Numerical Methods for Spectrum Computation of Monodromy Operators via Non-Causal Hold Discretization

Kentaro Hirata*, Tomomichi Hagiwara**, and Atsushi Itokazu*

Abstract: In this paper, numerical methods for the computation of the spectrum of the monodromy operator are investigated. This operator arises in a representation of time-delay systems from the discrete-time viewpoint and thus its spectrum computation is connected to their stability analysis directly. Theoretically, the proposed methods inherit the mathematical justification of the finite-dimensional approximation via the sample and hold discretization where the approximation is regarded (and justified) as a perturbation to the monodromy operator. The first key idea for the current extension is to relax the requirement on the causality of the hold operator, inspired by the fast-lifting approach. The second idea is to employ higher order holds based on polynomial interpolations. By combining these two, this paper derives efficient numerical methods in which the reduction to an eigenvalue problem is guaranteed to be completely rigorous.

Key Words: time-delay systems, approximation, non-casual hold, interpolation.

1. Introduction

Time-delay systems have been an important research field in control theory due to the practical significance as well as the mathematical depth. Because of its infinite-dimensional nature, the stability analysis is not straightforward as in the case of Finite-Dimensional Linear Time-Invariant (FDLTI) systems. Energy-based methods, manipulating a difference-differential equation or Delay Differential Equation (DDE) directly, often lead to individual sufficient conditions, e.g., various LMI conditions [1], Sum Of Squares (SOS) criterion [2] or Discretized Lyapunov Functional method (DLF) [3]. Their variations arise from the way how to relax the existence condition of Lyapunov-Krasovskii functional of complete-type to give finite-dimensional criteria.

On the other hand, the state space approach (e.g. [4]) based on the infinite-dimensional systems theory allows the same formalism as the FDLTI case. In this framework, a DDE is converted into an abstract differential (state) equation. The infinite-dimensional counterpart of A-matrix is given by the infinitesimal generator of a strongly continuous semigroup which describes the state transition on the corresponding function space. The spectrum of A-operator determines the stability precisely.

For retarded systems with a single delay, another infinite-dimensional representation can be derived by dividing the trajectory into pieces according to the delay length [5]. This technique called lifting is a powerful tool used in the sampled-data control literature, e.g., [6]. Such a representation was originally introduced in [7] by the first author to analyze the stability of a kind of hybrid dynamical systems, the passive dynamic walking, under the delayed feedback control [8],[9]. It immediately covers a representation of delay differential systems with the monodromy operator if we discard the discontinuous state transitions specific to such a hybrid dynamics. A potential merit of this expression is that one can circumvent the treatment of cumbersome unbounded differential operators as in [4].

A mathematical justification of a numerical procedure to compute the spectrum of the monodromy operator via the fast-sample/fast-hold (FSFH) approximation is given in [10],[11] on the basis of perturbation theory [12]. In [13],[14], stability of delay systems is investigated via operator inequalities in terms of this representation. The modified fast-sample/fast-hold (mFSFH) approximation developed in [15] is applied to the computation of the spectrum of the monodromy operator in [16]. It is shown that this approximation is effective in the numerical computation. This paper investigates the finite-dimensional approximation problem of the monodromy operator further from an operator-theoretic viewpoint and improves the numerical precision by using non-casual holds.

The monodromy operator is related to the solution operator treated in [17] and others. In that literature, the numerical computation methods for ODEs are applied to the stability analysis of DDEs. For example, [17] uses the algorithm in [18] consisting of an eigenvalue computation of a discretized solution operator and a local root finding of the characteristic equation with Newton’s method. The discretization is based on the Linear Multi-Step (LMS) method [19] with Lagrange interpolation to deal with the delayed term. In [20], another Runge-Kutta based approach is also investigated. A clear contrast between the viewpoints of this paper and the existing literature above is that the present discretization is brought by an explicit perturbation against a concrete integral operator (with initial values), while the discretized solution operator is introduced in an implicit manner as a consequence of numerical computations of trajectories. Although the ODE based method may be applicable for wider classes such as nonlinear systems, the present approach specific to the linear case brings sharp convergence analysis without involving the characteristic equation. The method like [21] locates in-between, i.e., convolution with piecewise constant function and delay approximation with Lagrange in-
terpolation are used. However, the emphasis of [21] is on num-
merical issues and no convergence proof is provided.

The following notations are used in the sequel. Given $h > 0$, $K_1$ denotes the function space $L_2(0, h)$. Define $K_2$ by

$$K_2 = \{ f \in C[0, h] \mid f \text{ has a left-hand limit at } h \}.$$  

Then $f \in K_2$ admits a unique extension to $\tilde{f}$ defined on $[0, h]$ such that $\tilde{f} \in C([0, h])$. In a similar manner, we define $K_3$ by

$$K_3 = \{ f \in C^1([0, h]) \mid f \text{ admits a unique extension to } \tilde{f} \text{ defined on } [0, h] \text{ such that } \tilde{f} \in C^1([0, h]) \}.$$  

The Kronecker product of two matrices $A$ and $B$ is expressed by $A \otimes B$. We use the same notation even when $B$ is an operator. The spectrum of the operator $F$ is denoted by $\sigma(F)$. For $x \in \mathbb{R}$, let $\lceil x \rceil$ denote the ceiling function of $x$.

### 2. Problem Formulation

This paper investigates the stability of delay differential equation (DDE)

$$\dot{x}(t) = Ax(t) + Gx(t-h), \quad h > 0$$  

with $A \in \mathbb{R}^{a \times n}$ and $G \in \mathbb{R}^{a \times n}$, under initial conditions

$$x(0) = x_0 \in \mathbb{R}^n, \quad x(\theta - h) = \varphi(\theta), \quad \theta \in [0, h).$$

First we consider the case $\varphi \in K_1^n$. Let us introduce the matrix factorization of $G$ such as $G = BC$ with some $B \in \mathbb{R}^{a \times \mu}$, $C \in \mathbb{R}^{\mu \times n}$. With this factorization, define the FDLTI system $\Sigma_F$ by

$$\Sigma_F: \dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t), \quad x \in \mathbb{R}^n, y, u \in \mathbb{R}^n.$$  

Let $\Sigma_d$ denote the pure delay

$$\Sigma_d: u(t) = y(t-h).$$

Then the dynamical behavior of (1) over $t \geq 0$ is equivalently represented by the feedback connection of $\Sigma_F$ and $\Sigma_d$ under the initial conditions $x(0) = x_0 \in \mathbb{R}^n$, $y(\theta - h) = C\varphi(\theta) = \psi(\theta) \in K_1^n$, $\theta \in [0, h]$. We denote this closed-loop system shown in Fig. 1 by $\Sigma$.

