Non-overlapping Schwarz methods in time for parabolic optimal control problems

Martin J. Gander and Liu-Di Lu

1 Introduction

The classical Schwarz method, originally introduced by Hermann Amandus Schwarz to prove existence and uniqueness of solutions to Laplace's equation [6], has since then been extensively studied and applied to a wide range of problems. A historical review can be found in [1]. It is well known that the method fails to converge when applied to non-overlapping spatial subdomains due to the repeated passing of identical "Dirichlet data" from one subdomain to the other. Several modified methods have then been proposed to address this issue, notably by Lions [5]. More recently, Schwarz methods in time have been proposed for the time-parallel solution of parabolic optimal control problems, as discussed in [2], and it was noted:

> We present a rigorous convergence analysis for the case of two subdomains, which shows that the classical Schwarz method converges, even without overlap! Reformulating the algorithm reveals that this is because imposing initial conditions for y and final conditions on λ is equivalent to using Robin transmission conditions between time subdomains for y.

To gain deeper insight into the convergence of the classical Schwarz method applied to parabolic optimal control problems with a non-overlapping time domain decomposition, we study the following model problem: for a given desired state $\hat{\boldsymbol{y}}(t)$ and parameters $\gamma, \nu > 0$, we want to solve the minimization problem

$$\begin{aligned} \min_{\boldsymbol{y},u} \frac{1}{2}\int_0^T \|\boldsymbol{y}-\hat{\boldsymbol{y}}\|^2 \mathrm{d}t + \frac{\gamma}{2}\|\boldsymbol{y}(T)-\hat{\boldsymbol{y}}(T)\|^2 + \frac{\nu}{2}\int_0^T \|u\|^2 \mathrm{d}t, \\ \text{subject to} \quad \dot{\boldsymbol{y}} + A\boldsymbol{y} = \mathbf{u}, \quad \boldsymbol{y}(0) = \boldsymbol{y}_0, \end{aligned} \tag{1}$$

Martin J. Gander
University of Geneva, Switzerland, e-mail: martin.gander@unige.ch

Liu-Di Lu
University of Geneva, Switzerland, e-mail: liudi.lu@unige.ch

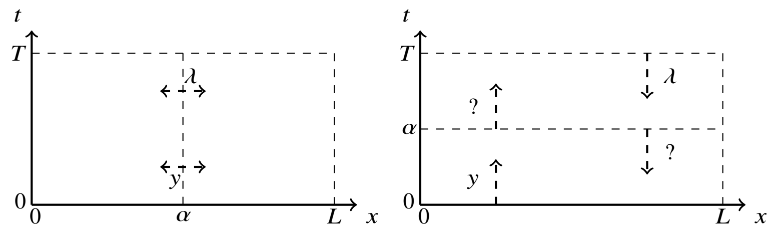

Fig. 1 One dimensional illustration of decomposing in space (left) and decomposing in time (right).

where $\dot{\boldsymbol{y}} + A\boldsymbol{y} = \mathbf{u}$ represents the semi-discretization of a parabolic partial differential equation (PDE) of the form $\partial_t y + \mathcal{L} y = u$. Here, $\mathbf{u}$ is the control variable, and $\boldsymbol{y}_0$ denotes the initial condition. By applying the Lagrange multiplier approach and eliminating the control variable $\mathbf{u}$, we derive from (1) the first-order optimality system (see, e.g., [3, Section 2])

$$\begin{cases} \begin{pmatrix} \dot{\boldsymbol{y}} \\ \dot{\boldsymbol{\lambda}} \end{pmatrix} + \begin{pmatrix} A & -\nu^{-1} I \\ -I & -A^T \end{pmatrix} \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} 0 \\ -\hat{\boldsymbol{y}} \end{pmatrix} \quad \text{in } (0,T), \\ \boldsymbol{y}(0) = \boldsymbol{y}_0, \\ \boldsymbol{\lambda}(T) + \gamma \boldsymbol{y}(T) = \gamma \hat{\boldsymbol{y}}(T), \end{cases} \tag{2}$$

where $\boldsymbol{\lambda}$ is the adjoint state. The system described by (2) is a forward-backward system of ordinary differential equations (ODEs), in which the state $\boldsymbol{y}$ propagates forward in time starting from an initial condition, while the adjoint state $\boldsymbol{\lambda}$ propagates backward in time with a final condition.

