ELSEVIER
Journal of Computational and Applied Mathematics 55 (1994) 31-50
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

Differentiable approximation by means of the Radon transformation and its applications to neural networks ☆

Yoshifusa Ito *
Toyohashi University of Technology, Tempaku-cho, Toyohashi, Aichi-ken, 441 Japan

Received 5 June 1992; revised 15 January 1993

Abstract

We treat the problem of simultaneously approximating a several-times differentiable function in several variables and its derivatives by a superposition of a function, say g, in one variable. In our theory, the domain of approximation can be either compact subsets or the whole Euclidean space $\mathbb{R}^d$. We prove that if the domain is compact, the function g can be used without scaling, and that even in the case where the domain of approximation is the whole space $\mathbb{R}^d$, g can be used without scaling if it satisfies a certain condition. Moreover, g can be chosen from a wide class of functions. The basic tool is the inverse Radon transform. As a neural network can output a superposition of g, our results extend well-known neural approximation theorems which are useful in neural computation theory.

Keywords: Uniform approximation; Approximation of derivatives; Plane wave; Differentiable approximation; Delta sequence; Radon transform; Hahn-Banach theorem; Neural network

1. Introduction

Several papers published in 1989 dealt with the problem of approximating continuous functions in several variables by linear combinations of a sigmoid function: [1,2,4,6,9]. So many papers on the same topic appeared one after another in that year, probably because they recognized simultaneously the importance of the approximation theorem in neural computation theory. They have proved that any continuous function f defined on a compact set of $\mathbb{R}^d$ can be approximated uniformly by a linear combination

$$\bar{f}(x) = \sum_{i=1}^{n} a_i g(c_i(\omega_i \cdot x - t_i)) \tag{1.1}$$

☆ This paper was supported in part by Grant 04246221 from the Ministry of Education and Culture, Japan.
* e-mail: ito@tantalus.tutkie.tut.ac.jp.

SSDI 0377-0427(93)E0178-O

with any accuracy, where $x = (x^{(1)}, \ldots, x^{(d)}) \in \mathbb{R}^d$, $\omega_i \in \mathbb{R}^d$, $|\omega_i| = 1$, and the function g is a sigmoid function fixed in advance. In other words, they have proved that any continuous function defined on compact sets can be approximated by a linear combination of scaled shifted rotations of a given sigmoidal plane wave $g(\omega \cdot x)$. Suppose that a three-layered neural network has d input layer units, n second layer units and an output layer unit. Then, the network can output the finite sum (1.1), provided that the input layer units receive the components of $x \in \mathbb{R}^d$ respectively and distribute them to the second layer, the second layer units have d input gates and output the value of $g(c_i(\omega_i \cdot x - t_i))$ respectively, and the output layer unit outputs the weighted sum of outputs of the second layer units. Since the input-output functions of biological neurons are sigmoidal, a sigmoid function is often used as an input-output function of neural units in neural computation theory.

In [12-14], it is proved that (1) the domain of uniform approximation of continuous functions can be extended to $\mathbb{R}^d$, (2) any sigmoid function can be used as the function g in the linear combination (1.1) for uniform approximation on $\mathbb{R}^d$ if it can be scaled, (3) any sigmoid function can be used as the function g without scaling if the domain of approximation is compact. Further, we have obtained a necessary and sufficient condition on sigmoid functions which ensures that they can be used as the function g without scaling for approximation on $\mathbb{R}^d$ (see the Note in Section 5).

In [10], the importance of simultaneous approximation of derivatives of functions by neural networks in economics and robotics is mentioned. It is expected that, in addition to these fields, there are many other fields where simultaneous approximations by neural networks are useful, because a number of phenomena are described by differential equations and natural laws are basically governed by differential equations. In the case where we treat differential equations by neural networks, functions and their derivatives must be approximated simultaneously. Consequently, extension of the neural approximation theorems to derivatives must be important. Since such differential equations are often on functions defined on $\mathbb{R}^d$, our extension of differentiable approximation to the whole space $\mathbb{R}^d$ may be worthwhile. In [10] is proved that each derivative $\partial^\alpha f$ of any m-times differentiable function f can be approximated by the corresponding derivative $\partial^\alpha \bar{f}$ of a linear combination $\bar{f}$ of the form (1.1) if the function g is m-times continuously differentiable and the integral of the mth derivative over $\mathbb{R}$ is nonzero and integrable. Later, Hornik [8] extended these results by weakening the integrability condition of the function g. In [10], the accuracy of the simultaneous uniform approximation on compact sets is estimated by the supremum norm but not extended to the whole space $\mathbb{R}^d$.

In this paper, we extend not only [12-14], but also [8,10]. Our previous results will be extended in a sense that the approximation is extended to derivatives, and the latter will be generalized because we prove that the domain of approximation can be extended to $\mathbb{R}^d$ and the approximation can be realized without scaling of the function g. Moreover, we extend the class of functions useful as g in (1.1). Hereafter, we call the simultaneous approximation of functions and their derivatives *differentiable approximation*. An input-output function of neural units, such as g in (1.1), is called an *activation function* in neural computation theory. Accompanying the extension of approximation to derivatives, we have to remove the restriction that activation functions are sigmoidal because derivatives of a sigmoid function are not sigmoidal. We use systematically the inverse Radon transform as well as delta sequences. The former was used in [12] and the latter will be defined in Section 2 of this paper. We show that these tools can be used for approximations both on $\mathbb{R}^d$ and on compact sets, although there are simpler proofs in the case of differentiable approximation on compact sets as is described in [15]. The existence of the delta sequence can be proved by means of the Hahn-Banach theorem, but in special cases it can be obtained constructively. In such cases, our proofs of the theorems can be regarded as algorithms for implementing differentiable approximations by neural networks because other parts can be proved constructively.

2. Construction of delta sequences

The discriminatory function is first used in [2] in the field of neural network theory. Modifying the original one, we have defined other kinds of discriminatory functions [13,14]. In this article, we use only one-dimensional versions of discriminatory functions for guaranteeing existence of delta sequences. In order to make this article self-contained, we describe the definitions of one-dimensional versions of the respective discriminatory functions, although the definitions in the higher-dimensional spaces are described in [13,14]. Even if not mentioned explicitly, the functions treated in this paper are measurable and the measures are signed Borel measures. Since sigmoid functions are important in neural network theory, we shall remark the relation of sigmoid functions to the respective lemmas of this section.

Let $g(t)$ be a function defined on the line $\mathbb{R}$. We call $g(t-t_0)$ a shift of g, $g(\omega t)$ a rotation of g and $g(ct)$ a scaling of g, where $t_0 \in \mathbb{R}$ is a shift, $\omega \in \{-1, 1\} = S^0$ is a rotator and $c \geqslant 0$ is a scalar. We write $g_c(t) = g(ct)$ and call $g_c(\omega t - t_0)$ a *scaled shifted rotation* of g, even in the special case where $c = 1$, $\omega = 1$ or $t_0 = 0$. In the higher-dimensional spaces, the rotator will be a unit vector.

Definition 2.1. Let g be a function defined on $\mathbb{R}$.

(i) We call g *discriminatory* if, for any measure ν on $\mathbb{R}$ with compact support,

$$\int g_c(\omega t - t_0)\,\mathrm{d}\nu(t) = 0, \tag{2.1}$$

for all $c \geqslant 0$, $\omega \in S^0$ and $t_0 \in \mathbb{R}$, implies $\nu = 0$.

(ii) We call g *strongly discriminatory* if, for any measure ν on $\mathbb{R}$ with compact support,

$$\int g(\omega t - t_0)\,\mathrm{d}\nu(t) = 0, \tag{2.2}$$

for all $t_0 \in \mathbb{R}$ and $\omega \in S^0$, implies $\nu = 0$.

(iii) We call g *completely discriminatory* if, for any finite measure ν on $\mathbb{R}$, (2.1) for all $c \geqslant 0$, $\omega \in S^0$ and $t_0 \in \mathbb{R}$ implies $\nu = 0$.

(iv) We call g *strongly completely discriminatory* if, for any finite measure ν defined on $\mathbb{R}$, (2.2) for all $t_0 \in \mathbb{R}$ and $\omega \in S^0$ implies $\nu = 0$.

The Fourier transform of a function f on $\mathbb{R}$ is defined by

$$\mathcal{F}f(s) = (2\pi)^{-1/2}\int f(t)\mathrm{e}^{-\sqrt{-1}st}\,\mathrm{d}t.$$

The Fourier transform can be extended to measures and distributions. We call g a *slowly increasing function* if it is slowly increasing in the sense of distribution. We denote by $\operatorname{supp}(f)$ the support of a function f. Constants can be regarded as polynomials.

Lemma 2.2. (i) *Every slowly increasing function g on $\mathbb{R}$ is discriminatory, if it is not a polynomial.*

(ii) *Every slowly increasing function g on $\mathbb{R}$ is strongly discriminatory, if* $\operatorname{supp}(\mathcal{F}g)$ *has an accumulation point.*

(iii) *Every square integrable function g on $\mathbb{R}$ is completely discriminatory, if $g \neq 0$.*

(iv) *Every square integrable function g on $\mathbb{R}$ is strongly completely discriminatory, if and only if* $\operatorname{supp}(\mathcal{F}g)$ *is dense on $\mathbb{R}$.*

Proof. We write $g_{c,\omega}(t) = g_c(\omega t)$ and $g_\omega(t) = g(\omega t)$. Recall that $\operatorname{supp}(\mathcal{F}g)$ and $\operatorname{supp}(\mathcal{F}\nu)$ are symmetric.

