IMA Journal of Numerical Analysis (2016) Page 1 of 20
doi:10.1093/imanum/drnxxx

An Exponential Integrator Based Discontinuous Galerkin Method for Linear Complementarity Systems

ZHENGYU WANG†
Department of Mathematics, Nanjing University, China
AND
XIAOJUN CHEN‡
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

[Received on 23 September 2016]

The linear complementarity system (LCS) is defined by a linear ordinary differential equation coupled with a finite-dimensional linear complementarity problem (LCP), which has many applications in engineering and economics. In this article we reformulate the LCS with the boundary condition as an LCP in the Hilbert space of square-integrable functions, and propose a new numerical method for the LCS by using exponential Euler integrator and discontinuous Galerkin approximation. The precision of the proposed method is better than that of the existing time stepping method in different magnitude of scale. Convergence analysis and numerical experiments are performed to support the arguments.

Keywords: the linear complementarity system; exponential integrators; discontinuous Galerkin approximation; boundary value problems.

1. Introduction

Given matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $Q \in \mathbb{R}^{m\times n}$, $M \in \mathbb{R}^{m\times m}$ and $E_0, E_T \in \mathbb{R}^{n\times n}$, a vector $b \in \mathbb{R}^n$, and vector-valued functions $f : [0,T] \to \mathbb{R}^n$ and $g : [0,T] \to \mathbb{R}^m$, the linear complementarity system (LCS) is to find a state-control pair (x,u) of functions: $x : [0,T] \to \mathbb{R}^n$, $u : [0,T] \to \mathbb{R}^m$, such that

$$\left\{\begin{array}{rcll} \dot{x}(t) & = & Ax(t)+Bu(t)+f(t) & t \in [0,T] \\ u(t) & \in & \mathrm{SOL}(M, Qx(t)+g(t)) & t \in [0,T] \\ b & = & E_0x(0)+E_Tx(T), & \end{array}\right. \tag{1.1}$$

where for a matrix $M \in \mathbb{R}^{m\times m}$ and a vector $q \in \mathbb{R}^m$, $\mathrm{SOL}(M,q)$ denotes the solution set of the linear complementarity problem (denoted by $\mathrm{LCP}(M,q)$):

$$\mathrm{SOL}(M,q) := \{u \in \mathbb{R}^m |\ 0 \leqslant u \perp Mu+q \geqslant 0\}.$$

The LCS is a powerful mathematical modeling tool and finds various applications in, e.g., electrical networks with switching structure, contact mechanical systems, dynamical transportation assignment (Brogliato (2003); Zhong & Sumalee & Friesz & Lam (2011)). For many other applications in engineering and economics refer to Heemels & Schumacher & Weiland (2000); Pang & Stewart (2008).

†Corresponding author. Email: zywang@nju.edu.cn. The author's work was supported in part by National Natural Science Foundation of China (Grant No.11571166).

‡Email: maxjchen@polyu.edu.hk. The author's work was supported in part by Hong Kong Research Grants Council PolyU153001/14P.

The existing numerical methods for LCSs normally utilize the time stepping scheme (Chen & Wang (2013); Han & Camlibel & Pang & Heemels (2012); Pang & Stewart (2008)). For a given mesh and step size

$$0 = t_0 < t_1 < \cdots < t_N = T, \qquad h = T/N, \tag{1.2}$$

the scheme computes the approximate solution (x^h, u^h), where x^h is piecewise linear continuous and u^h piecewise constant in $[0,T]$ with $x^h(t_k) = x^{h,k}$ and $u^h(t_k) = u^{h,k}$, such that

$$\left\{ \begin{array}{rcl} x^{h,k} & = & x^{h,k-1} + h(Ax^{h,k} + Bu^{h,k} + f(t_k)) \\ u^{h,k} & \in & \mathrm{SOL}(M, Qx^{h,k} + g(t_k)) \\ b & = & E_0 x^{h,0} + E_T x^{h,N}. \end{array} \right.$$

It was shown in Chen & Wang (2011) that if M is a P-matrix, $E_0 = I$ and $E_T = 0$, then the initial value problem (IVP) of (1.1) has a classic solution (x,u), where x is continuously differentiable and u is continuous on $[0,T]$, and the time stepping method has 1-order convergence[1]. In general, the LCS (1.1) does not have a classic solution, and one has to seek a weak solution (x,u), where x is absolutely continuous and u is integrable. The pair (x,u), besides the boundary/initial condition, satisfies

$$x(t) - x(s) = \int_s^t [Ax(\tau) + Bu(\tau) + f(\tau)]d\tau$$

and

$$u(t) \ \in \ \mathrm{SOL}(M, Qx(t) + g(t))$$

for almost every $0 \leqslant s < t \leqslant T$. Note that the solutions of the LCS usually do not have a good smoothness, and therefore applying an integrator of high order, like the collocation methods (Kunkel & Stöver (2002)) will not yield fast convergence.

Notice that for the boundary value problem (BVP) of the LCS, the initial state is not prescribed, and we need to find the one, which leads to the terminal state such that the boundary condition is fulfilled. The dependence of the terminal state on the initial one is hard to be tracked. The time stepping method was studied for the BVP of the LCS in Han & Camlibel & Pang & Heemels (2012), but the convergence rate was not established therein.

In this paper we propose a new numerical method for solving (1.1) by combining the exponential integrator and the discontinuous Galerkin approximation. Below we summarize the idea and our contributions.

At first, we reformulate the LCS (1.1) as an LCP$(\mathfrak{L}, \hat{g})$ in the Hilbert space $\mathscr{L}^2(0,T;\mathbb{R}^m)$ of the m-dimensional vector-valued square integrable functions over the interval $[0,T]$: finding $u \in \mathscr{L}^2(0,T;\mathbb{R}^m)$ such that the complementarity condition

$$0 \ \leqslant \ u(t) \perp (\mathfrak{L}u)(t) + \hat{g}(t) \ \geqslant 0 \tag{1.3}$$

holds a.e. in $[0,T]$. Here $\mathfrak{L}$ is a bounded linear operator on $\mathscr{L}^2(0,T;\mathbb{R}^m)$ w.r.t. $\|\cdot\|_{\mathscr{L}^2}$. Then we apply the Galerkin approximation to the equivalent variational inequality (VI) formulation of the LCP$(\mathfrak{L}, \hat{g})$ posed in the convex closed function family $\mathscr{L}^2_+(0,T;\mathbb{R}^m)$:

$$\mathscr{L}^2_+(0,T;\mathbb{R}^m) := \{u \in \mathscr{L}^2(0,T;\mathbb{R}^m) \mid u(t) \geqslant 0 \ \ a.e. \text{ in } [0,T]\}. \tag{1.4}$$

[1] We call the convergence is of order $p > 0$ if $\|u - u^h\| = O(h^p)$.

The resulted discretized problem is a finite-dimensional LCP(M^h, q^h), and its solution offers an approximate control u^h. An approximate state x^h is computed by applying the exponential Euler integrator to the ordinary differential equation (ODE) with the approximation u^h instead of u. To our knowledge, the idea of Galerkin approximation and exponential integrator has not been considered for the LCS. Refer to Hochbruck & Ostermann (2010) for comprehensive treatment of various exponential integrators. We establish the error estimate

$$\|u^h - u\|_{\mathscr{L}^2} \leqslant C\sqrt{\|\mathfrak{P}_+^h u - u\|_{\mathscr{L}^2}},$$

where $\mathfrak{P}_+^h u$ denotes the projection of u onto the family $\mathscr{U}_+^h$ of piecewise linear functions (normally discontinuous) that are componentwise nonnegative over the time interval. The choice of the function family actually yields the discontinuous Galerkin approximation, and gives the pair (x^h, u^h) with x^h absolutely continuous and $u^h \in \mathscr{U}_+^h$. If $u \in \mathscr{C}^2(0,T;\mathbb{R}^m)$, the space of functions that have continuous 2-order derivatives, then the above error estimate indicates 1-order convergence in $\|\cdot\|_{\mathscr{L}^2}$ due to $\|\mathfrak{P}_+^h u - u\|_{\mathscr{L}^2} = O(h^2)$. To our surprise, the numerical results may suggest 2-order convergence, both in $\|\cdot\|_{\mathscr{L}^2}$ and in $\|\cdot\|_2$ at the grid points. See Figure 2.

Of course the choice of the family of continuous piecewise linear functions gives the continuous Galerkin approximation. Note that the family of piecewise constant functions is involved in the time stepping method. The approximate solutions given by our method is better in different magnitude of scale than those offered by the continuous Galerkin method and the time stepping method, even for the meshes not so refined. See Figure 3.

Note that the time stepping method involves the evaluation of $(I - hA)^{-1}$, while our method needs the evaluation of $\varphi_k(hA)$, which is not necessarily time-consuming than that of the former, where $\varphi_k(hA)$ is some matrices related to the matrix exponential e^{hA}. Summarizing, the computational cost for our Galerkin approximation amounts to that for the time-stepping method, while the precision of the proposed method is much better, as illustrated by the numerical results. See Figure 2.

This paper is organized as follows: in Section 2 we study the reformulation of the LCS into an LCP in the space $\mathscr{L}^2(0,T;\mathbb{R}^m)$, we also investigate the solvability of the LCP therein. We present the Galerkin approximation in Section 3 and the convergence analysis in Section 4. In Section 5 we present the algorithmic details and report the numerical results for two examples coming from a switched mechanical system and a generalized Nash equilibrium problem.

2. Reformulation of the LCS as an LCP in the Hilbert Space

For a given LCS (1.1), we suppose throughout this article that the matrix $E_0 + E_T e^{TA}$ is nonsingular, where e^{TA} denotes the matrix exponential. Define matrix-valued kernel functions:

$$\begin{array}{rcll} K_1(t,s) & = & e^{(t-s)A}B & (\text{for } t \geqslant s) \\ K_2(t,s) & = & e^{tA}(E_0 + E_T e^{TA})^{-1} E_T e^{(T-s)A} B. & \end{array} \tag{2.1}$$

Setting $K_1(t,s) = 0$ for $0 \leqslant t \leqslant s \leqslant T$, we then have for any $(t,s) \in [0,T]^2$

$$\begin{array}{rcl} \|K_1(t,s)\|_F & \leqslant & e^{T\|A\|_F} \cdot \|B\|_F, \\ \|K_2(t,s)\|_F & \leqslant & e^{2T\|A\|_F} \|(E_0 + E_T e^{TA})^{-1} E_T\|_F \|B\|_F, \end{array}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Define the Volterra integral operators

$$(\mathfrak{L}_j u)(t) = \int_0^T K_j(t,s)u(s)ds, \qquad (j = 1,2). \tag{2.2}$$

Obviously, $\mathfrak{L}_1$ and $\mathfrak{L}_2$ are compact in $\mathscr{L}^2(0,T;\mathbb{R}^m)$, and therefore bounded, their norms are still denoted by $\|\cdot\|_{\mathscr{L}^2}$. Actually, from the bound $\|K_j(t,s)\|_F$ established above together with the inequality (see Conway (1985))

$$\|\mathfrak{L}_j\|_{\mathscr{L}^2} \leqslant \left(\iint_{[0,T]^2} \|K_j(t,s)\|_F^2 dsdt\right)^{\frac{1}{2}},$$

it follows the bound $\|\mathfrak{L}_j\|_{\mathscr{L}^2} \leqslant \kappa_j$, where

$$\kappa_1 = Te^{T\|A\|_F}\|B\|_F/\sqrt{2}, \quad \kappa_2 = Te^{2T\|A\|_F}\|(E_0+E_Te^{TA})^{-1}E_T\|_F\|B\|_F. \tag{2.3}$$

Let $b \in \mathbb{R}^n$, $f:[0,T]\to\mathbb{R}^n$ and $g:[0,T]\to\mathbb{R}^m$ be square integrable on $[0,T]$, and let the LCS (1.1) have a weak solution (x,u). Then x, the solution of the ODE in (1.1) with u, can be represented by the constant variation formula Hochbruck & Ostermann (2010):

$$x(t) = e^{tA}x(0) + \int_0^t e^{(t-s)A}[Bu(s)+f(s)]ds,$$

where the boundary condition $E_0x(0)+E_Tx(T)=b$ can be written as

$$b \;=\; E_0x(0)+E_Tx(T) \;=\; (E_0+E_Te^{TA})x(0)+E_T\int_0^T e^{(T-s)A}\left[Bu(s)ds+f(s)\right]ds.$$

It gives the initial state

$$x(0) \;=\; (E_0+E_Te^{TA})^{-1}\left(b-E_T\int_0^T e^{(T-s)A}\left[Bu(s)+f(s)\right]ds\right)$$

and the solution of the BVP of the ODE:

$$x(t) = \hat{f}(t)+(\mathfrak{L}_1u)(t)-(\mathfrak{L}_2u)(t), \tag{2.4}$$

where

$$\hat{f}(t) = e^{tA}(E_0+E_Te^{TA})^{-1}\left(b-E_T\int_0^T e^{(T-s)A}f(s)ds\right)+\int_0^t e^{(t-s)A}f(s)ds. \tag{2.5}$$

By plugging (2.4) into the LCP in (1.1), we can see that u is a solution of the following linear complementarity problem, denoted by LCP$(\mathfrak{L},\hat{g})$, which is to find $u\in\mathscr{L}^2(0,T;\mathbb{R}^m)$ such that for almost $t\in[0,T]$ it holds

$$0 \;\leqslant\; u(t) \perp (\mathfrak{L}u)(t)+\hat{g}(t) \;\geqslant 0, \tag{2.6}$$

where $\mathfrak{L}$ is the linear bounded operator

$$(\mathfrak{L}u)(t) := Q(\mathfrak{L}_1u)(t)-Q(\mathfrak{L}_2u)(t)+Mu(t) \tag{2.7}$$

and

$$\hat{g}(t) \;=\; g(t)+Q\hat{f}(t). \tag{2.8}$$

Now we have reformulated the LCS (1.1) into the LCP$(\mathfrak{L},\hat{g})$, an LCP posed in $\mathscr{L}^2(0,T;\mathbb{R}^m)$.

