Introduction

0.1 Why Extrapolation–Convergence Acceleration?

In many problems of scientific computing, one is faced with the task of finding or approximating limits of infinite sequences. Such sequences may arise in different disciplines and contexts and in various ways. In most cases of practical interest, the sequences in question converge to their limits very slowly. This may cause their direct use for approximating their limits to become computationally expensive or impossible.

There are other cases in which these sequences may even diverge. In such a case, we are left with the question of whether the divergent sequence represents anything, and if so, what it represents. Although in some cases the elements of a divergent sequence can be used as approximations to the quantity it represents subject to certain conditions, in most other cases it is meaningless to make direct use of the sequence elements for this purpose.

Let us consider two very common examples:

(i) Summation of infinite series: This is a problem that arises in many scientific disciplines, such as applied mathematics, theoretical physics, and theoretical chemistry. In this problem, the sequences in question are those of partial sums. In some cases, the terms a_k of a series $\sum_{k=0}^{\infty} a_k$ may be known analytically. In other cases, these terms may be generated numerically, but the process of generating more and more terms may become very costly. In both situations, if the series converges very slowly, the task of obtaining good approximations to its sum only from its partial sums $A_n = \sum_{k=0}^{n} a_k,\ n = 0, 1, \ldots,$ may thus become very expensive as it necessitates a very large number of the terms a_k. In yet other cases, only a finite number of the terms a_k, say $a_0, a_1, \ldots, a_N$, may be known. In such a situation, the accuracy of the best available approximation to the sum of $\sum_{k=0}^{\infty} a_k$ is normally that of the partial sum A_N and thus cannot be improved further. If the series diverges, then its partial sums have only limited direct use. Divergent series arise naturally in different fields, perturbation analysis in theoretical physics being one of them. Divergent power series arise in the solution of homogeneous ordinary differential equations around irregular singular points.

(ii) Iterative solution of linear and nonlinear systems of equations: This problem occurs very commonly in applied mathematics and different branches of engineering. When continuum problems are solved by methods such as finite differences and finite elements, large and sparse systems of linear and/or nonlinear equations are obtained. A very attractive way of solving these systems is by iterative methods. The sequences in question for this case are those of the iteration vectors that have a large dimension in general. In most cases, these sequences converge very slowly. If the cost of computing one iteration vector is very high, then obtaining a good approximation to the solution of a given system of equations may also become very high.

The problems of slow convergence or even divergence of sequences can be overcome under suitable conditions by applying *extrapolation methods* (equivalently, *convergence acceleration methods* or *sequence transformations*) to the given sequences. When appropriate, an extrapolation method produces from a given sequence $\{A_n\}$ a new sequence $\{\hat{A}_n\}$ that converges to the former's limit more quickly when this limit exists. In case the limit of $\{A_n\}$ does not exist, the new sequence $\{\hat{A}_n\}$ produced by the extrapolation method either diverges more slowly than $\{A_n\}$ or converges to some quantity called the *antilimit* of $\{A_n\}$ that has a useful meaning and interpretation in most applications. We note at this point that the precise meaning of the antilimit may vary depending on the type of the divergent sequence, and that several possibilities exist. In the next section, we shall demonstrate through examples how antilimits may arise and what exactly they may be.

Concerning divergent sequences, there are three important messages that we would like to get across in this book: (i) Divergent sequences can be interpreted appropriately in many cases of interest, and useful antilimits for them can be defined. (ii) Extrapolation methods can be used to produce good approximations to the relevant antilimits in an efficient manner. (iii) Divergent sequences can be treated on an equal footing with convergent ones, both computationally and theoretically, and this is what we do throughout this book. (However, everywhere-divergent infinite power series, that is, those with zero radius of convergence, are not included in the theoretical treatment generally.)

It must be emphasized that each $\hat{A}_n$ is determined from only a *finite* number of the A_m. This is a basic requirement that extrapolation methods must satisfy. Obviously, an extrapolation method that requires knowledge of all the A_m for determining a given $\hat{A}_n$ is of no practical value.

We now pause to illustrate the somewhat abstract discussion presented above with the Aitken Δ^2-process that is one of the classic examples of extrapolation methods. This method was first described in Aitken [2], and it can be found in almost every book on numerical analysis. See, for example, Henrici [130], Ralston and Rabinowitz [235], Stoer and Bulirsch [326], and Atkinson [13].

