arXiv:2303.01967v1 [math.NA] 3 Mar 2023

S形函数与奇点的多尺度分辨 *

DAAN HUYBRECHS† 和LLOYD N. TREFETHEN‡

摘要。在这篇简短的概念性论文中，我们发现，在三个文献各自独立的领域中，基本上适用的是同一套数学理论：(1) 对光滑函数进行 S形函数和径向基函数逼近，(2) 在奇异点附近对解析函数进行有理逼近，以及 (3) hp用于偏微分方程求解的网格细化。(1) 与 (2) 之间的关系，只需通过变量替换$s = \log(x)$ 即可建立；我们用“sigmoid $=$ log(ratapprox)”作为这一关系的非正式记忆口诀。

关键词：有理逼近，S形函数，逻辑斯谛函数，激活函数，径向基函数， hp 网格细化

数学学科分类代码：41A20，65D15，68T07

1. S形函数与有理逼近。具有分支点奇异性的函数可以通过有理函数实现根指数收敛逼近。例如，考虑

$$f(x) = \sqrt{x}, \quad x \in [0,1]. \tag{1.1}$$

自 Donald Newman 1964 [24] 以来，人们已知存在次数为 n 的有理函数

$$r_n(x) = a_0 + \sum_{k=1}^{n} \frac{a_k}{1 + x/\varepsilon_k} \tag{1.2}$$

使得

$$\|f - r_n\| = O(\exp(-C\sqrt{n})), \quad C > 0, \tag{1.3}$$

其中， $\|\cdot\|$ 是 $[0,1]$上的上确界范数。使这种根指数收敛成为可能的原因是，极点 $\{-\varepsilon_k\}$ 在 $x = 0$处的奇点附近呈指数聚集。这种效应适用于实函数或复函数的任意分支点奇异性 [12, 16, 33]附近，也是用于带角点区域 [1, 5, 12, 13]中拉普拉斯方程、双调和方程和亥姆霍兹方程的“闪电偏微分方程求解器”的基础。图 1 展示了这一模型问题中的根指数收敛和指数聚集。

式 (1.2) 中的函数 $1/(1 + x/\varepsilon_k)$ 在 $x \in [0,1]$ 上单调递减，且当 $x \ll \varepsilon_k$ 时取值为 ≈ 1，当 $x \gg \varepsilon_k$时取值为 ≈ 0。若引入变量替换

$$s = \log x \in [-\infty, 0], \quad x = e^s \in [0,1], \tag{1.4}$$

那么该函数变换为

$$\frac{1}{1 + e^s/\varepsilon_k} = \frac{1}{1 + e^{s - s_k}}, \tag{1.5}$$

*提交给编辑，日期。

†计算机科学系，鲁汶大学，比利时鲁汶 3001（daan.huybrechs@kuleuven.be）。

‡数学研究所，牛津大学，英国牛津 OX4 4DY（tre- fethen@maths.ox.ac.uk）。

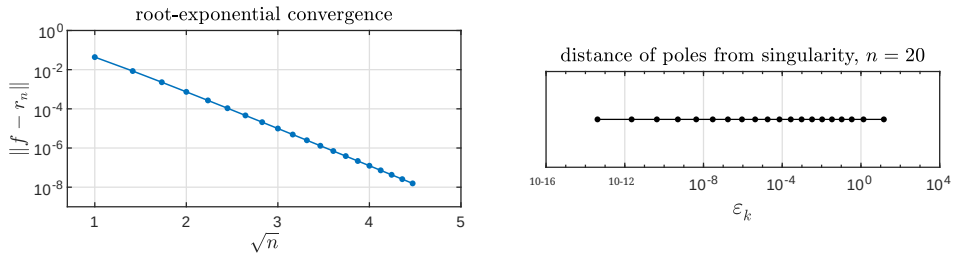


FIG. 1. *Root-exponential convergence (left, $0 \le n \le 20$) and exponential clustering of polesnear 0 in $(-\infty, 0)$ (right, $n = 20$) for degree n minimax rational approximation of $f(x) = \sqrt{x}$ on $[0, 1]$. Note that the horizontal axis in the first plot is $\sqrt{n}$. These effects generalize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