(i) Eq. (2.1) implies that $g_{c,\omega} * \nu = 0$. The Fourier transforms of $g_{c,\omega}$, ν and $g_{c,\omega} * \nu$ are well-defined. Since $\mathcal{F}g_{c,\omega}$ is a slowly increasing distribution and $\mathcal{F}\nu$ is a bounded analytic function, a product $\mathcal{F}g_{c,\omega}\mathcal{F}\nu$ is well-defined and an equation $\mathcal{F}(g_{c,\omega} * \nu) = \mathcal{F}g_{c,\omega}\mathcal{F}\nu$ holds. Since g is not a polynomial, $\operatorname{supp}(\mathcal{F}g)$ contains a point other than the origin. Hence, $\mathcal{F}g_{c,\omega}\mathcal{F}\nu \neq 0$ for a certain c if $\nu \neq 0$.

(ii) Let $\{s_n\}_{n=1}^{\infty}$ be points of $\operatorname{supp}(\mathcal{F}g)$ which converge to an accumulation point s_0 of $\operatorname{supp}(\mathcal{F}g)$. Then, $\mathcal{F}g_\omega\mathcal{F}\nu = 0$ implies that $\mathcal{F}\nu(s_n) = 0$ for $n = 0, 1, \ldots$. Hence, by the uniqueness theorem, $\mathcal{F}\nu = 0$, which in turn implies that $\nu = 0$.

(iii) Note that $\mathcal{F}\nu$ is bounded continuous and $\operatorname{supp}(\mathcal{F}g)$ has a point other than the origin. Hence, if $\nu \neq 0$, $\mathcal{F}g_{c,\omega}\mathcal{F}\nu \neq 0$ for a certain c.

(iv) It is obvious that $\mathcal{F}g_\omega\mathcal{F}\nu \neq 0$ for $\nu \neq 0$ if $\operatorname{supp}(\mathcal{F}g)$ is dense in $\mathbb{R}$. Conversely, if $\operatorname{supp}(\mathcal{F}g)$ is not dense in $\mathbb{R}$, there is a finite measure $\nu \neq 0$ such that $\operatorname{supp}(\mathcal{F}\nu) \cap \operatorname{supp}(\mathcal{F}g) = \emptyset$. Hence, (2.2) holds for $\nu \neq 0$. □

If g is slowly increasing and ν is a measure with compact support, the products $\mathcal{F}g_{c,\omega}\mathcal{F}\nu$ and $\mathcal{F}g_\omega\mathcal{F}\nu$ are well-defined. However, even if ν is a finite measure, these products are not necessarily defined. For this technical reason, we suppose that g is square integrable in (iii) and (iv).

If g is a polynomial, $\operatorname{supp}(\mathcal{F}g)$ contains only the origin. Hence, $\mathcal{F}g_c\mathcal{F}\nu = 0$ if $\mathcal{F}\nu(0) = 0$. This implies that any polynomial is neither discriminatory nor strongly discriminatory. An example of a discriminatory function which is not strongly discriminatory is $\sin t$. Every square integral nonzero function is completely discriminatory, but not necessarily strongly discriminatory. An example of a completely discriminatory which is not strongly completely discriminatory is $t^{-1}\sin t$.

We have supposed that g is slowly increasing in (i) and (ii) of Lemma 2.2 and square integrable in (iii) and (iv). However, this does not imply that functions such as rapidly increasing functions are irrelevant to our theory because we apply the conditions of the lemma to linear combinations of a given function. Even a linear combination of an exponentially increasing function can be strongly completely discriminatory. Let $g(t) = \mathrm{e}^{|t|}$ and set

$$G(t) = g(t-1) + g(t+1) - (\mathrm{e} + \mathrm{e}^{-1})g(t). \tag{2.3}$$

Then, G is a continuous function with compact support. It can be shown by direct calculation that $\operatorname{supp}(\mathcal{F}G) = \mathbb{R}$. Hence, owing to Lemma 2.2, G is strongly completely discriminatory, which implies that it is discriminatory, strongly discriminatory and completely discriminatory.

The function $g(t) = \mathrm{e}^{|t|}$ is an example of a strongly discriminatory function which increases exponentially. In fact, let ν be a measure with compact support and suppose that $\int g(t-t_0)\,\mathrm{d}\nu(t) = 0$ for all t_0. Then $\int G(t-t_0)\,\mathrm{d}\nu(t) = 0$ for all t_0. This implies that $\nu = 0$ because G is strongly discriminatory. Hence, g is strongly discriminatory. In this way, we can prove that if a linear combination of shifts of a function is slowly increasing and strongly discriminatory, the original function is strongly discriminatory. In the same way, the respective statements of Lemma 2.2 can be extended. Extension of Lemma 2.2 using a linear combination is important because sigmoid functions are often used as activation functions in neural computation theory. If $\lim_{t\to-\infty} h(t) = 0$ and $\lim_{t\to\infty} h(t) = 1$, and h is monotone increasing, then h is called a *sigmoid function*. Since the difference of two distinct shifts of any sigmoid function is nonpolynomial and square integrable, all sigmoid functions are discriminatory, strongly discriminatory and completely discriminatory. However, all sigmoid functions are not strongly completely discriminatory because the support of the Fourier transform of a sigmoid function is not necessarily dense in $\mathbb{R}$.

For a function g defined on $\mathbb{R}$, we set

$$\Sigma(g, c) = \left\{ \sum_{i=1}^{n} a_i g_{c_i}(\omega_i t - t_i); n \geqslant 1, a_i, t_i \in \mathbb{R}, c_i \geqslant 0, \omega_i \in S^0 \right\} \tag{2.4}$$

and

$$\Sigma(g) = \left\{ \sum_{i=1}^{n} a_i g(\omega_i t - t_i); n \geqslant 1, a_i, t_i \in \mathbb{R}, \omega_i \in S^0 \right\}. \tag{2.5}$$

Let us denote by $C(\mathbb{R})$ the space of continuous functions defined on $\mathbb{R}$, by $C(\boldsymbol{K}_1)$ the space of continuous functions defined on a compact set $\boldsymbol{K}_1 \subset \mathbb{R}$, by $C_c(\mathbb{R})$ the space of continuous functions defined on $\mathbb{R}$ with compact support and by $C_0(\mathbb{R})$ the completion of $C_c(\mathbb{R})$ by the uniform topology.

Lemma 2.3. *Let g be a function defined on $\mathbb{R}$, and $\boldsymbol{K}_1$ be a compact set of $\mathbb{R}$.*

(i) *If g is a slowly increasing nonpolynomial function of locally bounded variation, then any $f \in C(\boldsymbol{K}_1)$ can be uniformly approximated by a member of $\Sigma(g, c)$ on $\boldsymbol{K}_1$.*

(ii) *If g is a slowly increasing function of locally bounded variation and* $\operatorname{supp}(\mathcal{F}g)$ *has an accumulation point on $\mathbb{R}$, then any $f \in C(\boldsymbol{K}_1)$ can be uniformly approximated by a member of $\Sigma(g)$ on $\boldsymbol{K}_1$.*

(iii) *If g is a square integrable nonzero function of uniformly locally bounded variation, then any $f \in C_0(\mathbb{R})$ can be uniformly approximated by a member of $\Sigma(g, c)$ on $\mathbb{R}$.*

(iv) *If g is a square integrable nonzero function of uniformly locally bounded variation and* $\operatorname{supp}(\mathcal{F}g)$ *is dense in $\mathbb{R}$, then any $f \in C_0(\mathbb{R})$ can be uniformly approximated by a member of $\Sigma(g)$ on $\mathbb{R}$. Conversely, if* $\operatorname{supp}(\mathcal{F}g)$ *is not dense in $\mathbb{R}$, there is a function of $C_0(\mathbb{R})$ which cannot be approximated by a member of $\Sigma(g)$.*

Proof. We describe only the proof of the first statement fully and abbreviate others because higher-dimensional versions of these statements are proved in [13,14]. Let $\rho \in C_c(\mathbb{R})$ be a nonzero nonnegative function.

(i) We first prove that $\Sigma(g * \rho, c)|_{K_1}$ is dense in $C(\boldsymbol{K}_1)$, where $|_{K_1}$ stands for the restriction of each function of the set to $\boldsymbol{K}_1$. Suppose that there is a function $\varphi_1 \in C(\boldsymbol{K}_1)$ which does not belong to the closure of $\Sigma(g * \rho, c)|_{K_1}$ in the uniform topology. By the Hahn–Banach theorem [17, Theorem 3.5], there is a bounded functional Λ such that $\Lambda(\varphi_0) = 0$ for any $\varphi_0 \in \Sigma(g * \rho, c)$ and $\Lambda(\varphi_1) = 1$. By the Riesz representation theorem [16, Theorem 2.14], there is a measure ν with compact support for which $\Lambda(\varphi) = \int \varphi(t)\,\mathrm{d}\nu(t)$ for all $\varphi \in C(\boldsymbol{K}_1)$. Since $(g * \rho)_{c,\omega}(t - t_0)$ belongs to $\Sigma(g * \rho, c)$, we have that $\int (g * \rho)_{c,\omega}(t - t_0)\,\mathrm{d}\nu(t) = 0$ for all c, ω and t_0, which in turn implies that $\nu = 0$ as $g * \rho$ is discriminatory. This is a contradiction. Hence, f can be approximated by a member of $\Sigma(g * \rho, c)$. Since g is a function of locally bounded variation, $g * \rho$ can be uniformly approximated on $\boldsymbol{K}_1$ by a linear combination of shifts of g. Hence, (i) holds.

