

From PDEs constrained optimization to controllability problems via time domain decomposition

Pierre-Henri Cocquet^[0000-0003-0986-0800] and Liu-Di Lu^[0000-0003-3629-9879]

1 Introduction

Although partial differential equations (PDEs) constrained optimization and controllability problems share similarities, they originated from different historical contexts. Controllability problems can be traced back more than 300 years to Johann Bernoulli's brachistochrone problem; see [8] for a historical review. This problem initiated the development of optimal control theory, in which the system is often governed by ordinary differential equations (ODEs). PDE-constrained optimization, on the other hand, was developed in the early 1960s by the desire to extend control theory to systems governed by PDEs. This field has become an active research field following the monograph of Jacques-Louis Lions [6].

In this paper, we focus on applying domain decomposition techniques to solve both classes of problems. To illustrate these two problems, we consider the linear heat equation as the governing constraint in both cases. In the case of a PDE-constrained optimization problem, we have a given target \hat{y} and want to find optimal control strategies u such that the corresponding solution y of the heat equation is as close as possible to the given target. This leads to a constrained minimization problem:

$$\begin{aligned} \min_{y,u} \quad & \frac{\alpha}{2} \|y - \hat{y}\|_{L^2(Q)}^2 + \frac{\gamma}{2} \|y(T, \cdot) - \hat{y}(T, \cdot)\|_{L^2(\Omega)}^2 + \frac{\nu}{2} \|u\|_{L^2(Q)}^2, \\ \text{s.t.} \quad & \partial_t y - \kappa \Delta y = u \quad \text{in } Q := (0, T) \times \Omega, \\ & y = 0 \quad \text{on } \Sigma := (0, T) \times \partial\Omega \quad y = y_0 \quad \text{on } \Sigma_0 := \{0\} \times \Omega. \end{aligned} \tag{P1}$$

Here, $\alpha \geq 0, \gamma \geq 0$ and $\nu > 0$ are given penalization parameters, $\kappa > 0$ is the diffusion coefficient which is assumed to be constant in the spatial domain, y_0 is a

Pierre-Henri Cocquet
Université de Pau et des Pays de l'Adour, France, e-mail: pierre-henri.cocquet@univ-pau.fr

Liu-Di Lu
Lund University, Sweden, e-mail: liudi.lu@math.lu.se

given initial condition, $T > 0$ is the final time and $\Omega \subset \mathbb{R}^n$ with $n = 1, 2, 3$ is a bounded open set with Lipschitz boundary. For our controllability problem, we want to find optimal control strategies u with minimal L^2 -norm that pilot the solution of the heat equation y from a given initial condition y_0 to a zero final condition. This results in solving a similar minimization problem to (P1) as

$$\begin{aligned} & \min_u \frac{\nu}{2} \|u\|_{L^2(Q)}^2, \\ \text{s.t. } & \partial_t y - \kappa \Delta y = u \text{ in } Q, y|_{\Sigma} = 0, y = y_0 \text{ on } \Sigma_0, y = 0 \text{ on } \Sigma_T := \{T\} \times \Omega. \end{aligned} \quad (\text{P2})$$

The existence and uniqueness of a solution to (P1) are guaranteed by the strict-convexity of the linear quadratic optimization problem. The null-controllability of the classical heat equation is also well known, *e.g.*, see [2, Chapter 2.5]. The existence of solution $u \in L^2(Q)$ to our null-controllability problem (P2) has been proven in [7, Theorem 6.1.2] for Ω with $C^{2+\delta}$ boundary, $\delta > 0$. We emphasize that the results of [7, Theorem 6.1.2] have been obtained considering a weak solution to the heat equation. The similarity between these two problems (P1) and (P2) already appears in their formulations. Indeed, choosing $\hat{y}(T, \cdot) = 0$ and setting $\alpha = 0$, $\gamma \rightarrow +\infty$ in (P1) show that (P1) "converges" toward (P2). We make this convergence more rigorous in Section 2. On the other hand, if we apply separately time domain decomposition to the first-order optimality system issued from each problem, we show in Section 3 that under the same conditions, the convergence factor associated with (P1), ρ_{P1} , coincide with the convergence factor associated with (P2), ρ_{P2} . Our numerical experiments in Section 4 confirm these results.

