NEW TIME DOMAIN DECOMPOSITION METHODS FOR PARABOLIC OPTIMAL CONTROL PROBLEMS I: DIRICHLET-NEUMANN AND NEUMANN-DIRICHLET ALGORITHMS*

MARTIN J. GANDER† AND LIU-DI LU†

Abstract. We present new Dirichlet-Neumann and Neumann-Dirichlet algorithms with a time domain decomposition applied to unconstrained parabolic optimal control problems. After a spatial semi-discretization, we use the Lagrange multiplier approach to derive a coupled forward-backward optimality system, which can then be solved using a time domain decomposition. Due to the forward-backward structure of the optimality system, three variants can be found for the Dirichlet-Neumann and Neumann-Dirichlet algorithms. We analyze their convergence behavior and determine the optimal relaxation parameter for each algorithm. Our analysis reveals that the most natural algorithms are actually only good smoothers, and there are better choices which lead to efficient solvers. We illustrate our analysis with numerical experiments.

Key words. Time domain decomposition, Dirichlet-Neumann algorithm, Neumann-Dirichlet algorithm, Parallel in Time, Parabolic optimal control problems, Convergence analysis.

MSC codes. 65M12,65M55,65Y05,

1. Introduction.

PDE-constrained optimal control problems arise in various areas, often containing multiphysics or multiscale phenomena, and also high frequency components on different time scales. This requires very fine spatial and temporal discretizations, resulting in very large problems, for which efficient parallel solvers are needed; we refer to [14, 26] for a brief review. We present and analyze a new class of time domain decomposition methods based on Dirichlet-Neumann and Neumann-Dirichlet techniques. We consider as our model a parabolic optimal control problem: for a given target function $\hat{y} \in L^2(Q)$, $\gamma \geq 0$ and $\nu > 0$, we want to minimize the cost functional

$$J(y,u) := \frac{1}{2}\|y-\hat{y}\|^2_{L^2(Q)} + \frac{\gamma}{2}\|y(T)-\hat{y}(T)\|^2_{L^2(\Omega)} + \frac{\nu}{2}\|u\|^2_{U_{\mathrm{ad}}}, \tag{1.1}$$

subject to the linear parabolic state equation

$$\begin{aligned} \partial_t y - \Delta y &= u && \text{in } Q := \Omega \times (0,T), \\ y &= 0 && \text{on } \Sigma := \partial\Omega \times (0,T), \\ y(0) &= y_0 && \text{on } \Sigma_0 := \Omega \times \{0\}, \end{aligned} \tag{1.2}$$

where $\Omega \subset \mathbb{R}^d$, $d = 1,2,3$ is a bounded domain with boundary $\partial\Omega$, and T is the fixed final time. The control u on the right-hand side of the PDE is in an admissible set U_{ad}, and we want to control the solution of the parabolic PDE (1.2) towards a target state $\hat{y}$. For simplicity, we consider here homogeneous boundary conditions.

The parabolic optimal control problem (1.1)-(1.2) has a unique solution for the classical choice $u \in L^2(Q)$, which can be characterized by a forward-backward optimality system, see e.g. [4, 18, 26]. More recently, also energy regularization has been considered, see [23] for elliptic and [16] for parabolic cases. This is motivated by

*Submitted to the editors 2023.07.05.

Funding: This work was funded by the Swiss National Science Foundation Grant 192064.

†Section de Mathématiques, Université de Genève, rue du Conseil-Général 5-7, CP 64, 1205, Geneva, Switzerland (martin.gander@unige.ch, liudi.lu@unige.ch).

the fact that the state $y \in L^2(0,T;H_0^1(\Omega))$ is well-defined as the solution of the heat equation (1.2) for the control $z \in L^2(0,T;H^{-1}(\Omega))$, and thus offers an interesting alternative.

We are interested in applying Time Domain Decomposition methods (DDMs) to the forward-backward optimality system. DDMs were developed for elliptic PDEs and are very efficient in parallel computing environments, see e.g. [7, 25]. DDMs were extended to time-dependent problems using waveform relaxation techniques from [17], with a spatial decomposition and solving the problem on small space-time cylinders [12]. The extension of DDMs to elliptic optimal control problems is quite natural, see [1, 2, 5, 9], but less is known about DDMs applied to parabolic optimal control problems.

The role of the time variable in forward-backward optimality systems is key, and it is natural to seek efficient solvers through time domain decomposition. For classical evolution problems, the idea of time domain decomposition goes back to [24]. Parallel Runge Kutta methods were introduced in [22] with good small scale time parallelism. In [20, 27], the authors propose to combine multigrid methods with waveform relaxation. Parareal [19] uses a different approach, namely multiple shooting with an approximate Jacobian on a coarse grid, and Parareal techniques led to a new ParaOpt algorithm [10] for optimal control, see also [13]. In [8, 15], Schwarz methods are used to decompose the time domain for optimal control. Waveform relaxation techniques can also be applied to address such optimal control problems, for instance, using Dirichlet-Neumann waveform relaxation methods [21] and Optimized Schwarz waveform relaxation methods [6]. Note that the decomposition in [6, 21] is in space of the PDE constraint, in contrast to the approach presented in [8, 15], and also in contrast to our approach in time here.

We develop and analyze here new time domain decomposition algorithms to solve the PDE-constrained problem (1.1)-(1.2) using Dirichlet-Neumann and Neumann-Dirichlet techniques that go back to [3] for space parallelism. We introduce in Section 2 the optimality system and its semi-discretization. In Section 3 we present our new time parallel Dirichlet-Neumann and Neumann-Dirichlet algorithms and study their convergence. Numerical experiments are shown in Section 4, and we draw conclusions in Section 5.

2. Optimality system and its semi-discretization. The PDE-constrained optimization problem (1.1)-(1.2) can be solved using Lagrange multipliers [26, Chapter 3], see also [11] for a historical context. To obtain the associated optimality system, we introduce the Lagrangian function $\mathcal{L}$ associated with Problem (1.1)-(1.2),

$$\begin{aligned}\mathcal{L}(y,u,\lambda) =& J(y,u) + \langle \partial_t y - \Delta y - u, \lambda\rangle \\ =& \int_0^T \Big(\langle \partial_t y, \lambda\rangle_{V',V} + \int_\Omega \big(\frac{1}{2}|y-\hat{y}|^2 + \frac{\nu}{2}|u|^2 + \nabla y\cdot\nabla\lambda - u\lambda\big)\,\mathrm{d}\mathbf{x}\Big)\,\mathrm{d}t \\ &+ \frac{\gamma}{2}\int_\Omega |y(T)-\hat{y}(T)|^2\,\mathrm{d}\mathbf{x},\end{aligned}$$

with $y \in W(0,T) := L^2(0,T;V)\cap H^1(0,T;V')$, $u \in L^2(Q)$, $V := H_0^1(\Omega)$ and $V' := H^{-1}(\Omega)$ the dual space of V. Here $\lambda \in L^2(0,T;V)$ denotes the adjoint state (also called the Lagrange multiplier). Taking the derivative of $\mathcal{L}$ with respect to λ and equating this to zero, we find for all test functions $\chi \in L^2(0,T;V)$,

$$0 = \langle \partial_\lambda \mathcal{L}(y,u,\lambda), \chi\rangle = \int_0^T \big(\langle \partial_t y, \chi\rangle_{V',V} + \int_\Omega \big(\nabla y\cdot\nabla\chi - u\chi\big)\,\mathrm{d}\mathbf{x}\big)\,\mathrm{d}t,$$

which implies that $y \in V$ is the weak solution of the state equation (1.2) (also called the primal problem). Taking the derivative of $\mathcal{L}$ with respect to y and equating this to zero, and obtain for all $\chi \in W(0,T)$

$$\begin{aligned} 0 = \langle \partial_y \mathcal{L}(y,u,\lambda), \chi\rangle = \int_0^T & \Big(\langle \partial_t \chi, \lambda\rangle_{V',V} + \int_\Omega \big((y-\hat{y})\chi + \nabla\chi\cdot\nabla\lambda\big)\,\mathrm{d}\mathbf{x}\Big)\,\mathrm{d}t \\ =& \langle \chi(T), \lambda(T) + \gamma(y(T)-\hat{y}(T))\rangle_{L^2(\Omega)} - \langle \chi(0), \lambda(0)\rangle_{L^2(\Omega)} \\ &+ \int_0^T \langle -\partial_t\lambda - \Delta\lambda + (y-\hat{y}), \chi\rangle_{V',V}\,\mathrm{d}t, \end{aligned}$$

where we used integration by parts with respect to t in $\partial_t\chi$ and with respect to $\mathbf{x}$ in $\nabla\chi$. By choosing $\chi \in C_0^\infty(Q)$ and applying an argument of density, we find that the last integral is zero. Choosing then $\chi \in W(0,T)$ such that $\chi(0) = 0$, we obtain the adjoint equation (also called the dual problem)

$$\begin{aligned} \partial_t\lambda + \Delta\lambda &= y - \hat{y} && \text{in } Q, \\ \lambda &= 0 && \text{on } \Sigma, \\ \lambda(T) &= -\gamma(y(T)-\hat{y}(T)) && \text{on } \Sigma_T := \Omega\times\{T\}. \end{aligned} \tag{2.1}$$

Finally, taking the derivative of $\mathcal{L}$ with respect to u and equating this to zero, we obtain for all test functions $\chi \in L^2(Q)$, $0 = \langle \partial_u(y,u,p), \chi\rangle = \int_0^T\int_\Omega (\nu u - \lambda)\chi\,\mathrm{d}\mathbf{x}\,\mathrm{d}t$, which gives the optimality condition

$$\lambda = \nu u \quad \text{in } Q. \tag{2.2}$$

If a control u is optimal with the associated state y of the optimization problem (1.1)-(1.2), then the first-order optimality system (1.2), (2.1) and (2.2) must be satisfied. This is a forward-backward system, i.e., the primal problem is solved forward in time with an initial condition while the dual problem is solved backward in time with a final condition, and our new time decomposition algorithms solve this system. Since the time variable plays a special role, we consider a semi-discretization in space, and replace the spatial operator $-\Delta$ in the primal problem (1.2) by a matrix $A \in \mathbb{R}^{n\times n}$, for instance using a Finite Difference discretization in space. We then obtain as above the semi-discrete optimality system (dot denoting the time derivative)

$$\begin{cases} \dot{\boldsymbol{y}} + A\boldsymbol{y} = \boldsymbol{u} & \text{in } (0,T), \\ \boldsymbol{y}(0) = \boldsymbol{y}_0, \end{cases} \qquad \begin{cases} \dot{\boldsymbol{\lambda}} - A^T\boldsymbol{\lambda} = \boldsymbol{y} - \hat{\boldsymbol{y}} & \text{in } (0,T), \\ \boldsymbol{\lambda}(T) = -\gamma(\boldsymbol{y}(T) - \hat{\boldsymbol{y}}(T)), \end{cases}$$

where $\boldsymbol{\lambda}(t) = \nu\boldsymbol{u}(t)$ for all $t \in \Omega$. Eliminating $\boldsymbol{u}$, we obtain in matrix form

$$\begin{cases} \begin{pmatrix} \dot{\boldsymbol{y}} \\ \dot{\boldsymbol{\lambda}} \end{pmatrix} + \begin{pmatrix} A & -\nu^{-1}I \\ -I & -A^T \end{pmatrix} \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} 0 \\ -\hat{\boldsymbol{y}} \end{pmatrix} \text{ in } (0,T), \\ \boldsymbol{y}(0) = \boldsymbol{y}_0, \\ \boldsymbol{\lambda}(T) + \gamma\boldsymbol{y}(T) = \gamma\hat{\boldsymbol{y}}(T), \end{cases} \tag{2.3}$$

where I is the identity. If A is symmetric, $A = A^T$, which is natural for discretizations of $-\Delta$, then it can be diagonalized, $A = PDP^{-1}$, $D := \mathrm{diag}(d_1,\dots,d_n)$ with d_i the i-th eigenvalue of A. The system (2.3) can thus also be diagonalized

$$\begin{cases} \begin{pmatrix} \dot{\boldsymbol{z}} \\ \dot{\boldsymbol{\mu}} \end{pmatrix} + \begin{pmatrix} D & -\nu^{-1}I \\ -I & -D \end{pmatrix} \begin{pmatrix} \boldsymbol{z} \\ \boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} 0 \\ -\hat{\boldsymbol{z}} \end{pmatrix} \text{ in } (0,T), \\ \boldsymbol{z}(0) = \boldsymbol{z}_0, \\ \boldsymbol{\mu}(T) + \gamma\boldsymbol{z}(T) = \gamma\hat{\boldsymbol{z}}(T), \end{cases}$$

where $\boldsymbol{z} := P^{-1}\boldsymbol{y}$, $\mu := P^{-1}\boldsymbol{\lambda}$, $\hat{\boldsymbol{z}} := P^{-1}\hat{\boldsymbol{y}}$ and $\boldsymbol{z}_0 := P^{-1}\boldsymbol{y}_0$. This system then represents n independent 2×2 systems of ODEs of the form

$$(2.4)\qquad \begin{cases} \begin{pmatrix} \dot{z}_{(i)} \\ \dot{\mu}_{(i)} \end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1} \\ -1 & -d_i \end{pmatrix} \begin{pmatrix} z_{(i)} \\ \mu_{(i)} \end{pmatrix} = \begin{pmatrix} 0 \\ -\hat{z}_{(i)} \end{pmatrix} \text{ in } (0,T), \\ z_{(i)}(0) = z_{(i),0}, \\ \mu_{(i)}(T) + \gamma z_{(i)}(T) = \gamma \hat{z}_{(i)}(T), \end{cases}$$

where $z_{(i)}$, $\mu_{(i)}$, $\hat{z}_{(i)}$ are the i-th components of the vectors $\boldsymbol{z}$, $\boldsymbol{\mu}$, $\hat{\boldsymbol{z}}$. Isolating the variable in each equation in (2.4), we find the identities

$$(2.5)\qquad \mu_{(i)} = \nu(\dot{z}_{(i)} + d_i z_{(i)}), \qquad z_{(i)} = \dot{\mu}_{(i)} - d_i\mu_{(i)} + \hat{z}_{(i)}.$$

