TAME: Inverse Laplace Transform of Tame Functions

Nikita Deniskin[1,2] Federico Poloni[2]

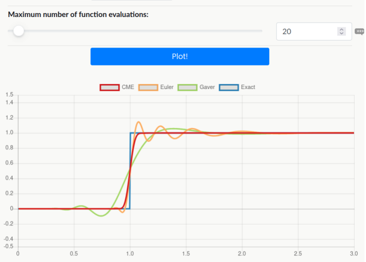
[1]Scuola Normale Superiore, Pisa
[2]University of Pisa

12th Conference on Matrix Analytic Methods
Toronto, June 2025

Laplace is Fourier's Evil Twin

Inverting Laplace transforms is an ill-posed problem. E.g., $f(x) = 0$ and $g(x) = \sin\left(\frac{1}{\varepsilon}t\right)$ have $\|\hat{f} - \hat{g}\|_\infty = \varepsilon$ but $\|f - g\|_\infty = 1$.

Lots of numerical attention, e.g., [Cohen book '07].



From inverselaplace.org, Horváth^2, Telek, Almousa, Talyigás.

But in queuing theory we often deal with very special functions:

- Phase-type distributions $f(t) = \boldsymbol{\alpha}^T \exp(tQ)\mathbf{q}$ [Neuts, Latouche–Ramaswami, Asmussen–o'Cinneide, Bean–Fackrell–Taylor, ...]
- First-return matrix $\Psi(t)$ in fluid queues [Bean–O'Reilly–Taylor '08].

Can we do better?

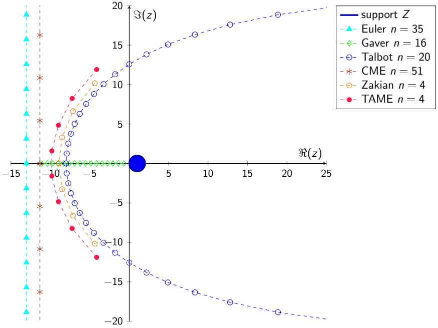
Setup: Abate–Whitt methods [Abate–Whitt '06]

Quadrature-like formulas to reconstruct f from a few evaluations of $\widehat{f}$:

Abate–Whitt Framework

$$f(t) \approx f_N(t) = \sum_{n=1}^{N} \frac{w_i}{t} \widehat{f}\left(\frac{\beta_i}{t}\right). \quad \text{(Often, } t = 1.\text{)}$$

Several classical methods exist, with different locations for the poles β_n.



Numerics?

Abate–Whitt gave us a quick recipe:

Table from [Abate–Whitt '06]:

Inversion Routine	Number of transform evaluations	system precision reqd. decimal digits	sig. digits produced for good transform	efficiency (SD/Prec.)
Gaver-Stehfest	2M	2.2 M	0.90 M	0.4
Euler	2M+1	1.0 M	0.60 M	0.6
Talbot	M	1.0 M	0.60 M	0.6

Table 1: A rough analysis of cost and efficiency for the three specified one-dimensional inversion routines.

- Focus on multiprecision rather than usual `floats`.
- No / mild dependence on the function, scaling. . . . ?

Our goal here:

- Identify relevant classes of functions and approximation properties.
- Give explicit error bounds.
- Suggest a new method (TAME) based on recent tools.

CME method [Horváth[2], Telek, Almousa, Talyigás]

Idea: construct a concentrated mixture of exponentials to approximate the delta distribution δ_1.

Figure from [Horváth[2], Telek '20]

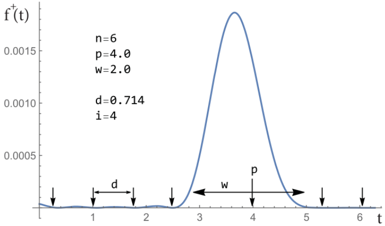


Figure 9. The spike and the zeros of $f^+(t)$.

$$f(1) = \int_0^\infty f(z)\delta_1(z)\,dz \approx \int_0^\infty f(z)\left(\sum_{n=1}^{N} w_n e^{-\beta_n z}\right)\,dz = \sum_{n=1}^{N} w_n \widehat{f}(\beta_n).$$

Zakian's method and rational approximation [Zakian '70]

A related idea: take the Laplace transform of $\delta_1(z) \approx \sum_{n=1}^{N} w_n e^{-\beta_n z}$ to get

$$e^{-y} \approx \sum_{n=1}^{N} \frac{w_n}{y+\beta_n} = \frac{p(y)}{q(y)}.$$

A classical problem in interpolation theory: rational approximation.

