L^2 AND L^∞ RATIONAL APPROXIMATION ON THE UNIT DISK*

MICHAEL S. ACKERMANN†, SEAN REITER‡, AND LLOYD N. TREFETHEN§

Abstract. Using recently developed algorithms, we compute and compare best L^2 and L^∞ rational approximations of analytic functions on the unit disk. Although there is some theory for these problems going back decades, this may be the first computational study. To compute the L^2 best approximations, we employ a new formulation of TF-IRKA in barycentric form.

Key words. rational approximation, IRKA, AAA

MSC codes. 41A20, 65D15

1. Introduction.

Rational approximation is an old subject, going back to the 19th century, but computations can be challenging. In this paper we compute and compare best L^2 and L^∞ rational approximations of scalar analytic functions on the unit disk D in the complex plane $\mathbb{C}$. So far as we are aware, no such comparisons have been published before. The algorithms we rely on for our computations are variants of TF-IRKA for L^2 [7] and AAA-Lawson for L^∞ [25]. They are not fail-safe, but they succeed in many cases.

Let $n \geq 0$ be an integer and let R_n be the set of rational functions of degree n that are analytic in the closed unit disk $\overline{D}$, meaning functions that can be written in the form $r(z) = p(z)/q(z)$ where p and q are polynomials of degree (at most) n and all the roots $\{\pi_k\}$ of q satisfy $|\pi_k| > 1$. Our 2- and ∞-norms are defined by

$$\|f\|_2 = \left(\frac{1}{2\pi}\int_S |f(z)|^2|dz|\right)^{1/2}, \quad \|f\|_\infty = \sup_{z\in S}|f(z)|,$$

where S is the unit circle. Note that the division by 2π makes $\|\cdot\|_2$ the root-mean-square norm, implying $\|f\|_2 \leq \|f\|_\infty$ for any f. We will be concerned with approximation of functions f analytic in $\overline{D}$, so by the maximum modulus principle, approximation on the circle implies approximation in the disk. Our functions belong to the spaces $L^2(S)$, $L^\infty(S)$, $H^2(D)$, and $H^\infty(D)$ as usually defined, and the problems we address would often be described as instances of H^2 or H^∞ approximation. Given a function f analytic in the closed unit disk, we let r_2^* and r_∞^* denote degree n *best approximations* defined by

$$\|f - r_2^*\|_2 = \inf_{r\in R_n}\|f - r\|_2, \quad \|f - r_\infty^*\|_\infty = \inf_{r\in R_n}\|f - r\|_\infty.$$

It is known that these approximations exist but need not be unique; see section 3. (Best approximations in L^∞ are also known as Chebyshev approximations.) When we write r_2^* or r_∞^*, it is to be understood that this refers to any best approximation if there are several. We denote the optimal errors by

$$E_2 = \|f - r_2^*\|_2, \quad E_\infty = \|f - r_\infty^*\|_\infty.$$

*Submitted to the editors DATE.

†Dept. of Mathematics, Virginia Tech, Blacksburg, VA 24061, USA (amike98@vt.edu)

‡Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA (s.reiter@nyu.edu)

§School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA (trefethen@seas.harvard.edu)

Note that r_2^*, r_∞^*, E_2, and E_∞ all depend on f and n, although our notation does not indicate this.

Instead of the closed unit disk, one could equivalently consider approximation on the closed *exterior* of the disk by assuming f is analytic there, including at ∞. This is the usual formulation in discrete-time systems theory [1, 2, 10, 14], and it is in some respects more elegant mathematically, including for investigating the distribution of poles of approximants, which accumulate as $n \to \infty$ on curves defined by orthogonality or energy-minimization conditions [6]. By a Möbius transformation, one could go further and transplant f and r to any other disk or disk-exterior or half-plane, but whereas L^∞ approximation is invariant under such transplantations, L^2 approximation is not. In particular, the problem of L^2 approximation in a half-plane (H^2 approximation) requires the interpolation condition $r(\infty) = f(\infty)$ in order for the approximation error to be finite. Half-plane domains are the natural setting for the model order reduction and reduced order modeling of continuous-time dynamical systems, which is the context in which our L^2 algorithms have most often been applied; see section 5. We mention that recent work generalizes some of these ideas to regions other than disks, disk exteriors, and half-planes [8, 9].

We begin in sections 2–4 with examples of best L^2 and L^∞ approximations and discussion of associated mathematical properties. Sections 5 and 6 then discuss algorithms for L^2 approximation, first based on the state-space formulation of the rational approximant that is so familiar in the model order reduction and reduced order modeling community, then based on a barycentric form of the Hermite rational approximant, for which there is less previous literature. We will not give details of our L^∞ algorithm, because it consists of applying the AAA-Lawson method exactly as described in [25]. For the examples of this paper we sampled at 200 roots of unity, a number which could of course be increased for more complicated problems. Alternatively, one could equally well use adaptive sampling by the continuum AAA-Lawson algorithm of [11]. The AAA-Lawson algorithm, like the AAA algorithm it is derived from (which computes near-best rather than best approximations), does not work with poles, but is just a greedy descent method involving support points and barycentric coefficients based on samples of f on the unit circle. Poles of the rational approximant are only computed after the fact, once good support points and corresponding optimal coefficients have been found. Our L^2 algorithms presented in sections 5 and 6, by contrast, do work with poles, exploiting the optimality conditions given in Theorem 3.2 below, and they adaptively sample f and its derivative at points in the disk rather than on the circle.

2. Examples. Given f and an approximation $r \in R_n$, the *error function* is

$$e(z) = r(z) - f(z),$$

which is defined at least on the closed unit disk, and the *error curve* is the closed curve $e(S)$. The error functions for the best degree n approximations are denoted by e_2^* and e_∞^*. Figure 1, the first of several of our figures in the same format, plots error curves for best degree 3 approximations of $f(z) = \exp(4z)$. Here and in all our examples, except as indicated otherwise, we believe (usually without proof) that the approximations we plot are unique best approximations correct to plotting accuracy. Throughout this paper, we plot curves for L^2 best approximations in green and those for L^∞ best approximations in blue.

