Numerische Mathematik (2025) 157:1793–1846
https://doi.org/10.1007/s00211-025-01493-4

Numerische Mathematik

Summing divergent matrix series

Rongbiao Wang[1] · **JungHo Lee**[1,2] · **Lek-Heng Lim**[1]

Received: 30 May 2025 / Revised: 6 June 2025 / Accepted: 8 June 2025 /
Published online: 26 September 2025

Abstract

We extend several celebrated methods in classical analysis for summing series of complex numbers to series of complex matrices. These include the summation methods of Abel, Borel, Cesàro, Euler, Lambert, Nörlund, and Mittag-Leffler, which are frequently used to sum scalar series that are divergent in the conventional sense. One feature of our matrix extensions is that they are fully noncommutative generalizations of their scalar counterparts—not only is the scalar series replaced by a matrix series, positive weights are replaced by positive definite matrix weights, order on $\mathbb{R}$ replaced by Loewner order, exponential function replaced by matrix exponential function, etc. We will establish the regularity of our matrix summation methods, i.e., when applied to a matrix series convergent in the conventional sense, we obtain the same value for the sum. Our second goal is to provide numerical algorithms that work in conjunction with these summation methods. We discuss how the block and mixed-block summation algorithms, the Kahan compensated summation algorithm, may be applied to matrix sums with similar roundoff error bounds. These summation methods and algorithms apply not only to power or Taylor series of matrices but to any general matrix series including matrix Fourier and Dirichlet series. We will demonstrate the utility of these summation methods: establishing a Fejér's theorem and alleviating the Gibbs phenomenon for matrix Fourier series; extending the domains of matrix functions and accurately evaluating them; enhancing the matrix Padé approximation and Schur–Parlett algorithms; and more.

Mathematics Subject Classification 35A01 · 65L10 · 65L12 · 65L20 · 65L70

✉ Lek-Heng Lim
lekheng@uchicago.edu

Rongbiao Wang
rbwang@uchicago.edu

JungHo Lee
junghol@andrew.cmu.edu

[1] Computational and Applied Mathematics, University of Chicago, 5747 S. Ellis Avenue, Chicago 60637, IL, USA

[2] Department of Statistics and Data Science, Carnegie Mellon University, 5000 Forbes Ave Pittsburg, Pittsburgh 15213, PA, USA

1 Introduction

As we learned in calculus or real analysis, whenever we have an expression

$$\sum_{k=0}^{\infty} a_k = s \tag{1.1}$$

for some $a_k \in \mathbb{C}$, $k = 0, 1, 2, \ldots$, and $s \in \mathbb{C}$, the meaning of '=' is *defined* to be the convergence of the sequence of partial sums $s_n := \sum_{k=0}^{n} a_k$ to the limit s in the standard Euclidean metric $|\cdot|$ on $\mathbb{C}$. In this case the series $\sum_{k=0}^{\infty} a_k$ is said to be convergent with value s; and if it does not meet this definition of convergence, then it is said to be divergent.

Because of its ubiquity and utility, we sometimes lose sight of the fact that such an interpretation of '=' in (1.1) is purely by convention, and not sacrosanct. A series divergent in the sense of the conventional definition may have a well-defined value under alternative definitions of '=' that are perfectly legitimate mathematically. Take the harmonic series $\sum_{k=1}^{\infty} 1/k$ for illustration, well-known to be divergent in the conventional sense but as soon as we change, say, the choice of the metric from Euclidean to p-adic $|\cdot|_p$, it becomes convergent in the sense that $|s_n - s|_p \to 0$ for some value $s \in \mathbb{C}$ that depends on p [1]. Indeed, a well-known result in p-adic analysis [2] is that with a p-adic metric, a series $\sum_{k=0}^{\infty} a_k$ is convergent if and only if $\lim_{k\to\infty} a_k = 0$, obviously false by the conventional definition of series convergence.

Even if we restrict ourselves to the Euclidean metric, which is what we will do in the rest of this article, the meaning of '=' still depends on a specific way to sum the values $a_k \in \mathbb{C}$, $k = 0, 1, 2, \ldots$. As was known to early analysts, there are many other reasonable ways to assign a value to a series that is divergent in the conventional sense, and such values are mathematically informative and useful in many ways [3]. As Hardy pointed out [3], a *summation method* just needs to be a function from the set of infinite series to values, assigning a sum to a series, which may or may not be convergent in the conventional sense.

The first and best-known summation method is likely Cesàro summation [4], that allows one to sum the Grandi series $1 - 1 + 1 - 1 + \cdots$ to $1/2$. The idea can be traced back even earlier to Leibniz, d'Alambert, Cauchy, and other predecessors of Cesàro [3, 5]. Cesàro summation has the property of being *regular*, i.e., for a series that is convergent in the conventional sense, the method gives an identical value for the sum. Regular summation methods have been studied extensively [3, 5–7] and applied in various fields from analytic number theory [8] to quantum field theory [9] to statistics [10]. Indeed summing divergent series is an important aspect of *renormalization*, a cornerstone of modern physics [11], particularly in the renormalization technique of *zeta function regularization* [12].

A main goal of our article is to show that many if not most of these summation methods for series of complex numbers extend readily and naturally to series of complex matrices. Take a toy example for illustration: the Neumann series

$$\sum_{k=0}^{\infty} X^k = (I - X)^{-1} \tag{1.2}$$

if and only if the spectral radius of $X \in \mathbb{C}^{d\times d}$ is less than 1. Again '=' here is interpreted in the sense of conventional summation, i.e., the sequence of partial sums $S_n := \sum_{k=0}^{n} X^k$ converges to $(I - X)^{-1}$ with respect to any matrix norm $\|\cdot\|$. Let $\lambda(X)$ denote the spectrum of X and $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$ the complex open unit disc. Depending on which method we use to sum the series on the left-hand side of (1.2), we obtain different interpretations of '=':

conventional: (1.2) holds if and only if $\lambda(X) \subseteq \mathbb{D}$;
Abel: (1.2) holds if and only if $\lambda(X) \subseteq \overline{\mathbb{D}} \setminus \{1\}$;
Cesàro: (1.2) holds if and only if $\lambda(X) \subseteq \overline{\mathbb{D}} \setminus \{1\}$ and the geometric and algebraic multiplicities are equal for each eigenvalue in $\lambda(X) \cap \partial\mathbb{D} \setminus \{1\}$;
Euler: (1.2) holds if and only if $\lambda((I + P)^{-1}(P + X)) \subseteq \mathbb{D}$ for some $P \succ 0$ commuting with X;
Borel: (1.2) holds if and only if $\lambda(X) \subseteq \{z \in \mathbb{C} : \mathrm{Re}(\lambda) < 1\}$.

The last four summation methods will be defined in due course. In case the reader is wondering, although the matrix $(I - X)^{-1}$ is well-defined as long as $1 \notin \lambda(X)$, we will see that there is no natural method that will extend the validity of (1.2) to all $X \in \mathbb{C}^{d\times d}$ with $\lambda(X) \subseteq \mathbb{C} \setminus \{1\}$.

In the toy example above, the series in question is a *power series* where the kth term is a scalar multiple of X^k. The matrix summation methods in our article will apply more generally to *any series* of matrices $\sum_{k=0}^{\infty} A_k$, where A_k may not be Taylor in nature, i.e., $(X-\alpha I)^k$, but may be Fourier $\sin(kX)$, Dirichlet $\exp(X \log k)$, Hadamard powers $X^{\circ k}$, or yet other forms not covered in this article, e.g., it could be defined by a recurrence relation $A_k = BA_{k-1}(I - A_{k-1})$ or randomly generated $A_k \sim \textsc{Wishart}(\Sigma, n)$.

So a second goal of our article is to provide practical numerical algorithms that complement our theoretical summation methods. These algorithms will allow us to compute, in standard floating point arithmetic, a matrix $\widehat{S} \in \mathbb{C}^{d\times d}$ that approximates the theoretical sum $S \in \mathbb{C}^{d\times d}$ of the series $\sum_{k=0}^{\infty} A_k$ given by the respective summation method.

These two aspects are complementary: There is no numerical method that would allow one to ascertain the convergence of a series, regardless of which summation method we use. A standard example is the harmonic series $\sum_{k=1}^{\infty} 1/k$; every numerical method would yield a finite value [13, Sect. 4.2], which is completely meaningless since its true value is $+\infty$. On the other hand, most of the matrix series we encounter will have no alternate closed-form expressions, again regardless of which summation method we use. The only way to obtain an approximate value would be through computing one in floating point arithmetic. In summary, the theoretical summation method permits us to determine convergence; and its corresponding numerical algorithm permits us to determine an approximate value. We provide an overview of these two aspects of our work.

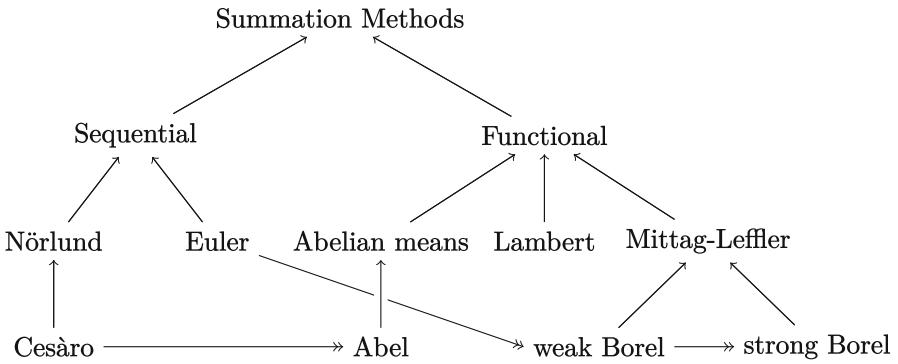


Fig. 1 Relations between various methods: $a \to b$ means a-summation is a special case of b-summation; $a \twoheadrightarrow b$ means a-summable implies b-summable

Theoretical: regular summation methods

As in the case of its scalar counterpart, a matrix summation method is a *partial function*, i.e., possibly defined on a subset of its stated domain, from the set of $d \times d$ complex matrix series to a sum in $\mathbb{C}^{d\times d}$. We will generalize five classes of summation methods for scalar-valued series to matrix-valued ones. Figure 1 organizes them in a tree.

The five summation methods fall under two broad categories, *sequential* and *functional* methods, discussed in Sects. 3 and 4 respectively. These terminologies follow those for scalar-valued series [5]. Basically, a sequential method transforms the terms of a series or its sequence of partial sums into another sequence, whereas a functional method would transform them into a function. We will generalize two of the most important sequential methods, Nörlund (of which Cesàro is a special case) and Euler; and three of the most important functional methods, Lambert, Abelian means, and Mittag-Leffler (Abel and Borel summations are respectively special cases of the latter two); showing that they also work for matrix series.

One feature of our generalizations that we wish to highlight is that they are truly matrix-valued to the fullest extent possible. For example, our generalization of Nörlund summation $\lim_{n\to\infty}(\rho_0+\cdots+\rho_n)^{-1}(\rho_n s_0+\rho_{n-1}s_1+\cdots+\rho_0 s_n)$ with $s_n := \sum_{k=0}^{n} a_k$ does not just replace $a_k \in \mathbb{C}$ by matrices $A_k \in \mathbb{C}^{d\times d}$ but also the positive scalars $\rho_0, \ldots, \rho_n$ by positive definite matrices $P_0, \ldots, P_n$. Our extension of Abel summation $\lim_{x\to 0}\sum_{k=0}^{\infty} a_k e^{-\rho_k x}$ does not merely replace $a_k \in \mathbb{C}$ by matrices $A_k \in \mathbb{C}^{d\times d}$ but also the increasing sequence $0 < \rho_0 < \rho_1 < \cdots$ by a sequence of matrices increasing in Loewner order $0 \prec P_0 \prec P_1 \prec \cdots$ and e^x by the matrix exponential function.

Practical: numerical summation algorithms

Once a matrix series is ascertained to be summable via one of the aforementioned theoretical methods, the corresponding numerical method would be used to provide an approximate value in the form of a finite sum. However, it is nontrivial to obtain an accurate value for this finite sum in finite precision arithmetic. Simply adding terms

in the finite sum in any fixed order would not give the most accurate result. In Sect. 6, we adapt three numerical summation algorithms in [13, 14] for sums of matrices:

(i) *block summation:* divide the finite sum into equally-sized blocks and sum the local blocks recursively, then sum the local sums recursively;
(ii) *compensated summation:* keep a running compensation term to extend the precision;
(iii) *mixed block summation:* divide the finite sum into equally-sized blocks and sum the local blocks with one algorithm, then sum the local sums with another algorithm.

We stress that the summation methods above apply to *any* series of matrices, not just power series of matrices like those commonly found in the matrix functions literature [15]. In Sect. 7, for the special case when we do have a matrix power series, we extend two algorithms for summing matrix power series [15] by enhancing them with the summation methods introduced in Sect. 3:

(iv) *Padé approximation*: best rational approximation of a matrix function at a given order;
(v) *Schur–Parlett algorithm*: Schur decomposition followed by block Parlett recurrence.

We present numerical experiments in Sect. 8 to illustrate the value and practicality of our summation methods:

(a) using Cesàro summation to alleviate Gibbs phenomenon in matrix Fourier series;
(b) using Euler and strong Borel summations to extend matrix Taylor series;
(c) using Euler summations for high accuracy evaluation of matrix functions;
(d) using Cesàro and Euler summations in Padé approximations;
(e) using Lambert summation to investigate matrix Dirichlet series;
(f) using compensated matrix summation for accurate evaluation of Hadamard power series.

There are some surprises. For example, for (1.2), using Euler sum to evaluate the Neumann series and using MATLAB's `inv` to invert $I - X$, Euler sum gives results that are an order of magnitude more accurate than MATLAB's `inv`; Gibbs phenomenon in matrix Fourier series happens only when the matrix involved is diagonalizable; a well-known property of the Riemann zeta function remains true for the matrix zeta function; etc.

As one of our goals is to compute an approximate value in floating arithmetic for the matrix series and summation methods studied in this article, so even though some of our theoretical results readily extend to Banach algebras we do not pursue this unnecessary generality.

2 Conventions and notations

Recall that it does not matter which matrix norm we use since all norms on a finite-dimensional space $\mathbb{C}^{d\times d}$ are equivalent and thus induce the same topology. Throughout

this article, we will use $\|\cdot\|$ to denote the Euclidean norm on $\mathbb{C}^d$ and spectral norm on $\mathbb{C}^{d\times d}$:

$$\|A\| := \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}.$$

Note that the norm notation is consistent if we adopt the standard convention of identifying vectors in $\mathbb{C}^d$ with single-column matrices in $\mathbb{C}^{d\times 1}$. For $P \in \mathbb{C}^{d\times d}$, we use the shorthand $P \succ 0$ for P positive definite, i.e., $x^*Px > 0$ for all nonzero $x \in \mathbb{C}^d$. Recall that this condition implies that P must also be Hermitian [16, p. 80] (but the analogous statement is not true over $\mathbb{R}$). More generally $\succ$ denotes the Loewner order, i.e., $A \succ B$ if and only if $A - B$ is positive definite. We write I for the identity matrix and $\mathbb{1}$ for the matrix of ones. We use $\lambda(X)$ to denote the spectrum of $X \in \mathbb{C}^{d\times d}$. A note of caution is that we do not treat $\lambda(X)$ as a multiset; whenever we write $\lambda(X) = \{\lambda_1, \ldots, \lambda_r\}$, the elements λ_i's are necessarily distinct. For example, $\lambda(I) = \{1\}$ always, not $\{1, \ldots, 1\}$.

We write $\mathrm{Re}(x)$ and $\mathrm{Im}(x)$ for the real and imaginary parts of $x \in \mathbb{C}$, respectively. We write $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$ for the open unit disc and $\mathbb{N} := \{0, 1, 2, \ldots\}$ for the nonnegative integers. Unless noted otherwise, all sequences, summands in a series, partial sums, will be indexed by $\mathbb{N}$ throughout this article. We denote closure and boundary of a set Ω by $\overline{\Omega}$ and $\partial\Omega$ respectively.

We also lay out some formal definitions and standard notations [17–20] related to sequences and series of matrices.

Definition 2.1 A *matrix sequence* $A_\bullet := (A_k)_{k=0}^\infty$ is a map from $\mathbb{N}$ to $\mathbb{C}^{d\times d}$ whose value at $k \in \mathbb{N}$ is denoted by $A_k \in \mathbb{C}^{d\times d}$. We say that the matrix sequence $A_\bullet$ converges to $A \in \mathbb{C}^{d\times d}$ if

$$\lim_{k\to\infty} \|A_k - A\| = 0,$$

and denote it by $\lim_{k\to\infty} A_k = A$. We denote the vector space of matrix sequences by

$$s(\mathbb{C}^{d\times d}) := \left\{A_\bullet : A_k \in \mathbb{C}^{d\times d}, k \in \mathbb{N}\right\}$$

and its subspace of convergent matrix sequences by

$$c(\mathbb{C}^{d\times d}) := \left\{A_\bullet : \lim_{k\to\infty} A_k = A \in \mathbb{C}^{d\times d}\right\}.$$

We speak of a *series* when we are interested in summing a sequence. Therefore, a series $\sum_{k=0}^\infty A_k$ and its underlying sequence $A_\bullet$ are one-and-the-same object and we will not distinguish them. While the convergence of a matrix sequence is unambiguous throughout this article, the summability of a matrix series is not and will take on multiple different meanings. Getting ahead of ourselves, we will be defining the R-sum S of a series $\sum_{k=0}^\infty A_k$ and writing

$$\sum_{k=0}^\infty A_k \overset{\mathsf{R}}{=} S \tag{2.1}$$

where different letters in place of R would refer to Nörlund means (N), Cesàro (C), Euler (E), Abelian means (A), Lambert (L), weak Borel (WB), strong Borel (SB), and Mittag-Leffler (M) summations, all of which will be defined in due course. We say that $A_\bullet$ is R-summable if there is a well-defined R-sum $S \in \mathbb{C}^{d\times d}$. The absence of a letter would denote conventional summation, i.e., S is the limit of its sequence of partial sums, $S_\bullet = (S_k)_{k=0}^\infty$, $S_n := \sum_{k=0}^n A_k$.

As usual, we write $C_b(\Omega) := C_b(\Omega, \mathbb{C})$ for the Banach space of complex-valued continuous functions equipped with the uniform norm; Ω will usually be an open interval in $\mathbb{R}$. We will often have to discuss matrices whose entries are in $C_b(\Omega)$, i.e., $A(x) = [a_{ij}(x)]$ with continuous and bounded $a_{ij} : \Omega \to \mathbb{C}$, $i, j = 1, \dots, d$. We denote the space of such matrices as $C_b(\Omega)^{d\times d}$. These may also be viewed as matrix-valued continuous maps $A : \Omega \to \mathbb{C}^{d\times d}$ or as tensor product of the two Banach spaces [21]:

$$C_b(\Omega)^{d\times d} = C_b(\Omega, \mathbb{C}^{d\times d}) = C_b(\Omega) \otimes \mathbb{C}^{d\times d}.$$

Indeed the tensor product view will be the neatest as we will also need to speak of $C_b(\Omega)^{d\times d}$-valued sequences:

$$s\big(C_b(\Omega)^{d\times d}\big) = s(\mathbb{C}^{d\times d}) \otimes C_b(\Omega), \qquad c\big(C_b(\Omega)^{d\times d}\big) = c(\mathbb{C}^{d\times d}) \otimes C_b(\Omega)$$

but we will avoid tensor products for fear of alienating readers unfamiliar with the notion.

We will occasionally use the notion of a *partial function* to refer to a map from a set $\mathcal{X}$ to a set $\mathcal{Y}$ defined on a subset $\mathcal{S} \subseteq \mathcal{X}$ called its *natural domain*. These are useful when we wish to speak loosely of a map from $\mathcal{X}$ to $\mathcal{Y}$ that may not be defined on all of $\mathcal{X}$, but whose natural domain may be difficult to specify a priori. A matrix summation method falls under this situation as we want to define a map R on $s(\mathbb{C}^{d\times d})$ that is only well-defined on its natural domain of R-summable series. Following convention in algebraic geometry, we write $\mathsf{R} : \mathcal{X} \dashrightarrow \mathcal{Y}$ to indicate that R may be a partial function.