![Fig. 1 Feedback configuration.](image)

The continuous-time signal $u(t)$, $t \in [0, \infty)$ can be expressed equivalently in a lifted form, i.e., by the sequence of signals with the duration $h$ as $\{ \tilde{u}_i \}_{i=0}^{\infty}$, where $\tilde{u}_i(\theta) = u(ih+\theta), \theta \in [0, h)$. Now let us define the state space $Z_1 := \mathbb{R}^n \oplus K_1^n$ for $i = 1, 2, 3$. Let $J_{12}$ denote the embedding operator from $K_2^n$ to $K_1^n$. With the notations $\mathcal{J}(\cdot) := \text{diag}[J_{12}, \ldots, (\cdot)]$ and $x_i := x(\theta_i)$, the state transition of $\Sigma$ is represented by

$$\begin{bmatrix} x_{i+1} \\ \tilde{u}_{i+1} \end{bmatrix} = \mathcal{F} \begin{bmatrix} x_i \\ \tilde{u}_i \end{bmatrix},$$

where

$$\mathcal{F} = \mathcal{I}(J_{12}) : Z_1 \rightarrow Z_1$$

is the monodromy operator and $\mathcal{F} : Z_1 \rightarrow Z_2$ is defined as

$$\mathcal{F} = \begin{bmatrix} \mathcal{F}_{11} & \mathcal{F}_{12} \\ \mathcal{F}_{21} & \mathcal{F}_{22} \end{bmatrix}$$

with

$$\mathcal{F}_{11} = e^{Ah} := A_h,$$

$$\mathcal{F}_{12} = \int_0^h e^{A(\theta-t)} \tilde{B}(\tau) \, d\tau,$$

$$\mathcal{F}_{21}(x)(\theta) = Ce^{Ah}f,$$

$$\mathcal{F}_{22}(x)(\theta) = \int_0^\theta e^{A(\theta-\tau)} \tilde{B}(\tau) \, d\tau.$$  

An important fact is that $\mathcal{F}$ is compact on $Z_1$ [10],[11]. As established in [13], the exponential stability of the time-delay system $\Sigma$ is characterized by the spectral radius of $\mathcal{F}$. Thus, in the following, we consider how to compute the spectrum of $\mathcal{F}$ numerically by way of finite-rank approximations and the eigenvalue problem of corresponding matrices.

### 3. Previous Works on Spectrum Computation

Before mentioning the main result of this paper, the previous works on the computation of the spectrum of $\mathcal{F}$ are summarized. From the structural difference between their matrix representations, we obtain an idea leading to new approximation schemes.

Let us denote the outside of the open disc centered at the origin with radius $\gamma > 0$ by $D_\gamma$. Since $\mathcal{F}$ is compact, $\sigma(\mathcal{F}) \cap D_\gamma$ constitutes a finite system of eigenvalues [12]. We denote it by $\sigma'(\mathcal{F})$.

#### 3.1 FSFH Approximation

In [10],[11], $\mathcal{F}$ is discretized via the FSFH (fast-sample and fast-hold) approximation of the functional part of the output of $\mathcal{F}$. The interval $[0, h)$ is subdivided into $N \in \mathbb{N}$ pieces and then each portion of the signal is approximated by a constant. Let $h' := h/N$, $\theta_k := kh'$. With the sampling operator $S: K_2^n \rightarrow \mathbb{R}^{\mu N}$ and the zero-th order hold operator $\mathcal{H}_0: \mathbb{R}^{\mu N} \rightarrow K_1^n$ defined by

$$Sf(\cdot) = \begin{bmatrix} f(\theta_0) \\ f(\theta_1) \\ \vdots \\ f(\theta_{N-1}) \end{bmatrix}, \quad \mathcal{H}_0: \begin{bmatrix} u_0 \\ u_2 \\ \vdots \\ u_{N-1} \end{bmatrix} \mapsto u(\theta),$$

$u(\theta) = u_k, \theta \in I_k := [\theta_k, \theta_{k+1})$, respectively, $\mathcal{F}$ is approximated by

$$\tilde{\mathcal{F}}_0 := \mathcal{I}(\mathcal{H}_0) \mathcal{I}(S) \tilde{\mathcal{F}}.$$  

 regard the difference between $\tilde{\mathcal{F}}_0$ and $\mathcal{F}$ as a perturbation to $\mathcal{F}$, i.e.,

$$\Delta_0 := \tilde{\mathcal{F}}_0 - \mathcal{F} = (\mathcal{I}(\mathcal{H}_0) \mathcal{I}(S) - \mathcal{I}(J_{12})) \tilde{\mathcal{F}}.$$  

Then it can be shown that $\Delta_0$ converges to zero in the general-
ized sense as $N \to \infty$. Since $F$ is a linear operator on $\mathcal{L}_2$, it is also closed. Thus the continuity of $\sigma'(F)$ ([12]) implies that the deviation of non-zero eigenvalues of $\hat{F}_0$ from those of $F$ is small for large $N$.

To be precise, their closeness is described as follows. For two unordered $N$-tuples of complex numbers $\hat{\mathfrak{z}} = \{\lambda_1, \ldots, \lambda_N\}$ and $\mathfrak{z} = \{\lambda_1, \ldots, \lambda_N\}$, the distance between them is defined by
\[
\text{dist}(\hat{\mathfrak{z}}, \mathfrak{z}) = \min \max_i |\lambda_i - \hat{\lambda}_i|
\]
where the min is taken over all possible renumberings of the elements of one of the $N$-tuples. Let the total multiplicity of $\sigma'(F)$ be $m$ and introduce the unordered $m$-tuple $\mathfrak{z}'(F)$ of repeated eigenvalues to represent finite systems of eigenvalues $\sigma'(F)$ and $\sigma'(\hat{F}_0)$. Then the distance between $\mathfrak{z}'(F)$ and $\mathfrak{z}'(\hat{F}_0)$ converges to zero as $N \to \infty$. This formulates the continuity of the repeated eigenvalues as a whole (see II-5.2 and IV-3.5 of [12]).

