Fast Step-response Evaluation of Linear Continuous-time Systems With Time Delay in the Feedback Loop

Yves Piguet * Philippe Mullhaupt **

* Calerga Sàrl, Av. de la Chablière 35, 1004 Lausanne, Switzerland (e-mail: yves.piguet@calerga.com).
** Laboratoire d’automatique, EPFL, 1015 Lausanne, Switzerland (e-mail: philippe.muellhaupt@epfl.ch).

Abstract: The simulation of step, impulse and ramp responses of linear continuous-time time-invariant single-input single-output systems with a pure time delay in the feedback loop is tackled. Contrary to systems without time delay where a conversion from Laplace transform to z transform with a zero-order hold allows for an exact response, delay in the feedback loop requires either an unbounded number of states or approximations. This paper proposes an approximation which fulfills a set of criteria, including exactness of the response over a specified multiple of the delay and the conservation of the stability or lack of stability. The approach is illustrated with examples.

1. INTRODUCTION

Continuous-time systems with time delays in the feedback loop are difficult to simulate, because they have an infinite number of states. They have both theoretical and practical importance, though. Time delays are simple to model and occur in many systems: properties of fluids transported in a pipe, such as temperature or concentrations, or electromagnetic emissions, exhibit a delay which is directly related to their speed and path length.

Related research has focused mainly on numerical solutions of delay differential equations (DDE). In Shampine and Thomson [2000], the implementation of function dde23 provided with MATLAB since version 6.5 is described; delayed states are interpolated from past results with cubic Hermite interpolation and discontinuities are propagated. In Simulink, time delays (“transport delay” blocks) are implemented as buffers of past samples. Output is obtained with linear interpolation, or extrapolation when the value of the delay is smaller than the integration time step.

In MATLAB’s control toolbox, objects representing linear time-invariant systems with delays cannot be connected in a feedback configuration; see The Mathworks, Inc. [2002]. The Time Delay System Toolbox for MATLAB, described in Kim et al. [2000], is a set of functions for simulation, design and analysis of systems with delays; its DDE functions for nonlinear and linear systems are both based on Kim and Pimenov [August 1997] which proposes a Runge-Kutta-like approach with fixed integration step.

This paper deals with the simulation of the step response of systems with a single time delay. It presents a method based on the Laplace and z transforms which exploits the system linearity. Impulse responses and ramp responses can be deduced simply by adding a derivator or an integrator to the transfer function (factor s or 1/s, respectively). Particular attention is paid to the amount of computation. Reducing the simulation time step to assure adequate accuracy, or multiple iterations over the whole simulation time span, should be avoided. The simulation time step should depend only on requirements for truthful display or further analysis. One of the goals is to permit interactive applications with software such as Sysquake, where the response is updated continuously when the user manipulates parameters with the mouse; see Piguet [2004].

The remainder of this paper is organized as follows. Section 2 enumerates the specifications which are considered as important. Section 3 defines the model and the kind of systems it can describes. Section 4 describes how the step response is computed in order to satisfy the specifications. Section 5 illustrates the method with different examples. Finally, Section 6 summarizes the results and gives possible extensions.

2. SPECIFICATIONS

The purpose of the simulation is to provide a response which matches the true response as well as possible both qualitatively and quantitatively. This includes the following characteristics:

- *Stability preservation.* The simulation should exhibit a diverging response if and only if the continuous-time system with delay is unstable.
- *Causality.* If the whole response is delayed, so should be the simulation, by the same value.
- *No smoothing.* This is especially important where the response or its derivative are discontinuous, typically for \( t = 0 \) and \( t = d \).
- *No artifact such as Gibbs effect.* Reducing the time step used for simulation should make the simulated response converge to the true response of the system.
• Approximation possible at any time scale. The simulation should be possible over time spans between fractions to large multiples of the delay.
• Error on each sample as small as possible. Among other quantities, overshoot should be preserved.
• Fast to compute. Simulation should allow to update continuously the step response when parameters are changed by the user in an interactive application.