(ii) The proof of this statement is similar to that of (i). If an appropriate ρ is chosen, $\operatorname{supp}(\mathcal{F}(g * \rho))$ has an accumulation point. Hence, $g * \rho$ is strongly discriminatory. Follow the proof of (i), replacing $\Sigma(g * \rho, c)$ and $\Sigma(g, c)$ by $\Sigma(g * \rho)$ and $\Sigma(g)$ respectively and using the fact that g and $g * \rho$ are strongly discriminatory. Then, we can confirm that (ii) holds.

(iii) Similarly to the case (i), we can prove that $\Sigma(g * \rho, c)$ is dense in $C_0(\mathbb{R})$, applying [16, Theorem 6.19] instead of [16, Theorem 2.14]. Since g is a function of uniformly locally bounded variation, $g * \rho$ can be uniformly approximated on $\mathbb{R}$ by a linear combination of g.

(iv) The "if" part of the proof is similar to that of (iii). Follow the proof of (iii), replacing $\Sigma(g * \rho)$ by $\Sigma(g)$. Then, it turns out to be the proof of the "if" part. Conversely, if $\operatorname{supp}(\mathcal{F}g)$ is not dense on $\mathbb{R}$, there is a rapidly decreasing nonzero function f such that $\operatorname{supp}(\mathcal{F}f) \cap \operatorname{supp}(\mathcal{F}g) = \emptyset$. Suppose that it can be uniformly approximated by $\bar{f} \in \Sigma(g)$. Then, we have

$$\int f(t)\big(f(t) - \bar{f}(t)\big)\,\mathrm{d}t = \int |\mathcal{F}f(t)|^2\,\mathrm{d}t.$$

The right-hand side of this equation is a positive constant, but the left-hand side can be arbitrarily small. This is a contradiction. □

If g is a sigmoid function, this lemma can be simplified.

Corollary 2.4. *Let h be a sigmoid function and $\boldsymbol{K}_1$ be a compact set of $\mathbb{R}$. Then,*

(i) *any $f \in C(\boldsymbol{K}_1)$ can be uniformly approximated by a member of $\Sigma(h, c)$ on $\boldsymbol{K}_1$;*

(ii) *any $f \in C(\boldsymbol{K}_1)$ can be uniformly approximated by a member of $\Sigma(h)$ on $\boldsymbol{K}_1$;*

(iii) *any $f \in C_0(\mathbb{R})$ can be uniformly approximated by a member of $\Sigma(h, c)$ on $\mathbb{R}$;*

(iv) *if $\operatorname{supp}(\mathcal{F}g)$ is dense in $\mathbb{R}$, then any $f \in C_0(\mathbb{R})$ can be uniformly approximated by a member of $\Sigma(h)$ on $\mathbb{R}$. Conversely, if $\operatorname{supp}(\mathcal{F}g)$ is not dense in $\mathbb{R}$, there is a function of $C_0(\mathbb{R})$ which cannot be approximated by a member of $\Sigma(h)$.*

Proof. The sigmoid function h satisfies the conditions on g of both (i) and (ii) in Lemma 2.3. Hence, the statements (i) and (ii) hold. Let g be a difference of two distinct shifts of h. Then, g satisfies the condition of the statement (iii). Hence, $f \in C_0(\mathbb{R})$ can be uniformly approximated by a member of $\Sigma(g, c)$, which implies that a member of $\Sigma(h, c)$ approximates the f. If $\operatorname{supp}(\mathcal{F}h) = \mathbb{R}$, the difference g satisfies the condition of the statement of (iv). Then, $f \in C_0(\mathbb{R})$ can be uniformly approximated by a member of $\Sigma(g)$ as well as by a member of $\Sigma(h)$. The rest of the statement (iv) can be proved as in Lemma 2.3. □

Now we define delta sequences.

Definition 2.5. (i) If a sequence $\{u_k\}_{k=1}^{\infty}$ of functions on $\mathbb{R}$ satisfies the conditions (a) and (b) below, we call it a *delta sequence on finite intervals*:

(a) $\int_{-1/k}^{1/k} u_k(t)\,dt$ converges to 1 as $k \to \infty$,

(b) $|u_k(t)| < \frac{1}{k}$, for $t \in \left[-k, -\frac{1}{k}\right] \cup \left[\frac{1}{k}, k\right]$,

for sufficiently large k.

(ii) If a sequence $\{u_k\}_{k=1}^{\infty}$ of functions on $\mathbb{R}$ satisfies the conditions (a) above and (c) below, we call it a *delta sequence on the line*:

(c) $|u_k(t)| < \frac{1}{k}$, for $t \in \left(-\infty, -\frac{1}{k}\right] \cup \left[\frac{1}{k}, \infty\right)$.

Of course, a delta sequence on the line is that on finite intervals. Elements of a delta sequence on finite intervals can be polynomials. The function below satisfies the conditions (a)–(c):

$$u_k(t) = (2\pi)^{-1/2} k^2 \exp(-\tfrac{1}{2}k^4 t^2). \tag{2.6}$$

Lemma 2.6. *Let g be a function defined on $\mathbb{R}$.*

(i) *If g is slowly increasing, nonpolynomial and of locally bounded variation, then $\Sigma(g, c)$ contains a delta sequence on finite intervals.*

(ii) *If g is slowly increasing and of locally bounded variation and* supp$(\mathcal{F}g)$ *has an accumulation point, then $\Sigma(g)$ contains a delta sequence on finite intervals.*

(iii) *If g is nonzero, square integrable and of uniformly locally bounded variation, then $\Sigma(g, c)$ contains a delta sequence on the line.*

(iv) *If g is nonzero, square integrable and of uniformly locally bounded variation and* supp$(\mathcal{F}g)$ *is dense on $\mathbb{R}$, then $\Sigma(g)$ contains a delta sequence on the line.*

Proof. Lemma 2.3 is useful for proving this lemma. The elements u_k of the respective delta sequences can be obtained by uniformly approximating functions such as defined by (2.6). □

Lemmas 2.3 and 2.6 guarantee that not only $\Sigma(g, c)$ but also $\Sigma(g)$ contains delta sequences both on finite intervals and on the line if the function g is one of the familiar functions such as $e^{-t^2/2}$, $(1+t^2)^{-1}$ and $e^{-|t|}$. Suppose that h is a sigmoid function. Then, by Corollary 2.4 and Lemma 2.6, $\Sigma(h, c)$ and $\Sigma(h)$ contain delta sequences both on finite intervals and on the line, if h is one of the familiar sigmoid functions such as $(1+e^{-t})^{-1}$, $(2\pi)^{-1}\int_{-\infty}^{t} e^{-s^2/2}\,ds$ and $\min(\max(0, t), 1)$.

Even if g is an exponentially increasing function, $\Sigma(g, c)$ and $\Sigma(g)$ can contain both kinds of delta sequences. It is now obvious that $\Sigma(G, c)$ and $\Sigma(G)$ contain both kinds of delta sequences if G is the function defined by (2.3) for $g(t) = e^{|t|}$. Hence, $\Sigma(e^{|t|}, c)$ as well as $\Sigma(e^{|t|})$ contain both kinds of delta sequences.

It is usually difficult to construct a delta sequence without scaling the function g. However, the author [12,13] has shown several examples of constructive methods of uniformly approximating continuous functions by linear combinations of an unscaled sigmoid function. They can be used for constructing delta sequences without scaling the sigmoid function. We present here a few examples of delta sequences, the elements of which are linear combinations of unscaled nonlinear functions.

Example 2.7. One of the sigmoid functions mentioned above is a ramp function, which is defined by

$$g(t)=\begin{cases}0, & t<0,\\ t, & 0\leqslant t\leqslant 1,\\ 1, & t>1.\end{cases}$$

If g can be scaled, it is easy to construct a delta sequence from g. Even if it cannot be scaled, we can construct delta sequences on finite intervals and on the line whose elements are linear combinations of unscaled shifts of g. Set

$$u_k(t)=\sum_{m=0}^{k}k^2\left(g\left(t+\frac{1}{k}-m\right)-2g(t-m)+g\left(t-\frac{1}{k}-m\right)\right).$$

Then, $\{u_k\}$ is a delta sequence on finite intervals. Set

$$u_k(t)=\sum_{m=0}^{2k^2}k^2\left(1-\frac{m}{2k^2}\right)\left(g\left(t+\frac{1}{k}-m\right)-2g(t-m)+g\left(t-\frac{1}{k}-m\right)\right).$$

Then, the sequence is a delta sequence on the line. Polygonal functions can also be used for constructing delta sequences. For details, see [13].

Using the following proposition, we can construct a delta sequence on the line.