Example 0.1.1 Let the sequence $\{A_n\}$ be such that

$$A_n = A + a\lambda^n + r_n \;\text{ with }\; r_n = b\mu^n + o(\min\{1, |\mu|^n\}) \;\text{ as } n \to \infty, \qquad (0.1.1)$$

where $A,\ a,\ b,\ \lambda$, and μ are in general complex scalars, and

$$a, b \neq 0, \quad \lambda, \mu \neq 0, 1, \quad \text{and} \quad |\lambda| > |\mu|. \qquad (0.1.2)$$

As a result, $r_n \sim b\mu^n = o(\lambda^n)$ as $n \to \infty$. If $|\lambda| < 1$, then $\lim_{n\to\infty} A_n = A$. If $|\lambda| \geq 1$, then $\lim_{n\to\infty} A_n$ does not exist, A being the antilimit of $\{A_n\}$ in this case. Consider now the Aitken Δ^2-process, which is an extrapolation method that, when applied to $\{A_n\}$, produces a sequence $\{\hat{A}_n\}$ with

$$\hat{A}_n = \frac{A_n A_{n+2} - A_{n+1}^2}{A_n - 2A_{n+1} + A_{n+2}} = \frac{\begin{vmatrix} A_n & A_{n+1} \\ \Delta A_n & \Delta A_{n+1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ \Delta A_n & \Delta A_{n+1} \end{vmatrix}}, \tag{0.1.3}$$

where $\Delta A_m = A_{m+1} - A_m$, $m \geq 0$. To see how $\hat{A}_n$ behaves for $n \to \infty$, we substitute (0.1.1) in (0.1.3). Taking into account the fact that $r_{n+1} \sim \mu r_n$ as $n \to \infty$, after some simple algebra it can be shown that

$$|\hat{A}_n - A| \leq \alpha |r_n| = O(\mu^n) = o(\lambda^n) \text{ as } n \to \infty, \tag{0.1.4}$$

for some positive constant α that is independent of n. Obviously, when $\lim_{n\to\infty} r_n = 0$, the sequence $\{\hat{A}_n\}$ converges to A whether $\{A_n\}$ converges or not. [If $r_n = 0$ for $n \geq N$, then $\hat{A}_n = A$ for $n \geq N$ as well, as implied by (0.1.4). In fact, the formula for $\hat{A}_n$ in (0.1.3) is obtained by requiring that $\hat{A}_n = A$ when $r_n = 0$ for all large n, and it is the solution for A of the equations $A_m = A + a\lambda^m$, $m = n, n+1, n+2$.] Also, in case $\{A_n\}$ converges, $\{\hat{A}_n\}$ converges more quickly and to $\lim_{n\to\infty} A_n = A$, because $A_n - A \sim a\lambda^n$ as $n \to \infty$ from (0.1.1) and (0.1.2). Thus, the rate of convergence of $\{A_n\}$ is enhanced by the factor

$$\frac{|\hat{A}_n - A|}{|A_n - A|} = O(|\mu/\lambda|^n) = o(1) \text{ as } n \to \infty. \tag{0.1.5}$$

A more detailed analysis of $\hat{A}_n - A$ yields the result

$$\hat{A}_n - A \sim b\frac{(\lambda - \mu)^2}{(\lambda - 1)^2}\mu^n \text{ as } n \to \infty, \tag{0.1.6}$$

that is more refined than (0.1.4) and asymptotically best possible as well. It is clear from (0.1.6) that, when the sequence $\{r_n\}$ does not converge to 0, which happens when $|\mu| \geq 1$, both $\{A_n\}$ and $\{\hat{A}_n\}$ diverge, but $\{\hat{A}_n\}$ diverges more slowly than $\{A_n\}$.

In view of this example and the discussion that preceded it, we now introduce the concepts of *convergence acceleration* and *acceleration factor*.

