Schwarz Methods for the Time-Parallel Solution of Parabolic Control Problems

Martin J. Gander and Felix Kwok

1 Introduction

Suppose we are interested in the following distributed control problem: given a system governed by the parabolic PDE $\dot{y}+\mathscr{L}y=u$ on the time interval $[0,T]$ (where $\dot{y}$ denotes the time derivative of y), we wish to choose the forcing term $u=u(t)$ to minimize the discrepancy between the trajectory and the desired state $\hat{y}=\hat{y}(t)$. After semi-discretization in space, we obtain for a given choice of parameters $\gamma,\nu>0$ the following minimization problem:

$$\begin{aligned} &\min_{y,u}\frac{1}{2}\int_0^T\|y-\hat{y}\|^2\,dt+\frac{\gamma}{2}\|y(T)-\hat{y}(T)\|^2+\frac{\nu}{2}\int_0^T\|u\|^2\,dt \\ &\qquad\text{subject to}\qquad \dot{y}+Ay=u,\qquad y(0)=y_0, \end{aligned} \tag{1}$$

where A is the matrix obtained by semi-discretization of the operator $\mathscr{L}$. While the PDE in (1) may resemble an initial-value problem, the minimization problem is in fact a two-point boundary value problem in time, since the first-order optimality conditions couple the PDE to an adjoint equation that is backwards in time and contains a final condition, see Section 2. To solve such systems in parallel, one can use multiple shooting methods, see [8] and references therein, or parareal-type algorithms in a reduced Hessian formulation, see [7, 4]. A Schwarz preconditioner in time for such systems was presented in [1], where on each subinterval $I_j=[T_j,T_{j+1}]$, one uses an initial condition for y from I_{j-1} and a final condition for the adjoint state λ from I_{j+1}. To the authors' knowledge, no convergence analysis is available for this method.

Martin J. Gander
Department of Mathematics, University of Geneva, e-mail: Martin.Gander@unige.ch

Felix Kwok
Department of Mathematics, Hong Kong Baptist University, e-mail: felix_kwok@hkbu.edu.hk

We study in this paper Schwarz methods for the time-parallel solution of (1). 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. Using well chosen linear combinations of y and λ as transmission conditions allows us to optimize the Robin conditions for performance, and leads to much faster Schwarz methods, especially when the spatial operator has eigenvalues close to zero. We illustrate our results with numerical experiments.

2 Schwarz Methods in Time

Using the Lagrange multiplier approach (see e.g. the historical review [5]), one can derive the forward and adjoint problems to be

$$\begin{cases} \dot{y} + Ay = u & \text{on } (0,T), \\ y(0) = y_0, \end{cases} \qquad \begin{cases} \dot{\lambda} - A^T\lambda = y - \hat{y} & \text{on } (0,T), \\ \lambda(T) = -\gamma(y(T) - \hat{y}(T)), \end{cases}$$

where the control u and adjoint state λ are related by the algebraic equation $\lambda(t) = \nu u(t)$ for all $t \in (0,T)$. Eliminating u, the above system can thus also be written as

$$\begin{bmatrix} \dot{y} \\ \dot{\lambda} \end{bmatrix} + \begin{bmatrix} A & -\nu^{-1}I \\ -I & -A^T \end{bmatrix} \begin{bmatrix} y \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ -\hat{y} \end{bmatrix}. \tag{2}$$

Suppose we wish to divide the time interval $(0,T)$ into two subintervals $I_1 = (0,\beta)$ and $I_2 = (\alpha,T)$ with $\alpha \le \beta$ in order to solve the two subdomain problems in parallel. Then for any choice of parameters $p,q \ge 0$, we propose the following parallel Schwarz algorithm: for $k = 1,2,\ldots$, solve

$$\begin{cases} \begin{bmatrix} \dot{y}_1^k \\ \dot{\lambda}_1^k \end{bmatrix} + \begin{bmatrix} A & -\nu^{-1}I \\ -I & -A^T \end{bmatrix} \begin{bmatrix} y_1^k \\ \lambda_1^k \end{bmatrix} = \begin{bmatrix} 0 \\ -\hat{y} \end{bmatrix} & \text{on } I_1 = (0,\beta), \\ y_1^k(0) = y_0, \\ \lambda_1^k(\beta) + p y_1^k(\beta) = \lambda_2^{k-1}(\beta) + p y_2^{k-1}(\beta), \end{cases} \tag{3a}$$

