Approximating the pth root by composite rational functions

Evan S. Gawlik[a], Yuji Nakatsukasa[b]

[a]*Department of Mathematics, University of Hawaii at Manoa*
[b]*Mathematical Institute, University of Oxford, and National Institute of Informatics*

Abstract

A landmark result from rational approximation theory states that $x^{1/p}$ on $[0,1]$ can be approximated by a type-(n,n) rational function with root-exponential accuracy. Motivated by the recursive optimality property of Zolotarev functions (for the square root and sign functions), we investigate approximating $x^{1/p}$ by composite rational functions of the form $r_k(x, r_{k-1}(x, r_{k-2}(\cdots(x, r_1(x,1)))))$. While this class of rational functions ceases to contain the minimax (best) approximant for $p \geq 3$, we show that it achieves approximately pth-root exponential convergence with respect to the degree. Moreover, crucially, the convergence is *doubly exponential* with respect to the number of degrees of freedom, suggesting that composite rational functions are able to approximate $x^{1/p}$ and related functions (such as $|x|$ and the sector function) with exceptional efficiency.

Keywords: Rational approximation, function composition, minimax, Zolotarev, square root, pth root, sign function, sector function
2010 MSC: 41A20, 41A25, 65D15, 41A50

1. Introduction

Classical results in polynomial and rational approximation theory concern the convergence of an approximant to a given function f as the degree of the approximant grows. In this paper we focus on approximants to the pth root function $x^{1/p}$ that take a *composite* form. Composing rational functions is an efficient way of generating a rational function $r(x) = r_k(\cdots r_2(r_1(x)))$ of high degree: if each r_i is of type (m,m), then r is of type (m^k, m^k), even though it is expressed by a small number $O(mk)$ of degrees of freedom. By choosing each r_i appropriately, one can often obtain a function r that approximates a desired function in a wide domain of interest.

There is no reason to expect—and it is generally not true—that $r_k(\cdots r_2(r_1(x)))$ can express the minimax rational approximant of a given type, say (m^k, m^k), to a given function. However, there is a remarkable property of the minimax rational approximants to the function $\operatorname{sign}(x) = x/|x|$ on $[-1,-\delta] \cup [\delta, 1]$ for $0 < \delta < 1$ (called Zolotarev functions):

Email addresses: egawlik@hawaii.edu (Evan S. Gawlik), nakatsukasa@maths.ox.ac.uk (Yuji Nakatsukasa)

appropriately composing Zolotarev functions gives another Zolotarev function of higher degree. In other words, the class of composite rational functions $r(x) = r_k(\cdots r_2(r_1(x)))$, with each r_i of type (m, m), contains the type-(m^k, m^k) minimax approximant to the sign function [19, 15, 4, 14]. Moreover, for a fixed δ, the convergence of Zolotarev functions is exponential in the degree. Since the degree is m^k, and the number of parameters necessary to express r is $d \approx 2km$, it follows that the convergence is $\exp(-m^k) = \exp(-\exp(Cd))$, a *double-exponential* convergence rate. This paper is about approximating the pth root $x^{1/p}$, for which the above minimax optimality under composition does not hold. Nonetheless, our main result constructs composite rational approximants to $x^{1/p}$, which are linked to Zolotarev functions and inherit some of their properties [8].

Functions related to the sign function, such as $|x|$ (via $|x| = x/\mathrm{sign}(x)$) and $\sqrt{x}$ (via $|x| \approx p(x^2)/q(x^2)$ then $\sqrt{x} \approx p(x)/q(x)$) can similarly be approximated by composite rational functions. Gawlik [7] does this for the square root and shows that a composite rational function yields the minimax rational approximant (in the relative sense) on intervals $[\delta, 1] \subset (0, 1]$, and that the approximation extends far into the complex plane. This observation generalizes earlier work on rational approximation of the square root with optimally scaled Newton iterations [2, 15, 19, 23]. Moreover, an extension was derived in [8], which shows that the pth root can be approximated efficiently on intervals $[\delta, 1] \subset (0, 1]$, although not with minimax quality.

Clearly, in the above papers the origin is excluded from the domain, as the functions have a singularity at $x = 0$. However, a landmark result from rational approximation theory [9, 21] states that x^β (for any real $\beta > 0$) on $[0, 1]$ can be approximated (in the absolute sense) by a type-(n, n) rational function with root-exponential accuracy. One might wonder, can this be done with a composite rational function? This is the question we address in this paper. We focus on the case in which $\beta = 1/p$ with $p \geq 2$ an integer.

We show that a rational function of the form $r(x) = r_k(x, r_{k-1}(x, r_{k-2}(\cdots(x, r_1(x, 1)))))$ can approximate $x^{1/p}$ on $[0, 1]$ with superalgebraic accuracy, with close to pth root-exponential convergence. Moreover—and crucially—the convergence is *doubly exponential* with respect to the number of degrees of freedom. That is, the error is $O(\exp(-c_1 \exp(c_2 d)))$ for some constants $c_1, c_2 > 0$, where d is the number of parameters needed to express the rational function. By "number of parameters" we mean $d = \sum_{i=1}^{k} m_i + \ell_i + 1$ if r_i has type (m_i, ℓ_i) for $i = 1, 2, \ldots, k$, so that d reflects the cost of evaluating r at a matrix argument.

Clearly, our result implies that any rational power of x can be approximated by a composite rational function. Moreover, since $|r(x) - x^{1/p}| \leq \epsilon$ on $[0, 1]$ implies $|r(x/s) - (x/s)^{1/p}| \leq \epsilon$ on $[0, s]$ for any $s > 0$, hence $|s^{1/p} r(x/s) - x^{1/p}| \leq s^{1/p}\epsilon$, our results also show that any rational power can be approximated efficiently on $[0, s]$ by a composite rational function. In addition, our approximants to $x^{1/p}$ immediately lead to approximants to the p-sector function $\mathrm{sect}_p(z) = z/(z^p)^{1/p}$, which takes the value $z/|z|$ on the p line segments $\exp(2\pi \mathrm{i} j/p)[0, \infty)$, $j = 0, 1, \ldots, p-1$. Such approximants lead to algorithms for the matrix sector function, which has been used in systems theory [20]. They can also form the basis of a spectral divide-and-conquer method for computing a matrix eigenvalue decomposition, generalizing the algorithm in [14] for matrices with nonreal eigenvalues.

Summary of Results. To summarize our results, let us introduce some terminology. We say that a univariate rational function $r(x) = p(x)/q(x)$ is of type (m, ℓ) if p and q are polynomials of degrees at most m and ℓ, respectively. We denote the set of all such rational functions by $\mathcal{R}_{m,\ell}$. We say that a bivariate rational function $r(x, y)$ is of type (m, ℓ) if $r(x, x)$ is of type (m, ℓ). We say that a univariate rational function r is (k, m, ℓ)-composite if r is a composition of k rational functions $r_i(x, y)$, $i = 1, 2, \ldots, k$, each of type (m, ℓ):

$$r(x) = r_k(x, r_{k-1}(x, r_{k-2}(\cdots(x, r_1(x, 1))))). \tag{1}$$

In this paper we only deal with two cases, $(m, \ell) = (p, p-1)$ and $(m, \ell) = (1, p)$. Here is the main result of this paper.

