MAPPING DISCRETE-TIME MODELS FOR DESCRIPTOR-SYSTEMS WITH CONSISTENT INITIAL CONDITIONS

Shin Kawai and Noriyuki Hori

Digital Control Laboratory, Division of Intelligent Interaction Technologies, University of Tsukuba, Tsukuba, Japan
Email: kawai@digicon-lab.esys.tsukuba.ac.jp; hori@iit.esys.tsukuba.ac.jp

Received March 2015, Accepted August 2015
No. 15-CSME-44, E.I.C Accession 3819

ABSTRACT
Discretization of a regular continuous-time descriptor-system, whose initial condition is consistent with its input, is considered using a general mapping method presented in our previous paper. The proposed mapping discrete-time model is shown to be a proper discretization under the definition explained in the paper. This assures that the response of the mapping model approaches that of the continuous-time descriptor system as the sampling period approaches zero. The consistency of initial conditions for the discrete-time model is also studied and the long-standing issue of ambiguities surrounding irregularities of discrete-time responses at the initial time are clarified with a simple solution. A proper range of design parameters are investigated and their suitable choices suggested. To illustrate the use of the proposed method, a simple circuit that cannot be expressed in the ordinary state-space form is considered. Its responses to a sinusoidal input when started from the consistent and inconsistent initial conditions are simulated to show that the irregularities at the initial time can be overcome easily. The proposed technique provides a convenient simulation and design environment for handling discrete-time systems in a unified manner with consistency and ease.

Keywords: descriptor systems; consistent initial conditions; discretization.
1. INTRODUCTION

Simulation and design of systems and controllers rely heavily on the use of digital devices lately, which require appropriate discrete-time models of continuous-time systems. Such models are well known for systems expressed in transfer-functions and state-space forms [1, 2]. However, as systems become increasingly more complex, models with more flexibilities are being developed to deal with various phenomena observed in such areas as large-scale systems [3], singularly perturbed systems [4], switched systems [5] and inverse systems [6], to name a few. These systems often involve subsystems that cannot be handled properly by state-space equations and require the descriptor-system expression [7], which is also called the generalized state-space form [8]. Their fundamental properties are investigated in [9] and control applications have been discussed in [10–12]. These are developed mostly for continuous-time descriptor systems and only few exist for the discrete-time case, even though there are numerous papers for discretization of systems expressed in transfer-function and state-space forms.

To work with descriptor systems in a discrete-time framework, a proper definition of discretization should be used to prove their veracity. Existing techniques of discretization for descriptor systems are ones based on numerical analyses [13–16]. In [13], the discrete-time model that is obtained by applying the forward difference approximation to the differentiator in the impulsive mode of the descriptor system is shown to be equivalent to the model obtained by applying the same approximation to the derivative of the input that appears in the solution of the system. Markov parameters are used in [14] to discretize the descriptor system, where again the forward difference approximation is applied to the derivative terms of the input. This resulted in the same model as one obtained in [13]. Discretization from the view-point of a solution to the system is attempted in [15], where the derivative of the input term is approximated by the backward difference. This model is identical to one proposed in [16], where the impulsive mode is discretized using the backward difference. Unlike the forward difference model [13, 14], the backward difference model [15, 16] is causal and can be used for on-line computations.

While the relationship among these studies are not clearly seen in their respective form, the model reported in [17] combines them in a unified framework and includes them as its special cases, giving a perspective view and suggesting new models. However, no proof is given there as to the validity of the mapping model from the discretization point of view. Moreover, no explanation is given there as to how an initial condition should be chosen, which resulted in impulsive response at the start of simulations. This issue on consistency of initial conditions for seamless numerical simulations has not been clarified for the discrete-time case. The present paper reviews the generalized discrete-time mapping model for a regular descriptor system, gives a proof that the model is valid in the sense of a given discretization definition, clarifies the consistent initial conditions to go with the model, and investigates how a discretization parameter can affect the performance and the stability of the resulting model.

The paper is organized as follows: After the introduction in Section 1, the problem to be solved is formulated in Section 2 where the descriptor system and the definition of discretization are reviewed briefly. Section 3 proposes a modification of the discretization definition so that the generalized mapping model is indeed a proper discretization for consistent initial conditions. Stability of the discrete-time model is studied as well. Section 4 applies the method to an electrical circuit and presents simulation results. Section 5 states conclusions.

2. PROBLEM STATEMENT

2.1. Descriptor System

It is assumed throughout the present study that the input is continuously differentiable for a sufficient number of times (this point will be made clear shortly) and the continuous-time descriptor system is regular. Let the
The system be given by

\[ \ddot{x}(t) = A\dot{x}(t) + B\dot{u}(t), \quad \dot{x}(0-) = x(0-) \]

where \( \dot{x}(t) \in \mathbb{R}^n \) is the descriptor vector, \( \dot{u}(t) \in \mathbb{R}^m \) the input, matrices \( \ddot{E}, A, B \) are constants with appropriate dimensions, and upper bars indicate continuous-time symbols in general. Matrix \( \ddot{E} \) can be singular and when it is non-singular, this equation becomes a standard state-space form. The initial condition is \( \dot{x}(0-) \), which is necessary and sufficient to completely determine \( \dot{x}(t) \) given \( \dot{u}(t) \), for \( t \geq 0 \).