In Figure 1 as in many examples, the L^2 and L^∞ error curves are topologically the same but quantitatively quite different. Both curves wind around the origin

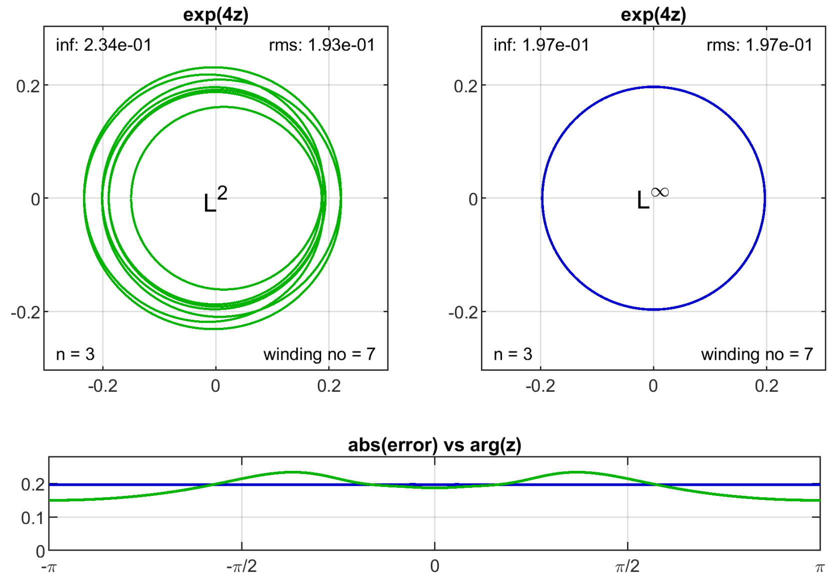

FIG. 1. *Error curves for best L^2 and L^∞ approximations of degree $n = 3$ to $f(z) = \exp(4z)$. Both curves have winding number $\omega = 2n + 1 = 7$. The error curve for the L^∞ case is not exactly circular but nearly so.*

$2n+1$ times, which is the generic number for degree n approximation though not the universal one. The L^2 curve is roughly circular, whereas the L^∞ curve is so close to circular that one cannot see in the plot that it is not exactly so. The near-circularity phenomenon was investigated in [30], and in this example the modulus $|e(z)|$ deviates from a constant value by about 3 parts in 1000. Theorem 6.3 of [30] indicates that for $f(z) = \exp(z)$ instead of $\exp(4z)$, the deviation from circularity would be less than one part in a million.

In the upper corners, Figure 1 lists the errors $\|e\|_2$ and $\|e\|_\infty$ for both the L^2 and L^∞ best approximations. Our definitions imply that these numbers are ordered as follows:

$$\|e_2^*\|_2 \le \|e_\infty^*\|_2 \le \|e_\infty^*\|_\infty \le \|e_2^*\|_\infty.$$

As our next examples, Figures 2 and 3 show approximations of $f(z) = \sqrt{1.1 - z}$ of degrees $n = 1$ and 7. It is interesting to note here the difference between the plots of error curves, in the upper part of each figure, and of $|e(z)|$ as a function of $\arg(z)$, in the lower part. From the error curves, one might be puzzled as to how the green approximations can be better in the 2-norm than the blue ones. The lower plots, however, make it clear that the L^2-optimal approximations have large errors only on a small portion of S near the singularity at $z = 1.1$.

It is also interesting to note in Figures 2 and 3 that, whereas L^∞ error curves tend to approach perfect circles as $n \to \infty$, this need not be the case in L^2. Thus in a relative sense, e_2^* and e_∞^* need not approach each other as $n \to \infty$. In an absolute sense, however, they do approach each other, for it is known that $\lim_{n\to\infty} E_2^{1/n}$ and $\lim_{n\to\infty} E_\infty^{1/n}$ are equal for a wide class of functions f [6, Thm. 5]. Such results are proved by methods related to the Hermite integral and potential functions discussed in section 4.

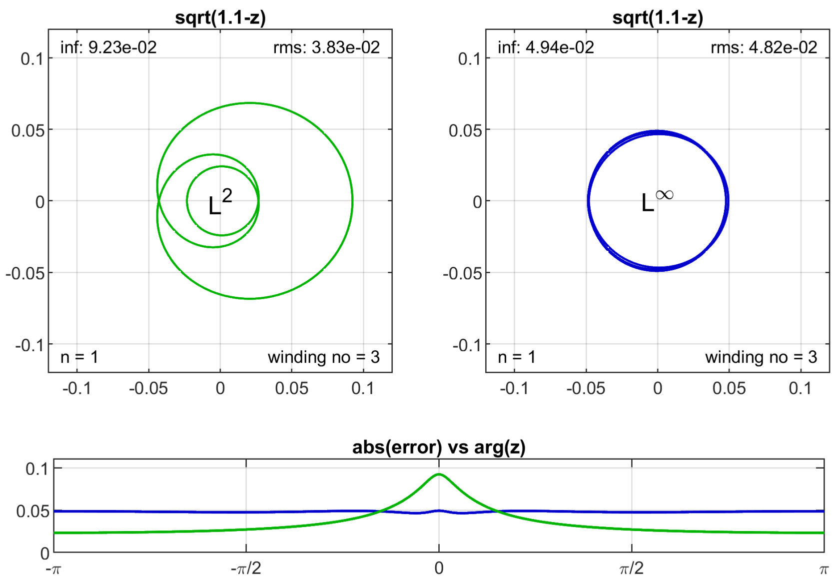

FIG. 2. *Degree* 1 *approximations of* $\sqrt{1.1-z}$.

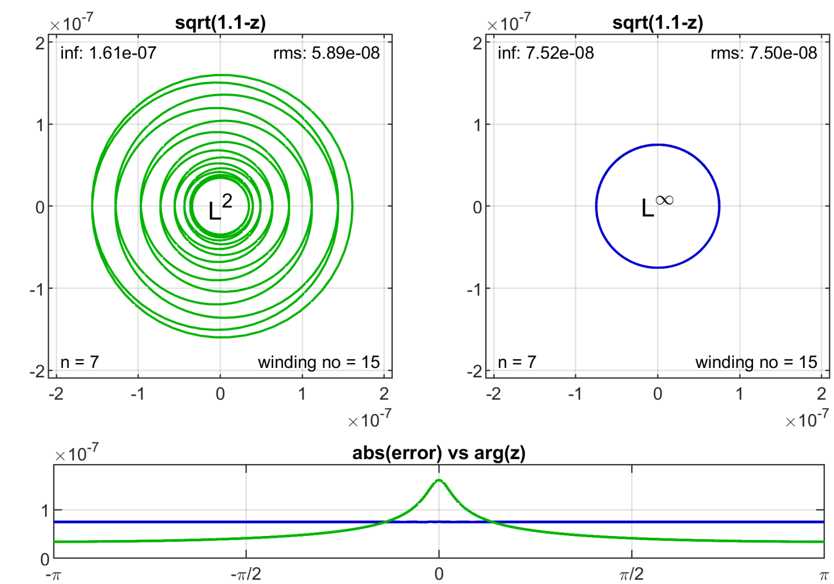

FIG. 3. *Degree* 7 *approximations of* $\sqrt{1.1-z}$.

