ESAIM: COCV 30 (2024) 62
https://doi.org/10.1051/cocv/2024051

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

A NEW DIAGONALIZATION BASED METHOD FOR PARALLEL-IN-TIME SOLUTION OF LINEAR-QUADRATIC OPTIMAL CONTROL PROBLEMS

Matthias Heinkenschloss[1,*] and Nathaniel J. Kroeger[2]

Abstract. A new diagonalization technique for the parallel-in-time solution of linear-quadratic optimal control problems with time-invariant system matrices is introduced. The target problems are often derived from a semi-discretization of a Partial Differential Equation (PDE)-constrained optimization problem. The solution of large-scale time dependent optimal control problems is computationally challenging as the states, controls, and adjoints are coupled to each other throughout the whole time domain. This computational difficulty motivates the use of parallel-in-time methods. For time-periodic problems our diagonalization efficiently transforms the discretized optimality system into n_t (=number of time steps) decoupled complex valued $2n_y \times 2n_y$ systems, where n_y is the dimension of the state space. These systems resemble optimality systems corresponding to a steady-state version of the optimal control problem and they can be solved in parallel across the time steps, but are complex valued. For optimal control problems with initial value state equations a direct solution *via* diagonalization is not possible, but an efficient preconditioner can be constructed from the corresponding time periodic optimal control problem. The preconditioner can be efficiently applied parallel-in-time using the diagonalization technique. The observed number of preconditioned GMRES iterations is small and insensitive to the size of the problem discretization.

Mathematics Subject Classification. 65M55, 65M12, 65M15, 65Y05, 49M41.

Received September 12, 2022. Accepted July 12, 2024.

1. Introduction

This paper introduces a new diagonalization technique for the parallel-in-time solution of linear-quadratic optimal control problems with time-invariant system matrices of the following type. Given

$$
\begin{aligned}
&\text{a symmetric positive semidefinite } Q \in \mathbb{R}^{n_y \times n_y}, \text{ a symmetric positive definite } R \in \mathbb{R}^{n_u \times n_u},\\
&\text{an invertible } M \in \mathbb{R}^{n_y \times n_y}, A \in \mathbb{R}^{n_y \times n_y}, B \in \mathbb{R}^{n_y \times n_u}, y_0 \in \mathbb{R}^{n_y},\\
&c \in L^2(0,T;\mathbb{R}^{n_y}), d \in L^2(0,T;\mathbb{R}^{n_u}), \text{ and } f \in L^2(0,T;\mathbb{R}^{n_y}),
\end{aligned}
\tag{1.1}
$$

Keywords and phrases: Optimal control, diagonalization, parallel-in-time, preconditioning.

[1] Department of Computational Applied Mathematics & Operations Research, MS-134, Rice University, 6100 Main Street, Houston, TX 77005-1892, USA.

[2] Department of Statistics, MS-138, Rice University, 6100 Main Street, Houston, TX 77005-1892, USA.

[*] Corresponding author: heinken@rice.edu

we want to solve

$$\min \quad \int_0^T \frac{1}{2} y(t)^T Q\, y(t) + c(t)^T y(t) + \frac{1}{2} u(t)^T R\, u(t) + d(t)^T u(t)\, \mathrm{d}t, \tag{1.2a}$$

$$\text{s.t.} \quad M \frac{\mathrm{d}}{\mathrm{d}t} y(t) + A\, y(t) + B\, u(t) = f(t), \quad t \in (0, T), \tag{1.2b}$$

$$M\, y(0) = M\, y_0, \tag{1.2c}$$

or (1.2) with (1.2c) replaced by the time periodic condition

$$M\, y(0) = M\, y(T). \tag{1.2d}$$

We are interested in large-scale versions of (1.2), which are often derived from a semi-discretization of a Partial Differential Equation (PDE)-constrained optimization problem. The solution of such problems is computationally challenging as the states, controls, and adjoints are coupled to each other throughout the whole time domain. This computational difficulty motivates the use of parallel-in-time methods. For time periodic problems, our diagonalization efficiently transforms the discretized optimality system into n_t (=number of time steps) decoupled complex valued $2n_y \times 2n_y$ systems, where n_y is the dimension of the state space. These systems resemble optimality systems corresponding to a steady-state version of the optimal control problem and they can be solved in parallel across the time steps. For optimal control problems with initial value state equations, a direct solution *via* diagonalization is not possible, but an efficient preconditioner can be constructed from the corresponding time periodic optimal control problem. The preconditioner can be efficiently applied parallel-in-time using the diagonalization technique. The observed number of preconditioned GMRES iterations is small and insensitive to the size of the problem discretization.

Our proposed approach is motivated by the diagonalization approach in [1], which requires $A = A^T$, symmetric positive definite $Q = M = M^T$, $B = \pm M$, and $R = \alpha M$, $\alpha > 0$. These requirements limit applicability of the diagonalization approach in [1] to PDE-constrained optimization problems with distributed control, full observation of the state, and PDEs that lead to symmetric A. In contrast, our approach is applicable to general linear-quadratic optimal control problems with time-invariant system matrices, and arguably simpler.

Like [1] we consider a time discretization of the optimality system using the θ-scheme, which includes the backward Euler method and the trapezoidal method as special cases. Because the optimal control problem is time-invariant and constant time steps are used, the optimality system can be written as a sum of Kronecker products of matrices related to the time-stepping and of coefficient matrices in the optimal control problem. If the optimal control problem is time periodic, then the matrices related to the time-stepping can be simultaneously diagonalized, which, using Kronecker product properties, directly leads to a block diagonalization of the discretized optimality system. This block diagonalization can be efficiently computed and used to solve the optimality system in parallel. The main computational step is the parallel solution of n_t (=number of time steps) complex valued $2n_y \times 2n_y$ systems, where n_y is the dimension of the state space. These systems are related to optimality systems corresponding to a steady-state version of the optimal control problem. These systems are a bit more expensive to solve than the $2n_t$ complex valued $n_y \times n_y$ systems that arise in the diagonalization of [1], and this extra expense is the price for the generality of our proposed scheme. For problems where both diagonalizations can be applied, the diagonalization of [1] requires a parametric scaling which is not needed in our approach. If the optimal control problem is not time periodic a direct diagonalization is not possible, but a diagonalization of a related time periodic problem can be used to derive an efficient parallel-in-time preconditioner.

The main contribution of this paper is a new diagonalization technique that is applicable to discretized optimality systems that arise from (1.2), without additional limitations on B, Q, and R. In addition, we also provide a number of theoretical results on the well-posedness of discretized optimality systems and systems related to the preconditioner.

A diagonalization approach related to [1], but for optimal control problems governed by the wave equation is studied in [2]. Like [1, 2] assumes a specific least squares objective, a specific spatial differential operator, and distributed control. Extension of our approach to the case of the wave equation but without the assumptions on the least squares objective, spatial differential operator, and control seems possible, but is beyond the scope of this paper.

While parallel-in-time diagonalization methods for optimal control are new and limited, these diagonalization approaches applied to time-dependent PDEs have a longer history. See, *e.g.*, [3–5]. Parallel-in-time diagonalization methods do not view the time stepping method as a series of time steps, but rather as an aggregated space-time system. If the PDE has time invariant coefficients and if a constant step size is used, this discretized PDE system can be written as a Kronecker product in which one of the factors is a matrix related to the time stepping. If this matrix can be diagonalized, Kronecker product properties can be used to block diagonalize the space-time system, which gives rise to a parallel-in-time direct algorithm to solve such discretized PDEs [3]. Specifically, if the PDE has time invariant coefficients, is periodic-in-time, and if a constant step size is used, this discretized PDE system can be written as a Kronecker product in which one of the factors is a circulant matrix. This circulant matrix can be efficiently diagonalized using the Fast Fourier Transform (FFT) as the change of basis matrix. By using this view, the matrix system corresponding to a periodic-in-time PDE can be block diagonalized, which gives rise to a direct algorithm to solve such periodic-in-time problems. For initial value problems, the periodic solution technique can be used as a preconditioner. For initial value problems the matrix arising from the time discretization is Toeplitz, not circulant, and therefore block diagonalization is not possible. However, by replacing the Toeplitz matrix with a suitable circulant matrix, one can derive a preconditioner that can be efficiently applied parallel-in-time using block diagonalization. Using circulant matrices as preconditioners for Toeplitz matrices is discussed, *e.g.*, [6] and [7], but the theory cannot be applied when they arise in the Kronecker product arising in the context of initial value problems. Instead, it can be argued that preconditioners based on Kronecker product structure and circulant matrices lead to preconditioned matrices that are (relatively) low rank perturbations of the identity. See, *e.g.*, [4], Thm. 1 and Section 3.3. The resulting preconditioners perform well in practice. In the general nonsymmetric case, however, the spectrum of the matrix may reveal little about the performance of iterative methods, see, *e.g.*, [8]. Therefore, for classes of initial value problems, [4] introduced a "flip" operator to turn the problem into a symmetric one, which allowed eigenvalue analysis to give theoretical backing to the time-periodic method's use as a preconditioner. A software package containing some of these diagonalization-based methods can be found in [5]. Variations of diagonalization ideas used to solve PDEs or incorporations of diagonalization methods into other solvers are presented, *e.g.*, in [9] and [10]. The afore mentioned papers [1, 2] and this paper extend these parallel-in-time diagonalization methods from the PDE to the optimal control context. The paper [11] also uses the Kronecker product view of the discretized optimality system but develops low-rank in time techniques for the iterative solution of this optimality system, which are not parallel-in-time.

This paper is organized as follows. In the next Section 2, we study the optimal control problem (1.2a,b,d) with periodic-in-time conditions, and we introduce our diagonalization for the θ-scheme discretization of the optimality system of the optimal control problem (1.2a,b,d) with periodic-in-time conditions. Invertibility of the matrix arising from the discretization of the optimality conditions is established by analyzing the diagonal blocks arising in our diagonalization of the optimality system matrix. We also review the approach of [1] to show where their additional assumptions enter. In Section 3, we study the optimal control problem (1.2a-c) with initial conditions. The well-known optimality conditions are reviewed in Section 3.1. The Kronecker product form of the discretized optimality system is derived in Section 3.2. Section 3.3 derives our preconditioner, provides basic estimates of the spectrum of the preconditioned system, and compares our preconditioner with that proposed in [1]. The numerical performance of our preconditioner is illustrated in Section 4. Appendices A and B contain additional information about the computational application of the proposed preconditioner and additional observations on the spectrum of the preconditioned discretized optimality system, respectively.

2. Periodic problem

We first consider the optimal control problem (1.2a,b,d) with periodic in time conditions. We begin by reviewing existence of solutions and optimality conditions. Then we introduce the θ-scheme for the discretization of the optimality conditions, and we write the discretized optimality system in Kronecker product form. The main result of this section is the diagonalization of the discretized optimality system, presented in Sections 2.3 and 2.4. This optimality system diagonalization is based on the diagonalization of circulant matrices and the structure of the Kronecker product form.

2.1. Problem specification and optimality system

To ensure that the state equation (1.2b,d) has a unique solution for every control u, right hand side f, and time $T > 0$, we require that the eigenvalues of $M^{-1}A \in \mathbb{R}^{n_y \times n_y}$ have strictly positive real parts.

Lemma 2.1. *If (1.1) holds and the eigenvalues of $M^{-1}A \in \mathbb{R}^{n_y \times n_y}$ have strictly positive real parts, then for every $u \in L^2(0,T;\mathbb{R}^{n_u})$, $f \in L^2(0,T;\mathbb{R}^{n_y})$, and $T > 0$, the state equation (1.2b,d) has a unique solution.*

Proof. Given $u \in L^2(0,T;\mathbb{R}^{n_u})$ the solution of the state equation (1.2b) is

$$y(t) = \exp(-M^{-1}A\,t)y(0) + \int_0^t \exp(-M^{-1}A(t-\tau))\big(M^{-1}f(\tau) - M^{-1}Bu(\tau)\big)\mathrm{d}\tau. \tag{2.1}$$

The periodicity condition (1.2d) reads

$$y(0) = \exp(-M^{-1}A\,T)y(0) + \int_0^T \exp(-M^{-1}A(t-\tau))\big(M^{-1}f(\tau) - M^{-1}Bu(\tau)\big)\mathrm{d}\tau.$$

Since the eigenvalues of $M^{-1}A \in \mathbb{R}^{n_y \times n_y}$ have strictly positive real parts, $I - \exp(-M^{-1}A\,T)$ is invertible for any $T > 0$. Hence

$$y(0) = \Big(I - \exp(-M^{-1}A\,T)\Big)^{-1} \int_0^T \exp(-M^{-1}A(t-\tau))\big(M^{-1}f(\tau) - M^{-1}Bu(\tau)\big)\mathrm{d}\tau.$$

□

Standard arguments can now be applied to prove the following existence and uniqueness result.

Theorem 2.2. *If (1.1) holds and the eigenvalues of $M^{-1}A \in \mathbb{R}^{n_y \times n_y}$ have strictly positive real parts, then the optimal control problem (1.2a,b,d) has a unique solution. Moreover, $y \in H^1(0,T;\mathbb{R}^{n_y})$, $u \in L^2(0,T;\mathbb{R}^{n_u})$ solve (1.2a,b,d) if and only if there exists $p \in H^1(0,T;\mathbb{R}^{n_y})$ such that*

$$u = -R^{-1}\left(B^T p + d\right), \tag{2.2}$$

and

$$\begin{aligned} M\frac{\mathrm{d}}{\mathrm{d}t}y(t) + A\,y(t) - BR^{-1}B^T p(t) &= f(t) + BR^{-1}d(t), && t \in (0,T), && \text{(2.3a)}\\ M\,y(0) &= M\,y(T), && \\ -M^T\frac{\mathrm{d}}{\mathrm{d}t}p(t) + A^T p(t) + Q\,y(t) &= -c(t), && t \in (0,T), && \text{(2.3b)}\\ M^T p(0) &= M^T p(T). && \end{aligned}$$

Proof. The result can be proven by adapting the results in, *e.g.*, [12, 13], Section I.1.3, [14], Section 1.5, [15, 16]. □

2.2. Discretized optimality system

We discretize the system (2.3) of necessary and sufficient optimality conditions using the θ-time stepping scheme, $\theta \in [0,1]$, with n_t equispaced steps with time step size $\delta t = T/n_t$. For $\theta = 1$ the scheme is the backward Euler method and for $\theta = 1/2$ the scheme is the trapezoidal method. Let $y_k, p_k,\ k = 1, \dots n_t$, denote the approximations of y, p at time step $t_k = k\,\delta t$, where $M\,y_0 = M\,y_{n_t}$, $M^T p_0 = M^T p_{n_t}$ due to the periodicity condition, and let $c_k = c(t_k), d_k = d(t_k), f_k = f(t_k)$. The time-discretized version of (2.3) is

$$\begin{aligned} M(y_1 - y_{n_t}) + \delta t A(\theta y_1 + (1-\theta) y_{n_t}) - \delta t B R^{-1} B^T (\theta p_1 + (1-\theta) p_{n_t}) \qquad & \qquad (2.4a)\\ = \delta t \theta (f_1 + B R^{-1} d_1) + \delta t (1-\theta)(f_0 + B R^{-1} d_0), & \\ M(y_k - y_{k-1}) + \delta t A(\theta y_k + (1-\theta) y_{k-1}) - \delta t B R^{-1} B^T (\theta p_k + (1-\theta) p_{k-1}) \qquad & \qquad (2.4b)\\ = \delta t \theta (f_k + B R^{-1} d_k) + \delta t (1-\theta)(f_{k-1} + B R^{-1} d_{k-1}), \qquad & k = 2, \dots, n_t, \end{aligned}$$

$$\begin{aligned} M^T(p_k - p_{k+1}) + \delta t A^T(\theta p_k + (1-\theta) p_{k+1}) + \delta t Q(\theta y_k + (1-\theta) y_{k+1}) \qquad & \qquad (2.4c)\\ = -\delta t (\theta c_k + (1-\theta) c_{k+1}), \qquad & k = 1, \dots, n_t - 1, \\ M^T(p_{n_t} - p_1) + \delta t A^T(\theta p_{n_t} + (1-\theta) p_1) + \delta t Q(\theta y_{n_t} + (1-\theta) y_1) \qquad & \qquad (2.4d)\\ = -\delta t (\theta c_{n_t} + (1-\theta) c_{n_t+1}). & \end{aligned}$$

The controls u_k can be recovered by using (2.2) at a discrete level:

$$u_k = -R^{-1}\left(B^T p_k + d_k\right), \qquad k = 1, \dots, n_t. \tag{2.5}$$

Next, we put the modified optimality conditions (2.4) into matrix form. The time discretization for periodic-in-time problems leads to particular circulant matrices. Define the $n_t \times n_t$ matrices

$$C_{-1} = \begin{pmatrix} 1 & & & -1 \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix}, \quad C_\theta = \begin{pmatrix} \theta & & & 1-\theta \\ 1-\theta & \theta & & \\ & \ddots & \ddots & \\ & & 1-\theta & \theta \end{pmatrix}, \quad \theta \in [0,1]. \tag{2.6}$$

Using the matrices (2.6), the optimality system (2.4) can be expressed in Kronecker product form

$$(C_{-1} \otimes M + C_\theta \otimes \delta t A)\, Y - (C_\theta \otimes \delta t B R^{-1} B^T) P = F_Y, \tag{2.7a}$$

$$(C_\theta^T \otimes \delta t Q) Y + \left(C_{-1}^T \otimes M^T + C_\theta^T \otimes \delta t A^T\right) P = F_P, \tag{2.7b}$$

where

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_{n_t} \end{pmatrix}, \qquad P = \begin{pmatrix} p_1 \\ \vdots \\ p_{n_t} \end{pmatrix}, \tag{2.7c}$$

$$F_Y = \begin{pmatrix} \delta t \theta (f_1 + B R^{-1} d_1) + \delta t (1-\theta)(f_0 + B R^{-1} d_0) \\ \vdots \\ \delta t \theta (f_{n_t} + B R^{-1} d_{n_t}) + \delta t (1-\theta)(f_{n_t-1} + B R^{-1} d_{n_t-1}) \end{pmatrix}, \quad F_P = \begin{pmatrix} -\delta t(\theta c_1 + (1-\theta) c_2) \\ \vdots \\ -\delta t(\theta c_{n_t} + (1-\theta) c_{n_t+1}) \end{pmatrix}. \tag{2.7d}$$

In addition, define the $2n_t \times 2n_t$ block matrices

$$\mathbf{C}_{1,1}^{(-1)} = \begin{pmatrix} C_{-1} & 0 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{C}_{2,2}^{(-1)} = \begin{pmatrix} 0 & 0 \\ 0 & C_{-1}^T \end{pmatrix} \in \mathbb{R}^{2n_t \times 2n_t} \tag{2.8a}$$

and, for $\theta \in [0,1]$,

$$\mathbf{C}_{1,1}^{(\theta)} = \begin{pmatrix} C_\theta & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{C}_{1,2}^{(\theta)} = \begin{pmatrix} 0 & C_\theta \\ 0 & 0 \end{pmatrix}, \quad \mathbf{C}_{2,1}^{(\theta)} = \begin{pmatrix} 0 & 0 \\ C_\theta^T & 0 \end{pmatrix}, \quad \mathbf{C}_{2,2}^{(\theta)} = \begin{pmatrix} 0 & 0 \\ 0 & C_\theta^T \end{pmatrix} \in \mathbb{R}^{2K \times 2K}. \tag{2.8b}$$

With these matrices the discretized optimality conditions (2.7) can be re-written in Kronecker product form as

$$\mathbf{K}_{\text{per}} \begin{pmatrix} Y \\ P \end{pmatrix} = \begin{pmatrix} F_Y \\ F_P \end{pmatrix}, \tag{2.9a}$$

where

$$\begin{aligned} \mathbf{K}_{\text{per}} = & \mathbf{C}_{1,1}^{(-1)} \otimes M + \delta t \mathbf{C}_{1,1}^{(\theta)} \otimes A - \delta t \mathbf{C}_{1,2}^{(\theta)} \otimes B R^{-1} B^T \\ & + \mathbf{C}_{2,2}^{(-1)} \otimes M^T + \delta t \mathbf{C}_{2,2}^{(\theta)} \otimes A^T + \delta t \mathbf{C}_{2,1}^{(\theta)} \otimes Q. \end{aligned} \tag{2.9b}$$

The system (2.4) and its equivalent form (2.9) are a discretization of the continuous-in-time optimality system (2.3). Under the assumptions of Theorem 2.2, the continuous-in-time optimality system (2.3) has a unique solution. However, we still have to prove that (2.4) and (2.9) are also well-posed.

