Some new developments on sum-of-exponential approximations

Motivated by applications in many-body physics

Zhen Huang

Department of Mathematics, University of California, Berkeley

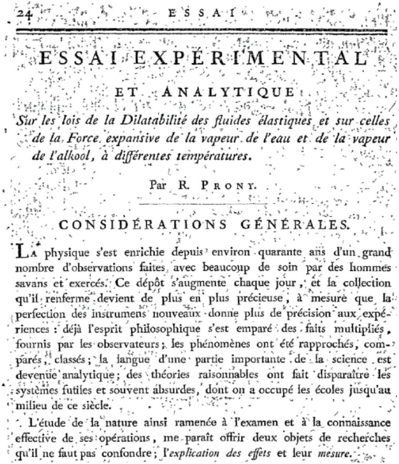
hertz@math.berkeley.edu

November 4, 2025

Table of Contents

Sum-of-exponential approximations

Long list of history, dating back to 1795:
Journal de l'Ecole Polytechnique, 1795, Vol.1 (2), pp. 24-36, by Gaspard Riche de Prony:



24 ESSAI

ESSAI EXPÉRIMENTAL

ET ANALYTIQUE:

Sur les lois de la Dilatabilité des fluides élastiques et sur celles de la Force expansive de la vapeur de l'eau et de la vapeur de l'alkool, à différentes températures.

Par R. PRONY.

CONSIDÉRATIONS GÉNÉRALES.

LA physique s'est enrichie depuis environ quarante ans d'un grand nombre d'observations faites avec beaucoup de soin par des hommes savans et exercés. Ce dépôt s'augmente chaque jour, et la collection qu'il renferme devient de plus en plus précieuse, à mesure que la perfection des instrumens nouveaux donne plus de précision aux expériences : déjà l'esprit philosophique s'est emparé des faits multipliés, fournis par les observateurs; les phénomènes ont été rapprochés, comparés, classés; la langue d'une partie importante de la science est devenue analytique; des théories raisonnables ont fait disparaître les systèmes futiles et souvent absurdes, dont on a occupé les écoles jusqu'au milieu de ce siècle.

L'étude de la nature ainsi ramenée à l'examen et à la connaissance effective de ses opérations, me paraît offrir deux objets de recherches qu'il ne faut pas confondre; *l'explication des effets* et leur *mesure*.

30 ESSAI

valeurs correspondantes de $x \ldots 0;\ x_{\prime};\ 2x_{\prime} \ldots nx_{\prime};\ (n+1)x_{\prime};$
$(n+2)x_{\prime} \ldots (2n-1)x_{\prime}$

les quantités $z_0, z_{\prime}, z_{\prime\prime}$, &c. doivent former une suite récurrente dont il faut trouver l'échelle de relation; soient $A_0, A_{\prime}, A_{\prime\prime} \ldots A_{(n)}$ des coëfficiens indéterminés, tels qu'on ait les équations de condition,

$$A_0 z_0 + A_{\prime} z_{\prime} + A_{\prime\prime} z_{\prime\prime} + \ldots + A_{(n)} z_{(n)} = 0$$
$$A_0 z_{\prime} + A_{\prime} z_{\prime\prime} + A_{\prime\prime} z_{\prime\prime\prime} + \ldots + A_{(n)} z_{(n+1)} = 0$$
$$A_0 z_{\prime\prime} + A_{\prime} z_{\prime\prime\prime} + A_{\prime\prime} z_{iv} + \ldots + A_{(n)} z_{(n+2)} = 0$$
$$A_0 z_{\prime\prime\prime} + A_{\prime} z_{iv} + A_{\prime\prime} z_{v} + \ldots + A_{(n)} z_{(n+3)} = 0$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$A_0 z_{(n-1)} + A_{\prime} z_{(n)} + A_{\prime\prime} z_{(n+1)} + \ldots + A_{(n)} z_{(2n-1)} = 0.$$

On pourra, pour plus de commodité, supposer $A_{(n)} = 1$ dans les applications numériques.