The regular descriptor system can be transformed into the Kronecker form \([18]\), as

\[ \ddot{E}\frac{d\ddot{x}(t)}{dt} = \tilde{A}\dot{x}(t) + \tilde{B}\dot{u}(t), \quad \dot{x}(0-) = x(0-) \]

where

\[ \tilde{E} = \tilde{Q}^{-1}\ddot{E}\tilde{P} = \begin{bmatrix} I & 0 \\ 0 & \tilde{N} \end{bmatrix}, \quad \tilde{A} = \tilde{Q}^{-1}A\tilde{P} = \begin{bmatrix} \tilde{A}_s & 0 \\ 0 & I \end{bmatrix}, \quad \tilde{B} = \tilde{Q}^{-1}B, \]

and \( I \) is the identity matrix. Matrix \( \tilde{P} \) and \( \tilde{Q} \) can be determined from a solution to a generalized eigenvalue problem. Research on their numerical computations can be found in \([19, 20]\). \( \tilde{N} \) is an \((n-d) \times (n-d)\) block Jordan matrix with null eigenvalues \([18]\), which is nilpotent with index \( h \). The input is assumed to be continuously-differential at least for \( h-1 \) times, which clarifies the statement given at the beginning of this section. The original and transformed descriptor vectors are related by the non-singular matrix as

\[ \tilde{x}(t) = \tilde{P}x(t). \]

Equation (2) is a decoupled system, which consists of the slow (exponential) and fast (static/impulsive) modes, and can be written as

\[ \frac{d\tilde{x}_s(t)}{dt} = \tilde{A}_s\tilde{x}_s(t) + \tilde{B}_s\dot{u}(t), \quad \tilde{x}_s(0-) = \tilde{x}(0-) \]

\[ \frac{d\tilde{x}_f(t)}{dt} = \tilde{x}_f(t) + \tilde{B}_f\dot{u}(t), \quad \tilde{x}_f(0-) = \tilde{x}(0-) \]

where \( \tilde{x} = [\tilde{x}_s^T \tilde{x}_f^T]^T \) with \( \tilde{x}_s \in \mathbb{R}^d \) being the vector of slow-mode variables and \( \tilde{x}_f \in \mathbb{R}^{n-d} \) the fast-mode variables. Figure 1 shows the block diagram of the continuous-time descriptor system in Kronecker form.

2.2. Consistent Initial Condition

The solutions of the descriptor system (5) and (6) are given, respectively, by \([18]\)

\[ \tilde{x}_s(t) = e^{\tilde{A}_s(t-\tau)}\tilde{x}(0-) + \int_0^t e^{\tilde{A}_s(t-\tau)}\tilde{B}_s\dot{u}(\tau) d\tau, \]

\[ \tilde{x}_f(t) = -\sum_{i=1}^{h-1} \tilde{N}_i^{(i-1)}(t)\tilde{x}_f(t-0-) - \sum_{i=0}^{h-1} \tilde{N}_i \tilde{B}_f \left( \dot{u}^{(i)}(t) + \sum_{k=0}^{i-1} \tilde{\delta}^{(i-k)}(t)\dot{u}^{(i-k-1)}(0) \right), \]

where \( \tilde{\delta} \) denotes the impulse function and \( \tilde{\delta}^{(i)} \) its \( i \)-th derivative. The slow mode solution (7) represents the dynamic response and is uniquely determined given an arbitrary initial condition and an input. The fast mode solution (8) shows the response of the static and impulsive modes and contains generalized functions for arbitrary initial conditions and inputs. In order for the response to contain no distributional components,
the initial condition on the fast-mode variables should satisfy the so called the consistency condition \( [9] \), which is

\[
\bar{x}_f(0-) = -\sum_{i=0}^{h-1} \bar{N}_i \bar{B}_f \bar{u}^{(i)}(0).
\] (9)

It should be noted that the consistent initial condition depends on the input being applied to the system. For continuous inputs, this is a necessary and sufficient condition for the response to be continuous at the initial time. Indeed, when this condition is satisfied, the fast-mode response (8) becomes

\[
\bar{x}_f(t) = -\sum_{i=0}^{h-1} \bar{N}_i \bar{B}_f \bar{u}^{(i)} (t),
\] (10)

and there will be no discontinuity at the initial time. For discontinuous inputs, the consistency condition is still required for the continuity of the response at the initial time. However, the condition itself involves distributions and it is practically impossible to realize it. In the rest of the present paper, it is assumed that the initial condition satisfies the above consistency requirement, which includes the smoothness assumption on the input as set out at the beginning of this section.

2.3. Discretization

A discrete-time signal is denoted by \( f(k, T) = f_k \), where \( k \) is integer that indicates the step number and \( T > 0 \) is the sampling period, as opposed to a continuous-time signal denoted by an upper bar as \( \bar{f}(t) = f \).

A conventional definition of discretization for ordinary signals is the following \([21]\):

**Definition 1** A discrete-time (vector) signal \( f_k \) is said to be a discretization of a continuous-time (vector) signal \( f \) if the following condition is satisfied: For each fixed \( \tau \) that satisfies \( kT \leq \tau < (k+1)T \), the following holds:

\[
\lim_{T \to 0} \| \bar{f}(\tau) - f(k, T) \| = 0.
\] (11)
This definition is applicable to continuous signals, although it can be modified to accommodate signals with discontinuity at a single time-instant [21]. In the present study, the continuity of signals is assumed.

By applying the definition of discretization of signals to inputs and outputs of systems, discretization of systems can be defined.

**Definition 2** Consider a continuous-time system whose input is sufficiently continuously differentiable. Then a discrete-time system is said to be a discrete-time model, or discretization, of this continuous-time system if its outputs are discretization for any input that is a discretization of the input to the continuous-time system.