Since non-zero eigenvalues of $\hat{F}_0$ coincide with those of the matrix $F_0: \mathbb{R}^{n \times p} \to \mathbb{R}^{n \times p}$ defined as
\[
F_0 := \mathcal{I}(S)\mathcal{F}I(\mathcal{H}_0),
\]
the computation of the spectrum of $F$ is reduced to the eigenvalue problem of a matrix.

For an operator (or matrix) $G$ and $\nu \in \mathbb{N}$, let us denote the Kronecker product $I_{\nu \times \nu} \otimes G$ by $(\mathcal{G})$. When the block size $\nu$ is obvious from the context, we simply write $\mathcal{G}$. Then, an explicit formula of $F_0$ is given as follows:
\[
F_0 = \mathcal{I}(\mathcal{G})\begin{bmatrix}
A_0 & F_{12} & 0 \\
F_{21} & F_{22} & X
\end{bmatrix} \mathcal{I}(\mathcal{M}_0),
\] (6)
\[
A_0 = e^{\mathcal{A}0}, \quad F_{12} = \begin{bmatrix}(\mathcal{A}_0)^{N-1} & \cdots & I \end{bmatrix},
\]
\[
F_{21} = \begin{bmatrix}
I \\
A_0 \\
\vdots \\
(\mathcal{A}_0)^{N-1}
\end{bmatrix}, \quad F_{22} = \begin{bmatrix}
I & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & I & 0
\end{bmatrix},
\]
\[
M_0 = \int_0^{\nu} e^{\mathcal{A}(\nu - \tau)} B \, d\tau.
\] (7)
See [10],[11] for details.

### 3.2 Fast-Lifting Approach

Another approximation is proposed in [16] to improve the slow convergence of the FSFH approach. As in the previous case, we subdivide the interval $[0, h)$, but this time, a series of functions defined on these smaller intervals is considered. Define $K'_1 := L_2[0, h')$ and $Z'_1 := \mathbb{R}^n \oplus (K'_1)^{\mu N}$ according to the subdivision. By using the fast-lifting operator $\mathcal{L}: K'_1 \to (K'_1)^{\mu N}$,
\[
(\mathcal{L}\hat{u})(\theta) = \begin{bmatrix}
\hat{u}_0(\theta_0 + \theta') \\
\hat{u}_0(\theta_1 + \theta') \\
\vdots \\
\hat{u}_0(\theta_{N-1} + \theta')
\end{bmatrix},
\] (8)
\[
\theta_2 = kh', \theta' \in [0, h'),
\]
define $F_{12}: Z'_1 \to Z'_1$ by
\[
F_{12} = \mathcal{I}(\mathcal{L})\mathcal{F}I(\mathcal{L})^{-1}.
\] (9)
Let $F_{21}, F_{22}$, and $F_{12}$ be the operators obtained from (3)-(5) by replacing $h$ with $h'$. Then a concrete expression of $F_{12}$ is given by
\[
F_{12} = \begin{bmatrix}
A_0 & F_{12} & 0 \\
F_{21} & F_{22} & X
\end{bmatrix} \mathcal{I}(\mathcal{M}_0),
\] (10)
Note that the spectrum is unchanged under the similarity transformation (9). In view of the structure of (10), it is convenient to approximate $F_{22}$ by the operator of the form $F_{22}X$ where $X \in \mathbb{R}^{n \times n}$ is a matrix. The problem how to choose $X$ to minimize the approximation error $\epsilon := F_{22} - F_{22}X$ in the sense of the Hilbert-Schmidt norm was first considered in [22] in the context of the sampled-data $H_{\infty}$ control. Later, it is extended to the case of operator norm in [15],[23]. Substitution of $F_{22} = F_{22}X + \epsilon$ into (10) yields
\[
F_{12} = \hat{F}_{12} + \hat{F}_{12}^X + \epsilon
\]
with
\[
\hat{F}_{12} = \int_0^{\nu} e^{\mathcal{A}(\nu - \tau)} B C e^{\mathcal{A} \tau} \, d\tau.
\] (11)
Since $F_{12}$ is compact and we can ensure $\|\epsilon\| \to 0$ as $N \to \infty$, the continuity of $\sigma'(F_{12})$ justifies the spectrum computation via $F_{12}$. By changing the order of the operators in (11), a matrix expression
\[
F_{12} = \begin{bmatrix}
A_0 & F_{12} & 0 \\
F_{21} & F_{22} + X
\end{bmatrix} \mathcal{I}(\mathcal{M}_0),
\] (12)
\[
\hat{Q}' = F_{12}X = \int_0^{\nu} e^{\mathcal{A}(\nu - \tau)} B C e^{\mathcal{A} \tau} \, d\tau
\]
is derived. This approach to computing $\sigma(F)$ is called the modified FSFH approximation.

### 3.3 Observation

As shown in [16], the modified FSFH approximation gives better numerical results than FSFH. However, we need additional computation to determine $X$. Thus it might be convenient if we can achieve the same level of accuracy by simple formulas like FSFH. By comparing (6) with (12), one can see a major structural difference in the (2,2)-block. While the (2,2)-block of $F_0$ is strictly lower triangular, the corresponding block of $F_{12}$ has non-zero diagonal blocks due to $X$. This difference between the resulting matrices of the discretization process is closely related to how we deal with the operator $F_{12}$. This is clear because the matrix $X$ in $F_{12}$ has arisen due to the approximation of $F_{12}$. In this respect, we see that while the original operator $F_{12}$ is strictly causal in continuous-time, its approximation given by $F_{12}X$ is non-causal because of the integral action in $F_{12}$. Roughly speaking, approximating $F_{12}$ corresponds to approximating its outputs with some amenable functions, and such treatment may be considered as an off-line approximation problem, in contrast to real-time problems in signal processing or digital feedback control. We can interpret that non-causal treatment of continuous-time signals, which provides more freedom, has led us to better approximation under the former context. This motivates us to develop the further non-causal approximation schemes based on non-causal holds discussed in the following sections.
4. Approximation with Non-Causal First Order Hold

Now we employ the non-causal first order hold (FOH) as the simplest case of non-causal hold operators. When deciding the slope of the hold function during the interval starting from the current time index \( k \), the non-causal FOH uses the future information, say, the value at the time index \( k+1 \). By introducing the augmented sampler \( S_a \) (defined below) followed by this hold, we approximate the functional part of the output signal of \( \hat{F} \) by a piecewise linear function. In the sequel, this method is referred to as the fast-sample/fast-FOH (FSFFOH) approximation.

