DECOMPOSE-THEN-OPTIMIZE: A NEW APPROACH TO DESIGN DOMAIN DECOMPOSITION METHODS FOR OPTIMAL CONTROL PROBLEMS. *

PIERRE-HENRI COCQUET†, MARTIN J. GANDER‡, AND ALEXANDRE VIEIRA§

Abstract. For optimal control problems there is a classical discussion of whether one should first optimize the problem and then discretize it, or the other way round. We are interested in exploring a similar question related to domain decomposition methods for optimal control problems which have received substantial attention over the past two decades, but new methods were mostly developed using the optimize-then-decompose approach. After a detailed introduction to this subject, we present and analyze a new domain decomposition method for optimal control problems that comes from the decompose-then-optimize strategy, which is less common. We use as our model problem a linear quadratic optimal control problem, which we decompose and then solve using an augmented Lagrangian optimization technique. This leads to a new domain decomposition algorithm for such problems that has very good scalability properties. We prove that, when the algorithm converges, it necessarily converges to an optimal point of the original, non-decomposed problem. We illustrate the efficiency of our new domain decomposition method with numerical examples from which we obtain very desirable properties for domain decomposition methods, namely that the convergence is independent of the meshsize and the number of subdomain.

Key words. Domain decomposition, Optimal control, Augmented Lagrangian

MSC codes. 49M27, 49M30

1. Introduction.

In optimal control and PDE constrained (equivalently constraint is commonly used) optimization problems of the form

$$\min_{y,u} \quad \mathcal{J}(y,u) \quad \text{s.t.} \quad g(y,u)=0, \tag{1.1}$$

a subject of discussion is the question of whether one should first optimize and then discretize the problem, i.e. compute the Lagrangian first order optimality conditions

$$\nabla\mathcal{L}(y,u,\lambda)=0, \tag{1.2}$$

and then discretize them (*optimize-then-discretize* or indirect method, path going from the middle down and then to the left in Figure 1.1), or if it is better to first discretize and then optimize the problem (*discretize-then-optimize* or direct method, path going left and then down in Figure 1.1), see e.g. [50]. There are advantages to both approaches: for discretize-then-optimize, one always obtains the true gradient of the discrete problem, even when the discretization is coarse, and symmetric formulations remain symmetric. Also, for time dependent problems, due to the Pontryagin maximum principle (see [23] for a historical introduction), the first order optimality conditions are of the form of a Hamiltonian differential equation, and if one discretizes first and then optimizes, this geometric structure can automatically be preserved [14],

*

Funding: A. Vieira and P.H. Cocquet have been supported by the "Agence Nationale de la Recherche" (ANR), Project O-TO-TT-FU number ANR-19-CE40-0011 (2019-2023). M.J. Gander is supported by the Swiss National Science Foundation. The authors also thank the CIRM for their hospitality and giving us the opportunity to work together during a "*Recherche en résidence*".

†Université de Pau et des Pays de l'Adour / E2S-UPPA, SIAME, PAU, 64000, France

‡UNIGE

§Physique et Ingénierie Mathématique pour l'Énergie et l'Environnement (PIMENT), Université de la Réunion, Sainte-Clotilde, France. Alexandre.Vieira@univ-reunion.fr

$$\begin{tikzcd}
\begin{array}{l}\mathcal{J}_h(\mathbf{y},\mathbf{u}) \longrightarrow \min \\ g_h(\mathbf{y},\mathbf{u}) = 0\end{array} \arrow[d, "\text{optimize}"] & \begin{array}{l}\mathcal{J}(y,u) \longrightarrow \min \\ g(y,u) = 0\end{array} \arrow[l, "\text{discretize}"'] \arrow[d, "\text{optimize}"] \arrow[r, "\text{decompose}"] \arrow[dr, phantom, "\text{FETI}"] & \begin{array}{l}\sum_i \mathcal{J}_i(y_i^{n+1}, u_i^{n+1}, y_j^n, u_j^n) \longrightarrow \min \\ \sum_i g_i(y_i^{n+1}, u_i^{n+1}, y_j^n, u_j^n) = 0\end{array} \arrow[d, shift left, "\text{optimize}"] \\
\nabla \mathcal{L}_h(\mathbf{y},\mathbf{u},\lambda) = 0 & \nabla \mathcal{L}(y,u,\lambda) = 0 \arrow[l, "\text{discretize}"'] \arrow[u] \arrow[r, shift left, "\text{decompose}"] & (y_i^{n+1}, u_i^{n+1}, \lambda_i^{n+1}) = DD(y_j^n, u_j^n, \lambda_j^n) \arrow[l, shift left, "\text{interpret}"] \arrow[u, shift left, "\text{interpret}"]
\end{tikzcd}$$

Fig. 1.1: *Discretize then optimize* versus *optimize then discretize* (left part) and decompose then optimize versus optimize then decompose (right part).

see also [29, 7] and [30, Chapter VI, exercises 14,15,16]. This is rather elegant, and can be important in low dimensions close to a critical value of the Hamiltonian, but not in general since the first order optimality condition is a boundary value problem and the time interval is not very large in general. In optimize-then-discretize, none of the above properties hold, but discretizations are much more flexible, one can adapt locally independently in the forward and backward problem, which can be an advantage, also in hyperbolic problems where CFL conditions need to be met. In some cases, the two approaches also commute.

We are interested here in a new, analogous question for optimal control problems and PDE constrained optimization when designing and analyzing parallel algorithms for their solution using domain decomposition (DD), namely the *decompose-then-optimize* and *optimize-then-decompose* approaches, see the right part of Figure 1.1. The situation here is more involved, and much less explored in the literature. Historically, the FETI method [19] was developed in this spirit, albeit for unconstrained optimization, and using the equivalence of the Laplace problem and the minimization of the Dirichlet integral, see the blue arrows in Figure 1.1: for FETI, one first considers the PDE problem as an equivalent minimization problem (blue arrow going up), then decomposes the minimization problem at the variational level (red arrow going to the right and blue arrow going down), which makes Neumann traces match automatically, and thus only the Dirichlet traces need to be imposed to match explicitly for the DD solution to be a solution of the underlying PDE which is done in FETI using Lagrange multipliers. FETI is also historically not written as an iteration, like the equivalent dual Schur complement method at the discrete level, and the Conjugate Gradient method is used to solve the decomposed system. Using however a stationary residual correction method for the FETI system shows that it is a classical Dirichlet-Dirichlet method (blue arrow going down), the dual of the Neumann-Neumann method [9] or primal Schur complement method at the discrete level, see e.g. [15, Section 4.8].

The *optimize-then-decompose* approach (black arrow going down in the middle and then right green arrow in Figure 1.1) has received quite some attention in the literature, since one can directly apply standard DD methods to the optimality system (1.2), and then study their convergence, which leads to interesting new results for DD methods, see for example [21, 22, 24, 25], and also [44]. In contrast to the *optimize-then-discretize* approach, in the *optimize-then-decompose* approach, it is also often shown that the resulting DD iterations can be interpreted as solving optimal control problems on the subdomains during the DD iteration, i.e. the green arrow going up on the right in Figure 1.1, see e.g. [42, 21, 22, 27, 17, 16, 43, 51]. DD methods can also be interpreted as an optimization problem, see for example [18], where optimization techniques are used to minimize the jumps in interface traces.

We focus here on the rather new *decompose-then-optimize* approach indicated by the red arrows going right and down in Figure 1.1. The idea here is to add some *continuity constraints* in the feasible set associated to the decomposed PDEs. These constraints must then be handled by different optimization techniques. Due to the many optimization techniques, one can discover new DD methods doing this, and even try to interpret their meaning as DD methods for the first order optimality system, as indicated by the left going red arrow in Figure 1.1. A very fruitful source for such new methods is the augmented Lagrangian technique introduced by Hestenes in [31], see also Powell [49]. In this approach one has, for (some of) the constraints, a penalization function composed of a quadratic penalty term and a scalar product involving a Lagrange multiplier. There is then a precise iterative algorithm on how to update the Lagrange multiplier and the penalty parameter in order to converge to an optimum, the augmented Lagrangian method. Without DD, this method has been successfully applied to many PDE-constrained optimization problems over the past decades, see [1, 3, 4, 5, 11, 12, 13, 32, 35, 36, 37], and the method is suitable for the analysis in infinite dimensional spaces, see for instance [38, 39, 40].

Our goal is to show how one can design and analyze new DD algorithms based on the principle of *decompose-and-optimize*, and we follow Emile Picard's recommendation [48] to do so: "Les méthodes d'approximation dont nous faisons usage sont théoriquement susceptibles de s'appliquer à toute équation, mais elles ne deviennent vraiment intéressantes pour l'étude des propriétés des fonctions définies par les équations différentielles que si l'on ne reste pas dans les généralités et si l'on envisage certaines classes d'équation[1]." We will thus choose a specific PDE-constrained optimization problem and use the augmented Lagrangian method to solve it to discover a new domain decomposition method based on the *decompose-then-optimize* approach.

Our paper is organized as follows: in Section 2, we present the decompose-then-optimize method based on the augmented Lagrangian algorithm for a linear quadratic optimization problem. In Section 3 we study the convergence of the new DD method for 2 subdomains, and explain how the results can be extended to more general decompositions ; we also use Fourier techniques to get more insight into the convergence behavior. Afterwards, we show numerically the excellent scalability properties of our algorithm in Section 4, and we draw some conclusions in Section 5.

Notations. We denote by ∇q the gradient of a real-valued function. Assuming that we have a Hilbert space $\mathcal{H}$ and a subspace $\mathcal{X}$ such that $\mathcal{X} \subset \mathcal{H} \subset \mathcal{X}^*$ is a Gelfand triple, the directional derivative of a function $F : x \in \mathcal{X} \mapsto F(x) \in \mathbb{R}$ is

$$\partial_x F(x)[\delta x] = \lim_{t \to 0} \frac{F(x + t\delta x) - F(x)}{t} = \langle \partial_x F(x), \delta x \rangle_{\mathcal{X}^*, \mathcal{X}} \, ,$$

where $\partial_x F(x)$ is the gradient of F. The notation $A \lesssim B$ means that there exists a constant C which can depend only the domain Ω, $C = C(\Omega)$, such that $A \leq C(\Omega)B$. For $\Gamma_\cap \subset \partial\Omega$, we denote by $H^s(\Gamma_\cap)$ the restriction to $\Gamma_\cap$ of distributions in $H^s(\partial\Omega)$, and by $H^{\frac{1}{2}}_{00}(\Gamma_\cap)$ the set of distributions defined on $\Gamma_\cap$ such that their extension with 0 on $\partial\Omega$ is in $H^{\frac{1}{2}}(\partial\Omega)$. We refer to [45] for more information about these trace spaces.

[1]The successive approximation methods we use can in theory be applied to any equation, but they only become really interesting for the study of properties of functions defined by differential equations if one does not remain in generalities, and considers certain classes of equations.

2. A decompose-then-optimize method based on the augmented Lagrangian algorithm. As indicated above, we use as our model problem the linear quadratic PDE constrained optimization problem[2]

$$
\begin{aligned}
&\min \quad \tilde{\mathcal{J}}(y,u) := \tfrac{1}{2}\|y - y_{\text{target}}\|^2_{L^2(\Omega)} + \tfrac{\alpha}{2}\|u\|^2_{L^2(\Omega)} \\
&s.t. \quad \begin{cases} -\Delta y = f + u \text{ on } \Omega, \\ y|_{\partial\Omega} = 0, \end{cases}
\end{aligned} \tag{2.1}
$$

where Ω is a bounded open subset of $\mathbb{R}^d$, $d \geq 1$, with Lipschitz boundary, $u \in L^2(\Omega)$ is our control and $f \in L^2(\Omega)$ is an imposed source term.

For FETI, an ad hoc augmented Lagrangian formulation penalizing the Dirichlet jumps has already been proposed in [46], which led to Robin transmission conditions when interpreted as a DD method, see also [2] for an augmented Lagrangian formulation with inequality constraints for FETI. For the PDE constrained optimization problem (2.1) however, we are only aware of [28, 34], where a quadratic penalization approach was studied with the PDE constraint decomposed into subdomains, and the trace jumps at the interfaces were then penalized in the augmented cost formulation. In [41], the authors also use an augmented Lagrangian method but to penalize the entire PDE, in addition to a decomposition technique, with no proof on convergence.

