Preconditioners based on ”Parareal” Time-Domain Decomposition for Time-dependent PDE-constrained Optimization

Stefan Ulbrich

Abstract We consider optimization problems governed by time-dependent parabolic PDEs and discuss the construction of parallel preconditioners based on the parareal method for the solution of quadratic subproblems which arise within SQP methods. In the case without control constraints, the optimality system of the subproblem is directly reduced to a symmetric PDE system, for which we propose a preconditioner that decouples into a forward and backward PDE solve. In the case of control constraints we apply a semismooth Newton method and apply the preconditioner to the semismooth Newton system. We prove bounds on the condition number of the preconditioned system which shows no or only a weak dependence on the size of regularization parameters for the control. We propose to use the parareal time domain decomposition method for the forward and backward PDE solves within the PDE preconditioner to construct an efficient parallel preconditioner. Numerical results show the efficiency of the approach.

1 Introduction

We consider parallel preconditioners for time-dependent PDE-constrained optimization problems of the form

$$\min_{y\in Y, u\in U} \quad f(y) + \frac{\alpha}{2}\|u\|_U^2 \quad \text{subject to} \quad E(y,u) = 0, \quad u \in U_{ad}, \tag{1}$$

where $u \in U$ is the control, $U_{ad} \subset U$ is closed and convex, $y \in Y \subset C([0,T];V)$ is a time-dependent state and $\alpha > 0$ is a regularization parameter. U is a Hilbert space, Y, V are Banach spaces, where $V \subset L^2(\Omega)$ with a

Department of Mathematics, Technische Universität Darmstadt, Dolivostr. 15, 64293 Darmstadt, Germany, e-mail: ulbrich@mathematik.tu-darmstadt.de

domain $\Omega \subset \mathbb{R}^n$. The state equation $E(y,u) = 0$ represents an appropriate (usually weak) formulation of a time-dependent PDE (or system of PDEs)

$$\begin{aligned} & y_t + F(t,x,y,u) = 0, \quad (t,x) \in \Omega_T := (0,T) \times \Omega \\ & y(0,x) = y_0(x), \quad x \in \Omega \end{aligned} \tag{2}$$

with initial data $y_0 \in V$. For convenience we assume that boundary conditions are incorporated in the state space Y. For notational convenience, we use the abbreviations $W = Y \times U$ and $w = (y,u)$.

Throughout the paper we will work under the following assumptions.

Assumption 1 *With $W := Y \times U$ the following holds for a given open convex set $W_0 = Y_0 \times U_0 \subset W$ containing the feasible set of* (1).

- *The mappings*

$$y \in Y \mapsto f(y), \quad (y,u) \in W \mapsto E(y,u) \in Z^*$$

  are continuously differentiable and the derivatives are uniformly bounded and Lipschitz on W_0.
- *For any $u \in U_0$ there exists a unique solution $y(u) \in Y_0$ of $E(y(u),u) = 0$.*
- *The derivative $E_y(y,u) \in \mathcal{L}(Y, Z^*)$ has an inverse that is uniformly bounded for all $(y,u) \in W_0$.*
- *$U_{ad} \subset U$ is closed and convex.*

In recent years, the design of efficient methods for the solution of PDE-constrained optimization problems (1) has received considerable attention, see for example [17, 29, 15, 11, 16, 13, 23, 28, 30, 31, 32, 33].

Usually, iterative solvers are applied to solve the arising linear systems and the inexactness is controlled by the globalization mechanism of the optimization method [13, 23, 33]. Optimization methods such as SQP- or interior point methods usually lead to auxiliary problems with a saddle point structure, e.g. the optimality system for the SQP subproblem or the primal-dual Newton system of interior point methods. To exploit the sparsity of these systems it is therefore of importance to have fast iterative solvers for these systems available, which are usually ill conditioned. Therefore, preconditioners are required to achieve fast convergence of iterative, often Krylov-based, solvers. The development of preconditioning and multigrid techniques for optimality systems in PDE-constrained optimization is an active reserach topic. First approaches for preconditioners of optimality systems have been proposed in [3, 4]. Block preconditioners for such systems have been proposed e.g. in [7, 25, 22, 34]. Multigrid preconditioners have for example been considered in [5, 6]. A time-domain While problems without inequality constraints are nowadays quite well understood, there are less results on the efficient preconditioning in the case of control and/or state constraints, see e.g. [14, 27, 21, 24]. While standard block preconditioners may depend strongly on critical parameters such as regularization or penalty parameter [14], the related preconditioners in

[21, 24] show only a weak dependence and in [24] estimates for the condition number are given, which apply to problems for control constraints and regularized state constraints.

Time-domain decomposition methods based on a block Gauss-Seidel iteration for linear-quadratic optimal control problems have been proposed and analyzed in [12] and block parareal preconditioners for such problems in [20].

In this work we build on the class of preconditioners proposed in [24] and extend their analysis to parabolic problems. The preconditioner decouples into two PDE solves. To obtain a parallel preconditioner, we use the time-domain decomposition method parareal to approximate the PDE solves within the preconditioner.

The parareal method was proposed in [18] as a parallel numerical scheme to solve evolution problems. The method is a time domain decomposition method and consists of a parallel predictor step based on a sufficiently exact propagator on the time slabs and a sequential corrector step computed by a coarse propagator. The algorithm has been successfully applied, e.g., to the Navier-Stokes equations and fluid-structure interaction problems [8, 9]. Its stability and convergence properties have for example been studied in [1, 10, 18, 19, 26]. In [10] it was shown that the parareal algorithm can be considered as a multiple shooting method as well as a time-multigrid method.

In this paper, we combine the class of preconditioners in [24], see also [21], with the parareal method to approximate the PDE solves within the preconditioner. We focus on the construction of preconditioners for the fast solution of subproblems arising in optimization methods for (1). For example SQP-type methods solve in the case $E_{yu} = 0$ (terms are otherwise often neglected) and $E_{uu} = 0$, which is often satisfied, given a current iterate $w^k = (y^k, u^k)$, $u^k \in U_{ad}$ subproblems of the form

$$\begin{aligned} &\min_{s=(s_y,s_u)\in W} q^k(s) := \langle f_y(y^k), s_y\rangle_{Y^*,Y} + \frac{1}{2}\langle s_y, H_k s_y\rangle_{Y,Y^*} + \frac{\alpha}{2}\|u^k + s_u\|_U^2 \\ &\text{subject to } \; E(w^k) + E_w(w^k)s = 0, \quad u^k + s_u \in U_{ad}, \end{aligned} \tag{3}$$

where $H^k \in \mathcal{L}(Y, Y^*)$ is an approximation of $L_{yy}(w^k, \lambda^k)$ with the Lagrangian function

$$L(y, u, \lambda) = f(y) + \alpha\|u\|_U^2 + \langle \lambda, E(y, u)\rangle_{Z,Z^*}.$$

In many practical algorithms H_k is chosen in such a way that the quadratic problem (3) is strictly convex. This is usually achieved by using $H_k = M_k^* M_k$, where $M_k \in \mathcal{L}(Y, Q)$, which we assume from now on. Under Assumption 1, the unique solution s_k satisfies with a Lagrange multiplier (adjoint state) $\lambda^k \in Z$ the following optimality system

$$E(w^k) + E_w(w^k)s^k = 0, \tag{4}$$

$$q_y^k(s^k) + E_y(w^k)^*\lambda^k = 0, \tag{5}$$

$$s_u^k \in U_{ad} - u^k, \langle q_u^k(s^k) + E_u(w^k)^*\lambda^k, s_u - s_u^k\rangle_{U^*,U} \geq 0 \;\forall\, s_u \in U_{ad} - u^k. \tag{6}$$

Since U is a Hilbert space and $U_{ad} \subset U$ is convex and closed, it is well known that with the identification $U = U^*$ the variational inequality (6) can equivalently be replaced by

$$s_u^k = P_{U_{ad}-u^k}(s_u^k - \gamma(q_u^k(s^k) + E_u(w^k)^*\lambda^k)) \tag{7}$$

with any fixed $\gamma > 0$ and the projection $P_{U_{ad}-u^k}$ in U onto $U_{ad} - u^k$.

We consider now two cases, the unconstrained case $U_{ad} = U$ and the box constrained case $U = L^2(\omega)$, $U_{ad} = \{u \in U;\, a \leq u \leq b \text{ a.e.}\}$ with $a, b \in U$, $a \leq b$.

In the case $U_{ad} = U$ optimality system (4)–(6) simplifies to the linear system

$$\begin{aligned} E(w^k) + E_w(w^k)s^k &= 0,\\ f_y(y^k) + H_k s_y^k + E_y(w^k)^*\lambda^k &= 0,\\ \alpha(u^k + s_u^k) + E_u(w^k)^*\lambda^k &= 0. \end{aligned}$$

Solving the last equation for s_u^k and inserting in the first equation yields the reduced optimality system

$$\begin{pmatrix} H_k & E_y(w^k)^* \\ E_y(w^k) & -\alpha^{-1}E_u(w^k)E_u(w^k)^* \end{pmatrix} \begin{pmatrix} s_y^k \\ \lambda^k \end{pmatrix} = \begin{pmatrix} -f_y(y^k) \\ -E(w^k) + E_u(w^k)u^k \end{pmatrix}.$$

In the box constrained case $U = L^2(\omega)$, $U_{ad} = \{u \in U;\, a \leq u \leq b \text{ a.e.}\}$ with $a, b \in U$, $a \leq b$, we set $\gamma = \alpha^{-1}$ in (7) and obtain

$$s_u^k = -u^k + \max(a, \min(b, -\alpha^{-1}E_u(w^k)^*\lambda^k)).$$

Inserting this in (4), (5), we arrive at the reduced optimality system

$$\left.\begin{aligned} f_y(y^k) + H_k s_y^k + E_y(w^k)^*\lambda^k = 0,\\ E(w^k) - E_u(w^k)u^k + E_y(w^k)s_y^k + \\ + E_u(w^k)\max(a, \min(b, -\alpha^{-1}E_u(w^k)^*\lambda^k)) = 0, \end{aligned}\right\} \quad G(s_y^k, \lambda^k) = 0.$$

It is well known that this is a semismooth system as long as $E_u(w^k)^* \in \mathcal{L}(Z, U)$ (note that we have identified $U^* = U$) satisfies in addition $E_u(w^k)^* \in \mathcal{L}(Z, L^p(\omega))$ for some $p > 2$, see [16, 30, 31]. If we now apply a semismooth Newton method then the iterates are given by $(s_y^{k,l+1}, \lambda^{k,l+1}) = (s_y^{k,l}, \lambda^{k,l}) + (\Delta y, \Delta\lambda)$, where $(\Delta y, \Delta\lambda)$ solves the linear system

$$\begin{pmatrix} H_k & E_y(w^k)^* \\ E_y(w^k) & -\alpha^{-1}E_u(w^k)D_{k,l}E_u(w^k)^* \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta\lambda \end{pmatrix} = -G(s_y^{k,l}, \lambda^{k.l}). \qquad (8)$$

Here, $D_{k,l} = d_{k,l} \cdot \mathrm{id}$ is a multiplication operator with $d_{k,l} \in L^\infty(\omega)$ defined by $d_{k,l} = 1_{\{a \le -\alpha^{-1}E_u(w^k)^*\lambda^{k,l} \le b\}}$.

