Some Density Problems in Multivariate Approximation

Allan Pinkus

Abstract. In this report we will review some density problems in multivariate approximation. Some of these problems are related to approximation by radial basis functions and ridge functions. Others are motivated by various mathematical models in the theory of (artificial) neural networks and computerized tomography.

§1. Introduction

We will look at a few elementary but not necessarily simple examples related to the approximation of *multivariate functions*, and ask one of the basic questions of approximation theory. Namely, is the subspace dense? If the answer is **yes**, then we can continue to ask for "good methods of approximation" and many of the other related standard questions of approximation theory. If the answer is **no**, then we have a problem, and so it is important that we know when it is yes.

Some of the questions we will look at were motivated by problems of (artificial) neural networks (ANN), or computerized tomography, and some by the study of radial basis functions, ridge functions and the like. Sometimes the problems seem historically to be more connected with harmonic analysis, integral geometry and even algebraic geometry. However all these problems can and should be considered as part of the repertoire of approximation theory.

We will talk around the following class of problems. Fix $f : \mathbb{R}_+ \to \mathbb{R}$. Let $\|\cdot\|$ denote the usual ℓ_2 norm on $\mathbb{R}^n$. We will consider:

A) $\text{span}\{f(\|\mathbf{x}-\mathbf{a}\|) : \mathbf{a} \in \mathbb{R}^n\}$
B) $\text{span}\{f(\rho\|\mathbf{x}-\mathbf{a}\|) : \mathbf{a} \in \mathbb{R}^n\,, \rho > 0\}$
C) $\text{span}\{g(\|\mathbf{x}-\mathbf{a}\|) : g \in C(\mathbb{R}_+)\,, \mathbf{a} \in \mathcal{A}\}$ (here g varies but the centers $\mathbf{a}$ come from a fixed set $\mathcal{A}$).

We ask when we have density. We will consider this question in $C(\mathbb{R}^n)$ with the topology of uniform convergence on compact subsets. That is, given *any* compact $K \subset \mathbb{R}^n$, when is the above dense in $C(K)$? We will also consider what is known, if anything, if we ask this same question in $L^2(\mathbb{R}^n)$.

§2. Problem A

We fix the function f and consider

$$\text{span}\{f(\|\mathbf{x}-\mathbf{a}\|) : \mathbf{a} \in \mathbb{R}^n\}\,.$$

This is a class of *radial basis functions* with arbitrary (all) centres, and we are asking for when this space has the density property. Note that this space is *translation and rotation invariant.*

(i) $L^2(\mathbb{R}^n)$. It is well-known that a translation invariant subspace is not dense if and only if there exists a set of positive Lebesgue measure E such that $\widehat{h}|_E = 0$ for all h in the space. This is a result due to N. Wiener. Thus if $g \in L^2(\mathbb{R}^n)$, then

$$\mathrm{span}\{g(\mathbf{x}-\mathbf{a}) : \mathbf{a} \in \mathbb{R}^n\}$$

is dense in $L^2(\mathbb{R}^n)$ if and only if $\widehat{g}$ does **not** vanish on a set of positive Lebesgue measure. (Somewhat similar results hold in $L^p(\mathbb{R}^n)$, $1 < p < \infty$, see Edwards [2].) How can we use the fact that we are here considering not the general situation, but only specific g? Namely radial functions, i.e., $f(\|\cdot\|)$ where f is a univariate functions defined on $\mathbb{R}_+$. It is known that in this case the Fourier transform is also radial and at $\mathbf{w}$ equals

$$\frac{2\pi}{\|\mathbf{w}\|^{(n-2)/2}} \int_0^\infty f(r) r^{n/2} J_{\frac{n-2}{2}}(\|\mathbf{w}\|r)\,dr$$

if $n \geq 2$, where $J_{\frac{n-2}{2}}$ is the Bessel function of order $\frac{n-2}{2}$. Aside from this, I know nothing more.

