Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Approximation of $1/x$ by Exponential Sums in $[1, \infty)$

by

Dietrich Braess and Wolfgang Hackbusch

Approximation of $1/x$ by Exponential Sums in $[1, \infty)$

Dietrich Braess and Wolfgang Hackbusch

Abstract

Approximations of $1/x$ by sums of exponentials are well studied for finite intervals. Here the error decreases like $\mathcal{O}(\exp(-ck))$ with the order k of the exponential sum. In this paper we investigate approximations of $1/x$ on the interval $[1, \infty)$. We prove estimates of the error by $\mathcal{O}(\exp(-c\sqrt{k}))$ and confirm this asymptotic estimate by numerical results. Numerical results lead to the conjecture that the constant in the exponent equals $c = \pi\sqrt{2}$.

1 Introduction

We consider the approximation of the function $f(x) = 1/x$ by sums of exponentials of the form

$$s_k(x) = \sum\nolimits_{j=1}^{k} \omega_j \exp(-\alpha_j x) \qquad (\omega_j, \alpha_j \in \mathbb{R}) \tag{1.1}$$

It is well-known that, given a fixed interval $[a, b]$ with $0 < a < b$, there is a unique exponential sum (1.1) minimising the Chebyshev norm $\|f - s_k\|_{\infty,[a,b]} := \max\{|f(x) - s_k(x)| : a \le x \le b\}$; see Braess [2, p. 194]. Some of the properties of the optimal approximation s_k^* are:

- $\omega_j > 0$ and $\alpha_j > 0$,

- the optimal s_k^* interpolates the function $f(x) = 1/x$ at exactly $2k$ points ξ_i satisfying $a < \xi_1 < \xi_2 < \ldots < \xi_{2k} < b$,

- the best approximation error $E_k := \|f - s_k^*\|_{\infty,[a,b]}$ decays exponentially like $E_k \lesssim C_1 \exp(-C_2 k)$ as $k \to \infty$.

More details will follow below.

The interest in approximations of the form (1.1) is manifold. One reason is that the d-variate function $1/(x_1 + \ldots + x_d)$ can be approximated with the same accuracy by the sum $\sum_{j=1}^{k} \omega_j \prod_{\nu=1}^{d} \exp(-\alpha_j x_\nu)$, where in each term the variables x_ν are separated in different factors. This fact can be used, e.g., in coupled cluster analysis in quantum chemistry [7]. Another reason is that (1.1) can be used to approximate the inverse A^{-1} of a positive definite operator A by $s_k^*(A)$. Here, the spectrum of A is to be contained in the interval $[a, b]$ above. An application of this kind can be found in Grasedyck [5].

Since $1/x$ and $\exp(-\alpha_j x)$ can easily be scaled by a factor, the left endpoint a of the interval $[a, b]$ will be replaced by $a = 1$ without loss of generality. Then b becomes b/a and is renamed by R, i.e., the new interval has the endpoints

$$a = 1, \quad b = R.$$

Now, k and R are the only two parameters of the approximation problem. The approximation error is denoted by

$$E_k(R) := \min_{\omega_j, \alpha_j} \|f - s_k\|_{\infty,[1,R]} = \left\|f - s_{k,R}^*\right\|_{\infty,[1,R]},$$

1

where

$$s_{k,R}^* := \operatorname{argmin}_{\omega_j, \alpha_j} \|f - s_k\|_{\infty, [1,R]}$$

is the optimal approximation. Two limits can be investigated:

- $k \to \infty$ with fixed R is the standard approach when the approximation error $E_k(R)$ is discussed (see §2).

- If k is fixed, the error $E_k(R)$ increases with R until a critical value R_k is reached. Then $E_k(R) = E_k(R_k)$ holds for *all* $R \geq R_k$, i.e., $E_k(R_k)$ is the error of the approximation in $[1, \infty)$.

The special interest of this paper concerns the "diagonal sequence" $\{E_k(R_k) : k \to \infty\}$, which is the best approximation error in $[1, \infty)$. The convergence behaviour of this approximation problem has been investigated less; for a weighted L_2 approximation see [3, 7]. We study three approaches and give theoretical estimates of the form $E_k \lesssim C_1 \exp(-c_2\sqrt{k})$; see §3. One of the approaches is related to Heron's method and is also appropriate for approximation in the complex plane. Although the analysis is based on sharp estimates for two problems of rational approximation, we get only suboptimal bounds for the coefficient c_2. The numerical results in §4 provide the size of the factor c_2 and confirm the $\sqrt{k}$ behaviour.

2 Approximation of $1/x$ in $[1, R]$

Given R, the best approximation s_k^* on $[1, R]$ is unique. The error function $\varepsilon_{k,R} := f - s_{k,R}^*$ has zeros at the interpolation points $1 < \xi_1 < \ldots < \xi_{2k} < R$ mentioned in the introduction. Moreover, there are $2k + 1$ points ζ_j with

$$\varepsilon_{k,R}(\zeta_j) = (-1)^j E_k(R) \qquad \text{at } 1 = \zeta_0 < \xi_1 < \zeta_1 < \ldots < \xi_{2k} < \zeta_{2k} \leq R. \tag{2.1}$$

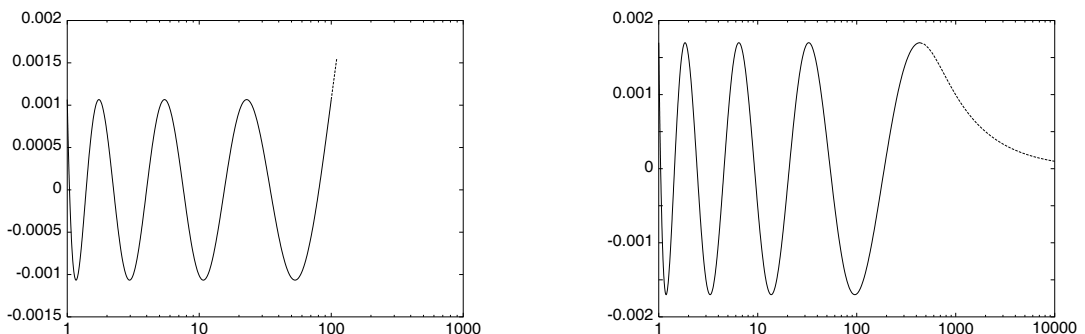For k large enough, $\zeta_{2k} = R$ holds as in the left part of Figure 2.1.