Some important theorems derived from this definition show that almost all existing discrete-time systems are discrete-time models in this sense for those systems expressible in the transfer-function and state-space forms [2, 21]. In these studies, the initial conditions of both the continuous-time and the discrete-time systems are set arbitrarily to zero. While this does not bring any serious problems to the slow-mode of the linear descriptor system, this can cause irregularities to the fast mode, which called for some modifications, as presented shortly. Under the circumstances, the problem to be solved in the present paper is formulated as follows:

**Problem Statement:** Clarify the definition of discretization applicable to a continuous-time, linear, regular, time-invariant, descriptor-system with a consistent initial condition, and prove that the generalized mapping model is indeed a discretization of the original descriptor system in this sense.

### 3. MAPPING DISCRETE-TIME MODELS

#### 3.1. Discretization Revisited

To facilitate the exposition, the following discrete-time operator is used, rather than the conventional shift-operator [1]:

\[
\delta f_k \triangleq \frac{q - 1}{T} f_k = \frac{f_{k+1} - f_k}{T},
\]

(12)

where \( q \) is the conventional shift-operator that satisfies \( q f_k = f_{k+1} \). The mapping method is one of the discretization techniques where a differentiator is replaced as \( \frac{df}{dt} \rightarrow \frac{\delta}{(\mu\delta + 1)} \), where \( \mu \) is an arbitrary real number that determines the relative weights of values at different sampling instants. In fact, \( \mu = 0 \) corresponds to the forward-difference model, \( \mu = 1 \) the backward-difference, and \( \mu = 1/2 \) the trapezoidal (Tustin) difference. It can be shown that for a function \( \bar{f} \) that is differentiable for a sufficient number of times,

\[
\lim_{T \to 0} \left\| \frac{\delta}{T} \left( \frac{\delta}{T\mu\delta + 1} \right)^i f_k \right\| = 0,
\]

(13)

and, thus, \( \left( \frac{\delta}{T\mu\delta + 1} \right)^i f_k \) is a discretization of the i-th derivatives of \( \bar{f} \).

**Definition 3** Consider a continuous-time system whose initial condition is consistent with the input. Then a discrete-time system is said to be a discrete-time model, or discretization, of this continuous-time system if its responses (whether internal-variables or outputs) are discretization (i) for any input that is a discretization of the input to the continuous-time system and (ii) for any initial condition on appropriate variable vectors that satisfy Eq. (11) at \( t = kT = 0 \).
In this definition, (a) internal variables are taken into account and (b) initial conditions are less restrictive than in Definition 1. As for (a), there are cases where the discrete-time system is a model from the input-output point of view, but not in terms of internal variables due to their non-uniqueness. However, this will allow more detailed analysis on internal aspects of the system. As for (b), it has previously been implicitly assumed that the initial conditions match exactly between the continuous and discrete time systems, independently of T. The present modification is more consistent with the definition of signal discretization.

3.2. Mapping Models

The generalized mapping discrete-time model of a continuous-time descriptor system was first presented in [17] and is one obtained by the substitution \( \frac{df}{dt} \rightarrow \frac{\delta}{(T \mu + 1)} \), where \( \mu \) must now be larger than \( \frac{1}{2} \) for a reason of certain stability as will be clarified later.

\[ \begin{align*}
  \delta v(k, T) &= \Gamma \left[ A v(k, T) + B u(k, T) \right], \\
  \bar{E} v(0, T) &= \Gamma^{-1} \bar{x}(0-) - T M \bar{B} \bar{u}(0), \\
  x(k, T) &= \Gamma \left[ \bar{E} v(k, T) + T M \bar{B} u(k, T) \right].
\end{align*} \]  

(14)  
(15)  
(16)

where

\[ \Gamma = (\bar{E} - T M \bar{A})^{-1}, \]  
(17)  
\[ M = \begin{bmatrix} \mu_s I & 0 \\ 0 & \mu_f I \end{bmatrix}, \quad \mu_s: \text{any real number, } \mu_f: \text{real number larger than } \frac{1}{2}. \]  
(18)

This model can be broken into the slow and fast modes as

\[ \begin{align*}
  \delta v_s(k, T) &= \Gamma_s \left[ \bar{A}_s v_s(k, T) + B u(k, T) \right], \quad v_s(0, T), \\
  x_s(k, T) &= \Gamma_s \left[ v_s(k, T) + T \mu_s B u(k, T) \right], \\
  \Gamma_s &= (I - T \mu_s \bar{A}_s)^{-1}, \\
  v_s(0, T) &= \Gamma_s^{-1} \bar{x}_s - T \mu_s \bar{B} \bar{u}(0).
\end{align*} \]  

(19)  
(20)  
(21)  
(22)

and

\[ \begin{align*}
  \delta v_f(k, T) &= \Gamma_f \left[ v_f(k, T) + B u(k, T) \right], \quad \tilde{N} v_f(0, T), \\
  x_f(k, T) &= \Gamma_f \left[ \tilde{N} v_f(k, T) + T \mu_f \tilde{B} u(k, T) \right], \\
  \Gamma_f &= (\tilde{N} - T \mu_f I)^{-1}, \\
  \tilde{N} v_f(0, T) &= \Gamma_f^{-1} \tilde{x}_f(0-) - T \mu_f \tilde{B} \tilde{u}(0).
\end{align*} \]  

(23)  
(24)  
(25)  
(26)

