Optimal Control of Volterra Difference Equations of the First Kind

Mohammed Alharthi[1,2] , Tim Hughes[2], and Markus Mueller[1,2]

***Abstract*— We consider optimal control for Volterra Difference Equations of the form**

$$x(n+1) = \sum_{i=0}^{n} B(i)x(n-i) + Cu(n), \quad n \in \mathbb{Z}^+. \tag{1}$$

We show that the optimal control problem can be solved via a Riccati equation or alternatively, and computationally less involved, by solving a linear equation. We consider an application from epidemiology where the optimal control problem admits an optimal solution using the theoretical result. However, the optimal control does not necessarily satisfy constraints of the system?s biology, i.e. non-negativity of the state.

I. INTRODUCTION

A. Types of Volterra Difference Equations

A Volterra Difference Equation (VDE) is a difference equation where the current state does not only depend on the previous state, but on past states typically starting from zero time. First, we introduce Volterra Difference Equation of the 'second kind' given by

$$x(n+1) = A(n)x(n) + \sum_{i=0}^{n} B(n,i)x(i), \quad n \in \mathbb{Z}^+, \tag{2}$$

where $A(n)$ and $B(n,i)$ are $k \times k$ real-valued matrices, for $k \in \mathbb{N}$ and $i, n \in \mathbb{Z}$ with $0 \leqslant i \leqslant n$. Here the term $\sum_{i=0}^{n} B(n,i)x(i)$ models the dependency of the current state on past states. If $B(n,i) = \widetilde{B}(n-i)$, then the equation is of convolution type.

For the so-called 'first kind' of Volterra difference systems, we set $A(n) = 0_{k\times k}$ for all $n \in \mathbb{Z}^+$:

$$x(n+1) = \sum_{i=0}^{n} B(n,i)x(i), \quad n \in \mathbb{Z}^+, \tag{3}$$

where $B(n,i)$ is a $k \times k$ real-valued matrix function in (n,i), for $i, n \in \mathbb{Z}$ with $0 \leqslant i \leqslant n$. Again, if $B(n,i) = \widetilde{B}(n-i)$, then the equation is of convolution type.

B. Stability of the scalar case

As for dynamical systems in general, stability of equilibrium states of a system is an important property.

Definition 1 (see, for example, [1, Definition 1.6.1]): An equilibrium state $x^* \in \mathbb{R}^k$ of the VDE (2) is said to be stable, if for a given $\varepsilon > 0$ and $n_0 \geqslant 0$ there exists $\delta = \delta(\epsilon, n_0)$ such that for any initial condition $x(n_0) = x_0$ with $\|x_0 - x^*\| < \delta$ it follows that the solution $x(n)$ of (2) satisfies $\|x(n) - x^*\| < \varepsilon$ for all $n \geqslant n_0$.

An equilibrium state $x^* \in \mathbb{R}^k$ of the VDE (2) is said to be uniformly stable, if δ may be chosen independently of n_0.

An equilibrium state $x^* \in \mathbb{R}^k$ of the VDE (2) is said to be (uniformly) attractive, if there exists $\gamma = \gamma(n_0) > 0$ ($\gamma > 0$ independently of n_0) such that $\lim_{n\to\infty} x(n) = x^*$ for all $x(n_0) = x_0$ such that $\|x(n) - x^*\| < \gamma$.

An equilibrium state $x^* \in \mathbb{R}^k$ of the VDE (2) is said to be (uniformly) asymptotically stable, if it is (uniformly) stable and (uniformly) attractive. $\Diamond$

To introduce characterisations of stability for Volterra Difference Equations, we first consider scalar convolution type VDEs of second kind:

$$x(n+1) = ax(n) + \sum_{i=0}^{n} b(n-i)x(i), \quad n \in \mathbb{Z}^+, \tag{4}$$

where $a \in \mathbb{R}$ and $b(n-i) \in \mathbb{R}$ for all $n - i \in \mathbb{N}$. The following result gives a sufficient condition for systems of the form (4).

Theorem 1 (see [1]): The zero solution of a system of form (4) is uniformly stable if

$$|a| + \sum_{i=0}^{n} |b(i)| \leqslant 1, \quad \text{for all } n \in \mathbb{Z}^+. \tag{5}$$

$\Diamond$

A similar result is true for scalar convolution type VDEs of first kind:

$$x(n+1) = \sum_{i=0}^{n} b(n-i)x(i), \quad n \in \mathbb{Z}^+. \tag{6}$$

A sufficient condition for global stability of zero solution is established by the following result by Kocic and Ladas (1993) [5].

Theorem 2: Assume that

$$\sum_{n=0}^{\infty} b(n) < 1. \tag{7}$$

Then the zero solution of a system of form (6) is globally asymptotically stable. $\Diamond$

II. OPTIMAL CONTROL

We follow the approach of the optimal control problem as discussed by Shah and George (2014) in [9].

