Uniform and unitary rational approximations of the matrix exponential

Pranav Singh (University of Bath)

ps2106@bath.ac.uk   www.pranavsingh.co.uk   @denoising   @brownadder

27th of June 2023

29th Biennial Numerical Analysis Conference
Strathclyde

Overview

Based on joint work with

Tobias Jawecki (TU Wien)

[JS 23] *Unitarity of some barycentric rational approximants*, Jawecki & S, arXiv:2205.10606 [math.NA]

Overview

Based on joint work with

Tobias Jawecki (TU Wien)

[JS 23] *Unitarity of some barycentric rational approximants*, Jawecki & S, arXiv:2205.10606 [math.NA]

Rational approximations to $\exp(\mathrm{i}x)$ given by

Overview

Based on joint work with

Tobias Jawecki (TU Wien)

[JS 23] *Unitarity of some barycentric rational approximants*, Jawecki & S, arXiv:2205.10606 [math.NA]

Rational approximations to $\exp(ix)$ given by

- AAA [NST 18]
  The AAA algorithm for rational approximation, Nakatsukasa, Sète & Trefethen, SIAM J. Sci. Comput., Vol. 40, Iss. 3 (2018).

Overview

Based on joint work with

Tobias Jawecki (TU Wien)

[JS 23] *Unitarity of some barycentric rational approximants*, Jawecki & S, arXiv:2205.10606 [math.NA]

Rational approximations to $\exp(ix)$ given by

- AAA [NST 18]

  The AAA algorithm for rational approximation, Nakatsukasa, Sète & Trefethen, SIAM J. Sci. Comput., Vol. 40, Iss. 3 (2018).

- AAA–Lawson [NT 20]

  An algorithm for real and complex rational minimax approximation, Nakatsukasa & Trefethen, SIAM J. Sci. Comput., Vol. 4, Iss. 5 (2020).

Overview

Based on joint work with

Tobias Jawecki (TU Wien)

[JS 23] *Unitarity of some barycentric rational approximants*, Jawecki & S, arXiv:2205.10606 [math.NA]

Rational approximations to $\exp(ix)$ given by

- AAA [NST 18]

  The AAA algorithm for rational approximation, Nakatsukasa, Sète & Trefethen, SIAM J. Sci. Comput., Vol. 40, Iss. 3 (2018).

- AAA–Lawson [NT 20]

  An algorithm for real and complex rational minimax approximation, Nakatsukasa & Trefethen, SIAM J. Sci. Comput., Vol. 4, Iss. 5 (2020).

Lead to geometric numerical integrators.

The matrix exponential

The matrix exponential is

$$e^{tA} = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!}$$

The matrix exponential

The matrix exponential is

$$\mathrm{e}^{tA} = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!} = V \mathrm{diag}(\mathrm{e}^{t\lambda_k}) V^{-1}, \quad \text{for} \quad A = V \mathrm{diag}(\lambda_k) V^{-1}$$

The matrix exponential

The matrix exponential is

$$\mathrm{e}^{tA} = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!} = V \mathrm{diag}(\mathrm{e}^{t\lambda_k}) V^{-1}, \quad \text{for} \quad A = V \mathrm{diag}(\lambda_k) V^{-1}$$

Geometric Numerical Integrators for ODEs/PDEs:

$\partial_t u = A\, u$ $\qquad$ $u_1 = \mathrm{e}^{hA} u_0$ $\qquad$ linear

The matrix exponential is

$$e^{tA} = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!} = V \text{diag}(e^{t\lambda_k}) V^{-1}, \quad \text{for} \quad A = V \text{diag}(\lambda_k) V^{-1}$$

Geometric Numerical Integrators for ODEs/PDEs:

$\partial_t u = A\, u$	$u_1 = e^{hA} u_0$	linear
$\partial_t u = A\, u + N(u)$	$u_1 = e^{hA} u_0 + \frac{e^{hA}-1}{A} N(u_0)$	non-linear

The matrix exponential

The matrix exponential is

$$\mathrm{e}^{tA} = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!} = V \mathrm{diag}(\mathrm{e}^{t\lambda_k}) V^{-1}, \quad \text{for} \quad A = V \mathrm{diag}(\lambda_k) V^{-1}$$

Geometric Numerical Integrators for ODEs/PDEs:

$\partial_t u = A\, u$	$u_1 = \mathrm{e}^{hA} u_0$	linear
$\partial_t u = A\, u + N(u)$	$u_1 = \mathrm{e}^{hA} u_0 + \frac{\mathrm{e}^{hA}-1}{A} N(u_0)$	non-linear
$\partial_t u = A(t)\, u$	$u_1 = \mathrm{e}^{\int_0^h A(s)\mathrm{d}s + \frac{1}{2}\int_0^h \int_0^s [A(r),A(s)]\mathrm{d}r\mathrm{d}s} u_0$	non-autonomous

The matrix exponential

The matrix exponential is

$$e^{tA} = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!} = V\text{diag}(e^{t\lambda_k})V^{-1}, \quad \text{for} \quad A = V\text{diag}(\lambda_k)V^{-1}$$

Geometric Numerical Integrators for ODEs/PDEs:

$\partial_t u = A\,u$	$u_1 = e^{hA}u_0$	linear
$\partial_t u = A\,u + N(u)$	$u_1 = e^{hA}u_0 + \frac{e^{hA}-1}{A}N(u_0)$	non-linear
$\partial_t u = A(t)\,u$	$u_1 = e^{\int_0^h A(s)\mathrm{d}s + \frac{1}{2}\int_0^h \int_0^s [A(r),A(s)]\mathrm{d}r\mathrm{d}s}u_0$	non-autonomous

Scalar approximations

$e^z \approx 1 + z$	$u_1 = (I + hA)u_0$	Forward Euler (F.E.)

The matrix exponential

The matrix exponential is

$$e^{tA} = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!} = V \text{diag}(e^{t\lambda_k}) V^{-1}, \quad \text{for} \quad A = V \text{diag}(\lambda_k) V^{-1}$$

Geometric Numerical Integrators for ODEs/PDEs:

$\partial_t u = A\, u$	$u_1 = e^{hA} u_0$	linear
$\partial_t u = A\, u + N(u)$	$u_1 = e^{hA} u_0 + \frac{e^{hA}-1}{A} N(u_0)$	non-linear
$\partial_t u = A(t)\, u$	$u_1 = e^{\int_0^h A(s)\mathrm{d}s + \frac{1}{2}\int_0^h \int_0^s [A(r),A(s)]\mathrm{d}r\mathrm{d}s} u_0$	non-autonomous

Scalar approximations

$e^z \approx 1 + z$	u_1	$=$	$(I + hA)u_0$	Forward Euler (F.E.)
$e^z \approx \frac{1}{1-z}$	$(I - hA)u_1$	$=$	u_0	Backward Euler (B.E.)

The matrix exponential

The matrix exponential is

$$e^{tA} = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!} = V\text{diag}(e^{t\lambda_k})V^{-1}, \quad \text{for} \quad A = V\text{diag}(\lambda_k)V^{-1}$$

Geometric Numerical Integrators for ODEs/PDEs:

$\partial_t u = A\, u$	$u_1 = e^{hA} u_0$	linear
$\partial_t u = A\, u + N(u)$	$u_1 = e^{hA} u_0 + \frac{e^{hA}-1}{A} N(u_0)$	non-linear
$\partial_t u = A(t)\, u$	$u_1 = e^{\int_0^h A(s)\mathrm{d}s + \frac{1}{2}\int_0^h \int_0^s [A(r),A(s)]\mathrm{d}r\mathrm{d}s} u_0$	non-autonomous

Scalar approximations

$e^z \approx 1 + z$	u_1	$=$	$(I + hA)u_0$	Forward Euler (F.E.)
$e^z \approx \frac{1}{1-z}$	$(I - hA)u_1$	$=$	u_0	Backward Euler (B.E.)
$e^z \approx \frac{1+z/2}{1-z/2}$	$(I - (h/2)A)\, u_1$	$=$	$(I + (h/2)A)\, u_0$	Trapezoidal Rule (T.R.)

Why compute the matrix exponential?

Schrödinger equation

$$\partial_t u = -\mathrm{i}\mathrm{H}u, \qquad u(0) = u_0, \qquad \mathrm{H}^* = \mathrm{H},$$

Why compute the matrix exponential?

Schrödinger equation

$$\partial_t u = -\mathrm{i}\mathrm{H}u, \qquad u(0) = u_0, \qquad \mathrm{H}^* = \mathrm{H},$$

$$u(t) \quad = \quad \mathrm{e}^{-\mathrm{i}t\mathrm{H}} u_0$$

Why compute the matrix exponential?

Schrödinger equation

$$\partial_t u = -\mathrm{i}\mathrm{H}u, \qquad u(0) = u_0, \qquad \mathrm{H}^* = \mathrm{H},$$

$$u(t) = \mathrm{e}^{-\mathrm{i}t\mathrm{H}}u_0$$

$$E(t) := \langle u(t), \mathrm{H}u(t)\rangle = \langle u(0), \mathrm{H}u(0)\rangle = E(0) \qquad \text{energy conservation}$$

Why compute the matrix exponential?

