Journal of Contemporary Applied Mathematics
V. 10, No 2, 2020, December
ISSN 2222-5498

Analysis of the Stability of Periodic Linear Difference Equation Systems on Extended Floating-Point Numbers

A. O. Çıbıkdiken, G. Çelik Kızılkan , A. Duman, K. Aydın

Abstract. The computation errors come out in the computers (uses floating point numbers) during calculations of mathematical problems. Therefore, the floating point numbers affect the results directly. In [1] and [2], the computation of fundamental matrix analyzed for the stability of periodic linear difference equation systems. These works used Wilkinson Model and Godunov Model. In this study, the computation of fundamental matrix has been studied on extended floating point numbers. The results have been applied to Schur stability of the system of linear difference equations with periodic coefficients. Furthermore, the effect of floating point arithmetics has been showed on numerical examples.

Key Words and Phrases: Floating Point, fundamental matrix, monodromy matrix, Schur stability, periodic linear difference equations.

2010 Mathematics Subject Classifications: 39A30, 65G50, 65Q10, 68Q01

1. Introduction

The computations are performed by floating point arithmetics in the computers. The numbers required in the computations are taken into memory and processed in accordance with the floating-point number structure. Many approaches have been introduced for the use of floating-point numbers ([17, 21, 5, 7, 12, 23]). In 1985, a standard was created by the IEEE (Institute of Electrical and Electronics Engineers) and ANSI (American National Standards Institute) for floating point numbers. This standard, called ANSI / IEEE Standard 754-1985, sets out the main principles of floating point numbers, based on binary systems. Errors are occurred during loading the real numbers to memory as floating-point numbers. The effect of these errors to computations has are examined by numerous researchers.

In this study, the computation of fundamental matrix will be used to analyze the stability of periodic linear difference equation systems. Therefore, the computation of fundamental matrix should be examined on floating-point numbers. In [1], the effect of computation error of monodromy matrix was examined using Wilkinson Model. Also, Godunov Model [2] has been used to determine the effect of computation error in applying to stability of periodic systems. Extended Floating Point Numbers (Extended Format) will be used to examine the effect of a computed fundamental matrix for analyzing stability in this study.

1.1. Fundamental Matrix of Periodic Linear Difference Equation System

The system (1.1) and for given $x_0 \in \mathbb{R}^N$ initial value

$$x_{n+1} = A_n x_n, x_0 - \text{initial vector}, n \geq 0 \tag{1.1}$$

is called *linear difference-Cauchy problem with periodic coefficients*. If I is an identity matrix then,

$$X_{n+1} = A_n X_n, X_0 = I, n \geq 0 \tag{1.2}$$

is a solution of Cauchy problem. In here,

$$X_n = \prod_{j=0}^{n-1} A_j = A_{n-1}A_{n-2}\cdots A_0, \quad \textit{and} \quad X_T = \prod_{j=0}^{T-1} A_j = A_{T-1}A_{T-2}\cdots A_0, \tag{1.3}$$

X_n is called *fundamental matrix* of the system (1.1) and X_T is called as *monodromy matrix* of the system (1.1) [14, 8, 9, 15, 10, 11]. The solution of the system (1.1) is

$$X_{kT+m} = X_m X_T^k x_0 \tag{1.4}$$

where $x_0 \in \mathbb{R}^N$ inital value $x_n = X_n x_0$, $n = kT + m$, $0 \leq m \leq T - 1$ [8, 9].

1.2. Extended Floating-Point Numbers (Extended Format, $\mathbb{F}^E$)

The set

$$\mathbb{F} = \mathbb{F}(\gamma, p_-, p_+, k) \tag{1.5}$$

where $p_- \in \mathbb{Z}^-, k, p_+ \in \mathbb{Z}^+$ for $p_- \leq p \leq p_+, p \in \mathbb{Z}$, is called as the set of computer numbers or set Format [17, 21, 14, 16, 3, 20, 22]. In [17, 14, 18], the operator

$$fl : \mathbb{D} \to \mathbb{F},\ fl(z) = z(1+\alpha) + \beta;\ \|\alpha\| \leq u,\ \|\beta\| \leq v,\ \alpha\beta = 0 \tag{1.6}$$

converts the elements of $\mathbb{D} = [-\varepsilon_\infty, \varepsilon_\infty] \cap \mathbb{R}$ to floating point numbers, where

$$u = \begin{cases} \frac{\varepsilon_1}{2}, & \text{rounding} \\ \varepsilon_1, & \text{chopping} \end{cases},\ v = \begin{cases} \frac{\varepsilon_0^d}{2}, & \text{rounding} \\ \varepsilon_0^d, & \text{chopping} \end{cases}. \tag{1.7}$$

In (1.7), $\varepsilon_0^d = \gamma^{p_- - k}, \varepsilon_1 = \gamma^{1-k}, \varepsilon_\infty = \gamma^{p_+}\left(1 - \frac{1}{\gamma^k}\right)$ are defined.

$\mathbb{F}^E$ has been called *Extended Format* that is characterized by the parameters $\varepsilon_0^d, \varepsilon_1, \varepsilon_\infty$ on $\mathbb{F}$.

