Padé Approximants via Matrix Elementary Row Operations: Algorithm, Error Bounds, and Optimality

WANNATIDA YONWILAD[1,a], EKKAWIT LAMPAI[2,b]
[1]Department of Learning Management Innovation,
Faculty of Education and Educational Innovation,
Kalasin University, Kalasin 46230,
THAILAND

[2]Department of Mathematics,
Faculty of Science and Technology,
Rajabhat Maha Sarakham University, Maha Sarakham 44000,
THAILAND

[a]ORCID: 0000-0002-0028-829X
[b]ORCID: 0009-0006-8462-7776

Abstract: We develop a matrix elementary row operations (ERO) framework for computing Padé approximants and analyzing their accuracy, extending an integer augmented-matrix device to the whole Padé table. The $[m/n]$ approximant of a formal power series is obtained by row-reducing an augmented Padé–Sylvester matrix, and a determinant criterion certifies nondegeneracy. For the exponential learning function $L(t) = 1 - e^{-at}$, a transfer lemma proves that the lower triangle $m \geq n$ of the Padé table is nondegenerate, and an exact Beta-integral remainder gives a uniform error bound whose diagonal constant $K_{n,n} = 1/\binom{2n}{n}$ is independent of a. The Padé-to-Taylor error ratio has an exact quotient-of-integrals representation; it decreases strictly, stays below $K_{n,n}$, and behaves like $(2n)!/(at)^{2n}$ as $t \to \infty$. We also give a shape analysis of the Padé curve, an order-selection theorem, and a numerical stability study. For optimal time allocation, the Padé surrogate recovers the exact optimum within a certified bound.

Key-Words: Padé approximants; matrix elementary operations; continued fractions; exponential learning functions; error bounds; Sylvester matrix; Taylor polynomial; numerical stability

Received: May 11, 2026. Revised: July 13, 2026. Accepted: August 14, 2026. Published: November 10, 2026.

1 Introduction

Rational approximation offers superior accuracy over polynomial approximation for functions with poles or slow polynomial decay, [1], [2], [3]. Among rational approximations, Padé approximants — fractions $P(t)/Q(t)$ whose power series agrees with $f(t)$ to the highest possible order — have wide applications in numerical analysis and mathematical physics, [1]; for e^z on the unit circle the diagonal ones converge, up to an absolute constant, as fast as best uniform rational approximants of equal degree, [4].

The connection between Padé approximants and continued fractions is classical: the main diagonal and the first subdiagonal of the Padé table correspond to convergents of continued fraction expansions of f, [1], [2]. This motivates extending matrix methods for continued fractions to Padé computation.

Inherited from [5]. (i) The idea of encoding a rational approximation problem in a matrix and solving it by elementary row operations; (ii) the matrix product formula for convergents of a finite continued fraction, reproduced here as Theorem 2. Since the setting of [5] is $\mathbb{Z}$, with partial quotients produced by the floor function, [6], which has no counterpart over $\mathbb{R}[[t]]$, the two settings are related by *analogy* — a unit of $\mathbb{Z}$ there, a Hankel determinant here, [7] — not by specialization.

New in the present paper. (a) The extension of the ERO reduction from the diagonal to every nondegenerate entry of the Padé table, with a proof that the row reduction is solution-preserving and with the pivot form of the classical determinant criterion (Theorems 4 and 5); (b) a transfer lemma proving that the lower triangle $m \geq n$ of the Padé table of $L(t) = 1 - e^{-at}$ is nondegenerate, and exhibiting explicit degenerate entries with $m < n$ (Lemma 7); (c) fully proved error bounds for L with an a-independent diagonal constant (Lemma 8, Theorem 9); (d) an exact representation of the Padé-to-Taylor error ratio, a self-contained proof of its strict monotone decay and sharp large-t asymptotics (Theorem 10); (e) an order-selection theorem distinguishing optimality of the error *bound* from that of the leading error *constant* (Theorem 11); (f) an exact shape analysis of the Padé learning curve (Proposition 13); and (g) a numerical stability analysis of the ERO reduction (Remark 6, Table 1). None of (a)–(g) appears in [5], which treats only finite continued fractions and does not consider the exponential learning function.

Relation to the Stieltjes companion paper. For

Stieltjes series, pivot positivity characterizes the moment condition and yields two-sided certified bounds, [7]; that theory excludes $L(t) = 1 - e^{-at}$ (Section 4), for which we supply sharp one-sided bounds instead.

Connections with control theory and dynamical systems. Diagonal Padé approximants to the exponential are the standard device for approximating the delay operator $e^{-\theta s}$ by a rational transfer function, for computing the matrix exponential by scaling-and-squaring, where $[n/n]$ is the workhorse kernel and the choice of n is an optimal-order problem of the type treated in Theorem 11, [8], [9], [10], and for model order reduction by moment matching, which is precisely the Padé condition (1). The poles of the $[n/n]$ approximant to e^z lie in the open right half-plane, [11], while the denominator of Section 4 is positive on $t \geq 0$ because its coefficients are; Section 5 treats the scalar instance $\dot{L} = a(1 - L)$.

Outline. Section 2 fixes notation and Section 3 develops the ERO reduction, its determinant criterion and its stability. Section 4 specializes to $L(t) = 1 - e^{-at}$ — transfer, an exact Beta-integral remainder, a ${}_1F_1$-free Padé–Taylor comparison and order selection — and Section 5 applies these to optimal time allocation.

2 Preliminaries

2.1 Padé Approximants

Let $f(t) = \sum_{k=0}^{\infty} c_k t^k$ be a formal power series with $c_k \in \mathbb{R}$.

Definition 1. The $[m/n]$ Padé approximant of f is the rational function $R_{m,n}(t) = P_m(t)/Q_n(t)$, where P_m has degree at most m, Q_n has degree at most n with $Q_n(0) = 1$, satisfying

$$f(t) \cdot Q_n(t) - P_m(t) = O(t^{m+n+1}), \qquad (1)$$

the last relation holding coefficientwise. The entry (m, n) is called *nondegenerate* when a pair (P_m, Q_n) satisfying (1) with $Q_n(0) = 1$ exists and is unique. This is weaker than *normality* in the block sense of [1], where an entry is normal iff it occurs exactly once in the table: for L, $R_{1,2} = R_{1,3} = R_{2,2}$, so $(2, 2)$ is nondegenerate but not block-normal.

The Padé table of $L(t) = 1 - e^{-at}$ does have degenerate entries: $c_0 = 0$ forces $\det(S_{0,n}) = 0$ for every $n \geq 1$. Lemma 7 below identifies the region in which nondegeneracy holds.

2.2 Continued Fractions and the Padé Staircase

The regular C-fraction (corresponding continued fraction) of f is $f(t) = c_0/(1 - \omega_1 t/(1 - \omega_2 t/\cdots))$, where the $\omega_k \in \mathbb{R}$ are determined by requiring each convergent to agree with f to the highest possible order, [1], [2]; the C-fractions of the classical functions, e^z among them, together with their convergence regions, are tabulated in [12]. Write $C_k = A_k/B_k$ for the k-th convergent. Theorem 2, and with it formula (2) below, is the $\mathbb{R}[[t]]$ counterpart of the integer product formula of [5]; it is not used below and, since $c_0 = 0$, does not apply to L.