Schrödinger equation

$$\partial_t u = -\mathrm{i}\mathrm{H}u, \qquad u(0) = u_0, \qquad \mathrm{H}^* = \mathrm{H},$$

$$\begin{aligned} u(t) &= \mathrm{e}^{-\mathrm{i}t\mathrm{H}} u_0 \\ E(t) := \langle u(t), \mathrm{H}u(t)\rangle &= \langle u(0), \mathrm{H}u(0)\rangle = E(0) \qquad \text{energy conservation} \\ \underbrace{\langle u(t), v(t)\rangle = \langle u(0), v(0)\rangle}_{\text{unitary evolution}} & \end{aligned}$$

Why compute the matrix exponential?

Schrödinger equation

$$\partial_t u = -\mathrm{i}\mathrm{H}u, \qquad u(0) = u_0, \qquad \mathrm{H}^* = \mathrm{H},$$

$$\begin{aligned} u(t) &= \mathrm{e}^{-\mathrm{i}t\mathrm{H}} u_0 \\ E(t) := \langle u(t), \mathrm{H}u(t) \rangle &= \langle u(0), \mathrm{H}u(0) \rangle = E(0) \qquad \text{energy conservation} \\ \underbrace{\langle u(t), v(t) \rangle = \langle u(0), v(0) \rangle}_{\text{unitary evolution}} &\implies \underbrace{\|u(t)\|_2 = \|u(0)\|_2 = 1}_{\text{mass or probability conservation}} \end{aligned}$$

Why compute the matrix exponential?

Schrödinger equation

$$\partial_t u = -\mathrm{i}\mathrm{H}u, \qquad u(0) = u_0, \qquad \mathrm{H}^* = \mathrm{H},$$

$$\begin{aligned} u(t) \quad &= \quad \mathrm{e}^{-\mathrm{i}t\mathrm{H}} u_0 \\ E(t) := \langle u(t), \mathrm{H}u(t)\rangle \quad &= \quad \langle u(0), \mathrm{H}u(0)\rangle = E(0) \qquad \text{energy conservation} \\ \underbrace{\langle u(t), v(t)\rangle = \langle u(0), v(0)\rangle}_{\text{unitary evolution}} \quad &\Longrightarrow \quad \underbrace{\|u(t)\|_2 = \|u(0)\|_2 = 1}_{\text{mass or probability conservation}} \end{aligned}$$

exp maps Lie algebra $\mathrm{i}\mathrm{H} \in \mathfrak{su}(n)$ to Lie group $\mathrm{e}^{-\mathrm{i}t\mathrm{H}} \in \mathrm{U}(n)$.

Why compute the matrix exponential?

Schrödinger equation

$$\partial_t u = -\mathrm{i}\mathrm{H}u, \qquad u(0) = u_0, \qquad \mathrm{H}^* = \mathrm{H},$$

$$\begin{aligned} u(t) &= \mathrm{e}^{-\mathrm{i}t\mathrm{H}} u_0 \\ E(t) := \langle u(t), \mathrm{H}u(t)\rangle &= \langle u(0), \mathrm{H}u(0)\rangle = E(0) \qquad \text{energy conservation} \\ \underbrace{\langle u(t), v(t)\rangle = \langle u(0), v(0)\rangle}_{\text{unitary evolution}} &\implies \underbrace{\|u(t)\|_2 = \|u(0)\|_2 = 1}_{\text{mass or probability conservation}} \end{aligned}$$

exp maps Lie algebra $\mathrm{i}\mathrm{H} \in \mathfrak{su}(n)$ to Lie group $\mathrm{e}^{-\mathrm{i}t\mathrm{H}} \in \mathrm{U}(n)$.

These properties are also desired from numerical approximations.

Why compute the matrix exponential?

Schrödinger equation

$$\partial_t u = -\mathrm{i}\mathrm{H}u, \qquad u(0) = u_0, \qquad \mathrm{H}^* = \mathrm{H},$$

$$\begin{aligned} u(t) &= \mathrm{e}^{-\mathrm{i}t\mathrm{H}} u_0 \\ E(t) := \langle u(t), \mathrm{H}u(t)\rangle &= \langle u(0), \mathrm{H}u(0)\rangle = E(0) \qquad \text{energy conservation} \\ \underbrace{\langle u(t), v(t)\rangle = \langle u(0), v(0)\rangle}_{\text{unitary evolution}} &\implies \underbrace{\|u(t)\|_2 = \|u(0)\|_2 = 1}_{\text{mass or probability conservation}} \end{aligned}$$

exp maps Lie algebra $\mathrm{i}\mathrm{H} \in \mathfrak{su}(n)$ to Lie group $\mathrm{e}^{-\mathrm{i}t\mathrm{H}} \in \mathrm{U}(n)$.

These properties are also desired from numerical approximations.

$\mathrm{e}^z \approx 1 + z$ $\qquad u_1 = (I - \mathrm{i}h\mathrm{H})u_0$ $\qquad$ F.E. $\qquad \|u_n\|_2 \to \infty$

Why compute the matrix exponential?

Schrödinger equation

$$\partial_t u = -\mathrm{i}\mathrm{H}u, \qquad u(0) = u_0, \qquad \mathrm{H}^* = \mathrm{H},$$

$$\begin{aligned} u(t) &= \mathrm{e}^{-\mathrm{i}t\mathrm{H}} u_0 \\ E(t) := \langle u(t), \mathrm{H}u(t)\rangle &= \langle u(0), \mathrm{H}u(0)\rangle = E(0) \qquad \text{energy conservation} \\ \underbrace{\langle u(t), v(t)\rangle = \langle u(0), v(0)\rangle}_{\text{unitary evolution}} &\implies \underbrace{\|u(t)\|_2 = \|u(0)\|_2 = 1}_{\text{mass or probability conservation}} \end{aligned}$$

exp maps Lie algebra $\mathrm{i}\mathrm{H} \in \mathfrak{su}(n)$ to Lie group $\mathrm{e}^{-\mathrm{i}t\mathrm{H}} \in \mathrm{U}(n)$.

These properties are also desired from numerical approximations.

$\mathrm{e}^z \approx 1 + z$	$u_1 = (I - \mathrm{i}h\mathrm{H})u_0$	F.E.	$\|u_n\|_2 \to \infty$
$\mathrm{e}^z \approx \frac{1}{1-z}$	$(I + \mathrm{i}h\mathrm{H})u_1 = u_0$	B.E.	$\|u_n\|_2 \to 0$

Why compute the matrix exponential?

Schrödinger equation

$$\partial_t u = -\mathrm{i}\mathrm{H}u, \qquad u(0) = u_0, \qquad \mathrm{H}^* = \mathrm{H},$$

$$\begin{aligned} u(t) &= \mathrm{e}^{-\mathrm{i}t\mathrm{H}} u_0 \\ E(t) := \langle u(t), \mathrm{H}u(t)\rangle &= \langle u(0), \mathrm{H}u(0)\rangle = E(0) \qquad \text{energy conservation} \\ \underbrace{\langle u(t), v(t)\rangle = \langle u(0), v(0)\rangle}_{\text{unitary evolution}} &\implies \underbrace{\|u(t)\|_2 = \|u(0)\|_2 = 1}_{\text{mass or probability conservation}} \end{aligned}$$

exp maps Lie algebra $\mathrm{i}\mathrm{H} \in \mathfrak{su}(n)$ to Lie group $\mathrm{e}^{-\mathrm{i}t\mathrm{H}} \in \mathrm{U}(n)$.

These properties are also desired from numerical approximations.

$\mathrm{e}^z \approx 1 + z$	$u_1 = (I - \mathrm{i}h\mathrm{H})u_0$	F.E.	$\|u_n\|_2 \to \infty$
$\mathrm{e}^z \approx \frac{1}{1-z}$	$(I + \mathrm{i}h\mathrm{H})u_1 = u_0$	B.E.	$\|u_n\|_2 \to 0$
$\mathrm{e}^z \approx \frac{1+z/2}{1-z/2}$	$(I + \mathrm{i}(h/2)\mathrm{H})\, u_1 = (I - \mathrm{i}(h/2)\mathrm{H})\, u_0$	T.R.	$\|u_n\|_2 = \|u_0\|_2$

Why compute the matrix exponential?

Schrödinger equation

$$\partial_t u = -\mathrm{i}\mathrm{H}u, \qquad u(0) = u_0, \qquad \mathrm{H}^* = \mathrm{H},$$

$$\begin{aligned} u(t) &= \mathrm{e}^{-\mathrm{i}t\mathrm{H}} u_0 \\ E(t) := \langle u(t), \mathrm{H}u(t)\rangle &= \langle u(0), \mathrm{H}u(0)\rangle = E(0) \quad \text{energy conservation} \\ \underbrace{\langle u(t), v(t)\rangle = \langle u(0), v(0)\rangle}_{\text{unitary evolution}} &\implies \underbrace{\|u(t)\|_2 = \|u(0)\|_2 = 1}_{\text{mass or probability conservation}} \end{aligned}$$

exp maps Lie algebra $\mathrm{i}\mathrm{H} \in \mathfrak{su}(n)$ to Lie group $\mathrm{e}^{-\mathrm{i}t\mathrm{H}} \in \mathrm{U}(n)$.

These properties are also desired from numerical approximations.