We conclude that in both cases we have for given $w = (y,u) \in W$, $\lambda \in Z$ and $H \in \mathcal{L}(Y, Y^*)$ to solve linear systems of the form

$$\begin{pmatrix} H & E_y(w)^* \\ E_y(w) & -\alpha^{-1}E_u(w)DE_u(w)^* \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta\lambda \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix},$$

where $r_1 \in Y^*$, $r_2 \in Z^*$ and $D = \mathrm{id}$ in the case $U_{ad} = U$ or $D = 1_{\{a \le -\alpha^{-1}E_u(w)^*\lambda \le b\}} \cdot \mathrm{id}$ in the case $U_{ad} = \{u \in U;\ a \le u \le b \text{ a.e.}\}$, $U = L^2(\omega)$.

Hence, introducing the operators

$$A := E_y(w), \quad CC^* := \alpha^{-1}E_u(w)DE_u(w)^*$$

and using that we consider Hessian approximations of the form $H = M^*M$, we arrive at saddle point systems of the form

$$\begin{pmatrix} M^*M & A^* \\ A & -CC^* \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta\lambda \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}.$$

We note that also the application of interior point methods leads to a system of this structure.

Our aim is to develop a preconditioner for this type of systems. To this end, we use the preconditioner in [24] that requires essentially the solution of two linear systems with operators of the form $A + CIM$ and $(A + CIM)^*$, respectively. We then use the parareal time domain decomposition technique within the preconditioner to approximately solve these linear PDEs in parallel.

The paper is organized as follows. In section 2 we introduce the general preconditioner of [24] and extend its analysis to parabolic problems. We will derive estimates for the condition number which show only a weak dependence on critical parameters such as the regularization parameter α. We will treat the case with and without control constraints. In section 3 we will recall the parareal method and its basic convergence properties. In section 4 we propose a parareal based preconditioner by using the parareal algorithm as aproximate PDE solver within the preconditioner of section 2. Moreover, we present numerical results for parabolic control problems without and with control constraints. We end in section 5 with some conclusions.

2 A preconditioner for optimality systems

As we have seen, the solution of (3) leads - either directly or after applying a semismooth Newton or interior point method - to linear systems of the form

$$\begin{pmatrix} M^*M & A^* \\ A & -CC^* \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta\lambda \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}. \tag{9}$$

Here, A corresponds to the linearized forward PDE operator $E_y(w)$ and A^* to its adjoint.

In order to apply the preconditioner and its analysis from [24], we make the following assumptions.

Assumption 2 *(Basic Assumptions)*

1. *Assume that the state space Y and the space of adjoints Z are reflexive Banach spaces and the control space U as well as the space Q are Hilbert spaces.*
2. *Let $A \in \mathcal{L}(Y, Z^*)$ be an isomorphism, which implies that its Banach-space adjoint $A^* : Z \to Y^*$ is an isomorphism as well.*
3. *Let $M \in \mathcal{L}(Y, Q)$ with dense range and let $C \in \mathcal{L}(U, Z^*)$.*
   We denote by the (Hilbert space) adjoint $C^* : Z \to U^* = U$ the mapping that satisfies
   $$(C^*p, u)_U = \langle \lambda, Cu \rangle_{Z,Z^*} \quad \forall u \in U$$
   It is continuous as well. Analogously, the (Hilbert space) adjoint $M^* : Q = Q^* \to Y^*$ is defined via
   $$\langle M^*q, y \rangle_{Y^*,Y} = (My, q)_Q \quad \forall y \in Y.$$
   It is continuous and injective since M has dense range.

Moreover, we need the following assumption which requires some compatibility between the control space and the objective function.

Assumption 3 *Let I be a non-zero continuous mapping*

$$I : Q \to U.$$

with $\|I\|_{Q,U} \le 1$.

2.1 Development of the preconditioner

To motivate the preconditioner, we start with some observations. As we have already noted, $M^* : Q \to Y^*$ is injective, and hence $M^* : Q \to \operatorname{ran} M^*$ is a bijective operator with inverse $M^{-*} : \operatorname{ran} M^* \to Q$. We define the new space $\hat{Z} \subset Z$

$$\hat{Z} := A^{-*}(\operatorname{ran} M^*) = \{z \in Z : A^*z \in \operatorname{ran} M^*\} \subset Z.$$

Then $M^{-*}A^*z$ is well defined for all $z \in \hat{Z}$ and $M^{-*}A^* : \hat{Z} \to Q$ is a bijective mapping. The following lemma shows that $\hat{Z}$ becomes a Hilbert space with the scalar product

$$(z_1, z_2) \in \hat{Z}^2 \mapsto (z_1, z_2)_{\hat{Z}} := (M^{-*}A^*z_1, M^{-*}A^*z_2)_Q. \tag{10}$$

Moreover, the bilinear form

$$(z_1, z_2) \in \hat{Z}^2 \mapsto (z_1, z_2)_K := (M^{-*}A^*z_1, M^{-*}A^*z_2)_Q + (C^*z_1, C^*z_2)_U \tag{11}$$

is a scalar product that defines an equivalent norm on $\hat{Z}$.

By using this we will in Lemma 2 reduce (9) to a system that will be used to construct our preconditioner.

Lemma 1. *$\hat{Z}$ is a Hilbert space with the scalar product $(\cdot,\cdot)_{\hat{Z}}$ in (10). Moreover, $(\hat{Z}, (\cdot,\cdot)_{\hat{Z}})$ is continuously embedded in Z and $(\cdot,\cdot)_K$ in (11) is a scalar product that defines an equivalent norm on $\hat{Z}$.*

Proof. For $z_1, z_2 \in \hat{Z} \subset Z$ we have $A^*z_i \in \operatorname{ran} M^*$, $i = 1, 2$, and thus there are $q_1.q_2 \in Q$ with $A^*z_i = M^*q_i$. Hence, $(z_1, z_2)_{\hat{Z}} = (q_1, q_2)_Q < \infty$. Moreover, for $z \in \hat{Z}$ we have $(z, z)_{\hat{Z}} \geq 0$ and since $M^{-*}A^* : \hat{Z} \to Q$ is injective, $(z, z)_{\hat{Z}} = 0$ implies $z = 0$. Hence, $(\hat{Z}, (\cdot,\cdot)_{\hat{Z}})$ is a pre Hilbert space.

Let $z \in \hat{Z}$ be arbitrary. Then $A^*z = M^*q$ with some $q \in Q$ and

$$\|z\|_{\hat{Z}} = (z, z)_{\hat{Z}}^{\frac{1}{2}} = (q, q)_Q^{\frac{1}{2}} = \|q\|_Q. \tag{12}$$

Hence, $\|z\|_Z = \|A^{-*}M^*q\|_Z \leq \|A^{-*}M^*\|_{Q,Z}\|q\|_Q = \|A^{-*}M^*\|_{Q,Z}\|z\|_{\hat{Z}}$ and thus the embedding $(\hat{Z}, (\cdot,\cdot)_{\hat{Z}}) \hookrightarrow Z$ is continuous.

Finally, $(\hat{Z}, (\cdot,\cdot)_{\hat{Z}})$ is complete and thus a Hilbert space. In fact, any Cauchy sequence (z_k) in $(\hat{Z}, (\cdot,\cdot)_{\hat{Z}})$ satisfies $A^*z_k = M^*q_k$ with $q_k \in Q$ and (12) shows that (q_k) is a Cauchy sequence in Q and hence $q_k \to q$ in Q. This implies $A^*z_k = M^*q_k \to M^*q$ in Y^*. By the continuous embedding $(\hat{Z}, (\cdot,\cdot)_{\hat{Z}}) \hookrightarrow Z$, (z_k) is also a Cauchy sequence in Z and thus $z_k \to z$ in Z which implies $A^*z_k \to A^*z$ in Y^*. We conclude that $A^*z = M^*q \in \operatorname{ran} M^*$ and thus $z \in \hat{Z}$. This shows that $(\hat{Z}, (\cdot,\cdot)_{\hat{Z}})$ is complete.

Since $C^* \in \mathcal{L}(Z, U)$ and $(\hat{Z}, (\cdot,\cdot)_{\hat{Z}}) \hookrightarrow Z$, we have $C^* \in \mathcal{L}(\hat{Z}, U)$ and thus $(\cdot,\cdot)_K = (\cdot,\cdot)_{\hat{Z}} + (C^*\cdot, C^*\cdot)_U$ is a bounded bilinear form on $(\hat{Z}, (\cdot,\cdot)_{\hat{Z}})$. Moreover, for all $z \in \hat{Z}$ holds

$$(z, z)_{\hat{Z}} \leq (z, z)_{\hat{Z}} + (C^*z, C^*z)_U = (z, z)_K \leq (1 + \|C^*\|_{\hat{Z},U}^2)(z, z)_{\hat{Z}}$$

and thus $(\cdot,\cdot)_K$ is a scalar product on $\hat{Z}$ that induces an equivalent norm. $\square$

Lemma 2. *Let Assumption 2 hold. Then the system*

$$\begin{aligned}(M^{-*}A^*\hat{z}, M^{-*}A^*w)_Q + (C^*\hat{z}, C^*w)_U = \\ = -\langle r_2, w\rangle_{Z^*,Z} - (C^*A^{-*}r_1, C^*w)_U \quad \forall\, w \in \hat{Z}\end{aligned} \tag{13}$$

has a unique solution $\hat{z} \in \hat{Z}$ *and the solution of* (9) *can be obtained by*

$$\Delta\lambda = \hat{z} + A^{-*}r_1, \tag{14}$$

$$\Delta y = A^{-1}(r_2 + CC^*\Delta\lambda). \tag{15}$$

Proof. A proof can be found in [24, Lem. 2.4]. We give here a more constructive proof. Inserting $\Delta\lambda = \hat{z} + A^{-*}r_1$ in (9) yields the following system for $\hat{z}$.

$$\begin{pmatrix} M^*M & A^* \\ A & -CC^* \end{pmatrix} \begin{pmatrix} \Delta y \\ \hat{z} \end{pmatrix} = \begin{pmatrix} 0 \\ r_2 + CC^*A^{-*}r_1 \end{pmatrix}. \tag{16}$$

This shows that $A^*\hat{z} \in \operatorname{ran} M^*$ and therefore $\hat{z} \in \hat{Z}$. To derive a reduced system for $\hat{z}$, we perform block elimination with the second equation. This yields the equation

$$(A^* + M^*MA^{-1}CC^*)\hat{z} = -M^*MA^{-1}(r_2 + CC^*A^{-*}r_1)$$

and we observe that all terms are in $\operatorname{ran} M^*$. Applying the bijective operator $M^{-*} : \operatorname{ran} M^* \to Q$ yields the equivalent system

$$(M^{-*}A^* + MA^{-1}CC^*)\hat{z} = -MA^{-1}(r_2 + CC^*A^{-*}r_1).$$

Since $M^{-*}A^* : \hat{Z} \to Q$ is bijective, we obtain an equivalent variational equation if we use $M^{-*}A^*w$ with $w \in \hat{Z}$ as test functions. This leads to

$$\begin{aligned}(M^{-*}A^*w, M^{-*}A^*\hat{z})_Q + (M^{-*}A^*w, MA^{-1}CC^*\hat{z})_Q = \\ = -(M^{-*}A^*w, MA^{-1}(r_2 + CC^*A^{-*}r_1))_Q \quad \forall\, w \in \hat{Z},\end{aligned}$$

and since $A^{-*}M^*M^{-*}A^*w = w$ for all $w \in \hat{Z}$, we obtain (13). Hence, we have shown that (9) is with (14) equivalent to (16). Moreover, (16) is equivalent to (13) and the second equation in (16), while the latter is by (14) equivalent to (15). We conclude that (13), (14) and (15) are equivalent to (9).