Theorem 2.3. *Assume (1.1) holds and that the eigenvalues of $M^{-1}A \in \mathbb{R}^{n_y \times n_y}$ have strictly positive real parts. If $\theta \in [1/2, 1]$ or if $\theta \in [0, 1/2)$ and $\delta t < 2Re(\lambda)/((1-2\theta)|\lambda|^2)$ for all eigenvalues λ of $M^{-1}A$, then $\mathbf{K}_{per}$ in (2.9b) is invertible.*

Proof. We will prove this result using the diagonalization of $\mathbf{K}_{\text{per}}$, which will be introduced in the next Section 2.3. Diagonalization of $\mathbf{K}_{\text{per}}$ and invertibility of the diagonal blocks, which will be established in Theorem 2.8, immediately imply this result. ☐

2.3. Diagonalization of the optimality system

Next, we discuss the block diagonalization of the $\mathbf{K}_{\text{per}}$ in (2.9b). This diagonalization is possible because the matrices (2.8) are simultaneously diagonalizable. The key result is the diagonalization of the circulant matrix. This result can be found, *e.g.*, in [17], Thm. 4.8.2, and is stated in the following theorem.

Theorem 2.4. *Let $\gamma \in \mathbb{C}$, $\omega = \mathrm{e}^{-2\pi i/n_t}$, $\lambda_k = 1 + \gamma\omega^{k-1}$, $k = 1, \ldots, n_t$, and*

$$V = \frac{1}{\sqrt{n_t}} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega & \cdots & \omega^{n_t-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n_t-1} & \cdots & \omega^{(n_t-1)^2} \end{pmatrix}, \quad D_{per}(\gamma) = diag(\lambda_1, \ldots, \lambda_{n_t}) \in \mathbb{C}^{n_t \times n_t}. \tag{2.10}$$

The circulant matrix

$$C_{per}(\gamma) = \begin{pmatrix} 1 & & & \gamma \\ \gamma & 1 & & \\ & \ddots & \ddots & \\ & & \gamma & 1 \end{pmatrix} \in \mathbb{C}^{n_t \times n_t}$$

and its conjugate transpose can be diagonalized as follows

$$C_{per}(\gamma) = V^{-1} D_{per}(\gamma) V, \quad C_{per}(\gamma)^* = V^{-1} D_{per}(\gamma)^* V,$$

where $D_{per}(\gamma)^*$ *denotes the complex conjugate of* $D_{per}(\gamma)$.

The previous result and the block structure of the matrices (2.8) imply the following result regarding the simultaneous diagonalization of the matrices in (2.8).

Corollary 2.5. *The matrices* (2.8) *can be simultaneously diagonalized as follows*

$$\mathbf{C}_{i,i}^{(-1)} = \mathbf{V}^{-1} \mathbf{D}_{i,i}^{(-1)} \mathbf{V}, \quad \mathbf{C}_{i,j}^{(\theta)} = \mathbf{V}^{-1} \mathbf{D}_{i,j}^{(\theta)} \mathbf{V}, \quad i, j = 1, 2,$$

where

$$\mathbf{V} = \begin{pmatrix} V & \\ & V \end{pmatrix}, \qquad \mathbf{D}_{1,1}^{(-1)} = \begin{pmatrix} D_{-1} & 0 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{D}_{2,2}^{(-1)} = \begin{pmatrix} 0 & 0 \\ 0 & D_{-1}^* \end{pmatrix}, \tag{2.11a}$$

$$\mathbf{D}_{1,1}^{(\theta)} = \begin{pmatrix} D_\theta & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{D}_{1,2}^{(\theta)} = \begin{pmatrix} 0 & D_\theta \\ 0 & 0 \end{pmatrix}, \quad \mathbf{D}_{2,1}^{(\theta)} = \begin{pmatrix} 0 & 0 \\ D_\theta^* & 0 \end{pmatrix}, \quad \mathbf{D}_{2,2}^{(\theta)} = \begin{pmatrix} 0 & 0 \\ 0 & D_\theta^* \end{pmatrix}, \quad \theta \in [0, 1], \tag{2.11b}$$

and where $D_{-1} = diag(0, 1 - \omega, \ldots, 1 - \omega^{n_t - 1})$, $D_\theta = diag(1, \theta + (1 - \theta)\omega, \ldots, \theta + (1 - \theta)\omega^{n_t - 1})$ *for* $\theta \in [0, 1]$, *and* $\omega = \mathrm{e}^{-2\pi i / n_t}$.

Proof. This result follows immediately from Theorem 2.4, the block diagonal structure of $\mathbf{V}$ and the block structures of the matrices (2.8). ☐

Next we use Corollary 2.5 to compute the block diagonalization of $\mathbf{K}_{\text{per}}$ in (2.9b).

Theorem 2.6. *The matrix* $\mathbf{K}_{per}$ *in* (2.9b) *can be re-written as*

$$\begin{aligned} \mathbf{K}_{per} = \left(\mathbf{V}^{-1} \otimes I_y\right) \Big(& \mathbf{D}_{1,1}^{(-1)} \otimes M + \delta t \mathbf{D}_{1,1}^{(\theta)} \otimes A - \delta t \mathbf{D}_{1,2}^{(\theta)} \otimes B R^{-1} B^T \\ & + \mathbf{D}_{2,2}^{(-1)} \otimes M^T + \delta t \mathbf{D}_{2,2}^{(\theta)} \otimes A^T + \delta t \mathbf{D}_{2,1}^{(\theta)} \otimes Q \Big) \left(\mathbf{V} \otimes I_y\right), \end{aligned} \tag{2.12}$$

$\theta \in [0, 1]$, *where* $\mathbf{V}$, $\mathbf{D}_{i,i}^{(-1)}$, $\mathbf{D}_{i,j}^{(\theta)}$, $i, j = 1, 2$, *are defined in* (2.11) *and* $I_y \in \mathbb{R}^{n_y \times n_y}$ *is the identity matrix.*

Proof. The identity (2.12) can be verified through direct calculation. The Kronecker product identity

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD), \tag{2.13}$$

implies that for arbitrary (appropriately sized) matrices $\mathbf{M}_1 \in \mathbb{R}^{2n_t \times 2n_t}$ and $M_2, I_y \in \mathbb{R}^{n_y \times n_y}$,

$$(\mathbf{V}^{-1} \otimes I_y)(\mathbf{M}_1 \otimes M_2)(\mathbf{V} \otimes I_y) = (\mathbf{V}^{-1} \mathbf{M}_1 \mathbf{V}) \otimes M_2. \tag{2.14}$$

Applying (2.14) to every term on the right hand side of (2.12) and using Corollary 2.5 recovers $\mathbf{K}_{\text{per}}$ in (2.9b). ☐

2.4. Diagonalization algorithm for solving periodic OCP

Theorem 2.6 provides an efficient way to solve the optimality system (2.9b). Let $\omega = \mathrm{e}^{-2\pi i/n_t}$. Recall from Corollary 2.5 that k-th diagonal entries of D_{-1}, D_{-1}^*, D_θ, $D_\theta^* \in \mathbb{C}^{n_t \times n_t}$, $\theta \in [0,1]$, are given by ($\overline{z}$ denotes the complex conjugate of the complex scalar z)

$$d_{-1,k} = 1 - \omega^{k-1}, \quad \overline{d}_{-1,k}, \quad d_{\theta,k} = \theta + (1-\theta)\omega^{k-1}, \quad \overline{d}_{\theta,k}, \quad k = 1, \ldots, n_t, \tag{2.15}$$

respectively. Because of the structure of the matrices $\mathbf{D}_{i,i}^{(-1)}$, $\mathbf{D}_{i,j}^{(\theta)}$, $i, j = 1, 2$, defined in (2.11), the matrix in the middle of the right hand side in (2.12) is block diagonal (after re-ordering).

The identity in (2.12) implies that the system (2.9b) can be solved in three steps:

1. Solve
$$(\mathbf{V}^{-1} \otimes I_y) \begin{pmatrix} X_y \\ X_p \end{pmatrix} = \begin{pmatrix} F_Y \\ F_P \end{pmatrix}.$$
2. Let
$$\begin{pmatrix} X_{y,k} \\ X_{p,k} \end{pmatrix} \in \mathbb{C}^{2n_y \times 2n_y}, \quad k = 1, \ldots, n_t,$$
denote the blocks corresponding to time step k. Solve
$$\begin{pmatrix} d_{-1,k} M + \delta t d_{\theta,k} A & -\delta t d_{\theta,k} B R^{-1} B^T \\ \delta t \overline{d}_{\theta,k} Q & \overline{d}_{-1,k} M^T + \delta t \overline{d}_{\theta,k} A^T \end{pmatrix} \begin{pmatrix} Z_{y,k} \\ Z_{p,k} \end{pmatrix} = \begin{pmatrix} X_{y,k} \\ X_{p,k} \end{pmatrix}, \quad k = 1, \ldots, n_t. \tag{2.16}$$
3. Solve
$$(\mathbf{V} \otimes I_y) \begin{pmatrix} Y \\ P \end{pmatrix} = \begin{pmatrix} Z_y \\ Z_p \end{pmatrix}.$$

All three of these steps are parallelizable. In step 1, due to the Kronecker product and the structure of the block matrix $\mathbf{V}$ in (2.11), the matrix $\mathbf{V}$ has to be applied to n_y blocks of length n_t of F_y and to n_y blocks of length n_t of F_p. These applications can be performed in parallel across the $2n_y$ blocks. Because of the structure of the matrix V the Fast Fourier Transform (FFT) can be used to apply V. Similarly, in step 3 the matrix V^{-1} has to be applied to n_y blocks of length n_t of Z_y and to n_y blocks of length n_t of Z_p. These applications can be performed in parallel across the number $2n_y$ of blocks and the Inverse Fast Fourier Transform (IFFT) can be used to apply V^{-1}. Step 2 can be performed in parallel across the n_t time steps.

For problems of moderate size, the systems (2.16) can be solved using sparse direct solvers. For larger problems they have to be solved iteratively. In that case, since the matrix V in (2.10) is unitary, the 2-norm residuals of the systems (2.16) and the condition numbers of the matrices in (2.16) determine the solution error. For the iterative solution of (2.16), the following reformulations are useful.

If $d_{\theta,k} \neq 0$, the systems (2.16) in Step 2 can be rearranged as Hermitian systems

$$\begin{pmatrix} \delta t Q & d_{\theta,k}^{-1}\big(\overline{d}_{-1,k} M^T + \delta t \overline{d}_{\theta,k} A^T\big) \\ \overline{d}_{\theta,k}^{-1}\big(d_{-1,k} M + \delta t d_{\theta,k} A\big) & -\delta t B R^{-1} B^T \end{pmatrix} \begin{pmatrix} \overline{d}_{\theta,k} Z_{y,k} \\ d_{\theta,k} Z_{p,k} \end{pmatrix} = \begin{pmatrix} X_{p,k} \\ X_{y,k} \end{pmatrix}, \quad k = 1, \ldots, n_t. \tag{2.17}$$

Defining $Z_{u,k} = -R^{-1}B^T(d_{\theta,k}Z_{p,k})$ allows one to write the block 2×2 Hermitian systems (2.17) as block 3×3 Hermitian systems

$$\begin{pmatrix} \delta t Q & 0 & d_{\theta,k}^{-1}\big(\overline{d}_{-1,k}M^T + \delta t\overline{d}_{\theta,k}A^T\big) \\ 0 & \delta t R & \delta t B^T \\ \overline{d}_{\theta,k}^{-1}\big(d_{-1,k}M + \delta t d_{\theta,k}A\big) & \delta t B & 0 \end{pmatrix} \begin{pmatrix} \overline{d}_{\theta,k}Z_{y,k} \\ Z_{u,k} \\ d_{\theta,k}Z_{p,k} \end{pmatrix} = \begin{pmatrix} X_{p,k} \\ 0 \\ X_{y,k} \end{pmatrix}, \quad k = 1,\dots,n_t, \tag{2.18}$$

which are complex, but share some of the properties of real saddle point (KKT) systems. Thus for larger systems, iterative methods and preconditioners for real saddle point (KKT) systems [18, 19] could be extended to this complex case.

Remark 2.7. Note that if $\theta = 1/2$ and n_t is even, the eigenvalue $d_{\theta,k}$ with index $k = 1 + n_t/2$ of C_θ shown in (2.15) is zero. Therefore, we need to require that the number of time steps n_t is odd if $\theta = 1/2$ to arrange the systems (2.16) in Step 2 as (2.17) or as (2.18). However, if (1.1) holds and the eigenvalues of $M^{-1}A \in \mathbb{R}^{n_y\times n_y}$ have strictly positive real parts, the systems (2.16) in Step 2 are invertible, *i.e.*, our diagonalization is applicable even if $\theta = 1/2$ and n_t is even.

Theorem 2.8. *Assume (1.1) holds and that the eigenvalues of $M^{-1}A \in \mathbb{R}^{n_y\times n_y}$ have strictly positive real parts. If $\theta \in [1/2, 1]$ or if $\theta \in [0, 1/2)$ and $\delta t < 2Re(\lambda)/((1-2\theta)|\lambda|^2)$ for all eigenvalues λ of $M^{-1}A$, then the systems (2.16) in Step 2 are invertible.*

Proof. From the definition (2.15) it follows that $d_{-1,k} = 0$ if and only if $k = 1$, and for $k = 1$, $d_{\theta,k} = 1$. Thus $d_{-1,k}$ and $d_{\theta,k}$ cannot both be zero.

If $d_{\theta,k} = 0$, the system matrix in (2.16) is

$$\begin{pmatrix} d_{-1,k}M & 0 \\ 0 & \overline{d}_{-1,k}M^T \end{pmatrix}$$

with $d_{-1,k} \neq 0$. Since M is invertible, the previous block 2×2 matrix is invertible.

If $d_{\theta,k} \neq 0$, the system matrix in (2.16) is invertible if and only if the block 3×3 Hermitian matrix in (2.18) is invertible. The block 3×3 Hermitian matrix in (2.18) is invertible if and only if $d_{-1,k}M + \delta t d_{\theta,k}A$ is invertible. (In this case, we can equivalently write (2.18) as a system in $Z_{u,k}$, and the assumptions on Q and R imply that the matrix in this system is Hermitian positive definite).

If $d_{\theta,k} \neq 0$ and $d_{-1,k} = 0$ (*i.e.*, $k = 1$), then $d_{-1,k}M + \delta t d_{\theta,k}A = \delta t d_{\theta,k}A$ is invertible, because $M^{-1}A$ and M are invertible.

Now, let $d_{\theta,k} \neq 0$ and $d_{-1,k} \neq 0$. The eigenvalues of $d_{-1,k}I + \delta t d_{\theta,k}M^{-1}A$ are $d_{-1,k} + \delta t d_{\theta,k}\lambda$, where λ are the eigenvalues of $M^{-1}A$. Now, let $\lambda = \alpha + i\beta$, $\alpha > 0$, be an eigenvalue of $M^{-1}A$. Using the definition (2.15) of $d_{-1,k}$ and $d_{\theta,k}$, the corresponding eigenvalue of $d_{-1,k}I + \delta t d_{\theta,k}M^{-1}A$ is

$$d_{-1,k} + \delta t d_{\theta,k}\lambda = 1 + \delta t\theta\lambda + \big(-1 + \delta t(1-\theta)\,\lambda\big)\omega^{k-1}.$$

If $d_{-1,k} + \delta t d_{\theta,k}\lambda = 0$, then $1 + \delta t\theta\lambda = -\big(-1 + \delta t(1-\theta)\,\lambda\big)\omega^{k-1}$ and

$$|1 + \delta t\theta\lambda|^2 = |-1 + \delta t(1-\theta)\,\lambda|^2|\omega^{k-1}|^2 = |-1 + \delta t(1-\theta)\,\lambda|^2.$$

However, using $\lambda = \alpha + i\beta$, $\alpha > 0$, gives

$$|1 + \delta t\theta\lambda|^2 = 1 + 2\delta t\theta\alpha + \delta t^2\theta^2(\alpha^2 + \beta^2)$$

and

$$\begin{aligned}|-1+\delta t(1-\theta)\,\lambda|^2 =&1+2\delta t\theta\alpha+\delta t^2\theta^2(\alpha^2+\beta^2)-2\delta t\alpha+\delta t^2(1-2\theta)(\alpha^2+\beta^2)\\ <&1+2\delta t\theta\alpha+\delta t^2\theta^2(\alpha^2+\beta^2)\end{aligned}$$

for $\theta\in[1/2,1]$ or for $\theta\in[0,1/2)$ and $\delta t<2\alpha/((1-2\theta)(\alpha^2+\beta^2))$. Hence $d_{-1,k}+\delta t d_{\theta,k}\lambda\neq 0$. □

2.5. Comparison with previous diagonalization approach

In this section, we compare our approach to that of [1] in the case where $A=A^T$, M is symmetric positive definite, and $Q=M$, $B=\pm M$, $R=\alpha M$, $\alpha>0$, which is the setting considered in [1]. As before, $\theta\in[0,1]$. Initially, we allow $A\neq A^T$ to highlight where the symmetry of A is needed. If $Q=M$, $B=\pm M$, $R=\alpha M$, the Kronecker product form (2.7) of the discretized optimality system (2.4) becomes

$$(C_{-1}\otimes M+C_\theta\otimes\delta tA)\,Y-(C_\theta\otimes\delta t\alpha^{-1}M)P=F_Y, \tag{2.19a}$$
$$(C_\theta^T\otimes\delta tM)Y+\left(C_{-1}^T\otimes M+C_\theta^T\otimes\delta tA^T\right)P=F_P, \tag{2.19b}$$

where Y and P are defined as in (2.7c), and F_Y and F_P are defined as in (2.7d) with $B=\pm M$ and $R=\alpha M$.