As a fourth example, Figure 4 shows degree $n = 3$ approximations of $f(z) = \tan(z^3)$, a function with six-fold angular symmetry about $z = 0$. A new feature here is that the winding number for both approximations is $\omega = 9$, which exceeds $2n + 1 = 7$. The unique L^∞ best approximation is actually a polynomial, $r_\infty^*(z) \approx 1.159500940306z^3$, so its error curve is six-fold symmetric. The best L^2 approximation, on the other hand, has three finite poles equally spaced along a circle, so it cannot

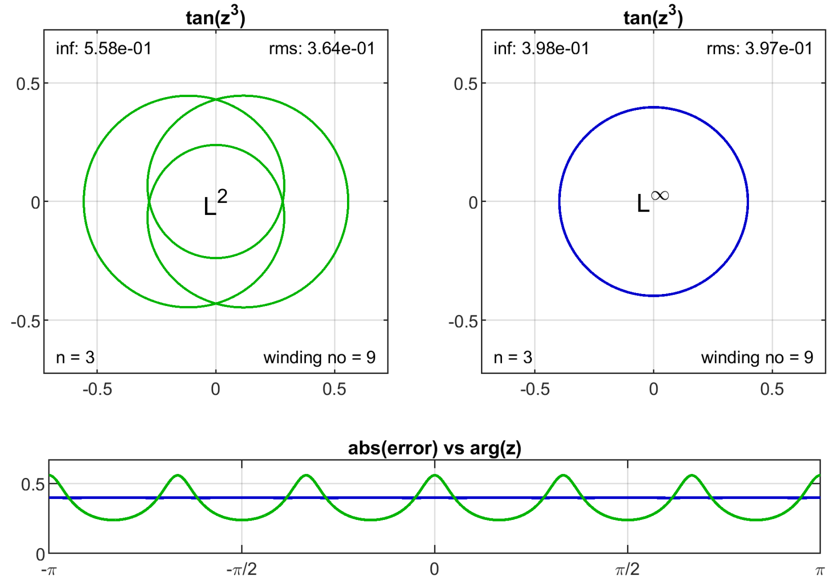

FIG. 4. *Degree 3 approximations of* $\tan(z^3)$. *The winding numbers exceed the generic value* $2n+1=7$. *The* L^2 *best approximation is nonunique, since it could be rotated by* $\pi/3$, *and the unique* L^∞ *best approximation is a polynomial,* $r^*_\infty(z) \approx 1.1596z^3$.

be exactly six-fold symmetric despite the impression given by the green curve of $|e(z)|$ against $\arg(z)$ in the lower image. The green error curve in the upper-left image appears to have winding number only 3, but because of the symmetry, it is actually traversed three times. This best approximation must be nonunique, since an equivalent approximation could be obtained by a rotation by an angle $\pi/3$ with a negation of sign. We comment on such degeneracies in the next section.

3. Optimality conditions. We make some comments on optimality conditions, without attempting a full discussion.

For L^∞, as mentioned in the introduction, best approximations on the unit disk are known to exist [32] but need not be unique [17].[1] The best known steps toward characterizing them are the *local Kolmogorov condition*, which is necessary for optimality (local as well as global), and the *global Kolmogorov condition*, also known as the *Meinardus-Schwedt condition*, which is sufficient for optimality [16, 22, 29]. We are not aware of a condition that is both necessary and sufficient, and it is notable that the AAA and AAA-Lawson algorithms, unlike previous less successful methods for L^∞ approximation and unlike our algorithms here for L^2 approximation, are not based on an attempt to enforce optimality conditions.

As a practical matter, best and near-best L^∞ approximations can usually be recognized by their nearly circular error curves, as illustrated in all the figures so far. In the simplest case, suppose $r \in R_n$ has an error curve $e(S)$ that happens to be a perfect circle with winding number $\omega \geq 2n+1$. Then if $\tilde{r}$ were a better approximation, the difference $r - \tilde{r}$ would also have to have winding number $\omega \geq 2n+1$, which is impossible since $r - \tilde{r} \in R_{2n}$: so r must be a best approximation. (This argument

[1]For example, Theorem 1 of [17] implies that $f(z) = z^2 + z^5$ has nonunique best approximations of degree $n = 1$.

amounts to an application of Rouché's theorem to the error functions $f-r$ and $f-\tilde{r}$.) This specialized conclusion generalizes in two important ways. First, if $r \in R_n$ is of degree less than n, belonging to R_ν for some $\nu < n$, then winding number $\omega \geq n+\nu+1$ is enough to imply optimality. Second, if the error curve is not perfectly circular but nearly circular, then the same reasoning ensures that r is a *near-best* approximant [30, Prop. 2.2]. This is a generalization to the unit disk of a familiar lemma in real approximation on an interval associated with the name of de la Vallée Poussin [22]:

THEOREM 3.1. NEARLY CIRCULAR $\Rightarrow$ NEAR-BEST FOR L^∞ APPROXIMATIONS. *Given a function f analytic in the closed unit disk, suppose $r \in R_\nu$ for some $\nu \leq n$ has an error curve $e(S)$ that does not pass through the origin and has winding number $\omega \geq \nu + n + 1$. Then*

$$\min_{z\in S} |e(z)| \leq E_\infty \leq \max_{z\in S} |e(z)|.$$

Theorem 4.1 of [16] generalizes this winding number condition to a related condition of alternation of signs along the unit circle, which can be used to prove the optimality of the best L^∞ approximation of Figure 4.

