SIGMOID FUNCTIONS, MULTISCALE RESOLUTION OF SINGULARITIES, AND HP-MESH REFINEMENT *

DAAN HUYBRECHS[†] AND LLOYD N. TREFETHEN[‡]

Abstract. In this short, conceptual paper we observe that closely related mathematics applies in four contexts with disparate literatures: (1) sigmoidal and RBF approximation of smooth functions, (2) rational approximation of analytic functions with singularities, (3) *hp*-mesh refinement for solution of PDEs, and (4) double exponential and generalized Gauss quadrature. The relationships start from the change of variables $s = \log(x)$, and they suggest possibilities for new analyses and new methods in several areas. Concerning (2) and (3), we show that both problems feature the same effect of "linear tapering" near the singularity—of clustered poles in rational approximation, and of polynomial orders in *hp*-mesh refinement. Concerning (4), we note that the tapering effect appears here too, and that the change of variables interpretation sheds new light on why the DE and generalized Gauss methods are effective at integrating arbitrary singularities.

Key words. rational approximation, sigmoid function, logistic function, activation function, radial basis function, *hp*-mesh refinement, DE quadrature, generalized Gauss quadrature

MSC codes. 41A20, 65D15, 65D32, 68T07

1. Introduction.

A longstanding theme in numerical computation is the effective treatment of singularities. In approximation theory, it is known that any branch point singularity can be approximated by rational functions with root-exponential convergence (i.e., convergence at a rate $O(\exp(-C\sqrt{n}))$ for some $C > 0$, where n is the degree of the rational function), provided the poles are exponentially clustered near the singularity. The first observation of this paper is that the mathematics of this effect is the same as that of a seemingly very different well-known phenomenon: that smooth functions can be approximated with great efficiency by translates of a fixed smooth function such as a sigmoid or a radial basis function (RBF). This equivalence is the subject of section 2, where we consider implications for both rational approximation and smooth approximation by sigmoids, and of section 6, where we generalize sigmoids to RBFs.

Section 3 turns to another relationship: between approximation of singularities by rational functions and resolution of singularities (typically at corners) in *hp*-mesh refinement in the finite element method (*hp*-FEM). Recently it has been recognized that best and near-best rational approximations exploit a "linear tapering" effect, in which poles are exponentially clustered near singularities with a density that tapers off linearly on a log scale [36]. In section 3 we show that this linear tapering appears for the same reason that *hp*- rather than just *h*- mesh refinement is advantageous in FEM calculations. The tapering is linear in both cases (in a 1D setting), and brings a factor of 2 improvement over nontapered distributions. These observations lead to a suggestion of an alternative method of *hp*-mesh refinement that should achieve the same asymptotically optimal convergence rate.

We then go on to consider further aspects of the exponential resolution of singularities. Section 4 discusses quadrature formulas designed for efficient integration of functions with endpoint singularities, in particular the double exponential (DE) and generalized Gauss methods. In both cases tapered exponential clustering is seen,

*Submitted to the editors DATE.
[†]Dept. of Computer Science, KU Leuven, 3001 Leuven, Belgium (daan.huybrechs@kuleuven.be).
[‡]Mathematical Institute, University of Oxford, Oxford OX4 4DY, UK (trefethen@maths.ox.ac.uk).

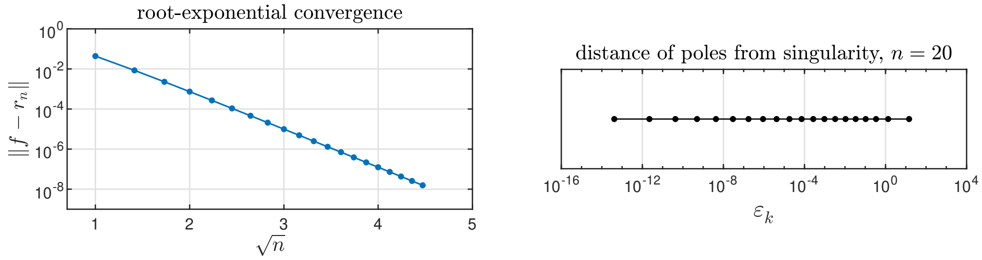

Fig. 1. *Root-exponential convergence (left, $0 \le n \le 20$) and exponential clustering of poles near 0 in $(-\infty, 0)$ (right, $n = 20$) for degree n minimax rational approximation of $f(x) = \sqrt{x}$ on $[0,1]$. Note that the horizontal axis in the first plot is $\sqrt{n}$. These effects generalize to rational approximation at any branch point singularity.*