Remark 2.9. As already noted in Remark 2.7, if $\theta=1/2$ and n_t is even, the eigenvalue $d_{\theta,k}$ with index $k=1+n_t/2$ of C_θ shown in (2.15) is zero. Because [1] use the inverse of C_θ they assume that the number of time steps n_t is odd if $\theta=1/2$.

The authors [1] then modify (2.19a) by multiplying the first equation by $(C_\theta^{-1}\otimes I_y)$ and multiplying the second equation by $(C_\theta^{-T}\otimes I_y)$. Using the Kronecker product identity (2.13), the system of equations (2.19) can be equivalently expressed as

$$\left(\begin{pmatrix}C_\theta^{-1}C_{-1} & -\delta t\alpha^{-1}I_{n_t}\\ \delta tI_{n_t} & C_\theta^{-T}C_{-1}^T\end{pmatrix}\otimes M+\delta t\mathbf{I}_{1,1}\otimes A+\delta t\mathbf{I}_{2,2}\otimes A^T\right)\begin{pmatrix}Y\\P\end{pmatrix}=\begin{pmatrix}\widetilde{F_Y}\\ \widetilde{F_P}\end{pmatrix}, \tag{2.20}$$

with

$$\widetilde{F_Y}=(C_\theta^{-1}\otimes I_y)F_Y,\quad \widetilde{F_P}=(C_\theta^{-T}\otimes I_y)F_P,\quad \mathbf{I}_{1,1}=\begin{pmatrix}I_{n_t}&0\\0&0\end{pmatrix}\in\mathbb{R}^{2n_t\times 2n_t},\quad \mathbf{I}_{2,2}=\begin{pmatrix}0&0\\0&I_{n_t}\end{pmatrix}\in\mathbb{R}^{2n_t\times 2n_t},$$

where $I_{n_t}\in\mathbb{R}^{n_t\times n_t}$ is an identity matrix.

Finally, [1] introduce a scaling of (2.20) which they use to improve the numerical properties of diagonalization introduced in the next steps. For given $\beta>0$, the first block of equations in (2.20) is multiplied by β^{-1} and Y is replaced by $\widehat{Y}=\beta^{-1}Y$ to obtain the following system equivalent to (2.20):

$$\Bigg(\underbrace{\begin{pmatrix}C_\theta^{-1}C_{-1} & -\delta t(\alpha\beta)^{-1}I_{n_t}\\ \delta t\beta\, I_{n_t} & C_\theta^{-T}C_{-1}^T\end{pmatrix}}_{=:\,\mathbf{G}(\beta)}\otimes M+\delta t\mathbf{I}_{1,1}\otimes A+\delta t\mathbf{I}_{2,2}\otimes A^T\Bigg)\begin{pmatrix}\widehat{Y}\\P\end{pmatrix}=\begin{pmatrix}\beta^{-1}\widetilde{F_Y}\\ \widetilde{F_P}\end{pmatrix}. \tag{2.21}$$

Using Theorem 2.4, $\mathbf{G}(\beta)$ can be decomposed into

$$\mathbf{G}(\beta) = \underbrace{\begin{pmatrix} V^{-1} & \\ & V^{-1} \end{pmatrix}}_{\mathbf{V}^{-1}} \underbrace{\begin{pmatrix} \Lambda(\theta) & -\delta t(\alpha\beta)^{-1} I_{n_t} \\ \delta t\beta\, I_{n_t} & \Lambda(\theta)^* \end{pmatrix}}_{=:\, \widetilde{\mathbf{G}}(\beta)} \underbrace{\begin{pmatrix} V & \\ & V \end{pmatrix}}_{\mathbf{V}}, \tag{2.22}$$

where

$$\Lambda(\theta) = D_\theta^{-1} D_{-1}, \tag{2.23}$$

V is given in (2.10), and D_{-1}, and D_θ are given in Corollary 2.5. Because of the structure of $\mathbf{I}_{1,1}$, $\mathbf{I}_{2,2}$ and $\mathbf{V}$, we can proceed as in Corollary 2.5 to diagonalize the matrix on the left hand side in (2.21):

$$\mathbf{G}(\beta) \otimes M + \delta t \mathbf{I}_{1,1} \otimes A + \delta t \mathbf{I}_{2,2} \otimes A^T = \big(\mathbf{V}^{-1} \otimes I_y\big)\Big(\widetilde{\mathbf{G}}(\beta) \otimes M + \delta t \mathbf{I}_{1,1} \otimes A + \delta t \mathbf{I}_{2,2} \otimes A^T\Big)\big(\mathbf{V} \otimes I_y\big). \tag{2.24}$$

Because all blocks of the 2×2 block matrix $\widetilde{\mathbf{G}}(\beta)$ in (2.22) are diagonal, $\widetilde{\mathbf{G}}(\beta)$ can be diagonalized. The matrix $\widetilde{\mathbf{G}}(\beta)$ can be symmetrically re-ordered into a block diagonal matrix with blocks in $\mathbb{C}^{2\times 2}$ and the eigendecompositions of the matrices $\mathbb{C}^{2\times 2}$ can be computed analytically. In [1] it is shown that

$$\widetilde{\mathbf{G}}(\beta) = \mathbf{W}(\beta)\,\mathbf{D}\,\mathbf{W}(\beta)^{-1}, \tag{2.25}$$

where the $\mathbf{W}(\beta)$ and $\mathbf{D}$ are specified below. The scaling by β in (2.21) is introduced to improve the conditioning of $\mathbf{W}(\beta)$, and [1], p. 10 suggests the choice $\beta = 1/\sqrt{\alpha}$. The matrix $\mathbf{W}(\beta)$ is not block diagonal for any $\beta \neq 0$ and, therefore, $\mathbf{W}(\beta)\,\mathbf{I}_{i,i}\,\mathbf{W}(\beta)^{-1} \neq \mathbf{I}_{i,i}$, $i = 1, 2$. However, if $A = A^T$, then $\mathbf{I}_{1,1} \otimes A + \mathbf{I}_{2,2} \otimes A^T = \mathbf{I} \otimes A$, where $\mathbf{I}$ is the $2n_t \times 2n_t$ identity matrix, and we can diagonalize

$$\begin{aligned} \mathbf{G}(\beta) \otimes M + \delta t \mathbf{I} \otimes A &= \big(\mathbf{V}^{-1} \otimes I_y\big)\Big(\widetilde{\mathbf{G}}(\beta) \otimes M + \delta t \mathbf{I} \otimes A\Big)\big(\mathbf{V} \otimes I_y\big) \\ &= \big(\mathbf{V}^{-1}\mathbf{W}(\beta) \otimes I_y\big)\Big(\mathbf{D} \otimes M + \delta t \mathbf{I} \otimes A\Big)\big(\mathbf{V}\mathbf{W}(\beta)^{-1} \otimes I_y\big). \end{aligned} \tag{2.26}$$

For completeness , we state the matrices in (2.25),

$$\mathbf{W}(\beta) = \begin{pmatrix} I_{n_t} & S_2(\beta) \\ S_1(\beta) & I_{n_t} \end{pmatrix} \widetilde{\mathbf{D}}^{-1}(\beta), \qquad \widetilde{\mathbf{D}}(\beta) = \begin{pmatrix} \sqrt{\Lambda(\theta)} + S_1(\beta) & \\ & \sqrt{\Lambda(\theta)^*} + S_2(\beta) \end{pmatrix},$$

$$\mathbf{D} = \begin{pmatrix} \mathrm{Re}(\Lambda(\theta)) + i\sqrt{\mathrm{Im}^2(\Lambda(\theta)) + \alpha^{-1}\delta t^2 I_{n_t}} & \\ & \mathrm{Re}(\Lambda(\theta)) - i\sqrt{\mathrm{Im}^2(\Lambda(\theta)) + \alpha^{-1}\delta t^2 I_{n_t}} \end{pmatrix},$$

$S_1(\beta) = i\frac{\alpha\beta}{\delta t}\Big(\mathrm{Im}(\Lambda(\theta)) - \sqrt{\mathrm{Im}^2(\Lambda(\theta)) + \alpha^{-1}\delta t^2 I_{n_t}}\Big)$, $S_2(\beta) = i\frac{1}{\beta\delta t}\Big(\mathrm{Im}(\Lambda(\theta)) - \sqrt{\mathrm{Im}^2(\Lambda(\theta)) + \alpha^{-1}\delta t^2 I_{n_t}}\Big)$. As shown in [1], Rem. 2.6, the introduction of the scaling by β and of the diagonal matrix $\widetilde{\mathbf{D}}(\beta)$ is crucial to ensure that $\mathbf{V}^{-1}\mathbf{W}(\beta)$ is well conditioned. As mentioned before, [1], p. 10 suggests the choice $\beta = 1/\sqrt{\alpha}$, but an optimal choice is not yet known. Our proposed diagonalization (2.12) does not depend on a parameter.

Remark 2.10. In [1], the authors use a version of (2.26) that is scaled by the mass matrix. That is, they set $L = M^{-1}A$ and consider the factorization

$$\begin{aligned} \mathbf{G}(\beta) \otimes I_y + \delta t \mathbf{I} \otimes L &= \big(\mathbf{V}^{-1} \otimes I_y\big)\Big(\widetilde{\mathbf{G}}(\beta) \otimes I_y + \delta t \mathbf{I} \otimes L\Big)\big(\mathbf{V} \otimes I_y\big) \\ &= \big(\mathbf{V}^{-1}\mathbf{W}(\beta) \otimes I_y\big)\Big(\mathbf{D} \otimes I_y + \delta t \mathbf{I} \otimes L\Big)\big(\mathbf{V}\mathbf{W}(\beta)^{-1} \otimes I_y\big). \end{aligned}$$

3. Initial value problem

With the periodic problem solved, we now turn our attention to the optimal control problem constrained by an initial-value problem (1.2a,b,c). First, we state the necessary and sufficient optimality conditions. Then, in Section 3.2, we apply the θ-scheme to discretize the optimality conditions and we write the discretized optimality conditions in Kronecker product form. In Section 3.3, we use a corresponding periodic in time problem to develop a preconditioner for the discretized optimality conditions corresponding to the initial-value problem (1.2a,b,c), and we provide basic estimates for the spectrum of the preconditioned system. The development of the preconditioner is analogous to that in [1], but it will be applied using the diagonalization developed in Sections 2.3, 2.4. Since this diagonalization is applicable to a broader class of problems our preconditioner for the discretized optimality conditions corresponding to the initial-value problem (1.2a,b,c) is applicable to a broader class of problems than the one in [1].

3.1. Problem specification and optimality system

The following is a standard result on the existence, uniqueness and characterization of the solution of (1.2a,b,c).

Theorem 3.1. *If* (1.1) *holds, then the optimal control problem (1.2a,b,c) has a unique solution. Moreover, $y \in H^1(0,T;\mathbb{R}^{n_u})$, $u \in L^2(0,T;\mathbb{R}^{n_u})$ solve (1.2a,b,d) if and only if there exists $p \in H^1(0,T;\mathbb{R}^{n_u})$ such that $u = -R^{-1}(B^Tp + d)$ and*

$$
\begin{aligned}
M\frac{\mathrm{d}}{\mathrm{d}t}y(t) + A\,y(t) - BR^{-1}B^Tp(t) &= f(t) + BR^{-1}d(t), && t \in (0,T), && \text{(3.1a)}\\
M\,y(0) &= M\,y_0, \\
-M^T\frac{\mathrm{d}}{\mathrm{d}t}p(t) + A^Tp(t) + Q\,y(t) &= -c(t), && t \in (0,T), && \text{(3.1b)}\\
M^Tp(T) &= 0.
\end{aligned}
$$

Proof. The result can be proven by adapting the results in, *e.g.*, [13], Section I.1.3, [14], Section 1.5, [15, 16]. ☐

3.2. Discretized optimality system

Let y_k, p_k, $k = 0,\dots n_t$, denote the approximations of y, p at time step $t_k = k\,\delta t$, where $M\,y_0$ is given from (1.2c) and $M^Tp_{n_t} = 0$, and let $c_k = c(t_k), d_k = d(t_k), f_k = f(t_k)$. The time-discretized version of (3.1) is

$$
\begin{aligned}
M(y_k - y_{k-1}) + \delta tA(\theta y_k + (1-\theta)y_{k-1}) - \delta tBR^{-1}B^T(\theta p_k + (1-\theta)p_{k-1}) \qquad\qquad \text{(3.2a)}\\
= \delta t\theta(f_k + BR^{-1}d_k) + \delta t(1-\theta)(f_{k-1} + BR^{-1}d_{k-1}), \qquad k = 1,\dots,n_t,
\end{aligned}
$$

$$
\begin{aligned}
M^T(p_{k-1} - p_k) + \delta tA^T(\theta p_{k-1} + (1-\theta)p_k) + \delta tQ(\theta y_{k-1} + (1-\theta)y_k) \qquad\qquad \text{(3.2b)}\\
= -\delta t(\theta c_{k-1} + (1-\theta)c_k), \qquad k = 1,\dots,n_t,
\end{aligned}
$$

where $M\,y_0$ is given from (1.2c) and $M^Tp_{n_t} = 0$. The system (3.2) is a system in the states $y_1,\dots,y_{n_t}$ and the adjoints $p_0,\dots,p_{n_t-1}$. Once these y_k's and p_k's are computed, the controls u_k can be recovered by using (2.5).

We now put the discretized optimality conditions (3.2) into matrix form. In contrast to the periodic case where circulant matrices arose, the time discretization now leads to particular Toeplitz matrices. Define the

$n_t \times n_t$ matrices

$$T_{-1} = \begin{pmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix}, \qquad T_\theta = \begin{pmatrix} \theta & & & \\ 1-\theta & \theta & & \\ & \ddots & \ddots & \\ & & 1-\theta & \theta \end{pmatrix}, \quad \theta \in [0,1], \tag{3.3a}$$

and

$$\widetilde{T}_\theta = \begin{pmatrix} 1-\theta & & & \\ \theta & 1-\theta & & \\ & \ddots & \ddots & \\ & & \theta & 1-\theta \end{pmatrix}, \quad \theta \in [0,1]. \tag{3.3b}$$

With these matrices, the discretized optimality conditions (3.2) can be re-written in Kronecker product form as

$$(T_{-1} \otimes M + T_\theta \otimes \delta t A)\, Y - (\widetilde{T}_\theta^T \otimes \delta t B R^{-1} B^T) P = F_Y, \tag{3.4a}$$

$$(\widetilde{T}_\theta \otimes \delta t Q) Y + \left(T_{-1}^T \otimes M^T + T_\theta^T \otimes \delta t A^T\right) P = F_P, \tag{3.4b}$$

where the unknowns are

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_{n_t} \end{pmatrix}, \qquad P = \begin{pmatrix} p_0 \\ \vdots \\ p_{n_t-1} \end{pmatrix}, \tag{3.4c}$$

and the right hand sides are

$$F_Y = \begin{pmatrix} \delta t\theta(f_1 + BR^{-1}d_1) + \delta t(1-\theta)(f_0 + BR^{-1}d_0) + My_0 - \delta t(1-\theta)Ay_0 \\ \vdots \\ \delta t\theta(f_{n_t} + BR^{-1}d_{n_t}) + \delta t(1-\theta)(f_{n_t-1} + BR^{-1}d_{n_t-1}) \end{pmatrix}, \tag{3.4d}$$

$$F_P = \begin{pmatrix} -\delta t(\theta c_0 + (1-\theta)c_1) - \delta t\theta Q y_0 \\ \vdots \\ -\delta t(\theta c_{n_t-1} + (1-\theta)c_{n_t}) \end{pmatrix}. \tag{3.4e}$$

Just as in the periodic case, we can re-write (3.4) using block 2×2 matrices. Define

$$\mathbf{T}_{1,1}^{(-1)} = \begin{pmatrix} T_{-1} & 0 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{T}_{2,2}^{(-1)} = \begin{pmatrix} 0 & 0 \\ 0 & T_{-1}^T \end{pmatrix} \in \mathbb{R}^{2n_t \times 2n_t}, \tag{3.5a}$$

$$\mathbf{T}_{1,1}^{(\theta)} = \begin{pmatrix} T_\theta & 0 \\ 0 & 0 \end{pmatrix}, \quad \widetilde{\mathbf{T}}_{1,2}^{(\theta)} = \begin{pmatrix} 0 & \widetilde{T}_\theta^T \\ 0 & 0 \end{pmatrix}, \quad \widetilde{\mathbf{T}}_{2,1}^{(\theta)} = \begin{pmatrix} 0 & 0 \\ \widetilde{T}_\theta & 0 \end{pmatrix}, \quad \mathbf{T}_{2,2}^{(\theta)} = \begin{pmatrix} 0 & 0 \\ 0 & T_\theta^T \end{pmatrix}, \quad \theta \in [0,1]. \tag{3.5b}$$

With these matrices, the discretized optimality conditions (3.4) can be written as

$$\mathbf{K}_{\text{init}} \begin{pmatrix} Y \\ P \end{pmatrix} = \begin{pmatrix} F_Y \\ F_P \end{pmatrix}, \tag{3.6a}$$

where the matrix is

$$\begin{aligned} \mathbf{K}_{\text{init}} =& \mathbf{T}_{1,1}^{(-1)} \otimes M + \delta t \mathbf{T}_{1,1}^{(\theta)} \otimes A - \delta t \widetilde{\mathbf{T}}_{1,2}^{(\theta)} \otimes BR^{-1}B^T \\ &+ \mathbf{T}_{2,2}^{(-1)} \otimes M^T + \delta t \mathbf{T}_{2,2}^{(\theta)} \otimes A^T + \delta t \widetilde{\mathbf{T}}_{2,1}^{(\theta)} \otimes Q. \end{aligned} \tag{3.6b}$$

The next result shows that the discretized optimality conditions (3.4) and the equivalent system (3.6a) have a unique solution if (1.1) holds and $M + \delta t\theta A$ is invertible.

Theorem 3.2. *If (1.1) holds and $M + \delta t\theta A$ is invertible, $\mathbf{K}_{init}$ in (3.6b) is invertible.*

Proof. Consider

$$\mathbf{K}_{\text{init}} \begin{pmatrix} Y \\ P \end{pmatrix} = \mathbf{0}, \tag{3.7}$$

where Y and P are given by (3.4c). We will prove that $Y = 0$ and $P = 0$, which implies the desired result.

Using the definition of $\mathbf{K}_{\text{init}}$, the equation (3.7) can be written as

$$M(y_k - y_{k-1}) + \delta t A(\theta y_k + (1-\theta) y_{k-1}) - \delta t BR^{-1}B^T(\theta p_k + (1-\theta)p_{k-1}) = 0, \qquad k = 1, \dots, n_t, \tag{3.8a}$$

$$M^T(p_{k-1} - p_k) + \delta t A^T(\theta p_{k-1} + (1-\theta)p_k) + \delta t Q(\theta y_{k-1} + (1-\theta) y_k) = 0, \qquad k = 1, \dots, n_t, \tag{3.8b}$$

where $y_0 = p_{n_t} = 0$.