Proposition 2.8. *Let u and g be square integrable continuous functions. Suppose that g is a bounded function of uniformly locally bounded variation. If the quotient $\mathcal{F}u/\mathcal{F}g$ is slowly increasing and its inverse Fourier transform is a square integrable continuous function, then a linear combination of shifts of g can approximate u uniformly on $\mathbb{R}$.*

Proof. Set $v=\mathcal{F}^{-1}(\mathcal{F}u/\mathcal{F}g)$. Then,

$$u(t)=v*g(t)=\int v(s)g(t-s)\,\mathrm{d}s. \tag{2.7}$$

For any $\varepsilon>0$, there is a continuous function w with compact support such that

$$\left|\int(v(s)-w(s))g(t-s)\,\mathrm{d}s\right|<\varepsilon,\quad \text{on } \mathbb{R}.$$

Since g is a function of uniformly locally bounded variation, the integral $\int w(s)g(t-s)\,\mathrm{d}s$ can be approximated uniformly on $\mathbb{R}$ by a finite sum:

$$\sum_{i=1}^{n}w(s_i)(s_i-s_{i-1})g(t-s_i). \tag{2.8}$$

This concludes the proof. □

Remark 2.9. Note that $\operatorname{supp}(\mathcal{F}g)$ must be dense in $\mathbb{R}$ in this proposition. This implies that g must be strongly completely discriminatory. This proposition suggests the importance of strongly discriminatory functions.

Example 2.10. Let $u_c(t) = (2\pi c^2)^{-1/2}e^{-t^2/2c^2}$ and $g(t) = \partial_t(1+e^{-t})^{-1} = e^{-t}(1+e^{-t})^{-2}$. Then $\mathcal{F}u_c(t) = (2\pi)^{-1/2}\exp(-\frac{1}{2}c^2t^2)$ and $\mathcal{F}g(t) = (2\pi)^{1/2}t(e^{\pi t} - e^{-\pi t})^{-1}$. Hence, $\mathcal{F}u_c/\mathcal{F}g$ is a rapidly decreasing C^∞-function, which implies that $v_c = \mathcal{F}^{-1}(\mathcal{F}u_c/\mathcal{F}g)$ is also rapidly decreasing. More precisely, since

$$\mathcal{F}v_c(t) = (2\pi t)^{-1}\exp(-\tfrac{1}{2}c^2t^2)(e^{\pi t} - e^{-\pi t}),$$

$$\frac{d}{dt}v_c(t) = -c^{-2}\sin(c^{-2}\pi t)\exp\left(\tfrac{1}{2}c^{-2}(\pi^2 - t^2)\right).$$

Hence,

$$v_c(t) = -c^{-2}\int_{-\infty}^{t}\sin(c^{-2}\pi s)\exp\left(\tfrac{1}{2}c^{-2}(\pi^2 - s^2)\right)ds.$$

Hence, u_c can be uniformly approximated on $\mathbb{R}$ by a linear combination of unscaled shifts of g. For an appropriate sequence $\{c_k\}$, $\{u_{c_k}\}$ is a delta sequence on the line. Hence, by approximating g by a difference of slightly shifted versions of the unscaled logistic function, we can construct a delta sequence on the line whose members are linear combinations of shifts of the logistic function. Thus, Proposition 2.8 and Example 2.10 suggest usefulness of the unscaled logistic function. See [3,14] for calculation.

3. Approximation of the inverse Radon transform

We simplify the notations of derivatives: $\partial_i = \partial/\partial x^{(i)}$. For the multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$, we write $|\alpha| = \alpha_1 + \cdots + \alpha_d$, $\alpha! = \alpha_1!\cdots\alpha_d!$, $\partial^\alpha = \partial_1^{\alpha_1}\ldots\partial_d^{\alpha_d}$ and $\alpha \geqslant \beta$ if $\alpha_1 \geqslant \beta_1, \ldots, \alpha_d \geqslant \beta_d$. We use the notation $\partial_t = \partial/\partial_t$ for a function in one variable t. Hence, $\partial_t^r g = g^{(r)}$. The definition of the space $C^m(\Omega)$ of m-times continuously differential functions is familiar, where $0 \leqslant m \leqslant \infty$ and Ω is the whole space $\mathbb{R}^d$ or a compact subset $\boldsymbol{K} \subset \mathbb{R}^d$. For simplicity, we suppose that a function of $C^m(\boldsymbol{K})$ is defined in an open set including $\boldsymbol{K}$. The space $C^m(\Omega)$ is endowed with a simultaneous uniform norm:

$$\|f\|_{m,\Omega} = \sup\{|\partial^\alpha f(x)|; x \in \Omega, |\alpha| \leqslant m\}. \tag{3.1}$$

We denote by $C_c^m(\mathbb{R}^d)$ the restriction of $C^m(\mathbb{R}^d)$ to functions with compact support and by $C_0^m(\mathbb{R}^d)$ its completion. The space of rapidly decreasing C^∞-functions defined on $\mathbb{R}^d$ is denoted by $\mathcal{S}(\mathbb{R}^d)$.

We denote by $W^m(\Omega)$ the space of functions defined on Ω which are m-times differentiable in the sense of distribution if the mth derivatives are functions. Let f be a function of $W^m(\Omega)$. We say that a function $f_1 \in W^m(\Omega)$ approximates f in $W^m(\Omega)$, if $\|f - f_1\|_{m,\Omega}$ is sufficiently small. The Heaviside function belongs to $W(\mathbb{R}) = W^0(\mathbb{R})$ and a polygonal sigmoid function belongs to $W^1(\mathbb{R})$. Any sigmoid function, not necessarily continuous, belongs to $W(\mathbb{R})$.

A unit vector ω is an element of the unit sphere S^{d-1}. Let $\omega_{\cdot 1}, \ldots, \omega_{\cdot d}$ be the components of ω and set $\omega^\alpha = \omega_{\cdot 1}^{\alpha_1}\cdots\omega_{\cdot d}^{\alpha_d}$. Denote by γ a directed grand arc on S^{d-1} and by ∂_γ the differentiation along γ. We denote by $C_0^m(\mathbb{R}\times S^{d-1})$ the space of C^m-functions defined on $\mathbb{R}\times S^{d-1}$ whose derivatives up to order m, $m \leqslant \infty$, converge to 0 as $|t| \to \infty$, and by $\mathcal{S}(\mathbb{R}\times S^{d-1})$ the space of rapidly decreasing

C^∞-functions defined on $\mathbb{R}\times S^{d-1}$. Let $m_{t,\omega}$ be a uniform measure with unit density on a hyperplane $\{x\in\mathbb{R}^d|\omega\cdot x=t\}$. The Radon transform of $f\in\mathcal{S}(\mathbb{R}^d)$ is defined by

$$\check{f}(t,\omega)=\int f(x)\,\mathrm{d}m_{t,\omega}(x). \tag{3.2}$$

Further, let μ be the uniform measure on S^{d-1} with unit density. Set $\Box_t=\partial_t^2$. For convenience, we define an operator L with respect to the variable t by

$$L\varphi(t,\omega)=\frac{1}{2^d\pi^{d-1}}(-\Box_t)^{(d-1)/2}\varphi(t,\omega). \tag{3.3}$$

Then we have that

$$f(x)=\int L\check{f}(\omega\cdot x,\omega)\,\mathrm{d}\mu(\omega). \tag{3.4}$$

For the details of these definitions, see [5,7,12].

The lemma below holds for the integrand of (3.4).

Lemma 3.1. *Suppose that* $f\in\mathcal{S}(\mathbb{R}^d)$. *Then,*

(a) $L\check{f}\in C_0^\infty(\mathbb{R}\times S^{d-1})$;

(b) $L\check{f}(\omega\cdot x,\omega)=L\check{f}(-\omega\cdot x,-\omega)$;

(c) *for any* $k\geqslant 0$ *and any directed grand arcs* $\gamma_1,\ldots,\gamma_q$ *there is a positive constant* $M_{k,\gamma_1,\ldots,\gamma_q}$ *such that*

$$|\partial_t^k\partial_{\gamma_1}\cdots\partial_{\gamma_q}L\check{f}(t,\omega)|<M_{k,\gamma_1,\ldots,\gamma_q}\big(|t|^{d+k}+1\big)^{-1},\quad \textit{for all } (t,\omega);$$

(d) *in particular,* $L\check{f}\in\mathcal{S}(\mathbb{R}\times S^{d-1})$ *for odd* d.

By (c) and (d) of the lemma, $L\check{f}$ is integrable over t for all d.