(ii) $C(\mathbb{R}^n)$. In this topology functions, for which the span of all their shifts are not dense, have been studied. They are called *mean-periodic* functions. There is no known characterization of mean-periodic functions in $C(\mathbb{R}^n)$ for $n \geq 2$. However many functions cannot be mean-periodic. For example

Proposition. *If* $g \in C(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$, $g \neq 0$, *then*

$$C(\mathbb{R}^n) = \overline{\mathrm{span}}\{g(\mathbf{x}-\mathbf{a}) : \mathbf{a} \in \mathbb{R}^n\}\,.$$

Proof: The continuous linear functionals on $C(\mathbb{R}^n)$ are represented by Borel measures of finite total variation and compact support. If the above space is not dense in $C(\mathbb{R}^n)$, then there exists such a measure μ such that

$$\int_{\mathbb{R}^n} g(\mathbf{x}-\mathbf{a})\,d\mu(\mathbf{x}) = 0\,,$$

for all $\mathbf{a} \in \mathbb{R}^n$. Both g and μ have "nice" Fourier transforms. Since the above is a convolution we must have

$$\widehat{g}(\mathbf{w})\widehat{\mu}(\mathbf{w}) = 0\,.$$

Now $\widehat{\mu}$ is an entire function, while $\widehat{g}$ is continuous. Since $\widehat{g}$ must vanish where $\widehat{\mu} \neq 0$, it follows that $\widehat{g} = 0$ and thus $g = 0$, a contradiction. ■

Thus if $f(\|\cdot\|) \in C(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$, then density holds. This is not in the least necessary. It is known that for $f(r) = r$ density holds. I am unaware of any complete characterization of all $f \in C(\mathbb{R}_+)$ for which density holds. There is a conjecture that density does hold except for the f of the form

$$f(r) = \sum_{k=0}^{\infty} p_k(r) J_{\frac{n-2}{2}}(\alpha_k r)$$

where the p_k are even polynomials and $\alpha_k > 0$ are of a certain sparsity. We will not here go into the reasons for this conjecture.

§3. Problem B

Here we consider

$$\text{span}\{f(\rho\|\mathbf{x}-\mathbf{a}\|) : \mathbf{a} \in I\!R^n\,, \rho > 0\}$$

where again the function f is fixed. I was motivated to consider this problem by a mathematical model in ANN which we will not describe here. There are simple characterizations of the f for which this set spans in both $L^2(I\!R^n)$ and in $C(I\!R^n)$, although the result and especially the proofs are rather different.

(i) $L^2(I\!R^n)$. *If* $f(\|\cdot\|) \in L^2(I\!R^n)\backslash\{0\}$, *then*

$$L^2(I\!R^n) = \overline{\text{span}}\{f(\rho\|\mathbf{x}-\mathbf{a}\|) : \mathbf{a} \in I\!R^n\,, \rho > 0\}\,.$$

Proof: Since this space is translation invariant it follows from the theorem of N. Wiener, mentioned earlier, that the space spans all $L^2(I\!R^n)$ if and only if there does not exist a set of positive Lebesgue measure E such that $\widehat{h}|_E = 0$ for all h in the space. Since $f \neq 0$, its Fourier transform is not identically zero. Furthermore the Fourier transform of f is also radial. Dilation by ρ, as above, essentially translates into dilation by $1/\rho$ in the Fourier transform. From this it follows that the Fourier transform of all functions in the space cannot all vanish on a set of positive Lebesgue measure. ■

(ii) $C(I\!R^n)$. The following result is to be found in [5].

If $f \in C(I\!R^n)$, *then*

$$C(I\!R^n) = \overline{\text{span}}\{f(\rho\|\mathbf{x}-\mathbf{a}\|) : \mathbf{a} \in I\!R^n\,, \rho > 0\}\,,$$

if and only if f *is not an even polynomial.*

Proof: If f is an even polynomial of degree $2m$, then $f(\rho\|\mathbf{x}-\mathbf{a}\|)$ is an even polynomial of degree $2m$ for every choice of $\mathbf{a} \in I\!R^n$ and $\rho > 0$. Thus the above space cannot possibly span all of $C(I\!R^n)$.

For the other direction we use a little known result of L. Schwartz [8] from 1944.

Theorem. *Let* $g \in C(I\!R^n)$. *The space*

$$\text{span}\{g(\lambda x_1 - a_1, \ldots, \lambda x_n - a_n) : a_1, \ldots, a_n \in I\!R\,, \lambda \in I\!R\}$$

is **not** *dense in* $C(I\!R^n)$ *if and only if there exists a homogeneous polynomial* p *such that*

$$p(D)g = 0$$

in the weak sense. If, in addition, the above space is rotation invariant, then it is not dense if and only if

$$\Delta^m g = 0$$

in the weak sense, for some m, where Δ is the Laplacian.