$\mathrm{e}^z \approx 1 + z$	u_1	$=$	$(I - \mathrm{i}h\mathrm{H})u_0$	F.E.	$\|u_n\|_2 \to \infty$
$\mathrm{e}^z \approx \frac{1}{1-z}$	$(I + \mathrm{i}h\mathrm{H})u_1$	$=$	u_0	B.E.	$\|u_n\|_2 \to 0$
$\mathrm{e}^z \approx \frac{1+z/2}{1-z/2}$	$(I + \mathrm{i}(h/2)\mathrm{H})\, u_1$	$=$	$(I - \mathrm{i}(h/2)\mathrm{H})\, u_0$	T.R.	$\|u_n\|_2 = \|u_0\|_2$

$\mathrm{cay}(z) = \frac{1+z/2}{1-z/2}$ maps Lie algebra $\mathrm{i}\mathrm{H} \in \mathfrak{su}(n)$ to Lie group $\mathrm{e}^{-\mathrm{i}t\mathrm{H}} \in \mathrm{U}(n)$.

Why compute the matrix exponential?

Schrödinger equation

$$\partial_t u = -\mathrm{i}\mathrm{H}u, \qquad u(0) = u_0, \qquad \mathrm{H}^* = \mathrm{H},$$

$$\begin{aligned} u(t) &= \mathrm{e}^{-\mathrm{i}t\mathrm{H}} u_0 \\ E(t) := \langle u(t), \mathrm{H}u(t)\rangle &= \langle u(0), \mathrm{H}u(0)\rangle = E(0) \quad \text{energy conservation} \\ \underbrace{\langle u(t), v(t)\rangle = \langle u(0), v(0)\rangle}_{\text{unitary evolution}} &\implies \underbrace{\|u(t)\|_2 = \|u(0)\|_2 = 1}_{\text{mass or probability conservation}} \end{aligned}$$

exp maps Lie algebra $\mathrm{i}\mathrm{H} \in \mathfrak{su}(n)$ to Lie group $\mathrm{e}^{-\mathrm{i}t\mathrm{H}} \in \mathrm{U}(n)$.

These properties are also desired from numerical approximations.

$\mathrm{e}^z \approx 1 + z$	u_1	$=$	$(I - \mathrm{i}h\mathrm{H})u_0$	F.E.	$\|u_n\|_2 \to \infty$
$\mathrm{e}^z \approx \frac{1}{1-z}$	$(I + \mathrm{i}h\mathrm{H})u_1$	$=$	u_0	B.E.	$\|u_n\|_2 \to 0$
$\mathrm{e}^z \approx \frac{1+z/2}{1-z/2}$	$(I + \mathrm{i}(h/2)\mathrm{H})\, u_1$	$=$	$(I - \mathrm{i}(h/2)\mathrm{H})\, u_0$	T.R.	$\|u_n\|_2 = \|u_0\|_2$

$\mathrm{cay}(z) = \frac{1+z/2}{1-z/2}$ maps Lie algebra $\mathrm{i}\mathrm{H} \in \mathfrak{su}(n)$ to Lie group $\mathrm{e}^{-\mathrm{i}t\mathrm{H}} \in \mathrm{U}(n)$.

Wave, KdV, NLS, Dirac, Liouville–von Neumann, Linblad, Pauli, MCTDHF, CCSD, TDDFT, ...

How to compute the matrix exponential?

C. Moler & C. V. Loan, *Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later*, SIAM Review 03.

How to compute the matrix exponential?

C. Moler & C. V. Loan, *Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later*, SIAM Review 03.

	Asymptotic $z \to 0$	Approximate e^z on spectrum $z \in [a, b] \subseteq \sigma(A)$	Iterative Use A and u_0
Polynomial	Taylor $\sum_{k=0}^{n} \frac{z^k}{k!}$	Chebyshev $J_0(\mathrm{i}) + 2\sum_{k=1}^{n} \mathrm{i}^k J_k(-\mathrm{i})\, T_k(z)$	Lanczos
Rational	Padé $\frac{1+\frac{1}{2}z+\frac{1}{12}z^2}{1-\frac{1}{2}z+\frac{1}{12}z^2}$	?	Rational Krylov

Other techniques: Diagonalisation, Scaling and Squaring, Splitting

How to compute the matrix exponential?

C. Moler & C. V. Loan, *Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later*, SIAM Review 03.

	Asymptotic $z \to 0$	Approximate e^z on spectrum $z \in [a, b] \subseteq \sigma(A)$	Iterative Use A and u_0
Polynomial	Taylor $\sum_{k=0}^{n} \frac{z^k}{k!}$	Chebyshev $J_0(\mathrm{i}) + 2\sum_{k=1}^{n} \mathrm{i}^k J_k(-\mathrm{i}) T_k(z)$	Lanczos
Rational	Padé $\frac{1+\frac{1}{2}z+\frac{1}{12}z^2}{1-\frac{1}{2}z+\frac{1}{12}z^2}$	?	Rational Krylov

Other techniques: Diagonalisation, Scaling and Squaring, Splitting

Since $\sigma(\mathrm{iH}) \subseteq \mathrm{i}\mathbb{R}$,

$$|f(\mathrm{i}x)| = 1 \qquad x \in \mathbb{R} \qquad \Longrightarrow \qquad f(\mathrm{iH}) \text{ is unitary}$$

How to compute the matrix exponential?

C. Moler & C. V. Loan, *Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later*, SIAM Review 03.

	Asymptotic $z \to 0$	Approximate e^z on spectrum $z \in [a, b] \subseteq \sigma(A)$	Iterative Use A and u_0
Polynomial	Taylor $\sum_{k=0}^{n} \frac{z^k}{k!}$	Chebyshev $J_0(\mathrm{i}) + 2\sum_{k=1}^{n} \mathrm{i}^k J_k(-\mathrm{i})\, T_k(z)$	Lanczos
Rational	Padé $\frac{1+\frac{1}{2}z+\frac{1}{12}z^2}{1-\frac{1}{2}z+\frac{1}{12}z^2}$	?	Rational Krylov

Other techniques: Diagonalisation, Scaling and Squaring, Splitting

Since $\sigma(\mathrm{iH}) \subseteq \mathrm{i}\mathbb{R}$,

$$|f(\mathrm{i}x)| = 1 \qquad x \in \mathbb{R} \qquad \Longrightarrow \qquad f(\mathrm{iH}) \text{ is unitary}$$

No non-constant polynomial method can be unitary. Proof: coercivity.

How to compute the matrix exponential?

C. Moler & C. V. Loan, *Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later*, SIAM Review 03.

	Asymptotic $z \to 0$	Approximate e^z on spectrum $z \in [a, b] \subseteq \sigma(A)$	Iterative Use A and u_0
Polynomial	Taylor $\sum_{k=0}^{n} \frac{z^k}{k!}$	Chebyshev $J_0(\mathrm{i}) + 2\sum_{k=1}^{n} \mathrm{i}^k J_k(-\mathrm{i})\, T_k(z)$	Lanczos
Rational	Padé $\frac{1+\frac{1}{2}z+\frac{1}{12}z^2}{1-\frac{1}{2}z+\frac{1}{12}z^2}$	AAA & AAA–Lawson	Rational Krylov

Other techniques: Diagonalisation, Scaling and Squaring, Splitting

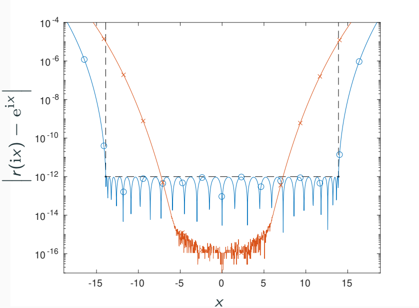
Since $\sigma(\mathrm{iH}) \subseteq \mathrm{i}\mathbb{R}$,

$$|f(\mathrm{i}x)| = 1 \qquad x \in \mathbb{R} \qquad \Longrightarrow \qquad f(\mathrm{iH}) \text{ is unitary}$$

No non-constant polynomial method can be unitary. Proof: coercivity.

Padé vs AAA

$(13, 13)$ Padé and AAA–Lawson approximants.



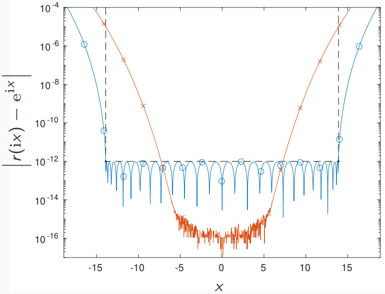
Padé methods are rational methods,

$$r_{m,n}(z) = \frac{p_m(z)}{q_n(z)} \approx e^z, \qquad p_m \in \mathcal{P}_m, q_n \in \mathcal{P}_n,$$

which approximate the Taylor expansion of e^z to the highest degree possible.

Padé vs AAA

(13, 13) Padé and AAA–Lawson approximants.



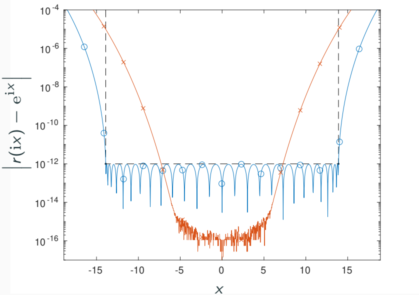
Padé methods are rational methods,

$$r_{m,n}(z) = \frac{p_m(z)}{q_n(z)} \approx \mathrm{e}^z, \qquad p_m \in \mathcal{P}_m, q_n \in \mathcal{P}_n,$$

which approximate the Taylor expansion of e^z to the highest degree possible. These are asymptotic in accuracy, i.e. under $z \to 0$, and need to focus most effort near the origin.

