When is approximation by Gaussian networks necessarily a linear process?

H. N. Mhaskar*
Department of Mathematics, California State University
Los Angeles, California, 90032, U.S.A.

Abstract

Let $s \geq 1$ be an integer. A Gaussian network is a function on $\mathbb{R}^s$ of the form $g(\mathbf{x}) = \sum_{k=1}^N a_k \exp(-\|\mathbf{x}-\mathbf{x}_k\|^2)$. The minimal separation among the centers, defined by $(1/2)\min_{1\leq j\neq k\leq N}\|\mathbf{x}_j-\mathbf{x}_k\|$, is an important characteristic of the network that determines the stability of interpolation by Gaussian networks, the degree of approximation by such networks, etc. Let (within this abstract only) the set of all Gaussian networks with minimal separation exceeding $1/m$ be denoted by $\mathcal{G}_m$. We prove that for functions $f \in L^2(\mathbb{R}^s)$ such that $\|f\|_{\mathbb{R}^s\setminus[-t,t]^s} = \mathcal{O}(t^{-\beta})$, if the degree of L^2 (nonlinear) approximation of f from $\mathcal{G}_m$ is $\mathcal{O}(m^{-\beta})$, then necessarily the degree of approximation of f by (rectangular) partial sums of degree m^2 of the Hermite expansion of f is also $\mathcal{O}(m^{-\beta})$. Moreover, Gaussian networks in $\mathcal{G}_m$ having fixed centers in a ball of radius $\mathcal{O}(m)$ and coefficients being linear functionals of f can be constructed to yield the same degree of approximation. Similar results are proved for the L^p norms, $1 \leq p \leq \infty$, but with the condition that the number of neurons N, should satisfy $\log N = \mathcal{O}(m^2)$.

1 Introduction

Let $s, N \geq 1$ be integers. A *Gaussian network* with N *neurons* is a function on the Euclidean space $\mathbb{R}^s$ of the form $\mathbf{x} \mapsto \sum_{k=1}^N a_k \exp(-\|\mathbf{x}-\mathbf{x}_k\|^2)$, where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^s$, the *centers* $\mathbf{x}_k$ are in $\mathbb{R}^s$, and $a_k \in \mathbb{R}$, $k = 1, \cdots, N$. These functions can be evaluated in hardware using parallel computation of the exponential terms, and are used extensively in many applications in pattern recognition, computer graphics, antenna array theory, probability density estimation, etc. A typical problem in all these applications is to approximate an unknown function (the *target function*) by such networks. The approximation power of Gaussian networks has also been well studied [13, 4, 8, 5, 6]. In [13], Park and Sandberg have proved that these networks possess universal approximation property; i.e., they can approximate an arbitrary continuous function on $\mathbb{R}^s$ uniformly and arbitrarily well on compact subsets of $\mathbb{R}^s$.

The complexity problem in the theory of function approximation is to determine the connection between the smoothness of the target function and the cost necessary to achieve an approximation with a desired accuracy. We describe this problem in some abstraction, before describing the goals of the present paper. The accuracy of approximation is usually measured with a norm on a normed linear space, X. The approximating functions are chosen from one of an increasing sequence of subsets of X, $V_0 \subset V_1 \subset \cdots \subset X$, where the index m of V_m denotes the cost associated with computing the approximation. For example, V_m may be the class of all polynomials of degree at most m. Since any element of V_m requires $m+1$ real parameters, and $2m$ floating point operations to evaluate it, m is the cost associated with the class V_m in this example. In the theory of cardinal spline approximation, the cost is usually taken to be the (reciprocal of the) separation between consecutive knots; in signal processing applications, one may consider the sampling frequency as the appropriate cost. The *degree of approximation* of an element $f \in X$ from V_m is the distance of f from V_m, measured in the norm of

---

*The research of this author was supported, in part, by grant DMS-0204704 from the National Science Foundation and grant DAAD19-01-1-0001 from the U.S. Army Research Office.

X. Although the target function f itself is typically not known, it is usual to assume that f satisfies some a priori conditions, encoded by the statement that f belongs to a compact subset of X. It is natural to expect that the stronger the conditions on f, the faster will be the rate at which the degree of approximation of f from V_m tends to 0 as $m \to \infty$. So, one considers a scale of compact sets W_α (for example, the Sobolev classes), where the parameter α denotes the "smoothness" of f. A typical *direct theorem* of approximation theory has the form: "If $f \in W_\alpha$ then the degree of approximation of f from V_m is $\mathcal{O}(m^{-\alpha})$ for all $m \geq 1$". The corresponding *converse theorem* is usually much deeper, and takes the form: "If the degree of approximation of f from V_m is $\mathcal{O}(m^{-\alpha})$ for all $m \geq 1$, then $f \in W_\alpha$". Direct and converse theorems are available in a variety of situations (cf. [7, 8]). Therefore, in approximation theory, one sometimes *defines* the smoothness class by the rate of decay of the degrees of approximation of its elements from a given sequence of sets V_m.

In this paper, we will explore the idea of measuring the cost of approximation by Gaussian networks in terms of the (reciprocal of the) minimal separation among the centers, as in the case of spline functions, rather than in terms of the number of neurons.

In the case of Gaussian networks (and radial basis function networks in general), a popular strategy for approximating the target function is to interpolate the given data, taking the centers from the data itself. In [11], Micchelli has proved that in contrast to polynomials, such an interpolation is always possible for a large class of radial basis functions, Gaussian networks in particular. The error estimates in this context are often obtained in the case of lattice data, and are in terms of the minimal separation between the points at which the interpolation takes place [2]. In [12], Narcowich and Ward have estimated the condition numbers of the interpolation matrices in the context of a general scattered data interpolation. Their estimates are in terms of the minimal separation between the interpolation points, independent of the number of points (and hence, of neurons) involved. In private conversations, many people have confirmed that the minimal separation is the main issue they have to face in order to compute efficient approximations.

Although the idea of thinking of the reciprocal of the minimal separation among the centers as the cost of approximation is motivated by these considerations, we will not be studying interpolation here. Instead, our main focus is to prove a converse theorem for approximation by Gaussian networks. We will show that there is a close connection between approximation by Gaussian networks and that by *weighted polynomials;* i.e., expressions of the form $P \exp(-\|\cdot\|^2/2)$, where P is a polynomial in s variables. Direct and converse theorems for weighted polynomial approximation are well studied ([3, 8, 9]). Therefore, we will define the smoothness classes in terms of the degrees of approximation by weighted polynomials. There are more explicit, albeit complicated, expressions for these smoothness classes. One example is given in the appendix.

An interesting consequence of this investigation is the following. We observe that, in general, approximation by Gaussian networks, even for a fixed number of neurons, is a nonlinear procedure, since the centers $\mathbf{x}_k$ may be chosen dependent on the target function. In contrast, the problem of approximation by weighted polynomials is a linear problem. Indeed, there are well known ([3, 8, 9]) linear operators which give the best order of magnitude for the degree of approximation by weighted polynomials of any given degree. Therefore, our converse theorem implies that if the degree of (the inherently nonlinear) approximation by Gaussian networks, measured in terms of the minimal separation among the centers, decays at a polynomial rate, then the same rate may be achieved using the linear operators of weighted polynomial approximation, with a commensurate cost, measured in terms of the degree of the polynomials involved. We note again, that there is no assumption about how the approximating Gaussian networks are arrived at in verifying the rate of decay of the degrees of approximation.

This raises the question of constructing Gaussian networks as linear operators, which yield the necessary degree of approximation. In this aspect, one can be more ambitious, and require that the network should have the right number of neurons, in addition to having the right minimal separation among the centers. In [8, Section 11.2], we have constructed Gaussian networks using linear procedures. If the degree of approximation is measured in terms of the number of neurons involved, then the operators yield the optimal order of magnitude for the degree of approximation, as measured by the theory of n-widths [1] for functions belonging to modified Sobolev classes. One criticism of these constructions was that the centers were required to cluster very close to each other. In this paper, we will remedy this concern, and construct networks as linear operators, which yield the optimal order of magnitude for the degree of approximation, measured both in terms of the number of neurons and the minimal separation among the centers.

The first drafts of this paper have been in circulation for at least 1.5 years at the time of this writing. Even though our focus in this paper is on the minimal separation rather than the number of neurons, some new results have emerged in this theory in the meantime, where the cost of approximation is measured, as usual, in terms of the number of neurons. One of the referees of this paper, as well as Professor Ward at Texas A&M University, have pointed out that our results implicitly contain an inequality, known as Bernstein inequality, in terms of the number of neurons, under some conditions on the minimal separation. Professor Erdelyi at Texas A&M University has kindly sent us a manuscript in preparation, where he proves this inequality purely in terms of the number of neurons, with no further conditions. This inequality leads to the converse theorems in terms of the number of neurons, matching our direct theorem in this theory. Our direct theorem in [8] is sharp in the sense of n-widths [1]. However, the converse theorem applies to individual functions rather than a class of functions. In particular, it appears that even if the cost of approximation is measured in terms of the number of neurons, if the degrees of approximation of a particular function by Gaussian networks decays polynomially, then a linear operator will yield the same order of magnitude in the error in approximating this function. We find it astonishing, because many people have told us, based on numerical experiments, that one can achieve a better degree of approximation by nonlinear procedures stacking the centers near the "bad points" of the target function.

The main results of this paper are discussed in Section 2. The proofs of the converse theorems, including all the auxiliary results, are given in Section 3. The proof of the direct theorem, including all the auxiliary results needed there, are given in Section 4. The construction of Gaussian networks as linear operators is reviewed in Section 5.

We would like to thank the referees for their numerous suggestions for the improvement of the presentation in this paper.