We decompose the domain of the PDE constraint in (2.1) into J non-overlapping subdomains, $\overline{\Omega} = \cup_{j=1}^J \overline{\Omega}_j$, $\overline{\Omega}_i \cap \overline{\Omega}_j = \emptyset$, and then add continuity constraints along the interfaces $\Gamma_{ij} := \partial\Omega_i \cap \partial\Omega_j = \Gamma_{ji}$. We denote the normal on Γ_{ij} outward from Ω_i into Ω_j by $\mathbf{n}_{ij}$. The decomposed optimization problem then reads

$$
\begin{aligned}
&\min_{y,u,g} \mathcal{J}(y,u,g) = \sum_{j=1}^{J} \frac{1}{2}\|y_j - y_{\text{target}}\|^2_{L^2(\Omega_j)} + \frac{\alpha}{2}\|u_j\|^2_{L^2(\Omega_j)} \\
&\text{s.t.} \begin{cases} -\Delta y_j = f + u_j & \text{in } \Omega_j, \\ y_j = 0 & \text{on } \partial\Omega \cap \Omega_j, \\ \partial_{\mathbf{n}_{ji}} y_j|_{\Gamma_{ji}} = g_{ji}, \ g_{ji} = -g_{ij} & \text{on } \Gamma_{ij} \neq \emptyset, \\ y_j = y_i & \text{on } \Gamma_{ij} \neq \emptyset, \\ (u_j, g_j) \in \mathcal{U}. \end{cases}
\end{aligned} \tag{2.2}
$$

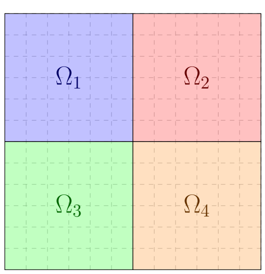


Fig. 2.1: Example decomposition of Ω into four subdomains Ω_j.

The controls $u = (u_1, u_2, \ldots, u_J)$ and the states $y = (y_1, y_2, \ldots, y_J)$ are now split into subdomain quantities, and $\mathcal{U}$ is the set of admissible controls which is defined below. Note that we added additional unknowns function $g := (g_{ji})_{ji}$ on the interfaces, which represent the normal derivative there that must match between the two subdomains, and also equations, because the matching conditions $\partial_{\mathbf{n}_{ji}} y_j|_{\Gamma_{ji}} = \partial_{\mathbf{n}_{ji}} y_i|_{\Gamma_{ji}}$ are replaced in (2.2) by the two equations stating that each normal trace must equal g_{ji}. Hence the new unknowns g_{ji}, like the unknown solutions y_j, must now appear in the minimization problem formulation in the first line in (2.2) on the left.

The set of admissible controls $\mathcal{U}$ is defined as

$$
\mathcal{U} := \left\{ (u_j, g_{ji}) \mid u_j \in L^2(\Omega_j), \ g_{ji} \in H^{-1/2}(\Gamma_{ji}) \text{ and } g_{ji} = -g_{ij} \right\}.
$$

[2] We use $\tilde{\mathcal{J}}$ here to have $\mathcal{J}$ for the decomposed optimization problem.

We can prove that the problems (2.1) and (2.2) have the same feasible sets and are therefore equivalent. Note that we could also have added an unknown for the Dirichlet traces instead, or even both Dirichlet and Neumann, which would lead to further new DD algorithms with the approach we describe below.

We next define a reduced problem associated to (2.2) by eliminating the subdomain solutions y_j (like static condensation),

$$\begin{aligned} \min_{u,g} \widehat{\mathcal{J}}(u,g) &:= \sum_{j=1}^{J} \frac{1}{2}\|y_j(u_j,g_{ji}) - y_{\text{target}}\|^2_{L^2(\Omega_j)} + \frac{\alpha}{2}\|u_j\|^2_{L^2(\Omega_j)} \\ &\text{s.t. } y_j|_{\Gamma_{ji}} - y_i|_{\Gamma_{ji}} = 0,\ (u,g) \in \mathcal{U}, \end{aligned} \tag{2.3}$$

where the statically condensed solutions $y_j(u_j, g)$ satisfy

$$\begin{cases} -\Delta y_j = u_j + f & \text{in } \Omega_j, \\ y_j = 0 & \text{on } \partial\Omega \cap \partial\Omega_j, \\ \partial_{\mathbf{n}_{ji}} y_j|_{\Gamma_{ji}} = g_{ji} & \text{on each part of } \partial\Omega_j \setminus \partial\Omega = \cup_i \Gamma_{ji}. \end{cases} \tag{2.4}$$

Our new DD method is obtained when applying an augmented Lagrangian algorithm to the reduced optimization problem (3.2). In the augmented Lagrangian algorithm, one does not only add the Lagrange multiplier term to the cost functional to form the Lagrangian, but in addition also penalizes (some of) the constraints in the Lagrangian. The augmented Lagrangian algorithm is then given by an iteration, in which at each iteration step an unconstrained optimization problem is solved for fixed values of the Lagrange multipliers and penalization parameters, and based on the result, these multipliers and parameters are updated, see [31]. In the context of (2.3), the continuity constraints $y_j|_{\Gamma_{ji}} - y_i|_{\Gamma_{ji}} = 0$ are those we penalize. The augmented cost functional we use for the reduced optimization problem (2.3) is

$$\begin{aligned} \widehat{\mathcal{J}}^+(u,g;\lambda,\rho) = \quad & \widehat{\mathcal{J}}(u,g) + \textstyle\sum_{(i,j)} \int_{\Gamma_{ij}} (y_i(u_i,g) - y_j(u_j,g))\lambda_{ij} \\ & + \textstyle\sum_{(i,j)} \frac{\rho_{ij}}{2}\|y_i(u_i,g) - y_j(u_j,g)\|^2_{L^2(\Gamma_{ij})}, \end{aligned} \tag{2.5}$$

where the sum is taken so that each interior interface is considered only once (there is a slight abuse of notation in (2.5) since, if $\lambda_{ij} \in H^{-1/2}(\Gamma_{ji})$, the integral $\int_{\Gamma_{ij}}$ must be understood as a duality bracket). The augmented Lagrangian algorithm then solves approximately at each iteration for given values $\lambda = \lambda^k$ and $\rho = \rho^k$ the reduced minimization problem

$$\min \widehat{\mathcal{J}}^+(u,g;\lambda^k,\rho^k) \text{ s.t. } (u,g) \in \mathcal{U}. \tag{2.6}$$

This iteration defines our new DD method for the optimal control problem (2.1) decomposed into subdomains as in (2.3). This approach is summarised in Algorithm 2.1. It should be noted that a closely related approach can be found in [8], but the authors treat the variables g_{ji} as *implicit* Lagrange multipliers for the continuity of the adjoint state over the interface $\Gamma_\cap$ (see [8, Remark 3.4.3]) and there is no convergence proof to the optimal solution.

REMARK 2.1. *Note that in Algorithm 2.1, the solution of* (2.6) *is parallelizable: for fixed u_j and g, the PDE-constraint can be solved in parallel, meaning each $y_j(u_j,g)$ can be found on independant processors. The same holds for the solution of the adjoint system (see Eq.* (3.5) *when 2-subdomains are considered). Also, during the update of*

Algorithm 2.1: New DD Algorithm based on Augmented Lagrangian

Data: Set initial tolerances $\{\eta_{ij}^0, \omega_{ij}^0\}_{ij} \subset (0,1)^2$, choose $\tau > 1$, and set $k = 0$.
Choose an initial guess u^0, g^0, and initial parameters $\{\lambda_{ij}^0\}_{ij} \subset \mathcal{V}_\cap^*$, $\{\rho_{ij}^0\}_{ij} \subset (1, \infty)$.

while $\sum_{\{ij,\ \Gamma_{ij}\neq\emptyset\}} \|y_i^{k-1} - y_j^{k-1}\|^2_{L^2(\Gamma_\cap)} \geq \max_{ij} \eta_{ij}^k$ **do**
  Solve approximately (2.6) for u^k, g^k s.t. $\|\partial_{u,g}\widehat{\mathcal{J}}^+(u^k, g^k)\|_{\mathcal{U}^*} \leq \max_{ij} \omega_{ij}^k$;
  for i, j *such that* $\Gamma_{ij} \neq \emptyset$ **do**
    if $\|y_i^k - y_j^k\|^2_{L^2(\Gamma_\cap)} \leq \eta_{ij}^k$ **then**
      $\lambda_{ij}^{k+1} = \lambda_{ij}^k + \rho_{ij}^k(y_i^k - y_j^k)$; `// update multiplier`
      $\rho_{ij}^{k+1} = \rho_{ij}^k$; `// keep penalization`
      $\omega_{ij}^{k+1} = (\rho_{ij}^k)^{-1}\omega_{ij}^k$; `// decrease tolerances`
      $\eta_{ij}^{k+1} = (\rho_{ij}^k)^{-1/2}\omega_{ij}^k$;
    else
      $\lambda_{ij}^{k+1} = \lambda_{ij}^k$; `// keep multiplier`
      $\rho_{ij}^{k+1} = \tau\rho_{ij}^k$; `// increase penalization`
      $\omega_{ij}^{k+1} = (\rho_{ij}^{k+1})^{-1}$; `// update tolerances`
      $\eta_{ij}^{k+1} = (\rho_{ij}^{k+1})^{-1/2}$;
    end
  end
  $k \leftarrow k + 1$
end

the descent method used in order to solve (2.1), an update of the u_j occurs and can be done in parallel (since it is based on the value of the adjoint variable which is defined only on the subdomain Ω_j), and most of the computation of $\widehat{\mathcal{J}}^+$ is decoupled on each subdomain. The only synchronization barrier arises for the computation of the cost on the virtual boundaries Γ_{ji} and the update of the virtual controls g_{ji}, since one needs to compute $y_j|_{\Gamma_{ji}} - y_i|_{\Gamma_{ji}}$ and thus this implies a synchronized communication between processors.

3. Convergence analysis. In this section, we will analyse the convergence of Algorithm 2.1 toward the solution of (2.1). We will only treat the two subdomains case in details, and comment on how these results can be extended to decomposition with more subdomains at the end of the current section.
Suppose we have two non-overlapping subdomains Ω_1, Ω_2 such that $\overline{\Omega} = \overline{\Omega_1} \cup \overline{\Omega_2}$, $\Omega_1 \cap \Omega_2 = \emptyset$ and $\partial\Omega \cap \partial\Omega_j \neq \emptyset$. We also need the functional spaces

$$\mathcal{V}_\cap := H_{00}^{\frac{1}{2}}(\Gamma_\cap), \quad \mathcal{U} := L^2(\Omega_1) \times L^2(\Omega_2) \times H^{-\frac{1}{2}}(\Gamma_\cap), \tag{3.1}$$

where $\Gamma_\cap = \partial\Omega_1 \cap \partial\Omega_2$ is the interface. The problem (2.3) with two subdomains then becomes

$$\begin{aligned} &\min_{u,g} \widehat{\mathcal{J}}(u, g) := \sum_{i=1}^{2} \frac{1}{2}\|y_i(u_i, g) - y_{\text{target}}\|^2_{L^2(\Omega_i)} + \frac{\alpha}{2}\|u_i\|^2_{L^2(\Omega_i)} \\ &\text{s.t. } y_1|_{\Gamma_\cap} - y_2|_{\Gamma_\cap} = 0,\ (u_1, u_2, g) \in \mathcal{U}, \end{aligned} \tag{3.2}$$

where the statically condensed solutions $y_i(u_i, g)$ satisfy

$$\left\{\begin{aligned} -\Delta y_i &= u_i + f && \text{in } \Omega_i,\\ y_i|_{\partial\Omega\cap\partial\Omega_i} &= 0, &&\\ \partial_{\mathbf{n}} y_i|_{\Gamma_\cap} &= (-1)^{i+1} g, && i = 1, 2. \end{aligned}\right. \tag{3.3}$$

The augmented Lagrangian associated to the reduced optimization problem (3.2) is defined as

$$\begin{aligned} \min_{u,g}\ \widehat{\mathcal{J}}^+(u, g; \lambda, \rho) &= \widehat{\mathcal{J}}(u, g)\\ &+ \int_{\Gamma_\cap} (y_1(u_1, g) - y_2(u_2, g))\lambda + \frac{\rho}{2}\|y_1(u_1, g) - y_2(u_2, g)\|^2_{L^2(\Gamma_\cap)}. \end{aligned} \tag{3.4}$$

The gradient of $\widehat{\mathcal{J}}^+$ with respect to u and g can be computed by introducing the Lagrangian associated to (3.2) and is given by

$$\partial_{u,g}\widehat{\mathcal{J}}^+(u, g; \lambda^k, \rho^k) = \begin{pmatrix} \alpha u_1 - \tilde{p}_1 \\ \alpha u_2 - \tilde{p}_2 \\ \tilde{p}_2|_{\Gamma_\cap} - \tilde{p}_1|_{\Gamma_\cap} \end{pmatrix},$$

where $\tilde{p}_i$ are the adjoint variables that satisfy

$$\left\{\begin{aligned} -\Delta\tilde{p}_i + y_i &= y_{\text{target}},\\ \tilde{p}_i|_{\partial\Omega} &= 0,\\ \partial_{\boldsymbol{n}_i}\tilde{p}_i &= (-1)^i \left(\lambda^k + \rho^k(y_1 - y_2)\right) \text{ on } \Gamma_\cap. \end{aligned}\right. \tag{3.5}$$

Note also that in this two subdomain case, $\eta^k_{ij} = \eta^k$ and $\omega^k_{ij} = \omega^k$.