Note that although error curves for L^∞ best approximations often look circular to the eye, they can never be exactly circular unless f is rational. The proof is that if the error curve is circular, then by the Schwarz reflection principle, $e = r - f$ can be analytically continued to the extended complex plane $\mathbb{C} \cup \{\infty\}$, making it a meromorphic function with finitely many poles, hence rational.[2]

Turning to L^2, best approximations again exist but need not be unique. A helpful pair of sources on this topic are [5] and [6]. In this case there is a necessary condition for best approximation that is the basis of our computations. First of all, a best approximation $r_2^* \approx f$ must always satisfy $r_2^*(0) = f(0)$. This follows from the property that the L^2 norm over S of an analytic function with Taylor series $a_0 + a_1 z + \cdots$ is the 2-norm of its coefficient vector $(a_0, a_1, \dots)^T$, as can be shown by orthogonality; thus if an approximation r does not interpolate f at $z = 0$, it can be improved by addition of a constant. In addition, there must be at least $2n$ further points of interpolation lying in pairs—that is, points of *Hermite interpolation*—located at the reflections $\{\bar{\pi}_k^{-1}\}$ in the unit circle of the poles $\{\pi_k\}$ of r_2^*. The following theorem appears to have been published first in [19], generalizing an earlier result restricted to the case of simple poles [13] (in Russian). Analogous results for L^2 approximation in a half-plane can be found in [21].

THEOREM 3.2. INTERPOLATION CONDITIONS FOR BEST L^2 APPROXIMATION. *If $f \notin R_{n-1}$ is analytic in the closed unit disk, suppose $r_2^* \in R_n$ is a best L^2 approximation of f. Then $r_2^* \notin R_{n-1}$, so r_2^* has exactly n finite or infinite poles $\pi_1, \dots, \pi_M$, counted with multiplicities $m_1, \dots, m_M$; thus $\sum_{k=1}^M m_k = n$. Moreover, r_2^* interpolates f at the origin,*

$$r_2^*(0) = f(0), \tag{3.1}$$

and in at least $2n$ additional points counted with multiplicity in the following sense:

$$r_2^{*(j)}(\bar{\pi}_k^{-1}) = f^{(j)}(\bar{\pi}_k^{-1}), \quad j = 0, 1, \dots, 2m_k - 1, \quad k = 1, 2, \dots, M. \tag{3.2}$$

[2]As an undergraduate at Harvard, the third author spent weeks trying to prove this without success. Finally he knocked at the office of Prof. Ahlfors, who delivered the argument instantly while standing in the doorway.

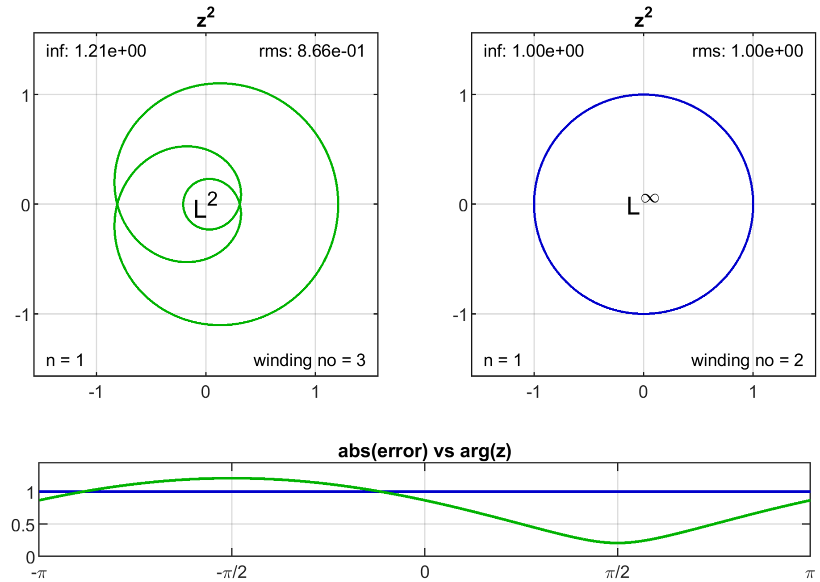

FIG. 5. *Degree* 1 *approximations of* z^2. *In* L^2, *this function has a continuum of nonunique best approximations, as described in the text, which have more points of interpolation than the (unique)* L^∞ *best approximant. The* L^∞ *error curve is our first example that is exactly rather than just nearly circular.*

If one of the poles is infinite, say $\pi_1 = \infty$, *then the* $k = 1$ *case of* (3.2) *becomes*

$$r_2^{*(j)}(0) = f^{(j)}(0), \quad j = 1, 2, \ldots, 2m_k. \tag{3.3}$$

Theorems 3.1 and 3.2 imply an interesting difference between L^∞ and L^2 approximation. For L^∞, the best approximation in R_n may be "defective," belonging to a space R_ν with $\nu < n$, but for L^2 it must always have exact degree n, implying that E_2 is a strictly decreasing sequence as $n \to \infty$ whenever f is not itself rational. As the simplest possible example, illustrated in Figure 5, consider best degree $n = 1$ approximation of $f(z) = z^2$ [19, p. 272]. In L^∞, the best approximation is $r_\infty^*(z) = 0$, with $E_\infty = 1$ and winding number 2, but in L^2, $r_2^*(z) = (ze^{i\theta}/2)/(ze^{-i\theta} - \sqrt{2}i)$ is a best approximation for any θ, with $E_2 = \sqrt{3}/2 \approx 0.866$ and winding number 3. Thus the best L^2 approximation is nonunique, as must hold more generally in any case where f is even and n is odd [5]. If f is perturbed slightly in this example, then instead of a nonunique global best approximation, we can expect to find that there may be a unique global best approximation and also one or more distinct approximations that are locally but not globally optimal [4]. As mentioned in the last section, the L^2 best approximation of Figure 4 is also nonunique.

4. Hermite integral and potential functions. The theory of rational approximations of an analytic function f on a complex domain K has been developed extensively in the past century. The fundamental tool is a Hermite contour integral involving a potential function $\phi(z)$. For a sharp form of the estimate applicable directly to the generic situation, we assume there are at least $2n+1$ interpolation points $\zeta_0, \ldots, \zeta_{2n} \in K$ with $r(\zeta_k) = f(\zeta_k)$, and we also assume that r has n poles, some of which may be at ∞. The poles and interpolation points are counted with multiplicity, and in the case of L^2 best approximation, Theorem 3.2 asserts that we can expect many interpolation points of multiplicity 2. As a physical interpretation of the potential function (more precisely, the log of its absolute value), we imagine that each interpolation point corresponds to a positive point charge of strength 1, and each pole is a negative point charge of strength 2. Here is the theorem as stated in section 12 of [26], a special case of [28, Lemma 2], with roots in earlier work of Walsh and Gonchar.