2 Relation between problems (P1) and (P2)

First of all, we introduce y_u as the unique weak solution to

$$\partial_t y_u - \kappa \Delta y_u = u \text{ in } (0, T) \times \Omega, y_u = 0 \text{ on } \Sigma, y_u = y_0 \text{ on } \Sigma_0. \quad (1)$$

Aubin's Lemma (see [3, Lemma 6.2]) ensures that any weak solution $y_u \in C([0, T], L^2(\Omega))$, hence the initial condition $y_0 \in L^2(\Omega)$ is meaningful. From [3, Theorem 6.6], there exists a unique weak solution $y_u \in X := L^2(0, T; H_0^1(\Omega))$ to (1) with $\partial_t y_u \in Y := L^\infty(0, T; L^2(\Omega))$ and $y_u|_{\Sigma_0} = y_0$, such that

$$\|\partial_t y_u\|_{L^2(0, T; H^{-1}(\Omega))} + \|y_u\|_Y + \|y_u\|_X \leq C \left(\|u\|_{L^2(Q)} + \|y_0\|_{L^2(\Omega)} \right), \quad (2)$$

for a suitable constant $C > 0$. For the relation between (P1) and (P2), we have the next result (see [1, Theorem 10] for a similar result in a finite-dimensional case).

Theorem 1 *Let $y_T := 0$ and u^* be a solution of (P2). Let u_ε^* be the unique solution of (P1) with $\alpha = 0$, $\gamma = 1/\varepsilon$ and $\hat{y}(T, \cdot) = y_T$ whose existence and uniqueness follows from the strict-convexity of (P1). Let $y_\varepsilon^* = y_{u_\varepsilon^*}$ then*

$$\|y_\varepsilon^*(T, \cdot) - y_T\|_{L^2(\Omega)} \leq \sqrt{v\varepsilon} \|u^*\|_{L^2(Q)}. \quad (3)$$

There exist accumulation points of $(u_\varepsilon^*)_\varepsilon \subset L^2(Q)$ for the weak topology of $L^2(Q)$. Any of such accumulation points is an optimal solution to (P2). If (P2) has the unique solution u^* , then the whole sequence $(u_\varepsilon^*)_\varepsilon$ converges to u^* strongly in $L^2(Q)$.

Proof. The optimality of u_ε^* ensures that $\forall u \in L^2(Q)$, one has

$$\frac{1}{2\varepsilon} \|y_\varepsilon^*(T, \cdot) - y_T\|_{L^2(\Omega)}^2 + \frac{\nu}{2} \|u_\varepsilon^*\|_{L^2(Q)}^2 \leq \frac{1}{2\varepsilon} \|y_u(T, \cdot) - y_T\|_{L^2(\Omega)}^2 + \frac{\nu}{2} \|u\|_{L^2(Q)}^2.$$

Now taking $u = u^*$ and using that y_{u^*} satisfies $y_{u^*}(T, \cdot) = y_T$ give

$$\frac{1}{2\varepsilon} \|y_\varepsilon^*(T, \cdot) - y_T\|_{L^2(\Omega)}^2 + \frac{\nu}{2} \|u_\varepsilon^*\|_{L^2(Q)}^2 \leq \frac{\nu}{2} \|u^*\|_{L^2(Q)}^2,$$