THEOREM 2.1 If (x,u) is a weak solution of the LCS (1.1), then u is a solution of the LCP$(\mathfrak{L},\hat{g})$. Conversely, if u is a solution of the LCP$(\mathfrak{L},\hat{g})$, then (x,u) is a weak solution of (1.1), where x is given by (2.4).

Proof. The first part has been shown above. Conversely, if u is a solution of the LCP$(\mathfrak{L},\hat{g})$, then x is well defined by (2.4), which is absolutely continuous since its weak derivative $Ax+Bu+f$ is integrable. □

REMARK 2.1 A specific case of the LCP$(\mathfrak{L},\hat{g})$ is the so-called convolution complementarity problem (CCP), which models some contact mechanical problems Stewart (2006):

$$0 \leqslant u(t) \perp \int_0^t K(t-\tau)u(\tau)d\tau + g(t) \geqslant 0,$$

where $K(\cdot)$ is a given kernel. The operator involved in CCP is compact and some properties of compact operators can be utilized in the algorithmic construction and convergence analysis, while the operator $\mathfrak{L}$ defined in (2.7) is not compact in general.

An obvious advantage of reformulating the LCS as the LCP$(\mathfrak{L},\hat{g})$ lies in that we can use rich theory and abundant numerical methods for operator equations (Chen & Nashed & Qi (1997); Pang & Qi (1993)) to treat the LCS since LCP$(\mathfrak{L},\hat{g})$ can be equivalently reformulated as

$$\min\{u(t),(\mathfrak{L}u)(t)+\hat{g}(t)\} = 0,$$

where "min" is taken componentwise. In order to develop a Galerkin approximation scheme we need the variational formulation of the LCP. It is easy to show that u is a solution of the LCP$(\mathfrak{L},\hat{g})$ if and only if $u \in \mathscr{L}^2_+(0,T;\mathbb{R}^m)$, and

$$\langle \mathfrak{L}u+\hat{g}, v-u\rangle_{\mathscr{L}^2} \geqslant 0, \qquad \forall v \in \mathscr{L}^2_+(0,T;\mathbb{R}^m) \tag{2.9}$$

where the set $\mathscr{L}^2_+(0,T;\mathbb{R}^m)$, defined as in (1.4), is convex and closed in $\|\cdot\|_{\mathscr{L}^2}$.

When M is positive semi-definite, $\mathfrak{L}(\cdot) := Q\mathfrak{L}_1(\cdot) - Q\mathfrak{L}_2(\cdot) + M(\cdot)$ is pseudo-monotone as $\mathfrak{L}_1$ and $\mathfrak{L}_2$ are linear compact (Figure 27.1 of Zeidler (1990) p.596). As a direct consequence of Theorem 32.C of Zeidler (1990) (II/B, p.875), we have the following results on the solvability of the LCS (1.1).

THEOREM 2.2 Let $f \in \mathscr{L}^2(0,T;\mathbb{R}^n)$ and $g \in \mathscr{L}^2(0,T;\mathbb{R}^m)$. Suppose that there is a $u_0 \in \mathscr{L}^2_+(0,T;\mathbb{R}^m)$ such that

$$\frac{\langle \mathfrak{L}u, u-u_0\rangle_{\mathscr{L}^2}}{\|u\|_{\mathscr{L}^2}} \to +\infty \quad \text{as } \|u\|_{\mathscr{L}^2} \to \infty. \tag{2.10}$$

(1) If M is positive semi-definite, then the LCP$(\mathfrak{L},\hat{g})$ has a solution $u \in \mathscr{L}^2(0,T;\mathbb{R}^m)$.
(2) If $\mathfrak{L}$ is monotone, then the solution set of the LCP$(\mathfrak{L},\hat{g})$ is convex and closed in $\|\cdot\|_{\mathscr{L}^2}$.

It is well known that if the operator $\mathfrak{L}$ is strongly monotone, then the LCP$(\mathfrak{L},\hat{g})$ has a unique solution in $\mathscr{L}^2(0,T;\mathbb{R}^m)$ (II/B, Zeidler (1990)). From (2.3) we know the norms of $\mathfrak{L}_j$ can be bounded by a function of T that is decreasing to 0 when $T \downarrow 0$. This indicates that $\mathfrak{L}$ is strongly monotone if M is positive definite and T is small enough. We refine it in the following theorem.

THEOREM 2.3 Let M be positive definite. If $T > 0$ is small enough, then $\mathfrak{L}$ defined in (2.7) is strongly monotone, and then the LCP$(\mathfrak{L},\hat{g})$ has a unique solution in $\mathscr{L}^2(0,T;\mathbb{R}^m)$.

Proof. Denote by σ the smallest eigenvalue of $\frac{1}{2}(M+M^T)$. Obviously, $\sigma > 0$ since M is positive definite. Let κ_j be the constants defined in (2.3). It is easy to see $\|Q\mathfrak{L}_j\|_{\mathscr{L}^2} \leqslant \|Q\|_F \cdot \kappa_j$ for $j = 1,2$. Then for any $u \in \mathscr{L}^2(0,T;\mathbb{R}^m)$ we have

$$\begin{aligned}\langle u, \mathfrak{L}u\rangle &= \langle u, Mu\rangle + \langle u, Q\mathfrak{L}_1 u\rangle - \langle u, Q\mathfrak{L}_2 u\rangle \\ &\geqslant \left(\sigma - \|Q\|_F \sum_{j=1,2} \|\mathfrak{L}_j\|_{\mathscr{L}^2}\right) \cdot \|u\|^2_{\mathscr{L}^2} = [\sigma - \|Q\|_F(\kappa_1+\kappa_2)] \cdot \|u\|^2_{\mathscr{L}^2},\end{aligned}$$

which follows that $\mathfrak{L}$ is strongly monotone for $T > 0$ small enough since κ_j decreases to 0 as $T \to 0^+$. □

REMARK 2.2 For the case of the initial value problem, namely, $E_0 = I$ and $E_T = 0$, we have $\kappa_2 = 0$. With the uniform mesh such that the step size h fulfills

$$\sqrt{2}\sigma > he^{h\|A\|_F}\|Q\|_F\|B\|_F,$$

Theorem 2.3 yields the unique existence of the solution of the LCP$(\mathfrak{L}, \hat{g})$ in $[0,h]$. Repeating the application of the theorem, one can establish the solvability of (1.1) for any $T > 0$.

The LCS is a special case of the so-called differential variational inequality, for which sufficient conditions for the existence of a solution (Theorem 2, pp.392) was given in Chen & Wang (2014). In the current setting, this theorem reads as follows.

THEOREM 2.4 Let E_0+E_T be nonsingular and $\hat{x}^0 = (E_0+E_T)^{-1}b$. Denote the solution set of the parameterized LCP by $\mathscr{S}(t,x) := \mathrm{SOL}(M, Qx(t)+g(t))$. Suppose that $\mathscr{S}(0,\hat{x}^0)$ is nonempty and bounded, and M is positive semi-definite. Denote $\mathscr{F}(t,x) := \{Ax+Bu+f(t) | u \in \mathscr{S}(t,x)\}$. If $\mathscr{S}(t,x)$ is lower semi-continuous near $(0,\hat{x}^0)$ or $\mathscr{F}(t,x)$ is singleton, then there exist $T, \delta_0, \zeta > 0$ such that the LCS (1.1) has a solution (x,u), where $x \in \mathscr{N}(\hat{x}^0, \delta_0 + \zeta T)$ is continuously differentiable with $x(0) \in \mathscr{N}(\hat{x}^0, \delta_0)$, and u is continuous and is the least norm element of $\mathscr{S}(t,x(t))$, namely, u is the least norm solution of the LCP$(M, Qx(t)+g(t))$.

3. Galerkin Approximation

3.1 *Approximation of* $\mathscr{L}^2_+(0,T;\mathbb{R}^m)$

We need an approximation of the function family $\mathscr{L}^2_+(0,T;\mathbb{R}^m)$. Let the mesh (1.2) be given, denote by $\mathscr{U}^h$ the space of piecewise linear functions. $\mathscr{U}^h$ is a $(2Nm)$-dimensional closed subspace of $\mathscr{L}^2(0,T;\mathbb{R}^m)$. Introduce

$$\hat{\psi}_1(t) = 1-t, \qquad \hat{\psi}_2(t) = t, \tag{3.1}$$

where $\chi_k(t)$ is the characteristic function in $I_k = (t_{k-1}, t_k]$ and e_i denotes the m-dimensional i-th unit coordinate vector. For $i = 1,\ldots,m$, $j = 1,2$ and $k = 1,\ldots,N$, define

$$\psi_{i,j,k}(t) := \hat{\psi}_j\left(\frac{t-t_{k-1}}{h}\right) \cdot \chi_k(t) \cdot e_i. \tag{3.2}$$

Obviously, the function family $\{\psi_{i,j,k} : 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant 2, 1 \leqslant k \leqslant N\}$ spans the subspace $\mathscr{U}^h$, and

$$\langle \psi_{i,j,k}, \psi_{i',j',k'}\rangle_{\mathscr{L}^2} = \begin{cases} h/3 & i = i',\ k = k',\ j = j' \\ h/6 & i = i',\ k = k',\ j \neq j' \\ 0 & \text{otherwise.}\end{cases} \tag{3.3}$$

Note that $u^h(t) = \sum_{i,j,k} c_{i,j,k}\psi_{i,j,k}(t) \in \mathscr{U}^h$ has the following form in the interval I_k

$$u^h(t) = \psi_1\left(\frac{t-t_{k-1}}{h}\right)\begin{pmatrix} c_{1,1,k} \\ \vdots \\ c_{m,1,k}\end{pmatrix} + \psi_2\left(\frac{t-t_{k-1}}{h}\right)\begin{pmatrix} c_{1,2,k} \\ \vdots \\ c_{m,2,k}\end{pmatrix}.$$

The set $\mathscr{U}_+^h$ is convex and closed. And since $u^h(t) \geqslant 0$ holds almost everywhere in $[0,T]$ if and only if $c_{i,j,k} \geqslant 0$ holds for all the coefficients, it has the following representation:

$$\mathscr{U}_+^h := \mathscr{U}^h \cap \mathscr{L}_+^2(0,T;\mathbb{R}^m) = \left\{ \sum_{i,j,k} c_{i,j,k}\cdot\psi_{i,j,k} : \ c_{i,j,k} \geqslant 0 \right\}.$$

Denote by $\mathscr{C}_+^2$ the set of functions in $\mathscr{L}^2(0,T;\mathbb{R}^m)$ that are componentwise nonnegative on $[0,T]$ and have continuous 2-order derivatives. Clearly, $\mathscr{C}_+^2$ is a dense subset of $\mathscr{U}_+^h$. Below we show that $\mathscr{U}_+^h$ is an approximation to $\mathscr{L}_+^2(0,T;\mathbb{R}^m)$ in a certain sense.

THEOREM 3.1 Let $u \in \mathscr{L}_+^2(0,T;\mathbb{R}^m)$, and $q_{i,j,k} = \langle u, \psi_{i,j,k}\rangle_{\mathscr{L}^2}/h$, and denote by "mid" the componentwise median operation. Then
(1) u has the unique projection $\mathfrak{P}_h^+ u = \sum_{i,j,k} c_{i,j,k}^* \psi_{i,j,k}$ onto the closed and convex set $\mathscr{U}_+^h$ with respect to the norm $\|\cdot\|_{\mathscr{L}^2}$, where the coefficients $c_{i,j,k}^*$ are given by

$$\begin{array}{rcl} c_{i,1,k}^* & = & mid\left\{0, 4q_{i,1,k} - 2q_{i,2,k}, 3q_{i,1,k}\right\} \\ c_{i,2,k}^* & = & mid\left\{0, 4q_{i,2,k} - 2q_{i,1,k}, 3q_{i,2,k}\right\}. \end{array} \tag{3.4}$$

(2) Moreover if $u \in \mathscr{C}_+^2$, then $\lim_{h\to 0^+} \|u - \mathfrak{P}_h^+ u\|_{\mathscr{L}^2} = 0$.

Proof. (1) We mention that the projection is well defined because the subset $\mathscr{U}_+^h$ is convex and closed. Noting $\langle u, \psi_{i,j,k}\rangle_{\mathscr{L}^2} = hq_{i,j,k}$, and by using the inner product of the basis functions given in (3.3), for any $u^h = \sum_{i,j,k} c_{i,j,k}\psi_{i,j,k}$ we compute

$$\begin{array}{rcl} \|u-u^h\|_{\mathscr{L}^2}^2 & = & \langle u,u\rangle_{\mathscr{L}^2} - 2\langle u,u^h\rangle_{\mathscr{L}^2} + \langle u^h,u^h\rangle_{\mathscr{L}^2} \\ = & \langle u,u\rangle_{\mathscr{L}^2} - 2\displaystyle\sum_{i,j,k} c_{i,j,k}u_{i,j,k} + \sum_{i,j,k}\sum_{i',j',k'} \langle \psi_{i,j,k}, \psi_{i',j',k'}\rangle_{\mathscr{L}^2} c_{i,j,k}c_{i',j',k'} \\ = & \langle u,u\rangle_{\mathscr{L}^2} + \dfrac{h}{3}\displaystyle\sum_{i,k}\left(c_{i,1,k}^2 + c_{i,2,k}^2 + c_{i,1,k}c_{i,2,k} - 6q_{i,1,k}c_{i,1,k} - 6q_{i,2,k}c_{i,2,k}\right). \end{array}$$

This objective function is separate in i and k, namely, the minimization of $\|u-u^h\|_{\mathscr{L}^2}^2$ in $\mathscr{U}_+^h$ is equivalent to solve the mN quadratic minimization problems of the form

$$\min_{c_{i,1,k},c_{i,2,k}\geqslant 0}\left\{c_{i,1,k}^2 + c_{i,2,k}^2 + c_{i,1,k}c_{i,2,k} - 6q_{i,1,k}c_{i,1,k} - 6q_{i,2,k}c_{i,2,k}\right\}$$

for $i = 1,\ldots,m$ and $k = 1,\ldots,N$, which can be equivalently reformulate as an LCP

$$0 \leqslant \begin{pmatrix} c_{i,1,k} \\ c_{i,2,k}\end{pmatrix} \perp \begin{pmatrix} 2 & 1 \\ 1 & 2\end{pmatrix}\begin{pmatrix} c_{i,1,k} \\ c_{i,2,k}\end{pmatrix} - 6\begin{pmatrix} q_{i,1,k} \\ q_{i,2,k}\end{pmatrix} \geqslant 0. \tag{3.5}$$

The LCP (3.5) has a unique solution. Noting that $u, \psi_{i,j,k} \in \mathscr{U}_+^h$, therefore $q_{i,j,k} = \langle u, \psi_{i,j,k}\rangle_{\mathscr{L}^2}/h \geqslant 0$. It is easy to test

- for the case $q_{i,2,k} \geqslant 2q_{i,1,k}$, the LCP (3.5) has the solution

$$c^*_{i,1,k} = 0, \quad c^*_{i,2,k} = 3q_{i,2,k};$$

- for the case $q_{i,2,k} < 2q_{i,1,k}$ but $q_{i,2,k} > \frac{1}{2}q_{i,1,k}$, the LCP (3.5) has the solution

$$c^*_{i,1,k} = 4q_{i,1,k} - 2q_{i,2,k}, \quad c^*_{i,2,k} = 4q_{i,2,k} - 2q_{i,1,k};$$

- for the case $q_{i,2,k} \leqslant \frac{1}{2}q_{i,1,k}$, the LCP (3.5) has the solution

$$c^*_{i,1,k} = 3q_{i,1,k}, \quad c^*_{i,2,k} = 0.$$

These three cases can be included in the form (3.4). The solutions $c^*_{i,j,k}$ give a global minimizer of $\|u-u^h\|^2_{\mathscr{L}^2}$ in $\mathscr{U}^h_+$.

