SIAM J. SCI. COMPUT.
Vol. 42, No. 3, pp. A1510–A1540

A PARALLEL-IN-TIME BLOCK-CIRCULANT PRECONDITIONER FOR OPTIMAL CONTROL OF WAVE EQUATIONS*

SHU-LIN WU† AND JUN LIU‡

Abstract. In this paper, we propose a new efficient preconditioner for iteratively solving the large-scale indefinite saddle-point sparse linear system, which arises from discretizing the optimality system in optimal control problems of wave equations with a one-shot second-order finite difference scheme in both space and time. The proposed preconditioner can be implemented in a parallel-in-time (PinT) manner via a carefully designed unitary diagonalization decomposition. Such an explicit unitary diagonalization is rarely seen in the literature. We also analyze the eigenvalue bounds of the preconditioned system, which are shown to be highly clustered around one. Moreover, a simple splitting algorithm that alternates between a linear complementarity problem (LCP) and a quasi-Newton iteration is discussed for handling the case with control constraints. Within the quasi-Newton iteration, our proposed PinT preconditioner can be directly used in preconditioning the Jacobian system of the same structure. Both 1D and 2D numerical examples are given to illustrate the promising convergence performance of our proposed PinT preconditioner in comparison with a recently proposed matching Schur complement (MSC) preconditioner.

Key words. optimal control, wave equation, circulant preconditioner, parallel-in-time (PinT), diagonalization, GMRES

AMS subject classifications. 65F08, 65F10, 65Y05

DOI. 10.1137/19M1289613

1. Introduction.

The development of fast and robust preconditioners in Krylov subspace solvers (e.g., GMRES) for solving optimization problems constrained by time-dependent PDEs is a very active topic [6, 12, 29]. However, in the last few decades, most of the related research in this direction has been concerned with the optimal control problems with parabolic and elliptic PDEs; see, e.g., [1, 2, 42, 43, 44, 45, 50]. Much less work has been devoted to the optimal control problems with (hyperbolic) wave equations; see [13, 18, 31, 32, 33, 34, 35, 37, 49, 56] and the references therein. In this paper, we propose and analyze a new parallel-in-time (PinT) preconditioner for solving such hyperbolic optimal control problems, which shows fast convergence in numerical experiments.

Let $\Omega \in \mathbb{R}^d$ with $d \geq 1$ be a bounded and open domain with Lipschitz boundary, and let $[0, T]$ be the time window of interest with $T > 0$. We consider the following distributed optimal control problem [36] of minimizing a tracking-type quadratic cost functional:

$$\min_{y,u} \qquad \mathcal{L}(y,u) := \frac{1}{2}\|y-g\|^2_{L^2(\Omega\times(0,T))} + \frac{\gamma}{2}\|u\|^2_{L^2(\Omega\times(0,T))}, \tag{1.1a}$$

*Submitted to the journal's Methods and Algorithms for Scientific Computing section September 25, 2019; accepted for publication (in revised form) March 2, 2020; published electronically May 6, 2020.

https://doi.org/10.1137/19M1289613

Funding: The work of the first author was supported by the NSF of China through grants 11771313 and 11671074, by the NSF of Sichuan Province through grant 2018JY0469, and by the Science Challenge Project through grant TZ2016002.

†School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, China (wushulin84@hotmail.com).

‡Corresponding author. Department of Mathematics and Statistics, Southern Illinois University Edwardsville, Edwardsville, IL 62026 (juliu@siue.edu).

subject to a linear wave equation with initial- and boundary-value conditions:

$$\text{(1.1b)} \qquad \begin{cases} y_{tt} - \Delta y = f + u & \text{in } \Omega \times (0,T), \quad y = 0, \text{ on } \partial\Omega \times (0,T), \\ y(\cdot,0) = y_0, \quad y_t(\cdot,0) = y_1 & \text{in } \Omega, \end{cases}$$

where $u \in L^2$ is the distributed control, $g \in L^2$ is the desired tracking trajectory or observation data, $\gamma > 0$ denotes the cost weight or regularization parameter, and f, y_0, and y_1 are the given functions. The existence, uniqueness, and regularity of the solution for problem (1.1a)–(1.1b) are well established [36] under suitable assumptions on the given data. Attributing to the strict convexity of the objective functional $\mathcal{L}$ and the linearity of the state equation, the optimal solution of (1.1a)–(1.1b) is characterized by the following first-order necessary (and also sufficient) optimality system:

$$\text{(1.2)} \qquad \begin{cases} y_{tt} - \Delta y - \frac{1}{\gamma} p = f & \text{in } \Omega \times (0,T), \qquad y = 0, \text{ on } \partial\Omega \times (0,T), \\ y(\cdot,0) = y_0, \quad y_t(\cdot,0) = y_1 & \text{in } \Omega, \\ p_{tt} - \Delta p + y = g & \text{in } \Omega \times (0,T), \qquad p = 0, \text{ on } \partial\Omega \times (0,T), \\ p(\cdot,T) = 0, \quad p_t(\cdot,T) = 0 & \text{in } \Omega, \end{cases}$$

where we have eliminated the control variable u from the optimality condition $\gamma u - p = 0$ in (1.2), leading to a reduced optimality system regarding only y and p. Notice that the state variable y evolves forward and the adjoint state p marches backward in time, which implies that we may have to solve all the time steps all at once (instead of time marching) in a one-shot discretization scheme. Such a one-shot scheme inevitably leads to a large-scale indefinite sparse linear system, whose approximation solution requires efficient iterative solvers with fast and robust convergence rates.

For iteratively solving such large-scale discretized linear systems by the Krylov subspace methods, the design of an effective and efficient preconditioner plays a critical role in reducing the overall computational cost. In particular, the development of a preconditioner highly depends on the underlying discretization schemes in space and time. Several discretization schemes for the optimal control of wave or hyperbolic PDEs are studied in the literature [13, 18, 31, 49], but there are fewer works [26, 35, 37] about the development of fast preconditioners. Numerical computation of such a resulting indefinite optimality system from hyperbolic PDEs is especially difficult due to the underlying Helmholtz-type operators with indefinite spectrum. On one hand, the constraint preconditioner in [35] shows a mesh-independent convergence rate, but its convergence rate deteriorates as the regularization parameter γ becomes smaller. On the other hand, the matching-type Schur complement (MSC) preconditioner in [37] has a parameter-robust spatial mesh-independent convergence rate, which, however, mildly depends on the temporal mesh step size. In particular, it was mentioned in [37] that the independence on the regularization parameter and both spatial and temporal mesh step sizes seems to be impossible to achieve simultaneously. We strive to tackle this seemingly impossible task and to at least get closer to partially fulfilling the desired goal.

With the popularity of massively parallel processors, it becomes increasingly more important to develop efficiently parallelizable numerical algorithms, such as domain decomposition algorithms and PinT algorithms, for solving time-dependent PDEs. In particular, the above-mentioned MSC preconditioner is not easily parallelized. We point out that the traditional time-marching schemes are intrinsically sequential,

which should be treated in a one-shot manner for better parallelization. Several all-at-once (or one-shot) schemes were studied in [41], where the resulting block-Toeplitz matrices are preconditioned by the corresponding Strang block-circulant matrices. The numerical results reported in [41] indicate very clear mesh-independent convergence rates. Some preliminary parallel numerical results given in [19] also show that such PinT diagonalization techniques indeed bring considerable speedup in computation time. Several block-circulant preconditioners based on different finite difference schemes in time were also developed in [19] for solving the wave equation. Such a time parallel diagonalization technique was first introduced in [40], and some novel modifications can be found in [53, 55], where the authors used this technique to improve the speedup of parareal and multigrid reduction in time (MGRIT) algorithms via a diagonalization-based coarse grid correction. Other PinT algorithms based on the diagonalization technique can be found in [3, 38, 39]. In these works, the authors studied the time-dependent PDEs—not the optimal control of time-dependent PDEs. We refer the interested readers to [15] for a recent survey of the PinT algorithms[1] and the books [7, 30] for general discussion on block-circulant preconditioners for block-Toeplitz matrices.

Inspired by the idea in [19, 41] for solving time-dependent PDEs in a one-shot manner, we propose a new PinT block-circulant preconditioner $\mathcal{P}$ for treating the optimality PDE system (1.2) discretized by a well-established all-at-once scheme. Significantly different from the cases discussed in [19, 41], we have to solve a two-by-two block matrix $\mathcal{M}$ upon full discretization, which becomes more difficult to diagonalize due to the coupled system. By exploiting the diagonal structure of the two-by-two block matrix $\mathcal{P}$, we are able to find explicit formulas for its diagonalization decomposition. More importantly, as one major contribution, we prove that the derived diagonalization decomposition is optimal in the sense that the eigenvector matrix is indeed unitary upon a scalar multiplication. We remark that the condition number of the eigenvector matrix has a serious influence on the roundoff error of the diagonalization procedure [16] and a unitary eigenvector matrix results in the minimal roundoff error. For a general nonsymmetric matrix, it is challenging, if not impossible, to achieve such minimal roundoff errors for its diagonalization decomposition. We attribute this perfect unitary diagonalization to our used implicit finite difference scheme in time that indeed admits a symmetric system reformulation. For instance, those one-shot schemes discussed in [16] do not provide a unitary diagonalization, which implies their nonunitary diagonalizations lead to larger roundoff errors.

By making use of the obtained unitary diagonalization decomposition and the fast Fourier transform (FFT) for factorizing the involved circulant matrices, for any input vector r, we can compute the preconditioning step $\mathcal{P}^{-1}r$ very efficiently via the three-step diagonalization technique [40], which yields a direct parallel implementation across all time points. To support the observed very satisfactory numerical results, we show that the preconditioned matrix has a very clustered spectrum, which indicates a robust fast convergence rate of the preconditioned GMRES method. As an important application, we also discussed the optimal control problems with control constraint, for which we formulate the optimality system in the framework of a dynamic complementarity system [23]. Within this framework, we can directly reuse $\mathcal{P}$ or its modification as an effective preconditioner for the chosen approximate Jacobian matrix of the same structure.

[1] We also recommend the website http://parallel-in-time.org for further PinT algorithms and applications.

The rest of this paper is organized as follows. In section 2, we introduce the used finite difference discretization of the optimality system and propose our PinT preconditioner with necessary details regarding its parallel implementation based on a special unitary diagonalization decomposition. The estimated roundoff error due to the direct diagonalization technique is also discussed. In section 3, we estimate the spectrum of the preconditioned matrix. We then show in section 4 an application of the PinT preconditioner for the optimal control problem with boxed control constraint. Numerical results are presented in section 5, and some concluding remarks are given in section 6. Throughout this paper, we denote by $^\mathsf{T}$ and * the (real) transpose and complex conjugate transpose of a matrix (vector), respectively. For any two vectors $v = (v_1, \ldots, v_m)^\mathsf{T}$ and $w = (w_1, \ldots, w_m)^\mathsf{T}$, the inequalities $0 \le v$ and $w \ge 0$ should be understood componentwise. The complementarity operator $\perp$ is also defined componentwise, i.e., $v \perp w$ if and only if $v_j w_j = 0$ for all $j = 1, 2, \ldots, m$. For brevity, we will also use the conventional notation $0 \le v \perp w \ge 0$, which is equivalent to $0 \le v$, $w \ge 0$, and $v \perp w$. For any diagonal matrix $D = \mathrm{diag}(d_1, d_2, \ldots, d_m)$, its square root $\sqrt{D}$ is also defined componentwise, i.e., $\sqrt{D} = \mathrm{diag}(\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_m})$.

2. A new PinT preconditioner. Given two positive integers N_x and N_t, we define the mesh step sizes $\tau = T/N_t$ and $h = 1/(N_x + 1)$. We partition the time interval $[0, T]$ uniformly by time points $\{t_n\}_{n=0}^{N_t}$ with $t_n = n\tau$. Let Δ_h be the second-order accurate discrete matrix approximation of the Laplacian operator Δ in (1.2) obtained by using the finite difference method (e.g., central difference) or the finite element method (e.g., P_1 element) with the given zero Dirichlet boundary condition. In particular, we will focus on studying an implicit leap-frog finite difference scheme established in [35], which fully discretizes (1.2) according to

$$\begin{aligned} &\frac{Y_{n+1} - 2Y_n + Y_{n-1}}{\tau^2} - \Delta_h \frac{Y_{n+1} + Y_{n-1}}{2} - \frac{1}{\gamma} P_n = F_n, \quad n = 1, 2, \ldots, N_t - 1, \\ &\frac{P_{n+1} - 2P_n + P_{n-1}}{\tau^2} - \Delta_h \frac{P_{n+1} + P_{n-1}}{2} + Y_n = G_n, \quad n = 1, 2, \ldots, N_t - 1, \end{aligned} \tag{2.1a}$$

where Y_n and P_n are lexicographic ordered vectors collecting the approximate solutions of $y(\cdot, t_n)$ and $p(\cdot, t_n)$ over all the space grids. Similar notations also apply to F_n and G_n. The initial conditions are approximated by using Taylor expansions together with the governing equations in (1.2):

$$\begin{aligned} &Y_0 = \mathrm{vec}(y_0), \qquad \left(1 - \frac{\tau^2 \Delta_h}{2}\right) Y_1 = \mathrm{vec}(y_0) + \tau \mathrm{vec}(y_1) + \frac{\tau^2}{2} \left(F_0 + \frac{1}{\gamma} P_0\right), \\ &P_{N_t} = \mathrm{vec}(0), \qquad \left(1 - \frac{\tau^2 \Delta_h}{2}\right) P_{N_t - 1} = \frac{\tau^2}{2} (-Y_{N_t} + G_{N_t}), \end{aligned} \tag{2.1b}$$

where $\mathrm{vec}(\cdot)$ denotes the lexicographic ordering vectorization of the corresponding function values over all the space grids. Using the Kronecker product notations, the above scheme can be formulated into a sparse linear system

$$\widehat{\mathcal{M}} \begin{bmatrix} y_h \\ p_h \end{bmatrix} := \left(\begin{bmatrix} B_1 & -\frac{\tau^2 \hat{I}_t}{\gamma} \\ \tau^2 \check{I}_t & B_1^\mathsf{T} \end{bmatrix} \otimes I_x - \frac{\tau^2}{2} \begin{bmatrix} B_2 & \\ & B_2^\mathsf{T} \end{bmatrix} \otimes \Delta_h \right) \begin{bmatrix} y_h \\ p_h \end{bmatrix} = \begin{bmatrix} f_h \\ g_h \end{bmatrix}, \tag{2.1c}$$

where $\hat{I}_t = \mathrm{diag}(\frac{1}{2}, 1, \ldots, 1)$, $\check{I}_t = \mathrm{diag}(1, \ldots, 1, \frac{1}{2}) \in \mathbb{R}^{N_t \times N_t}$, $I_x \in \mathbb{R}^{N_x \times N_x}$ is an identity matrix,

$$f_h = \tau^2 \begin{bmatrix} \frac{1}{2}F_0 + Y_1/\tau + Y_0/\tau^2 \\ F_1 - (I_x/\tau^2 - \frac{1}{2}\Delta_h)Y_0 \\ F_2 \\ \vdots \\ F_{N_t-1} \end{bmatrix}, \; g_h = \tau^2 \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_{N_t-1} \\ \frac{1}{2}G_{N_t} \end{bmatrix}, y_h = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{N_t-1} \\ Y_{N_t} \end{bmatrix}, p_h = \begin{bmatrix} P_0 \\ P_1 \\ \vdots \\ P_{N_t-2} \\ P_{N_t-1} \end{bmatrix},$$

and

$$B_1 = \begin{bmatrix} 1 & & & & \\ -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{bmatrix}, B_2 = \begin{bmatrix} 1 & & & & \\ 0 & 1 & & & \\ 1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 0 & 1 \end{bmatrix}. \tag{2.2}$$

Notice that $B_1, B_2 \in \mathbb{R}^{N_t \times N_t}$ are Toeplitz matrices representing the time discretization scheme. By letting $\tilde{y}_h = \sqrt{\gamma} y_h$ and $\tilde{f}_h = \sqrt{\gamma} f_h$, we can diagonally rescale the above system (2.1c) to get

$$\mathcal{M} \begin{bmatrix} \tilde{y}_h \\ p_h \end{bmatrix} := \left[\left(\begin{bmatrix} \sqrt{\gamma} I_t & \\ & I_t \end{bmatrix} \otimes I_x \right) \widehat{\mathcal{M}} \left(\begin{bmatrix} \frac{I_t}{\sqrt{\gamma}} & \\ & I_t \end{bmatrix} \otimes I_x \right) \right] \begin{bmatrix} \tilde{y}_h \\ p_h \end{bmatrix} = \begin{bmatrix} \tilde{f}_h \\ g_h \end{bmatrix},$$

with

$$\begin{aligned} \mathcal{M} &= \begin{bmatrix} B_1 & -\frac{\tau^2 \hat{I}_t}{\sqrt{\gamma}} \\ \frac{\tau^2 \check{I}_t}{\sqrt{\gamma}} & B_1^{\mathsf{T}} \end{bmatrix} \otimes I_x - \frac{\tau^2}{2} \begin{bmatrix} B_2 & \\ & B_2^{\mathsf{T}} \end{bmatrix} \otimes \Delta_h \\ &= \begin{bmatrix} B_1 \otimes I_x - \frac{\tau^2}{2} B_2 \otimes \Delta_h & -\frac{\tau^2 \hat{I}_t}{\sqrt{\gamma}} \\ \frac{\tau^2 \check{I}_t}{\sqrt{\gamma}} & B_1^{\mathsf{T}} \otimes I_x - \frac{\tau^2}{2} B_2^{\mathsf{T}} \otimes \Delta_h \end{bmatrix}. \end{aligned} \tag{2.3}$$

Based on the above discussion, we propose the following block-circulant preconditioner:

$$\mathcal{P} := \begin{bmatrix} C_1 & -\frac{\tau^2 I_t}{\sqrt{\gamma}} \\ \frac{\tau^2 I_t}{\sqrt{\gamma}} & C_1^{\mathsf{T}} \end{bmatrix} \otimes I_x - \frac{\tau^2}{2} \begin{bmatrix} C_2 & \\ & C_2^{\mathsf{T}} \end{bmatrix} \otimes \Delta_h, \tag{2.4}$$

where the Toeplitz matrices B_1 and B_2 within $\mathcal{M}$ are replaced by the Strang circulant matrices C_1 and C_2:

$$C_1 = \begin{bmatrix} 1 & & & 1 & -2 \\ -2 & 1 & & & 1 \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{bmatrix}, \; C_2 = \begin{bmatrix} 1 & & & 1 & 0 \\ 0 & 1 & & & 1 \\ 1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 0 & 1 \end{bmatrix}. \tag{2.5}$$

and the diagonal matrices $\hat{I}_t$ and $\check{I}_t$ are replaced by the identity matrix $I_t \in \mathbb{R}^{N_t \times N_t}$. We remark that replacing $\check{I}_t$ and $\hat{I}_t$ by the identity matrix I_t is necessary, since otherwise we cannot easily explicitly diagonalize $\mathcal{P}$ as needed in the efficient PinT implementation. As to be further explained below, there are two advantages of preconditioning $\mathcal{M}$ by $\mathcal{P}$: (1) the PinT implementation of $\mathcal{P}^{-1}r$ (with r being an input vector); and (2) the highly clustering of the complex eigenvalues of the nonsymmetric preconditioned matrix $\mathcal{P}^{-1}\mathcal{M}$.

