NONLINEAR APPROXIMATION USING GAUSSIAN KERNELS

THOMAS HANGELBROEK AND AMOS RON

ABSTRACT. It is well-known that non-linear approximation has an advantage over linear schemes in the sense that it provides comparable approximation rates to those of the linear schemes, but to a larger class of approximands. This was established for spline approximations and for wavelet approximations, and more recently for homogeneous radial basis function (surface spline) approximations. However, no such results are known for the Gaussian function. The crux of the difficulty lies in the necessity to vary the tension parameter in the Gaussian function spatially according to local information about the approximand: error analysis of Gaussian approximation schemes with varying tension are, by and large, an elusive target for approximators.

We introduce and analyze in this paper a new algorithm for approximating functions using translates of Gaussian functions with varying tension parameters. Our scheme is sophisticated to a degree that it employs even locally Gaussians with varying tensions, and that it resolves local singularities in a non-local way. We show that our algorithm is suitably optimal in the sense that it provides approximation rates similar to other established nonlinear methodologies like spline and wavelet approximations. As expected and desired, the approximation rates can be as high as needed and are essentially saturated only by the smoothness of the approximand.

1. INTRODUCTION

1.1. Nonlinear Radial Basis Function Approximation.

In this article we consider N-term approximation by Gaussian networks, an approximation technique widely used in statistics and engineering. This is an example of nonlinear approximation since we select d-variate functions residing in

$$\mathbb{G}_N := \left\{ \sum_{j=1}^{N} A_j \ \exp\left(- \left| \frac{\cdot - c_j}{\sigma_j} \right|^2 \right) : A \in \mathbb{C}^N, \sigma \in (0,\infty)^N, c \in \mathbb{R}^{dN} \right\}$$

which (failing to be closed under addition) is not a linear space. This stands in contrast to the linear approximation problem, often studied in radial basis function (RBF) theory, where the *centers* $(c_j)_j$, are predetermined and approximants are chosen from a linear space

$$\operatorname*{span}_{1\le j\le N} \phi_j(\cdot - c_j) = \operatorname*{span}_{1\le j\le N} \exp\left(- \left| \frac{\cdot - c_j}{\sigma_j} \right|^2 \right)$$

2000 *Mathematics Subject Classification.* 42C40, 46B70, 26B35, 42B25 .

Key words and phrases. Machine Learning, Gaussians, Kernels, Radial Basis Functions, Nonlinear Approximation, Besov Space, Triebel-Lizorkin Space.

Thomas Hangelbroek is supported by an NSF Postdoctoral Fellowship.

Amos Ron is supported by the National Science Foundation under grants DMS-0602837 and DMS-0914986, and by the National Institute of General Medical Sciences under Grant NIH-1-R01-GM072000-01.

that depends on the set of centers.

Heuristically, the benefit of the nonlinear approach is that by placing centers strategically, one may overcome defects, like discontinuities, cusps or other local deficiencies in smoothness, of the target function f. Because such defects may be manifested in a variety of ways, over regions or on lower dimensional manifolds, and may occur at different scales, finding a precise strategy is not at all straightforward. In this article, we present a sophisticated method for placing centers in a way that is suitable for creating effective nonlinear approximants.

An important distinction between the nonlinear and linear problems is in how convergence is measured. In the linear setting, the main approximation parameter measures density of the centers, usually by means of the "fill distance" $h = \max_{x\in\Omega} \operatorname{dist}\bigl(x, (c_j)_j\bigr)$; the underlying approximation problem is to measure the rate of convergence as h shrinks. In high dimensions, the assumption that centers fill a (high dimensional) region Ω with a small fill distance is computationally impractical. In nonlinear approximation the rate of convergence is measured against the parameter N, the cardinality of the set of centers. This approach lends itself to more frugal approximation in high dimensions.

The approximation scheme we introduce selects $s_{f,N}$ from $\mathbb{G}_N$, and is shown to have convergence rate $\|f - s_{f,N}\|_p = \mathcal{O}(N^{-s/d})$ for target functions f having L_τ smoothness s, with $\frac{1}{\tau} = \frac{s}{d} + \frac{1}{p}$. Generally speaking, such nonlinear estimates are sharp in the sense that they are similar to known results for nonlinear wavelet approximation, and one cannot expect to achieve a similar rate $\operatorname{dist}_p(f, \mathbb{G}_N) = \mathcal{O}(N^{-s/d})$ by decreasing either τ or the underlying smoothness.

To provide a more robust space of approximants, we permit the tension (aka shape or dilation) parameters σ_j to respond to the nonuniform distribution of the centers. The question of how to tune a tension parameter is of active interest to the Learning Theory community, [10, 11], as well as the RBF community [7, 1], but in most theoretical works, the tension parameter is taken to be constant for all centers. Although the spatially varying tension parameter is a natural idea, and is used in practice [6, 9], it has heretofore not been considered seriously in an approximation theoretic sense. Although it may be tempting to use tight dilations when the centers are dense, essentially setting σ_j proportional to a local spacing of centers around c_j, the manner in which our scheme sets the tension is more complicated, but one that is ultimately justified by the error estimates we provide. In any case, we note that there is some empirical evidence [4, Section 3] that Gaussian approximation is unstable without adjusting the tension.

Nonlinear approximation with RBFs has not been investigated with the same intensity as other basic elements of approximation theory (splines, wavelets, etc.). Recently, DeVore and Ron, [2] have made a first foray into nonlinear RBF approximation using RBFs that are fundamental solutions of elementary, homogeneous, elliptic PDEs. Such RBFs, which include the "surface splines," allow simple but elegant approximation schemes that are not burdened by the requirement that the target function must reside in the native space. In addition, the homogeneity of these RBFs means that the N-term approximation spaces are, essentially, invariant under rescaling and, thus, there is no need to select dilations σ_j – this is done automatically. However, many prominent RBFs, including the Gaussians, do not fall into this category. For the kernels considered by DeVore and Ron, the approximation order is *saturated*, meaning that for this method there is an upper bound on the rate of convergence: by increasing smoothness beyond a saturation level k (determined by the order of the elliptic differential operator inverted by the kernel) there is no corresponding increase in the rate of decay of the error. This is not so with Gaussian kernels. Furthermore, the kernels used by DeVore and Ron are dependent on the operator they invert, and, hence, (subtly) dependent on the spatial dimension. This is a hindrance which the Gaussians also avoid.