We use the identity of z to eliminate μ, and obtain a second-order ODE from (2.4),

$$(2.6)\qquad \begin{cases} \ddot{z}_{(i)} - (d_i^2 + \nu^{-1}) z_{(i)} = -\nu^{-1}\hat{z}_{(i)} \text{ in } (0,T), \\ z_{(i)}(0) = z_{(i),0}, \\ \dot{z}_{(i)}(T) + (\nu^{-1}\gamma + d_i) z_{(i)}(T) = \nu^{-1}\gamma\hat{z}_{(i)}(T). \end{cases}$$

Similarly, we can also eliminate z to get

$$(2.7)\qquad \begin{cases} \ddot{\mu}_{(i)} - (d_i^2 + \nu^{-1})\mu_{(i)} = -\dot{\hat{z}}_{(i)} - d_i\hat{z}_{(i)} \text{ in } (0,T), \\ \dot{\mu}_{(i)}(0) - d_i\mu_{(i)}(0) = z_{(i),0} - \hat{z}_{(i)}(0), \\ \gamma\dot{\mu}_{(i)}(T) + (1-\gamma d_i)\mu_{(i)}(T) = 0. \end{cases}$$

To simplify the notation in what follows, we define

$$(2.8)\qquad \sigma_i := \sqrt{d_i^2 + \nu^{-1}}, \quad \omega_i := \nu^{-1}\gamma + d_i, \quad \beta_i := 1 - \gamma d_i.$$

In our analysis for the error, $\hat{\boldsymbol{y}}$ will equal zero, which implies $\hat{\boldsymbol{z}} = 0$, and the solution of (2.6) and (2.7) is then

$$(2.9)\qquad z_{(i)}(t) \text{ or } \mu_{(i)}(t) = A_i\cosh(\sigma_i t) + B_i\sinh(\sigma_i t),$$

where A_i, B_i are two coefficients.

Remark 2.1. Our arguments above work for any diagonalizable matrix A, and thus our results will apply to more general parabolic optimal control problems than the heat equation. Note also that the diagonalization is only a theoretical tool for our convergence analysis, and not needed to run our new time domain decomposition algorithms.

3. Dirichlet-Neumann and Neumann-Dirichlet algorithms in time.

We now apply Dirichlet-Neumann (DN) and Neumann-Dirichlet (ND) techniques in time to obtain our new time domain decomposition algorithms to solve the system (2.4), and study their convergence. Focusing on the error equations, we set the initial condition $\boldsymbol{y}_0 = 0$ (i.e., $\boldsymbol{z}_0 = 0$) and the target functions $\hat{\boldsymbol{y}} = 0$ (i.e., $\hat{\boldsymbol{z}} = 0$). We decompose the time domain $\Omega := (0,T)$ into two non-overlapping time subdomains $\Omega_1 := (0,\alpha)$ and $\Omega_2 := (\alpha, T)$, where α is the interface. We denote by $z_{j,(i)}$ and $\mu_{j,(i)}$ the restriction to Ω_j, $j = 1,2$ of $z_{(i)}$ and $\mu_{(i)}$. Since system (2.4) is a forward-backward system, it appears natural at first sight to keep this property for the decomposed case, as illustrated in Figure 1: we expect to have a final condition for the adjoint state

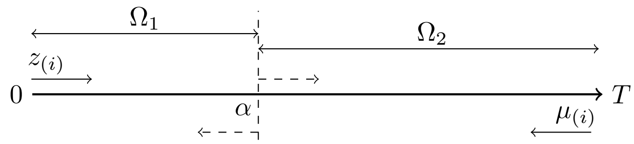

FIG. 1. *Illustration of the forward-backward system.*

$\mu_{(i)}$ in Ω_1 since we already have an initial condition for $z_{(i)}$; similarly, we expect to have an initial condition for the primal state $z_{(i)}$ in Ω_2 since we already have a final condition for $\mu_{(i)}$. Therefore, a natural DN algorithm in time solves for the iteration index $k = 1, 2, \ldots$

$$\begin{cases} \begin{pmatrix} \dot{z}^k_{1,(i)} \\ \dot{\mu}^k_{1,(i)} \end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1} \\ -1 & -d_i \end{pmatrix} \begin{pmatrix} z^k_{1,(i)} \\ \mu^k_{1,(i)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ in } \Omega_1, \\ z^k_{1,(i)}(0) = 0, \\ \mu^k_{1,(i)}(\alpha) = f^{k-1}_{\alpha,(i)}, \end{cases}$$
$$\begin{cases} \begin{pmatrix} \dot{z}^k_{2,(i)} \\ \dot{\mu}^k_{2,(i)} \end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1} \\ -1 & -d_i \end{pmatrix} \begin{pmatrix} z^k_{2,(i)} \\ \mu^k_{2,(i)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ in } \Omega_2, \\ \dot{z}^k_{2,(i)}(\alpha) = \dot{z}^k_{1,(i)}(\alpha), \\ \mu^k_{2,(i)}(T) + \gamma z^k_{2,(i)}(T) = 0, \end{cases} \tag{3.1}$$

and then the transmission condition is updated by

$$f^k_{\alpha,(i)} := (1-\theta) f^{k-1}_{\alpha,(i)} + \theta \mu^k_{2,(i)}(\alpha), \tag{3.2}$$

with a relaxation parameter $\theta \in (0, 1)$. However, there are many other ways to decouple in time using DN and ND techniques for problem (2.4): we can apply the technique to both states $(z_{(i)}, \mu_{(i)})$ as in (3.1), or we can apply it just to one of these two states in the reduced forms (2.6) and (2.7). And with the identities (2.5), we can transfer the Dirichlet and the Neumann transmission condition from one state to the other. We list in Table 1 all possible new time domain decomposition algorithms we can obtain, along with their equivalent representations in terms of other formulations. The algorithms can be classified into three main categories, and each category is composed of two blocks, the first block represents a DN technique applied to (2.4), whereas the second block represents a ND technique. Each block contains three rows: the first row is the algorithm applied to formulation (2.4), the second row the algorithm applied to formulation (2.6) and the third row the algorithm applied to formulation (2.7).

Remark 3.1. In Table 1, the transmission conditions $\ddot{z}_{(i)} + d_i \dot{z}_{(i)}$ and $\ddot{\mu}_{(i)} - d_i \dot{\mu}_{(i)}$ are in fact Robin type conditions, since, using the identity (2.5) of $z_{(i)}$ and $\mu_{(i)}$, we find $\dot{z}_{(i)} = \ddot{\mu}_{(i)} - d_i \dot{\mu}_{(i)}$ and $\dot{\mu}_{(i)} = \ddot{z}_{(i)} + d_i \dot{z}_{(i)}$. On the other hand, from the first equation of (2.6) and of (2.7), we have $\ddot{z}_{(i)} - \sigma_i^2 z_{(i)} = 0$ and $\ddot{\mu}_{(i)} - \sigma_i^2 \mu_{(i)} = 0$. Substituting $\ddot{z}_{(i)}$ and $\ddot{\mu}_{(i)}$ gives $\dot{\mu}_{(i)} = \ddot{z}_{(i)} + d_i \dot{z}_{(i)} = d_i \dot{z}_{(i)} + \sigma_i^2 z_{(i)}$ and $\dot{z}_{(i)} = \ddot{\mu}_{(i)} - d_i \dot{\mu}_{(i)} = \sigma_i^2 \mu_{(i)} - d_i \dot{\mu}_{(i)}$. Thus the transmission conditions containing a second derivative in Table 1 are indeed Robin type conditions. We decided to keep the notations $\ddot{z}_{(i)}$ and $\ddot{\mu}_{(i)}$ in Table 1 to show the direct link between the two states $z_{(i)}$ and $\mu_{(i)}$.

TABLE 1
Combinations of the DN and ND algorithms. The letter R stands for a Robin type condition.

	Problem	Ω_1	Ω_2	algorithm type
Category I: $(z_{(i)}, \mu_{(i)})$	(2.4)	$\mu_{(i)}$	$\dot{z}_{(i)}$	(DN)
	(2.6)	$\dot{z}_{(i)} + d_i z_{(i)}$	$\dot{z}_{(i)}$	(RN)
	(2.7)	$\mu_{(i)}$	$\ddot{\mu}_{(i)} - d_i \dot{\mu}_{(i)}$	(DR)
	(2.4)	$\dot{\mu}_{(i)}$	$z_{(i)}$	(ND)
	(2.6)	$\ddot{z}_{(i)} + d_i \dot{z}_{(i)}$	$z_{(i)}$	(RD)
	(2.7)	$\dot{\mu}_{(i)}$	$\dot{\mu}_{(i)} - d_i \mu_{(i)}$	(NR)
Category II: $z_{(i)}$	(2.4)	$z_{(i)}$	$\dot{z}_{(i)}$	(DN)
	(2.6)	$z_{(i)}$	$\dot{z}_{(i)}$	(DN)
	(2.7)	$\dot{\mu}_{(i)} - d_i \mu_{(i)}$	$\ddot{\mu}_{(i)} - d_i \dot{\mu}_{(i)}$	(RR)
	(2.4)	$\dot{z}_{(i)}$	$z_{(i)}$	(ND)
	(2.6)	$\dot{z}_{(i)}$	$z_{(i)}$	(ND)
	(2.7)	$\ddot{\mu}_{(i)} - d_i \dot{\mu}_{(i)}$	$\dot{\mu}_{(i)} - d_i \mu_{(i)}$	(RR)
Category III: $\mu_{(i)}$	(2.4)	$\mu_{(i)}$	$\dot{\mu}_{(i)}$	(DN)
	(2.6)	$\dot{z}_{(i)} + d_i z_{(i)}$	$\ddot{z}_{(i)} + d_i \dot{z}_{(i)}$	(RR)
	(2.7)	$\mu_{(i)}$	$\dot{\mu}_{(i)}$	(DN)
	(2.4)	$\dot{\mu}_{(i)}$	$\mu_{(i)}$	(ND)
	(2.6)	$\ddot{z}_{(i)} + d_i \dot{z}_{(i)}$	$\dot{z}_{(i)} + d_i z_{(i)}$	(RR)
	(2.7)	$\dot{\mu}_{(i)}$	$\mu_{(i)}$	(ND)

However, there are other interpretations of some transmission conditions in certain circumstances. For instance, let us take the Neumann condition $\dot{z}_{(i)}$ in the second block of Category II for the problem (2.4), it can also be interpreted as a Robin condition $\sigma_i^2 \mu_{(i)} - d_i \dot{\mu}_{(i)}$ using the above argument. Then, this algorithm can also be read as a Robin-Dirichlet (RD) type algorithm instead of a Neumann-Dirichlet type. Moreover, this interpretation is particularly useful in this case, since it reveals the fact that the forward-backward property of the problem (2.4) is still kept by this algorithm. Otherwise, we can also use the identity of $\mu_{(i)}$ in (2.5) to transfer this Neumann condition $\dot{z}_{(i)}$ to $\mu_{(i)} - d_i z_{(i)}$. This is also useful from a numerical point of view, since we can transfer a Neumann condition to a Dirichlet type condition. This will be used in detail in the following analysis.

3.1. Category I.

We start with the algorithms in Category I, which run on the pair $(z_{(i)}, \mu_{(i)})$ to solve (2.4), and study the DN and then the ND variant.

3.1.1. Dirichlet-Neumann algorithm (DN$_1$).

This is (3.1), at first sight the most natural method that keeps the forward-backward structure as in the original problem (2.4). To analyze the convergence behavior, we can choose any of the problem formulations (2.6), (2.7), since they are equivalent to (2.4). Choosing (2.6), the algorithm DN$_1$ for $i = 1, \ldots, n$, and iteration $k = 1, 2, \ldots$ is given by

$$(3.3) \qquad \begin{cases} \ddot{z}_{1,(i)}^k - \sigma_i^2 z_{1,(i)}^k = 0 \text{ in } \Omega_1, \\ z_{1,(i)}^k(0) = 0, \\ \dot{z}_{1,(i)}^k(\alpha) + d_i z_{1,(i)}^k(\alpha) = f_{\alpha,(i)}^{k-1}, \end{cases} \qquad \begin{cases} \ddot{z}_{2,(i)}^k - \sigma_i^2 z_{2,(i)}^k = 0 \text{ in } \Omega_2, \\ \dot{z}_{2,(i)}^k(\alpha) = \dot{z}_{1,(i)}^k(\alpha), \\ \dot{z}_{2,(i)}^k(T) + \omega_i z_{2,(i)}^k(T) = 0, \end{cases}$$

and the update of the transmission condition defined in (3.2) becomes

$$(3.4) \qquad f_{\alpha,(i)}^k = (1-\theta) f_{\alpha,(i)}^{k-1} + \theta\big(\dot{z}_{2,(i)}^k(\alpha) + d_i z_{2,(i)}^k(\alpha)\big).$$

This is a Robin-Neumann type algorithm applied to solve the problem (2.6). Using the general solution (2.9), and the initial and final condition, we find

$$z_{1,(i)}^k(t) = A_i^k \sinh(\sigma_i t),\; z_{2,(i)}^k(t) = B_i^k\Big(\sigma_i \cosh\big(\sigma_i(T-t)\big)+\omega_i \sinh\big(\sigma_i(T-t)\big)\Big), \tag{3.5}$$

where A_i^k and B_i^k are determined by the transmission conditions at α in (3.3). Note that we will use (3.5) in the analysis for all algorithms, since only the transmission conditions will change. Inserting (3.5) at the interface α into (3.3) and solving for A_i^k, B_i^k gives $A_i^k = \frac{f_{\alpha,(i)}^{k-1}}{\sigma_i\cosh(a_i)+d_i\sinh(a_i)}$ and $B_i^k = \frac{-f_{\alpha,(i)}^{k-1}\cosh(a_i)}{(\sigma_i\cosh(a_i)+d_i\sinh(a_i))(\sigma_i\sinh(b_i)+\omega_i\cosh(b_i))}$, where we let $a_i := \sigma_i\alpha$ and $b_i := \sigma_i(T-\alpha)$ to simplify the notations, and $a_i+b_i = \sigma_i T$. Using the update of the transmission condition (3.4), we obtain $f_{\alpha,(i)}^k = (1-\theta)f_{\alpha,(i)}^{k-1} + \theta f_{\alpha,(i)}^{k-1}\nu^{-1}\frac{\sigma_i\gamma+\beta_i\tanh(b_i)}{(\sigma_i+d_i\tanh(a_i))(\omega_i+\sigma_i\tanh(b_i))}$, which leads to the following result.