2.1. PinT implementation with diagonalization. We first explain the PinT implementation details of computing the preconditioning step $\mathcal{P}^{-1}r$. The starting point is the spectral diagonalization of the circulant matrices $C_{1,2}$: according to [5] the matrices $C_{1,2}$ can be diagonalized as $C_{1,2} = \mathbb{F}^*\Lambda_{1,2}\mathbb{F}$, where $\mathbb{F}$ is the discrete Fourier matrix defined by

$$\mathbb{F} = \frac{1}{\sqrt{N_t}}\left[\omega^{(j-1)(k-1)}\right]_{j,k=1}^{N_t}, \quad \omega = e^{\frac{2\pi \mathrm{i}}{N_t}}, \quad \mathrm{i} := \sqrt{-1}, \tag{2.6}$$

and $\Lambda_{1,2} = \mathrm{diag}\,(\mathbb{F}C_{1,2}(:,1))$ with $C_{1,2}(:,1)$ being the first column of $C_{1,2}$. Hence, the nth eigenvalue of C_2 is $1+\omega^{2(n-1)} = 1 + e^{\frac{4(n-1)\pi \mathrm{i}}{N_t}}$ and C_2 is invertible if the integer N_t is not a multiple of 4. If we are assuming C_2 is invertible, we have

$$\mathcal{P} = \underbrace{\left(\begin{bmatrix} C_1C_2^{-1} & -\frac{\tau^2(C_2^{-1})^{\mathsf{T}}}{\sqrt{\gamma}} \\ \frac{\tau^2 C_2^{-1}}{\sqrt{\gamma}} & C_1^{\mathsf{T}}(C_2^{-1})^{\mathsf{T}} \end{bmatrix} \otimes I_x - \frac{\tau^2}{2}\begin{bmatrix} I_t & \\ & I_t \end{bmatrix} \otimes \Delta_h\right)}_{=:\widetilde{\mathcal{P}}} \left(\begin{bmatrix} C_2 & \\ & C_2^{\mathsf{T}} \end{bmatrix} \otimes I_x\right). \tag{2.7}$$

Now, for any input vector r, we can compute $s = \mathcal{P}^{-1}r$ via

$$\eta := \begin{bmatrix}\eta_1 \\ \eta_2\end{bmatrix} = \widetilde{\mathcal{P}}^{-1}r, \ s = \begin{bmatrix} (C_2^{-1}\otimes I_x)\eta_1 \\ ((C_2^{-1})^{\mathsf{T}}\otimes I_x)\eta_2 \end{bmatrix}. \tag{2.8}$$

Once η is calculated, we can compute s with high efficiency by FFT. More specifically, recall that $\mathrm{vec}(Z)$ denotes the vectorization of the matrix Z formed by stacking its columns into a single column vector, and write $\eta_{1,2} = \mathrm{vec}(Z_{1,2})$. Then, by the Kronecker product property [20, p. 711], we have

$$(C_2^{-1}\otimes I_x)\eta_1 = \mathrm{vec}(Z_1(C_2^{-1})^{\mathsf{T}}) = \mathrm{vec}((C_2^{-1}Z_1^{\mathsf{T}})^{\mathsf{T}}) = \mathrm{vec}((\mathbb{F}^*\Lambda_2^{-1}\mathbb{F}Z_1^{\mathsf{T}})^{\mathsf{T}})$$

and

$$((C_2^{-1})^{\mathsf{T}}\otimes I_x)\eta_2 = \mathrm{vec}(Z_2C_2^{-1}) = \mathrm{vec}(Z_2\mathbb{F}^*\Lambda_2^{-1}\mathbb{F}),$$

where the multiplication of $\mathbb{F}$ and $\mathbb{F}^*$ can be computed via FFT and inverse FFT, respectively. Hence, the major computation lies in the first part, i.e., $\eta = \widetilde{\mathcal{P}}^{-1}r$. We now derive a special diagonalization of the matrix $\widetilde{\mathcal{P}}$ in (2.7). Using the mixed-product properties of Kronecker product, we can factorize $\widetilde{\mathcal{P}}$ as

$$\widetilde{\mathcal{P}} = \left(\begin{bmatrix}\mathbb{F}^* & \\ & \mathbb{F}^*\end{bmatrix}\otimes I_x\right)\left(\underbrace{\begin{bmatrix} \Lambda_1\Lambda_2^{-1} & -\frac{\tau^2(\Lambda_2^*)^{-1}}{\sqrt{\gamma}} \\ \frac{\tau^2\Lambda_2^{-1}}{\sqrt{\gamma}} & \Lambda_1^*(\Lambda_2^*)^{-1}\end{bmatrix}}_{=:\Lambda}\otimes I_x - \frac{\tau^2}{2}\begin{bmatrix} I_t & \\ & I_t\end{bmatrix}\otimes\Delta_h\right)\left(\begin{bmatrix}\mathbb{F} & \\ & \mathbb{F}\end{bmatrix}\otimes I_x\right), \tag{2.9}$$

where we have used two obvious facts that $C_{1,2}^{\mathsf{T}} = C_{1,2}^* = \mathbb{F}^*\Lambda_{1,2}^*\mathbb{F}$ and $\mathbb{F}^*\mathbb{F} = I_t = \mathbb{F}\mathbb{F}^*$. First, we note that the nth diagonal entries of Λ_1 and Λ_2 are explicitly given by

$$(\Lambda_1)_{n,n} = 1 - 2e^{\frac{2(n-1)\pi \mathrm{i}}{N_t}} + e^{\frac{4(n-1)\pi \mathrm{i}}{N_t}}, \qquad (\Lambda_2)_{n,n} = 1 + e^{\frac{4(n-1)\pi \mathrm{i}}{N_t}}, \qquad n = 1,\ldots,N_t,$$

respectively, which leads to (applying Euler's formula $e^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta$)

$$\begin{aligned}(\Lambda_1\Lambda_2^{-1})_{n,n} &= \frac{1 - 2e^{\frac{2(n-1)\pi \mathrm{i}}{N_t}} + e^{\frac{4(n-1)\pi \mathrm{i}}{N_t}}}{1 + e^{\frac{4(n-1)\pi \mathrm{i}}{N_t}}} \\ &= 1 - \frac{2}{e^{-\frac{2(n-1)\pi \mathrm{i}}{N_t}} + e^{\frac{2(n-1)\pi \mathrm{i}}{N_t}}} = 1 - \frac{1}{\cos\left(\frac{2(n-1)\pi}{N_t}\right)} \in \mathbb{R}.\end{aligned}$$

This implies the diagonal matrix $\Lambda_1\Lambda_2^{-1}$ is in fact real, and hence $\Lambda_1^*(\Lambda_2^*)^{-1} = (\Lambda_1\Lambda_2^{-1})^* = \Lambda_1\Lambda_2^{-1}$.

Second, since Λ_1 and Λ_2 are diagonal matrices with complex entries, the diagonalization decomposition of the key two-by-two block matrix Λ is very similar to the diagonalization of a two-by-two matrix. If Λ_2^{-1} exists, a routine calculation (see Remark 2.1 below) yields the following diagonalization decomposition of Λ:

$$\Lambda = \begin{bmatrix} \Lambda_1\Lambda_2^{-1} & -\frac{\tau^2(\Lambda_2^*)^{-1}}{\sqrt{\gamma}} \\ \frac{\tau^2\Lambda_2^{-1}}{\sqrt{\gamma}} & \Lambda_1\Lambda_2^{-1} \end{bmatrix} = \begin{bmatrix} I_t & S_2 \\ S_1 & I_t \end{bmatrix} \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \begin{bmatrix} I_t & S_2 \\ S_1 & I_t \end{bmatrix}^{-1} =: S\Sigma S^{-1}, \tag{2.10}$$

where $S_1 = \sqrt{-\Lambda_2^*\Lambda_2^{-1}}, S_2 = -\sqrt{-(\Lambda_2^*)^{-1}\Lambda_2}, \Sigma_1 = \Lambda_1\Lambda_2^{-1} + \Gamma,\ \Sigma_2 = \Lambda_1\Lambda_2^{-1} - \Gamma$ with $\Gamma = \mathrm{i}\frac{\tau^2}{\sqrt{\gamma}}|\Lambda_2^{-1}|$. We emphasize that such a spectral decomposition (2.10) of Λ is not unique, and its applicability highly depends on the condition number of the eigenvector matrix S and the operation cost of computing $S^{-1}v$. The next theorem shows that the proposed factorization (2.10) of the matrix Λ is *optimal* in the sense that $\mathrm{Cond}_2(S) = 1$. This nice property only depends on the used finite difference discretization scheme in time.

THEOREM 2.1. *Letting* $S = \left[\begin{smallmatrix} I_t & S_2 \\ S_1 & I_t \end{smallmatrix}\right]$ *with* S_1 *and* S_2 *being the matrices given by* (2.10)*, it holds that* $\frac{1}{\sqrt{2}}S$ *is a unitary matrix and thus* $\mathrm{Cond}_2(S) = 1$.

Proof. Given $S_1 = \sqrt{-\Lambda_2^*\Lambda_2^{-1}}$ and $S_2 = -\sqrt{-(\Lambda_2^*)^{-1}\Lambda_2}$ as in (2.10), we clearly have

$$S_1 + S_2^* = \sqrt{-\Lambda_2^*\Lambda_2^{-1}} - \sqrt{-\Lambda_2^{-1}\Lambda_2^*} = 0, \qquad S_1^* + S_2 = (S_1 + S_2^*)^* = 0,$$

and

$$S_1S_1^* = \sqrt{\Lambda_2^*\Lambda_2^{-1}\Lambda_2(\Lambda_2^*)^{-1}} = \sqrt{I_t} = I_t, \qquad S_2S_2^* = \sqrt{(\Lambda_2^*)^{-1}\Lambda_2\Lambda_2^{-1}\Lambda_2^*} = \sqrt{I_t} = I_t.$$

Hence,

$$\begin{aligned} SS^* &= \begin{bmatrix} I_t & S_2 \\ S_1 & I_t \end{bmatrix} \begin{bmatrix} I_t & S_1^* \\ S_2^* & I_t \end{bmatrix} = \begin{bmatrix} I_t + S_2S_2^* & S_1^* + S_2 \\ S_1 + S_2^* & I_t + S_1S_1^* \end{bmatrix} \\ &= \begin{bmatrix} I_t + S_2S_2^* & 0 \\ 0 & I_t + S_1S_1^* \end{bmatrix} = 2\begin{bmatrix} I_t & 0 \\ 0 & I_t \end{bmatrix}, \end{aligned}$$

which implies both $\frac{1}{\sqrt{2}}S$ and $\frac{1}{\sqrt{2}}S^*$ are unitary and hence $S^{-1} = \frac{1}{2}S^*$. This also shows that

$$\mathrm{Cond}_2(S) := \|S\|_2\|S^{-1}\|_2 = \sqrt{2}\left\|\frac{1}{\sqrt{2}}S\right\|_2\frac{1}{\sqrt{2}}\left\|\frac{1}{\sqrt{2}}S^*\right\|_2 = \left\|\frac{1}{\sqrt{2}}S\right\|_2\left\|\frac{1}{\sqrt{2}}S^*\right\|_2 = 1,$$

where we have used the fact that $\|\frac{1}{\sqrt{2}}S\|_2 = \|\frac{1}{\sqrt{2}}S^*\|_2 = 1$. ∎

Let $V = \left[\begin{smallmatrix} \mathbb{F}^* & \\ & \mathbb{F}^* \end{smallmatrix}\right]S$. From (2.9) and (2.10) we have

$$\widetilde{\mathcal{P}} = (V \otimes I_x)\left(\begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \otimes I_x - \frac{\tau^2}{2}\begin{bmatrix} I_t & \\ & I_t \end{bmatrix} \otimes \Delta_h\right)(V^{-1} \otimes I_x). \tag{2.11}$$

Since $\mathbb{F}$ is unitary and $S^{-1} = \frac{1}{2}S^*$, there holds

$$V^{-1} = \left(\begin{bmatrix} \mathbb{F}^* & \\ & \mathbb{F}^* \end{bmatrix} S\right)^{-1} = S^{-1}\begin{bmatrix} \mathbb{F} & \\ & \mathbb{F} \end{bmatrix} = \frac{1}{2}S^*\begin{bmatrix} \mathbb{F} & \\ & \mathbb{F} \end{bmatrix} = \frac{1}{2}\left(\begin{bmatrix} \mathbb{F}^* & \\ & \mathbb{F}^* \end{bmatrix} S\right)^* = \frac{1}{2}V^*.$$

With the decomposition (2.11), the vector $\eta = \widetilde{\mathcal{P}}^{-1}r$ in (2.8) can be computed via the following three steps:

$$\begin{aligned} &\text{Step (a)} && g = (V^{-1}\otimes I_x)r = \left(\frac{1}{2}V^*\otimes I_x\right)r, \\ &\text{Step (b)} && \left(\sigma_n I_x - \frac{\tau^2}{2}\Delta_h\right)w_n = g_n, \quad n = 1,2,\ldots,2N_t, \\ &\text{Step (c)} && \eta = (V\otimes I_x)w, \end{aligned} \tag{2.12}$$

where $g = [g_1^{\mathsf{T}},\ldots,g_{2N_t}^{\mathsf{T}}]^{\mathsf{T}}$, $w = [w_1^{\mathsf{T}},\ldots,w_{2N_t}^{\mathsf{T}}]^{\mathsf{T}}$, and σ_n denotes the nth diagonal entry of $\Sigma = [\begin{smallmatrix} \Sigma_1 & \\ & \Sigma_2 \end{smallmatrix}]$. Step (a) and Step (c) can be implemented efficiently (via FFT) by utilizing the special structure of the matrix V. Step (b) consists of $2N_t$ independent shifted linear systems with (complex) symmetric coefficient matrices, which can be solved in a highly parallel manner. In particular, each system can be approximately solved by the fast multigrid method [54] or the domain decomposition method. Fast iterative solvers for Step (b) deserve further investigation. The three-step procedure in (2.12) is the so-called *diagonalization* technique [16, 17, 19, 40, 41, 55]. This paper contributes to successfully extending such a diagonalization technique for solving a single PDE to handle the optimality PDE system arising from the optimal control of wave equations.

Remark 2.1. Let $E_{1,2,3} \in \mathbb{C}^{N_t\times N_t}$ be diagonal matrices and $E = [\begin{smallmatrix} E_1 & E_2 \\ E_3 & E_1 \end{smallmatrix}]$. If E_2 and E_3 are invertible, we can factorize E as

$$E = S\begin{bmatrix} E_1 + E_2S_1 & \\ & E_1 + E_3S_2 \end{bmatrix}S^{-1}, \quad \text{with } S = \begin{bmatrix} I_t & S_2 \\ S_1 & I_t \end{bmatrix},$$

where $S_1 = \sqrt{E_2^{-1}E_3}$, $S_2 = -\sqrt{E_2E_3^{-1}}$. To verify this factorization, it is sufficient to show

$$\begin{aligned} &\begin{bmatrix} E_1 + E_2S_1 & E_2 + E_1S_2 \\ E_3 + E_1S_1 & E_1 + E_3S_2 \end{bmatrix} = ES \\ &= S\begin{bmatrix} E_1 + E_2S_1 & \\ & E_1 + E_3S_2 \end{bmatrix} = \begin{bmatrix} E_1 + E_2S_1 & E_1S_2 + E_3S_2^2 \\ E_1S_1 + E_2S_1^2 & E_1 + E_3S_2 \end{bmatrix}, \end{aligned}$$

which is true since $E_2S_1^2 = E_3$ and $E_3S_2^2 = E_2$ by the choice of S_1 and S_2. Moreover, S is invertible since

$$S = \begin{bmatrix} I_t & S_2 \\ S_1 & I_t \end{bmatrix} = \begin{bmatrix} I_t & -\sqrt{E_2E_3^{-1}} \\ \sqrt{E_2^{-1}E_3} & I_t \end{bmatrix} = \begin{bmatrix} I_t & 0 \\ \sqrt{E_2^{-1}E_3} & I_t \end{bmatrix}\begin{bmatrix} I_t & -\sqrt{E_2E_3^{-1}} \\ 0 & 2I_t \end{bmatrix},$$

where we used the fact that $E_2^{-1}E_3E_2E_3^{-1} = E_2^{-1}E_2E_3E_3^{-1} = I_t$ as E_2 and E_3 are diagonal matrices.