We introduce an additive control term $Cu(n)$ to the system (3) of convolution type, where C is a $k \times m$ matrix for some $m \in \mathbb{N}$, i.e.

$$x(n+1) = \sum_{i=0}^{n} B(i)x(n-i) + Cu(n), \quad n \in \mathbb{Z}^+. \tag{8}$$

[1]Department of Mathematics, College of Engineering, Mathematics and Physical Sciences, Penryn Campus, University of Exeter, Penryn TR10 9FE, UK; ma784@exeter.ac.uk, m.mueller@exeter.ac.uk, t.h.hughes@exeter.ac.uk.

[2]Environment and Sustainability Institute, Penryn Campus, University of Exeter, Penryn TR10 9FE, UK.

We consider a quadratic performance index as cost function for the finite time process $(0 \leqslant n \leqslant N)$ as

$$J(u) = \frac{1}{2} x^*(N) S x(N) + \frac{1}{2} \sum_{n=0}^{N-1} [x^*(n) Q x(n) + u^*(n) R u(n))], \quad (9)$$

where S and Q are $k \times k$ positive semidefinite Hermitian matrices, and R is an $m \times m$ positive definite Hermitian matrix. We also introduce the set of Lagrange multipliers

$$\lambda(n) = Q x(n) + \sum_{i=0}^{N-n-1} B^*(i) \lambda(n+i+1), \quad n = 1, 2, \ldots, N-1, \quad (10)$$

with 'terminal state' $\lambda(N) = Sx(N)$. We can consider two optimal control problems, one with 'fixed' and one with 'free' terminal state $x(N)$. The following theorem and proof consider the case with free terminal state. We comment on the case with fixed terminal state in Remark 1.

Theorem 3 (see [9]): Minimization of the cost function (9) subject to the controlled VDE (8) has the optimal solution given by

$$\widehat{u}(n) = -R^{-1} C^* \lambda(n+1), \quad n = 0, 1, 2, \ldots, N-1, \quad (11)$$

where $\lambda(n)$ is defined by (10) with terminal condition $\lambda(N) = Sx(N)$. ◊

The following proof for Theorem 3 has been adapted from the one presented in [9] to allow its applicability to the case with free terminal state with minor adjustments, see Remark 1.

Proof of Theorem 3. We are aiming to minimize J defined in (9) subjected to the system (8) with an initial condition for state vector given by $x(0) = x^0$.

Using the set of Lagrange multipliers defined in (10), we define a second performance index adding zero terms $\lambda^*(n+1) \left[\sum_{i=0}^{n} B(i) x(n-i) + C u(n) - x(n+1)\right]$ and $\left[\sum_{i=0}^{n} B(i) x(n-i) + C u(n) - x(n+1)\right]^* \lambda(n+1)$, for all $n = 0, 1, \ldots, N-1$, to the cost function $J(u)$ to obtain $L(u)$ given by

$$\begin{aligned} L(u) &= \frac{1}{2} x^*(N) S x(N) \\ &+ \frac{1}{2} \sum_{n=0}^{N-1} \Bigg(\left[x^*(n) Q x(n) + u^*(n) R u(n))\right] \\ &+ \lambda^*(n+1) \left[\sum_{i=0}^{n} B(i) x(n-i) + C u(n) - x(n+1)\right] \\ &+ \left[\sum_{i=0}^{n} B(i) x(n-i) + C u(n) - x(n+1)\right]^* \lambda(n+1) \Bigg). \end{aligned} \quad (12)$$

We can see that the Hermitian transpose of $L(u)$ is equal to $L(u)$, i.e. $L(u) = L(u)^*$.

We follow Pontryagin's Maximum Principle (see, for example, [7, Ch I]) to minimize the function $L(u)$ over the input u. We obtain partial derivatives of $L(u)$ with respect to each component of the state vectors $x(t), u(t)$ and $\lambda(n)$, and set the respective partial derivatives equal to zero, that is

$$\begin{aligned} \frac{\partial L(u)}{\partial x(n)} &= 0_{k\times 1}, \quad n = 1, 2, \ldots, N, \\ \frac{\partial L(u)}{\partial u(n)} &= 0_{k\times 1}, \quad n = 0, 1, 2, \ldots, N-1, \\ \frac{\partial L(u)}{\partial \lambda(n)} &= 0_{k\times 1}, \quad n = 1, 2, \ldots, N. \end{aligned}$$

Here the partial derivatives for a vector $v = [v_1, \ldots, v_k]^\top$ is defined as

$$\frac{\partial}{\partial v} = \left[\frac{\partial}{\partial v_1}, \frac{\partial}{\partial v_2}, \ldots, \frac{\partial}{\partial v_k}\right]^\top.$$