2 Main Results

Throughout this paper, s will be considered a fixed integer, $s \geq 1$. Our main objective in this section is to compare the degree of approximation to a function of s variables by Gaussian networks with that given by weighted polynomials. For $x > 0$, let $\Pi_{x,s}$ be the class of all polynomials in s real variables with coordinatewise degree not exceeding x. The symbol $W\Pi_{x,s}$ denotes the class of all functions of the form $\mathbf{x} \to \exp(-\|\mathbf{x}\|^2/2)P(\mathbf{x})$, $P \in \Pi_{x,s}$, $\mathbf{x} \in \mathbb{R}^s$. In the appendix, we discuss the connection between the smoothness of the target function and the degree of its weighted polynomial approximation. The notation for the class of Gaussian networks will also involve different bounds on the centers as well as the number of neurons involved. Thus, for $m, M, N > 0$, the symbol $\mathbb{G}_{N,M,m,s}$ denotes the class of functions of the form

$$\mathbf{x} \mapsto \sum_{1\leq k\leq N,\ k\in\mathbb{Z}} a_k \exp(-\|\mathbf{x}-\mathbf{x}_k\|^2), \qquad \mathbf{x}, \mathbf{x}_k \in \mathbb{R}^s,\ a_k \in \mathbb{R},\ 1 \leq k \leq N, \tag{2.1}$$

where $\max_{1\leq k\leq N} \|\mathbf{x}_k\| \leq M$ and the *minimal separation,* $(1/2) \min\limits_{1\leq k,j\leq N, k\neq j} \|\mathbf{x}_j - \mathbf{x}_k\| \geq m^{-1}$. Also, the union of the class of networks over a certain parameter will be denoted by writing the symbol ∞ in place of that parameter; for example, $\mathbb{G}_{N,\infty,m,s} := \cup_{M>0}\mathbb{G}_{N,M,m,s}$, etc. If $1 \leq p \leq \infty$, and $f : \mathbb{R}^s \to \mathbb{R}$ is a Lebesgue measurable function, and $A \subseteq \mathbb{R}^s$ is a Lebsegue measurable set having positive measure, we write

$$\|f\|_{p,A} := \begin{cases} \{\int_A |f(\mathbf{x})|^p d\mathbf{x}\}^{1/p}, & \text{if } 1 \leq p < \infty, \\ \operatorname*{ess\,sup}\limits_{\mathbf{x}\in A} |f(\mathbf{x})|, & \text{if } p = \infty. \end{cases} \tag{2.2}$$

The set of all functions for which $\|f\|_{p,A} < \infty$ is denoted by $L^p(A)$, where, as usual, two functions that are equal almost everywhere on A are considered equal as elements of $L^p(A)$. If $f \in L^p(\mathbb{R}^s)$, and S is a subset of $L^p(\mathbb{R}^s)$, we write

$$\text{dist }(p; f, S) := \inf_{g\in S} \|f - g\|_{p,\mathbb{R}^s}. \tag{2.3}$$

In the sequel, we adopt the following convention regarding constants. The symbols $c, c_1, \cdots$ will denote positive constants depending only on s and p, but their values may be different at different occurrences, even within a single formula. The symbol S_f will denote a positive constant that may depend upon the target function f as well, and its value may also be different at different occurrences, even within a single formula.

For the clarity of presentation, we begin by first discussing our results in the case when $p = 2$. In light of the connection between the degree of weighted polynomial approximation of a function and its smoothness (Theorem A.1 in the appendix), the following theorem is a "converse theorem" for approximation by Gaussian networks.

Theorem 2.1 *Let* $f \in L^2(\mathbb{R}^s)$, $\beta, \beta_1 > 0$. *Suppose that*

$$m^{\beta_1}\|f\|_{2,\mathbb{R}^s\setminus[-cm^2,cm^2]^s} \le c_1 S_f, \qquad m \ge 1, \tag{2.4}$$

and

$$\text{dist }(2; f, \mathbb{G}_{\infty,\infty,m,s}) \le c_1 m^{-\beta} S_f, \qquad m \ge 1, \tag{2.5}$$

for some positive constant $S_f \ge \|f\|_{2,\mathbb{R}^s}$ *depending on* f, s, β_1, *and* β. *Then*

$$\text{dist }(2; f, W\Pi_{n,s}) \le cn^{-\min(\beta,\beta_1)/2} S_f, \qquad n = 1, 2, \cdots. \tag{2.6}$$

We remark that the above theorem requires no assumptions regarding the size of the networks or the different parameters involved in order to achieve the estimate (2.5), which may all depend nonlinearly upon the target function. Also, the theorem applies to individual functions, rather than a class of functions as in the theory of n-widths [1]. Indeed, the theory of n-widths is not applicable here, because we measure the cost of approximation in terms of the minimal separation among the centers, rather than the number of parameters. Finally, we remark that the condition (2.4) influences the degree of approximation by weighted polynomials. It is not difficult to construct polynomially decaying functions f in $L^2(\mathbb{R}^s)$ for which dist $(2; f, \mathbb{G}_{\infty,\infty,m,s}) = 0$, and dist $(2; f, W\Pi_{n,s})$ cannot decay too fast. One example in the univariate case is $\sum_{k=1}^{\infty} k^{-2}\exp(-(\cdot - k)^2)$, for which dist $(2; f, \mathbb{G}_{\infty,\infty,1,1})$ is obviously 0, but it can be shown using properties of Hermite polynomials that dist $(2; f, W\Pi_{n,1}) \ge cn^{-3/2}$.

The direct theorem corresponding to the converse theorem above, will be stated in greater generality below as Theorem 2.3. It is well known that the best $L^2(\mathbb{R}^s)$ approximation from $W\Pi_{n,s}$ is simply the n-th (rectangular) partial sum of the weighted Hermite expansion of the target function (cf. [14, Theorem 4.16]). The direct and converse theorems here can therefore be regarded as giving sufficient conditions under which approximation by Gaussian networks can be achieved using linear procedures, without sacrificing the order of magnitude of the degree of approximation.

Along with the direct theorem to be stated later (Theorem 2.3) below, Theorem 2.1 implies the following complete characterization of the functions with a polynomial decay near infinity and polynomial decay for the degree of approximation by Gaussian networks.

Corollary 2.1 *Let* $f \in L^2(\mathbb{R}^s)$, $\beta > 0$. *The following statements* (a), (b), *and* (c) *are equivalent.*
(a) *We have for every* $m \ge 1$,

$$m^{\beta}\|f\|_{2,\mathbb{R}^s\setminus[-cm^2,cm^2]^s} \le c_1 S_f, \tag{2.7}$$

and (2.5) holds for some positive constant $S_f \ge \|f\|_{2,\mathbb{R}^s}$ *depending on* f, s, *and* β.
(b) *We have*

$$\text{dist }(2; f, W\Pi_{n,s}) \le cn^{-\beta/2} S_f, \qquad n = 1, 2, \cdots. \tag{2.8}$$

(c) *We have*

$$\text{dist }(2; f, \mathbb{G}_{c_1 m^{2s}, c_2 m, m, s}) \le cm^{-\beta} S_f, \qquad m \ge 1. \tag{2.9}$$

In the case of the more general L^p spaces, in particular, in the case of uniform approximation, a converse theorem similar to Theorem 2.1 holds under a mild additional assumption on the number of neurons. However, we note that in spite of this assumption, the actual size of the networks yielding the estimates (2.11) below, as well as the other parameters of the networks, are allowed to be chosen possibly nonlinearly on the target function.

Theorem 2.2 *Let $1 \le p \le \infty$, $f \in L^p(\mathbb{R}^s)$, $\beta, \beta_1 > 0$. Suppose that*

$$m^{\beta_1} \|f\|_{p,\mathbb{R}^s \setminus [-cm^2, cm^2]^s} \le c_1 S_f, \qquad m \ge 1, \tag{2.10}$$

and

$$\text{dist }(p; f, \mathbb{G}_{N_m,\infty,m,s}) \le c_2 m^{-\beta} S_f, \qquad m \ge 1, \tag{2.11}$$

for some positive constant $S_f \ge \|f\|_{p,\mathbb{R}^s}$ depending on f, p, s, β_1, and β, and

$$N_m \le c_3 \exp(c_4 m^2), \qquad m \ge 1, \tag{2.12}$$

for some positive constants c_3, c_4 depending only on p, s, and β. Then

$$\text{dist }(p; f, W\Pi_{n,s}) \le c n^{-\min(\beta,\beta_1)/2} S_f, \qquad n = 1, 2, \cdots. \tag{2.13}$$

In [8], we have described a recipe for obtaining Gaussian networks to approximate any weighted polynomial arbitrarily closely. While the interest there was only in estimating the size of the network, the following direct theorem estimates the degree of approximation by Gaussian networks, both in terms of the number of neurons involved as well as the minimal separation among the centers, when the corresponding degree of approximation by weighted polynomials is known. Linear procedures for a near best weighted polynomial approximation are described in [9]. The operators there may also be evaluated using suitable quadrature formulas without disturbing the degree of approximation, thus yielding linear approximation schemes depending on samples of the target function. In Section 5, we will indicate how these operators can be used to obtain Gaussian networks that are linear operators of the target function, and yield the same order of magnitude for the degree of approximation as that given by weighted polynomial approximation. Thus, in the case of the general L^p spaces as well, Theorem 2.2 gives sufficient conditions under which the inherently nonlinear process of approximation by Gaussian networks can be replaced by a linear procedure, without sacrificing the degree of approximation.

The main interest in the following Theorem 2.3 is that it is a direct theorem that complements the converse theorems formulated above, where we keep track of the minimal separation, the maximum norm of the centers, and the number of centers, all at the same time.

Theorem 2.3 *Let $1 \le p \le \infty$, $f \in L^p(\mathbb{R}^s)$, and $\beta > 0$. Then (2.13) holds (with $\beta = \beta_1$) if and only if*

$$\text{dist }(p; f, \mathbb{G}_{\infty,cm,m,s}) \le c m^{-\beta} S_f, \qquad m \ge 1. \tag{2.14}$$

We note that the class $\mathbb{G}_{\infty,cm,m,s} \subseteq \mathbb{G}_{c_1 m^{2s},cm,m,s}$ for some $c_1 > 0$. Thus, if (2.13) holds, then (2.14) gives the degree of approximation by Gaussian networks both in terms of the number of parameters (which is comparable to that involved in weighted polynomial approximation), as well as in terms of the minimal separation among the centers. Viewed in terms of the number of parameters alone, the construction gives an optimal order of magnitude for the degree of approximation of functions in Sobolev classes, as measured by the nonlinear n-width of these classes. Further, the infinite-finite-range inequalities of weighted approximation (cf. [8, Theorem 6.1.6(c)]) show that an estimate of the form (2.13) implies (2.10) with the same exponent of $m = \sqrt{n}$ as in (2.13).

3 Proof of the converse theorems

3.1 Ideas behind the proof

In order to prove Theorems 2.1 and 2.2, we would like to demonstrate that for any Gaussian network $g = \sum a_k \exp(-\|\cdot - \mathbf{x}_k\|^2)$ with $\|g\|_{p,\mathbb{R}^s} \le 1$, and the minimal separation among the nodes exceeding $1/m$, dist $(p; g, W\Pi_{m^2,s})$ decays faster than any polynomial in $1/m$. In turn, this involves estimating $\sum |a_k|$ and dist $(p; \exp(-\|\cdot - \mathbf{x}_k\|^2), W\Pi_{m^2,s})$.

