Journal of Computational and Applied Mathematics 237 (2013) 529–541

Contents lists available at SciVerse ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

An algorithm for computing a Padé approximant with minimal degree denominator

O.L. Ibryaeva *, V.M. Adukov

South Ural State University, Lenin avenue 76, Chelyabinsk, Russia

ARTICLE INFO

Article history:
Received 24 January 2012
Received in revised form 10 June 2012

Keywords:
Padé approximant
Toeplitz matrix
Padé–Laplace method
Froissart doublets

ABSTRACT

In this paper, a new definition of a *reduced Padé approximant* and an algorithm for its computation are proposed. Our approach is based on the investigation of the kernel structure of the Toeplitz matrix. It is shown that the reduced Padé approximant always has nice properties which the classical Padé approximant possesses only in the normal case. The new algorithm allows us to avoid the appearance of Froissart doublets induced by computer roundoff in the non-normal Padé table.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Padé approximants are locally the best rational approximations that can easily be constructed from the coefficients of a given power series. They are closely related to continued fractions, orthogonal polynomials, and Gaussian quadrature methods [1]. They have been widely used in various problems of mathematics, physics, and engineering due to their property to effectively solve the problem of analytic continuation of the series beyond its disc of convergence [2].

There are many methods available to compute Padé approximants [3]. Some of them are implemented in computer algebra systems such as *Maple* and *Mathematica*, and their built-in utilities are frequently used in applied problems. For calculation with floating-point numbers in *Maple*, an algorithm due to Geddes [4] based on fraction-free symmetric Gaussian elimination is used. Recursive algorithms for computing Padé approximants are also widely distributed. At first, they allowed one to find Padé approximants in the case of a normal Padé table, but later some of them were generalized to the non-normal case; see, for example, [5].

It is well known that the Padé table of a rational function $R(z)$ is always non-normal since it contains an infinite singular block whose elements are identical to $R(z)$. Actually the entries inside that block usually differ from the rational function $R(z)$ through the appearance of supplementary common roots in the numerators and the denominators, but, after reducing the common factors, we get the rational function $R(z)$ (the Padé fraction in Gragg's terminology [6]). In any practical calculations (because of the computer roundoff and the noise presence in the input data from which Padé approximants are constructed), the paired roots in the numerators and the denominators will not be rigorously equal. This phenomenon of «pairing» of such zeros and poles was named the *Froissart phenomenon*, and the pairs are known as *Froissart doublets* [7]. For example, their appearance is inevitable in signal processing using the Padé–Laplace method [8]. In order to identify doublets, several Padé approximants (besides the desired one) are usually calculated. This requires additional coefficients of the Taylor series, and is undesirable, since the coefficients are often computed numerically and Padé approximants are known to be very sensitive to errors in the coefficients.

The *main purpose* of the present paper is to propose a new algorithm for computing Padé approximants. This algorithm finds the denominator with the minimal degree among all the denominators of the Padé approximant and is based on

* Corresponding author.
E-mail address: oli@6v6power.ru (O.L. Ibryaeva).

0377-0427/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2012.06.022

the results concerning the kernel structure of Toeplitz matrices [9]. In the singular block case it allows one to avoid the appearance of Froissart doublets induced by computer roundoff.

Recently, a similar algorithm for computing Padé approximants was proposed in [10]. The authors used the method of hopping across a square singular block of the Padé table to find its upper-left corner. Having found the rank of a matrix, they reduce the numerator and denominator degrees diagonally and move to another position in the singular block. There may be several such hops across the singular blocks before we get to the final position in the upper-left corner corresponding to the minimal degree denominator. In our algorithm, we find this denominator by another method, reducing the degree in the denominator alone. The use of singular value decomposition (SVD) for computing numerical ranks (in the present paper and in [10]) realizes a regularization of the ill-posed Padé approximant problem and leads to stable algorithms.

Note that the concept of minimality of polynomial degrees is thoroughly studied in the theory of matrix Padé approximant. This problem is much more complicated, since in the matrix case the concept of minimum degrees and irreducible polynomials loses the properties of uniqueness [11]. In some cases, a rational matrix function can be represented in irreducible form by different pairs of polynomials with different pairs of minimal degrees (in some sense). In the recent paper [12], the authors proposed rational approximations of a matrix formal power series, with both matrix polynomials, numerator and denominator, satisfying three conditions: (a) minimum row degrees for the numerator and denominator, (b) an invertible denominator at the origin, and (c) canonical representation (without free parameters).

The paper is organized as follows. In Section 2, some definitions and examples are given. These examples show that in order to avoid the appearance of Froissart doublets we should study the kernel of the Toeplitz matrix which gives us denominators of the Padé approximant. This is done in Section 3, which contains the main result on the parameterization of the set of all the denominators of the Padé approximant. This result allows us to give the definition of the modified Padé approximant in Section 4 and to establish its properties. In Section 5, we propose our algorithm for Padé approximant computation and provide some examples in order to show that it allows us to avoid Froissart doublets induced by computer roundoff.

2. Preliminaries

This section contains some definitions and numerical experiments obtained in *Maple* and *Mathematica*. We have chosen these packages because they are the most widespread symbolic computation systems used in research and applications.

The examples illustrate how Froissart doublets appear or do not appear when we deal with the singular block of a Padé table for a rational function. Note that our new algorithm for computing Padé approximants will be proposed in Section 5. We will implement it in the *Maple* system and solve some of the examples again in order to show that the new method allows us to avoid Froissart doublets induced by computer roundoff.

We start with the classical definition of Padé–Frobenius approximants.

Definition 2.1 (*Padé–Frobenius*). Let $f(z)$ be a (formal) power series $f(z) = \sum_{k=0}^{\infty} c_k z^k,\ c_k \in \mathbb{C}$. The (m, n) *Padé approximant* corresponding to $f(z)$ is the rational function $f_{m,n}(z) = \frac{P_{m,n}(z)}{Q_{m,n}(z)}$, where $P_{m,n}(z)$ and $Q_{m,n}(z)$ are polynomials in z such that

1. $Q_{m,n}(z) \not\equiv 0$, $\deg Q_{m,n}(z) \le n$, $\deg P_{m,n}(z) \le m$,
2. $f(z)Q_{m,n}(z) - P_{m,n}(z) = r_{m+n+1}z^{m+n+1} + r_{m+n+2}z^{m+n+2} + \cdots$.

Note that, in Gragg's paper [6], the Padé–Frobenius approximant is called a Padé form. The Padé form in which the denominator and numerator are coprime is called the Padé fraction in [6].

In a similar way, we can define the Padé–Frobenius approximant at the point $z = a$ for the series $\sum_{k=0}^{\infty} c_k(z-a)^k$. For the sake of brevity, we will call it the Padé approximant. Throughout the paper, we will use Padé approximants at the point $z = 0$.