(2) Denote by $\tilde{u}^h$ the piecewise linear interpolant of u:

$$\tilde{u}^h(t) = \varphi_1\left(\frac{t_k - t}{h}\right)u(t_{k-1}) + \varphi_2\left(\frac{t - t_{k-1}}{h}\right)u(t_k)$$

with the estimate $\|u-\tilde{u}^h\|_{\mathscr{L}^2} \leqslant C_u h^2$, where C_u is a constant independent of h (dependent of u). Obviously, $\tilde{u}^h$ is also nonnegative on $[0,T]$ and

$$\|u - \mathfrak{P}^+_h u\|_{\mathscr{L}^2} \leqslant \|u-\tilde{u}^h\|_{\mathscr{L}^2} \leqslant C_u h^2.$$

This completes the proof. □

REMARK 3.1 Let $\mathfrak{P}_h u$ be the orthogonal projection of $u \in \mathfrak{L}^2_+(0,T;\mathbb{R}^m)$ onto its subspace $\mathscr{U}^h$ with respect to $\langle\cdot,\cdot\rangle_{\mathscr{L}^2}$. We can see that

$$\|u - \max\{\mathfrak{P}_h u, 0\}\|_{\mathscr{L}^2} \leqslant \|u - \mathfrak{P}^+_h u\|_{\mathscr{L}^2}.$$

Normally the function $\max\{\mathfrak{P}_h u, 0\}$ is not an element of $\mathscr{U}^h_+$, but is in a more refined approximation function family. For example, let $u : [0,1] \to \mathbb{R}$ with $u(t) = 1$ when $t \in [0, \frac{1}{2}]$ and $u(t) = 0$ otherwise. Then one can compute $\mathfrak{P}_h u = \frac{5}{4}\varphi_1(t) - \frac{1}{4}\varphi_2(t)$, which takes negative values when $t \in (\frac{5}{6}, 1]$. Hence $\max\{\mathfrak{P}_h u, 0\} = \frac{5}{4}\varphi_1(t) - \frac{1}{4}\varphi_2(t) \geqslant 0$ in $[0, \frac{5}{6}]$ and $\max\{\mathfrak{P}_h u, 0\} = 0$ in $[\frac{5}{6}, 1]$. It is not linear on $[0,1]$ with the mesh $0 = t_0 < t_1 = 1$ and $h = 1$, but is piecewise linear on a refined mesh with $\frac{5}{6}$ being a mesh point.

REMARK 3.2 Suppose that u has finitely many discontinuous points (related to mode switches **?**), and suppose that the mesh is so refined that we have just two situations: all the components of u do not vanish in the interior of any subintervals except for those, where some components vanish constantly. Then $\mathfrak{P}^+_h u$ is the orthogonal projection of $u \in \mathfrak{L}^2_+(0,T;\mathbb{R}^m)$ onto the closed subspace $\mathscr{U}^h$ with respect to the inner product $\langle\cdot,\cdot\rangle_{\mathscr{L}^2}$, and $\lim_{h\to 0^+}\|u - \mathfrak{P}^+_h u\|_{\mathscr{L}^2} = 0$.

3.2 *Discretization of the LCP*$(\mathfrak{L}, \hat{g})$

We apply Galerkin approximation to the variational formulation (2.9), which is equivalent to the LCP$(\mathfrak{L}, \hat{g})$. Namely, we find $u^h = \sum_{l=1}^d z_l \varphi_l \in \mathscr{U}^h_+$ fulfilling (2.9) in $\mathscr{U}^h$, where $\{\varphi_1, \cdots, \varphi_d\}$ is a basis of $\mathscr{U}^h$,

$\varphi_l = \psi_{i,j,k}$ for $l = i + m(j-1) + 2m(k-1)$, $1 \leqslant i \leqslant m$, $j = 1,2$, $1 \leqslant k \leqslant N$, $d = 2Nm$, and where $\psi_{i,j,k}$ is defined by (3.2). That is, for any $v^h = \sum_{l=1}^{d} z'_l \varphi_l \in \mathscr{U}_+^h$, it holds true

$$\begin{aligned} 0 &\leqslant \frac{1}{h}\langle \mathfrak{L}u^h + \hat{g}, v^h - u^h\rangle_{\mathscr{L}^2} = \frac{1}{h}\langle \mathfrak{L}\sum_{i=1}^{d} z_i\varphi_i + \hat{g}, \sum_{j=1}^{d}(z'_j - z_j)\varphi_j\rangle_{\mathscr{L}^2} \\ &= \frac{1}{h}\sum_{i,j=1}^{d} z_i(z'_j - z_j)\langle \mathfrak{L}\varphi_i, \varphi_j\rangle_{\mathscr{L}^2} + \frac{1}{h}\sum_{j=1}^{d}(z'_j - z_j)\langle \hat{g}, \varphi_j\rangle_{\mathscr{L}^2} \\ &= (z'-z)^T(M^h z + q^h), \end{aligned}$$

where $z = (z_i)$, $z' = (z'_i)^T$, $q^h = (q^h_j) \in \mathbb{R}^d$, and $M^h = (M^h_{ji}) \in \mathbb{R}^{d\times d}$,

$$M^h_{ji} := \frac{1}{h}\langle \mathfrak{L}\varphi_i, \varphi_j\rangle_{\mathscr{L}^2}, \qquad q^h_j := \frac{1}{h}\langle \hat{g}, \varphi_j\rangle_{\mathscr{L}^2}. \tag{3.6}$$

We mention that $u^h \in \mathscr{U}_+^h$ if and only if $z \in \mathbb{R}^d_+$. Therefore the above variational formulation yields the d-dimensional LCP(M^h, q^h): find $z^h \in \mathbb{R}^d$ such that

$$0 \leqslant z^h \perp M^h z^h + q^h \geqslant 0. \tag{3.7}$$

Suppose that the LCP($\mathfrak{L}, \hat{g}$) has a solution $u \in \mathscr{L}^2(0,T;\mathbb{R}^m)$. Denote by $\mathfrak{P}^h_+ u$ the projection of u on to $\mathscr{U}_+^h$, denote

$$r(t) = \min\{u(t), (\mathfrak{L}u)(t) + \hat{g}(t)\}, \quad r^h(t) = \min\{(\mathfrak{P}^h_+u)(t), (\mathfrak{L}\mathfrak{P}^h_+u)(t) + \hat{g}(t)\}.$$

For fixed t we have a matrix $D = diag(d_i) \in \mathbb{R}^{m\times m}$ with $0 \leqslant d_i \leqslant 1$ such that

$$\begin{aligned} r(t) - r^h(t) &= \min\{u(t), (\mathfrak{L}u)(t) + \hat{g}(t)\} - \min\{(\mathfrak{P}^h_+u)(t), (\mathfrak{L}\mathfrak{P}^h_+u)(t) + \hat{g}(t)\} \\ &= (I-D)(u - \mathfrak{P}^h_+u)(t) + D\mathfrak{L}(u - \mathfrak{P}^h_+u)(t). \end{aligned}$$

See Alefeld & Chen & Potra (1999). Noting $\|I-D\|_2 \leqslant 1$ and $\|D\|_2 \leqslant 1$, we have

$$\begin{aligned} \|r(t) - r^h(t)\|_2^2 &= \|(I-D)(u(t) - (\mathfrak{P}^h_+u)(t)) + D\mathfrak{L}(u - \mathfrak{P}^h_+u)(t)\|_2^2 \\ &\leqslant \left(\|I-D\|_2\|u(t) - \mathfrak{P}^h_+u(t)\|_2 + \|D\|_2\|\mathfrak{L}(u - \mathfrak{P}^h_+u)(t)\|_2\right)^2 \\ &\leqslant 2\|u(t) - \mathfrak{P}^h_+u(t)\|_2^2 + 2\|\mathfrak{L}(u - \mathfrak{P}^h_+u)(t)\|_2^2. \end{aligned}$$

Since $u \in \mathscr{L}^2(0,T;\mathbb{R}^m)$ solves the LCP($\mathfrak{L}, \hat{g}$), it holds true $r(t) = 0$ a.e. over $[0,T]$. Therefore

$$\|r^h\|^2_{\mathscr{L}^2} = \int_0^T \|r(t) - r^h(t)\|_2^2 dt \leqslant 2\|u - \mathfrak{P}^h_+u\|^2_{\mathscr{L}^2} + 2\|\mathfrak{L}(u - \mathfrak{P}^h_+u)\|^2_{\mathscr{L}^2}.$$

Notice that $\mathfrak{L}$ is bounded and $\|u - \mathfrak{P}^h_+u\|^2_{\mathscr{L}^2} \to 0$ as $h \downarrow 0$. Then $\|r^h\|_{\mathscr{L}^2} \to 0$ as $h \downarrow 0$. We know $r^h = 0$ a.e. over $[0,T]$ when $\|r^h\|_{\mathscr{L}^2} = 0$. For this reason we can say $\mathfrak{P}^h_+u$ approximately solves the Galerkin approximation problem (3.7), and it justifies in a certain sense the Galerkin approximation in the LCS setting.

Of course, a very small measure $\|r^h\|_{\mathscr{L}^2}$ does not imply the solvability of (3.7). We study it below.

THEOREM 3.2 Let M be positive semi-definite, $f \in \mathscr{L}^2(0,T;\mathbb{R}^n)$ and $g \in \mathscr{L}^2(0,T;\mathbb{R}^m)$. Then under condition (2.10), the LCP (3.7) has a solution.

Proof. Fix the mesh and let $\mathfrak{P}_+^h u_0$ be the projection of u_0 onto $\mathscr{U}_+^h$. Let $u \in \mathscr{U}_+^h$ with $\|u\|_{\mathscr{L}^2} \to \infty$. Then from (2.10) it follows that

$$\begin{aligned} +\infty \leftarrow \frac{\langle \mathfrak{L}u, u-u_0\rangle_{\mathscr{L}^2}}{\|u\|_{\mathscr{L}^2}} &= \frac{\langle \mathfrak{L}u, u-\mathfrak{P}_+^h u_0\rangle_{\mathscr{L}^2}}{\|u\|_{\mathscr{L}^2}} + \frac{\langle \mathfrak{L}u, \mathfrak{P}_+^h u_0 - u_0\rangle_{\mathscr{L}^2}}{\|u\|_{\mathscr{L}^2}} \\ &\leqslant \frac{\langle \mathfrak{L}u, u-\mathfrak{P}_+^h u_0\rangle_{\mathscr{L}^2}}{\|u\|_{\mathscr{L}^2}} + \|\mathfrak{L}\|_{\mathscr{L}^2}\|\mathfrak{P}_+^h u_0 - u_0\|_{\mathscr{L}^2}. \end{aligned}$$

This follows that condition (2.10) holds true in $\mathscr{U}_+^2$ since $\|\mathfrak{L}\|_{\mathscr{L}^2}\|\mathfrak{P}_+^h u_0 - u_0\|_{\mathscr{L}^2}$ is finite. The conclusion is drawn again by Theorem 32.C of Zeidler (1990) (II/B, p.875). □

The solution $z^h = (z_i^h)$ of the LCP(M^h, q^h) gives a piecewise (discontinuous) linear function on the mesh, which has the following form on the subinterval $(t_{k-1}, t_k]$:

$$u^h(t) = \left(\frac{t_k - t}{h}\right) u_{k,1}^h + \left(\frac{t - t_{k-1}}{h}\right) u_{k,1}^h, \tag{3.8}$$

where

$$u_{k,1}^h = \begin{pmatrix} z_{m(2k-2)+1}^h \\ \vdots \\ z_{m(2k-2)+m}^h \end{pmatrix}, \quad u_{k,1}^h = \begin{pmatrix} z_{m(2k-1)+1}^h \\ \vdots \\ z_{m(2k-1)+m}^h \end{pmatrix}.$$

In subsequence, u^h is just called as the solution of the LCP(M^h, q^h) if no ambiguity caused.

4. Convergence Analysis

We study the convergence of the Galerkin approximation (3.7). We know that problem (2.9) and its Galerkin approximation (3.7) have the unique solution if the operator $\mathfrak{L}$ is strongly monotone, which is true if, for example, M is positive definite and T is small enough. For the case of strong monotonicity we prove the following result of the convergence rate by adapting the technique of Falk (Falk (1974)).