2.2. Estimation of the roundoff error due to the diagonalization of $\widetilde{\mathcal{P}}$. For any input vector r, let η be the exact solution of $\widetilde{\mathcal{P}}\eta = r$ and $\widetilde{\eta}$ be the approximate numerical solution obtained via the diagonalization procedure (2.12). Then, the roundoff error arising from these three steps will cause difference between η and $\widetilde{\eta}$, which can be characterized by the so-called *relative error*. There exists a permutation matrix Π (depending only on dimensions) such that $\Pi(A\otimes B)\Pi^{\mathsf{T}} = (B\otimes A)$ holds for any two square matrices A and B [25]. Hence, the system $\widetilde{\mathcal{P}}\eta = r$ can be reformulated into $(\Pi\widetilde{\mathcal{P}}\Pi^{\mathsf{T}})\Pi\eta = \Pi r$, that is,

$$\left(I_x \otimes \begin{bmatrix} C_1C_2^{-1} & -\frac{\tau^2(C_2^{-1})^{\mathsf{T}}}{\sqrt{\gamma}} \\ \frac{\tau^2 C_2^{-1}}{\sqrt{\gamma}} & C_1^{\mathsf{T}}(C_2^{-1})^{\mathsf{T}} \end{bmatrix} - \frac{\tau^2}{2}\Delta_h \otimes \begin{bmatrix} I_t & \\ & I_t \end{bmatrix}\right)\Pi\eta = \Pi r. \tag{2.13}$$

Assume that the matrix Δ_h is diagonalizable as $\Delta_h = \mathbb{V}_h^{-1}\mathrm{diag}(\sigma(\Delta_h))\mathbb{V}_h$ with $\sigma(\Delta_h)$ being the eigenvalues of Δ_h. Then, by letting

$$\widehat{\eta} = \left(\mathbb{V}_h \otimes \begin{bmatrix} I_t & \\ & I_t \end{bmatrix}\right)\Pi\eta,\ \widehat{r} = \left(\mathbb{V}_h \otimes \begin{bmatrix} I_t & \\ & I_t \end{bmatrix}\right)\Pi r, \tag{2.14}$$

the system (2.13) consists of N_x smaller systems corresponding to each eigenvalue $\lambda \in \sigma(\Delta_h)$:

$$\widetilde{\mathcal{P}}_\lambda\widehat{\eta}_\lambda := \left(\begin{bmatrix} C_1C_2^{-1} & -\frac{\tau^2(C_2^{-1})^{\mathsf{T}}}{\sqrt{\gamma}} \\ \frac{\tau^2 C_2^{-1}}{\sqrt{\gamma}} & C_1^{\mathsf{T}}(C_2^{-1})^{\mathsf{T}} \end{bmatrix} - \frac{\lambda\tau^2}{2}\begin{bmatrix} I_t & \\ & I_t \end{bmatrix}\right)\widehat{\eta}_\lambda = \widehat{r}_\lambda,\ \text{i.e., } (V\Sigma(\lambda)V^{-1})\widehat{\eta}_\lambda = \widehat{r}_\lambda, \tag{2.15}$$

where $\widehat{\eta}_\lambda$ and $\widehat{r}_\lambda$ are the corresponding sub-blocks of $\widehat{\eta}$ and $\widehat{r}$, respectively, and

$$\Sigma(\lambda) := \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} - \frac{\lambda\tau^2}{2}\begin{bmatrix} I_t & \\ & I_t \end{bmatrix}.$$

In view of the above discussion, we can analyze the relative error for each eigenvalue of the matrix Δ_h.

THEOREM 2.2. *Let $\lambda \in \sigma(\Delta_h)$ be an arbitrary eigenvalue of the space discrete matrix Δ_h and $\widehat{\eta}_\lambda$ be the exact solution of $\widetilde{\mathcal{P}}_\lambda\widehat{\eta}_\lambda = \widehat{r}_\lambda$ (cf.* (2.15)*) with any input vector $\widehat{r}_\lambda$ and $\widetilde{\eta}_\lambda$ be the approximate solution obtained via the diagonalization technique* (2.12)*. Assume Step* (b) *of* (2.12) *is solved in a direct manner (e.g., by LU factorization). Then, denoting the machine precision by ϵ, we have*

$$\frac{\|\widehat{\eta}_\lambda - \widetilde{\eta}_\lambda\|_2}{\|\widehat{\eta}_\lambda\|_2} \le \epsilon(4N_t+1)\mathrm{Cond}_2(\Sigma(\lambda)) + O(\epsilon^2). \tag{2.16}$$

Proof. For (2.15), according to the backward error analysis in [20, pp. 122–126], the solution obtained by the diagonalization technique (2.12) satisfies the perturbed systems

$$(V+\delta V_1)g_\lambda = \widehat{r}_\lambda,\ (\Sigma(\lambda)+\delta\Sigma)w_\lambda = g_\lambda,\ (V^{-1}+\delta V_2)\widetilde{\eta}_\lambda = w_\lambda,$$

where δV_1, δV_2, and $\delta\Sigma$ are the small perturbation of the matrices V, V^{-1}, and $\Sigma(\lambda)$, respectively. From [20, pp. 122–126], we have

$$\begin{aligned} &\|\delta V_1\|_2 \le 2\epsilon N_t\|V\|_2 + O(\epsilon^2),\ \|\delta V_2\|_2 \le 2\epsilon N_t\|V^{-1}\|_2 + O(\epsilon^2), \\ &\|\delta\Sigma(\lambda)\| \le \epsilon\|\Sigma(\lambda)\|_2 + O(\epsilon^2), \end{aligned} \tag{2.17a}$$

where the last inequality follows from the fact that $\Sigma(\lambda)$ is a diagonal matrix. Note that solving (2.15) by diagonalization is equivalent to exactly solving $(\widetilde{\mathcal{P}}_\lambda+\delta\widetilde{\mathcal{P}}_\lambda)\widetilde{\eta}_\lambda = \widehat{r}_\lambda$ with a suitable perturbation $\delta\widetilde{\mathcal{P}}_\lambda$. Moreover, from (2.15) we have

$$(V+\delta V_1)(\Sigma(\lambda)+\delta\Sigma)(V^{-1}+\delta V_2)\widetilde{\eta}_\lambda = \widehat{r}_\lambda. \tag{2.17b}$$

From (2.17a) and (2.17b), we can estimate $\delta\widetilde{\mathcal{P}}_\lambda$ as follows:

$$\|\delta\widetilde{\mathcal{P}}_\lambda\|_2 \le \epsilon(4N_t+1)\|V\|_2\|V^{-1}\|_2\|\Sigma(\lambda)\|_2 + O(\epsilon^2). \tag{2.18}$$

From [16, Lemma 2.6], it holds that $\frac{\|\widehat{\eta}_\lambda-\widetilde{\eta}_\lambda\|_2}{\|\widehat{\eta}_\lambda\|_2} \le \mathrm{Cond}_2(\widetilde{\mathcal{P}}_\lambda)\frac{\|\delta\widetilde{\mathcal{P}}_\lambda\|_2}{\|\widetilde{\mathcal{P}}_\lambda\|_2}$ and by using (2.18) we have

$$\begin{aligned}\frac{\|\widehat{\eta}_\lambda-\widetilde{\eta}_\lambda\|_2}{\|\widehat{\eta}_\lambda\|_2} &\le \epsilon(4N_t+1)\|V\|_2\|V^{-1}\|_2\|\Sigma(\lambda)\|_2\|\widetilde{\mathcal{P}}_\lambda^{-1}\|_2 + O(\epsilon^2)\\ &= \epsilon(4N_t+1)\mathrm{Cond}_2(V)\|\widetilde{\mathcal{P}}_\lambda^{-1}\|_2\|\Sigma(\lambda)\|_2 + O(\epsilon^2).\end{aligned} \tag{2.19}$$

Since $V = [\begin{smallmatrix}\mathbb{F}^* & \\ & \mathbb{F}^*\end{smallmatrix}]S$, $\widetilde{\mathcal{P}}_\lambda = V\Sigma(\lambda)V^{-1}$, and $\|\mathbb{F}\|_2 = \|\mathbb{F}^*\|_2 = 1$, we have $\mathrm{Cond}_2(V) = \mathrm{Cond}_2(S) = 1$ and

$$\|\widetilde{\mathcal{P}}_\lambda^{-1}\|_2 \le \|S\Sigma^{-1}(\lambda)S^{-1}\|_2 \le \mathrm{Cond}_2(S)\|\Sigma^{-1}(\lambda)\|_2 = \|\Sigma^{-1}(\lambda)\|_2.$$

Substituting these inequalities into (2.19) gives (2.16). ∎

The above theorem shows that the roundoff errors due to the diagonlization steps (2.12) only grow linearly with respect to N_t, where the other factor $\mathrm{Cond}_2(\Sigma(\lambda))$ corresponds to the unavoidable condition number of $\widetilde{\mathcal{P}}$ itself. For any matrix P, its spectral factorization $P = UDU^{-1}$ is not unique, where D and U are, respectively, the (diagonal) eigenvalue matrix and the eigenvector matrix. For any invertible diagonal matrix $\widetilde{D}$ it holds that $P = (U\widetilde{D})D(U\widetilde{D})^{-1}$. Hence, $\widetilde{U} = U\widetilde{D}$ is another eigenvector matrix. Therefore, a different spectral factorization leads to a different condition number of the eigenvector matrix, but it gives the same diagonal eigenvalue matrix D. In particular, the factor $\mathrm{Cond}_2(\Sigma(\lambda))$ in (2.16) cannot be further improved.

3. Eigenvalue analysis of the proposed PinT preconditioner. For the purpose of analysis, we will base our following eigenvalue analysis on an equivalent but symmetric reformulation of the discretized system. Interestingly, we find that the above diagonalization implementation cannot be directly transferred to the following symmetric system. Upon swapping the two rows and dividing both sides by τ^2, the above system (2.3) can be rewritten as an indefinite symmetric saddle-point system

$$\mathcal{A}\begin{bmatrix}\tilde{y}_h\\ p_h\end{bmatrix} := \begin{bmatrix}\frac{1}{\sqrt{\gamma}}\check{I}_{xt} & L^{\mathsf{T}}\\ L & -\frac{1}{\sqrt{\gamma}}\hat{I}_{xt}\end{bmatrix}\begin{bmatrix}\tilde{y}_h\\ p_h\end{bmatrix} = \begin{bmatrix}g_h/\tau^2\\ \tilde{f}_h/\tau^2\end{bmatrix} =: \begin{bmatrix}\bar{g}_h\\ \bar{f}_h\end{bmatrix}, \tag{3.1}$$

where $\hat{I}_{xt} = \hat{I}_t \otimes I_x, \check{I}_{xt} = \check{I}_t \otimes I_x$, and

(3.2)

$$L = \frac{1}{\tau^2}\left(B_1\otimes I_x - \frac{\tau^2}{2}B_2\otimes\Delta_h\right) = \frac{1}{\tau^2}\begin{bmatrix} G & 0 & 0 & 0 & \cdots & 0\\ -2I_x & G & 0 & 0 & \cdots & 0\\ G & -2I_x & G & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & 0 & 0\\ 0 & \cdots & G & -2I_x & G & 0\\ 0 & 0 & \cdots & G & -2I_x & G\end{bmatrix}$$

with $G = I_x - \frac{\tau^2}{2}\Delta_h$. The proposed preconditioner in (2.4) corresponds to the following symmetric indefinite preconditioner (here $\mathbb{P}$ is used only for the purpose of eigenvalue analysis):

$$\mathbb{P} := \begin{bmatrix} \frac{1}{\sqrt{\gamma}} I_{xt} & K^{\mathsf{T}} \\ K & -\frac{1}{\sqrt{\gamma}} I_{xt} \end{bmatrix}, \tag{3.3}$$

where $\check{I}_{xt}$ and $\hat{I}_{xt}$ in $\mathcal{A}$ are replaced by $I_{xt} = I_t \otimes I_x$ and the block Toeplitz matrix L is replaced by a block Strang circulant matrix K given as

(3.4)

$$K = \frac{1}{\tau^2}\left(C_1 \otimes I_x - \frac{\tau^2}{2} C_2 \otimes \Delta_h\right) = \frac{1}{\tau^2}\begin{bmatrix} G & 0 & 0 & \cdots & G & -2I_x \\ -2I_x & G & 0 & 0 & \cdots & G \\ G & -2I_x & G & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & G & -2I_x & G & 0 \\ 0 & 0 & \cdots & G & -2I_x & G \end{bmatrix}.$$

When a preconditioned Krylov subspace method (such as GMRES) is used, its convergence rate is often highly affected on the spectral distribution of the preconditioned matrix $\mathcal{P}^{-1}\mathcal{A}$, which will be explicitly estimated in the following subsections. Nevertheless, we mention the well-known negative fact [21] that the convergence rate of GMRES is not conclusively determined by the eigenvalues alone. To circumvent these theoretical difficulties of GMRES, one may alternatively use the preconditioned MINRES with a symmetric positive definite preconditioner; see [37] for related discussion and [46, 47] for the interesting idea of symmetrizing a nonsymmetric Toeplitz matrix. However, the generalization of such symmetrization techniques to our two-by-two block indefinite saddle-point system is not straightforward, and the construction of an effective symmetric positive definite preconditioner is also not trivial for our case. We highlight that in our preconditioner $\mathbb{P}$ we have modified all four blocks of $\mathcal{A}$ without resorting to any approximation of Schur complements, which was not commonly seen in the literature of preconditioning saddle-point systems.

We now proceed to estimate the eigenvalues of the preconditioned matrix $\mathbb{P}^{-1}\mathcal{A}$, that is,

$$\mathbb{P}^{-1}\mathcal{A} = \begin{bmatrix} \frac{1}{\sqrt{\gamma}} I_{xt} & K^{\mathsf{T}} \\ K & -\frac{1}{\sqrt{\gamma}} I_{xt} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{\sqrt{\gamma}} \check{I}_{xt} & L^{\mathsf{T}} \\ L & -\frac{1}{\sqrt{\gamma}} \hat{I}_{xt} \end{bmatrix}. \tag{3.5}$$

By noticing K is a rank-$2N_x$ perturbation of L, a standard low-rank perturbation argument [8] can be used to show that at least $2(N_t - 3)N_x$ eigenvalues of $\mathbb{P}^{-1}\mathcal{A}$ are exactly one, but the remaining $6N_x$ (nonunit) eigenvalues require further discussion to characterize. To facilitate the following analysis, we slightly modify the diagonal blocks of $\mathbb{P}$ into

$$\widehat{\mathbb{P}} = \begin{bmatrix} \frac{1}{\sqrt{\gamma}} \check{I}_{xt} & K^{\mathsf{T}} \\ K & -\frac{1}{\sqrt{\gamma}} \hat{I}_{xt} \end{bmatrix} = \mathbb{P} + \begin{bmatrix} \frac{1}{\sqrt{\gamma}} e_{N_t} e_{N_t}^{\mathsf{T}} & 0 \\ 0 & -\frac{1}{\sqrt{\gamma}} e_1 e_1^{\mathsf{T}} \end{bmatrix} \otimes \frac{I_x}{2} =: \mathbb{P} + \mathbb{P}_0, \tag{3.6}$$

which is a rank-$2N_x$ perturbation of $\mathbb{P}$. Here and hereafter, e_n is the nth column of the identity matrix I_t. Hence, we have the following matrix product relation:

$$\mathbb{P}^{-1}\mathcal{A} = (\mathbb{P}^{-1}\widehat{\mathbb{P}})(\widehat{\mathbb{P}}^{-1}\mathcal{A}) = \mathbb{P}^{-1}(\mathbb{P} + \mathbb{P}_0)(\widehat{\mathbb{P}}^{-1}\mathcal{A}) = (\mathcal{I} + \mathbb{P}^{-1}\mathbb{P}_0)(\widehat{\mathbb{P}}^{-1}\mathcal{A}),$$

where $\mathcal{I} = [\,{}^{I_t}_{\ I_t}\,] \otimes I_x$ is an identity matrix. Moreover, the eigenvalues of $\mathbb{P}^{-1}\widehat{\mathbb{P}}$ satisfy

$$|\lambda(\mathbb{P}^{-1}\widehat{\mathbb{P}}) - 1| = |\lambda(\mathbb{P}^{-1}\mathbb{P}_0)| \le \|\mathbb{P}^{-1}\mathbb{P}_0\|_2 \le \frac{1}{2},$$

where the last inequality will be shown below. From the matrix perturbation theory [28, Chap. 21], this implies the spectrum of $\mathbb{P}^{-1}\mathcal{A}$ is close to that of $\widehat{\mathbb{P}}^{-1}\mathcal{A}$. A typical 1D spectrum distribution (with $\gamma = 10^{-6}$) of $\mathbb{P}^{-1}\mathcal{A}$, $\mathbb{P}^{-1}\widehat{\mathbb{P}}$, and $\widehat{\mathbb{P}}^{-1}\mathcal{A}$ is shown in Figure 1, where we observe that the rectangular region containing all the eigenvalues of $\mathbb{P}^{-1}\mathcal{A}$ is only slightly different from that of $\widehat{\mathbb{P}}^{-1}\mathcal{A}$, which is relatively easier to analyze. In the following two subsections, we will estimate the eigenvalues of $\mathbb{P}^{-1}\widehat{\mathbb{P}}$ and $\widehat{\mathbb{P}}^{-1}\mathcal{A}$, respectively.

