SIAM J. SCI. COMPUT.
Vol. 32, No. 3, pp. 1180–1200
© 2010 Society for Industrial and Applied Mathematics

# ANALYSIS OF BLOCK PARAREAL PRECONDITIONERS FOR PARABOLIC OPTIMAL CONTROL PROBLEMS*

TAREK P. MATHEW†, MARCUS SARKIS‡, AND CHRISTIAN E. SCHAERER§

**Abstract.** In this paper, we describe block matrix algorithms for the iterative solution of a *large-scale* linear-quadratic optimal control problem involving a parabolic partial differential equation over a finite control horizon. We consider an "all at once" discretization of the problem and formulate three iterative algorithms. The first algorithm is based on preconditioning a symmetric positive definite reduced linear system involving only the unknown control variables; however inner-outer iterations are required. The second algorithm modifies the first algorithm to avoid inner-outer iterations by introducing an auxiliary variable. It yields a symmetric indefinite system with a positive definite block preconditioner. The third algorithm is the central focus of this paper. It modifies the preconditioner in the second algorithm by a parallel-in-time preconditioner based on the parareal algorithm. Theoretical results show that the preconditioned algorithms have optimal convergence properties and parallel scalability. Numerical experiments confirm the theoretical results.

**Key words.** parabolic PDE, preconditioners, parareal, multiple shooting, control problems, Krylov methods

**AMS subject classifications.** 49J20, 65F10, 65M55

**DOI.** 10.1137/080717481

## 1. Introduction.

In this paper, we describe several block matrix algorithms for solving an "all at once" discretization of a parabolic optimal control problem on a convex polygonal domain $\Omega \subset \mathbb{R}^d$ over a time interval $(t_0, t_f)$. The state variable $z(t, .)$ satisfies a parabolic equation with input control $v(.)$ and initial data $z_0 \in L^2(\Omega)$:

$$\begin{cases} z_t + \mathcal{A}z = \mathcal{B}v \ \text{ for } \ t_0 < t < t_f, \\ \quad z(0) = z_0, \end{cases} \tag{1.1}$$

where $\mathcal{A}$ denotes an elliptic operator. The optimal control problem seeks $v(.)$ which minimizes a cost functional $J(z(v), v)$, which measures the deviation of the state variable $z(t, .)$ from a desired target $\tilde{y}$. Formally, $\mathcal{A}$ is a coercive map from the Hilbert space $L^2(t_0, t_f; Y)$ to $L^2(t_0, t_f; Y')$, where $Y = H_0^1(\Omega)$ and $Y' = H^{-1}(\Omega)$, i.e., the dual of $Y$ with respect to the pivot space $H = L^2(\Omega)$. Here, the state variable space is defined as $\mathcal{Y} = \{z \in L^2(t_0, t_f; Y) : z_t \in L^2(t_0, t_f; Y')\}$. It can be shown that $\mathcal{Y} \subset \mathcal{C}^0([t_0, t_f]; H)$; see [14]. The distributed control $v$ belongs to an admissible space $\mathcal{U} = L^2(t_0, t_f; U)$, where in our application $U = L^2(\Omega)$, and $\mathcal{B}$ is an operator in $\mathcal{L}(\mathcal{U}, L^2(t_0, t_f; H))$. It can be shown that the problem (1.1) is well-posed (see [14]), and we indicate the dependence of $z$ on $v \in \mathcal{U}$ using the notation $z(v)$. Given a target

*Received by the editors March 2, 2008; accepted for publication (in revised form) January 13, 2010; published electronically April 16, 2010. This research was supported partially by CNPq Brazil project 500075/2002-6 and grant 300964/2006-4.
http://www.siam.org/journals/sisc/32-3/71748.html
†Instituto de Matemática Pura e Aplicada - IMPA, E. D. Castorina 110, 22460-320, Rio de Janeiro, RJ, Brazil (tmathew@poonithara.org).
‡Instituto de Matemática Pura e Aplicada - IMPA, E. D. Castorina 110, 22460-320, Rio de Janeiro, RJ, Brazil and Department of Mathematical Sciences at Worcester Polytechnic Institute, Worcester, MA 01609 (msarkis@impa.br, msarkis@wpi.br).
§Instituto de Matemática Pura e Aplicada - IMPA, E. D. Castorina 110, 22460-320, Rio de Janeiro, RJ, Brazil and Universidad Nacional de Asunción, Politécnica, P.O. Box 2111 SL, Paraguay (cschaer@impa.br, cschaer@pol.una.py).



Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

function $\tilde{y}$ in $L^2(t_0, t_f; H)$ and parameters $q > 0$, $r > 0$, and $s \geq 0$, we shall employ the following cost function, which we associate with the state equation (1.1):

$$J(z(v), v) := \frac{q}{2} \int_{t_0}^{t_f} \|z(v)(t,\cdot) - \tilde{y}(t,\cdot)\|^2_{L^2(\Omega)}\, dt + \frac{r}{2} \int_{t_0}^{t_f} \|v(t,\cdot)\|^2_{L^2(\Omega)}\, dt + \frac{s}{2} \|z(v)(t_f,\cdot) - \tilde{y}(t_f,\cdot)\|^2_{L^2(\Omega)}. \tag{1.2}$$

For simplicity, we assume that $z_0 \in Y$ and $\tilde{y} \in L^2(t_0, t_f; Y)$, and we normalize $q = 1$. The optimal control problem for (1.1) seeks a control $u \in \mathcal{U}$ which *minimizes* the cost function (1.2):

$$J(y, u) \;:=\; \min_{v \in \mathcal{U}} J(z(v), v). \tag{1.3}$$

Since $0 < r$ (regularization parameter), $0 < q$, and $0 \leq s$, the optimal control problem (1.3) is well-posed (see [14]).

Our focus in this paper is on block matrix iterative algorithms for solving an "all at once" discretization [5, 11, 12, 17, 20] of the optimal control problem (1.3). The problem (1.3) seeks a control function $u$ (optimal forcing term in the sense of the cost function (1.2)) such that the solution $z(v)$ to the parabolic equation closely matches the given "tracking" function $\tilde{y}$ on the interval $(t_0, t_f)$. We also require that $z(v)(t_f, \cdot)$ be close to $\tilde{y}(t_f, \cdot)$ by introducing the term $\frac{s}{2} \|z(v)(t_f,\cdot) - \tilde{y}(t_f,\cdot)\|^2_{L^2(\Omega)}$ in the cost function. The constrained minimization problem (1.3) minimizes (1.2) subject to the constraint (1.1), where the *quadratic* cost functional (1.2) is a square norm of the difference between the solution to the parabolic equation and the tracking function with appropriate *regularization*, while the *linear* constraint (1.1) is a parabolic equation involving the state variable.

Our discussion is organized as follows. In section 2, we describe the finite-dimensional linear-quadratic optimal control problem obtained using an *all at once* discretization of the control problem, where the spatial discretization is obtained by the finite element method and the temporal discretization by the $\theta$-scheme. The discretization of the functional is obtained by the finite element method [5, 6, 18]. This transforms the optimal control problem into a large algebraic constrained minimization problem, where optimality conditions yield a large saddle point system involving the state variables $\mathbf{y}$, the control variables $\mathbf{u}$, and the Lagrange multipliers $\mathbf{p}$; see [2, 1, 3].

In section 3, we describe the algorithms. Using a reduction approach employed in [19, 17], we obtain a symmetric ill-conditioned positive definite reduced system for the unknown *control variables* (with low dimension in realistic situations). We refer to this algorithm as "the reduction to $\mathbf{u}$ algorithm," and a preconditioner is described. We prove that the rate of convergence is independent of the spatial discretization parameter $h$ if $h^4/r$ is not large. If the parameter $s = 0$, the rate of convergence depends only on the parameter $r$. If $s > 0$, we prove that the rate of convergence depends on the time discretization parameter $\tau$ and the parameter $r$ (see expression (3.41)). As a result, the preconditioned conjugate gradient method (PCG) can be used to solve (3.1), but inner-outer iterations are required. To overcome this drawback, we introduce an auxiliary variable resulting in a symmetric indefinite ill-conditioned system on the auxiliary and control variables. For this expanded system we employ a symmetric positive definite block diagonal preconditioner [22]. We also prove that under the same conditions of the first preconditioner ($s = 0$), this preconditioner is independent of the $h$ and $\tau$ parameters when minimal residual (MINRES) acceleration

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

is used. Results analogous to the first preconditioner are obtained if $s \neq 0$; see Theorem 4.7.

In section 4, we present a saddle point preconditioner based on the *parareal* algorithm [13, 16, 21, 8, 10, 23] and iterative shooting methods [12, 24, 4]. This preconditioner yields a rate of convergence independent of the mesh size $h$ with a dependence on $\tau$. If the parareal preconditioner is used as an approximate solver in the context of the reduction to $\mathbf{u}$ algorithm and enough number of parareal iterations is performed, then the algorithm depends on $\tau$ for $s = 0$ and $\tau$ and $r$ for $s \neq 0$. Finally, in section 5, numerical tests confirm the theoretical results and show that the parareal preconditioner yields scalability when the number of subdomains is increased. To avoid proliferation of constants, we sometimes use the notation $a \preceq b$ to represent the inequality $a \leq (constant)b$, and $a \asymp b$ if $a \preceq b$ and $b \preceq a$, where the $(constant)$ does not depend on $\tau$, $h$, $\Delta T$, $r$, and $s$.

**2. The discretized problem and the saddle point system.** To discretize the state equation (1.1) in space, we apply the finite element method to its weak formulation for each fixed $t \in (t_0, t_f)$. We choose a quasi-uniform triangulation $\mathcal{T}_h(\Omega)$ of $\Omega$, employ the $\mathcal{P}_1$ conforming finite element space $Y_h \subset Y$ for $z(t,\cdot)$ and the $\mathcal{P}_0$ finite element space $U_h \subset U$ for approximating $v(t,\cdot)$. Let $\{\phi_j\}_{j=1}^{\hat{q}}$ and $\{\psi_j\}_{j=1}^{\hat{p}}$ denote the standard basis functions for $Y_h$ and $U_h$, respectively. For convenience, we shall employ the same notation $z \in Y_h$ and $z \in \mathbb{R}^{\hat{q}}$, or $v \in U_h$ and $v \in \mathbb{R}^{\hat{p}}$, to denote both a finite element function in space and its corresponding vector representation, and to indicate their time dependence, we shall denote $\underline{z}$ and $\underline{v}$, respectively.

A discretization in space of the continuous time linear-quadratic optimal control problem will seek to minimize the following quadratic functional:

$$\begin{aligned} J_h(\underline{z},\underline{v}) := \frac{1}{2}\int_{t_0}^{t_f} (\underline{z}-\underline{\tilde{y}})^T(t) M_h (\underline{z}-\underline{\tilde{y}})(t)\,dt + \frac{r}{2}\int_{t_0}^{t_f} \underline{v}^T(t) R_h \underline{v}(t)\,dt \\ + \frac{s}{2}(\underline{z}(t_f)-\underline{\tilde{y}}(t_f))^T M_h (\underline{z}(t_f)-\underline{\tilde{y}}(t_f)) \end{aligned} \tag{2.1}$$

subject to the *constraint* that $\underline{z}$ satisfies the discrete equation of state

$$M_h \dot{\underline{z}} + A_h\,\underline{z} = B_h \underline{v} \quad \text{for} \quad t_0 < t < t_f, \quad \text{and} \quad \underline{z}(t_0) = z_0^h. \tag{2.2}$$

Here $(\underline{z}-\underline{\tilde{y}})(t)$ and $(\underline{z}(t_f)-\underline{\tilde{y}}(t_f))$ denote the tracking and final errors, respectively, and $\underline{\tilde{y}}(t)$, $z_0^h \in Y_h$ are approximations of $\underline{\tilde{y}}(t)$, $z_0$ (for instance, $L^2(\Omega)$-projections into $Y_h$). The matrices $M_h \in \mathbb{R}^{\hat{q}\times\hat{q}}$, $A_h \in \mathbb{R}^{\hat{q}\times\hat{q}}$, $B_h \in \mathbb{R}^{\hat{q}\times\hat{p}}$, and $R_h \in \mathbb{R}^{\hat{p}\times\hat{p}}$ have entries $(M_h)_{ij} := (\phi_i,\phi_j)$, $(A_h)_{ij} := (\phi_i, \mathcal{A}\phi_j)$, $(B_h)_{ij} := (\phi_i, \mathcal{B}\psi_j)$, and $(R_h)_{ij} := (\psi_i,\psi_j)$, where $(\cdot,\cdot)$ denotes the $L^2(\Omega)$ inner product.