1.2. **The Methodology.** To construct the N-term approximant s_N, we begin with a wavelet decomposition of the target function $f = \sum_I f_I \psi_I$. Based on the size of the wavelet coefficient and the smoothness norm of the target function, the fixed budget of N terms is distributed over the elements in the expansion – into individual budgets N_I (many of which are zero). Each wavelet ψ_I is then approximated by a linear combination s_I of Gaussians that uses at most N_I terms. The full N term approximant is then $s_{f,N} = \sum f_I s_I$. The main idea is that we have a scheme for nonlinear approximation associated with this family of wavelets that can be lifted to the Gaussians by means of approximating the individual members of the family. Matters are simplified when we assume the entire family to be generated from a few prototypes via dilation and translation: our collection of Gaussians are invariant under these operations! This reduces the problem of efficiently approximating all members of the wavelet family to the problem of approximating a few fixed wavelets by linear combinations of Gaussians.

The crucial issue is to approximate a basic function ψ using a linear combination of N shifted Gaussians. We view the number N as the portion we are willing to invest in approximating ψ out of our total budget of centers. It is essential to understand how to apportion the budget, and this can only be accomplished when we have good N-term error estimates. Thus, we are interested in understanding how to approximate globally using only finitely many centers. This is a very hard problem for the Gaussian. We completely resolve this problem for a function ψ that is band-limited, and in addition, has rapid decay:

$$\text{for every } k \text{ there is a constant } C_k \text{ such that } |\psi(x)| \le C_k(1+|x|)^{-k}.$$

The trick we employ is to create an approximant $\sum_{\alpha\in h\mathbb{Z}^d} a(\alpha,h)\phi(\cdot-\alpha)$ that converges rapidly (globally) to ψ in the L_∞ norm, with coefficients $a(\alpha,h)$ that are roughly the same size as $\psi(\alpha)$. Then we modify this approximation scheme by throwing away centers from a region where ψ is small. This is where the two assumptions on the wavelet – that ψ is bandlimited and that it is rapidly decaying – come into play. Bandlimiting means that the "full" approximation scheme (using centers $h\mathbb{Z}^d$) has coefficients $a(\alpha,h)$ that can be expressed as the convolution of ψ with a Schwartz function. Rapid decay allows us to attribute polynomial decay of arbitrary orders to the coefficients.

1.3. **Organization.** In the Section 2 of this article, we develop the basic linear approximation scheme at the heart of our approach. First considered is the operator $T_h^\sharp$, which generates the 'full' approximant, an infinite series of Gaussians having the grid $h\mathbb{Z}^d$ as the set of centers. Second we develop the operator $T_h^\flat$, which generates the 'truncated' approximant – a linear combination of roughly h^{-2d} Gaussians . At the end of Section 2 we generalize $T_h^\flat$ to treat scaled wavelets using a fixed budget of N centers. This is the role of the map T_N. Corollary 4 gives the error for wavelets at all dilation levels.

Section 3 treats nonlinear approximation in L_p for $1 \le p < \infty$. Results match those obtained for surface splines in [2]. This involves a sophisticated strategy for distributing centers, which is expressed in Section 3.1. The main result is Theorem 9 in Section 3.3.

Section 4 treats nonlinear approximation in L_∞, a case was not considered in [2]. For technical reasons, we consider approximation of functions from Besov spaces in this section. The main result in that section is Theorem 12.

1.4. **Notation and Background.** We denote the ball with center c and radius R by $B(c, R)$. The symbol $I \subset \mathbb{R}^d$ will represent a cube with corner at $c(I) \in \mathbb{R}^d$ and sidelength $\ell(I) > 0$: it is the set $c(I) + [0, \ell(I)]^d$. We denote the volume of a set Ω in $\mathbb{R}^d$ by $|\Omega|$.

The natural affine change of variables associated with a cube I is denoted with the subscript I: i.e., for a function $g : \mathbb{R}^d \to \mathbb{C}$,

$$g_I(x) := g\left(\frac{x - c(I)}{\ell(I)}\right).$$

The symbol C, often with a subscript, will always represent a constant. The subscript is used to indicate dependence on various parameters. The value of C may change, sometimes within the same line.

For Schwarz functions, the d-dimensional Fourier transform is given by the formula $\widehat{f}(\xi) = \int_{\mathbb{R}^d} f(x)e^{-i\langle \xi,x\rangle}\, \mathrm{d}x$, and its inverse is $f(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} \widehat{f}(\xi)e^{-i\langle x,\xi\rangle}\, \mathrm{d}\xi$. An important property of the Gaussian functions

$$\phi_\sigma : x \mapsto \exp\bigl[-|x/\sigma|^2\bigr], \tag{1}$$

is that they satisfy $\widehat{\phi_\sigma} = (\sigma\sqrt{\pi})^d \phi_{(2/\sigma)}$.

2. Shift-invariant Gaussian approximation of band-limited functions

2.1. **Approximation using infinitely many centers.** Let $B \subset \mathbb{R}^d$ be a fixed ball centered at the origin. We denote by

$$H_B \tag{2}$$

the space of all Schwartz functions whose Fourier transform is supported in B. Let ϕ be the d-dimensional Gaussian function, dilated by a fixed (arbitrary) dilation $\sigma > 0$ (cf. (1)). Given $h > 0$, consider the linear space

$$S_h := S_h(\phi) := \operatorname{span}\{\phi(\cdot - \alpha) : \ \alpha \in h\mathbb{Z}^d\},$$

closed in the topology, say, of uniform convergence on compact sets.

We consider in this section approximation schemes and approximation errors for functions in H_B from the space S_h. We adopt to this end the approximation schemes of [1], and show that in our setup these schemes provide superb approximations to the class H_B: the error decays exponentially fast as the spacing parameter h tends to 0!

Let us fix now $f \in H_B$, and $h > 0$. We denote by f_ϕ the function whose Fourier transform is $\widehat{f}/\widehat{\phi}$. We note that f_ϕ is in H_B, since $f_\phi = f * \eta_\Phi$ for a Schwartz function η_ϕ (that depends only on ϕ and B) and H_B is an ideal in the Schwartz space. We then approximate f by $h^d T_h^\sharp f$, with

$$T_h^\sharp f := \left(\frac{1}{2\pi}\right)^d \sum_{\alpha \in h\mathbb{Z}^d} f_\phi(\alpha)\phi(\cdot - \alpha). \tag{3}$$

Our main result in this subsection is the following:

Proposition 1. *Let $B = B(0, R)$ be the ball of radius R centered at the origin. The uniform error in approximating $f \in H_B$ by $h^d T_h^\sharp f$ as above satisfies, for $h < \pi/R$,*

$$\|f - h^d T_h^\sharp f\|_\infty \leq C\|\widehat{f}\|_{L_1} h^d e^{-\frac{c}{h^2}}.$$

The constants C and c depend on R and the dilation parameter σ used in the definition of ϕ, but are independent of f and h.