Padé vs AAA

$(13, 13)$ Padé and AAA–Lawson approximants.



Padé methods are rational methods,

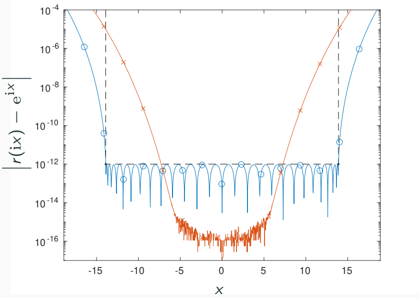
$$r_{m,n}(z) = \frac{p_m(z)}{q_n(z)} \approx \mathrm{e}^z, \qquad p_m \in \mathcal{P}_m, q_n \in \mathcal{P}_n,$$

which approximate the Taylor expansion of e^z to the highest degree possible. These are asymptotic in accuracy, i.e. under $z \to 0$, and need to focus most effort near the origin.

AAA and AAA–Lawson methods are adaptive algorithms that can produce rational approximants with uniform accuracy over a specified interval or test nodes x_k.

Padé vs AAA

$(13, 13)$ Padé and AAA–Lawson approximants.



Padé methods are rational methods,

$$r_{m,n}(z) = \frac{p_m(z)}{q_n(z)} \approx e^z, \qquad p_m \in \mathcal{P}_m, q_n \in \mathcal{P}_n,$$

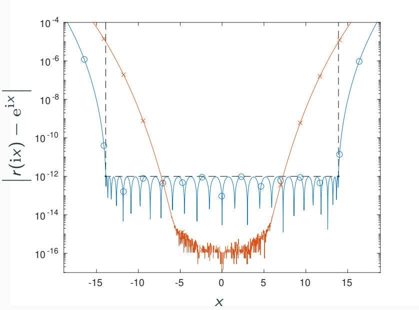
which approximate the Taylor expansion of e^z to the highest degree possible. These are asymptotic in accuracy, i.e. under $z \to 0$, and need to focus most effort near the origin.

AAA and AAA–Lawson methods are adaptive algorithms that can produce rational approximants with uniform accuracy over a specified interval or test nodes x_k.

$$r(x) = \underbrace{\sum_{j=1}^{m} \frac{e^{iy_j} w_j}{x - y_j}}_{n(x)} \Big/ \underbrace{\sum_{j=1}^{m} \frac{w_j}{x - y_j}}_{d(x)},$$

Padé vs AAA

(13, 13) Padé and AAA–Lawson approximants.



Padé methods are rational methods,

$$r_{m,n}(z) = \frac{p_m(z)}{q_n(z)} \approx \mathrm{e}^z, \qquad p_m \in \mathcal{P}_m, q_n \in \mathcal{P}_n,$$

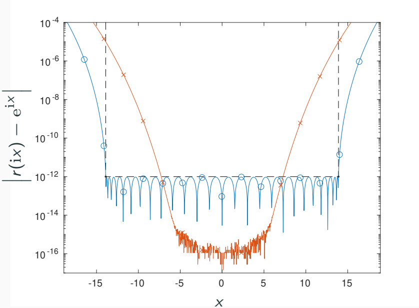
which approximate the Taylor expansion of e^z to the highest degree possible. These are asymptotic in accuracy, i.e. under $z \to 0$, and need to focus most effort near the origin.

AAA and AAA–Lawson methods are adaptive algorithms that can produce rational approximants with uniform accuracy over a specified interval or test nodes x_k.

$$r(x) = \underbrace{\sum_{j=1}^{m} \frac{\mathrm{e}^{\mathrm{i}y_j}\, w_j}{x - y_j}}_{n(x)} \Bigg/ \underbrace{\sum_{j=1}^{m} \frac{w_j}{x - y_j}}_{d(x)}, \qquad \text{find } w \text{ s.t. } r(x_k) = \frac{n(x_k)}{d(x_k)} \approx \mathrm{e}^{\mathrm{i}x_k}$$

Padé vs AAA

(13, 13) Padé and AAA–Lawson approximants.



Padé methods are rational methods,

$$r_{m,n}(z) = \frac{p_m(z)}{q_n(z)} \approx \mathrm{e}^z, \qquad p_m \in \mathcal{P}_m,\, q_n \in \mathcal{P}_n,$$

which approximate the Taylor expansion of e^z to the highest degree possible. These are asymptotic in accuracy, i.e. under $z \to 0$, and need to focus most effort near the origin.

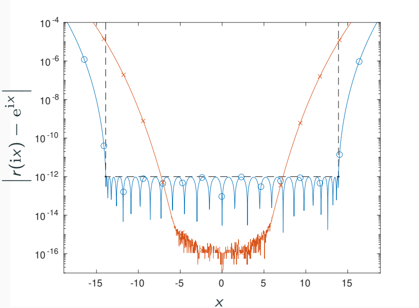
AAA and AAA–Lawson methods are adaptive algorithms that can produce rational approximants with uniform accuracy over a specified interval or test nodes x_k.

$$r(x) = \underbrace{\sum_{j=1}^{m} \frac{\mathrm{e}^{\mathrm{i}y_j}\, w_j}{x - y_j}}_{n(x)} \Bigg/ \underbrace{\sum_{j=1}^{m} \frac{w_j}{x - y_j}}_{d(x)}, \qquad \text{find } w \text{ s.t. } r(x_k) = \frac{n(x_k)}{d(x_k)} \approx \mathrm{e}^{\mathrm{i}x_k}$$

linearize & minimize $\quad \operatorname{argmin}_w \left(\sum_{k=1}^{n} \mu_k \left| n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k) \right|^2 \right)^{1/2}.$

Padé vs AAA

(13, 13) Padé and AAA–Lawson approximants.



Padé methods are rational methods,

$$r_{m,n}(z) = \frac{p_m(z)}{q_n(z)} \approx \mathrm{e}^z, \qquad p_m \in \mathcal{P}_m, q_n \in \mathcal{P}_n,$$

which approximate the Taylor expansion of e^z to the highest degree possible. These are asymptotic in accuracy, i.e. under $z \to 0$, and need to focus most effort near the origin.

AAA and AAA–Lawson methods are adaptive algorithms that can produce rational approximants with uniform accuracy over a specified interval or test nodes x_k.

$$r(x) = \underbrace{\sum_{j=1}^{m} \frac{\mathrm{e}^{\mathrm{i}y_j} w_j}{x - y_j}}_{n(x)} \Big/ \underbrace{\sum_{j=1}^{m} \frac{w_j}{x - y_j}}_{d(x)}, \qquad \text{find } w \text{ s.t. } r(x_k) = \frac{n(x_k)}{d(x_k)} \approx \mathrm{e}^{\mathrm{i}x_k}$$

linearize & minimize $\qquad \operatorname{argmin}_w \left(\sum_{k=1}^{n} \mu_k |n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k)|^2 \right)^{1/2}.$

Linear (weighted) least squares $\Rightarrow$ SVD

Loewner matrix (NST 18)

Loewner matrix (NST 18)

linear (weighted) least squares $\quad \operatorname{argmin}_w \left(\sum_{k=1}^{n} \mu_k |n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k)|^2 \right)^{1/2}.$

Loewner matrix (NST 18)

linear (weighted) least squares $\quad \operatorname{argmin}_w \Big(\sum_{k=1}^{n} \mu_k |n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k)|^2 \Big)^{1/2}.$

Loewner matrix $L_{kj} = \mu_k^{1/2} \frac{\mathrm{e}^{\mathrm{i}x_k} - \mathrm{e}^{\mathrm{i}y_j}}{x_k - y_j}$,

Loewner matrix (NST 18)

linear (weighted) least squares $\quad \operatorname{argmin}_w \left(\sum_{k=1}^{n} \mu_k |n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k)|^2 \right)^{1/2}.$

Loewner matrix $L_{kj} = \mu_k^{1/2} \frac{\mathrm{e}^{\mathrm{i}x_k} - \mathrm{e}^{\mathrm{i}y_j}}{x_k - y_j}$,

$$S_\mu = \mathrm{diag}(\mu_k^{1/2}),$$

Loewner matrix (NST 18)

linear (weighted) least squares $\quad \operatorname{argmin}_w \left(\sum_{k=1}^{n} \mu_k |n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k)|^2 \right)^{1/2}.$

Loewner matrix $L_{kj} = \mu_k^{1/2} \frac{\mathrm{e}^{\mathrm{i}x_k} - \mathrm{e}^{\mathrm{i}y_j}}{x_k - y_j}$,

$$S_\mu = \mathrm{diag}(\mu_k^{1/2}), \quad C_{kj} = (x_k - y_j)^{-1},$$

Loewner matrix (NST 18)

linear (weighted) least squares $\quad \operatorname{argmin}_w \left(\sum_{k=1}^{n} \mu_k |n(x_k) - e^{ix_k} d(x_k)|^2 \right)^{1/2}.$

Loewner matrix $L_{kj} = \mu_k^{1/2} \frac{e^{ix_k} - e^{iy_j}}{x_k - y_j}$,

$$S_\mu = \operatorname{diag}(\mu_k^{1/2}), \quad C_{kj} = (x_k - y_j)^{-1}, \quad M = S_\mu C$$