The unique solvability of (13) follows from Lemma 1 and the Riesz representation theorem, since the left hand side can be written as $(\hat{z}, w)_K$, and $(\hat{Z}, (\cdot,\cdot)_K)$ is a Hilbert space by Lemma 1. □

Hence, we have seen that (9) can by (14), (15) be reduced to (13), which can by (11) be written as

$$(\hat{z}, w)_K = \langle r, w\rangle_{Z^*,Z} \quad \forall\, w \in \hat{Z}. \tag{17}$$

Clearly, this system can be solved by a preconditioned conjugate gradient method.

Following [24] we construct a preconditioner by approximatly decoupling the operator K induced by $(\cdot,\cdot)_K$ into the product of two PDE operators.

By Assumptions 2, 3 we have $CIM \in \mathcal{L}(Y, Z^*)$ and therefore $(CIM)^* = M^*I^*C^* \in \mathcal{L}(Y^*, Z)$. We consider the preconditioner $\hat{K} : \hat{Z} \to \hat{Z}^*$ defined by

$$(z, w)_{\hat{K}} := (M^{-*}(A + CIM)^* z, M^{-*}(A + CIM)^* w)_Q \quad \forall\, w \in \hat{Z}. \tag{18}$$

The application of the preconditioner is described in Algorithm 2.1.

Algorithm 2.1 Application of the preconditioner $\hat{\mathbf{K}}$

Input: $\ell \in \hat{Z}^*$
Output: Solution $z \in \hat{Z}$ *of* $(z, w)_{\hat{K}} = \langle \ell, w\rangle_{\hat{Z}^*,\hat{Z}}$ $\forall\, w \in \hat{Z}$

1. *Let* $q \in Q$ *be the solution of*

$$(q, M^{-*}(A + CIM)^* w)_Q = \langle \ell, w\rangle_{\hat{Z}^*,\hat{Z}} \quad \forall\, w \in \hat{Z}.$$

2. *Let* $z \in Z$ *be the solution of*

$$\langle (A + CIM)^* z, w\rangle_{Y^*,Y} = \langle M^* q, w\rangle_{Y^*,Y} \quad \forall\, w \in Y. \tag{19}$$

Note that the solution z satisfies $A^*z = M^*q - M^*I^*C^*z \in \operatorname{ran}(M^*)$ and therefore $z \in \hat{Z}$.

The following Lemma estimates the condition number of K relative to the preconditioner $\hat{K}$.

Lemma 3. *Consider the preconditioner* $\hat{K} : \hat{Z} \to \hat{Z}^*$ *defined in* (18)*. Assume that the quantity*

$$\gamma_{\hat{K}} := \sup_{0 \neq z \in \hat{Z}} \frac{(C^*z, C^*z)_U}{(z, z)_{\hat{K}}}$$

is finite. Then we have the estimate

$$\frac{1}{2}(z, z)_{\hat{K}} \leq (z, z)_K \leq (2 + 3\gamma_{\hat{K}})(z, z)_{\hat{K}} \quad \forall\, \hat{z} \in \hat{Z}. \tag{20}$$

Hence, the condition number of K *relative to* $\hat{K}$ *is bounded by* $\kappa_{\hat{K}} \leq 4 + 6\gamma_{\hat{K}}$*. Moreover, if* $((M^{-*}A^* + \frac{1}{2}I^*C^*)z, I^*C^*z)_Q \geq 0$ *for all* $z \in \hat{Z}$ *then we obtain the improved estimate*

$$\frac{1}{2}(z, z)_{\hat{K}} \leq (z, z)_K \leq (1 + \gamma_{\hat{K}})(z, z)_{\hat{K}}.$$

Proof. The first part was already shown in [24]. For convenience, we present a complete proof. We use the inequality $2(q_1, q_2)_Q \leq \|q_1\|_Q^2 + \|q_2\|_Q^2$. This yields

$$(z,z)_{\hat{K}} = \|M^{-*}A^*z + I^*C^*z\|_Q^2 \le 2(\|M^{-*}A^*z\|_Q^2 + \|I^*C^*z\|_Q^2)$$
$$\le 2(\|M^{-*}A^*z\|_Q^2 + \|C^*z\|_U^2) = 2(z,z)_K.$$

This shows the first inequality in (20) and the second follows from

$$\begin{aligned}(z,z)_K &= \|M^{-*}A^*z + I^*C^*z - I^*C^*z\|_Q^2 + \|C^*z\|_U^2 \\ &= (z,z)_{\hat{K}} + \|I^*C^*z\|_Q^2 + \|C^*z\|_U^2 - 2(M^{-*}A^*z + I^*C^*z, I^*C^*z)_Q \\ &\le 2(z,z)_{\hat{K}} + 3\|C^*z\|_U^2 = 2(z,z)_{\hat{K}} + 3\|C^*z\|_U^2 \le (2+3\gamma_{\hat{K}})(z,z)_{\hat{K}}.\end{aligned}$$

If $((M^{-*}A^* + \frac{1}{2}I^*C^*)z, I^*C^*z)_Q \ge 0$ then we can improve the first estimate in the last line to

$$\begin{aligned}(z,z)_K &= (z,z)_{\hat{K}} + \|I^*C^*z\|_Q^2 + \|C^*z\|_U^2 - 2(M^{-*}A^*z + I^*C^*z, I^*C^*z)_Q \\ &\le (z,z)_{\hat{K}} + \|C^*z\|_U^2 \le (1+\gamma_{\hat{K}})(z,z)_{\hat{K}}.\end{aligned}$$

□

As we will see, for parabolic operators there exists an appropriate imbedding operator $J \in \mathcal{L}(\hat{Z}, Y)$ such that

$$\langle (A + CIM)^*z, Jz\rangle_{Y^*,Y} > 0 \quad \forall\, 0 \ne z \in \hat{Z}. \tag{21}$$

Then we have the following result that will be helpful to estimate $\gamma_{\hat{K}}$ for practical applications.

Lemma 4. *Let Assumption 2 and 3 hold and assume that* (21) *is satisfied with an imbedding operator $J \in \mathcal{L}(\hat{Z}, Y)$. Then $\gamma_{\hat{K}}$ in Lemma 3 can be estimated by*

$$\gamma_{\hat{K}} \le \sup_{0\ne z\in\hat{Z}} \frac{\|C^*z\|_U^2 \|MJz\|_Q^2}{(\langle Jz, A^*z\rangle_{Y,Y^*} + (C^*z, IMJz)_U)^2}. \tag{22}$$

Proof. Let $0 \ne z \in \hat{Z}$ be arbitrary. Then $(A^* + M^*I^*C^*)z \in \operatorname{ran}(M^*)$ and thus

$$\begin{aligned}0 &< \langle (A^* + M^*I^*C^*)z, Jz\rangle_{Y^*,Y} = (M^{-*}(A^* + M^*I^*C^*)z, MJz)_Q \\ &\le \|M^{-*}(A^* + M^*I^*C^*)z\|_Q \|MJz\|_Q = (z,z)_{\hat{K}}^{\frac{1}{2}} \|MJz\|_Q.\end{aligned}$$

Hence, we obtain

$$\gamma_{\hat{K}} = \sup_{0\ne z\in\hat{Z}} \frac{\|C^*z\|_U^2 \|MJz\|_Q^2}{(z,z)_{\hat{K}} \|MJz\|_Q^2} \le \sup_{0\ne z\in\hat{Z}} \frac{\|C^*z\|_U^2 \|MJz\|_Q^2}{(\langle A^*z, Jz\rangle_{Y^*,Y} + (C^*z, IMJz)_U)^2}.$$

□

2.2 Application to parabolic control problems

We consider a parabolic state equation of the form

$$\begin{aligned} y_t - \nabla\cdot(\sigma\nabla y) + a_1\cdot\nabla y + a_0 y = b + Bu \quad & \text{on } \Omega_T := (0,T)\times\Omega, \\ y(0,\cdot) = y_0 \quad & \text{on } \Omega, \\ y = 0 \quad & \text{on } (0,T)\times\partial\Omega, \end{aligned} \tag{23}$$

where $\sigma \in L^\infty(\Omega)$, $a_0 \in L^\infty(\Omega_T)$, $a_1 \in (H^1\cap L^\infty)(\Omega_T)^n$, $\sigma \geq \sigma_0 > 0$ and $y_0 \in L^2(\Omega)$.

We set $V = H_0^1(\Omega)$, $W(0,T) := \left\{v \in L^2(0,T;V) : v_t \in L^2(0,T;V^*)\right\}$ and $Y := W(0,T)$. Let $Z = Z_1 \times Z_2 := L^2(0,T;V)\times L^2(\Omega)$ and assume that $b \in Z_1^*$ and $B \in \mathcal{L}(U, Z_1^*)$. We work with the usual weak solutions and define the operator $A \in \mathcal{L}(Y, Z^*)$ by

$$\begin{aligned} \langle Ay, (\lambda,\mu)\rangle_{Z^*,Z} = (y(0),\mu)_{L^2(\Omega)} + \int_0^T \big(\langle y_t(t), \lambda(t)\rangle_{V^*,V} \\ + (\sigma\nabla y(t), \nabla\lambda(t))_{L^2(\Omega)} + (a_1\cdot\nabla y(t) + a_0 y(t), \lambda(t))_{L^2(\Omega)}\big)\, dt. \end{aligned}$$

Then the state equation is given by

$$\langle Ay, (\lambda,\mu)\rangle_{Z^*,Z} = \int_0^T \langle b(t), \lambda(t)\rangle_{V^*,V}\, dt + (y_0,\mu)_{L^2(\Omega)} \quad \forall\, (\lambda,\mu) \in Z.$$

It is well known that $A \in \mathcal{L}(Y, Z^*)$ has a bounded inverse $A^{-1} \in \mathcal{L}(Z^*, Y)$. For $z = (\lambda,\mu) \in W(0,T)\times L^2(\Omega_T)$ we have

$$\begin{aligned} \langle y, A^*(\lambda,\mu)\rangle_{Y,Y^*} = (y(T), \lambda(T))_{L^2(\Omega)} + (y(0), \mu - \lambda(0))_{L^2(\Omega)} \\ + \int_0^T \big(-\langle y(t), \lambda_t(t)\rangle_{V,V^*} + (\sigma\nabla y(t), \nabla\lambda(t))_{L^2(\Omega)} \\ + (a_1\cdot\nabla y(t) + a_0 y(t), \lambda(t))_{L^2(\Omega)}\big)\, dt. \end{aligned} \tag{24}$$

2.2.1 Tracking type functional and distributed control

Now consider for example the case

$$f(y) = \tfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega_T)}, \quad U = L^2(\Omega_T), \quad B = I_{L^2(\Omega_T), L^2(0,T;V^*)} \tag{25}$$

with the natural imbedding $I_{L^2(\Omega_T),L^2(0,T;V^*)}$. Then $Q = U = L^2(\Omega_T)$, $I = \mathrm{id}_{L^2(\Omega_T)}$, $M = I_{Y,L^2(\Omega_T)}$, $M^* = I_{L^2(\Omega_T),Y^*}$, $\operatorname{ran} M^* = L^2(\Omega_T) \subset Y^*$ and $M^{-*} = \mathrm{id}_{L^2(\Omega_T)}$. Using (24) we see that

$$\hat{Z} = \{z \in Z : A^* z \in L^2(\Omega_T)\} \subset \{(\lambda, \lambda(0)) : \lambda \in W(0,T), \lambda(T) = 0\}. \tag{26}$$

In particular

$$\bar{Z} := \left\{(\lambda, \lambda(0)) : \lambda \in L^2(0,T; V \cap H^2(\Omega)), \lambda_t \in L^2(0,T; L^2(\Omega)), \lambda(T) = 0\right\}$$

is a dense subset of $\hat{Z}$ (and under additional regularity assumptions on Ω and the initial data it conincides with $\hat{Z}$).