To investigate the application of a non-overlapping classical Schwarz method in time to solve this system, we decompose the time interval $(0,T)$ into two non-overlapping subdomains: $I_1 := (0,\alpha)$ and $I_2 := (\alpha,T)$, where α represents the interface. In Section 2, we first introduce four variants of the classical Schwarz algorithm and analyze their convergence behavior. In Section 3, we replace the Dirichlet transmission condition used in the classical Schwarz algorithm with a Neumann transmission condition and study the resulting convergence properties. Finally, we discuss our results in Section 4 and conclude with some comments.

2 Dirichlet transmission conditions

When decomposing in space, the standard way is to pass the values for the state y and the adjoint state λ from one subdomain to its neighbor, as illustrated in Figure 1 on the left. However, this becomes much more tricky when decomposing in time as shown in Figure 1 on the right. Since the system (2) is a forward-backward system, it initially seems natural to preserve this property in the decomposed case, and to transmit in I_1 a final condition for the adjoint state λ, while an initial condition for y

Table 1 Four non-overlapping Schwarz variants in time. Left: Dirichlet transmission conditions. Right: Neumann transmission conditions

name	SD_1	SD_2	SD_3	SD_4	SN_1	SN_2	SN_3	SN_4
$I_1 = (0, \alpha)$	$\boldsymbol{\lambda}$	$\boldsymbol{y}$	$\boldsymbol{y}$	$\boldsymbol{\lambda}$	$\dot{\boldsymbol{\lambda}}$	$\dot{\boldsymbol{y}}$	$\dot{\boldsymbol{y}}$	$\dot{\boldsymbol{\lambda}}$
$I_2 = (\alpha, T)$	$\boldsymbol{y}$	$\boldsymbol{\lambda}$	$\boldsymbol{y}$	$\boldsymbol{\lambda}$	$\dot{\boldsymbol{y}}$	$\dot{\boldsymbol{\lambda}}$	$\dot{\boldsymbol{y}}$	$\dot{\boldsymbol{\lambda}}$

is already present. Similarly, in I_2, where a final condition for λ is already present, it is natural to transmit an initial condition for the state y. A natural Schwarz algorithm in time hence solves for the iteration index $k = 1, 2, \ldots$

$$\begin{cases} \begin{pmatrix} \dot{\boldsymbol{y}}_1^k \\ \dot{\boldsymbol{\lambda}}_1^k \end{pmatrix} + \begin{pmatrix} A & -\nu^{-1} I \\ -I & -A^T \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_1^k \\ \boldsymbol{\lambda}_1^k \end{pmatrix} = \begin{pmatrix} 0 \\ -\hat{\boldsymbol{y}} \end{pmatrix} & \text{in } (0, \alpha), \\ \boldsymbol{y}_1^k(0) = \boldsymbol{y}_0, \\ \boldsymbol{\lambda}_1^k(\alpha) = \boldsymbol{\lambda}_2^{k-1}(\alpha), \end{cases}$$
$$\begin{cases} \begin{pmatrix} \dot{\boldsymbol{y}}_2^k \\ \dot{\boldsymbol{\lambda}}_2^k \end{pmatrix} + \begin{pmatrix} A & -\nu^{-1} I \\ -I & -A^T \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_2^k \\ \boldsymbol{\lambda}_2^k \end{pmatrix} = \begin{pmatrix} 0 \\ -\hat{\boldsymbol{y}} \end{pmatrix} & \text{in } (\alpha, T), \\ \boldsymbol{y}_2^k(\alpha) = \boldsymbol{y}_1^k(\alpha), \\ \boldsymbol{\lambda}_2^k(T) + \gamma \boldsymbol{y}_2^k(T) = \gamma \hat{\boldsymbol{y}}(T). \end{cases} \tag{3}$$

Here, $\boldsymbol{y}_j^k$ and $\boldsymbol{\lambda}_j^k$ represent the restriction of $\boldsymbol{y}^k$ and $\boldsymbol{\lambda}^k$ to the time subdomain I_j, $j = 1, 2$. The parallel version of this natural Schwarz algorithm (3) coincides with the optimized Schwarz algorithm (3a)-(3b) in [2], under the conditions $p = q = 0$ and $\alpha = \beta$ there.

