

Exact Discretization of a Scalar Differential Riccati Equation with Constant Parameters

N. Hori¹, C. A. Rabbath², and P. N. Nikiforuk³

1 Digital Control Laboratory, Graduate School of Systems and Information Engineering

University of Tsukuba

1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8573 Japan

Phone: +81-29-853-5139; Fax: +81-29-853-5207

E-mail: hori@kz.tsukuba.ac.jp

2 Defense Research and Development Canada - Valcartier

2459 Pie-XI Blvd. N., Val-Belair, Quebec, G3J 1X5 Canada

Phone: +1-418-844-4000

E-mail: camille-alain.rabbath@drdc-rddc.gc.ca

3 Department of Mechanical Engineering, University of Saskatchewan

57 Campus Drive, Saskatoon, Saskatchewan, S7N 5A9 Canada

Phone: +1-306-966-5440; Fax: +1-306-966-5427

E-mail: peter.nikiforu@usask.ca

Abstract — An exact, first-order, discrete-time model that gives correct values at the sampling instants for any sampling interval is derived for a nonlinear system whose dynamics are governed by a scalar Riccati differential equation with constant parameters. The model is derived by transforming the given differential equation into a stable linear form to which the invariant discretization is applied. This is in contrast with other existing methods which result in a second-order and usually unstable form and which is not suitable for on-line digital control purposes. Simulation results are presented to show that the proposed method is always exact at the sampling instants, whereas the popular forward difference model can be divergent unless the sampling interval is sufficiently small.

I. INTRODUCTION

The discretization of nonlinear systems has been one of the key issues in the implementation of digital control algorithms [1]. For linear systems, a number of useful methods have been developed [2], such as the exact discretization where the output of a discrete-time system matches that of a continuous-time system at the sampling instants for any sampling interval. Such models are exact for staircase types of inputs, which occur in sampled-data control systems involving zero-order-hold devices. While the exact discretization is well known for linear systems, no such method exists for general nonlinear systems [3]. For simulation purposes, accurate methods [4] can be used to solve a wide variety of nonlinear differential equations, although many of them are not suitable for on-line digital control purposes. However, those that can be implemented for on-line uses, such as the forward-difference model, can cause large errors in their results and, thus, require very small discretization intervals for computations. The accuracy issue in nonlinear cases is a critical one, since no matter how small the discretization interval may be chosen, a control system may not be stabilized [3]. Therefore, while the development of general approximate discretization techniques is very important, developing an exact model for a class of nonlinear systems and trying to expand the applicable class are equally important avenues to pursue.

Riccati differential equations are an extremely important topic in such areas as mathematics [5], physics [6], and engineering [7]. Although they can be solved by converting them into higher-order linear equations [7]-[8], these linear systems are usually unstable and not suitable for on-line control purposes. Moreover, these methods yield a first-order system only for a very restrictive and uninteresting class of Riccati equations. In contrast, the method proposed in [9] always leads to a class of stable, first-order, linear systems without such limitation. The present study is concerned with the generalization of the transformation method given in [9] and its application to the development of exact discrete-time models for a scalar Riccati differential equation with constant parameters. Two simulation examples are provided as well; one for discrete-time modelling of a system and the other for optimal control of a linear system.

II. LINEARIZATION OF RICCATI EQUATION

Consider a system whose dynamics are governed by a Riccati differential equation of the following form:

$$\dot{x} = f(x) = ax^2 + bx + c, \quad x(0) = x_0. \quad (1)$$

Using an appropriate variable transformation, this system can be converted into a linear second-order or first-order system, as follows:

A. Second Order Case

Theorem 1:

Consider the system (1). Introducing a free parameter K , this system can be represented by a linear state equation and a nonlinear output equation of the following form:

$$\begin{bmatrix} \dot{v} \\ \dot{u} \end{bmatrix} = \begin{bmatrix} K & \phi \\ \mu & \psi + K \end{bmatrix} \begin{bmatrix} v \\ u \end{bmatrix}, \quad \begin{bmatrix} v(0) \\ u(0) \end{bmatrix} = \begin{bmatrix} v_0 \\ \frac{\gamma x_0 + \delta}{\alpha x_0 + \beta} v_0 \end{bmatrix} \quad (2)$$

$$x = \frac{\delta v - \beta u}{-\gamma v + \alpha u}, \quad (3)$$

where α, β, γ and δ are arbitrary parameters that satisfy

$$\alpha\delta - \beta\gamma \neq 0 \quad (4)$$

and ϕ, μ and ψ are defined as

$$\begin{aligned} \phi &= \frac{a\beta^2 - b\alpha\beta + c\alpha^2}{\alpha\delta - \beta\gamma} \\ \mu &= -\frac{a\delta^2 - b\delta\gamma + c\gamma^2}{\alpha\delta - \beta\gamma} \\ \psi &= \frac{-b(\alpha\delta + \beta\gamma) + 2(a\beta\delta + c\alpha\gamma)}{\alpha\delta - \beta\gamma}. \end{aligned} \quad (5)$$

Proof: Let us assume the following form of the state equation and determine its coefficients to satisfy the condition:

$$\begin{cases} \dot{v} = \varepsilon_{11}v + \varepsilon_{12}u \\ \dot{u} = \varepsilon_{21}v + \varepsilon_{22}u. \end{cases} \quad (6)$$

Differentiation of eq. (3) and the use of the above relationship yield

$$\begin{aligned} \frac{dx}{dt} &= \frac{(\alpha\delta - \beta\gamma)(u\dot{v} - \dot{u}v)}{(\alpha u - \gamma v)^2} \\ &= \frac{(\alpha\delta - \beta\gamma)\{\varepsilon_{12}u^2 + (\varepsilon_{11} - \varepsilon_{22})uv - \varepsilon_{21}v^2\}}{(\alpha u - \gamma v)^2}. \end{aligned} \quad (7)$$

On the other hand, eqs. (1) and (3) give

$$\begin{aligned}\frac{dx}{dt} &= \frac{a(-\beta u + \delta v)^2 + b(-\beta u + \delta v)(\alpha u - \gamma v) + c(\alpha u - \gamma v)^2}{(\alpha u - \gamma v)^2} \\ &= \frac{(\alpha\delta - \beta\gamma)(\phi u^2 - \psi uv - \mu v^2)}{(\alpha u - \gamma v)^2}.\end{aligned}\quad (8)$$

Equating the corresponding terms of eqs. (7) and (8) yields

$$\varepsilon_{11} = K, \quad \varepsilon_{12} = \phi, \quad \varepsilon_{21} = \mu, \quad \varepsilon_{22} = K + \psi \quad (9)$$

where $\varepsilon_{11} = K$ is chosen here as a free parameter. (E.O.P.)

The fact that the variable x of the nonlinear system can be expressed as an algebraic function of the v and u variables of a linear system is very appealing, since the analytical solution is readily available and its properties are well understood for linear systems. Compared with the linear equations appearing in [7,8], eq. (2) contains a free parameter that can be adjusted to attain its stability. This can avoid the numerical problem in implementing discrete-time models, where a bounded signal is obtained as a ratio of two diverging signals as demonstrated in [9]. Furthermore, the form used in [7] can not be reduced to a first-order equation unless $c = 0$ in eq. (1), which is not useful. These problems are solved in the present study; the following lemma solves the former, while Theorem 2 solves the latter.

Lemma:

Sufficient conditions for the solution x , given by eqs. (2) and (3), to be computable using non-diverging variables v and u , are as follows:

(A) The system given by eq. (1) satisfies the following equilibrium condition:

$$b^2 - 4ac \geq 0. \quad (10)$$

(B) The gain K satisfies the following stability condition:

$$K < -\frac{\psi + \sigma}{2} \quad (11)$$

where

$$\sigma = \sqrt{b^2 - 4ac}. \quad (12)$$

Proof: So that the original variable x can be obtained as a ratio of bounded quantities, v and u , as in eq. (3), the value of κ must be selected such that all the eigenvalues of the system matrix for the linear system (2) have negative real parts. Since the characteristic equation is

$$\begin{aligned} \det \begin{bmatrix} \lambda - K & -\phi \\ -\mu & \lambda - \psi - K \end{bmatrix} \\ = \lambda^2 - (\psi + 2K)\lambda + (K^2 + \psi K - \mu\phi) = 0, \end{aligned} \quad (13)$$