In the case without control constraints we have

$$C = \alpha^{-\frac{1}{2}} I_{L^2(\Omega_T), L^2(0,T;V^*)}, \quad C^* = \alpha^{-\frac{1}{2}} I_{L^2(0,T;V), L^2(\Omega_T)}$$

and in the case with control constraints

$$C = \alpha^{-\frac{1}{2}} I_{L^2(\Omega_T), L^2(0,T;V^*)} D, \quad C^* = \alpha^{-\frac{1}{2}} D I_{L^2(0,T;V), L^2(\Omega_T)}, \tag{27}$$

with the multiplication operator $Dv = 1_{\mathcal{I}} v$ and the current estimate $\mathcal{I}$ of the inactive set. For a unified notation we set $\mathcal{I} = \Omega$ in the unconstrained case.

The application of the preconditioner in Algorithm 2.1 consists now of the following steps.

Algorithm 2.2 Preconditioner $\hat{\mathbf{K}}$ for parabolic problems

Input: $\ell \in \hat{Z}^*$
Output: Solution $z \in \hat{Z}$ *of* $(z, w)_{\hat{K}} = \langle \ell, w \rangle_{\hat{Z}^*, \hat{Z}}\ \forall\, w \in \hat{Z}$

1. *Compute the solution* $q \in Q = L^2(\Omega_T)$ *of*

$$(q, -w_t - \nabla\cdot(\sigma\nabla w + a_1 w) + (a_0 + \alpha^{-\frac{1}{2}} 1_{\mathcal{I}}) w)_{L^2(\Omega_T)} = \langle \ell, (w, w(0)) \rangle_{\hat{Z}^*, \hat{Z}} \quad \forall\, (w, w(0)) \in \bar{Z}.$$

   Note that we can use the dense subset $\bar{Z}$ *of* $\hat{Z}$ *as test space.*
2. *Compute a solution* $z = (\lambda, \lambda(0))$*,* $\lambda \in W(0,T)$*,* $\lambda(T) = 0$ *of*

$$\begin{aligned} &- \langle w, \lambda_t \rangle_{L^2(0,T;V), L^2(0,T;V^*)} + (\sigma\nabla w, \nabla\lambda)_{L^2(\Omega_T)} \\ &+ (a_1 \cdot \nabla w + (a_0 + \alpha^{-\frac{1}{2}} 1_{\mathcal{I}}) w, \lambda)_{L^2(\Omega_T)} = (q, w)_{L^2(\Omega_T)}\ \forall\, w \in Y. \end{aligned} \tag{28}$$

   Here we have already used the knowledge that the result $z \in Z$ *lives actually in* $\hat{Z}$*.*

We note that the application of the preconditioner decouples into the solution of two parabolic problems. Later we will apply the parareal time domain decomposition method to perform these two solves in parallel.

We obtain the following estimate for the condition number of the system (13) with preconditioner (18).

Theorem 1. *Consider the system* (8) *written in the form* (9) *for the linearized state equation* (23) *and the objective function* (25)*. Then the condition number* $\kappa_{\hat{K}}$ *of the operator* K *in the reduced system* (13) *relative to the* *preconditioner* $\hat{K}$ *in* (18) *is bounded by*

$$\kappa_{\hat{K}} \le 4 + 6\gamma_{\hat{K}}, \quad \gamma_{\hat{K}} \le \begin{cases} e^{cT} & \text{in the case without control constraints,} \\ \frac{e^{cT}}{2c_P\sigma_0\alpha^{\frac{1}{2}}} & \text{in the case with control constraints,} \end{cases}$$

where c_P *is a Poincaré constant on* Ω *and*

$$c = \max\left\{0, \tfrac{1}{\sigma_0}\|a_1\|^2_{L^\infty(\Omega_T)} - 2\,\mathrm{essinf}_{\Omega_T}\, a_0\right\} \tag{29}$$

In particular in the case $a_0 \ge 0$, $a_1 = 0$ *we have* $c = 0$.

The estimate for the control constrained case applies also to the case $U = L^2((0,T)\times\Omega_c)$ *with a smaller control domain* $\Omega_c \subsetneq \Omega$.

Proof. We apply Lemma 3 and Lemma 4, where we choose the operator $J \in \mathcal{L}(\hat{Z}, Y)$ in Lemma 4 by

$$J : (\lambda(t), \mu) \mapsto e^{ct}\lambda(t)$$

with c as in (29). Then we have with $z = (\lambda, \lambda(0)) \in \hat{Z}$

$$\begin{aligned}
\langle A^* z, Jz\rangle_{Y^*,Y} &= \frac{1}{2}\left(\langle A^* z, Jz\rangle_{Y^*,Y} + \langle z, AJz\rangle_{Z,Z^*}\right) = \frac{\|\lambda(0)\|^2_{L^2} + e^{cT}\|\lambda(T)\|^2_{L^2}}{2} \\
&+ \int_0^T e^{ct}\big((\sigma\nabla\lambda(t), \nabla\lambda(t))_{L^2} + (a_1\cdot\nabla\lambda(t) + (\tfrac{1}{2}c + a_0)\lambda(t), \lambda(t))_{L^2}\big)\,dt \\
&\ge \int_0^T e^{ct}\big(\tfrac{\sigma_0}{2}\|\nabla\lambda(t)\|^2_{L^2} + (\tfrac{1}{2}c + a_0 - \tfrac{\|a_1\|^2_{L^\infty}}{2\sigma_0})\|\lambda(t)\|^2_{L^2}\big)\,dt \\
&\ge \frac{\sigma_0}{2}\|e^{ct/2}\nabla\lambda(t)\|^2_{L^2(0,T;L^2(\Omega))} \ge \frac{c_P\sigma_0}{2}\|MJ^{\frac{1}{2}}z\|^2_Q
\end{aligned}$$

with a Poincaré constant c_P. Using the concrete definition of M and C (22) yields for the unconstrained case

$$\begin{aligned}
\gamma_{\hat{K}} &\le \sup_{0\ne z\in\hat{Z}} \frac{\|C^* z\|^2_U\|MJz\|^2_Q}{(\langle Jz, A^*z\rangle_{Y,Y^*} + (C^*z, IMJz)_U)^2} \\
&\le \sup_{0\ne(\lambda,\lambda(0))\in\hat{Z}} \frac{\alpha^{-1}\|\lambda\|^2_{L^2(\Omega_T)}\|e^{ct}\lambda\|^2_{L^2(\Omega_T)}}{(\frac{c_P\sigma_0}{2}\|e^{ct/2}\lambda\|^2_{L^2(\Omega_T)} + \alpha^{-\frac{1}{2}}\|e^{ct/2}\lambda\|^2_{L^2(\Omega_T)})^2} \\
&= \sup_{0\ne(\lambda,\lambda(0))\in\hat{Z}} \frac{\alpha^{-1}\|e^{-ct/2}\lambda\|^2_{L^2(\Omega_T)}\|e^{ct/2}\lambda\|^2_{L^2(\Omega_T)}}{(\frac{c_P\sigma_0}{2}\|\lambda\|^2_{L^2(\Omega_T)} + \alpha^{-\frac{1}{2}}\|\lambda\|^2_{L^2(\Omega_T)})^2} \le e^{cT}.
\end{aligned}$$

In the constrained case we obtain with the inactive set $\mathcal{I} \subset \Omega_T$

$$\begin{aligned}
\gamma_{\hat{K}} &\le \sup_{0\neq z\in\hat{Z}} \frac{\|C^*z\|_U^2\|MJz\|_Q^2}{(\langle Jz, A^*z\rangle_{Y,Y^*} + (C^*z, IMJz)_U)^2} \\
&\le \sup_{0\neq(\lambda,\lambda(0))\in\hat{Z}} \frac{\alpha^{-1}\|\lambda\|^2_{L^2(\mathcal{I})}\|e^{ct}\lambda\|^2_{L^2(\Omega_T)}}{(\frac{c_P\sigma_0}{2}\|e^{ct/2}\lambda\|^2_{L^2(\Omega_T)} + \alpha^{-\frac{1}{2}}\|e^{ct/2}\lambda\|^2_{L^2(\mathcal{I})})^2} \\
&= \sup_{0\neq(\lambda,\lambda(0))\in\hat{Z}} \frac{\alpha^{-1}\|e^{-ct/2}\lambda\|^2_{L^2(\mathcal{I})}\|e^{ct/2}\lambda\|^2_{L^2(\Omega_T)}}{(\frac{c_P\sigma_0}{2}\|\lambda\|^2_{L^2(\Omega_T)} + \alpha^{-\frac{1}{2}}\|\lambda\|^2_{L^2(\mathcal{I})})^2} \le \frac{e^{cT}}{2c_P\sigma_0\alpha^{\frac{1}{2}}}.
\end{aligned}$$

This estimate case applies also to the case $U = L^2((0,T)\times\Omega_c)$ with a smaller control domain $\Omega_c \subsetneq \Omega$, since we can choose $\mathcal{I} = \Omega_c$ in the unconstrained case and $\mathcal{I} \subset \Omega_c$ in the control constrained case. □

Now consider the case

$$f(y) = \tfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega_T)} + \tfrac{1}{2}\|y(T) - y_{d,T}\|^2_{L^2(\Omega)}, \quad U = L^2(\Omega_T). \tag{30}$$

Then $Q = L^2(\Omega_T) \times L^2(\Omega)$, $I = \mathrm{id}_{L^2(\Omega_T)}$, $M : y \in Y \mapsto (y, y(T)) \in Q$, $M^*q = (q_1, \cdot)_{L^2(\Omega_T)} + (q_2, \cdot(T))_{L^2(\Omega)}$. Using (24) we see that

$$\hat{Z} = \{z \in Z : A^*z \in \mathrm{ran}(M^*)\} \subset \{(\lambda, \lambda(0)) : \lambda \in W(0,T)\}.$$

In particular

$$\bar{Z} := \left\{(\lambda, \lambda(0)) : \lambda \in L^2(0,T; V \cap H^2(\Omega)), \lambda_t \in L^2(0,T; L^2(\Omega))\right\}$$

is a dense subset of $\hat{Z}$.

Theorem 2. *Under the assumptions of Theorem 1 but with the objective function* (30) *the condition number* $\kappa_{\hat{K}}$ *of the operator* K *in the reduced system* (13) *relative to the preconditioner* $\hat{K}$ *in* (18) *is in the case with and without control constraints bounded by*

$$\kappa_{\hat{K}} \le 4 + 6\gamma_{\hat{K}}, \quad \gamma_{\hat{K}} \le \frac{e^{cT}}{2\min\{1, c_P\sigma_0\}\alpha^{\frac{1}{2}}}$$

where c_P *is a Poincaré constant on* Ω *and* c *is given by* (29). *In particular in the case* $a_0 \ge 0$, $a_1 = 0$ *we have* $c = 0$.

The estimate applies also to the case $U = L^2((0,T)\times\Omega_c)$ *with a smaller control domain* $\Omega_c \subsetneq \Omega$.