Definition 0.1.2 Let $\{A_n\}$ be a sequence of in general complex scalars, and let $\{\hat{A}_n\}$ be the sequence generated by applying the extrapolation method ExtM to $\{A_n\}$, $\hat{A}_n$ being determined from A_m, $0 \leq m \leq L_n$, for some integer L_n, $n = 0, 1, \ldots$. Assume that $\lim_{n\to\infty} \hat{A}_n = A$ for some A and that, if $\lim_{n\to\infty} A_n$ exists, it is equal to this A. We shall say that $\{\hat{A}_n\}$ *converges more quickly* than $\{A_n\}$ if

$$\lim_{n\to\infty} \frac{|\hat{A}_n - A|}{|A_{L_n} - A|} = 0, \tag{0.1.7}$$

whether $\lim_{n\to\infty} A_n$ exists or not. When (0.1.7) holds we shall also say that the extrapolation method ExtM *accelerates the convergence* of $\{A_n\}$. The ratio $R_n = |\hat{A}_n - A|/|A_{L_n} - A|$ is called the *acceleration factor* of $\hat{A}_n$.

The ratios R_n measure the extent of the acceleration induced by the extrapolation method ExtM on $\{A_n\}$. Indeed, from $|\hat{A}_n - A| = R_n |A_{L_n} - A|$, it is obvious that R_n is the factor by which the acceleration process reduces $|A_{L_n} - A|$ in generating $\hat{A}_n$. Obviously, a good extrapolation method is one whose acceleration factors tend to zero quickly as $n \to \infty$.

In case $\{A_n\}$ is a sequence of vectors in some general vector space, the preceding definition is still valid, provided we replace $|A_{L_n} - A|$ and $|\hat{A}_n - A|$ everywhere with $\| A_{L_n} - A \|$ and $\| \hat{A}_n - A \|$, respectively, where $\| \cdot \|$ is the norm in the vector space under consideration.

0.2 Antilimits Versus Limits

Before going on, we would like to dwell on the concept of antilimit that we mentioned briefly above. This concept can best be explained by examples to which we now turn. These examples do not exhaust all the possibilities for antilimits by any means. We shall encounter more later in this book.

Example 0.2.1 Let A_n, $n = 0, 1, 2, \ldots$, be the partial sums of the power series $\sum_{k=0}^{\infty} a_k z^k$, that is, $A_n = \sum_{k=0}^{n} a_k z^k$, $n = 0, 1, \ldots$. If the radius of convergence ρ of this series is finite and positive, then $\lim_{n\to\infty} A_n$ exists for $|z| < \rho$ and is a function $f(z)$ that is analytic for $|z| < \rho$. Of course, $\sum_{k=0}^{\infty} a_k z^k$ diverges for $|z| > \rho$. If $f(z)$ can be continued analytically to $|z| = \rho$ and $|z| > \rho$, then the analytic continuation of $f(z)$ is the antilimit of $\{A_n\}$ for $|z| \geq \rho$.

As an illustration, let us pick $a_0 = 0$ and $a_k = -1/k$, $k = 1, 2, \ldots$, so that $\rho = 1$ and $\lim_{n\to\infty} A_n = \log(1 - z)$ for $|z| \leq 1$, $z \neq 1$. The principal branch of $\log(1 - z)$ that is analytic for all complex $z \notin [1, +\infty)$ serves as the antilimit of $\{A_n\}$ in case $|z| > 1$ but $z \notin [1, +\infty)$.

Example 0.2.2 Let A_n, $n = 0, 1, 2, \ldots$, be the partial sums of the Fourier series $\sum_{k=-\infty}^{\infty} a_k e^{\mathrm{i}kx}$; that is, $A_n = \sum_{k=-n}^{n} a_k e^{\mathrm{i}kx}$, $n = 0, 1, 2, \ldots$, and assume that $C_1 |k|^{\alpha} \leq |a_k| \leq C_2 |k|^{\alpha}$ for all large $|k|$ and some positive constants C_1 and C_2 and for some $\alpha \geq 0$, so that $\lim_{n\to\infty} A_n$ does not exist. This Fourier series represents a 2π-periodic *generalized function*; see Lighthill [167]. If, for x in some interval I of $[0, 2\pi]$, this generalized function coincides with an ordinary function $f(x)$, then $f(x)$ is the antilimit of $\{A_n\}$ for $x \in I$. (Recall that $\lim_{n\to\infty} A_n$, in general, exists when $\alpha < 0$ and a_n is monotonic in n. It exists unconditionally when $\alpha < -1$.)