4.1 Matrix Representation

Define the augmented sampling operator \( S_a : K^m_2 \to R^{m(N+1)} \) and the non-causal first order hold operator \( H_1 : R^{m(N+1)} \to K^m_2 \) as

\[
S_a f(\cdot) = \left[ S f(\cdot) \right]_{f(\theta_N - 0)} , \quad H_1 : \begin{bmatrix} u_0 \\ \vdots \\ u_N \end{bmatrix} \mapsto \begin{bmatrix} u(\theta) \\ \vdots \\ u_N \end{bmatrix},
\]

where

\[
u(\theta) = (u_{k+1} - u_k)\theta/h' + (k+1)u_k - ku_{k+1},
\]

when \( \theta \in I_k \). \( (13) \)

Then \( F \) is approximated by

\[
\hat{F}_1 = I(\hat{H}_1)I(S_a)F.
\]

As before, a change of the order yields a matrix representation \( F_1 = I(S_a)F I(\hat{H}_1) \) with the size \( n + \mu(N + 1) \). Describe the input-output relation of \( F_1 \) by

\[
\begin{bmatrix} \tilde{x}_1 \\ \tilde{y} \end{bmatrix} = F_1 \begin{bmatrix} x_0 \\ \tilde{u} \end{bmatrix}, \quad \begin{bmatrix} \tilde{y} \\ \tilde{u} \end{bmatrix} = \begin{bmatrix} u_0 \\ \vdots \\ u_N \end{bmatrix}.
\]

Then, from (13), the outputs in (14) are given by

\[
\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_n \end{bmatrix} = A_d \begin{bmatrix} x_0 + \sum_{\ell=0}^{N-1} (A'_\ell - I^{-1})P_\ell \xi_\ell \\ \vdots \\ x_0 + \sum_{\ell=0}^{k-1} (A'_\ell - I^{-1})P_\ell \xi_\ell \end{bmatrix}, \quad k = 0, \ldots, N,
\]

with

\[
\xi_\ell = \begin{bmatrix} u_\ell \\ u_{\ell+1} \end{bmatrix}, \quad \tilde{P} = \begin{bmatrix} 0 & P_1 \\ P_0 & 1 \end{bmatrix} = \begin{bmatrix} M_0 & M_1 \end{bmatrix} U,
\]

\[
U = \begin{bmatrix} I & 0 \\ -h^{-1}I & I \end{bmatrix},
\]

where \( M_0 \) and \( M_1 \) are given by (7) and

\[
M_1 = \int_0^{\theta} \tau e^{\theta h - \tau} B d\tau,
\]

respectively. Define

\[
F_{12} = \begin{bmatrix} F_{12} \\ A_d \end{bmatrix}, \quad F_{21} = \begin{bmatrix} F_{21} \\ A_d \end{bmatrix}.
\]

Then \( F_1 \) is given by

\[
F_1 = I(\hat{C}) \begin{bmatrix} A_d \\ F_{21} \\ F_{22} \end{bmatrix} R(P_0, P_1).
\]

4.2 Proof of Convergence

As in the previous case, one can justify the computation of the spectrum of \( F \) via the eigenvalues of \( F_1 \) based on the following lemma.

Lemma 1 The perturbation

\[
\Delta_1 := (I(\hat{H}_1)I(S_a) - I(J_{12}))\hat{F} = F_1 - F
\]

converges to zero as \( N \to \infty \) in the generalized sense.

Proof We use the following standard norm for \( z = \begin{bmatrix} x \\ \tilde{u} \end{bmatrix} \):

\[
||z||_z = \left( ||x||^2 + \int_0^1 |\tilde{u}(\tau)|^2 d\tau \right)^{1/2}.
\]

For linear operators, the convergence in the generalized sense is equivalent to that in the operator norm sense \( [12] \). Since

\[
\Delta_1 z = 0, \quad \forall z \in Z_1,
\]

it suffices to consider the approximation error caused by the sample and FOH for the family of functions

\[
\Phi = \begin{bmatrix} f \\ 0 \\ 1 \end{bmatrix} \hat{F}, \quad z \in U_1,
\]

\[
U_1 := \{ z \mid z \in Z_1, ||z||_z = 1 \},
\]

to evaluate the operator norm

\[
||\Delta_1||_z = \sup_{z \in U_1} ||\Delta_1 z||_z.
\]

For simplicity, we consider the case \( \mu = 1 \) here but the extension to the case \( \mu > 1 \) is straightforward. Let us take \( f \in \Phi \). Since \( f \in K_3 \), we can regard \( f \) to be continuous on \([0, h]\) and it makes no difference on the error bound. Let us denote the line segment connecting \((\theta_k, f(\theta_k))\) and \((\theta_{k+1}, f(\theta_{k+1}))\) by \( \ell_k(\theta) \). The function consisting of these segments is precisely the FSFFOH approximation of \( f \). We denote it by \( \hat{f} \), or simply we can write

\[
\hat{f} = H_1 S_a f.
\]

Let \( \hat{I}_k = [\theta_k, \theta_{k+1}] \) and define \( e_1(k) \) by

\[
e_1(k) := \max_{\theta \in \hat{I}_k} |f(\theta) - \ell_k(\theta)|, \quad k = 0, \ldots, N - 1.
\]

Obviouly,

\[
e_k(\theta) \leq \ell_k(\theta) \leq \hat{f}, \quad \forall \theta \in \hat{I}_k,
\]

(20)
where
\[ f_k := \min\{f(\theta_k), f(\theta_{k+1})\}, \quad \overline{f}_k := \max\{f(\theta_k), f(\theta_{k+1})\}. \]

Let
\[ e(\kappa) = \max_{\theta \in \hat{I}_k} |f(\theta) - f_k|, \quad \overline{e}(\kappa) = \max_{\theta \in \hat{I}_k} |f(\theta) - \overline{f}_k|. \quad (21) \]

