Block Diagonal Parareal Preconditioner for Parabolic Optimal Control Problems

Christian E. Schaerer[1], Tarek Mathew, and Marcus Sarkis[1]

[1] IMPA, Dona Castorina 110, Rio de Janeiro, RJ 22460-320, Brazil.
cschaer@fluid.impa.br, msarkis@impa.br and tmathew@poonithara.org

Summary. We describe an iterative algorithm for the solution of a large scale linear-quadratic parabolic optimal control problem. Unlike Ricatti equation based methods, we determine the control variable by an iterative procedure which solves a large saddle point system obtained by an *all at once* discretization strategy involving the state (primal) variables, the control variables and the adjoint (dual) variables. We derive a reduced symmetric indefinite linear system involving the control variables and auxiliary variables, and solve it using a preconditioned MINRES iteration, with a symmetric positive definite block diagonal preconditioner based on the parareal algorithm. Theoretical and numerical results show that the preconditioned algorithm has adequate convergence properties and parallel scalability.

1 Introduction

In this paper, we describe a block matrix iterative algorithm for solving an "all at once" discretization of a linear-quadratic parabolic optimal control problem (OCP) on a finite time interval, see [1, 3, 6]. The problem we consider seeks to determine a control function parameterizing the forcing term in a parabolic equation, so that the solution to the parabolic equation closely matches a given "tracking" function in a finite time interval. Formally this yields a constrained minimization problem in which the *quadratic* functional whose minimum is sought, 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 requires the state variable to solve the parabolic equation. The classical approach for determining the solution to a linear-quadratic optimal control problem consists of solving the Ricatti or Silvester equations, derived by an application of the Pontryagin maximum principle. However, this approach can be prohibitively expensive when the number of state variables is large. Instead, we employ an iterative algorithm to determine the solution to a combined spatial and temporal discretization of the optimal control problem, which yields a large saddle point system.

The spatial discretization of the parabolic equation is obtained by the finite element method, and its temporal discretization by the backward Euler scheme [1]. Using a reduction approach employed in [7, 6], we obtain a symmetric positive definite reduced system for the unknown *control variables* (with low dimension in realistic situations). This system can be solved using the CG method, but requires inner-outer iteration. To overcome this drawback, we introduce an auxiliary variable resulting in a symmetric indefinite ill-conditioned system. We employ a symmetric positive definite block diagonal preconditioner [6] based on the *parareal* algorithm of [4] and an iterative shooting method [3, 5]. It yields a rate of convergence independent of the mesh size h. In § 2, we describe the saddle point formulation of the optimal control problem and a reduced symmetric indefinite linear system. In § 3, we describe theoretical results on the parareal preconditioner. In § 4, numerical results verify the scalability of the algorithm.

2 The parabolic optimal control problem

For brevity, we shall omit a discussion of the parabolic optimal control problem and instead, begin with the finite dimensional linear-quadratic optimal control problem resulting from a spatial discretization of the parabolic optimal control problem. For $t \in [t_0, t_f]$ let $y(t) \in \mathbb{R}^{\hat{m}}$ denote a nodal vector representing a continuous piecewise linear finite element function approximating the solution to the parabolic equation, and let $u(t) \in \mathbb{R}^{\hat{p}}$ denote the nodal vector associated with the discrete control variable, which will be piecewise constant in space [3, 6]. The linear-quadratic optimal control problem will seek to minimize the following quadratic objective functional:

$$J(y,u) \equiv \frac{1}{2}\int_{t_o}^{t_f} \left(e(t)^T\, Q\, e(t) + u(t)^T\, R\, u(t)\right)\, dt + \frac{1}{2}\, e(t_f)^T\, C\, e(t_f), \qquad (1)$$

where $y_*(t) \in \mathbb{R}^{\hat{m}}$ denotes a given discrete "tracking" function at time t, and $e(t) = (y(t) - y_*(t))$ denotes the tracking error. Here $C \in \mathbb{R}^{\hat{m}\times\hat{m}}$, $Q \in \mathbb{R}^{\hat{m}\times\hat{m}}$, $R \in \mathbb{R}^{\hat{p}\times\hat{p}}$ are given symmetric positive definite matrices. The state variable $y(t)$ is required to solve the following system of differential equations obtained by spatial discretization of the parabolic equation:

$$\dot{y} = A\, y + B\, u, \quad \text{for} \quad t_o < t < t_f; \quad \text{and} \quad y(t_o) = y_0, \qquad (2)$$

where $A \in \mathbb{R}^{\hat{m}\times\hat{m}}$ denotes the symmetric *negative* definite matrix obtained from the discretization of the elliptic operator and $B \in \mathbb{R}^{\hat{m}\times\hat{p}}$ denotes the control matrix. To obtain a temporal discretization of (2), we partition $[t_o,\, t_f]$ into $\hat{l}$- sub-intervals with time step $\tau = (t_f - t_o)/\hat{l}$. We denote $t_l = l\,\tau$ and $y_l := y(t_l)$ for $0 \le l \le \hat{l}$. For simplicity, we assume that the state variable $y(t)$ is continuous and piecewise linear in each $(t_l,\, t_{l+1}]$, while the control $u(.)$ is constant on each interval $(t_l,\, t_{l+1}]$ with $u_{l+1/2} = u(t_{l+1/2})$.

Employing the backward Euler discretization of (2) in time, yields:

$$E\,\mathbf{y} + N\,\mathbf{u} = \mathbf{f}, \tag{3}$$

where $\mathbf{y} := [y_1, \ldots, y_{\hat{l}}]^T \in \Re^{\hat{l}\hat{m}}$ and $\mathbf{u} := [u_{1/2}, \ldots, u_{\hat{l}-1/2}]^T \in \mathbb{R}^{\hat{l}\hat{p}}$, denotes block vectors of state and control vectors, respectively, at all the discrete times. The input vector is $\mathbf{f} := [-F_0 y_o, 0, ..., 0]^T \in \mathbb{R}^{\hat{l}\hat{m}}$. Matrix $E =$ lower-bidiag$[F_0, -F_1] \in \mathbb{R}^{(\hat{l}\hat{m})\times(\hat{l}\hat{m})}$ involves the submatrices $F_0 := I \in \mathbb{R}^{\hat{m}\times\hat{m}}$ and $F_1 := I - \tau A \in \mathbb{R}^{\hat{m}\times\hat{m}}$, and the matrix $N = \tau I_{\hat{l}} \otimes B \in \mathbb{R}^{(\hat{l}\hat{p})\times(\hat{l}\hat{p})}$. Expression (1) is discretized analogously (see [6] for details) yielding:

$$J_h(y, u) = \frac{1}{2}(\mathbf{u}^T G \mathbf{u}^T + \mathbf{e}^T K \mathbf{e}), \tag{4}$$

where $G = r\tau I_{\hat{l}} \otimes h I_{\hat{p}} \in \mathbb{R}^{(\hat{l}\hat{p})\times(\hat{l}\hat{p})}$ and $K = \Gamma + Z \in \mathbb{R}^{(\hat{l}\hat{m})\times(\hat{l}\hat{m})}$. Here $\Gamma =$ diag$(0, 0, ..., 0, sM_h)$ and $Z = qM_\tau \otimes M_h$, where M_τ and M_h are appropriately defined mass matrices, I_j is an identity matrix of size j, $\{q, r, s\}$ are relaxation parameters, and $\otimes$ stands for Kronecker product. The Lagrangian $\mathcal{L}_h(\mathbf{y}, \mathbf{u}, \mathbf{p})$ for minimizing (4) subject to constraint (3) is:

$$\mathcal{L}_h(\mathbf{y}, \mathbf{u}, \mathbf{p}) = \frac{1}{2}(\mathbf{u}^T G \mathbf{u}^T + \mathbf{e}^T K \mathbf{e}) + \mathbf{p}^T(E\mathbf{y} + N\mathbf{u} - \mathbf{f}). \tag{5}$$