As an illustration, let us pick $a_0 = 0$ and $a_k = 1$, $k = \pm 1, \pm 2, \ldots$. Then the series $\sum_{k=-\infty}^{\infty} a_k e^{\mathrm{i}kx}$ represents the generalized function $-1 + 2\pi \sum_{m=-\infty}^{\infty} \delta(x - 2m\pi)$, where $\delta(z)$ is the Dirac delta function. This generalized function coincides with the ordinary function $f(x) = -1$ in the interval $(0, 2\pi)$, and $f(x)$ serves as the antilimit of $\{A_n\}$ for $n \to \infty$ when $x \in (0, 2\pi)$.

Example 0.2.3 Let $0 < x_0 < x_1 < x_2 < \cdots$, $\lim_{n\to\infty} x_n = \infty$, $s \neq 0$ and real, and let A_n be defined as $A_n = \int_0^{x_n} g(t) e^{\mathrm{i}st}\, dt$, $n = 0, 1, 2, \ldots$, where $C_1 t^{\alpha} \leq |g(t)| \leq C_2 t^{\alpha}$ for all large t and some positive constants C_1 and C_2 and for some $\alpha \geq 0$, so that

$\lim_{n\to\infty} A_n$ does not exist. In many such cases, the antilimit of $\{A_n\}$ is the *Abel sum* of the divergent integral $\int_0^\infty g(t)e^{\mathrm{i}st}\,dt$ (see, e.g., Hardy [123]) that is defined by $\lim_{\epsilon\to 0+}\int_0^\infty e^{-\epsilon t}g(t)e^{\mathrm{i}st}\,dt$. [Recall that $\int_0^\infty g(t)e^{\mathrm{i}st}\,dt$ exists and $\lim_{n\to\infty} A_n = \int_0^\infty g(t)e^{\mathrm{i}st}\,dt$, in general, when $\alpha < 0$ and $g(t)$ is monotonic in t for large t. This is true unconditionally when $\alpha < -1$.]

As an illustration, let us pick $g(t) = t^{1/2}$. Then the Abel sum of the divergent integral $\int_0^\infty t^{1/2}e^{\mathrm{i}st}\,dt$ is $e^{\mathrm{i}3\pi/4}\sqrt{\pi}/(2s^{3/2})$, and it serves as the antilimit of $\{A_n\}$.

Example 0.2.4 Let $\{h_n\}$ be a sequence in $(0, 1)$ satisfying $h_0 > h_1 > h_2 > \cdots$, and $\lim_{n\to\infty} h_n = 0$, and define $A_n = \int_{h_n}^1 x^\alpha g(x)\,dx,\ n = 0, 1, 2, \ldots$, where $g(x)$ is continuously differentiable on $[0, 1]$ a sufficient number of times with $g(0) \neq 0$ and α is in general complex and $\Re\alpha \leq -1$ but $\alpha \neq -1, -2, \ldots$. Under these conditions $\lim_{n\to\infty} A_n$ does not exist. The antilimit of $\{A_n\}$ in this case is the *Hadamard finite part* of the divergent integral $\int_0^1 x^\alpha g(x)\,dx$ (see Davis and Rabinowitz [63]) that is given by the expression

$$\sum_{i=0}^{m-1} \frac{1}{\alpha+i+1}\frac{g^{(i)}(0)}{i!} + \int_0^1 x^\alpha \left[g(x) - \sum_{i=0}^{m-1}\frac{g^{(i)}(0)}{i!}x^i\right]dx$$

with $m > -\Re\alpha - 1$ so that the integral in this expression exists as an ordinary integral. [Recall that $\int_0^1 x^\alpha g(x)\,dx$ exists and $\lim_{n\to\infty} A_n = \int_0^1 x^\alpha g(x)\,dx$ for $\Re\alpha > -1$.]

As an illustration, let us pick $g(x) = (1+x)^{-1}$ and $\alpha = -3/2$. Then the Hadamard finite part of $\int_0^1 x^{-3/2}(1+x)^{-1}\,dx$ is $-2 - \pi/2$, and it serves as the antilimit of $\{A_n\}$. Note that $\lim_{n\to\infty} A_n = +\infty$ but the associated antilimit is negative.

Example 0.2.5 Let s be the solution to the nonsingular linear system of equations $(I - T)x = c$, and let $\{x_n\}$ be defined by the iterative scheme $x_{n+1} = Tx_n + c,\ n = 0, 1, 2, \ldots$, with x_0 given. Let $\rho(T)$ denote the spectral radius of T. If $\rho(T) > 1$, then $\{x_n\}$ diverges in general. The antilimit of $\{x_n\}$ in this case is the solution s itself. [Recall that $\lim_{n\to\infty} x_n$ exists and is equal to s when $\rho(T) < 1$.]