THEOREM 4.1 Let $\mathfrak{L}$ be strongly monotone: $\langle \mathfrak{L}v, v\rangle_{\mathscr{L}^2} \geqslant \alpha\|v\|_{\mathscr{L}^2}^2$ holding with a constant $\alpha > 0$ for any $v \in \mathscr{L}^2(0,T;\mathbb{R}^m)$. Let u and u^h be the unique solutions of the LCP$(\mathfrak{L}, \hat{g})$ and the LCP(M^h, q^h), respectively. Then

$$\|u^h - u\|_{\mathscr{L}^2} \leqslant C\sqrt{\inf_{v^h \in \mathscr{U}_+^h} \|v^h - u\|_{\mathscr{L}^2}} = C\sqrt{\|\mathfrak{P}_h^+ u - u\|_{\mathscr{L}^2}}. \tag{4.1}$$

If $u \in \mathscr{C}^2(0,T;\mathbb{R}^m)$, the family of functions that have continuous 2-order derivatives, then $\|u^h - u\|_{\mathscr{L}^2} = O(h)$.

REMARK 4.1 The estimate (4.1) holds for any approximation of $\mathscr{L}_+^2(0,T;\mathbb{R}^m)$ that is convex and closed, for example the space of piecewise constant functions which are nonnegative over $[0,T]$.

Proof. Since u solves (2.9) and u^h solves (3.7), we have $\langle \mathfrak{L}u + \hat{g}, u\rangle_{\mathscr{L}^2} = 0$, $\langle \mathfrak{L}u + \hat{g}, u^h\rangle_{\mathscr{L}^2} \geqslant 0$, and

$\langle \mathfrak{L}u^h + \hat{g}, v^h \rangle_{\mathscr{L}^2} \geqslant 0$ for any $v^h \in \mathscr{U}_+^h$. These yield

$$\begin{aligned}
\langle \mathfrak{L}(u-u^h), u-u^h \rangle_{\mathscr{L}^2} &= \langle (\mathfrak{L}u+\hat{g}) - (\mathfrak{L}u^h+\hat{g}), u-u^h \rangle_{\mathscr{L}^2} \\
&= -\langle \mathfrak{L}u+\hat{g}, u^h \rangle_{\mathscr{L}^2} - \langle \mathfrak{L}u^h+\hat{g}, u \rangle_{\mathscr{L}^2} \\
&\leqslant \langle \mathfrak{L}u^h+\hat{g}, v^h \rangle_{\mathscr{L}^2} - \langle \mathfrak{L}u^h+\hat{g}, u \rangle_{\mathscr{L}^2} \\
&= \langle \mathfrak{L}(u^h-u), v^h-u \rangle_{\mathscr{L}^2} + \langle \mathfrak{L}u+\hat{g}, v^h-u \rangle_{\mathscr{L}^2} \\
&\leqslant \|\mathfrak{L}\|_{\mathscr{L}^2} \|u^h-u\|_{\mathscr{L}^2} \|v^h-u\|_{\mathscr{L}^2} + \|\mathfrak{L}u+\hat{g}\|_{\mathscr{L}^2} \|v^h-u\|_{\mathscr{L}^2}.
\end{aligned}$$

Using the condition $\alpha \|u-u^h\|_{\mathscr{L}^2} \leqslant \langle \mathfrak{L}(u-u^h), u-u^h \rangle_{\mathscr{L}^2}$ and the inequality

$$\|\mathfrak{L}\|_{\mathscr{L}^2} \|u^h-u\|_{\mathscr{L}^2} \|v^h-u\|_{\mathscr{L}^2} \leqslant \frac{\alpha}{2} \|u^h-u\|_{\mathscr{L}^2}^2 + \frac{\|\mathfrak{L}\|_{\mathscr{L}^2}^2}{2\alpha} \|v^h-u\|_{\mathscr{L}^2}^2$$

we obtain

$$\frac{\alpha}{2} \|u^h-u\|_{\mathscr{L}^2}^2 \leqslant \frac{\|\mathfrak{L}\|_{\mathscr{L}^2}^2}{2\alpha} \|v^h-u\|_{\mathscr{L}^2}^2 + \|\mathfrak{L}u+\hat{g}\|_{\mathscr{L}^2} \|v^h-u\|_{\mathscr{L}^2}.$$

This follows (4.1) since $v^h \in \mathscr{U}_+^h$ is arbitrary.

We take v^h as the piecewise linear interpolant of u, which, obviously, is an element of $\mathscr{U}_+^h$ with $\|u-\tilde{u}^h\|_{\mathscr{L}^2} \leqslant C_u h^2$ if $u \in \mathscr{C}_+$, where C_u is a constant independent of h (dependent of u). This completes the proof. □

THEOREM 4.2 Assume that the operator $\mathfrak{L}$ is monotone, and let u^h be the solution of the LCP(M^h, q^h). If $\{u^h\}$ is uniformly bounded for h small enough, then $\{u^h\}$ has a subsequence, which weakly converges to a solution of the LCP$(\mathfrak{L}, \hat{g})$.

Proof. If $\{u^h\}$ is uniformly bounded, then $\{u^h\}$ has a subsequence (still denoted by $\{u^h\}$ for avoiding the cumbersome presentation), converging to $u \in \mathscr{L}_+^2(0,T;\mathbb{R}^m)$ weakly. Now we prove that u is a solution of (2.9). Note that $\mathscr{C}_+^2$ is the set of functions that have continuous 2-order derivatives and are nonnegative in $[0,T]$. Since $\mathscr{C}_+^2$ is dense in $\mathscr{L}_+^2(0,T;\mathbb{R}^m)$, it is enough to show $\langle \mathfrak{L}u+\hat{g}, v-u \rangle_{\mathscr{L}^2} \geqslant 0$ for any $v \in \mathscr{C}_+^2$.

Let $v \in \mathscr{C}_+^2$ and take v^h as its piecewise linear interpolant. Then v^h is strongly convergent to v in $\|\cdot\|_{\mathscr{L}^2}$ as $h \downarrow 0$. Since u^h is weakly convergent to u, we have

$$\langle \mathfrak{L}u^h, v^h \rangle_{\mathscr{L}^2} \to \langle \mathfrak{L}u, v \rangle_{\mathscr{L}^2}, \qquad \langle \hat{g}, v^h-u^h \rangle_{\mathscr{L}^2} \to \langle \hat{g}, v-u \rangle_{\mathscr{L}^2}.$$

Therefore $\langle \mathfrak{L}u^h, v^h \rangle_{\mathscr{L}^2} + \langle \hat{g}, v^h-u^h \rangle_{\mathscr{L}^2} \to \langle \mathfrak{L}u, v \rangle_{\mathscr{L}^2} + \langle \hat{g}, v-u \rangle_{\mathscr{L}^2}$. And we obtain

$$0 \leqslant \langle \mathfrak{L}u^h, u^h \rangle_{\mathscr{L}^2} \leqslant \langle \mathfrak{L}u^h, v^h \rangle_{\mathscr{L}^2} + \langle \hat{g}, v^h-u^h \rangle_{\mathscr{L}^2}$$

since $\langle \mathfrak{L}u^h+\hat{g}, v^h-u^h \rangle_{\mathscr{L}^2} \geqslant 0$. This follows $\langle \mathfrak{L}u+\hat{g}, v-u \rangle_{\mathscr{L}^2} \geqslant 0$ for any $v \in \mathscr{C}_+^2$. □

5. Numerical Experiment

5.1 *Implementation details*

Let M^h and q^h be defined as in (3.6). Below we give the representation of the matrix M^h and the column vector q^h, where the linear interpolants f^h and g^h of f and g are used as their approximations. Below we

denote by $f_k = f(t_k)$ and $g_k = g(t_k)$ for $k = 0, \ldots, N$. The representations need the following matrices $G_k = G_k(hA)$, where $h > 0$ is the stepsize. Denote $G_0 = e^{hA}$ and

$$G_k = \sum_{n=k}^{\infty} \frac{1}{n!}(hA)^{n-k} = \frac{1}{k!}I + \frac{1}{(k+1)!}(hA) + \frac{1}{(k+2)!}(hA)^2 + \cdots.$$

The matrices G_k can be computed efficiently by using, e.g., the Pade, Krylov subspace approximations. Here we only need G_k for $0 \leqslant k \leqslant 4$, which can be evaluated in high precision by the existing software, e.g. the matrix function toolbox (Higham, http://www.ma.man.ac.uk/ higham/mftoolbox). Arguments on other software for computing the matrix functions can be found in Higham & Al-Mohy (2010).

Denote by $\otimes$ the Kronecker tensor product of two matrices. Then M^h has the forms:

$$\begin{aligned} M^h &= \frac{1}{6}I_N \otimes \begin{pmatrix} 2M & M \\ M & 2M \end{pmatrix} - h\sum_{k',k=1}^{N} e_{k'}e_k^T \otimes \begin{pmatrix} (J_{k',k})_{11} & (J_{k',k})_{12} \\ (J_{k',k})_{21} & (J_{k',k})_{22} \end{pmatrix} \\ &\quad + \frac{h}{2}\sum_{k'=2}^{N}\sum_{k=1}^{k'-1} e_{k'}e_k^T \otimes \begin{pmatrix} Q(G_1-G_2)B & QG_2B \\ Q(G_1-G_2)B & QG_2B \end{pmatrix} \\ &\quad + hI_N \otimes \begin{pmatrix} Q(G_3-G_4)B & QG_4B \\ Q(G_2-2G_3+G_4)B & Q(G_3-G_4)B \end{pmatrix}, \end{aligned}$$

where e_k is the k-th N-dimensional unit coordinate vector, and for $k', k = 1, \ldots, N$:

$$\begin{aligned} (J_{k',k})_{11} &= QG_0^{k'-1}G_2(E_0 + E_T e^{TA})^{-1}E_T G_0^{N-k}(G_1 - G_2)B \\ (J_{k',k})_{12} &= QG_0^{k'-1}G_2(E_0 + E_T e^{TA})^{-1}E_T G_0^{N-k}G_2B \\ (J_{k',k})_{21} &= QG_0^{k'-1}(G_1 - G_2)(E_0 + E_T e^{TA})^{-1}E_T G_0^{N-k}(G_1 - G_2)B \\ (J_{k',k})_{22} &= QG_0^{k'-1}(G_1 - G_2)(E_0 + E_T e^{TA})^{-1}E_T G_0^{N-k}G_2B. \end{aligned}$$

And we compute q^h in a naive manner. Given the approximate control u^h, we present the following formula for approximating the state x at the mesh points by $x_k^h = x^h(t_k)$ for $k = 0, 1, \cdots N$:

$$\begin{aligned} x_k^h &= G_0^k \hat{x}^{0,h} + \int_0^{t_k} e^{(t_k - s)A} f(s)ds + h\left(G_1 B\sum_{j=1}^{k} u_{j,1}^h + G_2 B\sum_{j=1}^{k}(u_{j,2}^h - u_{j,1}^h)\right) \\ &\quad - hG_0^k(E_0 + E_T e^{TA})^{-1}E_T\sum_{k=1}^{N} G_0^{N-k}\left[(G_1 - G_2)Bu_{k,1}^h + G_2 Bu_{k,2}^h\right], \end{aligned}$$

where

$$\hat{x}^{0,h} = (E_0 + E_T e^{TA})^{-1}\left(b - E_T\int_0^T e^{(T-s)A} f(s)ds\right).$$

Here we omit the proof of the form of M^h and the justification of the approximation of the state, which can be found in the supplement material (Wang & Chen (2016)).

5.2 *Numerical results*

In this subsection we apply the Galerkin approximation method and the time stepping method to two LCSs, which generate respectively the numerical solutions (x_g^h, u_g^h) and (x_e^h, u_e^h), where the subscript "e" stands for "Euler" since the time stepping method actually makes the use of implicit Euler method to

dicretize the ODEs involved in the LCSs. Here x_g^h is recovered in the manner stated above by using a solution u_g^h of the LCP(M^h, q^h). The exact solution of the LCS is always denoted by (x,u).

Here we are interested in the errors of state and control, in $\|\cdot\|_{\mathscr{L}^2}$ and in $\|\cdot\|_2$, namely we compute the values of

$$\begin{array}{llll} \|x_g^h - x\|_{\mathscr{L}^2}, & \|u_g^h - u\|_{\mathscr{L}^2}, & \|x_g^h(T) - x(T)\|_2, & \|u_g^h(T) - u(T)\|_2, \\ \|x_e^h - x\|_{\mathscr{L}^2}, & \|u_e^h - u\|_{\mathscr{L}^2}, & \|x_e^h(T) - x(T)\|_2, & \|u_e^h(T) - u(T)\|_2. \end{array}$$

If, for example, $\log(\|x_g^h - x\|_{\mathscr{L}^2})$ is affine w.r.t. $\log(h)$, then the slope gives the order of the state convergence in $\|\cdot\|_{\mathscr{L}^2}$. Therefore for the two examples we report the logarithms of the errors in different h.

Here we take the uniform grid, the LCP(M^h, q^h) is solved by using the PATH LCP solver Dirkse & Ferris & Munson (1994), numerical methods are coded and performed in the setting of Octave 4.0. Below we present the details of the numerical example, and thereafter present the comments on the numerical results.

EXAMPLE 5.1 The collapse of the Tacoma Narrows suspension bridge can be modeled by an ordinary differential equations of second order with nonsmooth data. The nonsmooth data is reformulated by a linear complementarity problem, which leads to an IVP of the LCS. The data of the problem and the exact solution can be found in Chen & Mahmoud (2008). This problem can be reformulated as an LCP$(\mathfrak{L}, \hat{g})$ with a strongly monotone $\mathfrak{L}$, and therefore has a unique solution (x,u).

Below we present some remarks on the numerical results.

(1) Take $h \approx 0.1$. The first components of the exact state x and their approximations $(x_g^h)_1$ and $(x_e^h)_1$ are plotted in the upper part of Figure 1, the counterpart results for the control are plotted in the lower part. The step size is not restrictive, while the numerical solutions offered by Galerkin approximation are very close to the exact one, and much more precise than the output of the time stepping method.