3.1. Existence of a unique solution. Before studying the convergence of Algorithm 2.1, we first prove that (3.4) admits a minimizer and that, as $\rho \to +\infty$, the solution of the decoupled PDE constraint (3.3) at the optimum (u, g) of (3.4) indeed satisfies the continuity at the interface. Note that Theorem 3.1 remains true for a number $J > 2$ of subdomains.

THEOREM 3.1. *For any $\lambda^k \in \mathcal{V}^*_\cap$ and $\rho > 0$, the optimization problem (3.4) admits a unique solution $(\overline{u}, \overline{g})$. If the associated $(\overline{y_1}, \overline{y_2})$ satisfy (3.3), then, for any $\lambda^k \in L^2(\Gamma_\cap)$, we have*

$$\|\overline{y_1} - \overline{y_2}\|^2_{L^2(\Gamma_\cap)} \lesssim \frac{1}{\rho} + \frac{1}{\rho^2}\left\|\lambda^k\right\|^2_{L^2(\Gamma_\cap)}, \tag{3.6}$$

where the non explicitly mentioned constants do neither depend on k nor on ρ.

Proof. First note that the admissible set $\mathcal{U}$ is convex, and the map $(u_i, g) \mapsto y_i(u_i, g)$ is affine in u_i, linear in g and injective. Therefore, we only need to prove the strict convexity of the associated function in the variables (y, u),

$$\mathcal{J}(y, u) := \frac{1}{2}\sum_{i=1}^{2}\left[\|y_i - y_{\text{target}}\|^2_{L^2(\Omega_i)} + \alpha\|u_i\|^2_{L^2(\Omega_i)}\right] + \int_{\Gamma_\cap}(y_1 - y_2)\lambda^k + \frac{\rho^k}{2}\|y_1 - y_2\|^2_{L^2(\Gamma_\cap)}.$$

$\mathcal{J}$ is twice differentiable, and its Hessian matrix is clearly positive definite. This proves that $\mathcal{J}$ is strictly convex and [33, Theorem 1.46] then gives existence and uniqueness of the solution.

Let $(\overline{u}, \overline{g})$ be the unique solution to (3.4). We then have

$$\mathcal{J}\left(\overline{y}(\overline{u}, \overline{g}), \overline{u}\right) \leq \mathcal{J}(y(u,g), u), \text{ for all admissible } u, g. \tag{3.7}$$

For any fixed $u \in L^2(\Omega)$, let $y_\Omega \in H_0^1(\Omega)$ be the unique solution to $-\Delta y_\Omega = u$ and g_Ω such that the restrictions $y_{\Omega,i} := y_\Omega|_{\Omega_i}$ satisfy $\partial_{\mathbf{n}_i} y_i|_{\Gamma_\cap} = (-1)^{i+1} g_\Omega$. Since $y_\Omega \in H^1(\Omega)$, we have $y_{\Omega,1} = y_{\Omega,2}$ on $\Gamma_\cap$, and (3.7) then yields

$$\begin{aligned}\int_{\Gamma_\cap} (\overline{y_1} - \overline{y_2})\lambda^k + \frac{\rho}{2}\|\overline{y_1} - \overline{y_2}\|^2_{L^2(\Gamma_\cap)} &\leq \mathcal{J}(y_\Omega(u, g_\Omega), u) \\ &= \frac{1}{2}\left[\|y_\Omega - y_{\text{target}}\|^2_{L^2(\Omega)} + \alpha\|f\|^2_{L^2(\Omega)}\right] =: C.\end{aligned}$$

Young's inequality together with the Cauchy-Schwarz inequality then gives

$$2\left|\int_{\Gamma_\cap} (\overline{y_1} - \overline{y_2})\lambda^k\right| \leq 2\left\|\lambda^k\right\|_{L^2(\Gamma_\cap)} \|\overline{y_1} - \overline{y_2}\|_{L^2(\Gamma_\cap)} \leq \frac{2}{\rho}\left\|\lambda^k\right\|^2_{L^2(\Gamma_\cap)} + \frac{\rho}{2}\|\overline{y_1} - \overline{y_2}\|^2_{L^2(\Gamma_\cap)}.$$

Combining the previous inequalities, we get

$$\rho\|\overline{y_1} - \overline{y_2}\|^2_{L^2(\Gamma_\cap)} \leq 2C - 2\int_{\Gamma_\cap} (\overline{y_1} - \overline{y_2})\lambda^k \leq 2C + \frac{2}{\rho}\left\|\lambda^k\right\|^2_{L^2(\Gamma_\cap)} + \frac{\rho}{2}\|\overline{y_1} - \overline{y_2}\|^2_{L^2(\Gamma_\cap)},$$

from which we finally infer the estimate (3.6). ◻

3.2. Convergence of the augmented Lagrangian. We now prove convergence of our new domain decomposition Algorithm 2.1. For ease of presentation, we denote by $x := (u_1, u_2, g)$. As in [39] (see also references therein), we rely on a so-called asymptotic KKT (AKKT) conditions (see e.g. [39, Definition 5.2]). We will prove that Algorithm 2.1 implies that the generated sequence complies with this AKKT condition and that this yields convergence to the optimal solution of (3.2). As exposed in [39], this approach is common for analyzing the convergence of algorithms based on the augmented Lagrangian.

Before introducing AKKT conditions, we first give the (standard) KKT conditions for (3.2). We introduce the Lagrangian associated to (3.2) which reads

$$\mathcal{L}(x, \lambda) = \widehat{\mathcal{J}}(x) + \langle \lambda, c_o(x)\rangle_{\mathcal{V}_\cap^*, \mathcal{V}_\cap}, \tag{3.8}$$

where $c_o : x = (u_1, u_2, g) \in \mathcal{U} \mapsto y_1(u_1, g)|_{\Gamma_\cap} - y_2(u_2, g)|_{\Gamma_\cap} \in \mathcal{V}_\cap$ is the linear map giving the continuity constraint and y_i satisfies (3.3). We recall that some $x \in \mathcal{U}$ is a KKT point of (3.2) if

$$\exists \lambda \in \mathcal{V}_\cap^* \text{ such that } \partial_x \mathcal{L}(x, \lambda) = 0 \text{ and } c_o(x) = 0. \tag{3.9}$$

To compute the derivative of $\mathcal{L}$ with respect to x, we introduce the linear operator $M(x) : \mathcal{V}_\cap^* \to \mathcal{U}^*$ such that $M(x)^* = \partial_x c_o(x)$. Since y_i are affine functions of x, the linear map $\partial_x c_o(x)$ does not depend on x and we use the notation $M := M(x)$. The differential of $\mathcal{L}$ with respect to x can then be written as

$$\partial_x \mathcal{L}(x, \lambda)[\delta x] = \partial_x \widehat{\mathcal{J}}(x)[\delta x] + \langle \lambda, M^* \delta x\rangle_{\mathcal{V}_\cap^*, \mathcal{V}_\cap} = \left\langle \partial_x \widehat{\mathcal{J}}(x) + M\lambda, \delta x\right\rangle_{\mathcal{U}^*, \mathcal{U}}.$$

We can then recast (3.9) so that $x \in \mathcal{U}$ is a KKT point for (3.2) if

$$\exists \lambda \in \mathcal{V}_\cap^* \text{ such that } \partial_x \widehat{\mathcal{J}}(x) + M\lambda = 0 \text{ and } c_o(x) = 0. \tag{3.10}$$

It is usual to ask for M^* to be surjective ; this condition is called *constraint qualification.* As shown in the Appendix, for a two subdomain decomposition, one has

THEOREM 3.2. *The map $M^* : \mathcal{U} \to \mathcal{V}_\cap$ is surjective and linear.*

This constraint qualification will now serve to analyse the convergence of iterates (x^k, λ^k), which we suppose to be AKKT points, defined as follows:

DEFINITION 3.3. *We say that a feasible*[3] *point $x \in \mathcal{U}$ respects the AKKT condition, if there are sequences $x_k \to x$ in $\mathcal{U}$ and $(\lambda^k) \subset \mathcal{V}_\cap^*$ such that*

$$\partial_x \mathcal{L}(x^k, \lambda^k) = \partial_x \widehat{\mathcal{J}}(x^k) + M\lambda^k \rightharpoonup 0 \text{ as } k \to +\infty \text{ in } \mathcal{U}^*. \tag{3.11}$$

This convergence property implies boundedness of the sequence of multipliers:

LEMMA 3.4. *Suppose (3.11) is verified at some point x^* (not necessarily feasible), and let (x^k, λ^k) be the associated sequence. Then (λ^k) is bounded.*

Proof. First, note that $x \in \mathcal{U} \mapsto \partial_x \widehat{\mathcal{J}}(x) \in \mathcal{U}^*$ and $\lambda \in \mathcal{V}_\cap^* \mapsto M\lambda \in \mathcal{U}^*$ are strongly continuous. Due to Theorem 3.2 and [10, Théorème II.2.20], we have the estimate

$$\forall \lambda \in \mathcal{V}_\cap^*, \ \|\lambda\|_{\mathcal{V}_\cap^*} \lesssim \|M\lambda\|_{\mathcal{U}^*}. \tag{3.12}$$

Therefore, for any $n \in \mathbb{N}$, we get

$$\begin{aligned}
\|\lambda^n - \lambda^0\|_{\mathcal{V}_\cap^*} &\lesssim \|M\lambda^n - M\lambda^0\|_{\mathcal{U}^*} \\
&\lesssim \|\partial_x \widehat{\mathcal{J}}(x^n) + M\lambda^n\|_{\mathcal{U}^*} + \|\partial_x \widehat{\mathcal{J}}(x^0) + M\lambda^0\|_{\mathcal{U}^*} + \|\partial_x \widehat{\mathcal{J}}(x^n) - \partial_x \widehat{\mathcal{J}}(x^0)\|_{\mathcal{U}^*} \\
&\leq C.
\end{aligned}$$

∎

We now prove that Algorithm 2.1 produces a sequence $(x^k, \lambda^k)_k$ which is an AKKT point.

PROPOSITION 3.5. *Let (x^k, λ^k, ρ^k) be the sequence generated by Algorithm 2.1, and suppose that $x^k \to x^*$ in $\mathcal{U}$. Then x^* is an AKKT point.*

Proof. To prove that x^* is an AKKT point, we start by noting that the cost function $\widehat{\mathcal{J}}^+(x; \lambda, \rho)$ defined in (3.4) can be written as

$$\widehat{\mathcal{J}}^+(x; \lambda, \rho) = \widehat{\mathcal{J}}(x) + \langle \lambda, c_o(x) \rangle_{\mathcal{V}_\cap^*, \mathcal{V}_\cap} + \frac{\rho}{2} \|c_o(x)\|^2_{L^2(\Gamma_\cap)},$$

and thus $\partial_x \widehat{\mathcal{J}}^+(x; \lambda, \rho) = \partial_x \widehat{\mathcal{J}}(x) + M(\lambda + \rho c_o(x))$. Therefore, in Algorithm 2.1, x^{k+1} complies with

$$\|\partial_x \widehat{\mathcal{J}}^+(x^{k+1}; \lambda^k, \rho^k)\|_{\mathcal{U}^*} = \|\partial_x \widehat{\mathcal{J}}(x^{k+1}) + M\lambda^{k+1}\|_{\mathcal{U}^*} \leq \omega^k. \tag{3.13}$$

As proved in [47, Lemma 3.1.1], $\lim \omega^k = \lim \eta^k = 0$. As a result, x^* verifies (3.11) and it remains to prove that x^* is a feasible point.

To prove that $c_o(x^k) \to 0$, we consider two cases: (ρ^k) bounded and $\rho^k \to +\infty$.

Case 1: Suppose (ρ^k) is bounded. Due to how (ρ^k) is updated, this implies that there exists $k_0 \in \mathbb{N}$ such that, for all $k \geq k_0$, $\rho^k = \rho^{k_0}$. Therefore, for all $k \geq k_0$, we have $\|y_1(x^k) - y_2(x^k)\|_{L^2(\Gamma_\cap)} \leq \eta^k$. Since $\eta^k \to 0$, this implies $y_1(x^k) - y_2(x^k) \to 0$ in $L^2(\Gamma_\cap)$.

[3] A point x is called feasible if $c_o(x) = 0$.

Case 2: Let us now suppose $\rho^k \to +\infty$. We recall that $\lambda^{k+1} = \lambda^k + \rho^k(y_1^k - y_2^k)$. Condition (3.13) can then be recast as

$$\left\| \left(\partial_x \widehat{\mathcal{J}}(x^{k+1}) + M\lambda^k \right) + \rho^k M \left(y_1(x^k) - y_2(x^k) \right) \right\|_{\mathcal{U}^*} \leq \omega^k.$$

Therefore, multiplying by $(\rho^k)^{-1}$, we obtain

$$\left\| M \left(y_1(x^k) - y_2(x^k) \right) \right\|_{\mathcal{U}^*} \leq \omega^k (\rho^k)^{-1} + (\rho^k)^{-1} \left\| \partial_x \widehat{\mathcal{J}}(x^{k+1}) + M\lambda^k \right\|_{\mathcal{U}^*}.$$

Since $(\partial_x \widehat{\mathcal{J}}(x^k))$ and (λ^k) are bounded, we find $M\left(y_1(x^k) - y_2(x^k)\right) \to 0$. Using now (3.12), one finds $y_1(x^k) - y_2(x^k) \to 0$ in $\mathcal{V}_\cap^*$.