As should become clear from these examples, the antilimit may have different meanings depending on the nature of the sequence $\{A_n\}$. Thus, it does not seem to be possible to define antilimits in a unique way, and we do not attempt to do this. It appears, though, that studying the asymptotic behavior of A_n for $n \to \infty$ is very helpful in determining the meaning of the relevant antilimit. We hope that what the antilimit of a given divergent sequence is will become more apparent as we proceed to the study of extrapolation methods.

0.3 General Algebraic Properties of Extrapolation Methods

We saw in Section 0.1 that an extrapolation method operates on a given sequence $\{A_n\}$ to produce a new sequence $\{\hat{A}_n\}$. That is, it acts as a mapping from $\{A_n\}$ to $\{\hat{A}_n\}$. In all cases of interest, this mapping has the general form

$$\hat{A}_n = \Phi_n(A_0, A_1, \ldots, A_{L_n}), \tag{0.3.1}$$

where L_n is some *finite* positive integer. (As mentioned earlier, methods for which $L_n = \infty$ are of no use, because they require knowledge of all the A_m to obtain $\hat{A}_n$ with finite n.) In addition, for most extrapolation methods there holds

$$\hat{A}_n = \sum_{i=0}^{K_n} \theta_{ni} A_i, \tag{0.3.2}$$

where K_n are some nonnegative integers and the θ_{ni} are some scalars that satisfy

$$\sum_{i=0}^{K_n} \theta_{ni} = 1 \tag{0.3.3}$$

for each n. (This is the case for all of the extrapolation methods we consider in this work.) A consequence of (0.3.2) and (0.3.3) is that such extrapolation methods act as *summability methods* for the sequence $\{A_n\}$.

When the θ_{ni} are independent of the A_m, the approximation $\hat{A}_n$ is linear in the A_m, thus the extrapolation method that generates $\{\hat{A}_n\}$ becomes a *linear summability method*. That is to say, this extrapolation method can be applied to every sequence $\{A_n\}$ with the same θ_{ni}. Both numerical experience and the different known convergence analyses suggest that linear methods are of limited scope and not as effective as nonlinear methods.

As the subject of linear summability methods is very well-developed and is treated in different books, we are not going to dwell on it in this book; see, for example, the books by Knopp [152], Hardy [123], and Powell and Shah [231]. We only give the definition of linear summability methods at the end of this section and recall the Silverman–Toeplitz theorem, which is one of the fundamental results on linear summability methods. Later in this work, we also discuss the Euler transformation that has been used in different practical situations and that is probably the most successful linear summability method.

When the θ_{ni} depend on the A_m, the approximation $\hat{A}_n$ is *nonlinear* in the A_m. This implies that if $C_m = \alpha A_m + \beta B_m$, $m = 0, 1, 2, \ldots$, for some constants α and β, and $\{\hat{A}_n\}$, $\{\hat{B}_n\}$, and $\{\hat{C}_n\}$ are obtained by applying a given nonlinear extrapolation method to $\{A_n\}$, $\{B_n\}$, and $\{C_n\}$, respectively, then $\hat{C}_n \neq \alpha\hat{A}_n + \beta\hat{B}_n$, $n = 0, 1, 2, \ldots$, in general. (Equality prevails for all n when the extrapolation method is linear.) Despite this fact, most nonlinear extrapolation methods enjoy a "sort of linearity" property that can be described as follows: Let $\alpha \neq 0$ and β be arbitrary constants and consider $C_m = \alpha A_m + \beta$, $m = 0, 1, 2, \ldots$. Then

$$\hat{C}_n = \alpha\hat{A}_n + \beta, \quad n = 0, 1, 2, \ldots. \tag{0.3.4}$$

In other words, $\{C_n\} = \alpha\{A_n\} + \beta$ implies $\{\hat{C}_n\} = \alpha\{\hat{A}_n\} + \beta$. This is called the *quasi-linearity* property and is a useful property that we want every extrapolation method to have. (All extrapolation methods treated in this book are quasi-linear.) A sufficient condition for this to hold is given in Proposition 0.3.1.