Proof. Using the fact that $\widehat{f_\phi} = \widehat{f}/\widehat{\phi}$, we write $h^d T_h^\sharp f$ as

$$\int_{\mathbb{R}^d} \widehat{f}(\theta) k_h(\theta, \cdot)\, \mathrm{d}\theta,$$

with

$$k_h(\theta, z) := (2\pi)^{-d} \frac{h^d}{\widehat{\phi}(\theta)} \sum_{\alpha \in h\mathbb{Z}^d} \phi(z - \alpha) e^{i\langle \theta, \alpha\rangle}.$$

Invoking the Poisson summation formula (which obviously is valid for the Gaussian function), we obtain that

$$k_h(\theta, z) = (2\pi)^{-d} \frac{e^{i\langle z, \theta\rangle}}{\widehat{\phi}(\theta)} \sum_{\beta \in 2\pi\mathbb{Z}^d/h} \widehat{\phi}(\theta + \beta) e^{i\langle z, \beta\rangle}.$$

When applying the above kernel to f, we are allowed to do the integration term-by-term, with the $(\beta = 0)$-term yielding the original function f. Therefore,

$$f(z) - h^d T_h^\sharp f(z) = \int_{\mathbb{R}^d} \widehat{f}(\theta) k'_h(\theta, z)\, \mathrm{d}\theta,$$

with

$$k'_h(\theta, z) := (2\pi)^{-d} \frac{e^{i\langle z, \theta\rangle}}{\widehat{\phi}(\theta)} \sum_{\beta \in 2\pi\mathbb{Z}^d/h \backslash \{0\}} \widehat{\phi}(\theta + \beta) e^{i\langle z, \beta\rangle}.$$

Note that the kernel is integrated only over $\theta \in B$, since $\operatorname{supp}(\widehat{f}) \subset B$ by assumption. Thus, we obtain that

$$\|f - h^d T_h^\sharp f\|_\infty \leq (2\pi)^{-d} \|\widehat{f}\|_1 K_h,$$

with

$$K_h := \sum_{\beta \in 2\pi\mathbb{Z}^d/h \backslash \{0\}} \|\widehat{\phi}(\cdot + \beta)/\widehat{\phi}\|_{L_\infty(B)}.$$

Let R denote the radius of B. If $2R < |\beta|$ then, for $\xi \in B$, $|\xi + \beta|^2 - |\xi|^2 \geq (|\beta| - 2|\xi|)^2$. Consequently

$$K_h \leq C_1 \widehat{\phi}(a),$$

for $a < \operatorname{dist}_2\big(2B, 2\pi\mathbb{Z}^d/h \backslash \{0\}\big) = 2(\pi/h - R)$. □

2.2. **Approximation using finitely many centers.** In this subsection, we modify the approximant of the previous subsection and use only a finite number of centers. This is a necessary step for us, since our budget of centers is finite. Our approximand is still a function $f \in H_B$.

Our setup is as follows. Given f and a mesh-scaling parameter h, we will approximate f by $h^d T_h^\flat f$, with

$$T_h^\flat f := (2\pi)^{-d} \sum_{\alpha \in h\mathbb{Z}^d \cap B_h} f_\phi(\alpha)\phi(\cdot - \alpha), \tag{4}$$

with f_ϕ and ϕ as in the previous subsection, and B_h is a ball of radius $1/h$. The crux here is the correspondence between the mesh size h, and the radius $1/h$ of the domain of the shifts we "preserve": $T_h^\flat f$ is obtained from $T_h^\sharp f$ by removing from the sum all shifts outside a ball of radius $1/h$. Note that the number of shifts $N := N(h)$ that are being used for a given h satisfies

$$N \sim h^{-2d}$$

with constants of equivalence depending on d only. At the end, we need to control the error in terms of the parameter N. For the time being, we still write the error in terms of the mesh size h.

Once the approximation operator uses the above truncated sum, one cannot expect the error to decay exponentially fast as in Proposition 1. However, the new error, measured in the uniform norm, still decays rapidly:[1]

Lemma 2. *Let $k > 0$, and $f \in H_B$. Then there exists $C_k > 0$ that does not depend on h such that for all small enough h*

$$\|f - h^d T_h^\flat f\|_\infty \le C_k h^k. \tag{5}$$

Proof. Thanks to Proposition 1, we only need to show that

$$h^d \|T_h^\sharp f - T_h^\flat f\|_\infty \le C h^k.$$

However, the norm $\|T_h^\flat f - T_h^\sharp f\|_\infty$ is bounded above by the sum

$$\sum_{|\alpha|>1/h, \alpha \in h\mathbb{Z}^d} |f_\phi(\alpha)|.$$

Since f_ϕ decays rapidly at ∞, the above sum is $O(h^k)$ for any fixed k, and our claim follows. □

The uniform error bound that we just obtained is not refined enough for our purposes. We will need better estimates for the error away from the origin, i.e., outside the ball B_h of radius $1/h$. Indeed, such estimates are valid, but require a different argument:

Lemma 3. *Let $k > 0$, and $f \in H_B$. Then there is a constant $C'_k > 0$ (depending on k, d and f but independent of h), so that the function $T_h^\flat f$ from Lemma 2 approximates f with pointwise error:*

$$|(f - h^d T_h^\flat f)(x)| \le C'_k h^k \big(1 + |x|\big)^{-k}. \tag{6}$$

---

[1]We could have made the dependence of C_k below on f more explicit. However, this is not needed for our subsequent applications.

Proof. If $|x| \le 2/h$, then

$$\big(1+|x|\big)^{-k} \ge (h/3)^k,$$

hence the requirement here follows from the inequality in Lemma 2 when k there is replaced by $2k$.