$$\begin{cases} \begin{bmatrix} \dot{y}_2^k \\ \dot{\lambda}_2^k \end{bmatrix} + \begin{bmatrix} A & -\nu^{-1}I \\ -I & -A^T \end{bmatrix} \begin{bmatrix} y_2^k \\ \lambda_2^k \end{bmatrix} = \begin{bmatrix} 0 \\ -\hat{y} \end{bmatrix} & \text{on } I_2 = (\alpha,T), \\ y_2^k(\alpha) - q\lambda_2^k(\alpha) = y_1^{k-1}(\alpha) - q\lambda_1^{k-1}(\alpha), \\ \lambda_2^k(T) = -\gamma(y_2^k(T) - \hat{y}(T)). \end{cases} \tag{3b}$$

For $p = q = 0$, the transmission conditions reduce to the classical conditions from [1]. To understand why we consider transmission conditions of this form, suppose that $A = A^T \in \mathbb{R}^{m\times m}$, so that A can be diagonalized as $A = QDQ^T$, with $Q^TQ = I$

and $D = \operatorname{diag}(d_1, \dots, d_m)$. Then the ODE system in (3a) can be written as

$$\begin{cases} \begin{bmatrix} \dot{z}_1^k \\ \dot{\mu}_1^k \end{bmatrix} + \begin{bmatrix} D & -\nu^{-1} I \\ -I & -D \end{bmatrix} \begin{bmatrix} z_1^k \\ \mu_1^k \end{bmatrix} = \begin{bmatrix} 0 \\ -\hat{z} \end{bmatrix} \qquad \text{on } I_1 = (0, \beta), \\ z_1^k(0) = z_0, \\ \mu_1^k(\beta) + p z_1^k(\beta) = \mu_2^{k-1}(\beta) + p z_2^{k-1}(\beta), \end{cases} \tag{4}$$

where $z_j^k = Q^T y_j^k$, $\mu_j^k = Q^T \lambda_j^k$ for $j = 1, 2$ and $\hat{z} = Q^T \hat{y}$, $z_0 = Q^T y_0$. Thus, we obtain m independent 2×2 systems of the form

$$\dot{z}_1^{(i),k} + d_i z_1^{(i),k} - \nu^{-1} \mu_1^{(i),k} = 0, \qquad \dot{\mu}_1^{(i),k} - d_i \mu^{(i),k} - z_1^{(i),k} = \hat{z}^{(i)}, \tag{5}$$

where $z_1^{(i),k}$ and $\mu_1^{(i),k}$ are the i-th components of z_1^k and μ_1^k respectively, and d_i is the i-th eigenvalue of A. By isolating μ from the first equation in (5) and substituting into the second, we obtain the second-order ODE

$$\ddot{z}_1^{(i),k} - (d_i^2 + \nu^{-1}) z_1^{(i),k} = -\nu^{-1} \hat{z}^{(i)}, \tag{6}$$

whereas the boundary conditions become

$$z_1^{(i),k}(0) = z_0^{(i)}(0), \;\; \dot{z}_1^{(i),k} + (d_i + p\nu^{-1}) z_1^{(i),k} \Big|_{t=\beta} = \dot{z}_2^{(i),k-1} + (d_i + p\nu^{-1}) z_2^{(i),k-1} \Big|_{t=\beta}.$$

Hence, once we eliminate the adjoint state, it becomes apparent that we are in fact imposing a Robin transmission condition on the elliptic boundary value problem (6), even with the classical Schwarz method $p = q = 0$ from [1]. With the additional parameter p and q, one can now optimize the convergence, as in optimized Schwarz methods [6]. Boundary conditions of the form $y - q\lambda$ in (3b) can be explained similarly; here, the minus sign is chosen so that the subdomain problem is well-posed for $q \geq 0$ whenever A is symmetric semi-positive definite.