Proof. We obtain exactly as in the proof of Theorem 1 for $z = (\lambda, \lambda(0)) \in \hat{Z}$

$$\langle A^*z, Jz\rangle_{Y^*,Y} = \frac{1}{2}\left(\langle A^*z, Jz\rangle_{Y^*,Y} + \langle z, AJz\rangle_{Z,Z^*}\right) = \frac{\|\lambda(0)\|^2_{L^2} + e^{cT}\|\lambda(T)\|^2_{L^2}}{2}$$
$$+ \int_0^T e^{ct}\big((\sigma\nabla\lambda(t), \nabla\lambda(t))_{L^2} + (a_1\cdot\nabla\lambda(t) + (\tfrac{1}{2}c + a_0)\lambda(t), \lambda(t))_{L^2}\big)\, dt$$
$$\geq \frac{\|\lambda(0)\|^2_{L^2} + e^{cT}\|\lambda(T)\|^2_{L^2}}{2} + \frac{c_P\sigma_0}{2}\|e^{ct/2}\lambda\|^2_{L^2}.$$

with a Poincaré constant c_P. Using the concrete definition of M and C (22) yields for $\mathcal{I} = \Omega_T$ in the unconstrained case and $\mathcal{I} \subset \Omega_T$ in the control constrained case

$$\gamma_{\hat K} \leq \sup_{0\neq z\in\hat Z} \frac{\|C^*z\|^2_U\|MJz\|^2_Q}{(\langle Jz, A^*z\rangle_{Y,Y^*} + (C^*z, IMJz)_U)^2}$$
$$\leq \sup_{0\neq(\lambda,\lambda(0))\in\hat Z} \frac{\alpha^{-1}\|\lambda\|^2_{L^2(\mathcal{I})}(\|e^{ct}\lambda\|^2_{L^2(\Omega_T)} + \|e^{cT}\lambda(T)\|^2_{L^2(\Omega)})}{(\frac{e^{cT}}{2}\|\lambda(T)\|^2_{L^2(\Omega)} + \frac{c_P\sigma_0}{2}\|e^{ct/2}\lambda\|^2_{L^2(\Omega_T)} + \alpha^{-\frac{1}{2}}\|e^{ct/2}\lambda\|^2_{L^2(\mathcal{I})})^2}$$
$$= \sup_{0\neq(\lambda,\lambda(0))\in\hat Z} \frac{\alpha^{-1}\|e^{-ct/2}\lambda\|^2_{L^2(\mathcal{I})}(\|e^{ct/2}\lambda\|^2_{L^2(\Omega_T)} + \|e^{cT/2}\lambda(T)\|^2_{L^2(\Omega)})}{(\frac{1}{2}\|\lambda(T)\|^2_{L^2(\Omega)} + \frac{c_P\sigma_0}{2}\|\lambda\|^2_{L^2(\Omega_T)} + \alpha^{-\frac{1}{2}}\|\lambda\|^2_{L^2(\mathcal{I})})^2}$$
$$\leq \frac{e^{cT}}{2\min\{1, c_P\sigma_0\}\alpha^{\frac{1}{2}}}.$$

This estimate case applies also to the case $U = L^2((0,T)\times\Omega_c)$ with a smaller control domain $\Omega_c \subsetneq \Omega$, since we can choose $\mathcal{I} = \Omega_c$ in the unconstrained case and $\mathcal{I} \subset \Omega_c$ in the control constrained case. □

2.2.2 An improved estimate of the condition number

We will now use regularity theory to improve the condition number estimate for the preconditioned system further. To this end, we extend the result of [24] for elliptic problems to parabolic problems.

We focus on the state equation (23) with objective function and control operator according to (25) and the case (27) of control constraints with current inactive set $\mathcal{I}$ or partial control domain $\mathcal{I} = (0,T)\times\Omega_c$ with $\Omega_c \subsetneq \Omega$. In this case we have as above

$$\gamma_{\hat K} = \sup_{0\neq z\in\hat Z} \frac{\|C^*z\|^2_{L^2(\Omega_T)}}{(z,z)_{\hat K}} = \sup_{0\neq(\lambda,\lambda(0))\in\hat Z} \frac{\|\alpha^{-\frac{1}{2}}1_{\mathcal{I}}\lambda\|^2_{L^2(\Omega_T)}}{\|A^*(\lambda,\lambda(0)) + \alpha^{-\frac{1}{2}}1_{\mathcal{I}}\lambda\|^2_{L^2(\Omega_T)}}. \tag{31}$$

We now derive an improved lower bound for $\|A^*(\lambda,\lambda(0)) + \alpha^{-\frac{1}{2}}1_{\mathcal{I}}\lambda\|_{L^2(\Omega_T)}$. We need the following regularity assumption.

Assumption 4 *The operator CIM is a multiplication operator of the form*

$$(CIM)v(t,x) = \phi(t,x)v(t,x), \quad \phi(t,x) = \phi_1 1_{\mathcal{I}}(t,x), \quad (t,x) \in \Omega_T,$$

where $\phi_1 > 0$ is a constant. Assume that $\mathcal{I} \subset \Omega$ and $\mathcal{A} := \Omega \backslash \mathcal{I}$ have Lipschitz boundary.

Moreover, we assume that $\hat{A}$ is cylinder-like, i.e., there exists a bi-Lipschitzian map $(t,x) \in \Omega_T \mapsto (t,\tau(t,x)) \in \Omega_T$ with $\tau(t,\mathcal{A}(t)) = \hat{\mathcal{A}}$ for all $t \in (0,T)$, where $\mathcal{A}(t) = \{x : (t,x) \in \mathcal{A}\}$, and denote by $\partial\mathcal{A}(t)$ the boundary of $\mathcal{A}(t)$ relative to Ω. Finally, we assume that for any $q \in L^2(\mathcal{A})$ the solution $(\lambda_{\mathcal{A}}, \mu_{\mathcal{A}}) \in L^2(0,T;H_0^1(\mathcal{A}(\cdot))) \times L^2(\mathcal{A}(0))$ of the problem

$$\langle w, A^*(\lambda_{\mathcal{A}}, \mu_{\mathcal{A}})\rangle_{Y,Y^*} = (w,q)_{L^2(\mathcal{A})} \ \forall\, w \in W_{\mathcal{A}}(0,T), \tag{32}$$

where $W_{\mathcal{A}}(0,T) = \left\{w \in L^2(0,T;H_0^1(\mathcal{A}(\cdot))) : w_t \in L^2(0,T;H_0^1(\mathcal{A}(\cdot)))^*\right\}$, satisfies $\lambda_{\mathcal{A}} \in W_{\mathcal{A}}(0,T) \cap L^2(0,T;H^{3/2+\varepsilon}(\mathcal{A}(\cdot))))$ and its normal trace the estimate

$$\|\partial_\nu \lambda_{\mathcal{A}}(t)\|_{L^2(\partial\mathcal{A}(t))} \le c_{tr,1}\|q(t)\|_{L^2(\mathcal{A}(t))}. \tag{33}$$

with a constant $c_{tr,1}$ independent of q and $t \in (0,T)$.

We note that for y_0, σ and $\mathcal{A}$ sufficiently regular this follows from parabolic regularity theory. We believe that also sets $\mathcal{A}$ that are not cylinder-like could be handled but leave this to future work.

Lemma 5. *Let $\mathcal{J}$ be an open domain with Lipschitz boundary. Then the following trace estimate holds for all $v \in H^1(\mathcal{J})$ with a constant $c_{tr,2}$*

$$\|v\|_{L^2(\partial\mathcal{J})} \le c_{tr,2}\sqrt{\|v\|_{H^1(\mathcal{J})}\|v\|_{L^2(\mathcal{J})}}.$$

Proof. See [24, Lem. 5]. □

In the following lemma we will for $q \in L^2(\Omega_T)$ consider the problem to find $(\lambda, \lambda(0)) \in \hat{Z}$ with

$$\langle w, A^*(\lambda,\lambda(0))\rangle_{Y,Y^*} + (w,\phi_1\lambda)_{L^2(\mathcal{I})} = (w,q)_{L^2(\Omega_T)} \ \forall\, w \in W(0,T), \tag{34}$$

which corresponds with $\phi_1 = \alpha^{-\frac{1}{2}}$ to (19) and its concrete form (28). As observed in (26), (24) yields $\lambda(T) = 0$.

Lemma 6. *Consider problem (34) where $\phi_1 > 0$ and $\mathcal{I}$, $\mathcal{A} = \Omega_T \setminus \mathcal{I}$ have the properties defined in Assumption 4. Assume that $q \in L^2(\Omega_T)$. Then the solution λ of (34) satisfies*

$$\phi_1\|e^{ct/2}\lambda\|_{L^2(\mathcal{I})} \le e^{cT/2}\left(\|q\|_{L^2(\mathcal{I})} + c_{\mathcal{I}}\phi_1^{\frac{1}{4}}\|q\|_{L^2(\mathcal{A})}\right). \tag{35}$$

Here, c is defined in (29) and $c_{\mathcal{I}}$ depends on $c_{\mathrm{tr},1}$, $c_{\mathrm{tr},2}$, σ_0, and the Poincaré constant c_P of Ω. Note that $c = 0$ if $a_0 \ge 0$ and $a_1 = 0$.

Proof. As in the proof for the elliptic case [24] we split λ into two parts $\lambda = \lambda_0 + \lambda_1$. Here, λ_0 is the extension by zero of the solution $\lambda_{\mathcal{A}}$ of the problem (32). Then $\lambda_0|_{\mathcal{A}} = \lambda_{\mathcal{A}}$ and $\lambda_0|_{\mathcal{I}} = 0$. By using (24), we observe similarly as above that (24) yields $\lambda_{\mathcal{A}}(T) = 0$ and $\mu_{\mathcal{A}} = \lambda_{\mathcal{A}}(0)$. Under the regularity asserted by Assumption 4, (32) reads

$$\begin{aligned} &- \langle w, (\lambda_{\mathcal{A}})_t\rangle_{L^2(0,T;H_0^1(\mathcal{A}(\cdot))),L^2(0,T;H_0^1(\mathcal{A}(\cdot))^*)} + (\sigma\nabla w, \nabla\lambda_{\mathcal{A}})_{L^2(\mathcal{A})} \\ &+ (a_1\cdot\nabla w + a_0 w, \lambda_{\mathcal{A}})_{L^2(\mathcal{A})} = (q, w)_{L^2(\mathcal{A})}\, \forall\, w \in W_{\mathcal{A}}(0,T), \end{aligned} \tag{36}$$

where $\lambda_{\mathcal{A}}(T) = 0$.