Ces équations étant en nombre n donneront les n rapports $\frac{A_0}{A_{(n)}}$; $\frac{A_{\prime}}{A_{(n)}}$; $\frac{A_{\prime\prime}}{A^{(n)}}$; .... $\frac{A_{(n-1)}}{A_{(n)}}$ qui composent l'échelle de relation demandée, et on aura

Sum-of-exponential approximations

$$C(t) \approx \sum_{j=1}^{N} w_j \mathrm{e}^{-\mathrm{i} z_j t}, \quad t \in [0, T].$$

Many previous works: [Beylkin & Monzón, Appl. Comput. Harmon. Anal., 19, 17 (2005).], [Beylkin & Monzón, Appl. Comput. Harmon. Anal., 27 (2009).], [Potts & Tasche, Linear Algebra Appl., 439 (2013).], [Wilber, Damle & Townsend, SIAM J. Sci. Comput., 2022.], etc.

Sum-of-exponential approximations

$$C(t) \approx \sum_{j=1}^{N} w_j \mathrm{e}^{-\mathrm{i} z_j t}, \quad t \in [0, T].$$

Many previous works: [Beylkin & Monzón, Appl. Comput. Harmon. Anal., 19, 17 (2005).], [Beylkin & Monzón, Appl. Comput. Harmon. Anal., 27 (2009).], [Potts & Tasche, Linear Algebra Appl., 439 (2013).], [Wilber, Damle & Townsend, SIAM J. Sci. Comput., 2022.], etc.

Relations to rational approximations via Fourier transform:

$$J(\omega) \approx -\frac{1}{\pi} \mathrm{Im} \left(\sum_{j=1}^{N} \frac{w_j}{\omega - z_j} \right)$$

AAA algorithm: [Nakatsukasa, Sète & Trefethen, SIAM J. Sci. Comput., 2018.]

Physical backgrounds

$C(t) \approx \sum_{j=1}^{N} w_j \mathrm{e}^{-\mathrm{i} z_j t}, \quad t \in [0, T].$

How does considering physical backgrounds change the game?

1. Restricted class of functions $C(t)$.
2. Constraints on weights w_j and poles z_j.

Physical backgrounds

$C(t) \approx \sum_{j=1}^{N} w_j \mathrm{e}^{-\mathrm{i} z_j t}, \quad t \in [0, T].$

How does considering physical backgrounds change the game?

1. Restricted class of functions $C(t)$.
2. Constraints on weights w_j and poles z_j.

- Class 1: imaginary-time correlation functions:

$$C(\tau) = \int_{\mathbb{R}} K_\beta(\tau, \omega) J(\omega) \mathrm{d}\omega, \quad \tau \in [0, \beta],$$

where $K_\beta(\tau, \omega) = \frac{\mathrm{e}^{-\tau\omega}}{1 \pm \mathrm{e}^{-\beta\omega}}$, $\mathrm{supp}(J) \subseteq [-W, W]$, $J \geq 0$.

- Class 2: real-time correlation functions:

$$C(t) = \int_{\mathbb{R}} J(\omega) f(\omega) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}\omega, \quad t \in [0, T],$$

Here $f(\omega) = \frac{1}{1 \pm \mathrm{e}^{\beta(\omega - \mu)}}$, $\mathrm{supp}(J) \subseteq [-W, W]$, $J \geq 0$.

- We will focus on Class 2 in this talk, since class 1 is relatively well-studied.

Imaginary-time correlation functions

$$C(\tau) = \int_{\mathbb{R}} K_\beta(\tau, \omega) J(\omega) \mathrm{d}\omega, \quad \tau \in [0, \beta],$$

where $K_\beta(\tau, \omega) = \frac{\mathrm{e}^{-\tau\omega}}{1 \pm \mathrm{e}^{-\beta\omega}}$, $\mathrm{supp}(J) \subseteq [-W, W]$, $J \geq 0$.
We want to approximate $C(\tau)$ by

$$C(\tau) \approx \sum_{j=1}^{N} w_j \mathrm{e}^{-z_j \tau}, \quad \tau \in [0, \beta], \quad z_j \in \mathbb{R},$$

Imaginary-time correlation functions

$$C(\tau) = \int_{\mathbb{R}} K_\beta(\tau,\omega) J(\omega)\mathrm{d}\omega, \quad \tau \in [0,\beta],$$

where $K_\beta(\tau,\omega) = \frac{\mathrm{e}^{-\tau\omega}}{1\pm\mathrm{e}^{-\beta\omega}}$, $\mathrm{supp}(J) \subseteq [-W,W]$, $J \geq 0$.
We want to approximate $C(\tau)$ by

$$C(\tau) \approx \sum_{j=1}^{N} w_j \mathrm{e}^{-z_j\tau}, \quad \tau \in [0,\beta], \quad z_j \in \mathbb{R},$$