For the following, we make use of the equalities

$$\frac{\partial}{\partial x} x^* M x = 2Mx,$$
$$\text{and} \quad \frac{\partial}{\partial x} x^* M y = My \quad \text{for any matrix } M = M^*,$$

We first consider the partial derivatives $\frac{\partial L(u)}{\partial x(n)}$ for $n = 1, 2, \ldots, N-1$:

$$\frac{\partial L(u)}{\partial x(n)} = Q x(n) + \sum_{i=0}^{N-1-n} B^*(i) \lambda(n+i+1) - \lambda(n), \quad n = 1, 2, \ldots, N-1, \quad (13)$$

and so, in view of $\frac{\partial L(u)}{\partial x(n)} = 0_{k\times 1}$ for $n = 1, 2, \ldots, N-1$, it follows that

$$0_{k\times 1} = Q x(n) + \sum_{i=0}^{N-1-n} B^*(i) \lambda(n+i+1) - \lambda(n), \quad n = 1, 2, \ldots, N-1. \quad (14)$$

Now, we consider the partial derivative $\frac{\partial L}{\partial x(N)}$:

$$0_{k\times 1} = \frac{\partial L(u)}{\partial x(N)} = Sx(N) - \lambda(N) \quad (15)$$

Similarly, considering the partial derivatives of L with respect to $u(n)$, we obtain the equations

$$0_{k\times 1} = \frac{\partial L(u)}{\partial u(n)} = R u(n) + C^* \lambda(n+1), \quad n = 0, 1, \ldots, N-1. \quad (16)$$

Lastly, considering the partial derivatives of L with respect to $\lambda(n)$, we obtain

$$0_{k\times 1} = \frac{\partial L(u)}{\partial \lambda(n)} = \sum_{i=0}^{n-1} B(i) x(n-i-1) + C u(n-1) - x(n), \quad n = 1, 2, \ldots, N. \quad (17)$$

Equation (17) is equivalent to the original Volterra Difference Equation (8) for the first $N-1$ non-negative integers.

Re-writing Equations (14) and Equation (15), we obtain

$$\lambda(n) = Qx(n) + \sum_{i=0}^{N-n-1} B^*(i)\lambda(n+i+1), \qquad n = 1, 2, \dots, N-1 \quad (18)$$

as required, see (10), and the final value of the Lagrange multipliers, i.e. $\lambda(N) = Sx(N)$.

Lastly, solving Equation (16) for $u(n)$ and recalling that R^{-1} exists because it was set to be a positive definite Hermitian, we get

$$u(n) = -R^{-1}C^*\lambda(n+1), \quad n = 0, 1, \dots, N-1 \quad (19)$$

Substituting (19) into (8) gives

$$x(n+1) = \sum_{i=0}^{n} B(i)x(n-i) - CR^{-1}C^*\lambda(n+1), \qquad n = 0, 1, 2, \dots, N-1, \quad (20)$$

with the initial condition $x(0) = x^0$.

To obtain the solution to the minimization problem, we need to solve (18) and (20) simultaneously. Note that for the system equation (20) the initial condition $x(0)$ is specified, while for the Lagrange multiplier equation (18), the final condition for $\lambda(N)$ is specified. Then,

$$\widehat{u}(n) = -R^{-1}C^*\lambda(n+1), \quad n = 0, 1, 2, \dots, N-1$$

where $\lambda(n)$ is the solution of the "VDE system and Lagrange multiplier boundary value problem". This completes the proof of the theorem. □

Remark 1: If we consider a fixed terminal state $x(N) = x_N \in \mathbb{R}^k$ for the optimal control problem "minimize $J(u)$ defined in (9) with respect to u and subject to the controlled VDE (8)", the cost function $J(u)$ reduces to

$$J(u) = \frac{1}{2}\sum_{n=0}^{N-1}[x^*(n)Qx(n) + u^*(n)Ru(n))],$$

because the term $\frac{1}{2}x^*(N)Sx(N)$ is a constant. Similarly the cost function $L(u)$ defined in (12) does not have the term $\frac{1}{2}x^*(N)Sx(N)$, and thus we do not need to consider the partial derivative $\frac{\partial L(u)}{\partial x(N)}$. The other partial derivatives lead to the same set of equations (18) and (20), and the optimal control given in (19). We do not require a specific terminal condition for the Lagrange multiplier equation (10) but $\lambda(N)$ can be determined implicitly, see the Case 2 of Corollary 1. ♢