It is not difficult to show that the only continuous solutions to this last equation for $g(\mathbf{x}) = f(\|\mathbf{x}\|)$ are the even polynomials. (This is an elliptic equation and so all (weak) solutions are in fact very smooth.) ■

(iii) One can also ask this same question in $L^1(\mathbb{R}^n)$. Because of the translation and rotation invariance it follows, from a different theorem of N. Wiener, that if $f(\|\cdot\|) \in L^1(\mathbb{R}^n)$ then

$$L^1(\mathbb{R}^n) = \overline{\text{span}}\{f(\rho\|\mathbf{x}-\mathbf{a}\|) : \mathbf{a} \in \mathbb{R}^n\,, \rho > 0\}\,,$$

if and only if

$$\int_{\mathbb{R}^n} f(\|\mathbf{x}\|)\, d\mathbf{x} \neq 0\,,$$

i.e., the Fourier transform of f does not vanish at zero.

We can also consider a slightly different problem. By

$$\text{span}\{f(\rho\|\mathbf{x}-\mathbf{a}\|) : \mathbf{a} \in \mathbb{R}^n\,, \rho > 0\}$$

we mean

$$\sum_{i=1}^{m} c_i f(\rho_i\|\mathbf{x}-\mathbf{a}^i\|)$$

for $c_i \in \mathbb{R}$, $\rho_i > 0$, $\mathbf{a}^i \in \mathbb{R}^n$ and any m finite. We could also demand that we restrict the ρ_i in that $\rho_i = \rho$ for $i = 1, \ldots, m$, (but the ρ is still variable). It is not difficult to show that even with this restrictive definition we get density in $L^1(\mathbb{R}^n)$ if and only if $\widehat{f}(0) \neq 0$. This follows using a standard mollifier argument wherein we let $\rho \to \infty$, see e.g. Park, Sandberg [4].

§4. Problem C

Here we consider

$$\text{span}\{g(\|\mathbf{x}-\mathbf{a}\|) : g \in C(\mathbb{R})\,, \mathbf{a} \in \mathcal{A}\}$$

where g varies over all of $C(\mathbb{R})$, but the centers $\mathbf{a}$ come from a fixed set $\mathcal{A}$. We are interested in determining conditions on the set $\mathcal{A}$ which guarantee density. This is a fundamentally different problem from (A) or (B). Let me first try to motivate it by explaining one philosophy behind this problem.

In a series of papers in the late 50's, Kolmogorov proved the following startling result which is considered to be the resolution, in the negative, of Hilbert's 13th problem.

Theorem. *There exist continuous functions of one variable h_{ij} ($i = 1, \ldots, 2n+1$, $j = 1, \ldots, n$) such that every continuous function f of n variables on $[0,1]^n$ can be represented in the form*

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{2n+1} g_i\Big(\sum_{j=1}^{n} h_{ij}(x_j)\Big)$$

where the g_i are continuous functions of one variable which depend on f.

There have been numerous generalizations of this theorem in various directions. Attempts to understand the nature of this theorem have also led to interesting concepts related to the complexity of functions. Nonetheless the theorem itself has had few, if any, direct applications. This seems rather surprising since it does state that we can and should consider multivariate functions as just compositions of univariate functions. In other words, we can represent (and not only approximate) multivariate functions as a finite (fixed number) of compositions of univariate functions. There have been a few, but very few, attempts to construct algorithms for approximating based on this theorem.

Perhaps one reason for the lack of success in this direction is related to the following result which is due to Vitushkin and Henkin (separately and together). It says that one cannot demand that the h_{ij} be C^1 functions. Since the answer is in the negative, I only formulate it for functions of two variables in $[0,1]^2$.

Theorem. *For any m fixed continuously differentiable functions h_i, $i = 1, \dots, m$, defined on $[0,1]^2$, the set of functions*

$$\Big\{ \sum_{i=1}^{m} g_i\big(h_i(x,y)\big) \;:\; g_i \in C(\mathbb{R}) \Big\}$$

is nowhere dense in the space of all functions continuous in $[0,1]^2$ with the topology of uniform convergence.