By our trace estimate (33) we conclude

$$\|\partial_\nu \lambda_{\mathcal{A}}(t)\|_{L^2(\partial\mathcal{A}(t))} \le c_{tr,1}\|q(t)\|_{L^2(\mathcal{A}(t))}. \tag{37}$$

Since λ_0 is an extension by zero of $\lambda_{\mathcal{A}}$ and (36) is only valid for testfunctions $w \in W_{\mathcal{A}}(0,T)$, integration by parts on $\mathcal{A}$ shows that λ_0 satisfies

$$\begin{aligned} \langle w, A^*(\lambda_0, \lambda_0(0))\rangle_{Y,Y^*} - \int_0^T (w(t), \sigma\partial_\nu\lambda_{\mathcal{A}}(t))_{L^2(\partial\mathcal{A}(t))}\, dt = \ &(w,q)_{L^2(\mathcal{A})} \\ &\forall\, w \in W(0,T). \end{aligned} \tag{38}$$

Hence, if we define $\lambda_1 \in W(0,T)$ as the solution of the problem

$$\begin{aligned} \langle w, A^*(\lambda_1, \lambda_1(0))\rangle_{Y,Y^*} + (w, \phi_1\lambda_1)_{L^2(\mathcal{I})} + \int_0^T (w(t), \sigma\partial_\nu\lambda_{\mathcal{A}}(t))_{L^2(\partial\mathcal{A}(t))}\, dt \\ = (w,q)_{L^2(\mathcal{I})}\ \forall\, w \in W(0,T) \end{aligned} \tag{39}$$

we see by adding (38) and (39) that $\lambda = \lambda_0 + \lambda_1$ solves the original problem (34) (note that $\lambda_0|_{\mathcal{I}} = 0$).

To obtain an estimate for λ_1 we test (39) with $e^{ct}\lambda_1(t)$, where c is defined in (29). Then we get as in the proof of Theorem 1

$$\begin{aligned} &\|q\|_{L^2(\mathcal{I})}\|e^{ct}\lambda_1\|_{L^2(\mathcal{I})} + \int_0^T \|\sigma\partial_\nu\lambda_{\mathcal{A}}(t)\|_{L^2(\partial\mathcal{A}(t))}\|e^{ct}\lambda_1(t)\|_{L^2(\partial\mathcal{A}(t))}\, dt \\ &\ge \langle e^{ct}\lambda_1, A^*(\lambda_1, \lambda_1(0))\rangle_+ + \phi_1\|e^{ct/2}\lambda_1\|^2_{L^2(\mathcal{I})} \\ &\ge \frac{\sigma_0}{2}\|e^{ct/2}\nabla\lambda(t)\|^2_{L^2(0,T;L^2(\Omega))} + \phi_1\|e^{ct/2}\lambda_1\|^2_{L^2(\mathcal{I})} \\ &\ge \frac{c_P\sigma_0}{2}\|e^{ct/2}\lambda(t)\|^2_{L^2(0,T;H^1(\Omega))} + \phi_1\|e^{ct/2}\lambda_1\|^2_{L^2(\mathcal{I})}. \end{aligned} \tag{40}$$

By Lemma 5 we obtain

$$\|e^{ct}\lambda_1(t)\|_{L^2(\partial\mathcal{A}(t))} \le c_{tr,2}\sqrt{\|e^{ct}\lambda_1(t)\|_{H^1(\mathcal{I}(t))}\|e^{ct}\lambda_1(t)\|_{L^2(\mathcal{I}(t))}}.$$

Division of (40) by the square-root of its right-hand side yields with (37)

$$\phi_1^{\frac{1}{2}} \|e^{ct/2}\lambda_1\|_{L^2(\mathcal{I})} \le \Bigg(\|q\|_{L^2(\mathcal{I})} \|e^{ct}\lambda_1\|_{L^2(\mathcal{I})}$$
$$+ \int_0^T \|\sigma \partial_\nu \lambda_{\mathcal{A}}(t)\|_{L^2(\partial\mathcal{A}(t))} c_{tr,2} \sqrt{\|e^{ct}\lambda_1(t)\|_{L^2(\mathcal{I}(t))} \|e^{ct}\lambda_1(t)\|_{L^2(\mathcal{I}(t))}}\, dt \Bigg) \cdot$$
$$\cdot \left(\frac{c_P\sigma_0}{2} \|e^{ct/2}\lambda(t)\|^2_{L^2(0,T;H^1(\Omega))} + \phi_1 \|e^{ct/2}\lambda_1\|^2_{L^2(\mathcal{I})} \right)^{-\frac{1}{2}}$$
$$\le e^{cT/2} \left(\|q\|_{L^2(\mathcal{I})} \phi_1^{-\frac{1}{2}} + c_{\mathcal{I}} \|q\|_{L^2(\mathcal{A})} \phi_1^{-\frac{1}{4}} \right).$$

Since $\lambda = \lambda_1$ on $\mathcal{I}$ we obtain from this the estimate (35). Tracing back the constant $c_{\mathcal{I}}$, we notice that it depends only on $c_{\mathrm{tr},1}$, $c_{\mathrm{tr},2}$, σ_0, and c_P. □

We obtain the following improved estimate for the condition number of the system (13) with preconditioner (18).

Theorem 3. *Consider the system* (8) *written in the form* (9) *for the linearized state equation* (23) *and the objective function* (25). *Assume that* $\mathcal{I} \subset \Omega_T$ *satisfies Assumption 4, where* $\mathcal{I}$ *is either the current estimate of the inactive set for control constraints or* $\mathcal{I} = (0,T) \times \Omega_c$ *in the case* $U = L^2((0,T) \times \Omega_c)$ *with a smaller control domain* $\Omega_c \subsetneq \Omega$. *Then the condition number* $\kappa_{\hat{K}}$ *of the operator* K *in the reduced system* (13) *relative to the preconditioner* $\hat{K}$ *in* (18) *is bounded by*

$$\kappa_{\hat{K}} \le 4 + 6\gamma_{\hat{K}}, \quad \gamma_{\hat{K}} \le e^{cT}(1 + c_{\mathcal{I}}\alpha^{-\frac{1}{8}})^2,$$

where $c_{\mathcal{I}}$ *is the constant in Lemma 6 and* c *is defined in* (29). *In the case* $a_0 \ge 0$, $a_1 = 0$ *we have* $c = 0$.

Remark 1. For the control constrained case or the case of a local control domain the condition number estimate improves from $O(\alpha^{-\frac{1}{2}})$ in Theorem 1 to $O(\alpha^{-\frac{1}{4}})$. Hence, the dependence on the regularization parameter is quite week, if the active set or control set is regular enough.

Proof. We apply Lemma 3. For our problem we have by (31)

$$\gamma_{\hat{K}} = \sup_{0 \ne (\lambda,\lambda(0)) \in \hat{Z}} \frac{\|\alpha^{-\frac{1}{2}} 1_{\mathcal{I}}\lambda\|^2_{L^2(\Omega_T)}}{\|A^*(\lambda,\lambda(0)) + \alpha^{-\frac{1}{2}} 1_{\mathcal{I}}\lambda\|^2_{L^2(\Omega_T)}}.$$

Applying Lemma 6 with $\phi_1 = \alpha^{-\frac{1}{2}}$ and $q = A^*(\lambda,\lambda(0)) + \alpha^{-\frac{1}{2}} 1_{\mathcal{I}}\lambda$ we have

$$\|\alpha^{-\frac{1}{2}} 1_{\mathcal{I}}\lambda\|_{L^2(\Omega_T)} \le \alpha^{-\frac{1}{2}} \|e^{ct/2}\lambda\|_{L^2(\mathcal{I})} \le e^{cT/2}(1 + c_{\mathcal{I}}\alpha^{-\frac{1}{8}}) \|q\|_{L^2(\Omega_T)}$$

and thus

$$\gamma_{\hat{K}} \le e^{cT}(1 + c_{\mathcal{I}}\alpha^{-\frac{1}{8}})^2.$$

□

3 Parareal time-domain decomposition

For parabolic problems, the application of the preconditioner $\hat{K}$ in Algorithm 2.2 decouples into two parabolic PDE solves. Therefore, time domain decomposition techniques are directly applicable within the preconditioner. In the following, we will use the parareal time-domain decomposition method as approximate solvers within the preconditioner.

The parareal algorithm was proposed by Lions, Maday, and Turinici [18] to speed up the numerical solution of time dependent partial differential equations by using parallel computers with a sufficiently large number of processors.

3.1 Description of the parareal method

Although the preconditioner requires only the solution of linearized PDEs we describe the parareal method more generally for nonlinear PDEs. We consider a time-dependent PDE (or a system) of the general form

$$\begin{aligned} &y_t + F(t,x,y) = 0, \quad (t,x) \in \Omega_T := (0,T) \times \Omega, \\ &y(0,x) = y_0(x), \quad x \in \Omega, \end{aligned} \tag{41}$$

where $y : [0,T] \to V$ maps time to a Banach space $V \subset L^2(\Omega)$, $y_0 \in V$ are initial data and $F(t,x,y)$ is a possibly time dependent partial differential operator in the variables $x \in \Omega$. The parareal technique, which was originally proposed in [18] and slightly modified in [1, 2], uses a time-domain decomposition

$$0 = t_0 < t_1 < \ldots < t_N = T, \tag{42}$$

which is uniform in the sense that

$$\eta_0 \Delta T \le t_{n+1} - t_n \le \Delta T, \quad \text{where} \;\; \Delta T := \max_{0 \le n < N} t_{n+1} - t_n.$$

We assume that on each time domain $[t_n, t_{n+1}]$, $0 \le n < N$, there exists a unique solution propagator

$$g(t_n, \cdot): \quad v \in V \mapsto g(t_n, v) := y(t_{n+1}) \in V,$$

where y is the solution of

$$\begin{aligned} &y_t + F(t,x,y) = 0, \quad (t,x) \in (t_n, t_{n+1}) \times \Omega, \\ &y(t_n,x) = v(x), \quad x \in \Omega. \end{aligned} \tag{43}$$

Moreover, we assume that we have a coarse grid approximation $g_c(t_n, v)$ of the exact propagator $g(t_n, v)$ available.

Example 1. We will later use a backward Euler step

$$\frac{g_c(t_n, v) - v}{t_{n+1} - t_n} + F(t_{n+1}, x, g_c(t_n, v)) = 0$$

as coarse propagator. As we will see the dissipativity of the backward Euler discretization is useful to stabilize the parareal scheme.

The parareal method uses a multiple shooting reformulation of the initial value problem (41) corresponding to the time domain decomposition (42). It combines parallel fine grid propagators with a coarse propagator for the iterative solution of the multiple shooting reformulation of (41).

More precisely, the parareal algorithm is defined as follows.

Algorithm 3.1 "Parareal" Time Integration Method:

Compute approximations y_n^k of $y_n = y(t_n)$, $1 \leq n \leq N$, for the solution y of (41) as follows.

1. Initialize by coarse scheme: $y_0^1 = y_0$,

$$y_{n+1}^1 = g_c(t_n, y_n^1), \quad 0 \leq n < N.$$

2. For $k = 1, \ldots, K-1$: $y_0^{k+1} = y_0$

$$y_{n+1}^{k+1} = \underbrace{g_c(t_n, y_n^{k+1})}_{\text{predictor}} + \underbrace{\big(g(t_n, y_n^k) - g_c(t_n, y_n^k)\big)}_{\substack{\text{corrector, computable} \\ \text{in parallel on time slabs}}}, \quad 0 \leq n < N. \tag{44}$$

The predictor step has to be computed sequentially to propagate the new approximation y_n^{k+1} by the coarse propagator. The error is corrected by using the approximation y_n^k of the previous iteration and can thus be computed in parallel.

It is obvious that the exact solution $y_n = y(t_n)$ is a fixed point of the parareal scheme, since $y_{n+1} = g(t_n, y_n)$. Moreover, after N iterations the exact solution is obtained. However, to obtain an efficient scheme, we want to apply only a few parareal iterations and are interested in the convergence speed of the parareal method. In practice the exact propagator $g(t_n, y_n^k)$ is replaced by an approximation $g_f(t_n, y_n^k)$ obtained by a sufficiently accurate fine grid scheme.