Vector $\mathbf{e} := [e_1^T, \ldots, e_{\hat{l}}^T]^T \in \mathbb{R}^{\hat{l}\hat{m}}$ is defined in terms of the discrete error vectors e_l for $l = 1, ..., \hat{l}$. To obtain a discrete saddle point formulation of (5), optimality conditions for $\mathcal{L}_h(\cdot, \cdot, \cdot)$ yields the symmetric indefinite linear system:

$$\begin{bmatrix} K & 0 & E^T \\ 0 & G & N^T \\ E & N & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} K\mathbf{g} \\ 0 \\ \mathbf{f} \end{bmatrix} \tag{6}$$

where $\mathbf{g} := [g_1, \ldots, g_{\hat{l}}]^T \in \mathbb{R}^{\hat{l}\hat{m}}$ for $g_l = y_*(l\tau)$. Eliminating $\mathbf{y} = E^{-1}(\mathbf{f} - N\,\mathbf{u})$ and $\mathbf{p} = E^{-T}(K\,\mathbf{g} - K\,\mathbf{y})$ in (6) yields the *reduced* Schur complement system:

$$(G + N^T E^{-T} K E^{-1} N)\mathbf{u} = \mathbf{b}_u, \tag{7}$$

see [6, 7], where $\mathbf{b}_u := N^T E^{-T} K \left(E^{-1}\mathbf{f} - \mathbf{g}\right)$ is pre-computed. Matrix $H := G + N^T E^{-T} K E^{-1} N$ will be symmetric positive definite and $(\mathbf{u}, G\mathbf{u}) \leq (\mathbf{u}, H\mathbf{u}) \leq \mu(\mathbf{u}, G\mathbf{u})$, where $\mu = O(1 + \frac{1+s/\tau}{r})$; see [6]. As a result, the PCG method can be used to solve (7), but double iteration will be required. To avoid double iteration, define auxiliary variables $\mathbf{w} := -E^{-T} K E^{-1} N\mathbf{u}$ and $\hat{\mathbf{b}} := -\mathbf{b}_u$. Then (7) will be equivalent to the symmetric indefinite system:

$$\begin{bmatrix} EK^{-1}E^T & N \\ N^T & -G \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \hat{\mathbf{b}} \end{bmatrix}. \tag{8}$$

System (8) will be ill-conditioned (see [6]), but can be solved using the MINRES algorithm with the symmetric positive definite block diagonal preconditioner $\mathcal{P} =$ diag$(\tilde{E}K^{-1}\tilde{E}^T, G)$, where matrix $\tilde{E}$ is any matrix spectrally

equivalent (or a preconditioner) to the evolution matrix E; see [6]. We next consider the parareal algorithm with n sweeps, denoted by E_n^{-1}, and analyze the spectral equivalence between $E^{-T}KE^{-1}$ and $E_n^{-T}KE_n^{-1}$, for use in $\mathcal{P}^{-1} = \mathrm{diag}(E_n^{-T}KE_n^{-1}, G^{-1})$.

3 Parareal preconditioner for $EK^{-1}E^T$

In this section, we first describe a parareal-shooting method preconditioner for $E\,\mathbf{z} = \mathbf{v}$, where $\mathbf{z} := [z_1^T, \ldots, z_{\hat{l}}^T]^T$ and $\mathbf{v} := [v_1^T, \ldots, v_{\hat{l}}^T]^T$ belong to $\mathbb{R}^{\hat{l}\hat{m}}$. Subsequently, we describe the parareal preconditioner for $E\,K^{-1}\,E^T$.

For simplicity of presentation we set $t_0 = 0$ and $z_0 = 0$. We adopt the notation $z(t_l) = z_l$, for $t_l = l\tau$ and $0 \le l \le \hat{l}$ and $z(t)$ for $t \in [t_0, t_f]$ via linear interpolation. We partition the time interval $[t_0, t_f]$ into $\hat{k}$ *coarse* subintervals of length $\Delta T = (t_f - t_0)/\hat{k}$, and set $T_k = k\Delta T$ for $0 \le k \le \hat{k}$. For each coarse time interval $[T_k, T_{k+1}]$, we introduce coarse and the local propagators. The coarse propagator $\mathcal{G}(T_{k+1}, Z_k)$ denotes the solution to $z_t = A\,z$ at T_{k+1} with initial data $z(T_k) = Z_k$, obtained by applying one step of the backward Euler method, i.e., $\mathcal{G}(T_{k+1}, Z_k) = \mathcal{G}\,Z_k$, where $\mathcal{G} = (I - A\Delta T)^{-1} \in \mathbb{R}^{\hat{m}\times\hat{m}}$. The fine (or local) propagator $\mathcal{F}(T_{k+1}, Z_k, v)$ denotes the solution to $z_t = A\,z + v$ at T_{k+1} with initial data $z(T_k) = Z_k$ and forcing v, obtained by applying the backward Euler method on the fine mesh with $t_l \in [T_k, T_{k+1}]$:

$$\mathcal{F}(t_l, Z_k, s) = \Phi^{\hat{\imath}} Z_k + \sum_{i=1}^{\hat{\imath}} \Phi^{\hat{\imath}-i} v_{\frac{T_k}{\tau}+i}$$

where $\hat{\imath} = \frac{t_l - T_k}{\tau}$ and $\Phi = (I - A\tau)^{-1}$. Imposing the continuity conditions at time T_k, i.e., $Z_k - \mathcal{F}Z_{k-1} = 0$ for $1 \le k \le \hat{k}$, we obtain the system:

$$\begin{bmatrix} I & & & \\ -\mathcal{F} & I & & \\ & \ddots & \ddots & \\ & & -\mathcal{F} & I \end{bmatrix} \begin{bmatrix} Z_0 \\ Z_1 \\ \vdots \\ Z_{\hat{k}} \end{bmatrix} = \begin{bmatrix} z_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{9}$$

In the parareal algorithm, the coarse propagator $\mathcal{G}$ is used as a tool for preconditioning the system (9) via the following Richardson iterative scheme:

$$\begin{bmatrix} Z_0 \\ Z_1 \\ \vdots \\ Z_{\hat{k}} \end{bmatrix}^{n+1} = \begin{bmatrix} Z_0 \\ Z_1 \\ \vdots \\ Z_{\hat{k}} \end{bmatrix}^{n} - \left(\begin{bmatrix} I & & & \\ -\mathcal{G} & I & & \\ & \ddots & \ddots & \\ & & -\mathcal{G} & I \end{bmatrix} \right)^{-1} \begin{bmatrix} r_0 \\ r_1 \\ \vdots \\ r_{\hat{k}} \end{bmatrix}^{n}, \tag{10}$$

where the residual vector $\mathbf{r}^n := \{r_k^n\}_{k=1}^{\hat{k}}$ is defined in the usual way from the equation (9). It is easy to see that the k^{th} row of equation (10) reduces to:

$$Z_{k+1}^{n+1} = \mathcal{G}(T_{k+1}, Z_k^{n+1}) + \mathcal{F}(T_{k+1}, Z_k^n, s) - \mathcal{G}(T_{k+1}, Z_k^n),$$

and can be shown that $\lim_{n\to\hat{k}} Z_{k+1}^n \to Z_{k+1}$, where $Z_{k+1} = \mathcal{F}(T_k, Z_k, s)$ for $0 \le k \le \hat{k}-1$; see [4, 2]. Let us define $\Phi_G := (I - A\tau)^{-1}$ and $\Phi_F := (I - A\tau)^{-\Delta T/\tau}$, with $\hat{\Phi}_G = I_{\hat{k}} \otimes \Phi_G$ and $\hat{\Phi}_F = I_{\hat{k}} \otimes \Phi_F$. The error $\mathbf{e}^n := \{e_k^n\}_{k=1}^{\hat{k}} \in \mathbb{R}^{\hat{k}\hat{m}}$, where $e_k^n := z(T_k) - Z_k^n$, at the n^{th} sweep of the parareal algorithm can be expressed as:

$$\mathbf{e}^{n+1} = \left(\hat{\Phi}_F - \hat{\Phi}_G\right) \mathcal{T} \mathbf{e}^n \tag{11}$$

where $\mathcal{T} \in \mathbb{R}^{(\hat{k}\hat{m})\times(\hat{k}\hat{m})}$ is the block upper triangular Toeplitz matrix defined by $\mathcal{T} := \left\{ \Phi_G^{j-i-1} \text{ if } j > i;\ 0 \text{ otherwise} \right\}$. Employing the spectral decomposition of matrix $A = U\Lambda U^T$, with $\Lambda := \text{diag}\{\lambda_m\}$, $1 \le m \le \hat{m}$, equation (11) can be rewritten as:

$$\boldsymbol{\xi}^{n+1} = \left(\hat{\Lambda}_F - \hat{\Lambda}_{\mathcal{G}}\right) \hat{\mathcal{T}} \boldsymbol{\xi}^n \text{ where } \hat{\mathcal{T}} := \left\{ \Lambda_G^{j-i-1} \text{ if } j > i;\ 0 \text{ otherwise} \right\}. \tag{12}$$

