THE FIRST FIVE YEARS OF THE AAA ALGORITHM

YUJI NAKATSUKASA*, OLIVIER SÈTE†, AND LLOYD N. TREFETHEN‡

Abstract. The AAA algorithm, introduced in 2018, computes best or near-best rational approximations to functions or data on subsets of the real line or the complex plane. It is much faster and more robust than previous algorithms for such problems and has been used in many applications since its appearance, including the numerical solution of Laplace, Poisson, and biharmonic PDE problems in irregular domains. AAA has also been extended in new directions and seems likely to be a tool of lasting importance in the future.

Key words. AAA algorithm, rational approximation, minimax, Laplace problem

MSC codes. 41A20, 65D15

The AAA algorithm is a numerical method for rational approximation of a function f on a real or complex domain. The three of us introduced the method in 2018 [29] together with a computer program `aaa` in the Chebfun package [11], which also runs standalone in MATLAB or Octave. AAA, pronounced "triple A" and derived from "adaptive Antoulas-Anderson," is much faster and more robust than previous methods in this area, and it is changing perceptions of how rational approximations can be used in applications.

For a simple example, the MATLAB code segment

```
Z = exp(2i*pi*(1:100)/100);
r = aaa(exp(Z),Z);
```

computes an approximation to $\exp(z)$ in 100 roots of unity with the default relative accuracy of 10^{-13}. The computation takes 0.002 s on our laptop and delivers a degree 7 rational function whose maximum deviation from e^z on the unit disk is 2.81×10^{-15}. For another example, the code segment

```
Z = linspace(4-50i,4+50i);
zeta = @(z) sum((1e4:-1:1).^(-z),2);
r = aaa(zeta(Z),Z);
```

evaluates the Riemann zeta function at 100 complex points with $\mathrm{Re}(z) = 4$ and then fits this data in 0.1 s by a rational function of degree 37. The phase portrait of r [40], shown in Figure 1, closely captures that of the zeta function itself in the striped region. The first two zeros of r in the upper half-plane are at $0.4999999987 + 14.1347251412i$ and $0.4999999987 + 21.0220396409i$, matching the corresponding zeros of $\zeta(z)$ to 8 digits. The pole of r near $z = 1$ has similar accuracy, falling at $z \approx 1.0000000069 + 0.0000000009i$.

AAA has become a standard method for many computations, with a rapidly growing literature. Applications to date include analytic continuation [9, 36], interpolation of equispaced data [21], Laplace problems with applications to magnetics [6, 7], conformal mapping [16], Stokes flow [43], nonlinear dynamics [9, 36], chemical

*Mathematical Institute, U. of Oxford, Oxford OX2 6GG, UK (nakatsukasa@maths.ox.ac.uk).

†Inst. Math. u. Informatik, University of Greifswald, 17489 Greifswald, Germany (olivier.sete@uni-greifswald.de).

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

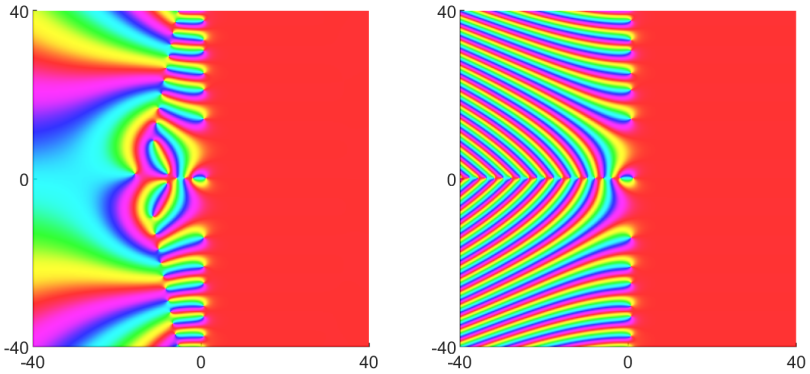

FIG. 1. *Phase portrait of the Riemann zeta function (right) compared with that of a rational approximation based on* 100 *sample points in the right half-plane (left). The AAA algorithm computes this approximation in* 0.1 *s on a laptop.*

