Partial-Padé prediction

Marc Prévost, Denis Vekemans

Partial-Padé prediction

M. Prévost, D. Vekemans *

June 9, 2006

Abstract

When the first terms of a sequence (called sequence to predict) are known, a prediction method is a method which gives us an approximation of the following terms (the so-constructed sequence is called the predicted sequence). In this paper, we expose two prediction methods respectively called ε_p-prediction which are generalizations of Aitken's Δ^2-prediction of C. Brezinski and M. Redivo Zaglia [6]- and Padé-prediction of J. Gilewicz [8]- which are very simple to use. In order to choose among the different partial-Padé-predictions, we study some of their properties.

However, the most important points of this paper are :

- the use of an extrapolation algorithm,(the ε-algorithm) to obtain a prediction algorithm for each partial-Padé-prediction (which avoids to solve a system).
- the results about consistency obtained for the partial-Padé-prediction , (i.e. under certain conditions, each term of the predicted sequence converges to the analogous term of the sequence to predict).

Introduction

The domains of application of the prediction methods are various : they are used in econometry, in statistics, in the solution of parabolic problems (see M. Morandi Cecchi, M. Redivo Zaglia and G. Scenna [10]), in the study of band structure in semi-conductors (see G. Allan [1] or A. Trias, M. Kiwi and M. Weissmann [16]), ...

The essential theme of this work is the systematical use of extrapolation methods to construct prediction ones. It is well known that from the knowledge of the first terms of a sequence (i.e. $S_0, S_1, \ldots, S_N$), we can estimate its limit S by using extrapolation methods. However, as we will see in the sequel, we can also predict the following terms of this sequence (i.e. $S_{N+1}, S_{N+2}, \ldots$).

We call prediction method a process that transforms a vector $(S_i)_{0\leq i\leq N}$ in a sequence $(S_{i,N})_{i\geq 0}$ that reproduces the terms of the vector, i.e. such that $S_i = S_{i,N}$, $\forall i \in \{0, 1, \ldots, N\}$. In this case, we will say that the sequence $(S_{i,N})_{i\geq 0}$ is a predicted sequence of the vector $(S_i)_{0\leq i\leq N}$.

The use of extrapolation process to construct a prediction method is an idea that has ever been used by using the E-algorithm (see C. Brezinski [4], C. Brezinski and M. Redivo Zaglia [6, pp. 392–395] and D. Vekemans [18]). However, this method (called E-prediction) has a drawback : for a good use, it needs the knowledge of a scale of comparison on which the sequence to predict has an aymptotic expansion.

*Laboratoire de Mathématiques Appliquées, Université du Littoral, Centre Universitaire de la Mi-Voix, Bâtiment H. Poincaré, 50 Rue F. Buisson B.P. 699, 62228 Calais Cédex, France, –e-mail prevost@lma.univ-littoral.fr, fax (33) 03 21 46 36 69

We begin this paper with two prediction methods respectively called ε_p-prediction and Padé-prediction. These two methods do not require a particular knowledge on the sequence to predict and are easy to use. The ε_p-prediction is built from the ε-algorithm, as the E-prediction is constructed from the E-algorithm. However, starting from an algebraic point of view, C. Brezinski and M. Redivo Zaglia [6, pp. 389–395] have defined Aitken's Δ^2-prediction which is a particular case of the ε_p-prediction, as Aitken's Δ^2-process is a particular case of the ε-algorithm. The Padé-prediction has already been studied by J. Gilewicz [8, pp. 424–439]. Moreover, we show how the Padé-prediction can be computed with the ε-algorithm.

In section 2, we generalize the Padé-prediction into the partial-Padé-prediction. The advantage of the partial-Padé-prediction is that we can fix the analytic behavior of the predicted sequence (for example, having a non-zero limit, or having a geometric asymptotic behaviour, ...). The link between the Padé-prediction and the partial-Padé-prediction leads to an algorithm for each partial-Padé-prediction based on the ε-algorithm.

As acceleration results are very important for extrapolation processes, we will establish Theorems concerning consistency of prediction methods. These are primordial in this analysis because the definition chosen for a prediction method is very weak.

At the end, this paper contains two numerical examples. The first one is intended to discuss the choice among prediction methods and to observe a numerical consistency. The second one presents the case of a non-convergent sequence.

In our paper, $\mathbb{K}=\mathbb{R}$ or $\mathbb{C}$.

1 The ε_p-prediction

As we did for the E-prediction, built from the E-algorithm (see D. Vekemans [18]), we can construct the ε_p-prediction by using the ε-algorithm. It is well known that the ε-algorithm is a particular case of the E-algorithm and we will naturally get that the ε_p-prediction is a particular case of the E-prediction (see D. Vekemans [17, p. 21]).

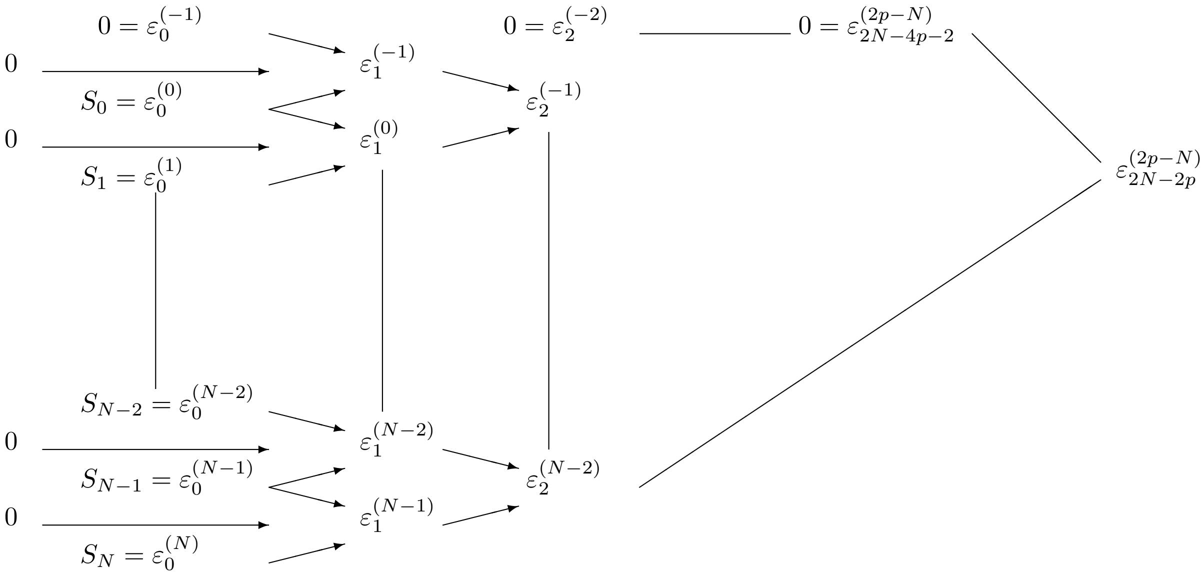
Let us remark here that other extrapolating algorithm, concisely referred in [6, pp. 57–58], could also be used to construct prediction methods, like the Richardson's process, the G-transform or summation processes. The so-obtained prediction methods are also particular cases of the E-prediction.

There are few papers dealing with the ε_p-prediction. In [6, pp. 389–395], C. Brezinski and M. Redivo Zaglia give a definition of the Δ^2-prediction which is a particular case of the ε_p-prediction (as the Aitken's Δ^2 process is a particular case of the ε-algorithm).

The ε-algorithm is an extrapolation algorithm which is due to P. Wynn [19]. For the definition of the ε-algorithm and results about its acceleration properties, the interested reader is referred to [4], [6], [8], [14], [19] or [20].

1.1 The algorithmic point of view of the ε_p-prediction

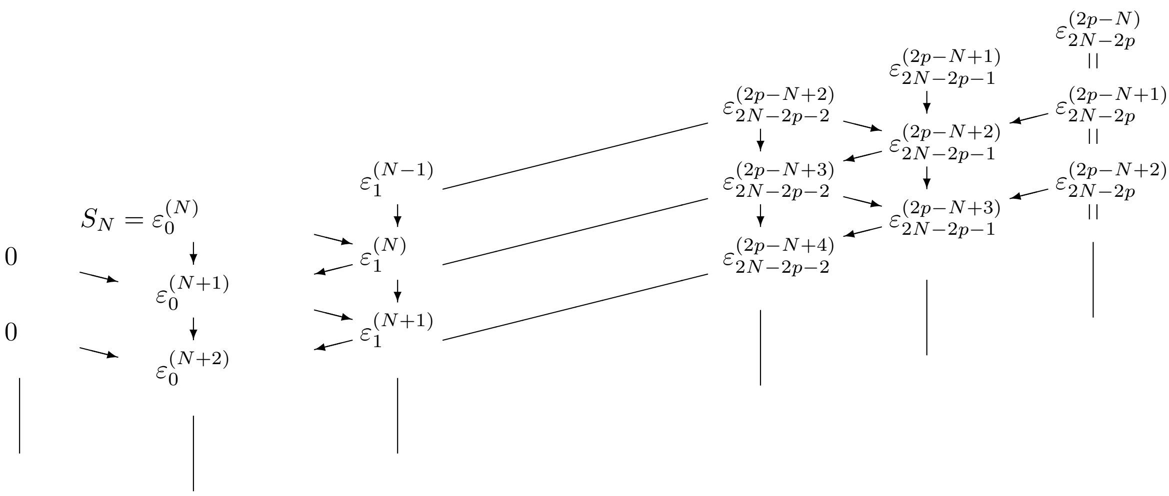
Let $(S_i)_{0\leq i\leq N} \in \mathbb{K}^{N+1}$ be a vector to predict and $p \in \{0, 1, \ldots, N\}$.
We form Table (1) :



with the classical rule of the ε-algorithm :

$$\left\{\begin{array}{l} \varepsilon_{-1}^{(i)} = 0,\ \forall i \in \{max(0,2p-N), max(0,2p-N)+1,\ldots,N\} \\ \varepsilon_0^{(i)} = S_i,\ \forall i \in \{max(0,2p-N), max(0,2p-N)+1,\ldots,N\} \\ \varepsilon_{2(i-1)}^{(-i)} = 0,\ \forall i \in \{1,2,\ldots,N-2p\} \\ \text{and } \varepsilon_j^{(i)} = \varepsilon_{j-2}^{(i+1)} + \dfrac{1}{\varepsilon_{j-1}^{(i+1)} - \varepsilon_{j-1}^{(i)}},\ \forall j \in \{1,2,\ldots,2N-2p\}, \\ \qquad \forall i \in \{max(2p-N,-1-[\frac{j-1}{2}]), max(2p-N,-1-[\frac{j-1}{2}])+1,\ldots,N-j\} \\ \qquad \text{(where, } \forall x \in I\!R,\ [x] \in I\!N \text{ is such that } [x] \le x < [x]+1\ ). \end{array}\right.$$