However there is a fundamental difference between the representation and the approximation of functions. In practise we should give up the expectation of exact representation. It is our hope that it might be possible, using smooth, calculable functions and the idea of superposition (composition), to develop good methods of approximating multivariate functions. This is one idea behind the following:

Consider a specific class Φ of functions $\phi : \mathbb{R}^n \to \mathbb{R}$ and study the linear space

$$\operatorname{span}\{g(\phi(\mathbf{x})) : \phi \in \Phi,\ g \in C(\mathbb{R})\}\,.$$

If Φ is composed of a finite number of smooth (C^1) functions, then this set is, by the above theorem, nowhere dense in the space of all functions continuous on any compact set in $\mathbb{R}^n$ (with interior) with the topology of uniform convergence. However if Φ is composed of an infinite set of functions, then this is no longer necessarily true.

What type of functions might we look at. One example is to look at the class

$$\Phi = \{ < \mathbf{a} \cdot \mathbf{x} > = \sum_{i=1}^{n} a_i x_i \;:\; \mathbf{a} \in \mathbb{R}^n \}\,.$$

We are then led to a consideration of the functions

$$\operatorname{span}\{g(< \mathbf{a} \cdot \mathbf{x} >) : \mathbf{a} \in \mathbb{R}^n, g \in C(\mathbb{R})\}\,.$$

This is the class of *ridge functions*. They are dense in $C(I\!R^n)$. In fact we can considerably restrict the set of admissible $\mathbf{a}$. It suffices that it runs over any set such that no non-trivial homogeneous polynomial vanishes thereon, see Lin, Pinkus [3].

Another example is to take a fixed function ϕ and consider Φ to be all (or a subset of) its shifts. That is, in this case we consider the approximating set

$$\text{span}\{g(\phi(\mathbf{x}-\mathbf{a})) : \mathbf{a} \in I\!R^n, g \in C(I\!R)\},$$

where ϕ is fixed. There are only a few things that one can say in this generality. One obvious necessary condition for density, in general, is that the set must separate points. (That is, not all functions in the space take on the same values at $\mathbf{x}$ and $\mathbf{y}$ for some $\mathbf{x} \neq \mathbf{y}$.) If ϕ is a polynomial then the demand of separation of points in this setting is equivalent to the linear independence of the first order partial derivatives of ϕ.

This necessary condition turns out to be also sufficient for $n = 2$ and $n = 3$. It is sufficient for $n = 4$ if the polynomial is homogeneous, but not in general. For $n \geq 5$ it is not the correct sufficient condition even if ϕ is homogeneous. However if ϕ is homogeneous and the first partial derivatives are algebraically independent, i.e., there does not exist a non-trivial polynomial p of n variables such that

$$p(\phi_{x_1}, \ldots, \phi_{x_n}) = 0,$$

then we necessarily have density. The above question is actually quite interesting and is connected to some work in algebraic geometry of more than a hundred years ago by L. O. Hesse, Max Noether, and Paul Gordon, (see Pinkus, Wajnryb [6] and [7]).

If ϕ (and it need not even be a polynomial) is such that

$$\{\mathbf{x} : c \leq \phi(\mathbf{x}) \leq d\}$$

is, for some $c < d$, a bounded nonempty set then we always have density. Thus, for example, the space

$$\text{span}\{g(\|\mathbf{x}-\mathbf{a}\|) : \mathbf{a} \in I\!R^n, g \in C(I\!R)\},$$

in dense in $C(I\!R^n)$. (This can also be proven by other methods.)

Thus we are led back to the problem we started this section with. We would like to characterize the sets $\mathcal{A}$ for which

$$\text{span}\{g(\|\mathbf{x}-\mathbf{a}\|) : \ g \in C(I\!R), \mathbf{a} \in \mathcal{A}\}$$

is dense in $C(I\!R^n)$.

Certain facts are relatively elementary to prove. For example, if $\mathcal{A}$ is not contained in the zero set of some polynomial (i.e., in an algebraic variety), then

$$C(I\!R^n) = \overline{\text{span}}\{g(\|\mathbf{x}-\mathbf{a}\|) : \ g \in C(I\!R), \mathbf{a} \in \mathcal{A}\}$$

In a recent preprint M. Agranovsky and E. T. Quinto [1] totally characterize the set of $\mathcal{A}$ for which density holds in $I\!R^2$. They prove what V. Lin and I had in fact conjectured. Their analysis is deep and lengthy. We will not delve into it here.