**Proof.** The slow mode part of the discrete-time model is the state-space equation [21] and, thus, the proof is omitted. As for the fast mode, Eq. (5), upon the mapping substitution, one obtains

\[ \tilde{N} \delta(T \mu_f \delta + 1)^{-1} x_f(k, T) = x_f(k, T) + \tilde{B} u(k, T), \]  
(27)

\[ \tilde{N} \delta(T \mu_f \delta + 1)^{-1} x_f(k, T) = x_f(k, T) + \tilde{B} u(k, T), \]  
(27)

which can be rearranged into the following two forms. The first is to rewrite the above as
\[ \delta \left\{ \left( \tilde{N} - T \mu \delta I \right) x_f(k, T) - T \mu \delta \tilde{B}_f u(k, T) \right\} = x_f(k, T) + \tilde{B}_f u(k, T), \]  
(28)
and define the new variable by
\[ \tilde{N} v_f(k, T) = \left( \tilde{N} - T \mu \delta I \right) x_f(k, T) - T \mu \delta \tilde{B}_f u(k, T). \]  
(29)
Applying the above to Eq. (28), Eqs. (23) to (25) can be obtained as a sufficient condition. The second is to rewrite Eq. (27) for \( x_f \) and use \( \tilde{N}^h = 0 \) as
\[ x_f(k, T) = \tilde{N} \delta (T \mu \delta + 1)^{-1} x_f(k, T) - \tilde{B}_f u(k, T) \]
\[ = \tilde{N} \delta (T \mu \delta + 1)^{-1} \left\{ \tilde{N} \delta (T \mu \delta + 1)^{-1} x_f(k, T) - \tilde{B}_f u(k, T) \right\} - \tilde{B}_f u(k, T) \]
\[ = - \sum_{i=0}^{h-1} \tilde{N}^i \tilde{B}_f (\delta (T \mu \delta + 1)^{-1})^i u(k, T). \]  
(30)
Since \( u_k \) is a discretization of \( \tilde{u} \) by assumption, this \( x_f \) is a signal discretization of \( \tilde{x}_f \) in Eq. (10) in view of eq. (13). To prove that the fast mode given by Eqs. (23) to (25), is a discretization of (6), it remains to be shown that the initial condition (26) satisfies Eq. (11) at \( t = kT = 0 \). This is done by choosing \( x_f(0, T) = \tilde{x}_f(0-) \) and \( u(0, T) = \tilde{u}(0) \) in Eq. (29) so that
\[ \tilde{N} v_f(0, T) = \left( \tilde{N} - T \mu \delta I \right) \tilde{x}_f(0-) - T \mu \delta \tilde{B}_f \tilde{u}(0), \]  
(31)
which is Eq. (26). This leads to \( \tilde{N} v_f(0, T) \to \tilde{N} \tilde{x}_f(0-) \) as \( T \to 0 \) and, thus, the initial condition part is proven by considering \( \tilde{f}(0-) = \tilde{N} \tilde{x}_f(0-) \) and \( f(0, T) = \tilde{N} v_f(0, T) \) in Eq. (11).

Remarks

- \( v_s \) and \( v_f \) are given by the discrete-time state-space equations and \( x_s \) and \( x_f \) are the output equations.
  These can be realized in a software form using the discrete-time integrator \( \frac{1}{\delta} \) as a building block.
• \( \Gamma_s \) is almost always non-singular and if not, it can always be made so by decreasing \( T \).

• \( \Gamma_f \) is always nonsingular as long as \( T \mu_f \) is nonzero.

• The mapping model equation (14) can be written as

\[
(\tilde{E} - TM\tilde{A})\delta v(k,T) = \tilde{A}v(k,T) + \tilde{B}u(k,T),
\]

which approaches the continuous-time equation (2) as \( T \to 0 \).

• While \( v_s \) and \( v_f \) approach \( x_s \) and \( x_f \), respectively, as \( T \to 0 \), \( x_s \) and \( x_f \) should be used as discretization of \( \tilde{x}_s \) and \( \tilde{x}_f \), rather than \( v_s \) and \( v_f \).

• The mapping model of the original system equation (1) can be obtained from the Kronecker form using Eq. (4).

### 3.3. Initial Condition

The necessary information on the initial condition of the discrete-time model given by (23) and (24) is \( \tilde{N}v_f(0,T) \), while all elements in the vector \( v_I(0,T) \) have to be specified at the beginning of numerical simulations. Therefore, the first element of vector \( v_{f,0}(0,T) \) that corresponds to the \( i \)-th Jordan block of \( \tilde{N} \) can be set to an arbitrary number. For this reason, the vector evaluated by

\[
v'_I(0,T) = \tilde{N}^T \{ \tilde{N} - T\mu_I I \} \tilde{x}_f(0-0) - T\mu_I B\tilde{u}(0) = \tilde{N}^T \tilde{N}v_I(0,T),
\]

(33)

could be used to specify the initial condition for simulations, in place of Eq. (26), since

\[
\tilde{N}v_I(0,T) = \tilde{N}v'_I(0,T).
\]