Figure 2.1: *Left:* Case of $k = 4$ and $R = 100 < R_k$ ($E_k(R) = 1.066_{10}$-3). *Right:* Case of $k = 4$ and $R_k = 436.06$ ($E_k(R_k) = 1.700_{10}$-3)

The asymptotic behaviour with respect to $k \to \infty$ can be obtained by using a result from rational approximation. The following estimate is obtained from an analysis of Heron's method for the computation of the square root of a given number; see Appendix A or [2, p. 145]. The typical quotient of two elliptic integrals can already be found in results for a wider class of approximation problems [4], but not with the constant 4.

Lemma 2.1 *Let $k \in \mathbb{N}$ and $0 < \kappa < 1$. There are polynomials p_k and q_{k-1} of degree k and $k - 1$, resp., such that*

$$\left| \frac{\sqrt{x} - p_k(x)/q_{k-1}(x)}{\sqrt{x}} \right| \leq 4\sigma^{-2k} \qquad \text{for } x \in [\kappa^2, 1] \tag{2.2}$$

2

where

$$\sigma = \sigma(\kappa) := \exp\left[\frac{\pi \mathbf{K}(\kappa)}{\mathbf{K}'(\kappa)}\right].$$

Here, $\mathbf{K}$ is the complete elliptic integral of the first kind with modulus κ (cf. Appendix A). Moreover $\mathbf{K}'(\kappa) = \mathbf{K}(\kappa')$ where κ' with $\kappa^2 + (\kappa')^2 = 1$ is the complementary modulus.

We are interested in large intervals, and the often cited formulae for the asymptotics $\lim_{\kappa \to 0} \mathbf{K}(\kappa) = \pi/2$ and $\lim_{\kappa \to 0} \mathbf{K}'(\kappa)/\log\frac{4}{\kappa} = 1$ are not sufficient. Instead, the sharper estimates from (A.4) and (A.5),

$$\mathbf{K}(\kappa) \geq \frac{\pi}{2}, \qquad \mathbf{K}'(\kappa) \leq \log(\frac{4}{\kappa} + 2),$$

will be used.

Instead of approximating $1/x$ on the interval $[1, R]$ we may approximate

$$f(x) := \frac{1}{x + \varepsilon}$$

on the interval $[1 - \varepsilon, R - \varepsilon]$ with ε being fixed later. This has the advantage that

$$|f(z)| \leq \frac{1}{\varepsilon} \qquad \text{whenever } \Re e\, z \geq 0.$$

Referring to Lemma 2.1 we know that

$$\left|\frac{\sqrt{x} - p_k(x)/q_{k-1}(x)}{\sqrt{x} + p_k(x)/q_{k-1}(x)}\right| \leq 4\sigma^{-2k} \qquad \text{for } x \in [\kappa^2, 1].$$

We set $x = z^2$ and

$$r_k(z) := \frac{p_k(z^2) - z q_{k-1}(z^2)}{p_k(z^2) + z q_{k-1}(z^2)}.$$

In particular we know that the numerator has $2k$ zeros in $[\kappa, 1]$. Similarly, all poles lie in the interval $[-1, -\kappa]$. Moreover,

$$\begin{aligned}
|r_k(z)| &\leq 4\sigma^{-2k}, & \text{for } z \in [\kappa, 1], \\
|r_k(z)| &= 1, & \text{if } \Re e\, z = 0, \\
r_k(z) &\to 1, & \text{as } |z| \to \infty.
\end{aligned} \qquad (2.3)$$

More generally, after replacing z by z/b we have that the expression in the first line of (2.3) lies in the interval $[\kappa b, b]$.

Now we follow the arguments in [2, p. 177], but adapt them to large intervals. Set $\kappa := (1 - \varepsilon)/(R - \varepsilon)$ and let the argument of r_k be rescaled for the interval $[1 - \varepsilon, R - \varepsilon]$. There is an exponential sum s_k of the form (1.1) that interpolates f at the $2k$ zeros of r_k in $[1 - \varepsilon, R - \varepsilon]$. Since $f - s_k$ has no more zeros in $[0, \infty)$, it follows that

$$0 < s_k(x) < f(x) \qquad \text{for } 0 \leq x < 1 - \varepsilon \text{ and } x > R - \varepsilon. \qquad (2.4)$$

Moreover,

$$|s_k(z)| \leq s_k(\Re e\, z) \leq s_k(0) < f(0) \qquad \text{for } \Re e\, z \geq 0 \qquad (2.5)$$

and

$$|f(z) - s_k(z)| \leq |f(z)| + |s_k(z)| \leq 2f(0) \qquad \text{for } \Re e\, z \geq 0.$$

The function $r_k^{-1}[f - s_k]$ is analytic in the right half-plane. We know that r_k has modulus 1 on the boundary of the half-plane and conclude from the maximum principle for holomorphic functions that

$$\left|\frac{f(z) - s_k(z)}{r_k(z)}\right| \leq 2f(0) \qquad \text{for } \Re e\, z \geq 0. \qquad (2.6)$$

3

It follows that

$$|f(x) - s_k(x)| \leq |r_k(x)| \, 2f(0) \leq 8\sigma^{-2k} f(0) \qquad \text{for } x \in [1 - \varepsilon, R - \varepsilon]. \tag{2.7}$$

Furthermore

$$|f(x) - s_k(x)| \leq f(x) \leq f(R) \qquad \text{for } x \geq R - \varepsilon. \tag{2.8}$$

The error (2.7) decreases exponentially, and it is known that the resulting parameter σ is correct for $R = 2$, i.e., when R is small [3]. Then it is reasonable to fix $\varepsilon = 1/k$. For large R we set $\varepsilon = 1/2$ and obtain with the asymptotics of the elliptic integrals and $1/\kappa = 2R - 1$ that

$$\left| \frac{1}{x} - s_k(x) \right| \leq 16 \exp\left[-\frac{k\,\pi^2}{\log(8R)} \right]. \tag{2.9}$$

We note that this estimate is not optimal. The right-hand side of (2.5) can be replaced by the decreasing function $3f(\Re e\, z)$, but unfortunately we are not able to profit from this sharper bound; see also Section 3.3.