where $\boldsymbol{\xi}^n := V^T \mathbf{e}^n$ and $V = I_{\hat{k}} \otimes U$. Here, $\Lambda_F := \text{diag}\{(1 - \lambda_m\tau)^{-\Delta T/\tau}\}$, $\Lambda_G := \text{diag}\{\beta_m\}$, and $\beta_m := (1 - \lambda_m \Delta T)^{-1}$ are $\mathbb{R}^{\hat{m}\times\hat{m}}$ diagonal matrices, while $\hat{\Lambda}_F := I_{\hat{k}} \otimes \Lambda_F$ and $\hat{\Lambda}_G := I_{\hat{k}} \otimes \Lambda_G$ are $\mathbb{R}^{(\hat{k}\hat{m})\times(\hat{k}\hat{m})}$. We now permute system (12) by reordering the eigenvalues, obtaining $\hat{m}$ systems of the form:

$$\zeta_m^{n+1} = \left((1 - \lambda_m\tau)^{-\Delta T/\tau} - \beta_m\right) \tilde{\mathcal{T}}(\beta_m) \zeta_m^n, \tag{13}$$

where the Toeplitz matrix $\tilde{\mathcal{T}}(\beta)$ of size $\hat{k}$ is given by

$$\tilde{\mathcal{T}}(\beta) := \left\{ \beta^{j-i-1} \text{ if } j > i,\ 0 \text{ otherwise} \right\}.$$

Remark 1. Note that each eigenvalue λ_m is negative, hence:

$$\beta_m = (1 - \lambda_m \Delta T)^{-1} \le (1 - \lambda_m \Delta T)^{-\Delta T/\tau} \le e^{\lambda_m \Delta T},$$

and therefore $|\,(1 - \lambda_m\tau)^{-\Delta T/\tau} - \beta_m| \le |e^{\lambda_m \Delta T} - \beta_m|$.

Remark 2. Let $\mathbf{z}$ denote the exact discrete solution to $E\,\mathbf{z} = \mathbf{v}$ and let $z^n(t_l) = \mathcal{F}(t_l, Z_k^n, v)$ for $t_l \in [T_k, T_{k+1}]$. The spectral decomposition of $z(t_l)$ and $z^n(t_l)$ yields $z(t_l) = \sum_{m=1}^{\hat{m}} \alpha_m(t_l) q_m$ and $z^n(t_l) = \sum_{m=1}^{\hat{m}} \alpha_m^n(t_l) q_m$, respectively. Hence, ζ_m^n in (13) can also be represented as $\zeta_m^n(T_k) = \alpha_m(T_k) - \alpha_m^n(T_k)$.

The following lemma states the convergence of the parareal algorithm for system (13); see [2].

Lemma 1. *Let $t_f < \infty$, $t_o = 0$, $\Delta T = t_f/\hat{k}$, $T_k = k\Delta T$ for $0 \le k \le \hat{k}$. Then*

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

where $\rho_n = \max_{0<\beta<1} \left(e^{1-1/\beta} - \beta\right)^n \frac{1}{n!} \left| \frac{d^{n-1}}{d\beta^{n-1}} \left(\frac{1-\beta^{\hat{k}-1}}{1-\beta} \right) \right| \le 0.2984^n$.

The next theorem shows spectral equivalence between the linear operators $E_n K^{-1} E_n^T$ and $E K^{-1} E^T$.

Theorem 1. *Let E_n be the n^{th} application of the parareal scheme and $\epsilon \in (0, 1/2)$. Then*

$$\gamma_{\min}\left(\mathbf{v}, E^{-T}KE^{-1}\mathbf{v}\right) \leq \left(\mathbf{v}, E_n^{-T}KE_n^{-1}\mathbf{v}\right) \leq \gamma_{\max}\left(\mathbf{v}, E^{-T}KE^{-1}\mathbf{v}\right) \quad (14)$$

where $\gamma_{\max} := \left(1 + \frac{t_f\rho_n}{4\tau\epsilon} + 2\epsilon\right)/(1-2\epsilon)$, $\gamma_{\min} := \left(1 - \frac{t_f\rho_n}{4\tau\epsilon} - 2\epsilon\right)/(1+2\epsilon)$.