Theorem 1.1. *Let $p \geq 2$ be an integer. There exists a positive constant N depending on p such that for every integer $n \geq N$, there exists a $(\lfloor \log_p n \rfloor + 1, p, p-1)$-composite rational function r of type $(n, n-1)$ such that*

$$\max_{x \in [0,1]} |r(x) - x^{1/p}| \leq 2 \exp(-bn^c), \tag{2}$$

where $b = \frac{1}{p}$ and

$$c = \frac{\log\left(\frac{p}{p-1}\right) \log 2}{\log\left(\frac{2p}{p-1}\right) \log p}. \tag{3}$$

Note that when $p = 2$, $c = \frac{1}{2}$, and as $p \to \infty$, $c \sim \frac{1}{p \log p}$.

Let us comment on the theorem. The bound (2) shows that by using a $(\lfloor \log_p n \rfloor + 1, p, p-1)$-composite rational function we can approximate the pth root with "$1/c$th root"-(nearly pth root) exponential accuracy with respect to the degree, which is suboptimal unless $p = 2$ (in which case a composite rational function on $[\delta, 1]$ is optimal in the relative sense).

However, the result is still striking in the following sense: the number of degrees of freedom used to express r is just $O(pk)$ for $n \approx p^k$ (see (13)-(14)), and therefore with respect to the degrees of freedom d, the convergence is

$$\max_{x \in [0,1]} |r(x) - x^{1/p}| \leq 2 \exp(-bp^{\tilde{c}d}), \tag{4}$$

indicating a *double-exponential* convergence with respect to d.

As a byproduct of our analysis, we will obtain analogous results for composite rational approximation of the p-sector function $\mathrm{sect}_p(z) = z/(z^p)^{1/p}$ on the set $S_p \subset \mathbb{C}$ given by

$$S_p = \{xe^{2\pi ij/p} \mid x \in [0, 1],\ j \in \{1, 2, \ldots, p\}\}. \tag{5}$$

We will also consider the subset $S_{p,\alpha}$ of S_p excluding the origin

$$S_{p,\alpha} = \{xe^{2\pi ij/p} \mid x \in [\alpha, 1],\ j \in \{1, 2, \ldots, p\}\}. \tag{6}$$

We say that a (k, m, ℓ)-composite rational function (1) is *pure* if the functions $r_j(x, y)$ appearing in (1) are univariate:

$$r(x) = r_k(r_{k-1}(\cdots(r_1(x)))).$$

Corollary 1.1. *Let $p \geq 2$ be an integer and $\alpha \in (0,1)$. There exists a positive constant N depending on p such that for every integer $n \geq N$, there exist pure $(\lfloor \log_p n \rfloor, 1, p)$-composite rational functions r and q of type $(n-p+1, n)$ such that*

$$\max_{z \in S_p} |z\,(r(z) - \operatorname{sect}_p(z))| \leq 2\exp(-bn^c), \tag{7}$$

where b and c are as in Theorem 1.1, and

$$\max_{z \in S_{p,\alpha}} |q(z) - \operatorname{sect}_p(z)| \leq \widehat{a}\exp(-\widehat{b}n^{\widehat{c}}), \tag{8}$$

where $\widehat{a} = \frac{p}{(p-1)^2}$, $\widehat{b} = \log 2 \left(\frac{\log \frac{p}{p-1}}{\log \frac{1}{\alpha}}\right)^{\frac{\log 2}{\log \frac{p}{p-1}}}$ and $\widehat{c} = \frac{\log 2}{\log p}$.

It is worth noting that the two rational functions r, q are generally different—they coincide for a particular value of α. The error in (7) is measured in a weighted norm, which is natural in view of the fact that $\operatorname{sect}_p(z)$ is discontinuous at $z = 0$. When $p = 2$ and $z \in S_2$, $z \operatorname{sect}_p(z) = |z|$ and $c = \frac{1}{2}$, so (7) recovers the root-exponential convergence of rational approximants to $|x|$ on $[-1,1]$ [22, Ch. 25]. By contrast, (8) shows that a better bound holds for the absolute error if one excludes the neighborhood of the origin. When $p = 2$, $\widehat{c} = 1$ and (8) recovers the exponential convergence of Zolotarev functions to the sign function on $[-1,-\alpha] \cup [\alpha, 1]$ [1, 3]. Note that $\widehat{b}$ decays like a negative power of $\log \frac{1}{\alpha}$ as $\alpha \to 0$.

Organization. This paper is organized as follows. In Section 2, we review some theory from [8] concerning composite rational approximants of the pth root on positive real intervals. In Section 3, we study the behavior of these approximants near the origin. We then prove Theorem 1.1 and Corollary 1.1 in Section 4, and we illustrate our results numerically in Section 5.

2. Composite rational approximation of the pth root

To approximate $x^{1/p}$ on an interval $[\alpha^p, 1] \subset (0,1]$, Gawlik [8] considers the recursively defined rational function

$$f_{k+1}(x) = f_k(x)\hat{r}_{m,\ell}\left(\frac{x}{f_k(x)^p}, \alpha_k, \sqrt[p]{\cdot}\right), \qquad f_0(x) = 1, \tag{9}$$

$$\alpha_{k+1} = \frac{\alpha_k}{\hat{r}_{m,\ell}\left(\alpha_k^p, \alpha_k, \sqrt[p]{\cdot}\right)}, \qquad \alpha_0 = \alpha, \tag{10}$$

where $\hat{r}_{m,\ell}(x, \alpha, \sqrt[p]{\cdot})$ is (a rescaling of) the *relative* minimax rational approximant of type $(m,\ell) \in \mathbb{N}_0 \times \mathbb{N}_0 \setminus \{(0,0)\}$ on the interval $[\alpha^p, 1]$:

$$\hat{r}_{m,\ell}(x, \alpha, \sqrt[p]{\cdot}) = \left(\frac{1+\alpha}{2\alpha}\right) r_{m,\ell}(x, \alpha, \sqrt[p]{\cdot}),$$

where

$$r_{m,\ell}(\,\cdot\,,\alpha,\sqrt[p]{\cdot}\,) = \operatorname*{arg\,min}_{r\in\mathcal{R}_{m,\ell}} \max_{x\in[\alpha^p,1]} \left|\frac{r(x)-x^{1/p}}{x^{1/p}}\right|. \tag{11}$$

Note that f_k is a composite rational function of the form (1). Gawlik shows that $f_k(x)$ is a rapidly convergent approximant to the pth root on $[\alpha^p, 1]$. With k recursions, the maximum relative error $|f_k(x) - x^{1/p}|/|x^{1/p}|$ on $[\alpha^p, 1]$ decays double exponentially in k: it is bounded above by $c_1 \exp(-c_2(m+\ell+1)^k)$ for some $c_1, c_2 > 0$ depending on m, ℓ, p, and α. Importantly, these constants depend very weakly on α; the analysis below will implicitly show that when $(m,\ell) = (1,0)$, c_1 is independent of α and c_2 decays like a negative power of $\log\frac{1}{\alpha}$ as $\alpha \to 0$, just like $\widehat{b}$ in (8).