Remark 2.2 *If an operator A^{-1} is to be approximated by $s_k(A)$ (as required in [5]) and if A contains complex eigenvalues, an approximation of $1/z$ is required in some subdomain of the complex plane. It is clear from (2.6) that we need bounds of*

$$r_k(z)$$

in that subdomain (and not only in the interval). Note that $\log r_k(z)$ is holomorphic in the complex plane without the intervals $[-R + 1/2, -1/2]$ and $[1/2, R - 1/2]$. In particular, $G(z) := \Re e \log r_k(z)$ is a solution of the Laplace equation with boundary values

$$
\begin{array}{lll}
G(z) & \leq -c & \text{for } z \in [\frac{1}{2}, R - \frac{1}{2}], \\
G(z) & \geq +c & \text{for } z \in [-R + \frac{1}{2}, -\frac{1}{2}], \\
G(z) & \to 0 & \text{as } |z| \to \infty,
\end{array}
$$

where $c = k\pi^2 / \log(8R) - \log 16$. The function G can be understood and computed as the potential induced by the load on a condenser whose plates are the intervals specified above [8].

3   Approximation of $1/x$ in $[1, \infty)$

If we fix k and let R increase to ∞, the right-hand side in (2.9) also increases, which does not reflect the correct asymptotic behaviour. Since $E_k(\cdot)$ is non-decreasing, we have $E_k(R^*) > 1/R^*$ for some positive number R^*. From (2.4) we know that $1/x > s_{k,R^*} > 0$ for $x > R^*$, and it follows that

$$|1/x - s^*_{k,R^*}(x)| < 1/x < 1/R^* \leq E_k(R^*) \qquad \text{for all } x > R^*.$$

Hence, $|1/x - s^*_{k,R^*}(x)| \leq E_k(R^*)$ holds for all $x \geq 1$. Consequently, s_{k,R^*} is the best approximation for any $R > R^*$.

We may be even more specific. Let R_k be the right-most extremum of $|1/x - s^*_{k,R^*}|$. It follows that $s^*_{k,R_k} = s^*_{k,R^*}$ is the best approximation for any $R > R_k$. These results are summarised in the following remark.

Remark 3.1 *There is a critical value $R_k \in (1, \infty)$ such that $E_k(\cdot)$ increases strictly in $(1, R_k)$ and is constant on $[R_k, \infty)$. As a consequence, $s^*_{k,R'} = s^*_{k,R''}$ holds for any $R', R'' \in [R_k, \infty)$. For $R = R_k$ we introduce the notations*

$$s^{**}_k := s^*_{k,R_k} = \operatorname{argmin}_{\omega_j, \alpha_j} \|f - s_k\|_{\infty, [1,\infty)} \qquad \text{and} \qquad E^*_k := E_k(R_k) = \|f - s^{**}_k\|_{\infty, [1,\infty)}.$$

The functions s^{**}_k are the best k-term approximations in $[1, \infty)$ and E^*_k are the related errors. In the case of $R > R_k$, the strict inequality $\zeta_{2k} < R$ holds in (2.1) (see the right-hand part of Figure 2.1).

In the following we will establish estimates of E^*_k of the form

$$E^*_k \leq C_1 \exp(-c_2 \sqrt{k}) \tag{3.1}$$

with certain constants $C_1, c_2 > 0$ and ask whether $\sqrt{k}$ (instead of k in (2.9)) is the optimal power of k. We will study three approaches.

3.1 Sinc approximation

A first estimate of E_k^* can be obtained from Sinc approximation results. For details of Sinc approximation and, in particular, Sinc quadrature we refer to the monograph of Stenger [10, §3.2].

Let F be an analytic, improperly integrable function in $(-\infty, \infty)$. Its Sinc interpolation $F \approx C_N(F, h)$ is defined by

$$C(F, h) := \sum_{\ell=-\infty}^{\infty} F(\ell h) S(\ell, h), \qquad S(\ell, h)(t) := \operatorname{sinc}\left(\frac{t}{h} - \ell\right) = \frac{\sin\left[\pi\left(t - \ell h\right)/h\right]}{\pi\left(t - \ell h\right)/h},$$

where $h > 0$ is the step size. Formal integration yields $\int_{-\infty}^{\infty} C(F, h)\mathrm{d}t = T(F, h) := h \sum_{\ell=-\infty}^{\infty} F(\ell h)$, which is the infinite trapezoidal rule. Thanks to the expected decay of F as $|t| \to \infty$, we may introduce the truncated quadrature formula

$$T_N(F, h) := h \sum_{\ell=-N}^{N} F(\ell h) \tag{3.2}$$

for some $N \in \mathbb{N}_0$. An estimation of the quadrature error is provided by the next lemma; for a proof see Stenger [10, (3.2.2)].

Lemma 3.2 *Let F be holomorphic in the strip $\mathfrak{D}_d := \{z \in \mathbb{C} : \Im m\, z \in (-d, d)\}$ with a finite value of $N(F, \mathfrak{D}_d) := \int_{\partial \mathfrak{D}_d} |F(z)|\,|\mathrm{d}z|$. Then*

$$\left|\int_{-\infty}^{\infty} F(t)\mathrm{d}t - T(F, h)\right| \leq \frac{\exp(-\pi d/h)}{2\sinh(\pi d/h)} N(F, \mathfrak{D}_d). \tag{3.3}$$

The additional error $|T_N(F, h) - T(F, h)|$ depends on the decay rate of F.