Remark on implementation. Since we are primarily interested in the behavior of the Schwarz method, we will regard solvers for the subdomain problems (3a) and (3b) as black boxes. We emphasize however that final conditions of the form $\lambda + py$ already appear when the objective function contains the target term $\frac{\gamma}{2}|y(T) - \hat{y}(T)|^2$, see (3b). Thus, existing solvers can be used as is or easily modified to handle the optimized conditions, see [2] or [3].

3 Convergence analysis

In this section, we assume A to be symmetric and semi-positive definite, so that (3a)–(3b) can be diagonalized as in Section 2 with $d_i \geq 0$. Moreover, since the problem is linear, we can analyze the error equation, which means setting y_0 and $\hat{y}$ to zero and studying how y_j^k and λ_j^k converge to zero as $k \to \infty$. After diagonalization, the first subdomain solution satisfies (6) with homogeneous initial condition:

$$\ddot{z}_1^{(i),k} - (d_i^2 + \nu^{-1}) z_1^{(i),k} = 0, \quad z_1^{(i),k}(0) = 0 \quad \Longrightarrow \quad z_1^{(i),k}(t) = A_i^k \sinh(\sigma_i t), \quad (7)$$

where $\sigma_i = \sqrt{d_i^2 + \nu^{-1}} > 0$, and A_i^k is a constant determined by the boundary condition $\nu \dot{z}_1^{(i),k} + (p + \nu d_i) z_1^{(i),k}|_{t=\beta} = g^{(i),k}$. Substituting the solution from (7) and isolating A_i^k yields $A_i^k = \frac{g_1^{(i),k}}{\nu[\sigma_i \cosh(\sigma_i \beta) + (d_i + p\nu^{-1}) \sinh(\sigma_i \beta)]}$. Next, we consider the subdomain $I_2 = (\alpha, T)$ at iteration $k+1$. The boundary data at $t = \alpha$ can be written as

$$\begin{aligned} h^{(i),k+1} := z_1^{(i),k} - q\mu_1^{(i),k}\Big|_{t=\alpha} &= -\nu q \dot{z}_1^{(i),k} + (1 - \nu q d_i) z_1^{(i),k}\Big|_{t=\alpha} \\ &= -g^{(i),k} \frac{\sigma_i q \cosh(\sigma_i \alpha) + (q d_i - \nu^{-1}) \sinh(\sigma_i \alpha)}{\sigma_i \cosh(\sigma_i \beta) + (d_i + p\nu^{-1}) \sinh(\sigma_i \beta)}. \end{aligned} \quad (8)$$

On the other hand, the ODE can be written as

$$\ddot{\mu}_2^{(i),k+1} - (d_i^2 + \nu^{-1}) \mu_2^{(i),k+1} = 0 \qquad \text{on } I_2 = (\alpha, T),$$
$$\mu_2^{(i),k+1}(T) + \gamma z_2^{(i),k+1}(T) = 0, \qquad z_2^{(i),k+1}(\alpha) - q\mu_2^{(i),k+1}(\alpha) = h^{(i),k+1}.$$

Since $z_2^{(i),k+1} = \dot{\mu}_2^{(i),k+1} - d_i \mu_2^{(i),k+1}$, the boundary conditions can be written as

$$\gamma \dot{\mu}_2^{(i),k+1}(T) + (1 - d_i \gamma) \mu_2^{(i),k+1}(T) = 0, \; \dot{\mu}_2^{(i),k+1}(\alpha) - (d_i + q) \mu_2^{(i),k+1}(\alpha) = h^{(i),k+1}.$$

The boundary condition at $t = T$ gives

$$\mu_2^{(i),k+1} = B_i^{k+1} \left[\sigma_i \gamma \cosh(\sigma_i (T - t)) + (1 - d_i \gamma) \sinh(\sigma_i (T - t)) \right],$$

where B_i^{k+1} is a constant. The boundary condition at $t = \alpha$ allows us to determine this constant (after some algebra) to be

$$B_i^{k+1} = \frac{-h^{(i),k+1}}{(\sigma_i (1 + q\gamma)) \cosh(\sigma_i (T - \alpha)) + (d_i (1 - q\gamma) + q + \nu^{-1}\gamma) \sinh(\sigma_i (T - \alpha))}.$$