A matrix $A \in M_N(\mathbb{D})$ can be stored to memory by $fl(A) = A + \psi$. ψ is storing error of matrix A in $\mathbb{F}^E$,

$$||\psi|| \leq u\sqrt{N}||A|| + vN. \tag{1.8}$$

A multiply of matrix A, $B \in M_N(\mathbb{F}^E)$ can be stored to memory by $fl(AB) = AB + \varphi$. φ is computation error of AB,

$$||\varphi|| \leq u\sqrt{N}||AB|| + vN. \tag{1.9}$$

Linear difference Cauchy problem is written

$$fl(A_{n-1}Y_{n-1}) = Y_n = (A_{n-1} + \psi_{n-1})Y_{n-1} + \varphi_n; Y_0 = I, n = 1, 2, 3, ... \tag{1.10}$$

where $Y_n = fl(A_{n-1}Yn - 1)$ is the computation of fundamental matrix X_n, $A \in M_N(D)$.

It is clear that matrix ψ_n is the storing error of matrix A_n , matrix φ_n is computation error of $(A_{n-1} + \psi_{n-1})Y_{n-1}$, $\varphi_1 = 0$.

The solution of (1.10) can be written

$$Y(n) = \prod_{i=0}^{n-1}(A_i + \psi_i) + \sum_{k=1}^{n}(\prod_{i=k}^{n-1}(A_i + \psi_i))\varphi_k. \tag{1.11}$$

2. Computation of Fundamental Matrix

In this section the computation of fundamental matrix X_n will be examined on floating point arithmetic in $\mathbb{F}^E$.

Let us investigate the upper bound of computation error matrix φ_n given by (1.10) according to

$$||\psi_{n-1}|| \leq u\sqrt{N}||A_{n-1}|| + vN \tag{2.1}$$

and

$$||\varphi_n|| \leq u\sqrt{N}||A_{n-1} + \psi_{n-1}||||Y_{n-1}|| + vN, Y_0 = I, n = 2, 3, ... \tag{2.2}$$

where $A_n \in M_N(D)$, ψ_n is placement error matrix and u is introduced in (1.7).

Theorem 2.1. *The following inequality holds:*

$$||\varphi_n|| \leq u\sqrt{N}(1 + u\sqrt{N})^{n-2}[(1 + u\sqrt{N})\max_{0\leq j\leq n-1}||A_j|| + vN]^n$$
$$+uvN^{3/2}\sum_{j=2}^{n-1}(1 + u\sqrt{N})^{n-j-1}[(1 + u\sqrt{N})\max_{2\leq k\leq n-1}||A_k|| + vN]^{n-j} + vN,$$

where φ_n is given by (2.2) for $n = 2, 3, ...$.

Proof. For $n = 2, 3, ...$, consider

$||\varphi_n|| \leq u\sqrt{N}(||A_{n-1}|| + ||\psi_{n-1}||)||Y_{n-1}|| + vN$; $||Y_n|| \leq (||A_{n-1}|| + ||\psi_{n-1}||)||Y_{n-1}|| + ||\varphi_n||$

from inequality (2.2). If we write Y_{n-1} and φ_{n-1} in inequality (2.2) we obtain

$$||\varphi_n|| \leq u\sqrt{N}(||A_{n-1}|| + ||\psi_{n-1}||)[(||A_{n-2}|| + ||\psi_{n-2}||)||Y_{n-2}|| + ||\varphi_{n-1}||] + vN$$
$$\leq u\sqrt{N}(||A_{n-1}|| + ||\psi_{n-1}||)(||A_{n-2}|| + ||\psi_{n-2}||)||Y_{n-2}||$$
$$+u^2N(||A_{n-1}|| + ||\psi_{n-1}||)(||A_{n-2}|| + ||\psi_{n-2}||)||Y_{n-2}||$$

$$+u\sqrt{N}(||A_{n-1}||+||\psi_{n-1}||)vN+vN$$
$$=u\sqrt{N}(1+u\sqrt{N})(||A_{n-1}||+||\psi_{n-1}||)(||A_{n-2}||+||\psi_{n-2}||)||Y_{n-2}||$$
$$+u\sqrt{N}(||A_{n-1}||+||\psi_{n-1}||)vN+vN.$$

Let's continue to iterate in the same way. So we obtain

$$||\varphi_n|| \leq u\sqrt{N}(1+u\sqrt{N})^{n-2}(||A_{n-1}||+||\psi_{n-1}||)(||A_{n-2}||+||\psi_{n-2}||)...(||A_0||+||\psi_0||)$$
$$+uvN^{3/2}\sum_{j=2}^{n-1}(1+u\sqrt{N})^{n-j-1}\prod_{r=j}^{n-1}(||A_r||+||\psi_r||)+vN. \tag{2.3}$$

Since $||A_i||+||\psi_i|| \leq (1+u\sqrt{N})||A_i||+vN$, we have

$$||\varphi_n|| \leq u\sqrt{N}(1+u\sqrt{N})^{n-2}[(1+u\sqrt{N})\max_{0\leq j\leq n-1}||A_j||+vN]^n$$
$$+uvN^{3/2}\sum_{j=2}^{n-1}(1+u\sqrt{N})^{n-j-1}[(1+u\sqrt{N})\max_{2\leq k\leq n-1}||A_k||+vN]^{n-j}+vN.$$