(2) The logarithms of the errors for the Galerkin approximation are plotted in Figure 2, w.r.t. the logarithms of different h. Note that the graphs are approximately 4 straight lines with the slopes all about of 2. This strongly indicates that the numerical solutions given by the Galerkin approximation method, have a convergence of order 2, both in norm $\|\cdot\|_{\mathscr{L}^2}$ and in $\|\cdot\|_2$ at the terminal time. Notice that Theorem 4.1 just gives the 1-order convergence.

(3) The logarithms of the errors $\|x_e^h(T) - x(T)\|_2$ are plotted in the upper part of Figure 3. The graphs are approximately 2 straight lines with the slopes about of 9.2334e-01. This indicates a convergence of the time stepping method in $\|\cdot\|_2$ at T of the order close to 1. As illustrated by the numerical results, the Galerkin approximation has a much better precision than the time-stepping method. For instance, for $h \approx 10^{-3}$, we have $\|x_e^h(T) - x(T)\|_2 / \|x_g^h(T) - x(T)\|_2 \approx 2.3565 \times 10^3$. The overperformance of the Galerkin approximation in $\|\cdot\|_{\mathscr{L}^2}$ is more obvious. In the same setting as above, our numerical results show

$$\frac{\|x_e^h - x\|_{\mathscr{L}^2}}{\|x_g^h - x\|_{\mathscr{L}^2}} \approx 2.2168 \times 10^7, \qquad \frac{\|u_e^h - u\|_{\mathscr{L}^2}}{\|u_g^h - u\|_{\mathscr{L}^2}} \approx 2.1610 \times 10^6.$$

(4) Theorem 4.1 indicates that the approximation error for the Galerkin approximation is bounded from above by $\|\mathfrak{P}_h^+ u - u\|_{\mathscr{L}^2}$, where $\mathfrak{P}_h^+ u$ is the projection of the true solution u onto the function family $\mathscr{U}_+^h$. Here we plot in the lower part of Figure 3 the logarithms of the errors of the projection onto three different function families: the family of piecewise constant, continuous and discontinuous piecewise linear functions that are nonnegative in $[0,T]$. The numerical results explain in another manner the

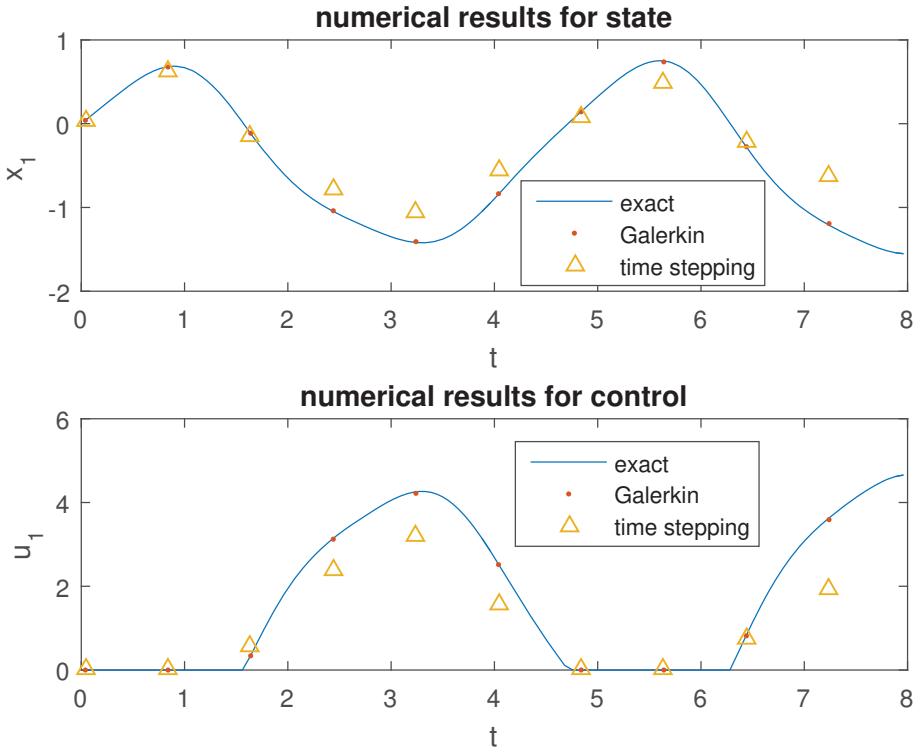

FIG. 1. Exact and numerical solutions of the LCS (Example 5.1 with $T = 3\pi$)

reason that the Galerkin approximation numerically overperforms the time stepping method (approximating u by a piecewise constant function), and justifies the application of the discontinuous Galerkin approximation instead of the continuous one (approximating u by a continuous piecewise linear function).

Notice that in the current case, the projection error for discontinuous piecewise linear function family is very small and not sensitive in h, it means that a high accuracy of our discontinuous Galerkin approximation can be achieved for an h not restrictive. We mention that the projection error is dependent of the geometry of the true solution u. If u is piecewise linear and its discontinuities are located at the grid points, then the error is zero.

EXAMPLE 5.2 Consider a dynamic Nash equilibrium problem with 2 players and zero-sum cost functionals. Denote by $y_i \in \mathbb{R}$ and u_i the i-th player's state and control variables, respectively, $i = 1, 2$. An

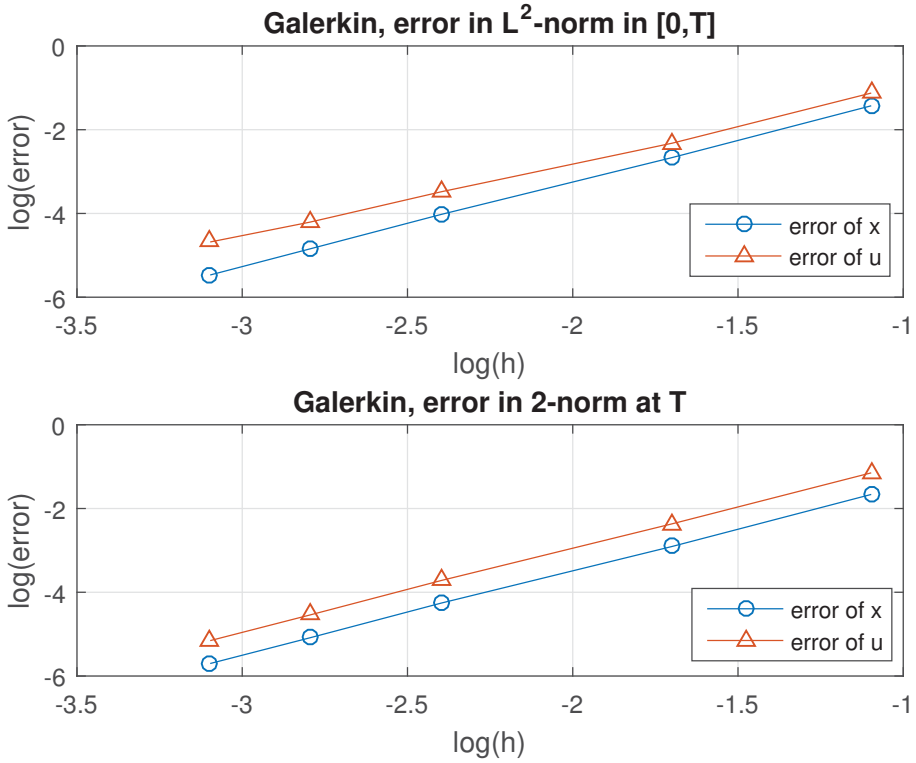

FIG. 2. Errors of (x_g^h, u_g^h) in $\|\cdot\|_{\mathscr{L}^2}$ and $\|\cdot\|_2$ at T (Example 5.1 with $T = 3\pi$)

equilibrium solution of this problem is a state-control pair $(y_1^*(t), y_2^*(t), u_1^*(t), u_2^*(t))$ satisfying

$$\begin{aligned} u_1^*(\cdot) \quad \in \quad & \text{argmin} \quad \theta(y_1, y_2^*, u_1, u_2^*) \\ & \text{s.t.} \quad \dot{y}_1(t) = -2 + 2y_1 + B_1 u_1 \\ & \qquad\quad y_1(0) = -1 \\ & \qquad\quad u_1 \geqslant 0,\, e^T(u_1 + u_2^*) \leqslant 1 \\ u_2^*(\cdot) \quad \in \quad & \text{argmax} \quad \theta(y_1^*, y_2, u_1^*, u_2) \\ & \text{s.t.} \quad \dot{y}_2(t) = -2t - y_2 + B_2 u_2 \\ & \qquad\quad y_2(0) = 2 \\ & \qquad\quad u_2 \geqslant 0,\, e^T(u_1^* + u_2) \leqslant 1, \end{aligned}$$

where $e = (1,1)^T$. Denote $y = (y_1, y_2)^T$ and $u = (u_1^T, u_2^T)^T$. The cost functional $\theta(y_1, y_2, u_1, u_2) = \theta(y, u)$ reads

$$y(T)^T L y(T) + l^T y(T) + \int_0^T y^T [Py + Su + h(t)] + u^T [Ru + d(t)]\, dt.$$

Here the matrices B_1, B_2, L, l, P, S and R are taken as in Chen & Wang (2014).

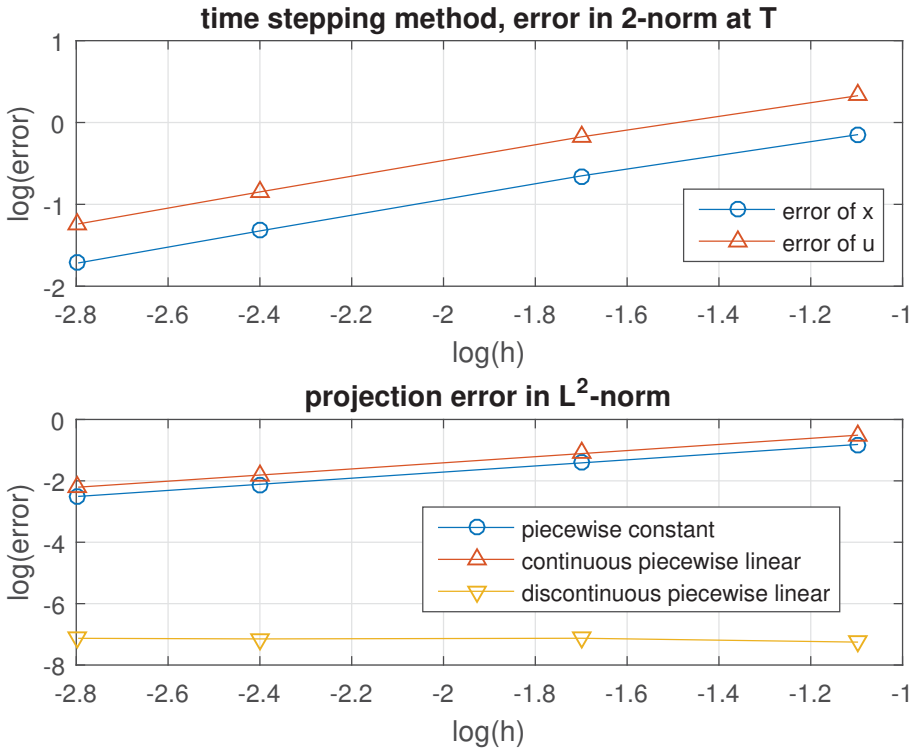

FIG. 3. Ratios of error for Galerkin and time stepping method (Example 5.1 with $T = 3\pi$)

The Pontryagin's minimum/maximum principle for each optimal control problem yields two coupled constrained Hamilton equations. By the vector $w \in \mathbb{R}^5$ of multipliers, it gives the following coupled system of ODE and a mixed linear complementarity problem (a little different to that of (1.1)):

$$\begin{array}{rcl} \dot{x}(t) & = & Ax(t)+Bu(t)+f(t) \\ 0 & = & Qx(t)+Mu(t)-C^T w(t)+g(t) \\ 0 & \leqslant & w(t) \perp Cu(t)-c \geqslant 0 \\ b & = & E_0 x(0)+E_T x(T). \end{array}$$

Here $x = (y_1, p_1, y_2, p_2)^T$, p_i is the costate of y_i. The detailed reformulation, and the data of the matrices A, B, Q, M, C, E_0, E_T, and the vectors b and c can be found in Chen & Wang (2014). Here we construct an exact solution (x,u) of the LCS by adapting the functions f and g. We mention that both the matrix M and the one defining the mixed LCP are positive semi-definite.

For this problem, our numerical method can be established in a very similar manner by using the variational formulation

$$\langle \hat{g} + \mathfrak{L}u, v-u \rangle_{\mathscr{L}^2} \geqslant 0, \qquad \forall v \in \mathscr{L}^2(0,T;\mathbb{R}^4) \times \mathscr{L}^2_+(0,T;\mathbb{R}^5).$$

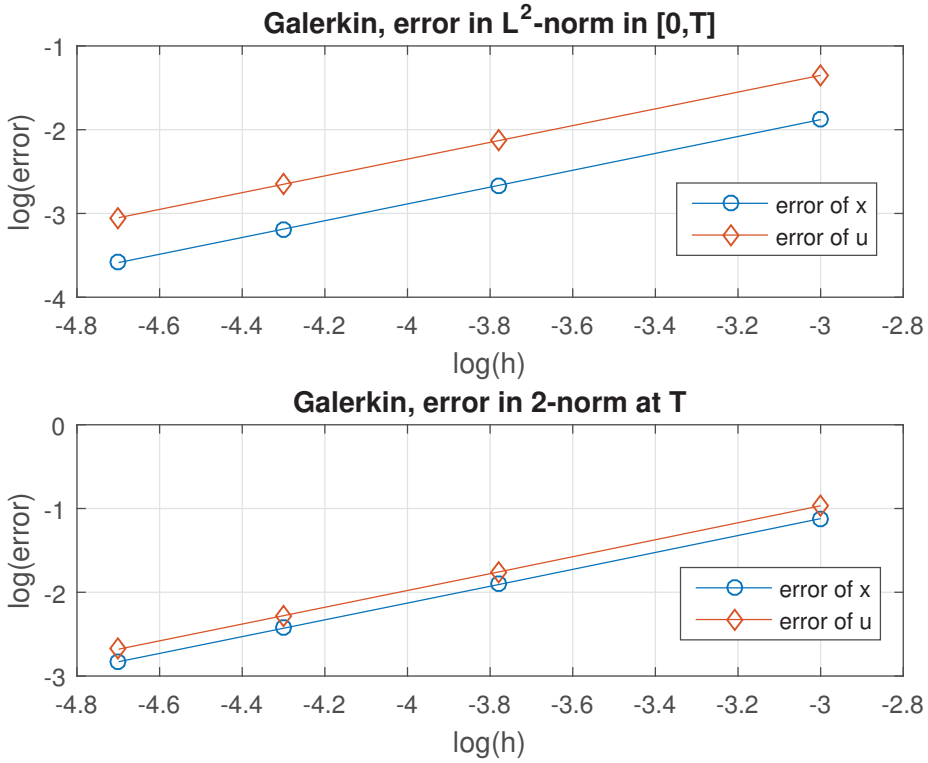

FIG. 4. Errors for Galerkin approximation (Example 5.2 with $T = 1$)

The Galerkin approximation yields a mixed linear complementarity problem of dimension $18N$, where N is the number of the subintervals.