3 Sequential summation methods

Many summation methods for scalar series are *sequential summation method* [3]. In this section we will extend them to matrix series, defining various partial functions that map a sequence $A_\bullet \in s(\mathbb{C}^{d\times d})$ into a suitably transformed sequence in $c(\mathbb{C}^{d\times d})$ and defining the corresponding sum[1] as the limit of the transformed sequence.

Let $C_{n,k} \in \mathbb{C}^{d\times d}$, $n, k \in \mathbb{N}$, and consider the partial function $\mathsf{R} : s(\mathbb{C}^{d\times d}) \dashrightarrow c(\mathbb{C}^{d\times d})$ given by

$$\mathsf{R}(A_\bullet)_n = \sum_{k=0}^{\infty} C_{n,k} S_k \tag{3.1}$$

for any $A_\bullet \in s(\mathbb{C}^{d\times d})$, $S_n = \sum_{k=0}^n A_k$, and $n \in \mathbb{N}$. The R-sum is defined to be $\lim_{n\to\infty} \mathsf{R}(A_\bullet)_n$, if the limit exists; and in which case we write (2.1) with $S = \lim_{n\to\infty} \mathsf{R}(A_\bullet)_n$.

[1] Sometimes called *antilimit* for easy distinction from the partial sums [3].

For a conventionally summable series, we expect our summation method to yield the same sum, i.e., $\sum_{k=0}^{\infty} A_k \overset{\mathsf{R}}{=} S$ whenever $\sum_{k=0}^{\infty} A_k = S$. This property is called *regularity*. All summation methods considered in our article will be shown to be regular. In fact they satisfy the slightly stricter but completely natural condition of *total regularity*: If a series sums to $+\infty$ conventionally, then the method also sums it to $+\infty$. Total regularity is the reason why the validity of (1.2) cannot be extended to all of $\mathbb{C} \setminus \{1\}$: For $d = 1$, $\sum_{k=0}^{\infty} x^k = +\infty$ whenever $x \in (1, +\infty)$ and so no totally regular method could ever yield $(1-x)^{-1}$ for all $x \in \mathbb{C}\setminus\{1\}$. We will not discuss total regularity in the rest of this article.

We provide a sufficient condition for the regularity of sequential summation methods (3.1), generalizing [3, Theorem 1] to matrices.

Theorem 3.1 *Let $C_{n,k} \in \mathbb{C}^{d\times d}$, $n, k \in \mathbb{N}$. Suppose*

(i) *there exists $\eta > 0$ such that $\sum_{k=0}^{\infty} \|C_{n,k}\| < \eta$ for each $n \in \mathbb{N}$;*
(ii) *$\lim_{n\to\infty} C_{n,k} = 0$ for each $k \in \mathbb{N}$;*
(iii) *$\lim_{n\to\infty} \sum_{k=0}^{\infty} C_{n,k} = I$.*

Then for any $A_\bullet \in s(\mathbb{C}^{d\times d})$ with $S_n = \sum_{k=0}^{n} A_k$ and $\lim_{n\to\infty} S_n = S$, the series $\sum_{k=0}^{\infty} C_{n,k} S_k$ is summable in the conventional sense for each $n \in \mathbb{N}$ and

$$\lim_{n\to\infty} \sum_{k=0}^{\infty} C_{n,k} S_k = S. \tag{3.2}$$

Proof Since $S_\bullet = (S_k)_{k=0}^{\infty}$ is convergent and therefore bounded, $\|S_k\| \le \beta$ for some $\beta > 0$ and all $k \in \mathbb{N}$. It follows from (i) that for each $n \in \mathbb{N}$,

$$\sum_{k=0}^{\infty} \|C_{n,k} S_k\| \le \sum_{k=0}^{\infty} \|C_{n,k}\| \cdot \|S_k\| \le \beta\eta.$$

Thus $\sum_{k=0}^{\infty} C_{n,k} S_k$ is summable. To show (3.2), first assume that $S = 0$. For $\varepsilon > 0$, choose $m \in \mathbb{N}$ sufficiently large so that $\|S_k\| < \varepsilon/2\eta$ for $k > m$. By (i) and (ii),

$$\lim_{n\to\infty} \sum_{k=0}^{m} C_{n,k} S_k = 0 \quad \text{and} \quad \Big\| \sum_{k>m} C_{n,k} S_k \Big\| \le \frac{\varepsilon}{2\eta} \sum_{k>m} \|C_{n,k}\| \le \frac{\varepsilon}{2}.$$

Hence

$$\lim_{n\to\infty} \sum_{k=0}^{\infty} C_{n,k} S_k = \lim_{n\to\infty} \left(\sum_{k=0}^{m} C_{n,k} S_k + \sum_{k>m} C_{n,k} S_k \right) = 0.$$

For $S \neq 0$, consider $A'_\bullet = (A'_k)_{k=0}^{\infty}$ with

$$A'_k := \begin{cases} A_0 - S & k = 0, \\ A_k & k = 1, 2, \ldots, \end{cases}$$

with partial sums $S'_k = S_k - S$, $k \in \mathbb{N}$. Since $\lim_{k\to\infty} S'_k = 0$, we get $\lim_{n\to\infty} \sum_{k=0}^{\infty} C_{n,k} S'_k = 0$. By (iii),

$$\lim_{n\to\infty} \sum_{k=0}^{\infty} C_{n,k}(S'_k + S) = \lim_{n\to\infty} \sum_{k=0}^{\infty} C_{n,k} S'_k + \lim_{n\to\infty} \sum_{k=0}^{\infty} C_{n,k} S = S.$$

□

Theorem 3.1 should be interpreted as follows: A summation method of the form (3.1) satisfying (iii) should be seen as taking matrix-weighted averages of the sequence of partial sums $S_\bullet$ for each n. Conditions (i) and (ii) are what one would expect for weights: absolutely summable and not biased towards any partial sum S_k respectively. Theorem 3.1 would be useful for establishing regularity of matrix summation methods involving various choices of matrix weights.

3.1 Nörlund means

The scalar version of this summation method was first introduced in [22] but named after Nörlund who rediscovered it [23]. Its most notable use is for summing Fourier series [24, 25]. Here we will extend it to series of matrices.

Definition 3.2 Let $P_\bullet := (P_k)_{k=0}^{\infty}$ be a sequence of positive definite matrices such that

$$\lim_{k\to\infty} \|(P_0 + \cdots + P_k)^{-1}\| \|P_k\| = 0. \tag{3.3}$$

For $A_\bullet = (A_k)_{k=0}^{\infty} \in s(\mathbb{C}^{d\times d})$, the series $\sum_{k=0}^{\infty} A_k$ is *Nörlund summable* to $S \in \mathbb{C}^{d\times d}$ with respect to $P_\bullet$ if

$$\lim_{n\to\infty} (P_0 + \cdots + P_n)^{-1}(P_n S_0 + P_{n-1} S_1 + \cdots + P_0 S_n) = S,$$

where $S_n = \sum_{k=0}^{n} A_k$, $n \in \mathbb{N}$. We denote this by

$$\sum_{k=0}^{\infty} A_k \overset{\mathrm{N}}{=} S$$

and call S the Nörlund mean of $A_\bullet$. There is an implicit choice of $P_\bullet$ not reflected in the notation.

Corollary 3.3 (Regularity of Nörlund mean) *Let $P_\bullet := (P_k)_{k=0}^{\infty}$ be a sequence of positive definite matrices satisfying (3.3). For $A_\bullet \in s(\mathbb{C}^{d\times d})$ and $S \in \mathbb{C}^{d\times d}$, if $\sum_{k=0}^{\infty} A_k = S$, then $\sum_{k=0}^{\infty} A_k \overset{\mathrm{N}}{=} S$.*

Proof The Nörlund mean is a sequential summation method with a choice of

$$C_{n,k} = \begin{cases} (P_0 + \cdots + P_n)^{-1} P_{n-k} & \text{if } k \le n, \\ 0 & \text{if } k > n, \end{cases}$$

in (3.1). To show regularity, we check the three conditions of Theorem 3.1: Since

$$\sum_{k=0}^{n} (P_0 + \cdots + P_n)^{-1} P_{n-k} = I,$$

the Conditions (i) and (iii) are satisfied. Since $P_0 + \cdots + P_n \succ P_0 + \cdots + P_{n-k}$, we have $\|(P_0 + \cdots + P_n)^{-1}\| \le \|(P_0 + \cdots + P_{n-k})^{-1}\|$ and so Condition (ii) is satisfied as

$$\lim_{n\to\infty} \|(P_0 + \cdots + P_n)^{-1}\| \|P_{n-k}\| \le \lim_{n\to\infty} \|(P_0 + \cdots + P_{n-k})^{-1}\| \|P_{n-k}\| = 0.$$

□

The well-known Cesàro summation is a special case of Nörlund summation [3, Sect. 5.13] with $P_\bullet$ given by

$$P_k = \binom{k+j-1}{j-1} I, \quad k \in \mathbb{N},$$

for $j \in \mathbb{N} \setminus \{0\}$. We extend the definition to matrices and write (C, j) for the jth order Cesàro summation. In particular, $(\mathsf{C}, 1)$ is Cesàro summation extended to a matrix series, defined formally below.

Definition 3.4 Let $A_\bullet = (A_k)_{k=0}^{\infty} \in s(\mathbb{C}^{d\times d})$ and $S_\bullet$ be its sequence of partial sums, $S_n = \sum_{k=0}^{n} A_k$. Define

$$\Sigma_n := \frac{1}{n} \sum_{k=0}^{n-1} S_k.$$

The series $\sum_{k=0}^{\infty} A_k$ is *Cesàro summable* to $S \in \mathbb{C}^{d\times d}$ if $\lim_{n\to\infty} \Sigma_n = S$. We denote this by

$$\sum_{k=0}^{\infty} A_k \overset{\mathsf{C}}{=} S$$

and call S the Cesàro sum of $A_\bullet$.

A standard example of a Cesàro summable (scalar-valued) series divergent in the usual sense is the Grandi series $1 - 1 + 1 - 1 + \cdots$, which sums to $1/2$ in the Cesàro sense. Indeed, this is a special case of $\sum_{k=0}^{\infty} x^k \overset{\mathsf{C}}{=} 1/(1-x)$ for any $x \in \overline{\mathbb{D}} \setminus \{1\}$, which follows from

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \sum_{k=0}^{m} x^k = \lim_{n\to\infty} \sum_{m=0}^{n-1} \frac{1-x^{m+1}}{n(1-x)} = \frac{1}{1-x}. \tag{3.4}$$

We establish the corresponding result for matrix Neumann series, which is more involved.

Proposition 3.5 (Cesàro summability of Neumann series) *Let $X \in \mathbb{C}^{d\times d}$. Then*

$$\sum_{k=0}^{\infty} X^k \overset{\mathsf{C}}{=} (I-X)^{-1} \tag{3.5}$$

if and only if $\lambda(X) \subseteq \overline{\mathbb{D}} \setminus \{1\}$ and the geometric and algebraic multiplicities are equal for each eigenvalue in $\lambda(X) \cap \partial\mathbb{D} \setminus \{1\}$.

Proof We start with the backward implication. The assumption falls into two cases:

(i) $\lambda(X) \subseteq \mathbb{D}$,
(ii) $\lambda(X) \cap \partial\mathbb{D} \setminus \{1\} \neq \varnothing$ and the geometric and algebraic multiplicities are equal for each eigenvalue in $\lambda(X) \cap \partial\mathbb{D} \setminus \{1\}$.

For the first case, we have $\sum_{k=0}^{\infty} X^k = (I-X)^{-1}$ and so (3.5) holds by Corollary 3.3 since Cesàro summation is a special case of Nörlund summation. For the second case, let $J_1, \ldots, J_r$ be a list of all Jordan blocks of X corresponding to eigenvalues $\lambda_1, \ldots, \lambda_j \in \partial\mathbb{D} \setminus \{1\}$ and $\lambda_{j+1}, \ldots \lambda_r \in \mathbb{D}$. By assumption, $\lambda_1, \ldots, \lambda_j$ have equal geometric and algebraic multiplicities and thus the Jordan blocks $J_1, \ldots, J_j$ are all 1×1. Hence the Jordan decomposition $X = WJW^{-1}$ has the form

$$J = \begin{bmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_j & & & \\ & & & J_{j+1} & & \\ & & & & \ddots & \\ & & & & & J_r \end{bmatrix},$$

where $J_i \in \mathbb{C}^{d_i \times d_i}$, $i = j+1, \ldots, r$. By (3.4),

$$\sum_{k=0}^{\infty} \lambda_i^k \overset{\mathsf{C}}{=} \frac{1}{1-\lambda_i} \qquad \text{for } i = 1, \ldots, j.$$

By the first case,

$$\sum_{k=0}^{\infty} J_i^k \overset{\mathsf{C}}{=} (1-J_i)^{-1} \qquad \text{for } i = j+1, \ldots, r.$$

Therefore,

$$\sum_{k=0}^{\infty}\begin{bmatrix}\lambda_1^k & & & & \\ & \ddots & & & \\ & & \lambda_j^k & & \\ & & & J_{j+1}^k & \\ & & & & \ddots & \\ & & & & & J_r^k\end{bmatrix} \overset{\mathsf{C}}{=} \begin{bmatrix}\frac{1}{1-\lambda_1} & & & & \\ & \ddots & & & \\ & & \frac{1}{1-\lambda_j} & & \\ & & & (I-J_{j+1})^{-1} & \\ & & & & \ddots & \\ & & & & & (I-J_r)^{-1}\end{bmatrix},$$

so

$$\sum_{k=0}^{\infty} X^k = \sum_{k=0}^{\infty} W J^k W^{-1} \overset{\mathsf{C}}{=} (I-X)^{-1}.$$

We next establish the forward implication. Suppose X has spectral radius $\rho(X) > 1$. As

$$\rho(X) = \lim_{k\to\infty} \|X^k\|^{\frac{1}{k}},$$

for any $\epsilon < \rho(X) - 1$, there is some $m \in \mathbb{N}$ such that $(\rho(X)-\epsilon)^k \leq \|X^k\|$ for all $k > m$. In other words, $\|X^k\|$ grows exponentially and thus

$$\lim_{n\to\infty} \|\Sigma_n\| = \lim_{n\to\infty} \left\| \frac{1}{n} \sum_{m=0}^{n-1} \sum_{k=0}^{m} X^k \right\| = \infty.$$

Also, observe that

$$\frac{1}{n} \sum_{m=0}^{n-1} \sum_{k=0}^{m} 1 = \frac{1}{n} \sum_{m=0}^{n-1} m = \frac{n-1}{2};$$

so if 1 is an eigenvalue of X, then its Neumann series cannot be Cesàro summable. Hence we must have $\lambda(X) \subseteq \overline{\mathbb{D}} \setminus \{1\}$. It remains to rule out the case where X has a Jordan block of size greater than 1×1 for an eigenvalue in $\partial\mathbb{D} \setminus \{1\}$. Suppose X has a Jordan block $J_\lambda \in \mathbb{C}^{d_i \times d_i}$ with $d_i \geq 2$ and corresponding eigenvalue $\lambda \in \partial\mathbb{D} \setminus \{1\}$. Dropping the subscript i to avoid clutter, we have

$$J_\lambda^k = \begin{bmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \dots & \binom{k}{d-1}\lambda^{k-(d-1)} \\ & \lambda^k & \binom{k}{1}\lambda^{k-1} & \dots & \binom{k}{d-2}\lambda^{k-(d-2)} \\ & & \ddots & \ddots & \vdots \\ & & & \lambda^k & \binom{k}{1}\lambda^{k-1} \\ & & & & \lambda^k \end{bmatrix} \quad \text{for } k > d. \tag{3.6}$$

Observe that the (1, 2)th entry,

$$\left(\frac{1}{n} \sum_{m=0}^{n-1} \sum_{k=0}^{m} J_\lambda^k\right)_{12} = \frac{1}{n} \sum_{m=0}^{n-1} \sum_{k=0}^{m} k\lambda^{k-1} = \sum_{k=0}^{n-1} \frac{(k-1)\lambda^k - k\lambda^{k-1} + 1}{n(1-\lambda)^2}$$

is divergent as $n \to \infty$. So the series $\sum_{k=0}^{\infty} J_\lambda^k$ is not Cesàro summable and neither is the Neumann series of X. □

The proof above shows that whenever there is a Jordan block of size greater than 1 with eigenvalues on $\partial\mathbb{D} \setminus \{1\}$, Cesàro summation will fail to sum the Neumann series. In Sect. 4.1, we will see how we may overcome this difficulty with Abel summation.

The best-known application of the scalar Cesàro summation is from Fourier Analysis [26–28]. It is well known that if $f \in L^2(-\pi, \pi)$, then its Fourier series

$$s_n(x) := \sum_{k=-n}^{n} \widehat{f}(k) e^{ikx}, \qquad \widehat{f}(k) := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx}\, dx,$$

converges to f in the L^2-norm, i.e., $\lim_{n\to\infty} \|s_n - f\|_2 = 0$. Fejér's theorem [29] gives the L^∞-norm analogue for continuous functions with one caveat—the series has to be taken in the Cesàro sense: If $f \in C(-\pi, \pi)$, then

$$\sigma_n(x) := \frac{1}{n} \sum_{m=0}^{n-1} \sum_{k=-m}^{m} \widehat{f}(k) e^{ikx},$$

converges uniformly to f, i.e., $\lim_{n\to\infty} \|\sigma_n - f\|_\infty = 0$.

A well-known consequence of Fejér's theorem is that if $f \in L^2(-\pi, \pi)$ is continuous at $x \in (-\pi, \pi)$, then its Cesàro sum converges pointwise to $f(x)$ [26, 28]. As an application of our notion of Cesàro summability for matrices, we extend Fejér's theorem to arbitrary matrices $X \in \mathbb{C}^{d\times d}$ with real eigenvalues and 2π-periodic functions $f \in C^{d-1}(\mathbb{R})$. We emphasize that we do not require diagonalizability of X. While we have assumed that f is $(d-1)$-times differentiable for simplicity, it will be evident from the proof that the result holds for any $f \in C(\mathbb{R})$ that is $(d_\lambda - 1)$-times differentiable at each $\lambda \in \lambda(X)$ where d_λ is the size of the largest Jordan block corresponding to λ. The exponential function, when applied to a matrix argument, refers to the matrix exponential [15, Sect. 10.8].

Proposition 3.6 (Fejér's theorem for matrix Fourier series) *Let $X \in \mathbb{C}^{d\times d}$ have all eigenvalues real. Let $f \in C^{d-1}(\mathbb{R})$ be 2π-periodic. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \sum_{k=-m}^{m} \widehat{f}(k) e^{ikX} = f(X). \tag{3.7}$$

Proof By the standard Fejér's theorem, for $j \le d-1$,

$$\lim_{n\to\infty} \sigma_n^{(j)}(\lambda) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \sum_{k=-m}^{m} (ik)^j \widehat{f}(k) e^{ik\lambda} = f^{(j)}(\lambda)$$

for any $\lambda \in \mathbb{R}$ and where the parenthetical superscripts denote jth derivative. For a Jordan block $J \in \mathbb{C}^{d\times d}$ with eigenvalue $\lambda \in \mathbb{R}$,

$$f(J) = \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} & \cdots & \frac{f^{(d-1)}(\lambda)}{(d-1)!} \\ & f(\lambda) & f'(\lambda) & \cdots & \frac{f^{(d-2)}(\lambda)}{(d-2)!} \\ & & \ddots & \ddots & \vdots \\ & & & f(\lambda) & f'(\lambda) \\ & & & & f(\lambda) \end{bmatrix}$$

and so

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \sum_{k=-m}^{m} \widehat{f}(k) e^{ikJ} &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \sum_{k=-m}^{m} \widehat{f}(k) \\ &\quad \times \begin{bmatrix} e^{ik\lambda} & ike^{ik\lambda} & \frac{(ik)^2 e^{ik\lambda}}{2!} & \cdots & \frac{(ik)^{d-1} e^{ik\lambda}}{(d-1)!} \\ & e^{ik\lambda} & ike^{ik\lambda} & \cdots & \frac{(ik)^{d-2} e^{ik\lambda}}{(d-2)!} \\ & & \ddots & \ddots & \vdots \\ & & & e^{ik\lambda} & ike^{ik\lambda} \\ & & & & e^{ik\lambda} \end{bmatrix} \\ &= \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} & \cdots & \frac{f^{(d-1)}(\lambda)}{(d-1)!} \\ & f(\lambda) & f'(\lambda) & \cdots & \frac{f^{(d-2)}(\lambda)}{(d-2)!} \\ & & \ddots & \ddots & \vdots \\ & & & f(\lambda) & f'(\lambda) \\ & & & & f(\lambda) \end{bmatrix} = f(J). \end{aligned}$$

Now let $X = W \operatorname{diag}(J_1, \ldots, J_r) W^{-1}$ be its Jordan decomposition with Jordan blocks $J_1, \ldots, J_r$. Then

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \sum_{k=-m}^{m} \widehat{f}(k) e^{ikX} &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \sum_{k=-m}^{m} \widehat{f}(k) W \begin{bmatrix} e^{ikJ_1} & & \\ & \ddots & \\ & & e^{ikJ_r} \end{bmatrix} W^{-1} \\ &= W \begin{bmatrix} f(J_1) & & \\ & \ddots & \\ & & f(J_r) \end{bmatrix} W^{-1} = f(X). \end{aligned}$$

□

Proposition 3.6 provides a way to remedy the Gibbs phenomenon for matrix Fourier series, which we will illustrate numerically in Sect. 8.1.