Suppose that the maximum in (19) is attained at \( \theta^* \in \hat{I}_k \). When \( f(\theta^*) \geq \ell_1(\theta^*) \), we have
\[ e(\kappa) = f(\theta^*) - \ell_1(\theta^*) \leq f(\theta^*) - f_k \leq e(\kappa) \]
from (20) and (21). On the other hand, when \( f(\theta^*) < \ell_1(\theta^*) \), a similar evaluation yields \( e(\kappa) \leq \overline{e}(\kappa) \). Thus the following inequality holds for each \( k \):
\[ e(\kappa) \leq \max\{e(\kappa), \overline{e}(\kappa)\}. \quad (22) \]

Since \( e(\kappa) \) and \( \overline{e}(\kappa) \) given by (21) represent the forward and backward ZOH approximation errors in the interval \( \hat{I}_k \), we refer to the result on the approximation error of FSFH. The discussion in [10],[11] for the special case \( C = I \) is readily extended to the current situation because the continuity of the integral kernel is also guaranteed for the case with general \( C \). Then, as in [10],[11], it can be shown that the family of functions \( \Phi \) is uniformly equicontinuous, i.e., given \( \forall \epsilon > 0, \exists \delta \) such that
\[ |f(\alpha) - f(\beta)| < \epsilon, \quad \forall f \in \Phi, \quad (23) \]
\[ \forall \alpha, \beta \in [0, h], \quad |\alpha - \beta| < \delta. \]
Then \( N_0 := [h/\delta] \), and \( N \geq N_0 \), we have the inequality \( h' = h/N < \delta \). By letting \( \alpha = \theta_k \) and \( \beta = \theta_{k+1} \), (23) guarantees
\[ e(\kappa) = \max_{\theta \in \hat{I}_k} |f(\theta) - f_k| < \epsilon, \quad \forall f \in \Phi, \quad (24) \]
for \( k = 0, \ldots, N - 1 \). Alternatively, by taking \( \beta = \theta_{k+1} \) and \( \alpha = \theta_k \), we obtain
\[ e(\kappa) = \max_{\theta \in \hat{I}_k} |f(\theta) - f(\theta_{k+1})| < \epsilon, \quad \forall f \in \Phi, \quad (25) \]
for \( k = 0, \ldots, N - 1 \). Since either \( (\overline{e}(\kappa), e(\kappa)) = (e(\kappa), e(\kappa)) \) or \( (\overline{e}(\kappa), e(\kappa)) = (e(\kappa), e(\kappa)) \) holds true for each \( k \) according to whether \( \overline{f}_k = f(\theta_k) \) or \( f(\theta_{k+1}) \), one can conclude
\[ e(\kappa) < \epsilon, \quad \forall f \in \Phi, \quad k = 0, \ldots, N - 1, \]
from (22). This implies
\[ \sup_{\kappa \in \hat{I}_k} \|\Delta z\|_{z_1} = \sup_{\kappa \in \hat{I}_k} \left( \int_0^h |f(\tau) - f(\tau')|^2 \, d\tau' \right)^{1/2} < \epsilon \sqrt{h}. \]
Thus from (18), the operator norm \( \|\Delta z\|_{z_1} \) converges to zero as \( N \to \infty \).

Thus Lemma 1 justifies the computation of the spectrum of \( \mathcal{F} \) from the eigenvalues of \( F_1 \) in the sense that there exists a true point spectrum of \( \mathcal{F} \) in the neighborhood of any nonzero eigenvalue of \( F_1 \).

4.3 Numerical Example

As in [11],[16], let us consider the delay differential equation
\[ \dot{x}(t) = -1/2 x(t) - x(t-h), \]
and analyze its stability with respect to \( h \). It is known that the spectral radius of \( \mathcal{F} \) is 1 when \( h = 4 \sqrt{3} \pi/9 \) and the corresponding eigenvalues on the unit circle can be obtained analytically as \( \lambda = -1/2 \pm j \sqrt{3}/2 \). Let \( \lambda_* = -1/2 + j \sqrt{3}/2 \) and \( \lambda'_* \) be the corresponding numerical value of each approximation. The computational error is quantified by \( E = |\lambda_* - \lambda'_*| \). The values of \( E \) obtained by FSFH, modified FSFH, and FSFFOH approximations for the number of division \( N = 8, 16, 32, 640 \) and 1280 are shown in Table 1. In the modified FSFH approximation, \( X \) is chosen to minimize the Hilbert-Schmidt norm \( \|E\|_{HS} \). Due to the relaxation of causality, FSFFOH achieves much better numerical accuracy over FSFH and the same level as modified FSFH. Thus the FSFFOH approximation can be an alternative of the modified FSFH.

5. Higher Order Approximation

Better performance of the FSFFOH approximation motivates trials of higher order non-causal schemes, e.g., [24]. Here we will introduce the Hermite interpolation, i.e., the third order polynomial approximation uniquely determined from two point values and two point derivatives.

For this purpose, we regard \( \mathcal{F} \) in (2) as an operator mapping \( Z_2 \to Z_1 \), which means that we assume the initial function \( \phi \) for (1) belongs to \( K^2_3 \). Let \( J_{23} \) denote the embedding operator from \( K^3_2 \) to \( K^3_2 \). The monodromy operator \( \mathcal{F} : Z_2 \to Z_2 \) is redefined by
\[ \mathcal{F} = I(J_{23}) \mathcal{F}. \quad (25) \]
This choice of the initial function, or the corresponding choice of the state space, \( Z_1 \) or \( Z_2 \), makes no difference on the stability condition given by the spectral radius of the state transition operator \((\mathcal{F},\mathcal{F})(\theta)\) thanks to the fact that the solution of (1) becomes smoother as it evolves in \( t \). Since \( \mathcal{F} \) is compact in \( Z_2 \) as well, e.g., [25], the approximation procedures mentioned earlier are also valid for \( \mathcal{F} \) on \( Z_2 \).

In what follows, the embedding operator in (25) is replaced by the sample and hold with the Hermite interpolation for an approximation purpose.