THEOREM 3.2. *The algorithm DN*$_1$ (3.1)-(3.2) *converges if and only if*

$$\rho_{DN_1} := \max_{d_i\in\lambda(A)}\left|1-\theta\Big(1-\nu^{-1}\frac{\sigma_i\gamma+\beta_i\tanh(b_i)}{\big(\sigma_i+d_i\tanh(a_i)\big)\big(\omega_i+\sigma_i\tanh(b_i)\big)}\Big)\right| < 1, \tag{3.6}$$

where $\lambda(A)$ is the spectrum of the matrix A.

Remark 3.3. Instead of focusing on the state $z_{(i)}$ for the analysis, we could also have focused on the state $\mu_{(i)}$, which gives the same result, see Appendix A.

To get more insight in the convergence behavior, we consider a few special cases.

COROLLARY 3.4. *If the matrix A is not singular, then the algorithm DN*$_1$ (3.1)-(3.2) *for $\theta = 1$ converges for all initial guesses.*

Proof. Substituting $\theta = 1$ into (3.6), we have

$$\rho_{\mathrm{DN}_1}|_{\theta=1} = \nu^{-1}\max_{d_i\in\lambda(A)}\left|\frac{\sigma_i\gamma+\beta_i\tanh(b_i)}{\big(\sigma_i+d_i\tanh(a_i)\big)\big(\omega_i+\sigma_i\tanh(b_i)\big)}\right|. \tag{3.7}$$

Using the definition of σ_i, β_i and ω_i from (2.8), the numerator can be written as $\sigma_i\gamma + \beta_i\tanh(b_i) = \gamma(\sigma_i - d_i\tanh(b_i)) + \tanh(b_i)$. Since $0 < \tanh(x) < 1$, $\forall x > 0$ and $\sigma_i - d_i\tanh(b_i) > 0$, both the numerator and the denominator in (3.7) are positive. Now the difference between the numerator and the denominator is $(\sigma_i + d_i\tanh(a_i))(\omega_i + \sigma_i\tanh(b_i)) - \nu^{-1}(\sigma_i\gamma+\beta_i\tanh(b_i)) = (1+\tanh(b_i)\tanh(a_i))(\sigma_i d_i+\omega_i d_i\tanh(\sigma_i T)) > 0$, meaning that for each eigenvalue d_i, $0 < \nu^{-1}\frac{\sigma_i\gamma+\beta_i\tanh(b_i)}{(\sigma_i+d_i\tanh(a_i))(\omega_i+\sigma_i\tanh(b_i))} < 1$. ☐

Remark 3.5. For the Laplace operator with homogeneous Dirichlet boundary conditions in our model problem (1.2), there is no zero eigenvalue for its discretization matrix A. If an eigenvalue $d_i = 0$, we have $\sigma_i|_{d_i=0} = \sqrt{\nu^{-1}}$, $\omega_i|_{d_i=0} = \gamma\nu^{-1}$ and $\beta_i|_{d_i=0} = 1$. Substituting these values into the convergence factor (3.7), we find $\rho_{\mathrm{DN}_1}|_{\theta=1,d_i=0} = \nu^{-1}\frac{\sqrt{\nu^{-1}}\gamma+\tanh(\sqrt{\nu^{-1}}(T-\alpha))}{\sqrt{\nu^{-1}}(\gamma\nu^{-1}+\sqrt{\nu^{-1}}\tanh(\sqrt{\nu^{-1}}(T-\alpha)))} = 1$, and convergence is lost. The convergence behavior of the algorithm DN$_1$ for small eigenvalues is thus not good. Furthermore, inserting $d_i = 0$ into (3.6) and using the above result, we find that $\rho_{\mathrm{DN}_1}|_{d_i=0} = 1$, independently of the relaxation parameter θ and the interface position α: relaxation can not fix this problem.

Remark 3.6. If some d_i goes to infinity, we have $\sigma_i \sim_\infty d_i$ and $\omega_i \sim_\infty d_i$, and therefore $\lim_{d_i\to\infty}\left|1-\theta\big(1-\nu^{-1}\frac{\sigma_i\gamma+\beta_i\tanh(b_i)}{(\sigma_i+d_i\tanh(a_i))(\omega_i+\sigma_i\tanh(b_i))}\big)\right| = |1-\theta|$, which is independent of α, so high frequency convergence is robust with relaxation. One can use $\theta = 1$ to get a good smoother, with the following convergence factor estimate.

COROLLARY 3.7. *If A is positive semi-definite, then the algorithm DN_1 (3.1)-(3.2) with $\theta = 1$ satisfies the convergence estimate $\rho_{DN_1}|_{\theta=1} \le \frac{1+\gamma\sigma_{\min}}{\nu d_{\min}^2}$, with $d_{\min} := \min \lambda(A)$ the smallest eigenvalue of A.*

Proof. Since for $\theta = 1$, Corollary 3.4 shows that the convergence factor is between 0 and 1 for each eigenvalue d_i, we can take (3.7) and remove the absolute value, $\rho_{\mathrm{DN}_1}|_{\theta=1} = \nu^{-1} \max_{d_i\in\lambda(A)} \frac{\tanh(b_i)+\gamma(\sigma_i-d_i\tanh(b_i))}{(\sigma_i+d_i\tanh(a_i))(\omega_i+\sigma_i\tanh(b_i))}$. Using the definition of σ_i and ω_i from (2.8), we have $\sigma_i > d_i \ge 0$ and $\omega_i \ge d_i \ge 0$. Since $0 < \tanh(x) < 1$, $\forall x > 0$, we obtain that $\sigma_i + d_i \tanh(a_i) \ge d_i$, $\omega_i + \sigma_i \tanh(b_i) \ge d_i$ and $\sigma_i - d_i \tanh(b_i) \le \sigma_i$. This implies $\frac{\tanh(b_i)+\gamma(\sigma_i-d_i\tanh(b_i))}{(\sigma_i+d_i\tanh(a_i))(\omega_i+\sigma_i\tanh(b_i))} \le \frac{1+\gamma\sigma_i}{d_i^2} = \frac{1}{d_i}(\frac{1}{d_i} + \gamma\frac{\sigma_i}{d_i})$. Using once again the definition of σ_i from (2.8), we find $\frac{\sigma_i}{d_i} = \sqrt{1+\frac{\nu^{-1}}{d_i^2}} \le \sqrt{1+\frac{\nu^{-1}}{d_{\min}^2}}$. Hence, we have $\frac{1+\gamma\sigma_i}{d_i^2} \le \frac{1+\gamma\sigma_{\min}}{d_{\min}^2}$, which concludes the proof. ∎

Since A comes from a spatial discretization, the smallest eigenvalue of A depends only little on the spatial mesh size, and convergence is thus robust under mesh refinement. Corollary 3.7 is however less useful when ν is small: for example for $\gamma = 0$, the bound is less than one only if $\nu > \frac{1}{d_{\min}^2}$, but we have also the following convergence result.

THEOREM 3.8. *The algorithm DN_1 (3.1)-(3.2) converges for all initial guesses under the assumption that the matrix A is not singular.*

Proof. From Corollary 3.4, we know that the convergence factor satisfies $0 < \rho_{\mathrm{DN}_1}|_{\theta=1} < 1$. Using its definition (3.6), we find for $\theta \in (0,1)$, $0 < 1-\theta < \rho_{\mathrm{DN}_1} = 1-\theta(1-\rho_{\mathrm{DN}_1}|_{\theta=1}) < 1$, which concludes the proof. ∎

Remark 3.9. As shown in the previous proof, the function $g(\theta) := 1-\theta(1-\rho_{\mathrm{DN}_1}|_{\theta=1})$ is decreasing for $\theta \in (0,1)$, which makes $\theta = 1$ the best relaxation parameter. This is further confirmed by our numerical experiments (see Figure 4). Due to the bad convergence behavior of the algorithm DN_1 for small eigenvalues, it only makes this most natural DN algorithm a good smoother but not a good solver.

3.1.2. Neumann-Dirichlet algorithm (ND_1). We now invert the two conditions, and apply the Neumann condition to the state $\mu_{(i)}$ in Ω_1 and the Dirichlet condition to the state $z_{(i)}$ in Ω_2, still respecting the forward-backward structure. For iteration index $k = 1, 2, \ldots$, the algorithm ND_1 computes

$$
\left\{\begin{aligned}
\begin{pmatrix} \dot z_{1,(i)}^k \\ \dot\mu_{1,(i)}^k \end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1} \\ -1 & -d_i \end{pmatrix}\begin{pmatrix} z_{1,(i)}^k \\ \mu_{1,(i)}^k \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ in } \Omega_1,\\
z_{1,(i)}^k(0) &= 0,\\
\dot\mu_{1,(i)}^k(\alpha) &= \dot\mu_{2,(i)}^k(\alpha),
\end{aligned}\right.
\tag{3.8}
$$
$$
\left\{\begin{aligned}
\begin{pmatrix} \dot z_{2,(i)}^k \\ \dot\mu_{2,(i)}^k \end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1} \\ -1 & -d_i \end{pmatrix}\begin{pmatrix} z_{2,(i)}^k \\ \mu_{2,(i)}^k \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ in } \Omega_2,\\
z_{2,(i)}^k(\alpha) &= f_{\alpha,(i)}^{k-1},\\
\mu_{2,(i)}^k(T) + \gamma z_{2,(i)}^k(T) &= 0,
\end{aligned}\right.
$$

and we update the transmission condition by

$$
f_{\alpha,(i)}^k := (1-\theta) f_{\alpha,(i)}^{k-1} + \theta z_{1,(i)}^k(\alpha), \quad \theta \in (0,1). \tag{3.9}
$$

For the convergence analysis, we choose to use the formulation (2.7), i.e.

$$\text{(3.10)}\qquad \begin{cases} \ddot{\mu}^k_{1,(i)} - \sigma_i^2\mu^k_{1,(i)} = 0 \text{ in } \Omega_1, \\ \dot{\mu}_{(i)}(0) - d_i\mu_{(i)}(0) = 0, \\ \dot{\mu}^k_{1,(i)}(\alpha) = \dot{\mu}^k_{2,(i)}(\alpha), \end{cases} \quad \begin{cases} \ddot{\mu}^k_{2,(i)} - \sigma_i^2\mu^k_{2,(i)} = 0 \text{ in } \Omega_2, \\ \dot{\mu}^k_{2,(i)}(\alpha) - d_i\mu^k_{2,(i)}(\alpha) = f^{k-1}_{\alpha,(i)}, \\ \gamma\dot{\mu}_{(i)}(T) + \beta_i\mu_{(i)}(T) = 0, \end{cases}$$

where the update of the transmission condition (3.9) becomes

$$\text{(3.11)}\qquad f^k_{\alpha,(i)} = (1-\theta)f^{k-1}_{\alpha,(i)} + \theta\big(\dot{\mu}^k_{1,(i)}(\alpha) - d_i\mu^k_{1,(i)}(\alpha)\big), \quad \theta \in (0,1).$$

The algorithm ND_1 (3.8) can thus be interpreted as a NR type algorithm (3.10). Using the general solution (2.9) and the initial and final conditions, we get

$$\text{(3.12)}\qquad \begin{aligned} \mu^k_{1,(i)}(t) &= A_i^k\big(\sigma_i\cosh(\sigma_i t) + d_i\sinh(\sigma_i t)\big), \\ \mu^k_{2,(i)}(t) &= B_i^k\Big(\gamma\sigma_i\cosh\big(\sigma_i(T-t)\big) + \beta_i\sinh\big(\sigma_i(T-t)\big)\Big), \end{aligned}$$

and from the transmission condition in (3.10) on each domain, and we obtain $A_i^k = \frac{f^{k-1}_{\alpha,(i)}(\sigma_i\gamma\sinh(b_i)+\beta_i\cosh(b_i))}{(\omega_i\sinh(b_i)+\sigma_i\cosh(b_i))(\sigma_i\sinh(a_i)+d_i\cosh(a_i))}$ and $B_i^k = \frac{-f^{k-1}_{\alpha,(i)}}{\omega_i\sinh(b_i)+\sigma_i\cosh(b_i)}$. Using the relation (3.11), we find $f^k_{\alpha,(i)} = (1-\theta)f^{k-1}_{\alpha,(i)} + \theta f^{k-1}_{\alpha,(i)}\nu^{-1}\frac{\sigma_i\gamma+\beta_i\coth(b_i)}{(\sigma_i+d_i\coth(a_i))(\omega_i+\sigma_i\coth(b_i))}$, which leads to the following result.