and our analysis sheds a new light on why this happens and on why both of these quadrature methods are effective at treating a range of singularities all at once, not just a single targeted singularity such as x^α.

The significance of the change of variables $s = \log(x)$ is not just algebraic. Section 5 explores its physical interpretation, showing how multiscale separation of scales in the x variable is equivalent to the effect of exponential decay of influences along channels for elliptic PDEs.

The aim of this paper is to point out new relationships and to suggest new explorations. With the change of variables $s = \log(x)$, multiscale analysis becomes translation. Rational approximation, adaptive mesh refinement, and numerical quadrature may all benefit from a consideration of the implications of this relationship.

2. Sigmoids $\leftrightarrow$ rational approximation.

An important result of rational approximation theory, which sets it far apart from polynomial approximation, is that functions with branch point singularities can be approximated with root-exponential convergence. For example, consider

$$f(x) = \sqrt{x}, \quad x \in [0,1]. \tag{2.1}$$

Since Donald Newman in 1964 [26] it has been known that there are degree n rational functions

$$r_n(x) = a_0 + \sum_{k=1}^{n} \frac{a_k}{1 + x/\varepsilon_k}, \quad a_k \in \mathbb{R},\ \varepsilon_k > 0 \tag{2.2}$$

such that

$$\|f - r_n\| = O(\exp(-C\sqrt{n})), \quad C > 0, \tag{2.3}$$

where $\|\cdot\|$ is the supremum norm on $[0,1]$. What makes this *root-exponential convergence* possible is that the poles $\{-\varepsilon_k\}$ are exponentially clustered near the singularity at $x = 0$. This effect and the estimate (2.3) apply at any branch point singularity [14, 18, 36] of a real or complex function and are the basis of "lightning PDE solvers" for the Laplace, biharmonic, and Helmholtz equations in domains with corners [1, 5, 14, 15, 37]. Figure 1 illustrates root-exponential convergence and exponential clustering for this model problem.

The function $1/(1 + x/\varepsilon_k)$ of (2.2) is monotonically decreasing for $x \in [0,1]$ and takes values ≈ 1 for $x \ll \varepsilon_k$ and ≈ 0 for $x \gg \varepsilon_k$. If we introduce the change of variables

$$s = \log(x) \in [-\infty, 0], \quad x = e^s \in [0,1], \tag{2.4}$$

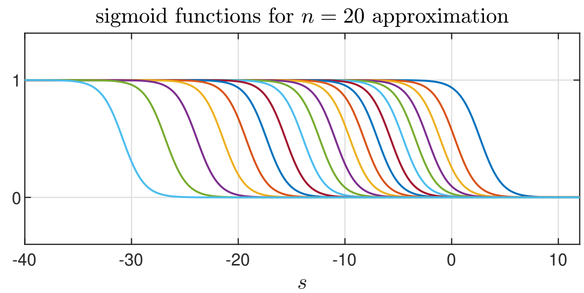

FIG. 2. *In the $s = \log(x)$ variable, the exponentially clustered poles of Figure 1 become sigmoid functions (2.5) translated to various center points s_k. Root-exponential convergence of rational approximations becomes a statement about approximation of smooth functions by linear combinations of translates of a fixed smooth function.*

then this function is transformed into

$$\frac{1}{1+e^s/\varepsilon_k} = \frac{1}{1+e^{s-s_k}}, \tag{2.5}$$

where $s_k = \log \varepsilon_k$. This function, or more properly its flipped form $1/(1+e^{s_k - s})$ with $s - s_k$ replaced by $s_k - s$, is the most basic example of a sigmoid function, known as the *logistic function.* In physics it goes by the name of the *Fermi* or *Fermi–Dirac function,* and it is an elementary transformation of the hyperbolic tangent. Functions of this kind are prototypical activation functions in neural networks, and the literature of this area is vast [7, 12, 19, 21, 28]. (We make no claim here to a serious link with the more recent surge in machine learning, since the approximations of this paper involve just a single hidden layer rather than a composition of several layers.)