We have the following observation on the numerical results.

(1) The logarithms of the errors for the Galerkin approximation and the time stepping method are plotted in Figures 4 and 5, respectively. Note that the four graphs in Figure 4 are approximately straight lines with the slopes all about 1.0076, this indicates the 1-order convergence of the Galerkin approximation. For this example, the precision of the Galerkin approximation is still much better than the time stepping method, whose convergence order is close to 0, as illustrated by the figures.

Notice that $\mathfrak{L}$ is not monotone, Theorem 4.1 can not be applied to provide error estimate and convergence order.

(2) If $\mathfrak{L}$ is not monotone, then the LCP(M^h, q^h) may have no solution or have multiple solutions, for which the LCP solvers may not perform well, their output could be far away from the true solution. This leads to the low order of convergence or even divergence.

Even when M is positive definite, $\mathfrak{L}$ is not necessarily monotone. The monotonicity is also dependent

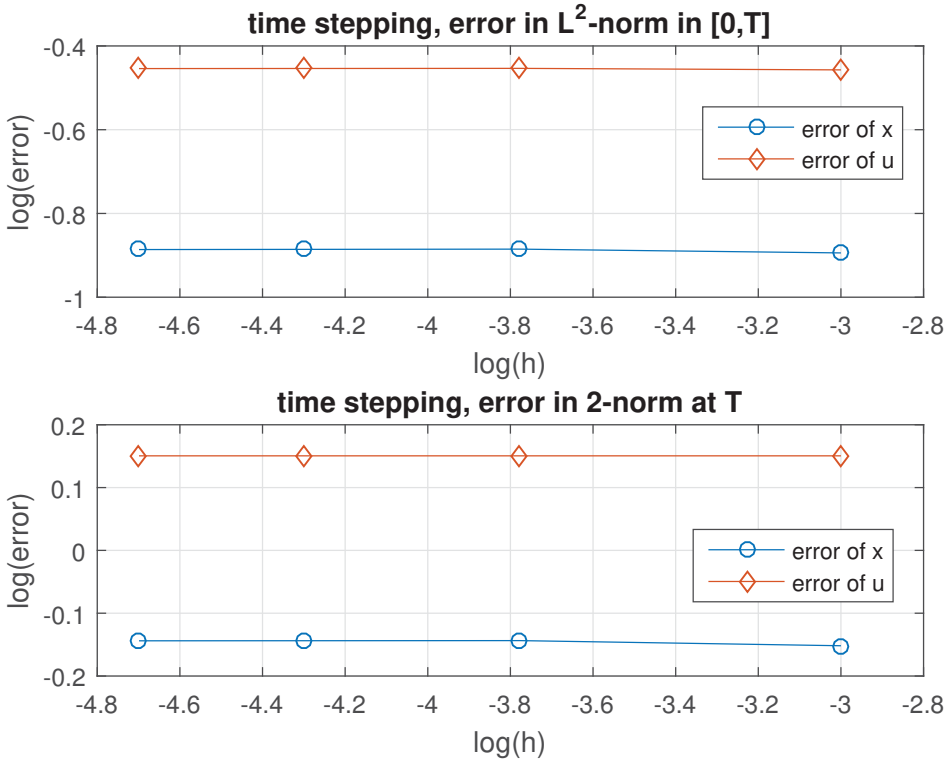

FIG. 5. Errors for time stepping method (Example 5.2 with $T = 1$)

of T. For illustrating this, we consider the regularized matrix $M + \lambda I$ with $\lambda = 0.1$, which is positive definite. It defines the operator $\mathfrak{L}_\lambda$ and then gives the matrix M^h_λ. If $\mathfrak{L}_\lambda$ is (strongly) monotone, then M^h_λ is positive (definite) semidefinite, and so its symmetric part has a nonnegative (positive) smallest eigenvalue $\sigma(T,h)$. We plot the values of $\sigma(T,h)$ in Figure 6 for different T and h. Note that for $T = 1, 0.5$, M^h_λ is indefinite, while positive definite for a smaller T.

6. Concluding remarks

This article reformulates the LCS into an LCP in an Hilbert space, and proposes a numerical method by using discontinuous Galerkin approximation. The method solves a finite-dimensional LCP(M^h, q^h) and uses the matrix exponential related matrices $\varphi_k(hA)$ to recover the state, these matrices are also used to construct the data of the LCP(M^h, q^h), which keeps a quite good fidelity to the LCS.

Numerical results show that the accuracy of the proposed method is much better than the time stepping method. Our method and the time stepping method need to evaluate $\varphi_k(hA)y$ and $(I - hA)^{-1}y$ respectively, for some column vector y. The evaluation of the former is not necessarily time-consuming

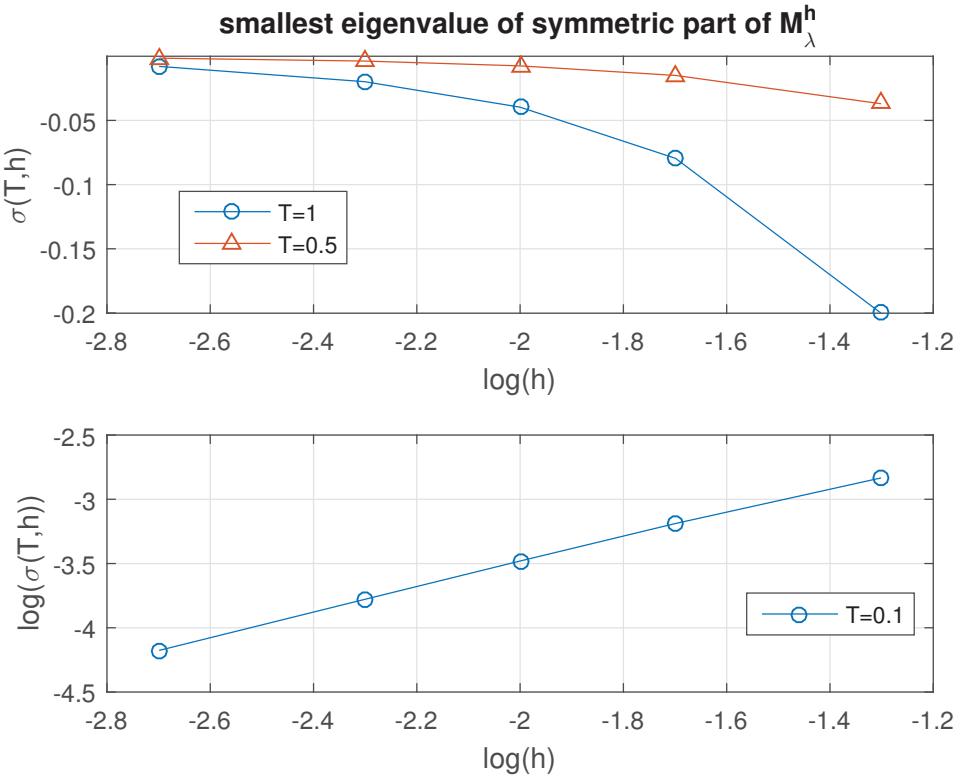

FIG. 6. smallest eigenvalue $\sigma(h)$ of the symmetric part of M^h (Example 5.2)

than that of the latter. See Higham & Al-Mohy (2010).

REFERENCES

ALEFELD, G.E. & CHEN, X. & POTRA, F.A. (1999) Numerical validation of solutions of linear complementarity problems. *Numer. Math.*, **83**, 1–23.

BROGLIATO, B. (2003) Some perspective on analysis and control of complementarity systems. *IEEE Trans. Autom. Control*, **48**, 918–935.

CHEN, X. & NASHED, Z. & QI, L. (1997) Convergence of Newton's method for singular smooth and nonsmooth equations using adaptive outer inverses. *SIAM J. Optim.*, **7**, 445–462.

CHEN, X. & MAHMOUD, S. (2008) Implicit Runge-Kutta methods for Lipschitz continuous ordinary differential equations. *SIAM J. Numer. Anal.*, **46**, 266–280.

CHEN, X. & WANG, Z. (2011) Error bounds for a differential linear variational inequality. *IMA J. Numer. Anal.*, **32**, 957–982.

CHEN, X. & WANG, Z. (2013) Convergence of regularized time-stepping methods for differential variational inequalities. *SIAM J. Optim.*, **23**, 1647–1671.

CHEN, X. & WANG, Z. (2014) Differential variational inequality approach to dynamic games with shared constraints. *Math. Program.*, **146**, 379–408.

CONWAY, J.B. (1985) *A Course in Functional Analysis*. Berlin: Springer.

DIRKSE, S.& FERRIS, M. & MUNSON, T. (2011) The PATH Solver. *http://pages.cs.wisc.edu/ ferris/path.html*.

FALK, R.S. (1974) Error estimates for the approximation of a class of variational inequalities. *Math. Compt.*, **28**, 963–971.

GWINNER, J. (2013) On a new class of differential variational inequalities and a stability result. *Math. Program.*, **139**, 205–221.

HAN, L. & CAMLIBEL, M.K. & PANG, J.-S. & HEEMELS. W.P.M.H. (2012) A unified numerical scheme for linear-quadratic optimal control problems with joint control and state constraints. *Optim. Meth. Softw.*, **27**, 761–799.

HEEMELS. W.P.M.H.& SCHUMACHER, J.M. & WEILAND, S. (2000) Linear complementarity systems. *SIAM J. Appl. Math.*, **60**, 1234–1269.

HIGHAM, N.J. & AL-MOHY, A.H. (2010) Computing matrix functions. *Acta Numer.*, **19**, 159–208.

HOCHBRUCK, M. & OSTERMANN, A. (2010) Exponential integrators. *Acta Numer.*, **19**, 209–286.

KUNKEL, P.& STÖVER, R. (2002) Symmetric collocation methods for linear differential-algebraic boundary value problems. *Numer. Math.*, **91**, 475–501.

PANG,J.-S. & QI, L. (1993) Nonsmooth equations: motivation and algorithms. *SIAM J. Optim.*, **3**, 443–465.

PANG,J.-S. & STEWART, D.E. (2008) Differential variational inequalities. *Math. Program.*, **113**, 345–424.

STEWART, D.E. (2006) Convolution complementarity problems with application to impact problems. *IMA J. Appl. Math.*, **71**, 92–119.

WANG, Z. & CHEN, X. (2016) Discretized form of the operator $\mathfrak{L}$ and state recovery. *supplement material to the submission*.

ZEIDLER, E. (1990) *Nonlinear Functional Analysis and Its Applications*. Leipzig: Teubner Verlag.

ZHONG, R.X.& SUMALEE, A.& FRIESZ, T.L.& LAM, WILLIAM H.K. (2011) Dynamic user equilibrium with side constraints for a traffic network: theoretical development and numerical solution algorithm. *Transpor. Res. B*, **45**, 1035–1061.

DISCRETIZED FORM OF THE OPERATOR $\mathfrak{L}$ AND STATE RECOVERY

ZHENGYU WANG* AND XIAOJUN CHEN†

1. Discretized form of the operator $\mathfrak{L}$. Given $A, E_0, E_T \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $Q \in \mathbb{R}^{m\times n}$, and $M \in \mathbb{R}^{m\times m}$. We remind us that the operator $\mathfrak{L}$, defined in $\mathscr{L}^2(0,T;\mathbb{R}^m)$, has the following form

$$(\mathfrak{L}u)(t) := Mu(t) + Q(\mathfrak{L}_1 u)(t) - Q(\mathfrak{L}_2 u)(t),$$

where

$$(\mathfrak{L}_1 u)(t) = \int_0^T e^{(t-s)A} B u(s) ds$$

and

$$(\mathfrak{L}_2 u)(t) = \int_0^T e^{tA}(E_0 + E_T e^{TA})^{-1} E_T e^{(T-s)A} B u(s) ds.$$

Below we derive the discretized form of $\mathfrak{L}$, represented by the matrix $M^h = (M^h_{ij})$ with $M^h_{ji} := \langle \mathfrak{L}\varphi_i, \varphi_j\rangle_{\mathscr{L}^2}/h$, in the finite-dimensional subspace $\mathscr{U}_h$ spanned by the basis functions

$$\psi_{i,j,k}(t) \;:=\; \hat{\psi}_j\left(\frac{t-t_{k-1}}{h}\right)\cdot \chi_k(t)\cdot e_i,$$

where $\chi_k(\cdot)$ is the characteristic function in I_k and e_i denotes the m-dimensional i-th unit coordinate vector, and $\hat{\psi}_1(t) = 1-t$, $\hat{\psi}_2(t) = t$. We show that M^h has the following form

$$\begin{aligned} M^h \;=\; & \frac{1}{6} I_N \otimes \begin{pmatrix} 2M & M \\ M & 2M \end{pmatrix} - h\sum_{k',k=1}^{N} e_{k'} e_k^T \otimes \begin{pmatrix} (J_{k',k})_{11} & (J_{k',k})_{12} \\ (J_{k',k})_{21} & (J_{k',k})_{22} \end{pmatrix} \\ & + \frac{h}{2}\sum_{k'=2}^{N}\sum_{k=1}^{k'-1} e_{k'} e_k^T \otimes \begin{pmatrix} Q(G_1-G_2)B & QG_2B \\ Q(G_1-G_2)B & QG_2B \end{pmatrix} \\ & + hI_N \otimes \begin{pmatrix} Q(G_3-G_4)B & QG_4B \\ Q(G_2-2G_3+G_4)B & Q(G_3-G_4)B \end{pmatrix}, \end{aligned}$$

where e_k is the k-th N-dimensional unit coordinate vector, and for $k', k = 1,\ldots,N$:

$$\begin{aligned} (J_{k',k})_{11} &= QG_0^{k'-1}G_2(E_0+E_Te^{TA})^{-1}E_TG_0^{N-k}(G_1-G_2)B \\ (J_{k',k})_{12} &= QG_0^{k'-1}G_2(E_0+E_Te^{TA})^{-1}E_TG_0^{N-k}G_2B \\ (J_{k',k})_{21} &= QG_0^{k'-1}(G_1-G_2)(E_0+E_Te^{TA})^{-1}E_TG_0^{N-k}(G_1-G_2)B \\ (J_{k',k})_{22} &= QG_0^{k'-1}(G_1-G_2)(E_0+E_Te^{TA})^{-1}E_TG_0^{N-k}G_2B. \end{aligned}$$

*Department of Mathematics, Nanjing University, China. email: zywang@nju.edu.cn
†Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong. email: maxjchen@polyu.edu.hk. The author's work was supported in part by Hong Kong Research Grants Council PolyU153001/14P.