Loewner matrix (NST 18)

linear (weighted) least squares $\quad \text{argmin}_w \Big(\sum_{k=1}^{n} \mu_k |n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k)|^2 \Big)^{1/2}.$

Loewner matrix $L_{kj} = \mu_k^{1/2} \frac{\mathrm{e}^{\mathrm{i}x_k} - \mathrm{e}^{\mathrm{i}y_j}}{x_k - y_j}$,

$$S_\mu = \mathrm{diag}(\mu_k^{1/2}), \quad C_{kj} = (x_k - y_j)^{-1}, \quad M = S_\mu C$$

$$S_F = \mathrm{diag}(\mathrm{e}^{\mathrm{i}x_k}),$$

Loewner matrix (NST 18)

linear (weighted) least squares $\quad \text{argmin}_w \left(\sum_{k=1}^{n} \mu_k |n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k)|^2 \right)^{1/2}.$

Loewner matrix $L_{kj} = \mu_k^{1/2} \frac{\mathrm{e}^{\mathrm{i}x_k} - \mathrm{e}^{\mathrm{i}y_j}}{x_k - y_j}$,

$$S_\mu = \mathrm{diag}(\mu_k^{1/2}), \quad C_{kj} = (x_k - y_j)^{-1}, \quad M = S_\mu C$$

$$S_F = \mathrm{diag}(\mathrm{e}^{\mathrm{i}x_k}), \quad S_f = \mathrm{diag}(\mathrm{e}^{\mathrm{i}y_j})$$

Loewner matrix (NST 18)

linear (weighted) least squares $\quad \operatorname{argmin}_w \left(\sum_{k=1}^{n} \mu_k |n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k)|^2 \right)^{1/2}.$

Loewner matrix $L_{kj} = \mu_k^{1/2} \frac{\mathrm{e}^{\mathrm{i}x_k} - \mathrm{e}^{\mathrm{i}y_j}}{x_k - y_j}$,

$$S_\mu = \operatorname{diag}(\mu_k^{1/2}), \quad C_{kj} = (x_k - y_j)^{-1}, \quad M = S_\mu C$$

$$S_F = \operatorname{diag}(\mathrm{e}^{\mathrm{i}x_k}), \quad S_f = \operatorname{diag}(\mathrm{e}^{\mathrm{i}y_j})$$

$$L = S_F M - M S_f.$$

Loewner matrix (NST 18)

linear (weighted) least squares $\quad \operatorname{argmin}_w \left(\sum_{k=1}^{n} \mu_k |n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k)|^2 \right)^{1/2}.$

Loewner matrix $L_{kj} = \mu_k^{1/2} \frac{\mathrm{e}^{\mathrm{i}x_k} - \mathrm{e}^{\mathrm{i}y_j}}{x_k - y_j}$,

$$\begin{aligned} S_\mu &= \operatorname{diag}(\mu_k^{1/2}), & C_{kj} &= (x_k - y_j)^{-1}, & M &= S_\mu C \\ S_F &= \operatorname{diag}(\mathrm{e}^{\mathrm{i}x_k}), & S_f &= \operatorname{diag}(\mathrm{e}^{\mathrm{i}y_j}) \end{aligned}$$

$$L = S_F M - M S_f.$$

$$(S_F C w)_k = \mathrm{e}^{\mathrm{i}x_k} \sum_{j=1}^{m} \frac{w_j}{x_k - y_j} = \mathrm{e}^{\mathrm{i}x_k} d(x_k),$$

Loewner matrix (NST 18)

linear (weighted) least squares $\quad \operatorname{argmin}_w \left(\sum_{k=1}^{n} \mu_k |n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k)|^2 \right)^{1/2}.$

Loewner matrix $L_{kj} = \mu_k^{1/2} \frac{\mathrm{e}^{\mathrm{i}x_k} - \mathrm{e}^{\mathrm{i}y_j}}{x_k - y_j}$,

$$\begin{aligned} S_\mu &= \operatorname{diag}(\mu_k^{1/2}), & C_{kj} &= (x_k - y_j)^{-1}, & M &= S_\mu C \\ S_F &= \operatorname{diag}(\mathrm{e}^{\mathrm{i}x_k}), & S_f &= \operatorname{diag}(\mathrm{e}^{\mathrm{i}y_j}) \end{aligned}$$

$$L = S_F M - M S_f.$$

$$(S_F C w)_k = \mathrm{e}^{\mathrm{i}x_k} \sum_{j=1}^{m} \frac{w_j}{x_k - y_j} = \mathrm{e}^{\mathrm{i}x_k} d(x_k), \quad \text{and} \quad (C S_f w)_k = \sum_{j=1}^{m} \frac{\mathrm{e}^{\mathrm{i}y_j} w_j}{x_k - y_j} = n(x_k).$$

Loewner matrix (NST 18)

linear (weighted) least squares $\quad \text{argmin}_w \left(\sum_{k=1}^{n} \mu_k |n(x_k) - e^{ix_k} d(x_k)|^2 \right)^{1/2}.$

Loewner matrix $L_{kj} = \mu_k^{1/2} \frac{e^{ix_k} - e^{iy_j}}{x_k - y_j}$,

$$\begin{aligned} S_\mu &= \operatorname{diag}(\mu_k^{1/2}), & C_{kj} &= (x_k - y_j)^{-1}, & M &= S_\mu C \\ S_F &= \operatorname{diag}(e^{ix_k}), & S_f &= \operatorname{diag}(e^{iy_j}) \end{aligned}$$

$$L = S_F M - M S_f.$$

$$(S_F C w)_k = e^{ix_k} \sum_{j=1}^{m} \frac{w_j}{x_k - y_j} = e^{ix_k} d(x_k), \quad \text{and} \quad (C S_f w)_k = \sum_{j=1}^{m} \frac{e^{iy_j} w_j}{x_k - y_j} = n(x_k).$$

$$\|Lw\|_2 = \left(\sum_{k=1}^{n} \mu_k |n(x_k) - e^{ix_k} d(x_k)|^2 \right)^{1/2}.$$

Loewner matrix (NST 18)

linear (weighted) least squares $\quad \operatorname{argmin}_w \left(\sum_{k=1}^{n} \mu_k |n(x_k) - e^{ix_k} d(x_k)|^2 \right)^{1/2}.$

Loewner matrix $L_{kj} = \mu_k^{1/2} \frac{e^{ix_k} - e^{iy_j}}{x_k - y_j}$,

$$
\begin{array}{lllllllll}
S_\mu & = & \operatorname{diag}(\mu_k^{1/2}), & C_{kj} & = & (x_k - y_j)^{-1}, & M & = & S_\mu C \\
S_F & = & \operatorname{diag}(e^{ix_k}), & S_f & = & \operatorname{diag}(e^{iy_j}) & & &
\end{array}
$$

$$L = S_F M - M S_f.$$

$$(S_F C w)_k = e^{ix_k} \sum_{j=1}^{m} \frac{w_j}{x_k - y_j} = e^{ix_k} d(x_k), \quad \text{and} \quad (C S_f w)_k = \sum_{j=1}^{m} \frac{e^{iy_j} w_j}{x_k - y_j} = n(x_k).$$

$$\|Lw\|_2 = \left(\sum_{k=1}^{n} \mu_k |n(x_k) - e^{ix_k} d(x_k)|^2 \right)^{1/2}.$$

$LV = US,$

Loewner matrix (NST 18)

linear (weighted) least squares $\quad \operatorname{argmin}_w \left(\sum_{k=1}^{n} \mu_k |n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k)|^2 \right)^{1/2}.$

Loewner matrix $L_{kj} = \mu_k^{1/2} \frac{\mathrm{e}^{\mathrm{i}x_k} - \mathrm{e}^{\mathrm{i}y_j}}{x_k - y_j}$,

$$\begin{aligned} S_\mu &= \operatorname{diag}(\mu_k^{1/2}), & C_{kj} &= (x_k - y_j)^{-1}, & M &= S_\mu C \\ S_F &= \operatorname{diag}(\mathrm{e}^{\mathrm{i}x_k}), & S_f &= \operatorname{diag}(\mathrm{e}^{\mathrm{i}y_j}) \end{aligned}$$

$$L = S_F M - M S_f.$$

$$(S_F C w)_k = \mathrm{e}^{\mathrm{i}x_k} \sum_{j=1}^{m} \frac{w_j}{x_k - y_j} = \mathrm{e}^{\mathrm{i}x_k} d(x_k), \quad \text{and} \quad (C S_f w)_k = \sum_{j=1}^{m} \frac{\mathrm{e}^{\mathrm{i}y_j} w_j}{x_k - y_j} = n(x_k).$$

$$\|Lw\|_2 = \left(\sum_{k=1}^{n} \mu_k |n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k)|^2 \right)^{1/2}.$$

$$LV = US, \qquad w = Ve_m, \qquad \|Lw\|_2 = \sigma_m$$

Loewner matrix (NST 18)

linear (weighted) least squares $\qquad \operatorname{argmin}_w \left(\sum_{k=1}^{n} \mu_k |n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k)|^2 \right)^{1/2}.$

Loewner matrix $L_{kj} = \mu_k^{1/2} \frac{\mathrm{e}^{\mathrm{i}x_k} - \mathrm{e}^{\mathrm{i}y_j}}{x_k - y_j}$,

$$\begin{array}{llllll} S_\mu & = \operatorname{diag}(\mu_k^{1/2}), & C_{kj} & = (x_k - y_j)^{-1}, & M & = S_\mu C \\ S_F & = \operatorname{diag}(\mathrm{e}^{\mathrm{i}x_k}), & S_f & = \operatorname{diag}(\mathrm{e}^{\mathrm{i}y_j}) & & \end{array}$$

$$L = S_F M - M S_f.$$

$$(S_F C w)_k = \mathrm{e}^{\mathrm{i}x_k} \sum_{j=1}^{m} \frac{w_j}{x_k - y_j} = \mathrm{e}^{\mathrm{i}x_k} d(x_k), \quad \text{and} \quad (C S_f w)_k = \sum_{j=1}^{m} \frac{\mathrm{e}^{\mathrm{i}y_j} w_j}{x_k - y_j} = n(x_k).$$

$$\|Lw\|_2 = \left(\sum_{k=1}^{n} \mu_k |n(x_k) - \mathrm{e}^{\mathrm{i}x_k} d(x_k)|^2 \right)^{1/2}.$$

$LV = US, \qquad w = Ve_m, \qquad \|Lw\|_2 = \sigma_m \quad \leq \|Lu\|_2$ for any $\|u\|_2 = 1$.