Obviously, the coefficients $q_0, \ldots, q_n$ of the denominator $Q_{m,n}(z)$ can be obtained by solving the following homogeneous system of linear equations with a Toeplitz $n \times (n+1)$ matrix:

$$\begin{pmatrix} c_{m+1} & c_m & \cdots & c_{m-n+1} \\ c_{m+2} & c_{m+1} & \cdots & c_{m-n+2} \\ \vdots & \vdots & & \vdots \\ c_{m+n} & c_{m+n-1} & \cdots & c_m \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ \vdots \\ q_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{1}$$

Here and below we assume that $c_k = 0$ if $k < 0$. From Condition 2 it follows that, with the $q_0, \ldots, q_n$ available, the coefficients $p_0, \ldots, p_m$ of the numerator $P_{m,n}(z)$ can be obtained from the formula

$$\begin{pmatrix} p_0 \\ p_1 \\ \vdots \\ p_m \end{pmatrix} = \begin{pmatrix} c_0 & 0 & \cdots & 0 \\ c_1 & c_0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ c_m & c_{m-1} & \cdots & c_{m-n} \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ \vdots \\ q_n \end{pmatrix}. \tag{2}$$

Thus, the (m, n) Padé approximant $f_{m,n}(z)$ always exists. Generally the polynomial $Q_{m,n}(z)$ (and hence $P_{m,n}(z)$) is not unique, since the rank of the matrix of the system (1) can be less than n. Nevertheless, it is easy to show that the rational function $f_{m,n}(z)$ is unique. Usually these rational functions are arranged in a double-entry table known as the *Padé table* $\{f_{m,n}(z)\}_{m,n=0,1,\dots}$ corresponding to $f(z)$. The Padé approximant $f_{m,n}(z)$ occupies the nth line and the mth column of the table. It is well known that identical Padé approximants can occur only in square blocks of the table. If the Padé table does not contain such blocks, it is said to be *normal*; otherwise, it is called *non-normal*.

Provided that there is denominator $Q_{m,n}(z)$ such that $Q_{m,n}(0) \neq 0$, we can rewrite Condition 2 as follows:

$$f(z) - \frac{P_{m,n}(z)}{Q_{m,n}(z)} = R_{m+n+1}z^{m+n+1} + R_{m+n+2}z^{m+n+2} + \cdots. \tag{3}$$

In this case, the rational function $\frac{P_{m,n}(z)}{Q_{m,n}(z)}$ is called a *Padé–Baker approximant*. Note that the Padé–Baker approximant does not always exist.

We now come to the second and the main part of this section, where we consider how Padé approximants can be obtained in such popular mathematical packages as *Maple*, *Mathematica*, and *Matlab*.

The built-in utility *pade* in *Matlab* approximates time delays by rational linear-time invariant models, i.e., actually it finds only diagonal Padé approximants for an exponential function by explicit formulas available for this case [13]. *Maple* and *Mathematica* compute Padé approximants in the general case, and in the examples below we will use their functions $pade(f, z = a, [m, n])$ (in *Maple*) and $Pade\ Approximant[expr,\ z,\ a,\ m, n]$ (in *Mathematica*), which compute the (m, n) Padé approximant of the function $f(z)$ at the point $z = a$.

To compare the results with the approximated rational functions we will find the zeros and poles of the approximants.

Example 2.1. Let us find by *Maple* and *Mathematica* the diagonal $(2, 2)$ Padé approximant $f_{2,2}$ for $f(z) = \frac{(z+1)(z-2)}{(z+2.1)(z-1)}$ at the point $z = 0$ via the following commands.

In *Maple*, we have

$f := \frac{(z+1)\cdot(z-2)}{(z+2.1)\cdot(z-1)}$:

$with(numapprox)$: $p := pade(f, z = 0, [2, 2])$;

$$\frac{0.9523809523 + 0.4761904764z - 0.4761904759z^2}{0.9999999999 - 0.5238095237z - 0.4761904759z^2}$$

$fsolve(numer(p), z, complex)$;

$$-1.000000000, \qquad 2.000000001$$

$fsolve(denom(p), z, complex)$;

$$-2.100000001, \qquad 1.000000000.$$

In *Mathematica*, we have

$In[1] := f := (z + 1) * (z - 2)/((z + 2.1) * (z - 1));$

$In[2] :=\ P := Pade\ Approximant[f, \{z, 0, \{2, 2\}\}]; P$

$$Out[2] := \frac{0.952381 + 0.47619z - 0.47619z^2}{1.000000000000000 - 0.52381z - 0.47619z^2}$$

$In[3] :=\ Roots[Numerator[P] == 0, z]$

$Out[3] :=\ z == -1. \parallel z == 2.$

$In[4] :=\ Roots[Denominator[P] == 0, z]$

$Out[4] :=\ z == -2.1 \parallel z == 1.$

As can be seen, the obtained approximant $f_{2,2}$ is identical to the function $f(z)$ since they have the same zeros and poles. Note that in this case the kernel dimension of the matrix of system (1) is equal to 1, i.e., the denominator of the Padé approximant is unique.

The following example shows that $f_{3,3}$ is again identical to the function $f(z)$ (both in *Maple* and in *Mathematica*), though the kernel is multidimensional.

Example 2.2. Let us find the $(3, 3)$ Padé approximant for the function $f(z) = \frac{(z+1)(z-2)}{(z+2.1)(z-1)}$ at the point $z = 0$.

In *Maple*, we have

$p := pade(f, z = 0, [3, 3])$;

$$\frac{0.9523809523 + 0.4761904764z - 0.4761904759z^2}{0.9999999999 - 0.5238095237z - 0.4761904759z^2}$$

$fsolve(numer(p), z, complex)$;

$$-1.000000000, \qquad 2.000000001$$

$fsolve(denom(p), z, complex)$;

$$-2.100000001, \qquad 1.000000000.$$

In *Mathematica*, we have

$In[1] := P := Pade\ Approximant[f, \{z, 0, \{3, 3\}\}]; P$

$$Out[1] := \frac{0.952381 + 0.47619z - 0.47619z^2 + 0.z^3}{1.000000000000000 - 0.52381z - 0.47619z^2 + 0.z^3}$$

$In[2] := Roots[Numerator[P] == 0, z]$

$Out[2] := z == -1. \parallel z == 2.$

$In[3] := Roots[Denominator[P] == 0, z]$

$Out[3] := z == -2.1 \parallel z == 1.$

However, approximant $f_{4,4}$ does not coincide with $f(z)$ because of the appearance of supplementary roots of the denominator and the numerator.

Example 2.3. Let us find the (4, 4) Padé approximant for the function $f(z) = \frac{(z+1)(z-2)}{(z+2.1)(z-1)}$ at the point $z = 0$.

In *Maple*, we have

$p := pade(f, z = 0, [4, 4])$;

$$\frac{0.9523809524 + 0.9750566894z - 0.2267573697z^2 - 0.2494331066z^3}{1. - 0.7505668935z^2 - 0.2494331066z^3}$$

$fsolve(numer(p), z, complex)$;

$$-\mathbf{1.909090909}, \qquad -0.9999999999, \qquad 2.000000000$$

$fsolve(denom(p), z, complex)$;

$$-2.100000001, \qquad -\mathbf{1.909090908}, \qquad 1.000000000.$$

In *Mathematica*, we have

$In[1] := P := Pade\ Approximant[f, \{z, 0, \{4, 4\}\}]; P$

$$Out[1] := \frac{0.952381 + 1.11851z - 0.155032z^2 - 0.321159z^3 - 3.33067 \times 10^{-16}z^4}{1.000000000000000 + 0.150624z - 0.829465z^2 - 0.321159z^3 + 0.z^4}$$

$In[2] := Roots[Numerator[P] == 0, z]$

$Out[2] := z == -9.64247 \times 10^{14} \parallel z == -\mathbf{1.48273} \parallel z == -1. \parallel z == 2.$

$In[3] := Roots[Denominator[P] == 0, z]$

$Out[3] := z == -2.1 \parallel z == -\mathbf{1.48273} \parallel z == 1.$

Thus, in Example 2.3, Froissart doublets induced by computer roundoff appear. Note that the doublets obtained in *Maple* and *Mathematica* are different (they are given in bold). The root -9.64247×10^{14} of the numerator (obtained in *Mathematica*) is caused by its small leading coefficient -3.33067×10^{-16}.