Before moving to our next method, we would like to point out that what may appear to be an innocuous change to a series could affect the value obtained using the summation methods in this article. For example, if we had added zeros to every third term of the Grandi's series $1-1+1-1+\cdots$ to obtain the series $1-1+0+1-1+0+\cdots$, its Cesàro sum decreases from $1/2$ to $1/3$.

3.2 Euler method

Euler summation methods are another class of sequential summation methods. Its name comes from the $(\mathsf{E}, 1)$-method for scalar series, which involves the Euler transform [3, 27]. Here we will extend Euler transform and Euler summation to matrices. Let $P \in \mathbb{C}^{d\times d}$ be a positive definite matrix. Emulating the calculation in [3, Sect. 8.2], for $A_\bullet = (A_k)_{k=0}^\infty \in s(\mathbb{C}^{d\times d})$,

$$\begin{aligned}\sum_{k=0}^\infty A_k &= \sum_{k=0}^\infty \big((I+P)^{-1}[I-P(I+P)^{-1}]^{-1}\big)^{k+1} A_k \\ &= \sum_{n=0}^\infty \sum_{k=0}^n \binom{n}{k} (I+P)^{-n-1} P^{n-k} A_k\end{aligned}$$

and thus we introduce the shorthand

$$\mathscr{E}_n^P(A_\bullet) := \sum_{k=0}^n \binom{n}{k} (I+P)^{-n-1} P^{n-k} A_k, \tag{3.8}$$

and call it the *P-Euler transform* of $\sum_{k=0}^\infty A_k$.

Definition 3.7 For $P \succ 0$, $A_\bullet \in s(\mathbb{C}^{d\times d})$, the matrix series $\sum_{k=0}^\infty A_k$ is *Euler summable* to $S \in \mathbb{C}^{d\times d}$ with respect to P or (E, P)*-summable* to S if

$$\sum_{n=0}^\infty \mathscr{E}_n^P(A_\bullet) = S.$$

We denote this by

$$\sum_{k=0}^\infty A_k \overset{(\mathsf{E},P)}{=\!=} S.$$

For the special case $P = \rho I$ where $\rho > 0$ is a scalar, we just write (E, ρ) instead of $(\mathsf{E}, \rho I)$.

Let $P \succ 0$. Then

$$
\begin{aligned}
&(I+P)^{n+1}\sum_{k=0}^{n}\mathscr{E}_k^P(A_\bullet)\\
&\quad=(I+P)^{n+1}\sum_{m=0}^{n}\mathscr{E}_m^P(A_\bullet)=\sum_{m=0}^{n}(I+P)^{n-m}\sum_{r=0}^{m}\binom{m}{r}P^{m-r}A_r\\
&\quad=\sum_{k=0}^{n}\sum_{r=0}^{k}\sum_{m=0}^{n}\binom{n-m}{k-r}\binom{m}{r}P^{n-k}A_r\\
&\quad=\sum_{k=0}^{n}\sum_{r=0}^{k}\binom{n+1}{k+1}P^{n-k}A_r=\sum_{k=0}^{n}\binom{n+1}{k+1}P^{n-k}S_k.
\end{aligned}
$$

Here we have used the Chu–Vandermonde's identity [30]: For any integers $0 \le r \le k \le n$,

$$
\sum_{m=0}^{n}\binom{n-m}{k-r}\binom{m}{r}=\binom{n+1}{k+1}.
$$

The Euler method is thus a sequential summation method (3.1) with a choice of

$$
C_{n,k}=\begin{cases}\binom{n+1}{k+1}P^{n-k}(I+P)^{-n-1} & \text{if } k\le n,\\ 0 & \text{if } k>n.\end{cases}
$$

Corollary 3.8 (Regularity of Euler summation) *For $A_\bullet \in s(\mathbb{C}^{d\times d})$ and $P, S \in \mathbb{C}^{d\times d}$ such that $P \succ 0$, if $\sum_{k=0}^{\infty}A_k = S$, then $\sum_{k=0}^{\infty}A_k \overset{(\mathrm{E},P)}{=\!=} S$.*

Proof This follows directly from Theorem 3.1, where the three conditions may be verified as follows: Since

$$
\sum_{k=0}^{\infty}\binom{n+1}{k+1}P^{n-k}(1+P)^{-n-1}=I-P^{n+1}(I+P)^{-n-1}\prec I
$$

and $\lim_{n\to\infty} I-P^{n+1}(I+P)^{-n-1}=I$, Conditions (i) and (iii) hold. Condition (ii) follows from

$$
\lim_{n\to\infty}\binom{n+1}{k+1}P^{n-k}(I+P)^{-n-1}=0.
$$

□

Euler summability depends highly on the choice of $P \succ 0$. The next result partially characterizes it for commuting P_1 and P_2 via Loewner order.

Theorem 3.9 *Let $A_\bullet \in s(\mathbb{C}^{d\times d})$ and $P_1, P_2 \in \mathbb{C}^{d\times d}$ be such that $P_2 \succ P_1 \succ 0$ and $P_1P_2 = P_2P_1$. If $\sum_{k=0}^{\infty}A_k \overset{(\mathrm{E},P_1)}{=\!=} S$, then $\sum_{k=0}^{\infty}A_k \overset{(\mathrm{E},P_2)}{=\!=} S$.*

Proof For any $P \in \mathbb{C}^{d\times d}$ with $P \succ 0$ and $P_1 P = P P_1$,

$$\sum_{n=0}^{m}\sum_{k=0}^{n}\binom{m}{n}\binom{n}{k}P^{m-n}(I+P)^{-m-1}P_1^{n-k}(I+P_1)^{-n-1}A_k$$
$$=\sum_{k=0}^{m}\binom{m}{k}(P_1+P+P_1P)^{m-k}(I+P_1+P+P_1P)^{-m-1}A_k = \mathscr{E}_m^{P_1+P+P_1P}(A). \tag{3.9}$$

Suppose $\sum_{k=0}^{\infty} A_k \overset{(\mathsf{E},P_1)}{=\!=} S$. Since $\sum_{n=0}^{\infty}\mathscr{E}_n^{P_1}(A) = S$, it is Euler summable for any $P \succ 0$. Set $P = (P_2 - P_1)(I+P_1)^{-1}$. By the regularity in Corollary 3.8, $\sum_{n=0}^{\infty}\mathscr{E}_n^{P_1}(A) \overset{(\mathsf{E},P)}{=\!=} S$, i.e.,

$$\sum_{m=0}^{\infty}\sum_{k=0}^{m}\binom{m}{k}P_2^{m-k}(I+P_2)^{-m-1}A_k = S.$$

Hence $\sum_{k=0}^{\infty} A_k \overset{(\mathsf{E},P_2)}{=\!=} S$. □

Equation (3.9) reveals the composition rule for Euler transforms: If P_1 and P_2 commutes, then the P_1-Euler transform of the P_2-Euler transform is the $(P_1 + P_2 + P_1P_2)$-Euler transform. To gain more insights, we apply it to the Neumann series.

Proposition 3.10 (Euler summability of Neumann series) *For $X \in \mathbb{C}^{d\times d}$, $P \succ 0$, and $PX = XP$,*

$$\sum_{k=0}^{\infty} X^k \overset{(\mathsf{E},P)}{=\!=} (I-X)^{-1}$$

if and only if $\lambda\big((I+P)^{-1}(P+X)\big) \subseteq \mathbb{D}$.

Proof By commutativity, the P-Euler transform of the Neumann series is

$$\mathscr{E}_n^P(X) = (I+P)^{-n-1}\sum_{k=0}^{n}\binom{n}{k}P^{n-k}X^k = (I+P)^{-n-1}(P+X)^n. \tag{3.10}$$

Therefore,

$$\sum_{n=0}^{\infty}\mathscr{E}_n^P(X) = (I+P)^{-1}\sum_{n=0}^{\infty}((I+P)^{-1}(P+X))^n,$$

which is conventionally summable to $(I-X)^{-1}$ if and only if $\lambda\big((I+P)^{-1}(P+X)\big) \subseteq \mathbb{D}$. □

The case of $P = \rho I$ for a scalar $\rho > 0$ is worth stating separately as they commute with all $X \in \mathbb{C}^{d\times d}$.

Corollary 3.11 *For $X \in \mathbb{C}^{d\times d}$, $\rho > 0$,*

$$\sum_{k=0}^{\infty} X^k \overset{(\mathsf{E},\rho)}{=\!=} (I - X)^{-1}$$

if and only if $\lambda(X) \subseteq \{z \in \mathbb{C} : |z + \rho| < 1 + \rho\}$.

Intuitively, choosing a "small" $P \in \mathbb{C}^{d\times d}$ ought to increase the rate of convergence. But it is difficult to obtain a universal relationship as the convergence rate invariably depends on the series. For scalar series, this is discussed in [31] and [32]. For matrix series, we will illustrate this numerically in Sect. 8.3.

Euler methods are generalized by the Borel methods. We will discuss their relationship in Sect. 4.3.

4 Functional summation methods

In sequential summation methods, we have a partial function $\mathsf{R} : s(\mathbb{C}^{d\times d}) \dashrightarrow c(\mathbb{C}^{d\times d})$ and the sum is the value of a limiting process as $n \to \infty$. In functional summation methods, we have a partial function $\mathsf{R} : s(\mathbb{C}^{d\times d}) \dashrightarrow c(C_b(\Omega)^{d\times d})$ and the sum is the value of *two* limiting processes—a sequential limit as $n \to \infty$ followed by a continuous limit as $x \to x_*$ in Ω.

We now lay out the details. Let $\Omega \subseteq \mathbb{R}$ and $x_* \in \overline{\Omega}$ or $x_* = \infty$. Let $\mathsf{R} : s(\mathbb{C}^{d\times d}) \dashrightarrow c(C_b(\Omega)^{d\times d})$ be a partial function defined by

$$\mathsf{R}(A_\bullet)(x) = \sum_{k=0}^{\infty} B_k(x) A_k \tag{4.1}$$

for some $B_k \in C_b(\Omega)^{d\times d}$, $k \in \mathbb{N}$. The R-sum is defined to be

$$S := \lim_{x\to x_*} \mathsf{R}(A_\bullet)(x)$$

if the limit exists, and in which case we write $\sum_{k=0}^{\infty} A_k \overset{\mathsf{R}}{=} S$. The careful reader might notice that while we wrote $\mathsf{R} : s(\mathbb{C}^{d\times d}) \dashrightarrow c(C_b(\Omega)^{d\times d})$, (4.1) seems to imply that $\mathsf{R} : s(\mathbb{C}^{d\times d}) \dashrightarrow C_b(\Omega)^{d\times d}$. The reason is that for a convergent series we do not distinguish between its sequence of partial sums in $c(C_b(\Omega)^{d\times d})$ and its limit in $C_b(\Omega)^{d\times d}$.

We begin with a sufficient condition for the regularity of functional summation methods (4.1). Unlike Theorem 3.1, the following result is not extended from any analogous result for scalar series but [3, Theorem 25] comes closest.

Theorem 4.1 *Let $\Omega \subseteq \mathbb{R}$, $x_* \in \overline{\Omega}$ or $x_* = \infty$, $B_k \in C_b(\Omega)^{d\times d}$, and $k \in \mathbb{N}$. Suppose*

(i) *there exists $\eta_0 > 0$ such that $\|B_0(x)\| \le \eta_0$ for all $x \in \Omega$;*
(ii) *$\lim_{x\to x_*} B_k(x) = I$ for each $k \in \mathbb{N}$;*
(iii) *there exists $\eta_1 > 0$ such that $\sum_{k=0}^{\infty} \|B_k(x) - B_{k+1}(x)\| < \eta_1$ for all $x \in \Omega$.*

Then for any $A_\bullet \in s(\mathbb{C}^{d\times d})$ *such that* $\sum_{k=0}^{\infty} A_k = S$, *the series* $\sum_{k=0}^{\infty} B_k(x)A_k$ *is summable in the conventional sense for each* $x \in \Omega$ *with*

$$\sum_{k=0}^{\infty} B_k(x)A_k \in C_b(\Omega)^{d\times d} \quad \textit{and} \quad \lim_{x\to x_*} \sum_{k=0}^{\infty} B_k(x)A_k = S. \tag{4.2}$$

Proof Let $S_n = \sum_{k=0}^{n} A_k$, $n \in \mathbb{N}$. Then

$$\sum_{k=0}^{n} B_k(x)A_k = \sum_{k=0}^{n-1} (B_k(x) - B_{k+1}(x))S_k + B_n(x)S_n.$$

By Conditions (i) and (iii), for each $x \in \Omega$ and $n \in \mathbb{N}$,

$$\lim_{k\to\infty} \|B_k(x) - B_{k+1}(x)\| = 0, \quad \|B_n(x)\| \le \|B_0(x)\| + \sum_{k=0}^{n-1} \|B_k(x) - B_{k+1}(x)\| \le \eta_0 + \eta_1.$$

So the sequence $(B_k(x))_{k=0}^{\infty}$ is convergent for each $x \in \Omega$ and

$$\Big\| \lim_{k\to\infty} B_k(x) \Big\| \le \eta_0 + \eta_1.$$

Thus there exists a bounded function $B(x)$ such that $\lim_{k\to\infty} B_k(x) = B(x)$ for each $x \in \Omega$.

We start with the left equality in (4.2). As $S_\bullet$ is convergent, it is bounded by some $\beta > 0$. By Condition (iii),

$$\begin{aligned}\sup_{x\in\Omega} \sum_{k=0}^{\infty} \|B_k(x)A_k\| &\le \sup_{x\in\Omega} \left[\sum_{k=0}^{\infty} \|B_k(x) - B_{k+1}(x)\| \cdot \|S_k\| + \|B(x)\| \cdot \|S\| \right] \\ &\le \sup_{x\in\Omega} \left[\sum_{k=0}^{\infty} \|B_k(x) - B_{k+1}(x)\| \cdot \|S_k\| \right] + \sup_{x\in\Omega} \|B(x)\| \cdot \|S\| < \infty.\end{aligned}$$

As absolute summability implies summability in a Banach space, we have $\sum_{k=0}^{\infty} B_k(x)A_k \in C_b(\Omega)^{d\times d}$.

For the limit in (4.2), assume first that $S = 0$. For $\varepsilon > 0$, choose $m \in \mathbb{N}$ sufficiently large so that $\|S_k\| < \varepsilon/2\eta_1$ for all $k > m$. By Condition (ii),

$$\lim_{x\to x_*} \sum_{k=0}^{m} (B_k(x) - B_{k+1}(x))S_k = 0.$$

By Condition (iii),

$$\Big\| \sum_{k\ge m} (B_k(x) - B_{k+1}(x))S_k \Big\| \le \frac{\varepsilon}{2\eta_1} \sum_{k\ge m} \|B_k(x) - B_{k+1}(x)\| \le \frac{\varepsilon}{2}.$$

Therefore,

$$\lim_{x\to x_*}\sum_{k=0}^{\infty}B_k(x)A_k=\lim_{x\to x_*}\sum_{k=0}^{\infty}(B_k(x)-B_{k+1}(x))S_k+B(x)S=0.$$

For $S\neq 0$, we just apply the same argument to $A'_\bullet=(A'_k)_{k=0}^{\infty}$ with

$$A'_k:=\begin{cases}A_0-S & k=0,\\ A_k & k=1,2,\ldots.\end{cases}$$

□

Theorem 4.1 may be interpreted as follows: A functional summation method (4.1) that satisfies Condition (ii) is a perturbation of the original series. If the perturbation functions B_k's are uniformly bounded, i.e., Condition (i) holds, and if the changes are small enough at each step in the sense of Condition (iii), then we have regularity. We will use Theorem 4.1 to establish regularity of two powerful matrix summation methods.

4.1 Abelian means

The scalar version of this class of summation methods gained its name from the Abel summation method, which contains the well-known Abel's Theorem for power series as a special case [33]. We will generalize it to matrix series.

Definition 4.2 Let $P_\bullet:=(P_k)_{k=0}^{\infty}$ be an unbounded sequence of positive definite matrices strictly increasing in the Loewner order, i.e., $0\prec P_0\prec P_1\prec\cdots$, and $\lim_{k\to\infty}\|P_k\|=\infty$. For $A_\bullet=(A_k)_{k=0}^{\infty}\in s(\mathbb{C}^{d\times d})$, the series $\sum_{k=0}^{\infty}A_k$ is *summable in Abelian means* to $S\in\mathbb{C}^{d\times d}$ with respect to $P_\bullet$ if $\sum_{k=0}^{\infty}A_ke^{-P_kx}$ is conventionally summable for all $x\in(0,\infty)$ and

$$\lim_{x\to 0}\sum_{k=0}^{\infty}A_ke^{-P_kx}=S.$$

We denote this by

$$\sum_{k=0}^{\infty}A_k\overset{(\mathrm{A},P_\bullet)}{=\!=}S.$$

Corollary 4.3 (Regularity of Abelian means) *Let* $P_\bullet:=(P_k)_{k=0}^{\infty}$ *be such that* $0\prec P_0\prec P_1\prec\cdots$ *and* $\lim_{k\to\infty}\|P_k\|=\infty$. *For any* $A_\bullet\in s(\mathbb{C}^{d\times d})$, *if* $\sum_{k=0}^{\infty}A_k=S$, *then* $\sum_{k=0}^{\infty}A_k\overset{(\mathrm{A},P_\bullet)}{=\!=}S$.

Proof Summation by Abelian means with respect to $P_\bullet$ is of the form (4.1). So we just need to check the conditions of Theorem 4.1. For any $x \in (0, \infty)$,

$$\sum_{k=0}^{\infty} \left(e^{-P_k x} - e^{-P_{k+1} x}\right) = e^{-P_0 x} - \lim_{k\to\infty} e^{-P_k x} = e^{-P_0 x} \preceq I.$$

So Conditions (i) and (iii) are satisfied. Condition (ii) is also satisfied as $\lim_{x\to 0^+} e^{-P_k x} = I$ for any $k \in \mathbb{N}$. □

The special case $P_\bullet = (0, I, 2I, \dots)$, which reduces to a power series under the change-of-variable $t = e^{-x}$, gives us the matrix analogue of the well-known Abel summability.

Definition 4.4 For $A_\bullet = (A_k)_{k=0}^{\infty} \in s(\mathbb{C}^{d\times d})$, the series $\sum_{k=0}^{\infty} A_k$ is *Abel summable* to $S \in \mathbb{C}^{d\times d}$ if $\sum_{k=0}^{\infty} A_k x^k$ is conventionally summable for all $x \in (0, 1)$ and

$$\lim_{x\to 1^-} \sum_{k=0}^{\infty} A_k x^k = S.$$

We denote this by

$$\sum_{k=0}^{\infty} A_k \overset{\mathsf{A}}{=} S.$$

We will see next that Abel summability is implied by Cesàro summability.

