On Approximation of Linear Functionals on L_p Spaces

Irwin W. Sandberg and Ajit Dingankar
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, Texas 78712

Abstract—
In a recent paper certain approximations to continuous nonlinear functionals defined on an L_p space ($1 < p < \infty$) are shown to exist. These approximations may be realized by sigmoidal neural networks employing a linear input layer that implements finite sums of integrals of a certain type. In another recent paper similar approximation results are obtained using elements of a general class of continuous linear functionals. In this note we describe a connection between these results by showing that every continuous linear functional on a compact subset of L_p may be approximated uniformly by certain finite sums of integrals.

I. INTRODUCTION

One of the earliest results in the area of neural networks is the proposition that any continuous real function defined on a compact subset of $\mathbb{R}^k$ (k an arbitrary positive integer) can be approximated arbitrarily well using a single-hidden-layer network with sigmoidal nonlinearities (see, for example, [2]). Among other results in the literature concerning arbitrarily good approximation that concern more general types of "target" functionals, different network structures, other nonlinearities, and various measures of approximation errors is the proposition in [5], [4] that any continuous real nonlinear functional on a compact subset of a real normed linear space can be approximated arbitrarily well using a single-hidden-layer neural network with a linear functional input layer and exponential (or polynomial or sigmoidal) nonlinearities.

In an interesting paper [1] by Chen and Chen similar results concerning single-hidden-layer neural networks with sigmoidal nonlinearities are proved for continuous real functionals on a compact subset of an L_p space ($1 < p < \infty$). One of the main results in [1] is the proposition that the linear functionals in the input layer may have the special form of a finite sum of integrals of a certain type. Here we connect these results by showing that every continuous linear functional can be approximated arbitrarily well on compact sets by finite sums of the type used in [1].

More specifically, the main result in [4] shows that any real continuous functional defined over a compact subset of a real normed linear space can be approximated arbitrarily well with a neural network employing an input layer of functionals that can be taken to be linear, and a single-hidden-layer implementing nonlinearities that, for example, can be taken to be sigmoidal. To describe the result proved there, let X be a real normed linear space, and let X^* be the set of bounded linear functionals on X (i.e., the set of bounded linear maps from X to the reals $\mathbb{R}$). Given a compact subset C of X, let Y be any set of continuous maps from X to $\mathbb{R}$ that is dense in X^* on C, in the sense that for each $\phi \in X^*$ and any $\rho > 0$ there is a $y \in Y$ such that $|\phi(x) - y(x)| < \rho$, $x \in C$. Also, let U be any set of continuous maps $u : \mathbb{R} \to \mathbb{R}$ such that given $\sigma > 0$ and any bounded interval $(\beta_1, \beta_2) \subset \mathbb{R}$ there exists a finite number of elements $u_1, \ldots, u_\ell$ of U for which $|\exp(\beta) - \sum_j u_j(\beta)| < \sigma$ for $\beta \in (\beta_1, \beta_2)$.[1] The result in [4] is this: Let g be a real-valued continuous map defined on C. Then given $\epsilon > 0$ there are a positive integer k, real numbers $c_1, \ldots, c_k$, elements $u_1, \ldots, u_k$ of U, and elements $y_1, \ldots, y_k$ of Y such that

$$|g(x) - \sum_j c_j u_j[y_j(x)]| < \epsilon$$

for $x \in C$.

The proof in [4] of the result described above shows that the result can be slightly sharpened in that "$y_1, \ldots, y_k$ of Y" can be replaced with "$y_1, \ldots, y_k$ of Y_ρ for some $\rho > 0$," where each Y_ρ is any set of continuous maps from X to $\mathbb{R}$ with the property that given $\phi \in X^*$ with $||\phi|| \leq 1$ there is a $y \in Y_\rho$ such that $|\phi(x) - y(x)| < \rho$, $x \in C$. Now suppose that $X = L_p$ with $1 < p < \infty$, and that $U = \{u : u(\beta) = cs(w\beta + \gamma),\ c, w, \gamma \in \mathbb{R}\}$, where s is a sigmoidal function. In this setting we obtain one of the main results in [1] (assuming continuous as opposed to generalized [1] sigmoids) as a special case of the above slightly-sharpened result, since our theorem below and a simple observation in Section 2.3 about its proof permit us to take the Y_ρ to be sets of finite linear combinations of integral functionals of the form considered in [1].[2] That is, in the L_p setting, for each ρ there is an $h > 0$ such that we can take the value at x of the elements y of Y_ρ to be given by certain finite linear combinations of integrals of the form

$$\frac{1}{2h} \int_{t-h}^{t+h} x(\tau)\, d\tau$$

with different values of t.

