [2]. The issue of how the error in the approximation is related to the number of sigmoidals used is one of paramount importance from the point of view of applications; it can be phrased in a more general way as the problem of approximating a given element (function) f in a Hilbert space H by means of an iterative sequence f_n, and has an enormous impact in establishing convergence results for projection pursuit algorithms [3], neural network training [4] and classification [5]. It has been shown that this problem can be given a constructive solution where the iterations taking place involve computations in a reduced subset G of H.

In this paper we formulate the problem in such a way that we can study the bounds for the error in the approximation, obtain the best possible trade-off between global and partial errors, and derive bounds for the global error when a prespecified partial error is fixed.

The rest of the paper is organized as follows: Section 2 will state the problem and highlight its practical implications. Section 3 will review Barron's and Dingankar's solutions to the problem, while Section 4 will provide the framework under which those solutions can be derived. Section 5 will analyze the limits of the global error and its relation to the partial errors at each step of the iterative process. Finally, Section 6 will close the paper with the conclusions and further work.

2 The Problem

Let G be a subset of a real or complex Hilbert space H, with norm $||.||$, such that its elements, g, are bounded in norm by some positive constant b. Let $\bar{co}(G)$ denote the convex closure of G (i.e. the closure of the convex hull of G in H). The first global bound result, attributed to Maurey [4], concerning the error in approximating an element of $\bar{co}(G)$ using convex combinations of n points in G, is the following:

Lemma 2.1 *Let f be an element of $\bar{co}(G)$ and c a constant such that $c > b^2 - ||f||^2 = b_f^2$. Then, for each positive integer n there is a point f_n in the convex hull of some n points of G such that: $||f - f_n||^2 \leq \frac{c}{n}$.*

The first constructive proof of this lemma was given by Jones [3] and refined by Barron [4]; it includes an algorithm to iterate the solution. The result is the following:

Theorem 2.1 *For each element f in $\bar{co}(G)$, let us define the parameter γ as follows:*

$$\gamma = \inf_{v \in H} \sup_{g \in G} \{||g - v||^2 - ||f - v||^2\}$$

[1]This work has been partially supported by AFOSR and BNR.

Let δ be a constant such that $\delta > \gamma$. Then, we can construct an iterative sequence f_n, f_n chosen as a convex combination of the previous iterate f_{n-1} and a $g_n \in G$, $f_n = (1-\lambda)f_{n-1} + \lambda g_n$, such that $||f - f_n||^2 \leq \frac{\delta}{n}$.

This new parameter, γ, is related to Maurey's b_f^2; just make $v = 0$ in the definition of γ to realize that $\gamma \leq b_f^2$.

The relation between this problem and the universal approximation property of sigmoidal networks was clearly established in [3], [4]; specifically, it has been proven that, under certain mild restrictions, continuous functions on compact subsets of $\mathcal{R}^d$ belong to the convex hull of the set of sigmoidal functions that one hidden layer neural networks can generate.

3 Constructive solutions

The proof of Theorem 2.1 given in [4] and [5] is based on the following lemma:

Lemma 3.1 *Given $f \in \bar{co}(G)$, for each element of $co(G)$, h, and $\lambda \in [0,1]$:*

$$\inf_{g \in G} ||f - (1-\lambda)h - \lambda g||^2 \leq (1-\lambda)^2 ||f - h||^2 + \lambda^2 \gamma \quad (1)$$

The main result is derived using an inductive argument. At step 1, find g_1 and ϵ_1 so that $||f - g_1||^2 \leq \inf_G ||f - g||^2 + \epsilon_1 \leq \delta$. This is guaranteed by (1), for $\lambda = 1$.

Let f_n be our iterative sequence in $co(G)$; assume that for $n \geq 2$, $||f - f_{n-1}||^2 \leq \delta/(n-1)$. It is then possible to choose among different values of λ and ϵ_n so that:

$$(1-\lambda)^2 ||f_{n-1} - f||^2 + \lambda^2 \gamma \leq \frac{\delta}{n} - \epsilon_n \quad (2)$$

At step n, select g_n such that $||f - (1-\lambda)f_{n-1} - \lambda g_n||^2 \leq$

$$\inf_{g \in G} ||f - (1-\lambda)f_{n-1} - \lambda g||^2 + \epsilon_n \quad (3)$$

Hence, using (1), (2), and (3) we get: $||f - f_n||^2 \leq \frac{\delta}{n}$, and that completes the proof of Theorem 2.1.