With the change of variables (2.4), we may follow (2.1) and (2.2) to define

$$F(s) = f(x) = e^{s/2} \tag{2.6}$$

and

$$R_n(s) = r_n(x) = a_0 + \sum_{k=1}^{n} \frac{a_k}{1+e^{s-s_k}}. \tag{2.7}$$

Equation (2.3) then implies that there are approximations (2.7) such that

$$\|F - R_n\| = O(\exp(-C\sqrt{n})), \quad C > 0, \tag{2.8}$$

where $\|\cdot\|$ is now the supremum norm on $[-\infty, 0]$. Equation (2.8) is nothing else than a claim about root-exponential approximation of a smooth function on $[-\infty, 0]$ by linear combinations of translates of a standard smooth function, plus a constant. Following results such as those of [9], this could be proved directly in this setting rather than in the setting of rational functions, where the standard proof technique is the relatively advanced Hermite contour integral formula [35]. The root-exponential rate results from balancing discretization errors associated with separations $\Delta s_k = O(1/\sqrt{n})$ against truncation errors associated with a grid extent $s_{\min} = \min s_k = O(\sqrt{n})$ [18]. Figure 2 illustrates the smooth functions in question for the problem of Figure 1.

To summarize this section: with the change of variables $s = \log(x)$, the approximation of a smooth function by linear combinations of translates of a logistic function becomes equivalent to the approximation of a function with a branch point singularity by rational functions with exponentially clustered poles.

3. Rational approximation $\leftrightarrow$ *hp*-mesh refinement. Rational approximation with poles exponentially clustered near singularities seems akin to the resolution of functions near singularities by piecewise polynomials on exponentially refined meshes. Such techniques of *mesh refinement* are well known in the literature of the finite element method and associated approximation theory [8, 16, 17, 22, 29, 30].

Exponential clustering of poles is reflected in the approximately uniform spacing on the semilogx scale in the right image of Figure 1, or equivalently, the approximately uniform spacing of the sigmoid functions (2.5) in Figure 2. However, it is notable that in both of these images the spacing is only approximately uniform, growing sparser toward the left. This is the phenomenon of *tapered exponential clustering* investigated in [36]. Quantitatively, one finds that the density of poles with respect to the s variable decreases linearly as s decreases to some value $s_{\min}$. This distribution brings a factor of 2 improvement in convergence rate as a function of n—because a uniform density would have the same convergence rate but twice as many poles. (The more local sparsification in the rightmost few points of Figures 1 and 2 is investigated in [18] with appeal to the asymptotic results of Stahl [31].)

Comparison reveals that this tapered exponential clustering corresponds closely to what is known as hp-mesh refinement (h stands for grid spacing, p for order of approximation). In particular, the standard hp-mesh refinement formula in one dimension has the same linear pattern just described, with polynomial order taking the role of pole density on the logarithmic scale. A singular function such as $\sqrt{x}$ on $[0, 1]$ is approximated by piecewise polynomials on intervals of lengths decreasing exponentially toward the singularity, with polynomial representations of linearly decreasing degrees $\dots, 3, 2, 1, 0$. It is the same pattern, and it brings the same factor of 2 speedup for the same reason.