□

Corollary 2.1. *If $A_n \in M_N(F)$ then it is clear that the placement error $\psi_n = 0$. In this case, it can be easily noticed that the inequality (2.3) is same as the inequality in [2].*

Theorem 2.2. *Let Y_n be the matrix defined by (1.10), then the following inequality holds:*

$$||Y_n|| \leq (1+u\sqrt{N})^{n-1}[(1+u\sqrt{N})\max_{0\leq j\leq n-1}||A_j||+vN]^n$$
$$+vN\sum_{j=2}^{n}(1+u\sqrt{N})^{n-j}[(1+u\sqrt{N})\max_{2\leq k\leq n-1}||A_k||+vN]^{n-j},\quad n=1,2,3,....$$

Proof. For $n=2,3,...$, let us write (2.2) in

$$||Y_n|| \leq (||A_{n-1}||+||\psi_{n-1}||)||Y_{n-1}||+||\varphi_n||.$$

Thus, we obtain

$$||Y_n|| \leq (||A_{n-1}||+||\psi_{n-1}||)||Y_{n-1}||+||\varphi_n||$$
$$\leq (||A_{n-1}||+||\psi_{n-1}||)||Y_{n-1}||+u\sqrt{N}(||A_{n-1}||+||\psi_{n-1}||)||Y_{n-1}||+vN$$
$$=(1+u\sqrt{N})(||A_{n-1}||+||\psi_{n-1}||)||Y_{n-1}||+vN$$

If we continue the iteration, we have

$$||Y_n|| \leq (1+u\sqrt{N})(||A_{n-1}||+||\psi_{n-1}||)[(1+u\sqrt{N})(||A_{n-2}||+||\psi_{n-2}||)||Y_{n-2}||+vN]+vN$$
$$=(1+u\sqrt{N})^2(||A_{n-1}||+||\psi_{n-1}||)(||A_{n-2}||+||\psi_{n-2}||)||Y_{n-2}||$$
$$+(1+u\sqrt{N})(||A_{n-1}||+||\psi_{n-1}||)vN+vN$$
$$\vdots$$
$$||Y_n|| \leq (1+u\sqrt{N})^{n-1}(||A_{n-1}||+||\psi_{n-1}||)(||A_{n-2}||+||\psi_{n-2}||)...(||A_0||+||\psi_0||)$$
$$+vN\sum_{j=2}^{n}(1+u\sqrt{N})^{n-j}\prod_{r=j}^{n-1}(||A_r||+||\psi_r||).$$

Since $||A_i||+||\psi_i|| \leq (1+u\sqrt{N})||A_i||+vN$, the inequality

$$||Y_n|| \leq (1+u\sqrt{N})^{n-1}[1+u\sqrt{N})\max_{0\leq j\leq n-1}||A_j||+vN]^n$$
$$+vN\sum_{j=2}^{n}(1+u\sqrt{N})^{n-j}[(1+u\sqrt{N})\max_{2\leq k\leq n-1}||A_k||+vN]^{n-j}$$

is obtained and this is the inequality that is sought. □

Theorem 2.3. *The inequality*

$$||Y_n-X_n|| \leq \sum_{k=1}^{n}\binom{n}{k}(\max_{0\leq j\leq n-1}||A_j||)^{n-k}(u\sqrt{N}\max_{0\leq j\leq n-1}||A_j||+vN)^k$$
$$+\sum_{k=2}^{n}[||A_{n-1}||...||A_{k+1}||||A_k||$$

$+\sum_{r=1}^{n-k}\binom{n-k}{r}(\max_{0\le j\le n-1}||A_j||)^{n-k-r}(u\sqrt{N}\max_{0\le j\le n-1}||A_j|| + vN)^r]$
$\times[u\sqrt{N}(1+u\sqrt{N})^{k-2}[(1+u\sqrt{N})\max_{0\le j\le k-1}||A_j||+vN]^k$
$+uvN^{3/2}\sum_{j=2}^{k-1}(1+u\sqrt{N})^{k-j-1}[(1+u\sqrt{N})\max_{0\le j\le n-1}||A_j||+vN]^{k-j}+vN]$
holds, where $A_n \in M_N(D)$, X_n is the fundamental matrix and Y_n is the computed fundamental matrix given by (1.10).