Note that the denominator does not vanish for any choice of $q, \gamma \geq 0$: if we define $\theta_i = \tanh^{-1}(d_i / \sigma_i)$, which is possible because $0 \leq d_i < \sigma_i$, then we can write the denominator as

$$\begin{aligned} &\sigma_i \cosh(\cdot) + d_i \sinh(\cdot) + q\gamma(\sigma_i \cosh(\cdot) - d_i \sinh(\cdot)) + (q + \nu^{-1}\gamma) \sinh(\cdot) \\ &\qquad = \nu^{-1/2} \left[\cosh(\cdot + \theta_i) + q\gamma \cosh(\cdot - \theta_i) \right] + (q + \nu^{-1}\gamma) \sinh(\cdot) > 0. \end{aligned}$$

If we now let $g^{(i),k+2} = \mu_2^{(i),k+1}(\beta) + p z_2^{(i),k+1}(\beta)$, we get

$$g^{(i),k+2}=h^{(i),k+1}\frac{\nu^{-1/2}\left[p\cosh(\sigma_i(T-\beta)+\theta_i)-\gamma\cosh(\sigma_i(T-\beta)-\theta_i)\right]-(1-\nu^{-1}p\gamma)\sinh(\sigma_i(T-\beta))}{\nu^{-1/2}\left[\cosh(\sigma_i(T-\alpha)+\theta_i)+q\gamma\cosh(\sigma_i(T-\alpha)-\theta_i)\right]+(q+\nu^{-1}\gamma)\sinh(\sigma_i(T-\alpha))}.$$

Substituting (8) into the above equations and taking absolute values, we obtain

Theorem 1. *The parallel Schwarz method* (3a)–(3b) *converges whenever* $\rho < 1$, *where*

$$\rho = \max_{d_i\in\lambda(A)}\left|\frac{\sigma_i q\cosh(\sigma_i\alpha)+(qd_i-\nu^{-1})\sinh(\sigma_i\alpha)}{\sigma_i\cosh(\sigma_i\beta)+(d_i+p\nu^{-1})\sinh(\sigma_i\beta)}\right.$$
$$\left.\cdot\frac{\nu^{-1/2}\left[p\cosh(\sigma_i(T-\beta)+\theta_i)-\gamma\cosh(\sigma_i(T-\beta)-\theta_i)\right]-(1-\nu^{-1}p\gamma)\sinh(\sigma_i(T-\beta))}{\nu^{-1/2}\left[\cosh(\sigma_i(T-\alpha)+\theta_i)+q\gamma\cosh(\sigma_i(T-\alpha)-\theta_i)\right]+(q+\nu^{-1}\gamma)\sinh(\sigma_i(T-\alpha))}\right|^{1/2},$$

where the maximum is taken over all the set of eigenvalues of A.

To gain a better understanding of the convergence, let us assume that $A = A^T$ is positive semi-definite (so that $d_i \geq 0$) and consider a few special cases.

Classical transmission conditions ($p = q = 0$)**.** Here the expression simplifies to

$$\rho^2 = \max_i\left(\frac{\sinh(\sigma_i\alpha)}{\cosh(\sigma_i\beta+\theta_i)}\cdot\frac{\nu^{1/2}\sinh(\sigma_i(T-\beta))+\gamma\cosh(\sigma_i(T-\beta)-\theta_i)}{\gamma\sinh(\sigma_i(T-\alpha))+\nu^{1/2}\cosh(\sigma_i(T-\alpha)+\theta_i)}\right).$$

If $\gamma \leq \sqrt{\nu}$, then $\rho < 1$ and the method converges; this is because

$$\sinh(\sigma_i\alpha) \leq \cosh(\sigma_i\alpha) \leq \cosh(\sigma_i\beta+\theta_i)$$

and, since $\sinh(\sigma_i(T-\beta)) \leq \cosh(\sigma_i(T-\beta)+\theta_i)$, we have

$$\begin{aligned}\nu^{1/2}\sinh(\sigma_i(T-\beta))&+\gamma\cosh(\sigma_i(T-\beta)-\theta_i)\\ &\leq \nu^{1/2}\sinh(\sigma_i(T-\alpha))+\gamma\cosh(\sigma_i(T-\alpha)+\theta_i)\\ &\leq \gamma\sinh(\sigma_i(T-\alpha))+\nu^{1/2}\cosh(\sigma_i(T-\alpha)-\theta_i).\end{aligned}$$