The parareal scheme can directly be applied to nonlinear equations and is then a nonlinear iteration. Thus, user-provided nonlinear fine grid solvers can be used directly. Besides the fact that nonlinear solvers can be more robust and efficient for some problems the nonlinear parareal algorithm has the advantage that the state in the time slabs has not to be stored in contrast to, e.g., a Newton iteration with inner parareal solver for the linearized equation.

3.2 Convergence properties of the parareal algorithm

In the following, we collect several convergence results for the parareal scheme in the context of time-dependent PDEs. The first result shows that the parareal after k iterations can be considered as a scheme of order km provided the solution is smooth enough and the coarse propagator has order m and is Lipschitz.

Theorem 4. ([1]) *Let $V_k \subset V_{k-1} \subset \ldots \subset V_0 = V$ be a scale of Banach spaces. Assume that*

1. (41) *is stable in all spaces V_j, $0 \le j \le k$, i.e.*

$$\|y(t)\|_{V_j} \le C\|v_0\|_{V_j} \quad \forall\, v_0 \in V_j, \quad t \in [0,T].$$

2. *The coarse propagator g_c is Lipschitz in the sense that for $0 \le j < k$*

$$\max_{0\le n<N} \|g_c(t_n, v) - g_c(t_n, w)\|_{V_j} \le (1 + C\Delta T)\|v - w\|_{V_j} \quad \forall\, v, w \in V_j.$$

3. *The coarse propagator g_c has order m in the sense that for $0 \le j < k$ with $\delta g(t_n, v) := g(t_n, v) - g_c(t_n, v)$ the estimate holds*

$$\max_{0\le n<N} \|\delta g(t_n, v) - \delta g(t_n, w)\|_{V_j} \le C(\Delta T)^{m+1}\|v - w\|_{V_{j+1}} \quad \forall\, v, w \in V_{j+1}.$$

Then for initial data $y_0 \in V_k$ the parareal scheme after k iterations has an accuracy of order mk, i.e.,

$$\max_{1\le n\le N} \|y(t_n) - y_n^k\|_{V_0} \le C(\Delta T)^{mk}\|y_0\|_{V_k}$$

with C independent of ΔT and y_0.

Proof. See [1, Thm. 1]. Our assumption 3. is a modification of assumption (H3) in [1] and is adjusted directly to the proof of [1, Thm. 1]. □

The above result requires additional regularity of the solution in order to achieve order km. By considering dissipative coarse propagators, Bal [1] was able to obtain an improved result for certain linear partial differential operators by using Fourier analysis.

Theorem 5. ([1]) *Consider a linear M-th order spatial operator F in 1D with constant coefficients and symbol $P(\xi) = \alpha_0 + \sum_{j=1}^{M-1} \alpha_j \xi^j + |\xi|^M$, where $\alpha_j \in \mathbb{C}$ and $\alpha_0 \ge 0$ such that $P(\xi) \ge 0$ or $\Re(P(\xi)) > 0$ (e.g., heat equation, advection-diffusion equation). If for $t_n = n\Delta T$, $\Delta T = T/N$, the coarse propagator is given by the θ-scheme*

$$\frac{g_c(t_n, v) - v}{\Delta T} + A(\theta g_c(t_n, v) + (1 - \theta)v) = 0$$

then for $\theta \in (1/2, 1]$ *and* $k = O(1)$ *the parareal scheme is stable and*

$$\max_{1 \le n \le N} \|y(t_n) - y_n^k\|_{H^s(\mathbb{R})} \le C(\Delta T)^k \|v_0\|_{H^s(\mathbb{R})}.$$

for all $s \in \mathbb{R}$*. Here,* $H^s(\mathbb{R})$ *denotes the usual Sobolev space of order* s*.*

Proof. As shown in [1], the result follows from Theorem 2 and 3 in [1]. □

The convergence behavior with respect to the iteration index k was analyzed in [10], where in particular the following result was shown.

Theorem 6. *Consider the heat equation in 1D, i.e., the spatial operator has the symbol* $P(\xi) = \xi^2$*. If for* $t_n = n\Delta T$*,* $\Delta T = T/N$*, the coarse propagator is a one-step method that has a stability function* $R(z)$ *such that*

$$\gamma_l := \sup_{z \le 0} \frac{|e^z - R(z)|}{1 - |R(z)|} \in (0, 1),$$

(for exampe the backward Euler scheme) then the parareal scheme is stable and

$$\max_{1 \le n \le N} \|y(t_n) - y_n^k\|_{H^s(\mathbb{R})} \le \gamma_l^k \max_{1 \le n \le N} \|y(t_n) - y_n^0\|_{H^s(\mathbb{R})}$$

for all $s \in \mathbb{R}$*.*

Proof. In [10, Thm. 4.9,Thm. 5.1] the estimate is shown pointwise for the Fourier transforms. Integration with the appropriate weights yields the estimate in the Sobolev norms.

The stability function for the backward Euler scheme is $R(z) = \frac{1}{1-z}$ and yields $\gamma_l = 0.2984256075$, see [10]. □

3.3 Parareal as preconditioned iteration

For our purposes it is useful to note that the parareal scheme can one one hand be seen as as a scheme of higher order on the coarse time grid but can on the other hand also be intepreted as a preconditioned iteration. Consider for simplicity the case that (41) is linear. Then

$$g(t_n, y_n) = G_n y_n + c_n, \quad g_c(t_n, y_n) = G_{c,n} y_n + c_{c,n}.$$

and the relation $y_{n+1} = g(t_n, y_n)$, $0 \le n < N$ can be written as

$$My := \begin{pmatrix} I & & & & \\ -G_1 & I & & & \\ & -G_2 & I & & \\ & & \ddots & \ddots & \\ & & & -G_{N-1} & I \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} c_0 + G_0 y_0 \\ c_1 \\ \vdots \\ c_{N-1} \end{pmatrix} =: c.$$

If we rewrite the parareal scheme (44) as

$$y_{n+1}^{k+1} - g_c(t_n, y_n^{k+1}) = y_{n+1}^k - g_c(t_n, y_n^k) + g(t_n, y_n^k) - y_{n+1}^k, \quad 0 \leq n < N,$$

we see that with the coarse grid approximation M_c of M

$$M_c = \begin{pmatrix} I & & & & \\ -G_{c,1} & I & & & \\ & -G_{c,2} & I & & \\ & & \ddots & \ddots & \\ & & & -G_{c,N-1} & I \end{pmatrix}$$

the parareal scheme (44) can be written as

$$y^{k+1} = y^k + M_c^{-1}(c - My^k). \tag{45}$$

Theorem 4 and 5 show that under their assumptions for fixed k the operator $I - M_c^{-1}M$ has for sufficiently large N (i.e., ΔT small enough) a condition number $\ll 1$. On the other hand, Theorem 6 shows that for the heat equation we have

$$\|I - M_c^{-1}M\|_{\max_n \|(\cdot)_n\|_{H^s(\mathbb{R})}, \max_n \|(\cdot)_n\|_{H^s(\mathbb{R})}} \leq \gamma_l,$$

where for example $\gamma_l = 0.2984256075$, if the backward Euler scheme is selected as coarse propagator.

Hence, we see that under appropriate assumptions M_c^{-1} is a good preconditioner for the operator M.

In the nonlinear case we introduce the operators

$$\mathcal{G}(y) = \begin{pmatrix} y_1 - g(t_0, y_0) \\ \vdots \\ y_N - g(t_{N-1}, y_{N-1}) \end{pmatrix}, \quad \mathcal{G}_c(y) = \begin{pmatrix} y_1 - g_c(t_0, y_0) \\ \vdots \\ y_N - g_c(t_{N-1}, y_{N-1}) \end{pmatrix}.$$

Then the parareal scheme can be written as the preconditioned iteration

$$\mathcal{G}_c(y^{k+1}) = \mathcal{G}_c(y^k) - \mathcal{G}(y^k)$$

which is (45) in the linear case. Hence, as already observed in [10] the parareal scheme can also be seen as a deferred correction method. In [10] it is also shown that the parareal method can be interpeted as a time-multigrid method.

4 A parareal-based preconditioner for optimality systems and numerical results

We use now the parareal method as an inexact implementation of the two PDE solves in the preconditioner (18), see Algorithm 2.1 and Algorithm 2.2. As in section 2.2 we focus on the case of parabolic state equations. In the following we always assume that the exact propagator $g(t_n, y_n)$ is replaced by a fine grid propagator consisting of a time stepping scheme on a finer time grid within the time slab $[t_n, t_{n+1}]$.

4.1 Strategies to combine the parareal scheme and the preconditioner

There are essentially two different approaches to apply the parareal scheme within the proposed preconditioner (18):

In the first approach the parareal scheme with a fixed iteration number K is used as time discretization scheme for the state equation in (1) and also for the linearized state equation in (3). Then using the same parareal scheme and its adjoint scheme within the preconditioner (18) in Algorithm 2.2 is just the application of the preconditioner (18) to the discretized problem.

In the second approach the state equation in (1) and also the linearized state equation in (3) is discretized by the fine grid propagator and the parareal scheme with a fixed iteration number K is used within the preconditioner to solve the corresponding optimality system. The resulting preconditioner is then an inexact version of the preconditioner (18) in Algorithm 2.2, since the two PDE solves are approximated by K parareal iterations.

In the following we present numerical results for the second approach.

4.2 Numerical results for a parabolic optimal control problem

As an example for (3) we consider the problem

$$\min_{\substack{y\in W(0,T)\\ u\in L^2((0,T)\times\Omega)}} \quad \frac{\beta}{2}\int_\Omega (y(T)-y_T)^2\,dx+\frac{1}{2}\int_0^T\int_\Omega (y-y_d)^2\,dx\,dt+\frac{\alpha}{2}\int_0^T\int_\Omega u^2\,dx\,dt$$

$$\begin{aligned} \text{s.t.} \qquad y_t - \Delta y &= u \quad \text{on } (0,T)\times\Omega,\\ y|_{(0,T)\times\partial\Omega} &= 0, \\ y(0,\cdot) &= y_0 \quad \text{on } \Omega, \end{aligned} \tag{46}$$

where $\beta \geq 0$ is a constant. The state and control space are given by

$$U = L^2((0,T) \times \Omega),$$
$$Y = W(0,T) = \left\{ y : y \in L^2(0,T; H_0^1(\Omega)), \quad y_t \in L^2(0,T; H^{-1}(\Omega)) \right\}.$$

4.2.1 Propagators in the parareal scheme

For the implementation of the parareal method we use the equidistant time decomposition

$$T_n = n\Delta T, \quad \Delta T = \frac{T}{N}.$$

For the coarse propagator $g_\Delta(t_n, v; u)$ we apply one backward Euler step in time and use a standard 5-point stencil for the Laplacian, i.e., $y_{n+1} = g_c(t_n, v; u)$ is given by

$$\frac{y_{n+1} - v}{\Delta T} - A y_{n+1} = u_n.$$

To approximate the exact propagator $g(t_n, v; u)$ we apply N_f steps of a Crank-Nicholson scheme with time step $\Delta t = \Delta T / N_f$ and use a standard 5-point stencil for the Laplacian. Hence, $y_{n+1}^f = g_f(t_n, v; u)$ is given by

$$v_0 = v,$$
$$\frac{v_{j+1} - v_j}{\Delta t} - A \frac{v_{j+1} + v_j}{2} = u_{n,j}, \quad j = 0, \ldots, N_f - 1,$$
$$y_{n+1}^f = v_{N_f}.$$

Remark 2. For the solution of the time step equations in the propagators, standard multigrid solvers can directly be used. It would be possible to control the inexactness of the multigrid solvers by the generalized SQP-methods.