Theorem 4.5 *For $A_\bullet \in s(\mathbb{C}^{d\times d})$ and $S \in \mathbb{C}^{d\times d}$, if $\sum_{k=0}^{\infty} A_k \overset{\mathsf{C}}{=} S$, then $\sum_{k=0}^{\infty} A_k \overset{\mathsf{A}}{=} S$.*

Proof Suppose $\sum_{k=0}^{\infty} A_k \overset{\mathsf{C}}{=} S$. For any $n \in \mathbb{N}$, let $S_n = \sum_{k=0}^{n} A_k$ and

$$\Sigma_n = \frac{1}{n} \sum_{k=0}^{n-1} S_k.$$

For any $x \in (0, 1)$, we have $S_n = (n+1)\Sigma_{n+1} - n\Sigma_n$ and

$$\sum_{k=0}^{n} A_k x^k = S_n x^n + \sum_{k=0}^{n-1} S_k x^k (1 - x). \tag{4.3}$$

Since $\lim_{n\to\infty} nx^n = 0$ and $\lim_{n\to\infty} \Sigma_n = S$, we get $\lim_{n\to\infty} S_n x^n = 0$. By (4.3),

$$\begin{aligned}\left\|\sum_{k=0}^{\infty} A_k x^k - S\right\| &= \left\|\sum_{k=0}^{\infty} S_k x^k(1-x) - \frac{S}{1-x}(1-x)\right\| \\ &= \left\|\sum_{k=0}^{\infty} S_k x^k(1-x) - \sum_{k=0}^{\infty} S x^k(1-x)\right\| \\ &= \left\|\sum_{k=0}^{\infty} (S_k - S)x^k(1-x)\right\| \le \sum_{k=0}^{\infty} \|S_k - S\| x^k(1-x).\end{aligned}$$

For $x \in (0,1)$, $\varepsilon > 0$, choose m sufficiently large such that if $k, p, q > m$,

$$\|\Sigma_k - S\| < \frac{(1-x)}{2}\varepsilon \quad \text{and} \quad \|\Sigma_p - \Sigma_q\| < \frac{(1-x)}{2}\varepsilon.$$

Then

$$\begin{aligned}\sum_{k=0}^{\infty} \|S_k - S\| x^k(1-x) &= (1-x)\left[\sum_{k=0}^{m} x^k \|S_k - S\| + \sum_{k=m+1}^{\infty} x^k \|S_k - S\|\right] \\ &= (1-x)\left[\sum_{k=0}^{m} x^k \|S_k - S\| + \sum_{k=m+1}^{\infty} x^k \|k(\Sigma_{k+1} - \Sigma_k) + (\Sigma_{k+1} - S)\|\right] \\ &\le (1-x)\left[\sum_{k=0}^{m} x^k \|S_k - S\| + \sum_{k=m+1}^{\infty} x^k \Big(k\|\Sigma_{k+1} - \Sigma_k\| + \|\Sigma_{k+1} - S\|\Big)\right] \\ &< (1-x)\left[\sum_{k=0}^{m} x^k \|S_k - S\| + \sum_{k=m+1}^{\infty} x^k (k+1)(1-x)\varepsilon\right] \\ &= (1-x)\sum_{k=0}^{m} x^k \|S_k - S\| + \varepsilon(1-x)^2 \sum_{k=m+1}^{\infty} (k+1)x^k \\ &< (1-x)\sum_{k=0}^{m} x^k \|S_k - S\| + \varepsilon.\end{aligned}$$

Taking limit $x \to 1^-$, we deduce the required Abel summability. □

Again, we will use the Neumann series as a test case. From the perspective of summing the Neumann series, Abel summation is the "right" generalization of Cesàro summation in that it overcomes the difficulty associated with Jordan blocks of size greater than 1, which we discussed after Proposition 3.5.

Lemma 4.6 *Let $J_\lambda \in \mathbb{C}^{d\times d}$ be a Jordan block with eigenvalue λ. Then*

$$\sum_{k=1}^{\infty} J_\lambda^k \overset{\text{A}}{=} (I - J_\lambda)^{-1}$$

if and only if $\lambda \in \overline{\mathbb{D}} \setminus \{1\}$.

Proof If $|\lambda| > 1$, then $\sum_{k=0}^{\infty} J_\lambda^k x^k$ is not summable for $x > 1/\lambda$, so the series is not Abel summable. If $\lambda = 1$, then $\lim_{x\to 1^-} \sum_{k=0}^{\infty} J_\lambda^k x^k$ does not exist, so the series is not Abel summable either.

For the converse, suppose $\lambda \in \overline{\mathbb{D}} \setminus \{1\}$. Let $0 < x < 1$. By (3.6), for $i, j = 1, \ldots, d$ with $j \geq i$, the (i, j)th entry of the matrix

$$\left(\sum_{k=0}^{\infty} J_\lambda^k x^k\right)_{ij} = x^{j-i} \sum_{k=0}^{\infty} \binom{k+j-i}{j-i} (\lambda x)^k = \frac{x^{j-i}}{(1-\lambda x)^{j-i+1}}.$$

Therefore,

$$\sum_{k=1}^{\infty} J_\lambda^k \overset{\mathsf{A}}{=} \begin{bmatrix} \frac{1}{1-\lambda} & \frac{1}{(1-\lambda)^2} & \frac{1}{(1-\lambda)^3} & \cdots & \frac{1}{(1-\lambda)^{d-1}} \\ & \frac{1}{1-\lambda} & \frac{1}{(1-\lambda)^2} & \cdots & \frac{1}{(1-\lambda)^{d-2}} \\ & & \ddots & \ddots & \vdots \\ & & & \frac{1}{1-\lambda} & \frac{1}{(1-\lambda)^2} \\ & & & & \frac{1}{1-\lambda} \end{bmatrix} = (I - J_\lambda)^{-1}.$$

□

We then apply the lemma to a Jordan decomposition to deduce what we are after.

Corollary 4.7 (Abel summability of Neumann series) *For $X \in \mathbb{C}^{d\times d}$,*

$$\sum_{k=0}^{\infty} X^k \overset{\mathsf{A}}{=} (I - X)^{-1}$$

if and only if $\lambda(X) \subseteq \overline{\mathbb{D}} \setminus \{1\}$.

Proof Let $\lambda(X) = \{\lambda_1, \ldots, \lambda_r\} \subseteq \overline{\mathbb{D}} \setminus \{1\}$ counting multiplicities and $X = WJW^{-1}$ be a Jordan decomposition with $J = \operatorname{diag}(J_{\lambda_1}, \ldots, J_{\lambda_r})$ and J_{λ_i} the Jordan block corresponding to λ_i, $i = 1, \ldots, r$. Then

$$\begin{aligned} \sum_{k=0}^{\infty} X^k &= \sum_{k=0}^{\infty} W \begin{bmatrix} J_{\lambda_1}^k & & \\ & \ddots & \\ & & J_{\lambda_r}^k \end{bmatrix} W^{-1} \\ &\overset{\mathsf{A}}{=} W \begin{bmatrix} (I - J_{\lambda_1})^{-1} & & \\ & \ddots & \\ & & (I - J_{\lambda_r})^{-1} \end{bmatrix} W^{-1} = (I - X)^{-1}. \end{aligned}$$

For the converse, suppose without loss of generality that $\lambda_1 \notin \overline{\mathbb{D}} \setminus \{1\}$. Then

$$\sum_{k=0}^{\infty} W \begin{bmatrix} J_{\lambda_1}^k & & \\ & \ddots & \\ & & J_{\lambda_r}^k \end{bmatrix} W^{-1}$$

is not Abel summable as the first Jordan subblock is not Abel summable. □

4.2 Lambert method

The original Lambert summation method, named after Johann Heinrich Lambert, played an important role in number theory [8, 27, 34], and is a particularly potent tool for summing Dirichlet series, as we will see in Sect. 8.5. Here we will generalize Lambert summation to matrix series.

Definition 4.8 For $A_\bullet = (A_k)_{k=0}^{\infty} \in s(\mathbb{C}^{d\times d})$, the series $\sum_{k=1}^{\infty} A_k$ is *Lambert summable* to $S \in \mathbb{C}^{d\times d}$ if $\sum_{k=1}^{\infty} kA_k x^k/(1 + x + \cdots + x^{k-1})$ is conventionally summable for every $x \in (0, 1)$ and

$$\lim_{x\to 1^-} (1-x) \sum_{k=1}^{\infty} \frac{kx^k}{1-x^k} A_k = S.$$

We denote this by

$$\sum_{k=1}^{\infty} A_k \overset{\text{L}}{=} S.$$

Corollary 4.9 (Regularity of Lambert summation) *For $A_\bullet \in s(\mathbb{C}^{d\times d})$ and $S \in \mathbb{C}^{d\times d}$, if $\sum_{k=0}^{\infty} A_k = S$, then $\sum_{k=0}^{\infty} A_k \overset{\text{L}}{=} S$.*

Proof Lambert summation is of the form (4.1), so we check the conditions of Theorem 4.1. Since $|x| < 1$ for $x \in (0, 1)$, Condition (i) is satisfied. For each $k \in \mathbb{N}$,

$$\lim_{x\to 1^-} \frac{kx^k}{1 + x + \cdots + x^{k-1}} = 1,$$

so Condition (ii) is satisfied. Condition (iii) is also satisfied as for any $x \in (0, 1)$,

$$\sum_{k=0}^{\infty} (1-x) \left| \frac{kx^k}{1-x^k} - \frac{(k+1)x^{k+1}}{1-x^{k+1}} \right| = x \le 1.$$

□

4.3 Borel and Mittag-Leffler methods

The scalar versions of Borel summation methods, named after Émile Borel [35], have important applications in physics [5, 9]. They come in two variants (weak and strong) and we will extend them to matrix series.

Definition 4.10 For $A_\bullet = (A_k)_{k=0}^\infty \in s(\mathbb{C}^{d\times d})$, its *weak Borel transform* is

$$\mathscr{WB}(A_\bullet)(x) = \sum_{k=0}^\infty S_k \frac{x^k}{k!}$$

for all $x > 0$, where $S_n = \sum_{k=0}^n A_k$. The series $\sum_{k=0}^\infty A_k$ is *weakly Borel summable* to $S \in \mathbb{C}^{d\times d}$ if $e^{-x}\,\mathscr{WB}(A_\bullet)(x)$ is conventionally summable for all $x > 0$ and

$$\lim_{x\to\infty} e^{-x}\,\mathscr{WB}(A_\bullet)(x) = S.$$

We denote this by

$$\sum_{k=0}^\infty A_k \overset{\mathsf{WB}}{=} S.$$

Theorem 4.11 (Regularity of weak Borel summation) *For* $A_\bullet \in s(\mathbb{C}^{d\times d})$, *if* $\sum_{k=0}^\infty A_k = S$, *then* $\sum_{k=0}^\infty A_k \overset{\mathsf{WB}}{=} S$.

Proof Let $S_n = \sum_{k=0}^n A_k$, $n \in \mathbb{N}$. As $S_\bullet$ is a convergent sequence, it is bounded by some $\beta > 0$. For each $x > 0$,

$$\|e^{-x}\,\mathscr{WB}(A_\bullet)(x)\| = \left\| e^{-x} \sum_{k=0}^\infty \frac{x^k}{k!} S_k \right\| \le \beta e^{-x} \sum_{k=0}^\infty \frac{x^k}{k!} = \beta.$$

Hence $e^{-x}\sum_{k=0}^\infty S_k x^k/k!$ is conventionally summable for all $x > 0$. Moreover,

$$\begin{aligned}\Big\| \lim_{x\to\infty} e^{-x}\,\mathscr{WB}(A_\bullet)(x) - S \Big\| &= \left\| \lim_{x\to\infty} e^{-x} \sum_{k=0}^\infty \frac{x^k}{k!} S_k - \lim_{x\to\infty} e^{-x} \sum_{k=0}^\infty \frac{x^k}{k!} S \right\| \\ &\le \lim_{x\to\infty} e^{-x} \sum_{k=0}^\infty \frac{x^k}{k!} \|S_k - S\| = 0.\end{aligned}$$

□

Definition 4.12 For $A_\bullet = (A_k)_{k=0}^\infty \in s(\mathbb{C}^{d\times d})$, its *strong Borel transform* is

$$\mathscr{SB}(A_\bullet)(x) := \sum_{k=0}^\infty A_k \frac{x^k}{k!}$$

for $x > 0$. The series $\sum_{k=0}^\infty A_k$ is *strongly Borel summable* to $S \in \mathbb{C}^{d\times d}$ if

$$\int_0^x e^{-t}\,\mathscr{SB}(A_\bullet)(t)\,dt$$

is conventionally summable for all $x > 0$ and

$$\int_0^\infty e^{-x}\, \mathscr{S}\mathscr{B}(A_\bullet)(x)\, dx = S. \tag{4.4}$$

We denote this by

$$\sum_{k=0}^{\infty} A_k \overset{\text{SB}}{=} S.$$

It turns out that, for a series of scalars, the strong Borel method is a special case of the Mittag-Leffler summation. We will show that the same is true for a series of matrices with the following matrix generalization of the latter.

Definition 4.13 For $A_\bullet = (A_k)_{k=0}^\infty \in s(\mathbb{C}^{d\times d})$, the series $\sum_{k=0}^\infty A_k$ is *Mittag-Leffler summable* to $S \in \mathbb{C}^{d\times d}$ with respect to $\alpha > 0$ if

$$\int_0^x e^{-t} \sum_{k=0}^{\infty} \frac{A_k t^{\alpha k}}{\Gamma(1+\alpha k)}\, dt$$

is conventionally summable for all $x > 0$ and

$$\int_0^\infty e^{-x} \sum_{k=0}^{\infty} \frac{A_k x^{\alpha k}}{\Gamma(1+\alpha k)}\, dx = S.$$

We denote this by

$$\sum_{k=0}^{\infty} A_k \overset{\text{M}}{=} S.$$

The implicit choice of α is not reflected in the notation. Here $\Gamma(x) := \int_0^\infty t^{x-1}e^{-t}\, dt$ is the Gamma function.

If we set $\alpha = 1$ above, we obtain the strong Borel method.

Theorem 4.14 (Regularity of Mittag-Leffler summation) *For $\alpha > 0$ and $A_\bullet \in s(\mathbb{C}^{d\times d})$, if $\sum_{k=0}^\infty A_k = S$, then $\sum_{k=0}^\infty A_k \overset{\text{M}}{=} S$. In particular, if $\sum_{k=0}^\infty A_k = S$, then $\sum_{k=0}^\infty A_k \overset{\text{SB}}{=} S$.*

Proof This follows from

$$S = \sum_{k=0}^{\infty} A_k = \sum_{k=0}^{\infty} \left(\int_0^\infty e^{-x} x^{\alpha k}\, dx \right) \frac{A_k}{\Gamma(1+\alpha k)} = \int_0^\infty e^{-x} \sum_{k=0}^{\infty} \frac{A_k x^{\alpha k}}{\Gamma(1+\alpha k)}\, dx.$$

□

We will next justify the 'weak' and 'strong' designations and see when they are equivalent [5, 36], generalizing [3, Theorem 123] to matrices.

Theorem 4.15 *Let* $A_\bullet = (A_k)_{k=0}^\infty \in s(\mathbb{C}^{d\times d})$. *If* $\sum_{k=0}^\infty A_k \overset{\text{WB}}{=} S$, *then* $\sum_{k=0}^\infty A_k \overset{\text{SB}}{=} S$. *The converse holds if and only if* $\lim_{x\to\infty} e^{-x}\, \mathscr{S}\mathscr{B}(A_\bullet)(x) = 0$.

For easy reference, we reproduce two lemmas used in the proof of [3, Theorem 122].

Lemma 4.16 (Hardy) *Let* $f : \mathbb{R} \to \mathbb{C}$ *be differentiable. If* $\lim_{x\to\infty} f(x) + f'(x) = 0$, *then* $\lim_{x\to\infty} f(x) = 0$.

Lemma 4.17 (Hardy) *Let* $a_\bullet \in s(\mathbb{C})$. *The series* $\mathscr{W}\mathscr{B}(a_\bullet)(x)$ *is conventionally summable for all* $x > 0$ *if and only if* $\mathscr{S}\mathscr{B}(a_\bullet)(x)$ *is conventionally summable for all* $x > 0$.

Proof of Thoerem 4.15 Suppose $\sum_{k=0}^\infty A_k \overset{\text{WB}}{=} S$. Then $\mathscr{W}\mathscr{B}(A_\bullet)(x)$ is conventionally summable for all $x > 0$. Applying Lemma 4.17 entrywise shows that $\mathscr{S}\mathscr{B}(A_\bullet)(x)$ is conventionally summable for all $x > 0$. Taking derivative,

$$\mathscr{S}\mathscr{B}(A_\bullet)'(x) = \sum_{k=0}^\infty A_{k+1}\frac{x^k}{k!} \quad \text{and} \quad \mathscr{W}\mathscr{B}(A_\bullet)'(x) = \sum_{k=0}^\infty S_{k+1}\frac{x^k}{k!}.$$

Therefore,

$$\begin{aligned} e^{-x}\,\mathscr{W}\mathscr{B}(A_\bullet)(x) - A_0 &= \int_0^x \frac{d}{dt}\big(e^{-t}\,\mathscr{W}\mathscr{B}(A_\bullet)(t)\big)\,dt \\ &= \int_0^x e^{-t}\big(\mathscr{W}\mathscr{B}(A_\bullet)'(t) - \mathscr{W}\mathscr{B}(A_\bullet)(t)\big)\,dt \\ &= \int_0^x \sum_{k=0}^\infty (S_{k+1} - S_k)\frac{e^{-t}t^k}{k!}\,dt = \int_0^x \sum_{k=0}^\infty A_{k+1}\frac{e^{-t}t^k}{k!}\,dt \\ &= \int_0^x e^{-t}\,\mathscr{S}\mathscr{B}(A_\bullet)'(t)\,dt \\ &= e^{-x}\,\mathscr{S}\mathscr{B}(A_\bullet)(x) - A_0 + \int_0^x e^{-t}\,\mathscr{S}\mathscr{B}(A_\bullet)(t)\,dt. \end{aligned}$$

Rearranging terms,

$$e^{-x}\,\mathscr{W}\mathscr{B}(A_\bullet)(x) = \int_0^x e^{-t}\,\mathscr{S}\mathscr{B}(A_\bullet)(t)\,dt + e^{-x}\,\mathscr{S}\mathscr{B}(A_\bullet)(x). \tag{4.5}$$

Taking limit $x \to \infty$, the left-hand side gives the weak Borel sum while the first term on the right-hand side gives the strong Borel sum. Since

$$\begin{aligned} \lim_{x\to\infty} e^{-x}\,\mathscr{W}\mathscr{B}(A_\bullet)(x) &= \lim_{x\to\infty}\left(\int_0^x e^{-t}\,\mathscr{S}\mathscr{B}(A_\bullet)(t)\,dt + e^{-x}\,\mathscr{S}\mathscr{B}(A_\bullet)(x)\right) \\ &= \lim_{x\to\infty}\left(\int_0^x e^{-t}\,\mathscr{S}\mathscr{B}(A_\bullet)(t)\,dt + \frac{d}{dx}\int_0^x e^{-t}\,\mathscr{S}\mathscr{B}(A_\bullet)(t)\,dt\right) = S, \end{aligned}$$

we may apply Lemma 4.16 entrywise so that

$$\lim_{x\to\infty} e^{-x}\,\mathscr{S}\mathscr{B}(A_\bullet)(x) = 0 \quad \text{and} \quad \lim_{x\to\infty}\int_0^x e^{-t}\,\mathscr{S}\mathscr{B}(A_\bullet)(t)\,dt = S.$$

Suppose $\sum_{k=0}^{\infty} A_k \overset{\mathsf{SB}}{=} S$. Then $\mathscr{S}\mathscr{B}(A_\bullet)(x)$ is conventionally summable for all $x > 0$ and so is $\mathscr{W}\mathscr{B}(A_\bullet)(x)$ by applying Lemma 4.17 entrywise. By (4.5),

$$\begin{aligned}\lim_{x\to\infty} e^{-x}\,\mathscr{W}\mathscr{B}(A_\bullet)(x) &= \lim_{x\to\infty}\int_0^x e^{-t}\,\mathscr{S}\mathscr{B}(A_\bullet)(t)\,dt + \lim_{x\to\infty} e^{-x}\,\mathscr{S}\mathscr{B}(A_\bullet)(x)\\ &= S + \lim_{x\to\infty} e^{-x}\,\mathscr{S}\mathscr{B}(A_\bullet)(x).\end{aligned}$$

Hence $\sum_{k=0}^{\infty} A_k \overset{\mathsf{WB}}{=} S$ if and only if $\lim_{x\to\infty} e^{-x}\,\mathscr{S}\mathscr{B}(A_\bullet)(x) = 0$. □

As we mentioned at the end of Sect. 3.2, both Borel methods generalize the Euler methods.