To explain the answer, assume that $\mathcal{A}$ is a *straight line* (or any subinterval thereof). Then the set of functions in the closure of

$$\mathrm{span}\{g(\|\mathbf{x}-\mathbf{a}\|) : \, g \in C(\mathbb{R}) \, , \mathbf{a} \in \mathcal{A}\}$$

is exactly the set of all continuous functions which are even about this straight line. Now if $\mathcal{A}$ is a *set of straight lines* in $\mathbb{R}^2$, then the above closure in not $C(\mathbb{R}^2)$ if and only if all these lines have a common intersection point, and the angles between each of the lines is a rational multiple of π.

If

$$C(\mathbb{R}^2) \neq \overline{\mathrm{span}}\{g(\|\mathbf{x}-\mathbf{a}\|) : \, g \in C(\mathbb{R}) \, , \mathbf{a} \in \mathcal{A}\}$$

then $\mathcal{A}$ is a subset of the set just described above, together with any finite number of points.

There is a dual version to this problem. It is connected with a uniqueness problem in the theory of spherical Radon transforms and thus to questions in integral geometry.

By duality, the space

$$\mathrm{span}\{g(\|\mathbf{x}-\mathbf{a}\|) : \, g \in C(\mathbb{R}) \, , \mathbf{a} \in \mathcal{A}\}$$

is not dense in $C(\mathbb{R}^n)$ if and only if there exists a non-trivial Borel measure μ of finite total variation and compact support such that

$$\int_{\mathbb{R}^n} g(\|\mathbf{x}-\mathbf{a}\|) d\mu(\mathbf{x}) = 0$$

for all $g \in C(\mathbb{R})$ and $\mathbf{a} \in \mathcal{A}$. This is easily seen to be equivalent to the fact that

$$\int_{B(\mathbf{a},r)} d\mu(\mathbf{x}) = 0$$

for all $\mathbf{a} \in \mathcal{A}$ and $r > 0$, where $B(\mathbf{a}, r)$ is the ball centered at $\mathbf{a}$ and of radius r. It can also be shown that the existence of such a μ implies the existence of an $h \in C(\mathbb{R}^n)$ with *compact support* such that

$$\int_{B(\mathbf{a},r)} h(\mathbf{x}) d\mathbf{x} = 0$$

for all $\mathbf{a} \in \mathcal{A}$ and $r > 0$. (I.e., we can assume that μ is sufficiently smooth.) This, in turn, is equivalent to

$$\int_{S(\mathbf{a},r)} h(\mathbf{x}) \, dA = 0$$

for all $\mathbf{a} \in \mathcal{A}$ and $r > 0$, where $S(\mathbf{a}, r)$ is the sphere centered at $\mathbf{a}$ and of radius r, and dA is the appropriate area measure on the sphere.

Our original problem is therefore equivalent to finding both necessary and sufficient conditions on $\mathcal{A}$ such that

$$\int_{S(\mathbf{a},r)} f(\mathbf{x}) \, dA = \int_{S(\mathbf{a},r)} g(\mathbf{x}) \, dA$$

for all $\mathbf{a} \in \mathcal{A}$ and $r > 0$, implies that $f = g$ if $f, g \in C(\mathbb{R}^n)$ and both (or their difference) have compact support.

References

1. Agranovsky, M. L., and E. T. Quinto, Injectivity sets for the Radon transform over circles and complete systems of radial functions, preprint.
2. Edwards, R. E., Spans of translates in $L^p(G)$, *J. Aust. Math. Soc.* **5** (1965), 216–233.
3. Lin, V. Ya., and A. Pinkus, Fundamentality of Ridge Functions, *J. Approx. Theory* **75** (1993), 295–311.
4. Park, J., and I. W. Sandberg, Universal approximation using radial-basis-function networks, *Neural Comp.* **3** (1991), 246–257.
5. Pinkus, A., TDI-Subspaces of $C(\mathbb{R}^d)$ and some density problems from neural networks, preprint.
6. Pinkus, A., and B. Wajnryb, Multivariate polynomials: a spanning question, to appear in *Constructive Approximation*.
7. Pinkus, A., and B. Wajnryb, A Problem of Approximation Using Multivariate Polynomials, to appear in *Uspekhi Mat. Nauk*.
8. Schwartz, L., Sur certaines familles non fondamentales de fonctions continues, *Bull. Soc. Math. France* **72** (1944), 141–145.

Acknowledgements. This research was supported by the Fund for the Promotion of Research at the Technion.

Allan Pinkus
Department of Mathematics
Technion, I. I. T.
Haifa, 32000
Israel