Quick review of known results:

1. Number of terms N required to reach ε-accuracy has the scaling $N \sim O(\log(\beta W/\varepsilon)^2)$.
2. Efficient algorithms: [Shinaoka, Otsuki et al 2017], [Kaye, Chen, Parcollet 2022], [Huang, Kaye, Strand, Denis, 2025].
3. Sometimes we have the additional physical constraint: $w_j \geq 0$. Or $w_j \succeq 0$ being positive semidefinite matrices when $C(\tau)$ is matrix-valued. Algorithms incorporating physical constraints: [Huang, Gull, Lin, 2023].

Real-time correlation functions

Focus on fermionic cases. Bosonic cases are similar.
Where does $C(t)$ come from?

$$C(t) = \int_{\mathbb{R}} J(\omega) f(\omega) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}\omega, \quad f(\omega) = \frac{1}{1 + \mathrm{e}^{\beta(\omega-\mu)}}, \quad t \in [0, T],$$

1. The spectral density function $J(\omega) \geq 0$, compactly supported or decaying exponentially.
2. The chemical potential μ and inverse temperature $\beta \in [0, +\infty]$.

What is $f(\omega)$?

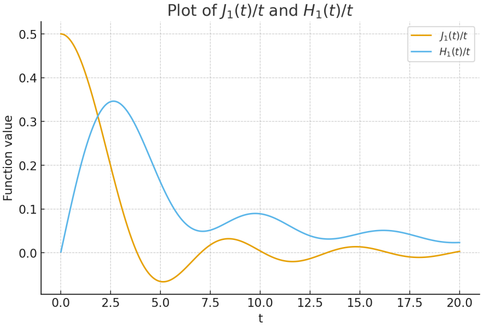
- If $\beta = 0$, then $f(\omega) = \frac{1}{2}$.
- If $\beta = +\infty$, then $f(\omega) = 1$ for $\omega < \mu$, $f(\omega) = 0$ for $\omega > \mu$.

Examples

$J(\omega) = \frac{2}{\pi}\sqrt{1-\omega^2}\cdot 1_{|\omega|\le 1}$. (Semi-circle).

- $\beta = +\infty$, $\mu = 0$.Then $f = 1$ for $\omega < 0$, $f = 0$ for $\omega > 0$.

$$C(t) = \frac{2}{\pi}\int_{-1}^{0}\sqrt{1-\omega^2}\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega = \frac{J_1(t)}{t} - \mathrm{i}\frac{H_1(t)}{t}.$$



Question: how many terms N are needed to approximate $C(t) \approx \sum_{j=1}^{N} w_j \mathrm{e}^{-\mathrm{i}z_j t}$ to accuracy ε on $[0, T]$?

What could be the factors affecting N?

$$C(t) = \int_{\mathbb{R}} J(\omega) f(\omega) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}\omega, \quad f(\omega) = \frac{1}{1+\mathrm{e}^{\beta(\omega-\mu)}}, \quad t \in [0,T],$$

$$C(t) \approx \sum_{j=1}^{N} w_j \mathrm{e}^{-\mathrm{i} z_j t}, \quad t \in [0,T].$$

1. Accuracy ε?
2. Maximum time T?
3. Temperature β? Bandwidth W? ($\mathrm{supp}(J) \subseteq [-W,W]$.)
4. Smoothness of J?

Previous analysis [Abanin 2024] with strong assumptions on analyticity of J, shows $N \sim O(\log(T/\varepsilon))^2$.

What could be the factors affecting N?

$$C(t) = \int_{\mathbb{R}} J(\omega) f(\omega) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}\omega, \quad f(\omega) = \frac{1}{1 + \mathrm{e}^{\beta(\omega - \mu)}}, \quad t \in [0, T],$$

$$C(t) \approx \sum_{j=1}^{N} w_j \mathrm{e}^{-\mathrm{i} z_j t}, \quad t \in [0, T].$$

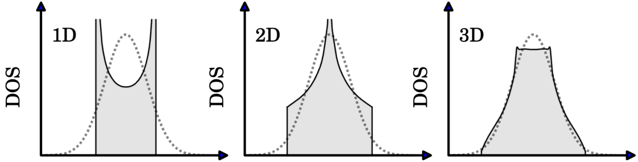
1. Accuracy ε?
2. Maximum time T?
3. Temperature β? Bandwidth W? ($\operatorname{supp}(J) \subseteq [-W, W]$.)
4. Smoothness of J?