3. MODEL

The following system is considered, in the domain of the Laplace transform:

**Definition 1.** A single-delay causal single-input single-output linear time-invariant continuous-time system is obtained by connecting the following sub-systems:

- a single time delay \( p_1(s) = e^{-ds} \) where \( d > 0 \) is the amount of time;
- causal rational transfer functions \( p_i(s), i = 2, \ldots, n \).

Sub-systems are connected in series (the input of a sub-system is equal to the output of another sub-system), in parallel (the outputs of multiple sub-systems are added), or with feedback (the output of one or multiple sub-systems are used in a direct or indirect way as their own inputs). A single exogenous signal \( U(s) \) enters the system (system input), and a single sub-system output \( Y(s) \) is considered (system output). The output and all sub-systems are connected directly or indirectly to the system input.

This system covers all responses of interest which occur in a single-input single-output system described by the serial connection of a time delay and a rational transfer function, controlled with a one- or two-degrees-of-freedom controller described by rational functions such as a PID (proportional-integral-derivative) controller.

**Theorem 2.** All single-delay causal single-input single-output linear time-invariant continuous-time systems between input \( U(s) \) and output \( Y(s) \) can be represented by the following transfer function:

\[
G(s) = \frac{Y(s)}{U(s)} = \frac{A(s) + B(s)e^{-ds}}{C(s) + D(s)e^{-ds}}
\]  

(1)

where \( A(s), B(s), C(s) \) and \( D(s) \) are polynomials of \( s \); and \( A(s)/C(s), B(s)/C(s), \) and \( D(s)/C(s) \) are all causal.

**Proof.** All single-delay causal single-input single-output linear time-invariant continuous-time systems can be written as

\[
Y(s) = G_{11}(s)U(s) + G_{12}(s)Y_d(s)
\]

\[
U_d(s) = G_{21}(s)U(s) + G_{22}(s)Y_d(s)
\]

where \( G_{ij}(s) \) are causal rational transfer functions, \( U_d(s) \) the input of the time delay and \( Y_d(s) = e^{-ds}U_d(s) \) its output. \( G(s) \) is obtained by eliminating \( U_d(s) \) and \( Y_d(s):\)

\[
G(s) = \frac{G_{11}(s) + (G_{12}(s)G_{21}(s) - G_{11}(s)G_{22}(s))e^{-ds}}{1 - G_{22}(s)e^{-ds}}
\]

This is equivalent to (1) with

![System without delay in direct path](image)

Fig. 1. System without delay in direct path.

\[
\frac{A(s)}{C(s)} = G_{11}(s)
\]

\[
\frac{A(s)}{C(s)} = G_{12}(s)G_{21}(s) - G_{11}(s)G_{22}(s)
\]

\[
\frac{A(s)}{C(s)} = -G_{22}(s)
\]

All transfer functions are causal.

Depending on the input and output, \( A(s) \) or \( B(s) \), and \( D(s) \), can be zero. The case where \( A(s) = B(s) = 0 \) is degenerated and not interesting. The case where \( C(s) \) is zero corresponds to a non-causal system where the output occurs with an advance of time \( d \) with respect to the input, which cannot occur in the system as defined above.

The case where \( D(s) \) is zero corresponds to the absence of delay in the feedback path; the response of \( G(s) \) is obtained by summing the response of \( A(s)/C(s) \) and the delayed response of \( B(s)/C(s) \). A single discrete-time transfer function is obtained by choosing a sampling period \( h \) such that the time delay is an integer multiple of \( h \) and by converting separately \( A(s)/C(s) \) and \( B(s)/C(s) \) with zero-order hold.

The remaining part of this paper will tackle the case where at most one of \( A(s) \) and \( B(s) \) is zero.

4. STEP RESPONSE

It is assumed that step responses of rational transfer functions can be computed without error. One method is to convert the Laplace-transform continuous-time transfer function to a z-transform discrete-time transfer function with the zero-order hold method, which requires to calculate the exponential of a matrix. Matrix exponentials can be evaluated numerically with arbitrary precision (see Moler and Loan [2003]). The discrete-time step response samples match exactly the continuous-time step response at the corresponding time instants.