Then we form Table (2) (it is a continuation of Table (1)):



by using the progressive rule of the ε-algorithm :

$$\left\{\begin{array}{l} \varepsilon_{2N-2p}^{(i)} = \varepsilon_{2N-2p}^{(2p-N)},\ \forall i \ge 2p-N+1 \\ \text{(the } \varepsilon-\text{prediction is obtained by giving the same value to each term of the right most column} \\ \text{and then going down in the table with the progressive rule of the } \varepsilon-\text{algorithm)} \\ \varepsilon_{-1}^{(i)} = 0,\ \forall i \ge N+1 \\ \text{and, } \varepsilon_j^{(i)} = \varepsilon_j^{(i-1)} + \dfrac{1}{\varepsilon_{j+1}^{(i-1)} - \varepsilon_{j-1}^{(i)}},\ \forall j \in \{0,1,\ldots,2N-2p-1\},\ \forall i \ge N-j+1. \end{array}\right.$$

We algorithmically define the ε_p-predicted sequence $(S_{i,N}^{(\varepsilon,p)})_{i\geq 0}$ from the vector $(S_i)_{0\leq i\leq N}$ by

$$S_{i,N}^{(\varepsilon,p)} = \varepsilon_0^{(i)},\ \forall i \geq max(0, 2p-N),$$

and

$$S_{i,N}^{(\varepsilon,p)} = S_i,\ \forall i \in \{0, 1, \ldots, 2p-N-1\}.$$

Note that $S_{i,N}^{(\varepsilon,p)} = S_i,\ \forall i \in \{0, 1, \ldots, N\}$ which implies that $(S_{i,N}^{(\varepsilon,p)})_{i\geq 0}$ is a predicted sequence of the vector $(S_i)_{0\leq i\leq N}$. We use the notation $(S_{i,N}^{(\varepsilon,p)})_{i\geq 0}$ with an upper right index ε because of the name of this prediction method and an upper right index p because this prediction method depends on this parameter.

Remarks

1. When $N = 2K$ and $p = K$, the ε_p-prediction will be called the ε-prediction. Then, the ε_p-prediction can be thought either as the ε-prediction of the initial vector to predict preceded by $N-2p$ zeros (if $N-2p > 0$) or as the ε-prediction of the initial vector to predict truncated of its first terms (if $N-2p < 0$).
2. The algorithm of the ε_p-prediction may be break down in some particular cases : for example, if in Table (1), the column of index N_0 ($N_0 \in \{0, 1, \ldots, N-2p\}$) is constant and equal to S, (because of division by zero) it is impossible to generate columns of index strictly greater than N_0. In this case, we may assume that the column of index N_0 in Table (2) is constant and equal to S in order to continue.
3. In the case of singularities (division by zero), the particular rules of ε-algorithm can be used (see [6, pp. 34–38] or [7]).
4. Instead of using the classical and the progressive rules of the ε-algorithm, we can use the cross rule of P. Wynn [20].

1.2 The algebraic point of view of the ε_p-prediction

Let $(S_i)_{0\leq i\leq N} \in I\!\!K^{N+1}$ be a vector to predict and $p \in \{0, 1, \ldots, N\}$.

We define $(b_0, b_1, \ldots, b_{N-p})$ as the solution (supposed to be existing) of the following system of linear equations:

$$\left\{\begin{array}{rcl} S_p &=& b_0 S_{2p-N} + b_1 S_{2p-N+1} + \ldots + b_{N-p-1} S_{p-1} + b_{N-p} \\ S_{p+1} &=& b_0 S_{2p-N+1} + b_1 S_{2p-N+2} + \ldots + b_{N-p-1} S_p + b_{N-p} \\ &\vdots& \\ S_N &=& b_0 S_p + b_1 S_{p+1} + \ldots + b_{N-p-1} S_{N-1} + b_{N-p} \end{array}\right.$$

with $S_i = 0$ when $i < 0$.

We algebraically define the ε_p-predicted sequence $(S_{i,N}^{(\varepsilon S,p)})_{i\geq 0}$ from the vector $(S_i)_{0\leq i\leq N}$ by

$$S_{i,N}^{(\varepsilon S,p)} = S_i,\ \forall i \in \{0, 1, \ldots, N\}$$

and the further terms of the predicted sequence by

$$S_{N+i,N}^{(\varepsilon S,p)} = b_0 S_{p+i,N}^{(\varepsilon S,p)} + b_1 S_{p+i+1,N}^{(\varepsilon S,p)} + \ldots + b_{N-p-1} S_{N+i-1,N}^{(\varepsilon S,p)} + b_{N-p},\ \forall i \geq 1.$$

We use the notation $(S_{i,N}^{(\varepsilon S,p)})_{i\geq 0}$ with an upper right index ε because of the name of this prediction method, an upper right index p because the prediction method depends on this parameter and an upper right index S because it needs solving of a linear system.

Theorem 1 *[17] Let $(S_i)_{0\leq i\leq N} \in I\!\!K^{N+1}$ be the vector to predict. If the quantities in the above equality are defined, then*

$$S_{i,N}^{(\varepsilon,p)} = S_{i,N}^{(\varepsilon S,p)},\ \forall i \in I\!\!N,$$

This Theorem establishes that the algorithmic and the algebraic points of view of the ε_p-prediction are the same.

Remark

In the algorithmic point of view of the ε_p-prediction, we choose to give the same value to the right most column. This choice leads to simplicity for the algebraic point of view. We could also choose to equal the right most column to a non constant sequence (for example, if we know the limit of the predicted sequence, we could equal the right most column to a sequence converging to the same limit). But, in this case, we surely lost the simplicity of the algebraic point of view.

From Cramer's identity, in the following Theorem, we can write $S_{N+i,N}^{(\varepsilon S,p)}$, $\forall i \in I\!\!N$, as a ratio of two determinants.

Theorem 2 *Let $(S_i)_{0\leq i\leq N} \in I\!\!K^{N+1}$ be the vector to predict, then*

$$S_{N+i,N}^{(\varepsilon S,p)} = -\frac{\begin{vmatrix} 0 & 1 & S_{p+i,N}^{(\varepsilon S,p)} & S_{p+i+1,N}^{(\varepsilon S,p)} & \cdots & S_{N+i-1,N}^{(\varepsilon S,p)} \\ S_p & 1 & S_{2p-N} & S_{2p-N+1} & \cdots & S_{p-1} \\ S_{p+1} & 1 & S_{2p-N+1} & S_{2p-N+2} & \cdots & S_p \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ S_N & 1 & S_p & S_{p+1} & \cdots & S_{N-1} \end{vmatrix}}{\begin{vmatrix} 1 & S_{2p-N} & S_{2p-N+1} & \cdots & S_{p-1} \\ 1 & S_{2p-N+1} & S_{2p-N+2} & \cdots & S_p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & S_p & S_{p+1} & \cdots & S_{N-1} \end{vmatrix}},\ \forall i \in I\!\!N,$$

with the convention $S_i = 0$ when $i < 0$.

1.3 Theoretical examples

We here give the expression of the terms of the ε_p-predicted sequence for small values of $N-p$.

First example

The Δ^2-prediction (when $N-p=1$) :

$S_{i,N}^{(\varepsilon,p)} = \dfrac{S_{N-2}S_N - S_{N-1}^2}{S_{N-2}-S_{N-1}} - S_{i-1,N}^{(\varepsilon,p)}\dfrac{S_N - S_{N-1}}{S_{N-2}-S_{N-1}}$ for $i \geq N+1$, when $S_{N-2}-S_{N-1} \neq 0$ and $S_{i,N}^{(\varepsilon,p)} = S_i$ for $i \in \{0,1,\ldots,N\}$ with $S_i = 0$ when $i < 0$.

Second example

When $N-p=2$:

$$S_{i,N}^{(\varepsilon,p)} = \frac{S_{N-4}S_{N-2}S_N + 2S_{N-3}S_{N-2}S_{N-1} - S_{N-1}^3 - S_{N-3}^2 S_N - S_{N-4}S_{N-1}^2}{S_{N-4}S_{N-2} + S_{N-3}S_{N-1} + S_{N-3}S_{N-2} - S_{N-2}^2 - S_{N-3}^2 - S_{N-4}S_{N-1}}$$

$$- S_{i-1,N}^{(\varepsilon,p)} \frac{S_{N-4}S_{N-3} + S_{N-2}S_{N-1} + S_{N-3}S_{N-2} - S_{N-4}S_{N-1} - S_{N-3}S_N - S_{N-2}^2}{S_{N-4}S_{N-2} + S_{N-3}S_{N-1} + S_{N-3}S_{N-2} - S_{N-2}^2 - S_{N-3}^2 - S_{N-4}S_{N-1}}$$
$$- S_{i-2,N}^{(\varepsilon,p)} \frac{S_{N-3}S_{N-1} + S_{N-2}S_N + S_{N-2}S_{N-1} - S_{N-3}S_N - S_{N-1}^2 - S_{N-2}^2}{S_{N-4}S_{N-2} + S_{N-3}S_{N-1} + S_{N-3}S_{N-2} - S_{N-2}^2 - S_{N-3}^2 - S_{N-4}S_{N-1}} \text{ for } i \geq N+1,$$

when $S_N S_{N-2} + S_{N-3}S_{N-1} + S_{N-3}S_{N-2} - S_{N-2}^2 - S_{N-3}^2 - S_{N-4}S_{N-1} \neq 0$,
and $S_{i,N}^{(\varepsilon,p)} = S_i$ for $i \in \{0, 1, \dots, N\}$ with $S_i = 0$ when $i < 0$.

1.4 Exponential extrapolation and ε_p-prediction

Transformation of systems of exponential equations into linear ones is known as Prony's method (see [9, pp. 272–280]).

Theorem 3 *[17]*

Let $(S_i)_{0 \leq i \leq N} \in \mathbb{K}^{N+1}$ be the vector to predict.