Theorem 2 (Classical, [1])**.** *Assume $c_0 \neq 0$ and that this fraction exists (equivalently, that the relevant Hankel determinants do not vanish, [1]). Then the convergents run along a staircase of the Padé table: $C_{2n+1} = R_{n,n}$ and $C_{2n+2} = R_{n,n+1}$. With the initial data $A_0 = 0$, $B_0 = 1$, $A_1 = c_0$, $B_1 = 1$ they satisfy the three-term recursions $A_k = A_{k-1} - \omega_{k-1} t A_{k-2}$, $B_k = B_{k-1} - \omega_{k-1} t B_{k-2}$ $(k \geq 2)$, equivalently the matrix product formula*

$$\begin{pmatrix} A_{k-1}(t) & A_k(t) \\ B_{k-1}(t) & B_k(t) \end{pmatrix} = M_1(t) \cdots M_k(t), \qquad (2)$$

where $M_1 = \left(\begin{smallmatrix} 0 & c_0 \\ 1 & 1 \end{smallmatrix}\right)$ and $M_k(t) = \left(\begin{smallmatrix} 0 & -\omega_{k-1} t \\ 1 & 1 \end{smallmatrix}\right)$ for $k \geq 2$.

2.3 Matrix Setup for Padé Computation

Definition 3 (Padé–Sylvester Matrix)**.** For series coefficients (c_k) and orders (m, n), the $n \times n$ Toeplitz matrix $S_{m,n}$ has (i, j)-entry c_{m+i-j} (with $c_k = 0$ for $k < 0$):

$$S_{m,n} = \begin{pmatrix} c_m & c_{m-1} & \cdots & c_{m-n+1} \\ c_{m+1} & c_m & \cdots & c_{m-n+2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m+n-1} & c_{m+n-2} & \cdots & c_m \end{pmatrix}. \qquad (3)$$

Theorem 4 shows that, with $S_{m,n}$ as in (3), the denominator coefficients solve $S_{m,n}\mathbf{q} = -\mathbf{c}$, where $\mathbf{q} = (q_1, \ldots, q_n)^\top$ and $\mathbf{c} = (c_{m+1}, \ldots, c_{m+n})^\top$.

3 The ERO Algorithm and Its Properties

3.1 Statement of the Reduction

Theorem 4 (ERO Algorithm)**.** *Let $f(t) = \sum_{k=0}^{\infty} c_k t^k$ and fix $m, n \geq 0$ with $\det(S_{m,n}) \neq 0$. Augment $A_{m,n} = [S_{m,n} \mid -\mathbf{c}] \in \mathbb{R}^{n \times (n+1)}$. Then:* (a) *$q_1, \ldots, q_n$ of Q_n are obtained by reducing $A_{m,n}$ to row echelon form via EROs;* (b) *each ERO preserves the solution set of $S_{m,n}\mathbf{q} = -\mathbf{c}$;* (c) *the numerator P_m is recovered by $p_k = \sum_{j=0}^{\min(k,n)} q_j c_{k-j}$, $k = 0, \ldots, m$ (with $q_0 = 1$).*

Proof. Write $Q_n(t) = 1 + \sum_{j=1}^{n} q_j t^j$ and $P_m(t) = \sum_{k=0}^{m} p_k t^k$. Expanding the product, the coefficient of t^k in $f\,Q_n - P_m$ equals $\sum_{j=0}^{\min(k,n)} q_j c_{k-j} - p_k$

for $0 \le k \le m$ and $\sum_{j=0}^{\min(k,n)} q_j c_{k-j}$ for $k > m$. Condition (1) requires all coefficients with $k \le m+n$ to vanish. The equations with $k \le m$ determine p_k and give (c). The equations with $m+1 \le k \le m+n$ read, after separating the term $j = 0$,

$$\sum_{j=1}^{n} c_{k-j}\, q_j = -c_k,\ k = m+1, \ldots, m+n, \quad (4)$$

which is exactly $S_{m,n}\mathbf{q} = -\mathbf{c}$ with the indexing of Definition 3 (put $k = m+i$, $i = 1, \ldots, n$).

For (b), every ERO is left multiplication of $[S_{m,n} \mid -\mathbf{c}]$ by an invertible elementary matrix E, so $S_{m,n}\mathbf{q} = -\mathbf{c}$ and $ES_{m,n}\mathbf{q} = -E\mathbf{c}$ have identical solution sets. Row echelon form is reached after finitely many EROs, and back substitution returns $\mathbf{q}$, unique precisely when $\det(S_{m,n}) \neq 0$, true by hypothesis; this gives (a) and (b). □

3.2 Determinant Characterization

Theorem 5 (Determinant Criterion; cf. Theorem 1.4.3 of [1])**.** *The entry (m, n) is nondegenerate in the sense of Definition 1 if and only if* $\det(S_{m,n}) \neq 0$*. If the leading principal minors of orders* $1, \ldots, n-1$ *are nonzero, Gaussian elimination without row interchanges applies and produces pivots* $d_1, \ldots, d_n$ *with* $\det(S_{m,n}) = \prod_{k=1}^{n} d_k$*. For arbitrary* $S_{m,n}$*, partial pivoting gives* $\Pi S_{m,n} = \Lambda U$ *with* Π *a permutation matrix and* $\det(S_{m,n}) = (-1)^r \prod_k U_{kk}$*,* r *the number of nontrivial interchanges.*

Proof. By the proof of Theorem 4, a normalized pair (P_m, Q_n) satisfying (1) exists and is unique if and only if (4) has a unique solution, since P_m is then determined by (c); and a square linear system has a unique solution iff its matrix is invertible.

For the factorization, elimination without interchanges produces $S_{m,n} = \Lambda U$ with Λ unit lower triangular and U upper triangular carrying $d_1, \ldots, d_n$ on its diagonal; taking determinants gives $\det(S_{m,n}) = \det\Lambda \cdot \det U = \prod_k d_k$. Each row interchange multiplies the determinant by -1, giving the signed formula. □

3.3 Complexity and Numerical Stability

Remark 6 (Cost and stability of the ERO reduction)*.* Reducing $A_{m,n}$ costs $\frac{2}{3}n^3 + O(n^2)$ flops, versus $O(n^2)$ for fast Toeplitz solvers of Levinson–Trench–Zohar type. The ERO reduction trades that asymptotic advantage, immaterial for $n \le 8$, for the backward stability of Gaussian elimination with partial pivoting (GEPP): the computed $\hat{\mathbf{q}}$ satisfies $(S_{m,n} + \Delta S)\hat{\mathbf{q}} = -\mathbf{c}$ with $\|\Delta S\|_\infty \le \beta_n \Gamma_n \varepsilon \|S_{m,n}\|_\infty$, ε the unit roundoff, Γ_n the growth factor and β_n a low-degree polynomial in n, [13]; Levinson-type solvers are proved weakly stable only in the symmetric positive definite case, [14], which $S_{m,n}$ is not, and can lose accuracy when leading principal minors are small — precisely the near-degenerate entries detected by Theorem 5, since the leading $k \times k$ principal submatrix of $S_{m,n}$ is $S_{m,k}$. The SVD-based reduction of [15] lowers the degrees when the matrix is numerically rank deficient, and barycentric schemes such as [16] avoid the Padé–Sylvester system altogether but do not target a prescribed order (m, n); we keep (m, n) fixed.

Table 1 quantifies this for $L(t) = 1 - e^{-0.35t}$: $\mathrm{cond}_2(S_{n,n})$ grows explosively ($\approx 1.2 \cdot 10^{22}$ at $n = 8$), yet the relative error of the GEPP solution against the exact rational solution stays below $2.1 \cdot 10^{-10}$. The ill-conditioning is almost entirely an artifact of scaling, since $c_k = (-1)^{k+1}a^k/k!$ ($k \ge 1$) decays factorially, and after equilibration $S \mapsto D_1 S D_2$ the condition number drops by 13 orders of magnitude at $n = 8$. What predicts the observed accuracy is the row-scaling-invariant Skeel condition number $\big\||S^{-1}||S||\mathbf{q}|\big\|_\infty / \|\mathbf{q}\|_\infty = 8.4 \cdot 10^6$ at $n = 8$, [13, Sec. 7.2]; as no significant element growth occurs, $3n\varepsilon\,\mathrm{cond}$ majorizes the last column of Table 1 at every n. We therefore recommend reducing the equilibrated matrix with partial pivoting and monitoring a condition estimate of it; the pivot product $\prod_k |d_k|$ is *not* scale-free, and at $a = 0.35$ the singular entry $(4, 5)$ returns a computed $3.2 \cdot 10^{-34}$, larger than the nonsingular $(6, 5)$ and $(7, 5)$.