Note that for $f_{3,3}$ and $f_{4,4}$ the denominator is not unique and, as we have seen, *Maple* and *Mathematica* may choose not the best one.

Let us take another rational function and find its approximation in its infinite singular block. As it can be seen below, *Maple* does not produce Froissart doublets, but they appear when using *Mathematica*.

Example 2.4. Let us find the (2, 3) Padé approximant for the function $f(z) = \frac{z+1.01}{(z+2)(z-2.01)}$ at the point $z = 0$.

In *Maple*, we have

$f := \frac{z+1.01}{(z+2)\cdot(z-2.01)}$:

$with(numapprox) : p := pade(f, z, [2, 3])$;

$$\frac{-0.2512437812 - 0.2487562190z}{1.000000000 + 0.002487561992z - 0.2487562188z^2}$$

fsolve(*numer*(*p*), *z*, *complex*);

-1.010000000

fsolve(*denom*(*p*), *z*, *complex*);

$-2.000000001, \qquad 2.010000000.$

In *Mathematica*, we have

$In[1] := f := (z + 1.01)/((z + 2) * (z - 2.01));$

$In[2] := P := PadeApproximant[f, \{z, 0, \{2, 3\}\}]; P$

$$Out[2] := \frac{-0.251244 - 0.122069z + 0.125433z^2}{1.000000000000000 - 0.501752z - 0.250011z^2 + 0.125433z^3}$$

$In[3] := Roots[Numerator[P] == 0, z]$

$Out[3] := z == -1.01 \parallel z == \mathbf{1.98318}$

$In[4] := Roots[Denominator[P] == 0, z]$

$Out[4] := z == -2. \parallel z == \mathbf{1.98318} \parallel z == 2.01.$

The following conclusions may be drawn from the examples. When the kernel dimension of the matrix of system (1) is equal to 1 (as in Example 2.1), we have one denominator of the Padé approximant, and Froissart doublets cannot appear. But they can sometimes appear when the kernel dimension is greater than 1. If an improper denominator is chosen, then supplementary (also called artificial) roots appear. These roots may have corresponding pairs among the roots of the numerator, and the cancelation of the common factors reduces the Padé approximant to the reduced form. But, in the presence of computer roundoff, the paired roots in the numerator and denominator will not be rigorously equal, and this can cause the appearance of Froissart doublets (as in Examples 2.3 and 2.4). Moreover, computer roundoff may produce artificial roots with large absolute value (as in Example 2.3), since the vanishing leading coefficients are small but not equal to zero.

To avoid these problems, we should choose the denominator with the minimal degree. As will be shown further, this denominator always exists. In order to find it, we have to study the kernel structure of the Toeplitz matrix of system (1). This will be done in the following section.

3. Parameterization of the denominator set

In this section, we study the structure of the set of all denominators for an (m, n) Padé approximant, i.e., the structure of the kernel $\ker T_{m+1}$ of the Toeplitz matrix

$$T_{m+1} = \|c_{i-j}\|_{\substack{i=m+1,\dots,m+n \\ j=0,1,\dots,n}}$$

from system (1).

First, let us prove that the minimal degree denominator exists, and establish its properties.

Proposition 3.1. *There exists the denominator $Q^0_{m,n}(z)$ with the minimal degree among all the denominators $Q_{m,n}(z)$. This denominator is unique up to a constant factor.*

Let $P^0_{m,n}(z)$ be the numerator of the Padé approximant corresponding to $Q^0_{m,n}(z)$. Then the polynomials $P^0_{m,n}(z)$ and $Q^0_{m,n}(z)$ do not have common non-zero roots.

Proof. Since all the polynomials $Q_{m,n}(z)$ satisfy the condition $\deg Q_{m,n}(z) \le n$, there exists a denominator with the minimal degree d.

Suppose that there are two denominators with degree d:

$$Q^0_{m,n}(z) = B_d z^d + \cdots + B_0, \qquad \widetilde{Q}^0_{m,n}(z) = \widetilde{B}_d z^d + \cdots + \widetilde{B}_0, \quad B_d \neq 0,\ \widetilde{B}_d \neq 0.$$

Let us introduce $Q(z) = \widetilde{B}_d Q^0_{m,n}(z) - B_d \widetilde{Q}^0_{m,n}(z)$. This is the denominator of the Padé approximant with degree less than d. This is possible if and only if $Q(z) \equiv 0$. Thus, $Q^0_{m,n}(z)$ and $\widetilde{Q}^0_{m,n}(z)$ are linearly dependent. The uniqueness is proved.

Let $P^0_{m,n}(z)$ be the numerator corresponding to $Q^0_{m,n}(z)$. Let us suppose that z_0 is the common non-zero root of $P^0_{m,n}(z)$ and $Q^0_{m,n}(z)$:

$$P^0_{m,n}(z) = (z - z_0)\widetilde{P}^0_{m,n}(z), \qquad Q^0_{m,n}(z) = (z - z_0)\widetilde{Q}^0_{m,n}(z).$$

By Definition 2.1, we have

$$f(z)\widetilde{Q}^0_{m,n}(z) - \widetilde{P}^0_{m,n}(z) = \frac{1}{(z - z_0)}\, O\left(z^{m+n+1}\right) = O\left(z^{m+n+1}\right).$$

Then $\widetilde{Q}^0_{m,n}(z)$ is the denominator with degree less than d. We have a contradiction, since d is the minimal degree. □

Now we describe the denominator set $\{Q_{m,n}(z)\}$, i.e., the kernel structure of T_{m+1}. This result will be used to construct the algorithm for obtaining the minimal degree denominator.

To find any denominator of the Padé approximant, we need the sequence of the Taylor coefficients $c_{m-n+1}, \ldots, c_{m+n}$ of $f(z)$. The sequence consists of the entries of matrix T_{m+1}. Let us introduce notions of indices and essential polynomials for $c_{m-n+1}, \ldots, c_{m+n}$. The notions were given in more general situation in [9]. Here, we formulate them specially for our case.

It is natural to include T_{m+1} in the family of Toeplitz matrices

$$T_k = \begin{pmatrix} c_k & c_{k-1} & \cdots & c_{m-n+1} \\ c_{k+1} & c_k & \cdots & c_{m-n+2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m+n} & c_{m+n-1} & \cdots & c_{2m-k+1} \end{pmatrix}, \quad m-n+1 \le k \le m+n, \tag{4}$$

which are constructed from the elements of $c_{m-n+1}, \ldots, c_{m+n}$.

Consider the sequence of spaces $\ker T_k$. It is more convenient to deal not with vectors $Q = (q_0, q_1, \ldots, q_{k-m+n-1})^t \in \ker T_k$ but with their generating polynomials $Q(z) = q_0 + q_1 z + \cdots + q_{k-m+n-1} z^{k-m+n-1}$. Instead of the spaces $\ker T_k$ we will use the isomorphic spaces $\mathcal{N}_k$ consisting of the generating polynomials.