Zakian suggests a Padé approximant, i.e., a rational function $r(y) = \frac{p(y)}{q(y)}$ which matches a certain number of derivatives

$$\left.\frac{d^j}{dy^j} e^{-y}\right|_{y=0} = \left.\frac{d^j}{dy^j} r(y)\right|_{y=0} \quad j = 0, 1, 2, \ldots N-1.$$

It's not immediately clear why this works: small approximation error in the "y-domain" does not necessarily mean small error in the "z-domain".

Sums of exponentials

Definition

We say that an Abate–Whitt method $(\beta_n, w_n)_{n=1}^N$ is ε-accurate on Ω when

$$\left| e^{-y} - \sum_{n=1}^{N} \frac{w_n}{y+\beta_n} \right| \leq \varepsilon \quad y \in \Omega.$$

Theorem

Let

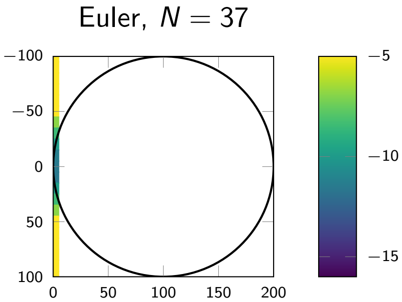
$$f(t) = \sum_{m=1}^{M} c_m e^{-\alpha_m t},$$

and take an AW method ε-accurate on $\Omega \supset \{\alpha_1 t, \dots, \alpha_M t\}$. Then, the error of the method is

$$|f(t) - f_N(t)| \leq (|c_1| + \cdots + |c_M|)\, \varepsilon.$$

Insights and generalizations

- Each method is accurate only in a limited region $\Omega \subseteq \mathbb{C}$
- The Ω you need depends on t and on the exponents in f.



Generalization: $\exp(tQ)$?

With a Jordan form $Q = VJV^{-1}$, the entries of V might be arbitrarily large.

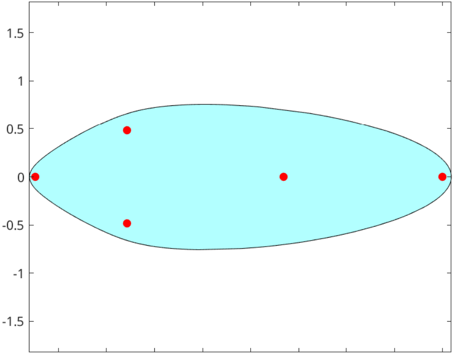
A technical tool lets us avoid this problem.

Field of values [Gustafson–Rao book '97]

Definition

The field of values (or *numerical range*) of $A \in \mathbb{C}^{d\times d}$ is

$$W(A) = \left\{ z \in \mathbb{C} \colon z = \frac{\mathbf{u}^* A\,\mathbf{u}}{\mathbf{u}^*\mathbf{u}},\ \mathbf{u} \in \mathbb{C}^d \setminus \{0\} \right\}.$$



A matrix version

Theorem [Crouzeix–Palencia '17]

Let ϕ be a holomorphic function; Then

$$\|\phi(A)\|_2 \leq (1+\sqrt{2}) \max_{z \in W(A)} |\phi(z)|.$$

Theorem

Let

$$f(t) = \exp(tQ),$$

and take an AW method ε-accurate on $\Omega \supset W(-tQ)$. Then, the error of the method is

$$\|f(t) - f_N(t)\|_2 \leq (1+\sqrt{2})\varepsilon.$$

The bound holds only for the 2-norm (but $\|\mathbf{v}\|_2 \leq \|\mathbf{v}\|_1$).

Results for phase-type distributions

A phase-type distribution is the absorption time of an absorbing CTMC:

$$\underbrace{f(t)}_{\text{pdf}} = \boldsymbol{\alpha}^T \exp(tQ)\mathbf{q}, \quad \underbrace{F(t)}_{\text{CDF}} = 1 - \boldsymbol{\alpha}^T \exp(tQ)\mathbf{1}, \quad \mathbf{q} = -Q\mathbf{1},\ \boldsymbol{\alpha}^T\mathbf{1} = 1.$$

Theorem

Let $(w_n, \beta_n)_{n=1}^N$ be ε-accurate on a region $\Omega \supset W(-tQ)$. Then,

$$|f(t) - f_N(t)| \leq (1+\sqrt{2})\varepsilon\|\mathbf{q}\|_1,$$
$$|F(t) - F_N(t)| \leq \varepsilon + (1+\sqrt{2})\varepsilon \underbrace{(-\boldsymbol{\alpha}^T Q^{-1}\mathbf{1})}_{\text{1st moment}}\|\mathbf{q}\|_1.$$

The tricky part: estimating $W(Q)$.