κ must be chosen to satisfy

$$\psi + 2K < 0 \quad (14)$$

and

$$K^2 + \psi K - \mu\phi = \left(K + \frac{\psi}{2}\right)^2 - \left(\frac{\sqrt{\psi^2 + 4\mu\phi}}{2}\right)^2 > 0. \quad (15)$$

Noting that

$$\begin{aligned} & \sqrt{\psi^2 + 4\mu\phi} \\ &= \frac{1}{(\alpha\delta - \beta\gamma)} \left[b^2(\alpha\delta + \beta\gamma)^2 - 4b(\alpha\delta + \beta\gamma)(a\beta\delta + c\alpha\gamma) \right. \\ & \quad \left. + 4(a\beta\delta + c\alpha\gamma)^2 - 4(a\delta^2 - b\delta\gamma + c\gamma^2)(a\beta^2 - b\alpha\beta + c\alpha^2) \right]^{1/2} \\ &= \frac{1}{(\alpha\delta - \beta\gamma)} \sqrt{(b^2 - 4ac)(\alpha\delta - \beta\gamma)^2} \\ &= \sqrt{b^2 - 4ac}, \end{aligned} \quad (16)$$

inequalities (14) and (15) lead to condition (11). Condition (10) is required so that σ is real; i.e., the equilibrium points of eq. (1) are real. (E.O.P)

B. First-Order Case

The variable transformation used in [9] was found basically by trial and error. Using the formulation presented in the previous section, the derivation can be viewed as its special case, such as the following:

Theorem 2:

Consider the Riccati equation given by eq. (1) that satisfies condition (10). By choosing the design parameters to satisfy the conditions given below, eq. (1) can always be expressed as the following class of linear first-order state equations:

$$\dot{v} = -\sigma v + \phi, \quad v(0) = \frac{\alpha x_0 + \beta}{\gamma x_0 + \delta}, \quad (17)$$

where ϕ is as defined in eq. (5). The output equation is given by

$$x = \frac{\delta v - \beta}{-\gamma v + \alpha}. \quad (18)$$

(I) The parameters δ and γ are chosen to satisfy

$$(b - \sigma)\gamma = 2a\delta \quad (19)$$

where

$$\gamma \neq 0, \quad (20)$$

that is, $-\delta/\gamma$ is a constant solution of eq. (1).

(II) The parameters α and β are so chosen that

$$(b - \sigma)\alpha \neq 2a\beta, \quad (21)$$

that is, $-\beta/\alpha$ is not a constant solution of eq. (1).

Proof: First, let $K = -\sigma$. By choosing δ as in eq.(19), it can be shown that $\mu = 0$, the condition (4) becomes (20) and (21), and the following relationship holds:

$$\begin{aligned} \psi &= \frac{-b\{\alpha(b - \sigma) + 2a\beta\} + 2a\beta(b - \sigma) + 4ac\alpha}{\alpha(b - \sigma) - 2a\beta} \\ &= \frac{\alpha b\sigma - (b^2 - 4ac)\alpha - 2a\beta\sigma}{\alpha(b - \sigma) - 2a\beta} \\ &= \sigma. \end{aligned} \quad (22)$$

Thus, eq. (2) reduces to

$$\begin{bmatrix} \dot{v} \\ \dot{u} \end{bmatrix} = \begin{bmatrix} -\sigma & \phi \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v \\ u \end{bmatrix}. \quad (23)$$

By setting $u(0)=1$, it follows that $u \equiv 1$ can be removed from the above dynamics, leaving the first-order equation as in eq. (17). This is equivalent to setting $u = 1$ so that

$$v = \frac{\alpha x + \beta}{\gamma x + \delta}, \quad (24)$$

which gives eq. (18).

(Q.E.D.)

Condition (19) ensures that $\mu = 0$, while eqs. (20) and (21) support (4). Furthermore, condition (19) automatically yields $\psi = \sigma$. Depending on how the parameters in (24) are chosen, a class of linear systems can be obtained, with ϕ in (17) being the difference. It should be emphasized that γ is a free parameter as long as it is nonzero and δ is uniquely determined by γ and the system parameters. The parameters α and β are also arbitrary as long as (21) holds.