Given that (9) is an approximant on $[\alpha^p, 1]$, which is an interval that excludes the singularity at $x = 0$, a natural question arises: can we approximate on $[0, 1]$? Intuitively, the function is still continuous at $x = 0$ (unlike e.g. the sign or sector function) with $0^{1/p} = 0$, and hence it is possible to approximate $x^{1/p}$ on the whole interval $[0, 1]$. Indeed Stahl [21] shows that $x^{1/p}$ on $[0, 1]$ can be approximated by a type-(n, n) rational function with root-exponential accuracy (we refer to [6, 16] for general results on classical rational approximation theory). Can a highly efficient rational approximant be constructed based on recursion as in (9)? It is important to note that we will necessarily switch to the (more natural) metric of absolute error $|r(x) - x^{1/p}|$ rather than the relative error $|r(x) - x^{1/p}|/|x^{1/p}|$ for this purpose.

It turns out that the rational function (9) does a good job approximating on $[0, 1]$, when α is chosen carefully: when it is too small, the error is large on $[\alpha^p, 1]$ (in fact it is maximal at $x = 1$ [8]). Conversely if α is too large, the error is large on $[0, \alpha^p]$ (in fact it is $O(\alpha)$ at $x = 0$, as we show below). A major task undertaken in what follows is to choose α so that the convergence is optimized, in that the error on $[0, \alpha^p]$ and $[\alpha^p, 1]$ are balanced to be approximately the same.

Our analysis will focus on the lowest-order version of the iteration (9-10), obtained by choosing $(m,\ell) = (1, 0)$. It is shown in [8, Proposition 5] (and elsewhere [11, 13]) that for this choice of m and ℓ,

$$\hat{r}_{1,0}(x,\alpha,\sqrt[p]{\cdot}\,) = \frac{1}{p}\left((p-1)\mu(\alpha) + \frac{x}{\mu(\alpha)^{p-1}}\right), \qquad \mu(\alpha) = \left(\frac{\alpha-\alpha^p}{(p-1)(1-\alpha)}\right)^{1/p}. \tag{12}$$

Thus, when $(m,\ell) = (1, 0)$, the iteration (9-10) reads

$$f_{k+1}(x) = \frac{1}{p}\left((p-1)\mu(\alpha_k) f_k(x) + \frac{x}{\mu(\alpha_k)^{p-1} f_k(x)^{p-1}}\right), \qquad f_0(x) = 1, \tag{13}$$

$$\alpha_{k+1} = \frac{p\alpha_k}{(p-1)\mu(\alpha_k) + \mu(\alpha_k)^{1-p}\alpha_k^p}, \qquad \alpha_0 = \alpha. \tag{14}$$

Note that f_k is $(k, p, p-1)$-composite since it is of the form (1) with

$$r_j(x,y) = \frac{1}{p}\left(\frac{(p-1)\mu(\alpha_{j-1})^p y^p + x}{\mu(\alpha_{j-1})^{p-1} y^{p-1}}\right)$$

for each j. It follows from this observation and an inductive argument that f_k has type $(p^{k-1}, p^{k-1}-1)$ for each $k \geq 1$.

We rely heavily on this explicit expression for the particular case $(m, \ell) = (1, 0)$, as it lets us analyze the functions in detail, which leads to a constructive proof for Theorem 1.1. The iteration (13-14) can be regarded as a scaled version of Newton's iteration [11, 13], which is commonly employed for matrix square roots [2],[10, Ch. 6]. We note that using larger values of (m, ℓ) may result in faster convergence, in particular a larger exponent c than (3). In view of (4), the convergence is still doubly exponential, with an improved constant $\tilde{c}$. However, we do not expect the improvement would be significant.

Moreover, composing low-degree rational functions is an extremely efficient way to construct high-degree rational functions of matrices, and we suspect that our choice $(m, \ell) = (1, 0)$ would give the fastest convergence in terms of the number of matrix operations needed to evaluate r at a matrix argument.

3. Bounding the error on $[0, \alpha^p]$

In this section, we analyze the absolute error committed by the function f_k defined by (13)–(14) on the interval $[0, \alpha^p]$. It will be convenient to consider not f_k but the scaled function

$$\widetilde{f}_k(x) = \frac{2\alpha_k}{1+\alpha_k} f_k(x), \tag{15}$$

which has the property that [8, Theorem 2]

$$\max_{x\in[\alpha^p,1]} \frac{\widetilde{f}_k(x) - x^{1/p}}{x^{1/p}} = -\min_{x\in[\alpha^p,1]} \frac{\widetilde{f}_k(x) - x^{1/p}}{x^{1/p}} = \frac{1-\alpha_k}{1+\alpha_k} \in (0,1). \tag{16}$$

We will prove the following estimate.

Theorem 3.1. *Let $\alpha \in (0,1)$. The function $\widetilde{f}_k$ defined by (13)–(14) and (15) satisfies*

$$\max_{x\in[0,\alpha^p]} |\widetilde{f}_k(x) - x^{1/p}| \leq 2\alpha \tag{17}$$

for every $k \geq 0$.

Experiments suggest that the bound (17) could be improved to $< \alpha$ for k large enough, but this does not affect what follows in any significant way.

We will prove Theorem 3.1 by a series of lemmas. Let

$$g_k(x) = \frac{x}{f_k(x^p)}.$$

Note that $g_0(x) = x$ and

$$g_{k+1}(x) = \frac{x}{f_k(x^p)\hat{r}_{1,0}\left(\frac{x^p}{f_k(x^p)^p}, \alpha_k, \sqrt[p]{\cdot}\right)} = \frac{g_k(x)}{\hat{r}_{1,0}(g_k(x)^p, \alpha_k, \sqrt[p]{\cdot})} = \hat{s}(g_k(x), \alpha_k), \tag{18}$$

where

$$\hat{s}(x,\alpha) = \frac{x}{\hat{r}_{1,0}(x^p, \alpha, \sqrt[p]{\cdot})} = \frac{px}{(p-1)\mu(\alpha) + \mu(\alpha)^{1-p}x^p}.$$

Also let

$$H(\alpha) = \hat{s}(\alpha,\alpha) = \frac{p\alpha}{(p-1)\mu(\alpha) + \mu(\alpha)^{1-p}\alpha^p},$$

so that $\alpha_{k+1} = H(\alpha_k)$.

Lemma 3.1. *For every $\alpha \in (0,1)$ and every $x \in [0,\alpha]$,*

$$0 \le x\hat{s}'(x,\alpha) \le \hat{s}(x,\alpha) \le H(\alpha),$$

where the prime denotes differentiation with respect to x.

Proof. A short calculation shows that

$$x\hat{s}'(x,\alpha) = w(x)\hat{s}(x,\alpha),$$

where

$$w(x) = \frac{(p-1)\left(1 - \left(\frac{x}{\mu(\alpha)}\right)^p\right)}{(p-1) + \left(\frac{x}{\mu(\alpha)}\right)^p}.$$

Since $0 \le w(x) \le 1$ for every $x \in [0,\mu(\alpha)]$, it follows that

$$0 \le x\hat{s}'(x,\alpha) \le \hat{s}(x,\alpha), \quad x \in [0,\mu(\alpha)].$$

In particular, the above inequalities hold on $[0,\alpha] \subset [0,\mu(\alpha)]$, and $\hat{s}(x,\alpha)$ is nondecreasing on $[0,\alpha]$. Thus,

$$\hat{s}(x,\alpha) \le \hat{s}(\alpha,\alpha) = H(\alpha), \quad x \in [0,\alpha].$$

□

Now let $\alpha \in (0,1)$ be fixed.