THEOREM 3.10. *The algorithm ND_1 (3.8)-(3.9) converges if and only if*

$$\text{(3.13)}\qquad \rho_{ND_1} := \max_{d_i\in\lambda(A)}\left|1-\theta\Big(1-\nu^{-1}\frac{\sigma_i\gamma+\beta_i\coth(b_i)}{\big(\sigma_i+d_i\coth(a_i)\big)\big(\omega_i+\sigma_i\coth(b_i)\big)}\Big)\right| < 1.$$

The convergence factor of the algorithm ND_1 (3.13) is very similar to that of DN_1 (3.6). For instance, the behavior for large and small eigenvalues shown in Remarks 3.5 and 3.6 still hold: when inserting $d_i = 0$ into (3.13) we find $\rho_{\mathrm{ND}_1}|_{d_i=0} = |1-\theta(1-\nu^{-1}\frac{\sqrt{\nu^{-1}}\gamma+\coth(\sqrt{\nu^{-1}}(T-\alpha))}{\sqrt{\nu^{-1}}(\gamma\nu^{-1}+\sqrt{\nu^{-1}}\coth(\sqrt{\nu^{-1}}(T-\alpha)))})| = 1$, again independent of the relaxation parameter θ and the interface position α; and when the eigenvalue d_i goes to infinity, we find $\lim_{d_i\to\infty}|1-\theta(1-\nu^{-1}\frac{\sigma_i\gamma+\beta_i\coth(b_i)}{(\sigma_i+d_i\coth(a_i))(\omega_i+\sigma_i\coth(b_i))})| = |1-\theta|$, again independent of the interface position α. Due however to the presence of the hyperbolic cotangent function in (3.13) instead of the hyperbolic tangent function in (3.6), we need further assumptions to obtain results like Corollaries 3.4 and 3.7. Indeed, substituting $\theta = 1$ into (3.13) and using the definition of σ_i, β_i from (2.8), the numerator reads $\sigma_i\gamma + \beta_i\coth(b_i) = \gamma(\sqrt{d_i^2+\nu^{-1}} - d_i\coth(\sqrt{d_i^2+\nu^{-1}}(T-\alpha))) + \coth(\sqrt{d_i^2+\nu^{-1}}(T-\alpha))$. Depending on γ, ν and α, this value could be negative. However, by setting $\gamma = 0$, the numerator is guaranteed to be positive, and we obtain the following results.

COROLLARY 3.11. *If A is not singular and the parameter $\gamma = 0$, then the algorithm ND_1 (3.8)-(3.9) for $\theta = 1$ converges for all initial guesses.*

Proof. Substituting $\theta = 1$ and $\gamma = 0$ into (3.13), we get

$$\text{(3.14)}\qquad \rho_{\mathrm{ND}_2}|_{\theta=1} = \nu^{-1}\max_{d_i\in\lambda(A)}\left|\frac{\coth(b_i)}{\big(\sigma_i+d_i\coth(a_i)\big)\big(d_i+\sigma_i\coth(b_i)\big)}\right|.$$

Since $\coth(x) > 1$, $\forall x > 0$, both the numerator and the denominator in (3.14) are positive, and the difference between them is $(\sigma_i + d_i\coth(a_i))(d_i + \sigma_i\coth(b_i)) - \nu^{-1}\coth(b_i) = (\coth(a_i) + \coth(b_i))(d_i^2 + \sigma_i d_i\coth(\sigma_i T)) > 0$, meaning that for each eigenvalue d_i, $0 < \nu^{-1}\frac{\coth(b_i)}{(\sigma_i+d_i\coth(a_i))(\omega_i+\sigma_i\coth(b_i))} < 1$, which concludes the proof. ∎

COROLLARY 3.12. *If A is positive semi-definite and the parameter $\gamma = 0$, then the algorithm ND$_1$ (3.8)-(3.9) for $\theta = 1$ satisfies the convergence estimate*

$$\rho_{ND_1}|_{\theta=1} \leq \frac{\coth\left(\sigma_{\min}(T-\alpha)\right)}{\nu(\sigma_{\min}+d_{\min})^2}. \tag{3.15}$$

Proof. Since we have shown in Corollary 3.11 that the convergence factor is between 0 and 1 for each eigenvalue d_i, we can take (3.14) and remove the absolute value, $\rho_{\mathrm{ND}_2}|_{\theta=1} = \nu^{-1}\max_{d_i\in\lambda(A)} \frac{\coth(b_i)}{(\sigma_i+d_i\coth(a_i))(d_i+\sigma_i\coth(b_i))}$. Since $\sigma_i = \sqrt{d_i^2+\nu^{-1}} \geq d_i \geq 0$ and $\coth(x) > 1$, $\forall x > 0$, we obtain that $\sigma_i + d_i\coth(a_i) \geq \sigma_i + d_i$ and $d_i + \sigma_i\coth(b_i) \geq \sigma_i + d_i$. This implies that $\frac{\coth(b_i)}{(\sigma_i+d_i\coth(a_i))(d_i+\sigma_i\coth(b_i))} \leq \frac{\coth(b_i)}{(\sigma_i+d_i)^2}$. Recalling $\coth(b_i) = \coth(\sigma_i(T-\alpha))$, and using the fact that $d_i \geq d_{\min}$ and $\sigma_i \geq \sigma_{\min} := \sqrt{d_{\min}^2+\nu^{-1}}$, we find $\frac{\coth(b_i)}{(\sigma_i+d_i)^2} \leq \frac{\coth(\sigma_{\min}(T-\alpha))}{(\sigma_{\min}+d_{\min})^2}$, which concludes the proof. ∎

Like for DN$_1$, the estimate (3.15) is independent of the spatial mesh size, and since for $\gamma = 0$, the convergence factor satisfies $0 < \rho_{\mathrm{ND}_1}|_{\theta=1} < 1$ as shown in Corollary 3.11, using the definition of the convergence factor (3.13), we obtain the following result.

THEOREM 3.13. *The algorithm ND$_1$ (3.8)-(3.9) converges for all initial guesses if $\gamma = 0$ and the matrix A is not singular.*

3.2. Category II. We now study algorithms in Category II which run only on the state $z_{(i)}$ to solve the problem (2.4), based on DN and ND techniques.

3.2.1. Dirichlet-Neumann algorithm (DN$_2$). As explained in Table 1, we apply the Dirichlet condition in Ω_1 and the Neumann condition in Ω_2 both on the primal state $z_{(i)}$. For the iteration index $k = 1, 2, \ldots$, the algorithm DN$_2$ solves

$$\begin{cases} \begin{pmatrix} \dot{z}_{1,(i)}^k \\ \dot{\mu}_{1,(i)}^k \end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1} \\ -1 & -d_i \end{pmatrix} \begin{pmatrix} z_{1,(i)}^k \\ \mu_{1,(i)}^k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ in } \Omega_1, \\ z_{1,(i)}^k(0) = 0, \\ z_{1,(i)}^k(\alpha) = f_{\alpha,(i)}^{k-1}, \end{cases}$$
$$\begin{cases} \begin{pmatrix} \dot{z}_{2,(i)}^k \\ \dot{\mu}_{2,(i)}^k \end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1} \\ -1 & -d_i \end{pmatrix} \begin{pmatrix} z_{2,(i)}^k \\ \mu_{2,(i)}^k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ in } \Omega_2, \\ \dot{z}_{2,(i)}^k(\alpha) = \dot{z}_{1,(i)}^k(\alpha), \\ \mu_{2,(i)}^k(T) + \gamma z_{2,(i)}^k(T) = 0, \end{cases} \tag{3.16}$$

and we update the transmission condition by

$$f_{\alpha,(i)}^k := (1-\theta) f_{\alpha,(i)}^{k-1} + \theta z_{2,(i)}^k(\alpha), \quad \theta \in (0,1). \tag{3.17}$$

At first glance, this algorithm does not have the forward-backward structure, with both an initial and a final condition on $z_{1,(i)}$ in Ω_1 and nothing on $\mu_{1,(i)}$. However, as mentioned in Remark 3.1, this is only a matter of interpretation: using the identity of $z_{(i)}$ from (2.5), we can rewrite the transmission condition $z_{1,(i)}^k(\alpha) = f_{\alpha,(i)}^{k-1}$ as $\dot{\mu}_{1,(i)}^k(\alpha) - d_i\mu_{1,(i)}^k(\alpha) = f_{\alpha,(i)}^{k-1}$, and define the update (3.17) as $f_{\alpha,(i)}^k := (1-\theta)f_{\alpha,(i)}^{k-1} + \theta(\dot{\mu}_{2,(i)}^k(\alpha) - d_i\mu_{2,(i)}^k(\alpha))$, to rediscover the forward-backward structure. Moreover, with the interpretation of $\mu_{1,(i)}^k$, the algorithm DN$_2$ (3.16) is a RN type algorithm.

For the analysis, we choose the state $z_{(i)}$ formulation: for $i = 1, \dots, n$ and iteration index $k = 1, 2, \dots,$ the equivalent algorithm reads

$$\begin{cases} \ddot{z}_{1,(i)}^k - \sigma_i^2 z_{1,(i)}^k = 0 \text{ in } \Omega_1, \\ z_{1,(i)}^k(0) = 0, \\ z_{1,(i)}^k(\alpha) = f_{\alpha,(i)}^{k-1}, \end{cases} \qquad \begin{cases} \ddot{z}_{2,(i)}^k - \sigma_i^2 z_{2,(i)}^k = 0 \text{ in } \Omega_2, \\ \dot{z}_{2,(i)}^k(\alpha) = \dot{z}_{1,(i)}^k(\alpha), \\ \dot{z}_{2,(i)}^k(T) + \omega_i z_{2,(i)}^k(T) = 0, \end{cases} \tag{3.18}$$

where we still update the transmission condition by (3.17). Note that (3.18) is still a DN type algorithm, like (3.16). Using the solutions (3.5) to determine the two coefficients A_i^k and B_i^k, we get from (3.18) $A_i^k = \frac{f_{\alpha,(i)}^{k-1}}{\sinh(a_i)}$ and $B_i^k = -\frac{f_{\alpha,(i)}^{k-1}\coth(a_i)}{\sigma_i\sinh(b_i)+\omega_i\cosh(b_i)}$. With (3.17), we find $f_{\alpha,(i)}^k = (1-\theta)f_{\alpha,(i)}^{k-1} - \theta f_{\alpha,(i)}^{k-1}\coth(a_i)\frac{\sigma_i\cosh(b_i)+\omega_i\sinh(b_i)}{\sigma_i\sinh(b_i)+\omega_i\cosh(b_i)}$, and thus obtain the following convergence results.

THEOREM 3.14. *The algorithm DN*$_2$ (3.16)-(3.17) *converges if and only if*

$$\rho_{DN_2} := \max_{d_i\in\lambda(A)}\left|1-\theta\Big(1+\coth(a_i)\frac{\sigma_i\coth(b_i)+\omega_i}{\sigma_i+\omega_i\coth(b_i)}\Big)\right| < 1. \tag{3.19}$$

COROLLARY 3.15. *The algorithm DN*$_2$ *for* $\theta = 1$ *does not converge if* $\alpha \le \frac{T}{2}$.

Proof. Substituting $\theta = 1$ into (3.19), we have

$$\rho_{\mathrm{DN}_2}|_{\theta=1} = \max_{d_i\in\lambda(A)}\left|\coth(a_i)\frac{\sigma_i\coth(b_i)+\omega_i}{\sigma_i+\omega_i\coth(b_i)}\right|. \tag{3.20}$$

Since $\coth(x) > 1$, $\forall x > 0$, both the numerator and the denominator in (3.20) are positive. When $a_i \le b_i$ (i.e., $\alpha \le T-\alpha$), we have $\coth(a_i) \ge \coth(b_i)$, and thus the difference between the numerator and the denominator is $\coth(a_i)(\omega_i+\sigma_i\coth(b_i)) - (\sigma_i+\omega_i\coth(b_i)) = \omega_i(\coth(a_i)-\coth(b_i)) + \sigma_i(\coth(b_i)\coth(a_i)-1) > 0$, meaning that $\coth(a_i)\frac{\sigma_i\coth(b_i)+\omega_i}{\sigma_i+\omega_i\coth(b_i)} > 1$, which concludes the proof. ∎

We need some extra assumptions to conclude for the case $\alpha > \frac{T}{2}$.

COROLLARY 3.16. *The algorithm DN*$_2$ *for* $\theta = 1$ *does not converge if* $\gamma = 0$.