In addition, a coarser space grid could be used for the coarse propagator.

4.2.2 Numerical results

We consider the specific problem

$$T = 1, \quad \Omega = (0,1)^2, \quad y_0(x) = 0,$$
$$y_d(t,x) = \sin(t)\, x_1 x_2 (1 - x_1)(1 - x_2), \quad y_T(x) = y_d(T,x).$$

For the discretization we use $N = 20$ time domains, coarse time step $\Delta T = \frac{1}{N}$, fine time step $\Delta t = \frac{\Delta T}{40}$ and a 100×100 equidistant space grid.

Unconstrained problem

We apply a preconditioned cg method to the discretized version of (13) and use the preconditioner (18), see Algorithm 2.2, where we replace the exact discrete PDE-solves on the fine grid by K parareal iterations according to Algorithm 3.1, where we add a parallel fine grid solve on the time slabs for the final initial data y_n^K obtained from the parareal iterations. Then the costs of the K parareal iterations are K sequential coarse solves and K parallel fine grid solves on the N time slabs.

The iteration history for different choices of α and β is shown in Table 1 for $K = 2$ and $K = 3$. In both cases the preconditioned cg method is stopped if the relative residual is $\leq 10^{-2}$. In the case $K = 2$, $\beta = 1e\text{-}3$ and $\alpha \in [1e\text{-}$

$K = 2$ parareal its.			$K = 3$ parareal its.		
α	β	pcgits	α	β	pcgits
1e−3	1	15	1e−3	1	8
1e−4	1	7	1e−4	1	6
1e−5	1	4	1e−5	1	4
1e−6	1	5	1e−6	1	5
1e−3	1e−3	3	1e−3	1e−3	3
1e−4	1e−3	3	1e−4	1e−3	3
1e−5	1e−3	3	1e−5	1e−3	3
1e−6	1e−3	2	1e−6	1e−3	2

Table 1 Preconditioned conjugate gradient iterations for the solution of (13) with preconditioner (18) for a relative stopping tolerance 10^{-2}. The two PDE solves within the preconditioner are approximated by K parareal iterations.

$6, 1e\text{-}3]$ we obtain with $N = 20$ processors we obtain the following time ratio between the time for the parallel solution of the SQP subproblem and the serial solve of the state equation.

$$\frac{\text{time(parallel SQP solve)}}{\text{time(serial state solve)}} = \frac{3 \times 4 \times (20 \text{ coarse} + 40 \text{ fine steps})}{800 \text{ fine steps}} = 0.9.$$

Control constrained problem

We add the control constraint

$$u \geq \frac{1}{10}$$

and apply a semismooth Newton method. The semismooth Newton system is reduced to the system (13). As above, we apply a preconditioned cg method to the discretized version of (13) and use the preconditioner (18), see Algorithm 2.2, where we replace the exact discrete PDE-solves on the fine grid by K

parareal iterations according to Algorithm 3.1. We set $\beta = 1$ and use $K = 3$ parareal iterations within the preconditioner. The praconditioned cg method is terminated if the relative residual is $\leq 10^{-1}$.

The iteration history for different choices of α is shown in Table 2.

$\alpha = 1\text{e}{-3}$		$\alpha = 1\text{e}{-4}$	
Residual	pcgits	Residual	pcgits
1.3e+1	5	1.3e+2	5
3.6e−1	6	7.5e+1	3
1.3e−2	4	1.6e+1	6
3.9e−4	5	1.1e−1	8
2.3e−5	4	5.4e−3	5
		2.3e−4	4

Table 2 Semismooth Newton method, where the semismooth Newton system is solved by a preconditioned conjugate gradient method with preconditioner (18). The two PDE solves within the preconditioner are approximated by 3 parareal iterations.

5 Conclusions

We have proposed a parallel preconditioner for optimality systems or semismooth Newton systems arising in parabolic optimal control problems without or with control constraints. The preconditioner is based on [24], see also [21]. We extend the estimates for the condition number to the case of parabolic optimal control problems. The estimates show no or only a weak dependence on the regularization parameter α. Since the preconditioner decouples into two PDE solves, we propose to obtain a parallel preconditioner by using the parareal method as approximate PDE solver within the preconditioner. The efficiency and very weak dependence on the regularization parameter α is confirmed by numerical results. Since the preconditioner requires only fine and coarse propagators for a slightly modified forward and backward PDE, it is quite easy to implement. The numerical tests show that already for 2-3 parareal iterations within the preconditoner the number of preconditioned cg iterations is very moderate. Hence, the preconditioner has potential for the parallel solution of large scale optimal control problems.

References

1. Bal, G.: On the convergence and the stability of the parareal algorithm to solve partial differential equations. In: Domain decomposition methods in science and engineering, *Lect. Notes Comput. Sci. Eng.*, vol. 40, pp. 425–432. Springer, Berlin (2005)

2. Bal, G., Maday, Y.: A "parareal" time discretization for non-linear PDE's with application to the pricing of an American put. In: Recent developments in domain decomposition methods (Zürich, 2001), *Lect. Notes Comput. Sci. Eng.*, vol. 23, pp. 189–202. Springer, Berlin (2002)
3. Battermann, A., Heinkenschloss, M.: Preconditioners for Karush-Kuhn-Tucker matrices arising in the optimal control of distributed systems. In: Control and estimation of distributed parameter systems (Vorau, 1996), *Internat. Ser. Numer. Math.*, vol. 126, pp. 15–32. Birkhäuser, Basel (1998)
4. Battermann, A., Sachs, E.W.: Block preconditioners for KKT systems in PDE-governed optimal control problems. In: Fast solution of discretized optimization problems (Berlin, 2000), *Internat. Ser. Numer. Math.*, vol. 138, pp. 1–18. Birkhäuser, Basel (2001)
5. Borzì, A.: Smoothers for control- and state-constrained optimal control problems. Comput. Vis. Sci. **11**(1), 59–66 (2008)
6. Borzi, A., Schulz, V.: Multigrid methods for PDE optimization. SIAM Rev. **51**(2), 361–395 (2009)
7. Dollar, H.S., Gould, N.I.M., Stoll, M., Wathen, A.J.: Preconditioning saddle-point systems with applications in optimization. SIAM J. Sci. Comput. **32**(1), 249–270 (2010)
8. Farhat, C., Chandesris, M.: Time-decomposed parallel time-integrators: theory and feasibility studies for fluid, structure, and fluid-structure applications. Internat. J. Numer. Methods Engrg. **58**(9), 1397–1434 (2003)
9. Fischer, P.F., Hecht, F., Maday, Y.: A parareal in time semi-implicit approximation of the Navier-Stokes equations. In: Domain decomposition methods in science and engineering, *Lect. Notes Comput. Sci. Eng.*, vol. 40, pp. 433–440. Springer, Berlin (2005)
10. Gander, M.J., Vandewalle, S.: Analysis of the parareal time-parallel time-integration method. SIAM J. Sci. Comput. **29**(2), 556–578 (2007)
11. Gong, W., Hinze, M., Zhou, Z.J.: Space-time finite element approximation of parabolic optimal control problems. J. Numer. Math. **20**(2), 111–145 (2012)
12. Heinkenschloss, M.: 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)
13. Heinkenschloss, M., Vicente, L.N.: Analysis of inexact trust-region SQP algorithms. SIAM J. Optim. **12**(2), 283–302 (2001/02)
14. Herzog, R., Sachs, E.: Preconditioned conjugate gradient method for optimal control problems with control and state constraints. SIAM J. Matrix Anal. Appl. **31**(5), 2291–2317 (2010). DOI 10.1137/090779127. URL http://dx.doi.org/10.1137/090779127
15. Hintermüller, M., Hinze, M.: A SQP-semismooth Newton-type algorithm applied to control of the instationary Navier-Stokes system subject to control constraints. SIAM J. Optim. **16**(4), 1177–1200 (2006)
16. Hintermüller, M., Ito, K., Kunisch, K.: The primal-dual active set strategy as a semismooth Newton method. SIAM J. Optim. **13**(3), 865–888 (2003) (2002)
17. Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S.: Optimization with PDE constraints, *Mathematical Modelling: Theory and Applications*, vol. 23. Springer, New York (2009)
18. Lions, J.L., Maday, Y., Turinici, G.: Résolution d'EDP par un schéma en temps "pararéel". C. R. Acad. Sci. Paris Sér. I Math. **332**(7), 661–668 (2001)
19. Maday, Y., Turinici, G.: A parareal in time procedure for the control of partial differential equations. C. R. Math. Acad. Sci. Paris **335**(4), 387–392 (2002)
20. Mathew, T.P., Sarkis, M., Schaerer, C.E.: Analysis of block parareal preconditioners for parabolic optimal control problems. SIAM J. Sci. Comput. **32**(3), 1180–1200 (2010)
21. Pearson, J.W., Stoll, M., Wathen, A.J.: Preconditioners for state-constrained optimal control problems with Moreau-Yosida penalty function. Numer. Linear Algebra Appl. **21**(1), 81–97 (2014)

22. Rees, T., Stoll, M., Wathen, A.: All-at-once preconditioning in PDE-constrained optimization. Kybernetika (Prague) **46**(2), 341–360 (2010)
23. Schiela, A., Günther, A.: An interior point algorithm with inexact step computation in function space for state constrained optimal control. Numer. Math. **119**(2), 373–407 (2011)
24. Schiela, A., Ulbrich, S.: Operator preconditioning for a class of inequality constrained optimal control problems. SIAM J. Optim. **24**(1), 435–466 (2014)
25. Schöberl, J., Zulehner, W.: Symmetric indefinite preconditioners for saddle point problems with applications to PDE-constrained optimization problems. SIAM J. Matrix Anal. Appl. **29**(3), 752–773 (electronic) (2007)
26. Staff, G.A., Rønquist, E.M.: Stability of the parareal algorithm. In: Domain decomposition methods in science and engineering, *Lect. Notes Comput. Sci. Eng.*, vol. 40, pp. 449–456. Springer, Berlin (2005)
27. Stoll, M., Wathen, A.: Preconditioning for partial differential equation constrained optimization with control constraints. Numer. Linear Algebra Appl. **19**(1), 53–71 (2012)
28. Tröltzsch, F.: On the Lagrange-Newton-SQP method for the optimal control of semilinear parabolic equations. SIAM J. Control Optim. **38**(1), 294–312 (electronic) (1999)
29. Tröltzsch, F.: Optimal control of partial differential equations, *Graduate Studies in Mathematics*, vol. 112. American Mathematical Society, Providence, RI (2010). Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels
30. Ulbrich, M.: Semismooth Newton methods for operator equations in function spaces. SIAM J. Optim. **13**(3), 805–842 (2003) (2002)
31. Ulbrich, M.: Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces, *MOS-SIAM Series on Optimization*, vol. 11. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA (2011)
32. Ulbrich, M., Ulbrich, S.: Primal-dual interior-point methods for PDE-constrained optimization. Math. Program. **117**(1-2, Ser. B), 435–485 (2009)
33. Ziems, J.C., Ulbrich, S.: Adaptive multilevel inexact SQP methods for PDE-constrained optimization. SIAM J. Optim. **21**(1), 1–40 (2011)
34. Zulehner, W.: Analysis of iterative methods for saddle point problems: a unified approach. Math. Comp. **71**(238), 479–505 (electronic) (2002)