According to [2], solving the boundary problem for (8) and (9) for infinite horizon, i.e. $N \to \infty$, is equivalent to solving the difference Riccati equation

$$\mathcal{F}(P(\omega)) = e^{-\omega}(Q + \upsilon^*(e^{-i\omega})P(\omega)) \times [I - C[R + C^\top P(\omega)C]^{-1}C^\top P(\omega)\upsilon(e^{-i\omega})], \quad (21)$$

where $\mathcal{F}(P)$ denotes the iteration of the discrete sequence $(P(\omega)_k)$, starting with an initial condition $P(\omega)_0 = P_0 \in \mathbb{R}^{k\times k}$, and where $\upsilon(e^{-i\omega})$ is defined as

$$\upsilon(e^{-i\omega}) = \sum_{i=0}^{\infty} B(n)e^{-i\omega n}, \, \omega \in [0, 2\pi]. \quad (22)$$

In [2], the authors consider the infinite-time problem, which requires to consider the discrete Fourier transform of $\widehat{u}(n)$, which is given as $V(\omega) = \sum_{n=0}^{\infty}\widehat{u}(n)e^{-i\omega n}$, for $\omega \in [0, 2\pi]$, and where $V(\omega)$ can be expressed as $V(\omega) = -[R + C^\top P(\omega)C]^{-1}C^\top P(\omega)\upsilon(e^{-i\omega})X(\omega)$, where $P(\omega)$ is derived from the Riccati equation (21) and $X(\omega)$ is the discrete Fourier transform of the optimal solution $\widehat{x}(n)$.

Under additional conditions the solution of Equation (21) can be determined by the successive approximation method $P_{k+1} = \mathcal{F}(P_k)$ starting from any initial value P_0 satisfying the condition $\ker P_0 \subset L$, where L is the maximal subspace that belongs to the null-space of Q and that is invariant to $\upsilon(e^{-i\omega})$.

Instead of aiming to solve a finite-time equivalent of the Riccati equation (21) to obtain the optimal control, we introduce a method using a linear system derived from the modified system equation (20) and Lagrange multipliers (18). This is summarized in the following corollaries.

For ease of notation we omit indices for zero and identity matrices in the following. Matrix entries denoted 0 and I have dimension $k \times k$, and vector entries denoted 0 have dimension $k \times 1$.

We define the $2Nk \times 2Nk$ matrix

$$\mathcal{A} = \begin{bmatrix} \mathcal{A}_{1,1} & \mathcal{A}_{1,2} \\ \mathcal{A}_{2,1} & \mathcal{A}_{2,2} \end{bmatrix} \quad (23)$$

and the $2Nk \times 1$ vector

$$\beta = -\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \beta_N \\ B(0)x_0 \\ B(1)x_0 \\ \vdots \\ B(N-1)x_0 \end{bmatrix} \quad (24)$$

where

$$\beta_N = 0 \quad (25)$$

for the free terminal state case (Case 1), and

$$\beta_N = x_N \quad (26)$$

for the fixed terminal state case (Case 2), and where the $Nk \times Nk$ matrix blocks $\mathcal{A}_{i,j}$, $i, j \in \{1, 2\}$ are defined as follows:

$$\mathcal{A}_{1,1} = \begin{bmatrix} -I & B(0)^* & B(1)^* & \dots & B(N-2)^* \\ 0 & -I & B(0)^* & \dots & B(N-3)^* \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & -I & B(0)^* \\ 0 & 0 & \dots & 0 & (\mathcal{A}_{1,1})_{N,N} \end{bmatrix}, \quad (27)$$

$$\mathcal{A}_{1,2} = \begin{bmatrix} Q & 0 & \dots & 0 & 0 \\ 0 & Q & \dots & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots & 0 \\ 0 & \vdots & \dots & Q & 0 \\ 0 & 0 & 0 & \dots & (\mathcal{A}_{1,2})_{N,N} \end{bmatrix}, \tag{28}$$

$$\mathcal{A}_{2,1} = \begin{bmatrix} -CR^{-1}C^* & 0 & \dots & 0 \\ 0 & -CR^{-1}C^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -CR^{-1}C^* \end{bmatrix}, \tag{29}$$

$$\mathcal{A}_{2,2} = \begin{bmatrix} -I & 0 & 0 & \dots & 0 \\ B(0) & -I & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ B(N-3) & \dots & B(0) & -I & 0 \\ B(N-2) & \dots & B(1) & B(0) & -I \end{bmatrix}, \tag{30}$$

where

$$(\mathcal{A}_{1,1})_{N,N} = -I, \quad \text{and} \quad (\mathcal{A}_{1,2})_{N,N} = S \tag{31}$$

for the free terminal state case (Case 1), and

$$(\mathcal{A}_{1,1})_{N,N} = 0, \quad \text{and} \quad (\mathcal{A}_{1,2})_{N,N} = -I \tag{32}$$

for the fixed terminal state case (Case 2).