For the case $|x| \ge 2/h$, we may prove that

$$|(f - h^d T_h^\flat f)(x)| \le C'_k\big(1+|x|\big)^{-2k},$$

since

$$h^k\big(1+|x|\big)^{-k} \ge C|x/2|^{-2k}.$$

To this end, we estimate the difference

$$f(x) - h^d T_h^\flat f(x)$$

directly. First, f decays rapidly, by assumption, hence certainly satisfies the required estimate. As to $h^d T_h^\flat f$, we note that, since f_ϕ decays rapidly, the sum

$$h^d \sum_{\alpha\in h\mathbb{Z}^d} |f_\phi(\alpha)|$$

is bounded, and the bound can be made independent of h (the bound is, essentially, the L_1-norm of f_ϕ). Thus, we can bound $T_h^\flat f(x)$, up to an h-independent constant, by

$$\max\{\phi(x-\alpha) :\ |\alpha| \le 1/h\}.$$

Since $|x| \ge 2/h$, $|x-\alpha| \ge |x|/2$, hence

$$h^d T_h^\flat f(x) \le C'_k \phi(x/2).$$

Thus we are left to show that

$$\phi(x) \le |x|^{-2k}, \text{ for } |x| \ge 1/h,$$

for small enough h, which is clearly valid due to the exponential decay of ϕ at ∞. ☐

2.3. **Gaussian approximation of a wavelet system.** We now assume that we have in hand a finite collection $\Psi \subset H_B$, with H_B as in the previous section. Then, Lemma 3 holds for each $f := \psi \in \Psi$. Considering Ψ as the set of mother wavelets in a suitable wavelet system, we need also to develop suitable approximation schemes for shifted dilations of ψ, i.e., we need approximation schemes and error bounds for functions of the form

$$\psi\big((\cdot - c)/\ell\big), \quad \psi\in\Psi,\ c\in\mathbb{R}^d,\ \ell > 0.$$

However, such schemes are trivial: since we are allowed to use shifted-dilated versions of our original Gaussian ϕ, we may simply use the approximation

$$\psi\big((\cdot - c)/\ell\big) \approx (h^d T_h^\flat \psi)\big((\cdot - c)/\ell\big).$$

Note that $T_h^\flat \psi$ employs $N \sim h^{-2d}$ centers. Fixing N momentarily, we define a new map, T_N, that is defined on all dilated shifts of each $\psi\in\Psi$ by

$$(T_N\psi)\big((\cdot - c)/\ell\big)\big) := N^{-1/2}(T^\flat_{N^{-1/(2d)}}\psi)\big((\cdot - c)/\ell\big). \tag{7}$$

The error bounds of the previous section apply directly here. We just need to replace each occurrence of h by $N^{-1/(2d)}$. Thus, we obtain:

Theorem 4. *Let $\psi \in H_B$ be given and finite. Let $k > 0$, and let I be a cube. Then, there exists a constant C independent of N and I such that, for every N sufficiently large, and for every I as above,*

$$|(\psi_I - T_N\psi_I)(x)| \leq CN^{-k/d}\left(1 + \frac{|x - c(I)|}{\ell(I)}\right)^{-2k}.$$

3. Nonlinear Approximation in L_p, $1 \leq p < \infty$

In the previous section, we derived error estimates for the approximation of each member of a bandlimited smooth wavelet system by suitably chosen N shifted-dilated Gaussians. Armed with these error estimates, we finally tackle in this section our central problem: approximating a general function by finitely many shifted-dilated Gaussians. Our approach follows [2] and is similarly based on approximating the wavelets in the wavelet expansion of the actual approximand. To this end, we choose first any, say orthogonal, wavelet system whose mother wavelets are all bandlimited Schwartz functions. We define below MRA systems and wavelets in the exact way that fits our needs. Let us stress that the actual definitions of wavelet systems are far more flexible.

Definition 5 (Wavelets)**.** *In this article a univariate wavelet system is an orthonormal MRA wavelet system whose generators are bandlimited Schwartz functions: a scaling function η_0 and a (mother) wavelet η_1, both bandlimited Schwartz functions. See* [8, 3.2] *or* [5] *for a possible construction. Multivariate wavelet systems are tensor products of a univariate one, hence its wavelets are indexed by (I, e), an ordered pair consisting of a dyadic cube, I, and a gender $e \in \mathcal{E} \in \{0,1\}^d \setminus \{\mathbf{0}\}$, corresponding to one of the (non-origin) corners of the unit cube $[0,1]^d$:*

$$\psi_e(x) = \prod_{j=1}^{d} \eta_{e_j}(x_j), \quad \psi_{I,e} := (\psi_e)_I.$$

Let $\mathcal{D}$ be the collection of all dyadic cubes, viz., with I_0 the unit cube,

$$\mathcal{D} := \{2^j(k + I_0) : \ j \in \mathbb{Z}, \ k \in \mathbb{Z}^d\}.$$

We denote by $\mathcal{D}_j$ the subset of dyadic cubes with common edgelength 2^j.

The wavelet $\psi_{I,e}$ is an affine change of variable (as in Section 1.4) of the mother wavelet $\psi_e = \psi_{I_0,e}$, for some $e \in \mathcal{E}$. Since we use more than one mother wavelet (indeed, we use $\#\mathcal{E} = 2^d - 1$), we regard $\mathcal{D}$ and $\mathcal{D}_j$ as multisets and we suppress dependence on the gender e. Thus, the notation ψ_I stands for the I-version of *any* of the mother wavelets, and a summation over $\mathcal{D}$ or over one of its subsets, unless otherwise noted, is assumed to take place over $\mathcal{E}$ as well. This does not cause any confusion, since in this section our algorithms and their analysis do not pay attention to the details of the actual mother wavelet that is employed.

Our problem is then the following basic one. We are given a smooth function f (from some smoothness class, see below) and a budget of N centers. We are then allowed to approximate f by a total of N shifted-dilated Gaussians. We carry out this approximation by distributing the centers across the wavelet system: for each $I \in \mathcal{D}$, we allocate N_I centers as "the I-budget" and use these budgeted centers for approximating the term $f_I\psi_I$ in the wavelet expansion

$$f = \sum_{I \in \mathcal{D}} f_I \psi_I. \tag{8}$$

The individual error when approximating $f_I\psi_I$ by N_I Gaussians was the subject of the previous section. Thus, our analysis here will focus on the estimation of the cumulative error. But, first and foremost, we need to devise an algorithm for distributing the budget across the different wavelets. We refer to this algorithm as the *cost distribution.*

3.1. **Triebel-Lizorkin Cost Distribution.** It is convenient to associate each wavelet with **cost** $\mathrm{c}_I > 0$ that is not necessarily an integer, and then to determine N_I from the formula

$$N_I := \begin{cases} \lfloor \mathrm{c}_I \rfloor, & \lfloor \mathrm{c}_I \rfloor \geq N_0, \\ 0, & \text{otherwise}, \end{cases}$$

where N_0 is a some fixed integer, that depends on the wavelet system and on nothing else.