If we define $(T_{i,N})_{i \geq 0}$ by

$$T_{i,N} = a_1 b_1^i + \dots + a_{N-p} b_{N-p}^i + c_1, \ \forall i \in \mathbb{N},$$

where $(a_1, a_2, \dots, a_{N-p}, b_1, b_2, \dots, b_{N-p}, c_1)$ is determined as a solution (supposed to exist) of the system of exponential equations :

$$\left\{ \begin{array}{rcl} S_{2p-N} & = & a_1 b_1^{2p-N} + \dots + a_{N-p} b_{N-p}^{2p-N} + c_1 \\ S_{2p-N+1} & = & a_1 b_1^{2p-N+1} + \dots + a_{N-p} b_{N-p}^{2p-N+1} + c_1 \\ & \vdots & \\ S_N & = & a_1 b_1^N + \dots + a_{N-p} b_{N-p}^N + c_1 \end{array} \right. ,$$

with the convention $S_i = 0$ when $i < 0$, then

$$T_{i,N} = S_{i,N}^{(\varepsilon,p)}, \ \forall i \in \mathbb{N},$$

where $S_{i,N}^{(\varepsilon,p)}$ is defined in Section 1.1.

In the above Theorem, $T_{i,N}$ is a linear combination of exponential terms that are $a_k b_k^i$'s, but it also contains an additional constant term c_1 which allows the predicted sequence $(T_{i,N})_{i \geq 0}$ to converge not only to zero or infinity contrarily to the Padé-prediction (see Section 2.4).

By considering the prediction methods as exponential extrapolation, we easily exhibit the analytic behavior of the predicted sequence, and when we want to choose between prediction methods, that is what we have to look at.

2 The Padé-prediction

The Padé-prediction has been studied by J. Gilewicz [8, chap. 8]. This prediction method has been obtained from Padé approximants. In this Section, we will recall the definition of the Padé-prediction and give some of its properties.

2.1 The Padé approximants

The Padé approximants have extensively been studied by H. Padé [11]. The properties of these approximants can be found in [8], [2] or in [3, p. 35], for the formal case.

If f is a formal series $f(t) = \sum_{i \geq 0} c_i t^i$, where $c_i \in \mathbb{C}$, $\forall i \in \mathbb{N}$, then the Padé approximant $[n/m]_f$ of f is defined by $f(t) - [n/m]_f(t) = O(t^{m+n-1})$, where $[n/m]_f$ is a rational function whose degree of numerator is lower or equal to n and whose degree of denominator is lower or equal to m.

2.2 Definition of the Padé-prediction

The Padé-predictionhas ever been defined and studied by J. Gilewicz [8, chap. 8].

Let $(S_i)_{0 \leq i \leq N} \in \mathbb{K}^{N+1}$ be the vector to predict. We set

$$P_N(t) = \sum_{i=0}^{N} S_i t^i \in \mathbb{K}_N[t].$$

Then, let $\dfrac{w(t)}{v(t)}$ (assumed to exist), be the Padé approximant [p/N-p] of $P_N(t)$.

J. Gilewicz defined the Padé-predicted sequence $(S_{i,N}^{(p)})_{i \geq 0}$ from the vector $(S_i)_{0 \leq i \leq N}$ such that, formally,

$$\sum_{i \geq 0} S_{i,N}^{(p)} t^i = \frac{w(t)}{v(t)}.$$

We use the notation $(S_{i,N}^{(p)})_{i \geq 0}$ where the upper index p is the degree of the numerator of the Padé approximant used.

The computation of $(S_{i,N}^{(p)})_{i \geq 0}$ is achieved by solving a system of $N - p$ linear equations. But, this solution of linear system can be avoided by using the ε-algorithm as explained in Section 2.5.

Remark
If $2p - N \geq 0$, $S_{i,N}^{(p)}$ is independent for $i \geq N + 1$ of the values of $S_0, S_1, \ldots, S_{2p-N-1}$ and S_{2p-N}.

2.3 Theoretical examples

We give below the expressions of the terms of the Padé-predicted sequence for small values of $N - p$.

First example
When $N - p = 1$,
$S_{i,N}^{(p)} = S_{i-1,N}^{(p)} \dfrac{S_N}{S_{N-1}}$ for $i \geq N + 1$, when $S_{N-1} \neq 0$
and $S_{i,N}^{(p)} = S_i$ for $i \in \{0, 1, \ldots, N\}$ with $S_i = 0$ when $i < 0$.

Second example
When $N - p = 2$,

$$S_{i,N}^{(p)} = S_{i-1,N}^{(p)}\frac{S_{N-1}}{S_{N-2}} - S_{i-2,N}^{(p)}\frac{S_{N-1}^2 - S_{N-2}S_N}{S_{N-2}^2} \text{ for } i \geq N+1, \text{ when } S_{N-2} \neq 0,$$

and $S_{i,N}^{(p)} = S_i$ for $i \in \{0, 1, \ldots, N\}$ with $S_i = 0$ when $i < 0$.

2.4 Exponential extrapolation and Padé-prediction

The exponential extrapolation (see [9, pp. 272–280]) and the Padé-prediction are simply connected by the following Theorem :

Theorem 4 *[17]*

Let $(S_i)_{0 \leq i \leq N} \in I\!\!K^{N+1}$ be the vector to predict.

If we define $(T_{i,N})_{i \geq 0}$ by

$$T_{i,N} = a_1 b_1^i + \ldots + a_{N-p} b_{N-p}^i, \ \forall i \in I\!\!N,$$

where $(a_1, a_2, \ldots, a_{N-p}, b_1, b_2, \ldots, b_{N-p})$ is determined as a solution (supposed to exist) of the system of exponential equations :

$$\begin{cases} S_{2p-N+1} &= a_1 b_1^{2p-N+1} + \ldots + a_{N-p} b_{N-p}^{2p-N+1} \\ S_{2p-N+2} &= a_1 b_1^{2p-N+2} + \ldots + a_{N-p} b_{N-p}^{2p-N+2} \\ & \vdots \\ S_N &= a_1 b_1^N + \ldots + a_{N-p} b_{N-p}^N \end{cases},$$

with the convention $S_i = 0$ when $i < 0$, then

$$T_{i,N} = S_{i,N}^{(p)}, \ \forall i \in I\!\!N.$$

2.5 Link between the ε_p-prediction and the Padé-prediction

Let us recall here that the calculation of a Padé approximant can be achieved by the ε-algorithm. This result comes out from the expression of the terms of the ε-algorithm as a ratio of two determinants. This last property has been established by Shanks [14].

Theorem 5 *Let $f(t) = \sum_{i \geq 0} c_i t^i$ be a formal series.*
We set
$\varepsilon_{-1}^{(i)} = 0, \ \forall i \in I\!\!N,$
$\varepsilon_{2i}^{(-i-1)} = 0, \ \forall i \in I\!\!N,$
$\varepsilon_0^{(i)} = \sum_{j=0}^{i} c_j t^j, \ \forall i \in I\!\!N,$
and we define $\varepsilon_{k+1}^{(i)}$ by

$$\varepsilon_{k+1}^{(i)} = \varepsilon_{k-1}^{(i+1)} + \frac{1}{\varepsilon_k^{(i+1)} - \varepsilon_k^{(i)}}, \ \forall k \in I\!\!N, \ \forall i \geq -1 - [\frac{k}{2}].$$

Then

$$\varepsilon_{2k}^{(n)} = [n + k/k]_{f(t)}, \ \forall k \in I\!\!N, \ \forall n \geq -k.$$

The above Theorem shows that the ε-algorithm applied to the partial sums of a series leads to the Padé table. The following Theorem can be viewed as a corollary of the preceding one. It allows to avoid solving of a system of linear equations in the computation of the Padé-prediction , only by using the ε-algorithm.

Theorem 6 *Let $(S_i)_{0\le i\le N} \in I\!K^{N+1}$ be the vector to predict.*
Let $(T_i)_{0\le i\le N} \in I\!K^{N+1}$ be given by $T_i = \sum_{j=0}^{i} S_j,\ \forall i \in \{0,1,\dots,N\}$, then

$$S_{i,N}^{(p)} = T_{i,N}^{(\varepsilon,p)} - T_{i-1,N}^{(\varepsilon,p)},\ \forall i \ge 1,$$

and

$$S_{0,N}^{(p)} = T_{0,N}^{(\varepsilon,p)}$$

where $(S_{i,N}^{(p)})_{i\ge 0}$ is the Padé-predicted sequence of $(S_i)_{0\le i\le N} \in I\!K^{N+1}$ and

$(T_{i,N}^{(\varepsilon,p)})_{i\ge 0}$ is the ε_p-predicted sequence of $(T_i)_{0\le i\le N} \in I\!K^{N+1}$.

Proof

We set
$g(t) = \sum_{i\ge 0} S_{i,N}^{(p)} t^i = [p/N-p]_f(t)$, with $f(t) = \sum_{i=0}^{N} S_i t^i$
$\varepsilon_{-1}^{(i)} = 0,\ \forall i \in I\!N$,
$\varepsilon_{2i}^{(-i-1)} = 0,\ \forall i \in I\!N$,
$\varepsilon_0^{(i)} = \sum_{j=0}^{i} S_{j,N}^{(p)} t^j,\ \forall i \in I\!N$,
and we define $\varepsilon_{k+1}^{(i)}$ by the classical rule of the ε-algorithm. Consequently, from Theorem 5,

$$\varepsilon_{2N-2p}^{(2p-N)} = [p/N-p]_f(t),\ \forall p \in I\!N,\ \forall N \ge p,$$

and

$$\varepsilon_{2N-2p}^{(2p-N+i)} = [p+i/N-p]_f(t),\ \forall p \in I\!N,\ \forall N \ge p,\ \forall i \in I\!N.$$

However $[p+i/N-p]_g(t) = [p/N-p]_g(t)$, because g is a rational function of degree (p/N-p).
For $t=1$, we have

$$\varepsilon_{2N-2p}^{(2p-N+i)} = [p+i/N-p]_g(1) = [p/N-p]_g(1) = \varepsilon_{2N-2p}^{(2p-N)},\ \forall p \in I\!N,\ \forall N \ge p,\ \forall i \in I\!N.$$

Thus the column of index $2N-2p$ is constant and this shows that

$$S_{i,N}^{(p)} = T_{i,N}^{(\varepsilon,p)} - T_{i-1,N}^{(\varepsilon,p)},\ \forall i \ge 1,$$

and

$$S_{0,N}^{(p)} = T_{0,N}^{(\varepsilon,p)}.$$

The conclusion is that the Padé-prediction can be computed by the ε-algorithm.

3 The partial-Padé-prediction

3.1 The partial-Padé approximants

Partial-Padé approximants have been defined by C. Brezinski in [5] and by M. Prévost in [12]. We give here a brief insight.