To obtain a temporal discretization of (2.1) and (2.2), we partition $[t_0,\, t_f]$ into $\hat{l}$ equal subintervals with time step size $\tau = (t_f - t_0)/\hat{l}$. We denote $t_l = t_0 + l\,\tau$ for $0 \leq l \leq \hat{l}$. Associated with this partition, we assume that the state variable $\underline{z}$ is continuous in $[t_0, t_f]$ and linear in each subinterval $[t_{l-1}, t_l]$, $1 \leq l \leq \hat{l}$, with associated basis functions $\{\vartheta_l\}_{l=0}^{\hat{l}}$ so that if $z_l \in \mathbb{R}^{\hat{q}}$ is the nodal representation of $\underline{z}(t_l)$, we have $\underline{z}(t) = \sum_{l=0}^{\hat{l}} z_l \vartheta_l(t)$. The control variable $\underline{v}$ is assumed to be time discontinuous and constant in each subinterval $(t_{l-1}, t_l)$ with basis functions $\{\chi_l\}_{l=1}^{\hat{l}}$ so that if $v_l \in \mathbb{R}^{\hat{p}}$ is the nodal representation of $\underline{v}(t_l - (\tau/2))$, then $\underline{v}(t) = \sum_{l=1}^{\hat{l}} v_l \chi_l(t)$.

The temporal-spatial discretization of expression (2.1) can be expressed as

$$J_h^\tau(\mathbf{z},\mathbf{v}) = \frac{1}{2}\,(\mathbf{z}-\tilde{\mathbf{y}})^T \mathbf{K} (\mathbf{z}-\tilde{\mathbf{y}}) + \frac{1}{2}\,\mathbf{v}^T \mathbf{G} \mathbf{v} + (\mathbf{z}-\tilde{\mathbf{y}})^T \mathbf{g}, \tag{2.3}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

where the block vectors $\mathbf{z} := [z_1^T, \ldots, z_{\hat{l}}^T]^T \in \mathbb{R}^{\hat{l}\hat{q}}$ and $\mathbf{v} := [v_1^T, \ldots, v_{\hat{l}}^T]^T \in \mathbb{R}^{\hat{l}\hat{p}}$ denote the state and control variables, respectively, at all the discrete times (in our notation, the dimension $\hat{l}\hat{q}$ means a vector of $\hat{l}$ subvectors of size $\hat{q}$). Here, the discrete target is denoted as $\tilde{\mathbf{y}} := [\tilde{y}_1^T, \ldots, \tilde{y}_{\hat{l}}^T]^T \in \mathbb{R}^{\hat{l}\hat{q}}$ with target error $e_l = (z_l - \tilde{y}_l)$ for $0 \le l \le \hat{l}$. In (2.3), matrix $\mathbf{K} := \mathbf{Z} + \mathbf{\Gamma}$, where $\mathbf{Z}, \mathbf{\Gamma} \in \mathbb{R}^{(\hat{l}\hat{q})\times(\hat{l}\hat{q})}$ with $\mathbf{\Gamma} = s\,\mathrm{diag}(0, 0, \ldots, M_h)$ and $\mathbf{Z} = D_\tau \otimes M_h$, $D_\tau \in \mathbb{R}^{\hat{l}\times\hat{l}}$ with entries $(D_\tau)_{ij} := \int_{t_0}^{t_f} \vartheta_i(t)\vartheta_j(t)dt$, for $1 \le i, j \le \hat{l}$, where $\otimes$ stands for the Kronecker product. Furthermore, matrix $\mathbf{G} = r\tau I_{\hat{l}} \otimes R_h \in \mathbb{R}^{(\hat{l}\hat{p})\times(\hat{l}\hat{p})}$, where $I_{\hat{l}} \in \mathbb{R}^{\hat{l}\times\hat{l}}$ is an identity matrix, and vector $\mathbf{g} := [g_1, g_2, \ldots, g_{\hat{l}}]^T \otimes M_h e_0 \in \mathbb{R}^{\hat{l}\hat{q}}$, where $g_l := \int_{t_0}^{t_f} \vartheta_l(t)\vartheta_0(t)dt$. Here, note that $g_1 \neq 0$, and $e_0$ does not necessary vanish. In our notation, a matriz in $\mathbb{R}^{(n_1 n_2)\times(n_3 n_4)}$ has $n_1$ block rows (and $n_3$ block columns) where each block has dimension $n_2 \times n_4$.

Employing the $\theta$-scheme discretization in time, (2.2) takes the form

$$F_1 z_{l+1} = F_0\, z_l + \tau B_h v_l \quad \text{for} \quad 1 \le l \le \hat{l}, \quad \text{and} \quad z_0 = z_0^h, \tag{2.4}$$

where $F_0, F_1 \in \mathbb{R}^{\hat{q}\times\hat{q}}$ are (fixed) matrices given by $F_0 := M_h - (1-\theta)\tau A_h$ and $F_1 := M_h + \theta\tau A_h$. A full discretization in time of (2.4) will have the matrix form

$$\mathbf{E}\,\mathbf{z} + \mathbf{N}\,\mathbf{v} = \mathbf{f}, \tag{2.5}$$

where the input vector is $\mathbf{f} := [(F_0 z_0^h)^T, 0^T, \ldots, 0^T]^T \in \mathbb{R}^{\hat{l}\hat{q}}$ and the block lower bidiagonal matrix $\mathbf{E} \in \mathbb{R}^{(\hat{l}\hat{q})\times(\hat{l}\hat{q})}$ is given by

$$\mathbf{E} = \begin{bmatrix} F_1 & & & \\ -F_0 & F_1 & & \\ & \ddots & \ddots & \\ & & -F_0 & F_1 \end{bmatrix}. \tag{2.6}$$

The block diagonal matrix $\mathbf{N} \in \mathbb{R}^{(\hat{l}\hat{q})\times(\hat{l}\hat{p})}$ is given by $\mathbf{N} = -\tau I_{\hat{l}} \otimes B_h$. The discrete Lagrangian functional $\mathcal{L}_h^\tau(\mathbf{z}, \mathbf{v}, \mathbf{q})$ for minimizing (2.3) subject to constraint (2.5) can be expressed in terms of variables $\mathbf{z}$, $\mathbf{v}$, and $\mathbf{q}$ as

$$\mathcal{L}_h^\tau(\mathbf{z}, \mathbf{v}, \mathbf{q}) = J_h^\tau(\mathbf{z}, \mathbf{v}) + \mathbf{q}^T(\mathbf{E}\mathbf{z} + \mathbf{N}\mathbf{v} - \mathbf{f}). \tag{2.7}$$

To obtain the stationary point $(\mathbf{y}, \mathbf{u}, \mathbf{p})$ of (2.7) we apply the KKT optimality conditions for $\mathcal{L}_h^\tau(\cdot, \cdot, \cdot)$. This yields the symmetric indefinite linear system

$$\begin{bmatrix} \mathbf{K} & \mathbf{0} & \mathbf{E}^T \\ \mathbf{0} & \mathbf{G} & \mathbf{N}^T \\ \mathbf{E} & \mathbf{N} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{K}\tilde{\mathbf{y}} - \mathbf{g} \\ \mathbf{0} \\ \mathbf{f} \end{bmatrix}, \tag{2.8}$$

where $\tilde{\mathbf{y}} := [(\tilde{y}_1)^T, \ldots, (\tilde{y}_{\hat{l}})^T]^T \in \mathbb{R}^{\hat{l}\hat{q}}$.

**3. Block matrix algorithms.** We shall now describe and analyze several preconditioned block matrix algorithms for iteratively solving saddle point system (2.8).

**3.1. Reduction to u algorithm.** We shall now describe an algorithm for solving the saddle point system (2.8) based on the solution of a reduced Schur complement system involving the control variable $\mathbf{u}$. Solving the third and first block rows in (2.8)

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

yield $\mathbf{y} = -\mathbf{E}^{-1}\mathbf{N}\mathbf{u} + \mathbf{E}^{-1}\mathbf{f}$ and $\mathbf{p} = -\mathbf{E}^{-T}\mathbf{K}\mathbf{y} + \mathbf{E}^{-T}\mathbf{K}\tilde{\mathbf{y}} - \mathbf{E}^{-T}\mathbf{g}$, respectively. System (2.8) can then be reduced to the following Schur complement system for $\mathbf{u}$:

$$(\mathbf{G} + \mathbf{N}^T\mathbf{E}^{-T}\mathbf{K}\mathbf{E}^{-1}\mathbf{N})\mathbf{u} = \mathbf{b}, \tag{3.1}$$

where $\mathbf{b}$ is given by

$$\mathbf{b} := \mathbf{N}^T\mathbf{E}^{-T}\left(\mathbf{K}\mathbf{E}^{-1}\mathbf{f} - \mathbf{K}\tilde{\mathbf{y}} + \mathbf{g}\right). \tag{3.2}$$

*Algorithm* 1 consists of solving the Schur complement system (3.1) by the conjugate gradient algorithm using $\mathbf{G}$ as a *preconditioner*. This algorithm is simple to implement. However, it requires *inner* and *outer* iterations, where the inner iterations deal with applications of $\mathbf{E}^{-1}$ (and $\mathbf{E}^{-T}$) to a vector. In principle, these applications must be performed very accurately. The implementation of inexact inner iterations (see [25]) is under research and will not be considered here.

**3.2. Reduction to u and w algorithms.** To avoid the expensive computations in Algorithm 1 involving the action of $\mathbf{E}^{-1}$ and $\mathbf{E}^{-T}$ each iteration, we formulate a new saddle point system equivalent to (3.1) by introducing an auxiliary variable $\mathbf{w}$:

$$\mathbf{w} := -\mathbf{E}^{-T}\mathbf{K}\mathbf{E}^{-1}\mathbf{N}\mathbf{u}.$$

Then, the solution to system (3.1) can be obtained by solving the system

$$\mathbf{H}\begin{bmatrix} \mathbf{w} \\ \mathbf{u} \end{bmatrix} := \begin{bmatrix} \mathbf{E}\mathbf{K}^{-1}\mathbf{E}^T & \mathbf{N} \\ \mathbf{N}^T & -\mathbf{G} \end{bmatrix}\begin{bmatrix} \mathbf{w} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ -\mathbf{b} \end{bmatrix}, \tag{3.3}$$

which is symmetric and *indefinite*. The action of $\mathbf{E}^{-1}$ is required only once in the assembly of the right-hand side input vector $\mathbf{b}$.

*Algorithm* 2 consists of solving linear system (3.3) by the MINRES algorithm using the block diagonal positive definite *preconditioner*

$$\mathbf{P} = \begin{bmatrix} \mathbf{E}\mathbf{K}^{-1}\mathbf{E}^T & \\ & \mathbf{G} \end{bmatrix}. \tag{3.4}$$

*Algorithm* 3 consists of solving linear system (3.3) by the MINRES algorithm using the block diagonal positive definite *preconditioner*

$$\mathbf{P}_n = \begin{bmatrix} \mathbf{E}_n\widehat{\mathbf{K}}^{-1}\mathbf{E}_n^T & \\ & \mathbf{G} \end{bmatrix}, \tag{3.5}$$

where matrix $\mathbf{E}_n\widehat{\mathbf{K}}^{-1}\mathbf{E}_n^T$ is described in detail in section 4. It is based on applying twice (one for $\mathbf{E}_n$ and another for $\mathbf{E}_n^T$) $n$ parallel-in-time iterations of a parareal algorithm.

We now focus on lower and upper bounds for the eigenvalues of the preconditioned matrices associated with Algorithms 1 and 2; see Theorems 3.2 and 3.4. To achieve this goal, we first study the spectral properties of $\mathbf{E}\mathbf{M}^{-1}\mathbf{E}^T$, where the block diagonal matrix $\mathbf{M} \in \mathbb{R}^{(\hat{l}\hat{q})\times(\hat{l}\hat{q})}$ is defined as $\mathbf{M} := I_{\hat{l}} \otimes M_h$.