Since the map $x \mapsto c_o(x)$ is continuous and we assumed that $x^k \to x^*$, we have $c_o(x^k) \to c_o(x^*)$ in $\mathcal{V}_\cap$. For both cases considered above, we end up having $c_o(x^k) = y_1(x^k) - y_2(x^k) \to 0$ in $L^2(\Gamma_\cap)$ (Case 1) or in $\mathcal{V}_\cap^*$ (Case 2). We recall that we have $\mathcal{V}_\cap \subset L^2(\Gamma_\cap) \subset \mathcal{V}_\cap^*$ with continuous embedding and since $c_o(x^k)$ converges to 0 in either $L^2(\Gamma_\cap)$ or $\mathcal{V}_\cap^*$, the uniqueness of the limit in these spaces ensure that $c_o(x^*) = 0$. ∎

We now prove that x^* is the optimal solution of (3.2).

PROPOSITION 3.6. *Let (x^k, λ^k) be the sequence generated by Algorithm 2.1. Suppose that $x^k \to x^*$ in $\mathcal{U}$, then x^* is the optimal solution of (3.2) with associated multiplier λ^* which is the weak-limit of (λ^k).*

Proof. Proposition 3.5 gives that x^* is an AKKT point. Since $\mathcal{V}_\cap^*$ is reflexive, its unit ball is weakly compact. Therefore, Lemma 3.4 implies that there exists λ^* such that $\lambda^n \rightharpoonup \lambda^*$ in $\mathcal{V}_\cap^*$. Therefore, taking the limit in (3.11), and since x^* is feasible, (x^*, λ^*) is such that

$$\partial_x \mathcal{L}(x^*, \lambda^*) = \partial_x \widehat{\mathcal{J}}(x^*) + M\lambda^* = 0, \; c_o(x^*) = 0,$$

which is exactly the KKT condition (3.10). Since the optimization problems (3.2) and the initial one (2.1) are equivalent, we obtain that λ^* is given by (3.16) and that x^* is the optimal solution of (3.2) which is unique since the latter is convex. ∎

REMARK 3.7. *The multiplier λ^*, associated to the continuity constraint of the original optimization problem (3.2), is also directly linked to the continuity constraint of the adjoint variable.*
The first order necessary conditions of optimality for (3.2) are given by

$$\begin{cases} \alpha u_i - p_i = 0 \text{ in } \Omega_i, \\ p_2|_{\Gamma_\cap} - p_1|_{\Gamma_\cap} = 0, \end{cases}, \; i = 1, 2, \; \begin{cases} -\Delta p_i + y_i = y_{target}, \\ \quad p_i|_{\partial\Omega} = 0, \\ \quad \partial_{\mathbf{n}_i} p_i = (-1)^{i+1}\lambda^* \text{ on } \Gamma_\cap, \end{cases} \tag{3.14}$$

where p_i are adjoint variables and $y_i = y_i(u_i, g)$ are solution to (3.3) that also comply with $y_1(u_1, g)|_{\Gamma_\cap} - y_2(u_2, g)|_{\Gamma_\cap} = 0$.

Introducing now $y, p \in H_0^1(\Omega)$ and $u \in L^2(\Omega)$ such that

$$y|_{\Omega_i} = y_i, \; p|_{\Omega_i} = p_i, \; u|_{\Omega_i} = u_i,$$

we get that (y, p, u) satisfy

$$\begin{cases} -\Delta y = f + u \quad \text{in } \Omega \\ \quad y \in H_0^1(\Omega), \end{cases}, \; \begin{cases} -\Delta p = -(y - y_{target}) \quad \text{in } \Omega, \\ \quad p \in H_0^1(\Omega), \end{cases}, \; \alpha u - p = 0 \text{ in } \Omega. \tag{3.15}$$

We emphasize that (3.15) are exactly the first order conditions of optimality of the undecomposed optimization problem (2.1). We thus obtain that

$$\lambda^* = (-1)^{i+1}\partial_{\boldsymbol{n}_i} p|_{\Gamma_\cap} \tag{3.16}$$

where p satisfies the coupled system (3.15).

Remark 3.8 (Extension to more than 2 subdomains). *The extension of the convergence results to more subdomains mainly consists in proving that Theorem 3.2 holds when more subdomains are involved. As shown in the appendix, the proof consists in proving that M is injective (which can be easily adapted to a more complex decomposition) and with closed range. This second part seems more challenging for several subdomains and depends on how the decomposition is made. For example, the proof of Lemma A.1 can be straightforwardly adapted to stripwise decomposition like those used in the numerical experiments from Section 4, mainly because we do not have subdomains with Neumann boundary condtions only. Furthermore, as long as one supposes M^* surjective, all the results leading to Proposition 3.6 can be easily adapted to a context with more subdomains. This extension to more subdomains can be seen more easily when one carries a similar analysis we did in section 3 once the problem (3.4) is discretized ; see Appendix B.*

3.3. Fourier analysis. To get more insight into the convergence behavior of our new DD Algorithm 2.1, we study its rate of convergence in the case where the penalization parameter ρ is kept constant during the iterations, and the domain is unbounded, $\Omega := \mathbb{R}^2$, with (x_1, x_2) a system of coordinates on $\mathbb{R}^2$. Let the subdomains $\Omega_1 := (-\infty, 0) \times \mathbb{R}$, $\Omega_2 := (0, +\infty) \times \mathbb{R}$, and the interface $\Gamma_\cap := \{0\} \times \mathbb{R}$. In this setting, we can study how the iterations $\{\lambda^k\}$ are converging, see e.g. [20, 26] for this approach to study the convergence of DD methods. The necessary and sufficient optimality conditions for (3.4) are

$$\left\{\begin{aligned} -\Delta y_i^k &= f + \alpha^{-1} p_i^k \qquad \text{in } \Omega_i,\\ y_i^k|_{\partial\Omega} &= 0,\\ \partial_{\mathbf{n}_i} y_i^k|_{\Gamma_\cap} &= (-1)^{i+1} g^k, \quad i = 1,2 \end{aligned}\right. \quad \left\{\begin{aligned} -\Delta p_i^k + y_i^k &= y_{\text{target}},\\ p_i^k|_{\partial\Omega} &= 0,\\ \partial_{\mathbf{n}_i} p_i^k &= (-1)^i \left(\lambda^k + \rho(y_1^k - y_2^k)\right) \text{ on } \Gamma_\cap,\\ p_1^k &= p_2^k \text{ on } \Gamma_\cap. \end{aligned}\right. \tag{3.17}$$

Since we are interested in the error in the new DD Algorithm 2.1, without loss of generality, we will suppose that $f \equiv 0$ and $y_{\text{target}} \equiv 0$, so that we can study convergence to zero, i.e. the error equations.

We assume that λ^0 has enough regularity so that its Fourier transform along the vertical axis is well-defined. Let $\omega \in \mathbb{R}$ be the Fourier variable and $(\hat{y}, \hat{p})$ be the Fourier transform in the x_2 direction of (y, p) along the interface. The Fourier transformed optimality conditions (3.17) then become

$$\left\{\begin{aligned} (-\partial_{x_1}^2 + \omega^2)\hat{y}_i^k &= \alpha^{-1}\hat{p}_i^k \qquad \text{in } \Omega_i,\\ \hat{y}_i^k|_{\partial\Omega} &= 0,\\ \partial_{x_1}\hat{y}_i^k|_{\Gamma_\cap} &= \hat{g}^k, \qquad i = 1,2 \end{aligned}\right. \quad \left\{\begin{aligned} (-\partial_{x_1}^2 + \omega^2)\hat{p}_i^k + \hat{y}_i^k &= 0,\\ \hat{p}_i^k|_{\partial\Omega} &= 0,\\ \partial_{x_1}\hat{p}_i^k &= -\left(\hat{\lambda}^k + \rho(\hat{y}_1^k - \hat{y}_2^k)\right) \text{ on } \Gamma_\cap,\\ \hat{p}_1^k &= \hat{p}_2^k \text{ on } \Gamma_\cap. \end{aligned}\right. \tag{3.18}$$

In order to solve this system of ordinary differential equations, we compute $\partial_{x_1}^4 \hat{y}_i^k$ and substitute,

$$
\begin{aligned}
\partial_{x_1}^4 \hat{y}_i^k =&\omega^2 \partial_{x_1}^2 \hat{y}_i^k - \alpha^{-1} \partial_{x_1}^2 \hat{p}_i^k \\
=&\omega^2 \partial_{x_1}^2 \hat{y}_i^k - \alpha^{-1}(\omega^2 \hat{p}_i^k + \hat{y}_i^k) \\
=&\omega^2 \partial_{x_1}^2 \hat{y}_i^k - \alpha^{-1} \hat{y}_i^k - \omega^2(-\partial_{x_1}^2 \hat{y}_i^k + \omega^2 \hat{y}_i^k) \\
=&2\omega^2 \partial_{x_1}^2 \hat{y}_i^k - (\omega^4 + \alpha^{-1}) \hat{y}_i^k.
\end{aligned}
$$

Using the boundary conditions at infinity, the solutions are thus of the form

$$
\begin{aligned}
\hat{y}_1^k =&C_1^k \exp\left(x_1 \sqrt{\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2}\right) + C_3^k \exp\left(x_1 \sqrt{-\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2}\right), \\
\hat{y}_2^k =&C_2^k \exp\left(-x_1 \sqrt{\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2}\right) + C_4^k \exp\left(-x_1 \sqrt{-\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2}\right),
\end{aligned}
$$

where C_j^k are constants determined by the transmission conditions. This implies for the adjoint variables

$$
\begin{aligned}
\hat{p}_1^k =&\mathrm{i}\alpha^{\frac{1}{2}} \left(-C_1^k \exp\left(x_1 \sqrt{\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2}\right) + C_3^k \exp\left(x_1 \sqrt{-\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2}\right)\right), \\
\hat{p}_2^k =&\mathrm{i}\alpha^{\frac{1}{2}} \left(-C_2^k \exp\left(-x_1 \sqrt{\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2}\right) + C_4^k \exp\left(-x_1 \sqrt{-\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2}\right)\right).
\end{aligned}
$$

The transmission conditions at $x_1 = 0$ then yield

$$
\begin{aligned}
\partial_{x_1} \hat{y}_1^k|_{x_1=0} =&C_1^k \sqrt{\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2} + C_3^k \sqrt{-\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2} = \hat{g}^k, \\
\partial_{x_1} \hat{y}_2^k|_{x_1=0} =&- C_2^k \sqrt{\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2} - C_4^k \sqrt{-\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2} = \hat{g}^k, \\
\partial_{x_1} \hat{p}_1^k|_{x_1=0} =&- \mathrm{i}\alpha^{\frac{1}{2}} \left(C_1^k \sqrt{\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2} - C_3^k \sqrt{-\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2}\right) = -\hat{\lambda}^{k+1}, \\
\partial_{x_1} \hat{p}_2^k|_{x_1=0} =&\mathrm{i}\alpha^{\frac{1}{2}} \left(C_2^k \sqrt{\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2} - C_4^k \sqrt{-\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2}\right) = -\hat{\lambda}^{k+1}.
\end{aligned}
$$

Denoting by $A := \sqrt{\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2}$ and $B := \sqrt{-\mathrm{i}\alpha^{-\frac{1}{2}} + \omega^2}$, we thus need to solve the linear system

$$
\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} AC_1^k \\ AC_2^k \\ BC_3^k \\ BC_4^k \end{pmatrix} = \begin{pmatrix} \hat{g}^k \\ \hat{g}^k \\ \mathrm{i}\alpha^{-\frac{1}{2}} \hat{\lambda}^{k+1} \\ -\mathrm{i}\alpha^{-\frac{1}{2}} \hat{\lambda}^{k+1} \end{pmatrix}.
$$

The inverse of the system matrix being simply half its transpose, we get

$$
\begin{aligned}
AC_1^k = -AC_2^k = \frac{1}{2}(\hat{g}^k + \mathrm{i}\alpha^{-\frac{1}{2}} \hat{\lambda}^{k+1}), \\
BC_3^k = -BC_4^k = \frac{1}{2}(\hat{g}^k - \mathrm{i}\alpha^{-\frac{1}{2}} \hat{\lambda}^{k+1}).
\end{aligned}
$$

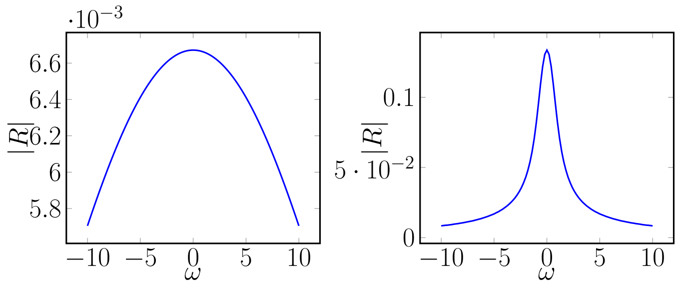


Fig. 3.1: $|R|$ for $\rho = 3$, $\alpha = 1e-5$ (Left) and $\alpha = 1$ (Right).