Proof For $t \le t_{k-1}$, we have for $j = 1, 2$:

$$\mathfrak{L}_1\psi_{i,j,k}(t) \;=\; \int_0^t e^{(t-s)A}B\psi_j\left(\frac{s-t_{k-1}}{h}\right)\cdot\chi_{I_k}(s)e_i ds \;=\; 0.$$

For $t \in (t_{k-1}, t_k)$, we have

$$\begin{aligned}
\mathfrak{L}_1\psi_{i,1,k}(t) &= \int_0^t e^{(t-s)A}B\psi_1\left(\frac{s-t_{k-1}}{h}\right)\chi_{I_k}(s)e_i ds \;=\; \int_{t_{k-1}}^t e^{(t-s)A}B\left(1-\frac{s-t_{k-1}}{h}\right)e_i ds\\
&= \left(\int_0^{t-t_{k-1}} e^{(t-t_{k-1}-\tau)A}\left(1-\frac{\tau}{h}\right)d\tau\right)Be_i\\
&= \left[(t-t_{k-1})\varphi_1((t-t_{k-1})A) - \frac{1}{h}(t-t_{k-1})^2\varphi_2((t-t_{k-1})A)\right]Be_i\\
&= h\left[\left(\frac{t-t_{k-1}}{h}\right)\varphi_1((t-t_{k-1})A) - \left(\frac{t-t_{k-1}}{h}\right)^2\varphi_2((t-t_{k-1})A)\right]Be_i,\\
\mathfrak{L}_1\psi_{i,2,k}(t) &= \int_0^t e^{(t-s)A}B\psi_2\left(\frac{s-t_{k-1}}{h}\right)\chi_{I_k}(s)e_i ds \;=\; \int_{t_{k-1}}^t e^{(t-s)A}B\frac{s-t_{k-1}}{h}e_i ds\\
&= \left(\int_0^{t-t_{k-1}} e^{(t-t_{k-1}-\tau)A}\frac{\tau}{h}d\tau\right)Be_i \;=\; \frac{1}{h}(t-t_{k-1})^2\varphi_2((t-t_{k-1})A)Be_i\\
&= h\left(\frac{t-t_{k-1}}{h}\right)^2\varphi_2((t-t_{k-1})A)Be_i.
\end{aligned}$$

For $t \ge t_k$, we have

$$\begin{aligned}
\mathfrak{L}_1\psi_{i,1,k}(t) &= \int_0^t e^{(t-s)A}B\psi_1\left(\frac{s-t_{k-1}}{h}\right)\chi_{I_k}(s)e_i ds\\
&= \int_{t_{k-1}}^{t_k} e^{(t-s)A}B\left(1-\frac{s-t_{k-1}}{h}\right)e_i ds \;=\; h\beta_1(hA)Be_i,\\
\mathfrak{L}_1\psi_{i,2,k}(t) &= \int_0^t e^{(t-s)A}B\psi_2\left(\frac{s-t_{k-1}}{h}\right)\chi_{I_k}(s)e_i ds\\
&= \int_{t_{k-1}}^{t_k} e^{(t-s)A}B\left(\frac{s-t_{k-1}}{h}\right)te_i ds \;=\; h\beta_2(hA)Be_i.
\end{aligned}$$

Now for $i = 1, \ldots, m$, $j = 1, 2$, and $k = 1, \ldots, N$, we compute

$$\begin{aligned}
\langle Q\mathfrak{L}_1\psi_{i,j,k}, \psi_{i',j',k'}\rangle_{\mathscr{L}^2} &= \int_0^T \chi_{I_{k'}}(t)e_{i'}^T\psi_{j'}\left(\frac{t-t_{k'-1}}{h}\right)Q\mathfrak{L}_1\psi_{i,j,k}(t)dt\\
&= e_{i'}^T\int_{t_{k'-1}}^{t_{k'}}\psi_{j'}\left(\frac{t-t_{k'-1}}{h}\right)Q\mathfrak{L}_1\psi_{i,j,k}(t)dt.
\end{aligned}$$

For $k > k'$, when $t \in [t_{k'-1}, t_{k'}]$, we have $t \le t_{k-1}$, and $Q\mathfrak{L}_1\psi_{i,j,k}(t) = 0$, and therefore

$$\langle Q\mathfrak{L}_1\psi_{i,j,k}, \psi_{i',j',k'}\rangle_{\mathscr{L}^2} \;=\; 0.$$

For $k = k'$, we obtain

$$\begin{aligned}
&\langle Q\mathfrak{L}_1\psi_{i,1,k}, \psi_{i',1,k}\rangle_{\mathscr{L}^2} \;=\; e_{i'}^T \int_{t_{k-1}}^{t_k} \psi_1\left(\frac{t-t_{k-1}}{h}\right) Q\mathfrak{L}_1\psi_{i,1,k}(t)dt \\
=\;& he_{i'}^T Q \int_{t_{k-1}}^{t_k} \psi_1\left(\frac{t-t_{k-1}}{h}\right)\left[\left(\frac{t-t_{k-1}}{h}\right)\varphi_1((t-t_{k-1})A) - \left(\frac{t-t_{k-1}}{h}\right)^2 \varphi_2((t-t_{k-1})A)\right] Be_i dt \\
=\;& h^2 e_{i'}^T Q\left(\int_0^1 (1-\tau)\left[\tau\varphi_1(\tau hA) - \tau^2\varphi_2(\tau hA)\right] d\tau\right) Be_i \\
=\;& h^2 e_{i'}^T Q\left(\int_0^1 (1-\tau)\left[\sum_{n=1}^{\infty}\frac{\tau^n}{n!}(hA)^{n-1} - \sum_{n=2}^{\infty}\frac{\tau^n}{n!}(hA)^{n-2}\right] d\tau\right) Be_i \\
=\;& h^2 e_{i'}^T Q\left(\sum_{n=1}^{\infty}\int_0^1 (1-\tau)\frac{\tau^n}{n!}(h_kA)^{n-1}d\tau - \sum_{n=2}^{\infty}\int_0^1 (1-\tau)\frac{\tau^n}{n!}(h_kA)^{n-2}d\tau\right) Be_i \\
=\;& h^2 e_{i'}^T Q\left(\sum_{n=1}^{\infty}\frac{1}{(n+2)!}(hA)^{n-1} - \sum_{n=2}^{\infty}\frac{1}{(n+2)!}(hA)^{n-2}\right) Be_i \;=\; h^2 e_{i'}^T Q\left(\varphi_3(hA) - \varphi_4(hA)\right) Be_i,
\end{aligned}$$

$$\begin{aligned}
&\langle Q\mathfrak{L}_1\psi_{i,1,k}, \psi_{i',2,k}\rangle_{\mathscr{L}^2} \;=\; e_{i'}^T \int_{t_{k-1}}^{t_k} \psi_2\left(\frac{t-t_{k-1}}{h}\right) Q\mathfrak{L}_1\psi_{i,1,k}(t)dt \\
=\;& he_{i'}^T Q \int_{t_{k-1}}^{t_k} \psi_2\left(\frac{t-t_{k-1}}{h}\right)\left[\left(\frac{t-t_{k-1}}{h}\right)\varphi_1((t-t_{k-1})A) - \left(\frac{t-t_{k-1}}{h}\right)^2 \varphi_2((t-t_{k-1})A)\right] Be_i dt \\
=\;& h^2 e_{i'}^T Q\left(\int_0^1 \tau\left[\tau\varphi_1(\tau hA) - \tau^2\varphi_2(\tau hA)\right] d\tau\right) Be_i \;=\; h^2 e_{i'}^T Q\left(\int_0^1 \tau\left[\sum_{n=1}^{\infty}\frac{\tau^n}{n!}(hA)^{n-1} - \sum_{n=2}^{\infty}\frac{\tau^n}{n!}(hA)^{n-2}\right] d\tau\right) Be_i \\
=\;& h^2 e_{i'}^T Q\left(\sum_{n=1}^{\infty}\frac{1}{n!(n+2)}(hA)^{n-1} - \sum_{n=2}^{\infty}\frac{1}{n!(n+2)}(hA)^{n-2}\right) Be_i \;=\; h^2 e_{i'}^T Q\left(\varphi_2(hA) - 2\varphi_3(hA) + \varphi_4(hA)\right) Be_i,
\end{aligned}$$

$$\begin{aligned}
&\langle Q\mathfrak{L}_1\psi_{i,2,k}, \psi_{i',1,k}\rangle_{\mathscr{L}^2} \;=\; e_{i'}^T \int_{t_{k-1}}^{t_k} \psi_1\left(\frac{t-t_{k-1}}{h}\right) Q\mathfrak{L}_1\psi_{i,1,k}(t)dt \\
=\;& he_{i'}^T Q \int_{t_{k-1}}^{t_k} \psi_1\left(\frac{t-t_{k-1}}{h}\right)\left(\frac{t-t_{k-1}}{h}\right)^2 \varphi_2((t-t_{k-1})A)Be_i dt \;=\; h^2 e_{i'}^T Q\left(\int_0^1 (1-\tau)\tau^2\varphi_2(\tau hA)d\tau\right) Be_i \\
=\;& h^2 e_{i'}^T Q\left(\int_0^1 (1-\tau)\sum_{n=2}^{\infty}\frac{\tau^n}{n!}(hA)^{n-2}d\tau\right) Be_i \;=\; h^2 e_{i'}^T Q\left(\sum_{n=2}^{\infty}\frac{1}{(n+2)!}(hA)^{n-2}\right) Be_i \;=\; h^2 e_{i'}^T Q\varphi_4(hA)Be_i,
\end{aligned}$$

and

$$\begin{aligned}
&\langle Q\mathfrak{L}_1\psi_{i,2,k}, \psi_{i',2,k}\rangle_{\mathscr{L}^2} \;=\; e_{i'}^T \int_{t_{k-1}}^{t_k} \psi_2\left(\frac{t-t_{k-1}}{h}\right) Q\mathfrak{L}_1\psi_{i,1,k}(t)dt \\
=\;& he_{i'}^T Q \int_{t_{k-1}}^{t_k} \psi_2\left(\frac{t-t_{k-1}}{h}\right)\left(\frac{t-t_{k-1}}{h}\right)^2 \varphi_2((t-t_{k-1})A)Be_i dt \;=\; h^2 e_{i'}^T Q\left(\int_0^1 \tau^3\varphi_2(\tau hA)d\tau\right) Be_i \\
=\;& h^2 e_{i'}^T Q\left(\int_0^1 \sum_{n=2}^{\infty}\frac{\tau^{n+1}}{n!}(hA)^{n-2}d\tau\right) Be_i \;=\; h^2 e_{i'}^T Q\sum_{n=2}^{\infty}\frac{1}{n!(n+2)}(hA)^{n-2}Be_i \\
=\;& h^2 e_{i'}^T Q\left(\sum_{n=2}^{\infty}\left(\frac{1}{(n+1)!} - \frac{1}{(n+2)!}\right)(hA)^{n-2}\right) Be_i \;=\; h^2 e_{i'}^T Q(\varphi_3(hA) - \varphi_4(hA))Be_i.
\end{aligned}$$

For $k < k'$, when $t \in [t_{k'-1}, t_{k'}]$, we have $t \geq t_k$, and therefore $\mathfrak{L}_1\psi_{i,j,k}(t) = h\beta_j(hA)Be_i$ is constant therein. We obtain

$$\begin{aligned}\langle Q\mathfrak{L}_1\psi_{i,j,k}, \psi_{i',j',k'}\rangle_{\mathscr{L}^2} &= \int_{t_{k'-1}}^{t_{k'}} e_{i'}^T\psi_{j'}\left(\frac{t-t_{k'-1}}{h}\right) Q\mathfrak{L}_1\psi_{i,j,k}(t)dt \\ = \; he_{i'}^TQ\beta_j(hA)Be_i\int_{t_{k'-1}}^{t_{k'}}\psi_{j'}\left(\frac{t-t_{k'-1}}{h}\right)dt &= \frac{1}{2}h^2e_{i'}^TQ\beta_j(hA)Be_i,\end{aligned}$$

namely,

$$\begin{aligned}\langle Q\mathfrak{L}_1\psi_{i,1,k}, \psi_{i',1,k'}\rangle_{\mathscr{L}^2} &= \frac{1}{2}h^2e_{i'}^TQ\left[\varphi_1(hA)-\varphi_2(hA)\right]Be_i \\ \langle Q\mathfrak{L}_1\psi_{i,1,k}, \psi_{i',2,k'}\rangle_{\mathscr{L}^2} &= \frac{1}{2}h^2e_{i'}^TQ\left[\varphi_1(hA)-\varphi_2(hA)\right]Be_i \\ \langle Q\mathfrak{L}_1\psi_{i,2,k}, \psi_{i',1,k'}\rangle_{\mathscr{L}^2} &= \frac{1}{2}h^2e_{i'}^TQ\varphi_2(hA)Be_i \\ \langle Q\mathfrak{L}_1\psi_{i,2,k}, \psi_{i',2,k'}\rangle_{\mathscr{L}^2} &= \frac{1}{2}h^2e_{i'}^TQ\varphi_2(hA)Be_i.\end{aligned}$$