Definition
Let $f(t) = \sum_{i \geq 0} c_i t^i$ be a formal series.
We consider $x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_s$, elements of $I\!\!K$ and we set

$$x(t) = (1 - x_1 t)(1 - x_2 t) \ldots (1 - x_r t) \in I\!\!K_r[t]$$

and

$$y(t) = (1 - y_1 t)(1 - y_2 t) \ldots (1 - y_s t) \in I\!\!K_s[t].$$

We define $(v(t), w(t)) \in I\!\!K_q[t] \times I\!\!K_p[t]$ (supposed to exist) such that

$$\begin{cases} v(0) = 1 \\ gcd(v(t), w(t)) = 1 \\ \dfrac{w(t)}{v(t)} = \dfrac{y(t)}{x(t)} f(t) + O(t^{p+q+1}) \end{cases}.$$

We will say that $\dfrac{w(t)x(t)}{v(t)y(t)}$ is the partial-Padé approximant [p/q] of $f(t)$ calculated from the polynomials $x(t)$ and $y(t)$.

Remark
The partial-Padé approximant [p/q] of $f(t)$ calculated from the polynomials $x(t)$ and $y(t)$ is exactly the Padé approximant [p/q] of $\dfrac{y(t)}{x(t)} f(t)$.

3.2 Definition of the partial-Padé-prediction

Let $(S_i)_{0 \leq i \leq N} \in I\!\!K^{N+1}$ be the vector to predict. We set

$$P_N(t) = \sum_{i=0}^{N} S_i t^i \in I\!\!K_N[t].$$

Let $x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_s$ be elements of $I\!\!K$ and let $\dfrac{w(t)x(t)}{v(t)y(t)}$ be the partial-Padé approximant [p/q] of $f(t)$ calculated from the polynomials $x(t)$ and $y(t)$.

We define the partial-Padé-predicted sequence $(S_{i,N}^{(p)})_{i \geq 0}$ calculated from the polynomials $x(t)$ and $y(t)$ for the vector $(S_i)_{0 \leq i \leq N}$ to predict such that, formally,

$$\sum_{i \geq 0} S_{i,N}^{(p)} t^i = \frac{w(t)x(t)}{v(t)y(t)}.$$

We use the same notation $(S_{i,N}^{(p)})_{i \geq 0}$ as for the Padé-predicted sequence, but we always specify for which polynomials this sequence is calculated.

Particular cases of the partial-Padé-prediction

- **The Padé-prediction.**
  When $x(t) = y(t) = 1$, the partial-Padé-prediction calculated from these polynomials is exactly the Padé-prediction.
- **The higher-Padé-type-prediction.**
  When $x(t) = 1$, the partial-Padé-prediction calculated from the polynomials $x(t)$ and $y(t)$ is called the higher-Padé-type-prediction calculated with generating polynomial $y(t)$.

- **The t-prediction** [15] A. Sidi and D. Levin used the Padé-type prediction with generating polynomials depending on the vector to predict. They also get results of consistency for this method.

3.3 Exponential extrapolation and higher-Padé-type-prediction

In this section we prove a Theorem generalizing Theorems 3 and 4. Moreover, from this Theorem one could conclude that the Padé-prediction and the ε_p-prediction are both particular cases of the higher-Padé-type-prediction.

Theorem 7 *[17]*

Let $(S_i)_{0\leq i\leq N} \in I\!\!K^{N+1}$ be the vector to predict.
We form the partial-Padé-predicted sequence $(S_{i,N}^{(p)})_{i\geq 0}$ calculated from the polynomials $x(t) = 1$ and $y(t) = (1-y_1t)^{\alpha_1}(1-y_2t)^{\alpha_2}\ldots(1-y_\sigma t)^{\alpha_\sigma}$ ($y_i \neq y_j$ when $i \neq j$, $\alpha_i \in I\!\!N$ and $s = \sum_{i=1}^{\sigma}\alpha_i$).
If we define $(T_{i,N})_{i\geq 0}$ by

$$\begin{cases} T_{i,N} = S_i, \ \forall i \in \{0,1,\ldots,2p-N-s\} \\ T_{i,N} = a_1b_1^i + \ldots + a_{N-p}b_{N-p}^i \\ +\sum_{l=1}^{\sigma}(\beta_{l,0} + \beta_{l,1}i + \ldots + \beta_{l,\alpha_l-1}i^{\alpha_l-1})y_l^i, \ \forall i \geq 2p-N-s+1 \end{cases}$$

where $(a_1, a_2, \ldots, a_{N-p}, \beta_{1,0}, \beta_{1,1}, \ldots, \beta_{1,\alpha_1-1}, \beta_{2,0}, \ldots, \beta_{\sigma-1,\alpha_{\sigma-1}-1}, \beta_{\sigma,0}, \beta_{\sigma,1}, \ldots, \beta_{\sigma,\alpha_\sigma-1})$ is determined as a solution (supposed to exist) of the system of exponential equations :

$$\begin{cases} S_{2p-N-s+1} &= a_1b_1^{2p-N-s+1} + \ldots + a_{N-p}b_{N-p}^{2p-N-s+1} \\ & +\sum_{l=1}^{\sigma}(\beta_{l,0} + \beta_{l,1}(2p-N-s+1) + \ldots + \beta_{l,\alpha_l-1}(2p-N-s+1)^{\alpha_l-1})y_l^{2p-N-s+1} \\ S_{2p-N-s+2} &= a_1b_1^{2p-N-s+2} + \ldots + a_{N-p}b_{N-p}^{2p-N-s+2} \\ & +\sum_{l=1}^{\sigma}(\beta_{l,0} + \beta_{l,1}(2p-N-s+2) + \ldots + \beta_{l,\alpha_l-1}(2p-N-s+2)^{\alpha_l-1})y_l^{2p-N-s+2} \\ & \vdots \\ S_N &= a_1b_1^N + \ldots + a_{N-p}b_{N-p}^N \\ & +\sum_{l=1}^{\sigma}(\beta_{l,0} + \beta_{l,1}N + \ldots + \beta_{l,\alpha_l-1}N^{\alpha_l-1})y_l^N \end{cases},$$

with the convention $S_i = 0$ when $i < 0$, then

$$T_{i,N} = S_{i,N}^{(p)}, \ \forall i \in I\!\!N.$$

Proof

We have

$$\begin{aligned} P_N(t) &= \sum_{i=0}^{N} S_it^i \\ &= \sum_{i=0}^{2p-N-s} S_it^i + \sum_{i=2p-N-s+1}^{N} S_it^i \\ &= \sum_{i=0}^{2p-N-s} S_it^i + \sum_{i=2p-N-s+1}^{N}(a_1b_1^i + \ldots + a_{N-p}b_{N-p}^i \\ &\quad +\sum_{l=1}^{\sigma}(\beta_{l,0} + \beta_{l,1}i + \ldots + \beta_{l,\alpha_l-1}i^{\alpha_l-1})y_l^i)t^i \\ &= \sum_{i=0}^{2p-N-s} S_it^i + t^{2p-N-s+1}\sum_{i=0}^{2N+s-2p-1}(a_1b_1^{i+2p-N-s+1} + \ldots + a_{N-p}b_{N-p}^{i+2p-N-s+1} \\ &\quad +\sum_{l=1}^{\sigma}(\beta_{l,0} + \beta_{l,1}(i+2p-N-s+1) + \ldots + \beta_{l,\alpha_l-1}(i+2p-N-s+1)^{\alpha_l-1})y_l^{i+2p-N-s+1})t^i, \end{aligned}$$

and, formally

$$P_N(t) = \sum_{i=0}^{2p-N-s} S_it^i + t^{2p-N-s+1}\left(\frac{a_1b_1^{2p-N-s+1}}{1-b_1t} + \ldots + \frac{a_{N-p}b_{N-p}^{2p-N-s+1}}{1-b_{N-p}t}\right)$$

$$+\sum_{l=1}^{\sigma} y_l^{2p-N-s+1}\left(\frac{\gamma_{l,0}}{1-y_l t}+\frac{\gamma_{l,1}}{(1-y_l t)^2}+\ldots+\frac{\gamma_{l,\alpha_l-1}}{(1-y_l t)^{\alpha_l}}\right)) + O(t^{N+1}),$$

where $\beta_{l,0}+\beta_{l,1}(i+2p-N-s+1)+\ldots+\beta_{l,\alpha_l-1}(i+2p-N-s+1)^{\alpha_l-1}$

$$= \gamma_{l,0}+\gamma_{l,1}(i+1)+\gamma_{l,2}\frac{(i+1)(i+2)}{2}+\ldots+\gamma_{l,\alpha_l-1}\frac{(i+1)(i+2)\ldots(i+\alpha_l-1)}{(\alpha_l-1)!}.$$

We now set

$$\left\{\begin{array}{l} u(t)=\sum_{i=0}^{2p-N-s} S_i t^i \in I\!K_{2p-N-s}[t] \\ \dfrac{a(t)}{b(t)}=\dfrac{a_1 b_1^{2p-N-s+1}}{1-b_1 t}+\ldots+\dfrac{a_{N-p} b_{N-p}^{2p-N-s+1}}{1-b_{N-p} t} \\ \text{where } a(t)\in I\!K_{N-p-1}[t] \text{ et } b(t)\in I\!K_{N-p}[t] \\ \text{and } \dfrac{c(t)}{y(t)}=\displaystyle\sum_{l=1}^{\sigma} y_l^{2p-N-s+1}\left(\frac{\gamma_{l,0}}{1-y_l t}+\frac{\gamma_{l,1}}{(1-y_l t)^2}+\ldots+\frac{\gamma_{l,\alpha_l-1}}{(1-y_l t)^{\alpha_l}}\right) \\ \text{where } c(t)\in I\!K_{s-1}[t] \text{ and } y(t)=(1-y_1 t)^{\alpha_1}(1-y_2 t)^{\alpha_2}\ldots(1-y_\sigma t)^{\alpha_\sigma}\in I\!K_s[t] \end{array}\right. .$$

Then, formally

$$P_N(t)=\frac{b(t)u(t)y(t)+t^{2p-N-s+1}(a(t)y(t)+b(t)c(t))}{b(t)y(t)}+O(t^{N+1}).$$

This implies that

$$[p/N-p]_{yP_N}(t)=\frac{b(t)u(t)y(t)+t^{2p-N-s+1}(a(t)y(t)+b(t)c(t))}{b(t)}$$

because $b(t)u(t)y(t)+t^{2p-N-s+1}(a(t)y(t)+b(t)c(t))\in I\!K_p[t]$, $b(t)\in I\!K_{N-p}[t]$ and from the definition of Padé approximants.
So, formally,

$$\sum_{i\geq 0} S_{i,N}^{(p)} t^i = \sum_{i=0}^{2p-N-s} S_i t^i + \sum_{i\geq 2p-N-s+1}(a_1 b_1^i+\ldots+a_{N-p} b_{N-p}^i)$$
$$+\sum_{l=1}^{\sigma}(\beta_{l,0}+\beta_{l,1} i+\ldots+\beta_{l,\alpha_l-1} i^{\alpha_l-1}) y_l^i) t^i.$$

Then,

$$\left\{\begin{array}{lcl} S_{i,N}^{(p)} & = & S_i,\ \forall i\in\{0,1,\ldots,2p-N-s\} \\ S_{i,N}^{(p)} & = & a_1 b_1^i+\ldots+a_{N-p} b_{N-p}^i \\ & & +\sum_{l=1}^{\sigma}(\beta_{l,0}+\beta_{l,1} i+\ldots+\beta_{l,\alpha_l-1} i^{\alpha_l-1}) y_l^i,\ \forall i\geq 2p-N-s+1 \end{array}\right.$$

and $T_{i,N}=S_{i,N}^{(p)},\ \forall i\in I\!N$.