Lemma 3.2. *For every $x \in [0,\alpha]$ and every $k \ge 0$,*

$$0 \le xg_k'(x) \le g_k(x) \le \alpha_k.$$

Proof. Since $g_0(x) = x$ and $\alpha_0 = \alpha$, the above inequalities hold when $k = 0$. Assume that they hold for some $k \ge 0$. Observe that

$$xg_{k+1}'(x) = xg_k'(x)\hat{s}'(g_k(x),\alpha_k).$$

Since $g_k(x) \in [0,\alpha_k]$ for $x \in [0,\alpha]$, Lemma 3.1 implies that $\hat{s}'(g_k(x),\alpha_k) \ge 0$. It follows from this and our inductive hypothesis that $xg_{k+1}'(x) \ge 0$ for $x \in [0,\alpha]$. In addition, since $xg_k'(x) \le g_k(x)$ and $g_k(x)\hat{s}'(g_k(x),\alpha_k) \le \hat{s}(g_k(x),\alpha_k)$,

$$xg_{k+1}'(x) \le \hat{s}(g_k(x),\alpha_k) = g_{k+1}(x).$$

Finally, since $\hat{s}(g_k(x),\alpha_k) \le H(\alpha_k) = \alpha_{k+1}$, it follows that $g_{k+1}(x) \le \alpha_{k+1}$. □

Lemma 3.3. *For every $x \in [0, \alpha^p]$ and every $k \geq 0$,*

$$0 < \widetilde{f}_k(x) \leq \alpha(1 + \varepsilon_k), \quad \varepsilon_k = \frac{1 - \alpha_k}{1 + \alpha_k}.$$

Proof. We first note that f_k is positive and nondecreasing on $[0, \alpha^p]$. Indeed, differentiating the relation

$$f_k(x^p) = \frac{x}{g_k(x)}$$

gives

$$px^{p-1} f_k'(x^p) = \frac{g_k(x) - x g_k'(x)}{g_k(x)^2},$$

so Lemma 3.2 implies that $f_k'(x^p) \geq 0$ for every $x \in [0, \alpha]$. Evaluating the recursion (13) at $x = 0$ gives

$$f_{k+1}(0) = f_k(0) \left(\frac{p-1}{p} \right) \mu(\alpha_k), \quad f_0(0) = 1,$$

so $f_k(0) > 0$ for every k. Since $\widetilde{f}_k(x)$ is a positive multiple of $f_k(x)$, it follows that $0 < \widetilde{f}_k(x) \leq \widetilde{f}_k(\alpha^p)$ for every $x \in [0, \alpha^p]$. Finally, taking $x = \alpha^p$ in (16) gives $\widetilde{f}_k(\alpha^p) \leq \alpha(1 + \varepsilon_k)$. □

By the lemma above,

$$|\widetilde{f}_k(x) - x^{1/p}| \leq \max\{|\widetilde{f}_k(x)|, |x^{1/p}|\} \leq \max\{\alpha(1 + \varepsilon_k), \alpha\} = \alpha(1 + \varepsilon_k) \leq 2\alpha, \quad x \in [0, \alpha^p],$$

so

$$\max_{x \in [0, \alpha^p]} |\widetilde{f}_k(x) - x^{1/p}| \leq 2\alpha.$$

This completes the proof of Theorem 3.1.

An estimate for the absolute error on $[0, 1]$ is now immediate: Combining the above theorem, (16), and the fact that $x^{1/p} \leq 1$ for $x \in [0, 1]$, we see that

$$\max_{x \in [0,1]} |\widetilde{f}_k(x) - x^{1/p}| \leq \max \left\{ 2\alpha, \frac{1 - \alpha_k}{1 + \alpha_k} \right\}. \tag{19}$$

3.1. Sector function approximation

We note that the function g_k in (18) approximates the p-sector function $\mathrm{sect}_p(z) = z/(z^p)^{1/p}$ (this observation appeared in [8, Sec. 4]), and g_k is a *pure* composite rational function of the form $g_k(z) = r_k(r_{k-1}(\cdots r_2(r_1(z))))$. In fact it is $(k, 1, p)$-composite, and an inductive argument shows that it has type $(p^k - p + 1, p^k)$. In the $p = 2$ case, this reduces to Zolotarev's best rational approximant to the sign function of type $(2^k - 1, 2^k)$. That is, as in the square root approximation, the minimax rational approximant is contained in the class of (here purely) composite rational functions.

Below we derive estimates for the maximum weighted error $|z(g_k(z) - \mathrm{sect}_p(z))|$ on the sets $S_p, S_{p,\alpha} \subset \mathbb{C}$ defined in (5) and (6). As before, it will be convenient to work not with $g_k(z)$ but with the rescaled function

$$\widetilde{g}_k(z) = \frac{2}{1 + \alpha_k} g_k(z) = \frac{4\alpha_k}{(1 + \alpha_k)^2} \frac{z}{\widetilde{f}_k(z^p)}.$$

As shown in [8, Sec. 4], the relative error $\frac{\widetilde{g}_k(z)-\mathrm{sect}_p(z)}{\mathrm{sect}_p(z)}$ is real-valued and equioscillates on each line segment $\{z \in \mathbb{C} \mid e^{-2\pi ij/p} z \in [\alpha, 1]\}$, $j = 0, 1, \ldots, p-1$. Note that here the relative and absolute errors are the same in modulus. The asymptotic convergence rate on $S_{p,\alpha}$ was analyzed in [8]. Here we quantify the non-asymptotic convergence on S_p.

Lemma 3.4. *For every $k \geq 0$,*

$$\max_{z \in S_p} |z(\widetilde{g}_k(z) - \mathrm{sect}_p(z))| \leq \max\left\{\alpha, \frac{1-\alpha_k}{1+\alpha_k}\right\}, \tag{20}$$

and

$$\max_{z \in S_{p,\alpha}} |\widetilde{g}_k(z) - \mathrm{sect}_p(z)| \leq \frac{1-\alpha_k}{1+\alpha_k}. \tag{21}$$

Proof. Let $z = x^{1/p} e^{2\pi ij/p}$ with $x \in [0,1]$ and $j \in \{1, \ldots, p\}$. Since $\widetilde{g}_k(z) = e^{2\pi ij/p} \widetilde{g}_k(x^{1/p})$ and $\mathrm{sect}_p(z) = e^{2\pi ij/p}$, we have

$$|z(\widetilde{g}_k(z) - \mathrm{sect}_p(z))| = |x^{1/p}(\widetilde{g}_k(x^{1/p}) - 1)|.$$

If $x \in [0, \alpha^p]$, then Lemma 3.2 implies that $0 \leq \widetilde{g}_k(x^{1/p}) \leq \frac{2\alpha_k}{1+\alpha_k} < 1$, so

$$|x^{1/p}(\widetilde{g}_k(x^{1/p}) - 1)| \leq x^{1/p} \leq \alpha, \quad x \in [0, \alpha^p].$$

On the other hand, if $x \in [\alpha^p, 1]$, then

$$|x^{1/p}(\widetilde{g}_k(x^{1/p}) - 1)| \leq |\widetilde{g}_k(x^{1/p}) - 1| = \left| \frac{4\alpha_k}{(1+\alpha_k)^2} \frac{x^{1/p}}{\widetilde{f}_k(x)} - 1 \right|. \tag{22}$$