Proposition 0.3.1 *Let a nonlinear extrapolation method be such that the sequence $\{\hat{A}_n\}$ that it produces from $\{A_n\}$ satisfies (0.3.2) with (0.3.3). Then the sequence $\{\hat{C}_n\}$ that it produces from $\{C_n = \alpha A_n + \beta\}$ for arbitrary constants $\alpha \neq 0$ and β satisfies the* *quasi-linearity property in (0.3.4) if the θ_{ni} in (0.3.2) depend on the A_m through the $\Delta A_m = A_{m+1} - A_m$ only and are homogeneous in the ΔA_m of degree 0.*

Remark. We recall that a function $f(x_1, \ldots, x_p)$ is *homogeneous of degree r* if, for every $\lambda \neq 0$, $f(\lambda x_1, \ldots, \lambda x_p) = \lambda^r f(x_1, \ldots, x_p)$.

Proof. We begin by rewriting (0.3.2) in the form $\hat{A}_n = \sum_{i=0}^{K_n} \theta_{ni}(\{A_m\})A_i$. Similarly, we have $\hat{C}_n = \sum_{i=0}^{K_n} \theta_{ni}(\{C_m\})C_i$. From (0.3.1) and the conditions imposed on the θ_{ni}, there exist functions $D_{ni}(\{u_m\})$ for which

$$\theta_{ni}(\{A_m\}) = D_{ni}(\{\Delta A_m\}) \text{ and } \theta_{ni}(\{C_m\}) = D_{ni}(\{\Delta C_m\}), \tag{0.3.5}$$

where the functions D_{ni} satisfy for all $\lambda \neq 0$

$$D_{ni}(\{\lambda u_m\}) = D_{ni}(\{u_m\}). \tag{0.3.6}$$

This and the fact that $\{\Delta C_m\} = \{\alpha \Delta A_m\}$ imply that

$$\theta_{ni}(\{C_m\}) = D_{ni}(\{\Delta C_m\}) = D_{ni}(\{\Delta A_m\}) = \theta_{ni}(\{A_m\}). \tag{0.3.7}$$

From (0.3.2) and (0.3.7) we have, therefore,

$$\hat{C}_n = \sum_{i=0}^{K_n} \theta_{ni}(\{A_m\})(\alpha A_i + \beta) = \alpha \hat{A}_n + \beta \sum_{i=0}^{K_n} \theta_{ni}(\{A_m\}). \tag{0.3.8}$$

The result now follows by invoking (0.3.3). ■

Example 0.3.2 Consider the Aitken Δ^2-process that was given by (0.1.3) in Example 0.1.1. We can reexpress $\hat{A}_n$ in the form

$$\hat{A}_n = \theta_{n,n} A_n + \theta_{n,n+1} A_{n+1}, \tag{0.3.9}$$

with

$$\theta_{n,n} = \frac{\Delta A_{n+1}}{\Delta A_{n+1} - \Delta A_n}, \quad \theta_{n,n+1} = \frac{-\Delta A_n}{\Delta A_{n+1} - \Delta A_n}. \tag{0.3.10}$$

Thus, $\theta_{ni} = 0$ for $0 \leq i \leq n-1$. It is easy to see that the θ_{ni} satisfy the conditions of Proposition 0.3.1 so that the Δ^2-process has the quasi-linearity property described in (0.3.4). Note also that for this method $L_n = n+2$ in (0.3.1) and $K_n = n+1$ in (0.3.2).

0.3.1 Linear Summability Methods and the Silverman–Toeplitz Theorem

We now go back briefly to linear summability methods. Consider the infinite matrix

$$M = \begin{bmatrix} \mu_{00} & \mu_{01} & \mu_{02} & \cdots \\ \mu_{10} & \mu_{11} & \mu_{12} & \cdots \\ \mu_{20} & \mu_{21} & \mu_{22} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}, \tag{0.3.11}$$

where μ_{ni} are some fixed scalars. The linear summability method associated with M is the linear mapping that transforms an arbitrary sequence $\{A_n\}$ to another sequence $\{A'_n\}$ through

$$A'_n = \sum_{i=0}^{\infty} \mu_{ni} A_i, \quad n = 0, 1, 2, \dots . \tag{0.3.12}$$

This method is *regular* if $\lim_{n\to\infty} A_n = A$ implies $\lim_{n\to\infty} A'_n = A$. The Silverman–Toeplitz theorem that we state next gives necessary and sufficient conditions for a linear summability method to be regular. For proofs of this fundamental result see, for example, the books by Hardy [123] and Powell and Shah [231].