We now discuss the cost distribution c_I, which depends on several factors. In addition to the volume of the dyadic cube, $|I|$, it depends on the wavelet coefficient f_I, the smoothness norm of f (defined below), and an estimate of the size of a partial reconstruction of f. To this end, we make the following definitions:

Definition 6. *Given $s, q > 0$, we define the maximal function $M_{s,q}f$ as*

$$M_{s,q}f(x) := \left(\sum_{I\in\mathcal{D}} |I|^{-sq/d} |f_I|^q \chi_I(x) \right)^{1/q}. \tag{9}$$

For a dyadic interval I, we define a partial function by

$$M_{s,q,I}f(x) := \left(\sum_{I\subset I'\in\mathcal{D}} |I'|^{-sq/d} |f_{I'}|^q \chi_{I'}(x) \right)^{1/q}. \tag{10}$$

Given now $\tau, s, q > 0$, we define the Triebel-Lizorkin space $F^s_{\tau,q}$ via the finiteness of the following quasi-seminorm:

$$|f|_{F^s_{\tau,q}} := \|M_{s,q}f\|_\tau. \tag{11}$$

We note that for any interval I, the partial maximal function $M_{s,q,I}f$ is nonnegative and always $\leq M_{s,q}f$. Furthermore, it achieves its maximum on the interval I, where it is constant. Thus the number $m_{s,q,I} := M_{s,q,I}f(x)$, $x \in I$, is well-defined, and $m_{s,q,I} = \sup_{y\in\mathbb{R}^d} M_{s,q,I}f(y) \leq M_{s,q}f(x)$. In the definition below, s stands for the smoothness of the function we approximate, and p for the norm in which we measure the error.

Definition 7 (Cost Distribution)**.** *Let $s > 0$, and $p \geq 1$. Define τ, q by $1/\tau := 1/p + s/d$ and $1/q := 1 + s/d$. Let $f \in F^s_{\tau,q}$, with wavelet expansion (8). We choose then* **the cost of a dyadic cube** *$I \in \mathcal{D}$ as*

$$\mathrm{c}_I := |f|^{-\tau}_{F^s_{\tau,q}}\, m^{\tau-q}_{s,q,I}\, |f_I|^q\, |I|^q N. \tag{12}$$

Let us first verify that the sum of all the costs is our budget N:

$$\sum_{I\in\mathcal{D}} \mathrm{c}_I = \sum_{I\in\mathcal{D}} |f|^{-\tau}_{F^s_{\tau,q}}\, m^{\tau-q}_{s,q,I}\, |f_I|^q\, |I|^{1-qs/d} N.$$

Since $|I| = \int_{\mathbb{R}^d} \chi_I(x)\,\mathrm{d}x$, we can write the right hand side as an integral, namely $|f|_{F^s_{\tau,q}}^{-\tau} N \int_{\mathbb{R}^d} \sum_I m_{s,q,I}^{\tau-q} |f_I|^q\, |I|^{-qs/d} \chi_I(x)\,\mathrm{d}x$. Invoking the fact that, for $x \in I$, $m_{s,q,I} \le M_{s,q} f(x)$ (and that $\tau > q$), gives

$$\begin{aligned} \sum_I c_I &\le |f|_{F^s_{\tau,q}}^{-\tau} N \int_{\mathbb{R}^d} \sum_I \big(M_{s,q} f(x)\big)^{\tau-q} |f_I|^q\, |I|^{-qs/d} \chi_I(x)\,\mathrm{d}x \\ &\le |f|_{F^s_{\tau,q}}^{-\tau} N \int_{\mathbb{R}^d} \big(M_{s,q} f(x)\big)^{\tau}\,\mathrm{d}x = N. \end{aligned}$$

3.2. **Approximating the Wavelet Expansion.** Once a budget of N_I centers is allocated for the approximation of the term $f_I \psi_I$ in the wavelet expansion of f, we appeal to Theorem 4 to conclude that the term can be approximated by N_I Gaussians with error that is bounded (up to a constant that depends only on the wavelet system and on the parameter k) by $|f_I| R_I$, where

$$\begin{aligned} R_I(x) &:= C_{k,d} \min(1, N_I^{-k/d}) \left(1 + \frac{\operatorname{dist}(x,I)}{\ell(I)}\right)^{-2k} \\ &\le C'_{k,d} \min(1, \mathrm{c}_I^{-k/d}) \left(1 + \frac{\operatorname{dist}(x,I)}{\ell(I)}\right)^{-2k}. \end{aligned} \tag{13}$$

The following lemma, which is proved in the next subsection, simplifies the above error:

Lemma 8. *Let $1 \le p < \infty$, then*

$$\left\| \sum_{I \in \mathcal{D}} |f_I|\, R_I \right\|_p \le C_{k,d} \left\| \sum_{I \in \mathcal{D}} \min(1, \mathrm{c}_I^{-k/d}) f_I \chi_I \right\|_p.$$

We are ready to state and prove our main result concerning the case $1 \le p < \infty$.

Theorem 9. *Given $s > 0$ and $1 \le p < \infty$, there is a constant $C_{p,s,d}$ so that for $f \in F^s_{\tau,q}$, with $1/\tau = 1/p + s/d$ and $1/q = 1 + s/d$, there is a linear combination of N Gaussians $s_f(x) := \sum_{j=1}^N A_j \exp\big[-\big(\frac{x-\xi_j}{\sigma_j}\big)^2\big]$ so that*

$$\|f - s_f\|_p \le C_{p,s,d} N^{-s/d} |f|_{F^s_{\tau,q}}.$$

Proof. Using the coefficients of the wavelet expansion (8), we can express s_f as

$$s_f := \sum f_I\, T_{N_I} \psi_I,$$

where each term, $[T_{N_I}\psi_I](x) = \sum_{j=1}^{N_I} a_{I,j} \exp\left[-\left(\frac{x - c_{I,j}}{\ell(I)}\right)^2\right]$, defined in (7), is composed of N_I Gaussians by the construction preceding Theorem 4 (note that the notation c_I stands for the I-cost, and is very different from the notation $c_{I,j}$ above). By the enumeration at the end of Section 3.1 ($\sum_{I\in\mathcal{D}} N_I \le N$), we know that no more than N Gaussians are used.