Theorem 4.18 *Let $A_\bullet = (A_k)_{k=0}^{\infty} \in s(\mathbb{C}^{d\times d})$ and $P \in \mathbb{C}^{d\times d}$ be such that $P \succ 0$. If $\sum_{k=0}^{\infty} A_k \overset{(\mathsf{E},P)}{=\!=} S$, then $\sum_{k=0}^{\infty} A_k \overset{\mathsf{WB}}{=} S$ and $\sum_{k=0}^{\infty} A_k \overset{\mathsf{SB}}{=} S$.*

Proof Let $S_n = \sum_{k=0}^{n} A_k, n \in \mathbb{N}$. By definition of Euler summability, $\sum_{k=0}^{\infty} A_k \overset{(\mathsf{E},P)}{=\!=} S$ if and only if $\lim_{n\to\infty} Z_n = S$ where

$$Z_n = \sum_{k=0}^{n} \mathscr{E}_k^P(A_\bullet) = (I+P)^{-n-1}\sum_{k=0}^{n}\binom{n+1}{k+1}P^{n-k}S_k.$$

Then

$$\begin{aligned}e^{Px}\sum_{k=0}^{\infty} S_k\frac{x^k}{k!} &= \left[\sum_{k=0}^{\infty}\frac{(Px)^k}{k!}\right]\left[\sum_{k=0}^{\infty} S_k\frac{x^k}{k!}\right]\\ &= \sum_{k=0}^{\infty}\left[\frac{S_k}{k!} + \frac{PS_{k-1}}{(k-1)!} + \frac{P^2S_{k-2}}{(k-2)!2!} + \cdots + \frac{P^kS_0}{k!}\right]x^k\\ &= \sum_{k=0}^{\infty}\frac{(I+P)^kx^k}{k!}Z_k.\end{aligned}$$

Thus,

$$e^{-x}\sum_{k=0}^{\infty}\frac{x^k}{k!}S_k = e^{-(I+P)x}\sum_{k=0}^{\infty}\frac{(I+P)^kx^k}{k!}Z_k. \tag{4.6}$$

Weak Borel summability follows as

$$\Big\| \lim_{x\to\infty} e^{-x} \sum_{k=0}^{\infty} \frac{x^k}{k!} S_k - S \Big\| = \Big\| \lim_{x\to\infty} e^{-(I+P)x} \sum_{k=0}^{\infty} \frac{(I+P)^k x^k}{k!} Z_k - \lim_{x\to\infty} e^{-(I+P)x} \sum_{k=0}^{\infty} \frac{(I+P)^k x^k}{k!} S \Big\|$$
$$\leq \lim_{x\to\infty} e^{-(I+P)x} \sum_{k=0}^{\infty} \frac{(I+P)^k x^k}{k!} \|Z_k - S\| = 0.$$

By Theorem 4.15, we obtain $\sum_{k=0}^{\infty} A_k \overset{\mathsf{SB}}{=} S$. □

As before we will use the Neumann series as our basic test case. The proof below sheds further light on how Borel summation generalize Euler summation.

Proposition 4.19 (Borel summability of Neumann series) *For $X \in \mathbb{C}^{d\times d}$, the following are equivalent:*

(i) $\sum_{k=0}^{\infty} X^k \overset{\mathsf{WB}}{=} (I-X)^{-1}$;
(ii) $\sum_{k=0}^{\infty} X^k \overset{\mathsf{SB}}{=} (I-X)^{-1}$;
(iii) $\lambda(X) \subseteq \{z \in \mathbb{C} : \mathrm{Re}(\lambda) < 1\}$.

Proof Let $\lambda(X) = \{\lambda_1, \ldots, \lambda_r\}$. We show that (ii) and (iii) are equivalent. If $1 \in \lambda(X)$, then the integral

$$\int_0^{\infty} e^{-x} \sum_{k=0}^{\infty} \frac{(xX)^k}{k!}\, dx$$

is divergent, so the Neumann series $\sum_{k=0}^{\infty} X^k$ is not strongly Borel summable. If $1 \notin \lambda(X)$, then

$$\int_0^{\infty} e^{-x} \sum_{k=0}^{\infty} \frac{(xX)^k}{k!}\, dx = \int_0^{\infty} e^{x(X-I)}\, dx = (X-I)^{-1} \Big(\lim_{x\to\infty} e^{x(X-I)} - I \Big), \quad (4.7)$$

so the Neumann series is strongly Borel summable to $(I-X)^{-1}$ if and only if

$$\lim_{n\to\infty} e^{n(X-I)} = 0. \quad (4.8)$$

As $\lim_{n\to\infty} A^n = 0$ if and only if $\lambda(A) \subseteq \mathbb{D}$; $|e^{\lambda-1}| < 1$ if and only if $\mathrm{Re}(\lambda) < 1$; and $\lambda(e^{X-I}) = \{e^{\lambda_1-1}, \ldots, e^{\lambda_r-1}\}$; we deduce that (4.8) holds if and only if (iii) holds.

We next show that (i) and (iii) are equivalent. The geometric series $\sum_{k=0}^{\infty} \lambda^k$ is not weakly Borel summable at $\lambda = 1$ as

$$\lim_{x\to\infty} e^{-x} \sum_{k=0}^{\infty} \frac{x^k}{(k-1)!} = \lim_{x\to\infty} e^{-x} \frac{d}{dx}(xe^x) = \lim_{x\to\infty} 1 + x = +\infty.$$

So the Neumann series is not weakly Borel summable if $1 \in \lambda(X)$. If $1 \notin \lambda(X)$, then

$$\begin{aligned}\lim_{x\to\infty} e^{-x}\sum_{k=0}^{\infty}(I-X)^{-1}(I-X^{k+1})\frac{x^k}{k!} &= \lim_{x\to\infty} e^{-x}(I-X)^{-1}\sum_{k=0}^{\infty}(I-X^{k+1})\frac{x^k}{k!}\\ &= \lim_{x\to\infty} e^{-x}(I-X)^{-1}(e^{x}I - Xe^{xX})\\ &= (I-X)^{-1} - \lim_{x\to\infty} Xe^{x(X-I)}.\end{aligned}$$

As in the case of (4.8), the last limit is zero if and only if (iii) holds. □

In this context, Borel summation may be viewed as a limiting case of Euler summation as $\rho \to \infty$: By Corollary 3.11, $\sum_{k=0}^{\infty} X^k \overset{(\mathsf{E},\rho)}{=\!=} (I-X)^{-1}$ if and only if $\lambda(X) \subseteq \{z \in \mathbb{C} : |z+\rho| < 1+\rho\}$; and

$$\bigcup_{\rho>0}\{z \in \mathbb{C} : |z+\rho| < 1+\rho\} = \{z \in \mathbb{C} : \mathrm{Re}(\lambda) < 1\}.$$

5 From theory to computations

In principle, every summation method discussed in Sects. 3 and 4 yields a numerical method for summing a matrix series. But in reality issues related to rounding errors will play an important role and have to be carefully treated. We will discuss these in the next two sections after making some observations.

The theoretical results in Sects. 3 and 4 are all about convergence (whether a method converges or diverges) but say nothing about the *rate* of convergence. Readers familiar with the matrix functions literature [15] may think that convergence rates should be readily obtainable but this is an illusion—the matrix series appearing in the matrix functions literature are invariably *power* or *Taylor series*, whereas the series appearing in Sects. 3 and 4 can be any arbitrary matrix series—Fourier, Dirichlet, Hadamard powers, etc.

For special matrix series whose kth term takes a simple fixed form like X^k, $\sin(kX)$, $\exp(X \log k)$, $X^{\circ k}$, there is some hope of deriving a 'remainder' that gives the convergence rate but even that may be a difficult undertaking. For an arbitrary matrix series $\sum_{k=0}^{\infty} A_k$, where the kth term can be any matrix $A_k \in \mathbb{C}^{d\times d}$, such 'remainder' do not generally exist even when it is a scalar series [3, 5–7].

To sum an arbitrary matrix series of complete generality, we may thus only assume that the truncated sum $\sum_{k=0}^{n} A_k \approx \sum_{k=0}^{\infty} A_k$ is ascertained to be a good approximation through some other means and that n is given as part of the input—there is no 'remainder' that allows one to estimate n a priori. We will have more to say about this issue in Sect. 6, where we will also discuss matrix adaptations of compensated summation [37], block and mixed block summations [14], methods that were originally developed for sums of scalars.

The special case of Taylor or power series, i.e., where $A_k = c_k A^k$, deserves special attention because of their central role in matrix functions [15]. In Sect. 7, we discuss

how one may adapt the Padé approximation [38] and Schur–Parlett [39] algorithms to work with any of the regular sequential summation methods R in Sect. 3, i.e., compute the R-sum $S \overset{\mathsf{R}}{=} \sum_{k=0}^{\infty} c_k A^k$ for a given $A \in \mathbb{C}^{d\times d}$ using these algorithms.

Henceforth we assume the standard model for floating-point arithmetic [13, Section 2.2]:

$$\mathrm{fl}(a * b) = (a * b)(1 + \delta), \quad |\delta| \leq \mathsf{u}, \quad * \in \{+, -, \times, \div\},$$

with $\mathrm{fl}(a)$ the computed value of $a \in \mathbb{R}$ in floating-point arithmetic and u the unit roundoff. For any computations in floating-point arithmetic involving more than a single operation, we denote by $\widehat{S}$ the final computed output of a quantity S. This is to avoid having to write, say, $\mathrm{fl}(a + \mathrm{fl}(b + \mathrm{fl}(c + d)))$, for the output of $a + b + c + d$, unless it is strictly necessary (like in Algorithm 2).

6 Accurate and fast numerical summation

In this section there will be no loss of generality in restricting our discussions to $\mathbb{R}$, since complex addition is performed separately for real and imaginary parts as real additions. As we alluded to in Sect. 5, for a general matrix series $S = \sum_{k=0}^{\infty} A_k$ where A_k has no special form, computing it means to approximate S up to some desired ε-accuracy by a partial sum $S_n = \sum_{k=0}^{n} A_k$, i.e., with $\|S_n - S\| < \varepsilon$. There are two considerations in choosing $n \in \mathbb{N}$.

Firstly, for a given $\varepsilon > 0$, the value of n depends on the summation method we choose. This is in fact an important motivation for the summation methods in Sects. 3 and 4, namely, they often require a smaller n to achieve the same ε-accuracy. For example, take any scalar alternating series $\sum_{k=0}^{\infty} a_k = s$ that is conventionally summable; it is known [32] that for

$$\left| s - \sum_{k=0}^{n_1} a_k \right| < \varepsilon, \qquad \left| s - \sum_{k=0}^{n_2} \mathscr{E}_k^1(a_\bullet) \right| < \varepsilon,$$

we need only $n_2 < n_1$ terms. Here $\mathscr{E}_k^1(a_\bullet)$ is the 1-Euler transform as defined in (3.8). In other words, Euler summation gets us to the same ε-accuracy with fewer terms than conventional summation. This advantage extends to series of matrices, as we will see with the Neumann series in Sect. 8.3.

Secondly, for a fixed choice of summation method and a fixed $\varepsilon > 0$, the value of n is highly sensitive to the order of summation and termination criteria. This is already evident in conventional summation of scalar series $s = \sum_{k=0}^{\infty} a_k$. Clearly we could not rely on $|s_n - s| = |\sum_{k=n+1}^{\infty} a_k| < \varepsilon$ as a termination criterion since the value of s is precisely what needs to be determined.

Suppose we use $|a_k| < \varepsilon$ (using $|a_k|/|s_k| < \varepsilon$ would not make much of a difference) as termination criterion with $\varepsilon = 10^{-6}$ and we use the geometric series with $a_k = 2^{-k}$ for illustration since we know $s = 2$. A straightforward summation algorithm is given by setting $s \leftarrow a_0 = 1/2$ and iteratively computing

$$s \leftarrow s + a_k \quad \text{for} \quad k = 1, \ldots, n, \tag{6.1}$$

until $|a_n| < 10^{-6}$, which gives the correct answer $\widehat{s} = 2$ in single precision. However, if we apply the same algorithm to what is essentially the same series with a single zero added as the first term:

$$b_k = \begin{cases} 0 & k = 0, \\ a_{k-1} & k \geq 1, \end{cases}$$

then although $s = \sum_{k=0}^{\infty} b_k = \sum_{k=0}^{\infty} a_k$, the computed sum is now $\widehat{s} = 0$ as it terminates at $n = 0$. The bottom line is that there is no universal termination criterion—n has to be ascertained on a case-by-case basis and for a general series we will have to assume that it is given as part of our input. Henceforth we will assume this and our goal is to compute $s := s_n = \sum_{k=1}^{n} a_k$ accurately.

The naïve algorithm in (6.1) is called *recursive summation* [13, Section 4.1]. It computes s with an error given by

$$\widehat{s} = \sum_{k=0}^{n} a_k(1 + \delta_k), \quad |\delta_k| \leq n\mathsf{u} + O(\mathsf{u}^2). \tag{6.2}$$

The algorithm extends immediately to matrix sums $S = \sum_{k=1}^{n} A_k \in \mathbb{R}^{d\times d}$. Since matrix addition is computed entrywise, if we write s_{ij} and a_{ijk} for the (i, j)th entry of S and A_k respectively, then (6.2) generalizes to

$$\widehat{s}_{ij} = \sum_{k=0}^{n} a_{ijk}(1 + \delta_{ijk}), \quad |\delta_{ijk}| \leq n\mathsf{u} + O(\mathsf{u}^2), \quad i, j = 1, \dots, d,$$

or, in terms of the Hadamard product $\circ$ and writing $\Delta_k := (\delta_{ijk}) \in \mathbb{R}^{d\times d}$,

$$\widehat{S} = \sum_{k=0}^{n} A_k \circ (\mathbb{1} + \Delta_k), \quad |\delta_{ijk}| \leq n\mathsf{u} + O(\mathsf{u}^2), \quad i, j = 1, \dots, d. \tag{6.3}$$

Since for $A, B \in \mathbb{R}^{d\times d}$,

$$\|A \circ B\| \leq \|A\| \max_{i,j=1,\dots,d} |b_{ij}|, \tag{6.4}$$

we obtain the forward error bound

$$\|S - \widehat{S}\| \leq \sum_{k=0}^{n} \|A_k \circ \Delta_k\| \leq n\mathsf{u} \sum_{k=0}^{n} \|A_k\| + O(\mathsf{u}^2).$$

This serves as a baseline bound—we will discuss three more accurate summation methods that can significantly reduce the coefficient $n\mathsf{u}$ to $O(\sqrt{n}\mathsf{u})$, $O(\mathsf{u})$, and even $O(\mathsf{u}^2)$.

For a scalar series, a simple strategy [13, Sect. 4.2] to improve accuracy of (6.1) is to reorder the summands in increasing magnitudes to minimize the rounding error at

each step. Note that this does not work for matrix series since there is no natural total order on $\mathbb{R}^{d\times d}$ and reordering often improves the accuracy of one entry at the expense of decreased accuracy in another.

6.1 Block summation algorithm

Assume without loss of generality that $b \in \mathbb{N}$ divides $n+1$. The block summation algorithm [14, Sect. 2.2] in Algorithm 1 modifies recursive summation (6.1) by dividing the sum into blocks of size b. In particular, it allows the block sums to be computed in parallel.

Algorithm 1 Block summation

INPUT: $A_0, \dots, A_n \in \mathbb{R}^{d\times d}$, block size b;
1: **for** $k = 1, \dots, (n+1)/b$ **do**
2: compute $S_i = \sum_{k=(i-1)b}^{ib-1} A_k$ with recursive summation (6.1);
3: **end for**
4: compute $S = \sum_{i=1}^{(n+1)/b} S_i$ with the recursive summation (6.1);
OUTPUT: S.

As in our discussion of (6.1), it is straightforward to extend [14, Equation 2.4] to a sum of matrices: Algorithm 1 satisfies

$$\widehat{S} = \sum_{k=0}^{n} A_k \circ (\mathbb{1} + \Delta_k), \quad |\delta_{ijk}| \le \Big(b + \frac{n+1}{b} - 2\Big)\mathsf{u} + O\big(\mathsf{u}^2\big), \quad i,j = 1,\dots,d,$$

with notations as in (6.3). By (6.4), we obtain the forward error bound for Algorithm 1:

$$\|S - \widehat{S}\| \le \Big(b + \frac{n+1}{b} - 2\Big)\mathsf{u} \sum_{k=0}^{n} \|A_k\| + O\big(\mathsf{u}^2\big).$$

The optimal bound $2\sqrt{n+1}-2$ is easily seen to be attained with $b = \sqrt{n+1}$ although in practice it is common to choose b to be a constant such as 128 or 256.

The parallelism in Algorithm 1 requires summands to be independent and may be lost in situations like computing a matrix polynomial $\sum_{k=0}^{n} c_k A^k$ with Horner's method (Algorithm 5).

6.2 Compensated summation algorithm

This is also known as Kahan summation [37] and is based on a clever exploitation of the floating point system. By observing that the rounding error in a floating-point addition of two matrices is itself a floating-point matrix, Algorithm 2 simply approximates this error with a correction term $C \in \mathbb{R}^{d\times d}$ at each step of recursive summation to adjust the computed sum.

Algorithm 2 Compensated summation

INPUT: $A_0, \ldots, A_n \in \mathbb{R}^{d\times d}$;
1: initialize $S \leftarrow 0$, $C \leftarrow 0$;
2: **for** $k = 0, \ldots, n$ **do**
3: $\quad Y \leftarrow \mathrm{fl}(A_k - C)$;
4: $\quad T \leftarrow \mathrm{fl}(S + Y)$;
5: $\quad C \leftarrow \mathrm{fl}(\mathrm{fl}(T - S) - Y)$;
6: $\quad S \leftarrow T$;
7: **end for**
OUTPUT: S.

Since the rounding error in floating point arithmetic is, by definition, the unit roundoff u, a straightforward matrix adaptation of [13, Equation 4.8] for Algorithm 2 yields

$$\widehat{S} = \sum_{k=0}^{n} A_k \circ (\mathbb{1} + \Delta_k), \quad |\delta_{ijk}| \le 2\mathsf{u} + O\big(\mathsf{u}^2\big), \quad i, j = 1, \ldots, d, \tag{6.5}$$

with notations as in (6.3). By (6.4), we obtain the forward error bound for Algorithm 2:

$$\|S - \widehat{S}\| \le 2\mathsf{u} \sum_{k=0}^{n} \|A_k\| + O\big(\mathsf{u}^2\big).$$

Remarkably, Algorithm 2 eliminates n from the error bound. This enhanced accuracy is achieved at the cost of three extra matrix additions per loop, and is often more expensive than simply switching to higher precision [14]. So compensated summation is usually deployed only when computations are already taking place at the highest available precision.

6.3 Mixed block summation algorithm

Assume without loss of generality that $b \in \mathbb{N}$ divides $n+1$. Algorithm 3 is a variant of Algorithm 1 that strikes a balance between a fast algorithm FastSum and an accurate algorithm AccurateSum.

Algorithm 3 Mixed block summation

INPUT: $A_0, \ldots, A_n \in \mathbb{R}^{d\times d}$, block size b, FastSum, AccurateSum;
1: **for** $k = 1, \ldots, (n+1)/b$ **do**
2: $\quad$ compute $S_i = \sum_{k=(i-1)b}^{ib-1} A_k$ with FastSum;
3: **end for**
4: compute $S = \sum_{i=1}^{(n+1)/b} S_i$ with AccurateSum;
OUTPUT: S.

When $b = 1$, Algorithm 3 is exactly AccurateSum and when $b = n + 1$, Algorithm 3 is exactly FastSum. The scalar version of this algorithm was proposed by Blanchard, Higham, and Mary in [14] and we merely adapted it for matrices. The

following corollary of [14, Theorem 3.1] follows from the same arguments used in Sects. 6.1 and 6.2. Recall that we write $\Delta_k := (\delta_{ijk}) \in \mathbb{R}^{d\times d}$.