Lemma 3.2. *For* $f\in\mathcal{S}(\mathbb{R}^d)$, *we have*

$$L\big((\partial^\alpha f)^\vee\big)(t,\omega)=\omega^\alpha\partial_t^{|\alpha|}L\check{f}(t,\omega),\quad \textit{on } \mathbb{R}\times S^{d-1}. \tag{3.5}$$

Proof. We have, for example, that, for $e_1=(1,0,\ldots,0)$,

$$\begin{aligned}\int\partial_1 f(x)\,\mathrm{d}m_{t,\omega}&=\lim_{h\to0}\frac{1}{h}\int\big(f(x+he_1)-f(x)\big)\,\mathrm{d}m_{t,\omega}\\&=\lim_{h\to0}\frac{1}{h}\Big(\int f(x)\,\mathrm{d}m_{t+h\omega_{\cdot1},\omega}-\int f(x)\,\mathrm{d}m_{t,\omega}\Big)\\&=\lim_{h\to0}\frac{1}{h}\big(\check{f}(t+h\omega_{\cdot1},\omega)-\check{f}(t,\omega)\big)=\omega_{\cdot1}\cdot\partial_t\check{f}(t,\omega).\end{aligned}$$

Repeating this procedure, we obtain

$$\int\partial^\alpha f(x)\,\mathrm{d}m_{t,\omega}=\omega^\alpha\partial_t^{|\alpha|}\check{f}(t,\omega).$$

In [12], it is proved that the order of the operators ∂_t and L can be interchanged. Hence,

$$L \int \partial^\alpha f(x) \, \mathrm{d}m_{t,\omega} = \omega^\alpha \partial_t^{|\alpha|} L\check{f}(t, \omega).$$

This concludes the proof. □

Lemma 3.3. (i) *Let $\{u_k\}_{k=1}^\infty$ be a delta sequence of W^m-functions on finite intervals. Then, we have*

$$\int \varphi(t-s) u_k^{(r)}(s) \, \mathrm{d}s = \int \varphi^{(r)}(t-s) u_k(s) \, \mathrm{d}s, \quad 0 \leqslant r \leqslant m, \tag{3.6}$$

for any $\varphi \in C_c^m(\mathbb{R})$ for sufficiently large k. This quantity converges uniformly to $\varphi^{(r)}(t)$ on any compact set of $\mathbb{R}$ for all $r \leqslant m$.

(ii) *Let $\{u_k\}_{k=1}^\infty$ be a delta sequence of W^m-functions on the line. If $u_k^{(r)}$, $r \leqslant m$, are bounded respectively, we then have* (3.6) *for any function $\varphi \in C_0^m(\mathbb{R})$ for sufficiently large k. In this case, the quantity* (3.6) *converges uniformly to $\varphi^{(r)}(t)$ on $\mathbb{R}$ for all $r \leqslant m$.*

Proof. Eq. (3.6) is obvious by integration by parts. We have

$$\int \varphi^{(r)}(t-s) u_k(s) \, \mathrm{d}s = \left(\int_{|s|<1/k} + \int_{|s|\geqslant 1/k} \right) \varphi^{(r)}(t-s) u_k(s) \, \mathrm{d}s.$$

In both cases, the first integral on the right-hand side converges to $\varphi^{(r)}$ uniformly on $\mathbb{R}$. In the case (i), the second integral is bounded by $k^{-1} \int |\varphi^{(r)}(t)| \, \mathrm{d}t$ on any compact set for sufficiently large k. In the case (ii), the second integral is bounded by $k^{-1} \int |\varphi^{(r)}(t)| \, \mathrm{d}t$ on $\mathbb{R}$. Hence, the lemma follows. □

In the lemma below, we use a sequence $\{\phi_N\}_{N=1}^\infty$ of infinitely differentiable symmetric functions defined on $\mathbb{R}$ by

$$\phi_N(t) = \begin{cases} 1, & \text{if } |t| \leqslant N, \\ \phi(|t| - N), & \text{if } |t| > N, \end{cases} \tag{3.7}$$

where ϕ is a symmetric C^∞-function such that $\phi(0) = 1$, $\phi(t) = 0$ for $t > 1$, $\phi^{(n)}(0) = 0$ for $n \geqslant 1$ and $0 \leqslant \phi(t) \leqslant 1$ on $\mathbb{R}$.

Proposition 3.4. (i) *Let $\{\phi_N\}_{N=1}^\infty$ be a sequence of functions defined by* (3.7) *and $\{u_k\}_{k=1}^\infty$ be a delta sequence of W^m-functions on finite intervals. For $f \in \mathcal{S}(\mathbb{R}^d)$, set*

$$f_{N,k}(x) = \int \int \phi_N(t) L\check{f}(t, \omega) u_k(\omega \cdot x - t) \, \mathrm{d}t \, \mathrm{d}\mu(\omega). \tag{3.8}$$

Then, for any $\varepsilon > 0$ and any compact set $\boldsymbol{K} \subset \mathbb{R}^d$, we have that if N and k are sufficiently large,

$$|\partial^\alpha f(x) - \partial^\alpha f_{N,k}(x)| < \varepsilon, \quad \text{on } \boldsymbol{K}, \tag{3.9}$$

for all α, $|\alpha| \leqslant m$.

(ii) *Let* $\{u_k\}$ *be a delta sequence of* W^m*-functions on the line and suppose that* $\partial_t^r u_k$ *are bounded for all* k *and* $r \leqslant m$. *For* $f \in \mathcal{S}(\mathbb{R}^d)$, *set*

$$f_k(x) = \int\int L\check{f}(t, \omega) u_k(x \cdot \omega - t)\, \mathrm{d}t\, \mathrm{d}\mu(\omega). \tag{3.10}$$

Then, for any $\varepsilon > 0$, *we have that if* k *is sufficiently large,*

$$|\partial^\alpha f(x) - \partial^\alpha f_k(x)| < \varepsilon, \quad \text{on } \mathbb{R}^d, \tag{3.11}$$

for all α, $|\alpha| \leqslant m$.

Proof. (i) Set

$$f_N(x) = \int \phi_N(\omega \cdot x) L\check{f}(\omega \cdot x, \omega)\, \mathrm{d}\mu(\omega). \tag{3.12}$$

For sufficiently large k, we have that, by Lemmas 3.2 and 3.3,

$$\begin{aligned}
&\partial^\alpha f_{N,k}(x) \\
&= \int\int \phi_N(t) L\check{f}(t, \omega) \partial^\alpha u_k(\omega \cdot x - t)\, \mathrm{d}t\, \mathrm{d}\mu(\omega) \\
&= (-1)^{|\alpha|} \int\int \phi_N(t) L\check{f}(t, \omega) \omega^\alpha \partial_t^{|\alpha|} u_k(\omega \cdot x - t)\, \mathrm{d}t\, \mathrm{d}\mu(\omega) \\
&= \int\int \sum_{i=0}^{|\alpha|} \frac{|\alpha|!}{i!\,(|\alpha| - i)!} (\partial_t^i \phi_N)(t) (\partial_t^{|\alpha|-i} L\check{f})(t, \omega) \omega^\alpha u_k(\omega \cdot x - t)\, \mathrm{d}t\, \mathrm{d}\mu(\omega) \\
&= \int\int \sum_{\beta \leqslant \alpha} \frac{\alpha!}{\beta!\,(\alpha - \beta)!} (\omega^\beta \partial_t^{|\beta|} \phi_N)(t) (\omega^{\alpha-\beta} \partial_t^{|\alpha|-|\beta|} L\check{f}(t, \omega)) u_k(\omega \cdot x - t)\, \mathrm{d}t\, \mathrm{d}\mu(\omega) \\
&= \sum_{\beta \leqslant \alpha} \frac{\alpha!}{\beta!\,(\alpha - \beta)!} \int\int \omega^\beta (\partial_t^{|\beta|} \phi_N)(t) L((\partial^{\alpha-\beta} f)^\vee)(t, \omega) u_k(\omega \cdot x - t)\, \mathrm{d}t\, \mathrm{d}\mu(\omega).
\end{aligned} \tag{3.13}$$

By Lemma 3.3, the right-hand side of (3.13) converges uniformly to

$$\sum_{\beta \leqslant \alpha} \frac{\alpha!}{\beta!\,(\alpha - \beta)!} \int \omega^\beta (\partial_t^{|\beta|} \phi_N)(\omega \cdot x) L((\partial^{\alpha-\beta} f)^\vee)(\omega \cdot x, \omega)\, \mathrm{d}\mu(\omega), \tag{3.14}$$

as $k \to \infty$ on any compact set. This limit coincides with $\partial^\alpha f_N(x)$. Hence, we have

$$|\partial^\alpha f_N(x) - \partial^\alpha f_{N,k}(x)| < \tfrac{1}{2}\varepsilon, \quad \text{if } |x| \leqslant N. \tag{3.15}$$

Each summand with $\beta \neq 0$ on the right-hand side of (3.14) converges uniformly to 0 on $\boldsymbol{K}$ as $N \to \infty$, because $\partial_t^r \phi_N(t) = 0$ on $\{|t| < N\} \cup \{|t| > N + 1\}$. Hence, we have

$$\lim_{N \to \infty} \partial^\alpha f_N(x) = \int L((\partial^\alpha f)^\vee)(\omega \cdot x, \omega)\, \mathrm{d}\mu(\omega), \quad \text{on } \boldsymbol{K}.$$

By (3.4), the right-hand side of this equality is equal to $\partial^\alpha f(x)$. Moreover, this convergence is uniform on $\boldsymbol{K}$. Hence,

$$|\partial^\alpha f(x) - \partial^\alpha f_N(x)| < \tfrac{1}{2}\varepsilon, \quad \text{on } \boldsymbol{K}, \tag{3.16}$$

for sufficiently large N. From (3.15) and (3.16), inequality (3.9) follows.