Table 1. Conditioning of $S_{n,n}$ for $L(t) = 1 - e^{-0.35t}$ and accuracy of the ERO (GEPP) solution. D_1 scales rows, then D_2 columns, to unit ∞-norm; rel. error is $\|\hat{\mathbf{q}} - \mathbf{q}\|_2 / \|\mathbf{q}\|_2$ against the exact rational $\mathbf{q}$; condition numbers in extended precision. For $a < 1$ GEPP makes no interchanges, so $\Gamma_n = 1$; entries at the 10^{-16} level sit at the unit-roundoff floor and are implementation-dependent.

n	$\mathrm{cond}_2(S_{n,n})$	$\mathrm{cond}_2(D_1 S_{n,n} D_2)$	rel. error
1	$1.00 \cdot 10^0$	$1.00 \cdot 10^0$	$1.6 \cdot 10^{-16}$
2	$1.04 \cdot 10^2$	$1.02 \cdot 10^1$	$< 10^{-16}$
3	$5.10 \cdot 10^4$	$1.61 \cdot 10^2$	$1.3 \cdot 10^{-15}$
4	$5.83 \cdot 10^7$	$3.05 \cdot 10^3$	$9.9 \cdot 10^{-16}$
5	$1.20 \cdot 10^{11}$	$6.50 \cdot 10^4$	$2.9 \cdot 10^{-13}$
6	$3.87 \cdot 10^{14}$	$1.50 \cdot 10^6$	$1.2 \cdot 10^{-12}$
7	$1.81 \cdot 10^{18}$	$3.66 \cdot 10^7$	$2.4 \cdot 10^{-11}$
8	$1.15 \cdot 10^{22}$	$9.33 \cdot 10^8$	$2.1 \cdot 10^{-10}$

Source: created by the authors.

4 Error Bounds for the Exponential Learning Function

The exponential learning function $L(t) = 1 - e^{-at}$, $a > 0$, solves $\dot{L} = a(1 - L)$, $L(0) = 0$, and models skill acquisition under constant instructional intensity; the empirical case for an exponential rather than a power law of practice is made in [17]. Its Taylor coefficients are $c_0 = 0$ and $c_k = (-1)^{k+1}a^k/k!$ for $k \geq 1$. The bounds below are one-sided: with $\mu_k = (-1)^k c_k$ one has $\mu_1 = -a < 0$, so L is not a Stieltjes series and the two-sided bracketing of that class is unavailable here, [7, Remark 4.3].

4.1 Transfer and Nondegeneracy

Throughout, write $E_n(s) = \sum_{j=0}^{n} \frac{(2n-j)!\,n!}{(2n)!\,j!\,(n-j)!} s^j$, so that the $[n/n]$ Padé approximant of e^z is $E_n(z)/E_n(-z)$, and, more generally, let $P_m^{\exp}$ and $Q_n^{\exp}$ denote the numerator and denominator of the $[m/n]$ Padé approximant of e^z, [1]; both depend on (m, n), the displayed index recording only the degree.

Lemma 7 (Transfer and nondegeneracy)**.** *Let $m \geq n \geq 0$ and $L(t) = 1 - e^{-at}$. Then:* (a) *the $[m/n]$ Padé approximant of L exists, is unique, and equals*

$$R_{m,n}(t) = 1 - \frac{P_m^{\exp}(-at)}{Q_n^{\exp}(-at)} = \frac{Q_n^{\exp}(-at) - P_m^{\exp}(-at)}{Q_n^{\exp}(-at)}; \tag{5}$$

(b) $\det(S_{m,n}) \neq 0$*, i.e. the lower triangle of the Padé table of L is nondegenerate;* (c) $L(t) - R_{m,n}(t) = -\big(e^{-at} - P_m^{\exp}(-at)/Q_n^{\exp}(-at)\big)$*, so the two remainders coincide in absolute value. For $m = n$, part (a) is the case $f \mapsto 1 - f$ of the classical invariance under homographic transformations of the function value, Theorem 1.5.3 of [1], stated there for the diagonal only; the proof below extends the affine case to the whole triangle $m \geq n$. The hypothesis $m \geq n$ cannot be dropped: for $m < n$ the entry may be degenerate, e.g.* $\det(S_{0,n}) = 0$ *for all $n \geq 1$ (because $c_0 = 0$) and* $\det(S_{2,3}) = \det(S_{2,5}) = \det(S_{4,5}) = 0$.

Proof. Put $g(t) = e^{-at}$, with power series coefficients $\gamma_k = (-a)^k/k!$, and let (P_m, Q_n) be any pair with $\deg P_m \leq m$, $\deg Q_n \leq n$, $Q_n(0) = 1$. Since $L = 1 - g$, putting $\tilde{P}_m := Q_n - P_m$ we have

$$L\,Q_n - P_m = -\big(g\,Q_n - \tilde{P}_m\big). \tag{6}$$

Because $m \geq n$ we have $\deg(Q_n - P_m) \leq \max(n, m) = m$ and, symmetrically, $\deg(Q_n - \tilde{P}_m) \leq m$; hence $P_m \mapsto \tilde{P}_m = Q_n - P_m$ is an involutive bijection of $\{\deg \leq m\}$ onto itself. By (6), (P_m, Q_n) satisfies the Padé condition (1) for L if and only if $(\tilde{P}_m, Q_n)$ satisfies it for g. The Padé table of the exponential is normal, hence nondegenerate, at every entry, [1, Sec. 1.2], so the g-problem has exactly one normalized solution; therefore so does the L-problem, which proves (a) and, via Theorem 5, (b). Assertion (c) is (6) divided by Q_n. The stated counterexamples are verified by direct evaluation of the Toeplitz determinants; e.g. $S_{0,n}$ has (i, j)-entry c_{i-j}, hence is lower triangular with every diagonal entry equal to $c_0 = 0$, so $\det(S_{0,n}) = c_0^n = 0$. □

For $m = n$, formula (5) gives at once

$$R_{1,1}(t) = \frac{at}{1 + \frac{a}{2}t}, \qquad R_{2,2}(t) = \frac{at}{1 + \frac{a}{2}t + \frac{a^2}{12}t^2}. \tag{7}$$

Those in (7) agree with the ERO reduction of $A_{1,1}$ and $A_{2,2}$.

4.2 An Exact Remainder and the Error Constant

Lemma 8 (Hermite's classical remainder; proof included for completeness)**.** *Let $m, n \geq 0$ and let $P_m^{\exp}/Q_n^{\exp}$ be the $[m/n]$ Padé approximant of e^z. Then for every z with $Q_n^{\exp}(z) \neq 0$,*

$$e^z - \frac{P_m^{\exp}(z)}{Q_n^{\exp}(z)} = \frac{(-1)^n z^{m+n+1}}{(m+n)!\,Q_n^{\exp}(z)} \times \int_0^1 e^{zu} u^n (1-u)^m du. \tag{8}$$

Consequently, with $s = at \geq 0$ and $z = -s$,

$$\left| e^{-s} - \frac{P_m^{\exp}(-s)}{Q_n^{\exp}(-s)} \right| \leq \frac{m!\,n!}{(m+n)!\,(m+n+1)!} \cdot \frac{s^{m+n+1}}{Q_n^{\exp}(-s)}, \tag{9}$$

and on the diagonal $m = n$ the constant is

$$\frac{(n!)^2}{(2n)!\,(2n+1)!} = \frac{K_{n,n}}{(2n+1)!}, \tag{10}$$

where $K_{n,n} := (n!)^2/(2n)! = 1/\binom{2n}{n}$ is independent of a. Numerically $K_{1,1} = 1/2$, $K_{2,2} = 1/6$, $K_{3,3} = 1/20$, $K_{4,4} = 1/70$, and by Stirling's formula $K_{n,n} \sim \sqrt{\pi n}\,4^{-n} \to 0$.