Proof. Let $\mathbf{z} = E^{-1}\mathbf{v}$ and $z(t) = \sum_{m=1}^{\hat{m}} \alpha_m(t) q_m$. Then:

$$(\mathbf{v}, E^{-T}KE^{-1}\mathbf{v}) = \|z\|^2_{L^2(0,t_f;L^2(\Omega))} = \sum_{m=1}^{\hat{m}} \|\alpha_m\|^2_{L^2(0,t_f)}.$$

Similarly, we have $(\mathbf{v}, E_n^{-T}KE_n^{-1}\mathbf{v}) = \|z^n\|^2_{L^2(0,t_f;L^2(\Omega))} = \sum_{m=1}^{\hat{m}} \|\alpha_m^n\|^2_{L^2(0,t_f)}$.

An upper bound for $\|E_n^{-1}\mathbf{v}\|^2_{L^2(0,t_f)}$ can be obtained as follows:

$$\begin{aligned}
\|\alpha_m^n\|^2_{L^2(0,t_f)} &= (\alpha_m^n - \alpha_m, \alpha_m^n + \alpha_m)_{L^2(0,t_f)} + \|\alpha_m\|_{L^2(0,t_f)} \\
&\leq \frac{1}{4\epsilon}\|\alpha_m^n - \alpha_m\|^2_{L^2(0,t_f)} + \epsilon\|\alpha_m^n + \alpha_m\|^2_{L^2(0,t_f)} + \|\alpha_m\|^2_{L^2(0,t_f)} \\
&\leq \frac{1}{4\epsilon}\|\alpha_m^n - \alpha_m\|^2_{L^2(0,t_f)} + 2\epsilon\|\alpha_m^n\|^2_{L^2(0,t_f)} + (1+2\epsilon)\|\alpha_m\|^2_{L^2(0,t_f)},
\end{aligned}$$

which reduces to

$$(1-2\epsilon)\|\alpha_m^n\|^2_{L^2(0,t_f)} \leq (1+2\epsilon)\|\alpha_m\|^2_{L^2(0,T)} + \frac{1}{4\epsilon}\|\alpha_m^n - \alpha_m\|^2_{L^2(0,t_f)}.$$

To obtain a bound for $\|\alpha_m^n - \alpha_m\|^2_{L^2(0,t_f)}$ we use that:

$$\|\alpha_m^n - \alpha_m\|^2_{L^2(0,t_f)} = \sum_{k=0}^{\hat{k}-1} \|\alpha_m^n - \alpha_m\|^2_{L^2(T_k,T_{k+1})} \leq \sum_{k=0}^{\hat{k}-1} \Delta T|\alpha_m^n(T_k) - \alpha_m(T_k)|^2$$

since $|\alpha_m^n(t_l) - \alpha_m(t_l)| = (1-\tau\lambda)^{-(t_l - T_k)/\tau}|\alpha_m^n(T_k) - \alpha_m(T_k)|$ and $(1-\tau\lambda)^{-(t_l-T_k)/\tau} \leq 1$. Hence:

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

Using Lemma 1 with the starting guess $\alpha_m^n(T_k) = 0$, $0 \leq k \leq \hat{k}$ we obtain:

$$\frac{t_f}{4\epsilon}\rho_n \max_{0\leq k\leq\hat{k}}|\alpha_m(T_k)|^2 \leq \frac{t_f}{4\epsilon\tau}\rho_n \max_{0\leq k\leq\hat{k}}\|\alpha_m(T_k)\|^2_{L^2(0,t_f)},$$

and

$$(1-2\epsilon)\sum_{m=1}^{\hat{m}}\|\alpha_m\|^2_{L^2(0,t_f)} \leq (1+2\epsilon+\frac{t_f\rho_n}{4\epsilon\tau})\sum_{m=1}^{\hat{m}}\|\alpha_m\|^2_{L^2(0,t_f)},$$

and so the upper bound (14) follows. The lower bound follows using the same procedure.

Table 1. Values of $\gamma_{max} - 1$ when τ is refined. The spatial discretization is $h = 1/10$ and $\Delta T = 1/20$.

$n \setminus \hat{l}$	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

Remark 3. Performing straightforward computations we obtain:

$$\min_{\epsilon} \gamma_{\max}(\epsilon) = 1 + \frac{4}{\sqrt{1 + \frac{4\tau}{\rho_n t_f}} - 1}.$$

Hence, for small values of ρ_n, we have $\gamma_{\max} - 1 \approx 2\sqrt{\frac{\rho_n t_f}{\tau}}$. The dependence of $\gamma_{max} - 1$ with respect to τ is sharp as evidenced in Table 1, see below, since it shows that $\gamma_{max} - 1$ increases by a $\sqrt{2}$ factor when the fine time step τ is refined by half.