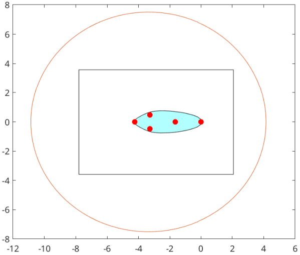
Field of values bounds

For any given Q, the field of values $W(Q)$ can be estimated using a few eigenvalue computations. Otherwise, this general result holds true:

Theorem

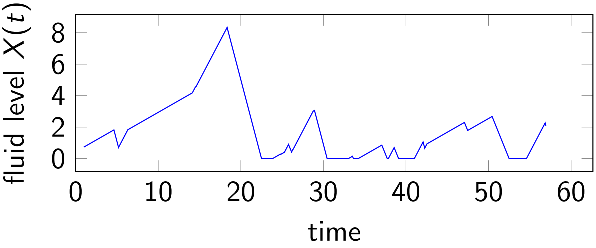
Let $Q \in \mathbb{R}^{d\times d}$ be a generator, λ a uniformization rate, i.e., $\lambda \geq \max|Q_{ii}|$. Then, $W(Q)$ lies inside a circle with center $-\lambda$ and radius $\sqrt{d}\lambda$.

One can reduce the area by using results in [Gau–Wang–Wu '16], but the region still depends on the dimension $\sim \frac{\sqrt{d}}{2}$.



The Ψ matrix in fluid queues [Asmussen, Ramaswami, ...]

Markov-modulated fluid models $Q \in \mathbb{R}^{d\times d}$, rates $r_1, \dots, r_d$.



τ = time of first return to starting level.

$$\Psi(t)_{ij} = P[\tau < t, \varphi(t) = j \in S_- \mid \varphi(0) = i \in S_+].$$

$$\underbrace{\psi(t)}_{\text{pdf}} = \frac{d}{dt}\underbrace{\Psi(t)}_{\text{CDF}}.$$

Its Laplace–Stieltjes transform $\widehat{\psi}(s)$ can be computed with various algorithms based on Riccati equations. [Bean–O'Reilly–Taylor '08]

Formulas for Ψ

Closed formula [Barbot–Sericola–Telek '01, Bean–Nguyen–Poloni '19]

$$\psi(t) = \sum_{k=1}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^{k-1}}{(k-1)!} B_k.$$

Here, λ is a uniformization rate, and

$$B_k = P[\text{first return is between } k\text{th and } k+1\text{st uniformization event}].$$

Equivalently, $\psi_{ij}(t)$ is the $\lim_{K\to\infty}$ of $\boldsymbol{\alpha}^T \exp(tQ)\mathbf{q}$, with $K \times K$

$$Q = \begin{bmatrix} -\lambda & \lambda & & & \\ & -\lambda & \lambda & & \\ & & \ddots & \ddots & \\ & & & -\lambda & \lambda \\ & & & & -\lambda \end{bmatrix}, \quad \boldsymbol{\alpha} = \begin{bmatrix} [B_K]_{ij} \\ [B_{K-1}]_{ij} \\ \vdots \\ [B_2]_{ij} \\ [B_1]_{ij} \end{bmatrix}, \quad \mathbf{q} = -Q\mathbf{1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \lambda \end{bmatrix}.$$

Results for Ψ

Theorem

Let $(w_n, \beta_n)_{n=1}^N$ be an AW method that is ε-accurate on a ball $B(t\lambda, t\lambda)$. Then,

$$|\psi(t) - \psi_N(t)| \leq \varepsilon\lambda(1+\sqrt{2})\Psi(\infty),$$
$$|\Psi(t) - \Psi_N(t)| \leq \varepsilon\Psi(\infty) + (1+\sqrt{2})\varepsilon\lambda \underbrace{\mathbb{E}[\psi(t)]}_{\text{1st moment}} \; dt.$$

This time, we have an explicit expression for $W(Q)$, and it does not depend on d.

The AAA algorithm [Nakatsukasa–Sète–Trefethen '18]

Idea: construct a rational approximant in the following form

Barycentric representation

$$r(z) = \frac{p(z)}{q(z)} = \frac{\displaystyle\sum_{k=1}^{K} \frac{u_k f_k}{z - z_k}}{\displaystyle\sum_{k=1}^{K} \frac{u_k}{z - z_k}}.$$

- The z_k are fake poles: clearing denominators, one sees that

$$(\deg p, \deg q) \leq (K-1, K-1) \quad \text{and} \quad \lim_{z \to z_k} r(z) = f_k.$$

- Hence $r(z)$ interpolates (z_k, f_k).
- More degrees of freedom to choose: u_k.

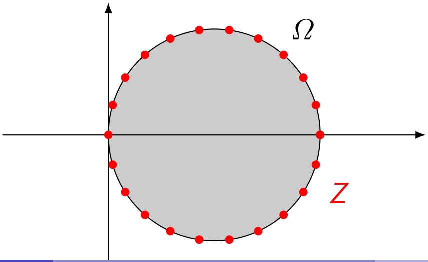
The AAA algorithm II

Inputs: f, threshold ε, a finite set of points $Z \subseteq \mathbb{C}$ where we want $r(z) \approx f(z)$.