LEMMA 3.1. *Let $A_h$ and $M_h$ be the $\hat{q} \times \hat{q}$ symmetric positive definite matrices introduced in* (2.2). *Let $\lambda_q$ for $1 \le q \le \hat{q}$ denote the generalized eigenvalues of matrix $A_h$ with respect to $M_h$. Let $\mathbf{E}$ be the evolution matrix defined in* (2.6) *where the matrices $F_0$ and $F_1$ are defined by*

$$F_0 := M_h - (1-\theta)\tau A_h \ \text{ and } \ F_1 := M_h + \theta\tau A_h, \tag{3.6}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

*respectively. Then, for* $\theta \geq \frac{1}{2}$*, the scheme* (2.4) *will be stable for all* $\tau > 0$*, while for* $\theta < \frac{1}{2}$*, it will be stable only if* $\tau \leq 2/\left((1-2\theta)\lambda_{\max}\right)$*. Furthermore, if the scheme* (2.4) *is stable, the following bound will hold:*

$$\rho_{\min}(\mathbf{w}, \mathbf{M}\mathbf{w}) \leq (\mathbf{E}\mathbf{M}^{-1}\mathbf{E}^T\mathbf{w}, \mathbf{w}) \leq \rho_{\max}(\mathbf{w}, \mathbf{M}\mathbf{w}), \tag{3.7}$$

*where*

$$\rho_{\min} := \min\left\{(\tau\lambda_{\min})^2, (2 + (2\theta - 1)\tau\lambda_{\max})^2\right\} \tag{3.8}$$

*and*

$$\rho_{\max} := 4\left(1 + \theta\tau\lambda_{\max}\right)^2, \tag{3.9}$$

*where* $\lambda_{\max} := \max_q \lambda_q$ *and* $\lambda_{\min} := \min_q \lambda_q$*.*

*Proof. Part* 1 (Stability condition). Consider the $\theta$-scheme for (2.2):

$$z_{l+1} = \Phi\, z_l + \tau\, F_1^{-1}\, B_h\, v_l, \tag{3.10}$$

where $\Phi \in \mathbb{R}^{\hat{q}\times\hat{q}}$ is the marching matrix given by

$$\Phi := (M_h + \theta\tau A_h)^{-1}(M_h - (1-\theta)\tau A_h). \tag{3.11}$$

Consider $X_h := [x_1, \ldots, x_{\hat{q}}]$ and $\Lambda_h := \mathrm{diag}\{\lambda_1, \ldots, \lambda_{\hat{q}}\}$ as the generalized eigenvectors and eigenvalues of $A_h$ with respect to $M_h$, i.e., $A_h X_h = M_h X_h \Lambda_h$, with $X_h^T M_h X_h = I$. Then it is easy to see that the stability condition for (3.10), i.e., $(\Phi z, M_h z) \leq (z, M_h z)$ for any $z \in \mathbb{R}^{\hat{q}}$, is given by

$$|\,(1 + \theta\tau\lambda_q)^{-1}(1 - (1-\theta)\tau\lambda_q)\,| \leq 1 \ \text{ for } q \in \{1, \ldots, \hat{q}\} \tag{3.12}$$

or equivalently

$$\left\{\begin{array}{rcl} 1 - (1-\theta)\tau\lambda_q & \leq & 1 + \theta\tau\lambda_q \\ -1 + (1-\theta)\tau\lambda_q & \leq & 1 + \theta\tau\lambda_q \end{array}\right\} \text{ for } q \in \{1, \ldots, \hat{q}\}. \tag{3.13}$$

From (3.13), we obtain $0 \leq \tau\lambda_q$ and $\tau(1-2\theta)\lambda_q \leq 2$ since $0 < \lambda_q$. In the case $\theta \geq 1/2$, there is no restriction on $\tau$; consequently the marching scheme will be unconditionally stable. On the other hand, if $\theta < 1/2$, then $0 < (1-2\theta)$, and in order for the scheme to be stable it is necessary that $\tau \leq 2/\left((1-2\theta)\lambda_{\max}\right)$, where $\lambda_{\max} = \max_q \lambda_q$. In this case, the marching scheme is conditionally stable.

*Part* 2 (Condition number of $\mathbf{E}\mathbf{M}^{-1}\mathbf{E}^T$). We diagonalize the blocks of $\mathbf{E}\mathbf{M}^{-1}\mathbf{E}^T$:

$$\mathbf{E}\mathbf{M}^{-1}\mathbf{E}^T = \begin{bmatrix} F_1 M_h^{-1} F_1^T & -F_1 M_h^{-1} F_0^T & & \\ -F_0 M_h^{-1} F_1^T & F_0 M_h^{-1} F_0^T + F_1 M_h^{-1} F_1^T & & \ddots \\ \ddots & \ddots & & -F_1 M_h^{-1} F_0^T \\ & -F_0 M_h^{-1} F_1^T & & F_0 M_h^{-1} F_0^T + F_1 M_h^{-1} F_1^T \end{bmatrix}. \tag{3.14}$$

Notice that $X_h^T M_h X_h = I$ and $X_h^T A_h X_h = \Lambda_h$ imply that $F_0$ and $F_1$ are diagonalized by $X_h$, yielding $\Lambda_0 = X_h^T F_0 X_h = X_h^T (M_h - (1-\theta)\tau A_h) X_h$ and matrix

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$\Lambda_1 = X_h^T F_1 X_h = X_h^T (M_h + \theta\tau A_h) X_h$. If $\mathbf{X} := \text{blockdiag}(X_h, \ldots, X_h)$, then the block matrix $\mathbf{X}^T\mathbf{E}\mathbf{M}^{-1}\mathbf{E}^T\mathbf{X}$ will have blocks which are diagonal matrices:

$$\mathbf{X}^T\mathbf{E}\mathbf{M}^{-1}\mathbf{E}^T\mathbf{X} = \begin{bmatrix} \Lambda_1{}^2 & -\Lambda_1\Lambda_0 & & & \\ -\Lambda_0\Lambda_1 & \Lambda_0^2+\Lambda_1^2 & \ddots & & \\ & -\Lambda_0\Lambda_1 & \Lambda_0^2+\Lambda_1^2 & -\Lambda_1\Lambda_0 & \\ & & \ddots & \ddots & -\Lambda_1\Lambda_0 \\ & & & -\Lambda_0\Lambda_1 & \Lambda_0^2+\Lambda_1^2 \end{bmatrix}, \tag{3.15}$$

where $\mathbf{X}^T\mathbf{E}\mathbf{M}^{-1}\mathbf{E}^T\mathbf{X} \in \mathbb{R}^{(\hat{l}\hat{q})\times(\hat{l}\hat{q})}$. Next, we permute the rows and columns of the block tridiagonal matrix (3.15) using a permutation matrix $\mathbf{\Pi}$. This yields the matrix $\mathbf{\Theta} := \mathbf{\Pi}(\mathbf{X}^T\mathbf{E}\mathbf{M}^{-1}\mathbf{E}^T\mathbf{X})\mathbf{\Pi}^T = \text{blockdiag}(\Theta_1, \ldots, \Theta_{\hat{q}})$, where each block submatrix $\Theta_q$ is a tridiagonal matrix with entries

$$\Theta_q := \begin{bmatrix} a_q^2 & -a_qb_q & & & \\ -a_qb_q & a_q^2+b_q^2 & -a_qb_q & & \\ & -a_qb_q & a_q^2+b_q^2 & -a_qb_q & \\ & & \ddots & \ddots & \ddots \\ & & & -a_qb_q & a_q^2+b_q^2 \end{bmatrix}, \tag{3.16}$$

where $b_q := (1-(1-\theta)\tau\lambda_q)$ and $a_q := (1+\theta\tau\lambda_q)$. Let $\mu(\Theta_q)$ denote an eigenvalue of submatrix $\Theta_q$ (and hence also of $\mathbf{\Theta}$). Then, Gershgorin's theorem [7, 26] yields

$$|\,\mu(\Theta_q) - a_q^2\,| \le |\,a_qb_q\,| \quad \text{or} \quad |\,\mu(\Theta_q) - a_q^2 - b_q^2\,| \le 2\,|\,a_qb_q\,|\,. \tag{3.17}$$

Using condition (3.12), we guarantee stability when $|\,b_q\,| \le |\,a_q\,|$, obtaining

$$\mu(\Theta_q) \le \max\Big(|\,a_q\,|\,(|\,a_q\,| + |\,b_q\,|)\,, (|\,a_q\,| + |\,b_q\,|)^2\Big) \le \max\ 4\,|\,a_q\,|^2 \tag{3.18}$$

and

$$\mu(\Theta_q) \ge \min\Big(\big(|\,a_q\,|^2 - |\,a_q\,||\,b_q\,|\big)\,, (|\,a_q\,| - |\,b_q\,|)^2\Big) \ge \min\,(|\,a_q\,| - |\,b_q\,|)^2\,. \tag{3.19}$$

An upper bound for $\mu(\Theta_q)$ in (3.18) is $\mu(\Theta_q) \le 4\,(1+\theta\tau\lambda_{\max})^2$, where $\lambda_{\max} = \max_q \lambda_q$. To obtain a lower bound for $\mu(\Theta_q)$, using (3.19) we employ the notation $\lambda_{\min} = \min_q \lambda_q$, obtaining $\mu(\Theta_q) \ge (\tau\lambda_{\min})^2$ or $\mu(\Theta_q) \ge (2+(2\theta-1)\tau\lambda_{\max})^2$. The estimate (3.7) then follows by using $(\mathbf{X}^{-T}\mathbf{\Pi}^{-1})(\mathbf{\Pi}^{-T}\mathbf{X}^{-1}) = \mathbf{M}$. ∎

The next theorem establishes bounds for the condition number of Algorithm 1, i.e., the spectrum of $\mathbf{G}^{-1}(\mathbf{G}+\mathbf{N}^T\mathbf{E}^{-T}\mathbf{K}\mathbf{E}^{-1}\mathbf{N})$. It uses the following straightforward properties of the matrices $\mathbf{Z}, \mathbf{G}, \mathbf{N}, \mathbf{M}$, and $\mathbf{\Gamma}$:

$$c_1\,\tau\,\mathbf{y}^T\mathbf{M}\mathbf{y} \le \mathbf{y}^T\mathbf{Z}\mathbf{y} \le c_2\,\tau\,\mathbf{y}^T\mathbf{M}\mathbf{y}, \tag{3.20}$$

$$c_3\,r\,\tau\,h^d\mathbf{u}^T\mathbf{u} \le \mathbf{u}^T\mathbf{G}\mathbf{u} \le c_4\,r\,\tau\,h^d\mathbf{u}^T\mathbf{u}, \tag{3.21}$$

$$c_5\,\tau^2\,h^{2d}\,\mathbf{u}^T\mathbf{u} \le \mathbf{u}^T\mathbf{N}^{\mathbf{T}}\mathbf{N}\mathbf{u} \le c_6\,\tau^2\,h^{2d}\,\mathbf{u}^T\mathbf{u}, \tag{3.22}$$

$$c_7\,h^d\,\mathbf{y}^T\mathbf{y} \le \mathbf{y}^T\mathbf{M}\mathbf{y} \le c_8\,h^d\,\mathbf{y}^T\mathbf{y}, \tag{3.23}$$

$$0 \le \mathbf{y}^T\mathbf{\Gamma}\mathbf{y} \le c_9\,s\,\mathbf{y}^T\mathbf{M}\mathbf{y}, \tag{3.24}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

where the $c_i$ are positive constants independent of $r$, $s$, $h$, and $\tau$. We note, however, for problems where the space of the control variables $\mathbf{u}$ is not dense enough within the space of the Lagrange multipliers variables $\mathbf{p}$, the constant $c_5$ may vanish.

THEOREM 3.2. *Let the matrices* $\mathbf{Z}$, $\mathbf{G}$, $\mathbf{N}$, *and* $\mathbf{M}$ *and* $\mathbf{\Gamma}$ *satisfy the bounds described in* (3.20)–(3.24). *Let* $\mathbf{K} = \mathbf{Z} + \mathbf{\Gamma}$. *Let the matrix* $\mathbf{E}$ *be defined by* (2.6). *Let* $\rho_{\max}$ *and* $\rho_{\min}$ *be as described in Lemma* 3.1. *Then, there exists* $\mu_{\min} > 0$ *and* $\mu_{\max} > 0$, *such that*

$$(1+\mu_{\min})\,\mathbf{u}^T\mathbf{G}\mathbf{u} \le \mathbf{u}^T(\mathbf{G}+\mathbf{N}^T\mathbf{E}^{-T}\mathbf{K}\mathbf{E}^{-1}\mathbf{N})\mathbf{u} \le (1+\mu_{\max})\,\mathbf{u}^T\mathbf{G}\mathbf{u}, \tag{3.25}$$

*where* $\mu_{\max} = \frac{(c_2\,\tau + c_9\,s)\,c_6\,\tau}{\rho_{\min}\,c_3 c_7\,r}$ *and* $\mu_{\min} = \frac{c_1\,c_5\,\tau^2}{\rho_{\max}\,c_4 c_8\,r}$.

*Proof.* Using the bounds (3.7) and (3.20)–(3.24) we obtain the following.