physics [24], microwave tubes [42], simulation of turbulence [25], nonlinear eigenvalue problems [19, 27, 33], finite element linearizations [8], design of preconditioners [3, 10], model order reduction [1, 17, 23, 32], and signal processing [10, 20, 28, 38, 41]. For signal processing, AAA has been adopted as the basis of the `rational` code in the MathWorks RF Toolbox [28], which includes practical enhancements that our own software lacks like the option to impose real symmetry. There have also been generalizations to multivariate approximation [4, 17, 20, 27].

The starting point of AAA is the representation of a rational function $r(z)$ not in the form $p(z)/q(z)$ with polynomials p and q, but by a *barycentric quotient* [2]

$$r(z) = \frac{n(z)}{d(z)} = \sum_{j=1}^{m} \frac{w_j f(s_j)}{z - s_j} \Bigg/ \sum_{j=1}^{m} \frac{w_j}{z - s_j}\,, \tag{1}$$

where $s_1, \dots, s_m$ are a set of m *support points* and $w_1, \dots, w_m$ are a set of nonzero *barycentric weights*. This perhaps counterintuitive formula can represent a degree $m-1$ rational function in a numerically stable manner even when the zeros and poles of r are clustered near singularities, which is one of the sources of the power of rational approximations. The p/q representation fails in such cases because p and q vary widely in magnitude over the approximation domain, causing loss of accuracy in floating point arithmetic even if they are expanded in the best possible basis.

The AAA algorithm starts from a real or complex discrete sample set Z and a set F of corresponding function values. It consists of an alternation between a nonlinear greedy step, in which the next support point z_m is chosen as the sample point where the error is largest, and a linear algebra step, in which a singular value decomposition (SVD) of a rectangular Loewner matrix A is computed to determine coefficients $\{w_j\}$ to minimize the discretized least-squares error $f(z)d(z) - n(z)$ for $z \in Z$. The rows of A correspond to sample points and the columns to support points. For details, including a 40-line prototype computer code, see [29].

AAA-Lawson algorithm for minimax approximation. In its original mode of operation, AAA computes a near-best as opposed to a best approximation (by which we mean a *minimax* approximation, minimizing the L^∞ error, as is common in approximation theory). However, with a second phase of iteratively reweighted least-squares

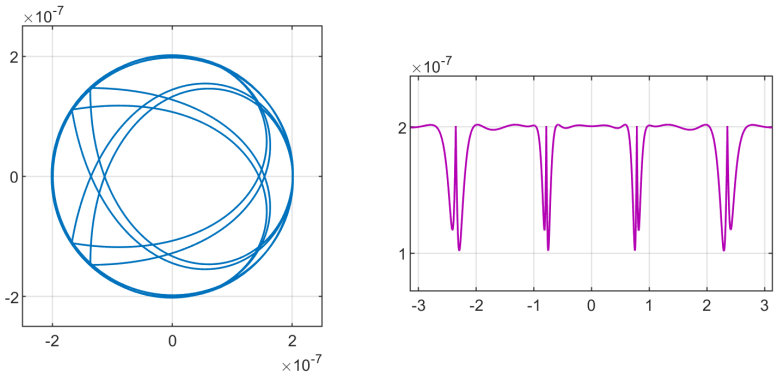

FIG. 2. *The image on the left shows the error curve* $(f-r)(Z)$ *for best degree* $n=4$ *approximation of* e^z *on the unit square by the AAA-Lawson algorithm* [12, 30]. *The image on the right shows* $|(f-r)(z)|$ *as a function of* $\arg(z)$. *The error curve is a curve of winding number* $2n+1=9$ *with nearly constant modulus except near the four corners* [34].