THEOREM 4.1. *Let f be analytic in the closure of a Jordan region Ω bounded by a Jordan contour Γ, and let the degree n rational function $r(z)$ interpolate f in $2n+1$ points $\zeta_0, \dots, \zeta_{2n}$ of a compact set $K \subset \Omega$, counted with multiplicity. Let $r \in R_n$ have n finite poles $\pi_1, \dots, \pi_n$ outside Ω, and define*

$$\phi(z) = \prod_{k=0}^{2n} (z - \zeta_k) \Bigg/ \prod_{k=1}^{n} (z - \pi_k)^2 . \tag{4.1}$$

Then for any $z \in K$,

$$f(z) - r(z) = \frac{1}{2\pi i} \int_\Gamma \frac{\phi(z)}{\phi(t)} \frac{f(t)}{t - z} \, dt. \tag{4.2}$$

If r has poles at ∞, then (4.2) *still holds with these poles dropped from the quotient* (4.1).

If *all* the poles are at ∞, then r reduces to a polynomial and we recover the Hermite integral for polynomial approximation presented for example in [31, Thm. 11.1].

The power of Theorem 4.1—or perhaps we should say the power of rational approximation—comes from the behavior of the quotient $\phi(z)/\phi(t)$ in (4.2). When the negative charges (poles) are well separated from the positive ones (interpolation points), this ratio will be exponentially small on K, establishing exponential decrease of the errors E_2 and E_∞ as $n \to \infty$. The theorem applies to any rational approximation $r \approx f$ with n finite poles that interpolates f in at least $2n + 1$ points of K. Having this many interpolation points is the generic situation for best and near-best approximations, and in cases when there are fewer, one may be able to apply the more general result of [28, Lemma 2], in which only linearized interpolation after multiplication by a denominator polynomial is required.

Figures 6–9 show contour plots of $|\phi(z)|$ for the examples of Figures 1–4. In each plot, red dots mark poles of $r(z)$, yellow dots mark interpolation points with $r(z) = f(z)$, and green dots mark Hermite interpolation points associated with Theorem 3.2 with $r(z) = f(z)$ and $r'(z) = f'(z)$. The contours are half-integer level curves of $\log_{10}(|\phi(z)|)$, and the first thing to note in the figures is that in each case the unit circle, the boundary of our approximation domain K, is approximately a level curve. This correlates with near-optimality, since the error (4.2) will be small when the points $\zeta_k \in K$ and $\pi_k \in \mathbb{C}\backslash\overline{\Omega}$ are in a near minimal-energy configuration. In the case of best L^2 approximation, by Theorem 3.2, the unit circle is *exactly* a level curve, for ϕ is a multiple of a finite Blaschke product (a rational function whose poles and zeros are reflections of each other in the unit circle).

Looking first at Figure 6, we see a fine illustration of the generic situation for L^2 and L^∞ best approximations. Both the L^2 and L^∞ approximants have three poles to the right of the unit circle, which lie at nearly but not exactly the same locations in the two cases. This approximate agreement of L^2 and L^∞ poles is common in our experience. The L^2 best approximation shows the poles reflected as three green dots,

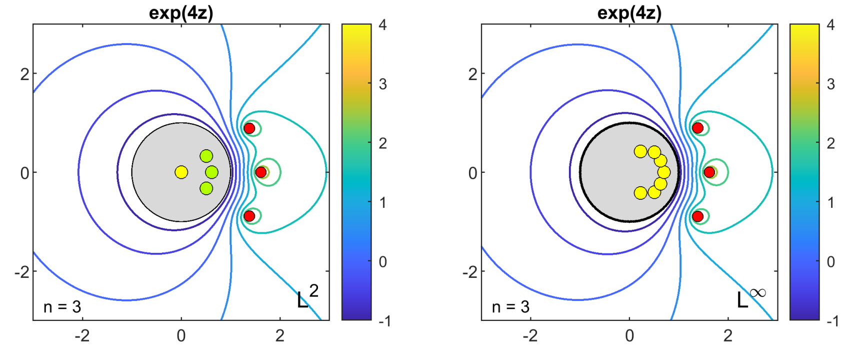

FIG. 6. *Potential contours for degree* 3 *approximations of* $\exp(4z)$*, following Figure* 1*. Contours show integer and half-integer levels of* $\log_{10}|\phi(z)|$*, with* ϕ *defined by* (4.1)*. Red dots mark poles of the approximant* r*, yellow dots mark interpolation points where* $r = f$*, and green dots marks double (Hermite) interpolation points with* $r = f$ *and* $r' = f'$*. By Theorem* 3.2*, for* L^2 *approximation there is a yellow dot at the origin and green dots at the points of reflection of the red dots in the unit circle.*

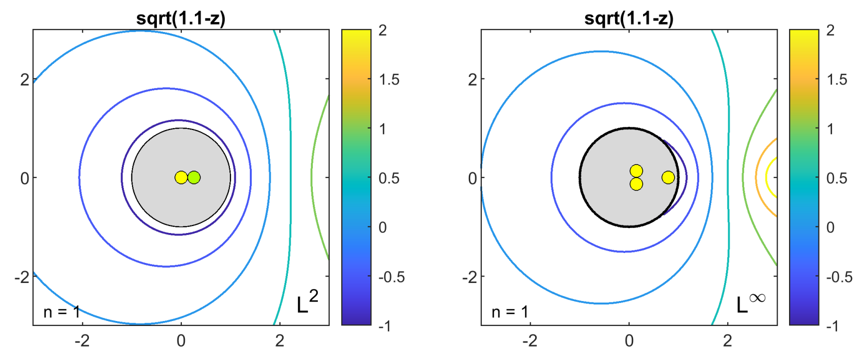

FIG. 7. *Potential contours for degree* 1 *approximations of* $\sqrt{1.1-z}$*, following Figure* 2*, with dots as in Figure* 6*. The pole* π_1 *is off-scale to the right at* $\pi_1 \approx 3.865$ *for* L^2 *and* $\pi_1 \approx 3.146$ *for* L^∞*.*

with double interpolation at each; in addition the yellow dot at the origin marks the seventh interpolation point. The L^∞ best approximation also has seven interpolation points in the disk, but now they are distinct and nonzero.