Lemma 3.3 *In addition to the assumptions of Lemma 3.2, assume the exponential decay $|F(t)| \leq c \cdot \mathrm{e}^{-\alpha|t|}$ for $t \in \mathbb{R}$ and choose the step size $h := \sqrt{\frac{2\pi d}{\alpha N}}$. Then*

$$\left|\int_{-\infty}^{\infty} F(t)\mathrm{d}t - T_N(F, h)\right| \leq \left(\frac{N(F, \mathfrak{D}_d)}{1 - \exp(-\sqrt{2\pi d\alpha N})} + \frac{2c}{\alpha}\right) \mathrm{e}^{-\sqrt{2\pi d\alpha N}} \tag{3.4}$$

holds.

The integral in

$$\frac{1}{x} = \int_0^{\infty} \exp(-xt)\mathrm{d}t$$

is not yet of the form $\int_{-\infty}^{\infty} F(t)\mathrm{d}t$. A suitable substitution is $t = \log(1 + \mathrm{e}^{\tau})$, which yields

$$\frac{1}{x} = \int_{-\infty}^{\infty} \mathrm{e}^{-x\log(1 + \mathrm{e}^{\tau})} \frac{\mathrm{d}\tau}{1 + \mathrm{e}^{-\tau}} \qquad \text{for } x > 0. \tag{3.5}$$

Let $d < \pi/2$. The integrand $F(t) = \exp(-x\log(1 + \mathrm{e}^t))/(1 + \mathrm{e}^{-t})$ behaves like $\mathcal{O}(\mathrm{e}^{-x\Re e\,\tau})$ for $\Re e\,\tau \geq 0$ ($\tau \in \mathfrak{D}_d$) and like $\mathcal{O}(\mathrm{e}^{-|\Re e\,\tau|})$ for $\Re e\,\tau \leq 0$ (cf. [6]). Hence, we have an exponential decay $|F(t)| \leq c \cdot \mathrm{e}^{-\alpha|t|}$ with $\alpha = \min\{1, x\}$. Furthermore, the integrand is holomorphic in $\mathfrak{D}_d$ with $N(F, \mathfrak{D}_d) = \mathcal{O}(1 + 1/x)$. Choose h as in Lemma 3.3. From Lemma 3.3 we obtain the estimate

$$|\eta_N(F, h)| \leq C\mathrm{e}^{-\sqrt{2\pi dN}}, \tag{3.6}$$

which holds uniformly with respect to $x \geq 1$. Note that the exponent is any number close to $-\pi\sqrt{N}$.

5

The quadrature formula has the form $h \sum_{\ell=-N}^{N} F(\ell h)(x) = \sum_{\ell=-N}^{N} \omega_\ell \exp(-\alpha_\ell x)$ with $\omega_\ell = h/(1 + \mathrm{e}^{-\ell h})$ and $\alpha_\ell = \log(1 + \mathrm{e}^{\ell h})$. Therefore we have proved the estimate

$$\left| \frac{1}{x} - s_{2N+1}(x) \right| \le C\mathrm{e}^{-\sqrt{2\pi dN}} \qquad (1 \le x < \infty)$$

for the special function $s_{2N+1} := h \sum_{\ell=-N}^{N} F(\ell h)$ in the set of functions (1.1). This estimate is an upper bound for the best approximation. Hence (3.1) holds for even $k = 2N + 1$ with

$$c_2 = \pi/\sqrt{2}.$$

Results of the Sinc quadrature are illustrated by two examples in Figure 3.1 with $N = 2$ ($h = \pi/\sqrt{N}$) and $N = 22$ ($h = 1.05 \cdot \pi/\sqrt{N}$). The accuracy of $1.193_{10}\text{-}2$ in the former and $2.63_{10}\text{-}7$ in the latter case can be compared with the much better results for $k = 5$ and $k = 45$ of the best approximation given in the first table of §4.
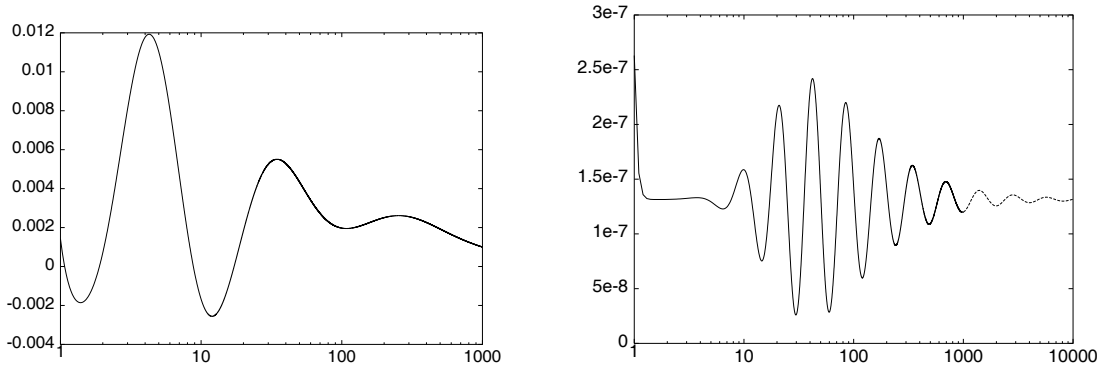

Figure 3.1: Approximation by means of Sinc quadrature. Left: $N = 2$ ($k = 5$). Right: $N = 22$ ($k = 45$)

3.2 Estimates using results for finite intervals

We use the results from the preceding section with

$$R := \frac{1}{8} \exp[\pi\sqrt{k}].$$

From (2.7) we obtain

$$\left| \frac{1}{x} - s_k(x) \right| \le 16 \exp[-\pi\sqrt{k}] \qquad \text{for } 1 \le x \le R,$$

and from (2.8)

$$\left| \frac{1}{x} - s_k(x) \right| \le \frac{1}{R} = 8 \exp[-\pi\sqrt{k}] \qquad \text{for } x \ge R.$$

Hence,

$$E_k^* \le 16\mathrm{e}^{-\pi\sqrt{k}}, \tag{3.7}$$

i.e., $c_2 = \pi$. This exponent is better than that from the Sinc approximation, but the numerical results in §3.3 show that there is still a gap of a factor of $\sqrt{2}$.