(ii) By Lemmas 3.2 and 3.3, we have that if $|\alpha| \leqslant m$,

$$\begin{aligned}\partial^\alpha f_k(x) &= \int\int L\check{f}(t,\omega)\partial^\alpha u_k(x\cdot\omega - t)\,\mathrm{d}t\,\mathrm{d}\mu(\omega)\\ &= (-1)^{|\alpha|}\int\int L\check{f}(t,\omega)\omega^\alpha\partial_t^{|\alpha|}u_k(x\cdot\omega - t)\,\mathrm{d}t\,\mathrm{d}\mu(\omega)\\ &= \int\int \omega^\alpha(\partial_t^{|\alpha|}L\check{f})(t,\omega)u_k(x\cdot\omega - t)\,\mathrm{d}t\,\mathrm{d}\mu(\omega)\\ &= \int\int L((\partial^\alpha f)^\vee)(t,\omega)u_k(x\cdot\omega - t)\,\mathrm{d}t\,\mathrm{d}\mu(\omega).\end{aligned} \tag{3.17}$$

Hence, by Lemma 3.3 and (3.4), $\partial^\alpha f_k$ converges to $\partial^\alpha f$ uniformly on $\mathbb{R}^d$. This concludes the proof. □

Thus, the function f and its derivatives can be approximated by the integral (3.8) on compact sets and by the integral (3.10) on the whole space.

4. Approximation theorems

In this section, we prove several approximation theorems. They are obtained by approximating the integrals (3.8) and (3.10) by finite sums of an element u_k of a delta sequence. Since u_k is obtained as a finite sum of scaled or unscaled shifted rotations of the component function, say g, the integrals are approximated by finite sums of g in conclusion. Recall that constants are regarded as polynomials and that a rotation of a function g in one variable is actually $g(t)$ or $g(-t)$. We use partitions of $[-N-1, N+1] \times S^{d-1}$ and $\mathbb{R} \times S^{d-1}$.

Theorem 4.1. *Let g be a function of $W^m(\mathbb{R})$ and $\boldsymbol{K}$ be a compact set of $\mathbb{R}^d$. Suppose that there is a linear combination G of scaled shifted rotations of g such that the derivative $\partial_t^m G$ is a slowly increasing nonpolynomial function of locally bounded variation. Then, for any $f \in C^m(\boldsymbol{K})$ and any $\varepsilon > 0$, there are a positive integer n, coefficients a_i, scalars c_i, rotators ω_i and shifts t_i, $i = 1, \ldots, n$, for which*

$$\bar{f}(x) = \sum_{i=1}^{n} a_i g_{c_i}(\omega_i\cdot x - t_i) \tag{4.1}$$

satisfies

$$|\partial^\alpha f(x) - \partial^\alpha \bar{f}(x)| < \varepsilon, \quad \textit{on } \boldsymbol{K}, \tag{4.2}$$

for all α, $|\alpha| \leqslant m$.

Proof. Since there is a function of $\mathcal{S}(\mathbb{R}^d)$ whose restriction to $\boldsymbol{K}$ approximates f in $C^m(\boldsymbol{K})$, we may suppose that $f \in \mathcal{S}(\mathbb{R}^d)$. By Lemma 2.6(i) there is a delta sequence $\{u_k\}$ on finite intervals whose members are linear combinations of scaled shifted rotations of G. Hence, f can be approximated in $W^m(\boldsymbol{K})$ by an integral of the form (3.8). Since $\partial_t^m G$ is a function of locally bounded variation, so are $\partial_t^r G$ and $\partial_t^r u_k$ for $r \leqslant m$. As the domain of f is restricted to $\boldsymbol{K}$, it is not difficult to prove that there is a finite partition $\{\Delta_i\}_{i=1}^n$ of $[-N-1, N+1] \times S^{d-1}$ for which

$$\bar{f}_{N,k}(x) = \sum_i \int \int_{\Delta_i} \phi_N(t) L\check{f}(t, \omega) \, \mathrm{d}t \, \mathrm{d}\mu(\omega) \, u_k(\omega_i \cdot x - t_i)$$

satisfies

$$|\partial^\alpha f_{N,k}(x) - \partial^\alpha \bar{f}_{N,k,\alpha}(x)| < \varepsilon, \quad \text{on } \boldsymbol{K},$$

for all α, $|\alpha| \leqslant m$, where (t_i, ω_i) is an arbitrary point of Δ_i. Since G is a linear combination of scaled shifted rotations of g, the proof is concluded. □

This theorem is an extension of the original neural approximation theorem proved in [4,9] and elsewhere. Under the restriction that $m = 0$ and g is a sigmoid function, this theorem is reduced to the original one.

Theorem 4.2. *Let g be a function of $W^m(\mathbb{R})$ and $\boldsymbol{K}$ be a compact set of $\mathbb{R}^d$. Suppose that there is a linear combination G of shifted rotations of g such that the derivative $\partial_t^m G$ is a slowly increasing nonpolynomial function of locally bounded variation. Then, for any $f \in C^m(\boldsymbol{K})$ and any $\varepsilon > 0$, there are a positive integer n, coefficients a_i, rotators ω_i and shifts t_i, $i = 1, \ldots, n$, for which*

$$\bar{f}(x) = \sum_{i=1}^n a_i g(\omega_i \cdot x - t_i) \tag{4.3}$$

satisfies (4.2) for all α, $|\alpha| \leqslant m$.

Proof. This theorem can be proved similarly to Theorem 4.1. By Lemma 2.6(ii), each member of the delta sequence on finite intervals can be approximated by a finite sum of unscaled shifted rotations of G. Hence, we obtain the theorem as in the proof of the Theorem 4.1. □

This theorem corresponds to [13, Theorem 2.6]. Now we treat simultaneous approximation of a function and its derivatives on $\mathbb{R}^d$.

Theorem 4.3. *Let $g \in W^m(\mathbb{R})$. Suppose that a certain linear combination G of scaled shifted rotations of g is a nonzero function such that $\partial_t^r G$, $0 \leqslant r \leqslant m$, are square integrable functions of uniformly locally bounded variation. Then, for any $f \in C_0^m(\mathbb{R}^d)$ and any $\varepsilon > 0$, there are a positive integer n, coefficients a_i, scalars c_i, rotators ω_i and shifts t_i, $i = 1, \ldots, n$, for which a finite sum*

$$\bar{f}(x) = \sum_{i=1}^n a_i g_{c_i}(\omega_i \cdot x - t_i) \tag{4.4}$$

satisfies

$$|\partial^\alpha f(x) - \partial^\alpha \bar{f}(x)| < \varepsilon, \quad \text{on } \mathbb{R}^d, \tag{4.5}$$

for all α, $|\alpha| \leqslant m$.

Proof. There is a function of $\mathcal{S}(\mathbb{R}^d)$ which approximates f uniformly on $\mathbb{R}^d$ in $C_0^m(\mathbb{R}^d)$. Hence, we may suppose that $f \in \mathcal{S}(\mathbb{R}^d)$. By Lemma 2.6(iii), there is a delta sequence $\{u_k\}$ on the line whose members are linear combinations of scaled shifted rotations of G. Hence, we can obtain a function f_k defined by (3.10) which satisfies (3.11). Since $\partial_t^r G$ are square integrable functions of bounded variation, so are $\partial_t^r u_k$. Set

$$f_{k,T}(x) = \int \int_{|t| \leqslant T} L\check{f}(t, \omega) u_k(\omega \cdot x - t) \, dt \, d\mu(\omega). \tag{4.6}$$

Then, by Lemma 3.2, we have that

$$\partial^\alpha f_{k,T}(x) = \int \int_{|t| \leqslant T} \omega^\alpha \partial_t^{|\alpha|} L\check{f}(t, \omega) u_k(\omega \cdot x - t) \, dt \, d\mu(\omega). \tag{4.7}$$

Since $\partial^\alpha L\check{f}(t, \omega)$ are integrable and u_k is bounded, there is $T > 1$ such that

$$|\partial^\alpha f(x) - \partial^\alpha f_{k,T}(x)| < \tfrac{1}{2}\varepsilon, \quad \text{on } \mathbb{R}^d, \tag{4.8}$$

for all α, $|\alpha| \leqslant m$. Suppose that N is sufficiently large and set

$$\boldsymbol{E}_{x,N} = \left\{ \omega \,\middle|\, \frac{|\omega \cdot x|}{|x|} \leqslant \frac{1}{N} \right\}.$$

Then, $\boldsymbol{E}_{x,N}$ is an equatorial belt of the unit sphere with poles at $\pm x/|x|$. Moreover, $\mu(\boldsymbol{E}_{x,N})$ can be arbitrarily small. If $|x| > 4NT$, we have

$$\left\{ \omega \,\middle|\, |\omega \cdot x - t| \leqslant \frac{1}{k}, |t| \leqslant T \right\} \subset \boldsymbol{E}_{x,2N}.$$

Since $\boldsymbol{E}_{x,2N} \subset \boldsymbol{E}_{x,N}$ and the distance between $\boldsymbol{E}_{x,2N}$ and $\boldsymbol{S}^{d-1} \backslash \boldsymbol{E}_{x,N}$ is positive, there is a fine partition $\{\Delta_{1i}\}$ of $[-T, T] \times \boldsymbol{S}^{d-1}$ such that

$$\boldsymbol{B}_{x,N} = \bigcup \{\Delta_{1i} | \Delta_{1i} \cap \boldsymbol{E}_{x,2N} \neq \emptyset\} \subset \boldsymbol{E}_{x,N}.$$