*Upper bound.*

$$\mathbf{u}^T\mathbf{N}^T\mathbf{E}^{-T}\mathbf{K}\mathbf{E}^{-1}\mathbf{N}\mathbf{u} \le (c_2\,\tau + c_9\,s)\,\mathbf{u}^T\mathbf{N}^T\mathbf{E}^{-T}\mathbf{M}\mathbf{E}^{-1}\mathbf{N}\mathbf{u} \tag{3.26}$$

$$\le \frac{(c_2\,\tau + c_9\,s)}{\rho_{\min}}\,\mathbf{u}^T\mathbf{N}^T\mathbf{M}^{-1}\mathbf{N}\mathbf{u} \tag{3.27}$$

$$\le \frac{(c_2\,\tau + c_9\,s)}{\rho_{\min}\,c_7 h^d}\,\mathbf{u}^T\mathbf{N}^T\mathbf{N}\mathbf{u} \tag{3.28}$$

$$\le \frac{(c_2\,\tau + c_9\,s)\,c_6\,\tau^2\,h^d}{\rho_{\min}\,c_7}\,\mathbf{u}^T\mathbf{u} \tag{3.29}$$

$$\le \frac{(c_2\,\tau + c_9\,s)\,c_6\,\tau}{\rho_{\min}\,c_3\,c_7\,r}\,\mathbf{u}^T\mathbf{G}\mathbf{u} \tag{3.30}$$

$$= \mu_{\max}\,\mathbf{u}^T\mathbf{G}\mathbf{u}; \tag{3.31}$$

therefore,

$$\mathbf{u}^T(G+\mathbf{N}^T\mathbf{E}^{-T}\mathbf{K}\mathbf{E}^{-1}\mathbf{N})\mathbf{u} \le (1+\mu_{\max})\,\mathbf{u}^T\mathbf{G}\mathbf{u}. \tag{3.32}$$

*Lower bound.*

$$\mathbf{u}^T\mathbf{N}^T\mathbf{E}^{-T}\mathbf{K}\mathbf{E}^{-1}\mathbf{N}\mathbf{u} \ge (c_1\,\tau)\,\mathbf{u}^T\mathbf{N}^T\mathbf{E}^{-T}\mathbf{M}\mathbf{E}^{-1}\mathbf{N}\mathbf{u} \tag{3.33}$$

$$\ge \frac{c_1\,\tau}{\rho_{\max}}\,\mathbf{u}^T\mathbf{N}^T\mathbf{M}^{-1}\mathbf{N}\mathbf{u} \tag{3.34}$$

$$\ge \frac{c_1\,\tau}{\rho_{\max}\,c_8 h^d}\,\mathbf{u}^T\mathbf{N}^T\mathbf{N}\mathbf{u} \tag{3.35}$$

$$\ge \frac{c_1\,c_5\,\tau^3\,h^d}{\rho_{\max}\,c_8}\,\mathbf{u}^T\mathbf{u} \tag{3.36}$$

$$\ge \frac{c_1\,c_5\,\tau^2}{\rho_{\max}\,c_4 c_8\,r}\,\mathbf{y}^T\mathbf{G}\mathbf{y} \tag{3.37}$$

$$= \mu_{\min}\,\mathbf{u}^T\mathbf{G}\mathbf{u}; \tag{3.38}$$

therefore,

$$\mathbf{u}^T(G+\mathbf{N}^T\mathbf{E}^{-T}\mathbf{K}\mathbf{E}^{-1}\mathbf{N})\mathbf{u} \ge (1+\mu_{\min})\,\mathbf{u}^T\mathbf{G}\mathbf{u}. \tag{3.39}$$

This yields the desired result. ∎

*Remark* 3.3. Henceforth, we consider only the choices $\theta = 1$ or $\theta = \frac{1}{2}$. For finite element discretizations on a polygonal domain of size $O(1)$ with quasi-uniform

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

meshes, the generalized eigenvalues $\lambda_q$ of $A_h$ with respect to $M_h$ satisfy the bounds $\alpha_1 \le \lambda_q \le \alpha_2 h^{-2}$, and therefore, for $\theta \ge 1/2$ and $h^2 \preceq \tau \preceq 1$, it follows that

$$\mu_{\min} \asymp \frac{h^4}{r} \text{ and } \mu_{\max} \asymp \frac{1+\frac{s}{\tau}}{r}. \tag{3.40}$$

Hence, from (3.25), it follows that the condition number of Algorithm 1 satisfies

$$\operatorname{cond}\left(\mathbf{G}^{-1}\left(\mathbf{G}+\mathbf{N}^T\mathbf{E}^{-T}\mathbf{K}\mathbf{E}^{-1}\mathbf{N}\right)\right) \preceq \left(1+\frac{1+\frac{s}{\tau}}{r}\right) \Big/ \left(1+\frac{h^4}{r}\right). \tag{3.41}$$

The next theorem establishes bounds for the condition number of Algorithm 2.

THEOREM 3.4. *Let matrices* $\mathbf{Z}$, $\mathbf{G}$, $\mathbf{N}$, $\mathbf{M}$, *and* $\mathbf{\Gamma}$ *satisfy* (3.20)–(3.24). *Define* $\mathbf{P} := \mathit{blockdiag}(\mathbf{E}\mathbf{K}^{-1}\mathbf{E}^T, \mathbf{G})$ *as a block diagonal preconditioner for the coefficient matrix* $\mathbf{H}$ *of system* (3.3). *Then the condition number of the preconditioned system* $\mathbf{P}^{-1}\mathbf{H}$ *satisfies the bound*

$$\mathit{cond}(\mathbf{P}^{-1}\mathbf{H}) \le \left(\frac{1+\mu_{\max}}{1+\mu_{\min}}\right)^{1/2}, \tag{3.42}$$

*where* $\mu_{\max}$ *and* $\mu_{\min}$ *are as defined in Theorem* 3.2. *Additionally, if the assumptions of Remark* 3.3 *hold, then*

$$\mathit{cond}(\mathbf{P}^{-1}\mathbf{H}) \preceq \left(1+\frac{1+\frac{s}{\tau}}{r}\right)^{1/2} \Big/ \left(1+\frac{h^4}{r}\right)^{1/2}. \tag{3.43}$$

*Proof.* Since $\mathbf{P}$ is positive definite, we consider the generalized eigenvalue problem

$$\begin{bmatrix} \mathbf{E}\mathbf{K}^{-1}\mathbf{E}^T & \mathbf{N} \\ \mathbf{N}^T & -\mathbf{G} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{u} \end{bmatrix} = \eta \begin{bmatrix} \mathbf{E}\mathbf{K}^{-1}\mathbf{E}^T & \\ & \mathbf{G} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{u} \end{bmatrix}. \tag{3.44}$$

We obtain the equations

$$(\eta-1)\mathbf{E}\mathbf{K}^{-1}\mathbf{E}^T\mathbf{w} = \mathbf{N}\mathbf{u} \text{ and } (\eta+1)\mathbf{G}\mathbf{u} = \mathbf{N}^T\mathbf{w}. \tag{3.45}$$

These equations yield $(\mathbf{G}+\mathbf{N}^T\mathbf{E}^{-T}\mathbf{K}\mathbf{E}^{-1}\mathbf{N})\mathbf{u} = \eta^2\mathbf{G}\mathbf{u}$, where $\eta^2$ are the generalized eigenvalues of $\mathbf{G}+\mathbf{N}^T\mathbf{E}^{-T}\mathbf{K}\mathbf{E}^{-1}\mathbf{N}$ with respect to $\mathbf{G}$. Using Theorem 3.2 we obtain bounds for $\eta$ as follows:

$$\max \eta \le (1+\mu_{\max})^{1/2}, \tag{3.46}$$

$$\min \eta \ge (1+\mu_{\min})^{1/2}. \tag{3.47}$$

The desired result now follows, since

$$\operatorname{cond}(\mathbf{P}^{-1}\mathbf{H}) \le \frac{\max|\eta|}{\min|\eta|}. \tag{3.48}$$

This yields the desired result. ∎

**4. Parareal approximation $\mathbf{E}_n^{-T}\widehat{\mathbf{K}}\mathbf{E}_n^{-1}$.** The parareal method is a parallel iterative method for solving an evolution equation based on a *decomposition* of its *temporal* domain $[t_0, t_f]$ into $\hat{k}$ *coarse* subintervals of length $\Delta T = (t_f - t_0)/\hat{k}$, setting $T_0 = t_0$ and $T_k = t_0 + k\Delta T$ for $1 \le k \le \hat{k}$; see [13]. It determines the solution at

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

the times $T_k$ for $1 \le k \le \hat{k}$ by using a *multiple-shooting* technique which requires solving the parabolic equation on each interval $(T_{k-1}, T_k)$ in *parallel*. To speed up the multiple-shooting iteration, the residual equations are "preconditioned" by solving a "coarse" time-grid discretization of the parabolic equation using the time step $\Delta T$.

In this section we formulate a preconditioner $\mathbf{E}_n$ for $\mathbf{E}$ based on $n$ Richardson iterations of the parareal algorithm. Using $\mathbf{E}_n$, an application of $\mathbf{E}_n^{-T}\widehat{\mathbf{K}}\mathbf{E}_n^{-1}$ to a vector $\mathbf{s} = [s_1^T, \ldots, s_{\hat{l}}^T]^T \in \mathbb{R}^{\hat{l}\hat{q}}$ can be computed in three steps: *step* 1, apply $\mathbf{E}_n^{-1}\mathbf{s} :\to \hat{\mathbf{z}}^n$ using $n$ applications of the parareal method (described below); *step* 2, multiply $\widehat{\mathbf{K}}\hat{\mathbf{z}}^n :\to \hat{\mathbf{t}}^n$ (see below); *step* 3, apply $\mathbf{E}_n^{-T}\hat{\mathbf{t}}^n :\to \mathbf{w}^n$, i.e., the transpose of *step* 1.

Let $\hat{m} = (T_k - T_{k-1})/\tau$ and $j_{k-1} = \frac{T_{k-1}-T_0}{\tau}$. Consider the solution $Z_k$ at $T_k$ defined by marching from $T_{k-1}$ to $T_k$ the $\theta$-scheme on the fine triangulation $\tau$ with an initial data $Z_{k-1}$ at $T_{k-1}$ with forcing term $[s_{j_{k-1}+1}^T, \ldots, s_{j_{k-1}+\hat{m}}^T]^T$. It is easy to see that

$$F_1\, Z_k = F_0^{\Delta}\, Z_{k-1} + S_k, \tag{4.1}$$

where $F_0^{\Delta} := (F_0 F_1^{-1})^{\hat{m}-1} F_0 \in \mathbb{R}^{\hat{q}\times\hat{q}}$, $S_k := \sum_{m=1}^{\hat{m}} \left(F_0 F_1^{-1}\right)^{\hat{m}-m} s_{j_{k-1}+m}$, $Z_0 = 0$, and $F_0$ and $F_1$ were introduced in (3.6). Imposing continuity $F_1\, Z_k - F_0^{\Delta}\, Z_{k-1} - S_k = 0$ at times $T_k$, for $1 \le k \le \hat{k}$, yields

$$\mathbf{C}\,\mathbf{Z} := \begin{bmatrix} F_1 & & & \\ -F_0^{\Delta} & F_1 & & \\ & \ddots & \ddots & \\ & & -F_0^{\Delta} & F_1 \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_{\hat{k}} \end{bmatrix} = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_{\hat{k}} \end{bmatrix} := \mathbf{S}. \tag{4.2}$$

In this paper we consider only the case where the coarse solution at $T_k$ with initial data $Z_{k-1} \in \mathbb{R}^{\hat{q}}$ at $T_{k-1}$ is obtained by applying one coarse time step of the backward Euler method $G_1 Z_k = G_0\, Z_{k-1}$, where matrix $G_1 := (M_h + A_h \Delta T)$ and $G_0 := M_h \in \mathbb{R}^{\hat{q}\times\hat{q}}$. We note, however, by using results from [10], the techniques developed here can be straightforwardly extended to the case where the backward Euler method on the coarse propagators is replaced by the implicit singly diagonally implicit Runge–Kutta or by the Radau IIA methods.