The estimate above has been performed in two steps. The inequality (2.9) applies to arbitrary completely monotone functions with the factor being adapted to the given function. When (3.7) is derived from (2.9), the decay of $1/x$ as $x \to \infty$ enters into the analysis.

6

3.3 Direct estimates

We note that there is also a one-step proof for the special function $1/x$. It is based on a result of Vjačeslavov [11] which in turn requires involved evaluations of some special integrals; see also [9]. *Given $\alpha > 0$ and $n \in \mathbb{N}$, there exists a polynomial p of degree n with n zeros in $[0,1]$ such that*

$$\left| x^\alpha \frac{p(x)}{p(-x)} \right| \le c_0(\alpha) \cdot \mathrm{e}^{-\pi\sqrt{\alpha n}} \qquad \text{for } 0 \le x \le 1.$$

Let p be the polynomial for $\alpha = 1/4$ as stated above. Since $p(\bar{z}) = \bar{p}(z)$, it follows that $p(z)/p(-z) = 1$ for $\Re e\, z = 0$ and

$$\left| \frac{p(z^2)}{p(-z^2)} \right| = 1 \qquad \text{for } \Re e\, z = |\Im m\, z| \ge 0. \tag{3.8}$$

We consider $P(z) := p^2(1/z^2)$ on the sector $\mathcal{S} := \{z \in \mathbb{C} : |\arg z| \le \pi/4\}$. By construction P has n double zeros in $[1, \infty)$, and from (3.8) it follows that

$$\left| \frac{P(z)}{P(-z)} \right| = 1 \quad \text{for } z \in \partial\mathcal{S}, \qquad \frac{P(x)}{x\, P(-x)} \le \left(c_0(1/4) \cdot \mathrm{e}^{-\pi\sqrt{n/4}} \right)^2 \quad \text{for } x \ge 1.$$

Now let s_n be the exponential sum interpolating $1/x$ and its first derivative at the (double) zeros of P. Since $1/x - s_n$ has no more zeros than the specified ones, we have $s_n(x) \le 1/x$ for $x \ge 0$. Hence,

$$|s_n(z)| \le s_n(\Re e\, z) \le 1/\Re e\, z \le \sqrt{2}/|z| \qquad \text{on the boundary of } \mathcal{S}.$$

Arguing as in Section §2, we introduce the auxiliary function $g(z) := (\frac{1}{z} - s_n(z))z\frac{P(-z)}{P(z)}$. We know that $|g(z)| \le 1 + \sqrt{2}$ holds on the boundary of $\mathcal{S}$ and therefore in $\mathcal{S}$. Finally,

$$\left| \frac{1}{z} - s_n(z) \right| = \left| g(z)\frac{P(z)}{z\, P(-z)} \right| \le (1 + \sqrt{2})\, c_0^2(1/4)\mathrm{e}^{-\pi\sqrt{n}}.$$

4 Numerical Results

The following results are based on the numerically computed coefficients of the best approximations s_k^{**} for $1 \le k \le 50$. Some of the results of R_k and E_k^* are given in the following table[1].

k	R_k	E_k^*	$E_k^*\, \mathrm{e}^{\pi\sqrt{2k}}/\log(2+k)$
1	8.667	8.55641E-02	6.62
2	41.54	1.78498E-02	6.89
5	1153	6.42813E-04	6.82
10	56502	1.31219E-05	6.67
15	1.175E+6	6.31072E-07	6.62
20	1.547E+7	4.79366E-08	6.60
25	1.514E+8	4.89759E-09	6.60
30	1.198E+9	6.18824E-10	6.61
35	8.064E+9	9.19413E-11	6.62
40	4.771E+10	1.55388E-11	6.64
45	2.540E+11	2.91895E-12	6.66
50	1.237E+12	5.99210E-13	6.68

The product $R_k \cdot E_k^*$ yields values between 0.741 and 0.742, i.e., nearly a constant. The last column of the table shows that $E_k^* \approx 6.7 \log(2+k)\mathrm{e}^{-\pi\sqrt{2k}}$ is a very good approximation.

---

[1]A complete list can be found in www.mis.mpg.de/scicomp/EXP_SUM/1_x/tabelle

We also tested the ansatz

$$E_k^* \approx C_1 \exp(-c_2\sqrt{k}) = \exp\{c_1 - c_2\sqrt{k}\}$$

in different parts of the k-interval. The least squares fit $\min\limits_{c_1,c_2} \sqrt{\sum_{k=k_1}^{k_2} \left(\frac{\log(E_k^*)}{c_1-c_2\sqrt{k}} - 1\right)^2}$ yields the coefficients in

k	1-10	11-20	21-30	31-40	41-50	11-50
c_1	1.588	2.328	2.510	2.619	2.697	2.477
c_2	4.014	4.288	4.329	4.349	4.361	4.326
minimum	3.2E-2	6.4E-4	1.5E-4	6.2E-5	3.2E-5	5.0E-3

The factor c_2 in $E_k^* \sim C_1 \exp(-c_2\sqrt{k})$ is at least $\pi/\sqrt{2} = 2.221$ due to §3.1, while we conjecture that the asymptotic limit is $c_2 = \sqrt{2}\pi \approx 4.4429$.

A Gauss' arithmetic-geometric mean and Heron's method

We can recognise the asymptotic behaviour from the numerical results above, and one may wonder that the deviation from that asymptotic law is very small also for small values of k. This is related to the fact that the result of Lemma 2.1 is obtained by a process that is often used for so-called *fast computations*. Moreover it sheds some light on Heron's famous algorithm. Additional facts are found in [1].

The discussion follows [2, p. 145], but it is directed to readers which are less familiar with approximation theory. Moreover, there are some completions and corrections of misprints.