from which the aforementioned inequality follows. We also get that the sequence $(u_\varepsilon^*)_\varepsilon$ is uniformly bounded in $L^2(Q)$ and thus admits a subsequence that converges weakly in $L^2(Q)$ toward some \tilde{u} . From (2), the sequence $(y_\varepsilon^*)_\varepsilon$ is uniformly bounded in some reflexive Banach space \mathcal{H} and thus admits a subsequence that weakly converges toward some $\tilde{y} \in \mathcal{H}$. Taking the limit $\varepsilon \rightarrow 0$ in the weak formulation of (1) and using that $u_\varepsilon^* \rightarrow \tilde{u}$, we obtain (up to a subsequence) that $y_\varepsilon^* \rightarrow \tilde{y} = y_{\tilde{u}}$. Owing to (3), $y_\varepsilon^*(T, \cdot) \rightarrow y_T$ strongly in $L^2(\Omega)$ so that $y_{\tilde{u}}(T, \cdot) = y_T$. The weak lower semicontinuity of the L^2 -norm finally yields $\|\tilde{u}\|_{L^2(Q)} \leq \liminf_{\varepsilon \rightarrow 0} \|u_\varepsilon^*\|_{L^2(Q)} \leq \|u^*\|_{L^2(Q)}$, and the optimality of u^* ensures that $\|\tilde{u}\|_{L^2(Q)} = \|u^*\|_{L^2(Q)}$. Hence, \tilde{u} is an optimal solution to (P2). Any accumulation point (for the weak topology of $L^2(Q)$) of $(u_\varepsilon^*)_\varepsilon$ is then an optimal solution to (P2). If (P2) has a unique optimal solution u^* , then $\tilde{u} = u^*$ and the whole sequence $(u_\varepsilon^*)_\varepsilon$ converges toward u^* . The uniqueness of the limit gives that $\liminf_{\varepsilon \rightarrow 0} \|u_\varepsilon^*\|_{L^2(Q)} = \lim_{\varepsilon \rightarrow 0} \|u_\varepsilon^*\|_{L^2(Q)} = \|u^*\|_{L^2(Q)}$, and then $u_\varepsilon^* \rightarrow u^*$ strongly. \square

3 Time domain decomposition

A classical way to treat these two problems (P1) and (P2) is to derive the first-order optimality system by applying the Lagrange multiplier approach. We can then obtain the reduced optimality system for (P1),

$$\begin{aligned} \partial_t y - \kappa \Delta y &= v^{-1} \lambda & \text{in } Q, & \quad y = y_0 & \text{on } \Sigma_0, \\ \partial_t \lambda + \kappa \Delta \lambda &= \alpha(y - \hat{y}) & \text{in } Q, & \quad \lambda = -\gamma(y - \hat{y}) & \text{on } \Sigma_T, \end{aligned} \quad (S1)$$

and the reduced optimality system for (P2),

$$\begin{aligned} \partial_t y - \kappa \Delta y &= v^{-1} \lambda & \text{in } Q, & \quad y = y_0 & \text{on } \Sigma_0, \quad y = 0 & \text{on } \Sigma_T, \\ \partial_t \lambda + \kappa \Delta \lambda &= 0 & \text{in } Q, & & & \end{aligned} \quad (S2)$$

where λ is the adjoint state. We refer to [6, Chapter 3 Section 3.1] for further details on the derivation of the reduced optimality system. We used the optimality condition $\nu u - \lambda = 0$ in Q in both cases to eliminate the control variable u in the state equation of y , and we omit the Dirichlet boundary conditions in (S1) and (S2). We also assume that both problems (P1), (P2) are solvable. We are now interested in applying domain decomposition methods to solve their associated optimality systems (S1) and (S2).

The alternating Schwarz algorithm, invented by H.A. Schwarz in 1869, is a way to solve a large problem by decomposing the original domain into smaller subdomains and then solving the problem on each subdomain one after another, while exchanging information between interfaces until the solution converges. We refer to [4] for a historical review. It is well known that the alternating Schwarz algorithm fails to converge when applied to non-overlapping spatial subdomains. This is due to the repeated passing of identical "Dirichlet data" from one subdomain to the other. However, it has been shown in [5] that there exist several variants of the alternating Schwarz algorithm in the time decomposition framework, and some of them may converge even without overlap. When decomposing in time and applying non-overlapping alternating Schwarz type algorithms to the reduced optimality system, a Dirichlet transmission condition at the time interface can be equivalently rewritten as a Robin type transmission condition, e.g., see [5, Equation 4]. This avoids passing identical "Dirichlet data" between subdomains and thus leads to the convergence. For this reason, we are interested here in studying the non-overlapping time-decomposed alternating Schwarz algorithms to solve (S1) and (S2). We decompose the space-time domain Q into two non-overlapping subdomains: $Q_1 := (0, \Gamma) \times \Omega$ and $Q_2 := (\Gamma, T) \times \Omega$, where $\Gamma \in (0, T)$ represents the interface.