By (16),

$$\frac{\widetilde{f}_k(x)}{x^{1/p}} \in \left[1 - \left(\frac{1-\alpha_k}{1+\alpha_k}\right), 1 + \left(\frac{1-\alpha_k}{1+\alpha_k}\right)\right] = \left[\frac{2\alpha_k}{1+\alpha_k}, \frac{2}{1+\alpha_k}\right], \quad x \in [\alpha^p, 1],$$

so

$$\frac{x^{1/p}}{\widetilde{f}_k(x)} \in \left[\frac{1+\alpha_k}{2}, \frac{1+\alpha_k}{2\alpha_k}\right], \quad x \in [\alpha^p, 1],$$

and hence

$$\frac{4\alpha_k}{(1+\alpha_k)^2} \frac{x^{1/p}}{\widetilde{f}_k(x)} - 1 \in \left[-\frac{1-\alpha_k}{1+\alpha_k}, \frac{1-\alpha_k}{1+\alpha_k}\right], \quad x \in [\alpha^p, 1]. \tag{23}$$

It follows that

$$|x^{1/p}(\widetilde{g}_k(x^{1/p}) - 1)| \leq \frac{1-\alpha_k}{1+\alpha_k}, \quad x \in [\alpha^p, 1].$$

For (21), we simply start from the second expression in (22) and use (23). □

4. Proof of Theorem 1.1 and Corollary 1.1

To examine the convergence of the recursion (13)-(14) on $[0,1]$, we first ask the question: given $\epsilon > 0$, what values of k and α are needed to get an error bounded by ϵ? In view of (19), we must choose $\alpha \leq \epsilon/2$ and k large enough so that $\frac{1-\alpha_k}{1+\alpha_k} \leq \epsilon$.

To determine k, it is convenient to define

$$\delta_k = \left(\frac{p-1}{2}\right)\left(\left(\frac{1}{\alpha_k}\right)^{1-1/p} - 1\right).$$

Lemma 4.1. *For every $k \geq 0$,*

$$\frac{1-\alpha_k}{1+\alpha_k} < \frac{p}{(p-1)^2}\delta_k.$$

Proof. The function $q(\alpha) = \frac{p}{2(p-1)}\left(\left(\frac{1}{\alpha}\right)^{1-1/p} - 1\right) - \frac{1-\alpha}{1+\alpha}$ satisfies $q(1) = 0$ and $q'(\alpha) = \frac{2}{(1+\alpha)^2} - \frac{1}{2\alpha^{2-1/p}} < 0$ for $\alpha \in (0,1)$, so $q(\alpha) > 0$ for $\alpha \in (0,1)$. In particular, $q(\alpha_k) > 0$. □

The above lemma shows that we can ensure $\frac{1-\alpha_k}{1+\alpha_k} \leq \epsilon$ by enforcing $\delta_k \leq (p-1)^2\epsilon/p$. Let us do so by selecting a scalar $\delta^* \in (0,1)$ and splitting the convergence of $\delta_k \to 0$ into two stages:

1. Find k_1 such that $\delta_{k_1} \leq \delta^*$.
2. Find k_2 such that $\delta_{k_1+k_2} \leq (p-1)^2\epsilon/p$.

Stage 1. We begin our analysis of the first stage with a lemma.

Lemma 4.2. *For every $\alpha \in (0,1)$,*

$$H(\alpha) > \alpha^{1-1/p}.$$

Proof. We have

$$\begin{aligned} H(\alpha) &= \frac{p\alpha\mu(\alpha)^{p-1}}{(p-1)\mu(\alpha)^p + \alpha^p} = \frac{p\alpha\mu(\alpha)^{p-1}}{\frac{\alpha-\alpha^p}{1-\alpha} + \alpha^p} = \frac{p\alpha\mu(\alpha)^{p-1}(1-\alpha)}{\alpha - \alpha^{p+1}} \\ &= \frac{p\mu(\alpha)^{p-1}(1-\alpha)}{1-\alpha^p} = \alpha^{1-1/p}\frac{g(\alpha)^{1-1/p}}{h(\alpha)}, \end{aligned}$$

where

$$\begin{aligned} g(\alpha) &= \frac{1-\alpha^{p-1}}{(p-1)(1-\alpha)} = \frac{1}{p-1}\sum_{j=0}^{p-2}\alpha^j, \\ h(\alpha) &= \frac{1-\alpha^p}{p(1-\alpha)} = \frac{1}{p}\sum_{j=0}^{p-1}\alpha^j. \end{aligned}$$

Since $0 < h(\alpha) < g(\alpha) < 1$ for every $\alpha \in (0,1)$, it follows that

$$\frac{g(\alpha)^{1-1/p}}{h(\alpha)} > \frac{g(\alpha)}{h(\alpha)} > 1.$$

□

Lemma 4.2 implies

$$\alpha_{k+1} \geq \alpha_k^{1-1/p} \tag{24}$$

for every k, so

$$\alpha_k \geq \alpha^{(1-1/p)^k}.$$

Thus, we will have $\delta_{k_1} \leq \delta^*$ if $\alpha^{(1-1/p)^{k_1}} \geq \left(\frac{p-1}{p-1+2\delta^*}\right)^{p/(p-1)}$, which means

$$k_1 \geq \frac{\log\log\frac{1}{\alpha} - \log\log\frac{p-1+2\delta^*}{p-1}}{\log\frac{p}{p-1}} - 1. \tag{25}$$

Stage 2. Next we determine k_2 such that $\delta_{k_1+k_2} \leq (p-1)^2\epsilon/p$.

Lemma 4.3. *For any $t \geq 1$,*

$$\frac{p-1}{p}t + \frac{1}{p}\frac{1}{t^{p-1}} - 1 \leq \frac{p-1}{2}(t-1)^2.$$

Proof. The function $q(t) = \frac{p-1}{p}t + \frac{1}{p}\frac{1}{t^{p-1}} - 1 - \frac{p-1}{2}(t-1)^2$ satisfies $q(1) = q'(1) = 0$ and $q''(t) = (p-1)(t^{-p-1} - 1) < 0$ for $t > 1$, so $q(t) \leq 0$ for $t \geq 1$. □

Lemma 4.4. *For every $k \geq 0$,*

$$\delta_{k+1} \leq \delta_k^2.$$

Proof. The recursion (14) can be written as

$$\frac{1}{\alpha_{k+1}} = \frac{p-1}{p}t_k + \frac{1}{p}\frac{1}{t_k^{p-1}}, \quad t_k = \frac{\mu(\alpha_k)}{\alpha_k}.$$

Using Lemma 4.3 and the observation that $\mu(\alpha_k) = \left(\frac{1}{p-1}\sum_{j=1}^{p-1}\alpha_k^j\right)^{1/p} \leq \alpha_k^{1/p} \implies t_k \leq (1/\alpha_k)^{1-1/p}$, we see that

$$\begin{aligned}
\left(\frac{1}{\alpha_{k+1}}\right)^{1-1/p} - 1 &\leq \frac{1}{\alpha_{k+1}} - 1 \\
&= \frac{p-1}{p}t_k + \frac{1}{p}\frac{1}{t_k^{p-1}} - 1 \\
&\leq \frac{p-1}{2}(t_k - 1)^2 \\
&\leq \frac{p-1}{2}\left(\left(\frac{1}{\alpha_k}\right)^{1-1/p} - 1\right)^2.
\end{aligned}$$

Multiplying by $(p-1)/2$ yields $\delta_{k+1} \leq \delta_k^2$. □

By the results above, we will have $\delta_{k_1+k_2} \le (p-1)^2\epsilon/p$ if $(\delta^*)^{2^{k_2}} \le (p-1)^2\epsilon/p$, which means

$$k_2 \ge \frac{\log\log\frac{p}{(p-1)^2\epsilon} - \log\log\frac{1}{\delta^*}}{\log 2}.$$

Finally, by taking $\alpha = \epsilon/2$ we ensure that the error on $[0, \alpha^p]$ is bounded by ϵ (recall (19)), so the error on $[0, 1]$ is bounded by ϵ.