A.1 The arithmetic-geometric mean

At an early age, Gauss became enamoured of a sequential procedure that is now known as the arithmetic-geometric process; see e.g. [1]. Given two numbers $0 < a_0 < b_0$, one successively takes the arithmetic mean and the geometric mean

$$a_{j+1} = \sqrt{a_j b_j}, \qquad b_{j+1} = \frac{1}{2}(a_j + b_j).$$

The common limit $\lim_{j\to\infty} a_j = \lim_{j\to\infty} b_j$ is called the arithmetic-geometric mean of a_0 and b_0 and is denoted as $m(a_0, b_0)$. It can be expressed in terms of a complete elliptic integral

$$I(a,b) = \int_0^\infty \frac{\mathrm{d}t}{\sqrt{(a^2+t^2)(b^2+t^2)}}. \tag{A.1}$$

The crucial observation for establishing the relation between $m(a,b)$ and $I(a,b)$ is that $I(a,b)$ is invariant under the transformation $(a,b) \mapsto (a_1,b_1) = (\sqrt{ab}, \frac{a+b}{2})$. We see this by the substitution $t = \frac{1}{2}(x - \frac{ab}{x})$. As x goes from 0 to ∞, the variable t increases from $-\infty$ to ∞. Moreover,

$$\mathrm{d}t = \frac{x^2+ab}{2x^2}\,\mathrm{d}x, \quad t^2 + \left(\frac{a+b}{2}\right)^2 = \frac{(x^2+a^2)(x^2+b^2)}{4x^2}, \quad t^2 + ab = \frac{(x^2+ab)}{4x^2}.$$

Hence,

$$I(a_1,b_1) = \frac{1}{2}\int_{-\infty}^\infty \frac{\mathrm{d}t}{\sqrt{(a_1^2+t^2)(b_1^2+t^2)}} = \int_0^\infty \frac{\mathrm{d}x}{\sqrt{(a^2+x^2)(b^2+x^2)}} = I(a,b) \tag{A.2}$$

yields the invariance.

Let $m = m(a,b)$, and set $a_0 = a$, $b_0 = b$. By induction it follows that $I(a_0, b_0) = I(a_j, b_j)$ for all j, and by continuity $I(a_0, b_0) = I(m, m)$. Obviously, $I(m,m) = \int_0^\infty \frac{\mathrm{d}t}{m^2+t^2} = \frac{\pi}{2m}$, and we conclude that

$$m(a,b) = \frac{\pi}{2I(a,b)}.$$

The *elliptic integrals* are defined by $\mathbf{K}'(\kappa) := I(\kappa, 1)$ and the relation between $\mathbf{K}$ and $\mathbf{K}'$ stated in Lemma 2.1. A scaling argument shows that

$$I(a, b) = b^{-1}\mathbf{K}'(a/b) \qquad \text{for } 0 < a \le b. \tag{A.3}$$

Since the arithmetic-geometric mean of 1 and κ lies between the arithmetic mean and the geometric mean, we get an estimate that is good for $\kappa \approx 1$.

$$\frac{\pi}{1+\kappa} \le \mathbf{K}'(\kappa) \le \frac{\pi}{2\sqrt{\kappa}}. \tag{A.4}$$

An estimate that is good for small κ is more involved:

$$\mathbf{K}'(\kappa) = 2\int_0^{\sqrt{\kappa}} \frac{\mathrm{d}t}{\sqrt{(1+t^2)(\kappa^2+t^2)}} \le 2\int_0^{\sqrt{\kappa}} \frac{\mathrm{d}t}{\sqrt{\kappa^2+t^2}} = 2\int_0^{1/\sqrt{\kappa}} \frac{\mathrm{d}t}{\sqrt{1+t^2}}$$

$$= 2\log\left(\sqrt{\frac{1}{\kappa}} + \sqrt{\frac{1}{\kappa}+1}\right) \le \log\left(4\left(\frac{1}{\kappa}+\frac{1}{2}\right)\right). \tag{A.5}$$

As a consequence, we have $(\pi/2)\mathbf{K}'(\kappa)/\mathbf{K}(\kappa) \le \log(\frac{4}{\kappa}+2)$ and

$$\frac{1}{\kappa} \ge \frac{1}{4}\exp\left[\frac{\pi\mathbf{K}'(\kappa)}{2\mathbf{K}(\kappa)}\right] - \frac{1}{2}. \tag{A.6}$$

A.2   The Landen transformation

The arithmetic-geometric mean of two given numbers is usually not determined via the integral (A.1). On the contrary, elliptic integrals are computed via the arithmetic-geometric process since its convergence is extremely fast. Set $\lambda_j := b_j/a_j$. Then

$$\lambda_{j+1} = \frac{a_j + b_j}{2\sqrt{a_j b_j}} = \frac{1+\lambda_j}{2\sqrt{\lambda_j}} = \frac{1}{2}\left(\sqrt{\lambda_j} + \frac{1}{\sqrt{\lambda_j}}\right).$$

Similarly,

$$\lambda_j = \left(\lambda_{j+1} + \sqrt{\lambda_{j+1}^2 - 1}\right)^2.$$

The mapping $\lambda \mapsto (\lambda + \sqrt{\lambda^2 - 1})^2$ is called *Landen transformation*. The following sequence (with rounded numbers) illustrates that either a few steps in the forward direction brings us to very large numbers or a few steps in the backward direction brings us close to 1:

$$6.825 \cdot 10^{14}, \ 1.306 \cdot 10^7, \ 1807.08, \ 21.26, \ 1+\sqrt{2}, \ 1.099, \ 1.0011, \ 1+1.53 \cdot 10^{-7}.$$

So in any case asymptotic formulae provide very good approximations in combination with the transformation. Finally, we mention that the numbers $(\lambda + 1)/(\lambda - 1)$ and λ' with $(\lambda')^{-2} + \lambda^{-2} = 1$ are moved by the same rule, but in the opposite direction.