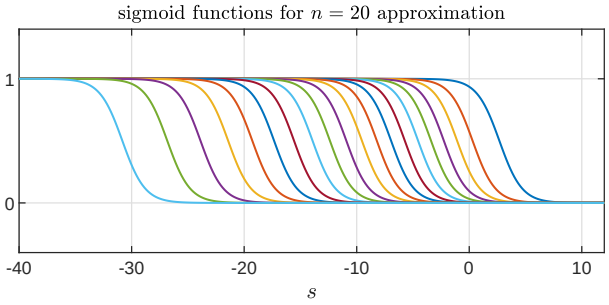


FIG. 2 *In the $s =$ x variable the exponentially clustered poles of Figure 1 become sigmoid . log, functions 1.5 translated to various center points s_k. Root-exponential convergence of rational () approximations becomes a statement about approximation of smooth functions by linear combinations of translates of a fixed smooth function.*

其中 $s_k = \log \varepsilon_k$。该函数，或者更准确地说，其反函数 $1/(1 + e^{s_k - s})$，是最基本的 sigmoid 函数示例，称为逻辑斯谛函数。在物理学中，它被称为费米函数或费米–狄拉克函数，也是双曲正切的一种基本变换。这类函数是神经网络中典型的激活函数，相关领域的文献浩如烟海 [7, 11,17, 19, 26]。

通过变量替换 (1.4)，我们可以仿照 (1.1) 和 (1.2) 来定义

$$F(s) = f(x) = e^{s/2} \tag{1.6}$$

and

$$R_n(s) = r_n(x) = a_0 + \sum_{k=1}^{n} \frac{a_k}{1 + e^{s - s_k}}. \tag{1.7}$$

方程（1.3）由此推出，存在满足下式的逼近（1.7）：

$$\|F - R_n\| = O(\exp(-C\sqrt{n})), \quad C > 0, \tag{1.8}$$

其中，$\|\cdot\|$ 现在表示 $[-\infty, 0]$上的上确界范数。方程（1.8）无非是在断言：通过标准光滑函数的平移的线性组合再加上一个常数，可以在 $[-\infty, 0]$ 上对一个光滑函数进行根指数逼近。根据 [9], 等人的结果，在这一情形下，这个结论或许可以用更简单的方法证明

相比之下，有理函数情形中的标准证明方法要复杂得多，需要用到相对高深的Hermite积分公式 [32]。根指数收敛率来自对以下两类误差的平衡：与间距 $\Delta s_k = O(1/\sqrt{n})$ 相关的离散误差，以及与网格范围 $s_{\min} = \min s_k = O(\sqrt{n})$ [16]相关的截断误差。图2展示了图1所述问题中涉及的光滑函数。

概括本节开篇内容：经过变量替换 $s = \log x$，用逻辑斯谛函数的平移的线性组合逼近光滑函数，等价于用极点呈指数聚集的有理函数逼近具有分支点奇异性的函数。

2. 有理逼近 $\leftrightarrow$ ***hp*** 网格细化。极点在奇异点附近呈指数聚集的有理逼近，显然与通过指数细化网格来分辨函数在奇异点附近的行为相似。这类网格细化技术在有限元方法及相关逼近理论的文献中广为人知 [8, 14, 15, 20, 27, 28]。

极点的指数聚集体现在图 1 右图的半对数横轴上近似均匀的间距中，或者等价地，体现在图 2 中 S 形函数 (1.5) 的近似均匀间距中。然而值得注意的是，在这两幅图中，间距都只是近似均匀，向左逐渐变稀。这就是 [33] 中研究的渐疏指数聚集现象。定量而言，随着 s 减小到某个值 $s_{\min}$，极点相对于 s 变量的密度大致呈线性下降。这种分布使收敛速度提高一倍（以 n 为函数）——因为均匀密度虽能达到相同的收敛速度，却需要两倍数量的极点。（图 1 和图 2 最右侧几个点更局部的稀疏化现象，则在 [16] 中借助 Stahl [29] 的渐近结果进行了研究。）

我们注意到，这种渐疏指数聚集与所谓的 hp 网格细化十分相似（h 表示网格间距， p 表示逼近阶数）。具体而言，一维中的标准 hp 网格细化公式具有刚才所描述的相同线性模式，其中多项式阶数取代了极点密度。像 $\sqrt{x}$ 这样定义在 $[0,1]$ 上的奇异函数，可以用分段多项式逼近：在趋近奇点时，区间长度按指数递减，而多项式表示的次数则线性递减.....3、2、1、0。模式相同，原因也相同，因此加速效果同样提升一倍。