The first estimate uses a theorem of Narcowich and Ward [12] (Lemma 3.2 below) regarding the lowest eigenvalue of the matrix $(\exp(-\|\mathbf{x}_k - \mathbf{x}_j\|^2)$. An interesting feature of this theorem is that the estimate is independent of the number of neurons involved; which motivated our desire to treat the minimal separation as the measurement of the cost of approximation. In the case when $p = 2$, this theorem yields an estimate on $\sum |a_k|$ in a trivial way. In the more general case, this involves the construction and estimation of an orthonormal basis for the span of $\{\exp(-\|\cdot - \mathbf{x}_k\|^2)\}$.

As to the quantity dist $(p; \exp(-\|\cdot - \mathbf{x}_k\|^2), W\Pi_{m^2,s})$, the infinite-finite range inequalities of weighted polynomial approximation (cf. [8, Theorem 6.1.6(c)]) imply that this quantity cannot tend to zero if $\|\mathbf{x}_k\| \ge cm$ for any $c > 1$. Therefore, we have to consider g as a sum of two networks, one with centers in the cube $[-am, am]^s$ for a suitable constant a, and the other with centers outside of this cube. When $\mathbf{x}_k \in [-am, am]^s$, one can estimate dist $(p; \exp(-\|\cdot - \mathbf{x}_k\|^2), W\Pi_{m^2,s})$ using the generating function for the Hermite polynomials in a simple manner. Together with the estimate on $\sum |a_k|$, this leads to the right approximation for this part of the network g. We will show that the other part has a small norm. This part is easy in the general case, when we have made an assumption regarding the number of neurons in the whole network. The case when $p = 2$ needs a more delicate estimate using the Narcowich-Ward estimate.

There are many other technical details in the proof, which require extensively the theory of weighted polynomial approximation [8, 9] and the properties of Hermite polynomials [15].

The proofs of the coefficient and subnetwork inequalities are different in the L^2 case from the more general case. The L^2 case is discussed in Section 3.2; the more general case is discussed in Section 3.3. The remaining parts of the proofs are the same for all the L^p spaces, $1 \le p \le \infty$. The approximation of Gaussian networks by weighted polynomials is discussed in Section 3.4. The proofs are concluded in Section 3.5.

3.2 Coefficient inequalities for $L^2(\mathbb{R}^s)$

In this subsection, we prove an estimate on the coefficients of a Gaussian network in terms of the L^2-norm of the network. We will also use this estimate to prove an inequality for the norm of a subnetwork with centers outside a "large" cube. The starting point of this investigation is the following observation.

Lemma 3.1 *Let* $g = \sum a_k \exp(-\|\cdot - \mathbf{x}_k\|^2) \in \mathbb{G}_{N,\infty,\infty,s}$, *and*

$$\mathcal{A} := (\sqrt{\pi/2})^s (\exp(-\|\mathbf{x}_j - \mathbf{x}_k\|^2/2))_{j,k=1}^N. \tag{3.1}$$

Then

$$\|g\|_{2,\mathbb{R}^s}^2 = \sum_{\ell,j=1}^N a_j a_\ell \mathcal{A}_{j,\ell}. \tag{3.2}$$

Proof. The equation (3.2) is clear from the fact that for any integers $1 \le \ell, j \le N$,

$$\begin{aligned}
&\int_{\mathbb{R}^s} \exp(-\|\mathbf{x} - \mathbf{x}_\ell\|^2) \exp(-\|\mathbf{x} - \mathbf{x}_j\|^2) d\mathbf{x} \\
&= \int_{\mathbb{R}^s} \exp(-\|\mathbf{x} - \mathbf{x}_\ell + \mathbf{x}_j\|^2) \exp(-\|\mathbf{x}\|^2) d\mathbf{x} \\
&= \int_{\mathbb{R}^s} \exp\left(-2\|\mathbf{x}\|^2 + 2\mathbf{x} \cdot (\mathbf{x}_\ell - \mathbf{x}_j) - \|\mathbf{x}_\ell - \mathbf{x}_j\|^2\right) d\mathbf{x} \\
&= \exp(-\|\mathbf{x}_\ell - \mathbf{x}_j\|^2/2) \int_{\mathbb{R}^s} \exp\left(-2\left\|\mathbf{x} - \frac{\mathbf{x}_\ell - \mathbf{x}_j}{2}\right\|^2\right) d\mathbf{x} \\
&= (\sqrt{\pi/2})^s \exp(-\|\mathbf{x}_\ell - \mathbf{x}_j\|^2/2).
\end{aligned} \tag{3.3}$$

□

In [12], Narcowich and Ward have given an important estimate on the minimal eigenvalue of the matrix $\mathcal{A}$.

Lemma 3.2 *Let* $N \ge 1$ *be an integer,* $\mathbf{x}_1, \cdots, \mathbf{x}_N$ *be any points in* $\mathbb{R}^s$,

$$q := \frac{1}{2} \min_{1 \le j,k \le N} \|\mathbf{x}_j - \mathbf{x}_k\| \tag{3.4}$$

be the minimal separation between the points, and

$$\delta := \delta_s := 12\left(\frac{\pi}{9}\Gamma^2\left(\frac{s+2}{2}\right)\right)^{1/(s+1)}. \tag{3.5}$$

The matrix $\mathcal{A}$, defined in (3.1), is positive definite, and its lowest eigenvalue λ_{min} satisfies

$$\lambda_{min} \geq cq^{-s}\exp(-2\delta^2 q^{-2}). \tag{3.6}$$

The following proposition gives some useful, although perhaps crude, estimates on the coefficients of the network g in terms of its norm.

Proposition 3.1 *Let $N \geq 1$ be an integer, $\mathbf{x}_1, \cdots, \mathbf{x}_N$ be any points in $\mathbb{R}^s$, q be the minimal separation as defined in (3.4), δ be as in (3.5), and*

$$g(\mathbf{x}) = \sum_{k=1}^{N} a_k \exp(-\|\mathbf{x}-\mathbf{x}_k\|^2). \tag{3.7}$$

Then

$$\sum_{k=1}^{N}|a_k|^2 \leq cq^s\exp(2\delta^2 q^{-2})\|g\|^2_{2,\mathbb{R}^s}, \tag{3.8}$$

and

$$\sum_{k=1}^{N}|a_k| \leq cN^{1/2}q^{s/2}\exp(\delta^2 q^{-2})\|g\|_{2,\mathbb{R}^s} \leq cN^{1/2}q^{s/2}\exp(\delta^2 q^{-2})\sum_{k=1}^{N}|a_k|. \tag{3.9}$$

Proof. Let the matrix $\mathcal{A}$ be defined as in (3.1). Since $\mathcal{A}$ is positive definite, its lowest eigenvalue satisfies

$$\frac{\sum_{j,\ell} a_\ell a_j \mathcal{A}_{\ell,j}}{\sum_{k=1}^{N}|a_k|^2} \geq \lambda_{\text{min}}.$$

In view of (3.2), the estimate (3.6) now implies (3.8). The first inequality in (3.9) is obtained by using the Schwarz inequality and (3.8). To prove the second inequality in (3.9), we observe that $\|\exp(-\|\cdot-\mathbf{y}\|^2)\|_{2,\mathbb{R}^s} = \pi^{s/2}$ for each $\mathbf{y}\in\mathbb{R}^s$. The desired inequality now follows from the triangle inequality. □

The following corollary of Proposition 3.1 will be used the proof of Theorem 2.1.

Corollary 3.1 *Let $N \geq 1$ be an integer, $\mathbf{x}_1, \cdots, \mathbf{x}_N$ be points in $\mathbb{R}^s$, $m \geq 1$, $g = \sum_{k=1}^{N} a_k \exp(-\|\cdot-\mathbf{x}_k\|^2) \in \mathbb{G}_{\infty,\infty,m,s}$, and $S \subseteq \{1, \cdots, N\}$. Then there exist positive constants c and γ dependent only on s, and in particular, independent of S, such that*

$$\|\sum_{k\in S} a_k\exp(-\|\cdot-\mathbf{x}_k\|^2)\|_{2,\mathbb{R}^s} \leq c|S|^{1/2}\exp(\gamma^2 m^2)\|g\|_{2,\mathbb{R}^s}. \tag{3.10}$$

Proof. We use the second inequality in (3.9), followed by Schwarz inequality, and (3.8) to obtain

$$\begin{aligned}\|\sum_{k\in S} a_k\exp(-\|\cdot-\mathbf{x}_k\|^2)\|_{2,\mathbb{R}^s} &\leq c\sum_{k\in S}|a_k| \leq c|S|^{1/2}\left\{\sum_{k\in S}|a_k|^2\right\}^{1/2} \\ &\leq c|S|^{1/2}\left\{\sum_{k=1}^{N}|a_k|^2\right\}^{1/2} \leq c|S|^{1/2}\exp(\gamma^2 m^2)\|g\|_{2,\mathbb{R}^s}.\end{aligned}$$

□

Finally, we prove a technical inequality regarding a subnetwork having centers outside of a cube.