In the above, the initial condition \( \tilde{N}v_I(0,T) \) was derived so that \( x_f(0,T) = \tilde{x}_f(0-0) \) holds. However, it could have been chosen as \( v_I(0,T) = \tilde{x}_f(0-0) \) just to satisfy Definition 3 of discretization. The reason why this is not used is that the former choice gives a consistent initial condition, while others do not, including the latter, and thus \( x_f(0,T) \neq \tilde{x}(0) \). This will give a transient response that does not exist in the continuous-time original, as pointed out in [15], which shows that the transient disappears in \( h \) steps for \( \mu_I = 1 \). This is generalized below. By the way, design parameter \( \mu_f \) characterizes the fast mode of the discrete-time model. The choice of \( \mu_f = 0 \) corresponds to the forward-difference model, which is identical to the model proposed in [13], while \( \mu_f = 1 \) gives the backward-difference model proposed in [15]. Although the forward-difference model can be defined, it cannot be implemented since \( \Gamma_f \) cannot be computed when \( \mu_f = 0 \). Eqs. (23) and (24) yield the response \( x_{f,IC}(k,T) \) to an initial condition as

\[
x_{f,IC}(k,T) = \Gamma_f\tilde{N}(T\Gamma_f + I)^k v_I(0,T)
\]

\[
= \Gamma_f\tilde{N}\Gamma_f^k \left\{ \tilde{N} - T(\mu_f - 1) I \right\}^k v_I(0,T)
\]

\[
= \tilde{N}\Gamma_f^{k+1} \left\{ \tilde{N} - T(\mu_f - 1) I \right\}^k v_I(0,T).
\]

(35)

This shows that the order of the response can be written as

\[
O\left( \frac{(-1)^k T^k (\mu_f - 1)^k}{T^{k+1} \mu_f^k} \right) = O\left( (-1)^k \frac{\tilde{N}}{T} \left( \frac{\mu_f - 1}{\mu_f} \right)^k \right),
\]

(36)

which approaches zero as \( k \to \infty \) when \(-1 < \frac{\mu_f - 1}{\mu_f} < 1\). The model given by (23) and (24) satisfies this condition when the stability condition \( \frac{1}{T} < \mu \) is satisfied.
It should be noted that when $\mu_I = 1$, the convergence of the transient response can be finite and one-step faster than that shown in [15]. When $\mu_I = 1$, the response equation (35) to the initial condition gives
\[
x_{t,IC}(k, T) = \Gamma_I \bar{N}^{k+1} \nu_I(0, T)
\]
\[
= (\Gamma_I \bar{N})^{k+1} \nu_I(0, T)
\]
\[
= \left[ -\sum_{i=1}^{h-1} T^{-i} \tilde{N}^i \right]^{k+1} \nu_I(0, T),
\]
whose order is $O\left( \left( \frac{\bar{N}}{T} \right)^{k+1} \right)$. Therefore, the response converges to zero in $k \geq h - 1$ steps. In comparison to this, the number of steps is $h$ in [15], where the initial condition is given as $x_I(0, T)$, whereas it is given as $\bar{N}x_I(0, T)$ in the present study.

It should also be noted that the descriptor variables in the static mode are the sampled signals $f_k = \bar{f}(t)|_{t=kT}$ and, thus, are discretized exactly independently of $\mu_I$ and $T$.

### 3.4. Discrete-Time Fast-Mode Poles

As far as the mapping discretization is concerned, the relationship of slow-mode poles between the continuous-time and the discrete-time systems is the same as that of the state-space form [2]. However, the fast-mode poles of the continuous-time descriptor system are infinite, while those of the discrete-time model are finite and all located at $-\frac{1}{T\mu_I}$. This can be seen from the roots of $\det(\varepsilon I - \Gamma_I) = 0$, where
\[
\Gamma_I = (\bar{N} - T\mu_I I)^{-1} = -\sum_{i=0}^{h-1} (T\mu_I)^{-i-1} \tilde{N}^i.
\]

The stability region of the complex $\varepsilon$ plain is known to be inside of the circle that is centered at $-\frac{1}{T}$ with the radius of $\frac{1}{T}$ [11]. Therefore, in order for the discretized, real, and fast-mode poles to be stable, $\mu_I$ should be chosen to be larger than $\frac{1}{T}$. To see how the pole locations depend on the value of $\mu_I$, let us convert the poles of the mapping model into those in the Laplace plain $s$. It should be emphasized that the operator used for system expression ($\varepsilon$ operator) and the method of discretization (mapping substitution) are two separate concepts and should not be confused; an operator used for discretization (substitution) can be different from one used for system expression. The shift operator $q$ is related to the time-shift of $T$ seconds and satisfies $q = e^{DT}$, where $D$ is the differential operator [11]. This is considered as a declaration that the smallest time period considered in the discrete-time context is $T$, so that its left-hand-side is a discrete-time shift operator, filling the gap between the continuous (right-hand-side $e^{DT}$) and discrete (left-hand-side $q$) time domains. Its transformed version is $z = e^{sT}$, where $s = \sigma + j\omega$. In the present study, the $\varepsilon$ operator, which is the transformed version of the $\delta$ operator, is used, where $\varepsilon = \left( \frac{e^{sT} - 1}{sT} \right) = \Sigma + j\Omega$. These two domains are related as follows:
\[
\begin{align*}
\Sigma &= \frac{e^{sT}\cos\omega - 1}{sT}, \\
\Omega &= \frac{e^{sT}\sin\omega}{sT}, \\
\sigma &= \frac{1}{sT} \ln \left\{ (\Sigma T + 1)^2 + (\Omega T)^2 \right\}, \\
\omega &= \frac{1}{T} \tan^{-1} \left( \frac{\Omega T}{\Sigma T + 1} \right).
\end{align*}
\]

Figure 3 shows the $\varepsilon$-domain (discrete-time) and the corresponding $s$-domain (continuous-time). The primary strip in the $s$-domain covers the entire $\varepsilon$-domain, where the negative portion is mapped to the inside of the circle and the positive to the outside. Other strips in the $s$-domain are piled onto the same $\varepsilon$-domain. It can be seen that the real axis of the $\varepsilon$-domain are divided into four sections depending on the value of $\mu_I$. It should be noted that the real section $(-\frac{1}{T}, \infty)$ in the $\varepsilon$-domain corresponds to the entire real axis in the
Fig. 3. Correspondence of fast-mode poles in the discrete and continuous time domains.