In the parareal algorithm, the following coarse propagator based on $G_0$ and $G_1$ is employed to precondition system (4.2) via

$$\begin{bmatrix} Z_1^{i+1} \\ Z_2^{i+1} \\ \vdots \\ Z_{\hat{k}}^{n+1} \end{bmatrix} = \begin{bmatrix} Z_1^{i} \\ Z_2^{i} \\ \vdots \\ Z_{\hat{k}}^{i} \end{bmatrix} + \left( \begin{bmatrix} G_1 & & & \\ -G_0 & G_1 & & \\ & \ddots & \ddots & \\ & & -G_0 & G_1 \end{bmatrix} \right)^{-1} \begin{bmatrix} R_1^{i} \\ R_2^{i} \\ \vdots \\ R_{\hat{k}}^{i} \end{bmatrix} \tag{4.3}$$

for $0 \le i \le n-1$, where the residual $\mathbf{R}^i := [{R_1^i}^T, \ldots, {R_{\hat{k}}^i}^T]^T \in \mathbb{R}^{\hat{k}\hat{q}}$ in (4.2) is defined as $\mathbf{R}^i := \mathbf{S} - \mathbf{C}\,\mathbf{Z}^i$, where $\mathbf{Z}^i := [{Z_1^i}^T, \ldots, {Z_{\hat{k}}^i}^T]^T \in \mathbb{R}^{\hat{k}\hat{q}}$, and $\mathbf{Z}^0 := [0^T, \ldots, 0^T]^T$.

We now define $\hat{\mathbf{z}}^n := \mathbf{E}_n^{-1}\mathbf{s}$. Let $\hat{\mathbf{z}}^n$ be the nodal representation of a piecewise linear function $\underline{\hat{z}}^n$ in time with respect to the fine triangulation $\tau$ on $[t_0, t_f]$ and continuous inside each coarse subinterval $[T_{k-1}, T_k]$; i.e., the function $\underline{\hat{z}}^n$ can be discontinuous across the coarse points $T_k$, $1 \le k \le \hat{k}-1$; therefore, $\hat{\mathbf{z}}^n \in \mathbb{R}^{(\hat{l}+\hat{k}-1)\hat{q}}$. On each subinterval $[T_{k-1}, T_k]$, $\underline{\hat{z}}^n$ is defined marching from $T_{k-1}$ to $T_k$ using the $\theta$-scheme with fine time steps $\tau$ and initial data $Z_{k-1}^n$ at $T_{k-1}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

We define the matrix $\widehat{\mathbf{K}} := \widehat{\mathbf{Z}} + \widehat{\mathbf{\Gamma}}$ with $\widehat{\mathbf{Z}}, \widehat{\mathbf{\Gamma}} \in \mathbb{R}^{((\hat{l}+\hat{k}-1)\hat{q})\times((\hat{l}+\hat{k}-1)\hat{q})}$, where $\widehat{\mathbf{\Gamma}} = s\,\mathrm{diag}(0,0,\ldots,M_h)$. Here, $\widehat{\mathbf{Z}} = \hat{D}_\tau \otimes M_h$, $\hat{D}_\tau := \mathrm{blockdiag}(\hat{D}_\tau^1,\ldots,\hat{D}_\tau^{\hat{k}})$, and the $\hat{D}_\tau^1 \in \mathbb{R}^{(\hat{m}\hat{q})\times(\hat{m}\hat{q})}$ and $\hat{D}_\tau^k \in \mathbb{R}^{((\hat{m}+1)\hat{q})\times((\hat{m}+1)\hat{q})}$ for $2 \le k \le \hat{k}$ are the time mass matrices associated to the subintervals $[T_{k-1}, T_k]$. Note that $\mathbf{K}$ can be obtained by assembling $\widehat{\mathbf{K}}$, and the equality $(\hat{l}+\hat{k}-1)\hat{q} = \hat{m}\hat{q} + (\hat{k}-1)(\hat{m}+1)\hat{q}$ holds.

*Remark* 4.1. In the following, we shall express using matrix notation, the parareal algorithm described in the preceding text. For convenience, we consider only *unique values* for the solution at the coarse times $T_k$, although the jumps can be computed using the evolution matrix. In matrix form, the parareal method to solve $\mathbf{E}\,\mathbf{z} = \mathbf{s}$ corresponds to a Schur complement-based Richardson iteration. We partition $\mathbf{z} = \left(\mathbf{z}_I^T, \mathbf{z}_B^T\right)^T$, where $\mathbf{z}_I = (\mathbf{z}_I^{(1)^T},\ldots,\mathbf{z}_I^{(\hat{k})^T})^T$ and $\mathbf{z}_I^{(k)} = (z_{(k-1)\hat{m}+1}^T,\ldots,z_{k\hat{m}-1}^T)^T$ are subvectors of $\mathbf{z}$ at the times $t_j = j\,\tau$ in $(T_{k-1}, T_k)$, while $\mathbf{z}_B = (z_{\hat{m}}^T,\ldots,z_{\hat{k}\hat{m}}^T)^T$ denotes subvectors of $\mathbf{z}$ at the times $T_1,\ldots,T_{\hat{k}}$. This block partitions system $\mathbf{E}\,\mathbf{z} = \mathbf{s}$:

$$\begin{bmatrix} \mathbf{E}_{II} & \mathbf{E}_{IB} \\ \mathbf{E}_{BI} & \mathbf{E}_{BB} \end{bmatrix} \begin{bmatrix} \mathbf{z}_I \\ \mathbf{z}_B \end{bmatrix} = \begin{bmatrix} \mathbf{s}_I \\ \mathbf{s}_B \end{bmatrix}, \tag{4.4}$$

where $\mathbf{E}_{II} = \mathrm{blockdiag}(\mathbf{E}_{II}^{(1)},\ldots,\mathbf{E}_{II}^{(\hat{k})})$ is a block diagonal matrix, with $\mathbf{E}_{II}^{(k)}$ denoting an evolution submatrix of $\mathbf{E}$ on $(T_{k-1}, T_k)$. The matrices $\mathbf{E}_{IB}$, $\mathbf{E}_{BI}$, and $\mathbf{E}_{BB}$ are also submatrices of $\mathbf{E}$ corresponding to the indices $B$ and $I$. The elimination of $\mathbf{y}_I$ yields a Schur complement system $\mathbf{C}\mathbf{y}_B = (\mathbf{s}_B - \mathbf{E}_{BI}\mathbf{E}_{II}^{-1}\mathbf{s}_I)$, where the Schur complement matrix $\mathbf{C} = (\mathbf{E}_{BB} - \mathbf{E}_{BI}\mathbf{E}_{II}^{-1}\mathbf{E}_{IB})$ can be shown to be block lower bidiagonal, as expressed below, where $F_0^\Delta := (F_0 F_1^{-1})^{\hat{m}-1} F_0$. In the parareal method, the following preconditioner $\mathbf{C}_0$ is also employed for $\mathbf{C}$, where $G_1 = (M_h + \Delta T\, A_h)$:

$$\mathbf{C} \equiv \begin{bmatrix} F_1 & & & \\ -F_0^\Delta & F_1 & & \\ & \ddots & \ddots & \\ & & -F_0^\Delta & F_1 \end{bmatrix} \quad \text{and} \quad \mathbf{C}_0 \equiv \begin{bmatrix} G_1 & & & \\ -G_0 & G_1 & & \\ & \ddots & \ddots & \\ & & -G_0 & G_1 \end{bmatrix}. \tag{4.5}$$

It can be shown that the spectral radius $\rho(I - C_0^{-1}C) < 1$, yielding a convergent iteration for the Schur complement system. The parareal preconditioner $\mathbf{E}_n$ for $\mathbf{E}$ corresponds to applying $n$ Richardson iterations based on the Schur complement system. To compute the action of $\hat{\mathbf{z}}^n = \mathbf{E}_n^{-1}\mathbf{s}$, in *step* 1 solve $\mathbf{E}_{II}\mathbf{z}_I^0 = \mathbf{s}_I$, then apply $n$ iterations of the iteration $\mathbf{z}_B^{i+1} = \mathbf{z}_B^i + \mathbf{C}_0^{-1}(\mathbf{s}_B - \mathbf{E}_{BI}\mathbf{z}_I^0 - \mathbf{C}\mathbf{z}_B^i)$ starting with $\mathbf{z}_B^0 = \mathbf{0}$. Compute $\mathbf{z}_I^n = \mathbf{E}_{II}^{-1}(\mathbf{s}_I - \mathbf{E}_{IB}\mathbf{z}_B^n)$, and define $\mathbf{z}^n = [(\mathbf{z}_I^n)^T, (\mathbf{z}_B^n)^T]^T$. The vector $\hat{\mathbf{z}}^n$ can be straightforwardly assembled from $\mathbf{z}^n$ and $\mathbf{z}_B^{n+1}$.

*Remark* 4.2. We note that numerical studies of parareal methodology have been considered in [16] for another class of parabolic optimal control problem. Here in (2.1) we consider the space $Y_h$ of state variables $\underline{z}$ to be continuous in time, while in [16] a discontinuous Galerkin–in-time discretization with a penalty term is considered to control the discontinuities of the state variables $\underline{z}$ at the coarse times $T_k$, $1 \le k \le \hat{k}-1$. The first term of (2.1) is also not included in [16]. The resulting linear-quadratic optimal control problem seeks to minimize the following quadratic functional:

$$\begin{aligned} J_h(\underline{z}, \underline{v}) := {} & \frac{1}{2\epsilon\Delta T} \sum_{k=1}^{\hat{k}-1} \|\underline{z}(T_k^-) - \underline{z}(T_k)\|_{L^2(\Omega)}^2 + \frac{1}{2}\int_{T_0}^{T_{\hat{k}}} \|\underline{v}(t)\|_{L^2(\Omega)}^2\, dt \\ & + \frac{s}{2}\|\underline{z}(T_{\hat{k}}^-) - \underline{\tilde{y}}(T_{\hat{k}})\|_{L^2(\Omega)}^2 \end{aligned} \tag{4.6}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

subject to the *constraint* that $\underline{z}$ satisfies the discrete equation of state

$$M_h \dot{\underline{z}} + A_h \underline{z} = B_h \underline{v} \quad \text{for} \quad [T_k, T_{k+1}), \quad 0 \le k \le \hat{k} - 1. \tag{4.7}$$

We note that the parareal methodology in [16] explores the presence of nonzero block diagonal matrices due to the discontinuous Galerkin penalty term, while here we use the auxiliary variable $\mathbf{w}$.

Let us decompose $Z_k = \sum_{q=1}^{\hat{q}} \alpha_q(T_k) x_q$ and $Z_k^n = \sum_{q=1}^{\hat{q}} \alpha_q^n(T_k) x_q$, where $x_q$ are the generalized eigenvectors of $A_h$ with respect to $M_h$, and let us introduce $\zeta_q^n(T_k) := \alpha_q(T_k) - \alpha_q^n(T_k)$. The convergence of the parareal algorithm for systems follows from the next lemma, which is an extension of the results presented in [9, 10]. From now on we focus on the case $\theta = 1$ for the fine mesh propagator. The case $\theta = 1/2$ can be treated similarly; see Remark 4.4 and Table 5.9.

LEMMA 4.3. *Let* $\Delta T = (t_f - t_0)/\hat{k}$ *and* $T_k = t_0 + k\Delta T$ *for* $0 \le k \le \hat{k}$. *Then*

$$\max_{1 \le k \le \hat{k}} |\alpha_q(T_k) - \alpha_q^n(T_k)| \le \sigma_n \max_{1 \le k \le \hat{k}} |\alpha_q(T_k) - \alpha_q^0(T_k)|,$$

*where* $\sigma_n := \sup_{0<\beta<1} \left(e^{1-1/\beta} - \beta\right)^n \frac{1}{n!} \left| \frac{d^{n-1}}{d\beta^{n-1}} \left( \frac{1-\beta^{\hat{k}-1}}{1-\beta} \right) \right| \le 0.2984256075^n$.