Proposition 3.2 *Let* $g := \sum_{k=1}^{N_m} a_k \exp(-\|\cdot - \mathbf{x}_k\|^2) \in \mathbb{G}_{\infty,\infty,m,s}$, *and* $a, b > 0$. *There exists a constant* A *depending only on* a, b, *and* s *(but not on* g*) with the following property: Let* $L := \{k \ : \ \|\mathbf{x}_k\| \ge Am\}$, *and* $h := \sum_{k \in L} a_k \exp(-\|\cdot - \mathbf{x}_k\|^2)$. *Then*

$$\|h\|_{2,[-am,am]^s} \le c\exp(-b^2m^2)\|g\|_{2,\mathbb{R}^s}. \tag{3.11}$$

Proof. We may assume that $\|g\|_{2,\mathbb{R}^s} = 1$. The estimate (3.8) implies that

$$\sum_{k=1}^{N_m} |a_k|^2 \le c\exp(2\alpha^2m^2), \tag{3.12}$$

where we may assume that $\alpha \ge \sqrt{s}a$. Let $L_k := \{j \ : \ 2^k\alpha m \le \|\mathbf{x}_j\| < 2^{k+1}\alpha m\}$. Then $|L_k| \le c(2^km^2)^s$. For $j \in L_k$ and $\mathbf{x} \in [-am, am]^s$, we have

$$\|\mathbf{x}_j - \mathbf{x}\| \ge 2^k\alpha m - \sqrt{s}am \ge (2^k - 1)\alpha m.$$

Therefore, for $k \ge 2$,

$$\begin{aligned}
\|\sum_{j\in L_k} a_j \exp(-\|\cdot - \mathbf{x}_j\|^2)\|_{2,[-am,am]^s} &\le (2am)^{s/2}\exp(-(2^k-1)^2\alpha^2m^2)\sum_{j\in L_k}|a_j| \\
&\le c(2^km^3)^{s/2}\exp(-(2^k-1)^2\alpha^2m^2)\left\{\sum_{j\in L_k}|a_j|^2\right\}^{1/2} \\
&\le c(2^km^3)^{s/2}\exp(-(2^k-1)^2\alpha^2m^2)\left\{\sum_{j=1}^{N_m}|a_j|^2\right\}^{1/2} \\
&\le c(2^km^3)^{s/2}\exp(-((2^k-1)^2-1)\alpha^2m^2) \le c(2^km^3)^{s/2}\exp(-(2^k\alpha m)^2/2)
\end{aligned} \tag{3.13}$$

We let K be an integer greater than $2 + \log_2(b/\alpha)$, and $L = \cup_{k=K}^{\infty} L_k$. Then

$$\|h\|_{2,[-am,am]^s} \le \sum_{k=K}^{\infty} \|\sum_{j\in L_k} a_j \exp(-\|\cdot - \mathbf{x}_j\|^2)\|_{2,[-am,am]^s} \le cm^{3s/2}\exp(-2b^2m^2).$$

This completes the proof. □

3.3 Coefficient inequalities for $L^p(\mathbb{R}^s)$

In this subsection, we prove the analogues of Propositions 3.1 and 3.2 in the case of the L^p norms. The conclusions of the propositions are qualitatively the same; the proofs are different, and sometimes require the additional condition (2.12) on the number of neurons.

Proposition 3.3 *Let* $N \ge 1$ *be an integer,* $\mathbf{x}_1, \cdots, \mathbf{x}_N$ *be any points in* $\mathbb{R}^s$, g *be as defined in (3.7),* q *be the minimal separation as defined in (3.4),* δ *be as in (3.5). Then for* $1 \le p \le \infty$,

$$\sum_{k=1}^{N}|a_k| \le cN^2q^s\exp(2\delta^2q^{-2})\|g\|_{p,\mathbb{R}^s} \le c_1N^2q^s\exp(2\delta^2q^{-2})\sum_{k=1}^{N}|a_k|. \tag{3.14}$$

The main difficulty in the proof of this proposition is the lack of an analogue of an explicit expression for $\|g\|_{p,\mathbb{R}^s}$ and Lemma 3.2. Our remedy is to find an orthonormal basis for the span of $\{\exp(-\|\cdot - \mathbf{x}_k\|^2)\}$, and use Lemma 3.2 to estimate the coefficients of the networks involved.

Lemma 3.3 *Let $N \ge 1$ be an integer, $\mathbf{x}_1, \cdots, \mathbf{x}_N$ be any points in $\mathbb{R}^s$, and V be the span of the functions $\exp(-\|\cdot - \mathbf{x}_j\|^2)$, $1 \le j \le N$. Let $\mathcal{A}$ be the matrix defined as in (3.1), and an upper triangular matrix R be found so that $\mathcal{A}^{-1} = R^T R$. Let*

$$p_\ell(\mathbf{x}) := \sum_{k=1}^{N} R_{\ell,k} \exp(-\|\mathbf{x} - \mathbf{x}_k\|^2), \qquad \ell = 1, \cdots, N. \tag{3.15}$$

Then $\{p_\ell\}_{1 \le \ell \le N}$ is an orthonormal basis for V. Further, with q, δ defined as in (3.4), (3.5) respectively, we have

$$\sum_{1 \le \ell, k \le N} R_{\ell,k}^2 \le cNq^s \exp(2\delta^2 q^{-2}). \tag{3.16}$$

Proof. First, we observe that $\mathcal{A}$ being a positive definite matrix, the upper triangular matrix R indeed exists as stated. Further, in view of (3.3), we verify that

$$\begin{aligned}
\int_{\mathbb{R}^s} p_\ell(\mathbf{x}) p_j(\mathbf{x}) d\mathbf{x} &= \sum_{1 \le k, \nu \le N} R_{\ell,k} R_{j,\nu} \mathcal{A}_{k,\nu} = (R\mathcal{A}R^T)_{\ell,j} \\
&= (RR^{-1}(R^T)^{-1}R^T)_{\ell,j} \\
&= \begin{cases} 1, & \text{if } \ell = j, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

This proves that $\{p_\ell\}_{\ell=1}^N$ is an orthonormal basis for V. Further, if $\lambda_1, \cdots, \lambda_N$ are the eigenvalues of $\mathcal{A}$ in decreasing order, we have

$$\begin{aligned}
\sum_{1 \le \ell, k \le N} R_{\ell,k}^2 &= \sum_{k=1}^{N} \sum_{\ell=1}^{N} R_{k,\ell}^T R_{\ell,k} = \sum_{k=1}^{N} \mathcal{A}_{k,k}^{-1} \\
&= \sum_{k=1}^{N} \frac{1}{\lambda_k} \le \frac{N}{\lambda_N}.
\end{aligned}$$

The estimate (3.16) follows from (3.6). □

Proof of Proposition 3.3. The second inequality in (3.14) is clear:

$$\|g\|_{p,\mathbb{R}^s} \le \sum_{k=1}^{N} |a_k| \|\exp(-\|\cdot - \mathbf{x}_k\|^2)\|_{p,\mathbb{R}^s} = \|\exp(-\|\cdot\|^2)\|_{p,\mathbb{R}^s} \sum_{k=1}^{N} |a_k|.$$

Let $\{p_\ell\}$ be the orthonormal set constructed in Lemma 3.3, and

$$b_\ell = \int_{\mathbb{R}^s} g(\mathbf{x}) p_\ell(\mathbf{x}) d\mathbf{x}, \qquad \ell = 1, \cdots, N.$$

Let p' be chosen so that $1/p + 1/p' = 1$. In view of Lemma 3.3, we see that

$$\|p_\ell\|_{p',\mathbb{R}^s} \le \sum_{k=1}^{N} |R_{\ell,k}| \|\exp(-\|\cdot - \mathbf{x}_k\|^2)\|_{p',\mathbb{R}^s} \le c \sum_{k=1}^{N} |R_{\ell,k}|.$$

Hence, Hölder's inequality implies that

$$|b_\ell| = \left| \int_{\mathbb{R}^s} g(\mathbf{x}) p_\ell(\mathbf{x}) d\mathbf{x} \right| \le c\|g\|_{p,\mathbb{R}^s} \sum_{k=1}^{N} |R_{\ell,k}|. \tag{3.17}$$

Since

$$g(\mathbf{x}) = \sum_{\ell=1}^{N} b_\ell p_\ell(\mathbf{x}) = \sum_{k=1}^{N} \{\sum_{\ell=1}^{N} b_\ell R_{\ell,k}\} \exp(-\|\mathbf{x} - \mathbf{x}_k\|^2),$$

we see that $a_k = \sum_{\ell=1}^N b_\ell R_{\ell,k}$, $k = 1, \cdots, N$. Therefore, (3.17) implies that

$$\begin{aligned}\sum_{k=1}^N |a_k| &\leq \sum_{k=1}^N \sum_{\ell=1}^N |b_\ell||R_{\ell,k}| \\ &\leq \quad c\|g\|_{p,\mathbb{R}^s} \sum_{k=1}^N \sum_{\ell=1}^N \left\{|R_{\ell,k}| \sum_{\nu=1}^N |R_{\ell,\nu}|\right\} = c\|g\|_{p,\mathbb{R}^s} \sum_{\ell=1}^N \left\{\sum_{k=1}^N |R_{\ell,k}|\right\}^2 . \qquad (3.18)\end{aligned}$$

Using Schwarz inequality and (3.16), we deduce that

$$\sum_{\ell=1}^N \left\{\sum_{k=1}^N |R_{\ell,k}|\right\}^2 \leq N \sum_{\ell=1}^N \sum_{k=1}^N |R_{\ell,k}|^2 \leq cN^2 q^s \exp(2\delta^2 q^{-2}).$$

Along with (3.18), this implies the first inequality in (3.14). □

The following analogue of Corollary 3.1 follows immediately from Proposition 3.3, and will be used in the proof of Theorem 2.2.

Corollary 3.2 *Let $1 \leq p \leq \infty$, $A, m \geq 1$, $1 \leq N \leq \exp(Am^2)$ be an integer, $\mathbf{x}_1, \cdots, \mathbf{x}_N$ be points in $\mathbb{R}^s$, $g = \sum_{k=1}^N a_k \exp(-\|\cdot - \mathbf{x}_k\|^2) \in \mathbb{G}_{N,\infty,m,s}$, and $S \subseteq \{1, \cdots, N\}$. Then there exist positive constants c and γ dependent only on A and s, and in particular, independent of S such that*

$$\|\sum_{k\in S} a_k \exp(-\|\cdot - \mathbf{x}_k\|^2)\|_{p,\mathbb{R}^s} \leq c \exp(\gamma^2 m^2)\|g\|_{p,\mathbb{R}^s}. \qquad (3.19)$$

Proof. In view of (3.14),

$$\|\sum_{k\in S} a_k \exp(-\|\cdot - \mathbf{x}_k\|^2)\|_{p,\mathbb{R}^s} \leq c \sum_{k\in S} |a_k| \leq \sum_{k=1}^N |a_k| \leq cN^2 \exp(2\delta^2 m^2)\|g\|_{p,\mathbb{R}^s}.$$

The inequality (3.19) now follows from the assumption that $N \leq \exp(Am^2)$. □

The following proposition is the analogue of Proposition 3.2.