在不同情形下，可以用多种方式解释线性渐疏。在 [33] 中，给出了一种基于位势理论的论证。下面概述一种较为简单的论证，它源自 DeVore 和 Scherer [8, 27] 对定义在 $[0,1]$上、采用指数分级网格的 x^α 分段多项式逼近的研究。为简便起见，我们以网格间距按 $1/2$倍细化来表述，不过最优因子实际上是 $(\sqrt{2}-1)^2 \approx 0.172$。几年后，Babuška 及其合作者 [14, 15]将 DeVore 和 Scherer 的观察推广到了常微分方程和偏微分方程的离散化。在多维情形下，具体细节会有所不同。

1. 在 $[1/2,1]$ 上逼近 $\sqrt{x}$ ，与在 $[1/4,1/2]$ 上逼近 $\sqrt{2x}$ 相同。2. 因此，在 $[1/4,1/2]$ 上逼近 $\sqrt{x}$ 也是同一个问题，只是精度标准放宽了 $\sqrt{2}$ 倍。3. 这类远离奇点且有界的函数，可以用多项式以指数收敛的速度逼近。4. 因此，将精度标准放宽 $\sqrt{2}$ 倍，可以使多项式次数降低一个固定增量，同时仍达到相同的精度。

5. 在 $[1/8, 1/4]$、$[1/16, 1/8]$ 上重复，.....

第二节总结如下：对于一维网格细化，hp 的标准公式要求多项式次数朝奇点方向线性递减；这对应于有理逼近中极点的渐疏指数聚集在对数尺度上的极点密度线性递减，因此带来相同的 2 倍加速。

我们注意到，这种标准的 hp 网格细化策略并非实现线性渐缩的唯一方法。另一种做法是固定多项式次数 p，转而在奇点处进行 h 超指数细化。我们不知道这种方案是否曾用于有限元计算。

3. 双指数求积与广义高斯求积。在求积或数值积分领域，人们已经开发出许多处理奇异点的方法。已知存在 x^α 这样的端点奇点时，可以推导出针对性的求积公式：其典型代表是高斯–雅可比求积。然而，为了处理更复杂或未知的奇异点，人们提出了更通用的技术。其中一种是双指数求积，即双曲正切-双曲正弦求积 [2, 23, 31]。如 [33], 图 14 所示，采用标准参数设置的双曲正切-双曲正弦公式会产生一种渐缩的指数聚集求积点分布，这表明它可能与有理逼近以及hp网格细化中观察到的现象有关。

Kirill Serkh（私人通信）告知我们，广义高斯求积公式与通用求积公式 [3, 4] 也会产生类似效应。这些求积公式通过与高斯求积相关的线性代数方法构造而成，能够高效地积分多种奇点，而不只是单个奇点（例如固定幂次 x^α），还包括一系列奇点，如 x^α、$\alpha \in [0, 1]$。同样，在一些重要情形下，节点似乎会以渐缩分布指数聚集在奇点附近。

4. $s = \log(x)$ 的物理学：尺度分离。变量替换 $s = \log x$ 不只是一个代数技巧，它还具有物理解释，这一点在 [33] 第 5 节中有所提及；例如，可参见该文的图 11。

甚至在引入变量替换之前，也可以从 x 变量中看出这一点。函数 $1/(1 + x/\varepsilon_k)$ 的极点位于 $x = 0$ 左侧距离为 ε_k 处；当 $x \ll \varepsilon_k$（取值为 1）和 $x \gg \varepsilon_k$（取值为 0）时，该函数在 $x = 0$ 右侧基本为常数，因此不起作用。只有在 $x \approx \varepsilon_k$ 时，这个函数才起作用。因此，指数分离的极点 $\{-\varepsilon_k\} \subseteq (-\infty, 0)$ 在物理上彼此解耦，各自在独立区间内发挥作用；位于 $-\varepsilon_k < 0$ 的每个极点仅在 $x \approx \varepsilon_k$ 时影响区间 $(0, \infty)$ 上的逼近。