Besides relating the two works [1] and [4], the theorem given in the next section is felt to be interesting in itself.

[1] Of course we can take U to be the set whose only element is $\exp(\cdot)$, or the set $\{u : u(\beta) = (\beta)^n/n!,\ n \in \{0, 1, \ldots\}\}$.

[2] We also extend the result in [1] to $p = 1$.

II. APPROXIMATION THEOREM

A. Notation and Definitions

Let L_p, $1 \le p < \infty$, denote the space of real-valued Lebesgue-measurable functions defined on a finite interval $[a,b]$ that are p^{th} power integrable, and let $||\cdot||$ denote the usual norm. For each integrable function $x\colon [a,b] \to I\!R$ and each $h > 0$, let the so-called *Steklov function* be defined by

$$x_h(t) = \frac{1}{2h}\int_{t-h}^{t+h} x(\tau)\,d\tau, \quad t \in [a,b]$$

where (abusing the notation) x in the integral is an extension of x that vanishes outside $[a,b]$. Note that x_h is continuous on $[a,b]$ and that, in particular, it belongs to L_p. For a given $n \in I\!N$ and $h > 0$, we use $\mathcal{S}_{n,h}$ to denote

$$\left\{ s\colon L_p \to I\!R \,|\, s(x) = \sum_{j=1}^{n} c_j\, x_h(\alpha_j), \quad c_j \in I\!R, \quad 1 \le j \le n \right\}$$

in which the α_j are the elements of the uniformly spaced grid of n points $\alpha_j = a + j(b-a)/n$, $1 \le j \le n$. The set $\mathcal{S}_{n,h}$ is a family of what might be called Steklov functionals of order n and width h.

The class $\mathcal{S}$ of all Steklov functionals is defined by

$$\mathcal{S} = \bigcup_{\substack{n\,\in\, I\!N \\ h\,>\,0}} \mathcal{S}_{n,h}.$$

B. The Theorem

Our result is the following.

Theorem: Let f be a real continuous linear functional defined on L_p and let C be a compact subset of L_p. Then for each $\epsilon > 0$ there exists an $s \in \mathcal{S}$ such that

$$|f(x) - s(x)| < \epsilon$$

for all $x \in C$.

Proof

Let $\epsilon > 0$ be given. By the characterization of relatively compact sets[3] inL_p, $1 < p < \infty$, due to Kolmogorov as extended to $p = 1$ by A. N. Tulaikov (see [3], pp. 212-216), for any $\delta > 0$ there exists an $h > 0$ such that $||x - x_h|| < \delta$ for all $x \in C$. Using the uniform continuity of f, choose h so that

$$|f(x) - f(x_h)| < \epsilon/3, \;\; x \in C. \tag{1}$$

Since f is a continuous linear functional on L_p, we have

$$f(x_h) = \int_a^b k(\tau)x_h(\tau)\,d\tau \tag{2}$$

for some $k \in Lq$ where q is the conjugate index of p. Since the map taking $x \in C$ to x_h in the normed space of continuous functions is continuous, we see that $\{x_h \,|\, x \in C\}$ is compact, and so $M = \sup_{x\,\in\, C}\ \sup_{t\,\in\,[a,b]} |x_h(t)|$ is finite. Using the integrability of k, let g be a step function on $[a,b]$ such that

$$M\int_a^b |k(\tau) - g(\tau)|\,d\tau < \epsilon/3. \tag{3}$$

By (1), (2), (3) and the triangle inequality in $I\!R$,

$$|f(x) - \int_a^b g(\tau)x_h(\tau)\,d\tau| < 2\epsilon/3, \;\; x \in C. \tag{4}$$

We will now show that there is an n and a $s \in \mathcal{S}_{n,h}$ such that

$$|\int_a^b g(\tau)x_h(\tau)\,d\tau - s(x)| < \epsilon/3, \;\; x \in C, \tag{5}$$

and this will complete the proof. (Recall that h has been chosen.) We use the following lemma.

Lemma ([4]): Let A and B be normed linear spaces, and let K be a compact subset of A. Suppose that $G: K \to B$ is Lipschitz. Let $\mathcal{F}$ be any family of Lipschitz maps $H: K \to B$ such that given any $\delta > 0$ and any finite set $\{x_1, \ldots, x_r\}$ of points of K there is an $H \in \mathcal{F}$ for which

$$||Gx_i - Hx_i|| < \delta, \quad 1 \le i \le r.$$

Then given $\epsilon > 0$ there is an $H \in \mathcal{F}$ such that $||Gx - Hx|| < \epsilon$ for all $x \in K$.