Although algorithm (3) preserves the forward-backward structure of the original system (2), studies in [3, 4] have shown that this structure is less important for the convergence behavior of Dirichlet–Neumann and Neumann–Neumann algorithms with time domain decomposition. Moreover, the forward-backward structure can always be recovered by using the linear system in (2), that is

$$\boldsymbol{\lambda} = \nu(\dot{\boldsymbol{y}} + A\boldsymbol{y}), \quad \boldsymbol{y} = \dot{\boldsymbol{\lambda}} - A^T \boldsymbol{\lambda} + \hat{\boldsymbol{y}}. \tag{4}$$

The two identities (4) also transform a Dirichlet transmission condition for one state into a particular Robin type transmission condition for the other state. We can therefore identify four variants of the classical Schwarz method applied to (2) with time domain decomposition, as summarized in Table 1 (left). We call these variants SD_1 to SD_4 (D for Dirichlet) in the first row. The second row (resp. third row) shows the transmission condition at the interface of I_1 (resp. I_2). All four variants use Dirichlet transmission conditions at the interface, but we can use (4) to recover the forward-backward structure for SD_2, SD_3, and SD_4 as explained above.

We now analyze the convergence of these four variants. For simplicity, we assume that A is symmetric, i.e., $A = A^T \in \mathbb{R}^{n\times n}$. This allows us to apply a diagonalization, which leads to n independent 2×2 reduced systems of ODEs. For the algorithm SD_1, this transforms (3) to,

$$
\begin{cases}
\begin{pmatrix}\dot{z}_{1,i}^k\\ \dot{\mu}_{1,i}^k\end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1}\\ -1 & -d_i\end{pmatrix}\begin{pmatrix} z_{1,i}^k\\ \mu_{1,i}^k\end{pmatrix} = \begin{pmatrix}0\\ -\hat{z}_i\end{pmatrix} & \text{in } (0,\alpha),\\
z_{1,i}^k(0) = z_{0,i}, \\
\mu_{1,i}^k(\alpha) = \mu_{2,i}^{k-1}(\alpha),
\end{cases}
$$
$$
\begin{cases}
\begin{pmatrix}\dot{z}_{2,i}^k\\ \dot{\mu}_{2,i}^k\end{pmatrix} + \begin{pmatrix} d_i & -\nu^{-1}\\ -1 & -d_i\end{pmatrix}\begin{pmatrix} z_{2,i}^k\\ \mu_{2,i}^k\end{pmatrix} = \begin{pmatrix}0\\ -\hat{z}_i\end{pmatrix} & \text{in } (\alpha,T),\\
z_{2,i}^k(\alpha) = z_{1,i}^k(\alpha), \\
\mu_{2,i}^k(T) + \gamma z_{2,i}^k(T) = \gamma \hat{z}_i(T),
\end{cases}
\tag{5}
$$

where $\boldsymbol{z}_j^k := P^{-1}\boldsymbol{y}_j^k$, $\hat{\boldsymbol{z}} := P^{-1}\hat{\boldsymbol{y}}$, $\boldsymbol{\mu}_j^k := P^{-1}\boldsymbol{\lambda}_j^k$ and $A = PDP^{-1}$ with $D := \operatorname{diag}(d_1,\dots,d_n)$ the eigenvalues of A. Furthermore, $z_{j,i}^k, \hat{z}_i$, and $\mu_{j,i}^k$ denote the ith components of the vectors $\boldsymbol{z}_j^k, \hat{\boldsymbol{z}}$, and $\boldsymbol{\mu}_j^k$. Note that the assumption of A being symmetric is only a theoretical tool for the convergence analysis and is not required to run these algorithms in practice.