To do this, we introduce a linear functional σ by the formula

$$\sigma\{z^j\} = c_{-j}, \qquad -m-n \le j \le -m+n-1.$$

In the theory of orthogonal polynomials this functional is called the Stieltjes functional.

Denote by $\mathcal{N}_k$ $(m-n+1 \le k \le m+n)$ the space of polynomials $Q(z)$ with formal degree $k-m+n-1$ satisfying the orthogonality conditions:

$$\sigma\left\{z^{-i}Q(z)\right\} = 0, \quad i = k, k+1, \ldots, m+n. \tag{5}$$

It is easily seen that $\mathcal{N}_k$ is the space of generating polynomials of vectors in $\ker T_k$. For convenience, we put $\mathcal{N}_{m-n} = 0$ and denote by $\mathcal{N}_{m+n+1}$ the $(2n+1)$-dimensional space of all polynomials with formal degree $2n$.

Let d_k be the dimension of the space $\mathcal{N}_k$, and let $\Delta_k = d_k - d_{k-1}$ $(m-n+1 \le k \le m+n+1)$. The following fact is crucial for our further considerations.

Proposition 3.2. *For any non-zero sequence $c_{m-n+1}, \ldots, c_{m+n}$, the inequalities*

$$0 = \Delta_{m-n+1} \le \Delta_{m-n+2} \le \cdots \le \Delta_{m+n} \le \Delta_{m+n+1} = 2 \tag{6}$$

are fulfilled.

Proof. It follows from Definition 2.1 that $\mathcal{N}_k$, $z\mathcal{N}_k$ are subspaces of $\mathcal{N}_{k+1}$ $(m-n \le k \le m+n)$, and that

$$\mathcal{N}_k \cap z\mathcal{N}_k = z\mathcal{N}_{k-1}.$$

Hence, by the Grassman formula, we have

$$\dim\left(\mathcal{N}_k + z\mathcal{N}_k\right) = 2d_k - d_{k-1}. \tag{7}$$

Let h_{k+1} be the dimension of any complement $\mathcal{H}_{k+1}$ of the subspace $\mathcal{N}_k + z\mathcal{N}_k$ in the whole space $\mathcal{N}_{k+1}$. From (7), we have

$$h_{k+1} = \Delta_{k+1} - \Delta_k, \tag{8}$$

i.e., $\Delta_{k+1} \ge \Delta_k$. By definition, $d_{m-n} = 0$, and d_{m-n+1} is also equal to zero. Hence, $\Delta_{m-n+1} = 0$. In a similar manner we can prove that $\Delta_{m+n+1} = 2$. □

It follows from inequalities (6) that there exist integers $\mu_1 \le \mu_2$ such that

$$\begin{aligned} &\Delta_{m-n+1} = \cdots = \Delta_{\mu_1} = 0, \\ &\Delta_{\mu_1+1} = \cdots = \Delta_{\mu_2} = 1, \\ &\Delta_{\mu_2+1} = \cdots = \Delta_{m+n+1} = 2. \end{aligned} \tag{9}$$

If the second row in these relations is absent, we assume that $\mu_1 = \mu_2$. In our case, $\mu_1 < \mu_2$, as will be shown later.

Definition 3.1. The integers μ_1, μ_2 defined in (9) will be called the essential indices (briefly, indices) of the sequence $c_{m-n+1}, \ldots, c_{m+n}$.

Proposition 3.3. *Let $\varkappa = \operatorname{rank} T_m$. Then the indices μ_1, μ_2 are found by the following formulas:*

$$\mu_1 = m-n+\varkappa, \qquad \mu_2 = m+n-\varkappa+1.$$

Proof. It follows from the definition of Δ_k that $\sum_{k=m-n+1}^{m+n+1} \Delta_k = d_{m+n+1} - d_{m-n} = 2n+1$.

On the other hand, from the relations (9), we have

$$\sum_{k=m-n+1}^{m+n+1} \Delta_k = 1 \cdot (\mu_2 - \mu_1) + 2 \cdot (m + n + 1 - \mu_2).$$

Hence, $\mu_1 + \mu_2 = 2m + 1$. This means that $\mu_1 \le m < \mu_2$.

In a similar manner, we obtain $\sum_{k=m-n+1}^{m} \Delta_k = d_m = n - \varkappa$ and $\sum_{k=m-n+1}^{m} \Delta_k = m - \mu_1$. Thus, $\mu_1 = m - n + \varkappa$.

Since $\mu_1 + \mu_2 = 2m + 1$, we get $\mu_2 = m + n - \varkappa + 1$. □

Now, we can describe the structure of the kernels of the matrices T_k. It follows from (8) and (9) that $h_{k+1} := \dim \mathcal{H}_{k+1} \neq 0$ iff $k = \mu_j$ $(j = 1, 2)$. In that case, $h_{k+1} = 1$. Therefore, for $k \neq \mu_j$,

$$\mathcal{N}_{k+1} = \mathcal{N}_k + z\mathcal{N}_k, \tag{10}$$

and, for $k = \mu_j$,

$$\mathcal{N}_{k+1} = \left(\mathcal{N}_k + z\mathcal{N}_k\right)\dot{+}\mathcal{H}_{k+1}. \tag{11}$$

Definition 3.2. Any polynomial $Q_j(z)$ that forms a basis for the one-dimensional complement $\mathcal{H}_{\mu_j+1}$ will be called the essential polynomial of the sequence $c_{m-n+1}, \ldots, c_{m+n}$ corresponding to the index μ_j, $j = 1, 2$.

It can be shown (see [9, Theorem 4.1]) that integers μ_1, μ_2 are the indices, and that polynomials $Q_1(z) \in \mathcal{N}_{\mu_1+1}$, $Q_2(z) \in \mathcal{N}_{\mu_2+1}$ are the essential polynomials iff

$$\sigma_0 := \sigma\{z^{-m-n-1}[Q_2(0)Q_1(z) - Q_1(0)Q_2(z)]\} \neq 0. \tag{12}$$

In the following theorem, the structure of the kernels of matrices T_k $(m - n + 1 \le k \le m + n)$ is described in terms of the indices and the essential polynomials.

Proposition 3.4. *Let μ_1, μ_2 be the indices and let $Q_1(z)$, $Q_2(z)$ be the essential polynomials of the sequence $c_{m-n+1}, \ldots, c_{m+n}$. Then*

$$\mathcal{N}_k = \begin{cases} 0, & m - n + 1 \le k \le \mu_1, \\ \{q_1(z)Q_1(z)\}, & \mu_1 + 1 \le k \le \mu_2, \\ \{q_1(z)Q_1(z) + q_2(z)Q_2(z)\}, & \mu_2 + 1 \le k \le m + n, \end{cases}$$

where $q_1(z)$, $q_2(z)$ are arbitrary polynomials of formal degree $k - \mu_j - 1$.

Proof. From (9), we have $d_k = 0$ for $k \in [m - n + 1, \mu_1)$, i.e., $\mathcal{N}_k = 0$.

Let $\mu_1 + 1 \le k \le \mu_2$. It follows from (10) and (11) that the polynomials

$$\left\{R_1(z),\ zR_1(z), \ldots, z^{k-\mu_1-1}R_1(z)\right\} \tag{13}$$

generate the space $\mathcal{N}_k$. The number of these polynomials is equal to $k - \mu_1$.

From the definition of Δ_j and relations (9), we have $d_k = \sum_{j=m-n+1}^{k} \Delta_j = k - \mu_1$.