Theorem 0.3.3 *(Silverman–Toeplitz theorem). The summability method associated with the matrix M in (0.3.11) is regular if and only if the following three conditions are fulfilled simultaneously:*

(i) $\lim_{n\to\infty} \sum_{i=0}^{\infty} \mu_{ni} = 1$.
(ii) $\lim_{n\to\infty} \mu_{ni} = 0, \ i = 0, 1, 2, \dots$.
(iii) $\sup_n \sum_{i=0}^{\infty} |\mu_{ni}| < \infty$.

Going back to the beginning of this section, we see that (0.3.3) is analogous to condition (i) of Theorem 0.3.3. The issue of numerical stability discussed in Section 0.5 is very closely related to condition (iii), as will become clear shortly.

The linear summability methods that have been of practical use are those whose associated matrices M are lower triangular, that is, those for which $A'_n = \sum_{i=0}^{n} \mu_{ni} A_i$. Excellent treatments of these methods from the point of view of convergence acceleration, including an extensive bibliography, have been presented by Wimp [363], [364], [365], [366, Chapters 2–4].

0.4 Remarks on Algorithms for Extrapolation Methods

A relatively important issue in the subject of extrapolation methods is the development of efficient algorithms (computational procedures) for implementing existing extrapolation methods. An efficient algorithm is one that involves a small number of arithmetic operations and little storage when storage becomes a problem.

Some extrapolation methods already have known closed-form expressions for the sequences $\{\hat{A}_n\}$ they generate. This is the case, for example, for the Aitken Δ^2-process. One possible algorithm for such methods may be the direct computation of the closed-form expressions. This is also the most obvious, but not necessarily the most economical, approach in all cases.

Many extrapolation methods are defined through systems of linear or nonlinear equations, that is, they are defined implicitly by systems of the form

$$\Psi_{n,i}(\hat{A}_n, \alpha_1, \alpha_2, \dots, \alpha_{q_n}; \{A_m\}) = 0, \quad i = 0, 1, \dots, q_n, \tag{0.4.1}$$

in which $\hat{A}_n$ is the main quantity we are after, and $\alpha_1, \alpha_2, \dots, \alpha_{q_n}$ are additional auxiliary unknowns. As we will see in the next chapters, the better sequences $\{\hat{A}_n\}$ are generated by those extrapolation methods with large q_n, in general. This means that we actually want to solve large systems of equations, which may be a computationally expensive proposition. In such cases, the development of good algorithms becomes especially important. The next example helps make this point clear.

Example 0.4.1 The Shanks [264] transformation of order k is an extrapolation method, which, when applied to a sequence $\{A_n\}$, produces the sequence $\{\hat{A}_n = e_k(A_n)\}$, where $e_k(A_n)$ satisfies the nonlinear system of equations

$$A_r = e_k(A_n) + \sum_{i=1}^{k} \alpha_i \lambda_i^r, \quad n \le r \le n + 2k, \tag{0.4.2}$$

where α_i and λ_i are additional (auxiliary) $2k$ unknowns. Provided this system has a solution with $\alpha_i \neq 0$ and $\lambda_i \neq 0, 1$ and $\lambda_i \neq \lambda_j$ if $i \neq j$, then $e_k(A_n)$ can be shown to satisfy the linear system

$$A_r = e_k(A_n) + \sum_{i=1}^{k} \beta_i \Delta A_{r+i-1}, \quad n \le r \le n + k, \tag{0.4.3}$$

where β_i are additional (auxiliary) k unknowns. Here $\Delta A_m = A_{m+1} - A_m$, $m = 0, 1, \ldots,$ as before. [In any case, we can start with (0.4.3) as the definition of $e_k(A_n)$.] Now, this linear system can be solved using Cramer's rule, giving us $e_k(A_n)$ as the ratio of two $(k+1) \times (k+1)$ determinants in the form

$$e_k(A_n) = \frac{\begin{vmatrix} A_n & A_{n+1} & \cdots & A_{n+k} \\ \Delta A_n & \Delta A_{n+1} & \cdots & \Delta A_{n+k} \\ \vdots & \vdots & & \vdots \\ \Delta A_{n+k-1} & \Delta A_{n+k} & \cdots & \Delta A_{n+2k-1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \Delta A_n & \Delta A_{n+1} & \cdots & \Delta A_{n+k} \\ \vdots & \vdots & & \vdots \\ \Delta A_{n+k-1} & \Delta A_{n+k} & \cdots & \Delta A_{n+2k-1} \end{vmatrix}}. \tag{0.4.4}$$