Proof. From (1.11), we can write
$||Y_n - X_n|| \le ||\prod_{i=0}^{n-1}(A_i+\psi_i) - \prod_{i=0}^{n-1}A_i|| + \sum_{k=2}^{n}||\prod_{i=k}^{n-1}(A_i+\psi_i)||||\varphi_k||$,
where X_n is fundamental matrix and Y_n is the computed fundamental matrix. For the expressions $||\prod_{i=0}^{n-1}(A_i+\psi_i) - \prod_{i=0}^{n-1}A_i||$ and $||\prod_{i=k}^{n-1}(A_i+\psi_i)||$, we obtain the following inequalities, respectively:

$$\begin{aligned}
||\prod_{i=0}^{n-1}(A_i+\psi_i) - \prod_{i=0}^{n-1}A_i|| &= ||(A_{n-1}+\psi_{n-1})...(A_1+\psi_1)(A_0+\psi_0) - A_{n-1}...A_1A_0||\\
&= ||A_{n-1}...A_1A_0 + A_{n-1}...A_1\psi_0 + ... + \psi_{n-1}...\psi_1\psi_0 - A_{n-1}...A_1A_0||\\
&\le ||A_{n-1}...A_1\psi_0|| + ... + ||\psi_{n-1}...\psi_1\psi_0||\\
&\le (||A_{n-1}||...||A_1||||\psi_0||) + ... + (||\psi_{n-1}||...||\psi_1||||\psi_0||)\\
&\le n(\max_{0\le j\le n-1}||A_j||)^{n-1}(\max_{0\le j\le n-1}||\psi_j||)+\tfrac{n^2-n}{2}(\max_{0\le j\le n-1}||A_j||)^{n-2}(\max_{0\le j\le n-1}||\psi_j||)^2\\
&\quad +... + n(\max_{0\le j\le n-1}||A_j||)(\max_{0\le j\le n-1}||\psi_j||)^{n-1} + (\max_{0\le j\le n-1}||\psi_j||)^n\\
&= \sum_{k=1}^{n}\binom{n}{k}(\max_{0\le j\le n-1}||A_j||)^{n-k}(\max_{0\le j\le n-1}||\psi_j||)^k,
\end{aligned}$$

$$\begin{aligned}
||\prod_{i=k}^{n-1}(A_i+\psi_i)|| &= ||(A_{n-1}+\psi_{n-1})...(A_{k+1}+\psi_{k+1})(A_k+\psi_k)||\\
&= ||A_{n-1}...A_{k+1}A_k + A_{n-1}...A_{k+1}\psi_k + ... + \psi_{n-1}...\psi_{k+1}\psi_k||\\
&\le ||A_{n-1}...A_{k+1}A_k|| + ||A_{n-1}...A_{k+1}\psi_k|| + ... + ||\psi_{n-1}...\psi_{k+1}\psi_k||\\
&\le ||A_{n-1}...A_{k+1}A_k|| + (||A_{n-1}||...||A_{k+1}||||\psi_k||) + ... + (||\psi_{n-1}||...||\psi_{k+1}||||\psi_k||)\\
&\le ||A_{n-1}...A_{k+1}A_k|| + (n-k)(\max_{0\le j\le n-1}||A_j||)^{n-k-1}(\max_{0\le j\le n-1}||\psi_j||)\\
&\quad +\tfrac{(n-k)-(n-k-1)}{2}(\max_{0\le j\le n-1}||A_j||)^{n-k-2}(\max_{0\le j\le n-1}||\psi_j||)^2 + ...\\
&\quad +(n-k)(\max_{0\le j\le n-1}||A_j||)(\max_{0\le j\le n-1}||\psi_j||)^{n-k-1} + (\max_{0\le j\le n-1}||\psi_j||)^{n-k}\\
&= ||A_{n-1}...A_{k+1}A_k||+\sum_{r=1}^{n-k}\binom{n-k}{r}(\max_{0\le j\le n-1}||A_j||)^{n-k-r}(\max_{0\le j\le n-1}||\psi_j||)^r.
\end{aligned}$$

Hence, the inequality

$$||Y_n - X_n|| \le \sum_{k=1}^{n}\binom{n}{k}(\max_{0\le j\le n-1}||A_j||)^{n-k}(\max_{0\le j\le n-1}||\psi_j||)^k+$$

$$\sum_{k=2}^{n}(||A_{n-1}||...||A_{k+1}||||A_k|| + \sum_{r=1}^{n-k}\binom{n-k}{r}(\max_{0\le j\le n-1}||A_j||)^{n-k-r}(\max_{0\le j\le n-1}||\psi_j||)^r)||\varphi_k|| \tag{2.4}$$

holds. If the upper bound of $||\varphi_n||$ from 2.1and the inequality (2.1) substitute in (2.4), we have

$$\begin{aligned}
||Y_n - X_n|| \le{}& \sum_{k=1}^{n}\binom{n}{k}(\max_{0\le j\le n-1}||A_j||)^{n-k}(u\sqrt{N}\max_{0\le j\le n-1}||A_j||+vN)^k\\
&+\sum_{k=2}^{n}[||A_{n-1}||...||A_{k+1}||||A_k||+\\
&\sum_{r=1}^{n-k}\binom{n-k}{r}(\max_{0\le j\le n-1}||A_j||)^{n-k-r}(u\sqrt{N}\max_{0\le j\le n-1}||A_j||+vN)^r]\\
&\times[u\sqrt{N}(1+u\sqrt{N})^{k-2}[(1+u\sqrt{N})\max_{0\le j\le k-1}||A_j||+vN]^k
\end{aligned}$$

$+uvN^{3/2}\sum_{j=2}^{k-1}(1+u\sqrt{N})^{k-j-1}[(1+u\sqrt{N})\max_{0\le j\le k-1}||A_j||+vN]^{k-j}+vN].$ □