One can explain linear tapering for hp-FEM in various ways in various settings. In [36] an argument is given based on potential theory. Here is an outline of the simpler argument that originates with DeVore and Scherer [8, 29] in the study of piecewise polynomial approximations of x^α on exponentially graded meshes on $[0, 1]$. We speak for simplicity in terms of mesh refinement by factors of $1/2$, though the optimal factor is actually $(\sqrt{2}-1)^2 \approx 0.172$. We show the reasoning for $\sqrt{x}$, though the same argument (with different constants) applies to x^α for any positive noninteger α. The observations of DeVore and Scherer were generalized to ODE and PDE discretizations a few years later by Babuška and his collaborators [16, 17]. In multiple dimensions, the details change.

1. Approximation of $\sqrt{x}$ on $[1/2, 1]$ is the same as approximation of $\sqrt{2x}$ on $[1/4, 1/2]$.
2. Therefore approximation of $\sqrt{x}$ on $[1/4, 1/2]$ is the same problem too, but with an accuracy criterion loosened by a factor $\sqrt{2}$.
3. Functions like these, bounded away from singularities, can be approximated by polynomials with exponential convergence.
4. Therefore, that loosening by the factor $\sqrt{2}$ allows one to lower the degree of the polynomial by a constant increment and still get the same accuracy.
5. Repeat on $[1/8, 1/4]$, $[1/16, 1/8], \dots$.

Note that if we skipped step (4) and fixed the same polynomial degree on all intervals, the same overall accuracy would still be achieved but in an unbalanced manner, with errors on $[1/2, 1]$ dominating those on the other subintervals. The number of free parameters would approximately double.

To summarize this section: the standard formula for hp-mesh refinement in 1D

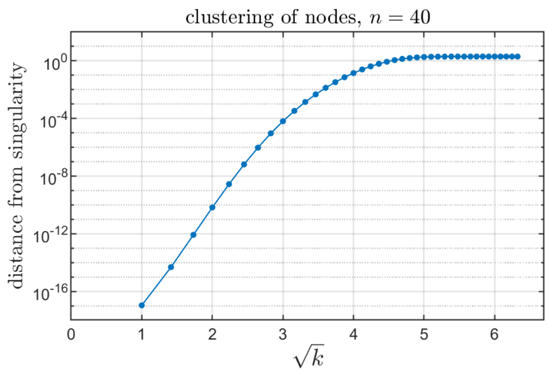

FIG. 3. *Nodes of a double exponential quadrature formula with standard parameters, showing tapered exponential clustering near a singularity. (Figure adapted from Figure* 13 *of* [36]*.)*

involves a linear decrease of the polynomial degree toward the singularity, and this corresponds to the linear decrease of the pole density on a logarithmic scale in tapered exponential clustering of poles in rational approximation, resulting in the same factor of 2 speedup.

The link with rational functions highlights that the standard *hp*-mesh refinement strategy is not the only way to achieve linear tapering. An alternative would be to hold the polynomial degree p fixed and instead refine h super-exponentially at the singularity. We do not know if such a prescription has been used in finite element calculations.

4. Double exponential and generalized Gauss quadrature. In the area of quadrature or numerical integration, many methods have been developed for dealing with singularities. When a fixed endpoint singularity like x^α is known, a targeted quadrature formula can be derived: the prototype is a Gauss–Jacobi formula. For dealing with more complicated, mixed, or unknown singularities, however, more general techniques have been proposed. One is *double exponential* or *tanh-sinh* quadrature [2, 25, 33]. As illustrated in Figure 3, the tanh-sinh formula with standard parameter choices produces a tapered exponentially clustered distribution of quadrature points. This figure is adapted from Figure 13 of [36], where full details can be found. Such results indicate that DE quadrature is probably related to what is seen with rational approximation and *hp*-mesh refinement, but so far as we know, no analysis of this effect has yet been carried out.

Kirill Serkh (private communication) has shown us that similar effects also arise with *generalized Gauss* and *universal* quadrature formulas [3, 4]. These are quadrature formulas that are constructed by linear algebra methods related to Gauss quadrature so as to be efficient at integrating not just a single singularity such as a fixed power x^α but a range of singularities such as x^α, $\alpha \in [0, 1]$. Again it appears that in important cases, the nodes are exponentially clustered near the singularity with a tapered distribution. The arguments of this paper reveal that the ability of a single such formula to handle a wide range of singularities is related to the ability of a fixed set of exponentially clustered poles for a rational function to resolve arbitrary branch point singularities [14, 18, 36].