It is easy to check that the map from $x \in C$ to $\int_a^b g(\tau)x_h(\tau)\,d\tau$ as well as the elements of $\cup_{n\,\in\, I\!N}\mathcal{S}_{n,h}$ are Lipschitz. Referring to the lemma, take $A = L_p, K = C, B = I\!R$, $\mathcal{F} = \cup_{n\,\in\, I\!N}\mathcal{S}_{n,h}$ and let G be defined by

$$Gx = \int_a^b g(\tau)x_h(\tau)\,d\tau. \tag{6}$$

Now choose any $x \in C$ and any $\delta > 0$. Since gx_h is bounded and sectionally continuous on $[a,b]$, there is an $n_0(x) \in I\!N$ with the following property. For $n \in I\!N$ with $n \ge n_0(x)$ the Riemann partition $\alpha_j = a + j(b-a)/n$, $0 \le j \le n$, of $[a,b]$ is such that

$$|\int_a^b g(\tau)x_h(\tau)\,d\tau - \sum_{j=1}^{n} g(\alpha_j)x_h(\alpha_j)(b-a)/n| < \delta.$$

Thus, given a set $\{x_1, \ldots, x_r\}$ of points of C we have

$$|Gx_i - \sum_{j=1}^{n} c_j\, x_{ih}(\alpha_j)| < \delta, \quad 1 \le i \le r.$$

for $n = \max_i\{n_0(x_i)\}$, in which $c_j = g(\alpha_j)(b-a)/n$. By the lemma there is an $s \in \cup_{n\,\in\, I\!N}\mathcal{S}_{n,h}$ such that (5) holds. This completes the proof of the theorem.[4]

[3] In [3] the term "compact" means relatively compact.

[4] It is possible to give a shorter proof of this result using deeper propositions from the theory of integration, but the approach used above is felt to be much more accessible to readers of these transactions.

C. Comments

A check of the proof shows that h (which depends on C and ϵ) can be chosen to be independent of f for f of norm at most unity, and that the theorem holds also for the case in which f is complex valued, L_p is a complex space, and the coefficients c_j are complex numbers.

The last part of the proof can be generalized to yield an interesting result concerning what might be called the "uniform Riemann integrability" of certain integrands. This result, which might be useful in other approximation studies, is discussed in the appendix.

APPENDIX

Uniform Integrability

Here let K be a compact subset of a normed linear space A, and consider Riemann-Stieltjes integrals of the form

$$\int_a^b u(x)(\tau)\, d\beta(\tau)$$

in which $x \in K$ and $u(x)(\cdot)$ and β are real valued. Suppose that the integral exists[5] for each x, that it is Lipschitz with respect to x, and that $u(\cdot)(\tau)$ is Lipschitz for each $\tau \in [a,b]$. Then by an argument similar to that at the end of the proof in Section 2.2 it follows that $u(x)(\cdot)$ is uniformly integrable in (for example) the sense that for each $\epsilon > 0$ there is a partition $\alpha_j = a + j(b-a)/n, \quad 0 \leq j \leq n$, of $[a,b]$ such that

$$|\int_a^b u(x)(\tau)\, d\beta(\tau) - \sum_{j=1}^{n} u(x)(\alpha_j)[\beta(\alpha_j) - \beta(\alpha_{(j-1)})]| < \epsilon$$

for all $x \in K$. This result can be extended in a few directions. For example, A need not be a normed linear space; it can be a metric space.[6] And $u(x)(\alpha_j)$ in the sum can be replaced with $u(x)(\lambda\alpha_j + (1-\lambda)\alpha_{(j-1)})$ for any given $\lambda \in [0,1]$.

REFERENCES

[1] T. Chen and H. Chen, "Approximations of continuous functionals by neural networks with application to dynamical systems," *IEEE Transactions on Neural Networks*, vol. 4, no. 6, pp. 910-918, November 1993.

[2] G. Cybenko, "Approximation by Superposition of a Single Function," *Mathematics of Control, Signals and Systems*, Vol. 2, pp. 303-314, 1989.

[3] I. Natanson, *Theory of Functions of a Real Variable*, Frederick Ungar Publishing Co., New York, 1955.

[4] I. W. Sandberg, "General structures for classification," *IEEE Transactions on Circuits and Systems - I*, vol. 41, no. 5, pp. 372-376, May 1994.

[5] I. W. Sandberg, "Approximations for nonlinear functionals," *IEEE Transactions on Circuits and Systems - I*, vol. 39, no. 1, pp. 65-67, January 1992.

[5] There is more than one standard definition of a Riemann-Stieltjes integral. We have in mind the definition in which the approximating finite sum must converge in $\mathbb{R}$ as the norm of the (not necessarily uniform) partition approaches zero.

[6] As mentioned in [4] the lemma in Section 2.2 has a natural extension to a metric space setting.