Theorem 2.4. *The inequality*

$$\begin{aligned}||Y_n-X_n|| \le \sum_{k=2}^{n}[(\textstyle\prod_{j=k}^{n-1}||A_j||)\\ &\times((u\sqrt{N}||A_{k-1}||+vN)(1+u\sqrt{N})^{k-2}[(1+u\sqrt{N})\max_{0\le j\le k-2}||A_j||+vN]^{k-1}\\ &+vN\sum_{j=2}^{k-1}(1+u\sqrt{N})^{k-j-1}[(1+u\sqrt{N})\max_{2\le i\le k-2}||A_i||+vN]^{k-j-1}\\ &+u\sqrt{N}(1+u\sqrt{N})^{k-2}[(1+u\sqrt{N})\max_{0\le j\le k-1}||A_j||+vN]^{k}\\ &+uvN^{3/2}\sum_{j=2}^{k-1}(1+u\sqrt{N})^{k-j-1}[(1+u\sqrt{N})\max_{2\le i\le k-1}||A_i||+vN]^{k-j}+vN)]\\ &+||A_{n-1}||||A_{n-2}||...||A_1||(u\sqrt{N}||A_0||+vN)\end{aligned}$$

holds, where $A_n\in M_N(D)$, X_n is the fundamental matrix and Y_n is the computed fundamental matrix given by (1.10).

Proof. Consider the Cauchy problem $Y_n-X_n=A_{n-1}(Y_{n-1}-X_{n-1})+\psi_{n-1}Y_{n-1}+\varphi_n$; $Y_0-X_0=0$. The solution of this Cauchy problem is obtained by iteration as

$$Y_n-X_n=A_{n-1}(A_{n-2}(Y_{n-2}-X_{n-2})+\psi_{n-2}Y_{n-2}+\varphi_{n-1})+\psi_{n-1}Y_{n-1}+\varphi_n.$$

Let us edit by iterating this solution as follows:

$$\begin{aligned}Y_n-X_n &= A_{n-1}A_{n-2}(Y_{n-2}-X_{n-2})+A_{n-1}\psi_{n-2}Y_{n-2}+\psi_{n-1}Y_{n-1}+A_{n-1}\varphi_{n-1})+\varphi_n\\ &= A_{n-1}A_{n-2}(A_{n-3}(Y_{n-3}-X_{n-3})+\psi_{n-3}Y_{n-3}+\varphi_{n-2})+A_{n-1}\psi_{n-2}Y_{n-2}+\psi_{n-1}Y_{n-1}\\ &\quad +A_{n-1}\varphi_{n-1}+\varphi_n\\ &= A_{n-1}A_{n-2}A_{n-3}(Y_{n-3}-X_{n-3})+A_{n-1}A_{n-2}\psi_{n-3}Y_{n-3}+A_{n-1}\psi_{n-2}Y_{n-2}+\psi_{n-1}Y_{n-1}\\ &\quad +A_{n-1}A_{n-2}\varphi_{n-2}+A_{n-1}\varphi_{n-1}+\varphi_n\\ &\ \vdots\\ Y_n-X_n &= A_{n-1}A_{n-2}A_{n-3}...A_0(Y_0-X_0)+A_{n-1}A_{n-2}...A_1\psi_0Y_0+A_{n-1}A_{n-2}...A_2\psi_1Y_1\\ &\quad +...+A_{n-1}\psi_{n-2}Y_{n-2}+\psi_{n-1}Y_{n-1}+A_{n-1}A_{n-2}...A_1\varphi_1\\ &\quad +...+A_{n-1}A_{n-2}\varphi_{n-2}+A_{n-1}\varphi_{n-1}+\varphi_n\\ &= \sum_{k=2}^{n}[(\textstyle\prod_{j=k}^{n-1}A_j)(\psi_{k-1}Y_{k-1}+\varphi_k)]+A_{n-1}A_{n-2}...A_1\psi_0.\end{aligned}$$

So, for $||Y_n-X_n||$ we can write

$$||Y_n-X_n|| \le \sum_{k=2}^{n}[(\textstyle\prod_{j=k}^{n-1}||A_j||)(||\psi_{k-1}||||Y_{k-1}||+||\varphi_k||)]+||A_{n-1}||||A_{n-2}||...||A_1||||\psi_0||.$$

If the upper bound of $||\varphi_n||$ from Theorem 2.1 and the upper bound of $||Y_n||$ from Theorem 2.2 substitute in the last inequality, we have

$$\begin{aligned}||Y_n-X_n|| \le \sum_{k=2}^{n}[(\textstyle\prod_{j=k}^{n-1}||A_j||)\\ &\times(||\psi_{k-1}||(1+u\sqrt{N})^{k-2}[(1+u\sqrt{N})\max_{0\le j\le k-2}||A_j||+vN]^{k-1}\\ &+vN\sum_{j=2}^{k-1}(1+u\sqrt{N})^{k-j-1}[(1+u\sqrt{N})\max_{2\le i\le k-2}||A_i||+vN]^{k-j-1}\\ &+u\sqrt{N}(1+u\sqrt{N})^{k-2}[(1+u\sqrt{N})\max_{0\le j\le k-1}||A_j||+vN]^{k}\\ &+uvN^{3/2}\sum_{j=2}^{k-1}(1+u\sqrt{N})^{k-j-1}[(1+u\sqrt{N})\max_{2\le i\le k-1}||A_i||+vN]^{k-j}+vN)]\end{aligned}$$