Proof. Put $w(u) = u^n(1-u)^m$, a polynomial of degree $m + n$, and

$$\Psi(z) := \frac{(-1)^n z^{m+n+1}}{(m+n)!} \int_0^1 e^{zu} w(u)\,du. \tag{11}$$

Repeated integration by parts gives, for any polynomial w of degree $m + n$ and any $z \neq 0$,

$$\int_0^1 e^{zu} w(u) du = \sum_{i=0}^{m+n} \frac{(-1)^i}{z^{i+1}} \times \big[e^z w^{(i)}(1) - w^{(i)}(0)\big]. \tag{12}$$

Because $w = u^n(1-u)^m$ vanishes to order m at $u = 1$ and to order n at $u = 0$, we have $w^{(i)}(1) = 0$ for $i < m$ and $w^{(i)}(0) = 0$ for $i < n$. Multiplying (12) by $(-1)^n z^{m+n+1}/(m+n)!$ therefore yields

$$\Psi(z) = \widetilde{Q}(z)\, e^z - \widetilde{P}(z), \tag{13}$$

where

$$\widetilde{Q}(z) = \frac{(-1)^n}{(m+n)!}\sum_{i=m}^{m+n}(-1)^i w^{(i)}(1) z^{m+n-i}, \tag{14}$$

$$\widetilde{P}(z) = \frac{(-1)^n}{(m+n)!}\sum_{i=n}^{m+n}(-1)^i w^{(i)}(0) z^{m+n-i}. \tag{15}$$

By (14) and (15), $\deg \widetilde{Q} \le n$ and $\deg \widetilde{P} \le m$; both formulas extend to $z = 0$ by continuity since Ψ in (11) is entire. The constant term of $\widetilde{Q}$ is the $i = m+n$ term, and $w^{(m+n)}(1) = (m+n)!\,(-1)^m$ because the leading coefficient of w is $(-1)^m$; hence $\widetilde{Q}(0) = (-1)^n(-1)^{m+n}(-1)^m = 1$. Finally, $\Psi(z) = O(z^{m+n+1})$ by its definition, so $(\widetilde{P}, \widetilde{Q})$ satisfies the Padé condition (1) for $f = e^z$ at order (m, n) with $\widetilde{Q}(0) = 1$. The Padé table of the exponential is normal, hence nondegenerate, [1, Sec. 1.2], so such a pair is unique: $\widetilde{Q} = Q_n^{\exp}$ and $\widetilde{P} = P_m^{\exp}$. Dividing (13) by $Q_n^{\exp}(z)$ gives (8).

For (9), set $z = -s$ with $s \ge 0$. By the explicit formula for $Q_n^{\exp}$, [1], every coefficient of $s \mapsto Q_n^{\exp}(-s)$ is positive and its constant term is 1, so $Q_n^{\exp}(-s) \ge 1$ and in particular it does not vanish. Moreover $0 < e^{-su} \le 1$ on $[0, 1]$, whence

$$\begin{aligned}\Big|\int_0^1 e^{-su}u^n(1-u)^m du\Big| &\le \int_0^1 u^n(1-u)^m du \\ &= \frac{m!\,n!}{(m+n+1)!},\end{aligned} \tag{16}$$

which combined with (8) gives (9). Setting $m = n$ gives (10), and $(n!)^2/(2n)! = 1/\binom{2n}{n}$ is immediate. Stirling's formula gives $\binom{2n}{n} \sim 4^n/\sqrt{\pi n}$. □

Remark (Relation to [18]). In [18], the error of the (m, n) Padé approximant of e^z and of its derivatives is bounded on discs about the origin, and convergence is shown along diagonal, row and column paths. Theorem 9 instead retains Q_n explicitly on the real half-line, and we are not aware of a counterpart of the ratio identity of Theorem 10 in that setting.

Remark (Relation to the Kummer function). Dividing (8) by $B(n+1, m+1)$ turns the integral into Kummer's ${}_1F_1(n+1; m+n+2; -s)$, whose modulus is at most 1 for $s \ge 0$ because the integrand is dominated by $u^n(1-u)^m$; inequality (16) is the elementary form of that estimate, and the first parameter is $n+1$, not 1.

Theorem 9 (Padé Error Bound). *Let $m \ge n \ge 0$ and let $R_{m,n}$ be the $[m/n]$ Padé approximant of $L(t) = 1 - e^{-at}$, $a > 0$, with denominator $Q_n(t) = Q_n^{\exp}(-at)$, a polynomial in t with nonnegative coefficients; in particular $Q_n(t) \ge 1 > 0$ for $t \ge 0$ and $R_{m,n}$ has no pole on $[0, \infty)$. Then for all $t \ge 0$*

$$|L(t) - R_{m,n}(t)| \le \frac{K_{m,n}}{Q_n(t)} \cdot \frac{(at)^{m+n+1}}{(m+n+1)!}, \tag{17}$$

where $K_{m,n} = m!\,n!/(m+n)!$, so that $K_{n,n} = (n!)^2/(2n)!$ on the diagonal. Moreover Q_n is nondecreasing on $[0, \infty)$, so $\inf_{[0,T]} Q_n = 1$ and

$$\sup_{[0,T]} |L - R_{m,n}| \le \frac{K_{m,n}\,(aT)^{m+n+1}}{(m+n+1)!}. \tag{18}$$

Proof. By Lemma 7(c) the remainder for L equals in modulus the remainder for e^{-at}, and $Q_n(t) = Q_n^{\exp}(-at)$ has positive coefficients by the explicit formula $Q_n^{\exp}(z) = \sum_{j=0}^{n} \frac{(m+n-j)!\,n!}{(m+n)!\,j!\,(n-j)!}(-z)^j$, so at $z = -at \le 0$ every term is ≥ 0 and the constant term is 1. Now apply Lemma 8, inequality (9), with $s = at$; regrouping $m!\,n!/[(m+n)!\,(m+n+1)!] = K_{m,n}/(m+n+1)!$ gives (17). Since Q_n has nonnegative coefficients and $Q_n(0) = 1$, it is nondecreasing on $[0, \infty)$ with minimum 1; substituting $t \le T$ and $Q_n \ge 1$ into (17) gives (18). □

4.3 Comparison: Padé vs. Taylor Approximation

Since $e^{-a\xi} \le 1$ for $\xi \ge 0$, the Lagrange remainder of e^{-at} gives the Taylor bound $(aT)^{N+1}/(N+1)!$ at degree N for L on $[0, T]$. The theorem below quantifies what the diagonal Padé approximant gains over it, without recourse to a mean-value parameter ξ or to unproved hypergeometric bounds. For the *matrix* exponential, where cost is counted in matrix products, polynomial schemes can reach a higher order than rational ones at equal cost, [19]; Theorem 10 is the scalar, equal-degree case.