Figure 7 is similar with $n = 1$, but perhaps less clear than Figure 6 since the poles are off-scale. Figure 8 shows the configuration with the degree increased to 7, with the poles clustered near the singularity at $z = 1.1$.

5. L^2 approximation algorithm: state space form. The starting point of our L^2 computations is the Transfer Function (TF) variant of the Iterative Rational Krylov Algorithm (IRKA) [15], called TF-IRKA [7]. IRKA was originally developed as a method for the H^2-optimal model order reduction of large-scale linear dynamical systems in continuous time. Thus, the approximation domain was the right half-plane; a closely related algorithm set on the disk exterior is called MIRIAm [10]. The input-to-output response of such a system with N ordinary differential equations, p input ports, and q output ports is characterized by a degree-N matrix-valued rational

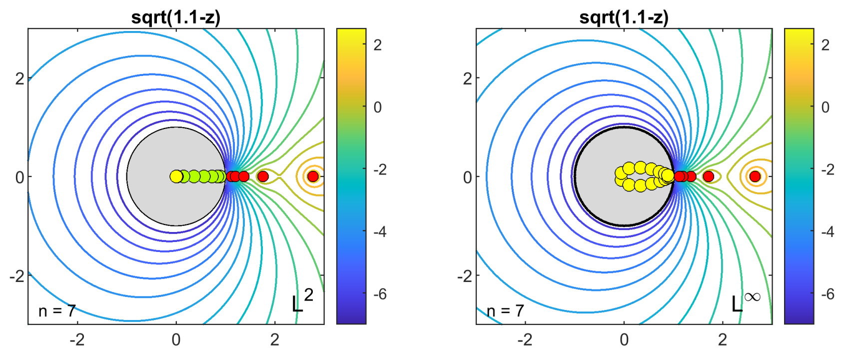

FIG. 8. *Potential contours for degree* 7 *approximations of* $\sqrt{1.1-z}$*, following Figure* 3*. In each case two of the seven poles* $\{\pi_k\}$ *are off-scale to the right.*

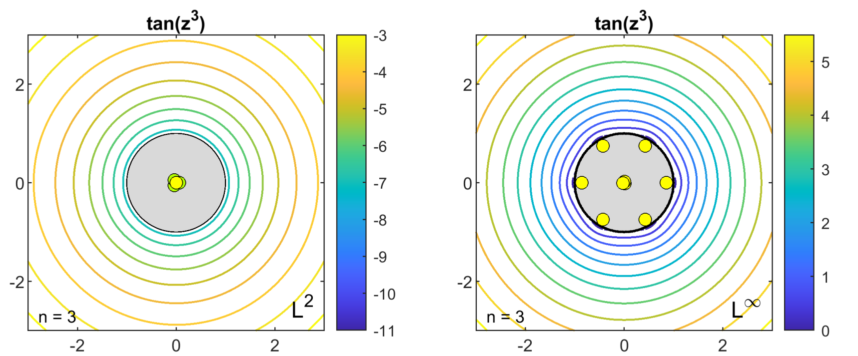

FIG. 9. *Potential contours for degree* 3 *approximations of* $\tan(z^3)$*, following Figure* 4*. The poles for* L^2 *approximation are off-scale with modulus about* 15.95*, leading to three Hermite interpolation points with modulus about* 0.06 *and three additional interpolation points at the origin. For* L^∞ *the approximation is a polynomial with a triple pole at* ∞*. Note that neither of these images shows* 6 *strings of poles radiating out from the disk to match the poles of* $\tan(z^3)$ *as one might expect, for the degree* $n = 3$ *is too small.*

function of dimensions $q \times p$, the *transfer function,*

$$\mathbf{H}(z) = \mathbf{C}^\top (z\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}, \tag{5.1}$$

where $\mathbf{E} \in \mathbb{R}^{N\times N}$ (nonsingular), $\mathbf{A} \in \mathbb{R}^{N\times N}$, $\mathbf{B} \in \mathbb{R}^{N\times p}$, $\mathbf{C} \in \mathbb{R}^{N\times q}$, and $\mathbf{D} \in \mathbb{R}^{q\times p}$ are the *state-space matrices* of $\mathbf{H}(z)$; see [1] for a general resource on system-theoretic model reduction. We refer to (5.1) as the state-space form of the rational function, and any degree N matrix-valued rational function of dimensions $q\times p$ can be expressed in this form [12, sec. 2.6.1]. (TF-)IRKA is an algorithm for computing low degree rational approximants of $\mathbf{H}$ that satisfy a set of Hermite interpolatory optimality conditions, akin to (3.2)[3]. In the context of model order reduction, the computed r_2^* is the transfer function of another linear dynamical system of degree n with the same structure and dimensions $q \times p$ as (5.1).

[3]In its original setting of [7, 15], the domain of analyticity is the open right half-plane, and the optimal interpolation points are the reflections of the poles in the imaginary axis, i.e., $\{-\overline{\pi}_k\}$.

For the following discussion, assume that the poles are simple, distinct, and finite. (TF-)IRKA computes r_2^* by an iterative process that adjusts the poles at each step until a fixed point is reached. Specifically, at each step, it produces a new rational approximant by interpolating doubly at the reflected poles of the previous iterate. Upon convergence—once the poles of the rational approximant stop changing up to a tolerance—the optimality conditions (3.2) are satisfied. Note that although these conditions are only necessary for optimality, since they will hold at maxima and stationary points as well as minima, in practice in our iteration they can be expected to behave as necessary and sufficient conditions for a local minimum. The reason is that stationary points other than minima act as repellers of the IRKA iteration, so the algorithm is unlikely to converge to them. (This repelling property has been published as a theorem in the case of L^2 approximation on a half-plane in the appendix of [18] and presumably holds on the disk too.)