Proof. We showed in Corollary 3.15 the result for $\alpha \le \frac{T}{2}$. Now $\alpha > \frac{T}{2}$ implies that $a_i > b_i$, thus $\coth(a_i) < \coth(b_i)$. Inserting $\gamma = 0$ into (3.20) and using the definition of σ_i from (2.8), the difference between the numerator and the denominator of (3.20) becomes $\coth(a_i)(d_i+\sigma_i\coth(b_i)) - (\sigma_i+d_i\coth(b_i)) = (\coth(a_i)-\coth(b_i))(d_i + \sigma_i\coth(b_i-a_i)) > 0$, where we use the fact that $d_i+\sigma_i\coth(b_i-a_i) < d_i-\sigma_i < 0$. This shows that DN$_2$ for $\theta = 1$ also does not converge for $\alpha > \frac{T}{2}$ when $\gamma = 0$. ∎

Unlike in Corollary 3.7 where we have an estimate of the convergence factor for DN$_1$, we cannot provide a general convergence estimate for the algorithm DN$_2$ (3.16)-(3.17), since we showed in Corollary 3.15 and Corollary 3.16 that it does not converge in some cases. However, we can still show the convergence behavior for extreme eigenvalues. In particular, if the eigenvalue $d_i = 0$, we find

$$\rho_{\mathrm{DN}_2}|_{d_i=0} = \left|1-\theta\Big(1+\coth(\sqrt{\nu^{-1}}\alpha)\frac{\coth\big(\sqrt{\nu^{-1}}(T-\alpha)\big)+\gamma\sqrt{\nu^{-1}}}{1+\gamma\sqrt{\nu^{-1}}\coth\big(\sqrt{\nu^{-1}}(T-\alpha)\big)}\Big)\right|. \tag{3.21}$$

When the eigenvalue goes to infinity, using Remark 3.6, we obtain $\lim_{\mathrm{d}_i\to\infty}\rho_{\mathrm{DN}_2} = |1-2\theta|$. By equioscillating the convergence factor for small (i.e., $\rho_{\mathrm{DN}_2}|_{d_i=0}$) and large eigenvalues (i.e., $\rho_{\mathrm{DN}_2}|_{d_i\to\infty}$), we obtain after some computations

$$\theta^*_{\mathrm{DN}_2} = \frac{2}{3+\coth(\sqrt{\nu^{-1}}\alpha)\frac{\coth\left(\sqrt{\nu^{-1}}(T-\alpha)\right)+\gamma\sqrt{\nu^{-1}}}{1+\gamma\sqrt{\nu^{-1}}\coth\left(\sqrt{\nu^{-1}}(T-\alpha)\right)}}. \tag{3.22}$$

THEOREM 3.17. *If we assume that the eigenvalues of A are anywhere in the interval $[0,\infty)$, then the optimal relaxation parameter $\theta^\star_{DN_2}$ for the algorithm DN_2 (3.16)-(3.17) with $\gamma=0$ is given by (3.22) and satisfies $\theta^\star_{DN_2}<\frac{1}{2}$.*

Proof. Taking the derivative of the convergence factor ρ_{DN_2} from (3.19) with respect to the eigenvalue d_i, we get $\frac{\mathrm{d}\rho_{\mathrm{DN}_2}}{\mathrm{d}d_i} = -\frac{d_i\alpha}{\sigma_i\sinh^2(a_i)}\frac{\sigma_i\coth(b_i)+\omega_i}{\sigma_i+\omega_i\coth(b_i)} - \frac{\nu^{-1}\coth(a_i)}{\sigma_i}\frac{\beta_i(\coth^2(b_i)-1)+\frac{d_i(T-\alpha)}{\sinh^2(b_i)}(1-\gamma^2\nu^{-1}-2d_i\gamma)}{(\sigma_i+\omega_i\coth(b_i))^2}$, where we used σ_i,ω_i and β_i from (2.8). The derivative becomes negative with $\gamma=0$, meaning that the convergence factor decreases with respect to the eigenvalue d_i. We can then deduce the optimal relaxation parameter using equioscillation: inserting $\gamma=0$ into (3.22), the denominator becomes $3+\coth(\sqrt{\nu^{-1}}\alpha)\coth(\sqrt{\nu^{-1}}(T-\alpha))<4$. ∎

For $\gamma>0$, it is not clear when the convergence factor ρ_{DN_2} is monotonic with respect to the eigenvalues, and thus the optimal relaxation parameter $\theta^\star_{\mathrm{DN}_2}$ could differ from (3.22).

3.2.2. Neumann-Dirichlet algorithm ($\mathbf{ND_2}$). We now invert the two conditions: for the iteration index $k=1,2,\dots$, the algorithm ND_2 to study is

$$\begin{cases} \begin{pmatrix}\dot z^k_{1,(i)}\\ \dot\mu^k_{1,(i)}\end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1}\\ -1 & -d_i\end{pmatrix}\begin{pmatrix} z^k_{1,(i)}\\ \mu^k_{1,(i)}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \text{ in } \Omega_1,\\ z^k_{1,(i)}(0)=0,\\ \dot z^k_{1,(i)}(\alpha) = f^{k-1}_{\alpha,(i)}, \end{cases}$$
$$\begin{cases} \begin{pmatrix}\dot z^k_{2,(i)}\\ \dot\mu^k_{2,(i)}\end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1}\\ -1 & -d_i\end{pmatrix}\begin{pmatrix} z^k_{2,(i)}\\ \mu^k_{2,(i)}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \text{ in } \Omega_2,\\ z^k_{2,(i)}(\alpha)=z^k_{1,(i)}(\alpha),\\ \mu^k_{2,(i)}(T)+\gamma z^k_{2,(i)}(T) = 0, \end{cases} \tag{3.23}$$

and then we update the transmission condition by

$$f^k_{\alpha,(i)} := (1-\theta)f^{k-1}_{\alpha,(i)} + \theta\dot z^k_{2,(i)}(\alpha), \quad \theta\in(0,1). \tag{3.24}$$

Similar to the algorithm DN_2 (3.16)-(3.17), we cannot see the forward-backward structure in Ω_1 for the algorithm ND_2 (3.23)-(3.24). But by interpreting the Neumann condition on $z_{1,(i)}$ in terms of $\mu_{1,(i)}$ as explained in Remark 3.1, the forward-backward structure is again revealed through a RD type algorithm.

We proceed for the convergence analysis using the formulation (2.6): for $i=1,\dots,n$ and iteration index $k=1,2,\dots$, we solve

$$\begin{cases} \ddot z^k_{1,(i)} - (d_i^2+\nu^{-1})z^k_{1,(i)} = 0 \text{ in } \Omega_1,\\ z^k_{1,(i)}(0)=0,\\ \dot z^k_{1,(i)}(\alpha) = f^{k-1}_{\alpha,(i)}, \end{cases} \quad \begin{cases} \ddot z^k_{2,(i)} - (d_i^2+\nu^{-1})z^k_{2,(i)} = 0 \text{ in } \Omega_2,\\ z^k_{2,(i)}(\alpha) = z^k_{1,(i)}(\alpha),\\ \dot z^k_{2,(i)}(T)+d_iz^k_{2,(i)}(T) = -\gamma\nu^{-1}z^k_{2,(i)}(T), \end{cases} \tag{3.25}$$

where we still update the transmission condition by (3.24). Note that both algorithms (3.23) and (3.25) are of ND type.

Using the solutions (3.5) and the transmission condition in (3.24), we obtain $A_i^k = \frac{f_{\alpha,(i)}^{k-1}}{\sigma_i \cosh(a_i)}$, $B_i^k = \frac{f_{\alpha,(i)}^{k-1} \tanh(a_i)/\sigma_i}{\sigma_i \cosh(b_i)+\omega_i \sinh(b_i)}$, and we therefore get for the update condition (3.24) $f_{\alpha,(i)}^k = (1-\theta) f_{\alpha,(i)}^{k-1} - \theta f_{\alpha,(i)}^{k-1} \tanh(a_i) \frac{\sigma_i \sinh(b_i)+\omega_i \cosh(b_i)}{\sigma_i \cosh(b_i)+\omega_i \sinh(b_i)}$.

THEOREM 3.18. *The algorithm ND_2 (3.23)-(3.24) converges if and only if*

$$\rho_{ND_2} := \max_{d_i \in \lambda(A)} \left| 1 - \theta \Big(1 + \tanh(a_i) \frac{\sigma_i \tanh(b_i) + \omega_i}{\sigma_i + \omega_i \tanh(b_i)} \Big) \right| < 1. \tag{3.26}$$

COROLLARY 3.19. *The algorithm ND_2 for $\theta = 1$ converges if $\alpha \leq \frac{T}{2}$.*

Proof. Substituting $\theta = 1$ into (3.26), we have

$$\rho_{\mathrm{ND}_2}|_{\theta=1} = \max_{d_i \in \lambda(A)} \left| \tanh(a_i) \frac{\sigma_i \tanh(b_i) + \omega_i}{\sigma_i + \omega_i \tanh(b_i)} \right|. \tag{3.27}$$

Since $0 < \tanh(x) < 1$, $\forall x > 0$, both the numerator and the denominator in (3.27) are positive. In the case where $a_i \leq b_i$ (i.e., $\alpha \leq T - \alpha$), we have $\tanh(a_i) \leq \tanh(b_i)$, and the difference between the numerator and the denominator is $\tanh(a_i)(\omega_i + \sigma_i \tanh(b_i)) - (\sigma_i + \omega_i \tanh(b_i)) = \omega_i(\tanh(a_i) - \tanh(b_i)) + \sigma_i(\tanh(b_i)\tanh(a_i) - 1) < 0$, meaning that $0 < \tanh(a_i) \frac{\sigma_i \tanh(b_i)+\omega_i}{\sigma_i+\omega_i \tanh(b_i)} < 1$. This concludes the proof. ∎

As shown in Corollary 3.15, the algorithm DN_2 (3.16)-(3.17) with $\theta = 1$ does not converge for $\alpha \leq \frac{T}{2}$, whereas the algorithm ND_2 (3.23)-(3.24) converges in this case. This reveals a symmetry behavior, since the only difference between these two algorithms is that we exchange the Dirichlet and the Neumann condition in the two subdomains. This symmetry is well-known for classical DN and ND algorithms.

COROLLARY 3.20. *For $\gamma = 0$, the algorithm ND_2 for $\theta = 1$ converges for all initial guesses.*

Proof. This is shown in Corollary 3.19 for $\alpha \leq \frac{T}{2}$. If $\alpha > \frac{T}{2}$, i.e. $a_i > b_i$, then $\tanh(a_i) > \tanh(b_i)$, and the difference between the numerator and the denominator is $\tanh(a_i)(d_i + \sigma_i \tanh(b_i)) - (\sigma_i + d_i \tanh(b_i)) = (\tanh(b_i)\tanh(a_i) - 1)(\sigma_i - d_i \tanh(a_i - b_i)) < 0$, where we use the fact that $0 < \sigma_i - d_i < \sigma_i - d_i \tanh(a_i - b_i)$. This shows that the algorithm ND_2 for $\theta = 1$ converge for $\alpha > \frac{T}{2}$ in the case $\gamma = 0$. ∎

Notice that the matrix A here can be singular, in contrast to the algorithm DN_1 in Corollary 3.4 where non-singularity is needed for A. As in the previous section, we can still show the convergence behavior for extreme eigenvalues. If the eigenvalue $d_i = 0$, we find

$$\rho_{\mathrm{ND}_2}|_{d_i=0} = \left| 1 - \theta \Big(1 + \tanh(\sqrt{\nu^{-1}}\alpha) \frac{\tanh\left(\sqrt{\nu^{-1}}(T-\alpha)\right) + \gamma\sqrt{\nu^{-1}}}{1 + \gamma\sqrt{\nu^{-1}} \tanh\left(\sqrt{\nu^{-1}}(T-\alpha)\right)} \Big) \right|. \tag{3.28}$$

The expression (3.28) is very similar to (3.21): when $\gamma = 0$, the convergence factor (3.21) becomes $\rho_{\mathrm{DN}_2}|_{d_i=0,\gamma=0} = |1 - \theta(1 + \coth(\sqrt{\nu^{-1}}\alpha)\coth(\sqrt{\nu^{-1}}\ (T-\alpha)))|$, whereas (3.28) becomes $\rho_{\mathrm{ND}_2}|_{d_i=0,\gamma=0} = |1 - \theta(1 + \tanh(\sqrt{\nu^{-1}}\alpha)\tanh(\sqrt{\nu^{-1}}(T-\alpha)))|$. We find again the symmetry between DN_2 and ND_2. In the case when the eigenvalue goes to infinity, using Remark 3.6, we obtain $\lim_{\mathrm{d}_i \to \infty} \rho_{\mathrm{ND}_2} = |1 - 2\theta|$, as for DN_2. By equioscillating the convergence factor again for small and large eigenvalues, we obtain after some computations the relaxation parameter

$$\theta^*_{\mathrm{ND}_2} = \frac{2}{3 + \tanh(\sqrt{\nu^{-1}}\alpha)\frac{\tanh\left(\sqrt{\nu^{-1}}(T-\alpha)\right)+\gamma\sqrt{\nu^{-1}}}{1+\gamma\sqrt{\nu^{-1}}\tanh\left(\sqrt{\nu^{-1}}(T-\alpha)\right)}}. \tag{3.29}$$

We thus obtain a similar result as Theorem 3.17.

THEOREM 3.21. *If we assume that the eigenvalues of A are anywhere in the interval $[0, \infty)$, then the optimal relaxation parameter $\theta^\star_{ND_2}$ for the algorithm ND_2 (3.23)-(3.24) with $\gamma = 0$ is given by (3.29), and satisfies $\frac{1}{2} < \theta^\star_{ND_2} < \frac{2}{3}$.*

Proof. As for Theorem 3.17, we take the derivative of ρ_{ND_2} with respect to d_i, $\frac{\mathrm{d}\rho_{\mathrm{ND}_2}}{\mathrm{d}d_i} = \frac{d_i\alpha}{\sigma_i\cosh^2(a_i)}\frac{\sigma_i\tanh(b_i)+\omega_i}{\sigma_i+\omega_i\tanh(b_i)} + \frac{\nu^{-1}\tanh(a_i)}{\sigma_i}\frac{\beta_i(1-\tanh^2(b_i))-\frac{d_i(T-\alpha)}{\cosh^2(b_i)}(\gamma^2\nu^{-1}+2d_i\gamma-1)}{(\sigma_i+\omega_i\tanh(b_i))^2}$, with σ_i, ω_i and β_i defined in (2.8). For $\gamma = 0$, the derivative is positive and thus ρ_{ND_2} increases with d_i. Therefore $\theta^*_{\mathrm{ND}_2}$ is determined by equioscillation. Inserting $\gamma = 0$ into (3.29), the denominator becomes $3 + \tanh(\sqrt{\nu^{-1}}\alpha)\tanh(\sqrt{\nu^{-1}}(T-\alpha)) < 4$. $\square$

As for DN_2 however, the monotonicity of the convergence factor ρ_{ND_2} is not guaranteed for $\gamma > 0$, and the optimal relaxation parameter $\theta^\star_{\mathrm{ND}_2}$ may differ from (3.29).