Theorem 10 (Quantitative Padé superiority over Taylor). *Let $L(t) = 1 - e^{-at}$ with $a > 0$ and $n \ge 1$, let $\mathcal{T}_{2n}$ be the degree-$2n$ Taylor polynomial of L at 0, let $R_{n,n}$ be the diagonal Padé approximant, and put $s = at$. Define*

$$\begin{aligned} I_1(s) &= \int_0^1 e^{-su}\big(u(1-u)\big)^n du, \\ I_2(s) &= \int_0^1 e^{-su}(1-u)^{2n} du. \end{aligned} \tag{19}$$

With I_1, I_2 as in (19)*, for every $t > 0$ the error ratio is* exactly

$$\rho_n(t) := \frac{|L(t) - R_{n,n}(t)|}{|L(t) - \mathcal{T}_{2n}(t)|} = \frac{1}{E_n(s)} \cdot \frac{I_1(s)}{I_2(s)}, \tag{20}$$

and: (a) $s \mapsto I_1(s)/I_2(s)$ *is strictly decreasing on* $[0,\infty)$*, with value* $K_{n,n}$ *at* $s = 0$*;* (b) *consequently, for all* $t > 0$*,*

$$\rho_n(t) < \frac{K_{n,n}}{E_n(at)} \le K_{n,n} = \frac{1}{\binom{2n}{n}} < 1, \qquad (21)$$

so $|L - R_{n,n}| < |L - \mathcal{T}_{2n}|$ *on* $(0,\infty)$*;* (c) ρ_n *is strictly decreasing,* $\rho_n(0^+) = K_{n,n}$*, and*

$$\rho_n(t) \sim \frac{(2n)!}{(at)^{2n}} \qquad (t \to \infty), \qquad (22)$$

so in particular $\rho_n(t) \to 0$*.*

Proof. Exact ratio. By Lemma 7(c) and Lemma 8 with $m = n$, $z = -s$,

$$|L - R_{n,n}| = \frac{s^{2n+1}}{(2n)!\, E_n(s)}\, I_1(s). \qquad (23)$$

For the Taylor error, the integral form of the remainder, $e^z - \sum_{k=0}^{N} z^k/k! = \frac{z^{N+1}}{N!}\int_0^1 e^{zu}(1-u)^N du$, applied with $N = 2n$, $z = -s$, and combined with $L - \mathcal{T}_{2n} = -\big(e^{-s} - \sum_{k=0}^{2n}(-s)^k/k!\big)$, gives

$$|L - \mathcal{T}_{2n}| = \frac{s^{2n+1}}{(2n)!}\, I_2(s). \qquad (24)$$

Dividing (23) by (24) yields (20); note that $I_2(s) > 0$ so the ratio is well defined for $t > 0$.

(a) Monotonicity. Fix $s \ge 0$ and let μ_s be the probability measure on $[0,1]$ with density proportional to $e^{-su}(1-u)^{2n}$, so that $I_2(s)$ is its normalizing constant. Put $g(u) = u^n(1-u)^n/(1-u)^{2n} = \big(u/(1-u)\big)^n$, which is strictly increasing on $[0,1)$. Then $I_1(s)/I_2(s) = \mathbb{E}_{\mu_s}[g]$ and, differentiating under the integral sign,

$$\begin{aligned}\frac{d}{ds}\frac{I_1}{I_2} &= \frac{I_1' I_2 - I_1 I_2'}{I_2^2} \\ &= -\big(\mathbb{E}_{\mu_s}[ug] - \mathbb{E}_{\mu_s}[u]\,\mathbb{E}_{\mu_s}[g]\big) \\ &= -\operatorname{Cov}_{\mu_s}\big(u, g(u)\big) < 0, \end{aligned} \qquad (25)$$

the covariance in (25) being strictly positive because $u \mapsto g(u)$ is strictly increasing, so that Chebyshev's association (FKG) inequality applies, [20]. At $s = 0$, $I_1(0) = B(n+1,n+1) = (n!)^2/(2n+1)!$ and $I_2(0) = 1/(2n+1)$, so $I_1(0)/I_2(0) = (n!)^2/(2n)! = K_{n,n}$.

(b) For $t > 0$ we have $s > 0$, hence by (a) $I_1(s)/I_2(s) < K_{n,n}$, while $E_n(s) \ge E_n(0) = 1$ since all coefficients of E_n are positive. Inserting both into (20) gives (21). Finally $\binom{2n}{n} \ge 2$ for $n \ge 1$, so $K_{n,n} \le 1/2 < 1$.

(c) Strict decrease of ρ_n follows from (a) together with the strict increase of E_n on $(0,\infty)$. For the asymptotics, Watson's lemma, [21], applied to I_1 and I_2 gives, as $s \to \infty$,

$$\begin{aligned} I_1(s) &= \frac{n!}{s^{n+1}}\big(1 + O(s^{-1})\big), \\ I_2(s) &= \frac{1}{s}\big(1 + O(s^{-1})\big), \end{aligned} \qquad (26)$$

since near $u = 0$ the integrands in (19) behave like u^n and 1, respectively. The leading coefficient of E_n is $n!/(2n)!$, so $E_n(s) = \frac{n!}{(2n)!}s^n(1 + O(s^{-1}))$. Substituting (26) into (20) gives

$$\begin{aligned}\rho_n(s) &= \frac{(2n)!}{n!\, s^n} \cdot \frac{n!}{s^n}\big(1 + O(s^{-1})\big) \\ &= \frac{(2n)!}{s^{2n}}\big(1 + O(s^{-1})\big), \end{aligned} \qquad (27)$$

so (27) is (22). □

Remark. Formula (22) is sharp. For $n = 2$ the computed ratio $\rho_n(s)s^{2n}/(2n)!$ equals 0.9227, 0.9920, 0.99920 at $s = 10^2, 10^3, 10^4$; for $n = 4$ it equals 0.7246, 0.9685, 0.99680 at the same points.

4.4 Optimal Approximation Order

Minimizing an error *bound* is not the same as minimizing the error itself. We therefore state both, and prove each.

Theorem 11 (Order selection)**.** *Fix* $a > 0$*, a horizon* $T > 0$*, a tolerance* $\delta > 0$*, and the work model* $W(m,n) = m + n =: N$*.* (a) (Bound-optimal diagonal order.) *On the diagonal the bound* (18) *is strictly decreasing in* n *once* $n > aT$*, so*

$$p^* = \min\Big\{p \in \mathbb{N} : \frac{K_{p,p}(aT)^{2p+1}}{(2p+1)!} \le \delta\Big\}. \qquad (28)$$

The order p^* *in* (28) *is well defined and is the least diagonal order whose* bound *meets the tolerance; every order above it does so too once* $p^* \ge aT$*. This is a statement about the bound, not a claim that* p^* *minimizes the true error at fixed work.* (b) (Genuine optimality of the leading constant.) *For fixed* $N = m + n$ *with* $m \ge n$*, write* $\kappa_{m,n} = m!\,n!/[(m+n)!\,(m+n+1)!]$*. Then*

$$L - R_{m,n} = (-1)^m \kappa_{m,n}(at)^{N+1}\big(1 + O(t)\big), \qquad (29)$$

so $\kappa_{m,n}$ *is the exact leading error constant, and*

$$\kappa_{m,n} = \frac{1}{\binom{N}{n}(N+1)!} \qquad (30)$$

is minimized precisely at $n = \lfloor N/2 \rfloor$*: for even* N *at* $m = n = N/2$*, and for odd* N *at* $m = (N+1)/2$*,* $n = (N-1)/2$ *(the transposed pair violates* $m \ge n$*, where* (29) *need not hold). Hence among the entries of total degree* N with $m \ge n$ *the (near-)diagonal one minimizes the true error asymptotically as* $t \to 0^+$*, not merely its bound. Outside that triangle this fails:* $R_{1,2} = R_{2,2}$*, so* $L - R_{1,2} = O(t^5)$ *beats* $L - R_{2,1} = O(t^4)$*.*

Proof. (a) Write $b_p = K_{p,p}(aT)^{2p+1}/(2p+1)!$. Then $b_{p+1}/b_p = \frac{K_{p+1,p+1}}{K_{p,p}} \frac{(aT)^2}{(2p+2)(2p+3)} = \frac{(p+1)^2(aT)^2}{(2p+1)(2p+2)^2(2p+3)} < 1$ for $p \geq aT$, and $b_p \to 0$ superexponentially, so the defining set is nonempty and has a least element.