The system (S1) has a forward-backward structure, where the state variable y propagates forward with an initial condition y_0 , and the adjoint variable λ propagates backward with a final condition $-\gamma(y(T, \cdot) - \hat{y}(T, \cdot))$. To retain the same forward-backward structure in subdomain Q_1 , one should impose a "final" condition at the interface $\Sigma_\Gamma := \{\Gamma\} \times \Omega$ for the adjoint variable λ . Similarly, one imposes an "initial" condition at the interface Σ_Γ for the state variable y to have the same forward-backward structure in Q_2 . We denote by AS1 this first type of alternating Schwarz algorithm which imposes λ at Σ_Γ in Q_1 and y at Σ_Γ in Q_2 . Then, for the iteration index $\ell = 1, 2, \dots$, the AS1 applied to solve (S1) reads

$$\begin{aligned} \partial_t y_1^\ell - \kappa \Delta y_1^\ell &= \nu^{-1} \lambda_1^\ell & \text{in } Q_1, & \quad \partial_t y_2^\ell - \kappa \Delta y_2^\ell = \nu^{-1} \lambda_2^\ell & \text{in } Q_2, \\ y_1^\ell &= y_0 & \text{on } \Sigma_0, & \quad y_2^\ell = y_1^\ell & \text{on } \Sigma_\Gamma, \\ \partial_t \lambda_1^\ell + \kappa \Delta \lambda_1^\ell &= \alpha(y_1^\ell - \hat{y}) & \text{in } Q_1, & \quad \partial_t \lambda_2^\ell + \kappa \Delta \lambda_2^\ell = \alpha(y_2^\ell - \hat{y}) & \text{in } Q_2, \\ \lambda_1^\ell &= \lambda_2^{\ell-1} & \text{on } \Sigma_\Gamma, & \quad \lambda_2^\ell + \gamma y_2^\ell = \gamma \hat{y} & \text{on } \Sigma_T. \end{aligned} \tag{S1-AS1}$$

For comparison, we also apply AS1 to solve (S2), which leads to the algorithm,

$$\begin{aligned}
\partial_t y_1^\ell - \kappa \Delta y_1^\ell = v^{-1} \lambda_1^\ell & \quad \text{in } Q_1, & \partial_t y_2^\ell - \kappa \Delta y_2^\ell = v^{-1} \lambda_2^\ell & \quad \text{in } Q_2, \\
y_1^\ell = y_0 & \quad \text{on } \Sigma_0, & y_2^\ell = y_1^\ell & \quad \text{on } \Sigma_\Gamma, \\
\partial_t \lambda_1^\ell + \kappa \Delta \lambda_1^\ell = 0 & \quad \text{in } Q_1, & y_2^\ell = 0 & \quad \text{on } \Sigma_T \\
\lambda_1^\ell = \lambda_2^{\ell-1} & \quad \text{on } \Sigma_\Gamma, & \partial_t \lambda_2^\ell + \kappa \Delta \lambda_2^\ell = 0 & \quad \text{in } Q_2.
\end{aligned} \tag{S2-AS1}$$