5.1 Matrix Representation

Let us introduce a generalized sampler \( \mathcal{S}_H \) which outputs the samples of \( f(\theta) \) and \( \frac{df}{d\theta}(\theta) \) from \( f \in K^3_2 \) simultaneously. Define the differential operator \( \mathcal{D}_H : K^3_2 \to K^3_2 \) by
\[ \mathcal{D}_H f = \frac{df}{d\theta}(\theta). \quad (26) \]
Then \( \mathcal{S}_H \) is written as
\[ \mathcal{S}_H = \left[ \begin{array}{c} \mathcal{S}_0 \\ \mathcal{S}_1 \mathcal{D}_H \end{array} \right]. \quad (27) \]
Given the data \( u_k, v_k \in \mathbb{R}^n, k = 0, \ldots, N \), the Hermite interpolation generates the piecewise cubic vector-valued polynomial

\[
q(\theta) = q_k(\theta), \quad \text{when } \theta \in I_k, k = 0, \ldots, N - 1,
\]

where

\[
q_k(\theta) = \sum_{i=0}^{3} a_i(k)(\theta - \theta_k)^i,
\]

(28)

\[
a_0(k) = u_k, \quad a_1(k) = v_k,
\]

(29)

\[
a_2(k) = -3h^{-2}u_k - u_{k+1} - h^{-1}(2v_k + v_{k+1}),
\]

(30)

\[
a_3(k) = 2h^{-3}u_k - u_{k+1} + h^{-2}(v_k + v_{k+1}).
\]

(31)

It is easy to see that

\[
q_k(\theta_k) = u_k, \quad \frac{dq_k}{d\theta}(\theta_k) = v_k, \quad \frac{dq_k}{d\theta}(\theta_{k+1}) = v_{k+1}.
\]

Since the data \( u_{k+1} \) and \( v_{k+1} \) are used to determine the function \( q_k(\theta) \) defined in the interval \( I_k = [\theta_k, \theta_{k+1}] \), this is precisely another non-causal scheme. Let \( \mathcal{H}_H \) be the hold operator describing this mapping, i.e.,

\[
\mathcal{H}_H : \bar{u} \mapsto q(\cdot),
\]

(32)

where \( \bar{u} = [u_0^T \cdots u_N^T v_0^T \cdots v_N^T]^T \). Then the Hermite interpolation-based approximation of \( \mathcal{F} \) can be written as

\[
\mathcal{F}_H = \mathcal{I}(\mathcal{H}_H)\mathcal{I}(S_H)\mathcal{F}.
\]

(33)

By changing the order of the operators in (33), a matrix representation \( F_H = \mathcal{I}(S_H)\mathcal{F}\mathcal{I}(\mathcal{H}_H) \) follows. The size of \( F_H \) is \( n + 2\mu(N+1) \). This scheme is referred to as the fast-sample/fast-Hermite-interpolation (FSFHII) approximation. Denote the input-output relation of \( F_H \) as

\[
[\bar{x}\bar{y}] = F_H [\bar{x}_0 \bar{u}],
\]

(34)

where \( \bar{y} = [y_0^T \cdots y_N^T \bar{y}_0^T \cdots \bar{y}_N^T]^T, \bar{x}_0, \bar{y} \in \mathbb{R}^{2n(N+1)} \). By (32), the right-hand side of (34) is transformed into

\[
F_H [\bar{x}_0 \bar{u}] = \mathcal{I}(S_H)\mathcal{F}\mathcal{I}(\mathcal{H}_H) [\bar{x}_0 \bar{u}] = \mathcal{I}(S_H)\mathcal{F} [\bar{x}_0 \bar{q}(\cdot)].
\]

Note that

\[
\bar{x}_1 = [I \quad 0] \mathcal{F} [\bar{x}_0 \bar{q}(\cdot)]
\]

\[
= e^{Ah} [\bar{x}_0 + \int_0^1 e^{-A\tau}B\bar{q}(\tau) \, d\tau],
\]

(35)

from (34). By substituting (28)-(31), the definite integral in (35) is calculated as

\[
\int_0^1 e^{-A\tau}B\bar{q}(\tau) \, d\tau = \sum_{l=0}^{N-1} \int_{\theta_l}^{\theta_{l+1}} e^{-A\tau}B\bar{q}(\tau) \, d\tau
\]

\[
= \sum_{l=0}^{N-1} \int_{\theta_l}^{\theta_{l+1}} e^{-A(\eta + \theta_l)}B\bar{q}(\eta) \, d\eta
\]

\[
= \sum_{l=0}^{N-1} (A_l')^{-l-1} \sum_{i=0}^{N} \left( \int_{\theta_l}^{\theta_{l+1}} \eta^i e^{A(\eta + \theta_l)}B \, d\eta \right) a_i(l),
\]

where we can express the coefficient vector of the Hermite interpolation as \( [a_0^T(l) \ a_1^T(l) \ a_2^T(l) \ a_3^T(l)]^T = V \xi \) with

\[
V = \begin{bmatrix}
I & 0 & 0 & 0 \\
0 & I & 0 & 0 \\
-3h^{-2}I & 3h^{-2}I & -2h^{-1}I & h^{-1}I \\
2h^{-3}I & -2h^{-2}I & h^{-2}I & -2h^{-1}I
\end{bmatrix},
\]

(36)

\( \hat{x}_1 \) is calculated as

\[
\hat{x}_1 = A_f \left[ \hat{x}_0 + \sum_{l=0}^{N-1} (A_l')^{-l-1} \hat{Q}_l \xi_l \right],
\]

(36)

\[
\hat{Q} = MV, \quad \hat{M} = \begin{bmatrix}
M_0 & M_1 & M_2 & M_3
\end{bmatrix},
\]

where \( M_i \) \( (i = 0, \ldots, 3) \) are given by (7), (15) and

\[
M_i = \int_0^1 \tau^i e^{Ah-\tau}B \, d\tau \quad (i = 2, 3).
\]

Denote the second output of \( \mathcal{F} \) \( \hat{\bar{x}}_0 \bar{q}(\cdot) \) by \( f \), i.e.,

\[
f := [0 \ I] \mathcal{F} [\bar{x}_0 \bar{q}(\cdot)].
\]

Then \( f \) is given by

\[
f(\theta) = C e^{Ah} \left[ \bar{x}_0 + \int_0^1 e^{-A\tau}B\bar{q}(\tau) \, d\tau \right], \quad \theta \in [0, h],
\]