Multiplying (3.8a) and (3.8b) by $((1-\theta)p_{k-1} + \theta p_k)^T$ and $((1-\theta)y_k + \theta y_{k-1})^T$, respectively, and subtracting the resulting equations gives

$$\begin{aligned} &((1-\theta)p_{k-1} + \theta p_k)^T M(y_k - y_{k-1}) - ((1-\theta)y_k + \theta y_{k-1})^T M^T(p_{k-1} - p_k) \\ &= -\delta t((1-\theta)p_{k-1} + \theta p_k)^T A(\theta y_k + (1-\theta)y_{k-1}) + \delta t((1-\theta)y_k + \theta y_{k-1})^T A^T(\theta p_{k-1} + (1-\theta)p_k) \\ &\quad + \delta t((1-\theta)p_{k-1} + \theta p_k)^T BR^{-1}B^T((1-\theta)p_{k-1} + \theta p_k) + \delta t((1-\theta)y_k + \theta y_{k-1})^T Q((1-\theta)y_k + \theta y_{k-1}). \end{aligned} \tag{3.9}$$

By straightforward calculations,

$$\begin{aligned} ((1-\theta)p_{k-1} + \theta p_k)^T M(y_k - y_{k-1}) - ((1-\theta)y_k + \theta y_{k-1})^T M^T(p_{k-1} - p_k) \\ = p_k^T M y_k - p_{k-1}^T M y_{k-1}, \end{aligned} \tag{3.10a}$$

$$\begin{aligned} -((1-\theta)p_{k-1} + \theta p_k)^T A(\theta y_k + (1-\theta)y_{k-1}) + ((1-\theta)y_k + \theta y_{k-1})^T A^T(\theta p_{k-1} + (1-\theta)p_k) \\ = (1-2\theta)\Big(p_k^T A y_k - p_{k-1}^T A y_{k-1}\Big). \end{aligned} \tag{3.10b}$$

Since $y_0 = p_{n_t} = 0$,

$$\sum_{k=1}^{n_t} p_k^T M y_k - p_{k-1}^T M y_{k-1} = p_{n_t}^T M y_{n_t} - p_0^T M y_0 = 0, \quad \sum_{k=1}^{n_t} p_k^T A y_k - p_{k-1}^T A y_{k-1} = p_{n_t}^T A y_{n_t} - p_0^T A y_0 = 0. \tag{3.11}$$

Inserting (3.10) into (3.9), summing up, and using (3.11) gives

$$
\begin{aligned}
0 =& \delta t \sum_{k=1}^{n_t} ((1-\theta)p_{k-1} + \theta p_k)^T BR^{-1}B^T((1-\theta)p_{k-1} + \theta p_k) \\
& \qquad + ((1-\theta)y_k + \theta y_{k-1})^T Q((1-\theta)y_k + \theta y_{k-1}) \geq 0,
\end{aligned} \tag{3.12}
$$

because R is symmetric positive definite and Q is symmetric positive semidefinite. Again because R is symmetric positive definite and Q is symmetric positive semidefinite, (3.12) implies

$$
B^T((1-\theta)p_{k-1} + \theta p_k) = 0 \quad \text{and} \quad Q((1-\theta)y_k + \theta y_{k-1}) = 0. \tag{3.13}
$$

With (3.13) the equations (3.8) are

$$
M(y_k - y_{k-1}) + \delta t A(\theta y_k + (1-\theta)y_{k-1}) = 0, \qquad k = 1, \ldots, n_t, \tag{3.14}
$$
$$
M^T(p_{k-1} - p_k) + \delta t A^T(\theta p_{k-1} + (1-\theta)p_k) = 0, \qquad k = 1, \ldots, n_t, \tag{3.15}
$$

where $y_0 = p_{n_t} = 0$. If $M + \delta t\theta A$ is invertible, the previous equations and $y_0 = p_{n_t} = 0$ imply $y_1 = y_2 = \ldots = y_{n_t} = 0$ and $p_{n_t-1} = p_{n_t-2} = \ldots = p_0 = 0$. $\square$

Remark 3.3. In [1], equation (3.2) a system which is different from (3.4) when $\theta \neq 1/2$ is considered. Instead of $\widetilde{\mathbf{T}}^{(\theta)}_{1,2}$, $\widetilde{\mathbf{T}}^{(\theta)}_{2,1}$ define

$$
\mathbf{T}^{(\theta)}_{1,2} = \begin{pmatrix} 0 & T_\theta^T \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{T}^{(\theta)}_{2,1} = \begin{pmatrix} 0 & 0 \\ T_\theta & 0 \end{pmatrix}. \tag{3.16}
$$

Note that $\widetilde{\mathbf{T}}^{(\theta)}_{1,2} = \mathbf{T}^{(\theta)}_{1,2}$ and $\widetilde{\mathbf{T}}^{(\theta)}_{2,1} = \mathbf{T}^{(\theta)}_{2,1}$ if and only if $\theta = 1/2$. Written in our notation the system [1], equation (3.2) is

$$
\mathbf{K}^{\mathrm{WZ}}_{\mathrm{init}} \begin{pmatrix} Y \\ P \end{pmatrix} = \begin{pmatrix} F_Y \\ F_P \end{pmatrix}, \tag{3.17a}
$$

where the matrix is

$$
\begin{aligned}
\mathbf{K}^{\mathrm{WZ}}_{\mathrm{init}} =& \mathbf{T}^{(-1)}_{1,1} \otimes M + \delta t \mathbf{T}^{(\theta)}_{1,1} \otimes A - \delta t \mathbf{T}^{(\theta)}_{1,2} \otimes BR^{-1}B^T \\
& + \mathbf{T}^{(-1)}_{2,2} \otimes M^T + \delta t \mathbf{T}^{(\theta)}_{2,2} \otimes A^T + \delta t \mathbf{T}^{(\theta)}_{2,1} \otimes Q,
\end{aligned} \tag{3.17b}
$$

the unknowns are (3.4c), the right hand side F_Y is given by (3.4d), and the right hand side F_Y is

$$
F_P = \begin{pmatrix} -\delta t(\theta c_0 + (1-\theta)c_1) - \delta t(1-\theta)Qy_0 \\ \vdots \\ -\delta t(\theta c_{n_t-1} + (1-\theta)c_{n_t}) \end{pmatrix}. \tag{3.17c}
$$

The system (3.17) is the matrix representation of

$$
\begin{aligned}
& M(y_k - y_{k-1}) + \delta t A(\theta y_k + (1-\theta)y_{k-1}) - \delta t BR^{-1}B^T((1-\theta)p_k + \theta p_{k-1}) \\
& \qquad = \delta t\theta(f_k + BR^{-1}d_k) + \delta t(1-\theta)(f_{k-1} + BR^{-1}d_{k-1}), \qquad k = 1, \ldots, n_t,
\end{aligned} \tag{3.18a}
$$

$$M^T(p_k - p_{k+1}) + \delta t A^T(\theta p_k + (1-\theta)p_{k+1}) + \delta t Q((1-\theta)y_k + \theta y_{k-1}) \\ = -\delta t(\theta c_k + (1-\theta)c_{k+1}), \qquad k = 0,\dots,n_t-1, \tag{3.18b}$$

where $M\,y_0$ is given from (1.2c) and $M^T p_{n_t} = 0$. Note the term $(\theta p_k + (1-\theta)p_{k-1})$ in (3.2a) is replaced by $((1-\theta)p_k + \theta p_{k-1})$ in (3.18a), and the term $(\theta y_k + (1-\theta)y_{k-1})$ in (3.2b) is replaced by $((1-\theta)y_k + \theta y_{k-1})$ in (3.18b). If $\theta = 1/2$, then (3.2) and (3.18) are identical, and the systems (3.4) and (3.17) are identical. Otherwise, they are different.

3.3. Diagonalization algorithm applied as preconditioner for initial-value problem

We want to construct a preconditioner for $\mathbf{K}_{\rm init}$ defined in (3.6b) that has the same Kronecker product structure as $\mathbf{K}_{\rm init}$. The general preconditioner design is motivated by the following observations.

Let $\mathbf{T}$ and M be two generic, nonsingular matrices. Given a square matrix M, we use I_M to denote the identity matrix of the same size as M. With a nonsingular matrix $\mathbf{C}$ of the same size as $\mathbf{T}$, we consider $\mathbf{C}\otimes M$ as a preconditioner for $\mathbf{T}\otimes M$. The difference between the identity matrix $I_{\mathbf{T}\otimes M} = I_{\mathbf{T}} \otimes I_M$ and the preconditioned Kronecker product is

$$I_{\mathbf{T}} \otimes I_M - \big(\mathbf{C}\otimes M\big)^{-1}\big(\mathbf{T}\otimes M\big) = I_{\mathbf{T}} \otimes I_M - \big(\mathbf{C}^{-1}\otimes M^{-1}\big)\big(\mathbf{T}\otimes M\big) = I_{\mathbf{T}} \otimes I_M - \big(\mathbf{C}^{-1}\mathbf{T}\otimes I_M\big) \\ = \big(I_{\mathbf{T}} - \mathbf{C}^{-1}\mathbf{T}\big)\otimes I_M = \mathbf{C}^{-1}\big(\mathbf{C}-\mathbf{T}\big)\otimes I_M.$$

The eigenvalues of $\mathbf{C}^{-1}\big(\mathbf{C}-\mathbf{T}\big)\otimes I_M$ are given by the eigenvalues of $\mathbf{C}^{-1}\big(\mathbf{C}-\mathbf{T}\big)$, each with multiplicity given by the size of M. In our application, $\mathbf{T}$ is one of the matrices (3.5) in which the non-zero block is defined by a Toeplitz matrix. For our preconditioner, we will choose a corresponding matrix of the type (2.8), in which the non-zero block is defined by a circulant matrix. With this choice of $\mathbf{C}$, the difference $\mathbf{C}-\mathbf{T}$ has only one non-zero entry (*i.e.*, is a rank one matrix). Consequently, with this choice of $\mathbf{C}$, $\mathbf{C}^{-1}\big(\mathbf{C}-\mathbf{T}\big)\otimes I_M$ has only one non-zero eigenvalue with multiplicity given by the size of M. Unfortunately, the situation is not quite as nice when we precondition $\mathbf{K}_{\rm init}$ defined in (3.6b), because $\mathbf{K}_{\rm init}$ is a sum of Kronecker products. However, our preconditioner design is motivated by these considerations. Our overall preconditioner design is also similar to that of [1]. However, [1] consider a scaled system (see Sect. 3.4 below). Moreover, the application of our preconditioner uses our diagonalization of Sections 2.3 and 2.4. Therefore, our preconditioner is applicable to a broader class of problems.

Our preconditioner is motivated by the previous observations. In addition to (2.6) and (2.8) define

$$\widetilde{C}_\theta = \begin{pmatrix} 1-\theta & & & \theta \\ \theta & 1-\theta & & \\ & \ddots & \ddots & \\ & & \theta & 1-\theta \end{pmatrix}, \quad \widetilde{\mathbf{C}}^{(\theta)}_{1,2} = \begin{pmatrix} 0 & \widetilde{C}_\theta^T \\ 0 & 0 \end{pmatrix}, \quad \widetilde{\mathbf{C}}^{(\theta)}_{2,1} = \begin{pmatrix} 0 & 0 \\ \widetilde{C}_\theta & 0 \end{pmatrix}, \quad \theta \in [0,1]. \tag{3.19}$$

Our preconditioner for $\mathbf{K}_{\rm init}$ defined in (3.6b) is given by

$$\widetilde{\mathbf{K}}_{\rm per} = \mathbf{C}^{(-1)}_{1,1}\otimes M + \delta t\mathbf{C}^{(\theta)}_{1,1}\otimes A - \delta t\widetilde{\mathbf{C}}^{(\theta)}_{1,2}\otimes BR^{-1}B^T \\ + \mathbf{C}^{(-1)}_{2,2}\otimes M^T + \delta t\mathbf{C}^{(\theta)}_{2,2}\otimes A^T + \delta t\widetilde{\mathbf{C}}^{(\theta)}_{2,1}\otimes Q. \tag{3.20}$$

Note that $\widetilde{\mathbf{K}}_{\rm per}$ is different from (2.9b) when $\theta \neq 1/2$, since in this case $\widetilde{\mathbf{C}}^{(\theta)}_{1,2} \neq \mathbf{C}^{(\theta)}_{1,2}$ and $\widetilde{\mathbf{C}}^{(\theta)}_{2,1} \neq \mathbf{C}^{(\theta)}_{2,1}$. However, $\widetilde{\mathbf{K}}_{\rm per}$ can be diagonalized in a similar way as $\mathbf{K}_{\rm per}$. See Appendix A for details. This diagonalization can also be used to prove the invertibility of the preconditioning matrix in (3.20).

Theorem 3.4. *Assume* (1.1) *holds and that the eigenvalues of* $M^{-1}A \in \mathbb{R}^{n_y \times n_y}$ *have strictly positive real parts. If* $\theta \in [1/2, 1]$ *or if* $\theta \in [0, 1/2)$ *and* $\delta t < 2Re(\lambda)/((1-2\theta)|\lambda|^2)$ *for all eigenvalues* λ *of* $M^{-1}A$*, then* $\widetilde{\mathbf{K}}_{per}$ *in* (3.20) *is invertible.*

Proof. We will prove this result using the diagonalization of $\widetilde{\mathbf{K}}_{\text{per}}$, which is worked out in Appendix A. Diagonalization of $\widetilde{\mathbf{K}}_{\text{per}}$ and invertibility of the diagonal blocks, which will be established in Theorem A.2, immediately imply this result. ☐

In practice, we will solve the preconditioned system using GMRES or another Krylov subspace method for general systems. Although in the nonsymmetric case, the spectrum of the preconditioned matrix alone does not characterize the convergence behavior of the iterative method, see, *e.g.*, [8], in practice the spectrum of the preconditioned matrix often still provides useful information. We prove that at least $2n_y(n_t-1)$ eigenvalues of the preconditioned matrix $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ are equal to one. This result is based on the general observations made at the beginning of this section and is also the type of result shown and proof used in [1], Section 3.2 for their preconditioned system.

Theorem 3.5. *For any* θ*, at least* $2n_y(n_t-1)$ *eigenvalues of the preconditioned matrix* $\widetilde{\mathbf{K}}_{per}^{-1}\mathbf{K}_{init}$ *are equal to one, i.e., at most* $2n_y$ *eigenvalues are not equal to one.*

Proof. Note that $\mathbf{I} - \widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}} = \widetilde{\mathbf{K}}_{\text{per}}^{-1}\big(\widetilde{\mathbf{K}}_{\text{per}} - \mathbf{K}_{\text{init}}\big)$, where

$$\widetilde{\mathbf{K}}_{\text{per}} - \mathbf{K}_{\text{init}} = \begin{pmatrix} (C_{-1} - T_{-1}) \otimes M + \delta t (C_\theta - T_\theta) \otimes A & 0 \\ (\widetilde{C}_\theta - \widetilde{T}_\theta) \otimes Q & (C_{-1} - T_{-1})^T \otimes M^T + \delta t (C_\theta - T_\theta)^T \otimes A^T \end{pmatrix}. \tag{3.21}$$

The matrix (3.21) has a rank of $2n_y$. This is because using the definitions in (3.3) and (3.19), the first block column

$$\begin{pmatrix} (C_{-1} - T_{-1}) \otimes M + \delta t (C_\theta - T_\theta) \otimes A \\ (\widetilde{C}_\theta - \widetilde{T}_\theta) \otimes Q \end{pmatrix}$$

has a rank of at most n_y because $(C_{-1} - T_{-1})$, $(C_\theta - T_\theta)$ and $(\widetilde{C}_\theta - \widetilde{T}_\theta)$ all contain their only nonzero entry in their last column. The second block column

$$\begin{pmatrix} 0 \\ (C_{-1} - T_{-1})^T \otimes M^T + \delta t (C_\theta - T_\theta)^T \otimes A^T \end{pmatrix}$$

also only has a rank of at most n_y because $(C_{-1} - T_{-1})^T$ and $(C_\theta - T_\theta)^T$ both contain their only nonzero entry in their first column. Thus, the whole matrix $\mathbf{I} - \widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ has a rank of at most $2n_y$, which implies the result. ☐

In Appendix B we provide additional observations on the spectrum of $\mathbf{I} - \widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$. Specifically, we conjecture that for large n_t (small δt) the nonzero eigenvalues of $\mathbf{I} - \widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ approximate the eigenvalues of a system defined *via* differential equations (the system (B.5)). We observe such a behavior in our numerical examples, but the detailed connection between the non-zero eigenvalues $\lambda \neq 0$ of $\mathbf{I} - \widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ and this systems defined *via* differential equations are missing. Such estimates of the eigenvalues of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ would be interesting, but may not provide an explanation of the convergence behavior of GMRES and other iterative methods. As we have mentioned before, for general nonsymmetric matrices there is no theoretical connection between the convergence behavior of GMRES and the eigenvalue distribution of the preconditioned system. See, *e.g.*, [8].

3.4. Comparison with previous diagonalization preconditioner approach

In this section, we compare our preconditioner to that of [1] in the case where $A = A^T$, M is symmetric positive definite, and $Q = M$, $B = \pm M$, $R = \alpha M$, $\alpha > 0$, which is the setting considered in [1]. As before, $\theta \in [0,1]$. The assumption $A = A^T$is only needed for the final diagonalization step of the preconditioner, analogously to (2.26), but is not needed to describe the basic approach. Therefore, we initially allow $A \neq A^T$ to highlight where symmetry of A is needed. If $Q = M$, $B = \pm M$, $R = \alpha M$, then the Kronecker product form (3.4) of the discretized optimality system (3.2) becomes

$$(T_{-1} \otimes M + T_\theta \otimes \delta t A)\, Y - (\widetilde{T}_\theta^T \otimes \delta t \alpha^{-1} M) P = F_Y, \tag{3.22a}$$
$$(\widetilde{T}_\theta \otimes \delta t M) Y + \left(T_{-1}^T \otimes M^T + T_\theta^T \otimes \delta t A^T\right) P = F_P, \tag{3.22b}$$

where Y and P are defined as in (3.4c), and F_Y and F_P are defined as in (3.4d) with $B = \pm M$ and $R = \alpha M$.

Since the preconditioner in [1] is built on the diagonalization ideas in Section 2.5, [1] scale the optimality system. This requires $\theta \in (0,1]$ to ensure invertibility of T_θ, and that the number of time steps K is odd if $\theta = 1/2$ to ensure invertibility of C_θ.