Thus, the number of polynomials in (13) is equal to the dimension of the space $\mathcal{N}_k$. Therefore, these polynomials form a basis of $\mathcal{N}_k$.

The case $\mu_2 + 1 \le k \le m + n$ can be considered in a similar manner. □

Now we apply the previous proposition to the case $k = m + 1$ and obtain the main result of this section on the parameterization of the denominator set.

Theorem 3.1. *Let $Q_{m,n}(z)$ be an arbitrary denominator of the (m, n) Padé approximant for the series $f(z) = \sum_{k=0}^{\infty} c_k z^k$. Form the sequence $\{c_{m-n+1}, \ldots, c_{m+n}\}$, which is necessary to find $Q_{m,n}(z)$.*

Let μ_1 be the first index, and let $Q_1(z)$ be the first essential polynomial of this sequence.

Then the denominator set is

$$\left\{Q_{m,n}(z)\right\} = \{q_1(z)Q_1(z)\}, \tag{14}$$

where $q_1(z)$ is an arbitrary polynomial with formal degree $m - \mu_1$. Thus, $Q_1(z)$ is the denominator $Q^0_{m,n}(z)$ with the minimal degree.

Remark 3.1. It follows from $Q_{m,n}(z) = q(z)Q_1(z)$ that the denominator $Q_{m,n}(z)$ such that $Q_{m,n}(0) \neq 0$ (Baker's condition) exists iff $Q_1(0) \neq 0$.

Hence, the Padé approximant with the minimal degree denominator is the Padé–Baker approximant if the latter exists.

Remark 3.2. Our parameterization (14) is similar to Gragg's parameterization in [6, Theorem 3.2, Item c]. However, to find the minimal degree denominator in Gragg's Theorem 3.2, we have to find any Padé form and reduce it by solving an ill-posed problem of computing the greatest common divisor (GCD) of two polynomials. Our Theorem 3.1 gives a GCD-free method for obtaining the minimal degree denominator $Q_1(z)$.

4. Reduced Padé approximant and its properties

In this section, we use the parameterization in order to modify the definition of the Padé approximants.

Definition 2.1 ignores the non-uniqueness (in the general case) of a denominator of a Padé approximant. It does not matter much for the normal Padé table, but in the singular case common factors of the denominator and the numerator cannot be cancelled because of roundoff errors. This is one of reasons why Froissart doublets can appear.

The main aim of this section is to modify the classical definition, Definition 2.1. We would like a Padé approximant to exist always, to have a unique denominator (up to a constant factor) and to coincide with the Padé–Baker approximant if the latter exists.

Let us add the minimality requirement to Definition 2.1.

Definition 4.1. Let $f(z)$ be a (formal) power series $f(z) = \sum_{k=0}^{\infty} c_k z^k,\ c_k \in \mathbb{C}$. The (m, n) *Padé approximant* corresponding to $f(z)$ is the rational function $f_{m,n}(z) = \frac{P_{m,n}(z)}{Q_{m,n}(z)}$, where $P_{m,n}(z)$ and $Q_{m,n}(z)$ are polynomials in z such that

1. $Q_{m,n}(z) \not\equiv 0$, $\deg Q_{m,n}(z) \le n$, $\deg P_{m,n}(z) \le m$,
2. $f(z)Q_{m,n}(z) - P_{m,n}(z) = r_{m+n+1}z^{m+n+1} + r_{m+n+2}z^{m+n+2} + \cdots$.
3. The polynomial $Q_{m,n}(z)$ has the minimal degree among all polynomials satisfying 1, 2.

Theorem 3.1 on parameterization of the denominator set gives a constructive method for finding the denominator with the minimal degree.

The next theorem shows that the Padé approximant has the desired properties.

Theorem 4.1. *For any power series $f(z)$, the following statements are fulfilled.*

1. *Any (m, n) Padé approximant exists and is unique.*
2. *The denominator $Q_{m,n}(z)$ of the Padé approximant is unique up to a constant factor, and $Q_{m,n}(z)$ is the first essential polynomial $Q_1(z)$ of the sequence $\{c_{m-n+1}, \ldots, c_{m+n}\}$.*
3. *$P_{m,n}(z)$ and $Q_{m,n}(z)$ do not have common non-zero roots.*
4. *The Padé–Baker approximant exists iff $Q_1(0) \neq 0$. The Padé approximant from Definition 4.1 is the Padé–Baker approximant if the latter exists.*
5. *If $Q_1(z)$ has root $z = 0$ of order $\delta_{m,n} > 0$, then*

$$f(z) - f_{m,n}(z) = Az^{m+n+1-\delta_{m,n}} + Bz^{m+n+2-\delta_{m,n}} + \cdots, \quad A \neq 0.$$

($\delta_{m,n}$ is called the deficiency index of the Padé approximant.)

Proof. Statements 1–4 evidently follow from Proposition 3.1, Theorem 3.1, and Remark 3.1.

We prove the last statement of the theorem. Let $z = 0$ be the root of order $\delta_{m,n}$ of $Q_1(z)$. Since $Q_1(0) = 0$, the number σ_0 from formula (12) is

$$\sigma_0 = -\sigma\{z^{-m-n-1}Q_1(z)\}Q_2(0) \neq 0.$$

Here $Q_2(z)$ is the second essential polynomial.

Thus $\sigma\{z^{-m-n-1}Q_1(z)\} \neq 0$.

It is easy to see that the number $\sigma\{z^{-m-n-1}Q_1(z)\}$ is the coefficient at z^{m+n+1} in the power series $f(z)Q_1(z)$. As it is not zero, we have

$$f(z)Q_1(z) - P_1(z) = A_1z^{m+n+1} + \cdots, \qquad A_1 = \sigma\{z^{-m-n-1}Q_1(z)\} \neq 0.$$

Here, $P_1(z)$ is the numerator corresponding to the denominator $Q_1(z)$.

After dividing this equation by $Q_1(z)$, we get

$$f(z) - f_{m,n}(z) = Az^{m+n+1-\delta_{m,n}} + \cdots,$$

where $A \neq 0$.

This means that the deficiency index of the Padé approximant $f_{m,n}(z)$ coincides with the multiplicity $\delta_{m,n}$ of the root $z = 0$ of the first essential polynomial $Q_1(z)$. □

Since the numerator and the denominator of Padé approximant from Definition 4.1 do not have common non-zero roots, we will call this approximant the *reduced Padé approximant*. Note that the numerator and the denominator of the reduced Padé approximant cannot be coprime because of their common zero roots.

Remark 4.1. It is impossible to improve Definition 4.1 in such a way that the denominator and the numerator are always coprime.

It is easy to see that, if $Q_{m,n}(z) = z^{\delta} Q^0_{m,n}(z)$, then $P_{m,n}(z) = z^{\delta} P^0_{m,n}(z)$. Hence $f_{m,n}(z) = \frac{P^0_{m,n}(z)}{Q^0_{m,n}(z)}$. This rational function is the Padé fraction in Gragg's terminology. However, the polynomials $P^0_{m,n}(z)$, $Q^0_{m,n}(z)$ cannot be considered as the numerator and the denominator of Padé approximant $f_{m,n}(z)$, since the Frobenius Condition 2 is not fulfilled:

$$f(z)Q^0_{m,n}(z) - P^0_{m,n}(z) = Az^{m+n-\delta+1} + \cdots.$$

To determine δ (the deficiency index of $f_{m,n}(z)$), we should know how many low-order coefficients of $Q_1(z)$ are equal to zero. Note that vanishing coefficients of the denominator and the numerator can be non-zero because of roundoff errors. In particular, the non-zero vanishing coefficients can cause the appearance of roots of the denominator and/or the numerator with the great absolute value (as in Example 2.3).