Summarizing, we have

$$\langle Q\mathfrak{L}_1\psi_{i,j,k}, \psi_{i',j',k'}\rangle_{\mathscr{L}^2} = \begin{cases} 0 & \text{when } k > k' \\ h^2\left[Q\left(\varphi_3(hA)-\varphi_4(hA)\right)B\right]_{i'i} & \text{when } k=k',\ j=1,\ j'=1 \\ h^2\left[Q\left(\varphi_2(hA)-2\varphi_3(hA)+\varphi_4(hA)\right)B\right]_{i'i} & \text{when } k=k',\ j=1,\ j'=2 \\ h^2\left[Q\varphi_4(hA)B\right]_{i'i} & \text{when } k=k',\ j=2,\ j'=1 \\ h^2\left[Q(\varphi_3(hA)-\varphi_4(hA))B\right]_{i'i} & \text{when } k=k',\ j=2,\ j'=2 \\ \frac{1}{2}h^2\left[Q\left(\varphi_1(hA)-\varphi_2(hA)\right)B\right]_{i'i} & \text{when } k<k',\ j=1,\ j'=1 \\ \frac{1}{2}h^2\left[Q\left[\varphi_1(hA)-\varphi_2(hA)\right]B\right]_{i'i} & \text{when } k<k',\ j=1,\ j'=2 \\ \frac{1}{2}h^2\left[Q\varphi_2(hA)B\right]_{i'i} & \text{when } k<k',\ j=2,\ j'=1 \\ \frac{1}{2}h^2\left[Q\varphi_2(hA)B\right]_{i'i} & \text{when } k<k',\ j=2,\ j'=2 \end{cases}$$

Collecting the entries $\langle Q\mathfrak{L}_1^h\psi_{i,j,k}, \psi_{i',j',k'}\rangle_{\mathscr{L}^2}$ in the matrix M_1^h by rearranging the

basis functions, and replacing $\varphi_k(hA)$ by G_k, we get

$$\begin{aligned} M_1^h &:= h^2 I_N \otimes \begin{pmatrix} Q(G_3-G_4)B & QG_4B \\ Q(G_2-2G_3+G_4)B & Q(G_3-G_4)B \end{pmatrix} \\ &\quad + \frac{h^2}{2} \begin{pmatrix} 0 & & & \\ 1 & \ddots & & \\ \vdots & \ddots & \ddots & \\ 1 & \cdots & 1 & 0 \end{pmatrix}_N \otimes \begin{pmatrix} Q(G_1-G_2)B & QG_2B \\ Q(G_1-G_2)B & QG_2B \end{pmatrix}. \end{aligned}$$

For the operator $\mathfrak{L}_2^h$, we can compute for $i = 1, \ldots, m$, $j = 1, 2$ and $k = 1, \ldots, N$:

$$\begin{aligned} (\mathfrak{L}_2\psi_{i,j,k})(t) &= e^{tA}(E_0 + E_T e^{TA})^{-1} E_T \int_0^T e^{(T-s)A} B\psi_{i,j,k}(s) ds \\ &= e^{tA}(E_0 + E_T e^{TA})^{-1} E_T \int_{t_{k-1}}^{t_k} e^{(T-s)A} B\psi_j\left(\frac{s-t_{k-1}}{h}\right) e_i ds \\ &= he^{tA}(E_0 + E_T e^{TA})^{-1} E_T \int_0^1 e^{(T-t_{k-1}-\tau h)A} B\psi_j(\tau) e_i d\tau \\ &= he^{tA}(E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \left(\int_0^1 e^{(1-\tau)h_k A}\psi_j(\tau) d\tau\right) Be_i \\ &= he^{tA}(E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \beta_j(hA) right) Be_i \end{aligned}$$

and therefore

$$\begin{aligned} \langle Q\mathfrak{L}_2\psi_{i,j,k}, \psi_{i',1,k'}\rangle_{\mathscr{L}^2} &= he_{i'}^T \int_{t_{k'-1}}^{t_{k'}} Qe^{tA}(E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \beta_j(hA) Be_i \psi_1\left(\frac{t-t_{k'-1}}{h}\right) dt \\ &= he_{i'}^T Q \left(\int_{t_{k'-1}}^{t_{k'}} e^{tA}\psi_1\left(\frac{t-t_{k'-1}}{h}\right) dt\right) (E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \beta_j(hA) Be_i \\ &= h^2 e_{i'}^T Q \left(\int_0^1 e^{(t_{k'-1}+\tau h)A}\psi_1(\tau) dt\right) (E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \beta_j(hA) Be_i \\ &= h^2 e_{i'}^T Q e^{t_{k'}A} \left(\int_0^1 e^{-(1-\tau)hA}\psi_1(\tau) dt\right) (E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \beta_j(hA) Be_i \\ &= h^2 e_{i'}^T Q e^{t_{k'}A} \beta_1(-hA)(E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \beta_j(hA) Be_i \\ &= h^2 \left[Qe^{t_{k'-1}A} \beta_2(hA)(E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \beta_j(hA) B\right]_{i'i}, \end{aligned}$$

and

$$\begin{aligned} \langle Q\mathfrak{L}_2\psi_{i,j,k}, \psi_{i',2,k'}\rangle_{\mathscr{L}^2} &= he_{i'}^T \int_{t_{k'-1}}^{t_{k'}} Qe^{tA}(E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \beta_j(hA) Be_i \psi_2\left(\frac{t-t_{k'-1}}{h}\right) dt \\ &= he_{i'}^T Q \left(\int_{t_{k'-1}}^{t_{k'}} e^{tA}\psi_2\left(\frac{t-t_{k'-1}}{h}\right) dt\right) (E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \beta_j(hA) Be_i \\ &= h^2 e_{i'}^T Q \left(\int_0^1 e^{(t_{k'-1}+\tau h)A}\psi_2(\tau) dt\right) (E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \beta_j(hA) Be_i \\ &= h^2 e_{i'}^T Q e^{t_{k'}A} \left(\int_0^1 e^{-(1-\tau)hA}\psi_2(\tau) dt\right) (E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \beta_j(hA) Be_i \\ &= h^2 e_{i'}^T Q e^{t_{k'}A} \beta_2(-hA)(E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \beta_j(hA) Be_i \\ &= h^2 \left[Qe^{t_{k'-1}A}(\varphi_1(hA) - \varphi_2(hA))(E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \beta_j(hA) B\right]_{i'i} \\ &= h^2 \left[Qe^{t_{k'-1}A} \beta_1(hA)(E_0 + E_T e^{TA})^{-1} E_T e^{(T-t_k)A} \beta_j(hA) B\right]_{i'i}. \end{aligned}$$

Collecting the entries $\langle Q\mathfrak{L}_2^h\psi_{i,j,k},\psi_{i',j',k'}\rangle_{\mathscr{L}^2}$ in a matrix by rearranging the basis functions, we get

$$M_2^h := h^2\sum_{k',k=1}^{N} e_{k'}e_k^T\otimes\begin{pmatrix}(J_{k',k})_{11} & (J_{k',k})_{12}\\ (J_{k',k})_{21} & (J_{k',k})_{22}\end{pmatrix},$$

where from $\beta_1(z)=\varphi_1(z)-\varphi_2(z)$ and $\beta_2(z)=\varphi_2(z)$, we have

$$\begin{aligned}
(J_{k',k})_{11} &= Qe^{t_{k'-1}A}\beta_2(hA)(E_0+E_Te^{TA})^{-1}E_Te^{(T-t_k)A}\beta_1(hA)B\\
&= QG_0^{k'-1}G_2(E_0+E_Te^{TA})^{-1}E_TG_0^{N-k}(G_1-G_2)B\\
(J_{k',k})_{12} &= Qe^{t_{k'-1}A}\beta_2(hA)(E_0+E_Te^{TA})^{-1}E_Te^{(T-t_k)A}\beta_2(hA)B\\
&= QG_0^{k'-1}G_2(hA)(E_0+E_Te^{TA})^{-1}E_TG_0^{N-k}G_2B\\
(J_{k',k})_{21} &= Qe^{t_{k'-1}A}\beta_1(hA)(E_0+E_Te^{TA})^{-1}E_Te^{(T-t_k)A}\beta_1(hA)B\\
&= QG_0^{k'-1}(G_1-G_2)(E_0+E_Te^{TA})^{-1}E_TG_0^{N-k}(G_1-G_2)B\\
(J_{k',k})_{22} &= Qe^{t_{k'-1}A}\beta_1(hA)(E_0+E_Te^{TA})^{-1}E_Te^{(T-t_k)A}\beta_2(hA)B\\
&= QG_0^{k'-1}(G_1-G_2)(E_0+E_Te^{TA})^{-1}E_TG_0^{N-k}G_2B.
\end{aligned}$$

Moreover, from

$$\begin{aligned}
&\langle M\psi_{i,j,k},\psi_{i',j',k'}\rangle_{\mathscr{L}^2}\\
=\;& e_{i'}^TMe_i\int_0^T\psi_j\left(\frac{t-t_{k-1}}{t_k-t_{k-1}}\right)\cdot\chi_{I_k}(t)\psi_{j'}\left(\frac{t-t_{k'-1}}{t_{k'}-t_{k'-1}}\right)\cdot\chi_{I_{k'}}(t)dt
\end{aligned}$$

it follows that $\langle M\psi_{i,j,k},\psi_{i',j',k'}\rangle_{\mathscr{L}^2}=0$ when $k\neq k'$, and

$$\begin{aligned}
\langle M\psi_{i,j,k},\psi_{i',j',k}\rangle_{\mathscr{L}^2} &= e_{i'}^TMe_i\int_{t_{k-1}}^{t_k}\psi_j\left(\frac{t-t_{k-1}}{t_k-t_{k-1}}\right)\psi_{j'}\left(\frac{t-t_{k-1}}{t_k-t_{k-1}}\right)dt\\
&= he_{i'}^TMe_i\int_0^1\psi_j(s)\psi_{j'}(s)ds,
\end{aligned}$$

and therefore

$$\langle M\psi_{i,j,k},\psi_{i',j',k'}\rangle_{\mathscr{L}^2}=\begin{cases}0 & \text{when } k\neq k'\\ \dfrac{1}{3}h(M)_{i'i} & \text{when } k=k',\ j=j'\\ \dfrac{1}{6}h(M)_{i'i} & \text{when } k=k',\ j\neq j'\end{cases}$$

Collecting the entries $\langle M)\psi_{i,j,k},\psi_{i',j',k'}\rangle_{\mathscr{L}^2}$ in a matrix by rearrangement, we get

$$M_3^h := \frac{h}{6}I_N\otimes\begin{pmatrix}2M & M\\ M & 2M\end{pmatrix}. \tag{1.1}$$

Finally, let $M^h=(M_1^h-M_2^h+M_3^h)/h$. This completes the proof for the form of the matrix M^h.

2. Justification of the state recovery. Denote

$$\hat{x}^{0,h} = (E_0+E_Te^{TA})^{-1}\left(b-E_T\int_0^T e^{(T-s)A}f(s)ds\right)$$

and

$$\hat{f}(t) \;=\; e^{tA}\hat{x}^{0,h} + \int_0^t e^{(t-s)A} f(s)ds.$$

Let the approximate control u^h be given by the solution of the LCP$(\mathfrak{L}, \hat{g})$. It is natural to approximate the state x by

$$x^h(t) = \hat{f}(t) + (\mathfrak{L}_1 u^h)(t) - (\mathfrak{L}_2 u^h)(t). \tag{2.1}$$

Note $(\mathfrak{L}_1\varphi_{i,j,k})(t_k) = 0$ when $k < k$. We compute for $k = 1, \ldots, N$:

$$\begin{aligned}
(\mathfrak{L}_1 u^h)(t_k) &= \sum_{i,j,k} c_{i,j,k}(\mathfrak{L}_1\psi_{i,j,k})(t_k) \;=\; h \sum_{i,j,k\le k} c_{i,j,k}\beta_j(hA)Be_i \\
&= h\left(\beta_1(hA)B\sum_{j=1}^{k} u^h_{j,1} + \beta_2(hA)B\sum_{j=1}^{k} u^h_{j,2}\right) \\
(\mathfrak{L}_2 u^h)(t_j) &= \sum_{i,j,k} c_{i,j,k}(\mathfrak{L}_2\psi_{i,j,k})(t_\nu) \\
&= he^{t_\nu A}(E_0 + E_T e^{TA})^{-1}E_T \sum_{i,j,k} c_{i,j,k} e^{(T-t_k)A}\beta_j(hA)Be_i \\
&= he^{t_\nu A}(E_0 + E_T e^{TA})^{-1}E_T \sum_{k=1}^{N} e^{(T-t_k)A}\left(\beta_1(hA)Bu^h_{k,1} + \beta_2(hA)Bu^h_{k,2}\right).
\end{aligned}$$

The above computation enables us to approximate the state x at the mesh points by $x^h_k = x^h(t_k)$, where $\nu = 0, 1, \cdots N$ and

$$\begin{aligned}
x^h_k \;=\; & G_0^k\hat{x}^{0,h} + \int_0^{t_k} e^{(t_k-s)A}f(s)ds + h\left(G_1B\sum_{j=1}^{k} u^h_{j,1} + G_2B\sum_{j=1}^{k}(u^h_{j,2} - u^h_{j,1})\right) \\
& -hG_0^k(E_0 + E_T e^{TA})^{-1}E_T\sum_{k=1}^{N} G_0^{N-k}\left[(G_1 - G_2)Bu^h_{k,1} + G_2Bu^h_{k,2}\right].
\end{aligned}$$