In a similar vein one can compute Gaussian quadrature rules for spaces of rational

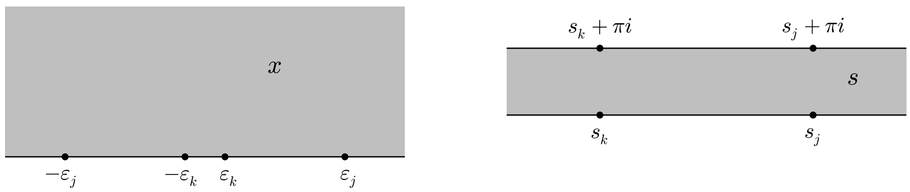

FIG. 4. *Sketch of the "physics" of the change of variables $s = \log(x)$. This is a conformal map of the upper half x-plane to an infinite strip in the s-plane, with poles $-\varepsilon_k$ and sample locations ε_k on exponentially separated scales mapping to poles $s_k + \pi i$ and sample locations s_k on opposite sides of the strip. The well-known exponential decay of influences along strips explains why a pole at $s_k + \pi i$ has a significant effect at s_k but a much smaller effect at a different point s_j with $|s_j - s_k| \gg 1$.*

functions with fixed clustered poles, such as those in Figure 1, using for example the algorithm for rational quadrature of Gautschi [11]. One again observes root-exponential convergence and, by now unsurprisingly, an exponential distribution of the quadrature points with linear tapering. The difference between DE and the Gaussian rules is that the former is optimized by tuning parameters, whereas Gaussian quadratures are inherently optimal or optimized by linear algebra calculations—yet both lead to similar distributions.

5. The physics of $s = \log(x)$: separation of scales. The change of variables $s = \log(x)$ is not just an algebraic trick. It also has a physical interpretation alluded to in section 5 of [36]; see for example Figure 11 of that paper.

Even before introducing the change of variables, the point can be seen in the x variable. The function $1/(1 + x/\varepsilon_k)$, with its pole at distance ε_k to the left of $x = 0$, is essentially constant and hence inactive to the right of $x = 0$ for $x \ll \varepsilon_k$ (taking the value 1) and $x \gg \varepsilon_k$ (taking the value 0). It is only for $x \approx \varepsilon_k$ that this function is active. Thus exponentially separated poles $\{-\varepsilon_k\} \subseteq (-\infty, 0)$ are physically decoupled, operating in independent regimes, with each pole at $-\varepsilon_k < 0$ affecting the approximation on $(0, \infty)$ nontrivially only for $x \approx \varepsilon_k$.

The change of variables $s = \log(x)$ suggests a physical explanation of this separation of scales effect. A problem with a singularity at $x = 0$ can be motivated as a model of a corner singularity in a PDE problem. Specifically, suppose a Laplace problem is posed in the upper half complex x-plane with a singularity at $x = 0$. Changing to $s = \log(x)$ transplants this problem to the infinite strip $0 < \mathrm{Im}s < \pi$ in the s-plane. The problem is now smooth, with the singularity moved to $-\infty$. Exponentially clustered poles $-\varepsilon_k \in (-\infty, 0)$ become well separated poles $s_k + \pi i$ on the upper side of the strip. And now, as sketched in Figure 4, it is a well-known effect of potential theory (or elasticity, where it is called the Saint-Venant principle [34]) that influences decay exponentially with distance along a strip. In the field of numerical conformal mapping, this goes by the name of the *crowding phenomenon*; see Theorems 2–5 of [13].