To analyze the convergence behavior of the algorithm SD_1, we solve analytically (5) by eliminating one variable to obtain a second-order ODE. If we choose to eliminate the adjoint state $\mu_{j,i}^k$ and use the first identity in (4), the Dirichlet transmission condition: $\mu_{1,i}^k(\alpha) = \mu_{2,i}^{k-1}(\alpha)$ will be transformed into a Robin transmission condition: $\nu(\dot{z}_{1,i}^k + d_i z_{1,i}^k) = \nu(\dot{z}_{1,i}^{k-1} + d_i z_{1,i}^{k-1})$. Note that although the natural classical Schwarz algorithm (3) uses Dirichlet transmission conditions at the interface, the convergence analysis actually evaluates a Robin–Dirichlet type algorithm. This transformation has also been observed in [2]. Solving the resulting second-order ODE allows us to determine the convergence factor for SD_1 as

$$
\rho_{\text{SD}_1} := \max_{d_i\in D}\left|\frac{1+\gamma(\sigma_i\coth(b_i) - d_i)}{\nu(\sigma_i\coth(a_i)+d_i)(\sigma_i\coth(b_i)+d_i+\gamma\nu^{-1})}\right|, \tag{6}
$$

where $\sigma_i := \sqrt{d_i^2+\nu^{-1}}$, $a_i = \sigma_i\alpha$ and $b_i = \sigma_i(T-\alpha)$.

Remark 1 Alternatively, one can eliminate the state $z_{j,i}^k$ using the second identity in (4), which results in solving a second-order ODE for $\mu_{j,i}^k$. This approach leads to a Dirichlet–Robin type algorithm instead, but with the same convergence factor (6), as observed also for Dirichlet–Neumann time decomposition methods, see [3, Appendix A].

To better understand the convergence of the algorithm SD_1, we now study the convergence factor (6) in detail. We can first remove the absolute value, since the denominator is positive and $\sigma_i\coth(b_i) > \sigma_i > d_i$. Next, for a given eigenvalue d_i, we have

$$
\begin{aligned}
&1+\gamma(\sigma_i\coth(b_i)-d_i) - \nu(\sigma_i\coth(a_i)+d_i)(\sigma_i\coth(b_i)+d_i+\gamma\nu^{-1})\\
=&-\nu d_i^2(\coth(a_i)\coth(b_i)+1) - (\coth(a_i)\coth(b_i)-1) - 2\gamma d_i\\
&-\nu\sigma_i d_i(\coth(a_i)+\coth(b_i)) + \gamma\sigma_i(\coth(b_i)-\coth(a_i)).
\end{aligned}
\tag{7}
$$

If $d_i \geq 0$ and $a_i \leq b_i$, then the latter expression is negative, which implies $\rho_{\mathrm{SD}_1} < 1$. Hence, we obtain the following result.

Theorem 1 *Assume that A is symmetric positive semi-definite (i.e., $d_i \geq 0$). Then the Schwarz algorithm (3) converges for all initial guesses if (i) $\alpha \leq \frac{T}{2}$, or (ii) $\gamma = 0$.*

The assumption on the matrix A is natural, for instance, if A is the finite-difference discretization of the Laplace operator $-\Delta$. Additionally, setting $\gamma = 0$ implies that we are not considering the final target in (1). In this case, the convergence factor reads $\frac{1}{\nu(\sigma_i \coth(a_i)+d_i)(\sigma_i \coth(b_i)+d_i)}$. Taking the derivative with respect to d_i, we find $-\frac{\sigma_i(\coth(a_i)+\coth(b_i))+2d_i}{\nu(\sigma_i \coth(a_i)+d_i)^2(\sigma_i \coth(b_i)+d_i)^2} - \frac{d_i(\cosh(a_i)\sinh(a_i)-a_i)}{\nu\sigma_i \sinh^2(a_i)(\sigma_i \coth(a_i)+d_i)^2(\sigma_i \coth(b_i)+d_i)} - \frac{d_i(\cosh(b_i)\sinh(b_i)-b_i)}{\nu\sigma_i \sinh^2(b_i)(\sigma_i \coth(a_i)+d_i)(\sigma_i \coth(b_i)+d_i)^2}$. This derivative is negative if $d_i \geq 0$, since $\cosh(x)\sinh(x) \geq x$, $\forall x \in \mathbb{R}$. Therefore, we can bound the convergence factor and find the following result.