(b) By Lemma 7(c) and (8) with $z = -at$,

$$\mathcal{J} := \int_0^1 e^{-atu} u^n (1-u)^m du,$$

$$L - R_{m,n} = -(-1)^n \frac{(-at)^{N+1}}{N!\, Q_n(t)} \mathcal{J} = (-1)^m \frac{(at)^{N+1}}{N!\, Q_n(t)} \mathcal{J}, \tag{31}$$

using $-(-1)^n(-1)^{N+1} = (-1)^m$ in (31), since $N = m + n$. As $t \to 0^+$, $Q_n(t) \to 1$ and the integral tends to $B(n+1, m+1) = m!\, n!/(N+1)!$, which gives (29) with $\kappa_{m,n} = m!\, n!/[N!\,(N+1)!]$. Identity (30) follows from $m!\, n!/N! = 1/\binom{N}{n}$. Since $\binom{N}{n}$ is maximized over $0 \leq n \leq N$ exactly at $n = \lfloor N/2 \rfloor$ and $n = \lceil N/2 \rceil$, and these coincide for even N, the constant $\kappa_{m,n}$ is minimized as claimed. □

Corollary 12 (Practical order selection)**.** *For a prescribed tolerance δ on $[0, T]$, the diagonal Padé order n suffices whenever*

$$\frac{(aT)^{2n+1}}{(2n+1)!} K_{n,n} \leq \delta. \tag{32}$$

That (32) *suffices follows from* (18)*. Let N^* be the least Taylor degree attaining the same uniform accuracy as the diagonal Padé approximant of total degree $2n$. For $a = 1$ and $n = 1, 2, 3, 4$ computation gives $N^* = 3, 6, 8, 11$ ($T = 1$); $5, 7, 10, 12$ ($T = 2$); $7, 9, 11, 13$ ($T = 3$); $11, 13, 15, 17$ ($T = 5$). The degree-reduction factor $N^*/(2n)$ is never below $4/3$ here and increases with aT for each fixed n (for $n = 1$: 1.5, 2.5, 3.5, 5.5), consistent with Theorem 10(c). ($N^* = 3$ at $T = 1$, $n = 1$ is a tie: both suprema are attained at $t = 1$, where $\mathcal{T}_3(1) = R_{1,1}(1) = 2/3$ exactly.)*

5 Application to Exponential Learning Functions

5.1 Shape of the Padé Learning Curve

Proposition 13 (Monotonicity and concavity)**.** *Let $s = at$, E_n as above, and define the even polynomial*

$$V_n(s) = E_n'(-s)E_n(s) + E_n(-s)E_n'(s). \tag{33}$$

The polynomial V_n of (33) *has degree $2n-2$, with leading coefficient $2(-1)^{n-1} n\, n!\, (n+1)!/[(2n)!]^2$. Then $R_{n,n}(0) = 0$ and*

$$R_{n,n}'(t) = a \frac{V_n(s)}{E_n(s)^2}, \qquad V_n(0) = 1. \tag{34}$$

Consequently $R_{n,n}$ is strictly increasing on $[0, \sigma_n/a]$, $\sigma_n \in (0, \infty]$ being the least positive zero of V_n. Exact computation gives $V_1 \equiv 1$ and $V_3(s) = (s^4 - 30s^2 + 600)/600 > 0$ for all real s (discriminant in s^2 equal to $-1500 < 0$), so $\sigma_1 = \sigma_3 = \infty$; likewise $\sigma_5 = \sigma_7 = \infty$, while $V_2(s) = (12 - s^2)/12$ gives $\sigma_2 = 2\sqrt{3}$, and $\sigma_4 = 6.1012$, $\sigma_6 = 8.7463$, $\sigma_8 = 11.3940$. Hence for every $1 \leq n \leq 8$,

$$\sigma_n \geq \sigma_2 = 2\sqrt{3} \approx 3.4641, \tag{35}$$

with equality exactly at $n = 2$. Similarly $R_{n,n}'' < 0$ on $[0, \upsilon_n/a)$, where υ_n is the least positive zero of the numerator of $R_{n,n}''$; exact computation gives $\upsilon_1 = \infty$ and $\upsilon_2, \ldots, \upsilon_8 = 6.8229$, 5.6101, 20.0243, 7.8371, 41.9971, 10.3752, 71.9940, so $\upsilon_n \geq \upsilon_3 = 5.6101 > 2\sqrt{3}$ for $1 \leq n \leq 8$; the two thresholds differ, $\sigma_3 = \infty$ but $\upsilon_3 = 5.6101$. In particular, for every $1 \leq n \leq 8$, $R_{n,n}$ is strictly increasing and strictly concave on $[0, 2\sqrt{3}/a]$.

Proof. By Lemma 7(a), $R_{n,n}(t) = 1 - E_n(-s)/E_n(s)$ with $s = at$; $R_{n,n}(0) = 1 - 1 = 0$. Differentiating in t and using $\frac{d}{ds} E_n(-s) = -E_n'(-s)$,

$$R_{n,n}'(t) = a \frac{E_n'(-s)E_n(s) + E_n(-s)E_n'(s)}{E_n(s)^2} = a \frac{V_n(s)}{E_n(s)^2}, \tag{36}$$

so that (36) is (34); $V_n(0) = 2E_n'(0)E_n(0) = 2 \cdot \frac{1}{2} \cdot 1 = 1$ since the coefficient of s in E_n is $n/(2n) = 1/2$. As $E_n(s) > 0$ for $s \geq 0$, the sign of $R_{n,n}'$ is that of V_n, so $R_{n,n}$ increases up to the first sign change of V_n, which is σ_n (σ_n is a simple root for $1 \leq n \leq 8$); the same argument applied to $R_{n,n}''$, whose denominator $E_n(s)^3$ is positive on $s \geq 0$, gives the statement about υ_n. The listed values of V_n, σ_n, and υ_n are obtained by exact symbolic computation of the polynomials V_n and of the numerator of $R_{n,n}''$ for $n = 1, \ldots, 8$, followed by certified root isolation; the assertions about V_1, V_2, V_3 are verified by inspection. The degree and leading coefficient hold for every n: writing e_j for the coefficient of s^j in E_n, the s^{2n-1} terms of (33) cancel and its s^{2n-2} terms sum to $2(-1)^{n-1} e_n e_{n-1}$. □

Remark. The threshold in (35) is necessary: for even n the leading coefficient of V_n is negative while $V_n(0) = 1$, so V_n has a positive zero and monotonicity of $R_{n,n}$ on all of $[0, \infty)$ fails — already $R_{2,2}$ attains its maximum at $s = 2\sqrt{3}$ — though it holds for $n = 1, 3, 5, 7$. Numerically σ_{2k} increases in k (3.46, 6.10, 8.75, 11.39, 14.04, 16.69 for $2k \leq 12$), consistent with $\sigma_n \geq 2\sqrt{2n-1}$.