Transfer function \( G(s) \) links input \( U(s) \) to output \( Y(s) \) with a direct path, delayed or not depending whether the time delay is in the feedback path or not. Fig. 4 shows the case without delay in the direct path, i.e. where \( B(s) = 0 \), and Fig. 4 the case with delay in the direct path, with \( A(s) = 0 \). The feedback loop adds a new term for every \( t = kd \), with \( k \in \mathbb{N}^+ \), which is filtered by \( L(s)^k \), where \( L(s) \) is the loop transfer function. It is possible to compute the exact response by summing these terms, simulated with zero-order hold conversion to discrete-time which gives exact results for step responses; however, the difficulty arises from the increasing degree of powers of \( L(s) \). The main idea is to simulate exactly the response over a few multiples of the delay (one is often sufficient).
and to approximate the subsequent effect of the feedback loop.

4.1 Partial decomposition

Eq. (1) can be seen as the sum of two components \( G_1(s) \) and \( G_2(s)e^{-ds} \), where \( G_1(s) \) has a non-delayed path from input to output, and \( G_2(s)e^{-ds} \) does not:

\[
G(s) = G_1(s) + G_2(s)e^{-ds}
\]  

(2)

For the sake of simplicity, it is assumed in this section that \( G(s) = G_1(s) \), i.e. that \( B(s) = 0 \). For the general case, \( G_1(s) \) and \( G_2(s) \) are handled separately.

The following theorem shows how \( G(s) \) can be decomposed as the sum of \( m \) transfer functions with delays in the numerator, and a transfer function with a delay of \( md \).

Theorem 3. Let

\[
G(s) = \frac{A(s)}{C(s)} e^{-ds} + \frac{D(s)}{C(s)} e^{-ds}
\]  

(3)

For any \( m \in \mathbb{N}^* \),

\[
G(s) = \sum_{k=0}^{m-1} \frac{A(s)}{C(s)} \left( -\frac{D(s)}{C(s)} \right)^k e^{-kd(s)} + \left( -\frac{D(s)}{C(s)} \right)^m \frac{A(s)e^{-mds}}{C(s) + D(s)e^{-ds}}
\]  

(4)

Proof. For \( m = 1 \), (4) becomes

\[
G(s) = \frac{A(s)}{C(s)} e^{-ds} - \frac{D(s)}{C(s)} \frac{A(s)e^{-ds}}{C(s) + D(s)e^{-ds}}
\]  

(5)

which is readily checked by subtracting \( A(s)/C(s) \) from (3). Proof for \( m > 1 \) proceeds recursively. Assume (4) is true for \( m - 1 \):

\[
G(s) = \sum_{k=0}^{m-2} \frac{A(s)}{C(s)} \left( -\frac{D(s)}{C(s)} \right)^k e^{-kd(s)} + \left( -\frac{D(s)}{C(s)} \right)^{m-1} \frac{A(s)e^{-(m-1)d(s)}}{C(s) + D(s)e^{-ds}}
\]

\[
= \sum_{k=0}^{m-2} \frac{A(s)}{C(s)} \left( -\frac{D(s)}{C(s)} \right)^k e^{-kd(s)} + \left( -\frac{D(s)}{C(s)} \right)^{m-1} G(s)e^{-(m-1)d(s)}
\]

Substituting \( G(s) \) by its value given by (5) yields

\[
G(s) = \sum_{k=0}^{m-2} \frac{A(s)}{C(s)} \left( -\frac{D(s)}{C(s)} \right)^k e^{-kd(s)} + \frac{A(s)}{C(s)} \left( -\frac{D(s)}{C(s)} \right)^{m-1} e^{-(m-1)d(s)}
\]

\[
+ \left( -\frac{D(s)}{C(s)} \right)^m \frac{A(s)e^{-mds}}{C(s) + D(s)e^{-ds}}
\]

which is equivalent to (4). \( \Diamond \)

Remarks:

- This decomposition is a particular case of polynomial long division, where the last term of (4) is the remainder.
- The number of terms in the sum can be arbitrarily large; when it becomes infinite,

\[
G(s) = \sum_{k=0}^{\infty} \frac{A(s)}{C(s)} \left( -\frac{D(s)}{C(s)} \right)^k e^{-kd(s)}
\]

However, this series cannot be used directly, because the term degree grows quickly and prevents numerical evaluation.