It should be emphasized that equation (23) is stable, while the second is only marginally stable. Therefore, the stability condition set forth in Lemma for the second-order system is not satisfied. However, the state of the marginally stable part is constant and excluded from any computation, so that no numerical problem is encountered in the first-order implementation.

III. EXACT DISCRETIZATION OF RICCATI EQUATION

In the following, the delta operator $\delta_T = (q-1)/T$ [2,10], where q is the usual shift-left operator and T is the sampling interval, is used, since it proves to be very convenient when relating discrete-time results to continuous-time results. The first-order linearized form is used below, since it is simpler to implement such systems for digital control applications, than the second-order form.

A. First-Order Model

Theorem 3:

For a system whose dynamics are governed by the Riccati equation (1), a discrete-time model that gives the exact value at the sampling instants for any sampling interval for stair-case type inputs, is given by the following:

$$\delta_T x_k = \Omega(x_k, T) f(x_k) \quad (25)$$

where the discrete-time integration gain is given by

$$\Omega(x_k, T) = \frac{\Gamma_\sigma}{1 + T\Gamma_\sigma \left(\frac{\sigma - b}{2} - ax_k \right)} \quad (26)$$

and

$$\Gamma_{\sigma} = \frac{e^{\sigma T} - 1}{\sigma T}, \quad (27)$$

both of which converge to unity as $T \rightarrow 0$.

Proof: The exact model of a linear system given by eq. (17) is known as the step-invariant-model (SIM) and is given [2] by

$$\delta_T v_k = \frac{v_{k+1} - v_k}{T} = \Gamma_{-\sigma} (-\sigma v_k + \phi). \quad (28)$$

Converting the variable v_k back into x_k using

$$v_k = \frac{\alpha x_k + \beta}{\gamma x_k + \delta}, \quad (29)$$

the first-order difference equation can be written, after some manipulations, in terms of the original variable x_k , as

$$x_{k+1} = \frac{\{\alpha\delta - \beta\gamma + T\Gamma_{-\sigma}\delta(\phi\gamma - \alpha\sigma)\}x_k + T\Gamma_{-\sigma}(\phi\delta - \sigma\beta)\delta}{\alpha\delta - \beta\gamma - T\Gamma_{-\sigma}(\phi\gamma\delta - \sigma\beta\gamma) - T\Gamma_{-\sigma}(\phi\gamma - \sigma\alpha)\gamma x_k}, \quad (30)$$

or in the delta form, as

$$\begin{aligned} \delta_T x_k &= \frac{\Gamma_{-\sigma}}{(\alpha\delta - \beta\gamma) - T\Gamma_{-\sigma}(\phi\delta - \sigma\beta)\gamma - T\Gamma_{-\sigma}(\phi\gamma - \sigma\alpha)\gamma x_k} \\ &\times \left[(\phi\gamma - \sigma\alpha)\gamma x_k^2 + \{2\phi\delta\gamma - \sigma(\alpha\delta + \beta\gamma)\}x_k + (\phi\delta - \sigma\beta)\delta \right] \end{aligned} \quad (31)$$

Using eq. (19) and the relationships in eq. (5), it can be shown that

$$\begin{aligned} (\phi\gamma - \sigma\alpha)\gamma &= \frac{\alpha^2(b\gamma\delta - c\gamma^2) - 2a\alpha\beta\delta + a\beta^2\gamma^2}{\alpha\delta - \beta\gamma} \\ &= a(\alpha\delta - \beta\gamma), \end{aligned} \quad (32)$$

$$\begin{aligned}
& 2\phi\delta\gamma - \sigma(\alpha\delta + \beta\gamma) \\
&= \frac{2(a\beta^2 - b\alpha\beta + c\alpha^2)\delta\gamma + b(\alpha^2\delta^2 + \beta^2\gamma^2) - 2(a\beta^2 + c\alpha^2)\gamma\delta}{\alpha\delta - \beta\gamma} \\
&= b(\alpha\delta - \beta\gamma),
\end{aligned} \tag{33}$$

$$\begin{aligned}
(\phi\delta - \sigma\beta)\delta &= \frac{-(b\delta\gamma - c\gamma^2)\beta^2 + c\alpha^2\delta^2 + b\beta^2\delta\gamma - 2c\alpha\beta\delta\gamma}{\alpha\delta - \beta\gamma} \\
&= c(\alpha\delta - \beta\gamma).
\end{aligned} \tag{34}$$