(IRLS) iteration, it is possible in most cases to improve near-best to best, as shown in 2020 [30]. We call this the *AAA-Lawson method,* since the iteration is a nonlinear barycentric analogue of the IRLS method introduced by Lawson in 1961 for linear approximation problems [26]. This is the only known robust fast algorithm for rational best approximation where the function or the domain or both are complex. (For real approximation on a real interval, there is the Remez algorithm, which for numerical stability must also be based on the barycentric representation [14].) An illustration is given in Figure 2. AAA-Lawson is a valuable tool for approximation theory explorations, but we do not recommend minimax approximation as the starting point for most applications, since standard AAA is more robust, especially for accuracies within a few orders of magnitude of machine precision.

AAA approximation on a continuum. The original AAA algorithm works with a discrete sample set Z, which in applications is typically the discretization of a continuum by hundreds or thousands of points. (Since poles and zeros will cluster near singularities, it is important to cluster the sample points too near corners of the domain.) It is natural to want an algorithm that can deal with a continuum more directly. Recently a variant of AAA has been developed for this and supported with open-source MATLAB and Julia software [12]. This adds speed, robustness, and simplicity to a AAA computation. For example, the command

```
r = aaax(@abs,80);
```

computes a near-best degree 80 approximation to $|x|$ on $[-1,1]$ in 0.2 s on our laptop based on the evaluation of $|x|$ in a total of only about 500 points. The accuracy is about 3.8×10^{-10}, not too far from the best approximation error of this degree, 4.4×10^{-12}. This is the classic problem made famous by Donald Newman in 1964 [31], a prototype of the general phenomenon of root-exponential convergence of rational approximations to functions with branch point singularities [35, chap. 25]. It is a very difficult problem computationally, for the zeros and poles of r cluster exponentially near the singularity at $x=0$ at distances as small as 4.3×10^{-10} [37]. Varga, Carpenter, and Ruttan required many hours of computing in 200-digit extended precision arithmetic thirty years ago to compute such approximations using the Remez algo-

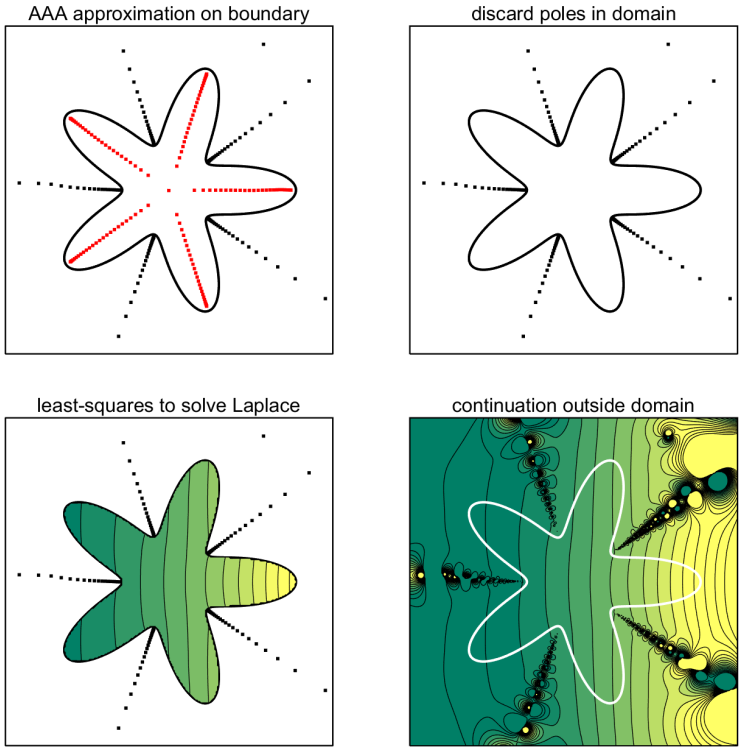

FIG. 3. *The AAA-LS method applied to solve a Laplace Dirichlet problem on a domain with five lobes. The solution computed in about* 1 s*, with accuracy about* 7×10^{-7}*, is represented as the real part of a rational function with* 125 *poles and is evaluable in about* 1 μs *per point. A polynomial representation of the same accuracy would require a degree in the tens or hundreds of thousands. The fourth panel shows evaluations outside the problem domain, revealing branch cuts of the harmonic continuation delineated by poles of the AAA approximation.*

rithm with a p/q representation [39]. The currently most robust implementation of the Remez algorithm, the Chebfun `minimax` command, takes 100 s to find this best approximation.