Corollary 1: The optimal control $\widehat{u}(n)$ defined in (11) can be derived by solving the linear system

$$\mathcal{A}\left[\lambda(1)^\top, \dots, \lambda(N)^\top, x(1)^\top, \dots, x(N)^\top\right]^\top = \beta, \tag{33}$$

where $\mathcal{A}$ is given by (23) with (27)–(30) and β is given by (24), and we consider (31) and (25) for the optimisation problem with free terminal state (Case 1), and (32) and (26) for the optimisation problem with fixed terminal state (Case 2). ◊

Proof. Case 1. Consider a free terminal state. In view of Theorem 3 and the Langrange multiplier boundary problem (10) we have that

$$\left.\begin{aligned} \lambda(N-1) &= Qx(N-1) + B(0)^*\lambda(N), \\ \lambda(N-2) &= Qx(N-2) \\ &\quad + B(0)^*\lambda(N-1) + B(1)^*\lambda(N), \\ &\vdots \\ \lambda(1) &= Qx(1) + B(0)^*\lambda(2) + B(1)^*\lambda(3) \\ &\quad + \dots + B(N-2)^*\lambda(N), \end{aligned}\right\} \tag{34}$$

and the terminal condition $\lambda(N) = Sx(N)$. Furthermore, in view of Equation (20) we have that

$$\left.\begin{aligned} x(1) &= B(0)x(0) - CR^{-1}C^*\lambda(1), \\ x(2) &= B(1)x(0) + B(0)x(1) - CR^{-1}C^*\lambda(2), \\ &\vdots \\ x(N) &= B(N-1)x(0) + \dots + B(0)x(N-1) \\ &\quad - CR^{-1}C^*\lambda(N). \end{aligned}\right\} \tag{35}$$

Hence, (34), (35) and $\lambda(N) = Sx(N)$ form a linear system in the $2N$ unknown vectors $\lambda(1), \dots, \lambda(N)$ and $x(1), \dots, x(N)$, which we can re-write as matrix equation

$$\mathcal{A}\Lambda = \beta \tag{36}$$

where the matrix $\mathcal{A}$ and the vector β are defined in (23) and (24), respectively, with (31) and (25), and

$$\Lambda = \left[\lambda(1)^\top, \dots, \lambda(N)^\top, x(0)^\top, \dots, x(N)^\top\right]^\top, \tag{37}$$

as required.

Case 2. Consider a fixed terminal state $x(N) = x_N \in \mathbb{R}^k$. Similarly to Case 1, the Langrange multiplier equation (10) and Equation (20) yield the sets of equations (34) and (35). In view of $x(N) = x_N$ fixed, we have only $2N - 1$ unknowns $\lambda(1), \dots, \lambda(N)$ and $x(1), \dots, x(N-1)$, and we augment the system of equations with the equality $x(N) = x_N$. This and the sets of equations (34) and (35) can be re-written as matrix equation (36), where the matrix $\mathcal{A}$ and the vector β are defined in (23) and (24), respectively, with (32) and (26), and Λ defined in (37). This completes the proof for the second case and the corollary. □

Corollary 2: We assume that the matrix

$$\mathcal{A}_{1,1} - \mathcal{A}_{1,2}\mathcal{A}_{2,2}^{-1}\mathcal{A}_{2,1} \tag{38}$$

is invertible for either assumption (31) (for Case 1: free terminal state) or (32) (for Case 2: fixed terminal state). Then, the optimal control $\widehat{u}(n)$ defined in (11) is unique and can be derived by calculating

$$\Lambda = \mathcal{A}^{-1}\beta, \tag{39}$$

where Λ is defined in (37), and $\mathcal{A}$ and β are defined in (23) and (24), respectively. ◊

Proof. First, note that for Cases 1 and 2 the matrix block $\mathcal{A}_{2,2}$ of $\mathcal{A}$ is equal. Note further that $\mathcal{A}_{2,2}$ is a lower triangular matrix with all diagonal entries equal to -1, hence $\mathcal{A}_{2,2}$ is invertible.

We need to show that $\mathcal{A} = \begin{bmatrix} \mathcal{A}_{1,1} & \mathcal{A}_{1,2} \\ \mathcal{A}_{2,1} & \mathcal{A}_{2,2} \end{bmatrix}$ is invertible. We have that

$$\begin{aligned} &\begin{bmatrix} \mathcal{A}_{1,1} & \mathcal{A}_{1,2} \\ \mathcal{A}_{2,1} & \mathcal{A}_{2,2} \end{bmatrix} \begin{bmatrix} I & 0 \\ -\mathcal{A}_{2,2}^{-1} & \mathcal{A}_{2,2}^{-1} \end{bmatrix} \\ &\qquad = \begin{bmatrix} \mathcal{A}_{1,1} - \mathcal{A}_{1,2}\mathcal{A}_{2,2}^{-1}\mathcal{A}_{2,1} & \mathcal{A}_{1,2}\mathcal{A}_{2,2}^{-1} \\ 0 & I \end{bmatrix}, \end{aligned}$$

which is a block upper triangular matrix with the invertible diagonal blocks in view of Assumption (38), hence invertible. Therefore, $\mathcal{A}$ is invertible. This concludes the proof of the corollary. □