Therefore,

$$\begin{aligned}
& [(\phi\gamma - \sigma\alpha)\gamma x_k^2 + \{2\phi\delta\gamma - \sigma(\alpha\delta + \beta\gamma)\}x_k + (\phi\delta - \sigma\beta)\delta] \\
&= (\alpha\delta - \beta\gamma)(ax_k^2 + bx_k + c).
\end{aligned} \tag{35}$$

As for the integrator gain, noting (19), (32), (34), and $\Gamma_{-\sigma} = e^{-\sigma T}\Gamma_{\sigma}$, it can be seen that

$$\begin{aligned}
& \frac{\Gamma_{-\sigma}}{(\alpha\delta - \beta\gamma) - T\Gamma_{-\sigma}(\phi\delta - \sigma\beta)\gamma - T\Gamma_{-\sigma}(\phi\gamma - \sigma\alpha)\gamma x_k} \\
&= \frac{\Gamma_{\sigma}}{e^{\sigma T}(\alpha\delta - \beta\gamma) - T\Gamma_{\sigma}c\frac{\gamma}{\delta}(\alpha\delta - \beta\gamma) - T\Gamma_{\sigma}a(\alpha\delta - \beta\gamma)x_k} \\
&= \frac{1}{\alpha\delta - \beta\gamma} \frac{\Gamma_{\sigma}}{1 + T\Gamma_{\sigma}\left(\sigma - \frac{2ac}{b - \sigma} - ax_k\right)} \\
&= \frac{1}{\alpha\delta - \beta\gamma} \frac{\Gamma_{\sigma}}{1 + T\Gamma_{\sigma}\left(\frac{\sigma - b}{2} - ax_k\right)}
\end{aligned} \tag{36}$$

From eqs. (35) and (36), it follows that eq. (31) can be written as eq. (25) with (26) and (27). It should be noted that the exact discrete-time model (25) is independent of parameters $\alpha, \beta, \gamma, \delta$, and K . (E.O.P)

The forward-difference model of (1) is given [1] by

$$\delta_T x_k = f(x_k), \tag{37}$$

which is very simple to implement. Although this model can be used for a wide variety of nonlinear systems, the sampling interval must be chosen to be very small for good accuracy. Compared with

this model, the integrator gain Ω in the exact model (25) is time-varying, which needs to be updated at every sampling instant. Although more complex than the forward-difference model, the exact model (25) gives superior performance, as demonstrated in the following simulations:

B. Example

Consider a nonlinear system whose dynamics are governed by the following equation:

$$\dot{x} = -2x^2 + 2x + 1, \quad x(0) = 1 \quad (38)$$

which gives $\sigma = 2\sqrt{3}$. The design parameters are chosen to be $\alpha = 0$, $\beta = \gamma = -2a = 4$, $\delta = \sigma - b = 2(\sqrt{3} - 1)$. (They yield $\phi = -a = 2$, $\psi = \sigma = 2\sqrt{3}$, and $\mu = 0$, so that the corresponding first-order linearized system is $\dot{v} = -2\sqrt{3}v + 2$.) The exact discrete-time model is given by

$$\delta_T x_k = \Omega_k \{-2x_k^2 + 2x_k + 1\}, \quad x_0 = 1, \quad (39)$$

where

$$\Omega(x_k, T) = \frac{\Gamma_\sigma}{1 + T\Gamma_\sigma \{(\sqrt{3} - 1) - 2x_k\}} \quad (40)$$

$$\Gamma_\sigma = \frac{e^{2\sqrt{3}T} - 1}{2\sqrt{3}T}. \quad (41)$$

The gain for the forward-difference model is constant at $\Omega = 1$.