We illustrate the process in Figure 1, where we fix integers[1] p and k, and numerically find the value of $\alpha \in (0,1)$ and accordingly $\epsilon = \frac{1-\alpha_k}{1+\alpha_k} = 2\alpha$ such that with the $(k, p, p-1)$-composite rational approximant $\widetilde{f}_k$ the error is $\max_{x\in[\alpha^p,1]} |x^{1/p} - \widetilde{f}_k(x)| \le \epsilon$, achieved at $x = 1$, and the error on $[0, \alpha^p]$ is bounded by ϵ. Observe that the maximum errors on $[0, \alpha^p]$ and $[\alpha^p, 1]$ are not equal but of the same order, suggesting the near optimality of our composite rational approximants.

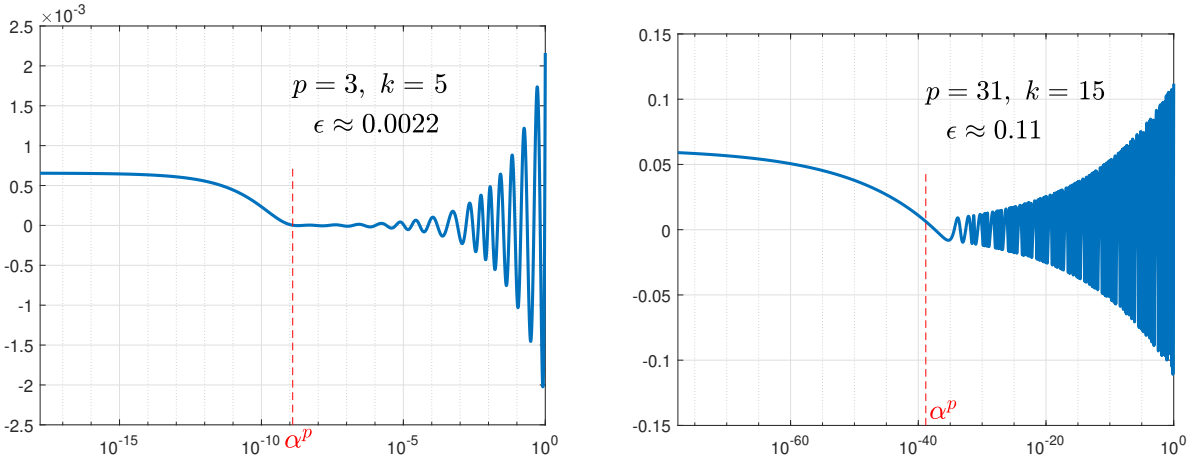

Figure 1: Error curves $\widetilde{f}_k(x) - x^{1/p}$. Note that the error on $[0, \alpha^p]$ is bounded by that on $[\alpha^p, 1]$, which is $\epsilon = 2\alpha$ in both cases.

Putting the above inequalities together, we conclude that

$$k = \frac{\log\log\frac{2}{\epsilon}}{\log\frac{p}{p-1}} + \frac{\log\log\frac{p}{(p-1)^2\epsilon}}{\log 2} + k_0 \tag{26}$$

recursions are enough to yield accuracy ϵ, where k_0 satisfies

$$k_0 \le -\frac{\log\log\frac{p-1+2\delta^*}{p-1}}{\log\frac{p}{p-1}} - \frac{\log\log\frac{1}{\delta^*}}{\log 2}. \tag{27}$$

Since k recursions translate into a rational function $\widetilde{f}_k$ of type $(p^{k-1}, p^{k-1}-1)$, it follows that the degree n of the rational function $\widetilde{f}_k$ achieving accuracy ϵ is

$$n = p^{\frac{\log\log\frac{2}{\epsilon}}{\log\frac{p}{p-1}} + \frac{\log\log\frac{p}{(p-1)^2\epsilon}}{\log 2} + k_0 - 1}.$$

[1] $p = 31$ is a somewhat arbitrary prime number, chosen in view of the number of days per month.

We rewrite this to express the error with respect to the degree n. Taking the logarithm and setting $\widetilde{k}_0 = k_0 - 1$, we get

$$\log n = \left(\frac{\log\log \frac{2}{\epsilon}}{\log \frac{p}{p-1}} + \frac{\log\log \frac{p}{(p-1)^2\epsilon}}{\log 2} + \widetilde{k}_0 \right) \log p \leq \left(\frac{\log\log \frac{2}{\epsilon}}{\log \frac{p}{p-1}} + \frac{\log\log \frac{2}{\epsilon}}{\log 2} + \widetilde{k}_0 \right) \log p. \quad (28)$$

Hence,

$$\log\log \frac{2}{\epsilon} \geq \frac{\log n - \widetilde{k}_0 \log p}{\log p(\frac{1}{\log \frac{p}{p-1}} + \frac{1}{\log 2})}.$$

Thus, defining

$$c := \frac{1}{\log p(\frac{1}{\log \frac{p}{p-1}} + \frac{1}{\log 2})} = \frac{\log 2 \log \frac{p}{p-1}}{\log p \log \frac{2p}{p-1}}, \quad (29)$$

we have

$$\log \frac{2}{\epsilon} \geq \left(\frac{n}{p^{\widetilde{k}_0}} \right)^c,$$

and therefore, writing $\widetilde{b} = 1/p^{c\widetilde{k}_0}$, we arrive at

$$\epsilon \leq 2\exp(-\widetilde{b} n^c).$$

This bound holds when n is a sufficiently large power of p. To handle the case in which $n \in \mathbb{N}$ is not a power of p, we note that $\lfloor \log_p n \rfloor + 1$ recursions yield a rational function of type $(p^{\lfloor \log_p n \rfloor}, p^{\lfloor \log_p n \rfloor} - 1)$, and for n large enough this function has error bounded above by

$$2\exp(-\widetilde{b}(p^{\lfloor \log_p n \rfloor})^c) \leq 2\exp(-\widetilde{b} p^{-c} n^c).$$

We will complete the proof of Theorem 1.1 by finding a lower bound for $\widetilde{b}$.