Theorem 2 *If A is symmetric positive semi-definite and $\gamma = 0$, we obtain the estimate $\rho_{SD_1} \leq \frac{1}{\nu(\sigma_{\min}\coth(\sigma_{\min}\alpha)+d_{\min})(\sigma_{\min}\coth(\sigma_{\min}(T-\alpha))+d_{\min})} < \frac{1}{\nu(\sigma_{\min}+d_{\min})^2}$, where $d_{\min}$ denotes the smallest eigenvalue of A and $\sigma_{\min} = \sqrt{d_{\min}^2 + \nu^{-1}}$.*

Studying the general form of (6), we observe that, for large eigenvalues d_i, the convergence factor approximates $\frac{1+\gamma(\sigma_i \coth(b_i)-d_i)}{\nu(\sigma_i \coth(a_i)+d_i)(\sigma_i \coth(b_i)+d_i+\gamma\nu^{-1})} \sim_\infty \frac{1}{4\nu d_i^2}$, implying that high-frequency components converge very fast. For $d_i = 0$, the convergence factor becomes $\rho_{\mathrm{SD}_1}|_{d_i=0} = \tanh(\sqrt{\nu^{-1}}\alpha)\frac{\gamma\sqrt{\nu^{-1}}\coth(\sqrt{\nu^{-1}}(T-\alpha))+1}{\coth(\sqrt{\nu^{-1}}(T-\alpha))+\gamma\sqrt{\nu^{-1}}}$, which is close to 1, especially when the control penalization parameter ν is small or $\gamma = 0$. Hence, low-frequency components converge very slowly. Based on the monotonicity of ρ_{SD_1} with respect to d_i when $\gamma = 0$, we can improve the convergence by introducing a relaxation parameter θ in the transmission condition to balance the convergence rates of low and high frequencies. For instance, we can replace the transmission condition on I_1 in (3) by $\lambda_1^k(\alpha) = f_\alpha^{k-1}$ with $f_\alpha^k := (1-\theta)f_\alpha^{k-1} + \theta\lambda_2^{k-1}(\alpha)$, and $\theta \in (0,1)$. The resulting convergence factor is $\rho_{\mathrm{SD}_1}^\theta := \max_{d_i \in D} |1-\theta(1+\frac{1+\gamma(\sigma_i \coth(b_i)-d_i)}{\nu(\sigma_i \coth(a_i)+d_i)(\sigma_i \coth(b_i)+d_i+\gamma\nu^{-1})})|$. Equioscillating between small and large eigenvalues, we determine the optimal relaxation parameter $\theta_{\mathrm{SD}_1}^* := \frac{2}{2+\tanh(\sqrt{\nu^{-1}}\alpha)\frac{\gamma\sqrt{\nu^{-1}}\coth(\sqrt{\nu^{-1}}(T-\alpha))+1}{\coth(\sqrt{\nu^{-1}}(T-\alpha))+\gamma\sqrt{\nu^{-1}}}}$. When $\gamma = 0$, the optimal relaxation parameter simplifies to $\frac{2}{2+\tanh(\sqrt{\nu^{-1}}\alpha)\tanh(\sqrt{\nu^{-1}}(T-\alpha))}$, which is approximately $\frac{2}{3}$.

For the algorithm SD_2, we reverse the transmission conditions $\mu_{1,i}^k(\alpha) = \mu_{2,i}^{k-1}(\alpha)$ and $z_{2,i}^k(\alpha) = z_{1,i}^k(\alpha)$ in (3), hence also in (5). We thus obtain for SD_2 the convergence factor $\rho_{\mathrm{SD}_2} := \max_{d_i \in D} |\frac{\nu(\sigma_i \coth(a_i)+d_i)(\sigma_i \coth(b_i)+d_i+\gamma\nu^{-1})}{1+\gamma(\sigma_i \coth(b_i)-d_i)}|$, which is the inverse of ρ_{SD_1}. Hence algorithm SD_2 diverges under the assumption of Theorem 1, and in particular, it diverges violently for high-frequency components, because $\frac{\nu(\sigma_i \coth(a_i)+d_i)(\sigma_i \coth(b_i)+d_i+\gamma\nu^{-1})}{1+\gamma(\sigma_i \coth(b_i)-d_i)} \sim_\infty 4\nu d_i^2$. For low-frequency components, the converge is poor, especially when ν is small, or it can even diverge when $\gamma = 0$, as