Corollary 6.1 (Error bound of mixed block summation algorithm) *Let the sum computed with* FastSum *satisfy*

$$\widehat{S}^{\mathsf{F}} = \sum_{k=0}^{n} A_k^{\mathsf{F}} \circ \left(\mathbb{1} + \Delta_k^{\mathsf{F}}\right), \quad |\delta_{ijk}^{\mathsf{F}}| \le \varepsilon^{\mathsf{F}}(n), \quad i, j = 1, \ldots, d,$$

and likewise for AccurateSum *with* A *in place of* F *in the superscript. Then the sum computed with Algorithm 3 satisfies*

$$\widehat{S} = \sum_{k=0}^{n} A_k \circ (\mathbb{1} + \Delta_k), \quad |\delta_{ijk}| \le \varepsilon(n,b) := \varepsilon^{\mathsf{F}}(b) + \varepsilon^{\mathsf{A}}(n/b) + \varepsilon^{\mathsf{F}}(b)\varepsilon^{\mathsf{A}}(n/b), \quad i, j = 1, \ldots, d,$$

and thus

$$\|S - \widehat{S}\| \le \varepsilon(n,b) \sum_{k=0}^{n} \|A_k\|.$$

In particular, if AccurateSum is calculated in double precision, i.e., $\varepsilon^{\mathsf{A}}(n) = O(\mathsf{u}^2)$, then the error bound is $\varepsilon(n,b) = \varepsilon^{\mathsf{F}}(b) + O(\mathsf{u}^2)$. Various options for the subroutines AccurateSum and FastSum are discussed in [14].

7 Summing matrix power series

Unlike the general matrix series considered in the last section, matrix power series admit more efficient algorithms. They are also intimately connected to the study of matrix functions [15]. The benefit of this connection goes both ways—the algorithms used to evaluate matrix functions, notably Padé approximation and Schur–Partlett algorithm, may be adapted to implement the summation methods in Sects. 3 and 4 numerically; the summation methods in Sects. 3 and 4 may in turn be used to enhance these algorithms and to extend the domains of matrix functions.

For these purposes, the following basic definition of a matrix function [15] suffices: If $X \in \mathbb{C}^{d\times d}$ and the power series $f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k$ converges in a neighborhood of $z_0 \in \mathbb{C}$, then

$$f(X) := \sum_{k=0}^{\infty} a_k (X - z_0 I)^k$$

whenever the matrix power series on the right is summable in the conventional sense. By definition, the domain of f is confined to

$$\Omega := \left\{ X \in \mathbb{C}^{d\times d} : \sum_{k=0}^{\infty} a_k (X - z_0 I)^k = S \text{ for some } S \in \mathbb{C}^{d\times d} \right\}.$$

With hindsight from Sects. 3 and 4, we may define

$$f(X) \overset{\mathsf{R}}{=} \sum_{k=0}^{\infty} a_k (X - z_0 I)^k$$

with respect to any regular summation method R, extending the domain of f to a potentially larger domain

$$\Omega_{\mathsf{R}} := \left\{ X \in \mathbb{C}^{d\times d} : \sum_{k=0}^{\infty} a_k (X - z_0 I)^k \overset{\mathsf{R}}{=} S \text{ for some } S \in \mathbb{C}^{d\times d} \right\} \supseteq \Omega.$$

This portends a new vista in the study of matrix functions but any further exploration would take us too far afield.

We will instead limit our attention to the numerics and only to regular sequential summation methods in Sect. 3 as these work hand-in-glove with numerical algorithms for matrix functions. In this regard, there is no loss of generality to assume that $z_0 = 0$. As is the case in Sect. 6, we begin by approximating $f(X)$ with its truncated Taylor Series $\sum_{k=0}^{n} a_k X^k$ for some $n \in \mathbb{N}$. But unlike the case of a general matrix series $\sum_{k=0}^{\infty} A_k$, working with a matrix power series $\sum_{k=0}^{n} a_k X^k$ permits us to ascertain n in advance to achieve a desired ε-accuracy,

$$\left\| f(X) - \sum_{k=0}^{n} a_k X^k \right\| < \varepsilon$$

as in [40, Theorem 11.2.4] or [41, Corollary 2] (see also [15, Theorem 4.8]).

We next see how we may add a summation method to the process. Let R be a regular sequential summation method (3.1) such that $C_{n,k} \in \mathbb{C}^{d\times d}$ for all $n, k \in \mathbb{N}$ and $C_{n,k} = 0$ for all $k > n$. The Nörlund means (with Cesàro summation as a special case) in Sect. 3.1 and Euler summation methods in Sect. 3.2 all meet this criterion. For any $A_k \in \mathbb{C}^{d\times d}$, $k \in \mathbb{N}$, and $S_n = \sum_{k=0}^{n} A_k$, observe that

$$\sum_{k=0}^{n} C_{n,k} S_k = \sum_{j=0}^{n} \left(\sum_{k=j}^{n} C_{n,k} \right) A_k. \tag{7.1}$$

So for matrix power series the summation is characterized by the sums

$$B_{n,k} := \sum_{j=k}^{n} a_k C_{n,j} \tag{7.2}$$

for $k \leq n$, $k, n \in \mathbb{N}$. Let $\varepsilon > 0$. If $\sum_{k=0}^{\infty} a_k X^k$ is R-summable to $f(X)$, then for some $n \in \mathbb{N}$,

$$\left\| f(X) - \sum_{k=0}^{n} B_{n,k} X^k \right\| = \left\| f(X) - \sum_{j=0}^{n} \left(\sum_{k=j}^{n} a_j C_{n,k} \right) X^j \right\| < \varepsilon. \tag{7.3}$$

Using this, we will generalize Padé approximation and the Schur–Parlett algorithm to work with Nörlund means and Euler summation. At this point, truncation error bounds like [40, Theorem 11.2.4] or [41, Corollary 2] that allow one to estimate n from a given ε are beyond our reach for (7.3). We will assume below, as we did in Sect. 6, that n is furnished as part of our inputs.

7.1 Padé approximation

This is one of the most powerful methods in matrix functions computations [15, Section 4.4.2]. The `expm` method in MATLAB, which implements the scaling-and-squaring method to compute the matrix exponential [38], is testament to one of the greatest wins[2] of Padé approximation. We will augment it with a regular sequential summation method R.

An (m, n)-*Padé approximant* of $f(z) = \sum_{k=0}^{\infty} a_k z^k$ with respect to R is a rational function $[p/q](z)$ where $p(z) = \sum_{k=0}^{m} \beta_k z^k$, $q(z) = \sum_{k=0}^{n} \gamma_k z^k$, $\gamma_0 = 1$, and

$$p(X)q(X)^{-1} = \sum_{k=0}^{m+n} B_{m+n,k} X^k, \tag{7.4}$$

with $B_{m+n,0}, B_{m+n,1}, \ldots, B_{m+n,m+n} \in \mathbb{C}^{d\times d}$ as defined in (7.1) and (7.2). By this definition, the standard Padé approximation in [15, Sect. 4.4.2] is then exactly the Padé approximation with respect to conventional summation.

Right multiplying $q(X)$ on both side of (7.4), we get

$$\sum_{k=0}^{m} \beta_k X^k = \sum_{k=0}^{m+n} \left(\sum_{j=0}^{k} \gamma_{k-j} B_{m+n,j} \right) X^k.$$

Since this holds for all $X \in \mathbb{C}^{d\times d}$, we may equate coefficients of X^k on both sides to get

$$\sum_{j=0}^{k} \gamma_{k-j} B_{m+n,j} = \begin{cases} \beta_k I & \text{if } k = 0, \ldots, m, \\ 0 & \text{if } k = m+1, \ldots, n. \end{cases} \tag{7.5}$$

For simplicity, we may choose a summation method R with $C_{n,k} = c_{n,k} I$ for some $c_{n,k} \in \mathbb{C}$ in (7.1) so that $B_{n,k} = b_{n,k} I$ for some $b_{n,k} \in \mathbb{C}$ in (7.2). This simplification is not overly restrictive as it includes important methods such as Cesàro summation and Euler summation with $P = \rho I$ for $\rho > 0$. The upside is that the coefficients of p/q may be easily determined by solving for β_k and γ_k in a system of $m+n+1$ linear equations (7.5). We summarize this in Algorithm 4.

[2] https://blogs.mathworks.com/cleve/2024/01/25/nick-higham-1961-2024/.

Algorithm 4 Padé approximation with sequential summation

INPUT: $X \in \mathbb{C}^{d\times d}$, $m, n \in \mathbb{N}$, $a_k, c_{m+n,k} \in \mathbb{C}$ for $k = 0, \ldots, m+n$;
1: **for** $k = 0, \ldots, m+n$ **do**
2: compute $b_{m+n,k} = a_k \sum_{j=k}^{m+n} c_{m+n,j}$;
3: **end for**
4: solve the linear system (7.5) for β_k for $k = 0, \ldots, m$ and γ_k for $k = 0, \ldots, n$;
5: compute $P = \sum_{k=0}^{m} \beta_k X^k$ and $Q = \sum_{k=0}^{n} \gamma_k X^k$ with Algorithm 5;
OUTPUT: PQ^{-1}.

Algorithm 5 Horner's method

INPUT: $a_0, \ldots, a_n \in \mathbb{R}$, $X \in \mathbb{R}^{d\times d}$;
1: initialize $P \leftarrow X$, $S \leftarrow a_0 I + a_1 X$;
2: **for** $k = 2, \ldots, n$ **do**
3: $P \leftarrow PX$;
4: $S \leftarrow S + a_k P$;
5: **end for**
OUTPUT: S.

Algorithm 5 in Algorithm 4 (and also in Algorithm 6 later) may be replaced by more sophisticated algorithms for evaluating matrix polynomials such as those in [42] or [43, Sect. 4.6.4], depending on whether one values stability or speed or yet other factors like parallelizability more.

Our approach in Algorithm 4 forms the Padé approximant $Y = PQ^{-1}$ in the usual way: solving a linear system with multiple right-hand sides $Q^{\mathsf{T}}Z = P^{\mathsf{T}}$ and taking $Y = Z^{\mathsf{T}}$. While there are alternative approaches via continued fractions and partial fractions, these are not necessarily stabler, as pointed out in [15, Sect. 4.4.3]. More importantly, while a representation of a Padé approximant in the form PQ^{-1} is readily available through solving the linear system (7.5), the same cannot be said of the other forms. Even for a function as standard as the matrix cosine function, it has been pointed out in [15, p. 290] that the Padé approximant has no convenient continued fraction or partial fraction form.

We favor the PQ^{-1} approach for yet a third reason: It is straightforward to incorporate the summation methods in Sect. 3, as we did in (7.4). We have added some experiments in Sect. 8.4 to show how Algorithm 4 works in conjunction with Cesàro and Euler summations, allowing us to sum a power series in regions far outside its usual range of convergence.

7.2 Schur–Parlett algorithm

The 'Schur' part of this algorithm is routine: To evaluate $f(X) = \sum_{k=0}^{\infty} a_k X^k$ for $X \in \mathbb{C}^{d\times d}$, a Schur decomposition $X = QRQ^*$ with unitary $Q \in \mathbb{C}^{d\times d}$ and upper triangular $R \in \mathbb{C}^{d\times d}$ yields $f(X) = Qf(R)Q^*$, thus reducing the problem to computing $f(R)$.

The 'Parlett' part of this algorithm is where the innovation lies: partition $R \in \mathbb{C}^{d\times d}$ into an $r \times r$ block matrix $R = (R_{ij})$, $i, j \in 1, \ldots, r$, with square diagonal blocks R_{ii}, $i = 1, \ldots, r$. Parlett [39] observed that the matrix $F = f(R)$ commutes with

R; has the same block structure $F = (F_{ij})$; and upon evaluating the diagonal blocks $F_{ii} = f(R_{ii})$, the superdiagonal blocks can be obtained from a system of Sylvester equations

$$R_{ii}F_{ij} - F_{ij}R_{jj} = F_{ii}R_{ij} - R_{ij}F_{jj} + \sum_{k=i+1}^{j-1}(F_{ik}R_{kj} - R_{ik}F_{kj}), \quad 1 \le i < j \le d. \tag{7.6}$$

The system (7.6) is nonsingular if and only if R_{ii} and R_{jj} have no eigenvalue in common [15, Sect. D.14]. Fortunately this may be guaranteed [44] by further transforming R into an identically partitioned upper triangular matrix $T = VRV^*$ with some unitary $V \in \mathbb{C}^{d\times d}$ such that for a fixed $\delta > 0$,

(i) $\min\{|\lambda_i - \lambda_j| : \lambda_i \in \lambda(T_{ii}), \lambda_j \in \lambda(T_{jj}), i \ne j\} > \delta$;
(ii) if the block T_{ii} has dimension greater than 1, then every $\lambda \in \lambda(T_{ii})$ has a corresponding $\mu \in \lambda(T_{ii})$ with $\mu \ne \lambda$ and $|\lambda - \mu| \le \delta$.

Essentially (i) says that between-block eigenvalues are well separated; and (ii) says that within-block eigenvalues are closely clustered. Since $X = (QV^*)T(QV^*)^*$ and $F = (QV^*)f(T)(QV^*)^*$, we may use T in place of R. This additional transformation from R to T carries other numerical advantages [15, Sect. 9.3] in the solution of (7.6).

We augment the Schur–Parlett algorithm with a regular sequential summation method R by computing the diagonal blocks F_{ii} as R-sums, i.e.,

$$F_{ii} = f(R_{ii}) \approx \sum_{k=0}^{n} B_{n,k}R_{ii}^k = \sum_{k=0}^{n}\Big(\sum_{j=k}^{n} a_k C_{n,j}\Big)R_{ii}^k, \quad i = 1, \dots, r, \tag{7.7}$$

where $B_{n,k}, C_{n,j} \in \mathbb{C}^{d\times d}$ are as defined in (7.1) and (7.2). Note that the diagonal blocks R_{ii}'s in (7.7) would in general have different dimensions for different i, which means that the matrices $B_{n,k}, C_{n,j}$ would need to have dimensions the same as R_{ii}'s and therefore chosen differently for each i. While there is no reason why we cannot do this we provide a simple workaround—we just set $C_{n,k} = c_{n,k}I$ for some $c_{n,k} \in \mathbb{C}$ as we did in Sect. 6.

This simplification in turn constraints us to use $C_{n,k} = c_{n,k}I$ for some $c_{n,k} \in \mathbb{C}$ in (7.1) but the result is both dimension-independent and computationally efficient as it only requires computing scalar coefficients $\sum_{k=j}^{n} a_j c_{n,k}$ as opposed to matrix coefficients. We summarize this in Algorithm 6.

Compared to directly summing (7.3), Algorithm 6 dramatically improves computational time, as we will see in Sect. 8.2.

8 Numerical experiments

We will present numerical experiments to illustrate the use of the summation methods in Sects. 3 and 4 in conjunction with the numerical algorithms in Sects. 6 and 7. Because of the large number of possible combinations, it is not possible to be exhaustive although we try to present a diverse selection. Our experiments will see four

Algorithm 6 Schur–Parlett algorithm with sequential summation

INPUT: $X \in \mathbb{C}^{d\times d}$, $n \in \mathbb{N}$, $a_k, c_{n,k} \in \mathbb{C}$ for $k = 0, \ldots, n$;
1: compute the Schur decomposition $X = QRQ^*$;
2: compute $T = VRV^*$ with block partition satisfying Conditions (i) and (ii);
3: $R \leftarrow T$;
4: $Q \leftarrow QV^*$;
5: **for** $i = 1, \ldots, r$ **do**
6: compute $F_{ii} = \sum_{k=0}^{n}(\sum_{j=k}^{n} a_k c_{n,j}) R_{ii}^k$ by Algorithm 5;
7: **end for**
8: **for** $j = 2, \ldots, r$ **do**
9: **for** $i = j-1, j-2, \ldots, 1$ **do**
10: solve for F_{ij} in the Sylvester equation (7.6);
11: **end for**
12: **end for**
OUTPUT: QFQ^*.

types of matrix series (Taylor, Fourier, Dirichlet, Hadamard); two sequential summations (Cesàro and Euler), two functional summations (Borel and Lambert); and three numerical algorithms (Schur–Parlett algorithm, recursive, and compensated summations). Each experiment is designed to showcase a different utility of these methods and algorithms.

Sect. 8.1: using Cesàro sums to alleviate Gibbs phenomenon in matrix Fourier series;
Sect. 8.2: using Euler and strong Borel sums to extend matrix Taylor series;
Sect. 8.3: using Euler sums for high accuracy evaluation of matrix functions;
Sect. 8.4: using Cesàro and Euler summations in Padé approximations;
Sect. 8.5: using Lambert sums to investigate matrix Dirichlet series;
Sect. 8.6: using compensated summation for accurate evaluation of Hadamard power series.

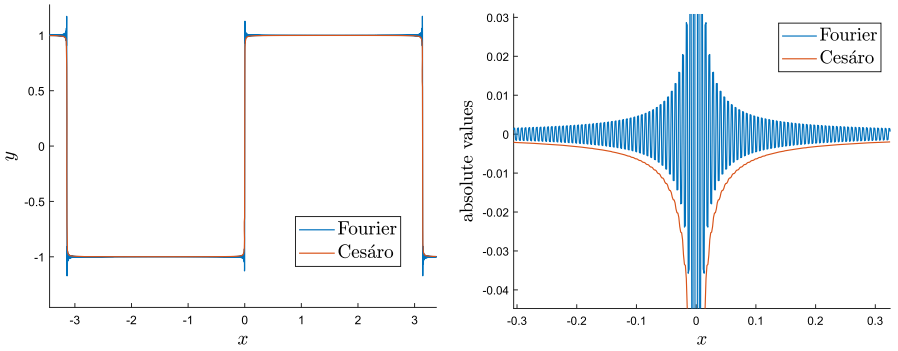
All experiments are performed with MATLAB R2023a in double precision ($\mathsf{u} = 2^{-52} \approx 2.22 \times 10^{-16}$) arithmetic unless noted otherwise. Plots presented in log scale would all be in base 10. All codes have been made available at

https://github.com/thomasw15/Summing-Divergent-Matrix-Series.

8.1 Avoiding Gibbs phenomenon with Cesàro summation

When one attempts to approximate a discontinuous function with its Fourier series, the Fourier approximation inevitably overshoots near a point of discontinuity—the notorious Gibbs phenomenon. The canonical example is given by the square wave function $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = \begin{cases} 1 & 2k\pi \le x < (2k+1)\pi, \ k \in \mathbb{Z}, \\ -1 & (2k-1)\pi \le x < 2k\pi, \ k \in \mathbb{Z}. \end{cases} \tag{8.1}$$



(a) Approximating square wave. (b) Absolute value of errors near 0.

Fig. 2 Gibbs phenomena in Fourier series corrected with Cesàro sum

Attempting to approximate f by its 1000-term Fourier series

$$f_{1000}(x) = \sum_{k=1}^{1000} \frac{2}{\pi k}\left(1 - (-1)^k\right) \sin(kx) \tag{8.2}$$

produces the blue curve in Fig. 2a

While the Gibbs phenomenon may be ameliorated with ad hoc recipes like the Lanczos factor [45], a superior remedy would be to use Cesàro summation. The same data in (8.2) yields the Cesàro partial sum

$$\sigma_{1000}(x) = \frac{1}{1000} \sum_{n=1}^{999} \sum_{k=1}^{n} \frac{2}{\pi k}\left(1 - (-1)^k\right) \sin(kx), \tag{8.3}$$

which nearly eliminates the wild oscillations completely, as shown in the red curves in Fig. 2a and b. While this is well known, the following matrix version is new, as far as we know.

Consider the following matrix Fourier series and its corresponding Cesàro sum:

$$\begin{aligned} F_{100}(X(t)) &= \sum_{k=1}^{100} \frac{2}{\pi k}\left(1 - (-1)^k\right) \sin(kX(t)), \\ \Sigma_{100}(X(t)) &= \frac{1}{99} \sum_{n=1}^{99} \sum_{k=1}^{n} \frac{2}{\pi k}\left(1 - (-1)^k\right) \sin(kX(t)) \end{aligned} \tag{8.4}$$

where $X : \mathbb{R} \to \mathbb{R}^{1000\times 1000}$ is a continuous matrix-valued function with $\lambda(X(0)) = 0$. Note that each summand involves a matrix sine function [15, Chapter 12], which we compute with the MATLAB function `funm(X,@sin)`.