We observe that in the first subdomain Q_1 on the left of the system (S2-AS1), it does not follow the same structure as in (S2). Indeed, the system (S2) is also a forward-backward system in which the initial and final conditions are all given for the state variable y . Hence, to retain the same forward-backward structure of (S2) when applying the alternating Schwarz algorithm, one should impose a "final" condition at the interface Σ_Γ for the state variable y in Q_1 and an "initial" condition at the interface Σ_Γ also for y in Q_2 . This then leads to a second type of alternating Schwarz algorithm denoted by AS2, which imposes y at Σ_Γ in both Q_1 and Q_2 . The AS2 applied to solve (S2) then reads

$$\begin{aligned}
\partial_t y_1^\ell - \kappa \Delta y_1^\ell = v^{-1} \lambda_1^\ell & \quad \text{in } Q_1, & \partial_t y_2^\ell - \kappa \Delta y_2^\ell = v^{-1} \lambda_2^\ell & \quad \text{in } Q_2, \\
y_1^\ell = y_0 & \quad \text{on } \Sigma_0, & y_2^\ell = y_1^\ell & \quad \text{on } \Sigma_\Gamma, \\
y_1^\ell = y_2^{\ell-1} & \quad \text{on } \Sigma_\Gamma, & y_2^\ell = 0 & \quad \text{on } \Sigma_T \\
\partial_t \lambda_1^\ell + \kappa \Delta \lambda_1^\ell = 0 & \quad \text{in } Q_1, & \partial_t \lambda_2^\ell + \kappa \Delta \lambda_2^\ell = 0 & \quad \text{in } Q_2,
\end{aligned} \tag{S2-AS2}$$

which now follows the same structure as (S2) in both subdomains. For comparison, we also apply AS2 to solve (S1), which leads to the algorithm,

$$\begin{aligned}
\partial_t y_1^\ell - \kappa \Delta y_1^\ell = v^{-1} \lambda_1^\ell & \quad \text{in } Q_1, & \partial_t y_2^\ell - \kappa \Delta y_2^\ell = v^{-1} \lambda_2^\ell & \quad \text{in } Q_2, \\
y_1^\ell = y_0 & \quad \text{on } \Sigma_0, & y_2^\ell = y_1^\ell & \quad \text{on } \Sigma_\Gamma, \\
y_1^\ell = y_2^{\ell-1} & \quad \text{on } \Sigma_\Gamma, & \partial_t \lambda_2^\ell + \kappa \Delta \lambda_2^\ell = \alpha(y_2^\ell - \hat{y}) & \quad \text{in } Q_2, \\
\partial_t \lambda_1^\ell + \kappa \Delta \lambda_1^\ell = \alpha(y_1^\ell - \hat{y}) & \quad \text{in } Q_1, & \lambda_2^\ell + \gamma y_2^\ell = \gamma \hat{y} & \quad \text{on } \Sigma_T.
\end{aligned} \tag{S1-AS2}$$

To get better insights into these four algorithms (S1-AS1)-(S1-AS2), we need to analyze their convergence. Since we focus here on the time domain decomposition, we can then apply a discretization in space and replace the spatial operator $-\Delta$ by a matrix $A \in \mathbb{R}^{N \times N}$, e.g., using the centered finite difference discretization in space. As we only have Dirichlet boundary conditions, this finite difference matrix A is symmetric and can be diagonalized with $PP^T = I_N$, $PAP^T = D := \text{diag}(d_1, \dots, d_N)$, and $d_i > 0$, $i = 1, \dots, N$. For instance, the eigenvalues in the one dimensional case ($n = 1$) are given by $d_i = \frac{4}{h^2} \sin^2(\frac{\pi i}{2(N+1)})$, $i = 1, \dots, N$, with $h := L/(N+1)$ the mesh size and L the length of the spatial domain. This diagonalization leads to N independent 2×2 systems of first-order ODEs in time.