Lemma A.1 *Let* λ_0, $\lambda_{-1}, \lambda_{-2}, \ldots$ *be a sequence generated by the Landen transformation. Then*

$$\lambda_{-j} \ge \frac{1}{4}\exp\left[2^j\,\frac{\pi}{2}\frac{\mathbf{K}'(\kappa_0)}{\mathbf{K}(\kappa_0)}\right] - \frac{1}{2} \qquad \text{with } \kappa_0 = \frac{1}{\lambda_0}. \tag{A.7}$$

Proof. Let $0 < \kappa < 1$ and $\kappa_1 = \frac{2\sqrt{\kappa}}{1+\kappa}$. From (A.2) and (A.3) it follows that

$$\mathbf{K}'(\kappa) = I(\kappa, 1) = I\left(\sqrt{\kappa}, \frac{1+\kappa}{2}\right) = \frac{2}{1+\kappa}\mathbf{K}'\left(\frac{2\sqrt{\kappa}}{1+\kappa}\right) = \frac{2}{1+\kappa}\mathbf{K}'(\kappa_1)$$

and

$$\mathbf{K}(\kappa_1) = \mathbf{K}'(\kappa_1') = I\left(\sqrt{1 - \frac{4\kappa}{(1+\kappa)^2}}, 1\right) = I\left(\frac{1-\kappa}{1+\kappa}, 1\right) = (1+\kappa)\, I(1-\kappa, 1+\kappa)$$

$$= (1+\kappa)\, I\left(\sqrt{1-\kappa^2}, 1\right) = (1+\kappa)\mathbf{K}(\kappa).$$

Hence,

$$\frac{\mathbf{K}'(\kappa)}{\mathbf{K}(\kappa)} = 2\,\frac{\mathbf{K}'(\kappa_1)}{\mathbf{K}(\kappa_1)}. \tag{A.8}$$

Now we set $\kappa_{-j} := 1/\lambda_{-j}$ and have by induction

$$\frac{\mathbf{K}'(\kappa_{-j})}{\mathbf{K}(\kappa_{-j})} = 2^j\,\frac{\mathbf{K}'(\kappa_0)}{\mathbf{K}(\kappa_0)},$$

and (A.6) yields the required estimate. ∎

In the following we will drop the term with $\frac{1}{2}$ in (A.7), since it can be replaced by a factor $1 + \mathcal{O}(\lambda_{-j}^{-1})$ that decreases exponentially.
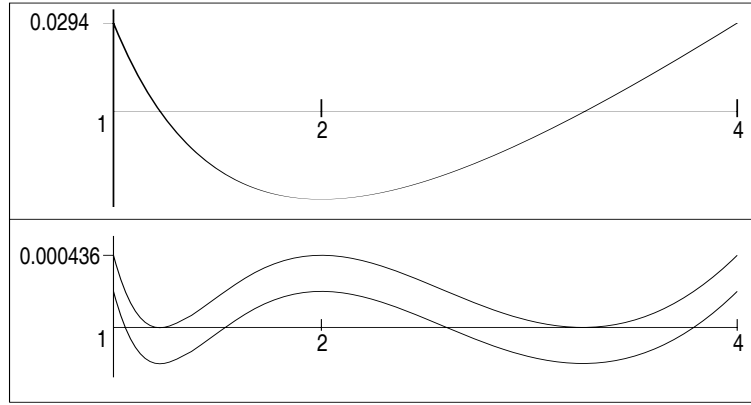

Figure A.1: In the upper figure the relative error of the best linear approximation (i.e., in $R_{0,1}$) on the interval $[1, 4]$ is depicted. Heron's method yields a one-sided rational approximation in $R_{2,1}$ with the number of oscillations doubled. The error of the one-sided approximation and of the rescaled one are depicted in the lower figure.

A.3 Heron's method and rational approximation

The square root of a given positive number is often computed by the iteration

$$v_{j+1} = \frac{1}{2}\left(v_j + \frac{x}{v_j}\right), \qquad j = 0, 1, 2, \ldots$$

The procedure is called *Heron's method* or sometimes the *Babylonian method*. The error reduction is described by

$$\frac{v_{j+1} - \sqrt{x}}{v_{j+1} + \sqrt{x}} = \frac{\frac{1}{2}(v_j + \frac{x}{v_j}) - \sqrt{x}}{\frac{1}{2}(v_j + \frac{x}{v_j}) + \sqrt{x}} = \left(\frac{v_j - \sqrt{x}}{v_j + \sqrt{x}}\right)^2. \tag{A.9}$$

We consider the weighted Chebyshev approximation on an interval $[a, b]$, i.e., we set $\|f\| := \sup_{a \leq x \leq b} \left| \frac{f(x)}{\sqrt{x}} \right|$. Assume that $u \in C[a, b]$ and

$$-\lambda \leq \frac{u(x) - \sqrt{x}}{\sqrt{x}} \leq +\lambda \qquad \text{for } x \in [a, b] \tag{A.10}$$

with some $\lambda < 1$. A simple calculation shows that $v := (1 - \lambda^2)^{-1/2} u$ satisfies

$$-\mu \leq \frac{v(x) - \sqrt{x}}{v(x) + \sqrt{x}} \leq +\mu \quad \text{where} \quad \frac{1 + \lambda}{1 - \lambda} = \left(\frac{1 + \mu}{1 - \mu} \right)^2. \tag{A.11}$$

Similarly $w := (1 - \lambda)^{-1} u$ provides a one-sided approximation

$$0 \leq \frac{w(x) - \sqrt{x}}{w(x) + \sqrt{x}} \leq +\lambda \qquad \text{for } x \in [a, b] \tag{A.12}$$

Moreover, the lower or upper bound in (A.10) is attained at some $x \in [a, b]$, if and only if the same holds for (A.11) and (A.12).