$\rho_{\mathrm{SD}_2}|_{d_i=0} = \coth(\sqrt{\nu^{-1}}\alpha)\frac{\coth(\sqrt{\nu^{-1}}(T-\alpha))+\gamma\sqrt{\nu^{-1}}}{\gamma\sqrt{\nu^{-1}}\coth(\sqrt{\nu^{-1}}(T-\alpha))+1}$. Based on these observations, algorithm SD_2 is not an efficient algorithm and can also not be improved with relaxation techniques.

We now study Algorithms SD_3 and SD_4. Since they pass Dirichlet data at the interface using only one state, they have similar behavior, and we just present the analysis for SD_3. We replace the transmission condition $\mu_{1,j}^k(\alpha) = \mu_{2,j}^{k-1}(\alpha)$ on I_1 with $z_{1,j}^k(\alpha) = z_{2,j}^{k-1}(\alpha)$ in (5). Using the second-order ODE for $z_{j,i}^k$, we find that the transmission conditions still remain, i.e., $z_{1,j}^k(\alpha) = z_{2,j}^{k-1}(\alpha)$ on I_1 and $z_{2,j}^k(\alpha) = z_{1,j}^k(\alpha)$ on I_2. This indicates that this is a Dirichlet–Dirichlet type algorithm in the classical Schwarz sense, thus suffering from the same non-convergence as the classical Schwarz algorithm without overlap. Indeed, its convergence factor ρ_{SD_3} equals 1 for all eigenvalues d_i, and this cannot be improved with relaxation. Similarly, using the second-order ODE for $\mu_{j,i}^k$, we obtain also $\rho_{\mathrm{SD}_4} = 1$. Hence, in contrast to the Dirichlet–Neumann algorithms in time [3], among all four Schwarz variants with Dirichlet transmission conditions, only algorithm SD_1 (3), which naturally preserves the forward-backward structure, exhibits good convergence behavior.

3 Neumann transmission condition

Schwarz methods with Neumann transmission conditions are not used for elliptic problems, since they are not convergent in general, as one can see from a simple 1D example. We investigate now if Schwarz methods in time for parabolic optimal control problems can be promising solvers. We identity once again four variants, see Table 1 (right), called SN_1 to SN_4. Similar to SD_1, algorithm SN_1 naturally retains the forward-backward structure. For the other three variants, this structure can be recovered using identities in (4).

To analyze the convergence behavior of the four variants, we follow the same approach as in Section 2. Algorithm SN_1 is similar to (3), but with transmission conditions replaced by $\dot{\boldsymbol{\lambda}}_{1,i}^k(\alpha) = \dot{\boldsymbol{\lambda}}_{2,i}^{k-1}(\alpha)$ on I_1 and $\dot{\boldsymbol{y}}_{2,i}^k(\alpha) = \dot{\boldsymbol{y}}_{1,i}^k(\alpha)$ on I_2. When analyzing its convergence using the second-order ODE for $z_{i,j}^k$, we are then examining a Robin–Neumann type algorithm. We find for SN_1 the convergence factor

$$\rho_{\mathrm{SN}_1} := \max_{d_i \in D} \left| \frac{1 + \gamma(\sigma_i \tanh(b_i) - d_i)}{\nu(\sigma_i \tanh(a_i) + d_i)(\sigma_i \tanh(b_i) + d_i + \gamma\nu^{-1})} \right|, \tag{8}$$

similar to ρ_{SD_1}, with hyperbolic cotangent functions in (6) replaced by hyperbolic tangent functions. However, unlike Theorem 1, we cannot directly obtain a similar result for SN_1, since the sign of (7) is less clear when replacing hyperbolic cotangent by hyperbolic tangent. Nevertheless, substituting $\gamma = 0$ into (8) yields $\frac{1}{\nu(\sigma_i \tanh(a_i)+d_i)(\sigma_i \tanh(b_i)+d_i)}$, which is a decreasing function of d_i, as σ_i and the hyperbolic tangent are both increasing functions of d_i when $d_i \geq 0$. Therefore, we obtain a similar result as Theorem 2.