4.2 Discrete-time rational transfer function

For simulation, the continuous-time transfer function in the Laplace domain must be converted to a discrete-time rational transfer function in the z domain.

Sampling period \( h \) is chosen such that \( d/h \) is an integer:

\[
n = \frac{d}{h} \in \mathbb{N}^*
\]  

(6)

Then \( e^{-ds} \) is converted to \( z^{-n} \).

Methods used to perform the conversion from Laplace transform to z transform have an effect on the result. Two of them are considered here: zero-order hold, which gives an exact response if the input is a step; and bilinear transform, which is preferred if the input is continuous, e.g. if the signal has been filtered by low-pass transfer functions. With a time delay in the feedback loop, it is impossible to obtain an exact result with rational functions of finite degree.

In (4), terms of the sum are rational transfer functions with delays in the direct path; since they are applied to an input step \( U(s) \), they can be converted with zero-order hold without approximation. The last term contains a delay in the feedback path. It can be rewritten as the serial connection of a pure delay, a rational transfer function, and a pure feedback block with unit direct path and a causal transfer function with delay in the feedback path:

\[
\left( -\frac{D(s)}{C(s)} \right)^m \frac{A(s)e^{-mds}}{C(s) + D(s)e^{-ds}} = e^{-mds} (-1)^m A(s)D_m(s) F(s)
\]

with

\[
F(s) = \frac{1}{1 + D(s)e^{-ds}/C(s)}
\]  

(7)
All rational transfer functions are converted with zero-order hold, and delays are represented by negative powers of \( z \); this is done without approximation. \( F(s) \), which is not rational, must be approximated.

4.3 Approximation of delayed feedback

Conversion from continuous-time to discrete-time with zero-order hold makes sense only when the input is actually held constant between samples. This is not the case in \( F(s) \), because the feedback signal is filtered recurrently by \( D(s)/C(s) \). The simplest alternative is the plain bilinear method, which is often used when converting a controller designed in continuous time for a digital implementation.

However, bilinear conversion introduces an approximation whose impact on the simulated step response must be evaluated. Reviewing the desired characteristics enumerated in Section 2, two of them are important here, stability and smoothing. A smoothing effect of the approximation of \( F(s) \) would be acceptable, because \( F \) is applied only to a signal already filtered by \( A(s)D^n(s)/C^{n+1}(s) \).

Stability (or instability) conservation is a different matter. The stability or instability property of the delayed feedback is entirely captured in \( F(s) \), not by other terms of (4). The theorem below gives a sufficient condition for conserving the stability or instability of \( G(s) \).

**Theorem 4.** If there is a single frequency \( \omega_s \geq 0 \) such that 
\[ |D(j\omega_s)/C(j\omega_s)| = 1 \]
then transfer function
\[ \tilde{G}(z) = \sum_{k=0}^{m-1} \frac{z^{-1}}{z} \left[ \mathcal{L}^{-1} \frac{A(s)}{sC(s)} \left( \frac{D(s)}{C(s)} \right)^k \right] z^{-kd/h} + \frac{z^{-1}}{z} \left[ \mathcal{L}^{-1} \frac{A(s)D^n(s)}{sC^{n+1}(s)} \right] \]
\[ 1 + \frac{z^{-d/h}}{(D(s)/C(s))} \left( \frac{z^{-1}}{z} \right)^{k+1} \]
where \( \mathcal{L}^{-1} \) denotes the inverse Laplace transform and \( Z \) the \( z \) transform, is a discrete-time approximation of \( G(s) \) which is stable if and only if \( G(s) \) is stable.