Thus, we face the problem of finding vanishing coefficients of the denominator and the numerator. The results of Theorem 4.1 allow us to solve this problem. This will be done in the following two theorems.

Recall that $\varkappa = \operatorname{rank} T_m$, $\mu_1 = m - n + \varkappa$, matrix T_{μ_1+1} has size $(2n-\varkappa) \times (\varkappa+1)$, and $\operatorname{rank} T_{\mu_1+1} = \varkappa$. The vector which is a basis for one-dimensional space $\ker T_{\mu_1+1}$ gives coefficients of the first essential polynomial $Q_1(z)$ which is the minimal degree denominator of the Padé approximant.

Further, we will denote by $T^{(k)}_{\mu_1+1}$ the matrix which is obtained from the matrix T_{μ_1+1} by deletion of the kth column, $k = 1, \ldots, \varkappa+1$.

Theorem 4.2. *Let $Q_1(z) = q_0 + q_1 z + \cdots + q_\varkappa z^\varkappa$ be the denominator with the minimal degree. Then $q_k = 0$ iff $\operatorname{rank} T^{(k+1)}_{\mu_1+1} = \varkappa - 1$, $k = 0, \ldots, \varkappa$.*

Proof. Let $q_k = 0$. We have $\operatorname{rank} T^{(k+1)}_{\mu_1+1} \le \varkappa$. Suppose $\operatorname{rank} T^{(k+1)}_{\mu_1+1} = \varkappa$. Then $T^{(k+1)}_{\mu_1+1}$ has a trivial kernel, i.e., $q_0 = q_1 = \cdots = q_{k-1} = q_{k+1} = \cdots = q_\varkappa = 0$. Thus, $Q_1(z) \equiv 0$. Since this is impossible, we have $\operatorname{rank} T^{(k+1)}_{\mu_1+1} \le \varkappa - 1$.

Assume that $\operatorname{rank} T^{(k+1)}_{\mu_1+1} < \varkappa - 1$. Then $T^{(k+1)}_{\mu_1+1}$ has a multidimensional kernel, which can be embedded in the kernel of T_{μ_1+1} in a natural way. But $\ker T_{\mu_1+1}$ is one dimensional; hence our assumption is false, and $\operatorname{rank} T^{(k+1)}_{\mu_1+1} = \varkappa - 1$.

On the other hand, if $\operatorname{rank} T^{(k+1)}_{\mu_1+1} = \varkappa - 1$, then $\ker T^{(k+1)}_{\mu_1+1}$ is one dimensional, and after the natural embedding in one-dimensional space $\ker T_{\mu_1+1}$, we get $q_k = 0$. □

Denote by $T^{[k]}_{\mu_1+1}$ the matrix which is obtained by inserting the row $(c_k\ c_{k-1} \ldots c_{k-\varkappa})$ at the beginning of the matrix T_{μ_1+1}, $k = 0, \ldots, m$.

Theorem 4.3. *Let $P_1(z) = p_0 + p_1 z + \cdots + p_m z^m$ be the numerator corresponding to the denominator $Q_1(z)$. Then $p_k = 0$ iff $\operatorname{rank} T^{[k]}_{\mu_1+1} = \varkappa$, $k = 0, \ldots, m$.*

Proof. It follows from (2) that $p_k = c_k q_0 + c_{k-1} q_1 + \cdots + c_{k-\varkappa} q_\varkappa$, $k = 0, \ldots, m$. Hence,

$$T^{[k]}_{\mu_1+1} \begin{pmatrix} q_0 \\ q_1 \\ \vdots \\ q_\varkappa \end{pmatrix} = \begin{pmatrix} p_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

If $p_k = 0$, then matrix $T^{[k]}_{\mu_1+1}$ has a non-trivial kernel; hence $\operatorname{rank} T^{[k]}_{\mu_1+1} < \varkappa + 1$. Since $\operatorname{rank} T_{\mu_1+1} = \varkappa$, we have $\operatorname{rank} T^{[k]}_{\mu_1+1} = \varkappa$.

On the other hand, if $\operatorname{rank} T^{[k]}_{\mu_1+1} = \varkappa$ then the inserted row $(c_k\ c_{k-1} \ldots c_{k-\varkappa})$ is a linear combination of the rows of T_{μ_1+1}. Therefore, $c_k q_0 + c_{k-1} q_1 + \cdots + c_{k-\varkappa} q_\varkappa = 0$, i.e., $p_k = 0$. □

From Theorems 4.2 and 4.3, we can obtain the multiplicities of $z = 0, z = \infty$ as the roots of $Q_1(z)$, $P_1(z)$, i.e., the deficiency index of $f_{m,n}(z)$ and the degrees of $Q_1(z)$, $P_1(z)$.

In particular, the following result on the deficiency index is now evident.

Corollary 4.1. *The number δ is the deficiency index of $f_{m,n}(z)$ iff*

$$\operatorname{rank} T^{(1)}_{\mu_1+1} = \cdots = \operatorname{rank} T^{(\delta-1)}_{\mu_1+1} = \varkappa - 1, \qquad \operatorname{rank} T^{(\delta)}_{\mu_1+1} = \varkappa.$$

We would like to end this section with the following remark. As is well known in the case of *normal* Padé approximant $f_{m,n}(z) = \frac{P_{m,n}(z)}{Q_{m,n}(z)}$, the degrees of $P_{m,n}(z)$ and $Q_{m,n}(z)$ are equal to m and n, respectively, and $\delta = 0$. Thus, in the normal case, the degrees of the numerator and the denominator and the deficiency index δ are known. Our modified definition (Definition 4.1) and the results of this section allow us to determine them not only in the normal case, but also in the non-normal one.

5. Algorithm

In this section, we present our algorithm for computing a reduced Padé approximant with the minimal degree denominator. As we have seen in the previous section, in order to realize the algorithm, we have to find the rank of matrix T_m, and the null space of T_{μ_1+1}. Moreover, in order to delete the vanishing coefficients of the denominator and the numerator we have to find the ranks of matrices $T^{(k)}_{\mu_1+1}$, $k = 1, \ldots, \varkappa + 1$, and $T^{[k]}_{\mu_1+1}$, $k = 0, \ldots, m$. In practice, due to rounding and measuring errors and finite computer precision, the elements of these matrices are perturbed with error, so we have to determine the rank and the null space of the original matrix from the perturbed matrix.

For computer calculations we need to define the numerical rank and the numerical null space of A. Recall these concepts which are crucial for the algorithm.

The definition of the numerical rank was first given in [14]. We will use the simplified definition (as in [15,13, p. 72]). The *numerical* ε *rank* of an $M \times N$ matrix A with respect to the threshold $\varepsilon > 0$ is defined as the smallest rank of all matrices within a 2-norm distance ε of A. Namely,

$$\operatorname{rank}(A, \varepsilon) = \min_B \{\operatorname{rank} B : \| A - B \|_2 \le \varepsilon\}.$$

The numerical ε rank may be characterized in terms of the singular value decomposition (SVD). According to Stewart [16], «singular value decomposition is the *crème de la crème* among rank-revealing decomposition».