Remark 2: The matrix defined in (38) of Corollary 2 is a sum of the upper block triangular matrix $\mathcal{A}_{1,1}$ and the lower block triangular matrix $\mathcal{A}_{1,2}\mathcal{A}_{2,2}^{-1}\mathcal{A}_{2,1}$, see (27)–(30), respectively. Note that $\mathcal{A}_{1,2}$ and $\mathcal{A}_{2,1}$ are block diagonal matrices and that $\mathcal{A}_{2,2}$ is lower block triangular, and thus we have that $\mathcal{A}_{2,2}^{-1}$ is lower block triangular, too.

The matrices Q, R and S are design parameters of the optimisation problem, see (9). Hence, we have the freedom to choose suitable parameters to obtain an invertible matrix in (38) if this is at all possible.

For the free terminal state case, we may choose Q, R and S in a way that $\mathcal{A}_{1,1}$ dominates the matrix in (38), thus we obtain a matrix dominated by the upper block triangular matrix $\mathcal{A}_{1,1}$, which is invertible. Hence, the resulting matrix $\mathcal{A}_{1,1} - \mathcal{A}_{1,2}\mathcal{A}_{2,2}^{-1}\mathcal{A}_{2,1}$ is then also invertible.

For the fixed terminal state case, the problem is more subtle. First, note that

$$\mathcal{A}_{2,2}^{-1} = \begin{bmatrix} -I & 0 & 0 & \dots & 0 \\ \Gamma_{2,1} & -I & 0 & \dots & 0 \\ \Gamma_{3,1} & \Gamma_{3,2} & -I & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \Gamma_{N,1} & \Gamma_{N,2} & \dots & \Gamma_{N,N-1} & -I \end{bmatrix},$$

where

$$\Gamma_{\mu,\mu-1} = -B(0), \quad \mu = 2, 3, \dots, N,$$
$$\Gamma_{\mu,\nu} = -B(\mu-2) - \sum_{i=0}^{\mu-3} B(\mu-3-i)\Gamma_{i+2,\nu},$$
$$\mu = 3, 4, \dots, N, \quad \nu = 1, \dots, \mu-2.$$

In view of (28), (29) and (32), we then have

$$\mathcal{A}_{1,2}\mathcal{A}_{2,2}^{-1}\mathcal{A}_{2,1} = \begin{bmatrix} Q\Lambda & 0 & 0 & \dots & 0 \\ -Q\Gamma_{2,1}\Lambda & Q\Lambda & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -Q\Gamma_{N-2,1}\Lambda & \dots & -Q\Gamma_{N-3,N-2}\Lambda & Q\Lambda & 0 \\ \Gamma_{N,1}\Lambda & \dots & \Gamma_{N,N-2}\Lambda & \Gamma_{N,N-1}\Lambda & -\Lambda \end{bmatrix},$$

where $\Lambda = CR^{-1}C^*$. We also have in view of (32), that the last row of block matrices of $\mathcal{A}_{1,1}$ contains only zero matrices of dimension $k \times k$. As for the free terminal state case, we may choose Q so that the first $N-1$ rows of $\mathcal{A}_{1,1} - \mathcal{A}_{1,2}\mathcal{A}_{2,2}^{-1}\mathcal{A}_{2,1}$ are dominated by the first $N-1$ rows of $\mathcal{A}_{1,1}$, however invertibility of $\mathcal{A}_{1,1} - \mathcal{A}_{1,2}\mathcal{A}_{2,2}^{-1}\mathcal{A}_{2,1}$ requires invertibility of the matrix $\Lambda = CR^{-1}C^*$, which may not be the case for a given matrix C. $\Diamond$

III. Application of Volterra Difference Equations: an Epidemiological Model

In this section, we consider an epidemic model introduced by Lauwerier (1981) (see [6] and also [4, Example 6.25]). This model is a scalar VDE of convolution type and models the fraction of susceptible individuals in a population, denoted $x(n)$ in the model, in relation to a time-varying measure of infectiousness, $b(i)$ in the model, see Equation (40). We consider the optimal control problem to minimise infectiousness in the population, and thus the fraction of infected individuals. This can be interpreted as a proportion of vaccinated individuals or by measures like lockdown or social distancing requirements, which effectively would reduce the time-varying measure of infectiousness $b(i)$.