However, it is possible for the method to diverge if $\gamma > \nu^{1/2}$, see Section 4. In the case when $\gamma = 0$, i.e., when the target state does not appear explicitly in the objective function, it is possible to estimate the convergence factor directly. Here we have

$$\rho^2 = \max_i\frac{\sinh(\sigma_i\alpha)\sinh(\sigma_i(T-\beta))}{\cosh(\sigma_i\beta+\theta_i)\cosh(\sigma_i(T-\alpha)+\theta_i)} < 1,$$

since $\alpha \leq \beta$. The term inside the maximum is a function of the eigenvalues d_i via $\sigma_i = \sqrt{d_i^2+\nu^{-1}}$ and $\theta_i = \operatorname{arctanh}(d_i/\sigma_i)$. It can be shown that this function is decreasing with respect to d_i on $[0,\infty)$, see also Figure 1; thus, if $d_{\min} \geq 0$ is the minimum eigenvalue of A and $\sigma_{\min}$ and $\theta_{\min}$ are the corresponding values, then one can estimate ρ by

$$\rho \leq \left(\frac{\exp(\sigma_{\min}(\alpha+T-\beta))}{\exp(\sigma_{\min}(\beta+T-\alpha)+2\theta_{\min})}\right)^{1/2} = e^{-\sigma_{\min}(\beta-\alpha)-\theta_{\min}},$$

where we used the bounds $\sinh(x) \leq \frac{1}{2}\exp(x)$ and $\cosh(x) \geq \frac{1}{2}\exp(x)$, valid for all $x \geq 0$. Thus, when the subdomains overlap, i.e., when $\beta - \alpha > 0$, the convergence factor decreases exponentially with respect to the overlap size $\beta - \alpha$. When there is no overlap, i.e., when $\alpha = \beta$, it is still possible to bound ρ by estimating $e^{-\theta_{\min}}$ since $\tanh(\theta_{\min}) = d_{\min}/\sigma_{\min}$ by definition, we get

$$\frac{e^{\theta_{\min}} - e^{-\theta_{\min}}}{e^{\theta_{\min}} + e^{-\theta_{\min}}} = \frac{1 - e^{-2\theta_{\min}}}{1 + e^{-2\theta_{\min}}} = \frac{d_{\min}}{\sigma_{\min}} \implies 1 - \frac{d_{\min}}{\sigma_{\min}} = e^{-2\theta_{\min}}\left(1 + \frac{d_{\min}}{\sigma_{\min}}\right).$$

This implies

$$e^{-2\theta_{\min}} = \frac{\sigma_{\min} - d_{\min}}{\sigma_{\min} + d_{\min}} = \frac{\nu^{-1}}{\left(\sqrt{d_j^2 + \nu^{-1}} + d_j\right)^2}.$$

Taking square roots, we obtain the following estimate:

Theorem 2. *Suppose A is symmetric positive definite and $\gamma = 0$. Then the parallel Schwarz method* (3a)–(3b) *with classical transmission conditions ($p = q = 0$) converges for all initial guesses with the estimate*

$$\rho \leq \frac{e^{-(\beta-\alpha)\sqrt{d^2+\nu^{-1}}}}{\sqrt{1+\nu d^2} + \nu^{1/2} d},$$

where $\beta - \alpha \geq 0$ is the overlap size and $d \geq 0$ is the smallest eigenvalue of A.

Note that if A arises from a spatial discretization of a differential operator, then the smallest eigenvalue of A typically does not vary much as the spatial grid is refined. Thus, *the convergence of the method is independent of the mesh parameter h*. However, if A is singular ($d = 0$) and there is no overlap, then convergence can be very slow, see the example in Section 4.