For each algorithm, solving the ODE system explicitly for each eigenvalue d_i first in Q_1 then in Q_2 leads to a recursive relation between two iterates $\mathbf{y}_{2,i}^\ell = \rho(d_i) \mathbf{y}_{2,i}^{\ell-1}$, where $\mathbf{y}_{2,i}^\ell(t) \approx y_2^\ell(t, x_i)$ is an approximation after the discretization and diagonalization in space, and $\rho(d_i)$ is the contraction factor associated with the

eigenvalue d_i . Each algorithm in (S1-AS1)-(S2-AS1) has an associated contraction factor $\rho(d_i)$ for each eigenvalue d_i , as will be shown in the sequel. The absolute value of the factor $|\rho(d_i)|$ measures the contraction of each algorithm as a function of d_i , and the algorithm converges if this absolute value is smaller than one for all eigenvalues, *i.e.*, $\max_{d_i} |\rho(d_i)| < 1$. More details of the diagonalization and computations to obtain the contraction factor can be found in [5].

The contraction factor associated with (S1-AS1) for each d_i is given by

$$\rho_{S1}^{AS1}(d_i) = \frac{\alpha + \gamma(\sigma_i \coth(\sigma_i(T - \Gamma)) - \kappa d_i)}{\nu(\sigma_i \coth(\sigma_i \Gamma) + \kappa d_i)(\sigma_i \coth(\sigma_i(T - \Gamma)) + \kappa d_i + \gamma \nu^{-1})}, \quad (4)$$

with $\sigma_i := \sqrt{\kappa^2 d_i^2 + \nu^{-1} \alpha}$. Similarly, the contraction factor of (S2-AS1) for each eigenvalue d_i is given by

$$\rho_{S2}^{AS1}(d_i) = \frac{\coth(\kappa d_i(T - \Gamma)) - 1}{\coth(\kappa d_i \Gamma) + 1}. \quad (5)$$

Setting $\alpha = 0$ in (4), we have

$$\rho_{S1}^{AS1}(d_i) \Big|_{\alpha=0} = \frac{\coth(\kappa d_i(T - \Gamma)) - 1}{\coth(\kappa d_i \Gamma) + 1} \frac{\gamma}{\nu \kappa d_i \coth(\kappa d_i(T - \Gamma)) + \nu \kappa d_i + \gamma}. \quad (6)$$

Then, when $\gamma \rightarrow \infty$, the second fraction goes to 1, thus $\lim_{\gamma \rightarrow \infty} \rho_{S1}^{AS1}(d_i) \Big|_{\alpha=0} = \rho_{S2}^{AS1}(d_i)$, meaning that these two contraction factors become the same in the limit case. Since this limit holds for each eigenvalue, we have the next result.

Theorem 2 *If $\alpha = 0$ in (P1), then the convergence behavior of two algorithms (S1-AS1) and (S2-AS1) becomes the same when $\gamma \rightarrow \infty$.*

In terms of two problems (P1) and (P2), when $\alpha = 0$, or more generally, taking a relatively small value of α w.r.t. γ and ν in (P1), it means that the discrepancy between the solution y of the heat equation and the given target \hat{y} is less important. In this context, taking large values of γ emphasizes that one searches for optimal control strategies in (P1) that can pilot the solution of the heat equation to the target only at the final time T . This is equivalent to the idea of the controllability problem (P2). In this case, it is relevant that we also obtain similar convergence behavior when applying the same type of alternating Schwarz algorithm to solve the corresponding optimality system. In particular, their convergence does not depend on the choice of the target function \hat{y} in (P1), nor on the final condition y_T in (P2).

In terms of two algorithms (S1-AS1) and (S2-AS1), it has been shown in [5, Theorem 1] that (S1-AS1) always converges for $\Gamma \leq T/2$ when $\alpha = 1$. This still holds for the algorithm (S2-AS1). Indeed, using two properties of the hyperbolic cotangent: $\coth(x) > 1$ and decreases for all $x > 0$, we find in (5) that $\coth(\kappa d_i(T - \Gamma)) \leq \coth(\kappa d_i \Gamma)$. Thus, $\rho_{S2}^{AS1}(d_i) < 1$ for each d_i , and the algorithm (S2-AS1) converges for $\Gamma \leq T/2$. Given the analytical form of (5), a more precise choice of Γ can be made. This is done in the next theorem, whose proof is omitted for brevity.