Proposition 3.4 *Let $g := \sum_{k=1}^{N_m} a_k \exp(-\|\cdot - \mathbf{x}_k\|^2) \in \mathbb{G}_{N_m,\infty,m,s}$, $1 \leq p \leq \infty$, and $a, b > 0$. There exists a constant A depending only on a, b, s, and p (but not on g) with the following property: Let $L := \{k \ : \ \|\mathbf{x}_k\| \geq Am\}$, and $h := \sum_{k\in L} a_k \exp(-\|\cdot - \mathbf{x}_k\|^2)$. If (2.12) holds then*

$$\|h\|_{p,[-am,am]^s} \leq c_3 \exp(-b^2 m^2)\|g\|_{p,\mathbb{R}^s}. \qquad (3.20)$$

Proof. We may assume that $\|g\|_{p,\mathbb{R}^s} = 1$. In view of (3.14) and (2.12), we conclude that there exists a positive constant, to be denoted (in this proof only) by α, such that

$$\sum_{k=1}^{N_m} |a_k| \leq c_1 \exp(\alpha^2 m^2). \qquad (3.21)$$

Let $A = \sqrt{s}a + \sqrt{\alpha^2 + 4b^2}$, and $\|\mathbf{w}\| \geq Am$. Then for $\mathbf{x} \in [-am, am]^s$,

$$\begin{aligned}\exp(-\|\mathbf{w} - \mathbf{x}\|^2) &\leq \exp\left(-(\|\mathbf{w}\| - \|\mathbf{x}\|)^2\right) \\ &\leq \exp\left(-(A - \sqrt{s}a)^2 m^2\right) = \exp\left(-(\alpha^2 + 4b^2)m^2\right).\end{aligned}$$

Applying this estimate to each of the basic Gaussians in h, we get from (3.21) that

$$\|h\|_{p,[-am,am]^s} \leq (2am)^{s/p} \exp\left(-(\alpha^2 + 4b^2)m^2\right) \sum_{k\in L} |a_k| \leq cm^{s/p} \exp(-4b^2 m^2).$$

□

3.4 Approximation of networks by weighted polynomials

In this section, we obtain estimates on the degree of approximation of Gaussian networks by weighted polynomials. The statement and proof for the following lemma are the same for all L^p norms.

Proposition 3.5 *Let $A, B > 0$, $1 \le p \le \infty$. There exists a constant C depending only on A, B, p, and s with the following property. For any $g \in \mathbb{G}_{\infty,Am,m,s}$, there exists a polynomial $P_g \in \Pi_{Cm^2,s}$ such that*

$$\|g - P_g \exp(-\|\cdot\|^2)\|_{p,\mathbb{R}^s} \le c_1 \exp(-B^2m^2)\|g\|_{p,\mathbb{R}^s}. \tag{3.22}$$

In the proof of this proposition, we will make extensive use of the classical Hermite polynomials $\{h_k\}$, defined formally by the generating function (cf. [15, formula (5.5.7)])

$$\exp(2yt - t^2) =: \pi^{1/4} \sum_{k=0}^{\infty} \frac{h_k(y)}{\sqrt{k!}} (\sqrt{2}t)^k. \tag{3.23}$$

The polynomial h_k is of precise degree k, and satisfies (cf. [15, formula 5.5.1])

$$\int_{\mathbb{R}} h_k(y)h_j(y)\exp(-y^2)dy = \begin{cases} 1, & \text{if } k = j,\ k, j = 0, 1, \cdots, \\ 0, & \text{otherwise.} \end{cases} \tag{3.24}$$

For a multi-integer $\mathbf{k}$, we define

$$h_{\mathbf{k}}(\mathbf{x}) = \prod_{j=1}^{s} h_{k_j}(x_j). \tag{3.25}$$

Writing $\mathbf{k}! := \prod_{j=1}^{s} k_j!$, and using standard multivariate notation, we have

$$\exp(-2\mathbf{x}\cdot\mathbf{w} - \|\mathbf{w}\|^2) = \pi^{s/4} \sum_{\mathbf{k}\ge 0} \frac{h_{\mathbf{k}}(\mathbf{x})}{\sqrt{\mathbf{k}!}} (\sqrt{2}\mathbf{w})^{\mathbf{k}}, \tag{3.26}$$

and for $\mathbf{k}, \mathbf{j} \ge 0$,

$$\int_{\mathbb{R}^s} h_{\mathbf{k}}(\mathbf{x})h_{\mathbf{j}}(\mathbf{x})\exp(-\|\mathbf{x}\|^2)d\mathbf{x} = \begin{cases} 1, & \text{if } \mathbf{k} = \mathbf{j}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.27}$$

The next lemma gives a useful estimate for the approximation of $\exp(-\|\cdot - \mathbf{w}\|^2)$ by weighted polynomials.

Lemma 3.4 *For integer $n \ge 1$, $\mathbf{w} \in \mathbb{R}^s$, let*

$$P_n(\mathbf{x}, \mathbf{w}) := \pi^{s/4} \sum_{0 \le |\mathbf{k}| \le n} \frac{h_{\mathbf{k}}(\mathbf{x})\exp(-\|\mathbf{x}\|^2)}{\sqrt{\mathbf{k}!}} \mathbf{w}^{\mathbf{k}}. \tag{3.28}$$

Then for any p, $1 \le p \le \infty$,

$$\|\exp(-\|\cdot - \mathbf{w}\|^2) - P_n(\cdot, \mathbf{w})\|_{p,\mathbb{R}^s} \le c_1 n^c \frac{(\sqrt{2s}\|\mathbf{w}\|)^{n+1}\exp(s\|\mathbf{w}\|^2)}{\sqrt{n!}}. \tag{3.29}$$

In order to prove this lemma, we first need a simple estimate.

Lemma 3.5 *For any $r \ge 0$ and $w \in [0, \infty)$, we have for $n \ge 4r - 3$,*

$$\sum_{k=n+1}^{\infty} \frac{k^r}{\sqrt{k!}} w^k \le c \frac{n^{r-1/2} w^{n+1}}{\sqrt{n!}} \exp(w^2/2), \tag{3.30}$$

where c is a positive constant depending only on r.

Proof. First, we observe that for any integer $M \ge 1$ and $u \ge 0$,

$$\sum_{k=M}^{\infty} \frac{u^k}{k!} = \frac{u^M}{M!}\sum_{k=0}^{\infty}\frac{u^k}{k!}\binom{k+M}{k}^{-1} \le \frac{u^M}{M!}e^u. \tag{3.31}$$

Hence, using Schwarz inequality and the fact (obtained from the Stirling approximation to the factorial) that

$$c\sqrt{k} \le \frac{2^{2k}k!^2}{(2k)!} \le c_1\sqrt{k},$$

we obtain for $w \in \mathbb{R}$ that

$$\begin{aligned}
\left(\sum_{k=M}^{\infty}\frac{w^{2k}}{\sqrt{(2k)!}}\right)^2 &\le \left\{\sum_{k=M}^{\infty}\frac{2^{2k}k!^2}{(2k)!}\frac{(w^2/2)^k}{k!}\right\}\left\{\sum_{k=M}^{\infty}\frac{(w^2/2)^k}{k!}\right\}\\
&\le c\frac{(w^2/2)}{\sqrt{M}}\frac{(w^2/2)^M}{M!}\exp(w^2/2)\sum_{k=M-1}^{\infty}\frac{(w^2/2)^k}{k!}\\
&\le c\frac{\sqrt{M}}{2^{2M}M!^2}(w^2)^{2M}\exp(w^2) \le c\frac{(w^2)^{2M}}{(2M)!}\exp(w^2) = c\frac{(w^{2M})^2}{(2M)!}\exp(w^2),
\end{aligned}$$

and

$$\left(\sum_{k=M}^{\infty}\frac{w^{2k+1}}{\sqrt{(2k+1)!}}\right)^2 \le \left(\frac{w}{\sqrt{2M+1}}\sum_{k=M}^{\infty}\frac{w^{2k}}{\sqrt{(2k)!}}\right)^2 \le c\frac{(w^{2M+1})^2}{(2M+1)!}\exp(w^2).$$

Thus, we conclude that

$$\sum_{k=N}^{\infty}\frac{w^k}{\sqrt{k!}} \le c\frac{w^N}{\sqrt{N!}}\exp(w^2/2). \tag{3.32}$$

Next, we assume without loss of generality that r is an integer. Since $k \ge 4r-2$ implies that

$$\frac{k!}{k^{2r}} \ge \left(1-\frac{2r-1}{k}\right)^{2r-1}(k-2r)! \ge (1/2)^{2r-1}(k-2r)!,$$

(3.32) leads to

$$\sum_{k=n+1}^{\infty}\frac{k^r}{\sqrt{k!}}w^k \le cw^{2r}\sum_{k=n+1-2r}^{\infty}\frac{w^k}{\sqrt{k!}} \le c\frac{w^{n+1}}{\sqrt{(n+1-2r)!}}\exp(w^2/2).$$

This implies (3.30). □

Proof of Lemma 3.4. From (3.26), we obtain that

$$\exp(-\|\mathbf{x}-\mathbf{w}\|^2) = \pi^{s/4}\sum_{\mathbf{k}\ge 0}\frac{h_{\mathbf{k}}(\mathbf{x})\exp(-\|\mathbf{x}\|^2)}{\sqrt{\mathbf{k}!}}(\sqrt{2}\mathbf{w})^{\mathbf{k}}. \tag{3.33}$$

Using the arithmetic-geometric inequality, we see that $|\mathbf{w}^{\mathbf{k}}| \le \|\mathbf{w}\|^{|\mathbf{k}|}$. Since $\|h_{\mathbf{k}}\exp(-\|\cdot\|^2/2)\|_{p,\mathbb{R}^s} \le c_1|\mathbf{k}|^c$ (cf. [8, Theorem 6.2.10]), we see that $\|h_{\mathbf{k}}\exp(-\|\cdot\|^2)\|_{p,\mathbb{R}^s} \le c_1|\mathbf{k}|^c$ as well. Therefore, (3.33) implies that

$$\begin{aligned}
\|\exp(-\|\cdot-\mathbf{w}\|^2) - P_n(\cdot,\mathbf{w})\|_{p,\mathbb{R}^s} &\le c_1\sum_{|\mathbf{k}|\ge n+1}\frac{|\mathbf{k}|^c(\sqrt{2}\|\mathbf{w}\|)^{|\mathbf{k}|}}{\sqrt{\mathbf{k}!}}\\
&= c_1\sum_{j=n+1}^{\infty}\frac{j^c(\sqrt{2}\|\mathbf{w}\|)^j}{\sqrt{j!}}\sum_{|\mathbf{k}|=j}\left(\frac{j!}{\mathbf{k}!}\right)^{1/2} \le c_1\sum_{j=n+1}^{\infty}\frac{j^c(\sqrt{2s}\|\mathbf{w}\|)^j}{\sqrt{j!}}.
\end{aligned}$$

The estimate (3.29) now follows from Lemma 3.5. □

Proof of Proposition 3.5. Let $g = \sum_{k=1}^{N_m} a_k \exp(-\|\cdot - \mathbf{x}_k\|^2) \in \mathbb{G}_{\infty,Am,m,s}$. Necessarily, $N_m \le cm^{2s}$. Therefore, (3.14) implies that

$$\sum_{k=1}^{N_m} |a_k| \le c_1 m^c \exp(c_2 m^2) \|g\|_{p,\mathbb{R}^s}. \tag{3.34}$$