Since the preconditioner in [1] is built on the diagonalization ideas in Section 2.5, [1] multiply the first equation by $(T_\theta^{-1} \otimes I)$, multiply the second equation by $(T_\theta^{-T} \otimes I)$, and scale by $\beta > 0$. With these scalings, the system of equations (3.22) can be equivalently expressed as

$$\underbrace{\left(\begin{pmatrix} T_\theta^{-1} T_{-1} & -\delta t (\alpha\beta)^{-1} T_\theta^{-1} \widetilde{T}_\theta^T \\ \delta t \beta \, T_\theta^{-T} \widetilde{T}_\theta & T_\theta^{-T} T_{-1}^T \end{pmatrix} \otimes M + \delta t \mathbf{I}_{1,1} \otimes A + \delta t \mathbf{I}_{2,2} \otimes A^T \right)}_{\mathbf{K}_{\text{init,scaled}}} \begin{pmatrix} \widehat{Y} \\ P \end{pmatrix} = \begin{pmatrix} \beta^{-1} \widetilde{F_Y} \\ \widetilde{F_P} \end{pmatrix}, \tag{3.23}$$

where $\widehat{Y} = \beta^{-1} Y$, $\widetilde{F_Y} = (T_\theta^{-1} \otimes I) F_Y$, $\widetilde{F_P} = (T_\theta^{-T} \otimes I) F_P$. The preconditioner of $\mathbf{K}_{\text{init,scaled}}$ is obtained by replacing the Toeplitz matrices by their corresponding circulant matrices. The preconditioner is

$$\widetilde{\mathbf{K}}_{\text{per,scaled}} = \underbrace{\begin{pmatrix} C_\theta^{-1} C_{-1} & -\delta t (\alpha\beta)^{-1} C_\theta^{-1} \widetilde{C}_\theta^T \\ \delta t \beta \, C_\theta^{-T} \widetilde{C}_\theta & C_\theta^{-T} C_{-1}^T \end{pmatrix}}_{\mathbf{G}_{\text{prec}}(\beta)} \otimes M + \delta t \mathbf{I}_{1,1} \otimes A + \delta t \mathbf{I}_{2,2} \otimes A^T. \tag{3.24}$$

Using Theorem 2.4, $\mathbf{G}_{\text{prec}}(\beta)$ can be decomposed into

$$\begin{pmatrix} C_\theta^{-1} C_{-1} & -\delta t (\alpha\beta)^{-1} C_\theta^{-1} \widetilde{C}_\theta^T \\ \delta t \beta \, C_\theta^{-T} \widetilde{C}_\theta & C_\theta^{-T} C_{-1}^T \end{pmatrix} = \begin{pmatrix} V^{-1} & \\ & V^{-1} \end{pmatrix} \begin{pmatrix} \Lambda(\theta) & -\delta t (\alpha\beta)^{-1} \widetilde{\Lambda}(\theta) \\ \delta t \beta \, \widetilde{\Lambda}(\theta)^* & \Lambda(\theta)^* \end{pmatrix} \begin{pmatrix} V & \\ & V \end{pmatrix}, \tag{3.25}$$

where $\Lambda(\theta) = D_\theta^{-1} D_{-1}$, $\widetilde{\Lambda}(\theta) = D_\theta^{-1} \widetilde{D}_\theta^*$, V is given in (2.10), and D_{-1}, and D_θ are given in Corollary 2.5, and $\widetilde{D}_\theta = \operatorname{diag}(1, (1-\theta) + \theta\omega, \ldots, (1-\theta) + \theta\omega^{K-1})$. The block 2×2 matrix on the right hand side of (3.25) can be diagonalized similarly to (2.25). Moreover, if $A = A^T$, then $\widetilde{\mathbf{K}}_{\text{per,scaled}}$ can be diagonalized analogously to (2.26). This diagonalization is used to apply the preconditioner $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}$.

Remark 3.6. In the following numerical results, we implement $\mathbf{K}_{\text{init,scaled}}$ and corresponding preconditioner $\widetilde{\mathbf{K}}_{\text{per,scaled}}$ as shown in (3.23) and (3.24). As already noted in Remark 2.10, the authors of [1] implement both

$\mathbf{K}_{\text{init,scaled}}$ and $\widetilde{\mathbf{K}}_{\text{per,scaled}}$ by scaling by the mass matrix. That is, they set $L = M^{-1}A$ and use

$$\mathbf{K}_{\text{init,scaled}} = \begin{pmatrix} T_\theta^{-1}T_{-1} & -\delta t(\alpha\beta)^{-1}T_\theta^{-1}\widetilde{T}_\theta^T \\ \delta t\beta\, T_\theta^{-T}\widetilde{T}_\theta & T_\theta^{-T}T_{-1}^T \end{pmatrix} \otimes I_y + \delta t\mathbf{I} \otimes L,$$
$$\widetilde{\mathbf{K}}_{\text{per,scaled}} = \mathbf{G}_{\text{prec}}(\beta) \otimes I_y + \delta t\mathbf{I} \otimes L.$$

We conclude the comparison with the previous approach by highlighting the similarities when $\theta = 1$.

Theorem 3.7. *If $\theta = 1$, and $A = A^T$, $Q = M = M^T$, $B = \pm M$, $R = \alpha M$, $\alpha > 0$, then the eigenvalues of $\widetilde{\mathbf{K}}_{per}^{-1}\mathbf{K}_{init}$ and $\widetilde{\mathbf{K}}_{per,scaled}^{-1}\mathbf{K}_{init,scaled}$ are identical for any β.*

Proof. When $\theta = 1$, $C_\theta = T_\theta = I_{n_t} \in \mathbb{R}^{n_t \times n_t}$. Thus, if $A = A^T$, $Q = M = M^T$, $B = \pm M$, $R = \alpha M$, then $\widetilde{\mathbf{K}}_{\text{per}}$ and $\mathbf{K}_{\text{init}}$ differ from $\widetilde{\mathbf{K}}_{\text{per,scaled}}$ and $\mathbf{K}_{\text{init,scaled}}$ solely by the factor β. Specifically, if

$$\mathbf{B} = \begin{pmatrix} I_{n_t} & \\ & \beta I_{n_t} \end{pmatrix} \otimes I_y,$$

then the matrices (3.6b), and (3.23) and (3.20), (3.24) satisfy $\mathbf{K}_{\text{init,scaled}} = \mathbf{B}\mathbf{K}_{\text{init}}\mathbf{B}^{-1}$, $\widetilde{\mathbf{K}}_{\text{per,scaled}} = \mathbf{B}\widetilde{\mathbf{K}}_{\text{per}}\mathbf{B}^{-1}$, and $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}} = \mathbf{B}(\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}})\mathbf{B}^{-1}$ for $\theta = 1$. Thus $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ and $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$ have the same eigenvalues. □

4. Numerical results

The first example is taken from [1] and is used to compare with eigenvalue computations in Figure 8 of [1]. Consider

$$\min \quad \frac{1}{2}\int_0^T\int_0^1 (y(t,x) - y_{des}(t,x))^2 \,\mathrm{d}x\,\mathrm{d}t + \frac{\alpha}{2}\int_0^T\int_0^1 u(t,x)^2 \,\mathrm{d}x\,\mathrm{d}t, \tag{4.1a}$$
$$\text{s.t.} \quad \frac{\partial}{\partial t}y(t,x) - \frac{\partial}{\partial x}(\kappa\frac{\partial}{\partial x}y(t,x)) = f(t,x) + u(t,x), \quad x \in (0,1),\ t \in (0,T), \tag{4.1b}$$
$$\kappa\frac{\partial}{\partial x}y(t,0) = \kappa\frac{\partial}{\partial x}y(t,1) = 0, \quad t \in (0,T), \qquad y(0,x) = y_0(x), \quad x \in (0,1). \tag{4.1c}$$

We use $\kappa = 1$, $T = 1$, $f(t,x) = 0$, $y_{des}(t,x) = 1$, $y_0(x) = 0$, and, unless noted otherwise, $\alpha = 10^{-5}$. To semi-discretize the problem we use a centered finite difference discretization on a uniform grid with mesh size $h = 1/(n_y - 1)$. This leads to a problem of the type (1.2a,b,c) with a tridiagonal matrix

$$A = \frac{\kappa}{h^2}\begin{pmatrix} 1 & -1 & & \\ 1 & -2 & 1 & \\ & \ddots & \ddots & \ddots \\ & 1 & -2 & 1 \\ & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{n_y \times n_y},$$

$M = I_y \in \mathbb{R}^{n_y \times n_y}$, $B = -I_y \in \mathbb{R}^{n_y \times n_y}$, $Q = I_y \in \mathbb{R}^{n_y \times n_y}$, and $R = \alpha\, I_y \in \mathbb{R}^{n_y \times n_y}$. (For the discretization of the objective one would use $Q = h\, I_y \in \mathbb{R}^{n_y \times n_y}$, and $R = \alpha h\, I_y \in \mathbb{R}^{n_y \times n_y}$, but we remove this scaling by h to be consistent with the example set-up in [1]. The scaling of the semi-discretized objective by a constant will not change the solution of the semi-discretized optimal control problem.)

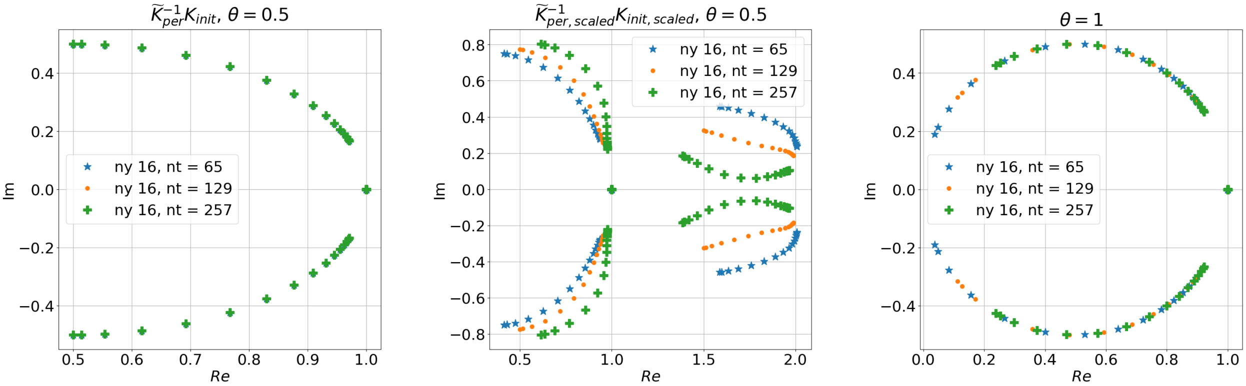

FIGURE 1. Eigenvalues of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ (left) and $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$ (center) for the discretization of the optimal control problem (4.1) with $\theta = 1/2$, $n_y = 16$ and different numbers of time steps n_t. The eigenvalues of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ do not change much as n_t increases and are better clustered than the eigenvalues of $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$ which change with the different numbers of time steps n_t shown. When $\theta = 1$, $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ and $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$ have the same eigenvalues (right).

The left and middle plots in Figure 1 show the eigenvalues of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ and $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$ (introduced in (3.24)) for $\theta = 1/2$ and different discretizations with constant spatial mesh size $n_y = 16$ and different numbers of time steps n_t. These discretization sizes are the ones used in [1]. Notice that the eigenvalues of the new $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ are better clustered than those of $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$. In addition, the eigenvalues of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ essentially do not change with increasing n_t while the eigenvalues of $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$ do. When $\theta = 1$, Theorem 3.7 proves that the eigenvalues of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ and $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$ are identical. These eigenvalues are shown in the right plot in Figure 1. In this $\theta = 1$ case, the eigenvalues cluster slightly better as n_t increases. The left and right plots in Figure 1 support our conjecture that the eigenvalues of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ in the limit, as the number of time steps n_t increases, approximate values (B.14) that are obtained from the eigenvalues of the matrix $\mathbf{S}$ in (B.10).

The results of Figure 1 are generated using the temporal and spatial discretizations applied in [1] for comparison with [1]. However, for a $\theta = 1/2$ time-stepping scheme, the discretization error is $O(h^2 + \delta t^2)$. Thus it is not practical to choose $\delta t \ll h$ (or in this setting, $n_t \gg n_y$), as was done here. Therefore, we explore the spectrum of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ for $\theta = 1/2$ and $\delta t \approx h$ ($n_t \approx n_y$) in Figure 2.

Figure 2 reveals that the eigenvalues of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ differ significantly between odd and even n_t before appearing to settle down to a limiting distribution for larger n_t. Recall that the preconditioner proposed by [1] cannot be used for an even number of time steps n_t for $\theta = 1/2$. The eigenvalue plots for $n_t = 32, 33$ are approaching the eigenvalue plots for $n_t = 65, 129, 257$ seen in the $\theta = 1/2$ case of Figure 1 (left plot). Again, this supports our conjecture that as the number of time steps n_t increases the eigenvalues of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ approximate the values (B.14) that are obtained from the eigenvalues of the matrix $\mathbf{S}$ in (B.10).

Since [1] provides plots of the singular values of the preconditioned matrices, we include them in Figure 3. As expected, most singular values are one.

We now investigate the performance of GMRES on the preconditioned systems with system matrices $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ and $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$. GMRES is stopped when the 2-norm of the relative preconditioned residual is below 10^{-6} or 1,000 iterations are reached.

Both preconditioners $\widetilde{\mathbf{K}}_{\text{per}}$ and $\widetilde{\mathbf{K}}_{\text{per,scaled}}$ applied to their respective systems substantially reduce the number of GMRES iterations. Figure 4 shows that the proposed preconditioner is slightly better in terms of the number

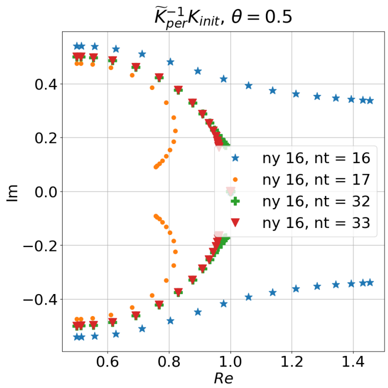

FIGURE 2. Eigenvalues of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ for $\theta = 1/2$, $n_y = 16$, $n_t = 16, 17, 32, 33$. While the eigenvalue plots for $n_t = 16$ and $n_t = 17$ wildly differ, the eigenvalue plots for $n_t = 32$ and $n_t = 33$ are nearly identical and resemble those seen in the $\theta = 1/2$ case of Figure 1.

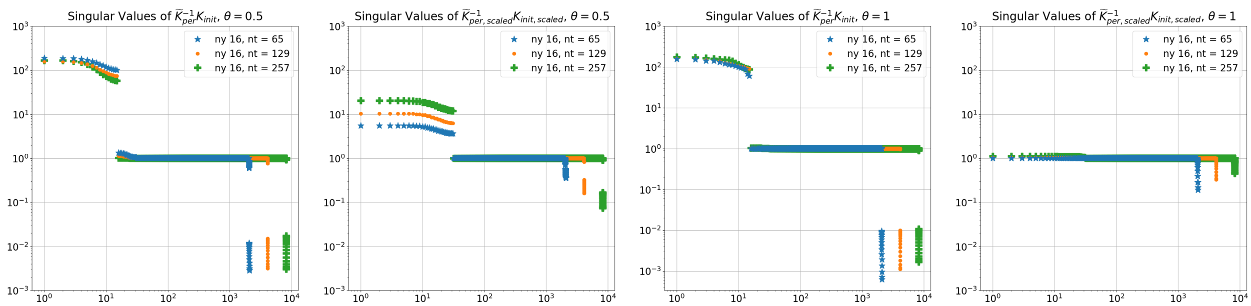

FIGURE 3. Singular values of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ (left, and second from right) and $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$ (second from left, and right) for the discretization of the optimal control problem (4.1) with $\theta = 1/2$ (left two plots), $\theta = 1$ (right two plots), $n_y = 16$, and different numbers of time steps n_t.

of preconditioned GMRES iterations. It is important for us to emphasize here that while we have plotted the relative preconditioned residuals of both problems together, the relative preconditioned residuals should not be directly compared - they are, as their name implies, relative to the initial preconditioned residual of their respective problem. Moreover, the preconditioned residuals are different when computed with the same vector. Moreover, the costs of the application of conditioners are different. Application of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}$ requires (parallel) solution of n_t complex valued $2n_y \times 2n_y$ systems, whereas application of $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}$ requires (parallel) solution of $2n_t$ complex valued $n_y \times n_y$ systems.

The convergence results in Figure 4 appear to correspond to the spectral plots seen in Figure 1. For the $\theta = 1/2$ case of Figure 1, it was observed that the eigenvalues of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ are better clustered than those of $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$. GMRES improves the relative residual faster for $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ in this case. For the $\theta = 1$ case from Figure 1, the eigenvalue distributions are the same. The GMRES convergence plots for the $\theta = 1$ case match closely. In addition, the performance of GMRES slightly improves as n_t increases. The eigenvalue plot for the $\theta = 1$ case from Figure 1 show that the eigenvalues get slightly more clustered as n_t increases.

We do not provide a GMRES analogue for Figure 2 in this example, but will explore the effect of odd vs. even time step number n_t for our preconditioner in next numerical examples.

Finally, we explore the effect of varying the control penalty α on the performance of GMRES for our preconditioned system $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$. Table 1 shows that as the regularization α approaches 0, the number of GMRES

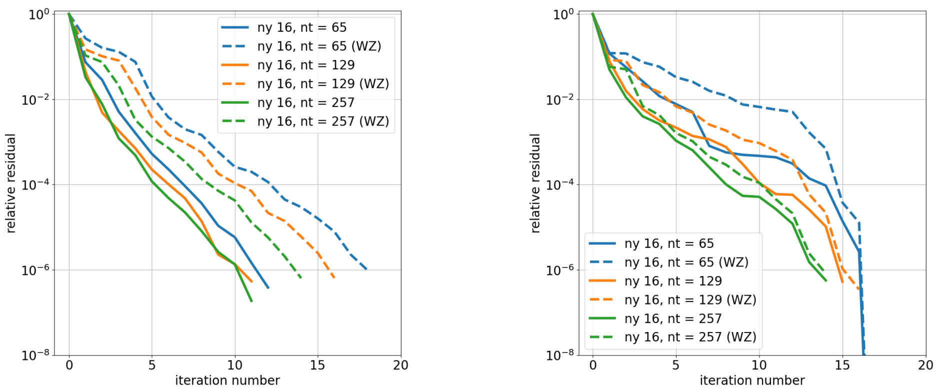

FIGURE 4. GMRES relative residual plots using $\widetilde{\mathbf{K}}_{\mathrm{per}}^{-1}\mathbf{K}_{\mathrm{init}}$ (solid) and $\widetilde{\mathbf{K}}_{\mathrm{per,scaled}}^{-1}\mathbf{K}_{\mathrm{init,scaled}}$ (dashed, introduced in [1]) for the discretization of the optimal control problem (4.1) with $\theta = 1/2$ (left) and $\theta = 1$ (right), $n_y = 16$ and different numbers of time steps n_t.

TABLE 1. Number of preconditioned GMRES iterations needed to reach a relative residual tolerance of 10^{-6} for the discretization of the optimal control problem (4.1) with $n_x = 16$, $n_t = 257$ and varying control regularization parameters α.

	$\alpha = 10^{-1}$	$\alpha = 10^{-3}$	$\alpha = 10^{-5}$
$\theta = 1/2$	5	7	11
$\theta = 1$	5	7	14

iterations needed stay small for both $\theta = 1$ and $\theta = 1/2$, but increases slightly. As α approaches 0, the conditioning of the discretized optimality system deteriorates. Recall that the computations reported in Figure 1 to 4 are performed with $\alpha = 10^{-5}$.