Modified AAA via a rotated Loewner matrix (JS 23)

$$K = \mathrm{diag}\Big((1 - \mathrm{e}^{-\mathrm{i}y_j})/|1 - \mathrm{e}^{-\mathrm{i}y_j}|\Big) \quad \text{and} \quad R = \mathrm{diag}\Big((1 - \mathrm{e}^{-\mathrm{i}x_k})/|1 - \mathrm{e}^{-\mathrm{i}x_k}|\Big)$$

Modified AAA via a rotated Loewner matrix (JS 23)

$$K = \operatorname{diag}\Big((1-\mathrm{e}^{-\mathrm{i}y_j})/|1-\mathrm{e}^{-\mathrm{i}y_j}|\Big) \quad \text{and} \quad R = \operatorname{diag}\Big((1-\mathrm{e}^{-\mathrm{i}x_k})/|1-\mathrm{e}^{-\mathrm{i}x_k}|\Big)$$

$$\mathrm{e}^{\mathrm{i}y_j} K_{jj} = \mathrm{e}^{\mathrm{i}y_j} \frac{(1-\mathrm{e}^{-\mathrm{i}y_j})}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = \frac{\mathrm{e}^{\mathrm{i}y_j}-1}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = -K_{jj}^{\,*}, \qquad \mathrm{e}^{\mathrm{i}x_k} R_{kk} = -R_{kk}^{\,*}$$

Modified AAA via a rotated Loewner matrix (JS 23)

$$K = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}y_j})/|1-\mathrm{e}^{-\mathrm{i}y_j}|\Big) \quad \text{and} \quad R = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}x_k})/|1-\mathrm{e}^{-\mathrm{i}x_k}|\Big)$$

$$\mathrm{e}^{\mathrm{i}y_j} K_{jj} = \mathrm{e}^{\mathrm{i}y_j} \frac{(1-\mathrm{e}^{-\mathrm{i}y_j})}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = \frac{\mathrm{e}^{\mathrm{i}y_j}-1}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = -{K_{jj}}^*, \qquad \mathrm{e}^{\mathrm{i}x_k} R_{kk} = -{R_{kk}}^*$$

$$S_f K = -K^*, \quad \text{and} \quad R S_F = -R^*.$$

We introduce a rotated version of the Loewner matrix, $\widehat{L} := -\mathrm{i}RLK$

Modified AAA via a rotated Loewner matrix (JS 23)

$$K = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}y_j})/|1-\mathrm{e}^{-\mathrm{i}y_j}|\Big) \quad \text{and} \quad R = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}x_k})/|1-\mathrm{e}^{-\mathrm{i}x_k}|\Big)$$

$$\mathrm{e}^{\mathrm{i}y_j} K_{jj} = \mathrm{e}^{\mathrm{i}y_j} \frac{(1-\mathrm{e}^{-\mathrm{i}y_j})}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = \frac{\mathrm{e}^{\mathrm{i}y_j}-1}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = -{K_{jj}}^*, \qquad \mathrm{e}^{\mathrm{i}x_k} R_{kk} = -{R_{kk}}^*$$

$$S_f K = -K^*, \quad \text{and} \quad R S_F = -R^*.$$

We introduce a rotated version of the Loewner matrix, $\widehat{L} := -\mathrm{i}RLK$

$$\widehat{L} = -\mathrm{i}RS_F MK + \mathrm{i}RMS_f K$$

Modified AAA via a rotated Loewner matrix (JS 23)

$$K = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}y_j})/|1-\mathrm{e}^{-\mathrm{i}y_j}|\Big) \quad \text{and} \quad R = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}x_k})/|1-\mathrm{e}^{-\mathrm{i}x_k}|\Big)$$

$$\mathrm{e}^{\mathrm{i}y_j} K_{jj} = \mathrm{e}^{\mathrm{i}y_j} \frac{(1-\mathrm{e}^{-\mathrm{i}y_j})}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = \frac{\mathrm{e}^{\mathrm{i}y_j}-1}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = -K_{jj}{}^*, \qquad \mathrm{e}^{\mathrm{i}x_k} R_{kk} = -R_{kk}{}^*$$

$$S_f K = -K^*, \quad \text{and} \quad R S_F = -R^*.$$

We introduce a rotated version of the Loewner matrix, $\widehat{L} := -\mathrm{i}RLK$

$$\widehat{L} = -\mathrm{i}RS_F MK + \mathrm{i}RMS_f K = \mathrm{i}(R^* MK - RMK^*)$$

Modified AAA via a rotated Loewner matrix (JS 23)

$$K = \operatorname{diag}\Big((1-\mathrm{e}^{-\mathrm{i}y_j})/|1-\mathrm{e}^{-\mathrm{i}y_j}|\Big) \quad \text{and} \quad R = \operatorname{diag}\Big((1-\mathrm{e}^{-\mathrm{i}x_k})/|1-\mathrm{e}^{-\mathrm{i}x_k}|\Big)$$

$$\mathrm{e}^{\mathrm{i}y_j} K_{jj} = \mathrm{e}^{\mathrm{i}y_j}\frac{(1-\mathrm{e}^{-\mathrm{i}y_j})}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = \frac{\mathrm{e}^{\mathrm{i}y_j}-1}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = -K_{jj}{}^*, \qquad \mathrm{e}^{\mathrm{i}x_k} R_{kk} = -R_{kk}{}^*$$

$$S_f K = -K^*, \quad \text{and} \quad RS_F = -R^*.$$

We introduce a rotated version of the Loewner matrix, $\widehat{L} := -\mathrm{i}RLK$

$$\widehat{L} = -\mathrm{i}RS_F MK + \mathrm{i}RMS_f K = \mathrm{i}(R^* MK - RMK^*) = 2\operatorname{Im}(RMK^*) \in \mathbb{R}^{n\times m},$$

Modified AAA via a rotated Loewner matrix (JS 23)

$$K = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}y_j})/|1-\mathrm{e}^{-\mathrm{i}y_j}|\Big) \quad \text{and} \quad R = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}x_k})/|1-\mathrm{e}^{-\mathrm{i}x_k}|\Big)$$

$$\mathrm{e}^{\mathrm{i}y_j} K_{jj} = \mathrm{e}^{\mathrm{i}y_j} \frac{(1-\mathrm{e}^{-\mathrm{i}y_j})}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = \frac{\mathrm{e}^{\mathrm{i}y_j}-1}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = -{K_{jj}}^*, \qquad \mathrm{e}^{\mathrm{i}x_k} R_{kk} = -{R_{kk}}^*$$

$$S_f K = -K^*, \quad \text{and} \quad R S_F = -R^*.$$

We introduce a rotated version of the Loewner matrix, $\widehat{L} := -\mathrm{i}RLK$

$$\widehat{L} = -\mathrm{i}RS_F MK + \mathrm{i}RMS_f K = \mathrm{i}(R^* MK - RMK^*) = 2\mathrm{Im}(RMK^*) \in \mathbb{R}^{n\times m},$$

$$\widehat{L}\widehat{V} = \widehat{U}S \quad \text{(real SVD)}$$

Modified AAA via a rotated Loewner matrix (JS 23)

$$K = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}y_j})/|1-\mathrm{e}^{-\mathrm{i}y_j}|\Big) \quad \text{and} \quad R = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}x_k})/|1-\mathrm{e}^{-\mathrm{i}x_k}|\Big)$$

$$\mathrm{e}^{\mathrm{i}y_j} K_{jj} = \mathrm{e}^{\mathrm{i}y_j}\frac{(1-\mathrm{e}^{-\mathrm{i}y_j})}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = \frac{\mathrm{e}^{\mathrm{i}y_j}-1}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = -{K_{jj}}^{*}, \qquad \mathrm{e}^{\mathrm{i}x_k} R_{kk} = -{R_{kk}}^{*}$$

$$S_f K = -K^*, \quad \text{and} \quad RS_F = -R^*.$$

We introduce a rotated version of the Loewner matrix, $\widehat{L} := -\mathrm{i}RLK$

$$\widehat{L} = -\mathrm{i}RS_F MK + \mathrm{i}RMS_f K = \mathrm{i}(R^*MK - RMK^*) = 2\mathrm{Im}(RMK^*) \in \mathbb{R}^{n\times m},$$

$$\widehat{L}\widehat{V} = \widehat{U}S \quad \text{(real SVD)} \Rightarrow LV = US, \;\; V = \mathrm{i}K\widehat{V} \text{ and } U = -R^*\widehat{U}.$$