Now we turn to the families $R_{m,m-1}$ that contain the quotients p_m / q_{m-1} of polynomials of degree m and $m - 1$, respectively and set

$$E_{m,m-1} := \inf_{v \in R_{m,m-1}} \| \sqrt{x} - v \|.$$

For convenience, $R_{0,-1} = R_{0,0}$ is the set of constant functions. Let u be the best approximation to $\sqrt{x}$ from $R_{m,m-1}$, and let $v = (1 - \|f - u\|^2)^{-1/2} u$ be the associated function in the sense of (A.11). One step of Heron's method yields a one-sided approximation $\tilde{w} = \frac{1}{2}(v + x/v) \in R_{2m,2m-1}$. By rescaling $\tilde{w}$ in the sense of (A.12) we determine w. From (A.9) and the preceding formulae we conclude that

$$\frac{1 - \|f - u\|}{1 + \|f - u\|} = \left(\frac{1 - \|f - w\|^{1/2}}{1 + \|f - w\|^{1/2}} \right)^2.$$

It was possibly first discovered by Rutishauser in 1963 that w is the best approximation from $R_{2m,2m-1}$, if u is optimal in $R_{m,m-1}$. Indeed, if u is optimal, the weighted error function $(u - \sqrt{x})/\sqrt{x}$ has $2m + 1$ extreme points (with alternating signs). These points and the zeros of $v - \sqrt{x}$ make $4m + 1$ extreme points of $(w - \sqrt{x})/\sqrt{x}$ and w is the best approximation from $R_{2m,2m-1}$. Hence,

$$\frac{1 + E_{m,m-1}}{1 - E_{m,m-1}} = \left(\frac{1 + E_{2m,2m-1}^{1/2}}{1 - E_{2m,2m-1}^{1/2}} \right)^2.$$

This formula may be rewritten as

$$E_{m,m-1}^{-1} = \frac{1}{2}(E_{2m,2m-1}^{1/2} + E_{2m,2m-1}^{-1/2}),$$

and we recognise that $E_{2m,2m-1}^{-1}$ is obtained from $E_{m,m-1}^{-1}$ by the Landen transformation. Finally, the error of the best constant function is easily derived from the best one-sided constant, i.e., $\sqrt{b}$. Recalling (A.12), we conclude that

$$E_{0,0}^{-1} = \frac{\sqrt{b} + \sqrt{a}}{\sqrt{b} - \sqrt{a}}. \tag{A.13}$$

Now we are in a position to prove Lemma 2.1 for the special numbers $m = 2^j$. The sequence

$$E_{0,0}^{-1}, \ E_{1,0}^{-1}, \ E_{2,1}^{-1}, \ldots, E_{2^j, 2^j - 1}^{-1}, \ldots$$

can be estimated since we know $E_{0,0}$. The extension to other values of m is performed in [2] via a successive reduction of the interval $[a, b]$ by the arithmetic-geometric process and approximation theoretic arguments for small intervals.

We rewrite (A.13) as $E_{0,0} = \frac{1-\kappa}{1+\kappa}$ with $\kappa := \sqrt{a/b}$. Given $m = 2^j$, we obtain from Lemma A.1

$$E_{m,m-1} \leq 4 \exp\left[-m\frac{\pi \mathsf{K}'(\frac{1-\kappa}{1+\kappa})}{2\mathsf{K}(\frac{1-\kappa}{1+\kappa})}\right] . \tag{A.14}$$

We move to the complementary modulus and recall (A.8),

$$\frac{\mathsf{K}'(\frac{1-\kappa}{1+\kappa})}{\mathsf{K}(\frac{1-\kappa}{1+\kappa})} = \frac{\mathsf{K}\left(\sqrt{1-(\frac{1-\kappa}{1+\kappa})^2}\right)}{\mathsf{K}'\left(\sqrt{1-(\frac{1-\kappa}{1+\kappa})^2}\right)} = \frac{\mathsf{K}(\frac{2\sqrt{\kappa}}{1+\kappa})}{\mathsf{K}'(\frac{2\sqrt{\kappa}}{1+\kappa})} = 2\frac{\mathsf{K}(\kappa)}{\mathsf{K}'(\kappa)} .$$

After inserting this equality into (A.14) the proof is complete.

References

[1] J.M. Borwein and P.B. Borwein: *Pi and the AGM.* John Wiley & Sons, 1987.

[2] D. Braess: *Nonlinear Approximation Theory.* Springer-Verlag, Berlin, 1986.

[3] D. Braess: *Asymptotics for the approximation of wave functions by exponential sums.* J. Approximation Theory **83** (1995), 93–103.

[4] A.A. Gončar: *On the speed of rational approximation of some analytic functions.* Mat. Sbornik **105** (1978), 147–163 = Math. USSR Sb. **34** (1978), 131–145.

[5] L. Grasedyck: *Existence and computation of a low Kronecker-rank approximant to the solution of a tensor system with tensor right-hand side.* Computing 72 (2004), 247–265.

[6] W. Hackbusch and B. Khoromskij. *Hierarchical Kronecker tensor-product approximation to a class of nonlocal operators in high dimensions.* Report 16, Max-Planck-Institut für Mathematik, Leipzig, 2004.

[7] W. Kutzelnigg: *Theory of the expansion of wave functions in a Gaussian basis.* Int. J. of Quantum Chemistry **51** (1994), 447–463.

[8] E. Saff and V. Totik: *Logarithmic Potentials with External Fields.* Springer-Verlag, Berlin, 1997.

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

[10] F. Stenger: *Numerical Methods Based on Sinc and Analytic Functions.* Springer-Verlag, New York, 1993.

[11] N.S. Vjačeslavov: *On the least deviation of the function* $\mathrm{sign}\, x$ *and its primitives from the rational functions in the L_p metrics, $0 < p < \infty$.* Math. USSR Sbornik **32** (1977), 19–31.

Dietrich Braess
Mathematisches Institut
Ruhr-Universität Bochum
D-44780 Bochum
email: braess@num.rub.de

Wolfgang Hackbusch
Max-Planck-Institut für Mathematik in den Naturwissenschaften
Inselstrasse 22-26
D-04103 Leipzig, Germany
email: wh@mis.mpg.de