4 Numerical Experiments

The optimal control problem we consider involves the 1D-heat equation:

$$y_t - y_{xx} = u, \;\; 0 < x < 1, \;\; t > 0$$

with boundary conditions $y(t, 0) = 0$ and $y(t, 1) = 0$ for $t \geq 0$, and initial data $y(0, x) = 0$ for $x \in [0, 1]$. The control variable u corresponds to the forcing term, and the performance function is $y_*(t, x) = x(1 - x)e^{-x}$ for $t \in [0, 1]$. Following [6], we employ a backward Euler discretization of the parabolic equation and choose $s = 0$, $q = 1$ and $r = 0.0001$. We let $\mathbf{r}_j$ denote the residual at iteration j and use $\|\mathbf{r}_j\|/\|\mathbf{r}_0\| \leq 10^{-6}$ as a stopping criteria for the iterative solvers.

Table 1 lists the value of $(\gamma_{\max} - 1)$ for different values of $\hat{l}$ and n (where $n = 7$ is equivalent to an exact solver). The results indicate that the method is *scalable* if "n" is kept constant. In addition, when "n" is increased, the number of MINRES iterations decreases. As a result, preconditioner $\mathcal{P}$ will yield optimal order convergence provided n is increases when τ is small.

Table 2 lists the number of MINRES iterations as $\Delta T/\tau$ is varied while mantaining constant τ. We chose $n = 2$. The number of iterations for the MINRES basically remains constant when h is refined and $\hat{k}$ is increased, and so the results indicate scalability. *Table 3* lists the number of MINRES iterations for $n = 2$ and $\tau = 1/512$ for different values of $\Delta T/\tau$. It indicates optimal order convergence.

Table 2. MINRES iterations using a parareal $n = 2$ as preconditioner. Parameters $q = 1$, $r = 0.0001$, $s = 0$, $t_f = 1$, $\Delta T/\tau = 16$. $n = 2/4/7$. Backward-Euler discretization is used in both fine grid and coarse grid.

$\hat{k}$	4	8	16	32
$\hat{l}$	64	128	256	512
$h = 1/16$	62 / 40 / 42	58 / 44 / 44	60 / 50 / 44	60 / 50 / 44
$h = 1/32$	60 / 42 / 42	58 / 44 / 44	60 / 50 / 44	62 / 50 / 44
$h = 1/64$	60 / 42 / 42	58 / 44 / 44	60 / 50 / 44	62 / 50 / 44

Table 3. MINRES iterations using a Parareal $n = 2$ as preconditioner. Parameters $q = 1$, $r = 0.0001$, $s = 0$, $t_f = 1$, $\tau = 1/512$. Backward-Euler discretization is used in both fine grid and coarse grid.

$\hat{k}$	8	16	32	64
$\Delta T/\tau$	64	32	16	8
$h = 1/16$	62	62	60	60
$h = 1/32$	62	62	62	60
$h = 1/64$	62	62	62	60

References

1. 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(2):687–713, 2005.
2. M. J. Gander and S. Vandewalle. On the super linear and linear convergence of the parareal algorithm. In *Lect. Notes Comput. Sci. Eng.*, 2005. Springer-Berlin.
3. M. Heinkenschloss. A time-domain decomposition iterative method for the solution of distributed linear quadratic optimal control problems. *J. Comput. Appl. Math.*, 173(1):169–198, 2005.
4. J. L. Lions, Y. Maday, and G. Turinici. Résolution d'edp par un schéma en temps pararéel. *C. R. Acad. Sci. Paris Sr. I Math.*, 332(7):661–668, 2001.
5. 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(8):1047–1053, 2001.
6. C. E. Schaerer, M. Sarkis, and T. Mathew. Block iterative algorithms for the solution of parabolic optimal control problems. In *Lecture Notes in Comput. Sci.*, 2006. Springer-Berlin.
7. M. Sarkis T. Mathew and C. E. Schaerer. Block matrix preconditioners foir elliptic optimal control problems. *To appear in Numer. Linear Algebra Appl.*, 2006.