3.4 Algorithm for the partial-Padé-prediction

We have already shown a link between the Padé-prediction and the ε_p-prediction (see Theorem 6).

It is now time to establish the self-evident link existing between the Padé-prediction and the partial-Padé-prediction. This link only uses applications called *elementary applications* avoiding the formal product of series.

At the end, we will just have to use these properties to construct the algorithm generating the partial-Padé-prediction. This algorithm will only need such elementary applications and the ε-algorithm.

The elementary applications

Let us consider the applications h_z and $\overline{h}_z$

$$h_z : \begin{cases} \mathbb{K}^{\mathbb{N}} & \longrightarrow & \mathbb{K}^{\mathbb{N}} \\ (S_0, S_1, \ldots, S_N, \ldots) & \longrightarrow & (S_0, zS_0 + S_1, \ldots, z^N S_0 + z^{N-1} S_1 + \ldots + S_N, \ldots). \end{cases}$$

$$\overline{h}_z : \begin{cases} \mathbb{K}^{N+1} & \longrightarrow & \mathbb{K}^{N+1} \\ (S_0, S_1, \ldots, S_N) & \longrightarrow & (S_0, zS_0 + S_1, \ldots, z^N S_0 + z^{N-1} S_1 + \ldots + S_N). \end{cases}$$

Then, the reverse applications are

$$h_z^{-1} : \begin{cases} \mathbb{K}^{\mathbb{N}} & \longrightarrow & \mathbb{K}^{\mathbb{N}} \\ (S_0, S_1, \ldots, S_N, \ldots) & \longrightarrow & (S_0, -zS_0 + S_1, \ldots, -zS_{N-1} + S_N, \ldots). \end{cases}$$

$$\overline{h}_z^{-1} : \begin{cases} \mathbb{K}^{N+1} & \longrightarrow & \mathbb{K}^{N+1} \\ (S_0, S_1, \ldots, S_N) & \longrightarrow & (S_0, -zS_0 + S_1, \ldots, -zS_{N-1} + S_N). \end{cases}$$

Now, with our notations, Theorem 6 becomes

$$\textbf{Padé-prediction} = h_1^{-1} \circ \varepsilon_p\textbf{-prediction} \circ \overline{h}_1$$

where the Padé-prediction is the transformation of the vector to predict into the Padé-predicted sequence (depending on the parameter p) and the ε_p-prediction is the transformation of the vector to predict into the ε_p-predicted sequence (depending on the same parameter p).

But we can also give the next trivial property.

When

$$x(t) = (1 - x_1 t)(1 - x_2 t) \ldots (1 - x_r t) \in \mathbb{K}_r[t],$$

and

$$y(t) = (1 - y_1 t)(1 - y_2 t) \ldots (1 - y_s t) \in \mathbb{K}_s[t],$$

$$\textbf{partial-Padé-prediction} = h_{x_1}^{-1} \circ h_{x_2}^{-1} \circ \ldots \circ h_{x_r}^{-1} \circ h_{y_1} \circ h_{y_2} \circ \ldots \circ h_{y_s} \circ \textbf{ Padé-prediction} \\ \circ \overline{h}_{x_1} \circ \overline{h}_{x_2} \circ \ldots \circ \overline{h}_{x_r} \circ \overline{h}_{y_1}^{-1} \circ \overline{h}_{y_2}^{-1} \circ \ldots \circ \overline{h}_{y_s}^{-1}$$

where the partial-Padé-prediction is the transformation of the vector to predict into the partial-Padé-predicted sequence (depending on the parameter p) calculated from the polynomials $x(t)$ and $y(t)$, and the Padé-prediction is the transformation of the vector to predict into the Padé-predicted sequence (depending on the same parameter p).

Thus we obtain the following property :
If

$$x(t) \in \mathbb{K}_r[t],$$

and

$$y(t) \in \mathbb{K}_s[t],$$

$$\textbf{partial-Padé-prediction} = h_{x_1}^{-1} \circ h_{x_2}^{-1} \circ \ldots \circ h_{x_r}^{-1} \circ h_{y_1} \circ h_{y_2} \circ \ldots \circ h_{y_s} \circ h_1^{-1} \circ \varepsilon_p\textbf{-prediction} \\ \circ \overline{h}_{x_1} \circ \overline{h}_{x_2} \circ \ldots \circ \overline{h}_{x_r} \circ \overline{h}_{y_1}^{-1} \circ \overline{h}_{y_2}^{-1} \circ \ldots \circ \overline{h}_{y_s}^{-1} \circ \overline{h}_1$$

where the partial-Padé-prediction is the transformation of the vector to predict into the partial-Padé-predicted sequence (depending on the parameter p) calculated from the polynomials $x(t)$ and $y(t)$, and the ε_p-prediction is the transformation of the vector to predict into the ε_p-predicted sequence (depending on the same parameter p).

This very simple algorithm can generate each partial-Padé-prediction without solving a system of linear equations (see Section 3.2) or of exponential ones (see Section 3.3).

4 Properties of the prediction of moments

In this section, we will consider that the vector to predict has components of the form $\int_\Omega x^i dg(x)$ where g is a nondecreasing function with bounded variation on Ω, an interval of $I\!R$.

Considering the Padé-prediction, we have the

Theorem 8 *For p such that $2p - N \geq -1$, let $(S_i)_{0 \leq i \leq N} \in I\!R^{N+1}$ be the vector to predict such that $S_i = \int_\Omega x^i dg(x),\ \forall i \in \{0, 1, \ldots, N\}$, where g is a nondecreasing function with bounded variation on Ω, an interval of $I\!R$.*

Then, the Padé-predicted sequence $(S_{i,N}^{(p)})_{i \geq 2p-N+1}$ satisfies

$$S_{i,N}^{(p)} = \int_\Omega x^i dh(x),\ \forall i \geq 2p - N + 1$$

where h is a nondecreasing function with bounded variation on Ω.

Proof

From hypothesis and Theorem 4, we obtain

$$S_{i,N}^{(p)} = a_1 b_1^i + \ldots + a_{N-p} b_{N-p}^i,\ \forall i \geq 2p - N + 1.$$

Then, from results on Gaussian quadrature, we get

1. $b_i \in \Omega,\ \forall i \in \{1, 2, \ldots, N-p\}$
2. $a_i \geq 0,\ \forall i \in \{1, 2, \ldots, N-p\}$.

Thus,

$$S_{i,N}^{(p)} = \int_\Omega x^i dh(x),\ \forall i \geq 2p - N + 1$$

where h is a nondecreasing function with bounded variation on Ω ($dh = \sum_{i=1}^{N-p} a_i d\delta_{b_i}$ where $d\delta_{b_i}$ is the Dirac measure at b_i).

Remarks

1. A particular case of this Theorem is the following : when the vector to predict is extracted from a totally monotonic sequence (Haussdorf's Theorem insures that totally monotonic sequences $(S_i)_{i \geq 0}$

are such that $S_i = \int_0^1 x^i dg(x),\ \forall i \in \mathbb{N}$ where g is a nondecreasing function with bounded variation on $[0,1]$), the Padé-predicted sequence will also be totally monotonic starting from the $(2p-N+1)^{th}$ term.

Also, there is a great difference for a vector $(S_i)_{0\le i\le N}$ to be extracted from a totally monotonic sequence or satisfying $(-1)^j\Delta^j S_i \ge 0,\ \forall(i,j)$ such that $i+j\le N$. Condition $(-1)^j\Delta^j S_i \ge 0,\ \forall(i,j),\ i+j\le N$, is not sufficient to obtain a totally monotonic Padé-predicted sequence starting from the $(2p-N+1)^{th}$term (for example, if we take $N=4,\ p=2,\ S_0=97,\ S_1=22,\ S_2=\frac{73}{9},\ S_3=\frac{13}{4},\ S_4=1$, then $(S_i)_{0\le i\le 4}$ satisfies $(-1)^j\Delta^j S_i\ge 0,\ \forall(i,j)$ such that $i+j\le 4$, but $(S^{(2)}_{i,4})_{i\ge 0}$ is not totally monotonic because $S^{(2)}_{5,4} = -\dfrac{2903409}{54760000} < 0$).

2. This Theorem can be used to construct a totally monotonic sequence with first terms given. Thus we have obtained an algorithmic construction of totally monotonic sequences.

3. The assumption $2p-N\ge -1$ cannot be removed (for example, if we take $N=2,\ p=0,\ S_0=6,\ S_1=3,\ S_2=2$, we have $2p-N=-2$ and $(S_i)_{0\le i\le 2}$ is extracted from the totally monotonic sequence $(\dfrac{6}{i+1})_{i\ge 0}$, but the Padé-predicted sequence $(S^{(0)}_{i,2})_{i\ge 0}$ is not totally monotonic because $\Delta^3 S^{(0)}_{1,2} = \frac{1}{24} > 0$).

Considering the higher-Padé-type-prediction calculated from the polynomial $y(t)=1-y_1t$, we have the

Theorem 9 *For p such that $2p-N\ge 0$, let $(S_i)_{0\le i\le N}\in \mathbb{R}^{N+1}$ be the vector to predict such that $S_i=\int_\Omega x^i dg(x),\ \forall i\in\{0,1,\ldots,N\}$, where g is a nondecreasing function with bounded variation on Ω, interval of $\mathbb{R}$.*

We assume that Ω is $[R,R'[$ or $]R',R]$ where R is finite and R' finite or infinite.

Then, the higher-Padé-type-predicted sequence $(S^{(p)}_{i,N})_{i\ge 2p-N+1}$ calculated from the polynomial $y(t)=1-Rt$ satisfies

$$S^{(p)}_{i,N} = \int_\Omega x^i dh(x),\ \forall i\ge 2p-N$$

where h is a nondecreasing function with bounded variation on Ω.