Optimized transmission conditions, no target state ($\gamma = 0$). To accelerate the convergence of the method when A is singular, let us consider choosing the parameters p and q to be equal but non-zero. Then the convergence factor becomes

$$\rho = \max_{d_i \in \lambda(A)} \left| \frac{\sigma_i p \cosh(\sigma_i\alpha) + (pd_i - \nu^{-1})\sinh(\sigma_i\alpha)}{\sigma_i \cosh(\sigma_i\beta) + (d_i + p\nu^{-1})\sinh(\sigma_i\beta)} \cdot \frac{p\sigma_i \cosh(\sigma_i(T-\beta)) + (pd_i - 1)\sinh(\sigma_i(T-\beta))}{\sigma_i \cosh(\sigma_i(T-\alpha)) + (p + d_i)\sinh(\sigma_i(T-\alpha))} \right|^{1/2}.$$

A plot of the right-hand side as a function of d_i for fixed $p > 0$ is shown in Fig. 1. We see that as $d_i \to \infty$, we have

$$\rho \longrightarrow p \cdot \lim_{d_i \to \infty} \left(\frac{\cosh(\sigma_i\alpha + \theta_i)\cosh(\sigma_i(T-\beta) + \theta_i)}{\cosh(\sigma_i\beta + \theta_i)\cosh(\sigma_i(T-\alpha) + \theta_i)} \right)^{1/2}.$$

Thus, if no overlap is used, then the method converges only if $0 \leq p < 1$. On the other hand, for $d_i = 0$, we have

$$\rho(d_i = 0) = \left| \frac{p\cosh(\sigma_i\alpha) - \nu^{-1/2}\sinh(\sigma_i\alpha)}{\cosh(\sigma_i\beta) + p\nu^{-1/2}\sinh(\sigma_i\beta)} \cdot \frac{\nu^{-1/2} p \cosh(\sigma_i(T-\beta)) - \sinh(\sigma_i(T-\beta))}{\nu^{-1/2}\cosh(\sigma_i(T-\alpha)) + p\sinh(\sigma_i(T-\alpha))} \right|^{1/2}.$$

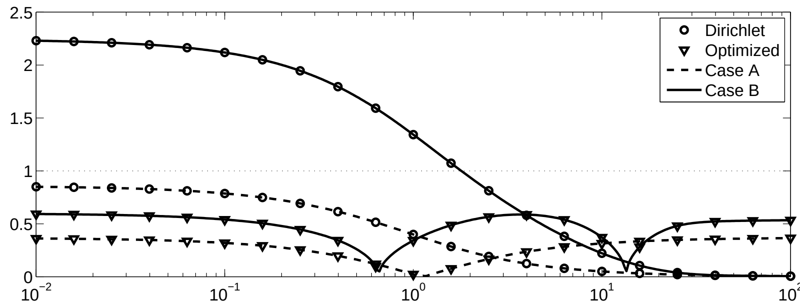

Fig. 1 A comparison of contraction factors as a function of eigenvalues d_i for classical ($p = q = 0$) and optimized transmission conditions (p and q obtained by equioscillation).

Thus, if we assume the eigenvalues of A can be anywhere in the interval $[0,\infty)$, then the smallest convergence factor is obtained when $|\rho(d_i = 0)| = |\rho(d_i \to \infty)|$, i.e., by equioscillation.

4 Numerical experiments

To understand how convergence depends on the different parameters, we consider for each ODE two different test cases:

Case A: The time interval $\Omega = [0,3]$ is subdivided into $\Omega_1 = (0,1)$, $\Omega_2 = (1,3)$ (no overlap), and the objective function has no explicit target term ($\gamma = 0$). The regularization parameter is $\nu = 1$.

Case B: The subdomains are $\Omega_1 = (0,2.9)$ and $\Omega_2 = (2.9,3)$, and the objective function has a target term with $\gamma = 10$. The regularization parameter is still $\nu = 1$.

For each test case, we plot in Figure 1 the convergence factor ρ as a function of the frequency d_i, both for classical ($p = q = 0$) and optimized transmission conditions. Based on the equioscillation criterion, we choose $p = q = 0.37$ for case A and $p = q = 0.55$ for case B. We see that when classical conditions are used, the method converges in case A for all frequencies, whereas in case B, the method only converges when the lowest eigenvalue of the spatial operator is larger than about 2. However, when optimized conditions are used, the parameters can be chosen so that the method converges for all frequencies, and the spectral radius can be made much smaller than in the classical case (0.37 versus about 0.9 for classical).