The difference between IRKA and its TF variant is in how it achieves the interpolation. In its traditional formulation, at each step, IRKA computes a rational interpolant by projecting the matrices $\mathbf{E}, \mathbf{A}, \mathbf{B}, \mathbf{C}$ onto particularly chosen rational Krylov subspaces. By contrast, TF-IRKA is data-driven, meaning that it only requires the ability to sample f and f' (not necessarily rational) at points in $\mathbb{C}$, and these data are used to construct r_2^* in the form (5.1). This is accomplished using the *Loewner-matrix* approach for interpolation. For the remainder of this section, we switch from TF-IRKA's original formulation to describe how we utilize it to compute best L^2 approximations to functions analytic in the disk. That is, the final approximation will satisfy both (3.1) and (3.2). For our scalar approximation problems, we take $p = q = 1$. Given samples $f_i = f(\sigma_i)$ and derivatives $f_i' = f'(\sigma_i)$ at the points $\sigma_i \in \mathbb{C}$, define the Hermite Loewner and shifted Hermite Loewner matrices $\mathbf{L} \in \mathbb{C}^{n\times n}$ and $\mathbf{M} \in \mathbb{C}^{n\times n}$ by

$$\mathbf{L}_{ij} = \begin{cases} \dfrac{f_i - f_j}{\sigma_i - \sigma_j} & \text{if } \ i \neq j \\ f_i' & \text{if } \ i = j \end{cases}, \quad \mathbf{M}_{ij} = \begin{cases} \dfrac{\sigma_i f_i - \sigma_j f_j}{\sigma_i - \sigma_j} & \text{if } \ i \neq j \\ f_i + \sigma_i f_i' & \text{if } \ i = j \end{cases} \tag{5.2}$$

as well as the data matrix $\mathbf{Y} = \begin{bmatrix} f_1 & \cdots & f_n \end{bmatrix}^\top$. Construct $\mathbf{E}$, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ as

$$\mathbf{E} = -\mathbf{L}, \quad \mathbf{A} = -\mathbf{M}, \quad \mathbf{B} = \mathbf{Y}, \quad \mathbf{C} = \mathbf{Y}.$$

This achieves the $2n$ interpolation conditions (3.2). For the other condition (3.1), define

$$\mathbf{D} = \frac{d_0}{d_1 d_2 + d_3 d_0},$$

where

$$d_0 = f(0) + \mathbf{C}^\top \mathbf{A}^{-1}\mathbf{B}, \quad d_1 = -\mathbf{1}^\top \mathbf{A}^{-1}\mathbf{B} - 1$$

and

$$d_2 = -\mathbf{C}^\top \mathbf{A}^{-1}\mathbf{1} - 1, \quad d_3 = -\mathbf{1}^\top \mathbf{A}^{-1}\mathbf{1},$$

and update the $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ matrices according to

$$\mathbf{A} \leftarrow \mathbf{A} + \mathbf{D}, \qquad \mathbf{B} \leftarrow \mathbf{B} - \mathbf{D}, \qquad \mathbf{C} \leftarrow \mathbf{C} - \mathbf{D},$$

where $\mathbf{1} \in \mathbb{R}^n$ is the vector of all ones. By construction, the computed interpolant $r(z)$ constructed as in (5.1) satisfies the $2n+1$ interpolation conditions

$$f(\sigma_k) = r(\sigma_k), \quad f'(\sigma_k) = r'(\sigma_k), \quad f(0) = r(0)$$

for $k = 1, \dots, n$.

Due to its origins in model order reduction, (TF-)IRKA has primarily been employed for computing low-degree rational approximations to high-degree rational functions. However, since the iteration is data-driven, it can equally be employed to compute L^2-optimal approximants to arbitrary functions analytic on a closed half-plane or disk or disk exterior. For further details on TF-IRKA and the Hermite Loewner framework, see [2].

6. L^2 approximation algorithm: barycentric form. In this section we provide an alternative method for computing the rational interpolants at each step of the TF-IRKA iteration. Recall that the interpolatory barycentric form of a degree n rational function is

$$r(z) = \sum_{i=1}^{n+1} \frac{w_i f_i}{z - t_i} \Bigg/ \sum_{i=1}^{n+1} \frac{w_i}{z - t_i}\ , \tag{6.1}$$

where $\{t_i\}$ are support points, $\{f_i\}$ are function values, and $\{w_i\}$ are barycentric weights. For any choice of nonzero weights, $r(t) \to f(t_i)$ as $t \to t_i$, and we remove the removable singularities by defining $r(t_i) = f(t_i)$ accordingly. This interpolatory property of r allows us to satisfy the $n + 1$ Lagrange interpolation conditions of Theorem 3.2 "for free" by taking $t_1, \dots, t_n$ as the reflections of the current poles $\{\pi_i\}$ in the unit circle and $t_{n+1} = 0$. We then use the weights to satisfy the remaining n Hermite conditions at $t_1, \dots, t_n$. A value $d_k \in \mathbb{C}$ is the derivative of a rational function in barycentric form (6.1) at $z = t_k$ if

$$0 = w_k d_k + \sum_{i \neq k}^{n+1} \frac{w_i (f_k - f_i)}{t_k - t_i},$$

as is readily derived from the Schneider–Werner formula for the derivative of a barycentric quotient [27]. Hence, we can enforce that the derivative of the barycentric form takes prescribed values $\{d_i\}$ at $t_1, \dots, t_n$ by requiring that the vector of weights $\mathbf{w} \in \mathbb{C}^{n+1}$ satisfies

$$\mathbf{w} \in \text{null}(\widehat{\mathbf{L}})$$

where $\widehat{\mathbf{L}} \in \mathbb{C}^{n \times (n+1)}$ is the rectangular Hermite Loewner matrix defined by deleting the last row of $\mathbf{L}$ in (5.2). Similar expressions appear in [3, 23], but to our knowledge this is the first time the conditions for an exact Hermite interpolating rational function in barycentric form have appeared.

For the computations of this paper, we have used the barycentric method of this section rather than the Loewner framework method of the last section. While we have not conducted extensive tests, we have not noticed significant practical differences between the two approaches.

7. Discussion. Sample Matlab codes used to generate the figures of this paper can be obtained from the authors. We have not currently developed robust software for such computations.

On regions other than disks or half-planes, the AAA-Lawson method for L^∞ still applies, but the L^2 methods of sections 5 and 6 do not. However, as mentioned in the introduction, recent work by Borghi and Breiten represents steps in this direction [8, 9].