Proof

We give the proof for $\Omega=[R,R'[$.

From the hypothesis and Theorem 7 used for $x(t)=1$ and $y(t)=1-Rt$, we obtain

$$S^{(p)}_{i,N} = a_1b_1^i+\ldots+a_{N-p}b_{N-p}^i+c_1R^i,\ \forall i\ge 2p-N.$$

Then, from results on Gauss-Radau quadratures, we get

1. $b_i\in\Omega\backslash\{R\},\ \forall i\in\{1,2,\ldots,N-p\}$
2. $a_i\ge 0,\ \forall i\in\{1,2,\ldots,N-p\}$
3. $c_1\ge 0$.

Thus,

$$S^{(p)}_{i,N} = \int_\Omega x^i dh(x),\ \forall i\ge 2p-N$$

where h is a nondecreasing function with bounded variation on Ω.

Remarks

1. When Ω is $[R, R'[$ or $]R', R]$, we obtain the previous Theorem for higher-Padé-type-predicted sequence calculated from the polynomial $y(t) = 1 - y_1 t$ where y_1 is an extremity of Ω. What does it happen if we consider

   (a) y_1 strictly included in Ω ?

   Theorem 9 cannot be applied (for example, if we take $N = 2,\ p = 1,\ y(t) = 1 - \dfrac{t}{2},\ S_0 = 2,\ S_1 = \dfrac{17}{15},\ S_2 = \dfrac{169}{225}$, we have $2p - N = 0$ and $(S_i)_{0 \leq i \leq 2}$ is extracted from the totally monotonic sequence $(\dfrac{1}{3^i} + \dfrac{4^i}{5^i})_{i \geq 0}$, but the higher-Padé-type-predicted sequence $(S_{i,2}^{(1)})_{i \geq 0}$ calculated from $y(t)$ is not totally monotonic because $\Delta S_{3,2}^{(1)} = \dfrac{60917}{1620000} > 0$).

   (b) y_1 strictly out of Ω ?

   Our Theorem 9 does not hold (for example, if we take $N = 2,\ p = 1,\ y(t) = 1 - 2 \cdot t,\ S_0 = 2,\ S_1 = \dfrac{5}{6},\ S_2 = \dfrac{13}{36}$, we have $2p - N = 0$ and $(S_i)_{0 \leq i \leq 2}$ is extracted from the totally monotonic sequence $(\frac{1}{2^i} + \frac{1}{3^i})_{i \geq 0}$, but the higher-Padé-type-predicted sequence $(S_{i,2}^{(1)})_{i \geq 0}$ calculated from $y(t)$ is not totally monotonic because $\Delta S_{4,2}^{(1)} = \dfrac{2908457}{53335584} > 0$).

2. If Ω is $[R, R'[$ or $]R', R]$, then Theorem 9 can be applied for the higher-Padé-type-predicted sequence calculated from the polynomial $y(t) = 1 - Rt$.
   Can we generalize this when $y(t) = (1 - Rt)^s$, for any s, only by changing the condition $2p - N \geq 0$ into $2p - N \geq s - 1$?
   The answer is no (for example, if we take $N = 3,\ p = 2,\ y(t) = (1 - t)^2,\ S_0 = 2,\ S_1 = \dfrac{5}{6},\ S_2 = \dfrac{13}{36},\ S_3 = \dfrac{35}{216}$, we have $2p - N = -1$ and $(S_i)_{0 \leq i \leq 3}$ is extracted from the totally monotonic sequence $(\dfrac{1}{2^i} + \dfrac{1}{3^i})_{i \geq 0}$, but the higher-Padé-type-predicted sequence $(S_{i,3}^{(2)})_{i \geq 0}$ calculated from $y(t)$ is not totally monotonic because $S_{6,3}^{(2)} = -\dfrac{8551439}{729000000} < 0$).

Then, we only have to consider the possibility when Ω is $[R, R']$ and $y(t) = (1 - Rt)(1 - R't)$ and we get the

Theorem 10 *For p such that $2p - N \geq 1$, let $(S_i)_{0 \leq i \leq N} \in I\!R^{N+1}$ be the vector to predict such that $S_i = \int_\Omega x^i dg(x),\ \forall i \in \{0, 1, \ldots, N\}$, where g is a nondecreasing function with bounded variation on Ω, an interval of $I\!R$.*

We assume that Ω is $[R, R']$ where R and R' are finite.

Then, the higher-Padé-type-predicted sequence $(S_{i,N}^{(p)})_{i \geq 2p-N+1}$ calculated from the polynomial $y(t) = (1 - Rt)(1 - R't)$ satisfies

$$S_{i,N}^{(p)} = \int_\Omega x^i dh(x),\ \forall i \geq 2p - N - 1$$

where h is a nondecreasing function with bounded variation on Ω.

Proof

From the hypothesis and Theorem 7 applied with $x(t) = 1$ and $y(t) = (1 - Rt)(1 - R't)$, we obtain

$$S_{i,N}^{(p)} = a_1 b_1^i + \ldots + a_{N-p} b_{N-p}^i + c_1 R^i + c_2 {R'}^i, \ \forall i \geq 2p - N - 1.$$

Then, from results on Gauss-Lobatto quadratures, we get

1. $b_i \in]R, R'[, \ \forall i \in \{1, 2, \ldots, N-p\}$
2. $a_i \geq 0, \ \forall i \in \{1, 2, \ldots, N-p\}$
3. $c_1 \geq 0$
4. $c_2 \geq 0$.

Thus,

$$S_{i,N}^{(p)} = \int_\Omega x^i dh(x), \ \forall i \geq 2p - N - 1$$

where h is a nondecreasing function with bounded variation on Ω.

5 Consistency in column and in diagonal for the higher-Padé-type-prediction

The term *consistency* means the following property : when the sequence to predict converges, then the predicted sequence converges to the same limit.

Let $(S_i)_{i\geq 0}$ be the sequence to predict and $P_{(y)}$ be the higher-Padé-type-prediction method calculated from the polynomial $y(t)$.

In order to explain the difference between consistency in column and consistency in diagonal, let us consider the following diagram :

$$\begin{array}{ccccc}
(S_{i,N,0}^{(p)})_{i\geq 0} & \xleftarrow{\quad t_0 \circ P_{(y)} \circ t_0 \quad} & \left[\begin{array}{c} S_0 \\ S_1 \\ \vdots \\ S_N \end{array}\right] & \xrightarrow{\quad P_{(y)} \quad} & (S_{i,N})_{i\geq 0} \\
 & & \begin{array}{c} S_{N+1} \\ \vdots \end{array} & & \\
(S_{i,N,m}^{(p)})_{i\geq m} & \xleftarrow{\quad t_m \circ P_{(y)} \circ t_{-m} \quad} & \left[\begin{array}{c} S_m \\ S_{m+1} \\ \vdots \\ S_{m+N} \end{array}\right. & & \\
 & & \vdots & & \\
 & & \text{sequence to predict} & &
\end{array}$$

where $t_{-j}((S_i)_{j\leq i\leq j+N}) = (T_i)_{0\leq i\leq N}$ such that $S_{j+i} = T_i, \ \forall i \in \{0, 1, \ldots, N\}$
and $t_j((T_{i,N}^{(p)})_{i\geq 0}) = (S_{i,N,j}^{(p)})_{i\geq j}$ such that $S_{j+i,N,j}^{(p)} = T_{i,N}^{(p)}, \ \forall i \in I\!N$.

We say that we have consistency in column for the higher-Padé-type-prediction when

$$\lim_{m\to\infty} S^{(p)}_{m+N+i,N,m} - S_{m+N+i} = 0, \text{ for } i \text{ and } N \text{ fixed}$$

and consistency in diagonal when

$$\lim_{N\to\infty} S^{(p)}_{m+N+i,N,m} - S_{m+N+i} = 0, \text{ for } i \text{ and } m \text{ fixed.}$$

The vocabulary used comes out from extrapolation methods. It seems to be natural to link the notion of convergence acceleration in column with the consistency in column and the notion of convergence acceleration in diagonal with the consistency in diagonal.

In short, consistency in column is consistency from a finite number of terms extracted farther and farther whereas consistency in diagonal is consistency from an increasing number of terms.

The problems appearing in the proofs, when we use the expression of the predicted terms as ratios of two determinants, are the study of the convergence of the determinantal expressions

1. of fixed dimension for the consistency in column.
2. of increasing dimension for the consistency in diagonal.

5.1 Consistency in column for the higher-Padé-type-prediction

In this subsection, we will only study the ε_p-prediction.

Let us begin with a result given by C. Brezinski and M. Redivo Zaglia in [6, p. 391].

Theorem 11 *Let $(S_i)_{i\geq 0}$ be the sequence to predict. It is assumed to converge to S.*
If $\exists M, \exists J$, such that $\forall j \geq J$, $\left|\dfrac{\Delta S_{j+1}}{\Delta S_j}\right| \leq M$, then the Δ^2-prediction is consistent in column.

We generalize this Theorem in the following.

Theorem 12 *Let $(S_i)_{i\geq 0}$ be the sequence to predict. It is assumed to converge to S.*
If $\forall N \in I\!N, \forall p \in I\!N, \exists M_1,\ \exists M_2,\ \exists J$, such that $\forall j \geq J$, we have :

$$M_1 \leq Abs\left(\frac{\begin{vmatrix} \Delta S_{2p-N+j} & \Delta S_{2p-N+1+j} & \dots & \Delta S_{p+j} \\ \Delta S_{2p-N+1+j} & \Delta S_{2p-N+2+j} & \dots & \Delta S_{p+1+j} \\ \vdots & \vdots & & \vdots \\ \Delta S_{p+j-1} & \Delta S_{p+j} & \dots & \Delta S_{N-1+j} \end{vmatrix}_{(l)}}{\begin{vmatrix} \Delta S_{2p-N+j} & \Delta S_{2p-N+1+j} & \dots & \Delta S_{p-1+j} \\ \Delta S_{2p-N+1+j} & \Delta S_{2p-N+2+j} & \dots & \Delta S_{p+j} \\ \vdots & \vdots & & \vdots \\ \Delta S_{p+j-1} & \Delta S_{p+j} & \dots & \Delta S_{N-2+j} \end{vmatrix}}\right) \leq M_2,\ \forall l \in \{1, 2, \dots, N-p\},$$

where $\begin{vmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n+1} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n+1} \\ \vdots & \vdots & & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n+1} \end{vmatrix}_{(m)}$, $\forall m \in \{1, 2, \dots, n+1\}$ *represents the determinant of the matrix* $(a_{k,l})_{1\le k\le n, 1\le l\le n+1, l\neq m}$*, then, the* ε_p*-prediction is consistent in column.*

Proof

From Theorem 2

$$S^{(\varepsilon S,p)}_{N+m+i,N,m} - S_{N+m} = -\frac{\begin{vmatrix} S_{N+m} & 1 & S^{(\varepsilon S,p)}_{p+m+i,N,m} & S^{(\varepsilon S,p)}_{p+m+i+1,N,m} & \dots & S^{(\varepsilon S,p)}_{N+m+i-1,N,m} \\ S_{p+m} & 1 & S_{2p-N+m} & S_{2p-N+1+m} & \dots & S_{p-1+m} \\ S_{p+1+m} & 1 & S_{2p-N+1+m} & S_{2p-N+2+m} & \dots & S_{p+m} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ S_{N+m} & 1 & S_{p+m} & S_{p+1+m} & \dots & S_{N-1+m} \end{vmatrix}}{\begin{vmatrix} 1 & S_{2p-N+m} & S_{2p-N+1+m} & \dots & S_{p-1+m} \\ 1 & S_{2p-N+1+m} & S_{2p-N+2+m} & \dots & S_{p+m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & S_{p+m} & S_{p+1+m} & \dots & S_{N-1+m} \end{vmatrix}}, \ \forall i \in I\!N,$$

with the convention $S_i = 0$ when $i < 0$.