Fig. 4. Real part of the discrete-time fast-mode pole as a function of $\mu_f$ ($T = 0.05\, s$).

$s$-domain, while the real section $(-\infty, -\frac{1}{T})$ in the $\varepsilon$-domain corresponds to the complex conjugate pairs along the boundary lines $\pm \frac{T}{2\mu_f}$, $l \in \mathbb{N}$ in the $s$-domain.

The mapping discretization maps the infinite continuous-time poles that corresponds to fast modes into the finite discrete-time ones at $\varepsilon_P = -\frac{1}{T\mu_f}$, which correspond to the continuous-time poles located for $\mu_f \neq 0$ at

$$
\sigma_P = \frac{1}{2T} \ln \left( \frac{\mu_f - 1}{\mu_f} \right)^2,
$$

(40)

$$
\omega_P = \begin{cases} 
\frac{2n}{T} \pi, & \mu < 0, 1 \leq \mu, \\
\frac{2n+1}{T} \pi, & 0 < \mu < 1,
\end{cases} \quad (n = 0, \pm 1, \pm 2, \ldots).
$$

(41)

Figure 4 shows the real part of the pole for $T = 0.05\, s$. It can be seen that the real part is negative when $\mu_f > \frac{1}{2}$ and, thus, the poles are stable.
3.5. Determination of Parameter $\mu_T$
Performances of the proposed model are affected by the choice of parameters $T$ and $\mu_T$. Of these, the former is restricted by such factors as the frequency components of the signals involved and often by constraints on hardware. The latter is practically the only design parameter and since there is no general guidelines on the choice of this parameter at present as far as we are aware, some of its effects are studied in the following.

Figure 5 shows the region from which the two parameters should be chosen, as explained below.

- In the proposed discrete-time model, an infinite pole of the continuous-time system is mapped into a finite pole located at $-\frac{1}{T\mu_T}$; i.e., the differentiator is mapped into $\frac{x}{T(T\mu_T+1)}$, so that the corner frequency at infinity is mapped to $\frac{1}{T\mu_T}$ rad/s. This sets an upper limit to the product $T\mu_T$. Denoting the maximum frequency of interests by $\omega_t$ rad/s, the product should satisfy $T\mu_T < \frac{1}{\ell_1\omega_t}$, where $\ell_1$ is usually set to be at least 1, and typically between 1 and 10. The product $T\mu_T$ plays a role of the time constant.

- There is also a limit set by the sampling theorem, as $\frac{2\pi}{T} > 2\omega_t$. In practice, where on-line processing is required, the factor 2 on the right-hand-side of the above inequality is inadequate and usually chosen.
to be a larger value, so that the sampling period is chosen to satisfy $T < \frac{\pi}{\ell_2\omega}$, where $\ell_2$ is larger than unity, typically between 5 and 10.

- The two items listed above also imply that the corner frequency $\frac{1}{T\mu_f}$ should be larger than or equal to the Nyquist frequency, $\frac{\pi}{\ell_2T}$. This yields $\frac{\pi}{\ell_2T} \leq \frac{1}{T\mu_f}$ and leads to $\mu_f \leq \frac{\ell_2}{\ell_1}$.

- The intersection of the vertical line set in the second item above and the horizontal line in the third item moves along the hyperbola set in the first item. It should be emphasized that the intersection of the hyperbola and the vertical line is determined from the choices of the time-constant and the sampling period. The horizontal line should therefore be considered as a consequence of these two requirements.

- In addition, the stability condition explained earlier requires $\mu_f > \frac{1}{2}$, which is a hard limit.

- It should be warned that for the highlighted regions in Fig. 5 to exist, so that a proper $\mu_f$ can be chosen, $\frac{1}{2} < \mu_f \leq \frac{\ell_2}{\ell_1}$ must be satisfied. Therefore, the ratio $\frac{\ell_2}{\ell_1}$ should be larger than $\frac{\pi}{2}$.

Figure 6 shows how the $\mu_f$-versus-$T$ plot changes as $\ell_1$ changes, while $\ell_2$ is kept constant. It can be seen that, using the same sampling interval, a smaller value has to be chosen for $\mu_f$ as $\ell_1$ increases. Figure 7 shows the change to $\ell_2$ with $\ell_1$ unchanged. As $\ell_2$ becomes larger, the vertical line $T = \frac{\pi}{\ell_2\omega}$ moves to the left, making $T$ smaller but $\mu_f$ larger. As $T$ becomes smaller, $\mu_f$ can be chosen to a larger value. In the method of [13, 14] and [15, 16], the value of $\mu_f$ is fixed at unity.