Let us recall the definition of SVD. Every $M \times N$ complex matrix A can be represented in the form

$$A = U \Sigma V^H,$$

where U, V are unitary matrices (^{H} means the Hermite conjugation), and Σ is an $M \times N$ matrix in which the upper $N \times N$ block is a diagonal matrix with all entries real and sorted in descending order. The diagonal entries $\sigma_1 \ge \cdots \ge \sigma_N$ of the matrix Σ are called the singular values.

In the case of exact calculations, we have $\sigma_1 \ge \cdots \ge \sigma_r > \sigma_{r+1} = \cdots = \sigma_N = 0$, where r is the rank of A. In practice, we have

$$\sigma_1 \ge \cdots \ge \sigma_r > \varepsilon \ge \sigma_{r+1} \ge \sigma_N \ge 0.$$

Let us denote $A_r = U \Sigma_r V^H$ with $\Sigma_r = \operatorname{diag}\{\sigma_1, \ldots, \sigma_r, 0, \ldots, 0\}$; then $\|A - A_r\|_2 = \sigma_{r+1}$ and $\operatorname{rank}(A, \varepsilon) = \operatorname{rank} A_r = r$ [14]. Moreover, A_r is the nearest matrix to A (with respect to the 2-norm) with rank r. Therefore the null space of A_r is called the numerical null space of A within ε. The null space of A_r is spanned by $\{v_{k+1}, \ldots, v_N\}$, where v_k is the kth column of matrix V. The ratio $\gamma = \frac{\sigma_r}{\sigma_{r+1}}$ is called the numerical rank gap. The numerical rank of the matrix A may be estimated reliably in the case of «well-defined numerical rank»; that is, when A has a single well-determined gap between large and small singular values.

Thus, in terms of SVD, the numerical ε rank is the number of singular values greater than the given threshold ε. There is no uniform threshold for all applications. The user must make a decision on the threshold ε based on the nature of the application.

The `rank(A,tol)` function in *Matlab* returns the number of singular values of A that are larger than the threshold (tolerance) `tol`. The default tolerance is $\max(M, N) * \operatorname{eps}(\operatorname{norm}(A))$. Here, $\operatorname{norm}(A)$ indicates the Euclidean norm of A and $\operatorname{eps}(\operatorname{norm}(A))$ is approximately 2.2×10^{-16} times $\operatorname{norm}(A)$. This choice is usually a good choice if the errors in the matrix elements are due to computer arithmetic and if there is a sufficiently large gap in the singular values around this tolerance.

In [10], the authors treat a singular value as zero if it is less than $\texttt{tol} \cdot \|c\|_2$, where $c = (c_0, \ldots, c_{m+n})$ and `tol` is a relative tolerance. The tolerance is applied also in the detection of zero coefficients in the numerator and the denominator. The chosen value of `tol` in [10] is 10^{-14}. Our first version of the algorithm [17] used two independent parameters (tolerances) for these two goals: (1) for determining if the singular value is zero and (2) for determining if the coefficient is zero. Now, due to Theorems 4.2 and 4.3, we can choose only one parameter. In some applications the use of two tolerances may be preferable.

Although SVD is the most widely used method for the determination of the numerical rank and the numerical null space, there are alternative methods such as URV decomposition, LU decomposition, or QR factorization with column pivoting.

Now we can present our algorithm for computing the (m, n) reduced Padé approximant $f_{m,n}(z)$ for $f(z)$ at the point $z = a$.

Algorithm.

Initialization:

$f(z) :=$ the approximated function;

$(m, n) :=$ the order of the Padé approximant;

$a :=$ the center of expansion;

$d := 0$, if we want to delete vanishing coefficients of numerator and the denominator, else $d := 1$.

1. Compute the Taylor coefficients $c_0, \ldots, c_{m+n}$ of $f(z)$ at the point $z = a$.

2. Form the Toeplitz matrix $T_m = \|c_{i-j+m}\|_{\substack{i=1,\ldots,n+1,\\ j=1,\ldots,n.}}$

3. Determine its rank $\varkappa = \operatorname{rank} T_m$ and the index $\mu_1 = m - n + \varkappa$.

4. Form matrix $T_{\mu_1+1} = \|c_{i-j+\mu_1+1}\|_{\substack{i=1,\ldots,m+n-\mu_1,\\ j=1,\ldots,n-m+\mu_1+1.}}$

5. *Test:* Determine if the rank of T_{μ_1+1} is equal to $\varkappa$. If not, interrupt calculation of the Padé approximant.
6. Find a basis for its one-dimensional kernel $\ker T_{\mu_1+1}$.
The obtained vector $(q_0, q_1, \ldots, q_{n-m+\mu_1})^T$ is a vector formed from the coefficients of the minimal degree denominator $Q_1(z) = \sum_{k=0}^{n-m+\mu_1} q_k(z-a)^k$ of the Padé approximant.
7. If $d = 0$, then, for $k = 0, \ldots, \varkappa$, do:
a. Form the matrix $T^{(k+1)}_{\mu_1+1}$.
b. Find its rank: $\operatorname{rank} T^{(k+1)}_{\mu_1+1}$.
c. If $\operatorname{rank} T^{(k+1)}_{\mu_1+1} = \varkappa - 1$, then replace q_k in $(q_0, q_1, \ldots, q_{n-m+\mu_1})^T$ by zero.
Else **step 6** should be omitted.
8. For $k = 0, \ldots, m$, do $p_k = c_k q_0 + c_{k-1}q_1 + \cdots + c_{k-\varkappa}q_\varkappa$ and form the vector $(p_0, \ldots, p_m)^T$ consisting of the coefficients of the numerator $P_1(z) = \sum_{k=0}^{m} p_k(z-a)^k$ corresponding to $Q_1(z)$.
9. If $d = 0$, then, for $k = 0, \ldots, m$, do:
a. Form the matrix $T^{[k]}_{\mu_1+1}$.
b. Find its rank: $\operatorname{rank} T^{[k]}_{\mu_1+1}$.
c. If $\operatorname{rank} T^{[k]}_{\mu_1+1} = \varkappa$, then replace p_k in $(p_0, \ldots, p_m)^T$ by zero.
Else **step 9** should be omitted.
End of **Algorithm.**
Output: $f_{m,n}(z) = \frac{P_1(z)}{Q_1(z)}$.
Comments:

(1) Step 1, consisting of computing the Taylor coefficients, is very important for the algorithm. It is well known (see, for example, [3]) that accuracy in the given coefficients c_k is essential for Padé approximants. Usually most computing effort goes into calculation of the coefficients rather than Padé approximants, and so the coefficients should be calculated with the greatest possible accuracy. The way of calculating them depends essentially on the problem under consideration. For example, in the Padé–Laplace method, coefficients c_k are found from experimental data by numerical integration. So the accuracy in the coefficients depends not only on the chosen method of integration but also on the quality of the experiment.
(2) Step 5 is the control step. If the rank of T_{μ_1+1} is not equal to $\varkappa$, then the computation of the Padé approximant should be interrupted. The problem is most likely ill posed, and ranks are found wrongly. We face the problem of «not well-defined numerical rank», or the tolerance `tol` is chosen improperly.
(3) In step 3, we find the essential index μ_1 according to Proposition 3.3. The first essential polynomial, which is the minimal degree denominator of the Padé approximant, is found in step 6 according to Theorem 3.1. Theorems 4.2 and 4.3 are used in steps 7 and 9, respectively. The output approximant $f_{m,n}(z)$ possesses properties listed in Theorem 4.1.