The second problem is derived from a semi-discretization of the following problem, which was also used in [1]:

$$\min \quad \int_0^T \int_\Omega \frac{1}{2}(y(t,x) - y_{des}(t,x))^2 + \frac{\alpha}{2} u(t,x)^2 \, \mathrm{d}x \, \mathrm{d}t, \tag{4.2a}$$

$$\text{s.t.} \quad \partial_t y(t,x) - \partial_{x_1}(\mathrm{e}^{-5x_1}\partial_{x_1} y(t,x)) - \partial_{x_2x_2} y(t,x) = u(t,x), \quad x \in \Omega, \ t \in (0,T), \tag{4.2b}$$

$$y(t,x) = 0, \quad t \in (0,T), x \in \partial\Omega, \qquad y(0,x) = y_0(x), \quad x \in \Omega. \tag{4.2c}$$

For this problem, $\Omega = (-1/2, 1/2)^2$, $T = 2$,

$$y_{des}(t,x) = \frac{(x_1^2 - \frac{1}{4})(x_2^2 - \frac{1}{4})}{x_1^2 + x_2^2 + \frac{1}{10}} \sin\left(5\pi x_1^2 \sqrt{\frac{1}{2} - x_2} + 2\pi t\right), \qquad y_0(x) = y_{des}(0,x),$$

and, unless noted otherwise, $\alpha = 10^{-5}$.

To semi-discretize the problem we use piecewise linear finite elements on a uniform triangulation of the unit square obtained as follows. First, we decompose $\Omega = (-1/2, 1/2)^2$ into squares of size $h \times h$ with $h = 1/(n_x + 1)$, and then we divide (south–west to north–east) each square into two triangles. This leads to a problem of the type (1.2a,b,c) with symmetric positive definite $Q = M$, $B = -M$, and $R = \alpha M$ and $n_y = n_x^2$. In particular, the preconditioner of [1] can be applied to this problem.

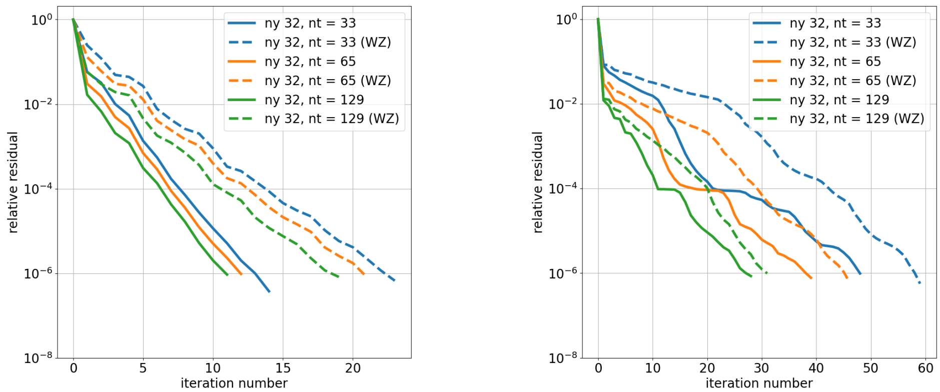

FIGURE 5. GMRES relative residual plots using $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ (solid) and $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$ (dashed, introduced in [1]) for the discretization of the optimal control problem (4.2) with $\theta = 1/2$ (left) and $\theta = 1$ (right), for constant spatial discretization $n_x = 32$ and variable temporal discretization.

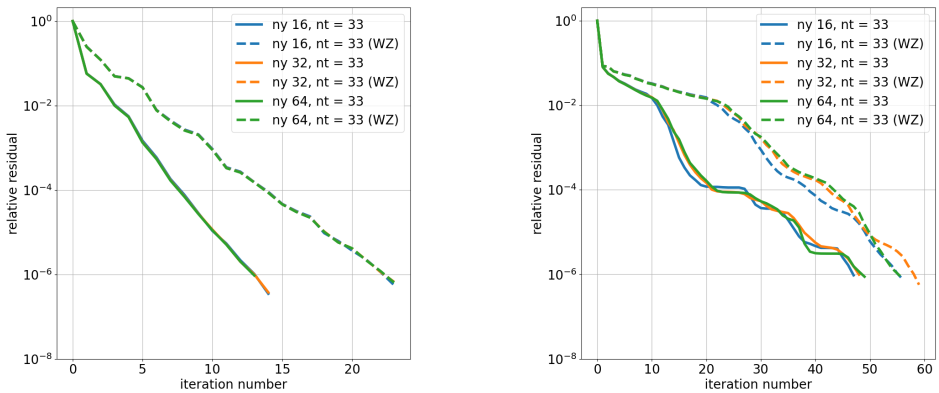

FIGURE 6. GMRES relative residual plots using $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ (solid) and $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$ (dashed, introduced in [1]) for the discretization of the optimal control problem (4.2) with $\theta = 1/2$ (left) and $\theta = 1$ (right), constant $n_t = 33$ and variable spatial discretization.

We begin by comparing our preconditioner $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ to the preconditioner $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$ proposed in [1]. Again, GMRES is stopped when the 2-norm of the relative preconditioned residual is below 10^{-6} or 1,000 iterations are reached.

Figure 5 demonstrates the same behavior as the behavior seen in Figure 4. In the $\theta = 1/2$ case, our preconditioner reduces the relative residual slightly faster, presumably due to the better eigenvalue clustering (Fig. 1). In the $\theta = 1$ case the two preconditioners reduce the relative residual at the same rate, and both get more efficient as n_t increases. Recall that this behavior for the $\theta = 1$ case was seen in Figure 4, but is much more pronounced here.

In addition to keeping the spatial discretization constant and changing the number of timesteps, we also kept the number of timesteps constant and varied the spatial discretization, as is shown in the following Figure 6.

Figure 6 shows that both preconditioners are insensitive to the spatial discretization. Note that all solid lines lie on each other and all dotted lines lie on each other in the $\theta = 1/2$ case. As we have noted before, we plot preconditioned residuals, which are different in cases $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ and $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}\mathbf{K}_{\text{init,scaled}}$. Moreover, the

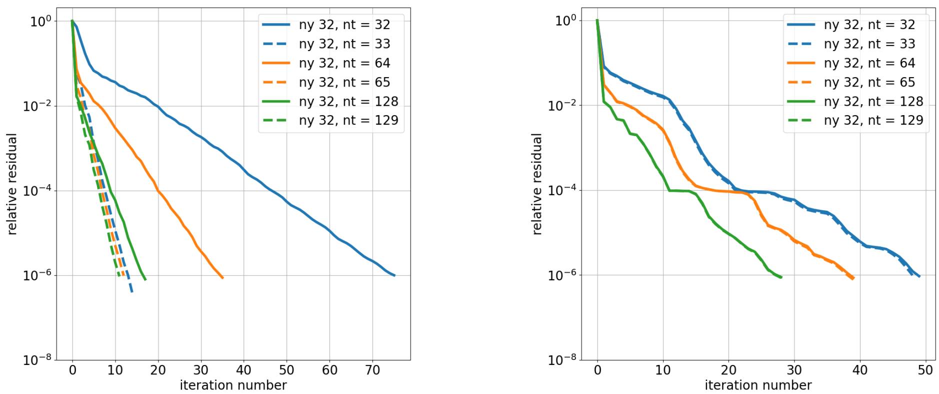

FIGURE 7. GMRES relative residual plots using $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ for the discretization of the optimal control problem (4.2) with $\theta = 1/2$ (left) and $\theta = 1$ (right), and different problem sizes $n_x = 32$ and varied n_t. When $\theta = 1/2$ and the number of time steps n_t is odd, the periodic preconditioner requires fewer iterations. When $\theta = 1$, there is little difference between odd and even time steps.

TABLE 2. Number of preconditioned GMRES iterations needed to reach relative residual tolerance of 10^{-6} for the discretization of the optimal control problem (4.2) with $n_x = 32$, $n_t = 65$ and varying control regularization parameters α.

	$\alpha = 10^{-1}$	$\alpha = 10^{-3}$	$\alpha = 10^{-5}$
$\theta = 1/2$	8	9	12
$\theta = 1$	5	9	39

application of $\widetilde{\mathbf{K}}_{\text{per}}^{-1}$ requires a (parallel) solution of n_t complex valued $2n_y \times 2n_y$ systems, whereas application of $\widetilde{\mathbf{K}}_{\text{per,scaled}}^{-1}$ requires (parallel) solution of $2n_t$ complex valued $n_y \times n_y$ systems.

The following Figure 7 displays GMRES applied to the preconditioned system $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ for even and odd numbers of time steps n_t. Recall that in Figure 2, for a 1D-in space optimal control problem, for $\theta = 1/2$ and for $\delta t \approx h$, the eigenvalue distribution between odd and even n_t differed significantly (with the odd n_t eigenvalues significantly better clustered than their even counterparts), but as n_t increased the spectrum changed little.

What was seen in Figure 2 is borne out in the GMRES convergence plot in Figure 7. While the odd vs even n_t matter a great deal for the convergence of GMRES in the $\theta = 1/2$ case when $h \approx \delta t$, as n_t increases the difference between even and odd n_t becomes increasingly irrelevant. The odd vs even n_t is not an issue in the $\theta = 1$ case. Note that in the $\theta = 1$ case, the dashed and dotted lines of the same color lie on top of each other. In the $\theta = 1$ case, the improvement to GMRES as n_t increases has also been seen in Figures 4 and 5.

Finally, as, was done in the previous numerical example, we examine the effects of varying the control regularization parameter α for our preconditioned system $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$. Table 2 shows that as the regularization α approaches 0, the number of GMRES iterations stays small for both $\theta = 1$ and $\theta = 1/2$, but increases. For $\theta = 1/2$, the number of iterations increases only slightly, but for $\theta = 1$, the increase in the number of iterations is larger as α decreases from 10^{-3} to 10^{-5}. Recall that the computations reported in Figures 5 to 7 are done with $\alpha = 10^{-5}$.

Our third and final problem is derived from a semi-discretization of

$$\min \quad \frac{1}{2}\int_0^T \int_{\Omega_{\text{obs}}} (y(t,x) - y_{des}(t,x))^2 \, \mathrm{d}x \, \mathrm{d}t + \frac{\alpha}{2}\int_0^T \int_{\Omega_c} u(t,x)^2 \, \mathrm{d}x \, \mathrm{d}t, \tag{4.3a}$$

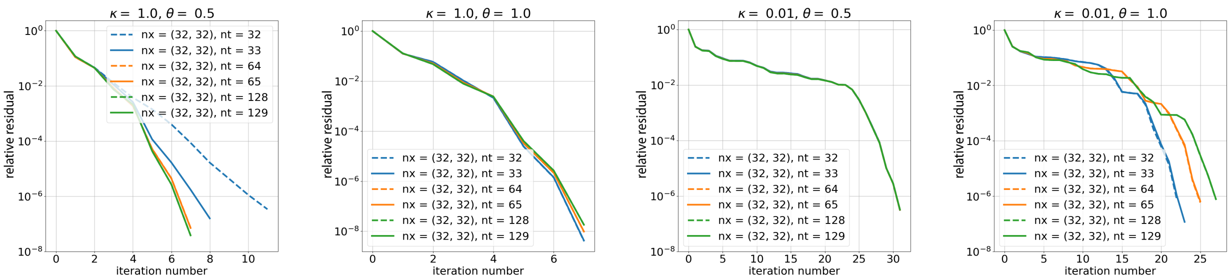

FIGURE 8. GMRES relative residual plots using $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ for the discretization of the optimal control problem (4.3) with $\theta = 1/2,\ 1,\ \kappa = 1,\ 0.01$, constant spatial discretization $n_x = 32$ and variable time discretization. For both values of θ, the GMRES iteration count is mostly independent of n_t.

$$\text{s.t.}\quad \frac{\partial}{\partial t}y(t,x) - \nabla\cdot(\kappa\nabla y(t,x)) + \beta^T\cdot\nabla y(t,x) = f(t,x) + u(t,x)\chi_{\Omega_c}(x), \quad x\in\Omega,\ t\in(0,T), \tag{4.3b}$$

$$y(t,x) = y_d(t,x), \quad t\in(0,T), x\in\partial\Omega, \qquad y(0,x) = y_0(x), \quad x\in\Omega. \tag{4.3c}$$

In this setting, the domain $\Omega = (0,1)^2$ is the unit square, the control region is $\Omega_c = (0,0.3]\times(0,0.3]$, and the observation region is $\Omega_{\text{obs}} = [0.5,1)\times[0.5,1)$. The function χ_{Ω_c} is the indicator function on Ω_c. We use $\kappa = 1$ or $\kappa = 10^{-2}$, $\beta = [1,1]$, $T = 1$, $f(t,x) = 0$, $y_{des}(t,x) = 1$, $y_d(t,x) = 1$, $y_0(x) = \cos(\pi x_1)\cos(\pi x_2)$, and unless noted otherwise $\alpha = 10^{-5}$. To semi-discretize the problem we use an upwind finite difference discretization on a uniform grid with mesh size $h_1 = 1/(n_{x_1}+1)$, $h_2 = 1/(n_{x_1}+1)$.

The preconditioner of [1] cannot be applied to this problem. In fact, this problem displays all properties that make the preconditioner proposed in [1] unusable: limited state observation ($Q \neq M$), limited control region ($B \neq M$, $R \neq \alpha M$), and the presence of an advection term ($A \neq A^T$).

We first study the impact of varying the number of timesteps on the convergence of preconditioned GMRES.

Note how Figure 8 differs from its counterparts in Figure 7. In Figure 7, recall for the $\theta = 1/2$ case that there was a large discrepancy between even and odd n_t that only resolved as n_t became large. In this advection-diffusion case, there is no difference between even and odd timesteps even for small n_t for the $\theta = 1/2$ case. In addition, for the $\theta = 1$ case, Figure 7 displays an improvement in the performance of GMRES as n_t in increased; here we see no such improvement, although we do note that the performance is already very good. For the $\kappa = 1.0$, $\theta = 1$ case and for the $\kappa = 0.01$, $\theta = 0.5$ case, all of the lines lay on each other.

To help explain these differences, we look at eigenvalue distributions of this problem (4.3). Note that because of the overall problem size, we compute the eigenvalues for coarser discretizations compared to those used for the GMRES runs in Figure 8. These eigenvalue plots are shown in the following Figure 9.

Figure 9 may help explain the convergence behavior of GMRES seen in Figure 8. For the $\kappa = 1$ case, the eigenvalues are tightly clustered for both $\theta = 1/2$ and $\theta = 1$ for larger n_t. This would explain the convergence of GMRES in 5 iterations for the $\kappa = 1$ case. In addition, while there still is a difference between even and odd n_t in the $\kappa = 1$ case, they are minor (note the scale on the real axis for the $\kappa = 1, \theta = 0.5$ cases). This explains why Figure 8 does not display the same even-odd phenomena in the $\theta = 1/2$ case that Figure 7 does. It also appears that for this advection-diffusion problem, the eigenvalues converge to their conjectured limiting distribution given *via* the eigenvalues of $\mathbf{S}$ in (B.10) faster than in the previous problems. One final observation is that the eigenvalue distributions are nearly identical between the $\theta = 1/2$ and $\theta = 1$ cases. This explains why GMRES needed roughly the same number of iterations to converge regardless of which θ was used, which was not the case with Figure 7 (there, $\theta = 1/2$ outperformed $\theta = 1$ by a considerable amount).

Next, we examine the effect of varying the spatial discretization while keeping the number of time steps constant for this problem. The following Figure 10 is the analogue to Figure 6 in the previous example.

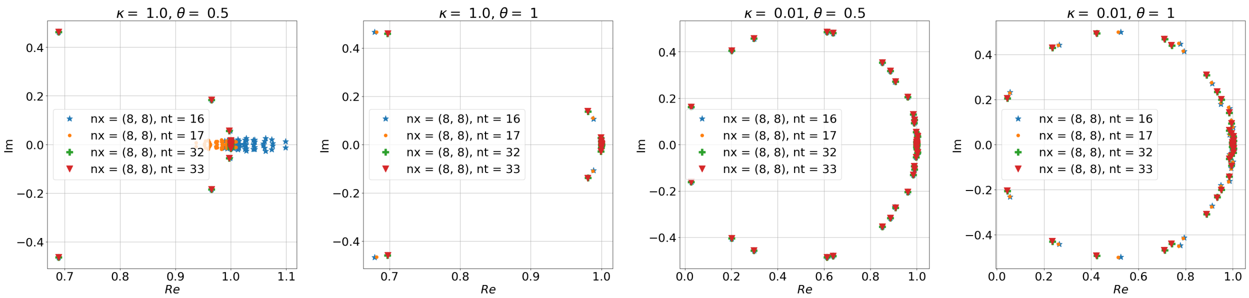

FIGURE 9. Eigenvalues of $\widetilde{\mathbf{K}}_{\text{per}}\mathbf{K}_{\text{init}}$ for the discretization of the optimal control problem (4.3) with $\theta = 1/2$ or $\theta = 1$, $\kappa = 1$ or $\kappa = 0.01$, and constant spatial discretization $n_x = 8$ and varied n_t. The eigenvalues are mostly constant with respect to n_t in the $\theta = 1$ case. In the $\theta = 1/2$ case, there is a slight difference between odd and even n_t that disappears as n_t increases.

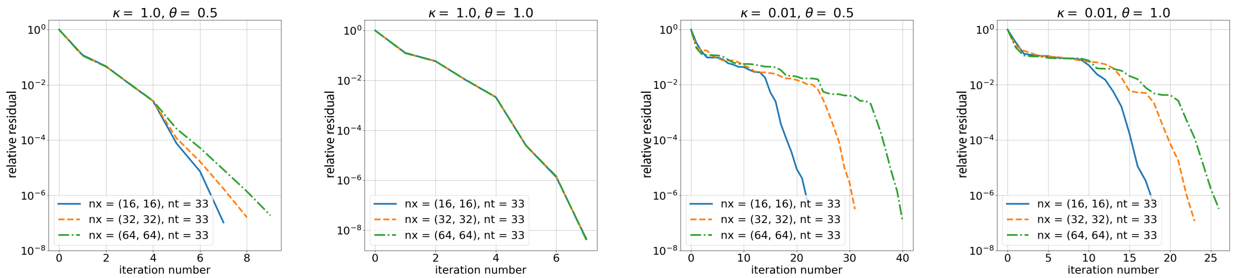

FIGURE 10. GMRES relative residual plots using $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ for the discretization of the optimal control problem (4.3) with $\theta = 1/2$ or $\theta = 1$, $\kappa = 1$ or $\kappa = 0.01$, with constant timestep $n_t = 33$ and varied spatial discretization. For both values of θ, the GMRES iteration count is mostly independent of the spatial discretization.

TABLE 3. Number of preconditioned GMRES iterations taken to reach relative residual tolerance of 10^{-6} for the discretization of the optimal control problem (4.3) with $n_x = 32$, $n_t = 33$, $\theta = 1$ (left) and $\theta = 1/2$ (right), and varying control regularization parameters α.