The argument just made is tied to the Laplace equation, whose solutions are conformally invariant. However, the essence of the matter will be the same for any problem whose highest order derivative is the Laplacian, because close to a singularity, this term will dominate. With the Helmholtz equation $\Delta u + k^2 u = 0$, for example, the influence of the $k^2 u$ term quickly shuts off to zero relative to that of the Δu term as one comes exponentially close to a corner [15].

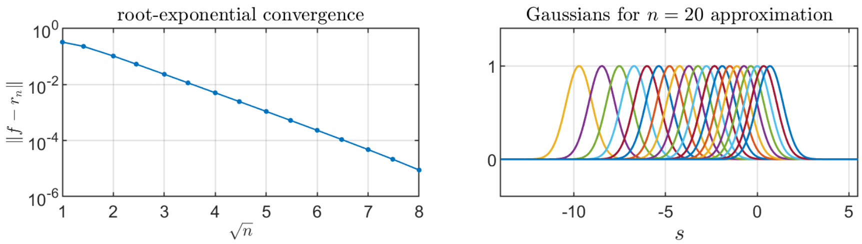

FIG. 5. *Curves as in Figures 1 and 2 for the approximation of $\sqrt{x}$ for $x \in [0,1]$. The approximating functions are transplants to $x \in [0,1]$ of Gaussians $\exp(-(s-s_k)^2)$ for $s \in [-\infty, 0]$, with the centers s_k distributed in a tapered manner. Clean root-exponential resolution of the singularity is observed.*

6. Radial basis functions and other activation functions. A rational function $r(x)$ as in (2.2) is a sum of simple poles, which in the s variable becomes a linear combination of sigmoids as in Figure 2. The picture changes little for other activation functions. For example, if the poles $(x + \varepsilon_k)^{-1}$ in (2.2) are replaced by powers $(x+\varepsilon_k)^{-a}$ for an arbitrary $a > 0$, Figure 2 does not change very much. This matches theoretical and experimental results in neural networks, where choices between activation functions are typically based more on the efficiency of learning algorithms such as stochastic gradient descent than on approximation power.

As many authors have noted, a closely related topic is that of approximation by radial basis functions. Here again one approximates a complicated function by a linear combination of translates of a simple fixed function, and convergence may be very fast when the latter is smooth [6, 9, 10, 23, 24, 27]. To illustrate the application of the duality between $x \in [0,1]$ and $s \in [-\infty, 0]$ to RBFs, Figure 5 constructs approximations to $\sqrt{x}$ on $[0,1]$ in the form of transplants to $x \in [0,1]$ of Gaussians $\exp(-(s-s_k)^2)$ for $s \in [-\infty, 0]$. The centers s_k are spaced in a tapered manner, as shown for the case $n = 20$ in the right-hand image. Clean root-exponential convergence is observed for this very nonstandard system of basis functions near a singularity.

As mentioned in section 2, although sigmoidal and other activation functions are important in neural networks and deep learning, the present paper touches only the surface of that discipline. The approximation (2.7) is not composite, but involves just what is conventionally called a single hidden layer. Smooth activation functions have accuracy advantages for single-layer approximation, but in the multi-layer setting of deep learning, that advantage diminishes and the simpler non-smooth function known as ReLU is used more often [21, 32]. We cannot resist mentioning that the universal approximation power of ReLU units was exploited by Henri Lebesgue at age 23 in his first published paper, in which he presented a new proof of the Weierstrass approximation theorem [20].

7. Conclusion. This paper started from a fundamental equivalence:

> $(*)$ *The change of variables $s = \log(x)$ gives an equivalence between smooth approximation for $s \in [-\infty, 0]$ and approximation with a singularity at $x = 0$ for $x \in [0,1]$.*

Considering this relationship in various contexts has led us to a number of observations and proposals:

1. *Rational approximation near singularities and smooth sigmoidal approxima-*

tion. In the simplest case (∗) gives an equivalence between rational approximation of x^α and other functions with branch point singularities on $[0,1]$ and approximation of smooth functions rapidly approaching a constant as $s \to -\infty$ on $[-\infty, 0]$ by linear combinations of translates of a smooth sigmoidal function (Figure 1). The literatures of these fields are largely disjoint, and this connection opens up the prospect that techniques used in one area, such as the Hermite contour integral in rational approximation, could be applied to obtain new results in the other.