4 Discussion

The values of λ and ϵ_n in [4] and [5] are related to the parameter α, $\alpha = \delta/\gamma - 1$, in the following way:

$$[4] \quad : \quad \lambda = \frac{||f - f_{n-1}||^2}{\gamma + ||f - f_{n-1}||^2}; \qquad \epsilon_n = \frac{\alpha \delta}{n(n+\alpha)}$$

$$[5] \quad : \quad \lambda = \frac{1}{n}; \qquad \epsilon_n = \frac{\alpha \gamma}{n^2}$$

It can be shown [6] that admissible values of λ satisfying inequality (2) for positive values of ϵ_n fall in the following interval:

$$\frac{||f - f_{n-1}||^2}{\gamma + ||f - f_{n-1}||^2} \pm$$

$$\frac{1}{\gamma + ||f - f_{n-1}||^2} \sqrt{||f - f_{n-1}||^4 - ||f - f_{n-1}||^2 + \frac{\delta}{n}}$$

To evaluate the possible choices for the bound ϵ_n we use the induction hypothesis in inequality (2); values of λ should now satisfy

$$(1-\lambda)^2 \frac{\delta}{n-1} + \lambda^2 \gamma \leq \frac{\delta}{n} - \epsilon_n$$

Admissible values of λ for positive values of ϵ_n fall then in the interval:

$$\frac{1+\alpha}{n+\alpha} \pm \frac{n-1}{n+\alpha} \sqrt{\frac{\alpha(1+\alpha)}{n(n-1)}}$$

Figure 1 shows the bounds of the interval for λ as a function of n. Bounds are plotted using solid lines, the center of the interval dotted line, and λ in [5] dash dotted line. Note how this last value approaches the limits of the interval, resulting in a poorer value for ϵ_n, as will be shown later.

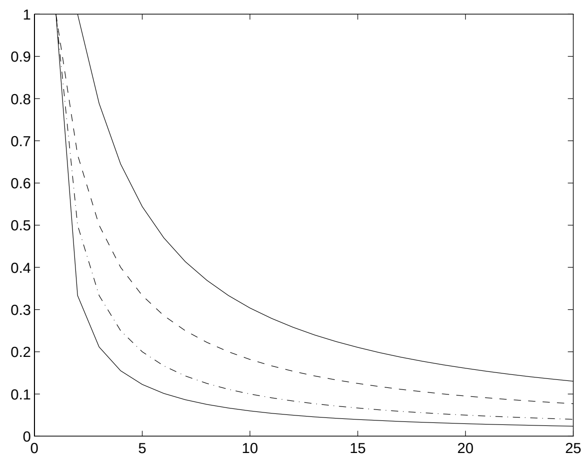

Figure 1: admissible values of λ as a function of n, for $\alpha = 1$

We now formulate the following questions:

1. Is it possible of achieve a further reduction in the error using convex combinations of n elements from G? What is the minimum bound for the global error assuming $\epsilon_n = 0$ for all n?

2. What is the optimal choice of λ for a given bound, so that ϵ_n is maximum, making the quasi-optimization problem at each step easier to solve?

3. For a prespecified partial error, ϵ_n, what is the bound for the global approximation problem?

Based on the assumptions made and in Lemma 3.1, let us formulate the problem again in a more general way: We look for a constructive approximation so that the overall error using n elements from G satisfies

$$||f - f_n||^2 \leq \frac{\delta}{b(n)} \tag{4}$$

$b(n)$ being a function of the parameter n which indicates the order of our approximation (i.e. $b(n) = n$ both in [3] and [5]) and δ the parameter related to γ as defined before. Let $f_n = (1-\lambda)f_{n-1} + \lambda g_n$; we want to find λ, ϵ_n, and the function $b(n)$ so that:

$$\begin{aligned} ||f - f_n||^2 &\leq \inf_{0<\lambda<1} \inf_{g \in G} ||f - (1-\lambda)f_{n-1} + \lambda g||^2 + \epsilon_n \\ &\leq \inf_{0<\lambda<1} (1-\lambda)^2 ||f - f_{n-1}||^2 + \lambda^2 \gamma + \epsilon_n \\ &\leq \inf_{0<\lambda<1} (1-\lambda)^2 \frac{\delta}{b(n-1)} + \lambda^2 \gamma + \epsilon_n \qquad (5) \\ &\leq \frac{\delta}{b(n)} \end{aligned}$$

Since $\delta = (1+\alpha)\gamma$, we can rewrite the last inequality:

$$\inf_{0<\lambda<1} (1-\lambda)^2 \frac{\delta}{b(n-1)} + \lambda^2 \delta + \epsilon_n - \lambda^2 \alpha \gamma \leq \frac{\delta}{b(n)}$$

This last expression represents the trade-off between the global error, $\delta / b(n)$, and the error at each of the sub-problems, ϵ_n. In reference [6] we have proved that setting $\epsilon_n = \lambda^2 \alpha \gamma$, then, for a given λ, the best rate of convergence of the approximation which can be achieved, measured in $b(n)$, is the one given in [4] and [5]; the value of λ which minimizes ϵ_n for that rate of convergence is precisely the value given in [5].

5 Limits and Bounds of the Approximation

Looking back at expression (5), we notice that, after using Lemma 3.1, we deal at each step with a quadratic problem in λ, which consists of minimizing

$$Q(\lambda_n) = (1-\lambda_n)^2 \frac{\delta}{b(n-1)} + \lambda_n^2 \gamma$$

provided that the induction hypothesis (4) is satisfied for $k < n$. We have introduced the notation λ_n to stress the variation of this parameter along the iterative process.

Writing $dQ(\lambda_n)/d\lambda_n = 0$, we get

$$\lambda_n = \frac{(1-\lambda_n)\delta}{\gamma b(n-1)} = \frac{1+\alpha}{1+\alpha+b(n-1)} \tag{6}$$

$$\begin{aligned} ||f - f_n||^2 &\leq (1-\lambda_n)^2 \left[\frac{\delta}{b(n-1)} + \frac{(1+\alpha)\delta}{b^2(n-1)} \right] + \epsilon_n \\ &= \frac{\delta}{1+\alpha+b(n-1)} + \epsilon_n = \frac{\delta}{b(n)} \qquad (7) \end{aligned}$$

$$\epsilon_n = \frac{\delta\,[1+\alpha+b(n-1)-b(n)]}{b(n)(1+\alpha+b(n-1))} \tag{8}$$

We conclude that there is a fundamental limitation in the rate of convergence that can be achieved under the hypothesis made so far, namely:

$$b(n) - b(n-1) \leq 1 + \alpha = \frac{\delta}{\gamma}$$

5.1 Minimum Global Error

Assuming that we can solve the partial approximation problems at each step of the iteration, so $\epsilon_n = 0, n \geq 1$, we have proved in [6] that the best rate of convergence that can be obtained follows the law c/n; the minimum value of the constant is $c = \gamma$.

Note that for this minimum to be reached we have

$$\lambda = \frac{1+\alpha}{(1+\alpha)n} = \frac{1}{n}$$

so the optimal convex combination would be the average of n elements from G, as in [5].

5.2 Fixing the rate of convergence

The problem of finding the maximum ϵ_n for a fixed convergence rate was also discussed in [6].

The value $\lambda = (1+\alpha)/(n+\alpha)$ solves the optimization problem; the best upper bound we can achieve for the partial error at each step of the iteration process coincides with Barron's bound, and is always greater than the bound found in [5].

5.3 Fixing ϵ_n

Given the nonlinear character of the recursion involved in (7), there is no analytical procedure to find a closed expression for $b(n)$. However, we can compute the bound of the approximation following the flow diagram of the optimal procedure, and derive from it some asymptotic results.