(37)

and thus \( \bar{y} \) is computed as

\[
\bar{y} = S_H f = \begin{bmatrix}
S_0 f \\
S_1 D_H f
\end{bmatrix},
\]

(38)

by using (27). From (26) and (37), we also obtain

\[
D_H f = CA_e^A \left[ \bar{x}_0 + \int_0^1 e^{-A\tau}B\bar{q}(\tau) \, d\tau \right] + CB\bar{q}(\cdot). \quad (39)
\]

Therefore, the same integral computation as (36) gives the elements of \( \bar{y} \) in (38) as

\[
y_k = C(A_j')^k \left[ \bar{x}_0 + \sum_{l=0}^{k-1} (A_l')^{-l-1} \hat{Q}_l \xi_l \right],
\]

(40)

\[
z_k = CA_e^A \left[ \bar{x}_0 + \sum_{l=0}^{k-1} (A_l')^{-l-1} \hat{Q}_l \xi_l \right] + CB\bar{u}_k,
\]

(41)

\( k = 0, \ldots, N \). Thus, a matrix representation for \( F_H \) in (34) is obtained by rearranging (36), (40), and (41). Let \( [Q_0 \ Q_1 \ Q_2 \ Q_3] := \bar{Q} \). By using the same notation for (16) and (17), \( F_H \) is given as

\[
F_H = \mathcal{I} \begin{bmatrix}
C \\
CA \\
F_{12} \mathcal{I}(R(Q_0, Q_1)) \\
F_{12} \mathcal{I}(R(Q_2, Q_3))
\end{bmatrix} + \begin{bmatrix}
0_{\mu \times \mu} \\
0_{\mu \times \mu} \\
0_{\mu \times \mu} \\
0_{\mu \times \mu}
\end{bmatrix}
\]

(41)

Note that as given in [26], the definite integrals \( M_i \) \( (i = 0, 1, 2, 3) \) can be calculated efficiently via the matrix exponential formula

\[
[ A'_0 \ M_0 \ M_1 \ M_2 / 2 \ M_3 / 6 ] = \begin{bmatrix}
I & 0 & 0 & 0 \\
A & B & 0 & 0 \\
0 & 0 & I_{3\mu \times 3\mu} & 0 \\
0 & 0 & 0 & I_{3\mu \times 3\mu}
\end{bmatrix} e^{Wh},
\]

with

\[
W = \begin{bmatrix}
A & B & 0 & 0 \\
0 & 0 & I_{3\mu \times 3\mu} & 0 \\
0 & 0 & 0 & I_{3\mu \times 3\mu}
\end{bmatrix}.
\]
5.2 Proof of Convergence

The following lemma establishes the generalized convergence of the perturbation on the operator $\mathcal{F}$ caused by the FSFHI approximation. This result immediately justifies the computation of the spectrum of $\mathcal{F}$ from the eigenvalues of $\mathcal{F}_H$.

**Lemma 2** The perturbation

$$\Delta_H := (\mathcal{I}(\mathcal{H}_I) \mathcal{I}(S_H) - \mathcal{I}(\mathcal{F}_H)) \mathcal{F}_H - \mathcal{F}$$

converges to zero as $N \to \infty$ in the generalized sense.

**Proof** Since $\mathcal{F}$ is now regarded acting on $Z_2$, first we specify the norm for $z = \left[ \frac{x}{\hat{a}(\cdot)} \right] \in Z_2$ by

$$\|z\|_{Z_2} = \max \left\{ |x|, \sup_{\theta \in [0,h]} |\hat{a}(\theta)| \right\}$$

and introduce the following family of functions:

$$\Phi = \left\{ f | f = [ 0 \ I ] \hat{F}_Z \mid z \in \mathcal{U}_2 \right\},$$

where $\mathcal{U}_2 = \{ z \mid z \in Z_2, \|z\|_{Z_2} = 1 \}$. Let $\hat{f}$ be the reconstruction of $f$ by the FSFHI, i.e.,

$$\hat{f}(\cdot) = \mathcal{H}_I S_H f(\cdot).$$

Similarly to the case of FSFFOH, it suffices to show that

$$\sup_{\theta \in [0,h]} |\hat{f}(\theta) - f(\theta)|$$

becomes arbitrarily small by taking $N$ large enough, irrespective of the choice of $f \in \Phi$. Define $e_H(k)$ by

$$e_H(k) := \max_{\theta \in I_k} |\hat{f}(\theta) - f(\theta)|, \quad k = 0, \ldots, N - 1.$$  \hspace{1cm} (42)

From (28)-(31), the difference in the right-hand side of (42) is written as

$$\hat{f}(\theta) - f(\theta) = \frac{2(u_k - u_{k+1}) + h'(v_k + v_{k+1})}{h^3}(\theta - \theta_k)^3$$

$$+ \frac{-3(u_k - u_{k+1}) - h'(2v_k + v_{k+1})}{h^2}(\theta - \theta_k)^2$$

$$+ v_k(\theta - \theta_k) + u_k - f(\theta),$$

for $\theta \in I_k$ where

$$u_k = f(\theta_k), \quad v_k = \frac{\partial f}{\partial \theta}(\theta_k).$$

Since $|\theta - \theta_k| \leq h'$ for $\theta \in I_k$, we obtain the bound

$$|\hat{f}(\theta) - f(\theta)| \leq 5|u_k - u_{k+1}|$$

$$+ 6h' \max(|v_k|, |v_{k+1}|) + |u_k - f(\theta)|.$$  \hspace{1cm} (43)

Even in the case for $\mathcal{U}_2$, given $\forall \varepsilon > 0$, one can guarantee the existence of $\delta$ such that $|a, b| \in [0, h]$ and $|a - b| < \delta$ imply (23) for $\forall f \in \Phi$, along the same lines as [10],[11]. Given $\varepsilon' > 0$, let $\varepsilon = \varepsilon'/12$. Then, for all number of the subdivision $N > N_1 := [h/\varepsilon']$, the forward ZOH approximation error $e_f(k)$ becomes smaller than $\varepsilon$ for $k = 0, \ldots, N - 1$ as expressed in (24). Therefore, $\forall N > N_1$, the following inequality holds:

$$5|u_k - u_{k+1}| + |u_k - f(\theta)| < \varepsilon'/2.$$  \hspace{1cm} (44)