For x, $|x| > 4NT$, set $\boldsymbol{B}'_{x,N} = \boldsymbol{S}^{d-1} \backslash \boldsymbol{B}_{x,N}$ and divide the integral (4.7) into two parts:

$$\partial^\alpha f_{k,T}(x) = \left\{ \int \int_{|t| \leqslant T, \boldsymbol{B}_{x,N}} + \int \int_{|t| \leqslant T, \boldsymbol{B}'_{x,N}} \right\} \omega^\alpha \partial_t^{|\alpha|} L\check{f}(t, \omega) u_k(x \cdot \omega - t) \, dt \, d\mu(\omega).$$

Let (t_i, ω_i) be an arbitrary point of Δ_{1i} and N be sufficiently large. We then have

$$\int \int_{|t| \leqslant T, \boldsymbol{B}_{x,N}} |\omega^\alpha \partial_t^{|\alpha|} L\check{f}(t, \omega) u_k(x \cdot \omega - t)| \, dt \, d\mu(\omega) \leqslant \tfrac{1}{4}\varepsilon$$

and

$$\sum\nolimits_1 \int \int_{\Delta_{1i}} |\omega^\alpha \partial_t^{|\alpha|} L\check{f}(t, \omega)| \, dt \, d\mu(\omega) \, |u_k(\omega_i \cdot x - t_i)| \leqslant \tfrac{1}{4}\varepsilon,$$

where $\sum_1$ stands for a sum over i such that $\Delta_{1i} \subset \boldsymbol{B}_{x,N}$. Since

$$\boldsymbol{B}'_{x,N} \subset \left\{ \omega \,\middle|\, |\omega \cdot x - t| > \frac{1}{k} \right\}, \quad \text{for all } t, \ |t| \leqslant T,$$

we have that, for sufficiently large k,

$$\int\int_{|t|\leqslant T, CB_{x,N}} |\omega^\alpha \partial_t^{|\alpha|} L\check{f}(t,\omega) u_k(x\cdot\omega - t)| \, \mathrm{d}t \, \mathrm{d}\mu(\omega) \leqslant \tfrac{1}{4}\varepsilon$$

and

$$\sum\nolimits_2 \int\int |\omega_{\Delta_{1i}}^\alpha \partial_t^{|\alpha|} L\check{f}(t,\omega)| \, \mathrm{d}t \, \mathrm{d}\mu(\omega) \, u_k(\omega_i\cdot x - t_i)| \leqslant \tfrac{1}{4}\varepsilon,$$

where $\sum_2$ stands for a sum over i such that $\Delta_{1i} \subset B'_{x,N}$. Hence,

$$\left|\partial^\alpha f_{k,T}(x) - \sum_i \int\int_{\Delta_{1i}} \omega^\alpha \partial_t^{|\alpha|} L\check{f}(t,\omega) \, \mathrm{d}t \, \mathrm{d}\mu(\omega) \, u_k(\omega_i\cdot x - t_i)\right| \leqslant \tfrac{1}{2}\varepsilon, \tag{4.9}$$

for x, $|x| > 4NT$, where $\sum = \sum_1 + \sum_2$.

Now let us suppose that $|x| \leqslant 4NT$. Since x is confined in the compact set and the integral (4.7) is over a compact set, there is a partition $\{\Delta_{2i}\}$ of $[-T,T]\times S^{d-1}$ such that

$$\left|\partial^\alpha f_{k,T}(x) - \sum_i \int\int_{\Delta_{2i}} \omega^\alpha \partial_t^{|\alpha|} L\check{f}(t,\omega) \, \mathrm{d}t \, \mathrm{d}\mu(\omega) \, u_k(x\cdot\omega_i - t_i)\right| \leqslant \tfrac{1}{2}\varepsilon.$$

Let $\{\Delta_i\}$ be a partition of $[-T,T]\times S^{d-1}$ finer than both $\{\Delta_{1i}\}$ and $\{\Delta_{2i}\}$. Then,

$$\left|\partial^\alpha f_{k,T}(x) - \sum_i \int\int_{\Delta_i} \omega^\alpha \partial_t^{|\alpha|} L\check{f}(t,\omega) \, \mathrm{d}t \, \mathrm{d}\mu(\omega) \, u_k(x\cdot\omega_i - t_i)\right| \leqslant \tfrac{1}{2}\varepsilon, \tag{4.10}$$

on $\mathbb{R}^d$ for all α, $|\alpha| \leqslant m$. Since u_k is a linear combination of scaled shifts of G which is a linear combination of scaled shifts of g, inequalities (4.8)–(4.10) conclude the proof. □

This theorem can be extended as below. For a neighborhood U_∞ of the infinity, we write $\dot{U}_\infty = U_\infty \backslash \{\infty\}$.

Theorem 4.4. *Let $f \in C^m(\mathbb{R}^d)$ and $g \in W^m(\mathbb{R})$. Suppose that a member G_1 of $\Sigma(g,c)$ as well as its derivatives $\partial_t^r G_1$, $1 \leqslant r \leqslant m$, are functions of bounded variation, and another member G_2 of $\Sigma(g,c)$ as well as its derivatives $\partial_t^r G_2$, $1 \leqslant r \leqslant m$, are square integrable functions of uniformly locally bounded variation. Then, statement* (1) *implies* (2):

(1) *there is a neighborhood U_∞ of infinity such that f can be approximated in $W^m(\dot{U})$ by a linear combination of scaled shifted rotations of G_1;*

(2) *f can be approximated in $W^m(\mathbb{R}^d)$ by a linear combination of scaled shifted rotations of g.*

Proof. Suppose that statement (1) holds and let $\bar{f}_1$ be a linear combination of scaled shifted rotations of G_1 such that $\|f - \bar{f}_1\|_{m,\dot{U}_\infty} < \varepsilon$. Since f and its derivatives are continuous, jump heights of $\bar{f}_1$ and its derivatives are less than 2ε at all discontinuity points in $\dot{U}_\infty$. Hence, using the fact that $\partial_t^r G_1$ are functions of bounded variation, we can prove that, for a mollifier ρ with sufficiently small support, $\|\bar{f}_1 - \bar{f}_1 * \rho\|_{m,\mathbb{R}^d} < 2\varepsilon$ holds. Since $f - \bar{f}_1 * \rho \in C^m(\mathbb{R}^d)$ and $\|f - \bar{f}_1 * \rho\|_{m,\dot{U}_\infty} < 3\varepsilon$, there is a function $f_2 \in C_0^m(\mathbb{R}^d)$ such that $\|f - \bar{f}_1 * \rho - f_2\|_{m,\mathbb{R}^d} < 4\varepsilon$. By Theorem 4.3, there is a linear combination $\bar{f}_2$ of the form (4.1) for which $\|f_2 - \bar{f}_2\|_{m,\mathbb{R}^d} < \varepsilon$ holds. From these results, we obtain that $\|f - \bar{f}_1 - \bar{f}_2\|_{m,\mathbb{R}^d} < 5\varepsilon$. Since $\bar{f}_1 + \bar{f}_2$ is a linear combination of the form (4.1), the proof is concluded. □

Denote by $\overline{\mathbb{R}}^d$ the one-point compactification of $\mathbb{R}^d$ and suppose that g satisfies the condition of this theorem for $m = 0$. For an appropriate t_0, $g(-t_0)$ is a nonzero constant, which implies that any $f \in C(\overline{\mathbb{R}}^d)$ can be approximated uniformly in a neighbourhood of the origin. Hence, by Theorem 4.4, such f can be approximated uniformly on $\overline{\mathbb{R}}^d$ by a linear combination of scaled shifted rotations of g. Since any sigmoid function satisfies the condition of the theorem, any $f \in C(\overline{\mathbb{R}}^d)$ can be approximated in $W(\overline{\mathbb{R}}^d)$ by any sigmoid function. This result is proved in [14].

Theorem 4.5. *Let $g \in W^m(\mathbb{R})$. Suppose that a certain linear combination G of shifted rotations of g is a nonzero function such that $\partial_t^r G$, $0 \leqslant r \leqslant m$, are square integrable functions of uniformly locally bounded variation and* supp$(\mathcal{F}G)$ *is dense on $\mathbb{R}$. Then, for any $f \in C_0^m(\mathbb{R}^d)$ and any $\varepsilon > 0$, there are a positive integer n, coefficients a_i, rotators ω_i and shifts t_i, $i = 1, \ldots, n$, for which a finite sum*

$$\bar{f}(x) = \sum_{i=1}^{n} a_i g(\omega_i \cdot x - t_i) \tag{4.11}$$

satisfies

$$|\partial^\alpha f(x) - \partial^\alpha \bar{f}(x)| < \varepsilon, \quad \text{on } \mathbb{R}^d, \tag{4.12}$$

for all α, $|\alpha| \leqslant m$.

Conversely, if supp$(\mathcal{F}G)$ *is not dense on $\mathbb{R}$, there is a function f of $C_0^m(\mathbb{R}^d)$ and $\varepsilon > 0$ for which any linear combination of shifted rotations of G never satisfies* (4.12).

Proof. By Lemma 2.6(iv), there is a delta sequence on the line, each member of which is a linear combination of unscaled shifted rotations of G. Hence, the proof of the first half of the theorem is similar to that of Theorem 4.3.