We have implemented the algorithm in procedure *ReducedPade*(f, m, n, a, d) in *Maple* (the code is available at http://olga.6v6power.ru/soft), and now we would like to repeat Examples 2.3 and 2.4 from Section 2.

Example 5.1. Let us find by *ReducedPade* the diagonal (4, 4) Padé approximant $f_{4,4}$ for $f(z) = \frac{(z+1)(z-2)}{(z+2.1)(z-1)}$ at the point $z = 0$ via the following commands.
$f := \frac{(z+1)\cdot(z-2)}{(z+2.1)\cdot(z-1)}$:
$app := ReducedPade(f, 4, 4, 0, 0)$;

$$\frac{0.7773220744 + 0.3886610375z - 0.3886610371z^2}{0.8161881781 - 0.427527140654618832z - 0.388661037357416250z^2}$$

$fsolve(numer(app), z, complex)$;

$$-0.9999999998, \qquad 2.000000001$$

$fsolve(denom(app), z, complex)$;

$$-2.099999999, \qquad 1.000000000.$$

Example 5.2. Let us find the (2, 3) Padé approximant for the function $f(z) = \frac{z+1.01}{(z+2)(z-2.01)}$ at the point $z = 0$.
$f := \frac{z+1.01}{(z+2)\cdot(z-2.01)}$:
$app := ReducedPade(f, 2, 3, 0, 0)$;

$$\frac{-0.2438127455 - 0.2413987580z}{0.9704230069 + 0.00241398781766979931z - 0.241398758054323787z^2}$$

$fsolve(numer(app), z, complex)$;

-1.010000000

$fsolve(denom(app), z, complex)$;

$-1.999999999, \qquad 2.010000000.$

As can be seen, the approximants obtained are identical to the approximated functions, and Froissart doublets do not appear. Note that in the given examples the value of the parameter d was equal to zero. Hence, the vanishing coefficients of the numerators and the denominators were deleted. The next example shows that such deletion is desirable.

Example 5.3. Let us find the $(3, 7)$ Padé approximant for the function $f(z) = \frac{z+1.01}{z^4+3z^2-4.01}$ at the point $z = 0$.

$f := \frac{z+1.01}{z^4+3z^2-4.01}$:

$app := ReducedPade(f, 3, 7, 0, 0)$;

$$\frac{0.1977728838 + 0.1958147364z}{-0.7852170931 + 0.58744421453119533 0z^2 + 0.19581473767293827 8z^4}$$

$fsolve(numer(app), z, complex)$;

-1.010000000

$fsolve(denom(app), z, complex)$;

$-1.000999098, \qquad -2.000499738I, \qquad 2.000499738I, \qquad 1.000999098$

$app := ReducedPade(f, 3, 7, 0, 1)$;

$$(0.1977728838 + 0.1958147377z - 1.48862691\,10^{-11}z^3)/(-0.7852170931 - 5.237107247868522 11\,10^{-9}z + 0.58744421453119466 4z^2 - 1.13198619922094679\,10^{-9}z^3 + 0.19581473767293805 6z^4 - 1.67404646568558580\,10^{-10}z^5)$$

$fsolve(numer(app), z, complex)$;

$-1.146906047\,10^5, \qquad -1.009999994, \qquad 1.146916147\,10^5$

$fsolve(denom(app), z, complex)$;

$-1.000999095, \qquad -1.728850777\,10^{-9} - 2.000499738I, \qquad -1.728850777\,10^{-9} + 2.000499738I,$
$1.000999101, \qquad 1.169709095\,10^9.$

6. Conclusion

This work was motivated by our intention of avoiding the appearance of Froissart doublets in signal processing using the Padé–Laplace method. To do this, we have modified the classical definition of a Padé approximant by adding the requirement of the minimal degree of its denominator. It turned out that this new definition of a *reduced Padé approximant* allows us to avoid Froissart doublets induced by computer roundoff.

The reduced Padé approximant can be easily obtained by the proposed algorithm and always has nice properties which the classical Padé approximant possesses only in the normal case: the denominator is unique up to a constant factor, the numerator and the denominator do not have common non-zero roots, and their degrees and the deficiency index are known exactly.

Acknowledgments

It is a pleasure to thank Professor A.L. Shestakov for his encouragement and guidance. We acknowledge the reviewer's comments and suggestions which were valuable in improving the quality of our manuscript. We are very grateful to L.N. Trefethen, P. Gonnet, and S. Guttel for pointing out Gragg's paper [6] to us and for useful comments on a preliminary version of this paper that helped to improve the exposition.

References

[1] C. Brezinski, From numerical quadrature to Padé approximation, Appl. Numer. Math. 60 (2010) 1209–1220.
[2] S.P. Suetin, Padé approximants and efficient analytic continuation of a power series, Russian Math. Surveys 57 (1) (2002) 43–141.
[3] G.A. Baker Jr., P.R. Graves-Morris, Padé Approximants, second ed., Cambridge University Press, Cambridge, 1996.

[4] K.O. Geddes, Symbolic computation of Padé approximants, ACM Trans. Math. Softw. 5 (2) (1979) 218–233.
[5] A. Bultheel, Recursive algorithms for nonnormal Padé tables, SIAM J. Appl. Math. 39 (1) (1980) 106–118.
[6] W.B. Gragg, The Padé table and its relation to certain algorithms of numerical analysis, SIAM Rev. 14 (1) (1972) 1–62.
[7] J. Gilewicz, Y. Kryakin, Froissart doublets in Padé approximation in the case of polynomial noise, J. Comput. Appl. Math. 153 (2003) 235–242.
[8] P. Claverie, A. Denis, E. Yeramian, The representation of functions through the combined use of integral transforms and Padé approximants: Padé–Laplace analysis of functions as sums of exponentials, Comput. Phys. Rep. 9 (1989) 247–299.
[9] V.M. Adukov, Generalized inversion of block Toeplitz matrices, Linear Algebra Appl. 274 (1998) 85–124.
[10] L.N. Trefethen, P. Gonnet, S. Guttel, Robust Padé approximation via SVD, SIAM Rev. (in press).
[11] G.L. Xu, A. Bultheel, Matrix Pade approximation: definitions and properties, Linear Algebra Appl. 137–138 (1990) 67–136.
[12] C. Pestano-Gabino, C. Gonzáles-Concepción, M.C. Gil-Fariña, A type of matrix Padé approximant inspired by scalar component models, J. Comput. Appl. Math. 236 (2012) 3360–3372.
[13] G.H. Golub, C.F. Van Loan, Matrix Computations, third ed., JHU, Baltimore, MD, 1996.
[14] G.H. Golub, V. Klema, G.W. Stewart, Rank degeneracy and least squares problems, Tech. Rep. TR 456, 1976.
[15] L.V. Foster, Rank and null space calculations using matrix decomposition without column interchanges, Linear Algebra Appl. 74 (1986) 47–71.
[16] G.W. Stewart, Determining rank in the presence of errors, Tech. Rep. TR-92-108, 1992.
[17] O.L. Ibryaeva, A new algorithm for calculating Padé approximants and its Matlab implementation, in: Proc. of South Ural State University. Ser. of Math. Simul. and Program., vol. 37, No. 254, 2011, pp. 99–107 (in Russian).