Next, we solve numerically the optimal control problem (1) with governing PDE $\partial_t u = \partial_{xx} u$ and regularization parameters $\nu = 1$, $\gamma = 0$. The problem is discretized using the second-order Crank–Nicolson method with spatial and temporal mesh size $h = 1/32$ and $1/64$. The problem is then solved in parallel using two time windows $\Omega_1 = (0,1)$ and $\Omega_2 = (1,3)$. Again we consider two cases: in the first case, we use Dirichlet boundary conditions in space, which means the operator A in (1) has

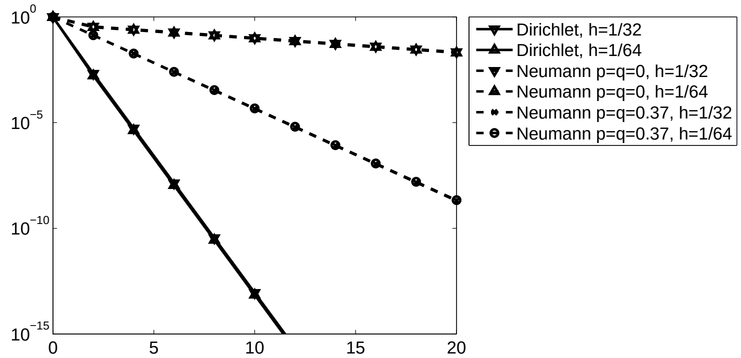

Fig. 2 Convergence of Algorithm (3a)–(3b) for different parameters and boundary conditions.

lowest eigenvalue $\pi^2 \approx 9.87$. From Figure 2, we see that the method converges very quickly with a rate that is indeed independent of h (see remark after Theorem 2). The fast convergence can be explained by Figure 1: the spectral radius curve beyond the point $d_i = 9.87$ is very close to zero, so the convergence is very quick indeed.

In the second case, we consider the same PDE, but with Neumann boundary conditions in space. In this case, zero is an eigenvalue of the spatial operator, meaning we have to minimize the convergence factor over the whole interval $d_i \in [0, \infty)$. Here, the method with classical transmission conditions ($p = q = 0$) converges very slowly, whereas convergence is much faster with optimized transmission conditions. Again the convergence is independent of the spatial mesh size, as expected.

5 Conclusions

We have presented a first analysis of Schwarz methods in time for parabolic control problems. We have shown that classical Schwarz methods already use Robin type transmission conditions, and introduced a parameter which can be chosen to obtain substantially faster convergence, especially when the spatial operator has eigenvalues close to zero. We are currently working on error estimates for the many-subdomain case and on higher order transmission conditions.

References

1. Barker, A.T., Stoll, M.: Domain decomposition in time for PDE-constrained optimization (2014). Submitted
2. Borzì, A., Schulz, V.: Computational optimization of systems governed by partial differential equations, vol. 8. SIAM (2012)

3. Dollar, H.S., Gould, N.I., Stoll, M., Wathen, A.J.: Preconditioning saddle-point systems with applications in optimization. SIAM Journal on Scientific Computing **32**(1), 249–270 (2010)
4. Du, X., Sarkis, M., Schaerer, C.E., Szyld, D.B.: Inexact and truncated parareal-in-time Krylov subspace methods for parabolic optimal control problems. ETNA **40**, 36–57 (2013)
5. Gander, M., Kwok, F., Wanner, G.: History of constrained optimization. In: Optimization with PDE Constraints, pp. 151–202. Springer Verlag (2014)
6. Gander, M.J.: Optimized Schwarz methods. SIAM Journal on Numerical Analysis **44**(2), 699–731 (2006)
7. Haber, E.: A parallel method for large scale time domain electromagnetic inverse problems. Applied Numerical Mathematics **58**(4), 422–434 (2008)
8. Heinkenschloss, M.: A time-domain decomposition iterative method for the solution of distributed linear quadratic optimal control problems. J. Comput. Appl. Math. **173**(1), 169–198 (2005)