2. The "tapering" effect of rational approximation is the same as the "p" part of hp-FEM. *hp*-mesh refinement dates back 40 years, whereas recognition of tapered exponential clustering of poles in rational approximation is recent [36]. We have shown that the mechanisms behind these effects are the same, and that in particular, each brings a factor of 2 speedup (in 1D), for the same reason (Figures 1 and 2). This analysis of the source of the factor of 2 reveals that other strategies of *hp*-FEM should be able to achieve the same optimal asymptotic convergence rate. In particular, instead of regular exponential mesh refinement combined with linearly tapered polynomial order, one could use fixed polynomial order combined with a tapered schedule of exponential refinement.

3. Multiscale separation of scales at a singularity can be interpreted as the Saint-Venant or "crowding" phenomenon of elliptic PDE (Figure 4). Two points at distances 10^{-10} and 10^{-5} from a corner singularity may seem to be close together, for example, but in fact they are only weakly coupled in the same way as the points $s = \log(10^{-10}) \approx -23.0$ and $s = \log(10^{-5}) \approx -11.5$ on the side of an infinite strip of width π (Figure 4).

4. Singularities can be approximated with root-exponential convergence by many different kinds of exponentially clustered RBFs. Sigmoid functions are the archetype, corresponding to poles $1/(x - x_j)$ of a rational function, but as shown in section 6, other smooth functions of s correspond to other singular functions of x, and the approximation powers are comparable (Figure 5).

5. DE and generalized Gauss quadrature. Quadrature of functions with endpoint singularities can exploit the same principles of tapered exponential clustering and insensitivity to the precise nature of a branch point singularity (Figure 3).

Our arguments have been univariate, whereas both approximation and solution of PDEs are important problems also in multiple dimensions. In the case of PDEs, this side of the subject is highly developed, and it would be interesting to see if results in this area could be transferred to new ideas for multivariate rational or related approximation.

Acknowledgments. We are grateful to Kirill Serkh of the University of Toronto for showing us the tapered exponentially clustered nodes of generalized Gauss and universal quadrature formulas.

REFERENCES

[1] P. J. Baddoo, *Lightning solvers for potential flows,* Fluids, 5 (2020), pp. 1–17.
[2] D. H. Bailey and J. Borwein, *Hand-to-hand combat with thousand-digit integrals,* J. Comput. Sci. 3 (2012), pp. 77–86.
[3] J. Bremer, Z. Gimbutas, and V. Rokhlin, *A nonlinear optimization procedure for generalized Gaussian quadratures,* SIAM J. Sci. Comput. 32 (2010), pp. 1761–1788.
[4] J. Bremer, V. Rokhlin, and I. Sammis, *Universal quadratures for boundary integral equations on two-dimensional domains with corners,* J. Comput. Phys. 229 (2010), pp. 8259–8280.