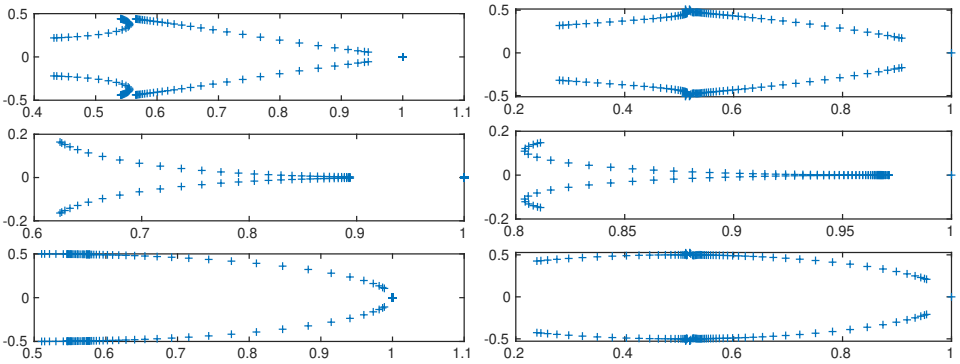

FIG. 1. *The eigenvalue distribution of* $\mathbb{P}^{-1}\mathcal{A}$ *(top row),* $\mathbb{P}^{-1}\widehat{\mathbb{P}}$ *(middle row), and* $\widehat{\mathbb{P}}^{-1}\mathcal{A}$ *(bottom row) for Example* 1 *with* $\gamma = 10^{-6}$*, and* $(N_x, N_t) = (32, 33)$ *(left) and* $(N_x, N_t) = (64, 65)$ *(right).*

3.1. Step I: Estimate the eigenvalues of $\mathbb{P}^{-1}\widehat{\mathbb{P}}$. Since $\mathbb{P}$ is symmetric, its singular values are given by the absolute value of all real eigenvalues. So we can estimate the norm $\|\mathbb{P}^{-1}\|_2 = 1/\sigma_{\min}(\mathbb{P})$ by estimating the smallest eigenvalue, in absolute value, of $\mathbb{P}$. Let $(\beta \in \mathbb{R}, z = [\,{}^{x}_{y}\,] \neq 0)$ be an eigenpair of $\mathbb{P}$ with $\|z\|_2 = 1$, that is,

$$\begin{bmatrix} \frac{1}{\sqrt{\gamma}} I_{xt} & K^{\mathsf{T}} \\ K & -\frac{1}{\sqrt{\gamma}} I_{xt} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \beta \begin{bmatrix} x \\ y \end{bmatrix}.$$

Clearly, we have $x \neq 0$ and $y \neq 0$; otherwise $z = 0$ since K is nonsingular. Writing both equations explicitly gives

$$\frac{1}{\sqrt{\gamma}} x + K^{\mathsf{T}} y = \beta x, \;\; Kx - \frac{1}{\sqrt{\gamma}} y = \beta y,$$

where the first equation gives $(\frac{1}{\sqrt{\gamma}} - \beta)x = -K^{\mathsf{T}} y$, and upon plugging this into the second equation we get

$$\beta^2 y - (KK^{\mathsf{T}} + \gamma^{-1} I_{xt}) y = 0. \tag{3.7}$$

Multiplying y^* from the left side of (3.7) and then dividing both sides by $y^* y$, we obtain

$$\beta^2 = \frac{y^*(KK^{\mathsf{T}} + \gamma^{-1} I_{xt}) y}{y^* y} \ge \sigma_{\min}^2(K) + \gamma^{-1} > 0,$$

where $\sigma_{\min}(K) > 0$ denotes the smallest singular value of K. Clearly, it holds that $|\beta| \geq \sqrt{\sigma_{\min}^2(K) + \gamma^{-1}}$, which implies $\sigma_{\min}(\mathbb{P}) \geq \sqrt{\sigma_{\min}^2(K) + \gamma^{-1}}$. Hence,

$$\|\mathbb{P}^{-1}\|_2 = \frac{1}{\sigma_{\min}(\mathbb{P})} \leq \frac{1}{\sqrt{\sigma_{\min}^2(K) + \gamma^{-1}}}.$$

By noticing $\|\mathbb{P}_0\|_2 = \frac{1}{2\sqrt{\gamma}}$, we therefore get

$$\begin{aligned}\|\mathbb{P}^{-1}\widehat{\mathbb{P}} - \mathcal{I}\|_2 &= \|\mathbb{P}^{-1}\mathbb{P}_0\|_2 \leq \|\mathbb{P}^{-1}\|_2\|\mathbb{P}_0\|_2 \\ &\leq \frac{1}{\sqrt{\sigma_{\min}^2(K) + \gamma^{-1}}}\frac{1}{2\sqrt{\gamma}} = \frac{1}{2\sqrt{\gamma\sigma_{\min}^2(K) + 1}} \leq \frac{1}{2},\end{aligned}$$

which implies $\mathbb{P}^{-1}\widehat{\mathbb{P}}$ is not too far away from an identity matrix. Also, we have the eigenvalue estimate

$$|\lambda(\mathbb{P}^{-1}\widehat{\mathbb{P}}) - 1| = |\lambda(\mathbb{P}^{-1}\widehat{\mathbb{P}} - \mathcal{I})| \leq \|\mathbb{P}^{-1}\widehat{\mathbb{P}} - \mathcal{I}\|_2 \leq \frac{1}{2},$$

which indicates the distances of those nonunit (including possible complex) eigenvalues of $\mathbb{P}^{-1}\widehat{\mathbb{P}}$ from one are uniformly bounded by $\frac{1}{2}$ (as clearly observed in Figure 1). This justifies our claim that the spectrum of $\mathbb{P}^{-1}\mathcal{A}$ is very close to that of $\widehat{\mathbb{P}}^{-1}\mathcal{A}$. Finally, we summarize the above conclusions into the following theorem.

THEOREM 3.1. *For the above given matrices* $\mathbb{P}$ *in* (3.5) *and* $\widehat{\mathbb{P}}$ *in* (3.6)*, there holds*

$$|\lambda(\mathbb{P}^{-1}\widehat{\mathbb{P}}) - 1| \leq \|\mathbb{P}^{-1}\widehat{\mathbb{P}} - \mathcal{I}\|_2 \leq \frac{1}{2}.$$

3.2. Step II: Estimate the eigenvalues of $\widehat{\mathbb{P}}^{-1}\mathcal{A}$. Following the analysis techniques in [52], we now discuss the eigenvalues of the modified preconditioned system $\widehat{\mathbb{P}}^{-1}\mathcal{A}$. Let $(\theta \in \mathbb{C}, z = [\begin{smallmatrix} x \\ y \end{smallmatrix}] \neq 0)$ be an eigenpair of $\widehat{\mathbb{P}}^{-1}\mathcal{A}$ with $\|z\|_2 = 1$, that is,

$$\begin{bmatrix} \frac{1}{\sqrt{\gamma}}\check{I}_{xt} & L^{\mathsf{T}} \\ L & -\frac{1}{\sqrt{\gamma}}\hat{I}_{xt} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \theta\begin{bmatrix} \frac{1}{\sqrt{\gamma}}\check{I}_{xt} & K^{\mathsf{T}} \\ K & -\frac{1}{\sqrt{\gamma}}\hat{I}_{xt} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}.$$

Writing both equations explicitly to get

$$\frac{1}{\sqrt{\gamma}}\check{I}_{xt}x + L^{\mathsf{T}}y = \theta\left(\frac{1}{\sqrt{\gamma}}\check{I}_{xt}x + K^{\mathsf{T}}y\right),\ Lx - \frac{1}{\sqrt{\gamma}}\hat{I}_{xt}y = \theta\left(Kx - \frac{1}{\sqrt{\gamma}}\hat{I}_{xt}y\right), \tag{3.8}$$

where the first equation gives

$$(1-\theta)x = \sqrt{\gamma}\check{I}_{xt}^{-1}(\theta K^{\mathsf{T}} - L^{\mathsf{T}})y,$$

and upon plugging this into the second equation in (3.8), we get (after multiplying both sides by $\sqrt{\gamma}$)

$$\theta^2(\gamma K\check{I}_{xt}^{-1}K^{\mathsf{T}} + \hat{I}_{xt})y - \theta(\gamma(K\check{I}_{xt}^{-1}L^{\mathsf{T}} + L\check{I}_{xt}^{-1}K^{\mathsf{T}}) + 2\hat{I}_{xt})y + (\gamma L\check{I}_{xt}^{-1}L^{\mathsf{T}} + \hat{I}_{xt})y = 0.$$

Multiplying y^* from the left side, we obtain a quadratic equation

$$a\theta^2 - b\theta + c = 0$$

with coefficients a, b, and c given by

$$a = y^*(\gamma K\check{I}^{-1}K^{\mathsf{T}}+\hat{I})y,\ b = y^*(\gamma(K\check{I}^{-1}L^{\mathsf{T}}+L\check{I}^{-1}K^{\mathsf{T}})+2\hat{I})y,\ c = y^*(\gamma L\check{I}^{-1}L^{\mathsf{T}}+\hat{I})y.$$

Since the case $\theta = 1$ is trivial to discuss, we assume $\theta \neq 1$, which implies both $x \neq 0$ and $y \neq 0$. Hence, $a > 0$ and $c > 0$, but here b is not necessarily positive. By the quadratic formula we have

$$\theta_{\pm} = \frac{b}{2a} \pm \sqrt{\left(\frac{b}{2a}\right)^2 - \frac{c}{a}} =: \Re(\theta) \pm i\Im(\theta).$$

(1) If $\Im(\theta) = 0$, then θ is real and $\theta^* = \theta$. Multiplying from the left side of the equations in (3.8) by x^* and y^*, respectively, we get

$$\begin{aligned}
\frac{1}{\sqrt{\gamma}}x^*\check{I}_{xt}x + x^*L^{\mathsf{T}}y &= \theta\left(\frac{1}{\sqrt{\gamma}}x^*\check{I}_{xt}x + x^*K^{\mathsf{T}}y\right),\\
y^*Lx - \frac{1}{\sqrt{\gamma}}y^*\hat{I}_{xt}y &= \theta\left(y^*Kx - \frac{1}{\sqrt{\gamma}}y^*\hat{I}_{xt}y\right),
\end{aligned}$$

which, upon subtracting the second equation from the conjugate of the first one, leads to

$$\frac{1}{\sqrt{\gamma}}x^*\check{I}_{xt}x + \frac{1}{\sqrt{\gamma}}y^*\hat{I}_{xt}y = \theta\left(\frac{1}{\sqrt{\gamma}}x^*\check{I}_{xt}x + \frac{1}{\sqrt{\gamma}}y^*\hat{I}_{xt}y\right),$$

where we have used the facts $(x^*L^{\mathsf{T}}y)^* = y^*Lx$ and $(x^*K^{\mathsf{T}}y)^* = y^*Kx$. This implies $\theta = 1$ since $x^*\check{I}_{xt}x \geq \frac{1}{2}\|x\|_2^2 > 0$ and $y^*\hat{I}_{xt}y \geq \frac{1}{2}\|y\|_2^2 > 0$. Hence, all the real eigenvalues must be exactly one. Moreover, the eigenvalue one of $\widehat{\mathbb{P}}^{-1}\mathcal{A} = \mathcal{I} - \widehat{\mathbb{P}}^{-1}(\widehat{\mathbb{P}} - \mathcal{A})$ actually has a multiplicity of at least $(2N_t - 4)N_x$, since the difference $\widehat{\mathbb{P}} - \mathcal{A}$ is only of rank $4N_x$ [24].

(2) If $\Im(\theta) \neq 0$, then $b^2 - 4ac < 0$, and we obviously have the expression

$$\Re(\theta) = \frac{b}{2a} = \frac{y^*(\gamma(K\check{I}_{xt}^{-1}L^{\mathsf{T}} + L\check{I}_{xt}^{-1}K^{\mathsf{T}}) + 2\hat{I}_{xt})y}{2y^*(\gamma K\check{I}_{xt}^{-1}K^{\mathsf{T}} + \hat{I}_{xt})y}$$

and also

$$|\theta|^2 = \Re(\theta)^2 + \Im(\theta)^2 = \frac{c}{a} = \frac{y^*(\gamma L\check{I}_{xt}^{-1}L^{\mathsf{T}} + \hat{I}_{xt})y}{y^*(\gamma K\check{I}_{xt}^{-1}K^{\mathsf{T}} + \hat{I}_{xt})y}.$$

To simplify the analysis, we write

$$\begin{aligned}
|\theta|^2 &= \frac{y^*(\gamma L\check{I}_{xt}^{-1}L^{\mathsf{T}} + \hat{I}_{xt})y}{y^*(\gamma K\check{I}_{xt}^{-1}K^{\mathsf{T}} + \hat{I}_{xt})y}\\
&= \underbrace{\frac{y^*(\gamma L\check{I}_{xt}^{-1}L^{\mathsf{T}} + \hat{I}_{xt})y}{y^*(\gamma LL^{\mathsf{T}} + I_{xt})y}}_{\rho_{LL}} \cdot \underbrace{\frac{y^*(\gamma LL^{\mathsf{T}} + I_{xt})y}{y^*(\gamma KK^{\mathsf{T}} + I_{xt})y}}_{\rho_{KL}} \cdot \underbrace{\frac{y^*(\gamma KK^{\mathsf{T}} + I_{xt})y}{y^*(\gamma K\check{I}_{xt}^{-1}K^{\mathsf{T}} + \hat{I}_{xt})y}}_{\rho_{KK}},
\end{aligned} \tag{3.9}$$

where we need to estimate three different positive ratios: ρ_{LL}, ρ_{KL}, and ρ_{KK}. We first notice that

$$\begin{aligned}
&\frac{y^*(L\check{I}_{xt}^{-1}L^{\mathsf{T}})y}{y^*(LL^{\mathsf{T}})y} \overset{v=L^{\mathsf{T}}y}{=\!=\!=} \frac{v^*\check{I}_{xt}^{-1}v}{v^*v} \in [1,2],\\
&\frac{y^*(KK^{\mathsf{T}})y}{y^*(K\check{I}_{xt}^{-1}K^{\mathsf{T}})y} \overset{v=K^{\mathsf{T}}y}{=\!=\!=} \frac{v^*v}{v^*\check{I}_{xt}^{-1}v} \in \left[\frac{1}{2},1\right], \quad \frac{y^*\hat{I}_{xt}y}{y^*I_{xt}y} \in \left[\frac{1}{2},1\right].
\end{aligned}$$

It is easy to show the following simple inequalities (for any positive numbers c_1, c_2, d_1, d_2):

$$\min\left\{\frac{c_1}{d_1}, \frac{c_2}{d_2}\right\} \le \frac{c_1+c_2}{d_1+d_2} \le \max\left\{\frac{c_1}{d_1}, \frac{c_2}{d_2}\right\},$$

with which we obtain

$$\frac{1}{2} \le \min\left\{\frac{y^*(L\check{I}_{xt}^{-1}L^{\mathsf{T}})y}{y^*(LL^{\mathsf{T}})y}, \frac{y^*\hat{I}_{xt}y}{y^*I_{xt}y}\right\} \le \rho_{LL} \le \max\left\{\frac{y^*(L\check{I}_{xt}^{-1}L^{\mathsf{T}})y}{y^*(LL^{\mathsf{T}})y}, \frac{y^*\hat{I}_{xt}y}{y^*I_{xt}y}\right\} \le 2,$$

and

$$\frac{1}{2} \le \min\left\{\frac{y^*(KK^{\mathsf{T}})y}{y^*(K\check{I}_{xt}^{-1}K^{\mathsf{T}})y}, \frac{y^*I_{xt}y}{y^*\hat{I}_{xt}y}\right\} \le \rho_{KK} \le \max\left\{\frac{y^*(KK^{\mathsf{T}})y}{y^*(K\check{I}_{xt}^{-1}K^{\mathsf{T}})y}, \frac{y^*I_{xt}y}{y^*\hat{I}_{xt}y}\right\} \le 2.$$

Hence, it holds that

$$\frac{1}{2}\sqrt{\rho_{KL}} \le |\theta| \le 2\sqrt{\rho_{KL}},$$

where the range of ρ_{KL} is determined by the min/max eigenvalues of the matrix

$$W := (\gamma KK^{\mathsf{T}} + I_{xt})^{-1}(\gamma LL^{\mathsf{T}} + I_{xt}). \tag{3.10}$$

To summarize, we have obtained the following theorem for estimating the eigenvalues of $\widehat{\mathbb{P}}^{-1}A$.