Note in particular that $C_3^k = -C_4^k$ and $C_1^k = -C_2^k$, and furthermore,

$$
\begin{aligned}
\hat{y}_1^k|_{x_1=0} - \hat{y}_2^k|_{x_1=0} &= C_1^k + C_3^k - (C_2^k + C_4^k) \\
&= 2(C_1^k + C_3^k) \\
&= (A^{-1} + B^{-1})\hat{g}^k + (A^{-1} - B^{-1})\mathrm{i}\alpha^{-\frac{1}{2}}\hat{\lambda}^{k+1}, \\
\hat{p}_1^k|_{x_1=0} - \hat{p}_2^k|_{x_1=0} &= -\mathrm{i}\alpha^{\frac{1}{2}}(C_1^k - C_3^k) + \mathrm{i}\alpha^{\frac{1}{2}}(C_2^k - C_4^k) \\
&= -\mathrm{i}\alpha^{\frac{1}{2}}(C_1^k - C_3^k - C_2^k + C_4^k) \\
&= 2\mathrm{i}\alpha^{\frac{1}{2}}(C_3^k - C_1^k) \\
&= \mathrm{i}\alpha^{\frac{1}{2}}(B^{-1} - A^{-1})\hat{g}^k + (A^{-1} + B^{-1})\hat{\lambda}^{k+1}.
\end{aligned}
$$

Since $\hat{p}_1^k|_{x_1=0} - \hat{p}_2^k|_{x_1=0} = 0$, this implies $\hat{g}^k = \mathrm{i}\alpha^{-\frac{1}{2}}\frac{(A+B)}{A-B}\hat{\lambda}^{k+1}$ from which we infer

$$
\hat{y}_1^k|_{x_1=0} - \hat{y}_2^k|_{x_1=0} = 4\mathrm{i}\alpha^{-\frac{1}{2}}(A-B)^{-1}\hat{\lambda}^{k+1}.
$$

Denote by $D := 4\mathrm{i}\alpha^{-\frac{1}{2}}(A-B)^{-1}$. Since in our new DD algorithm $\hat{\lambda}^k = \hat{\lambda}^{k+1} - \rho(\hat{y}_1^k|_{x_1=0} - \hat{y}_2^k|_{x_1=0})$, substituting the expression of $\hat{y}_1^k|_{x_1=0} - \hat{y}_2^k|_{x_1=0}$ gives $\hat{\lambda}^k = (1-\rho D)\hat{\lambda}^{k+1}$, and we thus obtain for the convergence factor

(3.19)

$$
R := \frac{\hat{\lambda}^{k+1}}{\hat{\lambda}^k} = (1-\rho D)^{-1} = \left(1 - \frac{4\mathrm{i}\rho}{\alpha^{1/2}\left(\sqrt{\omega^2 + \mathrm{i}\alpha^{-1/2}} - \sqrt{\omega^2 - \mathrm{i}\alpha^{-1/2}}\right)}\right)^{-1},
$$

where we substituted the expressions for D. Direct computations show that

$$
\lim_{\alpha\to 0} R = 0, \;\; R = \frac{\mathrm{i}\alpha^{1/2}}{4\rho}\left(\sqrt{\omega^2 + \mathrm{i}\alpha^{-1/2}} - \sqrt{\omega^2 - \mathrm{i}\alpha^{-1/2}}\right) + O\left(\frac{1}{\rho^2}\right).
$$

Note that, for all $\alpha \geq 0$ and any frequency ω, $|R|$ is strictly smaller than 1 as soon as ρ is large enough, which indicates geometric convergence of the new DD algorithm. We show a plot of $|R|$ for chosen parameters in Figure 3.1. We see that convergence is rather fast, even for low frequencies ω, and the method is a smoother: high frequencies converge extremely fast. Such behavior is also confirmed by our numerical results from Section 4 which show that the convergence of our decompose-then-optimize method is robust with respect to meshsize along with the number of subdomains. This Fourier analysis also indicates that these properties are going to hold regardless the discretization used.

4. Numerical experiments. We now test our algorithm on two examples. We will mainly pay attention to how the error evolves with the number of iterations and when the mesh is refined or the number of domains increases. An iteration here should be understood as the number of times one solves (2.6) and updates ρ or λ; see the **while** loop in Algorithm 2.1. We discretized the Laplacian using Q1 Finite Elements on a structured cartesian mesh with step size h in both x_1 and x_2 directions. The minimization problems in each iteration of our new DD algorithm are solved using the *quadprog* routine of MATLAB, which uses an interior-point algorithm, and we used $\tau = 5, \rho_{ij}^0 = 3$, and λ_{ij}^0 is a random initial guess with values between 0 and 10. The code used for the computations is available online[4].

4.1. Badly insulated room. Our first example is a badly insulated room, for which a floor heating should be designed, and we study the convergence of our new DD method when only 5 iterations are performed. The room geometry is $\Omega := (-1, 1) \times (0, 1)$, with non-insulated walls that are at temperature zero, modeled by homogeneous Dirichlet conditions, and we want to see how a floor heating system would have to operate to heat it up to 20 degrees hence $y_{\text{target}}(x_1, x_2) = 20$. The room has in addition and open window at the bottom left where cold air acts as a heat sink, and similarly a skylight on the right in the middle, see Figure 4.1 (top left). We model this with $f(x_1, x_2) = -3000\mathrm{e}^{-50((x_1-0.5)^2+(x_2-0.5)^2)} - 3000 \times \mathbf{1}_{[-0.75,-0.25]\times[0,0.25]}$, where $\mathbf{1}_{\mathcal{O}}$ is the indicator function of the set $\mathcal{O}$.

We decompose the room into two subdomains $\Omega_1 := (-1, 0) \times (0, 1)$ and $\Omega_2 := (0, 1)\times(0, 1)$, and apply our new domain decomposition algorithm in order to compute the optimal floor heating. We show in Figure 4.1 at the top the optimal control found for the floor heating by our new DD algorithm (middle), and the achieved room temperature (right), when using as penalization $\alpha = 1e - 5$. We see that such a badly insulated room is not at all suitable for floor heating, which needs to be evenly distributed; a more classical configuration with radiators along the walls, and especially in front of the window would be much more suitable. Floor heating systems are only good for well insulated buildings! In the bottom row we show the convergence of our new DD algorithm, on the left for the controls u_j and on the right for the solutions y_j. The error here is the difference between the iterates of our new DD algorithm and the solution computed on the whole domain directly, measured in the infinity norm in volume. We see that convergence is very fast: the error is reduces by 6 orders of magnitude in five iterations, and convergence is robust in the mesh size h: when the mesh is refined, convergence gets actually a little faster before approaching a limit.

We next increase the number of subdomains, using a decomposition into equal vertical strips $\Omega_j = (x_j, x_{j+1}) \times (0, 1),\ x_j = -1 + 2j/J,\ j = 0, ..., J$. We show in Figure 4.2 the errors for $J = 2, 4, 8, 16$ subdomains and meshsize $h = 1/64$. We see that convergence is still very fast and does not depend on the number of subdomains, the new DD method is scalable.

4.2. Convergence behavior with respect to α. We now test the convergence properties of our new DD algorithm as α varies. To do this, we consider a model problem inspired by [27], namely

$$y_{\text{target}}(x_1, x_2) := C \sin(l\pi x_1) \sin(l\pi x_2), \text{ and } f(x_1, x_2) := 2Cl^2\pi^2 \sin(l\pi x_1) \sin(l\pi x_2),$$

[4] https://gitlab.osureunion.fr/avieira/liquofeti

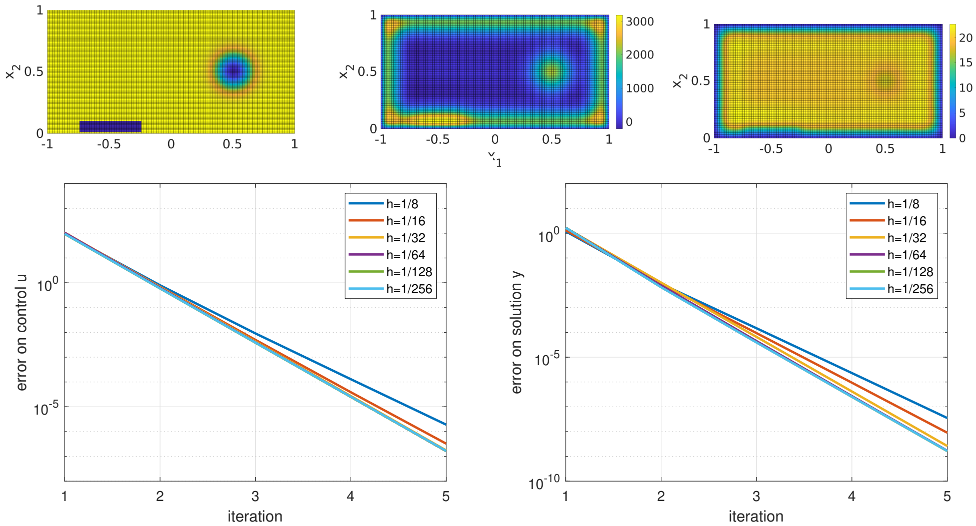


Fig. 4.1: Badly insulated room example: heat sinks (top left), best floor heating found by our new DD algorithm (top middle), temperature distribution achieved (top right). Bottom left: convergence of the new DD algorithm for the control u for different mesh sizes (log scale for the ordinate axis); bottom right: corresponding results for the solution y (log scale for the ordinate axis).

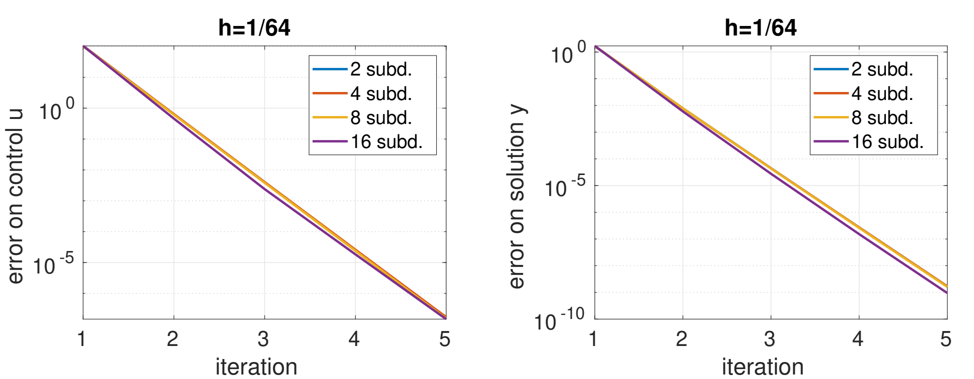
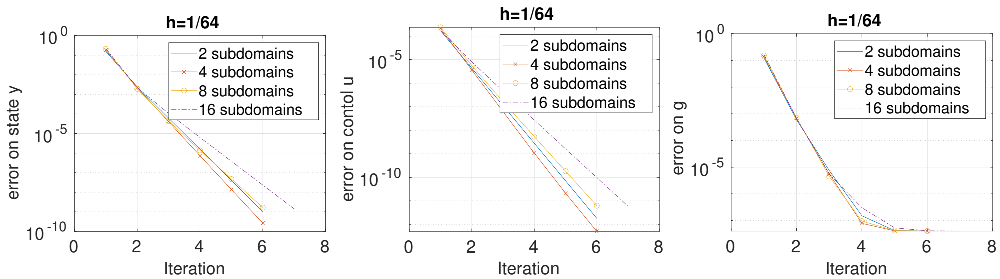


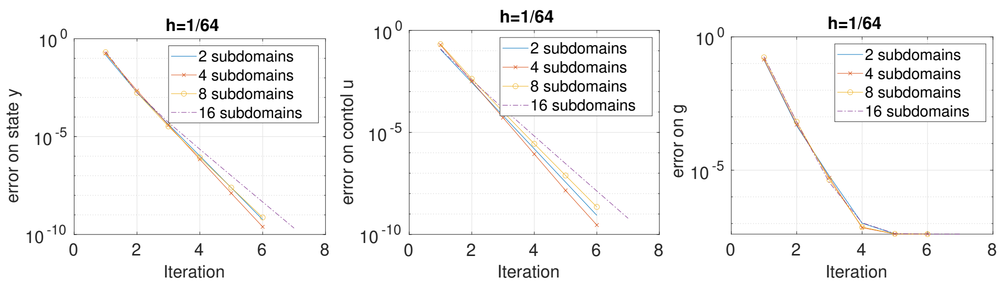
Fig. 4.2: Convergence of the new DD algorithm for the control u (left) and solution y (right) for different different numbers of subdomains for the badly insulated room example. Both figures are with log scale for the ordinate axis.

for different parameters $C \in \mathbb{R}$, and $l \in \mathbb{N}$. The optimal solution is $u^* = 0$ and $y^* = y_{\text{target}}$, for all admissible choices of parameters C, l, α. We solve this problem with our new DD Algorithm 2.1 using a decomposition into equal vertical strips as in the first physical example, and convergence parameters $\max_{ij} \omega_{ij}^k = \max_{ij} \eta_{ij}^k = 10^{-10}$.