$\theta = 1$	$\alpha = 10^{-1}$	$\alpha = 10^{-3}$	$\alpha = 10^{-5}$
$\kappa = 1$	4	4	7
$\kappa = 10^{-2}$	8	14	23

$\theta = 1/2$	$\alpha = 10^{-1}$	$\alpha = 10^{-3}$	$\alpha = 10^{-5}$
$\kappa = 1$	6	7	8
$\kappa = 10^{-2}$	9	17	31

Figure 10 reveals that our preconditioner is insensitive to the spatial discretization in the $\kappa = 1$ case for both values of θ, which is what was also observed in Figure 6 for the previous problem. Note that for the $\kappa = 1.0$, $\theta = 1.0$ case, all lines in the plot lie on each other. In the $\kappa = 0.01$ case, the preconditioner is somewhat sensitive to the spatial discretization for both values of θ.

Finally, we examine the effect of varying the control penalty α for $\kappa = 1$ and $\kappa = 10^{-2}$ for our preconditioned system $\widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$. Table 3 once again demonstrates what was observed from the previous numerical examples (Tables 1 and 2): as α decreases, the number of GMRES iterations needed increases slightly, but stays small overall. Recall that the computations reported in Figure 8 to 10 are done with $\alpha = 10^{-5}$.

5. Extensions

The optimal control problems that motivate this work are PDE-constrained optimization problems. In this case, the formulation (1.2) is obtained from a semi-discretization of the objective function and of the PDE constraints. This corresponds to the discretize-then-optimize approach. Alternatively, one can also set up the optimality conditions for the PDE-constrained optimization problem, and then semi-discretize the state PDE, the adjoint PDE, and other equations, to arrive at a semi-discretized optimality system. This is the optimize-then-discretize approach. For some spatial semi-discretizations, especially those used for advection dominated problems the discretize-then-optimize and the optimize-then-discretize approaches can lead to different systems. See, *e.g.*, [20–22].

Instead of (2.3) and (3.1) one obtains systems of the type

$$\begin{aligned} M\frac{\mathrm{d}}{\mathrm{d}t}y(t) + A\,y(t) - BR^{-1}\widetilde{B}p(t) &= f(t) + BR^{-1}d(t), && t \in (0,T), && \text{(5.1a)}\\ M\,y(0) &= M\,y(T), \\ -\widetilde{M}\frac{\mathrm{d}}{\mathrm{d}t}p(t) + \widetilde{A}p(t) + Q\,y(t) &= -c(t), && t \in (0,T), && \text{(5.1b)}\\ \widetilde{M}p(0) &= \widetilde{M}p(T). \end{aligned}$$

and

$$\begin{aligned} M\frac{\mathrm{d}}{\mathrm{d}t}y(t) + A\,y(t) - BR^{-1}\widetilde{B}p(t) &= f(t) + BR^{-1}d(t), && t \in (0,T), && \text{(5.2a)}\\ M\,y(0) &= M\,y_0, \\ -\widetilde{M}\frac{\mathrm{d}}{\mathrm{d}t}p(t) + \widetilde{A}p(t) + Q\,y(t) &= -c(t), && t \in (0,T), && \text{(5.2b)}\\ \widetilde{M}p(T) &= 0. \end{aligned}$$

In both cases the control is $u(t) = -R^{-1}(\widetilde{B}p(t) + d(t))$.

The developments in this paper extend to the case where (2.3) and (3.1) are replaced by (5.1) and (5.2). For example, (5.1) leads to a system (2.9) with (2.9b) replaced by

$$\begin{aligned} \mathbf{K}_{\mathrm{per}} =& \mathbf{C}_{1,1}^{(-1)} \otimes M + \delta t \mathbf{C}_{1,1}^{(\theta)} \otimes A - \delta t \mathbf{C}_{1,2}^{(\theta)} \otimes BR^{-1}\widetilde{B} \\ &+ \mathbf{C}_{2,2}^{(-1)} \otimes \widetilde{M} + \delta t \mathbf{C}_{2,2}^{(\theta)} \otimes \widetilde{A} + \delta t \mathbf{C}_{2,1}^{(\theta)} \otimes Q. \end{aligned} \tag{5.3}$$

Similarly, (5.2) leads to (3.6) with (3.6b) replaced by

$$\begin{aligned} \mathbf{K}_{\mathrm{init}} =& \mathbf{T}_{1,1}^{(-1)} \otimes M + \delta t \mathbf{T}_{1,1}^{(\theta)} \otimes A - \delta t \widetilde{\mathbf{T}}_{1,2}^{(\theta)} \otimes BR^{-1}\widetilde{B} \\ &+ \mathbf{T}_{2,2}^{(-1)} \otimes \widetilde{M} + \delta t \mathbf{T}_{2,2}^{(\theta)} \otimes \widetilde{A} + \delta t \widetilde{\mathbf{T}}_{2,1}^{(\theta)} \otimes Q. \end{aligned} \tag{5.4}$$

The preconditioner for (5.4) is

$$\begin{aligned} \widetilde{\mathbf{K}}_{\mathrm{per}} =& \mathbf{C}_{1,1}^{(-1)} \otimes M + \delta t \mathbf{C}_{1,1}^{(\theta)} \otimes A - \delta t \widetilde{\mathbf{C}}_{1,2}^{(\theta)} \otimes BR^{-1}\widetilde{B} \\ &+ \mathbf{C}_{2,2}^{(-1)} \otimes \widetilde{M} + \delta t \mathbf{C}_{2,2}^{(\theta)} \otimes \widetilde{A} + \delta t \widetilde{\mathbf{C}}_{2,1}^{(\theta)} \otimes Q \end{aligned} \tag{5.5}$$

instead of (3.20). The diagonalization approach proposed in this paper for the diagonalization of (5.3) and (5.5) extends straightforwardly. One has to make the substitutions $M^T \to \widetilde{M}$, $A^T \to \widetilde{A}$, and $B^T \to \widetilde{B}$. Such an extension is not possible with the approach of [1], because of its reliance on a special problem structure.

6. Conclusions

We have introduced a new diagonalization technique for the parallel-in-time solution of linear-quadratic optimal control problems with time-invariant system matrices. Our proposed approach is motivated by the diagonalization approach in [1], which is based on requirements that limit its applicability to a smaller subclass of applications. Our diagonalization approach relaxes these requirements at the cost of being slightly more expensive to execute each iteration. We have provided new results regarding the well-posedness of the discretized optimality system and the well-posedness of the proposed preconditioner. In addition, we have provided some new theoretical results regarding the spectrum of the preconditioned system with our diagonalization technique used as a preconditioner. However, a complete, detailed analysis of the spectrum of the preconditioned system is still missing. Finally, our numerical examples show that in terms of GMRES iterations, our diagonalization method and corresponding preconditioner are at least as competitive as the preconditioning approach seen in [1] on examples where the two can be directly compared, but our preconditioner can be used for a significantly broader class of problems.

Appendix A. Diagonalization of the preconditioner

The matrix $\widetilde{\mathbf{K}}_{\rm per}$ defined in (3.20) can be diagonalized in a similar way as $\mathbf{K}_{\rm per}$ defined in (2.9b). The process is a straightforward generalization of that described in Section 2.4. We state the main steps for completeness.

Theorem A.1. *The matrix $\widetilde{\mathbf{K}}_{per}$ in (3.20) can be re-written as*

$$\begin{aligned}\widetilde{\mathbf{K}}_{per} = \big(\mathbf{V}^{-1} \otimes I_y\big)\Big(&\mathbf{D}_{1,1}^{(-1)} \otimes M + \delta t \mathbf{D}_{1,1}^{(\theta)} \otimes A - \delta t \widetilde{\mathbf{D}}_{1,2}^{(\theta)} \otimes BR^{-1}B^T \\ &+ \mathbf{D}_{2,2}^{(-1)} \otimes M^T + \delta t \mathbf{D}_{2,2}^{(\theta)} \otimes A^T + \delta t \widetilde{\mathbf{D}}_{2,1}^{(\theta)} \otimes Q\Big)\big(\mathbf{V} \otimes I_y\big),\end{aligned} \tag{A.1}$$

where $\mathbf{V}$, $\mathbf{D}_{i,i}^{(-1)}$, $\mathbf{D}_{i,i}^{(\theta)}$, $i = 1, 2$, are defined in (2.11), $I \in \mathbb{R}^{n_y \times n_y}$ is the identity matrix, and, for $\theta \in [0, 1]$,

$$\widetilde{\mathbf{D}}_{1,2}^{(\theta)} = \begin{pmatrix} 0 & \widetilde{D}_\theta^* \\ 0 & 0 \end{pmatrix}, \quad \widetilde{\mathbf{D}}_{2,1}^{(\theta)} = \begin{pmatrix} 0 & 0 \\ \widetilde{D}_\theta & 0 \end{pmatrix}, \quad \text{where } \widetilde{D}_\theta = diag(1, (1-\theta) + \theta\omega, \dots, (1-\theta) + \theta\omega^{n_t - 1}). \tag{A.2}$$

Proof. The proof is a minor modification of the proof of Theorem 2.6. □

Theorem A.1 provides an efficient way to solve the optimality system

$$\widetilde{\mathbf{K}}_{\rm per} \begin{pmatrix} Y \\ P \end{pmatrix} = \begin{pmatrix} F_Y \\ F_P \end{pmatrix}. \tag{A.3}$$

Let $\omega = \mathrm{e}^{-2\pi i/n_t}$. The k-th diagonal entries of the matrices D_{-1}, D_{-1}^*, D_θ, D_θ^*, $\widetilde{D}_\theta$, $\widetilde{D}_\theta^* \in \mathbb{C}^{n_t \times n_t}$, $\theta \in [0, 1]$, are given by (again, $\overline{z}$ denotes the complex conjugate of the complex scalar z)

$$d_{-1,k} = 1 - \omega^{k-1}, \quad \overline{\mathrm{d}}_{-1,k}, \quad d_{\theta,k} = \theta + (1-\theta)\omega^{k-1}, \quad \overline{\mathrm{d}}_{\theta,k}, \quad \widetilde{\mathrm{d}}_{\theta,k} = (1-\theta) + \theta\omega^{k-1}, \quad \overline{\widetilde{\mathrm{d}}}_{\theta,k}, \tag{A.4}$$

$k = 1, \dots, n_t$, respectively. See Theorem 2.4. Because of the structure of the matrices $\mathbf{D}_{i,i}^{(-1)}$, $\mathbf{D}_{i,i}^{(\theta)}$, $\widetilde{\mathbf{D}}_{i,j}^{(\theta)}$, $i, j = 1, 2$, defined in (2.11) and (A.2), the matrix in the middle of the right hand side in (A.1) is block diagonal (after re-ordering).

The identity in (A.1) implies that the system (A.3) can be solved in three steps:

1. Solve

$$(\mathbf{V}^{-1}\otimes I_y)\begin{pmatrix}X_y\\X_p\end{pmatrix}=\begin{pmatrix}F_Y\\F_P\end{pmatrix}.$$

2. Let

$$\begin{pmatrix}X_{y,k}\\X_{p,k}\end{pmatrix}\in\mathbb{C}^{2n_y\times 2n_y},\quad k=1,\dots,n_t,$$

   denote the blocks corresponding to time step k. Solve

$$\begin{pmatrix}d_{-1,k}M+\delta t d_{\theta,k}A & -\delta t\overline{\widetilde{\mathrm{d}}}_{\theta,k}BR^{-1}B^T\\ \delta t\widetilde{\mathrm{d}}_{\theta,k}Q & \overline{\mathrm{d}}_{-1,k}M^T+\delta t\overline{\mathrm{d}}_{\theta,k}A^T\end{pmatrix}\begin{pmatrix}Z_{y,k}\\Z_{p,k}\end{pmatrix}=\begin{pmatrix}X_{y,k}\\X_{p,k}\end{pmatrix},\quad k=1,\dots,n_t. \tag{A.5}$$

3. Solve

$$(\mathbf{V}\otimes I_y)\begin{pmatrix}Y\\P\end{pmatrix}=\begin{pmatrix}Z_y\\Z_p\end{pmatrix}.$$

Again, all three of these steps are parallelizable as described in Section 2.4. For problems of moderate size, the systems (A.5) can be solved using sparse direct solvers. This is what we do in all examples in Section 4. For larger problems these systems have to be solved iteratively. As discussed in Section 2.4, the 2-norm residual of (A.5) and the condition numbers of the matrices in (A.5) determine the solution error in the iterative solution of (A.3). If the systems (A.5) are solved iteratively, the application of the preconditioner $\widetilde{\mathbf{K}}_{\rm per}^{-1}$ is inexact. If used in connection with Krylov subspace methods such as GMRES, the theory of inexact Krylov subspace methods can be applied. See, *e.g.*, [23, 24].

If the systems (A.5) have to be solved iteratively, their represenations as equivalent Hermitian systems is useful. If $\widetilde{d}_{\theta,k}\neq 0$, the systems (A.5) in Step 2 can be rearranged as Hermitian systems

$$\begin{pmatrix}\delta tQ & \overline{\widetilde{d}}_{\theta,k}^{\,-1}\big(\overline{d}_{-1,k}M^T+\delta t\overline{d}_{\theta,k}A^T\big)\\ \widetilde{d}_{\theta,k}^{-1}\big(d_{-1,k}M+\delta t d_{\theta,k}A\big) & -\delta tBR^{-1}B^T\end{pmatrix}\begin{pmatrix}\widetilde{d}_{\theta,k}Z_{y,k}\\ \overline{\widetilde{d}}_{\theta,k}Z_{p,k}\end{pmatrix}=\begin{pmatrix}X_{p,k}\\X_{y,k}\end{pmatrix},\quad k=1,\dots,n_t. \tag{A.6}$$

Defining $Z_{u,k}=-R^{-1}B^T(\overline{\widetilde{d}}_{\theta,k}Z_{p,k})$ allows one to write the block 2×2 Hermitian systems (A.6) as block 3×3 Hermitian systems

$$\begin{pmatrix}\delta tQ & 0 & \overline{\widetilde{d}}_{\theta,k}^{\,-1}\big(\overline{d}_{-1,k}M^T+\delta t\overline{d}_{\theta,k}A^T\big)\\ 0 & \delta tR & \delta tB^T\\ \widetilde{d}_{\theta,k}^{-1}\big(d_{-1,k}M+\delta t d_{\theta,k}A\big) & \delta tB & 0\end{pmatrix}\begin{pmatrix}\widetilde{d}_{\theta,k}Z_{y,k}\\ Z_{u,k}\\ \overline{\widetilde{d}}_{\theta,k}Z_{p,k}\end{pmatrix}=\begin{pmatrix}X_{p,k}\\0\\X_{y,k}\end{pmatrix}, \tag{A.7}$$

$k=1,\dots,n_t$. Note that if $\theta=1/2$ and n_t is even, the eigenvalue $\widetilde{d}_{\theta,k}=d_{\theta,k}$ with index $k=1+n_t/2$ of C_θ shown in (2.15) is zero. Therefore, we need to require that the number of time steps n_t is odd if $\theta=1/2$ to arrange the systems (A.5) in Step 2 as (A.6) or as (A.7). See also Remark 2.7. However, no restriction on the number of time steps if $\theta=1/2$ is needed to solve the systems (A.5) in Step 2. Analogously to Theorem 2.8 we have the following result.

Theorem A.2. *Assume (1.1) holds and that the eigenvalues of $M^{-1}A \in \mathbb{R}^{n_y \times n_y}$ have strictly positive real parts. If $\theta \in [1/2, 1]$ or if $\theta \in [0, 1/2)$ and $\delta t < 2Re(\lambda)/((1-2\theta)|\lambda|^2)$ for all eigenvalues λ of $M^{-1}A$, then the systems (A.5) in Step 2 are invertible.*

Proof. From the definition (A.4) it follows that $d_{-1,k} = 0$ if and only if $k = 1$, and for $k = 1$, $d_{\theta,k} = \widetilde{d}_{\theta,k} = 1$. Thus $d_{-1,k}$ and $\widetilde{d}_{\theta,k}$ cannot both be zero.

If $\widetilde{d}_{\theta,k} = 0$, the system matrix in (A.5) is

$$\begin{pmatrix} d_{-1,k}M + \delta t d_{\theta,k} A & 0 \\ 0 & \overline{d}_{-1,k}M^T + \delta t \overline{d}_{\theta,k} A^T \end{pmatrix}$$

with $d_{-1,k} \neq 0$. If $d_{\theta,k} = 0$, then invertibility of M implies invertibility of the previous block 2×2 matrix. If $d_{\theta,k} \neq 0$, then we have shown in the proof of Theorem 2.8 that $d_{-1,k}M + \delta t d_{\theta,k} A$ is invertible. Hence the previous block 2×2 matrix is invertible.

If $\widetilde{d}_{\theta,k} \neq 0$, the system matrix in (A.5) is invertible if and only if the block 3×3 Hermitian matrix in (A.7) is invertible. The block 3×3 Hermitian matrix in (A.7) is invertible if and only if $d_{-1,k}M + \delta t d_{\theta,k} A$ is invertible, an the invertibility of $d_{-1,k}M + \delta t d_{\theta,k} A$ was shown in the proof of Theorem 2.8. □

Appendix B. Spectrum of the preconditioned matrix

This appendix contains additional results about the spectrum of the preconditioned matrix.