[5] P. D. Brubeck and L. N. Trefethen, *Lightning Stokes solver,* SIAM J. Sci. Comp., 44 (2022), pp. A1205–A1226.
[6] M. D. Buhmann, Radial Basis Functions: Theory and Implementations, Cambridge, 2003.
[7] G. Cybenko, *Approximation by superpositions of a sigmoidal function,* Math. Control Signals Syst., 2 (1989), pp. 303–314.
[8] R. A. DeVore and K. Scherer, *Variable knot, variable degree spline approximation to* x^β, in R. A. DeVore and K. Scherer, eds., Quantitative Approximation, Academic Press, 1980, pp. 121–131.
[9] T. A Driscoll and B. Fornberg, *Interpolation in the limit of increasingly flat radial basis functions.* Computers Math. w. Applics., 43 (2002), pp. 413–422.
[10] G. E. Fasshauer, *Meshfree Approximation Methods with MATLAB,* World Scientific, 2007.
[11] W. Gautschi, *The use of rational functions in numerical quadrature,* J. Comput. Appl. Math., 133 (2001), pp. 111–126.
[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning,* MIT, 2016.
[13] A. Gopal and L. N. Trefethen, *Representation of conformal maps by rational functions,* Numer. Math., 142 (2019), pp. 359–382.
[14] A. Gopal and L. N. Trefethen, *Solving Laplace problems with corner singularities via rational functions,* SIAM J. Numer. Anal., 57 (2019), pp. 2074–2094.
[15] A. Gopal and L. N. Trefethen, *New Laplace and Helmholtz solvers,* Proc. Nat. Acad. Sci., 116 (2019), p. 10223.
[16] W. Gui and I. Babuška, *The h, p, and h-p version of the finite element method in* 1 *dimension.II: The error analysis of the h- and h-p versions,* Numer. Math., 49 (1986), pp. 613–657.
[17] B. Guo and I. Babuška, *The h-p version of the finite element method. I: The basic approximation results,* Comput. Mech. 1 (1986), pp. 21–41.
[18] A. Herremans, D. Huybrechs, and L. N. Trefethen, *Resolution of singularities by rational functions,* SIAM J. Numer. Anal., to appear.
[19] K. Hornik, M. Stinchcombe, and H. White, *Multilayer feedforward networks are universal approximators,* Neural Networks, 2 (1989), pp. 359–366.
[20] H. Lebesgue, *Sur l'approximation des fonctions,* Bull. Sci. Math., 22 (1898), pp. 278–287.
[21] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, Nature, 521 (2015), pp. 436–444.
[22] J. M. Melenk, *hp-Finite Element Methods for Singular Perturbations,* Springer, 2002.
[23] H. N. Mhaskar, *Neural networks for optimal approximation of smooth and analytic functions,* Neural Computation, 8 (1996), pp. 164–177.
[24] H. N. Mhaskar and C. A. Micchelli, *Approximation by superposition of a sigmoidal function and radial basis functions,* Adv. Appl. Math., 13 (1992), pp. 350–373.
[25] M. Mori and M. Sugihara, *The double-exponential transformation in numerical analysis,* J. Comput. Appl. Math. 127 (2001), pp. 287–296.
[26] D. J. Newman, *Rational approximation to* $|x|$, Mich. Math. J., 11 (1964), pp. 11–14.
[27] J. Park and I. W. Sandberg, *Universal approximation using radial basis function networks,* Neural Computation, 3 (1991), pp. 246–257.
[28] A. Pinkus, *Approximation theory of the MLP model in neural networks,* Acta Numer. 8 (1999), pp. 143–195.
[29] K. Scherer, *On optimal global error bounds obtained by scaled local error estimates,* Numer. Math., 36 (1981), pp. 151–176.
[30] C. Schwab, *p− and hp− Finite Element Methods: Theory and Application in Solid and Fluid Mechanics,* Clarendon Press, Oxford, 1998.
[31] H. R. Stahl, *Poles and zeros of best rational approximants of* $|x|$, Constr. Approx. 10 (1994), pp. 469–522.
[32] G. Strang, *The functions of deep learning,* SIAM News, 51 (2018), pp. 1–4.
[33] H. Takahasi and M. Mori, *Double exponential formulas for numerical integration,* Publ. RIMS Kyoto 9 (1974), pp. 721–741.
[34] R. A. Toupin, *Saint-Venant's principle,* Arch. Ratl. Mech. Anal., 18 (1965), pp. 83–96.
[35] L. N. Trefethen, *Numerical analytic continuation,* Japan J. Appl. Math. (2023).
[36] L. N. Trefethen, Y. Nakatsukasa, and J. A. C. Weideman, *Exponential node clustering at singularities for rational approximation, quadrature, and PDEs,* Numer. Math., 147 (2021), pp. 227–254.
[37] Y. Xue, S. L. Waters, and L. N. Trefethen, *Computation of 2D Stokes flows via lightning and AAA rational approximation,* arXiv:2306.13545v1, 2023.