Previous analysis [Abanin 2024] with strong assumptions on analyticity of J, shows $N \sim O(\log(T/\varepsilon))^2$.Our question:

1. Is there really a dependence on T? Unclear from numerics.
2. Generally unreasonable to assume analyticity of J. How to do complexity analysis in such cases?

Singularities of spectral density.

By the theory of van-Hove singularities, $J(\omega)$ could (provably) exhibit square-root, logarithmic, discontinuous, and even inverse square-root singularities.



A prototypical example beyond previous analysis: semicircular density:

$$J(\omega) = \begin{cases} \frac{2}{\pi}\sqrt{1-\omega^2}, & |\omega| \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Our result: clarify the origins of the T-dependence.

$$C(t) = \int_{\mathbb{R}} J_{\text{eff}}(\omega) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}\omega, \quad t \in [0, T],$$

$$C(t) \approx \sum_{j=1}^{N} w_j \mathrm{e}^{-\mathrm{i} z_j t}, \quad t \in [0, T].$$

Informal statement:

1. If measuing accuracy in L^∞ norm, then $N \sim O(\log^2(1/\epsilon))$, independent of T!
2. If measuring accuracy in L^1 norm, then:
   - If J_{eff} satisfies continuity, then $N \sim O(\log^2(1/\epsilon))$, independent of T!
   - If J_{eff} admits discontinuous singularities, then $N \sim O(\log^2(T/\epsilon))$.
3. In both cases, the dependence is independent of the inverse temperature β.

Our result: clarify the origins of the T-dependence.

$$C(t) = \int_{\mathbb{R}} J_{\text{eff}}(\omega)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega, \quad t \in [0,T],$$

$$C(t) \approx \sum_{j=1}^{N} w_j \mathrm{e}^{-\mathrm{i}z_j t}, \quad t \in [0,T].$$

Informal statement:

1. If measuing accuracy in L^∞ norm, then $N \sim O(\log^2(1/\epsilon))$, independent of T!
2. If measuring accuracy in L^1 norm, then:
    - If J_{eff} satisfies continuity, then $N \sim O(\log^2(1/\epsilon))$, independent of T!
    - If J_{eff} admits discontinuous singularities, then $N \sim O(\log^2(T/\epsilon))$.
3. In both cases, the dependence is independent of the inverse temperature β.

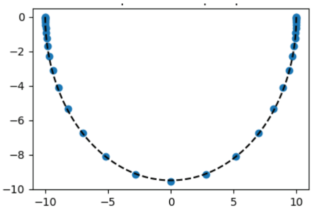
The T-dependence/independence is crucial, as it determines the overall cost of simulating an open quantum system to long times: how it scales with the final time T.

Motivation of the proof: numerical experiments

Consider the following example:

$$J(\omega) = \sqrt{1-(\omega/W)^2}, \quad W = 10, \quad C(t) = \int_{-W}^{W} J(\omega)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}\omega.$$

Numerically conducting an sum-of-exponential approximation of $C(t) \approx \sum_j \omega_j \mathrm{e}^{-\mathrm{i}z_j t}$ on $t \in [0, 100]$ to accuracy $\varepsilon = 10^{-8}$. Where are the poles z_j's found by numerical algorithms?

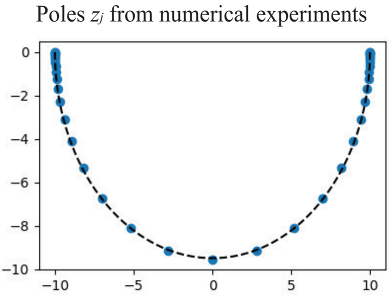


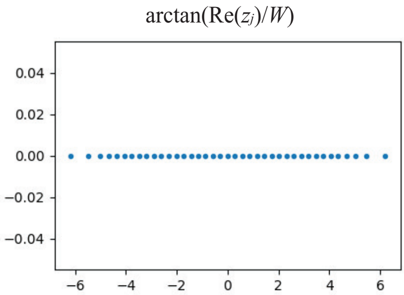
Looks like (1) forming a semi-eclipse; (2) exponentially clustered near $\omega = \pm 10$.