Let $x(n)$ denote the fraction of susceptible individuals in a certain population during the n^{th} day of an epidemic, and let $b(i) > 0$ be the measure of how infectious infected individuals are during the i^{th} day of their infection (note this is an average measure). Then the spread of an epidemic may be modelled by the equation

$$\ln \frac{1}{x(n+1)} = \sum_{i=0}^{n} (1 + \varepsilon - x(n-i))b(i), \quad n \in \mathbb{Z}^+, \quad (40)$$

where ε is a small non-negative number.

To transform (40) into a Volterra difference equation similar to Equation (4), we apply the state transformation

$$x(n) = e^{-y(n)}$$

which leads to the system equation

$$y(n+1) = \sum_{i=0}^{n} b(n-i)(1 + \varepsilon - e^{-y(i)}), \quad n \in \mathbb{Z}^+. \quad (41)$$

A biologically motivated requirement is that $x(n) \in [0, 1]$ (i.e. the fraction of susceptible individuals in a population is between 0% and 100%). Therefore, $y(n) \geq 0$ for all solutions of (41). Observe that during the early stages of an epidemic $x(n)$ is close to 1 and, consequently, $y(n)$ is close to zero. Hence it is reasonable to linearize (41) around zero. We replace $e^{-y(i)}$ by its linear approximation around zero, i.e. $e^{-y(i)} \approx 1 - y(i)$, Equation (41) becomes

$$y(n+1) = \sum_{i=0}^{n} b(n-i)(\varepsilon + y(i)), \quad y(0) \approx 0, \quad n \in \mathbb{Z}^+. \quad (42)$$

We consider $b(n)$ as a function of n which does not necessarily stabilize the system. The control problem is to stabilize the system and achieve a minimal outbreak of the epidemic during N time steps. We consider both cases: optimisation with free terminal state $y(N)$, and optimisation with fixed terminal state $y(N) = y_N \geqslant 0$. We apply the optimal control strategy and derive an optimal control input for System (42) with control term, that is

$$y(n+1) = \sum_{i=0}^{n} b(n-i)(\varepsilon + y(i)) + Cu(n),$$
$$y(0) = c, \quad n \in \mathbb{Z}^+, \quad (43)$$

for some small $c > 0$.

In the following subsections, we consider exemplar functions for $b(n)$ that are destabilizing. Applying the results from Corollaries 1 and 2, we derive input functions $u(n)$ which stabilize the system.

A. Example 1: 'weakly destabilizing' $b(n)$

We consider an infectiousness measure $b(n) = \frac{(0.65)^n}{2}$, see Panel (a) of Figure 1. The function does not satisfy the stability condition of Theorem 2, here $\sum_{n=0}^{\infty} b(n) \approx 1.4286 > 1$. For simplicity, we let $\varepsilon = 0$. We let $y(0) = c = 0.1$, i.e. 90% of the population are susceptible to the disease at the start of the epidemic. For the fixed terminal state optimisation problem we let $y(N) = y_N = 0.01$, i.e. the control objective in this case is to reduce the number of infected individuals to 1% of the population. The cost function parameters are $C = 2$, $R = 1$ and $Q = 1$, and

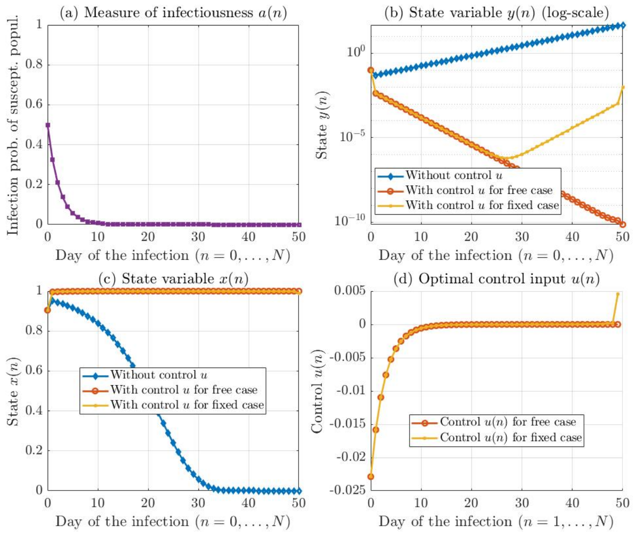

Fig. 1. The figures show simulation results for Example (III-A). For the chosen measure of infectiousness $b(n)$ (see panel (a)), the optimal control $u(n)$ (see panel (d)) leads to solutions $y(n)$ and the proportion of susceptible individuals $x(n)$ (See panels (b) and (c), respectively), that satisfy the biological constraints (i.e. non-negativity of $y(n)$ and hence proportions of susceptible population less than 100%).