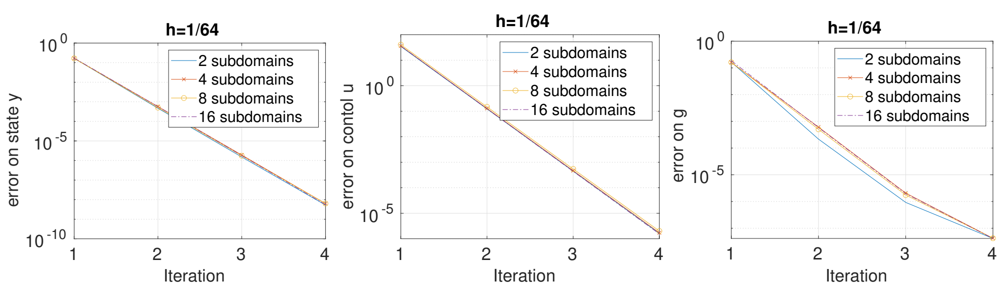
Results for different values of α, h and the number of subdomains are shown in Table 4.1 and Figure 4.3. We see that our new DD algorithm remains robust and fast



(a) $\alpha = 10$



(b) $\alpha = 10^{-2}$



(c) $\alpha = 10^{-6}$

Fig. 4.3: Convergence of the new DD algorithm with different values of α on the second example.

when changing the parameter α, the stepsize h and the number of subdomains. The number of iterations needed to achieve the prescribed precision is roughly constant in all our tests. Note nevertheless that, for large α, the rate slightly changes with a bigger number of subdomains.

5. Conclusions. We introduced a new concept for designing domain decomposition methods to solve optimal control problems in parallel: instead of following the optimize then decompose approach, we followed the decompose then optimize approach. Using then an augmented Lagrangian algorithm for the decomposed problem (or any other optimization technique) leads to new types of domain decomposition methods for such problems. We studied a particular example of such a new domain decomposition method, namely when the Dirichlet trace jump is also added as a penalization term in the augmented Lagrangian formulation.

Many other new domain decomposition methods can be obtain this way, for example by also including in the penalization the Neumann jump, or the entire PDE constraint. Our choice led to a domain decomposition method which is robust under mesh refinement, which lets us believe that this approach for designing new domain decomposition methods is a rather powerful one.

Appendix A. Surjectivity of operator M^*.

In this section, we will focus on proving that the derivative of the continuity constraint given by $c_o : (u_1, u_2, g) \in \mathcal{U} \mapsto y_1(u_1, g)|_{\Gamma_\cap} - y_2(u_2, g)|_{\Gamma_\cap} \in \mathcal{V}_\cap$, where $y_i(u_i, g)$ solves (3.3), is surjective. This condition (called regularity condition or constraint qualification) serves in order to prove that a minimizer of (3.2) respects the first order conditions of optimality (or KKT conditions). We recall that we denote by M^* the derivative of c_o with respect to (u_1, u_2, g). We first derive an explicit expression for M. To do this, we introduce the Lagrangian of (3.2),

$$\begin{aligned}\mathcal{L}(y, u, g, p, \lambda) = &\sum_{i=1}^{2} \int_{\Omega_i} \nabla y_i \cdot \nabla p_i - (-1)^{i+1} \int_{\Gamma_\cap} g p_i - \int_{\Omega_i} (f_i + u_i) p_i \\ &+ \frac{1}{2} \int_{\Omega_i} (y_i - y_{\text{target}})^2 + \frac{\alpha}{2} \int_{\Omega_i} u_i^2 + \int_{\Gamma_\cap} (y_1 - y_2)\lambda.\end{aligned}$$

where we set $u := (u_1, u_2)$, $y := (y_1, y_2)$ and $p := (p_1, p_2)$. In order to derive an explicit expression of M, we compute the derivative of $\mathcal{L}$ and obtain

$$\partial_{u,g} \mathcal{L}(y(u, g), u, g, p, \lambda) = \begin{pmatrix} \alpha u_1 - p_1 \\ \alpha u_2 - p_2 \\ p_1 - p_2|_{\Gamma_\cap} \end{pmatrix} = \partial_{u,g} \widehat{\mathcal{J}}(u, g) + M\lambda,$$

where $M : \mathcal{V}_\cap^* \to \mathcal{U}^*$ is the adjoint of M^*. Using classical computations, we can also prove that

$$\partial_{u,g} \widehat{\mathcal{J}}(u, g) = \begin{pmatrix} \alpha u_1 - \bar{p}_1 \\ \alpha u_2 - \bar{p}_2 \\ \bar{p}_2|_{\Gamma_\cap} - \bar{p}_1|_{\Gamma_\cap} \end{pmatrix} \text{ where } \begin{cases} -\Delta \bar{p}_i + y_i = y_{\text{target}}, \\ \bar{p}_i|_{\partial\Omega} = 0, \\ \partial_{\mathbf{n}} \bar{p}_i = 0 \text{ on } \Gamma_\cap. \end{cases}$$

Therefore, we get

$$M\lambda = \partial_{u,g} \mathcal{L}(y(u, g), u, g, p, \lambda) - \partial_{u,g} \widehat{\mathcal{J}}(u, g) = \begin{pmatrix} q_1 \\ q_2 \\ q_1|_{\Gamma_\cap} - q_2|_{\Gamma_\cap} \end{pmatrix},$$

where $q_i = \bar{p}_i - p_i \in H^1(\Omega_i)$ is the weak solution of

$$\begin{cases} -\Delta q_i = 0, \\ q_i|_{\partial\Omega} = 0, \\ \partial_{\mathbf{n}} q_i = (-1)^{i+1} \lambda \text{ on } \Gamma_\cap. \end{cases} \tag{A.1}$$

Note that M does not depend on the controls.

LEMMA A.1. *The operator $M : \lambda \in \mathcal{V}_\cap^* \mapsto M\lambda \in \mathcal{U}^*$ is injective with closed range.*

Proof. Injectivity: Let λ be such that $M\lambda = 0$. The associated q_i are then harmonic and verify the transmission condition (on $\Gamma_\cap$) $q_1|_{\Gamma_\cap} - q_2|_{\Gamma_\cap} = 0$ and $\partial_{\mathbf{n}_1} q_1 +$

$\partial_{\mathbf{n}_2} q_2 = 0$. As a result, $q_i = q|_{\Omega_i}$ where $q \in H_0^1(\Omega)$ is harmonic. Therefore $q = 0$ and thus $\lambda = \partial_{\mathbf{n}_1} q_1 = 0$.

Closed range: Let $M\lambda_n = (q_{1,n}, q_{2,n}, \varphi_n)^\intercal$ be a sequence of images such that $q_{i,n}$ converges toward some q_i in $L^2(\Omega_i)$ and φ_n converges toward some φ in $H_{00}^{1/2}(\Gamma_\cap)$. To show that M has closed range, we have to prove that there exists $\lambda \in H_{00}^{1/2}(\Gamma_\cap)^*$ such that $M\lambda = (q_1, q_2, \varphi)^\intercal$.

Step 1: Since $q_{i,n}$ is harmonic on Ω_i, we have

$$\forall \psi_i \in \mathcal{C}_c^\infty(\Omega_i) : \int_{\Omega_i} q_{i,n} \Delta \psi_i \, dx = 0.$$

Using the L^2 convergence of $q_{i,n}$ toward q_i, we obtain that q_i satisfies

$$\forall \psi_i \in \mathcal{C}_c^\infty(\Omega_i) : \int_{\Omega_i} q_i \Delta \psi_i \, dx = 0,$$

and Weyl's Lemma (see e.g. [52, p. 78, Theorem 18.G]) ensures that $q_i \in \mathcal{C}^\infty(\Omega_i)$ and satisfies $\Delta q_i = 0$ pointwise in Ω_i.

Step 2: Let us consider the spaces

$$H_{i,\Delta} := \left\{ \Phi_i \in H^1(\Omega_i) \mid \Delta \Phi_i \in L^2(\Omega_i),\ \Phi_i|_{\partial\Omega_i \setminus \Gamma_\cap} = 0 \right\}.$$

For any $\varphi \in H_{00}^{1/2}(\Gamma_\cap)$, we denote by $\widetilde{\varphi}$ its extension by zero to $\partial\Omega_i$. The latter is in $H^{1/2}(\partial\Omega_i)$ and the surjectivity of the trace operator gives the existence of some $\mathcal{E}\tilde{\varphi} \in H^1(\Omega_i)$ such that $\mathcal{E}\tilde{\varphi}|_{\partial\Omega_i} = \tilde{\varphi}$ and $\|\mathcal{E}\tilde{\varphi}\|_{H^1(\Omega_i)} \lesssim \|\tilde{\varphi}\|_{H_{00}^{1/2}(\partial\Omega_i)}$. From the Green's formula, for all $\Phi_i \in H_{i,\Delta}$,

$$\langle \partial_{\mathbf{n}_i} \Phi_i, \varphi \rangle_{H_{00}^{1/2}(\Gamma_\cap)^* \times H_{00}^{1/2}(\Gamma_\cap)} = \int_{\Omega_i} \nabla \Phi_i \cdot \nabla \mathcal{E}\tilde{\varphi} \, dx + \int_{\Omega_i} \Delta \Phi_i \, \mathcal{E}\tilde{\varphi} \, dx,$$

and thus, we can prove, using the Cauchy-Schwarz inequality, that

$$\left| \langle \partial_{\mathbf{n}_i} \Phi_i, \varphi \rangle_{H_{00}^{1/2}(\Gamma_\cap)^* \times H_{00}^{1/2}(\Gamma_\cap)} \right| \lesssim \|\varphi\|_{H_{00}^{1/2}(\Gamma_\cap)} \left(\|\Phi_i\|_{H^1(\Omega_i)} + \|\Delta \Phi_i\|_{L^2(\Omega_i)} \right).$$

As a result, taking the supremum over all φ such that $\|\varphi\|_{H_{00}^{1/2}(\Gamma_\cap)} = 1$, we prove that the linear map

$$\partial_{\mathbf{n}_i} : \Phi_i \in H_{i,\Delta} \mapsto \partial_{\mathbf{n}_i} \Phi_i \in H_{00}^{1/2}(\Gamma_\cap)^*$$

is continuous. In addition, for any harmonic $\Phi_i \in H_{i,\Delta}$, we have the bound

$$\|\partial_{\mathbf{n}_i} \Phi_i\|_{H_{00}^{1/2}(\Gamma_\cap)^*} \lesssim \|\Phi_i\|_{H^1(\Omega_i)} \lesssim \|\nabla \Phi_i\|_{L^2(\Omega_i)},$$

where we used the Poincaré inequality to get the last upper bound.