Modified AAA via a rotated Loewner matrix (JS 23)

$$K = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}y_j})/|1-\mathrm{e}^{-\mathrm{i}y_j}|\Big) \quad \text{and} \quad R = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}x_k})/|1-\mathrm{e}^{-\mathrm{i}x_k}|\Big)$$

$$\mathrm{e}^{\mathrm{i}y_j} K_{jj} = \mathrm{e}^{\mathrm{i}y_j}\frac{(1-\mathrm{e}^{-\mathrm{i}y_j})}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = \frac{\mathrm{e}^{\mathrm{i}y_j}-1}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = -K_{jj}{}^*, \qquad \mathrm{e}^{\mathrm{i}x_k} R_{kk} = -R_{kk}{}^*$$

$$S_f K = -K^*, \quad \text{and} \quad RS_F = -R^*.$$

We introduce a rotated version of the Loewner matrix, $\widehat{L} := -\mathrm{i}RLK$

$$\widehat{L} = -\mathrm{i}RS_F MK + \mathrm{i}RMS_f K = \mathrm{i}(R^* MK - RMK^*) = 2\mathrm{Im}(RMK^*) \in \mathbb{R}^{n\times m},$$

$$\widehat{L}\widehat{V} = \widehat{U}S \quad \text{(real SVD)} \Rightarrow LV = US, \; V = \mathrm{i}K\widehat{V} \text{ and } U = -R^*\widehat{U}.$$

$\widetilde{w} = Ve_m$ minimizes $\|Lu\|_2$.

Modified AAA via a rotated Loewner matrix (JS 23)

$$K = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}y_j})/|1-\mathrm{e}^{-\mathrm{i}y_j}|\Big) \quad \text{and} \quad R = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}x_k})/|1-\mathrm{e}^{-\mathrm{i}x_k}|\Big)$$

$$\mathrm{e}^{\mathrm{i}y_j} K_{jj} = \mathrm{e}^{\mathrm{i}y_j}\frac{(1-\mathrm{e}^{-\mathrm{i}y_j})}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = \frac{\mathrm{e}^{\mathrm{i}y_j}-1}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = -{K_{jj}}^*, \qquad \mathrm{e}^{\mathrm{i}x_k} R_{kk} = -{R_{kk}}^*$$

$$S_f K = -K^*, \quad \text{and} \quad R S_F = -R^*.$$

We introduce a rotated version of the Loewner matrix, $\widehat{L} := -\mathrm{i}RLK$

$$\widehat{L} = -\mathrm{i}RS_F MK + \mathrm{i}RMS_f K = \mathrm{i}(R^*MK - RMK^*) = 2\mathrm{Im}(RMK^*) \in \mathbb{R}^{n\times m},$$

$$\widehat{L}\widehat{V} = \widehat{U}S \quad \text{(real SVD)} \Rightarrow LV = US,\ \ V = \mathrm{i}K\widehat{V} \text{ and } U = -R^*\widehat{U}.$$

$\widetilde{w} = Ve_m$ minimizes $\|Lu\|_2$. Since $\widetilde{w} = \mathrm{i}K\widehat{V}e_m$,

$$\mathrm{e}^{\mathrm{i}y_j}\widetilde{w}_j = (S_f\widetilde{w})_j = (\mathrm{i}S_f K\widehat{V}e_m)_j = -(\mathrm{i}K^*\widehat{V}e_m)_j = {\widetilde{w}_j}^*.$$

Modified AAA via a rotated Loewner matrix (JS 23)

$$K = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}y_j})/|1-\mathrm{e}^{-\mathrm{i}y_j}|\Big) \quad \text{and} \quad R = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}x_k})/|1-\mathrm{e}^{-\mathrm{i}x_k}|\Big)$$

$$\mathrm{e}^{\mathrm{i}y_j} K_{jj} = \mathrm{e}^{\mathrm{i}y_j}\frac{(1-\mathrm{e}^{-\mathrm{i}y_j})}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = \frac{\mathrm{e}^{\mathrm{i}y_j}-1}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = -K_{jj}{}^*, \qquad \mathrm{e}^{\mathrm{i}x_k} R_{kk} = -R_{kk}{}^*$$

$$S_f K = -K^*, \quad \text{and} \quad RS_F = -R^*.$$

We introduce a rotated version of the Loewner matrix, $\widehat{L} := -\mathrm{i}RLK$

$$\widehat{L} = -\mathrm{i}RS_F MK + \mathrm{i}RMS_f K = \mathrm{i}(R^*MK - RMK^*) = 2\mathrm{Im}(RMK^*) \in \mathbb{R}^{n\times m},$$

$$\widehat{L}\widehat{V} = \widehat{U}S \quad \text{(real SVD)} \Rightarrow LV = US,\ V = \mathrm{i}K\widehat{V} \text{ and } U = -R^*\widehat{U}.$$

$\widetilde{w} = Ve_m$ minimizes $\|Lu\|_2$. Since $\widetilde{w} = \mathrm{i}K\widehat{V}e_m$,

$$\mathrm{e}^{\mathrm{i}y_j}\widetilde{w}_j = (S_f\widetilde{w})_j = (\mathrm{i}S_f K\widehat{V}e_m)_j = -(\mathrm{i}K^*\widehat{V}e_m)_j = \widetilde{w}_j{}^*.$$

$$\widetilde{n}(x) = \sum_{j=1}^m \frac{\mathrm{e}^{\mathrm{i}y_j}\widetilde{w}_j}{x-y_j} = \sum_{j=1}^m \frac{\widetilde{w}_j^*}{x-y_j} = \widetilde{d}(x)^*$$

Modified AAA via a rotated Loewner matrix (JS 23)

$$K = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}y_j})/|1-\mathrm{e}^{-\mathrm{i}y_j}|\Big) \quad \text{and} \quad R = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}x_k})/|1-\mathrm{e}^{-\mathrm{i}x_k}|\Big)$$

$$\mathrm{e}^{\mathrm{i}y_j} K_{jj} = \mathrm{e}^{\mathrm{i}y_j}\frac{(1-\mathrm{e}^{-\mathrm{i}y_j})}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = \frac{\mathrm{e}^{\mathrm{i}y_j}-1}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = -K_{jj}{}^*, \qquad \mathrm{e}^{\mathrm{i}x_k} R_{kk} = -R_{kk}{}^*$$

$$S_f K = -K^*, \quad \text{and} \quad RS_F = -R^*.$$

We introduce a rotated version of the Loewner matrix, $\widehat{L} := -\mathrm{i}RLK$

$$\widehat{L} = -\mathrm{i}RS_F MK + \mathrm{i}RMS_f K = \mathrm{i}(R^* MK - RMK^*) = 2\mathrm{Im}(RMK^*) \in \mathbb{R}^{n\times m},$$

$$\widehat{L}\widehat{V} = \widehat{U}S \quad \text{(real SVD)} \Rightarrow LV = US,\ V = \mathrm{i}K\widehat{V} \text{ and } U = -R^*\widehat{U}.$$

$\widetilde{w} = Ve_m$ minimizes $\|Lu\|_2$. Since $\widetilde{w} = \mathrm{i}K\widehat{V}e_m$,

$$\mathrm{e}^{\mathrm{i}y_j}\widetilde{w}_j = (S_f\widetilde{w})_j = (\mathrm{i}S_f K\widehat{V}e_m)_j = -(\mathrm{i}K^*\widehat{V}e_m)_j = \widetilde{w}_j{}^*.$$

$$\widetilde{n}(x) = \sum_{j=1}^{m}\frac{\mathrm{e}^{\mathrm{i}y_j}\widetilde{w}_j}{x-y_j} = \sum_{j=1}^{m}\frac{\widetilde{w}_j^*}{x-y_j} = \widetilde{d}(x)^* \quad \Rightarrow \quad \widetilde{r}(x) = \xi(x)^*\xi(x)^{-1}.$$

Modified AAA via a rotated Loewner matrix (JS 23)

$$K = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}y_j})/|1-\mathrm{e}^{-\mathrm{i}y_j}|\Big) \quad \text{and} \quad R = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}x_k})/|1-\mathrm{e}^{-\mathrm{i}x_k}|\Big)$$

$$\mathrm{e}^{\mathrm{i}y_j} K_{jj} = \mathrm{e}^{\mathrm{i}y_j} \frac{(1-\mathrm{e}^{-\mathrm{i}y_j})}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = \frac{\mathrm{e}^{\mathrm{i}y_j}-1}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = -K_{jj}{}^*, \qquad \mathrm{e}^{\mathrm{i}x_k} R_{kk} = -R_{kk}{}^*$$

$$S_f K = -K^*, \quad \text{and} \quad RS_F = -R^*.$$

We introduce a rotated version of the Loewner matrix, $\widehat{L} := -\mathrm{i}RLK$

$$\widehat{L} = -\mathrm{i}RS_F MK + \mathrm{i}RMS_f K = \mathrm{i}(R^* MK - RMK^*) = 2\mathrm{Im}(RMK^*) \in \mathbb{R}^{n\times m},$$

$$\widehat{L}\widehat{V} = \widehat{U}S \quad \text{(real SVD)} \Rightarrow LV = US,\ V = \mathrm{i}K\widehat{V} \text{ and } U = -R^*\widehat{U}.$$

$\widetilde{w} = Ve_m$ minimizes $\|Lu\|_2$. Since $\widetilde{w} = \mathrm{i}K\widehat{V}e_m$,

$$\mathrm{e}^{\mathrm{i}y_j}\widetilde{w}_j = (S_f\widetilde{w})_j = (\mathrm{i}S_f K\widehat{V}e_m)_j = -(\mathrm{i}K^*\widehat{V}e_m)_j = \widetilde{w}_j{}^*.$$

$$\widetilde{n}(x) = \sum_{j=1}^m \frac{\mathrm{e}^{\mathrm{i}y_j}\widetilde{w}_j}{x-y_j} = \sum_{j=1}^m \frac{\widetilde{w}_j^*}{x-y_j} = \widetilde{d}(x)^* \quad \Rightarrow \quad \widetilde{r}(x) = \xi(x)^*\xi(x)^{-1}.$$

Modified AAA: $\widetilde{r}$ is unitary, i.e. $|\widetilde{r}(x)| = 1$.