Set $G_{\omega,t}(x) = G(\omega \cdot x - t)$ and let $\mathcal{F}_d$ stand for the Fourier transform on $\mathbb{R}^d$. The support of $\mathcal{F}G$ is symmetric with respect to the origin. Hence, if it is not dense on $\mathbb{R}$, there is a spherically symmetric open set $\boldsymbol{B} \subset \mathbb{R}^d$ which has no intersection with the set

$$\bigcup_{\omega,t} \operatorname{supp}(\mathcal{F}_d G_{\omega,t}).$$

There is a function f of $\mathcal{S}(\mathbb{R}^d)$ whose support is contained in $\boldsymbol{B}$. Analogously to the proof of Lemma 2.3(iv), we can prove that f cannot be approximated by a linear combination of $G_{\omega,t}$'s. This concludes the proof. □

This theorem can be extended as follows.

Theorem 4.6. *Let $f \in C^m(\mathbb{R}^d)$ and $g \in W^m(\mathbb{R})$. Suppose that a member G_1 of $\Sigma(g)$ as well as its derivatives $\partial_t^r G_1$, $1 \leqslant r \leqslant m$, are functions of bounded variation, and another member G of $\Sigma(g)$ as well as its derivatives $\partial_t^r G$, $1 \leqslant r \leqslant m$, are square integrable functions of uniformly locally bounded variation. Then, if* supp$(\mathcal{F}G)$ *is dense in $\mathbb{R}$, statement* (1) *implies* (2):

(1) *there is a neighborhood* U_∞ *of infinity such that* f *can be approximated in* $W^m(\dot{U})$ *by a linear combination of scaled shifted rotations of* G_1;

(2) f *can be approximated in* $W^m(\mathbb{R}^d)$ *by a linear combination of scaled shifted rotations of* g.

Conversely, if $\operatorname{supp}(\mathcal{F}G)$ *is not dense on* $\mathbb{R}$, *there is a function* f *of* $C_0^m(\mathbb{R}^d)$ *which satisfies* (1) *but not* (2).

Proof. The proof of the first half is similar to that of Theorem 4.4 and the proof of the second half to that of Theorem 4.5. □

Let g be a sigmoid function. Then, a constant can be approximated uniformly in a neighbourhood of the infinity by a linear combination of shifted rotations of g. Hence, this theorem implies that if and only if $\operatorname{supp}(\mathcal{F}g)$ is dense on $\mathbb{R}$, any $f \in C(\overline{\mathbb{R}}^d)$ can be approximated uniformly on $\overline{\mathbb{R}}^d$ by a linear combination of shifted rotations of g. This fact is also proved in [14] by a distinct method.

5. Discussion

In this paper, we have shown systematically the usefulness of the inverse Radon transform for differentiable approximation both on compact sets and on the whole space $\mathbb{R}^d$ by neural networks. However, there are alternative easier proofs of Theorems 4.1 and 4.2 which guarantee the differential approximation on compact sets. As described in [13], approximation of functions on compact sets is easier than approximation on the whole space in many cases. This is true in the case where the approximation is extended to derivatives, too. By Nachbin's theorem, a function of $C^m(\boldsymbol{K})$ can be approximated in $C^m(\boldsymbol{K})$ by a polynomial for any compact set $\boldsymbol{K}$ of $\mathbb{R}^d$ and any polynomial can be expressed as a linear combination of powers of the form $(\omega \cdot x)^r$ [13]. Let $\{u_k\}_{k=1}^{\infty}$ be a delta sequence of functions of $W^m(\mathbb{R})$ on finite intervals, the derivatives of which are functions of locally bounded variation. For any finite closed interval $\boldsymbol{F} \subset \mathbb{R}$, there is a function $v_r \in C_c^\infty(\mathbb{R})$ which coincides with t^r in a neighbourhood of $\boldsymbol{F}$. Hence, by Lemma 3.3, t^r can be approximated in $C^m(\boldsymbol{K})$ by a convolution of v_r and an element u_k of the delta sequence. The convolution can be approximated in $C^m(\boldsymbol{K})$ by a linear combination of shifts of u_k as in the proof of Theorem 4.1. Hence, Theorems 4.1 and 4.2 can be proved without implication of the inverse Radon transform.

All theorems described in [12–14] are extended in this paper not only because the approximation is extended to derivatives, but also because the class of useful functions as the component function g in (1.1) is extended. Although we have used a different method here than in previous papers, the results of this paper contain all the previous results. Hence, the present method can be regarded as an extended alternative method of the proofs of the respective previous theorems. For example, in [14] a necessary and sufficient condition is obtained which ensures that a linear combination of an unscaled sigmoid function can approximate any function of $C_0(\mathbb{R}^d)$. The condition is that $\operatorname{supp}(\mathcal{F}\rho_h)$ is dense in $\mathbb{R}$, which coincides with the condition on $\operatorname{supp}(\mathcal{F}G)$. This implies that the proof of Theorem 4.5 is an alternative proof of the extension of [14, Theorem 3.5]. *Note*: There is a mistake in [14]: "$\operatorname{supp}(\mathcal{R}(\mathcal{F}\sigma_h))$" must be replaced by "$\operatorname{supp}(\mathcal{F}\sigma_h)$".

This paper not only extends [12–14], but also [8,10] which have dealt with differentiable approximation. When compared with these papers, the extension in this paper is threefold: (a) in the case of approximation on compact sets, the class of component functions is extended to slowly increasing function of $W^m(\mathbb{R}^d)$; (b) the domain of differentiable approximation is extended to the whole space $\mathbb{R}^d$; (c) another extension attained in this paper is the implementation of differentiable uniform approximation without scaling the activation function.

It must be noted that not any polynomial can be the activation function. This is obvious because a superposition of polynomials of order n is a polynomial of order equal to or less than n. From our point of view, the reason is that the Fourier transform of a polynomial has a support restricted at the origin. The statement of Lemma 2.2(i) excludes explicitly polynomials as component functions and the respective conditions of statements (ii)–(iv) of the lemma automatically exclude polynomials. It is interesting that the same conclusion was obtained in [15] by a distinct method.

The proofs in this paper can be regarded to be constructive, except for the proof of existence of the delta sequences. However, we can easily construct a delta sequence in many cases as we have described. Even if the component function cannot be scaled, construction of a delta sequence is sometimes possible. Then, a three-layered neural network, which can implement differentiable approximation, can be constructed in principle according to the proofs of this paper, although it may be a somewhat laborious work.

Acknowledgements

The author wishes to express his deep gratitude to Professors I. Kubo, T. Matsuzawa and C. Srinivasan for their helpful discussion and careful reading of the manuscript.

References

[1] B.W. Carroll and B.D. Dickinson, Construction of neural nets using the Radon transform, in: S. Amari, Ed., *Proc. IJCNN, Vol. 1* (IEEE Press, New York, 1989) 607–611.

[2] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control Signal Systems* **2** (1989) 303–314.

[3] A. Erdélyi, W. Magnus, F. Oberhettinger and F.G. Tricomi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954).

[4] K. Funahashi, On the approximate realization of continuous mapping by neural networks, *Neural Networks* **2** (1989) 183–192.

[5] I.M. Gel'fand, M.I. Greav and N.Ya. Vilenkin, *Generalized Function, Vol. 5* (Academic Press, New York, 1966) (English translation).

[6] R. Hecht-Nielsen, Theory of the back propagation neural network, in: S. Amari, Ed., *Proc. IJCNN, Vol. 1* (IEEE Press, New York, 1989) 593–605.

[7] S. Helgason, *The Radon Transform* (Birkhäuser, Basel, 1980).

[8] K. Hornik, Approximation capabilities of multilayer feedforward networks, *Neural Networks* **4** (1991) 251–257.

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

[10] K. Hornik, M. Stinchcombe and H. White, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, *Networks* **3** (1990) 551–560.

[11] B. Irie and S. Miyake, Capability of three-layered perceptrons, in: T. Kohonen, Ed., *IEEE Internat. Conf. on Neural Networks, Vol. 1* (IEEE Press, New York, 1988) 641–648.

[12] Y. Ito, Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory, *Neural Networks* **4** (1991) 385–394.

[13] Y. Ito, Approximation of functions on a compact set by finite sums of a sigmoid function without scaling, *Neural Networks* **4** (1991) 817–826.

[14] Y. Ito, Approximation of continuous functions of $\mathbb{R}^d$ by linear combinations of shifted rotations of a sigmoid function with and without scaling, *Neural Networks* **5** (1992) 105–115.
[15] Y. Ito, Approximations of differentiable functions and their derivatives on compact sets by neural networks, *Math. Sci.* **18** (1993) 11–19.
[16] W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1966).
[17] W. Rudin, *Functional Analysis* (McGraw-Hill, New York, 1973).
[18] L. Schwartz, *Théorie des Distributions* (Hermann, Paris, 1966).
[19] M. Stinchcombe and H. White, Universal approximation using feedforward networks with non-sigmoid hidden layer activation functions, in: S. Amari, Ed., *Proc. IJCNN, Vol. I* (IEEE Press, New York, 1989) 613–617.
[20] M. Stinchcombe and H. White, Approximating and learning unknown mappings using multilayer feedforward networks with bounded weights, in: J. Goodman and W. Karplus, Eds., *Proc. IJCNN, Vol. III*, San Diego, CA (IEEE Press, New York, 1990) 7–16.