3.3. Category III. We finally study algorithms in Category III which run only on the state $\mu_{(i)}$ to solve the problem (2.4), and use DN and ND techniques.

3.3.1. Dirichlet-Neumann algorithm ($\mathbf{DN}_3$). As shown in Table 1, we apply the Dirichlet condition in Ω_1 and the Neumann condition in Ω_2, both to the state $\mu_{(i)}$. For iteration index $k = 1, 2, \ldots$, the algorithm DN_3 solves

$$\begin{cases} \begin{pmatrix} \dot z^k_{1,(i)} \\ \dot\mu^k_{1,(i)} \end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1} \\ -1 & -d_i \end{pmatrix}\begin{pmatrix} z^k_{1,(i)} \\ \mu^k_{1,(i)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ in } \Omega_1, \\ z^k_{1,(i)}(0) = 0, \\ \mu^k_{1,(i)}(\alpha) = f^{k-1}_{\alpha,(i)}, \end{cases}$$
$$\begin{cases} \begin{pmatrix} \dot z^k_{2,(i)} \\ \dot\mu^k_{2,(i)} \end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1} \\ -1 & -d_i \end{pmatrix}\begin{pmatrix} z^k_{2,(i)} \\ \mu^k_{2,(i)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ in } \Omega_2, \\ \dot\mu^k_{2,(i)}(\alpha) = \dot\mu^k_{1,(i)}(\alpha), \\ \mu^k_{2,(i)}(T) + \gamma z^k_{2,(i)}(T) = 0, \end{cases} \tag{3.30}$$

and we update the transmission condition by

$$f^k_{\alpha,(i)} := (1-\theta)f^{k-1}_{\alpha,(i)} + \theta\mu^k_{2,(i)}(\alpha), \theta \in (0,1). \tag{3.31}$$

The forward-backward structure is now less present in Ω_2, where we would expect to have an initial condition for $z_{2,(i)}$ instead of $\mu_{2,(i)}$. By using the identity of $\mu_{(i)}$ in (2.5), we can interpret the Neumann condition $\dot\mu^k_{2,(i)}(\alpha) = \dot\mu^k_{1,(i)}(\alpha)$ as $d_i\dot z^k_{2,(i)}(\alpha) + \sigma_i^2 z^k_{2,(i)}(\alpha) = d_i\dot z^k_{1,(i)}(\alpha) + \sigma_i^2 z^k_{1,(i)}(\alpha)$, a Robin type condition on $z_{2,(i)}$. Therefore, the algorithm DN_3 can also be interpreted as a DR algorithm.

For the convergence analysis, it is natural to choose the interpretation in $\mu_{(i)}$, i.e., using (2.7), which gives

$$\begin{cases} \ddot\mu^k_{1,(i)} - \sigma_i^2\mu^k_{1,(i)} = 0 \text{ in } \Omega_1, \\ \dot\mu_{(i)}(0) - d_i\mu_{(i)}(0) = 0, \\ \mu^k_{1,(i)}(\alpha) = f^{k-1}_{\alpha,(i)}, \end{cases} \quad \begin{cases} \ddot\mu^k_{2,(i)} - \sigma_i^2\mu^k_{2,(i)} = 0 \text{ in } \Omega_2, \\ \dot\mu^k_{2,(i)}(\alpha) = \dot\mu^k_{1,(i)}(\alpha), \\ \gamma\dot\mu_{(i)}(T) + \beta_i\mu_{(i)}(T) = 0, \end{cases} \tag{3.32}$$

where we still update the transmission condition through (3.31). We observe that both (3.30) and (3.32) are DN type algorithms. Proceeding as before, we obtain:

THEOREM 3.22. *The algorithm DN_3 (3.30)-(3.31) converges if and only if*

$$\rho_{DN_3} := \max_{d_i\in\lambda(A)} \left|1-\theta\Big(1+\frac{\sigma_i+d_i\coth(a_i)}{\sigma_i\coth(a_i)+d_i}\frac{\gamma\sigma_i\coth(b_i)+\beta_i}{\gamma\sigma_i+\beta_i\coth(b_i)}\Big)\right| < 1. \tag{3.33}$$

To get more insight, we choose $\theta=1$ in (3.33), and find

$$\rho_{\mathrm{DN}_3}|_{\theta=1} = \max_{d_i\in\lambda(A)} \left|\frac{\sigma_i+d_i\coth(a_i)}{\sigma_i\coth(a_i)+d_i}\frac{\gamma\sigma_i\coth(b_i)+\beta_i}{\gamma\sigma_i+\beta_i\coth(b_i)}\right|. \tag{3.34}$$

It is less clear whether $\gamma\sigma_i+\beta_i\coth(b_i)$ is positive, since, using the definition of β_i and σ_i from (2.8), we have $\gamma\sigma_i+\beta_i\coth(b_i)=\gamma(\sqrt{d_i^2+\nu^{-1}}-d_i\coth(\sqrt{d_i^2+\nu^{-1}}(T-\alpha)))+\coth(\sqrt{d_i^2+\nu^{-1}}(T-\alpha))$, and depending on the values of ν,γ and α, this could be negative. However, we can simplify (3.34) by setting $\gamma=0$, and obtain:

COROLLARY 3.23. *If $\gamma=0$, then the algorithm DN_3 with $\theta=1$ converges for all initial guesses.*

Proof. Substituting $\theta=1$ into (3.34), we have

$$\rho_{\mathrm{DN}_3}|_{\theta=1} = \max_{d_i\in\lambda(A)} \left|\frac{\sigma_i\tanh(a_i)+d_i}{\sigma_i+d_i\tanh(a_i)}\tanh(b_i)\right|. \tag{3.35}$$

Both the numerator and the denominator are positive. Using $0<\tanh(x)<1$, $\forall x>0$, we get $(d_i+\sigma_i\tanh(a_i))-(\sigma_i+d_i\tanh(a_i))=(d_i-\sigma_i)(1-\tanh(a_i))<0$, meaning that $0<\tanh(b_i)\frac{\sigma_i\tanh(a_i)+d_i}{\sigma_i+d_i\tanh(a_i)}<1$, which concludes the proof. ∎

For $\gamma=0$, the algorithm DN_3 (3.30)-(3.31) converges for $\theta=1$ as well as the algorithm ND_2 (3.23)-(3.24), since their convergence factors are very similar. For extreme eigenvalues, inserting $d_i=0$ into (3.33), we find the identical formula as (3.28), and when the eigenvalue goes to infinity, we also obtain $\lim_{\mathrm{d}_i\to\infty}\rho_{\mathrm{DN}_3}=|1-2\theta|$. By equioscillating the convergence factor between small and large eigenvalues, we obtain thus the same relaxation parameter as (3.29), which leads to:

THEOREM 3.24. *If we assume the eigenvalues of A can be anywhere in the interval $[0,\infty)$, then the optimal relaxation parameter $\theta^\star_{DN_3}$ for the algorithm DN_3 (3.30)-(3.31) with $\gamma=0$ is identical to $\theta^\star_{ND_2}$.*

Proof. For $\gamma=0$, the convergence factors (3.27) and (3.35) become the same when exchanging a_i and b_i, and the result thus follows as for Theorem 3.21. ∎

3.3.2. Neumann-Dirichlet algorithm ($\mathbf{ND}_3$). We now exchange the Dirichlet and Neumann conditions on the two subdomains, and obtain

$$\begin{cases} \begin{pmatrix}\dot z^k_{1,(i)}\\ \dot\mu^k_{1,(i)}\end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1}\\ -1 & -d_i\end{pmatrix}\begin{pmatrix} z^k_{1,(i)}\\ \mu^k_{1,(i)}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \text{ in } \Omega_1,\\ z^k_{1,(i)}(0)=0,\\ \dot\mu^k_{1,(i)}(\alpha)=f^{k-1}_{\alpha,(i)},\end{cases}$$
$$\begin{cases} \begin{pmatrix}\dot z^k_{2,(i)}\\ \dot\mu^k_{2,(i)}\end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1}\\ -1 & -d_i\end{pmatrix}\begin{pmatrix} z^k_{2,(i)}\\ \mu^k_{2,(i)}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \text{ in } \Omega_2,\\ \mu^k_{2,(i)}(\alpha)=\mu^k_{1,(i)}(\alpha),\\ \mu^k_{2,(i)}(T)+\gamma z^k_{2,(i)}(T)=0,\end{cases} \tag{3.36}$$

where the transmission condition is updated by

$$f_{\alpha,(i)}^k := (1-\theta) f_{\alpha,(i)}^{k-1} + \theta \dot{\mu}_{2,(i)}^k(\alpha), \theta \in (0,1). \tag{3.37}$$

As for DN$_3$, we need to use the identity (2.5) and interpret $\mu_{2,(i)}^k(\alpha) = \mu_{1,(i)}^k(\alpha)$ as $\dot{z}_{2,(i)}^k(\alpha) + d_i z_{2,(i)}^k(\alpha) = \dot{z}_{1,(i)}^k(\alpha) + d_i z_{1,(i)}^k(\alpha)$ to reveal the forward-backward structure with a NR type algorithm. Using formulation (2.7), we get

$$\begin{cases} \ddot{\mu}_{1,(i)}^k - \sigma_i^2 \mu_{1,(i)}^k = 0 \text{ in } \Omega_1, \\ \dot{\mu}_{(i)}(0) - d_i \mu_{(i)}(0) = 0, \\ \dot{\mu}_{1,(i)}^k(\alpha) = f_{\alpha,(i)}^{k-1}, \end{cases} \qquad \begin{cases} \ddot{\mu}_{2,(i)}^k - \sigma_i^2 \mu_{2,(i)}^k = 0 \text{ in } \Omega_2, \\ \mu_{2,(i)}^k(\alpha) = \mu_{1,(i)}^k(\alpha), \\ \gamma \dot{\mu}_{(i)}(T) + \beta_i \mu_{(i)}(T) = 0. \end{cases} \tag{3.38}$$

THEOREM 3.25. *The algorithm ND$_3$ (3.36)-(3.37) converges if and only if*

$$\rho_{ND_3} := \max_{d_i \in \lambda(A)} \Big| 1 - \theta\Big(1 + \frac{\sigma_i + d_i \tanh(a_i)}{\sigma_i \tanh(a_i) + d_i} \frac{\gamma\sigma_i \tanh(b_i) + \beta_i}{\gamma\sigma_i + \beta_i \tanh(b_i)}\Big)\Big| < 1. \tag{3.39}$$

As in the previous section, we choose $\theta = 1$ in (3.39), and find

$$\rho_{\mathrm{ND}_3}|_{\theta=1} = \max_{d_i \in \lambda(A)} \Big| \frac{\sigma_i + d_i \tanh(a_i)}{\sigma_i \tanh(a_i) + d_i} \frac{\gamma\sigma_i \tanh(b_i) + \beta_i}{\gamma\sigma_i + \beta_i \tanh(b_i)} \Big|. \tag{3.40}$$

Again, using the definition of β_i and σ_i from (2.8), we have $\gamma\sigma_i \tanh(b_i) + \beta_i = \gamma(\sqrt{d_i^2 + \nu^{-1}} \tanh(\sqrt{d_i^2 + \nu^{-1}}(T-\alpha)) - d_i) + 1$, and depending on the values of ν, γ and α, this could be negative. However, we can simplify (3.40) by taking $\gamma = 0$, and then obtain the following result.

COROLLARY 3.26. *If $\gamma = 0$, then the algorithm ND$_3$ with $\theta = 1$ does not converge.*

Proof. Inserting $\gamma = 0$ into (3.40), we get

$$\rho_{\mathrm{DN}_3}|_{\theta=1} = \max_{d_i \in \lambda(A)} \Big| \frac{\sigma_i \coth(a_i) + d_i}{\sigma_i + d_i \coth(a_i)} \coth(b_i) \Big|. \tag{3.41}$$

Both the numerator and the denominator are positive. Using $\coth(x) \geq 1, \forall x > 0$, we find $(d_i + \sigma_i \coth(a_i)) - (\sigma_i + d_i \coth(a_i)) = (\sigma_i - d_i)(\coth(a_i) - 1) > 0$, implying that $\frac{\sigma_i \coth(a_i) + d_i}{\sigma_i + d_i \coth(a_i)} \coth(b_i) > 1$, which concludes the proof. ∎

Comparing Corollaries 3.23 and 3.26, we find again a symmetry if $\gamma = 0$, as for Corollaries 3.15 and 3.19, and with $\theta = 1$, ND$_3$ diverges like DN$_2$ when $\gamma = 0$. In fact, in this case, the convergence factor of ND$_3$ (3.41) is very similar to the convergence factor of DN$_2$ (3.20). Due to this divergence, we cannot provide a general estimate of the convergence factor. We can however still study the convergence behavior for extreme eigenvalues. Inserting $d_i = 0$ into (3.39), we find also (3.21), and thus for small eigenvalues ND$_3$ behaves like DN$_2$, like we observed for ND$_2$ and DN$_3$ earlier. When the eigenvalue goes to infinity, we also obtain $\lim_{\mathrm{d}_i \to \infty} \rho_{\mathrm{ND}_3} = |1 - 2\theta|$. Hence all the four algorithms DN$_2$, ND$_2$, DN$_3$ and ND$_3$ have the same limit for large eigenvalues. By equioscillation, we then obtain the same relaxation parameter as (3.22). This leads to a similar result as Theorem 3.17.