Similar exponential clustering near singularities in many topics in applied math. See Huybrechs, Trefethen, SIAM Rev., 2024.

Change of variables and quadrature

The poles z_j from numerical experiments, after change of variables (that maps $\pm W$ to $\pm\infty$),looks like uniform quadrature:

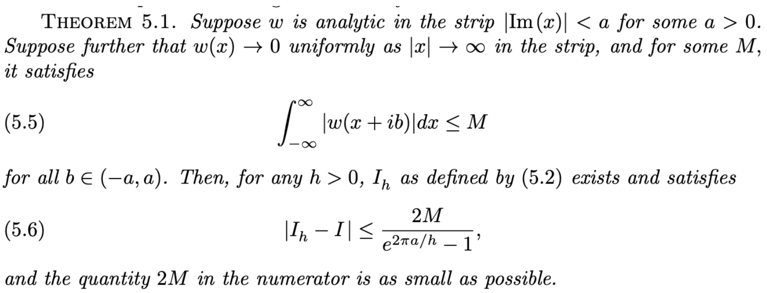

Poles z_j from numerical experiments


$\arctan(\mathrm{Re}(z_j)/W)$

(Also seen here, the linear tapering effect: the uniform grid becomes sparser towards the boundary. Similar phenomenon in rational approximation and finite element hp mesh. See Huybrechs, Trefethen, SIAM Rev., 2024.)

Uniform quadrature mesh on $\mathbb{R}$, assuming analyticity in the strip $|\mathrm{Im}(z)| < a$

Quadrature error with uniform mesh on $\mathbb{R}$ decays exponentially, with the rate depending on the width a of the strip of analyticity:



THEOREM 5.1. *Suppose w is analytic in the strip $|\mathrm{Im}\,(x)| < a$ for some $a > 0$. Suppose further that $w(x) \to 0$ uniformly as $|x| \to \infty$ in the strip, and for some M, it satisfies*

$$\int_{-\infty}^{\infty} |w(x+ib)|dx \le M \tag{5.5}$$

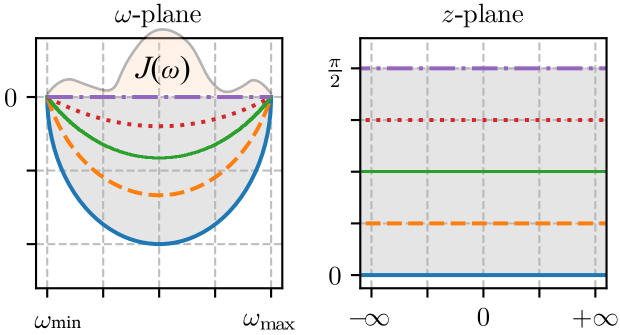
for all $b \in (-a, a)$. Then, for any $h > 0$, I_h as defined by (5.2) exists and satisfies

$$|I_h - I| \le \frac{2M}{e^{2\pi a/h} - 1}, \tag{5.6}$$

and the quantity $2M$ in the numerator is as small as possible.

Proof sketch

The above numerical observation motivates us to consider contour deformation + change of variables + uniform quadrature:



Open problems

1. Complexity analysis for SOE approximations with physical constraints:
$$C(t) \approx g^\dagger \mathrm{e}^{(-\mathrm{i}H-\Gamma)t} g, \quad t \in [0,T], \quad \Gamma \succeq 0.$$
How does N (size of H, Γ) scale with T and target accuracy ε?
2. Matrix-valued SOE approximation and rational approximation. ($W_j \in \mathbb{C}^{M\times M}, W_j \succeq 0$.)
$$C(t) \approx \sum_{j=1}^{N} W_j \mathrm{e}^{-z_j t}, \quad J(\omega) \approx \sum_k \frac{W_k}{\omega - z_k}.$$
Fast algorithms when both M and N are large?
3. Multi-variate SOE approximations and rational approximations. Fast and robust algorithms?
$$C(\boldsymbol{t}) \approx \sum_{j=1}^{N} w_j \mathrm{e}^{-\mathrm{i}\boldsymbol{z}\cdot\boldsymbol{t}}, \quad J(\boldsymbol{\omega}) \approx \sum_k \frac{w_k}{\boldsymbol{\omega} - \boldsymbol{z}_k}.$$

Thank you for your attention!