Step 3: Both $q_{i,n}$ satisfy the weak formulation

$$\forall \psi_i \in V_i : \int_{\Omega_i} \nabla q_{i,n} \cdot \nabla \psi_i \, dx = \langle \partial_{\mathbf{n}_i} q_{i,n}, \psi_i \rangle_{H_{00}^{1/2}(\Gamma_\cap)^* \times H_{00}^{1/2}(\Gamma_\cap)},$$

where $V_i := \left\{ \psi_i \in H^1(\Omega_i) \mid \psi_i|_{\partial\Omega_i \setminus \Gamma_\cap} = 0 \right\}$. Now taking $\psi_i = q_{i,n}$, one gets

$$\begin{aligned}
\sum_i \|\nabla q_{i,n}\|_{L^2(\Omega_i)}^2 &= \langle \partial_{\mathbf{n}_i} q_{1,n}, q_{1,n} \rangle_{H_{00}^{1/2}(\Gamma_\cap)^* \times H_{00}^{1/2}(\Gamma_\cap)} + \langle \partial_{\mathbf{n}_2} q_{2,n}, q_{2,n} \rangle_{H_{00}^{1/2}(\Gamma_\cap)^* \times H_{00}^{1/2}(\Gamma_\cap)} \\
&= \langle \partial_{\mathbf{n}_1} q_{1,n}, q_{1,n} - q_{2,n} \rangle_{H_{00}^{1/2}(\Gamma_\cap)^* \times H_{00}^{1/2}(\Gamma_\cap)} \\
&= \langle \partial_{\mathbf{n}_1} q_{1,n}, \varphi_n \rangle_{H_{00}^{1/2}(\Gamma_\cap)^* \times H_{00}^{1/2}(\Gamma_\cap)},
\end{aligned}$$

where we used that $\partial_{\mathbf{n}_1} q_{1,n}|_{\Gamma_\cap} + \partial_{\mathbf{n}_2} q_{2,n}|_{\Gamma_\cap} = 0$ and that, due to the definition of M, we have $\varphi_n = q_{1,n} - q_{2,n}$. Using now Step 2, we obtain that

$$\sum_i \|\nabla q_{i,n}\|^2_{L^2(\Omega_i)} \leq \|\partial_{\mathbf{n}_i} q_{1,n}\|_{H^{1/2}_{00}(\Gamma_\cap)^*} \|\varphi_n\|_{H^{1/2}_{00}(\Gamma_\cap)} \lesssim \|\nabla q_{1,n}\|_{L^2(\Omega_1)} \|\varphi_n\|_{H^{1/2}_{00}(\Gamma_\cap)}.$$

Using finally Young's inequality, we get

$$\sum_i \|\nabla q_{i,n}\|^2_{L^2(\Omega_i)} \lesssim \|\varphi_n\|^2_{H^{1/2}_{00}(\Gamma_\cap)}.$$

Since φ_n is assumed to converge toward φ, it is bounded and thus the sequences $(q_{i,n})_n$ are also bounded (uniformly with respect to n) in $H^1(\Omega_i)$. We can then extract subsequences that converge weakly in $H^1(\Omega_i)$ toward q_i. The trace operator being compact, we obtain that $q_i|_{\partial\Omega_i \setminus \Gamma_\cap} = 0$ and $\varphi = q_1|_{\Gamma_\cap} - q_2|_{\Gamma_\cap}$. Since $q_i \in V_i$ is harmonic, it also belongs to $H_{i,\Delta}$ and its normal derivative can be defined as an element of $H^{1/2}_{00}(\Gamma_\cap)^*$.

Step 4: We now identify the limit of the sequence λ_n. Let $\eta \in H^{1/2}_{00}(\Gamma_\cap)$, since its extension by zero (denoted by $\tilde{\eta}$) over $\partial\Omega_i$ is in $H^{1/2}(\partial\Omega_i)$, we have some $\mathcal{E}_i\tilde{\eta} \in V_i$ such that $\mathcal{E}_i\tilde{\eta}|_{\partial\Omega_i} = \tilde{\eta}$ and $\|\mathcal{E}_i\tilde{\eta}\|_{H^1(\Omega_i)} \lesssim \|\tilde{\eta}\|_{H^{1/2}_{00}(\Gamma_\cap)}$. Using that $\lambda_n = (-1)^{i+1}\partial_{\mathbf{n}_i} q_{i,n}$ and the weak formulation satisfied by $q_{i,n}$, we have

$$\begin{aligned}
\langle \lambda_n, \eta \rangle_{H^{1/2}_{00}(\Gamma_\cap)^* \times H^{1/2}_{00}(\Gamma_\cap)} &= \frac{1}{2} \langle \partial_{\mathbf{n}_1} q_{1,n}, \mathcal{E}_1\tilde{\eta}|_{\Gamma_\cap} \rangle_{H^{1/2}_{00}(\Gamma_\cap)^* \times H^{1/2}_{00}(\Gamma_\cap)} \\
&\quad - \frac{1}{2} \langle \partial_{\mathbf{n}_2} q_{2,n}, \mathcal{E}_2\tilde{\eta}|_{\Gamma_\cap} \rangle_{H^{1/2}_{00}(\Gamma_\cap)^* \times H^{1/2}_{00}(\Gamma_\cap)} \\
&= \frac{1}{2} \int_{\Omega_1} \nabla q_{1,n} \cdot \nabla \mathcal{E}_1\tilde{\eta}\, dx - \frac{1}{2} \int_{\Omega_2} \nabla q_{2,n} \cdot \nabla \mathcal{E}_2\tilde{\eta}\, dx.
\end{aligned}$$

Now passing to the limit (after extracting a subsequence), we obtain that

$$\begin{aligned}
\lim_{n\to+\infty} \langle \lambda_n, \eta \rangle_{H^{1/2}_{00}(\Gamma_\cap)^* \times H^{1/2}_{00}(\Gamma_\cap)} &= \frac{1}{2} \int_{\Omega_1} \nabla q_1 \cdot \nabla \mathcal{E}_1\tilde{\eta}\, dx - \frac{1}{2} \int_{\Omega_2} \nabla q_2 \cdot \nabla \mathcal{E}_2\tilde{\eta}\, dx \\
&= \left\langle \frac{1}{2} \left(\partial_{\mathbf{n}_1} q_1 - \partial_{\mathbf{n}_2} q_2 \right), \eta \right\rangle_{H^{1/2}_{00}(\Gamma_\cap)^* \times H^{1/2}_{00}(\Gamma_\cap)}.
\end{aligned}$$

We have then proved that $(\lambda_n)_n \subset H^{1/2}_{00}(\Gamma_\cap)^*$ has a subsequence converging toward $\lambda = \frac{1}{2}(\partial_{\mathbf{n}_1} q_1 - \partial_{\mathbf{n}_2} q_2)$. We emphasize that each q_i is unique since it is defined as the L^2-limit of $q_{i,n}$ and then the limit of the subsequence of (λ_n) is also unique. Urysohn's subsequence principle finally proves that the whole sequence λ_n converges toward λ. Since $M\lambda = (q_1, q_2, q_1|_{\Gamma_\cap} - q_2|_{\Gamma_\cap})^t = (q_1, q_2, \varphi)^t$ this proves that the range of M is closed. ∎

Applying now [10, Théorème II.20], we obtain Theorem 3.2.

Appendix B. Surjectivity of operator M^* when discretized with finite element.

Using a standard Galerkin method to discretize the continuous optimal control problem (3.2) following the notations in [19], we can go left the black arrow in the middle of Figure 1.1 and obtain the discrete optimal control problem

$$
\text{(B.1)} \qquad \begin{aligned} &\min \sum_{i=1}^{2} \frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{y}_{\text{target}_i})^\intercal N_i (\boldsymbol{y}_i - \boldsymbol{y}_{\text{target}_i}) + \frac{\alpha}{2}\boldsymbol{u}_i^\intercal N_i \boldsymbol{u}_i \\ &\text{s.t.} \begin{cases} K_i \boldsymbol{y}_i = \boldsymbol{F}_i + N_i \boldsymbol{u}_i + (-1)^{i+1} B_i^\intercal \boldsymbol{g}, \ i = 1, 2, \\ B_1 \boldsymbol{y}_1 - B_2 \boldsymbol{y}_2 = 0, \end{cases} \end{aligned}
$$

where $\boldsymbol{y}_i \in \mathbb{R}^{n_i^s + n^I}$ is the vector of degrees of freedom of the finite element approximation of y_i, n_i^s is the number of interior nodes in Ω_i and n^I the number of interface nodes on $\Gamma_\cap$, and B_i is of the form $B_i = [0_i, I_i]$, $i = 1, 2$, where 0_i is an $n^I \times n_i^s$ zero matrix, and I_i is the $n^I \times n^I$ identity matrix. K_i is the stiffness matrix and N_i the mass matrix, that are both invertible.

Solving this discrete optimal control problem (B.1) with the same new DD algorithm based on the augmented Lagrangian approach leads to the decoupled augmented discrete Lagrangian

(B.2)

$$
\begin{aligned} \min \sum_{i=1}^{2} &\frac{1}{2}(\boldsymbol{y}_i(\boldsymbol{u}_i, \boldsymbol{g}) - \boldsymbol{y}_{\text{target}_i})^\intercal N_i (\boldsymbol{y}_i(\boldsymbol{u}_i, \boldsymbol{g}) - \boldsymbol{y}_{\text{target}_i}) + \frac{\alpha}{2}\boldsymbol{u}_i^\intercal N_i \boldsymbol{u}_i \\ &+ (\boldsymbol{\lambda}^k)^\intercal (B_1 \boldsymbol{y}_1(\boldsymbol{u}_1, \boldsymbol{g}) - B_2 \boldsymbol{y}_2(\boldsymbol{u}_2, \boldsymbol{g})) + \frac{\rho^k}{2} \|B_1 \boldsymbol{y}_1(\boldsymbol{u}_1, \boldsymbol{g}) - B_2 \boldsymbol{y}_2(\boldsymbol{u}_2, \boldsymbol{g})\|_2^2, \end{aligned}
$$

where $\boldsymbol{y}_i(\boldsymbol{u}_i, \boldsymbol{g})$ satisfy

$$
K_i \boldsymbol{y}_i = \boldsymbol{F}_i + N_i \boldsymbol{u}_i + (-1)^{i+1} B_i^\intercal \boldsymbol{g}, \ i = 1, 2.
$$

As in the continuous framework, the convergence of our new DD method given by Algorithm 2.1 depends on the surjectivity of the derivative of $M_h^* : (\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{g}) \mapsto B_1 \boldsymbol{y}_1(\boldsymbol{u}_1, \boldsymbol{g}) - B_2 \boldsymbol{y}_2(\boldsymbol{u}_2, \boldsymbol{g})$ where $\boldsymbol{y}_i(\boldsymbol{u}_i, \boldsymbol{g})$ verify $K_i \boldsymbol{y}_i = \boldsymbol{F}_i + N_i \boldsymbol{u}_i + (-1)^{i+1} B_i^\intercal \boldsymbol{g}$, $i = 1, 2$ (see (3.12) and A for this property at the continuous level).

Proposition B.1. *The derivative of M_h^* w.r.t. $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{g})$ is onto.*

Proof. Note first that

$$
\begin{aligned} B_1 \boldsymbol{y}_1(\boldsymbol{u}_1, \boldsymbol{g}) - B_2 \boldsymbol{y}_2(\boldsymbol{u}_2, \boldsymbol{g}) = &\begin{pmatrix} B_1 K_1^{-1} N_1 & -B_2 K_2^{-1} N_2 & B_1 K_1^{-1} B_1^\intercal - B_2 K_2^{-1} B_2^\intercal \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \boldsymbol{g} \end{pmatrix} \\ &+ B_1 K_1^{-1} \boldsymbol{F}_1 - B_2 K_2^{-1} \boldsymbol{F}_2. \end{aligned}
$$

Therefore, the derivative of M_h^* is simply given by

$$
\partial M_h^* = \begin{pmatrix} B_1 K_1^{-1} N_1 & -B_2 K_2^{-1} N_2 & B_1 K_1^{-1} B_1^\intercal - B_2 K_2^{-1} B_2^\intercal \end{pmatrix}.
$$

In order to prove that ∂M_h^* is surjective, we will prove that (∂M_h) is injective. Let $\boldsymbol{\lambda} \in \mathbb{R}^{n^I}$. Suppose $(\partial M_h)\boldsymbol{\lambda} = 0$. This implies that

$$
\begin{aligned} K_1^{-\intercal} B_1^\intercal N_1 \boldsymbol{\lambda} &= 0 \\ -K_2^{-\intercal} B_2^\intercal N_2 \boldsymbol{\lambda} &= 0 \\ (B_1 K_1^{-\intercal} B_1^\intercal - B_2 K_2^{-\intercal} B_2^\intercal)\boldsymbol{\lambda} &= 0 \end{aligned} \iff \begin{aligned} B_1^\intercal N_1 \boldsymbol{\lambda} &= 0 \\ B_2^\intercal N_2 \boldsymbol{\lambda} &= 0 \end{aligned} \iff \boldsymbol{\lambda} = 0.
$$

∎

The proof of the convergence of our new discrete DD algorithm to the solutions of (B.1) can then be done as in the Section 3, or can be found in [6]. Note also that in this discrete context, the extension to a striped decomposition with more subdomains becomes easier, since only the injectivity of M needs to be proved .

REFERENCES

[1] Hedy Attouch and Mohamed Soueycatt. Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. applications to games, PDE's and control. *Pacific Journal of Optimization*, 5(1):17–37, 2008.

[2] Philip Avery and Charbel Farhat. The FETI family of domain decomposition methods for inequality-constrained quadratic programming: Application to contact problems with conforming and nonconforming interfaces. *Computer Methods in Applied Mechanics and Engineering*, 198(21-26):1673–1683, 2009.

[3] Maïtine Bergounioux. Augmented Lagrangian method for distributed optimal control problems with state constraints. *Journal of Optimization Theory and Applications*, 78(3):493–521, 1993.

[4] Maïtine Bergounioux and Mounir Haddou. A SQP-augmented Lagrangian method for optimal control of semilinear elliptic variational inequalities. In *Control and Estimation of Distributed Parameter Systems*, pages 57–72. Springer, 2003.

[5] Maïtine Bergounioux and Karl Kunisch. Augemented Lagrangian techniques for elliptic state constrained optimal control problems. *SIAM Journal on Control and Optimization*, 35(5):1524–1543, 1997.

[6] Dimitri P Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic press, 1982.

[7] J. Frédéric Bonnans and Julien Laurent-Varin. Computation of order conditions for symplectic partitioned Runge-Kutta schemes with application to optimal control. *Numerische Mathematik*, 103:1–10, 2006.