THEOREM 3.27. *If we assume that the eigenvalues of A are anywhere in the interval $[0, \infty)$, then the optimal relaxation parameter $\theta_{ND_3}^\star$ for the algorithm ND$_3$ (3.36)-(3.37) with $\gamma = 0$ is identical to $\theta_{DN_2}^\star$.*

Proof. In the case $\gamma = 0$, the convergence factors (3.20) and (3.41) are the same when exchanging a_i and b_i, and thus the proof follows as for Theorem 3.17. ∎

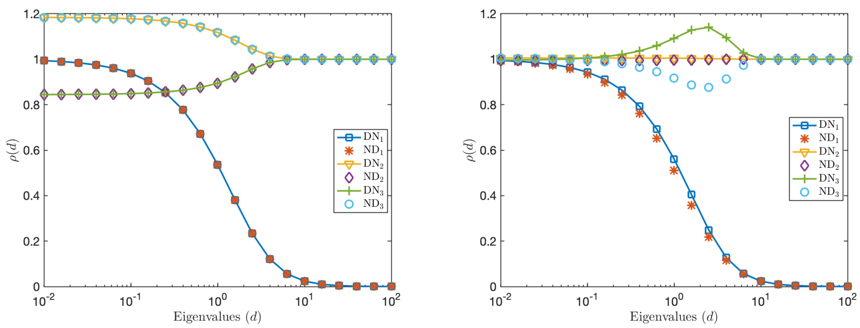

FIG. 2. *Convergence factor with $\theta = 1$ for a symmetric decomposition of the six new algorithms as function of the eigenvalues $d \in [10^{-2}, 10^2]$. Left: $\gamma = 0$. Right: $\gamma = 10$.*

4. Numerical experiments. We illustrate now our six new time domain decomposition algorithms with numerical experiments. We divide the time domain $\Omega = (0, 1)$ into two non-overlapping subdomains with interface α, and fix the regularization parameter $\nu = 0.1$. We will investigate the performance by plotting the convergence factor as function of the eigenvalues $d \in [10^{-2}, 10^2]$.

4.1. Convergence factor with $\theta = 1$ for a symmetric decomposition. We show in Figure 2 the convergence factors for all six algorithms for a symmetric decomposition, $\alpha = \frac{1}{2}$, with $\theta = 1$, on the left without final target state (i.e., $\gamma = 0$), and on the right with a final target state for $\gamma = 10$. Without final target state, the convergence factor of DN$_1$ and ND$_1$ coincide, as one can see also by substituting $\gamma = 0$ and $a_i = b_i$ into (3.7) and (3.14). The same also holds for the pairs DN$_2$ and ND$_3$, and DN$_3$ and ND$_2$. We also see the symmetry between DN$_2$ and ND$_2$, as well as DN$_3$ and ND$_3$. This changes when a final target state with $\gamma = 10$ is present: while the convergence behavior remains similar for DN$_1$ and ND$_1$, the symmetry between DN$_2$ and ND$_2$[1] and DN$_3$ and ND$_3$ remains. Furthermore, DN$_3$ converges with no final target but diverges with $\gamma = 10$, and vice versa for ND$_3$. In terms of the convergence speed, DN$_1$ and ND$_1$ are much better than the other four algorithms for high frequencies in both cases, and ND$_1$ is slightly better overall than DN$_1$ when $\gamma = 10$. The good high frequency behavior follows from our analysis: it depends for all 6 algorithms only on θ. In the case $\theta = 1$ here, the limit is $|1 - \theta| = 0$ for DN$_1$ and ND$_1$, and $|1 - 2\theta| = 1$ for DN$_2$, DN$_3$, ND$_2$ and ND$_3$. For the zero frequency, $d = 0$, the convergence factor for DN$_1$ and ND$_1$ equals 1 for all γ, but for DN$_2$, DN$_3$, ND$_2$ and ND$_3$ this depends on γ. Inserting $\theta = 1$ into (3.21) and (3.28), we obtain for DN$_2$ and ND$_3$ the convergence factor $\coth(\sqrt{\nu^{-1}}\alpha)\frac{\sqrt{\nu^{-1}}\coth(\sqrt{\nu^{-1}}\alpha)+\nu^{-1}\gamma}{\sqrt{\nu^{-1}}+\nu^{-1}\gamma\coth(\sqrt{\nu^{-1}}\alpha)}$, and for ND$_2$ and DN$_3$ $\tanh(\sqrt{\nu^{-1}}\alpha)\frac{\sqrt{\nu^{-1}}\tanh(\sqrt{\nu^{-1}}\alpha)+\nu^{-1}\gamma}{\sqrt{\nu^{-1}}+\nu^{-1}\gamma\tanh(\sqrt{\nu^{-1}}\alpha)}$. For $\gamma = 0$, the two convergence factors are approximately 1.185 for DN$_2$ and ND$_3$, 0.844 for ND$_2$ and DN$_3$, and for $\gamma = 10$, we get 1.005 for DN$_2$ and ND$_3$, and 0.995 ND$_2$ and DN$_3$.

4.2. Convergence factor with $\theta = 1$ for an asymmetric decomposition. For $\theta = 1$, we show on the left in Figure 3 the convergence factors with interface at $\alpha = 0.3$ and no final target state (i.e., $\gamma = 0$), and on the right $\alpha = 0.7$ with a final

[1]This is a bit hard to see on the right in Figure 2, but zooming in confirms that the convergence factor of DN$_2$ is above 1, and below 1 for ND$_2$.

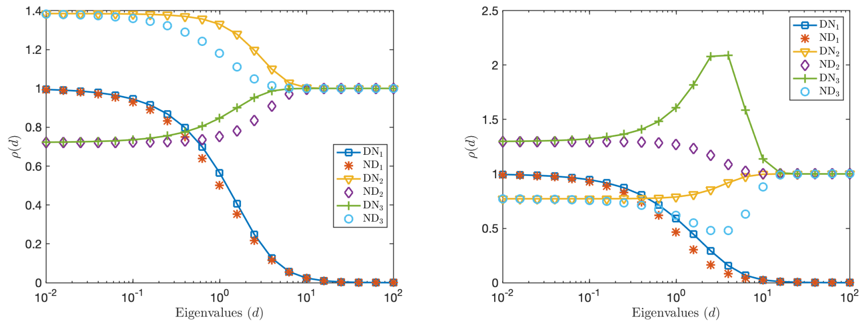

FIG. 3. *Convergence factor with* $\theta = 1$ *for an asymmetric decomposition of all six new algorithms as function of the eigenvalues* $d \in [10^{-2}, 10^{2}]$. *Left:* $\gamma = 0$ *and* $\alpha = 0.3$. *Right:* $\gamma = 10$ *and* $\alpha = 0.7$.

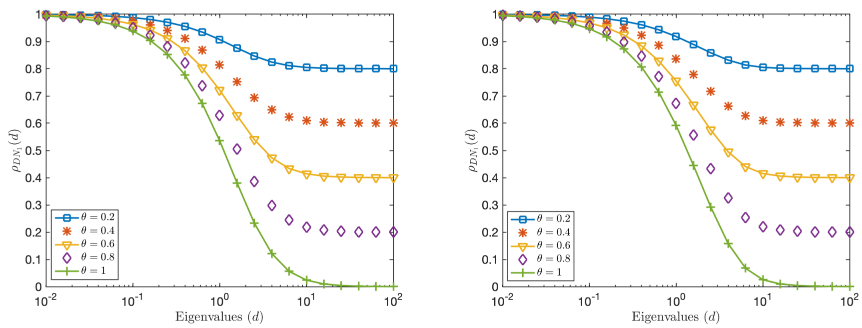

FIG. 4. *Convergence factor with different relaxation parameters of* DN_1 *as function of the eigenvalues* $d \in [10^{-2}, 10^{2}]$. *Left:* $\gamma = 0$ *and* $\alpha = 0.5$. *Right:* $\gamma = 10$ *and* $\alpha = 0.7$.

target state $\gamma = 10$. For DN_1 and ND_1, the convergence factor is similar in both cases, ND_1 being slightly better, and convergence is also similar to the symmetric case. This is because the convergence factor of the two algorithms for small and large eigenvalues is independent of the values of α, ν and γ. Their high frequency behavior is also much better compared to the other four algorithms in the two cases. For the other four algorithms, we see again the symmetry between DN_2 and ND_2, and DN_3 and ND_3. In general, DN_2 and ND_3 behave similarly, and also ND_2 and DN_3, but the influence of γ is more significant for DN_3 and ND_3 than DN_2 and ND_2. However their convergence factors all go to 1 for large eigenvalues, as for the symmetric decomposition. For the zero frequency, using the expressions (3.21) and (3.28) with $\theta = 1$, we obtain approximately 1.386 for DN_2 and ND_3, and 0.722 for ND_2 and DN_3 in the case $\gamma = 0$, $\alpha = 0.3$. For $\gamma = 10$, $\alpha = 0.7$, we get 0.771 for DN_2 and ND_3, and 1.296 for ND_2 and DN_3.

4.3. Convergence factor for Category I with different θ. Since DN_1 and ND_1 performed quite similarly, and much better than the others, we now investigate the dependence of DN_1 on θ in more detail. On the left in Figure 4 we show the convergence factor of DN_1 without final target state and a symmetric decomposition, and on the right with a final target state $\gamma = 10$ and an asymmetric decomposition. The convergence is very similar for these two settings, DN_1 is robust, and $\theta = 1$ gives the best performance.

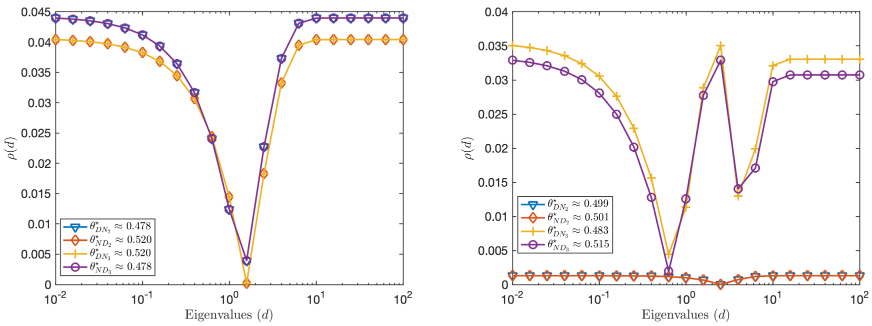

FIG. 5. *Convergence factor with $\theta^\star$ for a symmetric decomposition as function of the eigenvalues $d \in [10^{-2}, 10^2]$. Left: $\gamma = 0$. Right: $\gamma = 10$.*

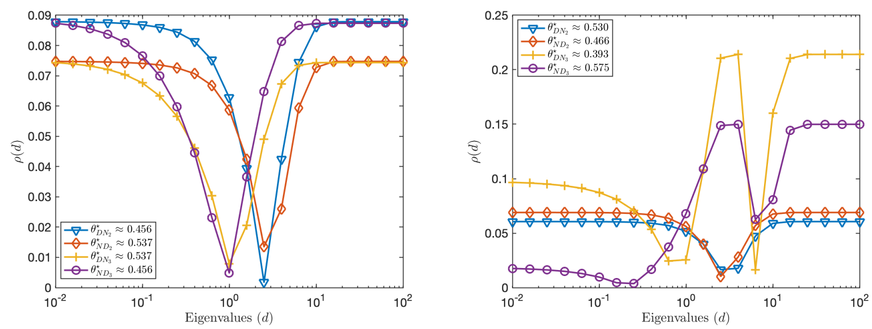

FIG. 6. *Convergence factor with $\theta^\star$ for an asymmetric decomposition as function of the eigenvalues $d \in [10^{-2}, 10^2]$. Left: $\gamma = 0$ and $\alpha = 0.3$. Right: $\gamma = 10$ and $\alpha = 0.7$.*

4.4. Convergence factor with optimal θ for a symmetric decomposition. Since the algorithms in Categories II and III are strongly related, we compare them now in Figure 5 for a symmetric decomposition using their optimal relaxation parameter $\theta^\star$, obtained numerically. On the left without final state, DN_2 and ND_3, and also ND_2 and DN_3, have the same convergence factor, and the optimal relaxation parameter satisfies $\theta^\star_{\mathrm{DN}_2} = \theta^\star_{\mathrm{ND}_3}$ and $\theta^\star_{\mathrm{ND}_2} = \theta^\star_{\mathrm{DN}_3}$ as proved in Theorem 3.24 and Theorem 3.27. These correspond to the value found using (3.22) and (3.29). In terms of the convergence speed, ND_2 and DN_3 are slightly better than DN_2 and ND_3. However, these similarities disappear when we add a final target state $\gamma = 10$. On the right in Figure 5, we see that now the convergence behavior of DN_2 and ND_2 is similar, and also DN_3 and ND_3 are rather similar, and DN_2 and ND_2 converge much faster compared to the others. We also see equioscillation between small and large eigenvalues. The theoretical results in (3.22) as well as in (3.29) still determine the optimal relaxation parameter $\theta^\star_{\mathrm{DN}_2}$ and $\theta^\star_{\mathrm{ND}_2}$ for DN_2 and ND_2, but not for DN_3 and ND_3, where we observe an equioscillation between small eigenvalues with some eigenvalues in the interval $[1, 10]$. Also ND_3 is slightly better than DN_3.

4.5. Convergence factor with optimal θ for an asymmetric decomposition. We show in Figure 6 the convergence factor with the optimal relaxation parameter $\theta^\star$ for the four algorithms in Categories II and III for an asymmetric decomposition. On the left with $\alpha = 0.3$ and no target state $\gamma = 0$ the convergence factors of the four algorithms are similar. This is consistent with the monotonicity we proved without final state. The optimal relaxation parameters satisfy $\theta^\star_{\mathrm{DN}_2} = \theta^\star_{\mathrm{ND}_3}$ and $\theta^\star_{\mathrm{ND}_2} = \theta^\star_{\mathrm{DN}_3}$, and we can use (3.22) and (3.29) to determine their values. Similar to the symmetric decomposition, ND_2 and DN_3 are slightly better than the others. However, these properties disappear again on the right in Figure 6 when there is a final state $\gamma = 10$. While DN_2 and ND_2 still equioscillate between the small and large eigenvalues, and the optimal relaxation parameter can be determined using (3.22) and (3.29), for DN_3 and ND_3 the equioscillation is between large eigenvalues and some eigenvalues in the interval $[1, 10]$. Hence, the optimal relaxation parameters for the algorithms DN_3 and ND_3 are different from DN_2 and ND_2. Also DN_2 and ND_2 converge much faster than the other two, and DN_2 is slightly faster than ND_2.