From Lemma 8 we have the error estimate

$$\|f - s_f\|_p \le C_{k,d} \left\| \sum_{I \in \mathcal{D}} \min(1, \mathrm{c}_I^{-k/d}) |f_I| \chi_I \right\|_p.$$

As long as k (which is arbitrary) is greater than s, we can estimate the error as the L_p norm of a series:

$$\|f - s_f\|_p \le C_{k,d} \left\| \sum_{I\in\mathcal{D}} E_I \right\|_p ,$$

where $E_I(x) := \mathrm{c}_I^{-s/d} |f_I| \chi_I(x)$. We now focus on estimating this series, pointwise.

By applying the definition of c_I, we obtain (after some elementary manipulation of exponents), $\mathrm{c}_I^{-s/d}|f_I| = |f|_{F^s_{\tau,q}}^{\tau s/d} m_{q,s,I}^{\tau/p-q} |f_I|^q |I|^{-qs/d} N^{-s/d}$. We recall that the I^{th} partial square-like function is constant on the cube I, where it equals $m_{q,s,I}$. This implies that $\chi_I(x) m_{q,s,I} = \chi_I(x) M_{q,s,I}(x)$, which shows that each term is

$$E_I(x) = N^{-s/d} |f|_{F^s_{\tau,q}}^{\tau s/d} M_{q,s,I}(x)^{\tau/p-q} |f_I|^q |I|^{-qs/d} \chi_I(x).$$

The series becomes much more manageable by making some simple substitutions. Writing the basic summand of the maximal function as $z_I := |f_I|^q |I|^{-qs/d}\chi_I(x)$, the partial sum of these basic summands, $Z_I := \sum_{I\subset I'} z_{I'}$, is observed to be the q^{th} power of the partial maximal function $Z_I = \big(M_{s,q,I}(x)\big)^q$, while the full sum of these, $Z := \sum_{I\in\mathcal{D}} z_I$, is simply the qth power of the (full) maximal function $Z = \big(M_{s,q}(x)\big)^q$. It is a simple observation that the full series under consideration now has the compact form

$$\sum_{I\in\mathcal{D}} E_I(x) = N^{-s/d} |f|_{F^s_{\tau,q}}^{\tau s/d} \sum_{I\in\mathcal{D}} z_I Z_I^{\frac{\tau}{pq}-1}.$$

It follows from the inequality $\sum_{I\in\mathcal{D}} z_I Z_I^{\epsilon-1} \le C_\epsilon Z^\epsilon$, valid for nonnegative sequences $(z_I)_{I\in\mathcal{D}}$ and $0<\epsilon$ with constant $C_\epsilon < \infty$ (this is [2][Lemma 6.3]), that

$$\sum_{I\in\mathcal{D}} E_I(x) \le C_{\frac{\tau}{pq}} N^{-s/d} |f|_{F^s_{\tau,q}}^{\tau s/d} \left(\big(M_{s,q}(x)\big)^q\right)^{\frac{\tau}{pq}} = C_{p,s,d} N^{-s/d} |f|_{F^s_{\tau,q}}^{\tau s/d} \big(M_{s,q}(x)\big)^{\tau/p}.$$

Taking the L_p norm controls the error:

$$\begin{aligned} \left(\int_{\mathbb{R}^d} \left(\sum_{I\in\mathcal{D}} E_I(x)\right)^p \mathrm{d}x\right)^{1/p} &\le C_{p,s,d} N^{-s/d} |f|_{F^s_{\tau,q}}^{\tau s/d} \left(\int_{\mathbb{R}^d} \big(M_{s,q}(x)\big)^\tau \mathrm{d}x\right)^{1/p} \\ &= C_{p,s,d} N^{-s/d} |f|_{F^s_{\tau,q}}^{\tau s/d+\tau/p} = C_{p,s,d} N^{-s/d} |f|_{F^s_{\tau,q}} \end{aligned}$$

since $\tau s/d + \tau/p = 1$. □

3.3. **On Lemma 8.** The vector-valued maximal inequality of Fefferman and Stein, [3, Theorem 1], controls the $L_r(\ell_s)$ norm of the sequence of functions $(MF_j)_j$ by the $L_r(\ell_s)$ norm of $(F_j)_j$, provided $1 < r, s < \infty$ (the operator M is the usual Hardy–Littlewood maximal operator $MF(x) := \sup_{x\in[a,b]^d} \frac{1}{(b-a)^d} \int_{[a,b]^d} |F(y)|\,\mathrm{d}y$):

$$\left\| \left(\sum |MF_j(x)|^s\right)^{1/s} \right\|_r \le C_{r,s} \left\| \left(\sum |F_j(x)|^s\right)^{1/s} \right\|_r$$

In the lemma we make use of a minor generalization of this for the modified maximal operator M_τ, defined for $0<\tau<\infty$ by

$$M_\tau f(x) := \sup_{x\in[a,b]^d} \left(\frac{1}{(b-a)^d} \int_{[a,b]^d} |f(y)|^\tau \,\mathrm{d}y \right)^{\frac{1}{\tau}}.$$

It is not difficult to show that for $\tau < p, q < \infty$,

$$\left\| \left(\sum |M_\tau f_j(x)|^q \right)^{1/q} \right\|_p \le K \left\| \left(\sum |f_j(x)|^q \right)^{1/q} \right\|_p . \tag{14}$$

Indeed, this follows by a direct application of the Fefferman–Stein inequality with $s = \frac{q}{\tau}$, $r = \frac{p}{\tau}$ (both greater than one), $F_j = f_j^\tau$ and $K = C_{r,s}^{1/\tau}$, because the modified maximal operator is related to the Hardy-Littlewood maximal operator by $MF_j^s = M_\tau f_j^q$ and the r and p norms are related by $\|g\|_r = \|g^{1/\tau}\|_p^\tau$.