Then

$$S^{(\varepsilon S,p)}_{N+m+i,N,m} - S_{N+m}$$

$$= -\frac{\begin{vmatrix} S_{N+m} & 1 & S^{(\varepsilon S,p)}_{p+m+i,N,m} & S^{(\varepsilon S,p)}_{p+m+i+1,N,m} & \dots & S^{(\varepsilon S,p)}_{N+m+i-1,N,m} \\ -\Delta S_{p+m} & 1 & -\Delta S_{2p-N+m} & -\Delta S_{2p-N+1+m} & \dots & -\Delta S_{p-1+m} \\ -\Delta S_{p+1+m} & 1 & -\Delta S_{2p-N+1+m} & -\Delta S_{2p-N+2+m} & \dots & -\Delta S_{p+m} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ -\Delta S_{N+m-1} & 1 & -\Delta S_{p+m-1} & -\Delta S_{p+m} & \dots & -\Delta S_{N+m-2} \\ 0 & 1 & S_{p+m} - S^{(\varepsilon S,p)}_{p+m+i,N,m} & S_{p+1+m} - S^{(\varepsilon S,p)}_{p+m+i+1,N,m} & \dots & S_{N-1+m} - S^{(\varepsilon S,p)}_{N+m+i-1,N,m} \end{vmatrix}}{\begin{vmatrix} 1 & S_{2p-N+m} & S_{2p-N+1+m} & \dots & S_{p-1+m} \\ 1 & S_{2p-N+1+m} & S_{2p-N+2+m} & \dots & S_{p+m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & S_{p+m} & S_{p+1+m} & \dots & S_{N-1+m} \end{vmatrix}}$$

$$= -\frac{\begin{vmatrix} \Delta S_{2p-N+m} & \Delta S_{2p-N+1+m} & \dots & \Delta S_{p-1+m} & \Delta S_{p+m} \\ \Delta S_{2p-N+m+1} & \Delta S_{2p-N+2+m} & \dots & \Delta S_{p+m} & \Delta S_{p+m+1} \\ \vdots & \vdots & & \vdots & \vdots \\ \Delta S_{p+m-1} & \Delta S_{p+m} & \dots & \Delta S_{N+m-2} & \Delta S_{N+m-1} \\ S^{(\varepsilon S,p)}_{p+m+i,N,m} - S_{p+m} & S^{(\varepsilon S,p)}_{p+m+i+1,N,m} - S_{p+1+m} & \dots & S^{(\varepsilon S,p)}_{N+m+i-1,N,m} - S_{N-1+m} & 0 \end{vmatrix}}{\begin{vmatrix} \Delta S_{2p-N+m} & \Delta S_{2p-N+1+m} & \dots & \Delta S_{p-1+m} \\ \Delta S_{2p-N+1+m} & \Delta S_{2p-N+2+m} & \dots & \Delta S_{p+m} \\ \vdots & \vdots & & \vdots \\ \Delta S_{p+m-1} & \Delta S_{p+m} & \dots & \Delta S_{N-2+m} \end{vmatrix}}$$

Then, $\forall m \geq J$,

$$|S^{(\varepsilon S,p)}_{N+i+m,N,m} - S_{N+m}| < \max(|M_1|, |M_2|) \sum_{j=0}^{N-p-1} \left| S^{(\varepsilon S,p)}_{p+m+i+j,N,m} - S_{p+m+j} \right|.$$

Then, recursively on i, it follows that

$$\lim_{m\to\infty} S^{(\varepsilon S,p)}_{N+i+m,N,m} - S_{N+m} = 0.$$

Consequently, as $(S_i)_{i\geq 0}$ is converging,

$$\lim_{m\to\infty} S^{(\varepsilon S,p)}_{N+i+m,N,m} - S_{N+m+i} = 0.$$

In this last Theorem, the condition concerning the ratio of two determinants is satisfied for totally monotonic sequences as we can see in the next Theorem.

Theorem 13 *Let $(S_i)_{i\geq 0}$ be a totally monotonic sequence.*
Then, $\forall j \geq N-2p$, we have :

$$0 \leq \frac{\begin{vmatrix} S_{2p-N+j} & S_{2p-N+1+j} & \dots & S_{p+j} \\ S_{2p-N+1+j} & S_{2p-N+2+j} & \dots & S_{p+1+j} \\ \vdots & \vdots & & \vdots \\ S_{p+j-1} & S_{p+j} & \dots & S_{N-1+j} \end{vmatrix}_{(l)}}{\begin{vmatrix} S_{2p-N+j} & S_{2p-N+1+j} & \dots & S_{p-1+j} \\ S_{2p-N+1+j} & S_{2p-N+2+j} & \dots & S_{p+j} \\ \vdots & \vdots & & \vdots \\ S_{p+j-1} & S_{p+j} & \dots & S_{N-2+j} \end{vmatrix}} \leq \begin{pmatrix} N-p \\ l-1 \end{pmatrix}, \ \forall l \in \{1,2,\dots,N-p\}.$$

Proof

From $S_i = \int_0^1 x^i d\alpha(x), \ \forall i \in I\!N$, the Cauchy-Binet Formula gives :

$$\begin{vmatrix} S_{2p-N+j} & S_{2p-N+1+j} & \dots & S_{p+j} \\ S_{2p-N+1+j} & S_{2p-N+2+j} & \dots & S_{p+1+j} \\ \vdots & \vdots & & \vdots \\ S_{p+j-1} & S_{p+j} & \dots & S_{N-1+j} \end{vmatrix}_{(l)}$$
$$= \int\int\dots\int_{0\leq x_0\leq x_1\leq\dots\leq x_n\leq 1} (x_0x_1\dots x_{N-p-1})^{2p-N+j} V(x_0,x_1,\dots,x_{N-p-1})$$
$$\times \begin{vmatrix} 1 & x_0 & \dots & x_0^{N-p} \\ 1 & x_1 & \dots & x_1^{N-p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{N-p-1} & \dots & x_{N-p-1}^{N-p} \end{vmatrix}_{(l)} d\alpha(x_1)d\alpha(x_2)\dots d\alpha(x_{N-p-1}),$$

where

$$V(x_0,x_1,\dots,x_n) = \begin{vmatrix} 1 & x_0 & \dots & x_0^n \\ 1 & x_1 & \dots & x_1^n \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \dots & x_n^n \end{vmatrix}$$

Then, by computing the coefficient of x^{l-1} in $V(x, x_0, x_1, \ldots, x_{N-p-1})$, we get :

$$\left|\begin{array}{cccc} 1 & x_0 & \ldots & x_0^{N-p} \\ 1 & x_1 & \ldots & x_1^{N-p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{N-p-1} & \ldots & x_{N-p-1}^{N-p} \end{array}\right|_{(l)} = V(x_0, x_1, \ldots, x_{N-p-1}) \times \sum_{0 \le i_1 < i_2 < \ldots < i_{N-p+1-l} \le N-p-1} x_{i_1} x_{i_2} \ldots x_{i_{N-p+1-l}}.$$

Then,

$$\left|\begin{array}{cccc} S_{2p-N+j} & S_{2p-N+1+j} & \ldots & S_{p+j} \\ S_{2p-N+1+j} & S_{2p-N+2+j} & \ldots & S_{p+1+j} \\ \vdots & \vdots & & \vdots \\ S_{p+j-1} & S_{p+j} & \ldots & S_{N-1+j} \end{array}\right|_{(l)} - \left(\begin{array}{c} N-p \\ l-1 \end{array}\right) \left|\begin{array}{cccc} S_{2p-N+j} & S_{2p-N+1+j} & \ldots & S_{p-1+j} \\ S_{2p-N+1+j} & S_{2p-N+2+j} & \ldots & S_{p+j} \\ \vdots & \vdots & & \vdots \\ S_{p+j-1} & S_{p+j} & \ldots & S_{N-2+j} \end{array}\right|$$

$$= \int\int\ldots\int_{0 \le x_0 \le x_1 \le \ldots \le x_n \le 1} (x_0 x_1 \ldots x_{N-p-1})^{2p-N+j} V(x_0, x_1, \ldots, x_{N-p-1})^2$$

$$\times \left(\sum_{0 \le i_1 < i_2 < \ldots < i_{N-p+1-l} \le N-p-1} x_{i_1} x_{i_2} \ldots x_{i_{N-p+1-l}} - \left(\begin{array}{c} N-p \\ l-1 \end{array}\right) \right) d\alpha(x_1) d\alpha(x_2) \ldots d\alpha(x_{N-p-1}) \le 0.$$

Since the sequence $(S_i)_{i \ge 0}$ is a totally monotonic sequence, (i.e. $S_i = \int_0^1 x^i dg(x),\ \forall i \in I\!N$, where g is a nondecreasing function with bounded variation on $[0, 1]$), $(\Delta S_i)_{i \ge 0}$ is also a totally monotonic sequence. A consequence is that the ε-prediction is consistent in column for totally monotonic sequences.

5.2 Consistency in diagonal for the higher-Padé-type-prediction

This subsection will concern only the Padé-prediction.

Let us begin with this obvious Theorem :

Theorem 14 *let us suppose that* $2p - N \ge -1$.