In a neighborhood of $x = 0$ the square wave function (8.1) is identical to the sign function, i.e., $\text{sign}(x) = 1$ if $x \geq 0$ and -1 if $x < 0$. It is therefore conceivable that the

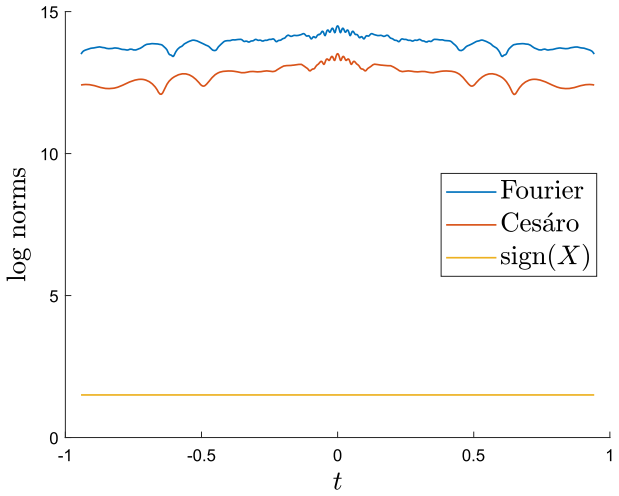


Fig. 3 Failure to approximate matrix sign function for nondiagonalizable X

same would hold for matrix functions and that the sums in (8.4) should approximate the matrix sign function $\operatorname{sign}(X)$ [15, Chapter 5] in a neighborhood of $X = 0$. Surprisingly this is only true if X is diagonalizable and false otherwise, a fact we discovered through the following numerical experiments.

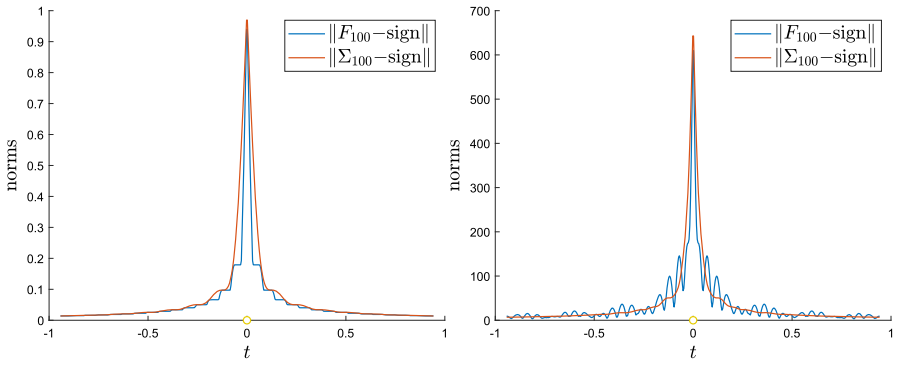
Consider the obviously nondiagonalizable matrix $X(t) = \operatorname{diag}(J_1(t), \ldots, J_{100}(t))$ where $J_i : \mathbb{R} \to \mathbb{R}^{10\times 10}$ is given by

$$J_i(t) = \begin{bmatrix} t\lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & t\lambda_i \end{bmatrix}.$$

Using the compensated summation in Algorithm 2, we compute the two sums in (8.4) and compare their norms with that of the matrix sign function in Fig. 3.

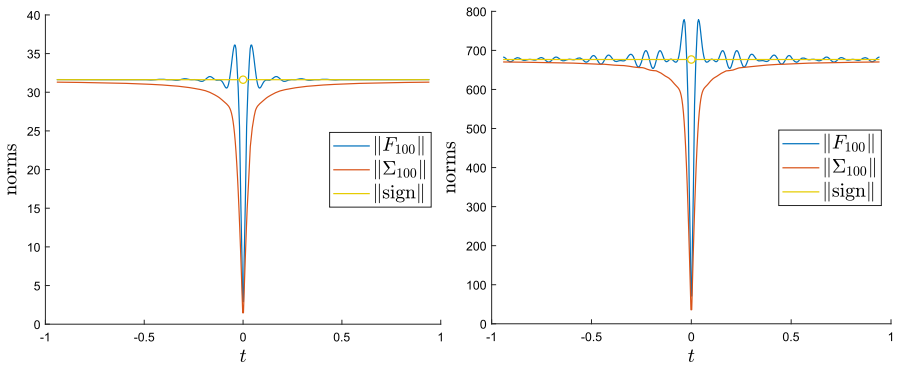
The result shows that the sums in (8.4) bear no resemblance to the matrix sign function—both $\|F_{100}(X(t))\|$ and $\|\Sigma_{100}(X(t))\|$ are orders of magnitude away from $\|\operatorname{sign}(X(t))\|$. With hindsight, the reason is clear, as the sums in (8.4) will always involve the superdiagonal of 1's, whereas these play no role in the matrix sign function. While we have chosen $X(t)$ above to accentuate this effect, the argument holds true as long as there is a single Jordan block of size at least 2×2, i.e., as long as the matrix is not diagonalizable.

On the other hand, the sums in (8.4) give a fair approximation of $\operatorname{sign}(X)$ for a diagonalizable matrix X and, as expected, we see prominent Gibbs phenomenon in $F_{100}(X)$ that is alleviated in $\Sigma_{100}(X)$. We will give a symmetric and a nonsymmetric example by randomly generating $\lambda_1, \ldots, \lambda_{1000} \in \mathbb{R}$, orthogonal Q and nonsingular



(a) symmetric matrices (b) nonsymmetric diagonalizable matrices

Fig. 4 Matrix sign function approximated by matrix Fourier and Cesàro sums



(a) symmetric matrices (b) nonsymmetric diagonalizable matrices

Fig. 5 Gibbs phenomena in matrix Fourier series corrected with Cesàro sum

tridiagonal $T \in \mathbb{R}^{1000\times 1000}$, and defining

$$Y(t) = Q\,\mathrm{diag}(t\lambda_1, \ldots, t\lambda_{1000})Q^{\mathsf{T}}, \qquad Z(t) = T\,\mathrm{diag}(t\lambda_1, \ldots, t\lambda_{1000})T^{-1}.$$

We approximate the square wave function with the matrix Fourier series and its Cesàro sum in (8.4), with $Y(t)$ and $Z(t)$ in place of $X(t)$, relying again on Algorithm 2 to compute the sums.

Outside $t = 0$, where the matrix sign function is undefined, both F_{100} and Σ_{100} provide fair approximations as quantified by $\|F_{100}(Y(t)) - \mathrm{sign}(Y(t))\|$ and $\|\Sigma_{100}(Y(t)) - \mathrm{sign}(Y(t))\|$ in Fig. 4. We expect the approximation errors to further decrease as the number of terms increases beyond 100. For comparison the more accurate approximations in Fig. 2a for the scalar series took a 1000-term approximation, which is beyond our reach here for 1000×1000 matrix series.

In a neighborhood of $t = 0$, we see the unmistakable mark of Gibbs phenomenon in F_{100}, reflected in the norms of $\|F_{100}(Y(t))\|$ and $\|F_{100}(Z(t))\|$, the blue curves in

Fig. 5a and b respectively. The oscillatory behavior vanishes when we instead look at the corresponding Cesàro sums Σ_{100}, whose norms are given by the red curves in Fig. 5. This indicates that for diagonalizable matrices, Cesàro summation is a remedy for Gibbs phenomenon in matrix Fourier series.

8.2 Accurate summation with Euler method and strong Borel method

These experiments accomplish two goals. We first verify numerically that the Euler and strong Borel methods indeed extend the domain of Neumann series beyond $\mathbb{D}$, which we demonstrated analytically in Corollary 3.11 and Proposition 4.19. The experiments for Euler methods are also used to show that the Schur–Parlett algorithm for Euler summation, i.e., Algorithm 6 with $c_{n,k} = \binom{n+1}{k+1}\rho^{n-k}(1+\rho)^{-n-1}$, is less accurate but dramatically faster than directly computing with Algorithm 2.

We generate fifty matrices $X \in \mathbb{C}^{1000\times 1000}$ such that $\lambda(X) \subseteq \{z \in \mathbb{C} : |z+\rho| < 1+\rho\}$ for $\rho = 10^4$ and $\lambda(X) \nsubseteq \mathbb{D}$. Note that the Neumann series $\sum_{k=0}^{\infty} X^k$ for such matrices will not be conventionally summable. Our goal is to verify that Euler method and strong Borel method will however yield the expected $(I-X)^{-1}$ numerically. For Euler method, we compute the truncated (E, ρ) sum as defined in (3.10),

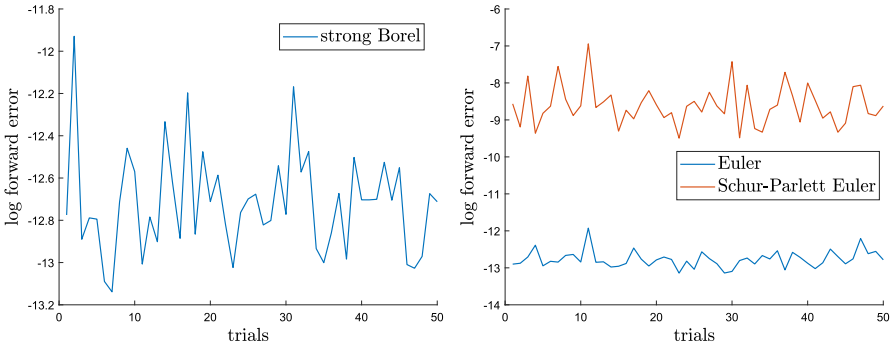
$$\widehat{S}_{(\mathsf{E},\rho)} := \sum_{k=0}^{10000} \mathscr{E}_k^{\rho}(X_\bullet)$$

where $X_\bullet = (X^k)_{k=0}^{\infty}$, first with compensated summation and then with the Schur–Parlett algorithm in single precision ($\mathsf{u} = 2^{-23} \approx 1.19 \times 10^{-7}$). For the strong Borel method, we use the MATLAB function `integral` with tolerance level 10^{-12} to compute the Borel sum $\widehat{S}_{\mathsf{SB}}$ as in (4.7) in single precision.

We plot the forward errors $\|\widehat{S} - (I-X)^{-1}\|$ in Fig. 6 and the backward errors $\|\widehat{S}(I-X) - I\|$ in Fig. 7, where $\widehat{S}$ is either $\widehat{S}_{(\mathsf{E},\rho)}$ or $\widehat{S}_{\mathsf{SB}}$. The near zero errors are strong numerical evidence that both Euler and strong Borel methods analytically extend the Neumann series to $(I-X)^{-1}$, which we of course know is true by virtue of Corollary 3.11 and Proposition 4.19.

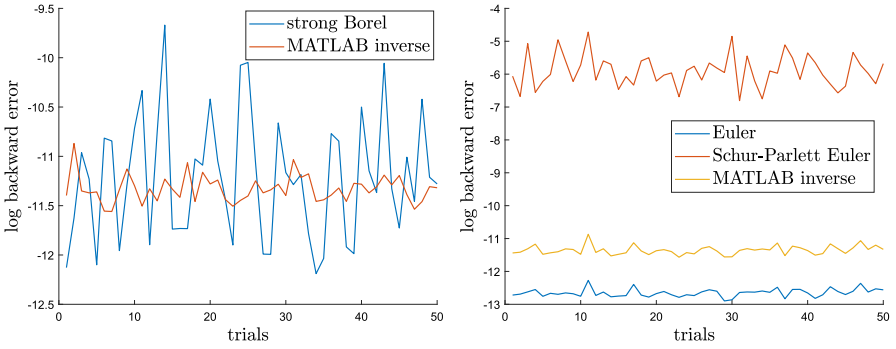
Observant readers might have noticed an issue here. We do not really have $(I-X)^{-1}$ exactly but only the output of the `inv` function in MATLAB, which is also subjected to floating point and approximation errors. Indeed our 'forward errors' here are simply a measure of deviation from $\widehat{S}_{\mathtt{inv}}$, the result of `inv` applied to $I-X$, computed in double precision. The backward errors $\|\widehat{S}(I-X) - I\|$ for $\widehat{S}_{\mathtt{inv}}$, $\widehat{S}_{(\mathsf{E},\rho)}$, $\widehat{S}_{\mathsf{SB}}$, computed in double precision, provide a more equitable comparison and therein lies a surprise—when computed with compensated summation, $\widehat{S}_{(\mathsf{E},\rho)}$, the result of Euler method, is more accurate than $\widehat{S}_{\mathtt{inv}}$, the result of MATLAB's `inv`, as is evident in Fig. 7b

As both forward and backward errors in Figs. 6b and 7b reveal, the Schur–Parlett algorithm gives less accurate results for Euler sums than compensated summation. However, a comparison of their running times in Fig. 8 shows that the former is significantly faster.



(a) strong Borel method (b) Euler method

Fig. 6 Log forward errors of strong Borel and Euler summations



(a) strong Borel method (b) Euler method

Fig. 7 Log backward errors of strong Borel and Euler summations

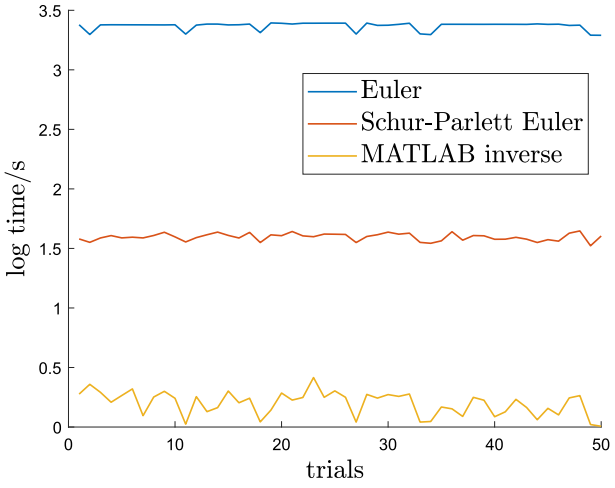


Fig. 8 Run time comparison of compensated summation and Schur–Parlett algorithm on Euler sums

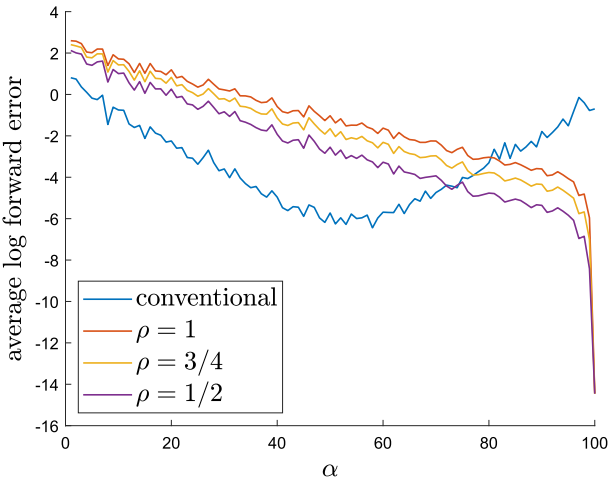


Fig. 9 Log errors from Euler methods

8.3 High accuracy sums with Euler methods

The surprising accuracy of Euler method computed with compensated summation uncovered in Sect. 8.2 deserves a more careful look. Here we will examine how the value of ρ impacts its accuracy.

We generate twenty bidiagonal matrices $X \in \mathbb{R}^{1000\times 1000}$ whose diagonal entries are negative with probability $\alpha \in \{0, 0.01, \ldots, 0.99, 1\}$. These matrices are generally not diagonalizable but we may readily prescribe their eigenvalue distribution. Again we will use the Neumann series $\sum_{k=0}^{\infty} X^k$, whose value $S = (I - X)^{-1}$ is known, as our test case. We approximate it with a 100-term truncated Taylor series and a 100-term Euler sum

$$\widehat{S} := \sum_{k=0}^{100} X^k \quad \text{and} \quad \widehat{S}_{(\mathsf{E},\rho)} := \sum_{k=0}^{100} \mathscr{E}_k^{\rho}(X_\bullet),$$

with $\rho \in \{1, 1/2, 1/4\}$, using compensated summation in Algorithm 2 to compute these sums. For a bidiagonal X we know $S = (I - X)^{-1}$ exactly in closed form and do not need to rely on MATLAB's `inv`, we may compute the forward errors $\|\widehat{S}_{(\mathsf{E},\rho)} - S\|$ and $\|\widehat{S} - S\|$. The logarithm of these values averaged over the twenty trials are plotted against α in Fig. 9.

We highlight two observations. Firstly, the downward trend of the curves for Euler method with increasing α shows that when the eigenvalues are predominantly negative, a truncated Euler sum gives a much higher level of accuracy with 100 terms than a truncated Taylor series with the same number of terms. This implies that Euler sums converge much faster than Taylor series. Secondly, when using Euler summation, smaller values of ρ lead to faster convergence than larger ones.

8.4 Extending summability range with Cesàro and Euler summations

For any $\alpha \in \mathbb{R}$, the binomial series on the left

$$\sum_{k=0}^{\infty} \binom{\alpha}{k} X^k = (I + X)^{\alpha} \tag{8.5}$$

converges to the value on the right in the conventional sense whenever $\lambda(X) \subseteq (-1, 1)$. Here $\binom{\alpha}{k} := \alpha(\alpha - 1) \cdots (\alpha - k + 1)/k!$ is the binomial coefficient, defined for any $\alpha \in \mathbb{R}$. This provides a good test case as the series on the left is an infinite series for any non-integer α but sums to the closed-form expression on the right conventionally if every eigenvalue λ of X satisfies $|\lambda| < 1$. However if there is any eigenvalue with $|\lambda| > 1$, then the series on the left necessarily diverges in the conventional sense. Our experiment will show that this range can be vastly extended.

We compute the (m, n)-Padé–Cesàro approximants, i.e., the $B_{m+n,k}$'s in (7.4) are given by

$$B_{m+n,k} = \binom{\alpha}{k} \frac{m + n + 1 - k}{m + n + 1} I.$$

We also compute the (m, n)-Padé–Euler approximants, i.e., the $B_{m+n,k}$'s in (7.4) are from the P-Euler transform in (3.8) with $P = \rho I$:

$$B_{m+n,k} = \binom{\alpha}{k} \sum_{j=k}^{m+n} \binom{m + n + 1}{j + 1} \frac{\rho^{m+n-j}}{(1 + \rho)^{m+n+1}} I.$$

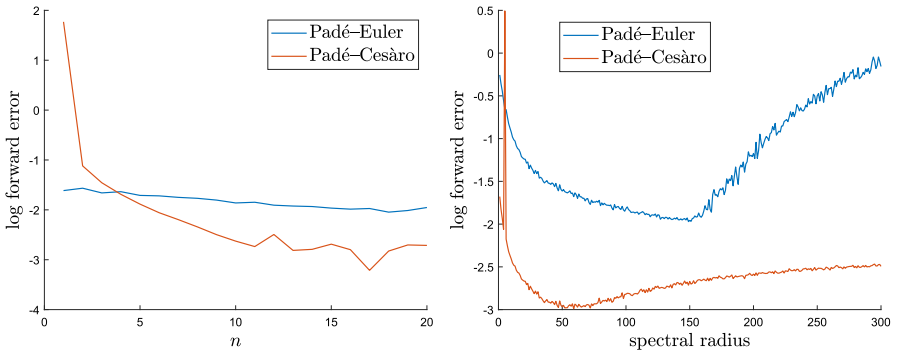
We set $m = n$ and denote these approximants by $\widehat{S}_{\mathsf{C},n}$ and $\widehat{S}_{\mathsf{E},n}$ respectively. We fix $\rho = 100$ and let α and n run over

$$\alpha = \pm\frac{1}{4}, \pm\frac{1}{2}, \pm\frac{3}{4}, \pm\frac{3}{5}, \pm\frac{4}{7}; \qquad n = 1, 2, \ldots, 20.$$

For each value of α and n, we repeat our experiments for ten matrices $X \in \mathbb{R}^{10\times 10}$, generated as $X = QRQ^{\mathsf{T}}$ with a random orthogonal matrix Q and a random upper triangular R with diagonal entries randomly chosen in [75, 150]. In other words, the spectrum $\lambda(X) \subseteq [75, 150]$, way beyond the range of convergence $(-1, 1)$ of the binomial series.