*Proof.* Using Theorem 2 from [9] we obtain

$$\zeta_q^n = \left( (1 + \tau\lambda_q)^{-\Delta T/\tau} - \beta_q \right) \mathcal{T}(\beta_q) \zeta_q^{n-1}, \tag{4.8}$$

where $\beta_q := (1 + \lambda_q \Delta T)^{-1}$ and $\mathcal{T}(\beta) := \{\beta^{j-i-1} \quad \text{if } j > i, \quad 0 \quad \text{otherwise}\}$ is a Toeplitz matrix of size $\hat{k}$. Applying (4.8) recursively we obtain

$$\max_{1 \le k \le \hat{k}} |\zeta_q^n| \le \sigma_n^q \max_{1 \le k \le \hat{k}} |\zeta_q^0|,$$

where

$$\sigma_n^q := \left\| \left( (1 + \tau\lambda_q)^{-\Delta T/\tau} - \beta_q \right)^n \mathcal{T}^n(\beta_q) \right\|_{L^\infty}. \tag{4.9}$$

Since $\lambda_q > 0$ and $e^{-\lambda_q \Delta T} \le (1 + \tau\lambda_q)^{-\Delta T/\tau} \le \beta_q$, we obtain

$$|\,(1 + \tau\lambda_q)^{-\Delta T/\tau} - \beta_q| \le |e^{-\lambda_q \Delta T} - \beta_q| = |e^{1-1/\beta_q} - \beta_q|,$$

which yields

$$\sigma_n^q \le |e^{1-1/\beta_q} - \beta_q|^n \|\mathcal{T}^n(\beta_q)\|_{L^\infty} \le \sup_{0<\beta<1} |e^{1-1/\beta} - \beta|^n \|\mathcal{T}^n(\beta)\|_{L^\infty}.$$

By considering $\|\mathcal{T}^n(\beta)\|_\infty \le \|\mathcal{T}(\beta)\|_\infty^n = |\frac{1-\beta^{\hat{k}-1}}{1-\beta}|^n$, a simpler upper bound for $\rho_n$ can be obtained:

$$\sup_{0<\beta<1} \left| e^{1-1/\beta} - \beta \right|^n \left| \frac{1 - \beta^{\hat{k}-1}}{1 - \beta} \right|^n \le \left( \sup_{0<\beta<1} \frac{e^{1-1/\beta} - \beta}{1 - \beta} \right)^n \approx 0.2984256075^n,$$

and the maximum $\beta_*$ is attained around 0.3528865, independently of $n$ and $\hat{k}$ ($\beta_*$ presents slight variation for $1 \le n$ and $6 \le \hat{k}$, cases of practical interest). ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

*Remark* 4.4. Lemma 4.3 can be extended to the case where $\theta = 1/2$ (Crank–Nicolson (CN) method) for the fine propagator and backward Euler for the coarse propagator. Indeed, it is easy to see that

$$\frac{e^{-\lambda_q \Delta T/2}}{(1-\lambda_q \Delta T/2)^{-1}} \le \left(\frac{1-\tau\lambda_q/2}{1+\tau\lambda_q/2}\right)^{\Delta T/\tau} \le \beta_q,$$

which yields

$$\left|\left(\frac{1-\tau\lambda_q/2}{1+\tau\lambda_q/2}\right)^{\Delta T/\tau} - \beta_q\right| \le \left|e^{-\lambda_q \Delta T/2}(1-\lambda_q\Delta T/2) - \beta_q\right| = \left|e^{\frac{1}{2}-\frac{1}{2\beta_q}}\left(\frac{3}{2}-\frac{1}{2\beta_q}\right) - \beta_q\right|,$$

and an upper bound for $\rho_n$ can be obtained:

$$\sup_{0<\beta<1}\left|e^{\frac{1}{2}-\frac{1}{2\beta}}\left(\frac{3}{2}-\frac{1}{2\beta}\right)-\beta\right|^n \left|\frac{1-\beta^{\hat{k}-1}}{1-\beta}\right|^n \le \left(\sup_{0<\beta<1}\frac{e^{\frac{1}{2}-\frac{1}{2\beta}}\left(\frac{3}{2}-\frac{1}{2\beta}\right)-\beta}{1-\beta}\right)^n$$
$$\approx 0.5030630435^n.$$

We now analyze the spectral equivalence between $\mathbf{E}_n^{-T}\hat{\mathbf{K}}\mathbf{E}_n^{-T}$ and $\mathbf{E}^{-T}\mathbf{K}\mathbf{E}^{-T}$.

LEMMA 4.5. *Let $\lambda_{\max}$ and $\sigma_n$ be defined as in Lemmas* 3.1 *and* 4.3, *respectively. Then for any $\mathbf{s} \in \mathbb{R}^{(\hat{l}\hat{q})\times(\hat{l}\hat{q})}$ we have*

$$\gamma_{\min}\left(\mathbf{E}^{-1}\mathbf{s}, \mathbf{K}\mathbf{E}^{-1}\mathbf{s}\right) \le \left(\mathbf{E}_n^{-1}\mathbf{s}, \widehat{\mathbf{K}}\mathbf{E}_n^{-1}\mathbf{s}\right) \le \gamma_{\max}\left(\mathbf{E}^{-1}\mathbf{s}, \mathbf{K}\mathbf{E}^{-1}\mathbf{s}\right),$$

*where*

$$\gamma_{\max} = 1 + \frac{4}{\sqrt{1+\frac{\tau}{\sigma_n^2(t_f-t_0+s)}}-1} \tag{4.10}$$

*and*

$$\gamma_{\min} = 1 - \frac{4}{\sqrt{1+\frac{\tau}{\sigma_n^2(t_f-t_0+s)}}+1}. \tag{4.11}$$

*Proof.* Let $X_h := [x_1, \ldots, x_{\hat{q}}]$ and $\Lambda_h := \operatorname{diag}\{\lambda_1, \ldots, \lambda_{\hat{q}}\}$ be the generalized eigenvectors and eigenvalues of $A_h$ with respect to $M_h$, i.e., $A_h = M_h X_h \Lambda_h X_h^{-1}$. Let $\mathbf{z} := \mathbf{E}^{-1}\mathbf{s}$ with $\underline{z}(t) = \sum_{q=1}^{\hat{q}} \alpha_q(t) x_q$, and let $\hat{\mathbf{z}}^n := \mathbf{E}_n^{-1}\mathbf{s}$ with $\underline{\hat{z}}^n(t) = \sum_{q=1}^{\hat{q}} \alpha_q^n(t) x_q$. We note that $\alpha_q^n$ might be discontinuous across the $T_k$. Then

$$(\mathbf{E}^{-1}\mathbf{s}, \mathbf{K}\mathbf{E}^{-1}\mathbf{s}) = \|\underline{z}\|^2_{L^2(t_0,t_f;L^2(\Omega))} + s\|\underline{z}(t_f)\|^2_{L^2(\Omega)} = \sum_{q=1}^{\hat{q}} \|\alpha_q\|^2_{L^2(t_0,t_f)} + s|\alpha_q(t_f)|^2,$$

$$(\mathbf{E}_n^{-1}\mathbf{s}, \widehat{\mathbf{K}}\mathbf{E}_n^{-1}\mathbf{s}) = \|\underline{\hat{z}}^n\|^2_{L^2(t_0,t_f;L^2(\Omega))} + s\|\underline{\hat{z}}(t_f)^n\|^2_{L^2(\Omega)} = \sum_{q=1}^{\hat{q}} \|\alpha_q^n\|^2_{L^2(t_0,t_f)} + s|\alpha_q^n(t_f)|^2.$$

*First part* (Estimation of $\|\alpha_q^n\|^2$). For estimating $\|\alpha_q^n\|^2$ we obtain for $\epsilon \in (0, 1/2)$

$$\begin{aligned}\|\alpha_q^n\|^2_{L^2(t_0,t_f)} &= \left(\alpha_q^n - \alpha_q, \alpha_q^n + \alpha_q\right)_{L^2(t_0,t_f)} + \|\alpha_q\|^2_{L^2(t_0,t_f)} \\ &\le \frac{1}{4\epsilon}\|\alpha_q^n - \alpha_q\|^2_{L^2(t_0,t_f)} + \epsilon\|\alpha_q^n + \alpha_q\|^2_{L^2(t_0,t_f)} + \|\alpha_q\|^2_{L^2(t_0,t_f)} \\ &\le \frac{1}{4\epsilon}\|\alpha_q^n - \alpha_q\|^2_{L^2(t_0,t_f)} + 2\epsilon\|\alpha_q^n\|^2_{L^2(t_0,t_f)} + (1+2\epsilon)\|\alpha_q\|^2_{L^2(t_0,t_f)},\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

which reduces to

$$(1-2\epsilon)\|\alpha_q^n\|^2_{L^2(t_0,t_f)} \le (1+2\epsilon)\|\alpha_q\|^2_{L^2(t_0,t_f)} + \frac{1}{4\epsilon}\|\alpha_q^n - \alpha_q\|^2_{L^2(t_0,t_f)}. \tag{4.12}$$

For each $t_l \in [T_{k-1}, T_k]$ we have

$$|\alpha_q^n(t_l) - \alpha_q(t_l)| = \left((1+\tau\lambda_q)^{-1}\right)^{(t_l - T_{k-1})/\tau} |\alpha_q^n(T_{k-1}) - \alpha_q(T_{k-1})|,$$

and since $\lambda_q > 0$ implies $\left((1+\tau\lambda_q)^{-1}\right)^{(t_l-T_{k-1})/\tau} \le 1$, we obtain

$$\|\alpha_q^n - \alpha_q\|^2_{L^2(T_{k-1},T_k)} \le \Delta T |\alpha_q^n(T_{k-1}) - \alpha_q(T_{k-1})|^2.$$

Hence

$$(1-2\epsilon)\|\alpha_q^n\|^2_{L^2(t_0,t_f)} \le (1+2\epsilon)\|\alpha_q\|^2_{L^2(t_0,t_f)} + \frac{t_f - t_0}{4\epsilon} \max_{0\le k\le \hat{k}} |\alpha_q^n(T_k) - \alpha_q(T_k)|^2.$$

Using Lemma 4.3 with $\alpha_q(T_0) = 0$ and initial guess $\alpha_q^0(T_k) = 0$ and using

$$\max_{0\le k\le \hat{k}} |\alpha_q(T_k)|^2 = |\alpha_q(T_{k'})|^2 \le \frac{4}{\tau} \min_{\beta} \|\alpha_q(T_{k'}) + \beta t\|^2_{L^2(T_{k'}, T_{k'}+\tau)},$$

where $|\alpha_q(T_{k'})| := \max_{0\le k\le \hat{k}} |\alpha_q(T_k)| = |\alpha_q(T_{k'})|$, we obtain

$$\max_{0\le k\le \hat{k}} |\alpha_q^n(T_k) - \alpha_q(T_k)|^2 \le \sigma_n^2 \max_{0\le k\le \hat{k}} |\alpha_q(T_k)|^2 \le \frac{4\sigma_n^2}{\tau}\|\alpha_q\|^2_{L^2(t_0,t_f)}, \tag{4.13}$$

and the upper bound follows as

$$\|\alpha_q^n\|^2_{L^2(t_0,t_f)} \le \left(1 + \frac{\sigma_n^2(t_f - t_0)}{\tau\epsilon} + 2\epsilon\right)/(1-2\epsilon)\|\alpha_q\|^2_{L^2(t_0,t_f)}. \tag{4.14}$$

The lower bound for $\|\alpha_q^n\|^2$ follows similarly and is given by

$$\|\alpha_q^n\|^2_{L^2(t_0,t_f)} \ge \left(1 - \frac{\sigma_n^2(t_f - t_0)}{\tau\epsilon} - 2\epsilon\right)/(1+2\epsilon)\|\alpha_q\|^2_{L^2(t_0,t_f)}. \tag{4.15}$$

*Second part* (Estimation of $s\,|\alpha_q^n(t_f)|^2$). An expression similar to (4.12) holds; i.e.,

$$(1-2\epsilon)|\alpha_q^n(t_f)|^2 \le (1+2\epsilon)|\alpha_q(t_f)|^2 + \frac{1}{4\epsilon}|\alpha_q(t_f) - \alpha_q^n(t_f)|^2. \tag{4.16}$$

Using (4.13) to bound the second term of the right-hand side we obtain

$$(1-2\epsilon)|\alpha_q^n(t_f)|^2 \le (1+2\epsilon)|\alpha_q(t_f)|^2 + \frac{\sigma_n^2}{\tau\epsilon}\|\alpha_q\|^2_{L^2(t_0,t_f)}. \tag{4.17}$$

The lower bound for $|\alpha_q^n(t_f)|$ similarly follows, and it is given by

$$(1+2\epsilon)|\alpha_q^n(t_f)|^2 \ge (1-2\epsilon)|\alpha_q(t_f)|^2 - \frac{\sigma_n^2}{\tau\epsilon}\|\alpha_q\|^2_{L^2(t_0,t_f)}. \tag{4.18}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Combining expressions (4.14) and (4.17), and (4.15) and (4.18), yields the following upper and lower bounds:

$$\gamma_{\max}(\epsilon) = \left(1 + 2\epsilon + \frac{\sigma_n^2(t_f - t_0 + s)}{\tau\epsilon}\right)/(1 - 2\epsilon)$$

and

$$\gamma_{\min}(\epsilon) = \left(1 - 2\epsilon - \frac{\sigma_n^2(t_f - t_0 + s)}{\tau\epsilon}\right)/(1 + 2\epsilon).$$

Defining $\gamma_{\max} = \inf_{\epsilon\in(0,1/2)} \gamma_{\max}(\epsilon)$ and $\gamma_{\min} = \sup_{\epsilon\in(0,1/2)} \gamma_{\min}(\epsilon)$ and performing straightforward computations yields (4.10) and (4.11).