Proof of Lemma 8. From (13) follows $R_I \le C_{k,d} \min(1, \mathrm{c}_I^{-k/d}) \left(1 + \frac{\mathrm{dist}(x,I)}{\ell(I)}\right)^{-2k}$. Observe that there is a constant, C_d, depending only on d so that

$$\left(1 + \frac{\mathrm{dist}(x,I)}{\ell(I)}\right)^{-d} \le M(\chi_I)(x).$$

We can assume $k > d/2$ without loss. For τ between $d/2k$ and 1 we have

$$\left(1 + \frac{\mathrm{dist}(x,I)}{\ell(I)}\right)^{-2k} \le \left(1 + \frac{\mathrm{dist}(x,I)}{\ell(I)}\right)^{-d/\tau} \le C_d M_\tau(\chi_I)(x),$$

since $M_\tau(\chi_I) = M(\chi_I)^{1/\tau}$. It follows from the modified Fefferman-Stein inequality (14), that

$$\begin{aligned} \left\| \sum_{I \in \mathcal{D}} |f_{I,e}|\, R_I \right\|_p &\le C_{k,d} \left\| \sum_{I \in \mathcal{D}} |f_{I,e}| \, \min(1, \mathrm{c}_I^{-k/d}) M_\tau(\chi_I) \right\|_p \\ &\le C_{k,d} \left\| \min(1, \mathrm{c}_I^{-k/d}) |f_{I,e}| \chi_I \right\|_p . \end{aligned}$$

□

4. Nonlinear Approximation in L_∞

Although the basic strategy for nonlinear RBF approximation in L_∞ is, at heart, the same as in L_p, there are some complications that require us to give it a slightly different treatment. The fundamental difference is that the Hardy-Littlewood maximal inequality (and, hence, its vector valued analogue, the Fefferman-Stein inequality, used in the previous section) does not hold for $p = \infty$. For this reason, we choose to work with family of smoothness spaces that do not require us to explicitly work with maximal operators. Smoothness is measured using a Besov norm, and we use a Besov space based cost distribution to determine how to distribute the budget.[2]

Definition 10. *For $\tau = d/s \in (0, \infty)$ and $q \in (0, \infty)$, the Besov space $B^s_{\tau,q}$ is the space of L_τ functions for which the (quasi-)seminorm $|f|_{B^s_{\tau,q}}$ is finite, where*

$$|f|_{B^s_{\tau,q}} := \left\| k \mapsto \left(\sum_{I \in \mathcal{D}_k} |f_I|^\tau \right)^{1/\tau} \right\|_{\ell_q(\mathbb{Z})} .$$

Here, the coefficients (f_I) are as in (8).

[2]The Besov space approach is valid for the case $p < \infty$ that was analysed in the previous section, too. However, the Triebel-Lizorkin space $F^s_{\tau,q}$ is slightly larger than the Besov space of the same parameters.

Note that for $q \le \tau \le 1$, $f \in B^s_{\tau,q}$ implies that the wavelet coefficients f_I are absolutely summable. Since the wavelets ψ_I are uniformly bounded, this means that the wavelet expansion (8) is absolutely convergent for $s \ge d$ and $f \in B^s_{\tau,q}$ (meaning that the main issue for L_∞ approximation is resolved in this case). For $1 < \tau < \infty$ and $q \le 1$, we also have unconditional convergence of the wavelet expansion, since

$$\sum_{k\in\mathbb{Z}} \sum_{I\in\mathcal{D}_k} |f_I||\psi_I(x)| \le \sum_{k\in\mathbb{Z}} \left(\sum_{I\in\mathcal{D}_k} |f_I|^\tau\right)^{1/\tau} \left(\sum_{I\in\mathcal{D}_k} |\psi_I(x)|^{\tau'}\right)^{1/\tau'}.$$

Because $|\psi_I(x)|^{\tau'} \le C\big(1+|x-c(I)|/\ell(I)\big)^{-(d+1)}$, the second factor is bounded with a constant depending on d and totally independent of k and x. Thus, the right hand side is controlled by $|f|_{B^s_{\tau,q}}$. This is a reflection of the fact that $B^s_{\tau,q}$ is embedded in L_∞ for $\tau = d/s$ and $q \le \min(\tau, 1)$. Although L_∞ has no unconditional basis, the Besov space does; the wavelet expansion (8) converges unconditionally in these cases.

4.1. **Besov Cost Distribution:** The approach we take for treating L_∞ error is to alter the strategy for budgeting slightly. As before, for each wavelet ψ_I, we create an approximant $T_{N_I}\psi_I$ using a portion N_I of the total budget N, but the precise distribution of this budget follows different rules. We rely again on a cost distribution. In this case, it is:

$$\mathrm{c}_I = N|f|^{-q}_{B^s_{\tau,q}} A_j^{q-\tau}|f_I|^\tau = N\left(\frac{A_j}{|f|_{B^s_{\tau,q}}}\right)^q \left(\frac{|f_I|}{A_j}\right)^\tau. \tag{15}$$

The indices τ and q are determined by $1/\tau = s/d$ and $1/q = 1 + s/d$. The quantity A_j is a sort of "energy" of f at the dyadic level j:

$$A_j := \left(\sum_{I\in\mathcal{D}_j} |f_I|^\tau\right)^{1/\tau}.$$

We do not invest in the wavelet corresponding to I if $\lfloor \mathrm{c}_I\rfloor \le N_0$ (the constant from Lemma 3). Thus, we set

$$N_I = \begin{cases} \lfloor \mathrm{c}_I\rfloor = \Big\lfloor N|f|^{-q}_{B^s_{\tau,q}} A_j^{q-\tau}|f_I|^\tau \Big\rfloor, & \lfloor \mathrm{c}_I\rfloor \ge N_0 \\ 0 & \text{otherwise.}\end{cases} \tag{16}$$

With this choice at most N Gaussians are used:

$$\begin{aligned}\sum_{I\in\mathcal{D}} N_I &\le \sum_{I\in\mathcal{D}} \mathrm{c}_I \le N|f|^{-q}_{B^s_{\tau,q}} \sum_{j=0}^\infty A_j^{q-\tau} \sum_{I\in\mathcal{D}_j} |f_I|^\tau \\ &= N|f|^{-q}_{B^s_{\tau,q}} \sum_{j=0}^\infty A_j^{q-\tau} A_j^\tau = N.\end{aligned}$$

4.2. **Approximating the Wavelet Expansion:** The following lemma is a rough analogue to Lemma 8, where we show that the investment of centers made in one "energy level" gives a suitable error.