There is not much variation across different values of α and so we will present a typical case $\alpha = -3/4$. We will treat $(I+X)^{-3/4}$ computed using MATLAB's `mpower` function as the true value. With this, we obtain forward errors $\|\widehat{S}_{\mathsf{C},m} - (I+X)^{-3/4}\|_{\infty,1}$ and $\|\widehat{S}_{\mathsf{E},m} - (I+X)^{-3/4}\|_{\infty,1}$; note that we are computing absolute errors in the $(\infty, 1)$-norm $\|X\|_{\infty,1} = \max_{i,j=1,\ldots,n} |x_{ij}|$ in order to show the number of correct decimal digits. We plot their average values on a log scale in Fig. 10a

To see how far we can go before these approximants begin to show signs of divergence, we generate X as described above but now with $\lambda(X) \subseteq [\frac{1}{2}r, r]$ for r going up to 300. We show the forward error for the (15, 15)-Padé–Cesàro and Padé–Euler approximants on a log scale in Fig. 10b



(a) increasing order of approximants (b) increasing spectral range of X

Fig. 10 Log forward errors of Padé–Cesàro and Padé–Euler approximants

8.5 Matrix Dirichlet series with Lambert summation

A Dirichlet series is a scalar series

$$\sum_{n=0}^{\infty} \frac{a_n}{n^z}$$

where $a_\bullet \in s(\mathbb{C})$ and z is a complex variable. The best-known Dirichlet series is the Riemann zeta function

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}.$$

Another well-known Dirichlet series is one whose coefficients are given by $a_n = \mu(n)$, where

$$\mu(n) = \begin{cases} 1 & n \text{is square-free with an even number of prime factors,} \\ -1 & n \text{is square-free with an odd number of prime factors,} \\ 0 & n \text{is not square-free,} \end{cases}$$

is the Möbius function. It turns out that for any $z \in \mathbb{C}$ with $\mathrm{Re}(z) > 1$,

$$\sum_{n=1}^{\infty} \frac{\mu(n)}{n^z} = \frac{1}{\zeta(z)}$$

and

$$\lim_{z \to 1} \sum_{n=1}^{\infty} \frac{\mu(n)}{n^z} = 0. \tag{8.6}$$

An important application of the scalar Lambert summation [46, Lemma 2.3.7] is to show that

$$\sum_{n=1}^{\infty} \frac{\mu(n)}{n} \overset{\text{L}}{=} 0 \tag{8.7}$$

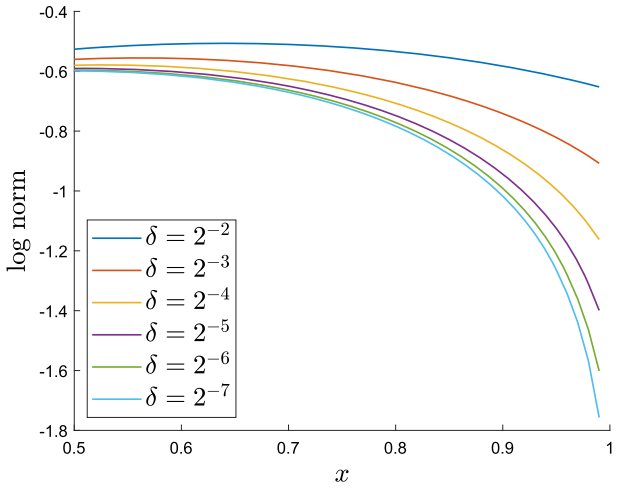


Fig. 11 Lambert approximation of the Dirichlet series

and our goal here is to verify a matrix analogue numerically.

It is straightforward to extend the definitions above. A matrix Dirichlet series is a matrix series

$$\sum_{n=0}^{\infty} a_n n^{-X}$$

where X is a complex matrix variable that takes values in $\mathbb{C}^{d\times d}$ and

$$n^X := \exp(\log(n)X),$$

with exp the matrix exponential function [15, Chapter 10]. Our numerical experiments show that if $X \in \mathbb{C}^{d\times d}$ has $\mathrm{Re}(X) \succeq I$, then

$$\sum_{n=1}^{\infty} \mu(n) n^{-X} \tag{8.8}$$

is Lambert summable in the sense of Definition 4.8 and

$$\lim_{X\to I} \left(\sum_{n=1}^{\infty} \mu(n) n^{-X} \right) \overset{\text{L}}{=} 0. \tag{8.9}$$

This is a matrix analogue of (8.6) and (8.7). Unlike the scalar version in (8.6), which is conventionally summable, our matrix version in (8.9) requires Lambert summation as the matrix Dirichlet series (8.8) is not conventionally summable if $1 \in \lambda(X)$, but is nevertheless Lambert summable.

To verify (8.9) numerically, we generate random matrices $X \in \mathbb{C}^{1000\times 1000}$ with $\mathrm{Re}(X) \succeq I$ and $\|X - I\| = \delta$ for $\delta = 2^{-2}, 2^{-3}, \ldots, 2^{-7}$, and compute

$$\widehat{S} = (1-x)\sum_{n=1}^{10000} \frac{nx^n}{1-x^n}\mu(n)n^{-X}$$

to approximate the Lambert sum as $x \to 1^-$. As shown in Fig. 11, for each δ, $\|\widehat{S}\|$ approaches a limiting value as $x \to 1^-$, and $\|\widehat{S}\| \to 0$ as $\delta \to 0$ or, equivalently, $X \to I$.

8.6 Recursive versus compensated summations

We present two sets of experiments to compare recursive summation in (6.1) with compensated summation in Algorithm 2, focusing on how the errors scale with respect to series length and matrix dimensions.

For $X \in \mathbb{R}^{d\times d}$, we consider the n-term Neumann series

$$\sum_{k=0}^{n-1} X^k = (I - X^n)(I - X)^{-1} =: S \tag{8.10}$$

for fixed $n = 1000$ and $d = 1, \ldots, 1000$. We also consider its n-term Hadamard analogue, i.e., with power taken with respect to the Hadamard product

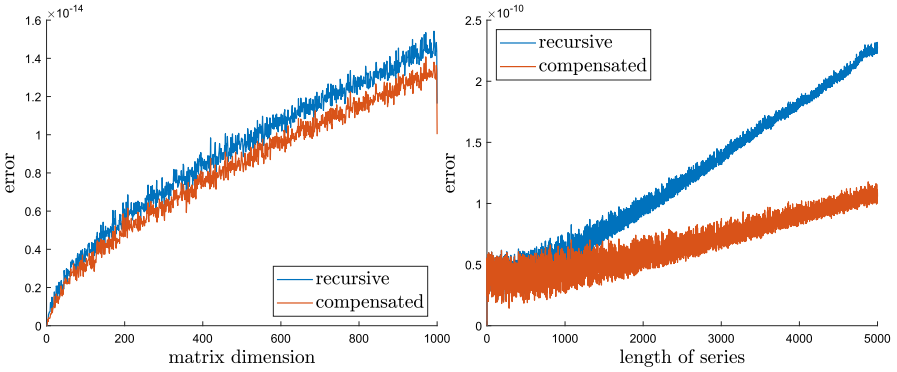
$$\sum_{k=0}^{n-1} X^{\circ k} = S_\circ, \quad s^\circ_{ij} = \frac{1 - x_{ij}^n}{1 - x_{ij}}, \quad i, j = 1, \ldots, d, \tag{8.11}$$

for fixed $d = 1000$ and $n = 1, \ldots, 5000$. In both cases we have the respective closed-form expressions for S and $S_\circ$ on the right of (8.10) and (8.11) that give their exact values and thereby permit calculation of forward errors.

We compute $\widehat{S}$, the sum on the left of (8.10), and $\widehat{S}_\circ$ the sum on the left of (8.11) using both recursive summation in (6.1) and compensated summation in Algorithm 2. The forward errors $\|\widehat{S} - S\|_\mathrm{F}$ and $\|\widehat{S}_\circ - S_\circ\|_\mathrm{F}$ are shown in Fig. 12a and b respectively. The result is clear: compensated summation is consistently more accurate than recursive summation, particularly with respect to increasing series length n, where the increase in errors follow significantly different trends.

While our forward error bound (6.5) predicts that the errors in compensated summation should be free of any dependence on n, this is assuming that we know the kth term *exactly*. In our sum (8.11), the kth term $X^{\circ k}$ is *computed*, and the increase in errors we see in Fig. 12b is a result of the multiplication errors accumulating in $\widehat{S}_\circ$ as n increases.

In case the reader is wondering why we did two different sets of experiments with respect to standard and Hadamard products: Hadamard products will not reveal the dependence on d in Fig. 12a as they are computed entrywise; whereas standard products will result in the multiplicative errors masking the trend in Fig. 12b showing



(a) increasing matrix dimension d (b) increasing series length n

Fig. 12 Errors for recursive and compensated summation algorithms

dependence on n, as computing X^k requires an order of magnitude more multiplications than computing $X^{\circ k}$.

9 Conclusion

This article is likely the first systematic study of summation techniques, both theoretical and numerical, for matrix series. Indeed we are unable to find much discussion of general numerical algorithms even for summing *conventionally convergent* matrix series, let alone the more convoluted summation methods for matrix series divergent in the conventional sense. The handful of previous works we found [15, 38, 44] had all focused on conventional summation of specific matrix Taylor series related to matrix functions, and said nothing of other summation methods or more general matrix series. Despite the length of our article, it still leaves significant room for future work, with several immediate open problems that we will briefly describe.

Our extensions of matrix Abelian mean in Definition 4.2, matrix Lambert sum in Definition 4.8, weak and strong matrix Borel sums in Definitions 4.10 and 4.12, leave open the question of whether one may further extend them by replacing the scalar parameter x in these definitions by a positive definite matrix. One may also ask a similar question of the matrix Mittag-Leffler sum in Definition 4.13: Could the gamma function be replaced by the matrix gamma function [47].

Another aspect beyond the scope of this article is that of conditioning, which likely explains the surprising accuracy of Euler method over matrix inversion uncovered in Sect. 8.2. Note that the left- and right-hand sides of (1.2), despite being equal in value, involve two different computational problems and almost surely have entirely different condition numbers. What is lacking is a study of the condition numbers of the summation methods in Sects. 3 and 4.

The numerical methods in Sects. 6 and 7 are mainly designed with accuracy in mind. They work well when adapted for matrix series and in conjunction with the summation methods in Sects. 3 and 4. When it comes to speed, there are many acceleration methods

for scalar series such as Aitken's δ^2-process and the vector ε-algorithm [48–50], but these involve nonlinear transforms and adapting them for matrix series is a challenge we save for the future.

As we alluded to in the introduction, these summation methods will allow for numerical investigations of "random matrix series," one that has its kth term A_k randomly generated according to some distributions like Wishart or GUE [51]. Many celebrated results in random matrix theory were indeed discovered first through numerical experiments and only rigorously proved much later.

Acknowledgements This work is partially supported by the DARPA grant HR00112190040, the NSF grants DMS 1854831 and ECCS 2216912, and a Vannevar Bush Faculty Fellowship ONR N000142312863. We thank the two anonymous reviewers for their very helpful comments and suggestions.

References

1. Boyd, D.W.: A p-adic study of the partial sums of the harmonic series. Experim. Math. **3**(4), 287–302 (1994)
2. Robert, A.M.: A course in p-adic analysis. Graduate texts in mathematics, vol. 198, p. 437. Springer, Heidelberg (2000) https://doi.org/10.1007/978-1-4757-3254-2
3. Hardy, G.H.: Divergent Series, Éditions Jacques Gabay, Sceaux, (1992) With a preface by J. E. Littlewood and a note by L. S. Bosanquet, Reprint of the revised p. 396. (1963)
4. Cesàro, E.: Sur la multiplication des series. Bull. Sci. Math. **14**, 114–120 (1890)
5. Shawyer, B., Watson, B.: Borel's methods of summability: theory and applications. Oxford Mathematical Monographs, p. 242. The Clarendon Press, Oxford Science Publications, Oxford University Press, New York (1994)
6. Boos, J.: Classical and Modern methods in summability. Oxford Mathematical Monographs, p. 586. Oxford Assisted by Peter Cass, Oxford Science Publications, Oxford University Press (2000)
7. Peyerimhoff, A.: Lectures on Summability. Lecture Notes in Mathematics, Vol. 107, p. 111. Springer (1969)
8. Wiener, N.: Tauberian theorems. Ann. Math. **33**(1), 1–100 (1932). https://doi.org/10.2307/1968102
9. Glimm, J., Jaffe, A.: Quantum physics: a functional integral point of view. 2nd Edn., p. 535. Springer (1987) https://doi.org/10.1007/978-1-4612-4728-9
10. Gurau, R.G., Krajewski, T.: Analyticity results for the cumulants in a random matrix model. Ann. Inst. Henri Poincaré D **2**(2), 169–228 (2015). https://doi.org/10.4171/AIHPD/17
11. Weinberg, S.: The Quantum Theory of Fields. Cambridge University Press, Cambridge (2005)
12. Hawking, S.W.: Zeta function regularization of path integrals in curved spacetime. Comm. Math. Phys. **55**(2), 133–148 (1977)
13. Higham, N.J.: Accuracy and stability of numerical algorithms, 2nd edn., p. 680. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2002). https://doi.org/10.1137/1.9780898718027
14. Blanchard, P., Higham, N.J., Mary, T.: A class of fast and accurate summation algorithms. SIAM J. Sci. Comput. **42**(3), 1541–1557 (2020). https://doi.org/10.1137/19M1257780
15. Higham, N.J.: Functions of matrices, p. 425. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2008). https://doi.org/10.1137/1.9780898717778. Theory and computation
16. Zhang, F.: Matrix Theory: basic results and techniques, 2nd Edn. Universitext, p. 399. Springer (2011). https://doi.org/10.1007/978-1-4614-1099-7.

17. Coskun, I., Riedl, E.: Normal bundles of rational curves in projective space. Math. Z. **288**(3–4), 803–827 (2018). https://doi.org/10.1007/s00209-017-1914-z
18. Dunford, N., Schwartz, J.T.: Linear operators. I. General Theory. Pure and applied mathematics, vol. 7, p. 858. Interscience Publishers Inc, New York; Interscience Publishers Ltd., London, With the assistance of W. G. Bade and R. G, Bartle (1958)
19. Manin, Y.I.: A course in mathematical logic for mathematicians, 2nd Edn. Graduate texts in Mathematics, vol. 53, p. 384. Springer. Chapters I–VIII translated from the Russian by Neal Koblitz, With new chapters by Boris Zilber and the author (2010). https://doi.org/10.1007/978-1-4419-0615-1
20. Rudin, W.: Principles of mathematical analysis, 3rd Edn. International Series in Pure and Applied Mathematics, p. 342. McGraw-Hill Book Co., New York-Auckland-Düsseldorf (1976)
21. Lim, L.-H.: Tensors in computations. Acta Numer **30**, 555–764 (2021). https://doi.org/10.1017/S0962492921000076
22. Woronoi, G.F.: Extension of the notion of the limit of the sum of terms of an infinite series. Ann. Math. **33**(3), 422–428 (1932). https://doi.org/10.2307/1968525
23. Nörlund, N.E.: Sur une application des fonctions permutables. Universitets Arsskrift, (N.F.) **avd. 2 16**(3) (1920)
24. Hille, E., Tamarkin, J.D.: On the summability of Fourier series. I. Trans. Amer. Math. Soc. **34**(4), 757–783 (1932). https://doi.org/10.2307/1989428
25. Sahney, B.N.: On the Nörlund summability of Fourier series. Pacific J. Math. **13**, 251–262 (1963)
26. Katznelson, Y.: An Introduction to Harmonic Analysis, 3rd Edn. Cambridge Mathematical Library, p. 314. Cambridge University Press, Cambridge (2004). https://doi.org/10.1017/CBO9781139165372
27. Korevaar, J.: Tauberian Theory. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], A century of developments, vol. 329, p. 483. Springer (2004). https://doi.org/10.1007/978-3-662-10225-1
28. Young, N.: An introduction to Hilbert space. Cambridge Mathematical Textbooks, p. 239. Cambridge University Press, Cambridge (1988). https://doi.org/10.1017/CBO9781139172011
29. Fejér, L.: Untersuchungen über Fouriersche Reihen. Math. Ann. **58**(1–2), 51–69 (1903). https://doi.org/10.1007/BF01447779
30. Askey, R.: Orthogonal polynomials and special functions, p. 110. Society for Industrial and Applied Mathematics, Philadelphia, PA (1975)
31. Knopp, K.: Theorie und Anwendung der Unendlichen Reihen. 4th ed. p. 583. Springer (1947)
32. Rosser, J.B.: Transformations to speed the convergence of series. J. Res. Nat. Bur. Stand. **46**, 56–64 (1951)
33. Hairer, E., Wanner, G.: Analysis by its history. Undergraduate texts in mathematics, readings in mathematics. p. 374. Springer (1996). https://doi.org/10.1007/978-0-387-77036-9
34. Hardy, G.H., Littlewood, J.E.: On a Tauberian theorem for lambert's series, and some fundamental theorems in the analytic theory of numbers. Proc. London Math. Soc **19**(1), 21–29 (1920). https://doi.org/10.1112/plms/s2-19.1.21
35. Borel, E.: Mémoire sur les séries divergentes. Ann. Sci. École Norm. Sup. **3**(16), 9–131 (1899)
36. Borwein, D., Shawyer, B.L.R.: On Borel-type methods. Tohoku Math. J. **18**, 283–298 (1966). https://doi.org/10.2748/tmj/1178243418
37. Kahan, W.: Pracniques: further remarks on reducing truncation errors. Commun. ACM **8**(1), 40 (1965). https://doi.org/10.1145/363707.363723
38. Higham, N.J.: The scaling and squaring method for the matrix exponential revisited. SIAM Rev. **51**(4), 747–764 (2009). https://doi.org/10.1137/090768539
39. Parlett, B.N.: A recurrence among the elements of functions of triangular matrices. Linear Algebra Appl. **14**(2), 117–121 (1976). https://doi.org/10.1016/0024-3795(76)90018-5
40. Golub, G.H., Van Loan, C.F.: Matrix computations, 3rd Edn. Johns Hopkins Studies in the Mathematical Sciences, p. 698. Johns Hopkins University Press, Baltimore, MD (1996)
41. Mathias, R.: Approximation of matrix-valued functions. SIAM J. Matrix Anal. Appl. **14**(4), 1061–1063 (1993). https://doi.org/10.1137/0614070
42. Paterson, M.S., Stockmeyer, L.J.: On the number of nonscalar multiplications necessary to evaluate polynomials. SIAM J. Comput. **2**, 60–66 (1973). https://doi.org/10.1137/0202007
43. Knuth, D.E.: The Art of Computer Programming. Seminumerical Algorithms, Vol. 2, 3rd Edn., p. 762. Addison-Wesley, Reading, MA (1998)
44. Davies, P.I., Higham, N.J.: A Schur-Parlett algorithm for computing matrix functions. SIAM J. Matrix Anal. Appl. **25**(2), 464–485 (2003). https://doi.org/10.1137/S0895479802410815

45. Acton, F.S.: Numerical methods that work. Mathematical Association of America, p. 549 Washington, DC. Corrected reprint of the 1970 edition (1990)
46. Mursaleen, M.: Applied summability methods. SpringerBriefs in Mathematics, p. 124. Springer (2014). https://doi.org/10.1007/978-3-319-04609-9
47. Jódar, L., Cortés, J.C.: Some properties of gamma and beta matrix functions. Appl. Math. Lett. **11**(1), 89–93 (1998). https://doi.org/10.1016/S0893-9659(97)00139-0
48. Brezinski, C., Redivo Zaglia, M.: Extrapolation methods. Studies in Computational Mathematics, vol. 2, p. 464. North-Holland Publishing Co., Amsterdam. Theory and practice, With 1 IBM-PC floppy disk (5.25 inch) (1991)
49. Graves-Morris, P.R., Roberts, D.E., Salam, A.: The epsilon algorithm and related topics. vol. 122, pp. 51–80. Numerical analysis, Interpolation and extrapolation vol. II (2000). https://doi.org/10.1016/S0377-0427(00)00355-1
50. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical recipes in Fortran, 2nd Edn. The art of scientific computing, with a separately available computer disk. p. 963. Cambridge University Press, Cambridge (1992)
51. Borodin, A., Corwin, I., Guionnet, A. (Eds.): Random matrices. IAS/Park City Mathematics Series. vol. 26, p. 498. American Mathematical Society, Providence, RI (2019). https://doi.org/10.1090/pcms/026