4. EXAMPLE

4.1. The Circuit Model

To illustrate the technique described so far, a simple circuit model shown in Fig. 8 is considered. The system can be modeled by the following descriptor form:

$$
\begin{bmatrix}
1 & 0 \\
0 & 0
\end{bmatrix}
\frac{d}{dt}
\begin{bmatrix}
\tilde{q} \\
\tilde{i}
\end{bmatrix}
= \begin{bmatrix}
0 & 1 \\
-1 & 0
\end{bmatrix}
\begin{bmatrix}
\tilde{q} \\
\tilde{i}
\end{bmatrix}
+ \begin{bmatrix}
0 & \tilde{C}
\end{bmatrix}
\hat{u}(t),
$$

$$
\begin{bmatrix}
\tilde{q}(0-) \\
\tilde{i}(0-)
\end{bmatrix},
$$

(42)

where $\tilde{q}$ is the electric charge, $\tilde{i}$ the current, $\hat{u}$ the input voltage, $\tilde{C}$ the capacitance ($\tilde{C} = 100\mu F$), and the descriptor vector is

$$
\tilde{x}' = \begin{bmatrix}
\tilde{q} \\
\tilde{i}
\end{bmatrix}^T.
$$

(43)
The system is regular and has nilpotency of \( h = 2 \), i.e., has no exponential mode, so that its Kronecker form consists only of the fast mode, as

\[
\begin{bmatrix}
  0 & 1 \\
  0 & 0 \\
\end{bmatrix}
\frac{d}{dt}
\begin{bmatrix}
  \bar{i} \\
  \bar{q} \\
\end{bmatrix}
= \begin{bmatrix}
  \bar{i} \\
  \bar{q} \\
\end{bmatrix}
+ \begin{bmatrix}
  0 \\
  -\bar{C} \\
\end{bmatrix} \bar{u}(t),
\]

(44)

where the new descriptor vector is

\[
\bar{x} = \bar{x}_f = \begin{bmatrix}
  \bar{i} \\
  \bar{q} \\
\end{bmatrix}^T.
\]

(45)

The first element in this vector is an impulsive mode and the second a static mode. The original and new descriptor vectors are related by

\[
\bar{x}' = \begin{bmatrix}
  0 & 1 \\
  1 & 0 \\
\end{bmatrix} \bar{x}.
\]

(46)

The response under the consistent initial condition is given by

\[
\bar{x}_f(t) = -\sum_{i=0}^{1} N_i \bar{B}_i \bar{u}^{(i)}(t) = \begin{bmatrix}
  \bar{C} \bar{u}^{(1)}(t) \\
  \bar{C} \bar{u}(t) \\
\end{bmatrix}.
\]

(47)

4.2. The Mapping Model

The mapping model of Eq. (44) is given by

\[
\begin{bmatrix}
  -T \mu_f & 1 \\
  0 & -T \mu_f \\
\end{bmatrix} \delta \mathbf{v}(k, T) = \mathbf{v}(k, T) + \begin{bmatrix}
  0 \\
  -\bar{C} \\
\end{bmatrix} u(k, T),
\]

(48)

or, in the simulation form (the state-space form), by

\[
\delta \mathbf{v}(k, T) = \begin{bmatrix}
  -\frac{1}{T \mu_f} - \left( \frac{1}{T \mu_f} \right)^2 \\
  0 \\
\end{bmatrix} \left\{ \mathbf{v}(k, T) + \begin{bmatrix}
  0 \\
  -\bar{C} \\
\end{bmatrix} u(k, T) \right\},
\]

(49)

and

\[
\mathbf{x}(k, T) = \begin{bmatrix}
  -\frac{1}{T \mu_f} - \left( \frac{1}{T \mu_f} \right)^2 \\
  0 \\
\end{bmatrix} \left\{ \begin{bmatrix}
  0 & 1 \\
  0 & 0 \\
\end{bmatrix} \mathbf{v}(k, T) + T \mu_f \begin{bmatrix}
  0 \\
  -\bar{C} \\
\end{bmatrix} u(k, T) \right\},
\]

(50)

where all of its eigenvalues are clustered at \(-\frac{1}{T \mu_f}\).
4.3. Sinusoidal Input

Let us consider the input signal given by

\[ \tilde{u}(t) = \sin(2\pi t), \quad -\infty < t < \infty, \]  

which is continuously differentiable at least for \( h - 1 \) times and is different from \( 1(t) \sin(2\pi t) \), where \( 1(t) \) is the unit step function. Let us also assume that the simulation be started at an arbitrary initial time \( t = 0 \). To achieve this, the consistent response is needed, which is given by

\[ \tilde{x}_f(t) = \begin{bmatrix} 2\pi \tilde{C} \cos(2\pi t) \\ \tilde{C} \sin(2\pi t) \end{bmatrix}, \]

so that the consistent initial condition is found to be

\[ \tilde{x}_f(0-) = \begin{bmatrix} 2\pi \tilde{C} \\ 0 \end{bmatrix}. \]

Simulations are carried out using Simulink for \( \tilde{C} = 100\mu F \) under different conditions. The response of the discrete-time system is shown as the output through the zero-order-hold. Figures 9 to 15 show those for different values of \( \mu_f \) using the same \( T = 0.05 \) s. In all cases, \( x_2(k, T) \) is the exact discretization, giving the identical response as that shown in Fig. 9, and the plots for other cases are omitted. It can be seen that the mapping model is stable for \( \mu_f > \frac{1}{2} \) (Figs. 9–12) and unstable for \( \mu_f < \frac{1}{2} \) (Fig. 13–15). The response \( x_1(k, T) \) in Fig. 9 shows that there is a transient stage that subsides within a second, after which the steady-state appears. For \( \mu_f \) greater than 1, there is a small phase-shift and amplitude attenuation. For \( \mu_f \) between \( 1/2 \) and 1, the response seems to be fairly close to the continuous-time response. When \( \mu_f \) is reduced below \( 1/2 \) but positive, the response becomes oscillatory and divergent. If it is reduced further to negative values, the response diverges without oscillation. In contrast, \( x_2 \) is exact at the sampling instants for any sampling period, irrespective of \( \mu_f \).