THEOREM 3.2. *For the above given matrices $\mathcal{A}$ in* (3.1) *and $\widehat{\mathbb{P}}$ in* (3.6)*, let $(\theta \in \mathbb{C}, z = [\begin{smallmatrix} x \\ y \end{smallmatrix}] \neq 0)$ be any eigenpair of $\widehat{\mathbb{P}}^{-1}\mathcal{A}$. Then*

(1) *If θ is real, then $\theta = 1$. The eigenvalue $\theta = 1$ has a multiplicity of at least $(2N_t - 4)N_x$.*
(2) *If θ is not real, then we have the expression*

$$|\theta|^2 = \frac{y^*(\gamma L\check{I}_{xt}^{-1}L^{\mathsf{T}} + \hat{I}_{xt})y}{y^*(\gamma K\check{I}_{xt}^{-1}K^{\mathsf{T}} + \hat{I}_{xt})y}.$$

Moreover, there holds

$$\frac{1}{2}\sqrt{\rho_{KL}} \le |\theta| \le 2\sqrt{\rho_{KL}},$$

where ρ_{KL} is bounded by the min/max eigenvalues of the matrix W given in (3.10).

We can discuss more about the range of ρ_{KL}. Clearly, it holds that

$$|\rho_{KL} - 1| = \left|\frac{y^*(\gamma LL^{\mathsf{T}} - \gamma KK^{\mathsf{T}})y}{y^*(\gamma KK^{\mathsf{T}} + I_{xt})y}\right| \le \gamma\left|\frac{y^*(LL^{\mathsf{T}} - KK^{\mathsf{T}})y}{y^*y}\right| \le \gamma\|LL^{\mathsf{T}} - KK^{\mathsf{T}}\|_\infty.$$

In the 1D case, using a uniform mesh with $\tau = 2h$, it holds that $\|G\|_\infty = \|I_x - \frac{\tau^2}{2}\Delta_h\|_\infty = 1 + 2(\frac{\tau}{h})^2 = 9$. Recall from (3.2) and (3.4) that

$$K = L + \frac{1}{\tau^2}R := L + \frac{1}{\tau^2}\begin{bmatrix} 0 & \cdots & 0 & G & -2I_x \\ & 0 & \cdots & 0 & G \\ & & 0 & \cdots & 0 \\ & & & \ddots & \vdots \\ & & & & 0 \end{bmatrix},$$

which implies

$$KK^{\mathsf{T}} - LL^{\mathsf{T}} = \left(L + \frac{1}{\tau^2}R\right)\left(L + \frac{1}{\tau^2}R\right)^{\mathsf{T}} - LL^{\mathsf{T}} = \frac{1}{\tau^4}(\tau^2 RL^{\mathsf{T}} + \tau^2 LR^{\mathsf{T}} + RR^{\mathsf{T}}),$$

where

$$\tau^2 RL^{\mathsf{T}} + \tau^2 LR^{\mathsf{T}} = \begin{bmatrix} 0 & \cdots & 0 & G^2 & -4G \\ 0 & 0 & \cdots & 0 & G^2 \\ 0 & 0 & 0 & \cdots & 0 \\ & & \vdots & \ddots & \vdots \\ G^2 & 0 & & & \\ -4G & G^2 & 0 & \cdots & 0 \end{bmatrix}, RR^{\mathsf{T}} = \begin{bmatrix} G^2 + 4I_x & -2G & 0 & \cdots & 0 \\ -2G & G^2 & 0 & \cdots & 0 \\ & & \ddots & \ddots & \\ 0 & 0 & & & 0 \\ & & \ddots & \ddots & \\ 0 & \vdots & & & 0 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Hence, it is straightforward to show

$$\begin{aligned} \|LL^{\mathsf{T}} - KK^{\mathsf{T}}\|_\infty &\le \frac{1}{\tau^4}(\|G^2 + 4I_x\|_\infty + \| - 2G\|_\infty + \|G^2\|_\infty + \| - 4G\|_\infty) \\ &\le \frac{1}{\tau^4}(4 + 6\|G\|_\infty + 2\|G\|_\infty^2) \le \frac{220}{\tau^4}. \end{aligned}$$

This implies that ρ_{KL} (or the eigenvalues of W given by (3.10)) becomes closer to 1 as the regularization parameter $\gamma \to 0$, which predicts a faster GMRES convergence rate with smaller γ for fixed mesh step sizes. However, according to this estimate with a fixed γ, the eigenvalue distribution of W indeed gets worse as τ becomes smaller, which matches with the numerical observations (see Figure 2). Interestingly, a similar connection between γ and τ^4 was also observed in [37, Thm. 3.2], where the derived eigenvalue upper bound becomes nonuniform when $\tau^4 < \gamma$, indicating the possibility of a slower convergence rate with a very small τ.

It also holds that

$$-1 < \frac{-y^*(\gamma KK^{\mathsf{T}})y}{y^*(\gamma KK^{\mathsf{T}} + I_{xt})y} < \rho_{KL} - 1$$

and

$$\rho_{KL} - 1 < \frac{y^*(LL^{\mathsf{T}})y}{y^*(KK^{\mathsf{T}} + \gamma^{-1}I_{xt})y} < \min\left\{\frac{y^*(LL^{\mathsf{T}})y}{y^*(KK^{\mathsf{T}})y}, \gamma\sigma_{\max}^2(L)\right\},$$

which, however, does not lead to uniform bounds for ρ_{KL}, since the eigenvalues of $(KK^{\mathsf{T}})^{-1}LL^{\mathsf{T}}$ are not uniformly bounded. Using the data in Example 1, a typical eigenvalue distribution of W and $\widehat{\mathbb{P}}^{-1}\mathcal{A}$ is shown in Figures 2 and 3, respectively. As expected, the eigenvalues of W do not seem to be uniformly bounded as the mesh is refined for a fixed γ. Nevertheless, the eigenvalues indeed become more clustered around 1 as γ gets smaller. Numerically, we do observe very tight bounds for the magnitude of $|\theta|$, which may explain the observed fast convergence rate of the preconditioned GMRES solver with our proposed PinT preconditioner. However, it remains open to prove such tight bounds for estimating $|\theta|$, which cannot be easily achieved based on the spectrum information of $(KK^{\mathsf{T}})^{-1}LL^{\mathsf{T}}$.

As a further illustration, we also plotted in Figure 4 the nonunity eigenvalues of $(KK^{\mathsf{T}})^{-1}LL^{\mathsf{T}}$, which seem to be highly clustered around 1, but they are not uniformly bounded due to a few outliers. Such an observed highly clustered but nonuniformly bounded spectrum is well known [7, 30] for block- (or two-level) circulant type preconditioners, which, however, often show very fast convergence rates for the preconditioned Krylov subspace solvers in practice.

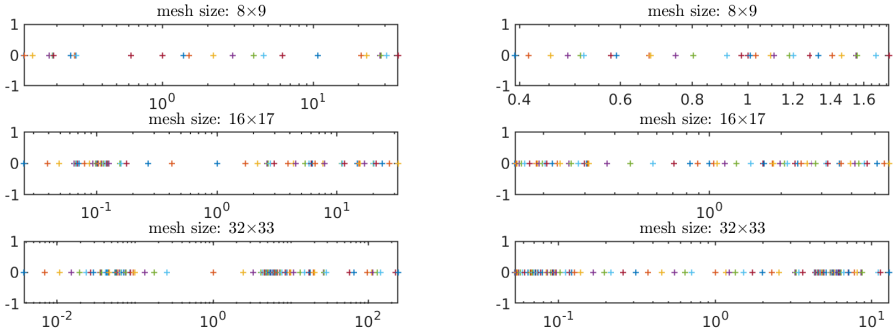

FIG. 2. *The eigenvalue distribution of W given by* (3.10) *for Example* 1 *with different mesh sizes (left: $\gamma = 10^{-2}$; right: $\gamma = 10^{-4}$).*

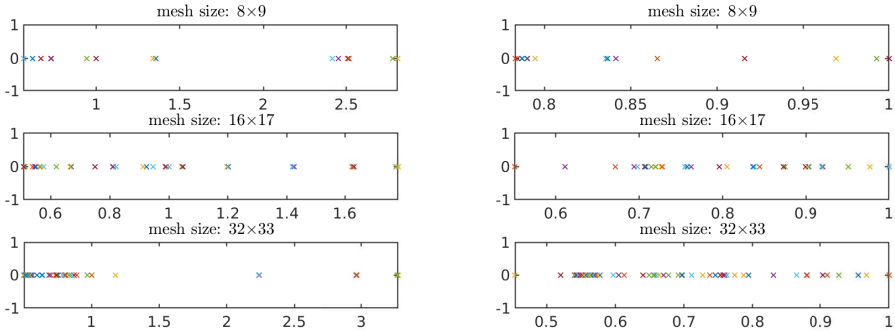

FIG. 3. *The absolute eigenvalue $|\theta|$ distribution of $\widehat{\mathbb{P}}^{-1}\mathcal{A}$ for Example* 1 *with different mesh sizes (left: $\gamma = 10^{-2}$; right: $\gamma = 10^{-4}$).*

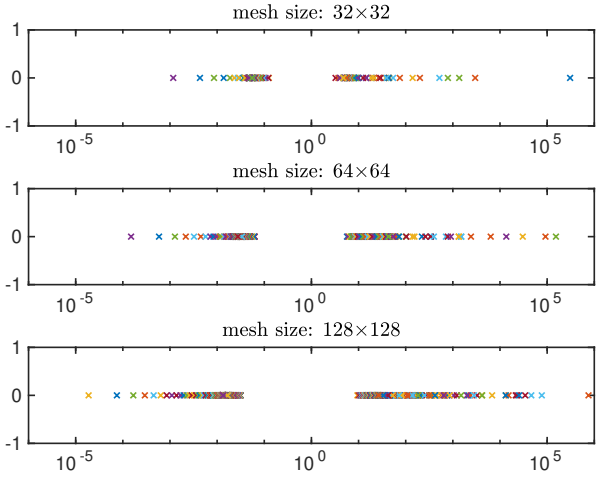

FIG. 4. *The nonunity eigenvalue distribution of $(KK^{\mathsf{T}})^{-1}LL^{\mathsf{T}}$ for Example* 1 *with different mesh sizes.*

4. Application to the control constraint situation. In real-world applications, many PDE optimal control problems are equipped with boxed constraints of the control variable, i.e.,

$$u_a \le u(x,t) \le u_b \quad \text{in } \Omega \times (0,T), \tag{4.1}$$

where $u_a \le u_b$ are two given constants. For this kind of control-constrained problem, the semismooth Newton iterations (or equivalently the primal-dual active set method) together with the preconditioned Krylov subspace solvers (as inner loop) are the mainstream method; see, e.g., [22, 27, 33]. At the kth semismooth Newton iteration, we need to solve a linear system with coefficient matrix of the form

$$\mathcal{M}_k = \begin{bmatrix} B_1 & -\frac{\tau^2\chi_k}{\sqrt{\gamma}} \\ \frac{\tau^2 \check{I}_t}{\sqrt{\gamma}} & B_1^{\mathsf{T}} \end{bmatrix} \otimes I_x - \frac{\tau^2}{2} \begin{bmatrix} B_2 & \\ & B_2^{\mathsf{T}} \end{bmatrix} \otimes \Delta_h, \tag{4.2}$$

where $\check{I}_t$ is given in (2.1c), B_1 and B_2 are given in (2.2), and χ_k is a 0-1 binary diagonal matrix due to the active/inactive control constraints. It is not easy to directly apply our proposed block-circulant preconditioner to handle $\mathcal{M}_k$. On one hand, if we simply ignore χ_k and continue to use $\mathcal{P}$ given by (2.4) as the preconditioner, the preconditioned GMRES method performs poorly. On the other hand, if we keep χ_k as a component of $\mathcal{P}$, there is no explicit formula available for the spectral decomposition of $\mathcal{P}$. Hence, an alternative approach is needed for treating the control-constrained case.

In this section, we propose a new algorithm to handle this kind of problem, where the preconditioning technique proposed in section 2 can be used directly. First, we represent the constraint $u_a \le u \le u_b$ as

$$Qu + q \ge 0, \qquad \text{with} \qquad Q = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad q = \begin{bmatrix} -u_a \\ u_b \end{bmatrix}.$$

Then, according to the methodology developed in [23], the optimality condition of the control problem (1.1a)–(1.1b) with constraint (4.1) can be formulated as

$$\begin{cases} \gamma u - p - Q^{\mathsf{T}}\mu = 0, & 0 \le \mu \perp Qu + q \ge 0, \qquad \text{in } \Omega \times (0,T), \\ y_{tt} - \Delta y = f + u & \text{in } \Omega \times (0,T), \qquad y = 0, \text{ on } \partial\Omega \times (0,T), \\ y(\cdot,0) = y_0, \quad y_t(\cdot,0) = y_1 & \text{in } \Omega, \\ p_{tt} - \Delta p + y = g & \text{in } \Omega \times (0,T), \qquad p = 0, \text{ on } \partial\Omega \times (0,T), \\ p(\cdot,T) = 0, \quad p_t(\cdot,T) = 0 & \text{in } \Omega, \end{cases} \tag{4.3}$$

where μ is the Lagrange multiplier of the control constraints and p is the adjoint state. We can again eliminate the control variable $u = \gamma^{-1}(p + Q^{\mathsf{T}}\mu)$ and rewrite (4.3) as follows:

$$\begin{cases} 0 \le \mu \perp \gamma^{-1}Qp + q + \gamma^{-1}QQ^{\mathsf{T}}\mu \ge 0 & \text{in } \Omega \times (0,T), \\ y_{tt} - \Delta y = f + \gamma^{-1}(p + Q^{\mathsf{T}}\mu) & \text{in } \Omega \times (0,T), \qquad y = 0, \text{ on } \partial\Omega \times (0,T), \\ y(\cdot,0) = y_0, \quad y_t(\cdot,0) = y_1 & \text{in } \Omega, \\ p_{tt} - \Delta p + y = g & \text{in } \Omega \times (0,T), \qquad p = 0, \text{ on } \partial\Omega \times (0,T), \\ p(\cdot,T) = 0, \quad p_t(\cdot,T) = 0 & \text{in } \Omega. \end{cases} \tag{4.4}$$

To solve (4.4) efficiently via a splitting algorithm, we make the following change of variable:

$$\mu = \gamma\phi, \tag{4.5}$$

which transforms (4.4) to

(4.6)
$$\begin{cases} 0 \le \phi \perp \gamma^{-1}Qp + q + QQ^{\mathsf{T}}\phi \ge 0 & \text{in } \Omega\times(0,T), \\ y_{tt} - \Delta y - \gamma^{-1}p = Q^{\mathsf{T}}\phi + f & \text{in } \Omega\times(0,T), \quad y = 0, \text{ on } \partial\Omega\times(0,T), \\ y(\cdot,0) = y_0, \quad y_t(\cdot,0) = y_1 & \text{in } \Omega, \\ p_{tt} - \Delta p + y = g & \text{in } \Omega\times(0,T), \quad p = 0, \text{ on } \partial\Omega\times(0,T), \\ p(\cdot,T) = 0, \quad p_t(\cdot,T) = 0 & \text{in } \Omega. \end{cases}$$

Let Φ_n denote the lexicographic ordered column vector collecting the approximate solutions of $\phi(\cdot,t_n)$ over all the space grids. Discretizing the above two PDEs in (4.6) via the same scheme and notations used in section 2 gives

$$\begin{aligned} &0 \le \Phi_n \perp \gamma^{-1}Q_hP_n + q_h + Q_h{Q_h}^{\mathsf{T}}\Phi_n \ge 0, \ n = 0,1,2,\ldots,N_t-1, \\ &\widehat{\mathcal{M}}\begin{bmatrix} y_h \\ p_h \end{bmatrix} = \begin{bmatrix} (\hat{I}_t \otimes {Q_h}^{\mathsf{T}})\phi_h + f_h \\ g_h \end{bmatrix}, \end{aligned} \tag{4.7}$$

where $\phi_h = \left(\Phi_0^{\mathsf{T}}, \Phi_1^{\mathsf{T}}, \ldots, \Phi_{N_t-1}^{\mathsf{T}}\right)^{\mathsf{T}}$, $Q_h = \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right] \otimes I_x$, $q_h = \left[\begin{smallmatrix} -u_a \\ u_b \end{smallmatrix}\right] \otimes (1,1,\ldots,1)^{\mathsf{T}}$, and

$$\widehat{\mathcal{M}} = \begin{bmatrix} L & -\gamma^{-1}(\hat{I}_t \otimes I_x) \\ (\check{I}_t \otimes I_x) & L^{\top} \end{bmatrix}, \text{ with } L = \frac{1}{\tau^2}B_1 \otimes I_x - \frac{1}{2}B_2 \otimes \Delta_h.$$

The discretized problem (4.7) is a large-scale nonlinear algebraic system consisting of a linear complementarity system (LCP) and a system of linear equations. Clearly, the unconstrained case corresponds to vanishing LCP with $\phi_h = 0$. We propose solving (4.7) via the following splitting quasi-Newton iterations ($k = 0,1,\ldots$ denotes the iteration index):

$$\begin{aligned} &0 \le \Phi_n^{k+1} \perp \gamma^{-1}Q_hP_n^k + q_h + Q_h{Q_h}^{\mathsf{T}}\Phi_n^{k+1} \ge 0, \ n = 0,1,2,\ldots,N_t-1, \\ &\begin{bmatrix} y_h^{k+1} \\ p_h^{k+1} \end{bmatrix} = \begin{bmatrix} y_h^k \\ p_h^k \end{bmatrix} - \mathcal{J}^{-1}r^k, \qquad \text{with } r^k := \widehat{\mathcal{M}}\begin{bmatrix} y_h^k \\ p_h^k \end{bmatrix} - \begin{bmatrix} (\hat{I}_t \otimes {Q_h}^{\mathsf{T}})\phi_h^{k+1} + f_h \\ g_h \end{bmatrix}, \end{aligned} \tag{4.8a}$$

where the approximate Jacobian matrix $\mathcal{J}$ is chosen as

$$\mathcal{J} = \begin{bmatrix} L & -\hat{I}_t \otimes I_x \\ \check{I}_t \otimes I_x & L^{\mathsf{T}} \end{bmatrix}. \tag{4.8b}$$

Since $Q_hQ_h^{\mathsf{T}} = \left[\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right] \otimes I_x$ is a semipositive definite matrix, according to [9, pp. 4–5, Chapter 1] the LCP in (4.8a) can be reformulated into the following least-norm problem (for fixed k and n):

(4.9)
$$\Phi_n^{k+1} = \operatorname{argmin}\left\{\|z\|_2 : z \ge 0, \ \gamma^{-1}Q_hP_n^k + q_h + Q_hQ_h^{\mathsf{T}}z \ge 0\right\}, \ n = 1,2,\ldots,N_t,$$

which again can be solved efficiently via a quadratic programming solver (e.g., `quadprog` in MATLAB). Other more specialized algorithms (e.g., the Newton-type method [14]) for solving LCPs can also be used.