$+||A_{n-1}||||A_{n-2}||...||A_1||||\psi_0||.$

Also, if we use (2.1) the following result is achieved:

$||Y_n - X_n|| \leq \sum_{k=2}^{n}[(\prod_{j=k}^{n-1}||A_j||)$
$\times((u\sqrt{N}||A_{k-1}|| + vN)(1 + u\sqrt{N})^{k-2}[(1 + u\sqrt{N})\max_{0\leq j\leq k-2}||A_j|| + vN]^{k-1}$
$+vN\sum_{j=2}^{k-1}(1 + u\sqrt{N})^{k-j-1}[(1 + u\sqrt{N})\max_{2\leq i\leq k-2}||A_i|| + vN]^{k-j-1}$
$+u\sqrt{N}(1 + u\sqrt{N})^{k-2}[(1 + u\sqrt{N})\max_{0\leq j\leq k-1}||A_j|| + vN]^{k}$
$+uvN^{3/2}\sum_{j=2}^{k-1}(1 + u\sqrt{N})^{k-j-1}[(1 + u\sqrt{N})\max_{2\leq i\leq k-1}||A_i|| + vN]^{k-j} + vN)]$
$+||A_{n-1}||||A_{n-2}||...||A_1||(u\sqrt{N}||A_0|| + vN).$

□

3. Applying The Results To Schur Stability of Periodic Systems

The results on Schur stability of linear difference Cauchy problem with periodic coefficients were given in [1] and [2] for Wilkinson and Godunov Models. In this section, these results will be applied in extended format and the changes due to differences in models will be examined in the computations. Let

$$Y_{n+1} = (A_n + B_n)Y_n, n \in \mathbb{Z}, \tag{3.1}$$

where $A_n = A_{n+T}$ and $B_n = B_{n+T}$, N- dimensional periodic (T- period). It is called perturbed system of the system (1.2). Continuity theorem on the monodromy matrix in [15] guarantees Schur stability of the system (3.1) when the system (1.2) or matrix X_T is Schur stable ([19, 6, 13]). The following theorem which is application of continuity theorem can easily be obtained as same to Theorem in [1]. For $T = 1$, the system (1.2) transforms the system

$$x_{n+1} = Ax_n, n \in \mathbb{Z},$$

and it is called linear difference equation system with constant coefficients. Therefore, $\omega_1(A, T)$ can be written

$$\omega_1(A, T) = \omega(A),\ \omega(A) = ||\sum_{k=0}^{\infty}(A^*)^k A^k||.$$

Furthermore, in this case $\omega_1(A, 1)$ is equal to $\omega(X_1) = \omega(A)$ ([9, 10]).

Theorem 3.1. *If the matrix Y_T is Schur stable and the inequality*

$$||Y_T - X_T|| \leq \sqrt{||Y_T||^2 + \frac{1}{\omega(Y_T)}} - ||Y_T|| \tag{3.2}$$

holds, then the matrix X_T is Schur stable, where the matrix Y_T is computed monodromy matrix of X_T and the matrix X_T is perturbed matrix of Y_T.

We can obtain following corollary by $n = T$ in Theorem 2.3 and Theorem 2.4.

Corollary 3.1. *The following inequalities hold, where the matrix X_T is the monodromy matrix of the system (1.2) and the matrix Y_T is the computed matrix of the monodromy matrix X_T.*

$$||Y_T - X_T|| \leq \sum_{k=1}^{T} \binom{T}{k} (\max_{0\leq j\leq T-1} ||A_j||)^{T-k} (u\sqrt{N} \max_{2\leq i\leq k-2} ||A_i|| + vN)^k$$
$$+ \sum_{k=2}^{T} [\prod_{j=k}^{T-1} ||A_j|| + \sum_{r=1}^{T-k} \binom{T-k}{r} (\max_{0\leq j\leq T-1} ||A_j||)^{T-k-r} (u\sqrt{N} \max_{0\leq j\leq T-1} ||A_j|| + vN)^r]$$
$$\times [u\sqrt{N}(1+u\sqrt{N})^{k-2}((1+u\sqrt{N}) \max_{0\leq j\leq k-1} ||A_j|| + vN)^k +$$
$$uvN^{3/2} \sum_{j=2}^{k-1} (1+u\sqrt{N})^{k-j-1}((1+u\sqrt{N}) \max_{0\leq j\leq T-1} ||A_j|| + vN)^{k-j} + vN] \tag{3.3}$$