Theorem 3 *If A is symmetric positive semi-definite and $\gamma = 0$, we obtain the estimate* $\rho_{SN_1} \leq \frac{1}{\nu(\sigma_{\min}\tanh(\sigma_{\min}\alpha)+d_{\min})(\sigma_{\min}\tanh(\sigma_{\min}(T-\alpha))+d_{\min})}$.

Moreover, for large eigenvalues, the convergence factor for SN_1 is approximately $\frac{1+\gamma(\sigma_i\tanh(b_i)-d_i)}{\nu(\sigma_i\tanh(a_i)+d_i)(\sigma_i\tanh(b_i)+d_i+\gamma\nu^{-1})} \sim_\infty \frac{1}{4\nu d_i^2}$, meaning that SN_1 is a very good smoother for high-frequency components. For a zero eigenvalue $d_i = 0$, $\rho_{\mathrm{SN}_1}|_{d_i=0} = \coth(\sqrt{\nu^{-1}}\alpha)\frac{\gamma\sqrt{\nu^{-1}}\tanh(\sqrt{\nu^{-1}}(T-\alpha))+1}{\tanh(\sqrt{\nu^{-1}}(T-\alpha))+\gamma\sqrt{\nu^{-1}}}$. Thus, low-frequency components converge very slowly, especially when ν is small, or it can diverge when $\gamma = 0$. As with SD_1, one can use the monotonicity of ρ_{SN_1} in the case $\gamma = 0$ and improve the convergence with a relaxation parameter θ. The convergence factor with relaxation is $\rho^\theta_{\mathrm{SN}_1} := \max_{d_i\in D}|1-\theta(\frac{1+\gamma(\sigma_i\tanh(b_i)-d_i)}{\nu(\sigma_i\tanh(a_i)+d_i)(\sigma_i\tanh(b_i)+d_i+\gamma\nu^{-1})})|$. Using the equioscillation principle, we determine the optimal relaxation parameter $\theta^*_{\mathrm{SN}_1} := \frac{2}{2+\coth(\sqrt{\nu^{-1}}\alpha)\frac{\gamma\sqrt{\nu^{-1}}\tanh(\sqrt{\nu^{-1}}(T-\alpha))+1}{\tanh(\sqrt{\nu^{-1}}(T-\alpha))+\gamma\sqrt{\nu^{-1}}}}$. For $\gamma = 0$, this becomes $\frac{2}{2+\coth(\sqrt{\nu^{-1}}\alpha)\coth(\sqrt{\nu^{-1}}(T-\alpha))}$ and is also bounded by $\frac{2}{3}$.

For algorithm SN_2, we reverse the transmission condition in SN_1 and obtain again the inverse of the convergence factor of SN_1 in (8). As for SD_2, SN_2 is hence not an efficient algorithm and cannot be improved using relaxation. For algorithms SN_3 and SN_4, they pass Neumann data at the interface using only one state. Similarly as for SD_3 and SD_4, we find that $\rho_{\mathrm{SN}_3} = \rho_{\mathrm{SN}_4} = 1$ for all eigenvalues d_i, indicating stagnation and no improvement with relaxation. Hence, among the four variants with Neumann transmission conditions, only algorithm SN_1, which naturally preserves the forward-backward structure, has good convergence behavior, and this even though it is a Schwarz method with Neumann transmission conditions, which do not work in the elliptic case!