By the same calculation leading to (39), $\frac{\partial f}{\partial \theta}(\theta)$ is expressed as

$$\frac{\partial f}{\partial \theta}(\theta) = D_H \left[ 0 \ I \right] \hat{F}_Z$$

$$= C A e^{A\theta} \left\{ x + \int_{0}^{\theta} e^{-A t} B \tilde{u}(\tau) \, d\tau \right\} + C B \tilde{u}(\theta),$$

for some $z = \left[ \frac{x}{\hat{a}(\cdot)} \right] \in \mathcal{U}_2$. Similarly to the proof of the claim (a) of Lemma 1 in [11], it can be shown that $\frac{\partial f}{\partial \theta}(\theta) \in K^2_2$ above is uniformly bounded on $\mathcal{U}_2$, i.e., $\forall K_1 > 0$ such that

$$\|\frac{\partial f}{\partial \theta}(\theta)\| \leq K_1, \quad \forall \theta \in [0,h], \quad \forall f \in \Phi.$$

Let $N_2 := \lfloor (12K_1h)/\varepsilon' \rfloor$. Then for $\forall N \geq N_2$, we have the following inequality:

$$6h' \max(|v_k|, |v_{k+1}|) \leq 6h/NK_1 \leq \varepsilon'/2.$$  \hspace{1cm} (45)

By combining (43), (44) and (45), we can conclude that whenever $N > N_3 := \max(N_1, N_2)$, it holds that

$$e_H(k) < \varepsilon', \quad \forall f \in \Phi, \quad k = 0, \ldots, N - 1.$$  \hspace{1cm} (46)

This implies

$$\|\Delta_H\| = \sup_{z \in \mathcal{U}_2} \|\Delta_H z\|_{Z_2} = \sup_{f \in \Phi} \left\{ \max_{\theta \in [0,h]} |\hat{f}(\theta) - f(\theta)| \right\}$$

$$= \sup_{f \in \Phi} \left\{ \max_{k} e_H(k) \right\} < \varepsilon', \quad \forall N > N_3.$$  \hspace{1cm} (47)

\[ \square \]

5.3 Numerical Example

Consider the stability of the delayed feedback control system

$$\begin{cases} \dot{x}(t) &= 0 \quad 1 \quad x(t) + [0 \ 1] u(t), \\ u(t) &= -0.4 \quad -0.2 \quad \{ x(t) - x(t - h) \}. \end{cases}$$

The closed-loop dynamics is governed by the delay differential equation (1) with

$$A = \begin{bmatrix} 0 & 1 \\ -4 & -0.4 \end{bmatrix}, \quad G = BC, \quad B = [0 \ 1], \quad C = [0.4 \ 0.2].$$

As discussed in [27], this time-delay system reveals interval stability in terms of $h$. The stability regions of $h$ are given by $(h_1(k), h_2(k))$, $k = 0, 1, 2$, where $h_1(k) = \pi(4k + 1)/4$, $h_2(k) = \sqrt{30\pi(2k + 1)/12}$. At the end points of stable intervals, $\mathcal{F}$ has eigenvalues on the unit circle. Those values can be obtained analytically. At $h = h_1(k)$, they are located at $\pm j$. When $h = h_2(k)$, they are at $-1 \pm j0$ (multiple eigenvalues).

As in the previous example, we measure the numerical precision by the absolute error about the eigenvalues on the unit circle corresponding to a critical delay length $h$. Consider the case $h = h_2(2)$. Figure 2 shows the absolute errors of FSFH, modified FSFHI, FSFFOH and FSFHI approximations versus the number of division $N = 2^k$ ($k = 6, 7, \ldots, 10$). The computational error with the FSFHI approximation is drastically smaller than other three methods for the same $N$. To evaluate the computational efficiency, we must take the fact that the matrix size of FSFHI is almost doubled by the use of the derivative information into consideration. Figure 3 illustrates the relative\(^2\) CPU

\(^2\) The CPU time of FSFHI for $N = 2^k$ is chosen as the base unit.
time of each method versus $N$. As expected, the CPU time of FSFHI is the largest at the same $N$ but is at the same level as (or even smaller than) those of other three methods at $2N$. Since FSFHI is much more accurate than other methods with doubled $N$ (Fig. 2), one can conclude that the FSFHI shows the best computational efficiency among others.

![Fig. 2 Computational error.](image)

![Fig. 3 Relative CPU time.](image)

6. Conclusions

In this paper, the numerical computation of the spectrum of the monodromy operator is considered. Based on the structural observation of the existing methods, a new finite-dimensional approximation with non-causal FOH is introduced. Not only it gives better numerical results over ZOH, but also it admits a mathematical justification which guarantees the convergence. It is also shown that a higher order polynomial approximation with Hermite interpolation effectively improves the numerical accuracy. The convergence proof for the case of this interpolation is also provided. Although the introduction of the derivative information almost doubles the size of the matrix representation for the same number of the division, one can say that FSFHI is still effective among others due to its overwhelming precision even if this size issue is taken into account.

References


Kentaro Hirata (Member)

He received his B.E., M.E., and Ph.D. degrees from Kyoto University, Japan, in 1988, 1990, and 1997, respectively. In 2005, he joined the faculty of Nara Institute of Science and Technology, where he is currently an Associate Professor of Graduate School of Information Science. His research interests include control systems theory for periodic motions, human-machine systems and industrial applications. He is a member of ISCIE, ISIJ and IEEE.

Tomomichi Hagiwara (Member)

He received his B.E., M.E., and Ph.D. degrees from Kyoto University, Japan, in 1984, 1986, and 1990, respectively. Since 1986, he has been with the Department of Electrical Engineering, Kyoto University, where he is currently a Professor. His research interests include dynamical system theory and control theory such as analysis and design of sampled-data systems, time-delay systems and two-degrees-of-freedom control systems. He is a member of ISCIE, IEEJ and IEEE.

Atsushi Itokazu

He received his B.A. and M.E. degrees from University of the Ryukyus and Nara Institute of Science and Technology, Japan, in 2007 and 2009, respectively. His research topic at NAIST was numerical analysis of delay systems. From 2009, he is working for Mitsubishi Electric Corporation as an electrical engineer.