This completes the proof. ∎

*Remark* 4.6. For sufficiently small $\sigma_n$ the expressions (4.10) and (4.11) become

$$\gamma_{\max} - 1 \asymp 4\,\sigma_n \sqrt{\frac{t_f - t_0 + s}{\tau}} \tag{4.19}$$

and

$$\gamma_{\min} - 1 \asymp -4\,\sigma_n \sqrt{\frac{t_f - t_0 + s}{\tau}}. \tag{4.20}$$

Additionally, if the assumptions of Remark 3.3 hold, for $(t_f - t_0) \asymp 1$, we obtain

$$\gamma_{\max} - 1 \asymp \sigma_n \sqrt{\frac{1 + s}{\tau}} \tag{4.21}$$

and

$$\gamma_{\min} - 1 \asymp -\sigma_n \sqrt{\frac{1 + s}{\tau}}. \tag{4.22}$$

The next theorem estimates the condition number of Algorithm 3.

THEOREM 4.7. *Let us take* $\mathbf{P}_n := \text{blockdiag}(\mathbf{E}_n\widehat{\mathbf{K}}^{-1}\mathbf{E}_n^T, \mathbf{G})$ *as a block diagonal preconditioner for the coefficient matrix* $\mathbf{H}$ *of system* (3.3). *Assume the same conditions of Theorem* 3.4 *hold. Then the condition number of* $\mathbf{P}_n^{-1}\mathbf{H}$ *satisfies*

$$cond(\mathbf{P}_n^{-1}\mathbf{H}) \le \frac{\gamma_{\max}}{\gamma_{\min}} cond(\mathbf{P}^{-1}\mathbf{H}), \tag{4.23}$$

*where* $\gamma_{\max}$ *and* $\gamma_{\min}$ *are defined by* (4.10) *and* (4.11), *respectively.*

*Proof.* We obtain an upper bound for the eigenvalues $\eta$ of $\mathbf{P}_n^{-1}\mathbf{H}$ as follows:

$$\max|\eta\left(\mathbf{P}_n^{-1}\mathbf{H}\right)| = \max\left|\eta\left(\mathbf{P}_n^{-1/2}\mathbf{H}\mathbf{P}_n^{-1/2}\right)\right| \tag{4.24}$$

$$= \max_{\mathbf{v}\neq 0}\left(\frac{\mathbf{v}^T\mathbf{P}_n^{-1/2}\mathbf{H}\mathbf{P}_n^{-1/2}\mathbf{v}}{\mathbf{v}^T\mathbf{v}}\right) \tag{4.25}$$

$$\le \max_{\mathbf{v}\neq 0}\left(\frac{\mathbf{v}^T\mathbf{H}\mathbf{v}}{\mathbf{v}^T\mathbf{P}\mathbf{v}}\right)\max_{\mathbf{v}\neq 0}\left(\frac{\mathbf{v}^T\mathbf{P}\mathbf{v}}{\mathbf{v}^T\mathbf{P}_n\mathbf{v}}\right) \tag{4.26}$$

$$\le \gamma_{\max}\max|\eta\left(\mathbf{P}^{-1}\mathbf{H}\right)|. \tag{4.27}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Analogously, we obtain a lower bound

$$\min|\eta\left(\mathbf{P}_n^{-1}\mathbf{H}\right)| \geq \gamma_{\min}\min|\eta\left(\mathbf{P}^{-1}\mathbf{H}\right)|. \tag{4.28}$$

The bound (4.23) then follows. ∎

*Remark* 4.8. If the assumptions of Remark 3.3 hold, we obtain

$$\eta_{\max}\left(\mathbf{P}_n^{-1}\mathbf{H}\right) \preceq \gamma_{\max}\left(1+\frac{1+s/\tau}{r}\right) \tag{4.29}$$

and

$$\eta_{\min}\left(\mathbf{P}_n^{-1}\mathbf{H}\right) \succeq \gamma_{\min}\left(1+\frac{h^4}{r}\right). \tag{4.30}$$

**5. Numerical experiments.** In this section, we describe numerical results on tests of an optimal control problem involving the following one-dimentional heat equation:

$$\begin{cases} z_t - z_{xx} = v, & 0 < x < 1, \quad t > 0, \\ z(t,0) = 0, & 0 \leq t, \\ z(t,1) = 0, & 0 \leq t, \\ z(0,x) = 0, & 0 < x < 1. \end{cases} \tag{5.1}$$

We choose the performance function $\tilde{y}(x) = x(1-x)e^{-x}$ for $t \in [0,1]$. Following [15], we normalize the parameter $q = 1$. The numerical results in Tables 5.1–5.8 are obtained by discretizing (5.1) using the backward Euler method ($\theta = 1$), while the results in Table 5.9 are obtained by discretizing (5.1) using the CN method ($\theta = 1/2$). As a stopping criteria for the iterative solvers, we choose $\|\mathbf{r}_k\|/\|\mathbf{r}_0\| \leq 10^{-9}$, where $\mathbf{r}_k$ denotes the residual at the $k$th iteration.

TABLE 5.1

*Number of PCG iterations for Algorithm 1 for fixed $q = 1$, $r = 10^{-4}$, and $s = 0$ ($s = 1$). Backward Euler discretization was used for different time steps $\tau = 1/\hat{l}$ and mesh sizes $h = 1/\hat{m}$.*

| $\hat{m} \setminus \hat{l}$ | 32 | 64 | 128 | 256 | 512 |
|---|---|---|---|---|---|
| 32 | 36 (41) | 38 (44) | 40 (46) | 40 (47) | 40 (48) |
| 64 | 36 (41) | 38 (44) | 40 (46) | 40 (47) | 40 (48) |
| 128 | 36 (41) | 38 (44) | 40 (46) | 40 (47) | 40 (48) |
| 256 | 36 (41) | 38 (44) | 40 (46) | 40 (47) | 40 (48) |
| 512 | 36 (41) | 38 (44) | 40 (46) | 40 (47) | 40 (48) |

TABLE 5.2

*Number of PCG iterations for Algorithm 1 for varying values of $\tau = 1/\hat{l}$ and $s$, with fixed $q = 1$ and $r = 10^{-4}$. Backward Euler is employed with mesh size $h = 1/32$ and time step $\tau = 1/\hat{l}$.*

| $\hat{l} \setminus s$ | 0 | 1 | 10 | 100 |
|---|---|---|---|---|
| 32 | 36 | 41 | 50 | 68 |
| 64 | 38 | 44 | 52 | 73 |
| 128 | 40 | 46 | 57 | 83 |
| 256 | 40 | 47 | 62 | 89 |
| 512 | 40 | 48 | 64 | 96 |

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

TABLE 5.3
*Condition number associated with Algorithm 1 for varying values of $s$ and $r$ and fixed $q = 1$. Backward Euler discretization was employed with mesh size $h = 1/32$ and time step $\tau = 1/64$.*

| $r \setminus s$ | $10^4$ | $10^2$ | 1 | $10^{-2}$ | 0 |
|---|---|---|---|---|---|
| $10^{-2}$ | $4.9\ 10^4$ | $5.0\ 10^2$ | 6.2 | 1.9 | 1.9 |
| $10^{-4}$ | $4.7\ 10^6$ | $4.8\ 10^4$ | $5.2\ 10^2$ | 93 | 93 |
| $10^{-6}$ | $4.2\ 10^8$ | $4.8\ 10^6$ | $5.1\ 10^4$ | $9.2\ 10^3$ | $9.1\ 10^3$ |

TABLE 5.4
*Number of MINRES iterations for Algorithm 2 when $q = 1$, $r = 10^{-4}$, and $s = 0$ ($s = 1$). Backward Euler discretization is used with time step $\tau = 1/\hat{l}$ and mesh size $h = 1/\hat{m}$.*

| $\hat{m} \setminus \hat{l}$ | 32 | 64 | 128 | 256 | 512 |
|---|---|---|---|---|---|
| 32 | 56 (60) | 58 (62) | 58 (64) | 60 (66) | 60 (68) |
| 64 | 56 (60) | 58 (62) | 58 (64) | 60 (66) | 60 (68) |
| 128 | 56 (60) | 58 (62) | 58 (64) | 60 (66) | 60 (68) |
| 256 | 56 (60) | 58 (62) | 58 (64) | 60 (66) | 60 (68) |
| 512 | 56 (60) | 58 (62) | 58 (64) | 60 (66) | 60 (68) |

**Algorithm 1**. We employ matrix $\mathbf{G}$ (see section 2) as the preconditioner for system (3.1) and use the PCG method to solve the resulting preconditioned system. Table 5.1 presents the number of PCG iterations for different temporal and spatial mesh sizes. We choose the parameters $s = 0$ ($s = 1$) and $r = 10^{-4}$. As predicted by Theorem 3.2 (see also (3.41)), the number of PCG iterations remains constant as $h$ is refined; note that $O(h^4/r) \preceq 1$ in all the numerical tests. Table 5.1 also shows that the number of iterations deteriorates very weakly when $\tau$ is refined. As expected from (3.41), this deterioration is more noticeable for larger $s$; see Tables 5.1 and 5.2. The sharpness of the condition number estimates (3.41) with respect to the parameters $r$ and $s$ are shown in Table 5.3.

**Algorithm 2.** Table 5.4 presents the number of iterations required to solve system (3.3) preconditioned by matrix (3.4) using the MINRES method. We vary the temporal and spatial mesh sizes, and we choose the parameters $s = 0$ ($s = 1$) and $r = 10^{-4}$. As predicted by the analysis (see (3.43)), as the spatial mesh is refined, the number of iterations remains bounded. We note that the numerical experiments show (not all are included here) that Algorithms 1 and 2 have similar properties when we consider other values of $s$, $r$, and $\tau$. This similarity is expected, since both algorithms have similar condition number estimates.

**Algorithm 3** (parareal based). Table 5.5 lists the MINRES iterations required to solve system (3.3) preconditioned by matrix (3.5) using $n = 2$ parareal iterations in $E_n^{-1}$ and $E_n^{-T}$, with coarse propagation based on the backward Euler method. Different values of mesh sizes $h$ and coarse time steps $\Delta T$ are tested, with fixed $\tau = 1/512$ and parameters $s = 0$, $r = 10^{-3}$, $10^{-4}$, $10^{-5}$. Comparing Table 5.5 (for $r = 10^{-4}$) with Table 5.4 (for $s = 0$ and $\tau = 1/512$), we see that for a variety of mesh sizes $h$, the number of MINRES iterations in Table 5.5 is only 21% more than in Table 5.4. This is remarkable, since preconditioner (3.5) can distribute $F_1^{\Delta}$ into $(t_f - t_0)/\Delta T$ processes in parallel. The scalability of preconditioner (3.5) with respect to the number of subintervals $(t_f - t_0)/\Delta T$ is also seen in Table 5.5. The number of iterations not depending on $(t_f - t_0)/\Delta T$ follows from Remark 4.6, while the nondependence on $h$ follows from Remark 4.8, since $h^4/r \preceq 1$ in the tests.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

TABLE 5.5
*Number of MINRES iterations for Algorithm 3 when $q = 1$, $s = 0$, and $r = 10^{-3}/10^{-4}/10^{-5}$. Backward Euler discretization is used with $\tau = 1/512$ and $h = 1/\hat{m}$. The preconditioner employed $n = 2$ parareal iterations and a backward Euler coarse propagator with $\hat{k} = 1/(\Delta T)$.*

| $\hat{k}$ | 8 | 16 | 32 | 64 |
|---|---|---|---|---|
| $\Delta T/\tau$ | 64 | 32 | 16 | 8 |
| $h = 1/16$ | 40 / 78 / 300 | 40 / 78 / 164 | 40 / 78 / 160 | 40 / 76 / 158 |
| $h = 1/32$ | 40 / 78 / 300 | 40 / 78 / 164 | 40 / 78 / 160 | 40 / 76 / 158 |
| $h = 1/64$ | 40 / 78 / 300 | 40 / 78 / 164 | 40 / 78 / 160 | 40 / 76 / 158 |

TABLE 5.6
*Values of $\kappa_{\max} - 1$ for $\tau = (1/\hat{l})$ and $n$ parareal iterations.*

Parameters $h = 1/10$, $\Delta T = 1/20$, and $s = 0$.

| $n \setminus \hat{l}$ | 200 | 400 | 800 | 1600 |
|---|---|---|---|---|
| $n = 1$ | 0.864415 | 1.449299 | 2.473734 | 4.371709 |
| $n = 2$ | 0.070835 | 0.097852 | 0.136802 | 0.193845 |
| $n = 3$ | 0.007760 | 0.010765 | 0.015141 | 0.021165 |
| $n = 4$ | 0.000865 | 0.001224 | 0.001715 | 0.002397 |