AAA-LS algorithm for Laplace and related problems. Perhaps the most fundamentally important and surprising extension and application of AAA methods was introduced in an arXiv paper by Stefano Costa in 2020 [5] and is being actively extended in various directions [6, 7, 43]. Since the days of Joseph Walsh and Mstislav Keldysh nearly a century ago, it has been recognized that if rational functions are good at approximating analytic functions, then their real parts will be good at approximating harmonic functions (i.e., solutions of the Laplace equation). With AAA as a new tool for rational approximation, it was natural to expect this to lead to new methods for solving Laplace and related problems. However, this is not straightforward. For example, suppose we wish to solve $\Delta u = 0$ in a simply-connected domain Ω subject to a real boundary condition $u(z) = h(z)$ on the boundary $\partial\Omega$. AAA enables us quickly to calculate an approximation $r \approx h$, but r will have poles in Ω as well as outside. In other words, it approximates an analytic extension of h to a neighborhood of $\partial\Omega$, but not to all of Ω. So how can it help in solving the Laplace problem?

Costa realized that this challenge can be addressed by separating the poles of the

AAA approximant into those inside and those outside Ω. The functions $1/(z - s_j)$ defined by poles $\{s_j\}$ outside Ω span a space that contains excellent approximations to the solution of the Laplace problem, which can be found numerically by a linear least-squares calculation. Thus the AAA-least squares (AAA-LS) method consists of these steps, as illustrated in Figure 3: (1) approximate the boundary data by a rational function, (2) collect just the poles outside Ω to form a basis for an approximation space, (3) solve a least-squares problem in this basis. Each pole leads to two columns of the least-squares matrix, corresponding to the real and imaginary parts of $1/(z-s_j)$. The method proves extremely powerful in practice, as is shown most recently in the computations of Y. Xue of solutions to the biharmonic equation for Stokes flow [43] and in the solution of certain Poisson problems. Initial theoretical justifications appear in [7], but there is much more to be said, and further work on the theoretical side is underway.

The separation of poles across two sides of a contour has many potential applications beyond the solution of Laplace problems. This is the starting point of effective numerical methods for Wiener–Hopf and more general Riemann–Hilbert problems. Closely related topics are the Hilbert transform and Dirichlet-to-Neumann maps, and the AAA-LS approach is effective for all of these. Examples can be found in [5, 6, 7], and there are many further possibilities. In all these applications, rational functions offer the prospect of strikingly fast and accurate computations in the presence of singularities or near-singularities, when other methods tend to converge very slowly or require case-by-case analysis of detailed behavior at singularities.

Discussion. Unlike other algorithms for rational approximation, such as the Ellacott–Williams and Istace–Thiran algorithms or the well known method of Vector Fitting [13, 18, 22], AAA is not based on an attempt to enforce a condition of optimality. (Discussions of these and other algorithms can be found in [29] and [30].) This appears to be one of the secrets of its success. As often happens in optimization, the most difficult part of the problem may be not what to do near the solution, but how to get near it in the first place. AAA seems extraordinarily good at this, for reasons not yet fully understood. There is an interesting analogy with training algorithms for deep learning, where optimization is carried out by methods such as stochastic gradient descent that have little to do with optimality conditions [15].

As a generalization of polynomials to functions with poles not constrained to lie at infinity, rational functions have a fundamental importance in computational science, and AAA is opening many doors. Rational approximation, however, is only one long-established nonlinear optimization problem among many. Another important one is the approximation of a function $f(x)$ by a linear combination of translates of Gaussians of unknown positions and widths. Before AAA, one would have imagined that these were problems of comparable (and considerable) difficulty. With the appearance of AAA, rational approximation suddenly seems rather easy in practice. Are there analogous algorithms for other problems such as fitting by Gaussians? We simply do not know, and we regard this as a fascinating and important challenge.