**Proof.** Considering the Nyquist criterion, one way to maintain the stability or instability is to use the bilinear conversion with prewarping at the crossover frequency \( \omega_s \) to convert the rational part of the open-loop transfer function, \( D(s)/C(s) \). This guarantees that the phase introduced by the delay at the crossover frequency is the same in continuous time as in discrete time. Hence the number of encirclements of \(-1\) is the same since the frequency response crosses the unit circle only once. \( \diamond \)

**Remark:** the result still holds when the open-loop frequency response crosses the unit circle multiple times provided that other crossings of the unit circle are far enough from \(-1\).

Bilinear conversion with prewarping at \( \omega_s \) maps \( D(s)/C(s) \) to \( D_{bp}(z)/C_{bp}(z) \):
\[ \frac{D_{bp}(z)}{C_{bp}(z)} = \frac{D(s)}{C(s)} \left( \frac{\omega_s}{\tan(\omega_s h/2)} \right) \frac{z^{-1}}{z^{1/2}+1} \]
Polynomials \( D_{bp}(z) \) and \( C_{bp}(z) \) are obtained by canceling denominators appearing in the right-hand term.

Cases where \( |D(j\omega)/C(j\omega)| \neq 1 \) for any real \( \omega \) correspond to degenerate cases where the bilinear transform without prewarping should be used:
\[ \frac{D_{bp}(z)}{C_{bp}(z)} = \frac{D(\frac{x}{h} \cdot \frac{z}{z-1})}{C(\frac{x}{h} \cdot \frac{z}{z-1})} \]

Since there is no intersection between the Nyquist diagram and the unit circle, the error in the phase shift produced by the delay cannot stabilize or destabilize the system when it is converted to discrete time.

4.4 Exact response over delay length

The method described above gives an exact response for \( t \leq md \) and a good approximation for \( t > md \). It is often sufficient to have \( m = 1 \): in the worst case where the numerator’s degree is one less than the denominator’s degree, the discontinuity of the step response first derivative is conserved. This section gives the discrete-time transfer function for \( m = 1 \) corresponding to the general form of (1) where both \( A(s) \) and \( B(s) \) can be non-zero.

Eq. (1), (2) and (7) give
\[ G_1(s) = \frac{A(s)}{C(s)} F(s) \]
\[ G_2(s) = \frac{B(s)}{C(s)} F(s) \]
Transfer function \( F(s) \) is the effect of the delayed feedback. It should be noted that its step response is 1 for \( 0 \leq t < d \).

Rational transfer functions \( A(s)/C(s) \), \( B(s)/C(s) \), and \( D(s)/C(s) \) are converted to discrete-time transfer functions in \( z \) with zero-order hold or bilinear mapping with prewarping as explained above (the frequency \( \omega_s \) used for prewarping is the crossover frequency of the open-loop transfer function associated with \( F(s) \), i.e. \( D(s)/C(s) \), which is common to \( G_1(s) \) and \( G_2(s) \)), and the step response of \( G(s) \) is computed with difference equations.

Let
\[ \frac{A_{zoh}(z)}{C_{zoh}(z)} = \frac{z-1}{z} \left[ \mathcal{L}^{-1} \frac{A(s)}{sC(s)} \right] \]
\[ \frac{B_{zoh}(z)}{C_{zoh}(z)} = \frac{z-1}{z} \left[ \mathcal{L}^{-1} \frac{B(s)}{sC(s)} \right] \]
The poles of the zero-order hold conversion of \( A(s)/C(s) \) and \( B(s)/C(s) \) are the same, hence the common denominator \( C_{zoh}(z) \). Then \( H(z) \), the discrete-time transfer function used to simulate the step response of \( G(s) \), is
\[ H(z) = z^n \frac{A_{zoh}(z) + B_{zoh}(z)}{C_{zoh}(z)} \cdot \frac{C_{bp}(z)}{z^n C_{bp}(z) + D_{bp}(z)} \]

5. EXAMPLES

Examples illustrate the properties of the step response approximation.
5.1 First-order system with time delay in the direct path