Let $(S_i)_{0 \le i \le N} \in I\!R^{N+1}$ *be the vector to predict such that* $S_i = \int_a^b x^i dg(x),\ \forall i \in \{0, 1, \ldots, N\}$, *where* g *is a nondecreasing function with bounded variation on* $[a, b]$ *(*a *and* b *are assumed to be finite).*

Then the Padé-predicted sequence $(S_{i,N}^{(p)})_{i \ge 0}$ *satisfies*

$$\lim_{N \to \infty} \frac{S_{N+i+m,N,m}^{(p)} - S_{N+i+m}}{\rho^N} = 0,\ \forall \rho > max(|a|, |b|),\ \forall i \in I\!N.$$

Proof

From Theorem 8,

$$S_{i,N}^{(p)} = \int_a^b x^i dh(x),\ \forall i \ge 2p - N + 1$$

where h is a nondecreasing function with bounded variation on $[a,b]$. Then, obviously,

$$\lim_{N\to\infty} \frac{S^{(p)}_{N+i+m,N,m}}{\rho^N} = 0,\ \forall \rho > R,\ \forall k \in I\!N.$$

And, as

$$\lim_{N\to\infty} \frac{S_{N+i+m}}{\rho^N} = 0,\ \forall \rho > R,\ \forall k \in I\!N,$$

it follows that

$$\lim_{N\to\infty} \frac{S^{(p)}_{N+k,N} - S_{N+k}}{\rho^N} = 0,\ \forall \rho > R,\ \forall k \in I\!N.$$

We directly deduce the three following Theorems (in these Theorems, the value of ρ, defined in the Theorem 14, can be fixed equal to one) :

Theorem 15 $2p - N \geq -1$.

Let $(S_i)_{0\leq i\leq N} \in I\!R^{N+1}$ be the vector to predict extracted from a totally monotonic sequence such that

$$\lim_{i\to\infty} \frac{\Delta S_{i+1}}{\Delta S_i} = \mu \qquad (\mu \neq 1),$$

i.e. linearly converging.

Then the Padé-prediction is consistent in diagonal.

Theorem 16 $2p - N \geq -1$.

Let $(S_i)_{0\leq i\leq N} \in I\!R^{N+1}$ be the vector to predict extracted from a totally oscillating sequence (i.e. $S_i = \int_{-1}^{0} x^i dg(x),\ \forall i \in I\!N$, where g is a nondecreasing function with bounded variation on $[-1,0]$), such that

$$\lim_{i\to\infty} \frac{\Delta S_{i+1}}{\Delta S_i} = \mu \qquad (\mu \neq -1).$$

Then the Padé-prediction is consistent in diagonal.

Theorem 17 $2p - N \geq -1$.

Let $(S_i)_{0\leq i\leq N} \in I\!R^{N+1}$ be the vector to predict such that $S_i = \int_a^b x^i dg(x),\ \forall i \in \{0,1,\ldots,N\}$, where g is a nondecreasing function with bounded variation on $[a,b]$.

If we assume $-1 < a < b < 1$, then the Padé-prediction is consistent in diagonal.

6 Numerical applications

Prediction methods, more precisely the Padé-prediction and the Aitken's Δ^2-prediction has already been used on physical problems as we can see in[16], [1] or in [10].

However, we choose here to exhibit sequences for which each term is known. So, we can make a comparison between the sequence to predict and the predicted sequence.

6.1 First Example

The sequence we will study here is given by :

$$S_0 = 2,\ S_{i+1} = \sqrt{1 + S_i},\ \forall i \in \mathbb{N}.$$

Generally, if we consider that nothing is known on the sequence to predict, we apply the Padé-prediction or the ε_p-prediction. Let us here display the differences $|S_{i,2p}^{(\varepsilon,p)} - S_i|$, where $S_{i,2p}^{(\varepsilon,p)}$ is the ε-predicted vector $(S_i)_{0 \le i \le 2p}$.

We display the difference $|S_{i,2p}^{(\varepsilon,p)} - S_i|$, for different values of p and i.

i	$\lvert S_{i,4}^{(\varepsilon,2)} - S_i\rvert * 10^7$	$\lvert S_{i,6}^{(\varepsilon,3)} - S_i\rvert * 10^{12}$	$\lvert S_{i,8}^{(\varepsilon,4)} - S_i\rvert * 10^{18}$	$\lvert S_{i,10}^{(\varepsilon,5)} - S_i\rvert * 10^{25}$	$\lvert S_{i,12}^{(\varepsilon,6)} - S_i\rvert * 10^{34}$
0	0.	0.	0.	0.	0.
1	0.	0.	0.	0.	0.
2	0.	0.	0.	0.	0.
3	0.	0.	0.	0.	0.
4	0.	0.	0.	0.	0.
5	1.40	0.	0.	0.	0.
6	2.57	0.	0.	0.	0.
7	3.21	1.66	0.	0.	0.
8	3.50	3.12	0.	0.	0.
9	3.62	3.93	1.29	0.	0.
10	3.66	4.31	2.44	0.	0.
11	3.68	4.46	3.09	0.70	0.
12	3.69	4.52	3.39	1.33	0.
13	3.69	4.55	3.52	1.68	2.79
14	3.69	4.55	3.57	1.85	5.28
15	3.69	4.56	3.58	1.91	6.69
16	3.69	4.56	3.59	1.94	7.34
17	3.69	4.56	3.59	1.95	7.61
18	3.69	4.56	3.59	1.95	7.72
19	3.69	4.56	3.59	1.96	7.76
20	3.69	4.56	3.59	1.96	7.77

Of course, since the predicted sequence reproduces the known terms, the difference between the first known terms and the terms of the predicted sequence is zero.

6.2 Second Example

Let us now deal with the moments on the Cantor set on [0,1]. These moments are linked by a recurrence relation with an increasing numbers of terms:

$$S_0 = 1, S_n = \frac{1}{3^n - 1} \sum_{j=0}^{n} \binom{n}{j} 2^{n-j-1} S_j.$$

We display the relative error (in percentage) between the exact terms of the sequence $(S_n)_n$ and the Padé-predicted one. $S_{i,m}^{(p)}$ is the predicted term of S_i with the Padé-prediction method based on the knowledge of $S_0, \ldots, S_{m+p}$. (we recall that $S_{i,m}^{(p)}$ is the i-th coefficient of the Taylor series of the Padé approximant $[p/m-p]$ to the polynomial $\sum_{j=0}^{m+p} S_j t^j$). $S_{i,m}^{(p)}$ is computed with ε-algorithm as explained in Section 5.4. See Section 4 for the notations.

i	$(S_{i,4}^{(2)} - S_i)/S_i * 100$	$(S_{i,6}^{(3)} - S_i)/S_i * 100$	$(S_{i,8}^{(4)} - S_i)/S_i * 100$	$(S_{i,10}^{(5)} - S_i)/S_i * 100$
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	-0.77	0	0	0
6	-2.6	0	0	0
7	-5.3	-0.056	0	0
8	-8.9	-0.24	0	0
9	-13.	-0.6	-0.0033	0
10	-17.	-1.2	-0.017	0
11	-22.	-2.	-0.053	-0.00019
12	-27.	-3.	-0.12	-0.0012
13	-32.	-4.3	-0.24	-0.0043
14	-36.	-5.7	-0.41	-0.012
15	-41.	-7.3	-0.65	-0.025
16	-45.	-9.	-0.96	-0.049
17	-49.	-11.	-1.4	-0.086
18	-53.	-13.	-1.8	-0.14
19	-57.	-15.	-2.4	-0.21
20	-60.	-17.	-3.	-0.31

This prediction method has been proved to be consistent in diagonal for sequence of moments, i.e. $\forall i$ fixed $\in I\!N, \lim_{p\to\infty} |S_{2p+i,2p}^{(p)} - S_{2p+i}| = 0$.

References

[1] G. ALLAN, A linear prediction of the recursion coefficients, *J. Phys. C*, *Solid State Physics*, **17** (1984), 3945-3955.

[2] G.A.BAKER, P. GRAVES MORRIS,"Padé Approximants", Second ed., Cambridge University Press, Cambridge,1996.

[3] C.BREZINSKI, " Padé-type Approximants and General Orthogonal Polynomials", Birkhäuser, Basel, 1980.

[4] C. BREZINSKI, Prediction properties of some extrapolation methods, *Appl. Numer. Math.*, **1** (1985), 457-462.

[5] C. BREZINSKI, Partial Padé approximation, *J. Approx. Theory*, **54** (1988), 210-233.

[6] C. BREZINSKI, M. REDIVO ZAGLIA, "Extrapolation Methods. Theory and Practice", North Holland, Amsterdam, 1991.

[7] F. CORDELLIER, Démonstration algébrique de l'extension de l'identité de Wynn aux tables de Padé non normales, in *Padé Approximation and its Applications*, L.Wuytack ed., LNM 765, Springer-Verlag, Berlin, (1979), pp.36-60.

[8] J. GILEWICZ, "Approximants de Padé", LNM 667, Springer-Verlag, Berlin, 1978.

[9] C. LANCZOS," Applied Analysis", Prentice Hall, Englewood Cliffs, 1956.

[10] M. MORANDI CECCHI, M. REDIVO ZAGLIA et G.SCENNA, Approximation of the numerical solution of parabolic problems, *Computational and Applied Mathematics I*, (1992), 71-80.

[11] H. PADE, Sur la représentation approchée d'une fonction par des fractions rationnelles, *Ann. Ec. Norm. Sup.*,**9** (1892), 1-93, 1892.

[12] M. PREVOST, Determinantal expression for partial Padé approximants, *Appl. Numer. Math.*, **6** (1989-1990), 221-224.

[13] M. PREVOST, Approximation of weight function and approached Padé approximants, *JCAM*,**32**(1990), 237-252.

[14] D. SHANKS, Nonlinear transformations of divergent and slowly convergent sequences, *J. Math. Phys.*, **34**(1955), 1-42.

[15] A. SIDI, D. LEVIN, Prediction properties of the t-transform, *SIAM J. Numer. Anal.*, **20** (1983), 589-598.

[16] A. TRIAS, M. KIWI, M.WEISSMANN, Reconstruction of the density of states from its moments, *Phys. Rev. B*, **28**(1983), 1859-1863.

[17] D. VEKEMANS, "Algorithmes pour Méthodes de Prédiction", *Thèse, Université de Lille 1*, 1995.

[18] D. VEKEMANS, Algorithm for the E-prediction, *to appear in JCAM.*

[19] P. WYNN, On a device for computing the $e_m(S_n)$ transformation, *MTAC*, **10**(1956), 91-96.

[20] P. WYNN, Upon systems of recursions which obtain among the quotients of the Padé table, *Numer. Math.*, **8**(1966), 264-269.