Modified AAA via a rotated Loewner matrix (JS 23)

$$K = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}y_j})/|1-\mathrm{e}^{-\mathrm{i}y_j}|\Big) \quad \text{and} \quad R = \mathrm{diag}\Big((1-\mathrm{e}^{-\mathrm{i}x_k})/|1-\mathrm{e}^{-\mathrm{i}x_k}|\Big)$$

$$\mathrm{e}^{\mathrm{i}y_j} K_{jj} = \mathrm{e}^{\mathrm{i}y_j} \frac{(1-\mathrm{e}^{-\mathrm{i}y_j})}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = \frac{\mathrm{e}^{\mathrm{i}y_j}-1}{|1-\mathrm{e}^{-\mathrm{i}y_j}|} = -K_{jj}{}^*, \qquad \mathrm{e}^{\mathrm{i}x_k} R_{kk} = -R_{kk}{}^*$$

$$S_f K = -K^*, \quad \text{and} \quad RS_F = -R^*.$$

We introduce a rotated version of the Loewner matrix, $\widehat{L} := -\mathrm{i}RLK$

$$\widehat{L} = -\mathrm{i}RS_F MK + \mathrm{i}RMS_f K = \mathrm{i}(R^* MK - RMK^*) = 2\mathrm{Im}(RMK^*) \in \mathbb{R}^{n\times m},$$

$$\widehat{L}\widehat{V} = \widehat{U}S \quad \text{(real SVD)} \Rightarrow LV = US,\ V = \mathrm{i}K\widehat{V} \text{ and } U = -R^*\widehat{U}.$$

$\widetilde{w} = Ve_m$ minimizes $\|Lu\|_2$. Since $\widetilde{w} = \mathrm{i}K\widehat{V}e_m$,

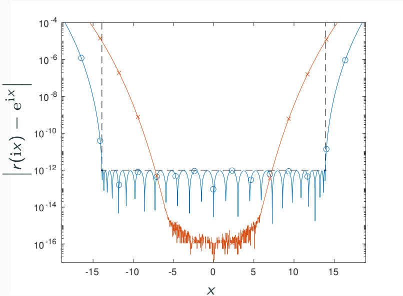
$$\mathrm{e}^{\mathrm{i}y_j}\widetilde{w}_j = (S_f\widetilde{w})_j = (\mathrm{i}S_f K\widehat{V}e_m)_j = -(\mathrm{i}K^*\widehat{V}e_m)_j = \widetilde{w}_j{}^*.$$

$$\widetilde{n}(x) = \sum_{j=1}^{m} \frac{\mathrm{e}^{\mathrm{i}y_j}\widetilde{w}_j}{x-y_j} = \sum_{j=1}^{m} \frac{\widetilde{w}_j^*}{x-y_j} = \widetilde{d}(x)^* \quad \Rightarrow \quad \widetilde{r}(x) = \xi(x)^*\xi(x)^{-1}.$$

Modified AAA: $\widetilde{r}$ is unitary, i.e. $|\widetilde{r}(x)| = 1$. AAA: $w = \mathrm{e}^{\mathrm{i}\phi}Ve_m$, $r = \mathrm{e}^{2\mathrm{i}\phi}d^*/d$.

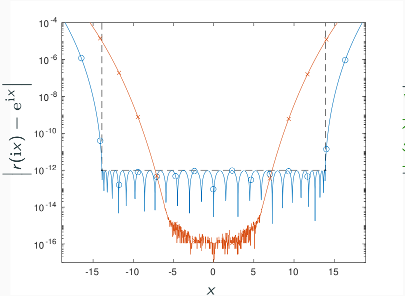
Unitarity to machine precision

Error in approximation of e^{ix}
(Padé vs AAA–Lawson)

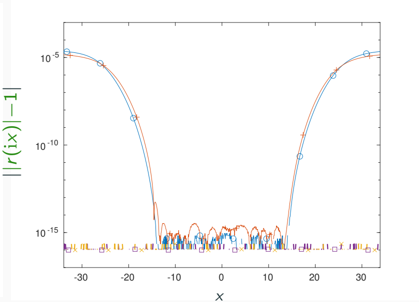


Unitarity to machine precision

Error in approximation of $\mathrm{e}^{\mathrm{i}x}$
(Padé vs AAA–Lawson)

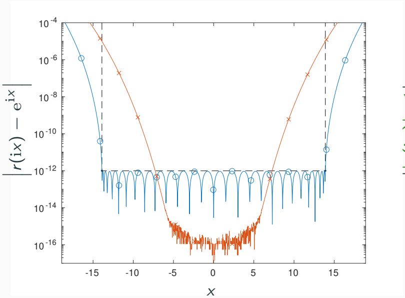


Error in unitarity of approximant r
(original and modified AAA/AAA–Lawson)

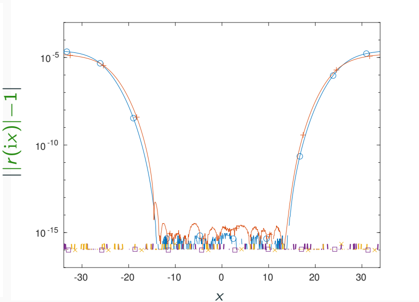


Unitarity to machine precision

Error in approximation of $\mathrm{e}^{\mathrm{i}x}$
(Padé vs AAA–Lawson)



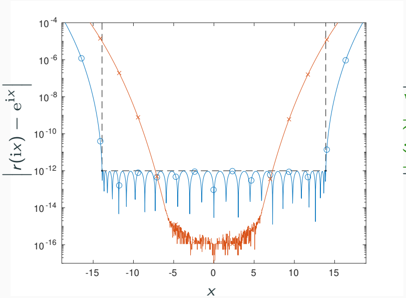
Error in unitarity of approximant r
(original and modified AAA/AAA–Lawson)



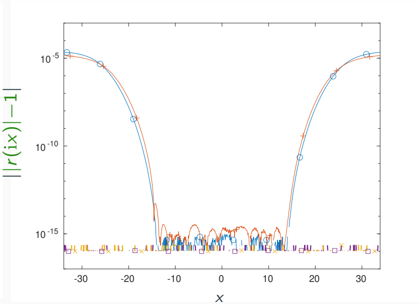
While theoretically AAA and AAA–Lawson should produce unitary approximants, this not true in computer arithmetic.

Unitarity to machine precision

Error in approximation of e^{ix}
(Padé vs AAA–Lawson)



Error in unitarity of approximant r
(original and modified AAA/AAA–Lawson)



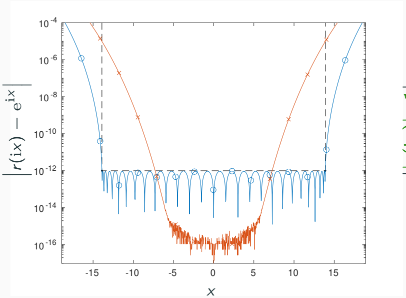
While theoretically AAA and AAA–Lawson should produce unitary approximants, this not true in computer arithmetic. The Cayley form

$$\widetilde{r}(x) = \xi(x)^* \xi(x)^{-1}$$

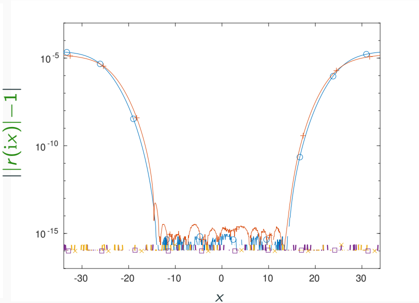
of modified AAA and AAA–Lawson ensures unitarity to machine precision.

Unitarity to machine precision

Error in approximation of $\mathrm{e}^{\mathrm{i}x}$
(Padé vs AAA–Lawson)



Error in unitarity of approximant r
(original and modified AAA/AAA–Lawson)



While theoretically AAA and AAA–Lawson should produce unitary approximants, this not true in computer arithmetic. The Cayley form

$$\widetilde{r}(x) = \xi(x)^* \xi(x)^{-1}$$

of modified AAA and AAA–Lawson ensures unitarity to machine precision.

[JS 23] *Unitarity of some barycentric rational approximants*, Jawecki & S, arXiv:2205.10606