We can use this determinantal representation to compute $e_k(A_n)$, but this would be very expensive for large k and thus would constitute a bad algorithm. A better algorithm is one that solves the linear system in (0.4.3) by Gaussian elimination. But this algorithm too becomes costly for large k. The ε-algorithm of Wynn [368], on the other hand, is very efficient as it produces all of the $e_k(A_n)$, $0 \le n + 2k \le N$, that are defined by $A_0, A_1, \ldots, A_N$ in only $O(N^2)$ operations. It reads

$$\begin{aligned} &\varepsilon_{-1}^{(n)} = 0, \quad \varepsilon_0^{(n)} = A_n, \quad n = 0, 1, \ldots, \\ &\varepsilon_{k+1}^{(n)} = \varepsilon_{k-1}^{(n+1)} + \frac{1}{\varepsilon_k^{(n+1)} - \varepsilon_k^{(n)}}, \quad n, k = 0, 1, \ldots, \end{aligned} \tag{0.4.5}$$

and we have

$$e_k(A_n) = \varepsilon_{2k}^{(n)}, \quad n, k = 0, 1, \ldots. \tag{0.4.6}$$

Incidentally, $\hat{A}_n$ in (0.1.3) produced by the Aitken Δ^2-process is nothing but $e_1(A_n)$. [Note that the Shanks transformations are quasi-linear extrapolation methods. This can be seen either from the equations in (0.4.2), or from those in (0.4.3), or from the determinantal representation of $e_k(A_n)$ in (0.4.4), or even from the ε-algorithm itself.]

Finally, there are extrapolation methods in the literature that are defined exclusively by recursive algorithms from the start. The θ-algorithm of Brezinski [32] is such an extrapolation method, and it is defined by recursion relations very similar to those of the ε-algorithm.

0.5 Remarks on Convergence and Stability of Extrapolation Methods

The analysis of convergence and stability is the most important subject in the theory of extrapolation methods. It is also the richest in terms of the variety of results that exist and still can be obtained for different extrapolation methods and sequences. Thus, it is impossible to make any specific remarks about convergence and stability at this stage. We can, however, make several remarks on the approach to these topics that we take in this book. We start with the topic of convergence analysis.

0.5.1 Remarks on Study of Convergence

The first stage in the convergence analysis of extrapolation methods is formulation of conditions that we impose on the $\{A_n\}$. In this book, we deal with sequences that arise in common applications. Therefore, we emphasize mainly conditions that are relevant to these applications. Also, we keep the number of the conditions imposed on the $\{A_n\}$ to a minimum as this leads to mathematically more valuable and elegant results. The next stage is analysis of the errors $\hat{A}_n - A$ under these conditions. This analysis may lead to different types of results depending on the complexity of the situation. In some cases, we are able to give a full asymptotic expansion of $\hat{A}_n - A$ for $n \to \infty$; in other cases, we obtain only the most dominant term of this expansion. In yet other cases, we obtain a realistic upper bound on $|\hat{A}_n - A|$ from which powerful convergence results can be obtained. An important feature of our approach is that we are not content only with showing that the sequence $\{\hat{A}_n\}$ converges more quickly than $\{A_n\}$, that is, that convergence acceleration takes place in accordance with Definition 0.1.2, but instead we aim at obtaining the precise asymptotic behavior of the corresponding acceleration factor or a good upper bound for it.

0.5.2 Remarks on Study of Stability

We now turn to the topic of stability in extrapolation. Unlike convergence, this topic may not be common knowledge, so we start with some rather general remarks on what we mean by stability and how we analyze it. Our discussion here is based on those of Sidi [272], [300], [305], and is recalled in relevant places throughout the book.

When we compute the sequence $\{\hat{A}_n\}$ in finite-precision arithmetic, we obtain a sequence $\{\tilde{A}_n\}$ that is different from $\{\hat{A}_n\}$, the exact transformed sequence. This, of course,