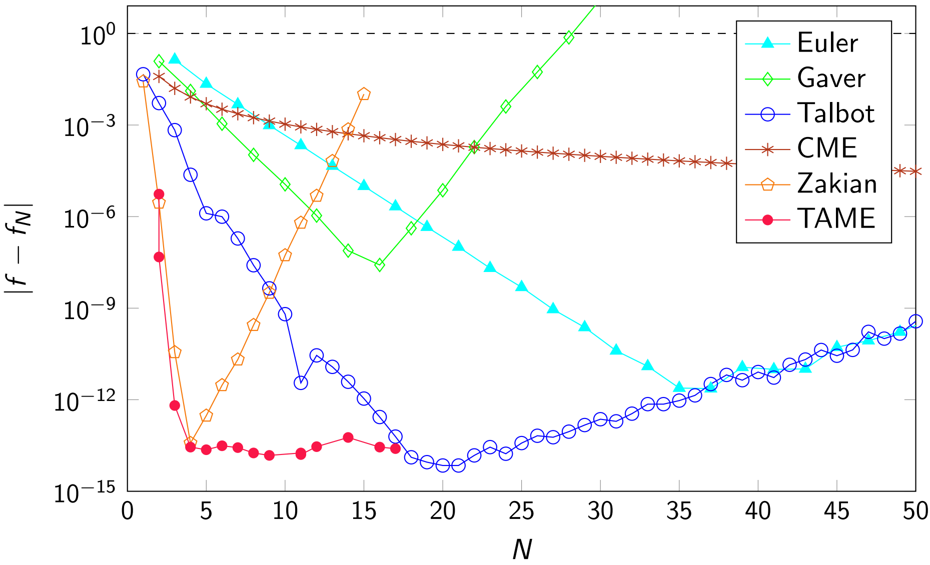
- Start with one interpolation point ($K = 1$): $z_1 \in Z, f_1 = f(z_1)$.
- Find u_k so that $r(z) \approx f(z)$ on the other points of Z, by solving the least-squares problem

$$\min \sum_{z \in Z} |f(z)q(z) - p(z)|^2$$

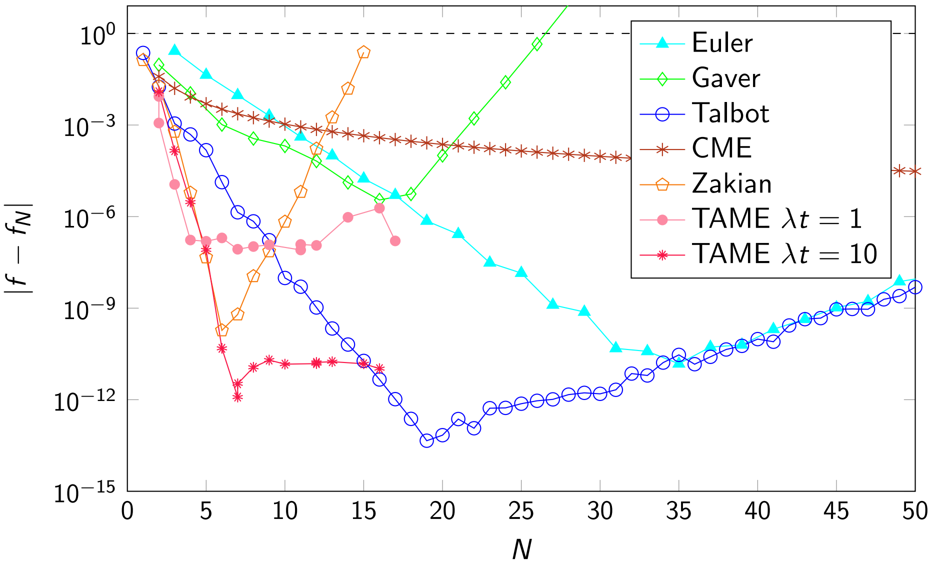
- Add one more interpolation point: $z_{K+1} = \arg\max_{z \in Z} |f(z) - r(z)|$.
- Continue until error $\leq \varepsilon$ on all points in Z.



Experiments: $\Psi(t)$ with $\lambda t = 1$ (Drift = -1.24e-01)



Experiments: $\Psi(t)$ with $\lambda t = 10$ (Drift = -3.89e-02)



Conclusions: how to TAME your transform

- Error bounds for Abate–Whitt methods.
- New method: TAME.
- Poles/weights can (and should) be adapted to domain Ω.

Questions?

Conclusions: how to TAME your transform

- Error bounds for Abate–Whitt methods.
- New method: TAME.
- Poles/weights can (and should) be adapted to domain Ω.

Questions?