变量替换 $s = \log x$ 为这种尺度分离效应提供了一种物理解释。具有奇点 $x = 0$ 的问题，可以作为偏微分方程问题中角点奇异性的模型。具体而言，假设在上半复 x-平面上提出一个拉普拉斯问题，且奇点位于 $x = 0$。通过变量替换 $s = \log x$，可将该问题映射到 $0 < \mathrm{Im} s < \pi$ 位于 s-平面中的无限条带上。此时问题变得光滑，奇点移至 $-\infty$。指数聚集极点 $-\varepsilon_k \in (-\infty, 0)$ 变成了条带上侧彼此充分分离的极点 $s_k + \pi i$。现在，正如图3所示，位势理论（或弹性力学中称为圣维南原理）的一个众所周知的效应是：影响会沿条带随距离增加而呈指数衰减。

刚才的论证与拉普拉斯方程有关，因为它的解

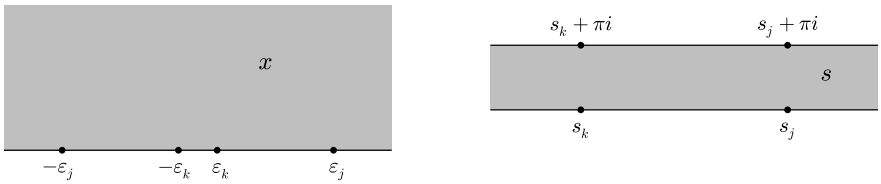


FIG. 3。 *Sketch of the "physics" of the change of variables* $s = \log x$. *This is a conformal map ofthe upper half* x*-plane to an infinite strip in the* s*-plane, exponentially separated scales mapping to poles* $s_k + \pi i$ *and sample locations* s_k *on opposite sides of the strip. The wall-to-wall 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* $|s_j - s_k| \gg 1$.

在共形映射下保持不变。不过，对于最高阶导数为拉普拉斯算子的任何问题，核心道理都相同，因为在奇点附近，这一项会占主导地位。例如，对于亥姆霍兹方程 $\Delta u + k^2 u = 0$，当距离角点以指数速度迅速趋近于零时， $k^2 u$ 项的影响也会迅速消失。

5. 径向基函数和其他激活函数。正如式（1.2）中的有理函数 $r(x)$ 是简单极点之和，在 s 变量中，它可表示为如图 2 所示的 S形函数线性组合。对于其他各种激活函数，图像变化不大。例如，如果将式（1.2）中的极点（$x + \varepsilon_k$）$^{-1}$ 替换为幂（$x + \varepsilon_k$）$^{-a}$，其中 $a > 0$ 任意，图 2 也不会有太大变化。这与神经网络中的理论和实验结果相符：在神经网络中，激活函数的选择通常更多取决于随机梯度下降等学习算法的效率，而非逼近能力。

正如许多作者所指出的，一个密切相关的课题是径向基函数的逼近。同样，人们用一个较简单的固定函数的平移的线性组合来逼近复杂函数；当后者光滑时，收敛速度可能非常快 [6, 9, 10, 21, 22, 25]。

最后需要指出，尽管 S 形函数及其他激活函数在神经网络和深度学习中十分重要，但本文对这一领域仅触及皮毛。逼近式 (1.7) 不是复合形式，而只涉及通常所说的单个"隐藏层"。光滑激活函数在单层逼近中具有精度优势，但在深度学习的多层情形下，这种优势会减弱，因此人们更常使用一种较简单的非光滑函数 ReLU [19, 30]。我们还忍不住要提一句：亨利·勒贝格在 23 岁时，就在他的第一篇已发表论文中利用了 ReLU 单元的通用逼近能力；该论文给出了魏尔斯特拉斯逼近定理的一种新证明 [18]。