Now, let $n = Cm^2$, where $C \ge 1$ is a constant to be chosen later. Using Lemma 3.4, we obtain $P_n(\cdot, \mathbf{x}_k)$ such that for $k = 1, \cdots, N_m$,

$$\begin{aligned} \|\exp(-\|\cdot - \mathbf{x}_k\|^2) - P_n(\cdot, \mathbf{x}_k)\|_{p,\mathbb{R}^s} &\le c_1 n^c \frac{(\sqrt{2s}\|\mathbf{x}_k\|)^{(n+1)} \exp(s\|\mathbf{x}_k\|^2)}{n!^{1/2}} \\ &\le c_1 m^c \frac{(\sqrt{2s}Am)^{(n+1)} \exp(sA^2m^2)}{n!^{1/2}}. \end{aligned}$$

Therefore, the function $h = \sum_{k=1}^{N_m} a_k P_n(\cdot, \mathbf{x}_k)$ satisfies

$$\begin{aligned} \|g - h\|_{p,\mathbb{R}^s} &\le c_1 \|g\|_{p,\mathbb{R}^s} m^c \frac{(c_2 m)^{(n+1)} \exp(c_3 m^2/2)}{n!^{1/2}} \\ &\le \|g\|_{p,\mathbb{R}^s} \frac{(c_2 m)^{Cm^2} \exp(c_3 m^2)}{(Cm^2)^{Cm^2/2+1/4} \exp(-Cm^2/2)} \\ &\le \|g\|_{p,\mathbb{R}^s} \exp\left(m^2(c + c_1 C - c_2 C \log C)\right). \end{aligned} \tag{3.35}$$

We may now choose C large enough to ensure that $c + c_1 C - c_2 C \log C \le -B^2$. □

3.5 Proof of Theorems 2.1 and 2.2

In order to prove these theorems, we need one more fact from the theory of weighted polynomial approximation.

Lemma 3.6 *Let $n \ge 1$ be an integer, $1 \le p \le \infty$, and $B > 0$. There exists a positive constant c_B depending only on B, p, and s such that for every $P \in \Pi_{n,s}$,*

$$\|P \exp(-\|\cdot\|^2)\|_{p,\mathbb{R}^s \setminus [-c_B\sqrt{n}, c_B\sqrt{n}]^s} \le \exp(-Bn) \|P \exp(-\|\cdot\|^2)\|_{p,\mathbb{R}^s}. \tag{3.36}$$

Proof. The lemma follows easily from the univariate infinite-finite-range inequalities [8, Theorem 6.1.6(c)]. □

We will prove Theorem 2.2 first.

Proof of Theorem 2.2. Suppose $g = \sum_{k=1}^{N_m} a_k \exp(-\|\cdot - \mathbf{x}_k\|^2)$ is chosen so that

$$\|f - g\|_{p,\mathbb{R}^s} \le cm^{-\beta} S_f, \tag{3.37}$$

where N_m satisfies (2.12). Necessarily, $\|g\|_{p,\mathbb{R}^s} \le S_f$.

Let $\varphi : \mathbb{R} \to [0,1]$ be an infinitely many times differentiable function such that $\varphi(x) = 1$ if $x \in [-1/2, 1/2]$ and $\varphi(x) = 0$ if $x \notin [-1, 1]$. Set $\phi(\mathbf{x}) := \prod_{k=1}^{s} \varphi(x_k/m)$, $\mathbf{x} \in \mathbb{R}^s$. Then ϕ is an infinitely many times continuously differentiable function on $\mathbb{R}^s$, $\phi(\mathbf{x}) = 1$ if $\mathbf{x} \in [-m/2, m/2]^s$, and $\phi(\mathbf{x}) = 0$ if $\mathbf{x} \notin [-m, m]^s$. It is not difficult to see that $\|(1 + \|\mathbf{x}\|^2)^{j/2} \mathcal{D}\phi\|_{\infty,\mathbb{R}^s} < c_j$ for any partial derivative $\mathcal{D}$ of order j, with a positive constant c_j independent of m. Therefore, the results in [9] imply that there exists a polynomial $P \in \Pi_{m^2,s}$ such that

$$\|\phi - P \exp(-\|\cdot\|^2)\|_{\infty,\mathbb{R}^s} \le cm^{-\beta}. \tag{3.38}$$

We will write $\mathcal{P} := P \exp(-\|\cdot\|^2)$. Using Lemma 3.6, we find $a > 1$ such that

$$\|\mathcal{P}\|_{\infty,\mathbb{R}^s \setminus [-am, am]^s} \le c_1 \exp(-3\gamma^2 m^2), \tag{3.39}$$

where γ is the constant appearing in Corollary 3.2.

Next, we use Proposition 3.4 with $\sqrt{3}\gamma$ in place of b to obtain a number A, the set L and the subnetwork h of g such that

$$\|h\|_{p,[-am,am]^s} \le c_1 \exp(-3\gamma^2 m^2) S_f. \tag{3.40}$$

Clearly, the network $g-h$ contains at most cm^{2s} neurons. In view of (3.19), we observe that

$$\|g-h\|_{p,\mathbb{R}^s} \le c\exp(2\gamma^2 m^2) S_f. \tag{3.41}$$

Since $a>1$ and $\phi(\mathbf{x})=0$ outside of $[-m,m]^s$, we see from (3.40) that

$$\|\phi h\|_{p,\mathbb{R}^s} \le c\exp(-3\gamma^2 m^2) S_f. \tag{3.42}$$

From (3.40) and the fact that $\|\mathcal{P}\|_{\infty,\mathbb{R}^s} \le c$, we obtain that

$$\|\mathcal{P}h\|_{p,[-am,am]^s} \le c\exp(-3\gamma^2 m^2) S_f. \tag{3.43}$$

In view of (3.39) and the fact that $\|g\|_{p,\mathbb{R}^s} \le S_f$, we see that

$$\|\mathcal{P}g\|_{p,\mathbb{R}^s\setminus[-am,am]^s} \le c\exp(-3\gamma^2 m^2) S_f.$$

Similarly, (3.39) and (3.41) imply that

$$\|(g-h)\mathcal{P}\|_{p,\mathbb{R}^s\setminus[-am,am]^s} \le c\exp(-\gamma^2 m^2) S_f.$$

Consequently,

$$\|\mathcal{P}h\|_{p,\mathbb{R}^s\setminus[-am,am]^s} \le c\exp(-\gamma^2 m^2) S_f.$$

Together with (3.43), this implies that

$$\|\mathcal{P}h\|_{p,\mathbb{R}^s} \le c\exp(-\gamma^2 m^2) S_f.$$

Therefore, (3.42) implies that

$$\|(\phi-\mathcal{P})h\|_{p,\mathbb{R}^s} \le c\exp(-\gamma^2 m^2) S_f.$$

Since (cf. (3.38))

$$\|(\phi-\mathcal{P})g\|_{p,\mathbb{R}^s} \le cm^{-\beta} S_f,$$

we conclude that

$$\|(\phi-\mathcal{P})(g-h)\|_{p,\mathbb{R}^s} \le cm^{-\beta} S_f. \tag{3.44}$$

Since $g-h \in \mathbb{G}_{cm^{2s},Am,m,s}$, we may use Proposition 3.5 to obtain a polynomial $Q \in \Pi_{cm^2,s}$, such that with $\mathcal{Q} = Q\exp(-\|\cdot\|^2)$,

$$\|g-h-\mathcal{Q}\|_{p,\mathbb{R}^s} \le c_1 \exp(-\gamma^2 m^2) S_f. \tag{3.45}$$

Since

$$\|\phi g - \mathcal{P}\mathcal{Q}\|_{p,\mathbb{R}^s} \le \|\phi h\|_{p,\mathbb{R}^s} + \|(\phi-\mathcal{P})(g-h)\|_{p,\mathbb{R}^s} + \|(g-h-\mathcal{Q})\mathcal{P}\|_{p,\mathbb{R}^s},$$

the estimates (3.42), (3.44), (3.45) lead to

$$\|\phi g - \mathcal{P}\mathcal{Q}\|_{p,\mathbb{R}^s} \le cm^{-\beta} S_f. \tag{3.46}$$

Now, using (2.10) and the facts that $0 \le \phi(\mathbf{x}) \le 1$ for $\mathbf{x} \in \mathbb{R}^s$ and $\phi(\mathbf{x}) = 1$ for $\mathbf{x} \in [-m/2, m/2]^s$, we obtain

$$\begin{aligned} \|f - \mathcal{P}\mathcal{Q}\|_{p,\mathbb{R}^s} &\le \|f - \phi f\|_{p,\mathbb{R}^s} + \|(f-g)\phi\|_{p,\mathbb{R}^s} + \|\phi g - \mathcal{P}\mathcal{Q}\|_{p,\mathbb{R}^s} \\ &\le cm^{-\min(\beta,\beta_1)} S_f. \end{aligned}$$

Since $\mathcal{P}\mathcal{Q}$ has the form $R\exp(-2\|\cdot\|^2)$ with $R \in \Pi_{cm^2}$, we have proved that for every integer $\nu \ge 1$, there exists $R_\nu \in \Pi_{\nu,s}$ such that

$$\|f - R\exp(-2\|\cdot\|^2)\|_{p,\mathbb{R}^s} \le c\nu^{-\min(\beta,\beta_1)/2} S_f.$$

According to Corollary 3.1 in [9], this statement is equivalent to (2.13), which involves $\exp(-\|\cdot\|^2/2)$ rather than $\exp(-2\|\cdot\|^2)$. □

PROOF OF THEOREM 2.1. The proof of this theorem is verbatim the same as that of Theorem 2.2, except that we use Corollary 3.1 (and (3.10)) in place of Corollary 3.2 (respectively, (3.19)) and Proposition 3.2 in place of Proposition 3.4. We observe that these results in the case of L^2 norm do not require the condition (2.12). □

4 Proof of the direct theorem

In this section, we prove Theorem 2.3. As expected, an essential ingredient is to approximate weighted polynomials by Gaussian networks.