REFERENCES

[1] A. C. Antoulas, C. A. Beattie, and S. Güğercin, *Interpolatory Methods for Model Reduction,* SIAM (2020).

[2] J.-P. Berrut and L. N. Trefethen, *Barycentric Lagrange interpolation,* SIAM Rev., 46 (2004), pp. 501–517.

[3] A. Budiša, X. Hu, M. Kuchta, K.-A. Mardal, and L. Zikatanov, *Rational approximation pre-*

conditioners for multiphysics problems, Intl. Conf. Numer. Methods and Applics. Springer (2022), pp. 100–113.
[4] A. Carracedo Rodriguez, L. Balicki, and S. Güğercin, *The p-AAA algorithm for data driven modeling of parametric dynamical systems,* SIAM J. Sci. Comput., 45 (2023), pp. A1332–A1358.
[5] S. Costa, *Solving Laplace problems with the AAA algorithm,* arXiv:2001.09439v1, 2020.
[6] S. Costa, E. Costamagna, and P. Di Barba, *Modelling permanent-magnet excited uniform fields with rational approximations,* manuscript submitted to ISEF 2023 (Pavia), 2023.
[7] S. Costa and L. N. Trefethen, *AAA-least squares rational approximation and solution of Laplace problems,* In A. Hujdurović, et al., eds., European Congress of Mathematics, EMS Press, 2023.
[8] E. Deckers, S. Jonckheere, and K. Meerbergen, *Time integration of finite element models with nonlinear frequency dependencies,* arXiv:2206.09617 (2022).
[9] G. Deng and C. J. Lustri, *Exponential asymptotics of woodpile chain nanoptera using numerical analytic continuation,* Stud. Appl. Math., 150 (2023), pp. 520–557.
[10] N. Derevianko, G. Plonka, and M. Petz, *From ESPRIT to ESPIRA: estimation of signal parameters by iterative rational approximation,* IMA J. Numer. Anal., 43 (2023), pp. 789–827.
[11] T. A. Driscoll, N. Hale, and L. N. Trefethen, Chebfun Guide, Pafnuty Press, Oxford, 2014; see also `www.chebfun.org`.
[12] T. A. Driscoll, Y. Nakatsukasa, and L. N. Trefethen, *The AAA algorithm on a continuum,* SIAM J. Sci. Comput., to appear.
[13] S. Ellacott and J. Williams, *Rational Chebyshev approximation in the complex plane,* SIAM J. Numer. Anal., 13 (1976), pp. 310–323.
[14] S.-I. Filip, Y. Nakatsukasa, L. N. Trefethen, and B. Beckermann, *Rational minimax approximation via adaptive barycentric representations,* SIAM J. Sci. Comput., 40 (2018), pp. A2427–A2455.
[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning,* MIT Press, 2016.
[16] A. Gopal and L. N. Trefethen, *Representation of conformal maps by rational functions,* Numer. Math., 142 (2019), pp. 359–382.
[17] I. V. Gosea and S. Güttel, *Algorithms for the rational approximation of matrix-valued functions,* SIAM J. Sci. Comput., 43 (2021), pp. A3033–A3054.
[18] B. Gustavsen and A. Semlyen, *Rational approximation of frequency domain responses by vector fitting,* IEEE Trans. Power Deliv., 14 (1999), pp. 1052–1061.
[19] S. Güttel, G. M. Negri Porzio, and F. Tisseur, *Robust rational approximations of nonlinear eigenvalue problems,* SIAM J. Sci. Comput., 44 (2022), pp. A2439–A2463.
[20] A. Hochman, *FastAAA: A fast rational-function fitter,* in 2017 IEEE 26th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), IEEE (2017), pp. 1–3.
[21] D. Huybrechs and L. N. Trefethen, *AAA interpolation of equispaced data,* BIT Numer. Math., 2023.
[22] M.-P. Istace and J.-P. Thiran, *On computing best Chebyshev complex rational approximants,* Numer. Algorithms, 5 (1993), pp. 299-308.
[23] W. R. Johns, L. Monzón, M. Reynolds, and L. Davis, *Iterative stability enforcement in adaptive Antoulas-Anderson algorithms for H2 model reduction,* SIAM J. Sci. Comput., 45 (2023), A1844–A1861.
[24] M. Kehry, W. Klopper, and C. Holzer, *Robust relativistic many-body Green's function based approaches for assessing core ionized and excited states,* J. Chem. Phys., 159 (2023), pp. 044116-1–044116–14.
[25] B. Keith, U. Khristenko, and B. Wohlmuth, *A fractional PDE model for turbulent velocity fields near solid walls,* J. Fluid Mech., 916 (2021), pp. A21-1–30.
[26] C. L. Lawson, *Contributions to the Theory of Linear Least Maximum Approximations,* PhD thesis, UCLA, 1961.
[27] P. Lietaert, K. Meerbergen, J. Pérez, and B. Vandereycken, *Automatic rational approximation and linearization of nonlinear eigenvalue problems,* IMA J. Numer. Anal., 42 (2022), pp. 1087–1115.
[28] The MathWorks, Inc., *RF Toolbox*, `https://www.mathworks.com`.
[29] Y. Nakatsukasa, O. Sète, and L. N. Trefethen, *The AAA algorithm for rational approximation,* SIAM J. Sci. Comput., 40 (2018), pp. A1494–A1522.
[30] Y. Nakatsukasa and L. N. Trefethen, *An algorithm for real and complex rational minimax approximation,* SIAM J. Sci. Comput., 42 (2020), pp. A3157–A3179.
[31] D. J. Newman, *Rational approximation to $|x|$,* Mich. Math. J., 11 (1964), pp. 11–14.