致谢。我们感谢多伦多大学的 Kirill Serkh 向我们介绍广义高斯求积公式与通用求积公式中渐缩的指数聚集节点。

参考文献

[1] P. J. Baddoo, *Lightning solvers for potential flows*,《Fluids》5 (2020), 第 1–17 页。
[2] D. H. Bailey and J. Borwein, *Hand-to-hand combat with thousand-digit integrals*,《J. Comput. Sci.》3 (2012), 第 77–86 页。[3] J. Bremer, Z. Gimbutas, and V. Rokhlin, *A nonlinear optimization procedure for generalized Gaussian quadratures* 32 2010 1761 1788 ,《SIAM J. Sci. Comput.》（），第 – 页。[4] J. Bremer, V. Rokhlin, and I. Sammis, *Universal quadratures for boundary integral equations on two-dimensional domains with corners* 229 2010 8259 8280 ,《J. Comput. Phys.》（），第 – 页。[5] P. D. Brubeck and L. N. Trefethen, *Lightning Stokes solver*,《SIAM J. Sci. Comp.》44 (2022), 第 A1205–A1226 页。[6] M. D. Buhmann,《径向基函数：理论与实现》，剑桥，2003。[7] G. Cybenko, *Approximation by superpositions of a sigmoidal function*,《Math. Control Signals Syst.》2 (1989), 第 303–314 页。[8] R. A. DeVore and K. Scherer, *Variable knot, variable degree spline approximation to x^β*，载于 R. A. DeVore 和 K. Scherer 编，《定量逼近》，Academic Press，1980，第 121–131 页。
[9] T. A Driscoll and B. Fornberg, *Interpolation in the limit of increasingly flat radial basis functions.* 43 2002 413 422 《Computers Math. w. Applics.》（），第 – 页。[10] G. E. Fasshauer,《MATLAB 无网格逼近方法》，World Scientific，2007。
[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*，MIT，2016。
[12] A. Gopal and L. N. Trefethen, *Solving Laplace problems with corner singularities via rational functions* 57 2019 2074 2094 ,《SIAM J. Numer. Anal.》（），第 – 页。[13] A. Gopal and L. N. Trefethen, *New Laplace and Helmholtz solvers*,《Proc. Nat. Acad. Sci.》116 (2019), 第 10223 页。[14] 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* 49 1986 613 657 ,,,《Numer. Math.》（），第 – 页。[15] B. Guo and I. Babuška, *The h-p version of the finite element method. I: The basic approximation results* 1 1986 21 41 ,《Comput. Mech.》（），第 – 页。[16] A. Herremans, D. Huybrechs, and L. N. Trefethen, *Resolution of singularities by rational functions* ,《SIAM J. Numer. Anal.》，已投稿。[17] K. Hornik, M. Stinchcombe, and H. White, *Multilayer feedforward networks are universal approximators* 2 1989 359 366 ,《神经网络》（），第 – 页。[18] H. Lebesgue, *Sur l'approximation des fonctions*,《Bull. Sci. Math.》22 (1898), 第 278–287 页。
[19] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*,《Nature》521 (2015), 第 436–444 页。
[20] J. M. Melenk, *hp-Finite Element Methods for Singular Perturbations*，Springer，2002。[21] H. N. Mhaskar, *Neural networks for optimal approximation of smooth and analytic functions*,《Neural Computation》8 (1996), 第 164–177 页。[22] H. N. Mhaskar and C. A. Micchelli, *Approximation by superposition of a sigmoidal function and radial basis functions* 13 1992 350 373 ,《Adv. Appl. Math.》（），第 – 页。[23] M. Mori and M. Sugihara, *The double-exponential transformation in numerical analysis*,《J. Comput. Appl. Math.》127 (2001), 第 287–296 页。[24] D. J. Newman, *Rational approximation to $|x|$*,《Mich. Math. J.》11 (1964), 第 11–14 页。[25] J. Park and I. W. Sandberg, *Universal approximation using radial basis function networks*,《Neural Computation》3 (1991), 第 246–257 页。
[26] A. Pinkus, *Approximation theory of the MLP model in neural networks*,《Acta Numer.》8 (1999), 第 143–195 页。
[27] K. Scherer, *On optimal global error bounds obtained by scaled local error estimates*,《Numer. Math.》36 (1981), 第 151–176 页。[28] C. Schwab, *p- and hp- Finite Element Methods: Theory and Application in Solid and Fluid Mechanics* 1998 ，Clarendon Press，牛津，。
[29] H. R. Stahl, *Poles and zeros of best rational approximants of $|x|$*,《Constr. Approx.》10 (1994), 第 469–522 页。
[30] G. Strang, *The functions of deep learning*,《SIAM News》51 (2018), 第 1–4 页。
[31] H. Takahasi and M. Mori, *Double exponential formulas for numerical integration*,《Publ. RIMS Kyoto》9 (1974), 第 721–741 页。[32] L. N. Trefethen, *Numerical analytic continuation*,《Japan J. Appl. Math.》，已投稿。
[33] L. N. Trefethen, Y. Nakatsukasa, and J. A. C. Weideman, *Exponential node clustering at singularities for rational approximation quadrature and PDEs* 147 2021 ,,,《Numer. Math.》（），第 227–254 页。