For $n = 1, 2, \cdots$, a discrete (signed) measure ν_n will be called a *quadrature measure* if ν_n is supported at finitely many points, $\int_{\mathbb{R}^s} \exp(\|\mathbf{x}\|^2)|d\nu_n(\mathbf{x})| \le c_1 n^c$, and the following quadrature formula holds.

$$\int_{\mathbb{R}^s} P(\mathbf{x})d\nu_n(\mathbf{x}) = \int_{\mathbb{R}^s} P(\mathbf{x})\exp(-\|\mathbf{x}\|^2)d\mathbf{x}, \qquad P \in \Pi_{3n,s}. \tag{4.1}$$

A well known example [15, Section 3.4] is the following. Let $\{x_{k,2n}\}$ be the zeros of the polynomial h_{2n}, and

$$\lambda_{k,2n} := \left\{\sum_{\ell=0}^{2n-1} h_\ell^2(x_{k,2n})\right\}^{-1}$$

be the corresponding Cotes numbers. For a multi-integer $\mathbf{k}$, we write $\mathbf{x}_{\mathbf{k},2n} := (x_{k_1,2n}, \cdots, x_{k_s,2n})$, and $\lambda_{\mathbf{k},2n} := \prod_{j=1}^s \lambda_{k_j,2n}$. Using the well known univariate Gauss quadrature formula, we can verify that the measure ν_n^* that associates the mass $\lambda_{\mathbf{k},2n}$ with each of the points $\mathbf{x}_{\mathbf{k},2n}$ satisfies (4.1). The fact that it satisfies the condition $\int_{\mathbb{R}^s} \exp(\|\mathbf{x}\|^2)|d\nu_n^*(\mathbf{x})| \le c_1 n^c$ follows from the corresponding univariate theorem [8, Theorem 8.2.7]. Thus, the measure ν_n^* defined in this way is a quadrature measure. Other examples can be constructed, based on "arbitrary" nodes, by imitating the arguments in [10]. In the following discussion in this section, one may assume that the quadrature measure is given by ν_n^*.

The first step in our estimate is the following lemma.

Lemma 4.1 *Let $n \ge 1$ be an integer, $0 < w < 1$, $1 \le p \le \infty$. Let ν_n be a quadrature measure, and*

$$\begin{aligned}\mathcal{T}_{\mathbf{m},n,w}(\mathbf{x}) &:= w^{-|\mathbf{m}|}(\pi(1-w^2))^{-s/2}\int_{\mathbb{R}^s}\exp\left(\frac{w\|\mathbf{y}\|^2}{1+w^2}\right)\\ &\times \quad h_{\mathbf{m}}(\mathbf{y})\exp\left(-\frac{1+w^2}{2(1-w^2)}\left\|\mathbf{x}-\frac{2w\mathbf{y}}{1+w^2}\right\|^2\right)d\nu_n(\mathbf{y}).\end{aligned} \tag{4.2}$$

Then there exist positive constants c_1, c depending only on s and w such that

$$\|h_{\mathbf{m}}\exp(-\|\cdot\|^2/2) - \mathcal{T}_{\mathbf{m},n,w}\|_{p,\mathbb{R}^s} \le c_1 n^c w^n, \qquad 0 \le \mathbf{m} \le n. \tag{4.3}$$

We observe that with the choice $w = 1/\sqrt{3}$, $n = m^2$, and $\nu_n = \nu_n^*$, $\mathcal{T}_{\mathbf{m},n,1/\sqrt{3}} \in \mathbb{G}_{cm^{2s},c_1m,c_2m,s}$.

Proof. The proof is based on the Mehler's formula [15, Problem 24, p. 380]

$$\sum_{k=0}^\infty h_k(x)h_k(y)w^k = (\pi(1-w^2))^{-1/2}\exp\left(\frac{2xyw - x^2w^2 - y^2w^2}{1-w^2}\right).$$

Multiplying this identity several times yields

$$\sum_{\mathbf{k}\ge 0} h_{\mathbf{k}}(\mathbf{x})h_{\mathbf{k}}(\mathbf{y})w^{|\mathbf{k}|} = (\pi(1-w^2))^{-s/2}\exp\left(\frac{2\mathbf{x}\cdot\mathbf{y}w - \|\mathbf{x}\|^2w^2 - \|\mathbf{y}\|^2w^2}{1-w^2}\right). \tag{4.4}$$

Since $0 < w < 1$ and $|h_{\mathbf{k}}(\mathbf{y})| \le c\exp(\|\mathbf{y}\|^2/2)$ for all $\mathbf{k}$ (cf. [8, Theorem 6.2.10]), this series converges uniformly and absolutely on compact subsets of $\mathbb{R}^s \times \mathbb{R}^s$. We observe that

$$\frac{2\mathbf{x}\cdot\mathbf{y}w - \|\mathbf{x}\|^2w^2 - \|\mathbf{y}\|^2w^2}{1-w^2} - \frac{\|\mathbf{x}\|^2}{2} = \frac{w^2\|\mathbf{y}\|^2}{1+w^2} - \frac{1+w^2}{2(1-w^2)}\left\|\mathbf{x} - \frac{2w\mathbf{y}}{1+w^2}\right\|^2.$$

Hence, (4.4) and the definition (4.2) imply that for $\mathbf{m} \ge 0$,

$$\sum_{\mathbf{k}\ge 0} h_{\mathbf{k}}(\mathbf{x})\exp(-\|\mathbf{x}\|^2/2)w^{|\mathbf{k}|}\int_{\mathbb{R}^s} h_{\mathbf{k}}(\mathbf{y})h_{\mathbf{m}}(\mathbf{y})d\nu_n(\mathbf{y}) = w^{|\mathbf{m}|}\mathcal{T}_{\mathbf{m},n,w}(\mathbf{x}).$$

Since the quadrature formula (4.1) is valid for polynomials of degree at most $3n$, we conclude that for $0 \le \mathbf{m} \le n$,

$$\begin{aligned}
&\mathcal{T}_{\mathbf{m},n,w}(\mathbf{x}) - h_{\mathbf{m}}(\mathbf{x})\exp(-\|\mathbf{x}\|^2/2) \\
&= w^{-|\mathbf{m}|} \sum_{\mathbf{k}+\mathbf{m}\not\le 3n} h_{\mathbf{k}}(\mathbf{x})\exp(-\|\mathbf{x}\|^2/2) w^{|\mathbf{k}|} \int_{\mathbb{R}^s} h_{\mathbf{k}}(\mathbf{y}) h_{\mathbf{m}}(\mathbf{y}) d\nu_n(\mathbf{y}). \qquad (4.5)
\end{aligned}$$

We recall again that $|h_{\mathbf{k}}(\mathbf{y})| \le c\exp(\|\mathbf{y}\|^2/2)$ and the condition that

$$\int_{\mathbb{R}^s} \exp(\|\mathbf{y}\|^2)|d\nu_n(\mathbf{y})| \le c_1 n^c.$$

Further, $\|h_{\mathbf{k}}\exp(-\|\cdot\|^2/2)\|_{p,\mathbb{R}^s} \le c_1|\mathbf{k}|^c$. Hence, (4.5) leads to

$$\|h_{\mathbf{m}}\exp(-\|\cdot\|^2/2) - \mathcal{T}_{\mathbf{m},n,w}\|_{p,\mathbb{R}^s} \le c_1 n^c w^{-|\mathbf{m}|} \sum_{|\mathbf{k}|\ge 2n} |\mathbf{k}|^c w^{|\mathbf{k}|} \le c_1 n^c w^{2n-|\mathbf{m}|} \le c_1 n^c w^n.$$

□

The following proposition shows the construction of a Gaussian network to approximate a weighted polynomial.

Proposition 4.1 *Let* $P = \sum_{0\le\mathbf{k}\le n} b_{\mathbf{k}} h_{\mathbf{k}} \in \Pi_{n,s}$, $m = \sqrt{n}$, *and* $g_P := \sum_{0\le\mathbf{k}\le n} b_{\mathbf{k}} \mathcal{T}_{\mathbf{k},n,1/\sqrt{3}}$, *where the networks* $\mathcal{T}_{\mathbf{k},n,1/\sqrt{3}}$ *are constructed as in Lemma 4.1. Then for* $1 \le p \le \infty$,

$$\|P\exp(-\|\cdot\|^2/2) - g_P\|_{p,\mathbb{R}^s} \le c\exp(-c_2 m^2)\|P\exp(-\|\cdot\|^2/2)\|_{p,\mathbb{R}^s}. \qquad (4.6)$$

PROOF. Since ([8, Theorem 6.2.10])

$$\|h_k\exp(-(\cdot)^2/2)\|_{q,\mathbb{R}} \le c_1(k+1)^c, \qquad k = 0, 1, \cdots, \quad 1 \le q \le \infty,$$

we get for $0 \le \mathbf{k} \le n$,

$$|b_{\mathbf{k}}| = \left|\int_{\mathbb{R}^s} P(\mathbf{x}) h_{\mathbf{k}}(\mathbf{x})\exp(-\|\mathbf{x}\|^2) d\mathbf{x}\right| \le c_1 n^c \|P\exp(-\|\cdot\|^2/2)\|_{p,\mathbb{R}^s},$$

and hence,

$$\sum_{0\le\mathbf{k}\le n} |b_{\mathbf{k}}| \le c_1 n^c \|P\exp(-\|\cdot\|^2/2)\|_{p,\mathbb{R}^s} = c_1 m^{c_2} \|P\exp(-\|\cdot\|^2/2)\|_{p,\mathbb{R}^s}. \qquad (4.7)$$

From the estimates (4.7) and (4.3), we conclude that

$$\|P\exp(-\|\cdot\|^2/2) - g_P\|_{p,\mathbb{R}^s} \le c\exp(-c_2 m^2)\|P\exp(-\|\cdot\|^2/2)\|_{p,\mathbb{R}^s}.$$

□

Remark. We recall that the centers of the networks $\mathcal{T}_{\mathbf{k},n,1/\sqrt{3}}$ are independent of $\mathbf{k}$. Hence, if the quadrature measure in the construction of these networks is ν_n^*, then the network $g_P \in \mathbb{G}_{cm^{2s},c_1m,c_2m,s}$.

PROOF OF THEOREM 2.3. Suppose that (2.13) holds, and let $P = \sum_{0\le\mathbf{k}\le n} b_{\mathbf{k}} h_{\mathbf{k}} \in \Pi_{n,s}$ be found so that

$$\|f - P\exp(-\|\cdot\|^2/2)\|_{p,\mathbb{R}^s} \le n^{-\beta/2} S_f.$$

Necessarily, $\|P\exp(-\|\cdot\|^2/2)\|_{p,\mathbb{R}^s} \le S_f$. Constructing the network g_P as in Proposition 4.1 (where the quadrature measure is chosen to be ν_n^*), we conclude from (4.6) that $\|f - g_P\|_{p,\mathbb{R}^s} \le m^{-\beta} S_f$.