5. Conclusion. We introduced and analyzed six new time domain decomposition methods based on Dirichlet-Neumann and Neumann-Dirichlet techniques for parabolic optimal control problems. Our analysis shows that while at first sight it might be natural to preserve the forward-backward structure in the time subdomains as well, there are better choices that lead to substantially faster algorithms. We find that the algorithms in Categories II and III with optimized relaxation parameter are much faster than the algorithms in Category I, and they can still be identified to be of forward-backward structure using changes of variables. We also found many interesting mathematical connections between these algorithms. Algorithms in Category I are natural smoothers, while algorithms in Categories II and III with optimized relaxation parameter are highly efficient solvers.

Our study was restricted to the two subdomain case, but the algorithms can all naturally be written for many subdomains, and then one can also run them in parallel. They can also be used for more general parabolic constraints than the heat equation. Extensive numerical results will appear elsewhere.

REFERENCES

[1] R. A. Bartlett, M. Heinkenschloss, D. Ridzal, and B. G. van Bloemen Waanders, *Domain decomposition methods for advection dominated linear-quadratic elliptic optimal control problems*, Computer Methods in Applied Mechanics and Engineering, 195 (2006), pp. 6428–6447, https://doi.org/10.1016/j.cma.2006.01.009.
[2] J.-D. Benamou, *A domain decomposition method with coupled transmission conditions for the optimal control of systems governed by elliptic partial differential equations*, SIAM Journal on Numerical Analysis, 33 (1996), pp. 2401–2416, https://doi.org/10.1137/S0036142994267102.
[3] P. E. Bjørstad and O. B. Widlund, *Iterative methods for the solution of elliptic problems on regions partitioned into substructures*, SIAM Journal on Numerical Analysis, 23 (1986), pp. 1097–1120, https://doi.org/10.1137/0723075.
[4] A. Borzì and V. Schulz, *Computational Optimization of Systems Governed by Partial Differential Equations*, Society for Industrial and Applied Mathematics, 2011, https://doi.org/10.1137/1.9781611972054.
[5] H. Chang and D. Yang, *A Schwarz domain decomposition method with gradient projection for optimal control governed by elliptic partial differential equations*, Journal of Computational and Applied Mathematics, 235 (2011), pp. 5078–5094, https://doi.org/10.1016/j.cam.2011.04.037.
[6] G. Ciaramella, L. Halpern, and L. Mechelli, *Convergence analysis and optimization of a robin schwarz waveform relaxation method for time-periodic parabolic optimal control problems*, Journal of Computational Physics, 496 (2024), p. 112572, https://doi.org/10.1016/j.jcp.2023.112572.
[7] V. Dolean, P. Jolivet, and F. Nataf, *An Introduction to Domain Decomposition Methods: Algorithms, Theory, and Parallel Implementation*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2015, https://doi.org/10.1137/1.9781611974065.

[8] M. J. Gander and F. Kwok, *Schwarz methods for the time-parallel solution of parabolic control problems*, in Domain Decomposition Methods in Science and Engineering XXII, T. Dickopf, M. J. Gander, L. Halpern, R. Krause, and L. F. Pavarino, eds., Cham, 2016, Springer International Publishing, pp. 207–216, https://doi.org/10.1007/978-3-319-18827-0_19.
[9] M. J. Gander, F. Kwok, and B. C. Mandal, *Convergence of substructuring methods for elliptic optimal control problems*, in Domain Decomposition Methods in Science and Engineering XXIV, Cham, 2019, Springer International Publishing, pp. 291–300, https://doi.org/10.1007/978-3-319-93873-8_27.
[10] M. J. Gander, F. Kwok, and J. Salomon, *Paraopt: A parareal algorithm for optimality systems*, SIAM Journal on Scientific Computing, 42 (2020), pp. A2773–A2802, https://doi.org/10.1137/19M1292291.
[11] M. J. Gander, F. Kwok, and G. Wanner, *Constrained Optimization: From Lagrangian Mechanics to Optimal Control and PDE Constraints*, Springer International Publishing, Cham, 2014, pp. 151–202, https://doi.org/10.1007/978-3-319-08025-3_5.
[12] M. J. Gander and A. M. Stuart, *Space-time continuous analysis of waveform relaxation for the heat equation*, SIAM Journal on Scientific Computing, 19 (1998), pp. 2014–2031, https://doi.org/10.1137/S1064827596305337.
[13] M. Heinkenschloss, *A time-domain decomposition iterative method for the solution of distributed linear quadratic optimal control problems*, Journal of Computational and Applied Mathematics, 173 (2005), pp. 169–198, https://doi.org/10.1016/j.cam.2004.03.005.
[14] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints*, Springer Dordrecht, 2009, https://doi.org/10.1007/978-1-4020-8839-1.
[15] F. Kwok, *On the time-domain decomposition of parabolic optimal control problems*, in Domain Decomposition Methods in Science and Engineering XXIII, C.-O. Lee, X.-C. Cai, D. E. Keyes, H. H. Kim, A. Klawonn, E.-J. Park, and O. B. Widlund, eds., Cham, 2017, Springer International Publishing, pp. 55–67, https://doi.org/10.1007/978-3-319-52389-7_5.
[16] U. Langer, O. Steinbach, F. Tröltzsch, and H. Yang, *Space-time finite element discretization of parabolic optimal control problems with energy regularization*, SIAM Journal on Numerical Analysis, 59 (2021), pp. 675–695, https://doi.org/10.1137/20M1332980.
[17] E. Lelarasmee, A. Ruehli, and A. Sangiovanni-Vincentelli, *The waveform relaxation method for time-domain analysis of large scale integrated circuits*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 1 (1982), pp. 131–145, https://doi.org/10.1109/TCAD.1982.1270004.
[18] J.-L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, 170, Springer-Verlag Berlin Heidelberg, 1 ed., 1971.
[19] J.-L. Lions, Y. Maday, and G. Turinici, *A parareal in time procedure for the control of partial differential equations*, Comptes Rendus Mathematique, 335 (2002), pp. 387–392, https://doi.org/10.1016/S1631-073X(02)02467-6.
[20] C. Lubich and A. Ostermann, *Multi-grid dynamic iteration for parabolic equations*, BIT Numerical Mathematics, 27 (1987), pp. 216–234, https://doi.org/10.1007/BF01934186.
[21] B. C. Mandal, *Substructuring waveform relaxation methods for parabolic optimal control problems*, in Soft Computing for Problem Solving, J. C. Bansal, K. N. Das, A. Nagar, K. Deep, and A. K. Ojha, eds., Singapore, 2019, Springer Singapore, pp. 485–494, https://doi.org/10.1007/978-981-13-1595-4_39.
[22] W. L. Miranker and W. Liniger, *Parallel methods for the numerical integration of ordinary differential equations*, Mathematics of Computation, 21 (1967), pp. 303–320, https://doi.org/10.2307/2003233.
[23] M. Neumüller and O. Steinbach, *Regularization error estimates for distributed control problems in energy spaces*, Mathematical Methods in the Applied Sciences, 44 (2021), pp. 4176–4191, https://doi.org/10.1002/mma.7021.
[24] J. Nievergelt, *Parallel methods for integrating ordinary differential equations*, Commun. ACM, 7 (1964), p. 731–733, https://doi.org/10.1145/355588.365137.
[25] A. Toselli and O. B. Widlund, *Domain Decomposition Methods - Algorithms and Theory*, Springer Berlin, Heidelberg, 1 ed., 2005, https://doi.org/10.1007/b137868.
[26] F. Tröltzsch, *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*, vol. 112, Graduate Studies in Mathematics, 2010, https://doi.org/10.1090/gsm/112.
[27] S. Vandewalle and E. Velde, *Space-time concurrent multigrid waveform relaxation*, Annals of Numerical Mathematics, 1-4 (1994), pp. 347–363, https://doi.org/10.13140/2.1.1146.1761.

Appendix A. Convergence analysis using $\mu_{(i)}$.

We can also use formulation (2.7) to analyze the convergence behavior of the algorithm DN_1 (3.1)-(3.2), we then need to study

$$\left\{\begin{aligned} \ddot\mu_{1,(i)}^k-\sigma_i^2\mu_{1,(i)}^k&=0 \text{ in } \Omega_1,\\ \dot\mu_{(i)}(0)-d_i\mu_{(i)}(0)&=0,\\ \mu_{1,(i)}^k(\alpha)&=f_{\alpha,(i)}^{k-1}, \end{aligned}\right. \qquad \left\{\begin{aligned} \ddot\mu_{2,(i)}^k-\sigma_i^2\mu_{2,(i)}^k&=0 \text{ in } \Omega_2,\\ \ddot\mu_{2,(i)}^k(\alpha)-d_i\dot\mu_{2,(i)}^k(\alpha)&=\ddot\mu_{1,(i)}^k(\alpha)-d_i\dot\mu_{1,(i)}^k(\alpha),\\ \gamma\dot\mu_{(i)}(T)+\beta_i\mu_{(i)}(T)&=0, \end{aligned}\right. \tag{A.1}$$

with the update of the transmission condition

$$f_{\alpha,(i)}^k=(1-\theta)f_{\alpha,(i)}^{k-1}+\theta\mu_{2,(i)}^k(\alpha)\quad \theta\in(0,1). \tag{A.2}$$

This is a DR type algorithm applied to solve (2.7). Using (3.12), we determine the two coefficients A_i^k and B_i^k from the transmission condition from (A.1). Using then relation (A.2), we find

$$f_{\alpha,(i)}^k=(1-\theta)f_{\alpha,(i)}^{k-1}+\theta\nu^{-1}f_{\alpha,(i)}^{k-1}\frac{\gamma\sigma_i+\beta_i\tanh(b_i)}{\big(\sigma_i+d_i\tanh(a_i)\big)\big(\omega_i+\sigma_i\tanh(b_i)\big)},$$

which is exactly the same convergence factor as (3.6).

Appendix B. 1D Advection-diffusion problems. We can also consider the operator $\partial_x-\kappa\partial_{xx}$, and use a finite difference scheme to discretize it, for instance, an upwind discretization for the advection part ∂_x and the standard centred discretization for the diffusion part ∂_{xx}. With mesh size h, the eigenfunctions in this case are $e^{in\pi jh}$ with eigenvalues $d_n:=2(\frac{1}{h}+\kappa\frac{2}{h^2})\sin^2(\frac{n\pi h}{2})+i\frac{1}{h}\sin(n\pi h)$. As presented in Section 4, we can then check the convergence behavior of the proposed algorithms for advection-diffusion problems. As an example, we keep the same setting as for Figure 5, but now use the eigenvalues from above. We show in Figure 7 the convergence factor with respect to the eigenvalues for diffusion coefficient $\kappa=10^{-1}$ and $\kappa=10^{-2}$. Comparing with the pure diffusion case in Figure 5, we see that adding an advection term leads to slower convergence, while the order from best to worst algorithm is maintained as for pure diffusion, both for $\gamma=0$ (left) and $\gamma=10$ (right). For $\gamma=10$, the slower algorithm variants even tend to stagnate as the problem becomes advection dominant, but the fast algorithms remain fast in that case, see Figure 5 (right). We also see that the optimized relaxation parameters depend on the presence of advection.

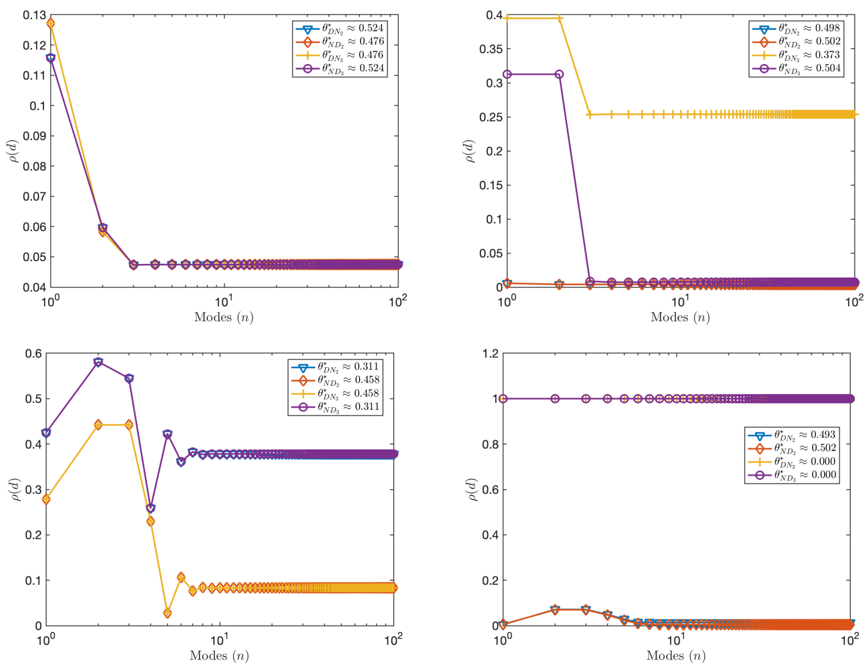

FIG. 7. *Convergence factor with $\theta^\star$ for a symmetric decomposition as function of the eigenvalues $d_n = 2(\frac{1}{h} + \kappa\frac{2}{h^2})\sin^2(\frac{n\pi h}{2}) + i\frac{1}{h}\sin(n\pi h)$, $n \in [10^0, 10^2]$. Top: $\kappa = 10^{-1}$. Bottom: $\kappa = 10^{-2}$. Left: $\gamma = 0$. Right: $\gamma = 10$.*