Figure 1 shows the step response of the continuous-time system which approaches $(1 + \sqrt{3})/2 = 1.366$ as time elapses, that of the proposed exact discrete-time model, and that of the forward-difference model, for $T = 0.2, 0.4$ and 0.8 seconds. It can be seen from these plots that the proposed model gives the discrete-time sequence that matches the continuous-time solution at sampling instants for all the sampling intervals tested, while the forward-difference model gives sequences that are not exact for any sampling interval. Although the error at the sampling instants becomes smaller with the forward-difference model as the sampling interval approaches zero, the sequence is divergent for $T = 0.8$ second, for which the chaotic behaviour can be observed (not shown here). In contrast, a chaotic behaviour does not occur with the exact model, no matter how large the sampling interval may be.

IV. CONCLUSIONS

A stable, first-order, and exact, discrete-time model was derived for the scalar Riccati differential equation with constant parameters. Exact models have long been sought in such areas as physics and biology, but these models are unstable, second-order, and not suitable for digital control purposes. Since the accuracy of nonlinear discrete-time models can be highly critical [3], the proposed model with no error in the discrete-time solution for any (slow) sampling rates, as demonstrated by two examples, can be an important scheme in the computation of nonlinear controllers and filter gains. This model also paves the way to the extension of the plant-input-mapping method [11], which is a digital redesign method that guarantees stability for any non-pathological sampling interval using the exact discrete-time model of a linear plant, into nonlinear cases.

Since most nonlinear systems that appear in the control areas can be immersed into differential equations in at most quadratic form [12], an extension of the proposed method into higher-order Riccati equations or matrix Riccati equations is important. The scalar equation (1) could be integrated directly and, therefore, system (25) could be derived by considering values at discrete-time instants. However, this approach is difficult to extend to higher-order cases and methods based-on linearization may be more useful, since linear systems can be integrated easily.

Acknowledgment

The work reported in this paper was supported, in part, by JSPS grant #17560202. The authors wish to express their gratitude for the valuable remarks and suggestions made by anonymous reviewers.

REFERENCES

1. Yuz, J. I. and Goodwin, G. C., "On Sampled-Data Models for Nonlinear Systems," *IEEE Trans. on Automatic Control*, 2005, **AC-50**, pp. 1477-1489.
2. Hori, N., Mori, T., and Nikiforuk, P. N., "A New Perspective for Discrete-Time Models of a Continuous-Time System," *IEEE Trans. on Automatic Control*, 1992, **AC-37**, pp. 1013-1017.
3. Nešić, D. and Teel, A., "A Framework for Stabilization of Nonlinear Sampled-Data Systems based on their Approximate Discrete-time Models," *IEEE Trans. on Automatic Control*, 2004, **AC-49**, pp. 1103-1122.
4. Lambert, J. D., *Computational Methods in Ordinary Differential Equations*, (John Wiley & Sons, New York, 1973).

5. Zelikin, M. I., *Control Theory and Optimization*, Encyclopaedia of Mathematical Sciences, Vol. 86, (Springer, Berlin, 2000).
6. Nowakowski, M. and Rosu, H. C., "Newton's Laws of Motion in the Form of a Riccati Equation," *Physical Review E*, Vol. 65, 047602-1~4, 2002.
7. Reid, W. T., *Riccati Differential Equations*, Mathematics in Science and Engineering, Vol. 86, (Academic Press, New York, 1972).
8. Hirota, R., *Lectures on Difference Equations – From Continuous to Discrete Domains*, (Science Publishing, Tokyo, 2000, in Japanese).
9. Hori, N., and Rabbath, C. A., "Application of Digital Control Theory to Exact Discretization of a Logistic Equation with a Constant Term," *Proc. of the IEEE Conference on Control Applications*, Toronto, 2005, pp. 303-307.
10. Middleton, R. H. and Goodwin, G. C., *Digital Control and Estimation – A Unified Approach*, (Prentice-Hall, Englewood Cliffs, NJ, 1990).
11. Markazi, A. H. D. and Hori, N., "Discretization of Continuous-Time Control Systems with Guaranteed Stability," *IEE Proc. on Control Theory and Applications*, 1995, **142**, pp. 323-328.
12. Ohtsuka, T., "Model Structure Simplification of Nonlinear Systems via Immersion," *IEEE Trans. on Automatic Control*, 2005, **AC-50**, pp. 607-618.

List of Figures and Captions

Figure 1

- (a) Simulation results with $T=0.2s$.
- (b) Simulation results with $T=0.4s$.
- (c) Simulation results with $T=0.8s$.