The step response of the following delayed first-order system with unit negative feedback is computed.

\[ P(s) = \frac{2e^{-ds}}{s+1} \]

The closed-loop transfer function is

\[ G(s) = \frac{2e^{-ds}}{s+1+2e^{-ds}} \]

Matching (1) gives

\[ \frac{A(s)}{C(s)} = 0, \quad \frac{B(s)}{C(s)} = \frac{D(s)}{C(s)} = \frac{2}{s+1} \]

The sampling period is chosen as \( h = 0.2 \), on purpose large with respect to the time constant of \( P(s) \). \( B(s)/C(s) \) is converted to discrete-time with zero-order hold and \( D(s)/C(s) \) with bilinear transform:

\[ \frac{B_{zoh}(z)}{C_{zoh}(z)} = \frac{0.181}{z-0.819} \]

\[ \omega_n = 1.732 \]

\[ \frac{D_{bp}(z)}{C_{bp}(z)} = \frac{0.183z + 0.183}{z - 0.817} \]

Then (8) gives

\[ H(z) = \frac{0.181}{z - 0.819} \frac{z - 0.817}{z^n(z - 0.8165) + 0.183z + 0.183} \]

Its step response with \( n = 5 \), which approximates the step response of \( G(s) \) with \( d = 1 \), is shown in Fig. 3. While the slow sampling is noticeable, the response starts exactly at \( t = 1 \) and is exactly a first-order response for \( 1 < t < 2 \).

To evaluate whether the arbitrary choice of \( h \) has an important effect, Fig. 4 shows the error between the response simulated with \( h = 0.2 \) and a reference response simulated with \( h = 0.02 \), i.e. a sampling frequency multiplied by 10 (solid line). The maximum error is 0.02, i.e. a smoothing is observed at \( t = 1 \). Fig. 4 shows also the error of the response obtained with Simulink in dash line. A noticeable error in the effect of this loop, i.e. oscillations, which have a smaller damping; zero-order hold introduces an additional delay, so this is not surprising. So the choice of using different methods of conversions from continuous-time to discrete time is validated by this example.

5.2 Effect of conversion methods

Was the choice of zero-order hold conversion and bilinear transform with prewarping sound? A naive approach would consist in using the zero-order hold conversion everywhere. Fig. 6 shows the response with zero-order hold conversion for all transfer functions. The use of zero-order hold for the feedback loop results in an important error in the effect of this loop, i.e. oscillations, which have a smaller damping; zero-order hold introduces an additional delay, so this is not surprising. So the choice of using different methods of conversions from continuous-time to discrete time is validated by this example.
5.3 Impulse response

While the approach described above provides a solution for the step response, it can readily be adapted to the impulse response case, by multiplying \( G(s) \) by \( s \). Fig. 7 shows the impulse response of \( G(s) \) with \( h = 0.02 \).

5.4 Large time delay

The system of Example 5.1 is simulated with a time delay \( d = 5 \) and a sampling period \( h = 0.2 \) (Fig. 8). The larger time delay makes the system unstable.

6. CONCLUSION

A method for computing the step, impulse and ramp responses of a continuous-time linear time-invariant single-input single-output systems with time delay in the loop has been designed. A general transfer function, which covers all the cases where a single time delay is connected to rational transfer functions, is converted to a discrete-time transfer function with mixed zero-order hold and bilinear with prewarping methods. In its simplest version, the result is exact for \( t \leq d \) (\( t \leq 2d \) if the whole response is delayed by \( d \)) and the stability or instability of the system is conserved; exact response over a longer time range is possible with higher-order transfer functions. Examples validate the approach for different cases, both qualitatively and quantitatively, and show its effectiveness with respect to Simulink.

This method could be extended in at least two directions. First, systems under consideration could include multiple time delays. If these delays are small multiples of a common divisor, a partial fraction expansion leads to the sum of independent feedback loops with delays. Second, linear state-space models with delays could be used, with possible benefits from a numerical point of view.

REFERENCES