Lemma 4.5. *If $\delta^* = \frac{1}{2}$, then*

$$\widetilde{b} \geq \frac{1}{p^{1-c}}.$$

Proof. It suffices to show that $c\widetilde{k}_0 \leq 1 - c$. When $\delta^* = \frac{1}{2}$, we see from (27) that

$$1 - c\widetilde{k}_0 = 1 - c(k_0 - 1) \geq \frac{\log p \log \frac{2p}{p-1} + \log \frac{p}{p-1} \log\log 2 + \log 2 \log\log \frac{p}{p-1}}{\log p \log \frac{2p}{p-1}} + c.$$

The denominator in the fraction above is positive, and the numerator satisfies

$$\begin{aligned} &\log p \log \frac{2p}{p-1} + \log \frac{p}{p-1} \log\log 2 + \log 2 \log\log \frac{p}{p-1} \\ &= (\log p + \log\log 2) \log \frac{p}{p-1} + \left(\log p + \log\log \frac{p}{p-1} \right) \log 2. \end{aligned}$$

The first term above is manifestly positive for $p \geq 2$. The second term is also positive since $\log \frac{p}{p-1} = \int_{p-1}^{p} \frac{1}{x}\, dx > \frac{1}{p}$ for $p \geq 2$. It follows that $1 - c\widetilde{k}_0 \geq c$. □

Remark. The lower bound on $\widetilde{b}$ can be slightly improved by choosing δ^* to minimize $\widetilde{k}_0$ in the above lemma. Although the minimizer of $\widetilde{k}_0$ cannot be solved for explicitly, numerical evidence suggests that the approximation $\delta^* = \left(\frac{p}{p-1}\right)^2$ can improve the bound to $\widetilde{b} \geq F(p)$ for some function $F(p)$ asymptotic to $\frac{2}{p^{1-c}}$.

It is easy to see by comparing (20) with (19) that the same analysis, this time choosing $\alpha = \epsilon$ rather than $\alpha = \epsilon/2$, also yields (7) in Corollary 1.1.

It remains to establish (8). For this, we take α fixed and use a similar argument. In this case k_1 can be regarded as a constant independent of ϵ, since the error in the interval $[0, \alpha^p]$ is irrelevant. Therefore we write $\widehat{k} := \frac{\log\log\frac{1}{\alpha}}{\log\frac{p}{p-1}} + \widetilde{k}_0$, and in place of (28), the lowest degree n required for ϵ accuracy on $S_{p,\alpha}$ satisfies $\log n \leq \left(\widehat{k} + \frac{\log\log\frac{p}{(p-1)^2\epsilon}}{\log 2}\right)\log p$. Thus defining $\widehat{a} := \frac{p}{(p-1)^2}$ and

$$\widehat{c} := \frac{\log 2}{\log p} (> c), \tag{30}$$

we have $\log\frac{\widehat{a}}{\epsilon} \geq \left(\frac{n}{p^{\widehat{k}}}\right)^{\widehat{c}}$, and so setting $\widehat{b} = 1/p^{\widehat{c}\widehat{k}}$ we obtain $\epsilon \leq \widehat{a}\exp(-\widehat{b}n^{\widehat{c}})$, as required. Like before, this upper bound weakens to

$$\widehat{a}\exp(-\widehat{b}p^{-\widehat{c}}n^{\widehat{c}}) = \widehat{a}\exp\left(-\frac{1}{2}\widehat{b}n^{\widehat{c}}\right)$$

when n is not a power of p.

We will complete the proof of (8) by finding a lower bound for $\widehat{b}$.

Lemma 4.6. *If $\delta^* = \frac{1}{2}$, then*

$$\widehat{b} \geq 2\log 2 \left(\frac{\log\frac{p}{p-1}}{\log\frac{1}{\alpha}}\right)^{\frac{\log 2}{\log\frac{p}{p-1}}}.$$

Proof. Substituting $\delta^* = \frac{1}{2}$ into (27) gives

$$\begin{aligned}
\widehat{c}(\widehat{k}+1) &= \frac{\log 2}{\log p}\left(\frac{\log\log\frac{1}{\alpha}}{\log\frac{p}{p-1}} + k_0\right) \\
&\leq \frac{\log 2}{\log p}\left(\frac{\log\log\frac{1}{\alpha}}{\log\frac{p}{p-1}} - \frac{\log\log 2}{\log 2} - \frac{\log\log\frac{p}{p-1}}{\log\frac{p}{p-1}}\right) \\
&= \frac{\log 2}{\log p}\left(\frac{\log\frac{\log\frac{1}{\alpha}}{\log\frac{p}{p-1}}}{\log\frac{p}{p-1}}\right) - \frac{\log\log 2}{\log p}.
\end{aligned}$$

Thus,

$$\widehat{b} = \frac{1}{p^{\widehat{c}\widehat{k}}} = \frac{2}{p^{\widehat{c}(\widehat{k}+1)}} \geq 2\log 2\left(\frac{\log\frac{p}{p-1}}{\log\frac{1}{\alpha}}\right)^{\frac{\log 2}{\log\frac{p}{p-1}}}.$$

□

5. Examples

In Figure 2 we illustrate our main result (2) on approximation of $x^{1/p}$. For integers $k = 1, 2, \ldots$, we compute the error ϵ of the composite rational approximants as in Figure 1, and plot the errors against $p^{ck} (\approx n^c)$ for $p \in \{2, 5, 31\}$ in log-scale. The plots also show least-squares affine fits to the convergence data for each p. The fact that the affine fits closely trace the data suggests the exponent c in (29) is sharp, especially for small values of p. For the $p = 31$ plot, which ends early because computing further data was infeasible (note e.g. that $\alpha^p < 10^{-70}$ for $k \geq 15$), there is a slight bend in the convergence, which suggests that our c in (3) might be a slight underestimate for large p.

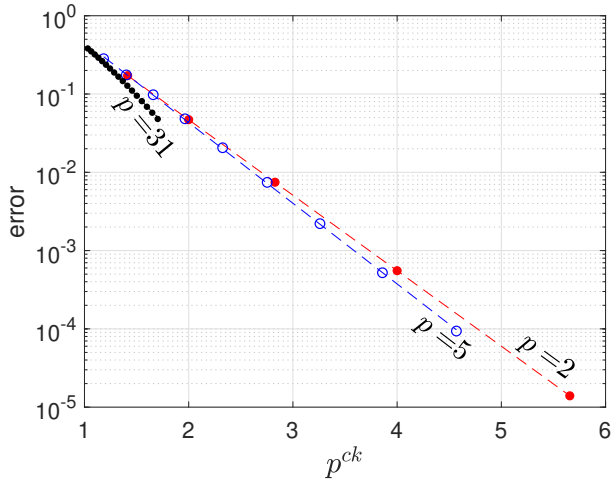

Figure 2: Error history $\max_{x\in[0,1]} |\widetilde{f}_k(x) - x^{1/p}|$ for varying k for $p \in \{2, 5, 31\}$, along with linear fits shown as dashed lines.

Finally, Figure 3 shows the error of the approximant $\widetilde{g}_k(z)$ to $\mathrm{sect}_p(z)$, which clearly exhibits equioscillation. Note how increasing k results in progressively smaller error (in log-scale), reflecting the double-exponential convergence. The error curves $|\widetilde{g}_k(z) - \mathrm{sect}_p(z)|$ look identical on each of the segments $[\alpha, 1] \exp(2\pi \mathrm{i} j/p)$ for $j = 0, \ldots, p-1$.

6. Discussion

We have seen that a large number of well-known functions can be approximated by composite rational functions. This is perhaps counterintuitive given that composite functions form a small subclass of functions of the same degree, as investigated in Ritt's classical work [18] for polynomials (see also a more recent work by Rickards [17]), and Bogatyrev [4] for rational functions[2]. More generally, we think composite (rational) functions are a non-standard but powerful tool in approximation theory, and we regard this as a contribution

[2]It is worth noting that these papers study *pure* composite functions, and our definition (1) is more general.