Figures 16 to 18 show the responses of the model started from \( v(0, T) = \tilde{x}_f(0-) \), which makes the initial condition \( x(0, T) \) inconsistent. In this case, as discussed in Section 3.2, there will be a transient response. When \( \mu_f > 1 \), the response catches up the continuous-time response without oscillation (Fig. 16), while the response overshoots and oscillates when \( \mu_f < 1 \) (Fig. 18). When \( \mu_f = 1 \), the transient response disappears in \( k = h - 1 \) steps (Fig. 17).
Fig. 10. State responses $x_1$ for $\mu_f = 2$ at $T = 0.05$ s.

Fig. 11. State responses $x_1$ for $\mu_f = 1$ at $T = 0.05$ s.

Fig. 12. State responses $x_1$ for $\mu_f = 0.7$ at $T = 0.05$ s.

Fig. 13. State responses $x_1$ for $\mu_f = 0.4$ at $T = 0.05$ s.

Fig. 14. State responses $x_1$ for $\mu_f = -0.1$ at $T = 0.05$ s.

Fig. 15. State responses $x_1$ for $\mu_f = -10$ at $T = 0.05$ s.
Fig. 16. State responses $x_1$ for $\mu_f = 2$ at $T = 0.05$ s.

Fig. 17. State responses $x_1$ for $\mu_f = 1$ at $T = 0.05$ s.

Fig. 18. State responses $x_1$ for $\mu_f = 0.7$ at $T = 0.05$ s.

Fig. 19. State responses $x_1$ for $\mu_f T = 0.05$ s.

Fig. 20. State responses $x_1$ for $\mu_f T = 0.05$ s.

Fig. 21. State responses $x_1$ for $\mu_f T = 0.05$ s.
Figures 11 and 19 to 21 show the results where the time constant $T\mu_f$ is constant at 0.05 for $\mu_f$ of 5, 2, 1, and 0.7, and started from the consistent initial condition of Eq. (53). They all have a similar speed of response, while the shape is smoother using a smaller sampling period. However, a closer look at these plots reveals that the values at the sampling instants are actually closer for a larger value of $T$. A possible reason for this will be stated shortly.

Figure 22 illustrates the choices of parameters $T$ and $\mu_f$ used for simulations, for which the model is stable, for the case of $\ell_1 = 1$ and $\ell_2 = 5$. The choice of $\ell_1 = 1$ implies that the corner frequency of the differentiator is the same as the input frequency, while $\ell_2 = 5$ means that the sampling frequency is 5 times faster than the Nyquist frequency. The range of suitable choices is indicated by a rectangle with its borders emphasized with hatching.
In Fig. 9, the value of $T \mu_f$ is larger than the constant hyperbola curve of $\frac{1}{2\pi}$. Therefore, although $T$ satisfies the sampling theorem, its performance is not acceptable. For Fig. 10, $T \mu_f$ is smaller than $\frac{1}{2\pi}$ but $\mu_f$ is larger than $\frac{\ell_2}{\ell_1}$. Although better than Fig. 9, its performance is not acceptable either.

Figures 19, 20, 11, and 21 compare responses of models whose time-constant $T \mu_f$ are identical. As hinted earlier, among these models, ones given in Fig. 11 and 21 are better in the sense that they give closer values at the sampling instants.

While a fixed sampling period of $T = 0.05$ s is used in all the simulations shown above, it was observed, although not shown here, that the discrete-time response of the stable mapping model becomes closer to the continuous-time response as the sampling period becomes smaller.

5. CONCLUSION

For a regular descriptor-system expressed in Kronecker form with an initial condition that is consistent with the input, a class of discrete-time models, called the generalized mapping model, has been presented and shown to be valid models in the sense of discretization defined in the present paper, but missing in our previous work [17]. Furthermore, the longstanding issue of proper initial conditions for discrete-time descriptor systems has been investigated and a simple solution presented. With the initial condition satisfied, the discrete-time model ensures that no transient responses that do not correspond to the continuous-time responses can occur. The choice of a proper initial condition is even more critical for a nonlinear system, whose response can stray into a different equilibrium point, thus drastically changing the response. The mapping model, which was proposed in our previous study [17] and includes the existing methods [13–16] as special cases, may be tuned for a particular need by adjusting the value of design parameter $\mu_f$. Therefore, a suitable region of values for $\mu_f$ and $T$ has been investigated in the present paper. It was found that the lower bound is set by the stability condition and the upper bound basically by the maximum frequency of interests in the input signal $\omega_t$, with some room to play using the integers $\ell_1$ and $\ell_2$. With such an insight, these parameters may be chosen on an educated trials-and-errors basis rather than a blind guess work.

While the fast-mode portion of a continuous-time descriptor-system is nonproper and cannot be implemented exactly, its mapping model can always be made bi-proper and implemented in discrete-time form. This is achieved by converting a singular matrix $\bar{E}$ into a non-singular $\bar{E} - T M \bar{A}$, where $T M$ can be zero for the slow mode but should be non-zero for the fast mode. In this manner, the fast mode of the mapping model transforms the infinite continuous-time poles into the finite discrete-time ones. Therefore, $\mu_s$ and $\mu_f$ should be chosen differently in general. This is easy to do for the Kronecker form where the slow and the fast modes are decoupled. Once the response is obtained for the Kronecker form, the original descriptor vector can be recovered by a matrix transformation as needed.

REFERENCES