Theorem 3 Assume that $\Gamma < \frac{1}{2\kappa d_{\min}} \ln \frac{1}{2}(1 + e^{2\kappa d_{\min} T})$ where d_{\min} is the smallest eigenvalue of A . Then $\max_{d_i} |\rho_{S_2}^{AS_1}(d_i)| < 1$ hence Algorithm (S2-AS1) converges.

In the one dimensional case with only Dirichlet boundary conditions, the smallest eigenvalue is $d_{\min} = \frac{4}{h^2} \sin^2(\frac{\pi}{2(N+1)}) = \frac{\pi^2}{L^2} - h^2 \frac{\pi^4}{12L^4} + O(h^4)$. Note that for the function f , $\lim_{a \rightarrow 0} f(a) = 1/2$ and $\lim_{a \rightarrow \infty} f(a) = 1$. We thus have the bound $\frac{T}{2} < \frac{T}{2\kappa d_{\min}} \ln \frac{1}{2}(1 + e^{2\kappa d_{\min} T}) < T$, which also holds in higher dimensions, with eigenvalues $d_i \in (0, \infty)$, $i = 1, \dots, N$. On the other hand, the contraction factor (5) decreases as a function of d_i , which is bounded above by $\rho_{S_2}^{AS_1}(d_{\min})$. The convergence behavior then only depends on the smallest eigenvalue.

Following the same approach, we can solve the ODE systems associated with (S2-AS2) and (S1-AS2) to find their contraction factors. For each d_i , these two contraction factors are $\rho_{S_1}^{AS_2}(d_i) = 1$, and $\rho_{S_2}^{AS_2}(d_i) = 1$. This is not surprising because of the transmission conditions used in (S2-AS2) and (S1-AS2). For both algorithms, we have $y_1^\ell = y_2^{\ell-1}$ on Σ_Γ in Q_1 , and $y_2^\ell = y_1^\ell$ on Σ_Γ in Q_2 . For any given initial guess y_2^0 , the same information is constantly passed back and forth between two subdomains, solving the same problem at each iteration. Thus, both algorithms stagnate. This does not depend on the eigenvalues or on any parameter values, since these transmission conditions are imposed to decompose (S1) and (S2) at the continuous level. Recall that these transmission conditions are the natural choice to decompose the optimality system (S2), since we maintain the same structure in both subdomains in (S2-AS2) as in the original optimality system (S2). This reveals once again the importance of choosing properly the transmission condition in the non-overlapping alternating Schwarz in time framework as discussed in [5].

4 Numerical experiments

To illustrate the convergence behavior of these alternating Schwarz algorithms, we first plot the contraction factors (4) and (5) as functions of eigenvalues. We set $\alpha = 1$, $\nu = 1$, $\kappa = 1$, $T = 1$ and an equal decomposition with $\Gamma = T/2$. Fig. 1 on the top left illustrates two contraction factors (4) and (5) for eigenvalues $d_i \in [10^{-2}, 10^2]$ and different values of $\gamma \in \{1, 10, 10^2\}$.

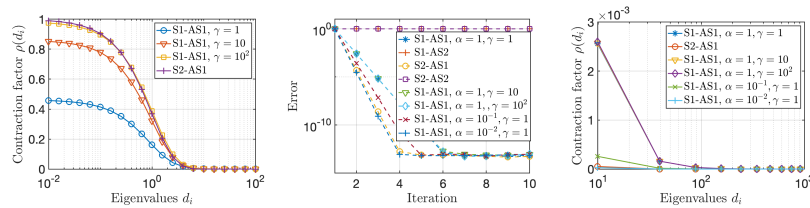


Fig. 1 Left: contraction factors (4) and (5) as functions of eigenvalues $d_i \in [10^{-2}, 10^2]$ with different values of γ . Middle: error decay of four algorithms (S1-AS1)-(S1-AS2). Right: contraction factors for eigenvalues $d_i \in [10, 10^3]$ with different values of γ and α