4 Numerical experiments and comments

We first plot the convergence factor ρ as a function of the eigenvalues d_i in Figure 4 (left). We set the parameters $\nu = 0.1, \gamma = 10, T = 1$ and $\alpha = 0.4$, and observe that both SD_2 and SN_2 diverge for $d_i \geq 1$. Algorithms SD_1 and SN_1 are two good smoothers for high-frequency components, but they exhibit poor convergence for low-frequency components. This is significantly improved when using relaxation techniques. We find numerically $\theta^*_{\mathrm{SD}_1} \approx 0.692$ and $\theta^*_{\mathrm{SN}_1} \approx 0.640$, which are consistent with their theoretical values. To evaluate the numerical performance of these variants, we apply them to solve the one-dimensional heat control problem $\partial_t y - \partial_{xx} y = u$ with homogenous Dirichlet boundary conditions and a zero initial condition. We keep the same parameter values and choose the target state $\hat{y} = \sin(\pi x)(2t^2 + t)$. We use the Crank-Nicolson scheme with mesh size $h_t = h_x = 1/32$. The error decay as a function of the number of iterations is shown in Figure 4 (right). As expected, SD_2 diverge violently, and both SD_3 and SD_4 stagnate. The convergence of SD_1 and SN_1 is already efficient without relaxation, due to the smallest eigenvalue in this test case

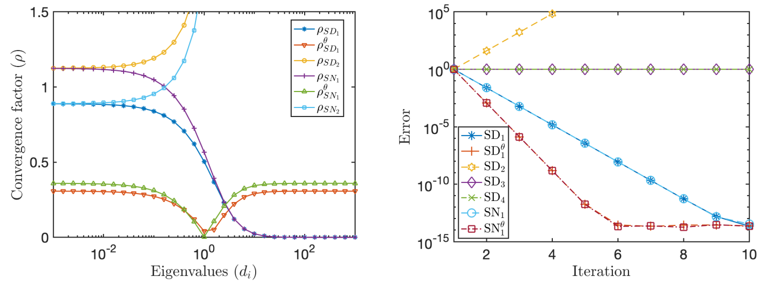

Fig. 2 Convergence factor as a function of eigenvalues (left) and error decay as a function of the number of iterations (right).

being around 10. The convergence can be improved from 10 to 6 iterations with a relaxation parameter $\theta = 0.975$ for both algorithms.

Unlike our observation in [3, 4] for Dirichlet–Neumann and Neumann–Neumann algorithms in time, classical Schwarz algorithms with only Dirichlet or Neumann transmission conditions are much more sensitive to the forward-backward structure. We observe that only SD_1 and SN_1, which naturally preserve this structure, have good convergence behavior. All other variants preform poorly and cannot be improved even with relaxation techniques. For SD_1 and SN_1, we also provided estimates and closed-form expressions for the optimal relaxation parameters. In [2], the authors used transmission conditions of the form $\boldsymbol{\lambda} + p\boldsymbol{y}$ on I_1 and $\boldsymbol{y} - q\boldsymbol{\lambda}$ on I_2, with two parameters $p, q \geq 0$. It would be interesting to extend this approach using transmission conditions of the form $\dot{\boldsymbol{\lambda}} + p\dot{\boldsymbol{y}}$ and $\dot{\boldsymbol{y}} - q\dot{\boldsymbol{\lambda}}$ to further improve the convergence in the case of Neumann transmission conditions only.

References

1. M. J. Gander. Schwarz methods over the course of time. *Electronic Transactions on Numerical Analysis*, 31:228–255, 2008.
2. M. J. Gander and F. Kwok. Schwarz methods for the time-parallel solution of parabolic control problems. In T. Dickopf, M. J. Gander, L. Halpern, R. Krause, and L. F. Pavarino, editors, *Domain Decomposition Methods in Science and Engineering XXII*, pages 207–216, Cham, 2016. Springer.
3. M. J. Gander and L.-D. Lu. New time domain decomposition methods for parabolic optimal control problems I: Dirichlet–Neumann and Neumann–Dirichlet algorithms. *SIAM Journal on Numerical Analysis*, 62(4):2048–2070, 2024.
4. M. J. Gander and L.-D. Lu. New time domain decomposition methods for parabolic optimal control problems II: Neumann–Neumann algorithms. To appear in SIAM Journal on Numerical Analysis, 2024.
5. P.-L. Lions. On the Schwarz Alternating Method III: A Variant for Nonoverlapping Subdomains. In T. F. Chan, R. Glowinski, J. Périaux, and O. Widlund., editors, *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 202–223. SIAM, 1990.

6. H. A. Schwarz. Über einen Grenzübergang durch alternierendes Verfahren. *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, 15:272–286, 1870.