$$||Y_T - X_T|| \leq \sum_{k=2}^{T} [(\prod_{j=k}^{T-1} ||A_j||)((u\sqrt{N}||A_{k-1}|| + vN)$$
$$\times (1+u\sqrt{N})^{k-2}[(1+u\sqrt{N}) \max_{0\leq j\leq k-2} ||A_j|| + vN]^{k-1}$$
$$+ vN \sum_{j=2}^{k-1} (1+u\sqrt{N})^{k-j-1}[(1+u\sqrt{N}) \max_{2\leq i\leq k-2} ||A_i|| + vN]^{k-j-1}$$
$$+ u\sqrt{N}(1+u\sqrt{N})^{k-2}[(1+u\sqrt{N}) \max_{0\leq j\leq k-1} ||A_j|| + vN]^k$$
$$+ uvN^{3/2} \sum_{j=2}^{k-1} (1+u\sqrt{N})^{k-j-1}[(1+u\sqrt{N}) \max_{2\leq i\leq k-1} ||A_i|| + vN]^{k-j} + vN)]$$
$$+ ||A_{T-1}|| ||A_{T-2}|| ... ||A_1|| (u\sqrt{N}||A_0|| + vN) \tag{3.4}$$

The Corollary 3.1 guarantees Schur stability of the system(1.2) (or monodromy matrix X_T) when the *computed matrix* Y_T is Schur stable.

Corollary 3.2. *Let X_T be monodromy matrix of the system (1.2) and Y_T be the computed matrix of the matrix X_T on floating point arithmetic. If the computed matrix Y_T is Schur stable and the inequality*

$\Delta_i < \Delta_s$, $i = 1, 2$

holds then monodromy X_T is Schur stable, where

- Δ_1 *is the upper bound of (3.2),*
- Δ_2 *is the upper bound of (3.3),*
- $\Delta_s = \sqrt{||Y_T||^2 + \frac{1}{\omega(Y_T)}} - ||Y_T||$.

Corollary 3.3. *If $A_n \in M_N(F)$ then it is clear that Δ_i is same as in [2].*

4. Numerical Examples

Numerical examples have been computed using the MVC (Matrix Vector Calculator). The function QdaStab [4] has been used to compute the value $\omega(A)$ of matrix A.

Example 1. Let $\mathbb{F}^E = \mathbb{F}^E(10, -3, 3, 3)$ and matrices

$$A_0 = \begin{pmatrix} 0,9995 & 0,01 \\ 0,0005 & 0,0002 \end{pmatrix} \text{ and } A_1 \begin{pmatrix} 0,9987 & 0,02 \\ 0,0001 & -0,001 \end{pmatrix}$$

where A_0, $A_1 \notin M_2(\mathbb{F}^E)$. Let us investigate Schur stability, where $T = 2$. Monodromy matrix X_2 of the system (1.1) has been computed with

$$X_2 = \begin{pmatrix} 0,99921 & 0,010001 \\ 9,945.10^{-5} & 8.10^{-7} \end{pmatrix}.$$

And the monodromy matrix X_2 is Schur stable, since $\omega(X_2) = 634,143$. If the matrix Y_2 is computed matrix in $\mathbb{F}^E$, the matrices

$$Y_2^r = \begin{pmatrix} 1 & 0,01 \\ 0 & 0 \end{pmatrix}, Y_2^c = \begin{pmatrix} 0,997 & 0,00998 \\ 0 & 0 \end{pmatrix}$$

are obtained. $\omega(Y_2^r) = \infty$, $\omega(Y_2^c) = 166,94$ and so, computed matrix Y_2 is Schur stable by chopping, but is not Schur stable by rounding.

Example 2. Let $\mathbb{F}^E = \mathbb{F}^E(10, -3, 3, 3)$ and matrices

$$A_0 = \begin{pmatrix} 0.002987 & 0.1029 & 0.0675 \\ 0.05416 & 0.01246 & 0.001856 \\ 0.002137 & 0.06438 & 0.002454 \end{pmatrix}, A_1 = \begin{pmatrix} 0.005467 & 0.06435 & 0.0418 \\ 0.003464 & 0.004829 & 0.001463 \\ 0.01248 & 0.04352 & 0.0002547 \end{pmatrix}$$

$$\text{and } A_2 = \begin{pmatrix} 0.02149 & 0.01298 & 0.03145 \\ 0.002367 & 0.03564 & 0.004758 \\ 0.0022657 & 0.004321 & 0.0001468 \end{pmatrix}$$

where A_0, A_1, $A_2 \notin M_2(\mathbb{F}^E)$. Let us investigate Schur stability, where $T = 3$. The matrices

$$Y_3^r = \begin{pmatrix} 0.000156 & 0.00015 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, Y_3^c = \begin{pmatrix} 0.000155 & 0.000149 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

are computed matrices in $\mathbb{F}^E$. So, the values

$$\omega(Y_3^r) = 1,\ \Delta_s^r = 0.9997836067,\ \Delta_1^r = 0.9875218373 * 10^{-4}\ ,\ \Delta_2^r = 0.506766165 * 10^{-4}$$
$$\omega(Y_3^c) = 1,\ \Delta_s^c = 0.9997850207,\ \Delta_1^c = 0.2009159367 * 10^{-3}\ ,\ \Delta_2^c = 0.1036855603 * 10^{-3}$$

are obtained. It seems that $\Delta_i^r < \Delta_s^r$ and $\Delta_i^c < \Delta_s^c$ for $i = 1, 2$. Therefore, in both cases, Corollary 3.2guarentees Schur stability of the monodromy matrix X_3 in $\mathbb{F}^E(10, -3, 3, 3)$.