There are two advantages of the algorithm (4.8a). First, we can solve both the linear algebraic system and the linear complementarity system in a PinT pattern. Precisely, we can compute all the components of ϕ_h^{k+1}, i.e., $\{\Phi_n^{k+1}\}_{n=1}^{N_t}$, simultaneously. Moreover, we can compute $\mathcal{J}^{-1}r^k$ by using our proposed PinT block-circulant preconditioner. Second, as we will illustrate in Examples 3 and 4 of section 5, the iteration (4.8a)–(4.8b) converges rapidly, and the convergence rate is very robust with respect to the regularization and discretization parameters. We also want to mention that the variable change (4.5) and the special choice of the Jacobian matrix $\mathcal{J}$ play equally important roles in achieving the robust convergence of iteration (4.8a)–(4.8b). Without (4.5) or (4.8b), the convergence rate deteriorates as the regularization parameter or the discretization parameter becomes small. However, a complete convergence analysis of (4.8a)–(4.8b) requires further efforts, which is beyond the scope of our current paper and will be left as future work.

5. Numerical examples. In this section, we provide several numerical results for our proposed PinT preconditioner. Here, we mainly focus on the convergence properties of the preconditioner, and parallel computation results will be given in the near future. As given in (2.1a)–(2.1b), the optimality system (1.2) is discretized in space using the central finite difference method and in time by the implicit leap-frog finite difference scheme in [35], with a uniform mesh in both space and time. All simulations are implemented using MATLAB on a Dell Precision Workstation with Intel(R) Core(TM) i7-7700K CPU@4.2GHz and 32GB RAM. The CPU time (in seconds) is estimated using the timing functions `tic/toc`. We remark that the reported CPU times are based on one-time simulation, which may show some variance. We employ the right-preconditioned GMRES solver (without restarts) provided by the IFISS package [10, 11, 51] and choose a zero initial guess and a stopping tolerance `tol` based on the reduction in relative residual norms. According to the error estimate in [35], we will measure the discrete $L^\infty((0,T);L^2(\Omega))$ error norms of the state and adjoint state approximation as e_y and e_p, respectively, and then estimate the experimental order of accuracy by calculating the logarithmic ratio of the approximation errors between two successive refined meshes, i.e.,

$$\mathrm{R} = \log_2\left(\frac{e_y(h,\tau)}{e_y(2h,2\tau)}\right) \quad \text{or} \quad \log_2\left(\frac{e_p(h,\tau)}{e_p(2h,2\tau)}\right),$$

which should be close to 2 for a second-order accuracy in approximation errors. It is worthwhile to mention that the GMRES solver without preconditioning would require a much greater number of iterations, which can easily exceed several thousands even for the simple 1D case. For brevity, we choose to omit such results and focus on comparing our proposed PinT preconditioner with the state-of-the-art MSC preconditioner.

Example 1. Let $\Omega = (0,1)$ and $T = 2$. We choose $y_0(x) = \sin(\pi x)$, $y_1(x) = 0$,

$$f = -\frac{1}{\gamma}\sin(\pi x)(e^t - e^T)^2,$$

and

$$g(x,t) = 2(2e^{2t} - e^{T+t})\sin(\pi x) + \pi^2\sin(\pi x)(e^t - e^T)^2 + \sin(\pi x)\cos(\pi t),$$

such that the exact solution is $y(x,t) = \sin(\pi x)\cos(\pi t)$ and $p(x,t) = \sin(\pi x)(e^t - e^T)^2$.

In Table 1, we report the required number of GMRES iterations (denoted by "It") and CPU times for our proposed PinT preconditioner. The number of iterations is

TABLE 1
Numbers of GMRES iterations and CPU times for Example 1 *with our proposed PinT preconditioner.*

$\texttt{tol} = 10^{-7}$	$\gamma = 1$		$\gamma = 10^{-2}$		$\gamma = 10^{-4}$		$\gamma = 10^{-6}$		$\gamma = 10^{-8}$	
(N_x, N_t)	It	CPU	It	CPU	It	CPU	It	CPU	It	CPU
(128, 129)	5	0.1	5	0.1	5	0.1	5	0.1	5	0.1
(256, 257)	5	0.2	7	0.3	5	0.2	5	0.3	5	0.2
(512, 513)	9	1.4	15	2.4	5	0.8	5	0.8	5	0.8
(1024,1025)	9	5.6	23	15.4	9	5.7	5	3.3	5	3.3

TABLE 2
Numbers of GMRES iterations and CPU times for Example 1 *with the MSC preconditioner* ($\texttt{tol} = 10^{-7}$).

$\texttt{tol} = 10^{-7}$	$\gamma = 1$		$\gamma = 10^{-2}$		$\gamma = 10^{-4}$		$\gamma = 10^{-6}$		$\gamma = 10^{-8}$	
(N_x, N_t)	It	CPU	It	CPU	It	CPU	It	CPU	It	CPU
(128, 129)	6	0.1	11	0.2	14	0.2	11	0.2	3	0.1
(256, 257)	6	0.2	11	0.3	18	0.6	17	0.5	4	0.1
(512, 513)	6	0.7	11	1.3	20	2.5	26	3.4	10	1.2
(1024,1025)	5	2.3	10	4.6	21	10.4	38	21.8	19	9.2

TABLE 3
Comparison of error results for Example 1 *with the PinT and MSC preconditioners with different* γ.

	$\texttt{tol} = 10^{-7}$	Our PinT preconditioner				The MSC preconditioner			
γ	(N_x, N_t)	e_y	R	e_p	R	e_y	R	e_p	R
1	(128, 129)	3.5e-03	–	8.3e-03	–	3.5e-03	–	8.3e-03	–
	(256, 257)	8.7e-04	2.0	2.1e-03	2.0	8.7e-04	2.0	2.1e-03	2.0
	(512, 513)	2.2e-04	2.0	5.3e-04	2.0	2.2e-04	2.0	5.3e-04	2.0
	(1024,1025)	5.5e-05	2.0	1.3e-04	2.0	5.4e-05	2.0	1.4e-04	1.9
10^{-2}	(128, 129)	3.1e-02	–	2.0e-03	–	3.1e-02	–	2.0e-03	–
	(256, 257)	7.7e-03	2.0	5.0e-04	2.0	7.7e-03	2.0	5.0e-04	2.0
	(512, 513)	1.9e-03	2.0	1.3e-04	2.0	1.9e-03	2.0	1.3e-04	2.0
	(1024,1025)	4.9e-04	2.0	3.1e-05	2.0	5.1e-04	1.9	2.9e-05	2.1
10^{-4}	(128, 129)	5.5e-02	–	4.0e-04	–	5.5e-02	–	4.0e-04	–
	(256, 257)	1.4e-02	2.0	1.0e-04	2.0	1.4e-02	2.0	1.0e-04	2.0
	(512, 513)	3.5e-03	2.0	2.5e-05	2.0	3.5e-03	2.0	2.5e-05	2.0
	(1024,1025)	8.7e-04	2.0	6.3e-06	2.0	9.0e-04	2.0	6.5e-06	1.9
10^{-6}	(128, 129)	3.0e-01	–	1.3e-04	–	3.0e-01	–	1.3e-04	–
	(256, 257)	7.8e-02	1.9	3.3e-05	1.9	7.7e-02	2.0	3.3e-05	1.9
	(512, 513)	2.0e-02	2.0	8.4e-06	2.0	2.0e-02	1.9	8.9e-06	1.9
	(1024,1025)	4.9e-03	2.0	2.1e-06	2.0	4.8e-03	2.1	3.0e-06	1.6
10^{-8}	(128, 129)	8.9e-01	–	2.5e-05	–	8.8e-01	–	2.3e-05	–
	(256, 257)	2.8e-01	1.7	8.9e-06	1.5	2.0e-01	2.2	5.9e-06	2.0
	(512, 513)	7.5e-02	1.9	2.6e-06	1.8	5.0e-02	2.0	2.0e-06	1.6
	(1024,1025)	1.9e-02	2.0	6.6e-07	2.0	2.5e-02	1.0	2.0e-06	-0.0

very robust with respect to the mesh step sizes and the regularization parameter γ. Based on our eigenvalue analysis, we would expect to achieve faster convergence rates (or require fewer iterations) when γ becomes very small, which is confirmed by the results in the last few columns. Nevertheless, we do observe a slightly deteriorated convergence rate of our PinT preconditioner for a relatively large $\gamma \geq 10^{-4}$ with a very fine mesh, which is anticipated since our shown eigenvalue distribution is indeed not uniformly bounded. In addition, the preconditioning step is unavoidably affected by the introduced roundoff errors during the diagonalization process. For comparison, the convergence results using the matching Schur complement (MSC) preconditioner

(based on the same diagonally rescaled system) are given in Table 2, where we notice slower convergence rates for very small γ. Nevertheless, the MSC preconditioner indeed performs better with larger $\gamma \geq 10^{-2}$, which can be roughly explained by the underlying mechanism of matching two Schur complement terms. Overall, our PinT preconditioner has faster and more robust convergence rates for the cases with $\gamma \leq 10^{-4}$, which are of our particular interest.

TABLE 4
Comparison of error results for Example 1 *with the MSC preconditioner with different* γ *and* $\texttt{tol} = 10^{-8}$ or $\texttt{tol} = 10^{-10}$.

		The MSC preconditioner ($\texttt{tol} = 10^{-8}$)					The MSC preconditioner ($\texttt{tol} = 10^{-10}$)				
γ	(N_x, N_t)	e_y	R	e_p	R	It	e_y	R	e_p	R	It
1	(128, 129)	3.5e-03	–	8.3e-03	–	6	3.5e-03	–	8.3e-03	–	8
	(256, 257)	8.7e-04	2.0	2.1e-03	2.0	7	8.7e-04	2.0	2.1e-03	2.0	8
	(512, 513)	2.2e-04	2.0	5.3e-04	2.0	6	2.2e-04	2.0	5.3e-04	2.0	8
	(1024,1025)	5.4e-05	2.0	1.3e-04	2.0	6	5.5e-05	2.0	1.3e-04	2.0	7
10^{-2}	(128, 129)	3.1e-02	–	2.0e-03	–	11	3.1e-02	–	2.0e-03	–	13
	(256, 257)	7.7e-03	2.0	5.0e-04	2.0	11	7.7e-03	2.0	5.0e-04	2.0	13
	(512, 513)	1.9e-03	2.0	1.3e-04	2.0	11	1.9e-03	2.0	1.3e-04	2.0	13
	(1024,1025)	4.9e-04	2.0	3.1e-05	2.0	11	4.9e-04	2.0	3.1e-05	2.0	12
10^{-4}	(128, 129)	5.5e-02	–	4.0e-04	–	17	5.5e-02	–	4.0e-04	–	22
	(256, 257)	1.4e-02	2.0	1.0e-04	2.0	21	1.4e-02	2.0	1.0e-04	2.0	27
	(512, 513)	3.5e-03	2.0	2.5e-05	2.0	23	3.5e-03	2.0	2.5e-05	2.0	29
	(1024,1025)	8.7e-04	2.0	6.3e-06	2.0	24	8.7e-04	2.0	6.3e-06	2.0	31
10^{-6}	(128, 129)	3.0e-01	–	1.3e-04	–	14	3.0e-01	–	1.3e-04	–	20
	(256, 257)	7.8e-02	1.9	3.3e-05	1.9	21	7.8e-02	1.9	3.3e-05	1.9	31
	(512, 513)	2.0e-02	2.0	8.4e-06	2.0	34	2.0e-02	2.0	8.4e-06	2.0	49
	(1024,1025)	4.9e-03	2.0	2.1e-06	2.0	49	4.9e-03	2.0	2.1e-06	2.0	65
10^{-8}	(128, 129)	8.8e-01	–	2.5e-05	–	4	8.9e-01	–	2.5e-05	–	8
	(256, 257)	2.8e-01	1.7	8.6e-06	1.5	7	2.8e-01	1.7	8.9e-06	1.5	12
	(512, 513)	7.4e-02	1.9	2.5e-06	1.8	14	7.5e-02	1.9	2.6e-06	1.8	23
	(1024,1025)	1.8e-02	2.0	6.9e-07	1.8	26	1.9e-02	2.0	6.6e-07	2.0	39

Besides the number of iterations, there are also some significant differences in approximation errors between the two preconditioners. The combined error results are given in Table 3, where we notice the chosen tolerance ($\texttt{tol} = 10^{-7}$) is sufficient for our PinT preconditioner to attain the desired second-order accuracy for most cases. However, the MSC preconditioner fails to achieve the expected second-order accuracy in the approximations for many cases. This indicates a better numerical robustness of our PinT preconditioner, especially when γ is very small. Such a remarkable discrepancy in approximation errors is not surprising, since a very small γ leads to a more ill-conditioned system that is very sensitive to any numerical errors from early termination of iterative solvers. We believe the degraded approximation accuracy of the MSC preconditioner with a smaller γ is likely due to its own very large condition number [48]. For further improving the accuracy of MSC preconditioner, it is necessary to use a much smaller tolerance (e.g., $\texttt{tol} = 10^{-10}$), and hence it requires significantly more iterations. For a complete illustration, the results of the MSC preconditioner with smaller $\texttt{tol} = 10^{-8}$ and $\texttt{tol} = 10^{-10}$ are presented in Table 4, where the approximation accuracy using $\texttt{tol} = 10^{-10}$ indeed matches with the results of our PinT preconditioner in Table 3. Hence, for a highly ill-conditioned system, a smaller stopping tolerance should be used to obtain accurate approximations.

Example 2 [35]. Now, we consider a 2D problem with $\Omega = (0,1)^2$, $T = 2$, and

$$y_0(x_1,x_2) = \sin(\pi x_1)\sin(\pi x_2), \quad y_1(x_1,x_2) = \sin(\pi x_1)\sin(\pi x_2),$$
$$f(x_1,x_2,t) = (1+2\pi^2)e^t \sin(\pi x_1)\sin(\pi x_2) - \frac{1}{\gamma}(t-T)^2\sin(\pi x_1)\sin(\pi x_2),$$
$$g(x_1,x_2,t) = (e^t + 2 + 2\pi^2(t-T)^2)\sin(\pi x_1)\sin(\pi x_2).$$

The exact solution of the optimal control problem is

$$y(x_1,x_2,t) = e^t \sin(\pi x_1)\sin(\pi x_2) \quad \text{and} \quad p(x_1,x_2,t) = (t-T)^2\sin(\pi x_1)\sin(\pi x_2).$$

Similar to Example 1, as shown in Tables 5 and 6, the convergence rates of our proposed PinT preconditioner are faster and more robust than that of the MSC preconditioner. With $\texttt{tol} = 10^{-7}$, the MSC preconditioner fails to provide second-order approximation accuracy for a very small γ. To recover the anticipated second-order accuracy, based on our numerical experiments, it is necessary to use a smaller tolerance $\texttt{tol} = 10^{-10}$. The corresponding convergence results are reported in Table 7, where more iterations are used. In this case, our PinT preconditioner clearly outperforms the MSC preconditioner.

TABLE 5
Numbers of GMRES iterations and CPU times for Example 2 *with our proposed PinT preconditioner.*

$\texttt{tol} = 10^{-7}$	$\gamma = 10^{-2}$		$\gamma = 10^{-4}$		$\gamma = 10^{-6}$		$\gamma = 10^{-8}$		$\gamma = 10^{-10}$	
(N_x, N_x, N_t)	It	CPU	It	CPU	It	CPU	It	CPU	It	CPU
(64,64,65)	5	0.7	5	1.1	5	0.8	5	0.8	4	0.6
(128,128,129)	11	13.9	5	7.2	5	6.7	5	6.4	5	6.7
(256,256,257)	17	226.6	5	59.6	5	60.3	5	61.0	5	60.7

TABLE 6
Numbers of GMRES iterations and CPU times for Example 2 *with the MSC preconditioner* ($\texttt{tol} = 10^{-7}$).

$\texttt{tol} = 10^{-7}$	$\gamma = 10^{-2}$		$\gamma = 10^{-4}$		$\gamma = 10^{-6}$		$\gamma = 10^{-8}$		$\gamma = 10^{-10}$	
(N_x, N_x, N_t)	It	CPU	It	CPU	It	CPU	It	CPU	It	CPU
(64,64,65)	12	1.6	15	2.0	10	1.4	4	0.6	1	0.2
(128,128,129)	12	13.1	19	21.7	17	18.8	6	6.9	2	2.8
(256,256,257)	12	119.4	23	230.6	23	236.5	9	90.3	2	25.2

TABLE 7
Numbers of GMRES iterations and CPU times for Example 2 *with the MSC preconditioner* ($\texttt{tol} = 10^{-10}$).