5.2 The Optimal Learning Sequencing Problem

Consider the mastery-constrained minimum-time problem, the Optimal Learning Sequencing Problem (OLSP): given content items $C_1, \dots, C_J$ with learning functions $L_i(t) = 1 - e^{-a_i t}$ and target mastery $\tau \in (0,1)$, minimize $\sum_{i=1}^{J} T_i$ subject to $L_i(T_i) \geq \tau$. Since $L_i(T_i) \geq \tau \iff T_i \geq -\ln(1-\tau)/a_i$, the exact optimum is $T_i^* = -\ln(1-\tau)/a_i$; we ask what happens when L_i is replaced by the Padé model $R_{n,n}^{(i)}$ on the box $0 \leq T_i \leq 2\sqrt{3}/a_i$. Two hypotheses are needed, and both are explicit. By Proposition 13, for $1 \leq n \leq 8$ each $R_{n,n}^{(i)}$ is continuous, strictly increasing and strictly concave on $[0, 2\sqrt{3}/a_i]$ with $R_{n,n}^{(i)}(0) = 0$; and the surrogate reaches the target there provided

$$\tau \leq \tau_{\max} := \min_{1 \leq n \leq 8} R_{n,n}(2\sqrt{3}/a) \approx 0.9282, \quad (37)$$

the minimum being attained at $n = 2$, where $R_{2,2}(2\sqrt{3}/a) = 0.928203$; $\tau_{\max}$ lies below the value $1 - e^{-2\sqrt{3}} \approx 0.9687$ of L there, so the exact-model threshold would not suffice. Under (37) the superlevel set $\{T_i \in [0, 2\sqrt{3}/a_i] : R_{n,n}^{(i)}(T_i) \geq \tau\}$ equals $[\hat{T}_i^*, 2\sqrt{3}/a_i]$, a nonempty closed interval, so the surrogate feasible set is convex and $\hat{T}_i^*$ is the unique root of $R_{n,n}^{(i)}(T) = \tau$ there. The surrogate OLSP is therefore a convex program separable across i, solved by $T_i = \hat{T}_i^*$. Below $\tau = 0.85$, comfortably inside (37).

Proposition 14 (Time allocation error)**.** *Let $\tau \in (0,1)$, let $T_i^* = -\ln(1-\tau)/a_i$ be the exact optimal time above, and let $\hat{T}_i^*$ be the surrogate optimum. Let δ be any uniform bound for $|L_i - R_{n,n}^{(i)}|$ on an interval containing both T_i^* and $\hat{T}_i^*$ (only $|L_i(\hat{T}_i^*) - R_{n,n}^{(i)}(\hat{T}_i^*)| \leq \delta$ is used). If $\delta < (1-\tau)/2$, then*

$$|T_i^* - \hat{T}_i^*| \leq \frac{2\delta}{a_i(1-\tau)}. \quad (38)$$

Proof. Drop the index i. By construction $R_{n,n}(\hat{T}^*) = \tau = L(T^*)$, so $|L(\hat{T}^*) - L(T^*)| = |L(\hat{T}^*) - R_{n,n}(\hat{T}^*)| \leq \delta$. Hence $|L(\hat{T}^*) - \tau| \leq \delta$, so L takes values in $[\tau - \delta, \tau + \delta]$ between T^* and $\hat{T}^*$, and by the mean value theorem there is ξ strictly between them with

$$\delta \geq |L(\hat{T}^*) - L(T^*)| = L'(\xi)\,|\hat{T}^* - T^*|. \quad (39)$$

In (39), $L'(t) = ae^{-at} = a\big(1 - L(t)\big)$ and $L(\xi) \leq \tau + \delta$, so $L'(\xi) \geq a(1 - \tau - \delta)$. The hypothesis $\delta < (1-\tau)/2$ gives $1-\tau-\delta > (1-\tau)/2 > 0$, whence $|\hat{T}^* - T^*| \leq \delta/[a(1-\tau-\delta)] \leq 2\delta/[a(1-\tau)]$. □

5.3 Padé Table for $L(t) = 1 - e^{-0.35t}$

Table 2 lists the first entries of the Padé table for $L_4(t) = 1 - e^{-0.35t}$ (the rate $a_4 = 0.35$ of the four-topic example below), computed via Theorem 4 and checked against Lemma 7(a).

Table 2. Selected Padé approximants $R_{m,n}(t)$ for $L(t) = 1 - e^{-0.35t}$, computed via Theorem 4.

$[m/n]$	$P_m(t)/Q_n(t)$
$[1/0]$	$0.35t$
$[1/1]$	$0.35t \,/\, (1 + 0.1750t)$
$[2/1]$	$(0.35t - 0.02042t^2)/(1 + 0.1167t)$
$[2/2]$	$0.35t \,/\, (1 + 0.1750t + 0.01021t^2)$
$[3/3]$	$\frac{0.35t + 0.0007146t^3}{1 + 0.1750t + 0.01225t^2 + 0.0003573t^3}$

Source: created by the authors.

5.4 Numerical Comparison

Table 3 compares Taylor and Padé approximation for the hardest topic ($a_4 = 0.35$) of a four-topic calculus curriculum with $a_1 = 0.80$, $a_2 = 0.62$, $a_3 = 0.47$, $a_4 = 0.35$ and target mastery $\tau = 0.85$. The rates are *illustrative* values ordered by increasing conceptual demand; no empirical estimation is claimed, and the conclusions below hold for any strictly decreasing $a_1 > a_2 > a_3 > a_4 > 0$, only the tabulated values depending on them. Evaluation is at the exact optimal time $T_4^* = -\ln(0.15)/0.35 = 5.420343$, where $a_4 T_4^* = 1.897120$ and $L_4(T_4^*) = 0.850000$. Entries are compared *at equal total degree*: the Taylor polynomial of degree N against the diagonal Padé approximant of total degree $2n = N$.

Table 3. Taylor vs. Padé approximation of $L_4(T_4^*) = 0.850000$ at equal total degree $N = 2n$ ($a_4 = 0.35$, $T_4^* = 5.420343$). The column ρ_n is the exact error ratio of Theorem 10; its theoretical ceiling is $K_{n,n} = 0.5000,\ 0.1667,\ 0.0500,\ 0.0143$ for $n = 1, 2, 3, 4$.

	Taylor, deg N		Padé $[n/n]$		
N	value	error	value	error	ρ_n
2	0.09759	$7.5 \cdot 10^{-1}$	0.97360	$1.2 \cdot 10^{-1}$	0.164
4	0.69584	$1.5 \cdot 10^{-1}$	0.84373	$6.3 \cdot 10^{-3}$	0.041
6	0.83588	$1.4 \cdot 10^{-2}$	0.85015	$1.5 \cdot 10^{-4}$	0.011
8	0.84926	$7.4 \cdot 10^{-4}$	0.85000	$2.1 \cdot 10^{-6}$	0.003
10	0.84998	$2.5 \cdot 10^{-5}$	—	—	—

Source: created by the authors.

Three features of Table 3 confirm the theory. First, at every total degree the Padé error is smaller, and the measured ρ_n lies strictly below the ceiling $K_{n,n}$ of Theorem 10(b), the gap being the factor $E_n(s)^{-1} I_1 / I_2 / K_{n,n}$, substantial already at $s = 1.897120$. Second, the error ratio Taylor/Padé at equal degree is 6.1, 25, 94, 353 for $N = 2, 4, 6, 8$, so the accuracy gain per unit of degree grows. Third,

the $[3/3]$ approximant reaches an error of $1.5 \cdot 10^{-4}$ at total degree 6, whereas at T_4^* Taylor needs degree 9 (degree 8 gives only $7.4 \cdot 10^{-4}$) — a degree-reduction factor $9/6 = 1.5$, rising with the horizon as quantified in Corollary 12.

Finally, Proposition 14 is tight up to the factor 2 of its last step, δ being the supremum of $|L_4 - R_{n,n}|$ over $[0, \max(T_4^*, \hat{T}_4^*)]$, well below $(1-\tau)/2$. At $n = 3$ the surrogate optimum is $\hat{T}_4^* = 5.417476$, so $|T_4^* - \hat{T}_4^*| = 2.87 \cdot 10^{-3}$, against the guaranteed bound (38), $2\delta/[a(1-\tau)] = 5.75 \cdot 10^{-3}$, with $\delta = 1.51 \cdot 10^{-4}$; at $n = 4$ the values are $3.97 \cdot 10^{-5}$ and $7.95 \cdot 10^{-5}$.