Parameters $h = 1/10$, $\Delta T = 1/20$, and $s = 0.01$.

| $n \setminus \hat{l}$ | 200 | 400 | 800 | 1600 |
|---|---|---|---|---|
| $n = 1$ | 0.864419 | 1.449305 | 2.473744 | 4.371725 |
| $n = 2$ | 0.070852 | 0.097885 | 0.136851 | 0.193918 |
| $n = 3$ | 0.007784 | 0.010800 | 0.015180 | 0.021208 |
| $n = 4$ | 0.000869 | 0.001229 | 0.001721 | 0.002436 |

Parameters $h = 1/10$, $\Delta T = 1/20$, and $s = 1$.

| $n \setminus \hat{l}$ | 200 | 400 | 800 | 1600 |
|---|---|---|---|---|
| $n = 1$ | 3.779956 | 6.866334 | 12.933472 | 24.992363 |
| $n = 2$ | 0.226120 | 0.333472 | 0.500244 | 0.768025 |
| $n = 3$ | 0.024360 | 0.034622 | 0.049315 | 0.070448 |
| $n = 4$ | 0.002846 | 0.004026 | 0.005699 | 0.008069 |

TABLE 5.7
*Values of $\kappa_{\max} - 1$ for $s = 0/1/10/100$ and $n$ parareal iterations.*

Parameters $h = 1/10$, $\Delta T = 1/20$, and $\tau = 1/800$

| $n \setminus s$ | 0 | 1 | 10 | 100 |
|---|---|---|---|---|
| $n = 2$ | 0.136802 | 0.500244 | 2.239858 | 16.307768 |
| $n = 3$ | 0.015141 | 0.049315 | 0.157836 | 0.581009 |
| $n = 4$ | 0.001715 | 0.005699 | 0.017459 | 0.056029 |

In the next set of experiments, we compute the largest eigenvalue $\kappa_{\max}$ of

$$\kappa_{\max} := \max_{\mathbf{s}\neq\mathbf{0}} \frac{(\mathbf{E}_n^{-1}\mathbf{s}, \widehat{\mathbf{K}}\mathbf{E}_n^{-1}\mathbf{s})}{(\mathbf{E}^{-1}\mathbf{s}, \mathbf{K}\mathbf{E}^{-1}\mathbf{s})}. \tag{5.2}$$

The goal is to compare the sharpness of estimate $\gamma_{\max} - 1$ in (4.6) with the exact value $\kappa_{\max} - 1$. Tables 5.6 and 5.7 list the values of $\kappa_{\max} - 1$. We use a backward Euler fine time step discretization from $t_0 = 0$ to $t_f = 1$, the spatial mesh size is $h = 1/10$, and the coarse time step in the parareal (with backward Euler coarse time propagator) is $\Delta T = 1/20$. The results in Table 5.6 confirm that the $1/\sqrt{\tau}$ term dependence in (4.21) is sharp. The results in Table 5.7 confirm that the $\sqrt{1+s}$ term dependence in (4.21) is sharp for $\sigma_n$ sufficiently small, i.e., $n$ sufficiently large. Tables 5.6 and 5.7

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

TABLE 5.8

*Number of MINRES iterations for Algorithm 3. The parameters are $q = 1$, $s = 0$, and $r = 10^{-4}$. Backward Euler discretization is used with $\tau = 1/\hat{l}$ and $h = 1/\hat{m}$. The number of parareal iterations is $n = 1/2/4/7$, and the number of coarse time subintervals is $\hat{k}$. $\Delta T/\tau = 16$.*

| $\hat{k}$ | 4 | 8 | 16 | 32 |
|---|---|---|---|---|
| $\hat{l}$ | 64 | 128 | 256 | 512 |
| $h = 1/16$ | 76 / 58 / 58 / 58 | 83 / 74 / 58 / 58 | 89 / 76 / 66 / 60 | 96 / 78 / 66 / 60 |
| $h = 1/32$ | 76 / 58 / 58 / 58 | 83 / 74 / 58 / 58 | 89 / 76 / 66 / 60 | 96 / 78 / 66 / 60 |
| $h = 1/64$ | 76 / 58 / 58 / 58 | 83 / 74 / 58 / 58 | 89 / 76 / 66 / 60 | 96 / 78 / 66 / 60 |

TABLE 5.9

*Number of MINRES iterations for Algorithm 3. The parameters are $q = 1$, $s = 0$, and $r = 10^{-4}$. CN discretization is used with $h = 1/\hat{m}$ and $\tau = 1/\hat{l}$. The number of parareal iterations is $n = 1/2/4/7$, and the number of coarse time subintervals is $\hat{k}$. $\Delta T/\tau = 16$.*

| $\hat{k}$ | 4 | 8 | 16 | 32 |
|---|---|---|---|---|
| $\hat{l}$ | 64 | 128 | 256 | 512 |
| $h = 1/16$ | 97 / 58 / 58 / 58 | 86 / 76 / 60 / 60 | 91 / 78 / 66 / 60 | 97 / 78 / 66 / 60 |
| $h = 1/32$ | 176 / 58 / 58 / 58 | 170 / 90 / 60 / 60 | 106 / 80 / 66 / 60 | 97 / 78 / 66 / 60 |
| $h = 1/64$ | 259 / 60 / 58 / 58 | 433 / 211 / 60 / 60 | 375 / 147 / 66 / 60 | 189 / 95 / 66 / 60 |

confirm the geometrical decay with respect to $n$, however, with a geometrical factor smaller than the value 0.2984256075 established in Lemma 4.3.

In Tables 5.8 and 5.9 we study the convergence associated with preconditioner (3.5) as we increase the number of iterations $n$ of the parareal algorithm. We consider two discretizations for (5.1). The backward Euler method ($\theta = 1$) is considered in Table 5.8, while the CN method ($\theta = 1/2$) is considered in Table 5.9. The coarse propagation operator for the parareal algorithm is the backward Euler method. The parameters are chosen as $s = 0$ and $r = 10^{-4}$. The number of fine temporal subintervals inside each coarse temporal subintervals is set equal to $\Delta T/\tau = 16$. The tables present the number of MINRES iterations required to solve system (3.3) preconditioned by the block diagonal matrix (3.5). Different values of mesh sizes $h$, coarse time step $\Delta T$, and parareal iterations $n$ are tested. Comparing Table 5.4 (for $s = 0$ and $\tau = 1/512$) with Table 5.8 ($\theta = 1$), we conclude that Algorithms 2 and 3 (with $n = 7$) are nearly identical. We can also see in Table 5.8 that for the backward Euler discretization Algorithm 3 is robust and scalable for $n = 1$ or $n = 2$ parareal iterations, while in Table 5.9 the CN discretization requires $n = 4$ parareal iterations. More iterations are expected for the CN method since $\sigma_n(CN) = 0.503063035^n$ and $\sigma_n(BE) = 0.2984256075^n$; see Remark 4.4 and Lemma 4.3.

**6. Conclusion.** Our central focus in this paper has been the parallel-in-time preconditioner described in Algorithm 3. Its convergence rate is described in Theorem 4.7 and depends on the parameters $\gamma_{\min}$ and $\gamma_{\max}$; see Remarks 4.6 and 4.8. Since $\sigma_n \to 0$ as $n \to \infty$, this yields an asymptotically sharp bound. However, numerical tests indicate effective convergence even for significantly smaller $n$ (such as $n = 2$).

**Acknowledgments.** We thank the referees for their comments on an earlier version of this paper, which helped improve our presentation.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

## REFERENCES


[1] A. Battermann and M. Heinkenschloss, *Preconditioners for Karush-Kuhn-Tucker matrices arising in the optimal control of distributed systems*, in Optimal Control of Partial Differential Equations, W. Desch, F. Kappel, and K. Kunisch, eds., Birkhäuser, Basel, 1998, pp. 15–32.

[2] A. Battermann and E. W. Sachs, *Block preconditioner for KKT systems in PDE-governed optimal control problems*, in Workshop on Fast Solutions of Discretized Optimization Problems, H. W. Hoppe, K.-H. Hoffmann, and V. Schulz, eds., Birkhäuser, Basel, 2001, pp. 1–18.

[3] M. Benzi, G. H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–37.

[4] M. Berggren and M. Heinkenschloss, *Parallel solution of optimal-control problems by time-domain decomposition*, in Computational Science for the 21st Century, M.-O. Bristeau, G. Etgen, W. Fitzgibbon, J. L. Lions, J. Periaux, and M. F. Wheeler, eds., Wiley, Chichester, 1997, pp. 102–112.

[5] G. Biros and O. Gattas, *Parallel Lagrange–Newton–Krylov–Schur methods for PDE-constrained optimization. Part* I: *The Krylov–Schur solver*, SIAM J. Sci. Comput., 27 (2005), pp. 687–713.

[6] G. Biros and O. Ghattas, *Parallel Lagrange–Newton–Krylov–Schur methods for PDE constrained optimization. Part* II: *The Lagrange–Newton solver and its application to optimal control of steady viscous flows*, SIAM J. Sci. Comput., 27 (2005), pp. 714–739.

[7] J. W. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[8] C. Farhat and M. Chandesris, *Time-decomposed parallel time-integrators: Theory and feasibility studies for fluid, structure, and fluid-structure applications*, Internat. J. Numer. Methods Engrg., 58 (2003), pp. 1397–1434.

[9] M. J. Gander and S. Vandewalle, *On the super linear and linear convergence of the parareal algorithm*, in Domain Decomposition Methods in Science and Engineering, Lect. Notes Comput. Sci. Eng. 55, D. E. Keyes and O. B. Widlund, eds., Springer, New York, 2005, pp. 291–298.

[10] M. J. Gander and S. Vandewalle, *Analysis of the parareal time-parallel time-integration method*, SIAM J. Sci. Comput., 29 (2007), pp. 556–578.

[11] E. Haber and U. M. Ascher, *Preconditioned all-at-once methods for large, sparse parameter estimation problems*, Inverse Problems, 17 (2001), pp. 1847–1864.

[12] M. Heinkenschloss, *A time-domain decomposition iterative method for the solution of distributed linear quadratic optimal control problems*, J. Comput. Appl. Math., 173 (2005), pp. 169–198.

[13] J. L. Lions, Y. Maday, and G. Turinici, *Résolution d'edp par un schéma en temps pararéel*, C. R. Acad. Sci. Paris Ser. I Math., 332 (2001), pp. 661–668.

[14] J. L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer, New York, 1971.

[15] A. Locatelli, *Optimal Control: An Introduction*, Birkhäuser, Berlin, 2001.

[16] Y. Maday and G. Turinici, *A parareal in time procedure for the control of partial differential equations*, C. R. Math. Acad. Sci. Paris, 335 (2002), pp. 387–392.

[17] T. P. Mathew, M. Sarkis, and C. E. Schaerer, *Block iterative algorithms for the solution of parabolic optimal control problems*, in 7th International Meeting High Performance Computing for Computational Science - VECPAR 2006, Lecture Notes in Comput. Sci. 4395, Springer, New York, 2007, pp. 452–465.

[18] T. Mathew, P. L. Polyakov, G. Russo, and J. Wang, *Domain decomposition operator splittings for the solution of parabolic equations*, SIAM J. Sci. Comput., 19 (1998), pp. 912–932.

[19] T. Mathew, M. Sarkis, and C. E. Schaerer, *Analysis of block matrix preconditioners for elliptic optimal control problems*, Numer. Linear Algebra Appl., 14 (2007), pp. 257–279.

[20] P. Neittaanäki and D. Tiba, *Optimal Control of Linear Parabolic Systems: Theory, Algorithms and Applications*, Marcel Dekker, New York, 1994.

[21] J. Nievergelt, *Parallel methods for integrating ordinary differential equations*, Comm. ACM, 7 (1964), pp. 731–733.

[22] T. Rusten and R. Winther, *A preconditioned iterative method for saddlepoint problems*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 887–904.

[23] M. Sarkis, C. E. Schaerer, and T. Mathew, *Block diagonal parareal preconditioner for parabolic optimal control problems*, in Domain Decomposition Methods in Science and Engineering XVII, Lect. Notes Comput. Sci. Eng. 60, Springer, New York, 2008, pp. 409–416.


Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[24] C. E. Schaerer and E. Kaszkurewicz, *The shooting method for the numerical solution of ordinary differential equations: A control theoretical perspective*, Internat. J. Systems Sci., 32 (2001), pp. 1047–1053.
[25] V. Simoncini and D. B. Szyld, *Theory of inexact Krylov subspace methods and applications to scientific computing*, SIAM J. Sci. Comput., 25 (2003), pp. 454–477.
[26] R. S. Varga, *Matrix Iterative Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1962.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.