As expected, increasing the value of γ in (4) leads to behavior more similar to that of (5) for all eigenvalues. We also observe a slight difference in the range of $d_i = 1$ between two curves ρ_{S2}^{AS1} and ρ_{S1}^{AS1} with $\gamma = 10^2$, due to the value $\alpha = 1$ in (4). Taking $\alpha = 0$, the curve of $\rho_{S1}^{AS1}|_{\alpha=0}$ is then always below that of ρ_{S2}^{AS1} , since the second fraction in (6) is always smaller than one. More generally, we observe that all convergence factors are monotonically decreasing, and their values are especially small for large eigenvalues. This indicates that both algorithms (S1-AS1) and (S2-AS1) are good smoothers, and their convergence behavior for small eigenvalues can be further improved by introducing a relaxation parameter.

Next, we test the error decay of four algorithms (S1-AS1)-(S1-AS2) in the one-dimensional case. We keep the same setting as in the previous test and consider the length $L = 1$, the initial condition $y_0(x) = \frac{1}{2\pi^2}(1 - \exp(\pi^2 T)) \sin(\pi x)$, the target function $\hat{y}(t, x) = \frac{1}{2\pi^2}(\exp(\pi^2 t) - \exp(\pi^2 T)) \sin(\pi x)$ in (S1), and the final condition $y_T(x) = 0$ in (S2). In particular, this target function is also the solution of the optimality system (S2) with $\lambda(t, x) = (\exp(\pi^2 t) - \exp(\pi^2 T)/2) \sin(\pi x)$. We use the Crank-Nicolson scheme to discretize both systems with mesh size $h_t = h_x = 1/32$. Fig. 1 on the top right illustrates the error decay of four algorithms. As explained in our theoretical analysis, both algorithms (S2-AS2) and (S1-AS2) stagnate as $\rho_{S1}^{AS2} = \rho_{S2}^{AS2} = 1$. Regarding algorithms AS1, we observe that the algorithm (S2-AS1) converges faster than (S1-AS1) in the case $\alpha = 1$ and $\gamma = 1$. However, when increasing the value of γ , the convergence of (S1-AS1) stays the same. Indeed, the smallest eigenvalue in this case is approximately π^2 , and Fig. 1 on the bottom further illustrates contraction factors for large eigenvalues. We observe that ρ_{S2}^{AS1} is much smaller in this case, and increasing γ in ρ_{S1}^{AS1} cannot track the behavior of ρ_{S2}^{AS1} . However, decreasing α in ρ_{S1}^{AS1} can better track the convergence behavior of ρ_{S2}^{AS1} , which has also been testified in Fig. 1 on the top right.

References

1. Ciaramella, G., Borzi, A., Dirr, G., Wachsmuth, D.: Newton methods for the optimal control of closed quantum spin systems. *SIAM J. Sci. Comput.* **37**(1), A319–A346 (2015)
2. Coron, J.M.: Control and Nonlinearity, *Mathematical Surveys and Monographs*, vol. 136. American Mathematical Society (2007)
3. Ern, A., Guermond, J.L.: Theory and Practice of Finite Elements, *Applied Mathematical Sciences*, vol. 159. Springer New York (2004)
4. Gander, M.J.: Schwarz methods over the course of time. *Electronic Transactions on Numerical Analysis* **31**, 228–255 (2008)
5. Gander, M.J., Lu, L.D.: Non-overlapping Schwarz methods in time for parabolic optimal control problems. *Domain Decomposition Methods in Science and Engineering XXVIII* (2025)
6. Lions, J.L.: Optimal Control of Systems Governed by Partial Differential Equations. Springer Berlin, Heidelberg (1971)
7. Puel, J.P.: Control of partial differential equations: Theoretical aspects. *Control and Inverse Problems for Partial Differential Equations* **22**, 1–46 (2019)
8. Sussmann, H., Willems, J.: 300 years of optimal control: from the brachistochrone to the maximum principle. *IEEE Control Systems Magazine* **17**(3), 32–44 (1997)