$\texttt{tol} = 10^{-10}$	$\gamma = 10^{-2}$		$\gamma = 10^{-4}$		$\gamma = 10^{-6}$		$\gamma = 10^{-8}$		$\gamma = 10^{-10}$	
(N_x, N_x, N_t)	It	CPU	It	CPU	It	CPU	It	CPU	It	CPU
(64,64,65)	14	2.1	21	3.4	16	2.3	7	1.0	3	0.5
(128,128,129)	14	17.7	27	37.3	24	28.4	11	12.1	4	4.9
(256,256,257)	14	154.8	35	400.0	36	396.1	16	162.1	7	71.5

Example 3 [37]. In this example, we consider the control-constrained case with

the following data:

$$\Omega = (0,1),\ T = 2,\ u_a = 5, u_b = 10,\ y_0(x) = \sin(\pi x),\ y_1(x) = 0,$$
$$f = -\max\{u_a, \gamma^{-1}\min\{u_b, \sin(\pi x)(t-T)^2\}\},$$
$$g = 2\sin(\pi x) + \pi^2 \sin(\pi x)(t-T)^2 + \sin(\pi x)\cos(\pi t).$$

The exact solution is $y(x,t) = \sin(\pi x)\cos(\pi t)$ and $p(x,t) = \sin(\pi x)(t-T)^2$. The initial guess of the quasi-Newton iteration (4.8a) is chosen as zeros, and the stopping condition is based on the reduction of relative residual $\|r^k\|_2/\|r^0\|_2 \le \texttt{tol}$, where r^k denotes the residual vector at the kth iteration; see (4.8a). In each iteration of (4.8a), we compute $\mathcal{J}^{-1}r^k$ by the preconditioned GMRES solver. Numerically, we find that it is unnecessary to solve the linear system very accurately in order to achieve the desired fast convergence. Hence, we used a relatively larger stopping tolerance $\sqrt{\texttt{tol}}$ in the inner preconditioned GMRES solver for better computational efficiency. The computation of the LCP systems for the N_t steps can be highly parallelizable, for which we use a damped and perturbed Newton-type algorithm in [14], with a stopping tolerance of 10^{-8}.

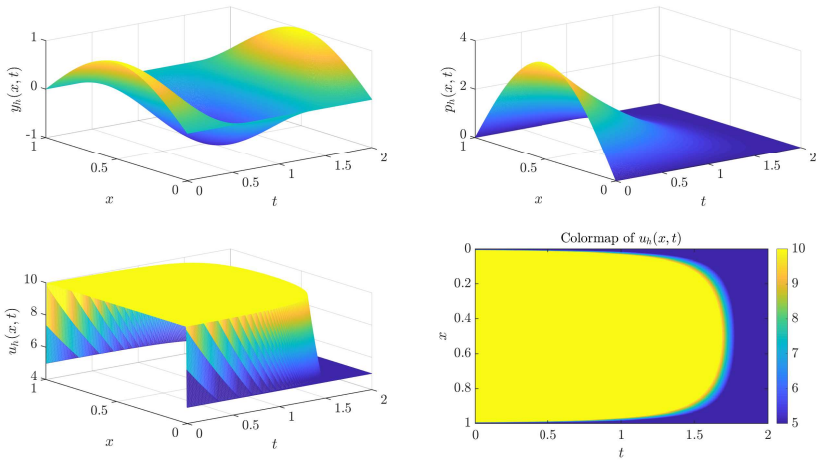

FIG. 5. *Computed y_h, p_h, and u_h with box constraints $5 \le u \le 10$ and $\gamma = 10^{-2}$ in Example 3 ($N_x = 512, N_t = 513$).*

Figure 5 depicts the computed optimal state, adjoint state, and optimal control (with its colormap) under the control constraint $5 \le u \le 10$, where the optimal control was reconstructed from the computed adjoint state. In Table 8, we report the error and the convergence results of solving Example 3 by the proposed quasi-Newton iterations (4.8a) with our PinT preconditioner. The desired second-order accuracy is observed in all cases. The columns $\mathrm{It_N}$, $\mathrm{It_L}$, and $\mathrm{It_G}$ provide the number of outer quasi-Newton iterations, averaged inner optimization iterations for solving LCP, and averaged inner preconditioned GMRES iterations for solving the Jacobian system defined by (4.8b), respectively. We notice that both the outer quasi-Newton solver and the inner LCP solver show clear mesh-independent convergence rates that are also very robust with respect to the regularization parameter. However, we indeed observe mildly deteriorated convergence rates in the inner preconditioned GMRES solver with our current PinT preconditioner.

TABLE 8

Results for Example 3 solved by the proposed quasi-Newton iteration with our PinT preconditioner ($\texttt{tol} = 10^{-7}$*).*

γ	(N_x, N_t)	e_y	R	e_p	R	$\mathrm{It_N}$	$\mathrm{It_L}$	$\mathrm{It_G}$	CPU
10^{-2}	(128,129)	1.8e-03	–	2.6e-04	–	7	5	41	4.5
	(256,257)	4.5e-04	2.0	6.5e-05	2.0	7	5	54	20.5
	(512,513)	1.1e-04	2.0	1.6e-05	2.0	7	5	80	137.5
10^{-4}	(128,129)	1.8e-03	–	2.5e-04	–	6	5	41	4.0
	(256,257)	4.5e-04	2.0	6.4e-05	2.0	6	5	56	18.5
	(512,513)	1.1e-04	2.0	1.6e-05	2.0	6	5	84	128.1
10^{-6}	(128,129)	1.8e-03	–	2.5e-04	–	7	5	47	5.2
	(256,257)	4.5e-04	2.0	6.5e-05	2.0	7	5	67	25.5
	(512,513)	1.1e-04	2.0	1.8e-05	1.9	7	5	98	188.8
10^{-8}	(128,129)	1.8e-03	–	2.5e-04	–	6	5	37	3.7
	(256,257)	4.5e-04	2.0	6.3e-05	2.0	6	6	57	19.0
	(512,513)	1.1e-04	2.0	1.6e-05	2.0	7	6	96	172.3
10^{-10}	(128,129)	1.8e-03	–	2.5e-04	–	6	6	39	3.8
	(256,257)	4.5e-04	2.0	6.4e-05	2.0	6	6	57	19.2
	(512,513)	1.1e-04	2.0	1.6e-05	2.0	6	6	82	122.5

This numerical observation in Table 8 is unexpected, since the Jacobian matrix $\mathcal{J}$ just corresponds to the same system in Example 1 with $\gamma = 1$. We conjecture that such a less satisfactory convergence behavior of the inner GMRES solver is mainly caused by the *nonsmoothness* of the solution ϕ_h of the LCP system.[2] Such a nonsmooth ϕ_h^k leads to a nonsmooth right-hand side r^k (defined by (4.8a)). But the product $\mathcal{J}^{-1}r^k$ should be smooth in the asymptotic sense (as k gets larger), because upon convergence the solutions y_h and p_h are indeed smooth. This may partially explain why it takes more iterations to find a good approximation of the solution $\mathcal{J}^{-1}r^k$ in the Krylov subspace $\mathrm{span}\{r^k, \mathcal{J}r^k, \mathcal{J}^2r^k, \dots\}$. We highlight that a similar worse convergence behavior of the GMRES solver is also reported in [19] when applying the block-circulant preconditioner to a one-shot scheme for the wave equation with nonsmooth initial data.

To ameliorate the convergence rate of the inner GMRES solver, we need to design a more effective preconditioner for the approximate Jacobian matrix $\mathcal{J}$. Inspired by the idea in [4, 38, 39], we consider a slight generalization of the preconditioner $\mathcal{P}$ in (2.4) with $\gamma = 1$, that is,

$$\mathcal{P}(\alpha) = \begin{bmatrix} C_1(\alpha) & -\tau^2 I_t \\ \tau^2 I_t & C_1^{\mathsf{T}}(\alpha) \end{bmatrix} \otimes I_x - \frac{\tau^2}{2} \begin{bmatrix} C_2(\alpha) & \\ & C_2^{\mathsf{T}}(\alpha) \end{bmatrix} \otimes \Delta_h, \tag{5.1}$$

where $\alpha \in (0, 1]$ is a free parameter and $C_{1,2}(\alpha)$ are the α-*circulant* matrices defined by

$$C_1(\alpha) = \begin{bmatrix} 1 & & & \alpha & -2\alpha \\ -2 & 1 & & & \alpha \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{bmatrix}, \; C_2(\alpha) = \begin{bmatrix} 1 & & & \alpha & 0 \\ 0 & 1 & & & \alpha \\ 1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 0 & 1 \end{bmatrix}. \tag{5.2}$$

[2] The nonsmoothness of the solution of an LCP system is well known; see, e.g., [9, pp. 668–670, Chapter 7].

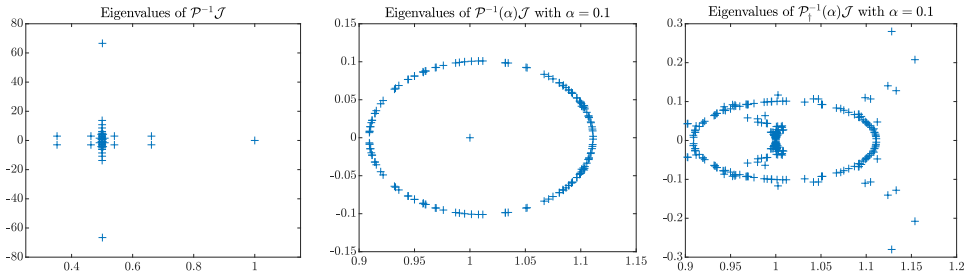

FIG. 6. *The eigenvalue distribution of $\mathcal{P}^{-1}\mathcal{J}$ (i.e., $\mathcal{P}^{-1}(\alpha)\mathcal{J}$ with $\alpha = 1$) (left panel), $\mathcal{P}^{-1}(\alpha)\mathcal{J}$ with $\alpha = 0.1$ (middle panel), and $\mathcal{P}_{\dagger}^{-1}(\alpha)\mathcal{J}$ with $\alpha = 0.1$ (right panel) in Example* 3 *(with $N_x = 64$ and $N_t = 65$).*

The preconditioner used for Table 8 corresponds $\alpha = 1$ in (5.1). In Figure 6, we plot the distribution of the eigenvalues of $\mathcal{P}^{-1}(\alpha)\mathcal{J}$ with two values of α: $\alpha = 1$ and $\alpha = 0.1$. We observe that a slightly small parameter α leads to more clustered eigenvalues. Unfortunately, an explicit formula, as given for the case $\alpha = 1$, for the diagonalization of $\mathcal{P}(\alpha)$ with $\alpha \in (0, 1)$ in section 2.1, is not available at the moment. Alternatively, we can choose the block diagonals of $\mathcal{P}(\alpha)$ and get the block-diagonal preconditioner

$$\mathcal{P}_{\dagger}(\alpha) = \begin{bmatrix} C_1(\alpha) & \\ & C_1^{\mathsf{T}}(\alpha) \end{bmatrix} \otimes I_x - \frac{\tau^2}{2} \begin{bmatrix} C_2(\alpha) & \\ & C_2^{\mathsf{T}}(\alpha) \end{bmatrix} \otimes \Delta_h, \tag{5.3}$$

for which an explicit diagonalization is indeed available, since the α-circulant matrices $C_{1,2}(\alpha)$ can be diagonalized simultaneously [5, Theorem 2.10] as

$$C_{1,2}(\alpha) = V(\alpha) D_{1,2}(\alpha) V^{-1}(\alpha), \tag{5.4}$$

where

$$V(\alpha) = \Lambda(\alpha)\mathbb{F}^* \quad \text{and} \quad D_{1,2}(\alpha) = \sqrt{N_t}\,\mathbb{F}\Lambda^{-1}(\alpha) C_{1,2}(\alpha)(:, 1)$$

with $\Lambda(\alpha) = \mathrm{diag}(1, \alpha^{-\frac{1}{N_t}}, \alpha^{-\frac{2}{N_t}}, \ldots, \alpha^{-\frac{N_t-1}{N_t}})$ and $C_{1,2}(\alpha)(:, 1)$ being the first column of the matrix $C_{1,2}(\alpha)$. With (5.4) and letting $U(\alpha) = \begin{bmatrix} V(\alpha) & \\ & V^*(\alpha) \end{bmatrix} \otimes I_x$, we can diagonalize $\mathcal{P}_{\dagger}(\alpha)$ in a PinT manner as

$$\mathcal{P}_{\dagger}(\alpha) = U(\alpha) \left(\begin{bmatrix} D_1(\alpha) & \\ & D_1^*(\alpha) \end{bmatrix} \otimes I_x - \frac{\tau^2}{2} \begin{bmatrix} D_2(\alpha) & \\ & D_2^*(\alpha) \end{bmatrix} \otimes \Delta_h \right) U(\alpha)^{-1}. \tag{5.5}$$

In Figure 6, we also show the eigenvalue distribution of $\mathcal{P}_{\dagger}^{-1}(\alpha)\mathcal{J}$ with $\alpha = 0.1$, and this illustrates that the clustering of the eigenvalues is only slightly worse than $\mathcal{P}^{-1}(\alpha)\mathcal{J}$ and is much better than that of $\mathcal{P}^{-1}\mathcal{J}$. It is worthwhile to mention that the Jacobian matrix $\mathcal{J}$ does not depend on the regularization parameter γ, which allows $\mathcal{P}_{\dagger}(\alpha)$ to be a good preconditioner of $\mathcal{J}$. In particular, the matrix $\mathcal{P}_{\dagger}(\alpha)$ will not be a good preconditioner of $\mathcal{M}$ (or $\widehat{\mathcal{M}}$) whenever $\gamma \ll 1$. Hence, our choice of the Jacobian matrix $\mathcal{J}$ is very crucial.

In comparison with Table 8, we report in Table 9 the approximation error and convergence results of solving Example 3 by using $\mathcal{P}_{\dagger}(\alpha)$ with $\alpha = 0.1$ as the preconditioner. We notice that the fast and robust convergence rates of the quasi-Newton iterations and the LCP solver are not affected, but the convergence rate of the inner

TABLE 9

Results for Example 3 *by the quasi-Newton iteration with the block α-circulant preconditioner $\mathcal{P}_\dagger(\alpha = 0.1)$ ($tol = 10^{-7}$).*

γ	(N_x, N_t)	e_y	R	e_p	R	$\mathrm{It_N}$	$\mathrm{It_L}$	$\mathrm{It_G}$	CPU
10^{-2}	(128,129)	1.8e-03	–	2.6e-04	–	7	5	4	0.8
	(256,257)	4.5e-04	2.0	6.5e-05	2.0	7	5	5	2.9
	(512,513)	1.1e-04	2.0	1.6e-05	2.0	6	5	5	9.1
	(1024,1025)	2.8e-05	2.0	4.1e-06	2.0	6	5	5	36.4
10^{-4}	(128,129)	1.8e-03	–	2.5e-04	–	6	5	4	0.7
	(256,257)	4.5e-04	2.0	6.4e-05	2.0	6	5	4	2.4
	(512,513)	1.1e-04	2.0	1.6e-05	2.0	5	5	4	7.7
	(1024,1025)	2.8e-05	2.0	4.0e-06	2.0	6	5	5	34.7
10^{-6}	(128,129)	1.8e-03	–	2.5e-04	–	6	5	4	0.7
	(256,257)	4.4e-04	2.0	6.4e-05	2.0	6	5	4	2.5
	(512,513)	1.1e-04	2.0	1.6e-05	2.0	5	5	4	7.8
	(1024,1025)	2.5e-05	2.2	2.9e-06	2.4	5	6	4	30.0
10^{-8}	(128,129)	1.8e-03	–	2.5e-04	–	6	6	4	0.8
	(256,257)	4.5e-04	2.0	6.4e-05	2.0	5	6	4	2.2
	(512,513)	1.1e-04	2.0	1.6e-05	2.0	5	6	4	7.9
	(1024,1025)	2.5e-05	2.2	3.0e-06	2.4	5	6	4	30.5
10^{-10}	(128,129)	1.8e-03	–	2.5e-04	–	6	6	4	0.8
	(256,257)	4.5e-04	2.0	6.4e-05	2.0	5	6	4	2.1
	(512,513)	1.1e-04	2.0	1.6e-05	2.0	5	6	4	8.0
	(1024,1025)	2.5e-05	2.2	3.0e-06	2.4	5	6	4	30.4

GMRES solver is dramatically improved to be clearly mesh-independent. More in-depth eigenvalue analysis of such a block (diagonal) α-circulant preconditioner will be carried out in our future work. It is worthwhile to point out that the simplified $\mathcal{P}_\dagger(\alpha)$ is much easier to diagonalize than both $\mathcal{P}(\alpha)$ and $\mathcal{P}$.