1. Select a constant δ such that $\delta \geq \gamma$; let $\delta = (1+\alpha)\gamma$.
2. Find $g_1 \in G$ so that $||f - g_1||^2 < \delta$. Set $f_1 = g_1$.
3. For $n > 1$, evaluate:
   (a) $\lambda_n = (1+\alpha)/\left(1+\alpha+b(n-1)\right)$ from (6)
   (b) Find $g_n \in G$ so that
   $$||f - (1-\lambda_n)f_{n-1} - \lambda_n g_n||^2 \leq$$
   $$\inf_G ||f - (1-\lambda_n)f_{n-1} - \lambda_n g||^2 + \epsilon_n$$
   (c) Make $f_n = (1-\lambda_n)f_{n-1} + \lambda_n g_n$
   (d) Compute $b(n)$ from (7)

In order to make the appropriate comparisons with previous results, we will set $\epsilon_n = (\alpha\gamma/n^2)$, as in [5]. Then, again under the induction hypothesis,

$$\frac{1}{b(n)} = \frac{1}{1+\alpha+b(n-1)} + \frac{\alpha}{(1+\alpha)n^2}$$

To predict the asymptotic behavior of $b(n)$, let us assume that, at step $n-1$, $b(n-1) \geq \beta(1+\alpha)(n-1)$, we will then prove that, for some values of the constant β, we can imply that also $b(n) \geq \beta(1+\alpha)n$.

Since $b(n-1) \geq \beta(1+\alpha)(n-1)$,

$$\frac{1}{b(n)} \leq \frac{1}{(1+\alpha)(1+\beta(n-1))} + \frac{\alpha}{(1+\alpha)n^2} \Rightarrow$$

$$\frac{b(n)}{n(1+\alpha)} \geq \frac{n(1+\beta(n-1))}{n^2+\alpha(1+\beta(n-1))} \Rightarrow$$

$$\frac{b(n)}{n(1+\alpha)} \geq \beta \quad \Leftrightarrow \quad n(1-\beta) \geq \beta\alpha(1+\beta(n-1)) \Leftrightarrow$$

$$n(1-\beta-\beta^2\alpha) \geq \beta\alpha(1-\beta)$$

This last inequality is asymptotically fulfilled for any value of β such that:

$$0 \leq \beta \leq \frac{\sqrt{4\alpha+1}-1}{2\alpha}$$

Then, for the value of ϵ_n selected in [5], the asymptotic value for $b(n)$ is:

$$b(n) = (1+\alpha)n\frac{\sqrt{4\alpha+1}-1}{2\alpha}$$

which is a better rate than the one obtained in [5].

6 Conclusions and Further Work

We have studied in this paper a framework where constructive algorithms based on convex combinations of elements from a subset of Hilbert space can be derived. We have obtained the optimal values for the coefficients in the convex expansions to guarantee a desired convergence rate. We have also studied the trade-off between global and partial errors for that optimal value.

Our further work includes the study of different convex combinations, which might include back-fitting of previously computed values, and the design of a practical algorithm that achieves the bounds obtained.

References

[1] G. Cybenko, "Approximation by superpositions of a sigmoidal function", *Math. Contr. Signals, Syst.*, **2** 313–314 (1989).

[2] K. Hornik, M. Stinchcombe and H. White, "Multilayer feedforward networks are universal approximators", *Neural Networks*, **2** 359–366 (1989).

[3] L.K. Jones, "A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training", *The Annals of Statistics* vol. **20** 608–613 (1992)

[4] A.R. Barron, "Universal approximation bounds for superpositions of a sigmoidal function", *IEEE Trans. on Inf. Theory* vol. **39**930–945(1993)

[5] A.T. Dingankar and I.W. Sandberg, "A note on Error Bounds for Approximations in Inner Product Spaces", To appear in *Circuits, Systems and Signal Processing*, (1996)

[6] D. Docampo, C.T. Abdallah and D. Hush, "Error Bounds in Constructive Approximation", to appear in *Proceedings of Fourth COST Bayona Workshop on Intelligent Methods in Signal Processing and Communications.*