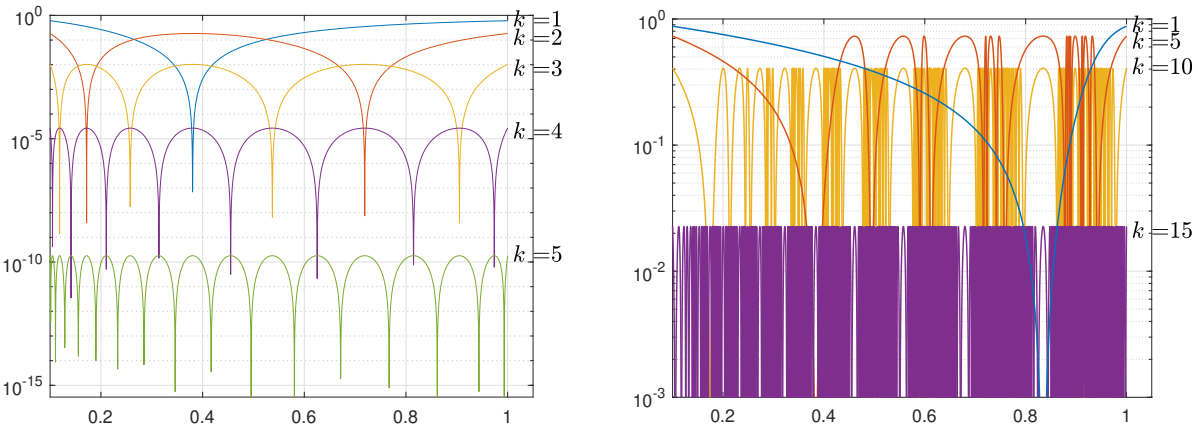

Figure 3: Error $|\widetilde{g}_k(z) - \mathrm{sect}_p(z)|$ on $[\alpha, 1]$ for $\alpha = 0.1$, $p = 3$ (left) and $p = 31$ (right). The fact that the plots do not appear to go down to 0 between equioscillation points is simply an artifact of the plotting scheme, which is based on 10^4 equispaced sample points.

towards demonstrating their effectiveness and practicality. Indeed, one might say they are already used extensively in scientific computing:

1. Composite rational functions are implicitly employed in most algorithms for computing matrix functions [10], in which approximating a function on the spectrum of the matrix is required. This includes the all-important matrix exponential [10]. For the pth root, a standard algorithm [10, Ch. 7] employs Newton's method, which ultimately approximates $A^{1/p}$ with a sequence of rational functions f_k of A given recursively by $f_{k+1}(x) = \frac{1}{p}((p-1)f_k(x) + x/f_k(x)^{p-1})$, $f_0(x) = 1$. The function f_k is composite rational and similar to the approximants we use, but not the same (it is unscaled), and it exhibits exponential rather than double-exponential convergence on $[0, 1]$, which can be easily verified by examining the convergence at $x = 0$. Generally speaking, Newton's method for computing a matrix function $f(A)$ (or more generally for various nonlinear problems, e.g. rootfinding) can often be interpreted as approximating $f(A)$ (or the solution) by a composite rational function of A. We also note that for evaluating matrix functions, the composite structure can be beneficial in terms of numerical stability in addition to efficiency; this is because the composite structure can avoid involving very small coefficients, and hence ill-conditioning [14].
2. The rapidly growing subject of deep learning is based on composing a large number of nonlinear activation functions [12]. A recent work [5] builds upon this observation to propose a network based on rational activation functions, leading to a high-degree composite rational function to approximate the input-output map. It is shown to often outperform popular networks based on ReLU activation functions.

For these reasons we believe that understanding the power and limitations of composite (rational) functions may have important ramifications in scientific computing. Future work includes identifying the class of functions that can (and equally interestingly, cannot) be approximated efficiently by composite rational functions.

Acknowledgment

We thank Alex Townsend, a discussion with whom inspired this work. EG was supported partially by NSF grants DMS-1703719 and DMS-2012427. YN was supported partially by the JSPS grant no. 18H05837.

References

[1] N. I. Akhiezer. *Elements of the Theory of Elliptic Functions*, volume 79 of *Translations of Mathematical Monographs*. American Mathematical Society, 1990.

[2] B. Beckermann. Optimally scaled Newton iterations for the matrix square root. *FUN13: Advances in Matrix Functions and Matrix Equations workshop*, 2013.

[3] B. Beckermann and A. Townsend. On the singular values of matrices with displacement structure. *SIAM J. Matrix Anal. Appl.*, 38(4):1227–1248, 2017.

[4] A. Bogatyrev. Rational functions admitting double decompositions. *Transactions of the Moscow Mathematical Society*, 73:161–165, 2012.

[5] N. Boullé, Y. Nakatsukasa, and A. Townsend. Rational neural networks. In *Advances in Neural Information Processing Systems*, volume 33, pages 14243–14253, 2020.

[6] D. Braess. *Nonlinear Approximation Theory*. Springer, 1986.

[7] E. S. Gawlik. Zolotarev iterations for the matrix square root. *SIAM J. Matrix Anal. Appl.*, 40(2):696–719, 2019.

[8] E. S. Gawlik. Rational minimax iterations for computing the matrix pth root. *Constr. Approx.*, 2021.

[9] A. Gončar. On the rapidity of rational approximation of continuous functions with characteristic singularities. *Mathematics of the USSR-Sbornik*, 2(4):561, 1967.

[10] N. J. Higham. *Functions of Matrices: Theory and Computation*. SIAM, Philadelphia, PA, USA, 2008.

[11] R. F. King. Improved Newton iteration for integral roots. *Mathematics of Computation*, 25(114):299–304, 1971.

[12] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436, 2015.

[13] G. Meinardus and G. Taylor. Optimal partitioning of Newton's method for calculating roots. *Mathematics of Computation*, 35(152):1221–1230, 1980.

[14] Y. Nakatsukasa and R. W. Freund. Computing fundamental matrix decompositions accurately via the matrix sign function in two iterations: The power of Zolotarev's functions. *SIAM Rev.*, 58(3):461–493, 2016.

[15] I. Ninomiya. Best rational starting approximations and improved Newton iteration for the square root. *Math. Comp.*, 24(110):391–404, 1970.

[16] P. P. Petrushev and V. A. Popov. *Rational Approximation of Real Functions*. Cambridge University Press, 2011.

[17] J. Rickards. When is a polynomial a composition of other polynomials? *Amer. Math. Monthly*, 118(4):358–363, 2011.

[18] J. F. Ritt. Prime and composite polynomials. *Transactions of the American Mathematical Society*, 23(1):51–66, 1922.

[19] H. Rutishauser. Betrachtungen zur Quadratwurzeliteration. *Monatshefte für Mathematik*, 67(5):452–464, 1963.

[20] L. S. Shieh, Y. T. Tsay, and C. T. Wang. Matrix sector functions and their applications to systems theory. In *IEE Proceedings D (Control Theory and Applications)*, volume 131, pages 171–181. IET, 1984.

[21] H. R. Stahl. Best uniform rational approximation of x^α on [0, 1]. *Acta Math.*, 190(2):241–306, 2003.

[22] L. N. Trefethen. *Approximation Theory and Approximation Practice*. SIAM, Philadelphia, 2013.

[23] E. Wachspress. Positive definite square root of a positive definite square matrix. *Unpublished*, 1962.