Lemma 11. *For a finitely supported sequence of coefficients $(a_I)_{I\in\mathcal{D}_j}$, we have the estimate, which holds for $2k>d$:*

$$\left\| \sum_{I\in\mathcal{D}_j} a_I\psi_I - \sum_{I\in\mathcal{D}_j} a_I[T_{N_I}\psi_I]\right\|_\infty \le C_{k,d}\sup_{I\in\mathcal{D}_j}\left|a_I N_I^{-k/d}\right|.$$

Proof. We treat the estimate by considering the error termwise. Theorem 4 gives the pointwise bound

$$\sum_{I\in\mathcal{D}_j} |a_I|\,|\psi_I(x)-[T_{N_I}\psi_I](x)| \le C_{k,d}\sum_{I\in\mathcal{D}_j}|a_I|N_I^{-k/d}\left(1+\frac{\mathrm{dist}(x,I)}{\ell(I)}\right)^{-2k}.$$

By applying Hölder's inequality, the lemma follows, since for $2k>d$, the series $\sum_{I\in\mathcal{D}_j}\left(1+\frac{\mathrm{dist}(x,I)}{\ell(I)}\right)^{-2k}$ is bounded by a finite constant $C_{k,d}$ that is independent of both j and x. □

We are now in a position to prove our main result for L_∞ approximation.

Theorem 12. *Given $s>0$, there is a constant $C_{s,d}$ so that for $f\in B^s_{\tau,q}$, with $1/\tau = s/d$ and $1/q = 1+s/d$, there is a linear combination of N Gaussians $s_f(x) := \sum_{j=1}^N A_j\exp\left[-(\frac{x-\xi_j}{\sigma_j})^2\right]$ so that*

$$\|f-s_f\|_\infty \le C_{s,d}N^{-s/d}|f|_{B^s_{\tau,q}}.$$

Proof. Using the budget (16), the approximant is

$$s_f := \sum_{I\in\mathcal{D}_k} f_I\,T_{N_I}\psi_I,$$

where each term, $[T_{N_I}\psi_I]$ is composed of N_I Gaussians as in Theorem 4.

We estimate $\|f-s_f\|_\infty$, recalling the unconditional convergence of the wavelet expansion for functions coming from the Besov space for this choice of τ and q.

$$\begin{aligned}
\|f-s_f\|_\infty &\le \sum_{j=-\infty}^{\infty}\left\|\sum_{I\in\mathcal{D}_j} f_I(\psi_I - T_{N_I}\psi_I)\right\|_\infty\\
&\le C(k,d)\sum_{j=-\infty}^{\infty}\left\|I\mapsto N_I^{-k/d}|f_I|\right\|_{\ell_\infty(\mathcal{D}_j)}\\
&\le C(k,d)\sum_{j=-\infty}^{\infty}\left\|I\mapsto \mathrm{c}_I^{-s/d}|f_I|\right\|_{\ell_\infty(\mathcal{D}_j)}.
\end{aligned}$$

The first inequality is simply the triangle inequality, since the sums considered are all finite, while the second is Lemma 11. The final inequality holds for $s<k$, because $\mathrm{c}_I \le 1+N_I \le CN_I$.

By invoking the definition of c_I, and by manipulating exponents (specifically, using the facts that $\tau s/d = 1$ and that $s/d + 1 = 1/q$) we arrive at

$$\begin{aligned}
\|f - s_f\|_\infty &\le C(k,d)N^{-s/d}|f|_{B^s_{\tau,q}}^{qs/d} \sum_{j=-\infty}^{\infty} A_j^{(\tau-q)s/d} \left\| I \mapsto |f_I|^{1-\tau s/d} \right\|_{\ell_\infty(\mathcal{D}_j)} \\
&= C(k,d)N^{-s/d}|f|_{B^s_{\tau,q}}^{qs/d} \sum_{j=-\infty}^{\infty} A_j^{(\tau-q)s/d} \\
&= C(k,d)N^{-s/d}|f|_{B^s_{\tau,q}}^{qs/d} \sum_{j=-\infty}^{\infty} A_j^{q} = C(k,d)N^{-s/d}|f|_{B^s_{\tau,q}}.
\end{aligned}$$

□

References

[1] C. de Boor and A. Ron, *Fourier analysis of the approximation power of principal shift-invariant spaces*, Constr. Approx., 8 (1992), pp. 427–462.

[2] R. A. DeVore and A. Ron, *Approximation using scattered shifts of a multivariate function*. To appear in Trans. Amer. Math. Soc.

[3] C. Fefferman and E. M. Stein, *Some maximal inequalities*, Amer. J. Math., 93 (1971), pp. 107–115.

[4] B. Fornberg and J. Zuev, *The Runge phenomenon and spatially variable shape parameters in RBF interpolation*, Comput. Math. Appl., 54 (2007), pp. 379–398.

[5] L. Grafakos, *Classical and modern Fourier analysis*, Pearson Education, Inc., Upper Saddle River, NJ, 2004.

[6] E. J. Kansa and R. E. Carlson, *Improved accuracy of multiquadric interpolation using variable shape parameters*, Comput. Math. Appl., 24 (1992), pp. 99–120. Advances in the theory and applications of radial basis functions.

[7] X. Li and C. A. Micchelli, *Approximation by radial bases and neural networks*, Numer. Algorithms, 25 (2000), pp. 241–262. Mathematical journey through analysis, matrix theory and scientific computation (Kent, OH, 1999).

[8] Y. Meyer, *Wavelets and operators*, vol. 37 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 1992. Translated from the 1990 French original by D. H. Salinger.

[9] B. Schlkopf, B. Scholkopf, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, *Comparing support vector machines with gaussian kernels to radial basis function classifiers*, IEEE Transactions on Signal Processing, 45 (1997), pp. 2758–2765.

[10] I. Steinwart and C. Scovel, *Fast rates for support vector machines using Gaussian kernels*, Ann. Statist., 35 (2007), pp. 575–607.

[11] Y. Ying and D.-X. Zhou, *Learnability of Gaussians with flexible variances*, J. Mach. Learn. Res., 8 (2007), pp. 249–276 (electronic).

Thomas Hangelbroek, Department of Mathematics, Texas A&M University, College Station, TX 77843

E-mail address: hangelbr@math.tamu.edu

Amos Ron, Computer Science Department, University of Wisconsin, Madison, WI 53706

E-mail address: amos@cs.wisc.edu