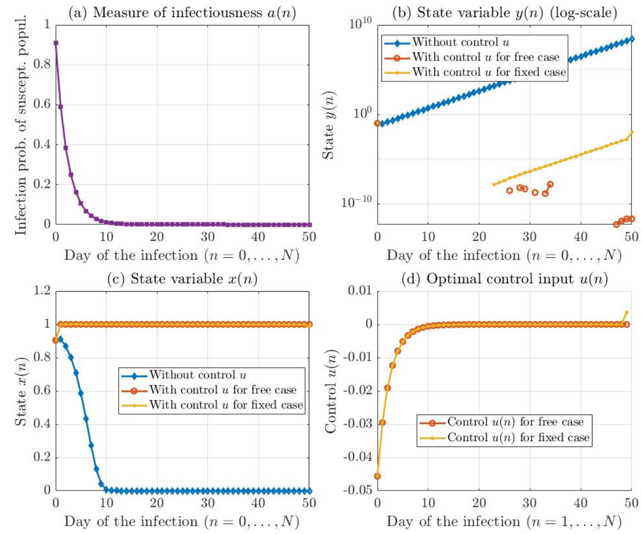

Fig. 2. The figures show simulation results for Example (III-B). The optimal control problem for the chosen function $b(n)$ (see panel (a)) returns negative values for state $y(n)$ (see panel (b) with missing values for both cases), which in return means that the population has a proportion larger than 100% of the susceptible individuals.

$S = 2$ for the free terminal state. We consider simulations of $N = 50$ days from the initial infection rate of 10% of the population.

The simulation results show that the objective is achieved and the state variables remain within biologically realistic constraints, see Panels (b) and (c) of Figure 1.

B. Example 2: 'strongly destabilizing' $b(n)$

We consider an infectiousness measure $b(n) = \frac{(0.65)^n}{1.1}$, see Panel (a) of Figure 2. Note that the trend is equivalent to the measure $b(n)$ from Example III-A, however the rates are scaled by a smaller value. As for $b(n)$ from Example III-A, the function $b(n)$ defined above does not satisfy the stability condition of Theorem 2, here $\sum_{n=0}^{\infty} b(n) \approx 2.5974 > 1$. As for Example III-A, we let $\varepsilon = 0$ and $y(0) = c = 0.1$, and $y(N) = 0.01$ for the fixed terminal state optimisation problem. As for Example III-A, the cost function parameters are $C = 2$, $R = 1$ and $Q = 1$, and $S = 2$ for the free terminal state. Again, we consider a simulation of $N = 50$ days from the initial infection infection rate of 10% of the population.

The simulation results show that the optimal control $u(n)$ does not lead to state variables that remain within biologically realistic constraints, see Panels (b) of Figure 2 where negative values for $y(n)$ are omitted from the log-scale plot.

From these examples we may conclude that the linear system given by (33) allows to derive control inputs that solve the optimal control problem for a given $b(n)$. However, solutions may not necessarily satisfy constraints that are required in view of modelling assumption, here proportions required to remain between 0 and 1.

IV. Conclusion

This paper explores optimal control problems for convolution type Volterra difference equations of first kind. We show that the optimal control problem can be solved by a system of linear equations. Example applications to an epidemiological model show that the approach yields solutions for the optimal control problem. However, for certain parameter choices the optimal solutions do not satisfy constraints implied by the biology of the system.

Further work will resolve this problem by develop approaches the allow for the inclusion of constraints, for example, dynamical programming methods and other optimization algorithms for nonlinear optimisation problems.

V. Acknowledgements

The first author thanks University of Bisha, Saudi Arabia, and Saudi Arabian Cultural Bureau in the UK, for their continued support and funding of his PhD studies.

References

[1] S. A. Al-Garni. *Numerical study of volterra difference equations of the second kind*. PhD thesis, King Fahd University of Petroleum and Minerals, 2003.

[2] M. Dymkov and I. Gaishun. Control problems for discrete volterra systems. 2000.

[3] S. Elaydi. Stability and asymptoticity of volterra difference equations: A progress report. *Journal of Computational and Applied Mathematics*, 228(2):504–513, 2009.

[4] S. N. Elaydi. *An Introduction to Difference Equations*. Springer-Verlag, Berlin, Heidelberg, 1996.

[5] G. Ladas and V. Kocic. Global behavior of nonlinear difference equations of higher order with applications. 1995.

[6] H. A. Lauwerier. Mathematical models of epidemics. *MC Tracts*, 1981.

[7] L. S. Pontryagin, V. G. Boltyanksii, R. V. Gamkrelidze, and E. F. Mishchenko. *Mathematical theory of optimal processes*. Gordon and Breach Science Publishers, 1986.

[8] Y. N. Raffoul. *Qualitative Theory of Volterra Difference Equations*. Springer, 2018.

[9] T. P. Shah and R. K. George. Optimal control of discrete volterra system-a classical approach. *International Journal of Applied Mathematics & Computation*, 5(4), 2014.