TABLE 10

Results for Example 4 *by the quasi-Newton iteration with the block α-circulant preconditioner $\mathcal{P}_\dagger(\alpha = 0.1)$ ($tol = 10^{-7}$).*

γ	(N_x, N_t)	e_y	R	e_p	R	$\mathrm{It_N}$	$\mathrm{It_L}$	$\mathrm{It_G}$	CPU
10^{-2}	(64,64,65)	5.80e-04	–	4.97e-04	–	6	5	4	19.7
	(128,128,129)	1.49e-04	2.0	1.27e-04	2.0	6	5	4	97.6
	(256,256,257)	3.76e-05	2.0	3.19e-05	2.0	6	5	5	827.3
10^{-4}	(64,64,65)	5.80e-04	–	4.97e-04	–	6	5	4	13.3
	(128,128,129)	1.49e-04	2.0	1.27e-04	2.0	6	5	4	100.0
	(256,256,257)	3.76e-05	2.0	3.19e-05	2.0	6	5	5	850.3
10^{-6}	(64,64,65)	5.80e-04	–	4.97e-04	–	6	5	4	14.4
	(128,128,129)	1.49e-04	2.0	1.27e-04	2.0	6	6	4	104.6
	(256,256,257)	3.76e-05	2.0	3.19e-05	2.0	6	6	5	902.0
10^{-8}	(64,64,65)	5.80e-04	–	4.97e-04	–	6	6	4	25.9
	(128,128,129)	1.49e-04	2.0	1.27e-04	2.0	6	6	4	110.2
	(256,256,257)	3.76e-05	2.0	3.19e-05	2.0	6	6	5	926.7
10^{-10}	(64,64,65)	5.80e-04	–	4.97e-04	–	6	6	4	15.1
	(128,128,129)	1.49e-04	2.0	1.27e-04	2.0	6	6	4	108.6
	(256,256,257)	3.76e-05	2.0	3.19e-05	2.0	6	6	5	929.6

Example 4. In the last example, we consider a control-constrained 2D problem

with the data

$$\Omega = (0,1)^2, T = 2, u_a = -10, u_b = -5, y_0(x_1,x_2) = 0, y_1(x_1,x_2) = \nu(x_1,x_2),$$
$$f(x_1,x_2,t) = -\frac{1}{(t+1)^2}\nu(x_1,x_2) - \ln(t+1)\Delta\nu(x_1,x_2) - \max\left\{u_a, \min\left\{u_b, \frac{1}{\gamma}(t-T)^2\nu(x_1,x_2)\right\}\right\},$$
$$g(x_1,x_2,t) = 2\nu(x_1,x_2) - (t-T)^2\Delta\nu(x_1,x_2) + \ln(t+1)\nu(x_1,x_2),$$

where $\nu(x_1,x_2) = (e^{x_1}-1)(e^{x_1}-e)(e^{x_2}-1)(e^{x_2}-e)$. The exact solution of this optimal control problem is $y(x_1,x_2,t) = \ln(t+1)\ \nu(x_1,x_2)$ and $p(x_1,x_2,t) = (t-T)^2\ \nu(x_1,x_2)$. With the same solver setting as in Example 3, we report in Table 10 the approximation errors and convergence results of solving Example 4 by the proposed quasi-Newton iterations (4.8a) with the block diagonal α-circulant preconditioner $\mathcal{P}_\dagger(\alpha)$. Similar to the 1D case, we again observe clearly mesh-independent convergence rates of the outer quasi-Newton iterations, the LCP solver, and the inner GMRES solver, which numerically demonstrate the effectiveness of our proposed algorithm in higher spatial dimensions. In particular, the number of iterations is almost the same as in Table 9. The expected second-order approximation accuracy is also obviously achieved.

6. Conclusions. In this paper, we proposed a new parallel-in-time (PinT) block-circulant preconditioner for iteratively solving the discretized optimality system, which stems from the distributed optimal control of wave equations. Based on a novel unitary diagonalization of the preconditioner, its detailed parallel implementation steps are described. The eigenvalues of the preconditioned system are discussed in detail, which are shown to get closer to one as the regularization parameter becomes smaller. This indicates a faster convergence rate for smaller regularization parameters (of particular interest in practice), which was verified by numerical examples. In comparison with the recently developed matching Schur complement (MSC) preconditioner, our proposed PinT preconditioner shows more robust convergence rates with respect to both mesh step sizes and the regularization parameter. More importantly, the designed diagonalization provides a great potential to efficiently parallelize the computation of the preconditioning step in time, which is very desirable for large-scale practical applications. Our paper largely generalizes some known diagonalization techniques originally developed for solving time-dependent PDEs to the corresponding PDE-constrained optimal control problems. We are getting closer to solving the open problem posed in [37] to design a preconditioner that achieves a convergence rate independent of both spatial and temporal mesh step-sizes and the regularization parameter. Parallel numerical results for validating the practical parallel efficiency of our proposed PinT preconditioner are currently undertaken and will be reported elsewhere.

Application of the PinT preconditioner to the optimal control problem with a control constraint brings us several interesting topics for further research. First, it deserves a complete convergence analysis to get a general principle for choosing the variable change (4.5) and the Jacobian matrix $\mathcal{J}$ to achieve a rapid and robust convergence in the quasi-Newton iteration (4.8a). Second, it is worth explaining why the block-circulant preconditioner shows degraded convergence rates in the GMRES solver in the presence of nonsmooth data. It is also worth exploring the diagonalization of the generalized preconditioner $\mathcal{P}(\alpha)$ (cf. (5.1)) when $\alpha \in (0,1)$, since we find numerically that a smaller α indeed leads to more clustered eigenvalues. We believe

that more sophisticated problem formulations, such as different regularization norms or boundary control problems, could also be tackled using our PinT preconditioner upon appropriate modification.

Acknowledgments. The authors would like to thank the editor and two anonymous referees for their valuable comments and constructive revision suggestions, which have greatly improved the quality of this paper.

REFERENCES

[1] Z.-Z. Bai, *Block preconditioners for elliptic PDE-constrained optimization problems*, Computing, 91 (2010), pp. 379–395, https://doi.org/10.1007/s00607-010-0125-9.

[2] Z.-Z. Bai, M. Benzi, F. Chen, and Z.-Q. Wang, *Preconditioned MHSS iteration methods for a class of block two-by-two linear systems with applications to distributed control problems*, IMA J. Numer. Anal., 33 (2012), pp. 343–369, https://doi.org/10.1093/imanum/drs001.

[3] L. Banjai and D. Peterseim, *Parallel multistep methods for linear evolution problems*, IMA J. Numer. Anal., 32 (2012), pp. 1217–1240, https://doi.org/10.1093/imanum/drq040.

[4] D. Bini, *Parallel solution of certain Toeplitz linear systems*, SIAM J. Comput., 13 (1984), pp. 268–276, https://doi.org/10.1137/0213019.

[5] D. A. Bini, G. Latouche, and B. Meini, *Numerical Methods for Structured Markov Chains*, Oxford University Press, New York, 2005.

[6] A. Borzì and V. Schulz, *Multigrid methods for PDE optimization*, SIAM Rev., 51 (2009), pp. 361–395, https://doi.org/10.1137/060671590.

[7] R. Chan and X. Jin, *An Introduction to Iterative Toeplitz Solvers*, Fundam. Algorithms 5, SIAM, Philadelphia, 2007, https://doi.org/10.1137/1.9780898718850.

[8] R. H. Chan, J. G. Nagy, and R. J. Plemmons, *FFT-based preconditioners for Toeplitz-block least squares problems*, SIAM J. Numer. Anal., 30 (1993), pp. 1740–1768, https://doi.org/10.1137/0730089.

[9] R. W. Cottle, J.-S. Pang, and R. E. Stone, *The Linear Complementarity Problem*, Classics Appl. Math. 60, SIAM, Philadelphia, 2009, https://doi.org/10.1137/1.9780898719000.

[10] H. Elman, A. Ramage, and D. Silvester, *Algorithm* 866: *IFISS, a MATLAB toolbox for modelling incompressible flow*, ACM Trans. Math. Softw., 33 (2007), pp. 2–14, https://doi.org/10.1145/1236463.1236469.

[11] H. C. Elman, A. Ramage, and D. J. Silvester, *IFISS: A computational laboratory for investigating incompressible flow problems*, SIAM Rev., 56 (2014), pp. 261–273, https://doi.org/10.1137/120891393.

[12] H. C. Elman, D. J. Silvester, and A. J. Wathen, *Finite Elements and Fast Iterative Solvers: With Applications in Incompressible Fluid Dynamics*, 2nd ed., Oxford University Press, Oxford, UK, 2014.

[13] M. H. FarahI, J. E. Rubio, and D. A. Wilson, *The optimal control of the linear wave equation*, Internat. J. Control, 63 (1996), pp. 833–848, https://doi.org/10.1080/00207179608921871.

[14] A. Fischer, *A Newton-type method for positive-semidefinite linear complementarity problems*, J. Optim. Theory Appl., 86 (1995), pp. 585–608, https://doi.org/10.1007/bf02192160.

[15] M. J. Gander, 50 *years of time parallel time integration*, in Multiple Shooting and Time Domain Decomposition Methods, T. Carraro, M. Geiger, S. Körkel, and R. Rannacher, eds., Springer, Cham, 2015, pp. 69–113, https://doi.org/10.1007/978-3-319-23321-5_3.

[16] M. J. Gander, L. Halpern, J. Rannou, and J. Ryan, *A direct time parallel solver by diagonalization for the wave equation*, SIAM J. Sci. Comput., 41 (2019), pp. A220–A245, https://doi.org/10.1137/17m1148347.

[17] M. J. Gander and S.-L. Wu, *Convergence analysis of a periodic-like waveform relaxation method for initial-value problems via the diagonalization technique*, Numer. Math., 143 (2019), pp. 489–527, https://doi.org/10.1007/s00211-019-01060-8.

[18] M. Gerdts, G. Greif, and H. J. Pesch, *Numerical optimal control of the wave equation: Optimal boundary control of a string to rest in finite time*, Math. Comput. Simul., 79 (2008), pp. 1020–1032, https://doi.org/10.1016/j.matcom.2008.02.014.

[19] A. Goddard and A. Wathen, *A note on parallel preconditioning for all-at-once evolutionary PDEs*, Electron. Trans. Numer. Anal., 51 (2019), pp. 135–150, https://doi.org/10.1553/etna_vol51s135.

[20] G. Golub and C. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 2012.

[21] A. Greenbaum, V. Pták, and Z. Strakoš, *Any nonincreasing convergence curve is possible for GMRES*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 465–469, https://doi.org/10.1137/s0895479894275030.

[22] M. Gugat and V. Grimm, *Optimal boundary control of the wave equation with pointwise control constraints*, Comput. Optim. Appl., 49 (2011), pp. 123–147, https://doi.org/10.1007/s10589-009-9289-7.

[23] L. Han, M. K. Camlibel, J.-S. Pang, and W. M. H. Heemels, *A unified numerical scheme for linear-quadratic optimal control problems with joint control and state constraints*, Optim. Methods Softw., 27 (2012), pp. 761–799, https://doi.org/10.1080/10556788.2011.593624.

[24] G. Heidel and A. Wathen, *Preconditioning for boundary control problems in incompressible fluid dynamics*, Numer. Linear Algebra Appl., 26 (2018), e2218, https://doi.org/10.1002/nla.2218.

[25] H. V. Henderson and S. R. Searle, *The vec-permutation matrix, the vec operator and Kronecker products: A review*, Linear Multilinear Algebra, 9 (1981), pp. 271–288, https://doi.org/10.1080/03081088108817379.

[26] R. Herzog, J. W. Pearson, and M. Stoll, *Fast iterative solvers for an optimal transport problem*, Adv. Comput. Math., 45 (2018), pp. 495–517, https://doi.org/10.1007/s10444-018-9625-5.

[27] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2003), pp. 865–888, https://doi.org/10.1137/s1052623401383558.

[28] L. Hogben, *Handbook of Linear Algebra*, CRC Press, Boca Raton, FL, 2013.

[29] K. Ito and K. Kunisch, *Semi-smooth Newton methods for state-constrained optimal control problems*, Systems Control Lett., 50 (2003), pp. 221–228, https://doi.org/10.1016/s0167-6911(03)00156-7.

[30] X. Jin, *Developments and Applications of Block Toeplitz Iterative Solvers*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.

[31] A. Kröner, K. Kunisch, and B. Vexler, *Semismooth Newton methods for optimal control of the wave equation with control constraints*, SIAM J. Control Optim., 49 (2011), pp. 830–858, https://doi.org/10.1137/090766541.

[32] A. Kröner, *Numerical Methods for Control of Second Order Hyperbolic Equations*, Ph.D. thesis, Technische Universität München, Munich, Germany, 2011, https://mediatum.ub.tum.de/node?id=1084592.

[33] A. Kröner, K. Kunisch, and B. Vexler, *Semismooth Newton methods for optimal control of the wave equation with control constraints*, SIAM J. Control Optim., 49 (2011), pp. 830–858, https://doi.org/10.1137/090766541.

[34] K. Kunisch and D. Wachsmuth, *On time optimal control of the wave equation and its numerical realization as parametric optimization problem*, SIAM J. Control Optim., 51 (2013), pp. 1232–1262, https://doi.org/10.1137/120877520.

[35] B. Li, J. Liu, and M. Xiao, *A fast and stable preconditioned iterative method for optimal control problem of wave equations*, SIAM J. Sci. Comput., 37 (2015), pp. A2508–A2534, https://doi.org/10.1137/15m1020526.

[36] J.-L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer, Berlin, Heidelberg, 1971.

[37] J. Liu and J. W. Pearson, *Parameter-robust preconditioning for the optimal control of the wave equation*, Numer. Algorithms, 83 (2019), pp. 1171–1203, https://doi.org/10.1007/s11075-019-00720-y.

[38] X. Lu, H.-K. Pang, and H.-W. Sun, *Fast approximate inversion of a block triangular Toeplitz matrix with applications to fractional sub-diffusion equations*, Numer. Linear Algebra Appl., 22 (2015), pp. 866–882, https://doi.org/10.1002/nla.1972.

[39] X. Lu, H.-K. Pang, H.-W. Sun, and S.-W. Vong, *Approximate inversion method for time-fractional subdiffusion equations*, Numer. Linear Algebra Appl., 25 (2018), e2132, https://doi.org/10.1002/nla.2132.

[40] Y. Maday and E. M. Rønquist, *Parallelization in time through tensor-product space–time solvers*, C. R. Math. Acad. Sci. Paris, 346 (2008), pp. 113–118, https://doi.org/10.1016/j.crma.2007.09.012.

[41] E. McDonald, J. Pestana, and A. Wathen, *Preconditioning and iterative solution of all-at-once systems for evolutionary partial differential equations*, SIAM J. Sci. Comput., 40 (2018), pp. A1012–A1033, https://doi.org/10.1137/16m1062016.

[42] J. W. Pearson and M. Stoll, *Fast iterative solution of reaction-diffusion control problems arising from chemical processes*, SIAM J. Sci. Comput., 35 (2013), pp. B987–B1009, https://doi.org/10.1137/120892003.

[43] J. W. Pearson, M. Stoll, and A. J. Wathen, *Regularization-robust preconditioners for time-dependent PDE-constrained optimization problems*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 1126–1152, https://doi.org/10.1137/110847949.

[44] J. W. Pearson and A. J. Wathen, *A new approximation of the Schur complement in preconditioners for PDE-constrained optimization*, Numer. Linear Algebra Appl., 19 (2011), pp. 816–829, https://doi.org/10.1002/nla.814.

[45] J. W. Pearson and A. J. Wathen, *Fast iterative solvers for convection-diffusion control problems*, Electron. Trans. Numer. Anal, 40 (2013), pp. 294–310, http://etna.mcs.kent.edu/vol.40.2013/pp294-310.dir/pp294-310.pdf.

[46] J. Pestana, *Preconditioners for symmetrized Toeplitz and multilevel Toeplitz matrices*, SIAM J. Matrix Anal. Appl., 40 (2019), pp. 870–887, https://doi.org/10.1137/18m1205406.

[47] J. Pestana and A. J. Wathen, *A preconditioned MINRES method for nonsymmetric Toeplitz matrices*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 273–288, https://doi.org/10.1137/140974213.

[48] Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2003, https://doi.org/10.1137/1.9780898718003.

[49] Z. Sabeh, M. Shamsi, and I. M. Navon, *An indirect shooting method based on the POD/DEIM technique for distributed optimal control of the wave equation*, Internat. J. Numer. Methods Fluids, 86 (2017), pp. 433–453, https://doi.org/10.1002/fld.4460.

[50] J. Schöberl, R. Simon, and W. Zulehner, *A robust multigrid method for elliptic optimal control problems*, SIAM J. Numer. Anal., 49 (2011), pp. 1482–1503, https://doi.org/10.1137/100783285.

[51] D. Silvester, H. Elman, and A. Ramage, *Incompressible Flow and Iterative Solver Software (IFISS) Version* 3.6, February 2019, http://www.manchester.ac.uk/ifiss/.

[52] V. Simoncini and M. Benzi, *Spectral properties of the Hermitian and skew-Hermitian splitting preconditioner for saddle point problems*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 377–389, https://doi.org/10.1137/s0895479803434926.

[53] S.-L. Wu, *Toward parallel coarse grid correction for the parareal algorithm*, SIAM J. Sci. Comput., 40 (2018), pp. A1446–A1472, https://doi.org/10.1137/17m1141102.

[54] S.-L. Wu, H. Zhang, and T. Zhou, *Solving time-periodic fractional diffusion equations via diagonalization technique and multigrid*, Numer. Linear Algebra Appl., 25 (2018), e2178, https://doi.org/10.1002/nla.2178.

[55] S.-L. Wu and T. Zhou, *Acceleration of the two-level MGRIT algorithm via the diagonalization technique*, SIAM J. Sci. Comput., 41 (2019), pp. A3421–A3448, https://doi.org/10.1137/18m1207697.

[56] E. Zuazua, *Propagation, observation, and control of waves approximated by finite difference methods*, SIAM Rev., 47 (2005), pp. 197–243, https://doi.org/10.1137/s0036144503432862.