Consider the eigenvalue problem $\big(\mathbf{I} - \widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}\big)\mathbf{v} = \lambda\,\mathbf{v}$, or its equivalent form

$$\lambda\,\widetilde{\mathbf{K}}_{\text{per}}\mathbf{v} = \big(\widetilde{\mathbf{K}}_{\text{per}} - \mathbf{K}_{\text{init}}\big)\mathbf{v}. \tag{B.1}$$

Note that

$$\begin{aligned}\widetilde{\mathbf{K}}_{\text{per}} - \mathbf{K}_{\text{init}} = & \big(\mathbf{C}_{1,1}^{(-1)} - \mathbf{T}_{1,1}^{(-1)}\big) \otimes M + \delta t\big(\mathbf{C}_{1,1}^{(\theta)} - \mathbf{T}_{1,1}^{(\theta)}\big) \otimes A - \delta t\big(\widetilde{\mathbf{C}}_{1,2}^{(\theta)} - \widetilde{\mathbf{T}}_{1,2}^{(\theta)}\big) \otimes BR^{-1}B^T \\ & + \big(\mathbf{C}_{2,2}^{(-1)} - \mathbf{T}_{2,2}^{(-1)}\big) \otimes M^T + \delta t\big(\mathbf{C}_{2,2}^{(\theta)} - \mathbf{T}_{2,2}^{(\theta)}\big) \otimes A^T + \delta t\big(\widetilde{\mathbf{C}}_{2,1}^{(\theta)} - \widetilde{\mathbf{T}}_{2,1}^{(\theta)}\big) \otimes Q. \end{aligned} \tag{B.2}$$

Let

$$\mathbf{v} = \begin{pmatrix} Y \\ P \end{pmatrix},$$

where Y and P are given by (3.4c), but are generally complex valued. Using the definitions (3.20), (2.8), (3.19), and (3.5), as well as the identity (B.2), the detailed version of equation (B.1) is

$$\begin{aligned}
\lambda\Big(M(y_1 - y_{n_t}) + \delta t A(\theta y_1 + (1-\theta)y_{n_t}) - \delta t BR^{-1}B^T((1-\theta)p_0 + \theta p_1)\Big) &= -My_{n_t} + \delta t(1-\theta)Ay_{n_t}, \\
\lambda\Big(M(y_k - y_{k-1}) + \delta t A(\theta y_k + (1-\theta)y_{k-1}) - \delta t BR^{-1}B^T((1-\theta)p_{k-1} + \theta p_k)\Big) &= 0, \qquad k = 2, \dots, n_t - 1, \\
\lambda\Big(M(y_{n_t} - y_{n_t-1}) + \delta t A(\theta y_{n_t} + (1-\theta)y_{n_t-1}) - \delta t BR^{-1}B^T((1-\theta)p_{n_t-1} + \theta p_0)\Big) &= -\delta t\theta\, BR^{-1}B^T p_0, \\
\lambda\Big(M^T(p_0 - p_1) + \delta t A^T(\theta p_0 + (1-\theta)p_1) + \delta t Q((1-\theta)y_1 + \theta y_{n_t})\Big) &= \theta\delta t\, Qy_{n_t}, \\
\lambda\Big(M^T(p_{k-1} - p_k) + \delta t A^T(\theta p_{k-1} + (1-\theta)p_k) + \delta t Q((1-\theta)y_k + \theta y_{k-1})\Big) &= 0, \qquad k = 2, \dots, n_t - 1, \\
\lambda\Big(M^T(p_{n_t-1} - p_0) + \delta t A^T(\theta p_{n_t-1} + (1-\theta)p_0) + \delta t Q((1-\theta)y_{n_t} + \theta y_{n_t-1})\Big) &= -M^T p_0 + \delta t(1-\theta)\, A^T p_0.
\end{aligned} \tag{B.3}$$

If $\lambda = 0$, the system (B.3) collapses to

$$\big(-M + \delta t(1-\theta)A\big)y_{n_t} = \theta\delta t\, Q y_{n_t} = 0 \quad \text{and} \quad \big(-M + \delta t(1-\theta)A\big)^T p_0 = \delta t\theta\, BR^{-1}B^T p_0 = 0.$$

Thus, the eigenspace corresponding to $\lambda = 0$ contains the vectors with $y_{n_t} = p_0 = 0$ and $y_1, \dots, y_{n_t-1}$, $p_1, \dots, p_{n_t-1}$ arbitrary, which implies that $\lambda = 0$ is an eigenvalue of the $2n_t n_y \times 2n_t n_y$ matrix $\mathbf{I} - \widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ with multiplicity at least $2(n_t - 1)n_y$. This is another proof of Theorem 3.5

If we consider eigenvalues $\lambda \neq 0$ of $\mathbf{I} - \widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$, then the system (B.3) can be written in a compact form. If $\lambda \neq 0$, we define

$$y_0 = (1-\lambda^{-1})y_{n_t}, \qquad p_{n_t} = (1-\lambda^{-1})p_0, \tag{B.4a}$$

and write the system (B.3) as

$$M(y_k - y_{k-1}) + \delta t A(\theta y_k + (1-\theta)y_{k-1}) - \delta t BR^{-1}B^T((1-\theta)p_{k-1} + \theta p_k) = 0, \qquad k = 1, \dots, n_t, \tag{B.4b}$$

$$M^T(p_{k-1} - p_k) + \delta t A^T(\theta p_{k-1} + (1-\theta)p_k) + \delta t Q((1-\theta)y_k + \theta y_{k-1}) = 0, \qquad k = 1, \dots, n_t. \tag{B.4c}$$

Bounding the non-zero eigenvalues $\lambda \neq 0$ of $\mathbf{I} - \widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ *via* the analysis of (B.4) has not yet been successful. However, we can shed some light this problem by considering (B.4) for large n_t (small δt).

The system (B.4), which characterizes the nonzero eigenvalues of $\mathbf{I} - \widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ can be interpreted as a discretization of the system

$$M\frac{\mathrm{d}}{\mathrm{d}t}y(t) + A\,y(t) - B\,R^{-1}B^T p(t) = 0, \qquad t \in (0,T), \tag{B.5a}$$

$$M\,y(0) = (1-\lambda^{-1})M\,y(T), \tag{B.5b}$$

$$-M^T\frac{\mathrm{d}}{\mathrm{d}t}p(t) + A^T p(t) + Q\,y(t) = 0, \qquad t \in (0,T), \tag{B.5c}$$

$$M^T p(T) = (1-\lambda^{-1})M^T p(0). \tag{B.5d}$$

We want to find $\lambda \in \mathbb{C} \setminus \{0\}$ and functions $(y,p) \in H^1(0,T;\mathbb{C}^{n_y}) \times H^1(0,T;\mathbb{C}^{n_y})$ with $(y(0), p(T)) \neq (0,0)$, such that (B.5) holds. We expect that for large n_t (small δt) the nonzero eigenvalues of $\mathbf{I} - \widetilde{\mathbf{K}}_{\text{per}}^{-1}\mathbf{K}_{\text{init}}$ approximate the λ's obtained by finding $\lambda \neq 0$ and functions $(y,p) \in H^1(0,T;\mathbb{C}^{n_y}) \times H^1(0,T;\mathbb{C}^{n_y})$ with $(y(0), p(T)) \neq (0,0)$, such that (B.5) holds.

We will write the problem (B.5) as a matrix eigenvalue problem. This matrix eigenvalue problem will be stated in (B.10), (B.11) below, and is based on the mapping of $(y_0, p_T) \in \mathbb{C}^{n_y} \times \mathbb{C}^{n_y}$ to the solution y, p of

$$M\frac{\mathrm{d}}{\mathrm{d}t}y(t) + A\,y(t) - B\,R^{-1}B^T p(t) = 0, \qquad t \in (0,T), \tag{B.6a}$$

$$M\,y(0) = M\,y_0, \tag{B.6b}$$

$$-M^T\frac{\mathrm{d}}{\mathrm{d}t}p(t) + A^T p(t) + Q\,y(t) = 0, \qquad t \in (0,T), \tag{B.6c}$$

$$M^T p(T) = M^T p_T. \tag{B.6d}$$

Theorem B.1. *If (1.1) holds, then for every $(y_0, p_T) \in \mathbb{C}^{n_y} \times \mathbb{C}^{n_y}$ there exists a unique solution $(y,p) \in H^1(0,T;\mathbb{C}^{n_y}) \times H^1(0,T;\mathbb{C}^{n_y})$ of (B.6).*

Proof. Introduce the function u via

$$Ru(t) = -B^T p(t). \tag{B.7}$$

The solutions of (B.6a,b) and (B.6c,d) are

$$y(t) = \exp(-M^{-1}A\,t)y_0 - \int_0^t \exp(-M^{-1}A(t-\tau))\; M^{-1}B\,u(\tau)\mathrm{d}\tau, \tag{B.8a}$$

$$p(t) = \exp(-M^{-T}A^T\,(T-t))p_T - \int_t^T \exp(-M^{-T}A^T(s-t))\; M^{-T}Qy(s)\mathrm{d}s. \tag{B.8b}$$

Inserting (B.8b) and (B.8a) into (B.7) gives

$$\begin{aligned}
Ru(t) =& - B^T\exp(-M^{-T}A^T(T-t))p_T + \int_t^T B^T\exp(-M^{-T}A^T(s-t))\; M^{-T}Q\exp(-M^{-1}A\,s)y_0\mathrm{d}s \\
&- \int_t^T B^T\exp(-M^{-T}A^T(s-t))\; M^{-T}Q\int_0^s \exp(-M^{-1}A(s-\tau))\; M^{-1}B\,u(\tau)\mathrm{d}\tau\mathrm{d}s \\
=& - B^T\exp(-M^{-T}A^T(T-t))p_T + \int_t^T B^T\exp(-M^{-T}A^T(s-t))\; M^{-T}Q\exp(-M^{-1}A\,s)y_0\mathrm{d}s \\
&- \int_t^T B^T M^{-T}\exp(-A^T M^{-T}(s-t))\; Q\int_0^s \exp(-M^{-1}A(s-\tau))\; M^{-1}B\,u(\tau)\mathrm{d}\tau\mathrm{d}s.
\end{aligned}$$

Thus, given $(y_0, p_T) \in \mathbb{C}^{n_y} \times \mathbb{C}^{n_y}$, the functions $(y, p) \in H^1(0,T;\mathbb{C}^{n_y}) \times H^1(0,T;\mathbb{C}^{n_y})$ defined by (B.8) solve (B.6) if and only if $u \in L^2(0,T;\mathbb{C}^{n_u})$ solves

$$Ru(t) + (Hu)(t) = g(t) \quad \text{for almost all } t \in [0,T], \tag{B.9}$$

where H is the bounded linear operator from $L^2(0,T;\mathbb{C}^{n_u})$ to $L^2(0,T;\mathbb{C}^{n_u})$ defined by

$$u \mapsto Hu := \int_\bullet^T \int_0^s B^T M^{-T}\exp(-A^T M^{-T}(s-\bullet))\; Q\exp(-M^{-1}A(s-\tau))\; M^{-1}B\,u(\tau)\mathrm{d}\tau\mathrm{d}s$$

and $g \in L^2(0,T;\mathbb{C}^{n_u})$ is defined by

$$g(t) = -B^T\exp(-M^{-T}A^T(T-t))p_T + \int_t^T B^T\exp(-M^{-T}A^T(s-t))\; M^{-T}Q\exp(-M^{-1}A\,s)y_0\mathrm{d}s.$$

Because $\int_0^T\int_t^T \ldots.\mathrm{d}s\mathrm{d}t = \int_0^T\int_0^s \ldots.\mathrm{d}t\mathrm{d}s$ and Q is symmetric positive definite, the operator H satisfies

$$\begin{aligned}
&\int_0^T u(t)^*(Hu)(t)\mathrm{d}t \\
&= \int_0^T\int_t^T u(t)^* B^T M^{-T}\exp(-A^T M^{-T}(s-t))\; Q\int_0^s \exp(-M^{-1}A(s-\tau))\; M^{-1}B\,u(\tau)\mathrm{d}\tau\mathrm{d}s\mathrm{d}t \\
&= \int_0^T\int_0^s u(t)^* B^T M^{-T}\exp(-A^T M^{-T}(s-t))\mathrm{d}t\; Q\int_0^s \exp(-M^{-1}A(s-\tau))\; M^{-1}B\,u(\tau)\mathrm{d}\tau\mathrm{d}s \geq 0.
\end{aligned}$$

Since R is symmetric positive definite,

$$\int_0^T u(t)^*\big(Ru(t) + (Hu)(t)\big)\mathrm{d}t \geq \lambda_{min}(R)\int_0^T u(t)^* u(t)\mathrm{d}t,$$

where $\lambda_{min}(R) > 0$ is the smallest eigenvalue of R. This implies that (B.9) has a unique solution $u \in L^2(0,T;\mathbb{C}^{n_u})$ for every $g \in L^2(0,T;\mathbb{C}^{n_u})$. $\square$

Corollary B.2. *Let* (1.1) *hold. If* $\lambda \in \mathbb{C} \setminus \{0\}$ *and* $(y,p) \in H^1(0,T;\mathbb{C}^{n_y}) \times H^1(0,T;\mathbb{C}^{n_y})$ *with* $(y(0),p(T)) \neq (0,0)$ *satisfy* (B.5)*, then* $\lambda \neq 1$.

Proof. If $\lambda = 1$, the problem (B.5) is equivalent to (B.6) with $y_0 = 0$ and $p_T = 0$. The unique solution of (B.6) with $y_0 = p_T = 0$ is $y = p = 0$, which does not satisfy $(y(0),p(T)) \neq (0,0)$. $\square$

The mapping $(y_0,p_T) \mapsto (y,p)$ that maps initial/final conditions into the solution of (B.6) is linear. Given this mapping we can define the matrix $\mathbf{S} \in \mathbb{C}^{2n_y \times 2n_y}$ via

$$\mathbb{C}^{n_y} \times \mathbb{C}^{n_y} \ni (y_0,p_T) \mapsto \mathbf{S}(y_0,p_T) = (y(T),p(0)) \in \mathbb{C}^{n_y} \times \mathbb{C}^{n_y}. \tag{B.10}$$

The problem (B.5) can now be written as the following matrix eigenvalue problem: Find $\lambda \in \mathbb{C} \setminus \{0\}$ and $(y_0,p_T) \in \mathbb{C}^{n_y} \times \mathbb{C}^{n_y}$, $(y_0,p_T) \neq (0,0)$, such that

$$(1-\lambda^{-1})\,\mathbf{S}(y_0,p_T) = (y_0,p_T). \tag{B.11}$$

Recall from Corollary B.2 that $\lambda \neq 1$ and we can equivalently state the problem (B.11) as follows: If $\mu \neq 0$ is an eigenvalue of $\mathbf{S}$, with eigenvector $(y_0,p_T) \in \mathbb{C}^{n_y} \times \mathbb{C}^{n_y}$,

$$\mathbf{S}(y_0,p_T) = \mu\,(y_0,p_T), \tag{B.12}$$

then $\mu = (1-\lambda^{-1})^{-1}$ implies that

$$\lambda = (1-\mu^{-1})^{-1} \tag{B.13}$$

and $(y,p) \in H^1(0,T;\mathbb{C}^{n_y}) \times H^1(0,T;\mathbb{C}^{n_y})$ satisfy (B.5), where the functions y,p are the solutions of (B.6). Moreover, we expect that for large n_t (small δt) the nonzero eigenvalues of $\mathbf{I} - \widetilde{\mathbf{K}}_{\mathrm{per}}^{-1}\mathbf{K}_{\mathrm{init}}$ approximate the eigenvalues (B.13), and the eigenvalues not equal to one of $\widetilde{\mathbf{K}}_{\mathrm{per}}^{-1}\mathbf{K}_{\mathrm{init}}$ approximate

$$1-(1-\mu^{-1})^{-1} = \frac{1}{1-\mu}. \tag{B.14}$$

We observe this in our numerical examples. However, bounds in the eigenvalues μ of $\mathbf{S}$ and resulting bounds on (B.14), as well as the detailed connection between the non-zero eigenvalues $\lambda \neq 0$ of $\mathbf{I} - \widetilde{\mathbf{K}}_{\mathrm{per}}^{-1}\mathbf{K}_{\mathrm{init}}$ which are characterized by (B.4) and (B.11) are missing.

Acknowledgements

We thank the reviewers for their careful comments, which have led to improvements in the presentation of the paper.

Funding

This research was supported in part by grants DE-NA0003525, NSF DMS-1819144, AFOSR FA9550-22-1-0004, and by NSF Graduate Research Fellowship grant 1842494.

References

[1] S.-L. Wu and T. Zhou, Diagonalization-based parallel-in-time algorithms for parabolic PDE-constrained optimization problems. *ESAIM Control Optim. Calc. Var.* **26** (2020) Paper No. 88, 26.

[2] S.-L. Wu and J. Liu, A parallel-in-time block-circulant preconditioner for optimal control of wave equations. *SIAM J. Sci. Comput.* **42** (2020) A1510–A1540.

[3] Y. Maday and E.M. Rønquist. Parallelization in time through tensor-product space-time solvers. *C. R. Math. Acad. Sci. Paris* **346** (2008) 113–118.

[4] E. McDonald, J. Pestana and A.J. Wathen, Preconditioning and iterative solution of all-at-once systems for evolutionary partial differential equations. *SIAM J. Sci. Comput.* **40** (2018) A1012–A1033.

[5] M.J. Gander, J. Liu, S.-L. Wu, X. Yue and T. Zhou, ParaDiag: parallel-in-time algorithms based on the diagonalization technique. arXiv:2005.09158v4 92021). URL: http://arxiv.org/abs/2005.09158.

[6] R.H. Chan, Circulant preconditioners for Hermitian Toeplitz systems. *SIAM J. Matrix Anal. Appl.* **10** (1989) 542–550.

[7] R.H. Chan and G. Strang, Toeplitz equations by conjugate gradients with circulant preconditioner. *SIAM J. Sci. Statist. Comput.* **10** (1989) 104–119.

[8] A. Greenbaum, V. Pták and Z. Strakoš, Any nonincreasing convergence curve is possible for GMRES. *SIAM J. Matrix Anal. Appl.* **17** (1996) 465–469.

[9] J. Liu and Z. Wang, A ROM-accelerated parallel-in-time preconditioner for solving all-at-once systems in unsteady convection-diffusion PDEs. *Appl. Math. Comput.* **416** (2022) Paper No. 126750, 18.

[10] S.-L. Wu and T. Zhou, Acceleration of the two-level MGRIT algorithm *via* the diagonalization technique. *SIAM J. Sci. Comput.* **41** (2019) A3421–A3448.

[11] M. Stoll and T. Breiten, A low-rank in time approach to PDE-constrained optimization. *SIAM J. Sci. Comput.* **37** (2015) B1–B29.

[12] V. Barbu and N.H. Pavel, Periodic optimal control in Hilbert space. *Appl. Math. Optim.* **33** (1996) 169–188.

[13] A. Bensoussan, G. Da Prato, M.C. Delfour and S.K. Mitter, Representation and Control of Infinite Dimensional Systems. Systems & Control: Foundations & Applications, 2nd edn. Birkhäuser Boston Inc., Boston, MA (2007).

[14] M. Hinze, R. Pinnau, M. Ulbrich and S. Ulbrich, Optimization with PDE Constraints. Vol. 23 of *Mathematical Modelling, Theory and Applications*. Springer Verlag, Heidelberg, New York, Berlin (2009).

[15] J. Jahn, Introduction to the Theory of Nonlinear Optimization, 3rd edn. Springer Verlag, Berlin, Heidelberg, New York (2007).

[16] J.-L. Lions, Optimal Control of Systems Governed by Partial Differential Equations. Die Grundlehren der mathematischen Wissenschaften, Band 170. Springer-Verlag, New York-Berlin (1971).

[17] G.H. Golub and C.F. Van Loan, Matrix Computations. Johns Hopkins Studies in the Mathematical Sciences, 4th edn. Johns Hopkins University Press, Baltimore, MD (2013).

[18] M. Benzi, G.H. Golub and J. Liesen, Numerical solution of saddle point problems. *Acta Numer.* **14** (2005) 1–137.

[19] A.J. Wathen, Preconditioning. *Acta Numer.* **24** (2015) 329–376.

[20] S.S. Collis and M. Heinkenschloss, Analysis of the streamline upwind/Petrov Galerkin method applied to the solution of optimal control problems. Technical Report TR02–01, Department of Computational and Applied Mathematics, Rice University, Houston, TX (2002).

[21] M. Heinkenschloss and D. Leykekhman, Local error estimates for SUPG solutions of advection-dominated elliptic linear-quadratic optimal control problems. *SIAM J. Numer. Anal.* **47** (2010) 4607–4638.

[22] D. Leykekhman, Investigation of commutative properties of discontinuous Galerkin methods in PDE constrained optimal control problems. *J. Sci. Comput.* **53** (2012) 483–511.

[23] V. Simoncini and D.B. Szyld, Theory of inexact Krylov subspace methods and applications to scientific computing. *SIAM J. Sci. Comput.* **25** (2003) 454–477.

[24] V. Simoncini and D.B. Szyld, Recent computational developments in Krylov subspace methods for linear systems. *Numer. Linear Algebra Appl.* **14** (2007) 1–59.