References

[1] A.O. Çıbıkdiken, K. Aydın, *Computation of the monodromy matrix in floating point arithmetic with the Wilkinson Model*, Comput Math Appl, 67, Issue 5 (2014), 1186-1194.

[2] A.O. Çıbıkdiken, K. Aydın, *Computation of the monodromy matrix on floating point arithmetic with Godunov Model*, Konuralp Journal of Mathematics, 4, Issue 1 (2016), 23-32.

[3] G. Bohlender, *Floating-point computation of functions with maximum accuracy*, IEEE Trans. Comput., C-26 (7) (1977) 621-632.

[4] H. Bulgak, D. Eminov, *Computer dialogue system MVC*, Selçuk J. Appl. Math. 2(2) (2001) 17-38.

[5] J.H. Wilkinson, Error analysis of floating-point computation, Numerische Mathematik 2, (1960) 319-340.

[6] J. Rohn, *Positive definiteness and stability of interval matrices*, SIAM J. Matrix Anal. Appl., 15(1) (1994) 175-184.

[7] J. Von Neumann, H. Goldstine, Numerical inverting of matrices of high order, Bull. Amer. Math. Soc., 53 (1947) 1021-1099.

[8] K. Aydın, *The Condition Number for the Asymptotic Stability of the Periodic Ordinary Differential Systems*, Ph.D. Thesis, Selçuk University Graduate Natural and Applied Sciences, Konya, 1996 (Turkish).

[9] K. Aydın, H. Bulgak, G.V. Demidenko, *Numeric characteristics for asymptotic stability of solutions to linear difference equations with periodic coefficients*, Sib. Math. J. 41(6) (2000) 1005-1014.

[10] K. Aydın, H. Bulgak, G.V. Demidenko, *Continuity of numeric characteristics for asymptotic stability of solutions to linear difference equations with periodic coefficients*, Selçuk J. of Appl. Math. 2(2) (2001) 5-10.

[11] K. Aydın, H. Bulgak, G.V. Demidenko, *Asymptotic stability of solutions to perturbed linear difference equations with periodic coefficients*, Sib. Math. J. 43(3)(2002) 389-401.

[12] L.F. Shampine, Jr.R.C. Allen, S. Pruess, Fundamentals of numerical computing, John Wiley and Sons Inc., 1997.

[13] M. Voicu, O. Pastravanu, *Generalized matrix diagonal stability and linear dynamical systems*, Linear Algebra Appl. 419(2006) 299-310.

[14] Ö. Akın, H. Bulgak, *Linear Difference Equations and Stability Theory*, Selçuk University Research Centre of Applied Mathematics, No.2, Konya, 1998 (Turkish).

[15] R.P. Agarwal, *Difference Equations and Inequalities*, 2nd ed., Pure and Applied Mathematics Series, Vol. 228, National University of Singapore, 2000.

[16] S.K. Godunov, *The solution of systems of linear equations*, Nauka, Moscow, 1980 (in Russian).

[17] S.K. Godunov, A.G. Antonov, O.P. Kiriluk, V.I. Kostin, *Guaranteed Accuracy in Mathematical Computations*, Englewood Cliffs, N. J., Prentice- Hall, 1993.

[18] S.K. Godunov, *Modern aspects of linear algebra*, Translations of Mathematical Monographs, 175, Providence, RI : American Mathematical Society, 1998.

[19] S.N. Elaydi, *An Introduction to Difference Equations*, Springer- Verlag, New York, 1996.

[20] U.W. Kulisch, *Mathematical foundation of computer arithmetic*, IEEE Trans. Comput., C-26(7) (1977) 610-621.

[21] U.W. Kulisch, W.L. Miranker, *Computer Arithmetic in Theory and Practice*, Academic Press Inc., 1981.

[22] U.W. Kulisch, *Rounding Near Zero*, Electrotechnical Laboratory, Agency of Industrial Science and Technology, MITI, Tsukuba, Japan, 1999.

[23] W. Kahan, Lecture notes on the status of IEEE standart 754 for binary floating point arithmetic, University of California, 1997.

A. O. Chibikdiken
Department of Computer Engineering, KTO Karatay University, Konya, Türkiye
E-mail: ali.osman.cibikdiken@karatay.edu.tr

G. Ch. Kizilkan
Department of Mathematics and Computer Sciences, Faculty of Sciences, Necmettin Erbakan University, Konya, Türkiye
*E-mail:*gckizilkan@erbakan.edu.tr

A. Duman
Department of Mathematics and Computer Sciences, Faculty of Sciences, Necmettin Erbakan University, Konya, Türkiye
*E-mail:*aduman@erbakan.edu.tr

K. Aydin
Department of Mathematics, Faculty of Sciences, Selçuk University, Konya, Türkiye
*E-mail:*kaydin@selcuk.edu.tr

Received 29 October 2020
Accepted 14 November 2020