Rational approximations are not only useful for rational approximation. They can play a role in many other computational problems, and for a survey of dozens of such applications, see [26]. A theme in applications is that rational functions are often much more powerful than polynomials, especially when dealing with unbounded or nonconvex domains, or with functions having singularities on the domain boundary or nearby. On the other hand, we are not aware that there is a great difference in the approximation power of L^2 and L^∞ approximations on bounded domains.

Data availability statement. This paper does not make use of any data.

Acknowledgments. This project originated in April 2025 in Banff, Alberta at the BIRS Workshop on Challenges, Opportunities, and New Horizons in Rational Approximation organized by Anil Damle, Serkan Gugercin, and Heather Wilber, where MSA and SR met LNT and first began to compare L^2 and L^∞. We are also grateful for helpful suggestions from Laurent Baratchart and Maxim Yattselev.

REFERENCES

[1] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*, SIAM, 2005.

[2] A. C. Antoulas, C. Beattie, and S. Gugercin, *Interpolatory Methods for Model Reduction*, SIAM, 2020.

[3] L. Balicki, I. V. Gosea, and S. Gugercin, *Including derivatives to the barycentric representation of rational interpolants,* unpublished report (2022).

[4] L. Baratchart, *Existence and generic properties of L^2 approximants for linear systems,* IMA J. Math. Control & Infor., 3 (1986), pp. 89–101.

[5] L. Baratchart, *A remark on uniqueness of best rational approximants of degree* 1 *in L^2 of the circle,* Elect. Trans. Numer. Anal., 25 (2006), pp. 54–66.

[6] L. Baratchart, H. Stahl, and M. Yattselev, *Weighted extremal domains and best rational approximation*, Adv. Math., 209 (2012), pp. 357–407.

[7] C. Beattie and S. Gugercin, *Realization-independent $\mathcal{H}_2$ approximation*, in 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), IEEE, pp. 4953–4958.

[8] A. Borghi and T. Breiten, *Generalizing the optimal interpolation points for IRKA,* in MATHMOD Short Contribution volume 2025: 11th Vienna International Conference on Mathematical Modeling (2025), pp. 7–8.

[9] A. Borghi and T. Breiten, *Data-driven optimal approximation on Hardy spaces in simply connected domains,* Adv. Comput. Math., 51 (2025), pp. 59-1–59-33.

[10] A. Bunse-Gerstner, D. Kubalińska, G. Vossen, and D. Wilczek, *h_2-norm optimal model order reduction for large scale discrete dynamical MIMO systems,* J. Comput. Appl. Math., 233 (2010), pp. 1202–1216.

[11] T. A. Driscoll, Y. Nakatsukasa, and L. N. Trefethen, *AAA rational approximation on a continuum,* SIAM J Sci. Comput., 46 (2024), pp. A929–A952.

[12] G. Dullerud and F. Paganini, *A Course in Robust Control Theory: A Convex Approach,* Springer Science & Business Media (2013).

[13] V. D. Erohin, *On the best approximation of analytic functions by rational functions with free poles,* Dokl. Akad. Nauk SSSR, 128 (1959), pp. 29–32. (Russian)

[14] J. Grimm, *Rational approximation of transfer functions in the Hyperion software,* Technical Report 4002, INRIA (2000).

[15] S. Gugercin, A. C. Antoulas, and C. Beattie, *$\mathcal{H}_2$ model reduction for large-scale linear dynamical systems*, SIAM J. Matrix Anal. Applics., 30 (2008), pp. 609–638.

[16] M. H. Gutknecht, *On complex rational approximation part I: The characterization problem,* Computational Aspects of Complex Analysis: Proceedings of the NATO Advanced Study Institute held at Braunlage, Harz, Germany, July 26–August 6, 1982, Springer Netherlands (1983).

[17] M. H. Gutknecht and L. N. Trefethen, *Nonuniqueness of best rational approximations on the unit disk,* J. Approx. Th., 39 (1983), pp. 275–288.

[18] W. Krajewski, A. Lepschy, M. Redivo-Zaglia, and U. Viaro, *A program for solving the L_2 reduced-order model problem with fixed denominator degree,* Numer. Algs., 9 (1995), pp. 355–377.

[19] A. L. Levin, *The distribution of poles of rational functions of best approximation and related questions,* Math. USSR-Sbornik, 9 (1969), pp. 267–274.

[20] A. J. Mayo and A. C. Antoulas, *A framework for the solution of the generalized realization problem*, Lin. Alg. Applics., 425 (2007), pp. 634–662.

[21] L. Meier and D. Luenberger, *Approximation of linear constant systems,* IEEE Trans. Aut. Control, 2 (1967), pp. 585–588.
[22] G. Meinardus, *Approximation of Functions: Theory and Numerical Methods,* Springer (1967).
[23] W. Mitchell, *Two intriguing variants of the AAA algorithm for rational approximation,* arXiv:2508.11169 (2025).
[24] Y. Nakatsukasa, O. Sète, and L. N. Trefethen, *The AAA algorithm for rational approximation*, SIAM J. Sci. Comput., 40 (2018), pp. A1494–A1522.
[25] Y. Nakatsukasa and L. N. Trefethen, *An algorithm for real and complex rational minimax approximation*, SIAM J. Sci. Comput., 42 (2020), pp. A3157–A3179.
[26] Y. Nakatsukasa and L. N. Trefethen, *Applications of AAA rational approximation*, Acta Numer., 35 (2026), to appear.
[27] C. Schneider and W. Werner, *Some new aspects of rational interpolation,* Math. Comput. 47 (1986), pp. 285–299.
[28] H. Stahl, *On the convergence of generalized Padé approximants,* Constr. Approx., 5 (1989), pp. 221–240.
[29] J.-P. Thiran and M.-P. Istace, *Optimality and uniqueness conditions in complex rational Chebyshev approximation with examples,* Constr. Approx., 9 (1993), pp. 83–103.
[30] L. N. Trefethen, *Rational Chebyshev approximation on the unit disk*, Numer. Math., 37 (1981), pp. 297–320.
[31] L. N. Trefethen, *Approximation Theory and Approximation Practice, extended edition,* SIAM (2019).
[32] J. L. Walsh, *The existence of rational functions of best approximation*, Trans. AMS, 33 (1931), pp. 668–689.