[32] D. Pradovera, *Toward a certified greedy Loewner framework with minimal sampling,* arXiv:2303.01015v1 (2023).
[33] Y. Saad, M. El-Guide, and A. Miedlar, *A rational approximation method for the nonlinear eigenvalue problem,* arXiv:1901.01188v2 (2020).
[34] L. N. Trefethen, *Near-circularity of the error curve in complex Chebyshev approximation,* J. Approx. Thy., 31 (1981), pp. 344–367.
[35] L. N. Trefethen, *Approximation Theory and Approximation Practice, Extended Edition,* SIAM, 2019.
[36] L. N. Trefethen, *Numerical analytic continuation,* Japan J. Appl. Math. (2023), pp. 1–50.
[37] L. N. Trefethen, Y. Nakatsukasa, and J. A. C. Weideman, *Exponential node clustering at singularities for rational approximation, quadrature, and PDEs,* Numer. Math., 147 (2021), pp. 227–254.
[38] A. Valera-Rivera and A. E. Engin, *AAA algorithm for rational transfer function approximation with stable poles,* Lett. Electr. Compatibility Prac. and Applics., 3 (2021), pp. 92–95.
[39] R. S. Varga, A. Ruttan, and A. J. Carpenter, *Numerical results on best uniform approximation of $|x|$ on $[-1,1]$,* Math. USSR-Sb., 74 (1993), pp. 271–290.
[40] E. Wegert, *Visual Complex Functions: An Introduction with Phase Portraits,* Springer Science & Business Media, 2012.
[41] H. Wilber, A. Damle, and A. Townsend, *Data-driven algorithms for signal processing with trigonometric rational functions,* SIAM J. Sci. Comput., 44 (2022), pp. C185–C209.
[42] C. Xue et al., *A rational function approximation algorithm for the frequency sweep of microwave tube,* 24th Intern. Vaccum Elect. Conf. (IVEC) (2023).
[43] Y. Xue, S. L. Waters, and L. N. Trefethen, *Computation of 2D Stokes flows via lightning and AAA rational approximation,* arXiv:2306.13545v1, 2023.