[8] Aïcha Bounaim. *Méthodes de décomposition de domaine: Application à la résolution de problèmes de contrôle optimal*. PhD thesis, Université Joseph-Fourier-Grenoble I, 1999.

[9] Jean-François Bourgat, Roland Glowinski, Patrick Le Tallec, and Marina Vidrascu. Variational formulation and algorithm for trace operator in domain decomposition calculations. In Tony Chan, Roland Glowinski, Jacques Périaux, and Olof Widlund, editors, *Domain Decomposition Methods*, pages 3–16, Philadelphia, PA, 1989. SIAM.

[10] Haïm Brézis. *Analyse fonctionnelle: théorie et applications*. Collection Mathématiques appliquées pour la maîtrise. Masson, 1983.

[11] Eduardo Casas. Control of an elliptic problem with pointwise state constraints. *SIAM Journal on Control and Optimization*, 24(6):1309–1318, 1986.

[12] Bastien Chaudet-Dumas and Jean Deteix. Shape derivatives for an augmented Lagrangian formulation of elastic contact problems. *ESAIM: Control, Optimisation and Calculus of Variations*, 27:S14, 2021.

[13] Zhiming Chen and Jun Zou. An augmented Lagrangian method for identifying discontinuous parameters in elliptic systems. *SIAM Journal on Control and Optimization*, 37(3):892–910, 1999.

[14] Monique Chyba, Ernst Hairer, and Gilles Vilmart. The role of symplectic integrators in optimal control. *Optimal control applications and methods*, 30(4):367–382, 2009.

[15] Gabriele Ciaramella and Martin J. Gander. *Iterative methods and preconditioners for systems of linear equations*. SIAM, 2022.

[16] Gabriele Ciaramella, Laurence Halpern, and Luca Mechelli. Convergence analysis and optimization of a Robin Schwarz waveform relaxation method for periodic parabolic optimal control problems. Technical report, MOX Lab, Politecnico di Milano, 2022.

[17] Gabriele Ciaramella, Felix Kwok, and Georg Müller. Nonlinear optimized Schwarz preconditioner for elliptic optimal control problems. In *Domain Decomposition Methods in Science and Engineering XXVI*. Springer, 2020.

[18] Marco Discacciati, Paola Gervasio, and Alfio Quarteroni. The interface control domain decomposition (ICDD) method for elliptic problems. *SIAM Journal on Control and Optimization*, 51(5):3434–3458, 2013.

[19] Charbel Farhat and Francois-Xavier Roux. A method of finite element tearing and interconnecting and its parallel solution algorithm. *International journal for numerical methods in engineering*, 32(6):1205–1227, 1991.

[20] Martin J. Gander. Optimized Schwarz methods. *SIAM Journal on Numerical Analysis*,

44(2):699–731, 2006.

[21] Martin J. Gander and Felix Kwok. Schwarz methods for the time-parallel solution of parabolic control problems. In *Domain decomposition methods in science and engineering XXII*, pages 207–216. Springer, 2016.

[22] Martin J. Gander, Felix Kwok, and Bankim C. Mandal. Convergence of substructuring methods for elliptic optimal control problems. In *Domain Decomposition Methods in Science and Engineering XXIV 24*, pages 291–300. Springer, 2018.

[23] Martin J. Gander, Felix Kwok, and Gerhard Wanner. Constrained optimization: From Lagrangian mechanics to optimal control and PDE constraints. *Optimization with PDE Constraints: ESF Networking Program'OPTPDE'*, pages 151–202, 2014.

[24] Martin J. Gander and Liu-Di Lu. Dirichlet-Neumann and Neumann-Neumann methods for elliptic control problems. In *Domain decomposition methods in science and engineering XXVII*, pages 207–216. Springer, 2023.

[25] Martin J. Gander and Liu-Di Lu. Dirichlet-Neumann and Neumann-Neumann methods for parabolic control problems. *in preparation*, 2023.

[26] Martin J. Gander and Hui Zhang. Schwarz methods by domain truncation. *Acta Numerica*, 31:1–134, 2022.

[27] Wei Gong, Felix Kwok, and Zhiyu Tan. Convergence analysis of the Schwarz alternating method for unconstrained elliptic optimal control problems, 2022.

[28] Max Gunzburger and Jeehyun Lee. A domain decomposition method for optimization problems for partial differential equations. *Computers & Mathematics with Applications*, 40(2-3):177–192, 2000.

[29] William W. Hager. Runge- Kutta methods in optimal control and the transformed adjoint system. *Numerische Mathematik*, 87:247–282, 2000.

[30] Ernst Harier, Christian Lubich, and Gerhard Wanner. *Geometric numerical integration.* Springer, 2002.

[31] Magnus R Hestenes. Multiplier and gradient methods. *Journal of optimization theory and applications*, 4(5):303–320, 1969.

[32] Michael Hintermüller and Karl Kunisch. Feasible and noninterior path-following in constrained minimization with low multiplier regularity. *SIAM Journal on Control and Optimization*, 45(4):1198–1221, 2006.

[33] Michael Hinze, René Pinnau, Michael Ulbrich, and Stefan Ulbrich. *Optimization with PDE constraints*, volume 23. Springer Science & Business Media, 2008.

[34] L.S. Hou and Jangwoon Lee. A Robin-Robin non-overlapping domain decomposition method for an elliptic boundary control problem. *International Journal of Numerical Analysis & Modeling*, 8(3), 2011.

[35] Kazufumi Ito and Karl Kunisch. The augmented Lagrangian method for equality and inequality constraints in Hilbert spaces. *Mathematical programming*, 46(1):341–360, 1990.

[36] Kazufumi Ito and Karl Kunisch. The augmented Lagrangian method for parameter estimation in elliptic systems. *SIAM Journal on Control and Optimization*, 28(1):113–136, 1990.

[37] Kazufumi Ito and Karl Kunisch. Augmented Lagrangian–SQP methods for nonlinear optimal control problems of tracking type. *SIAM journal on control and optimization*, 34(3):874–891, 1996.

[38] Alfredo Iusem and Rolando Gárciga Otero. Inexact version of proximal point and augmented Lagrangian algorithms in Banach spaces. *Numerical Functional Analysis and Optimization*, 22(5-6):609–640, 2001.

[39] Christian Kanzow, Daniel Steck, and Daniel Wachsmuth. An augmented Lagrangian method for optimization problems in Banach spaces. *SIAM Journal on Control and Optimization*, 56(1):272–291, 2018.

[40] Phan Quoc Khanh, Tran Hue Nuong, and Michel Théra. On duality in nonconvex vector optimization in Banach spaces using augmented Lagrangians. *Positivity*, 3(1):49–64, 1999.

[41] Karl Kunisch and Xue-Cheng Tai. Nonoverlapping domain decomposition methods for inverse problems. In *Proceedings of 9th International Conference on Domain Decompostion Methods, Editors P. Bjorstard, M. Espedal and D. Keyes, John Wiley and Sons*. Citeseer, 1997.

[42] John E. Lagnese and Günter Leugering. *Domain in Decomposition Methods in Optimal Control of Partial Differential Equations*, volume 148. Springer Science & Business Media, 2004.

[43] Günter Leugering. Domain decomposition of optimal control problems for dynamic networks of elastic strings. *Computational Optimization and Applications*, 16:5–27, 2000.

[44] Günter Leugering. Nonoverlapping domain decomposition for instantaneous optimal control of friction dominated flow in a gas-network. Technical report, 2020. preprint.

[45] Jacques Louis Lions and Enrico Magenes. *Non-homogeneous boundary value problems and*

applications: Vol. 1, volume 181. Springer Science & Business Media, 2012.

[46] Frédéric Magoulès and François-Xavier Roux. Lagrangian formulation of domain decomposition methods: A unified theory. *Applied Mathematical Modelling*, 30(7):593–615, 2006. Parallel and Vector Processing in Science and Engineering.

[47] Jan Hendrik Maruhn. An augmented Lagrangian algorithm for optimization with equality constraints in Hilbert spaces. Master's thesis, Virginia Tech, 2001.

[48] Émile Picard. Sur l'application des méthodes d'approximations successives à l'étude de certaines équations différentielles ordinaires. *Journal de mathématiques pures et appliquées*, 9:217–271, 1893.

[49] Michael JD Powell. A method for nonlinear constraints in minimization problems. *Optimization*, pages 283–298, 1969.

[50] Anil V Rao. A survey of numerical methods for optimal control. *Advances in the Astronautical Sciences*, 135(1):497–528, 2009.

[51] Alexandre Vieira and Pierre-Henri Cocquet. Optimized schwarz method for coupled direct-adjoint problems applied to parameter identification in advection-diffusion equation. In *Domain Decomposition Methods in Science and Engineering XXVII*, pages 417–425, Cham, 2024. Springer Nature Switzerland.

[52] Eberhard Zeidler. *Nonlinear functional analysis and its applications: II/B: nonlinear monotone operators*. Springer Science & Business Media, 2013.

#subd.	h	α	Err. on y	Err. on u	Err. on g	Cost	# it.
2	1/64	1.0e+01	1.03e-09	1.81e-12	4.02e-08	4.18e-21	6
		1.0e+00	1.07e-09	1.87e-11	4.02e-08	4.49e-21	6
		1.0e-02	5.78e-10	8.58e-10	4.02e-08	1.14e-21	6
		1.0e-04	1.50e-09	4.69e-08	4.03e-08	2.06e-21	5
		1.0e-06	5.30e-09	1.62e-06	4.14e-08	2.67e-21	4
	1/256	1.0e+01	1.11e-09	1.97e-12	4.92e-11	4.91e-21	6
		1.0e+00	1.12e-09	1.98e-11	4.93e-11	4.98e-21	6
		1.0e-02	7.35e-10	1.09e-09	4.65e-11	1.83e-21	6
		1.0e-04	1.43e-09	4.45e-08	7.58e-11	1.84e-21	5
		1.0e-06	4.68e-09	1.48e-06	6.34e-10	2.64e-21	4
4	1/64	1.0e+01	2.64e-10	5.06e-13	4.01e-08	2.06e-22	6
		1.0e+00	2.11e-10	3.85e-12	4.01e-08	1.45e-22	6
		1.0e-02	2.46e-10	2.88e-10	4.01e-08	2.00e-22	6
		1.0e-04	1.42e-09	4.17e-08	4.03e-08	4.59e-21	5
		1.0e-06	6.33e-09	1.72e-06	4.22e-08	9.48e-21	4
	1/256	1.0e+01	3.12e-10	6.09e-13	4.20e-11	3.07e-22	6
		1.0e+00	3.59e-10	6.68e-12	4.24e-11	4.29e-22	6
		1.0e-02	2.14e-10	2.50e-10	4.13e-11	1.97e-22	6
		1.0e-04	1.53e-09	4.47e-08	7.43e-11	4.68e-21	5
		1.0e-06	4.71e-09	1.49e-06	6.58e-10	7.82e-21	4
8	1/64	1.0e+01	1.66e-09	6.25e-12	4.02e-08	3.00e-20	6
		1.0e+00	1.83e-09	6.10e-11	4.02e-08	3.36e-20	6
		1.0e-02	7.39e-10	2.23e-09	4.02e-08	4.28e-21	6
		1.0e-04	7.98e-10	2.36e-08	4.02e-08	1.01e-21	5
		1.0e-06	6.35e-09	2.01e-06	4.22e-08	2.23e-20	4
	1/256	1.0e+01	8.14e-10	3.05e-12	4.59e-11	7.94e-21	6
		1.0e+00	1.31e-09	4.43e-11	4.96e-11	1.78e-20	6
		1.0e-02	3.64e-10	1.18e-09	4.26e-11	1.31e-21	6
		1.0e-04	1.01e-09	3.25e-08	4.28e-11	1.61e-21	5
		1.0e-06	5.09e-09	1.65e-06	6.69e-10	2.01e-20	4
16	1/64	1.0e+01	1.41e-09	5.91e-12	4.02e-08	4.37e-20	7
		1.0e+00	1.05e-09	4.70e-11	4.02e-08	2.57e-20	7
		1.0e-02	2.11e-10	6.16e-10	4.01e-08	4.69e-22	7
		1.0e-04	4.50e-10	2.40e-08	4.02e-08	5.22e-22	5
		1.0e-06	5.56e-09	1.55e-06	4.14e-08	1.43e-20	4
	1/256	1.0e+01	1.25e-09	5.35e-12	5.00e-11	3.53e-20	7
		1.0e+00	1.54e-09	6.75e-11	5.30e-11	5.62e-20	7
		1.0e-02	2.78e-10	1.17e-09	4.24e-11	1.70e-21	7
		1.0e-04	5.00e-10	2.56e-08	4.21e-11	6.82e-22	5
		1.0e-06	4.59e-09	1.37e-06	5.92e-10	1.26e-20	4

Table 4.1: Errors on the optimal solution for the second example using different values of α and number of subdomains. The parameters were fixed as $l = 2$, $C = 1$.