6 Conclusion

We have established a matrix elementary row operations framework that transports the augmented-matrix device of [5] from $\mathbb{Z}$ to $\mathbb{R}[[t]]$, with self-contained proofs of every step. The ERO algorithm (Theorem 4) covers off-diagonal orders as well as the diagonal; the determinant criterion (Theorem 5) certifies nondegeneracy, and Remark 6 shows that the apparent catastrophic conditioning of the Padé–Sylvester system is an artifact of scaling. For $L(t) = 1 - e^{-at}$ the transfer lemma (Lemma 7) gives a region of nondegeneracy and exhibits degenerate entries outside it; the exact Beta-integral remainder (Lemma 8) yields error bounds with the a-independent diagonal constant $K_{n,n} = 1/\binom{2n}{n}$ (Theorem 9); and Theorem 10 replaces a heuristic error-ratio argument by an exact quotient-of-integrals identity, a covariance proof of strict decay and the sharp asymptotics $\rho_n \sim (2n)!/(at)^{2n}$. Theorem 11 separates optimality of the bound from that of the leading error constant, and Proposition 13 determines, for $1 \le n \le 8$, the exact interval of monotonicity and concavity of the Padé learning curve, which makes the surrogate OLSP a convex program. Natural extensions include multivariate Padé approximants, [1], matrix-valued learning states via the matrix exponential, [8], [9], [10], and spectral questions for operator-valued models, for which the machinery of [22] offers a template. Proposition 13 raises the open question of proving $\sigma_n \ge 2\sqrt{2n-1}$ for all n, an equality at $n = 2$.

Acknowledgment:
The authors thank Kalasin University for institutional support and the three anonymous referees, whose criticism of the original proofs of Lemma 8 and Theorem 10 shaped this revision.

References:

[1] G. A. Baker Jr. and P. Graves-Morris, *Padé Approximants*, 2nd ed., Encyclopedia of Mathematics and its Applications, Vol. 59, Cambridge University Press, Cambridge, 1996, ISBN 978-0-521-45007-2, https://doi.org/10.1017/CBO9780511530074.

[2] C. Brezinski, *Padé-Type Approximation and General Orthogonal Polynomials*, International Series of Numerical Mathematics, Vol. 50, Birkhäuser, Basel, 1980, ISBN 978-3-0348-6559-3, https://doi.org/10.1007/978-3-0348-6558-6.

[3] L. N. Trefethen, *Approximation Theory and Approximation Practice, Extended Edition*, Society for Industrial and Applied Mathematics, Philadelphia, 2019, ISBN 978-1-611975-93-2, https://doi.org/10.1137/1.9781611975949.

[4] P. B. Borwein, On Padé and best rational approximation, *Canadian Mathematical Bulletin*, Vol. 26, No. 1, 1983, pp. 50–57, https://doi.org/10.4153/CMB-1983-009-x.

[5] W. Yonwilad, N. Thongmual, C. Sahatsathatsana, W. Pimpasalee, and S. Hobanthad, A matrix elementary operations framework for best rational approximation via continued fractions, *International Journal of Mathematics and Computer Science*, Vol. 21, No. 2, 2026, pp. 535–539, https://doi.org/10.69793/ijmcs/02.2026/yonwilad.

[6] A. Ya. Khinchin, *Continued Fractions*, English translation edited by H. Eagle, Dover Publications, Mineola, NY, 1997, ISBN 978-0-486-69630-0, https://store.doverpublications.com/products/9780486696300 [Access Date: August 14, 2026].

[7] W. Yonwilad, E. Lampai, and W. Nakkhasen, A Hankel-matrix ERO framework for Padé approximation of Stieltjes series, *Asia Pacific Journal of Mathematics*, Vol. 13, 2026, article 116, https://doi.org/10.28924/APJM/13-116.

[8] C. Van Loan, Computing integrals involving the matrix exponential, *IEEE Transactions on Automatic Control*, Vol. 23, No. 3, 1978, pp. 395–404, https://doi.org/10.1109/TAC.1978.1101743.

[9] C. B. Moler and C. F. Van Loan, Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later, *SIAM Review*, Vol. 45, No. 1, 2003, pp. 3–49, https://doi.org/10.1137/S00361445024180.

[10] N. J. Higham, The scaling and squaring method for the matrix exponential revisited, *SIAM Journal on Matrix Analysis and Applications*, Vol. 26, No. 4, 2005, pp. 1179–1193, https://doi.org/10.1137/04061101X.

[11] E. B. Saff and R. S. Varga, On the zeros and poles of Padé approximants to e^z, *Numerische Mathematik*, Vol. 25, No. 1, 1975, pp. 1–14, https://doi.org/10.1007/BF01419524.

[12] A. Cuyt, V. B. Petersen, B. Verdonk, H. Waadeland, and W. B. Jones, *Handbook of Continued Fractions for Special Functions*, Springer, Dordrecht, 2008, ISBN 978-1-4020-6948-2, https://doi.org/10.1007/978-1-4020-6949-9.

[13] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., Society for Industrial and Applied Mathematics, Philadelphia, 2002, ISBN 978-0-89871-521-7, https://doi.org/10.1137/1.9780898718027.

[14] J. R. Bunch, Stability of methods for solving Toeplitz systems of equations, *SIAM Journal on Scientific and Statistical Computing*, Vol. 6, No. 2, 1985, pp. 349–364, https://doi.org/10.1137/0906025.

[15] P. Gonnet, S. Güttel, and L. N. Trefethen, Robust Padé approximation via SVD, *SIAM Review*, Vol. 55, No. 1, 2013, pp. 101–117, https://doi.org/10.1137/110853236.

[16] Y. Nakatsukasa, O. Sète, and L. N. Trefethen, The AAA algorithm for rational approximation, *SIAM Journal on Scientific Computing*, Vol. 40, No. 3, 2018, pp. A1494–A1522, https://doi.org/10.1137/16M1106122.

[17] A. Heathcote, S. Brown, and D. J. K. Mewhort, The power law repealed: The case for an exponential law of practice, *Psychonomic Bulletin & Review*, Vol. 7, No. 2, 2000, pp. 185–207, https://doi.org/10.3758/BF03212979.

[18] C. Underhill and A. Wragg, Convergence properties of Padé approximants to $\exp(z)$ and their derivatives, *IMA Journal of Applied Mathematics*, Vol. 11, No. 3, 1973, pp. 361–367, https://doi.org/10.1093/imamat/11.3.361.

[19] J. Sastre, J. Ibañez, J. M. Alonso, and E. Defez, Beyond Paterson–Stockmeyer: advancing matrix polynomial computation, *WSEAS Transactions on Mathematics*, Vol. 24, 2025, pp. 684–693, https://doi.org/10.37394/23206.2025.24.68.

[20] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre, Correlation inequalities on some partially ordered sets, *Communications in Mathematical Physics*, Vol. 22, No. 2, 1971, pp. 89–103, https://doi.org/10.1007/BF01651330.

[21] R. Wong, *Asymptotic Approximations of Integrals*, Classics in Applied Mathematics, Vol. 34, Society for Industrial and Applied Mathematics, Philadelphia, 2001, ISBN 978-0-89871-497-5, https://doi.org/10.1137/1.9780898719260.

[22] A. A. Kutsenko, Application of matrix-valued integral continued fractions to spectral problems on periodic graphs with defects, *Journal of Mathematical Physics*, Vol. 58, No. 6, 2017, article 063516, https://doi.org/10.1063/1.4989987.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Wannatida Yonwilad: Conceptualization, Formal Analysis (Theorems 4–5, Lemmas 7–8, Theorem 9), Writing – original draft, Supervision. Ekkawit Lampai: Formal Analysis (Theorems 10–11, Corollary 12, Propositions 13–14, Tables 1–3), Validation, Writing – review and editing.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the authors used Grammarly to check grammar and spelling and Claude (Anthropic) to cross-check numerical calculations and suggest wording for individual passages. The authors reviewed, revised and approved all AI-assisted output, and take full responsibility for the content of the article. All mathematical ideas, theorems, proofs, arguments and conclusions are the authors' own.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflicts of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0, https://creativecommons.org/licenses/by/4.0/deed.en_US