Next, we prove the converse. Let $g = \sum_{k=1}^{N_m} a_k \exp(-\|\cdot - \mathbf{x}_k\|^2) \in \mathbb{G}_{N_m,cm,m,s}$ be found so that

$$\|f - g\|_{p,\mathbb{R}^s} \le m^{-\beta} S_f.$$

Necessarily, $\|g\|_{p,\mathbb{R}^s} \le S_f$, and $N_m \le cm^{2s}$. Therefore, Proposition 3.5 yields a polynomial $P_g \in \Pi_{cm^2,s}$ such that

$$\|f - P_g \exp(-\|\cdot\|^2)\|_{p,\mathbb{R}^s} \le \|f-g\|_{p,\mathbb{R}^s} + \|g - P_g\exp(-\|\cdot\|^2)\|_{p,\mathbb{R}^s} \le c_1 m^{-\beta} S_f.$$

We have now proved that for each integer $\nu \ge 1$, there exists a polynomial $R_\nu \in \Pi_{\nu,s}$ such that

$$\|f - R_\nu \exp(-\|\cdot\|^2)\|_{p,\mathbb{R}^s} \le \nu^{-\beta/2} S_f.$$

According to Corollary 3.1 in [9], this statement is equivalent to (2.13), which involves $\exp(-\|\cdot\|^2/2)$ rather than $\exp(-\|\cdot\|^2)$. $\square$

5 A construction for Gaussian networks

In this section, we recall a construction of linear operators which yield a near best degree of approximation by weighted polynomials, and hence, construct Gaussian network approximations which are linear operators. In the case when the degree of weighted polynomial approximation to the target function decays polynomially with the degree, the degree of approximation given by our networks has the same order of magnitude, both in terms of the number of neurons, as well as the (square of the reciprocal of) the minimal separation among the centers.

In this section, let

$$\phi(t) = \begin{cases} 1, & \text{if } |t| \le 1/2, \\ 2(1-|t|), & \text{if } 1/2 < |t| \le 1, \\ 0, & \text{if } |t| > 1, \end{cases}$$

and

$$\Phi(\mathbf{x}) := \prod_{j=1}^{s} \phi(x_j), \qquad \mathbf{x} = (x_1, \cdots, x_s).$$

Let $1 \le p \le \infty$. For $f \in L^p(\mathbb{R}^s)$, and multiinteger $\mathbf{k} \ge 0$, we define the Hermite coefficient of f by

$$b_{\mathbf{k}}(f) := \int_{\mathbb{R}^s} f(\mathbf{t}) h_{\mathbf{k}}(\mathbf{t}) \exp(-\|\mathbf{t}\|^2/2) d\mathbf{t}. \tag{5.1}$$

For integer $n \ge 1$, the polynomial operator $V_n(f)$ is defined by

$$V_n(f, \mathbf{x}) := \sum_{\mathbf{k} \ge 0} \Phi(\mathbf{k}/n) b_{\mathbf{k}}(f) h_{\mathbf{k}}(\mathbf{x}) \exp(-\|\mathbf{x}\|^2/2). \tag{5.2}$$

We define the networks $\mathcal{T}_{\mathbf{k},m^2,1/\sqrt{3}}$ as in Lemma 4.1 using the measure ν_n^*, and the operator

$$\mathcal{G}_m(f, \mathbf{x}) := \sum_{\mathbf{k} \ge 0} \Phi(\mathbf{k}/m^2) b_{\mathbf{k}}(f) \mathcal{T}_{\mathbf{k},m^2,1/\sqrt{3}}(\mathbf{x}). \tag{5.3}$$

Then $\mathcal{G}_m(f) \in \mathbb{G}_{cm^{2s},c_1 m, c_2 m, s}$. We have proved in [9] that

$$\|f - V_n(f)\|_{p,\mathbb{R}^s} \le c\text{dist }(p; f, W\Pi_{n/2,s}), \qquad f \in L^p(\mathbb{R}^s),\ n = 1, 2, \cdots.$$

Consequently, Proposition 4.1 shows that if dist $(p; f, W\Pi_{n/2,s}) \le n^{-\beta/2} S_f$, then

$$\|f - \mathcal{G}_m(f)\|_{p,\mathbb{R}^s} \le cm^{-\beta} S_f.$$

References

[1] R. DeVore, R. Howard and C. A. Micchelli, *Optimal nonlinear approximation,* Manuscripta Mathematica, **63** (1989), 469-478.

[2] N. Dyn, *Interpolation and approximation by radial and related functions,* in "Approximation Theory,VI" (C. K. Chui, L. L. Schumaker, and J. D. Ward Eds.), Academic Press, 1989, pp. 211–234.

[3] G. Freud, *Direct and converse theorems in the theory of weighted polynomial approximation,* Math. Zeitschrift, **126** (1972), 123–134.

[4] F. Girosi and G. Anzellotti, *Rates of convergence for radial basis functions and neural networks,* in "Artificial Neural Networks for Speech and Vision", (R.J. Mammone, editor), Chapman Hall, London, 1993, pp. 97-113.

[5] V. Kurkova and M. Sanguineti, *Bounds on rates of variable basis and neural network approximation,* IEEE Trans. Inform. Theory, **47** September, 2001, 2659–2665.

[6] V. Kurkova and M. Sanguineti, *Comparison of worst case errors in linear and neural network approximation,* IEEE Trans. Inform. Theory, **48**, January, 2002, 264–275.

[7] G. G. Lorentz, "Approximation of functions", Holt, Rinehart and Winston, New York, 1966.

[8] H. N. Mhaskar, "Introduction to the theory of weighted polynomial approximation", World Scientific, Singapore, 1996.

[9] H. N. Mhaskar, *On the degree of approximation in multivariate weighted approximation,* in "Advanced Problems in Constructive Approximation" (Proceedings of the IDOMAT 2001 conference) (M.D. Buhmann and D.H. Mache Eds.), ISNM **142**, Birkhäuser, Basel, 2002, pp.129–141.

[10] H. N. Mhaskar, *On the representation of band limited functions using finitely many bits,* Journal of Complexity, **18** (2002), no. 2, 449–478.

[11] C. A. Micchelli, *Interpolation of scattered data: distance matrices and conditionally positive definite functions,* Constructive Approximation, **2** (1986), 11–22.

[12] F. Narcowich and J. D. Ward, *Norm estimates for the inverses of a genral class of scattered data radial function interpolation matrices,* Journal of Approximation Theory, **69** (1992), 84–109.

[13] J. Park and I. W. Sandberg, *Universal approximation using radial basis function networks,* Neural Computation, **3** (1991), 246-257.

[14] W. Rudin, "Real and complex analysis", McGraw Hill, New York, 1974.

[15] G. Szegö, "Orthogonal polynomials", Amer. Math. Soc. Colloq. Publ. **23**, Amer. Math. Soc., Providence, 1975.

A Appendix

For the convenience of the reader, we summarize here some facts regarding direct and converse theorems of weighted polynomial approximation theory. The degree of weighted polynomial approximation depends both on the decay of the target function near infinity as well as its smoothness. In [3], Freud introduced a modified modulus of smoothness ((A.3) below) that measures both. For a function $f : \mathbb{R} \to \mathbb{R}$ and $h \in \mathbb{R}$, we define the forward differences of f recusively by

$$\Delta_h^1(f,x) := f(x+h) - f(x), \quad \Delta_h^r(f,x) := \Delta_h^{r-1}(\Delta_h^1(f,x)),\ r = 2,3,\cdots. \tag{A.1}$$

One defines $\Delta_h^0(f,x) := f(x)$. The decay of the function will be measured by the auxiliary function

$$Q_h(x) := \begin{cases} h(1+x^2)^{1/2}, & \text{if } |x| \le h^{-1}, \\ 1, & \text{otherwise.} \end{cases} \tag{A.2}$$

Freud introduced the following modification of the usual modulus of smoothness for $f \in L^p(\mathbb{R})$ $(1 \le p \le \infty)$ and $\delta > 0$.

$$\omega_r(p; f, \delta) := \sum_{j=0}^{r} \sup_{|h|\le\delta} \|Q_\delta^{r-j} \Delta_h^j f\|_{p,\mathbb{R}}, \qquad r = 1,2,\cdots. \tag{A.3}$$

The following *equivalence theorem* is a reformulation of the corresponding theorem by Freud.

Theorem A.1 *Let $\beta > 0$ and $r > \beta$ be an integer, $1 \le p \le \infty$, $f \in L^p(\mathbb{R})$ (respectively, $f \in C_0(\mathbb{R})$ if $p = \infty$). The following statements* (a), (b), *and* (c) *are equivalent.*
(a) *We have*

$$\text{dist}\,(p; f, W\Pi_{n,1}) \le cn^{-\beta/2} S_f, \qquad n = 1,2,\cdots. \tag{A.4}$$

(b) *We have*

$$\omega_r(p; f, \delta) \le c\delta^\beta S_f, \qquad \delta > 0. \tag{A.5}$$

(c) *If k and j are positive integers and $0 < \beta - j < k$ then f has a derivative of order j, $f^{(j)} \in L^p(\mathbb{R})$ ($C_0(\mathbb{R})$ if $p = \infty$). Denoting by $g(x) = \exp(-x^2/2)f(x)$, we have*

$$\omega_k(p; \exp((\cdot)^2/2)g^{(j)}, \delta) \le c\delta^{\beta-j} S_f, \qquad \delta > 0. \tag{A.6}$$

In [9], we have defined a multivariate analogue of the modulus of smoothness (A.3), and obtained the analogue of the above theorem. We did not formulate the analogue of part (c) there, because it is thought to be routine. We observe that if an estimate of the form (A.5) is valid for *some* $r > \beta$, then it is valid also for *all* $r > \beta$, because the equivalent statement (A.4) does not involve r.

While Theorem A.1 shows that the degree of weighted polynomial approximation of *individual functions* is equivalent to their smoothness as measured by the modified modulus of smoothness, the theory of nonlinear n-widths [1] can be used to show that weighted polynomials are, in the sense of order of magnitude, the optimal class of approximants for functions in the class $W_{p,\beta}$ defined by

$$W_{p,\beta} := \{f \in L^p(\mathbb{R}) \ :\ \sup_{\delta>0} \delta^{-\beta}\omega_r(p; f, \delta) \le 1\}, \tag{A.7}$$

where $C_0(\mathbb{R})$ is understood in place of $L^p(\mathbb{R})$ in the case when $p = \infty$.