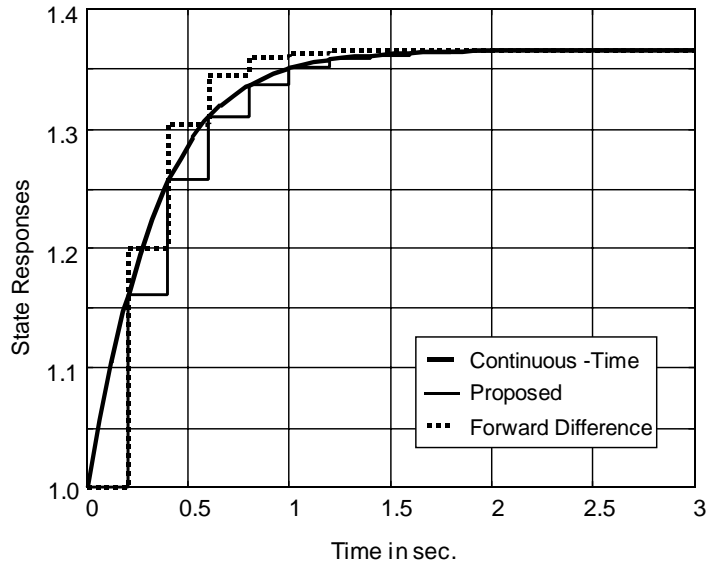


Figure 1 (a) Simulation results with $T=0.2s$.

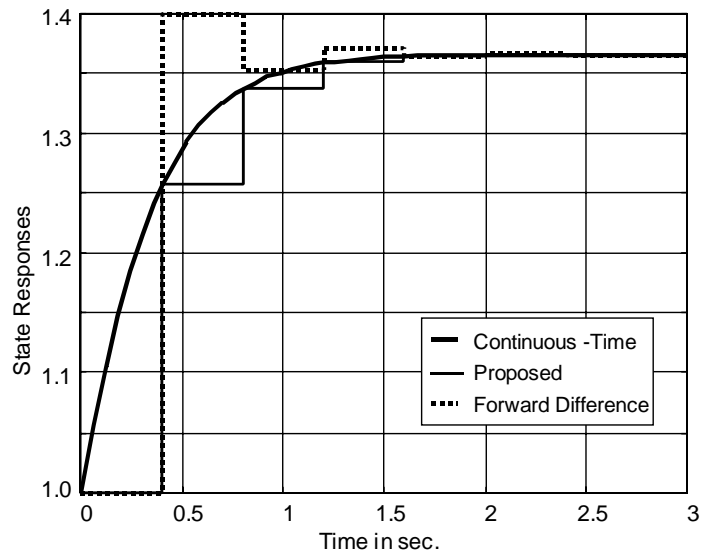


Figure 1 (b) Simulation results with $T=0.4s$.

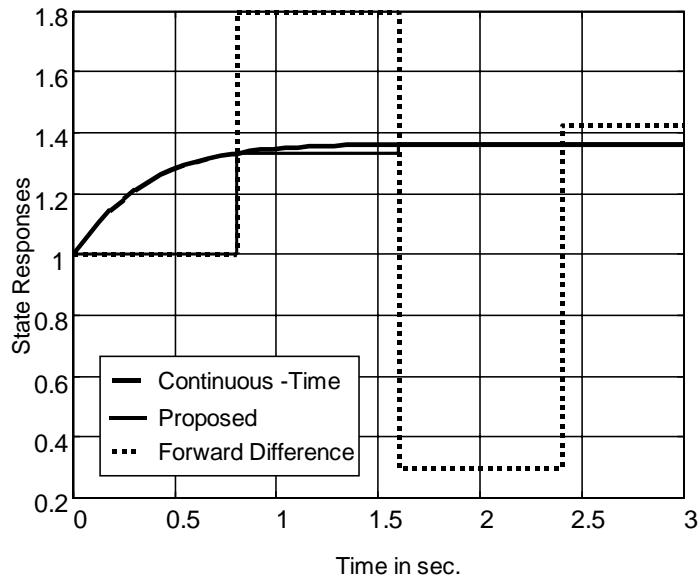


Figure 1 (c) Simulation results with $T=0.8s$.