1. A Non-Overlapping Optimized Schwarz Method which Converges with Arbitrarily Weak Dependence on $h$

M.J. Gander$^1$, G.H. Golub$^2$

1. Introduction. Optimized Schwarz methods have been introduced in [11] to correct the uneven convergence properties of the classical Schwarz method. In the classical Schwarz method high frequency components converge very fast, whereas low frequency components are only converging very slowly and hence slow down the performance of the overall method. This can be corrected by replacing the Dirichlet transmission conditions in the classical Schwarz method by Robin or higher order transmission conditions which approximate the classical absorbing boundary conditions used to truncate infinite domains for numerical computations on bounded domains. The new methods are called optimized Schwarz methods because the new transmission conditions are obtained by optimizing their coefficients for the performance of the method.

Using transmission conditions different from the Dirichlet ones is however not new. P.-L. Lions proposed in [14] to use Robin transmission conditions to obtain a converging non-overlapping variant of the Schwarz method, a result not possible with Dirichlet transmission conditions. But it was in the context of a particular problem, namely the Helmholtz equation, where the importance of radiation conditions was first realized in the PhD thesis of Deprès [3]. Several publications for the Helmholtz equation followed; in the context of control [1], for an overlapping variant in [2], and a first approach to optimize the transmission conditions without overlap in [4]. An interesting variant of a Schwarz method using perfectly matched layers can be found in [17]. Fully optimized transmission conditions were published in [8, 12] for the non-overlapping variant of the Schwarz method and a first approach for the overlapping case can be found in [11]. Very soon it was realized that approximations to absorbing boundary conditions were very effective for other types of equations as well. For the convection-diffusion equation, the first paper proposing optimized transmission conditions for a non-overlapping variant of the Schwarz method is [3]. Around the same time, a discrete version of such a Schwarz method was developed at the algebraic level in [16, 15], but it proved to be difficult to optimize the free parameters. Second order optimized transmission conditions for convection-diffusion were explored in [13] for the non-overlapping case and for symmetric positive definite problems in [7] with the first asymptotic results of the performance of those methods. Such transmission conditions are also crucial in the case of evolution problems, as shown in [10], and for systems of equations, for the Euler equations, see [6]. A complete survey for symmetric positive definite problems with all the asymptotic performances for overlapping and non-overlapping variants, is in preparation [9].

We show in this paper that the transmission conditions in the optimized Schwarz methods can be chosen such that the convergence rate of the method has an arbitrarily weak asymptotic dependence on the mesh parameter $h$, even if no overlap is used. This result is obtained by choosing a sequence of transmission conditions which is

---

$^1$McGill University, mgander@math.mcgill.ca, supported in part by NSERC grant 228061.
$^2$Stanford University, golub@scm.stanford.edu, supported in part by DOE DE-FC02-01ER41777.
applied cyclically in the optimized Schwarz iteration. Closed form expressions for the transmission conditions are derived which give an asymptotic convergence rate \( \rho = 1 - O(h^{1/m}) \) for \( m \) an arbitrary power of 2.

2. The Model Problem. We consider for this paper the self adjoint coercive model problem

\[
\mathcal{L}(u) := (\eta - \Delta)u = f \quad \text{in } \Omega = \mathbb{R}^2
\]

and we assume that the solution \( u(x, y) \) stays bounded at infinity. We can pose an equivalent problem on \( \mathbb{R}^2 \) decomposed into two overlapping subdomains \( \Omega_1 = (-\infty, L) \times \mathbb{R} \), \( \Omega_2 = (0, \infty) \times \mathbb{R} \), \( L > 0 \), with boundaries \( \Gamma_L \) at \( x = L \) and \( \Gamma_0 \) at \( x = 0 \) as shown in Figure 2.1, namely

\[
(\eta - \Delta)v = f \quad \text{in } \Omega_1, \quad (\eta - \Delta)w = f \quad \text{in } \Omega_2, \quad v = w \quad \text{on } \Gamma_L, \quad w = v \quad \text{on } \Gamma_0.
\]

Then the restriction of the solution \( u \) of the original problem to \( \Omega_1 \) coincides with the solution \( v \) of the partitioned problem and the restriction of the solution \( u \) to \( \Omega_2 \) coincides with \( w \) of the partitioned problem. If the overlap however becomes zero, \( L = 0 \), then the subdomain problems do not necessarily coincide with the solution \( u \) of the original problem any more. one has to introduce the additional condition that the derivatives need to match.

\[
\partial_x w = \partial_x v \quad \text{on } \Gamma_0 = \Gamma_L.
\]

To make the coupling more robust with respect to small overlap, we introduce the subdomain coupling

\[
(\partial_x + S_v)v = (\partial_x + S_v)w \quad \text{on } \Gamma_L, \quad (\partial_x + S_w)w = (\partial_x + S_w)v \quad \text{on } \Gamma_0,
\]

where \( S_v \) and \( S_w \) are for the moment undetermined linear operators acting in the \( y \) direction. Note that for example choosing \( S_v = -S_w = p \) for some constant \( p > 0 \) leads to subdomain solutions \( v \) and \( w \) which coincide with the solution \( u \) of the original problem even if the overlap is zero. \( L = 0 \), since then the conditions \( (\partial_x + p)v = (\partial_x + p)w \) and \( (\partial_x - p)w = (\partial_x - p)v \) on \( \Gamma_0 \) imply both continuity of the subdomain solution and its derivative at \( x = 0 \). The subdomain problems are then coupled by a Robin transmission condition, an idea introduced in [14]. The goal of optimized Schwarz methods is to determine good choices for the operators \( S_v \) and \( S_w \) to obtain fast domain decomposition methods at a computational cost comparable to the classical Schwarz method.
3. An Optimized Schwarz Method. We introduce a Schwarz relaxation to the system coupled with the new conditions,

\[
\begin{align*}
(\eta - \Delta)v^n &= f, & \text{in } \Omega_1, \\
(\eta - \Delta)w^n &= f, & \text{in } \Omega_2, \\
(\partial_x + S_v)v^n &= (\partial_x + S_v)w^{n-1} & \text{on } \Gamma_L, \\
(\partial_x + S_w)w^n &= (\partial_x + S_w)v^{n-1} & \text{on } \Gamma_0.
\end{align*}
\]

This iteration can be analyzed using Fourier analysis, see for example [11]. The convergence rate of this algorithm is

\[
\rho(k) = \frac{\sqrt{\eta + k^2 - \sigma_v(k)}}{\sqrt{\eta + k^2 + \sigma_v(k)}} \cdot \frac{\sqrt{\eta + k^2 + \sigma_w(k)}}{\sqrt{\eta + k^2 - \sigma_w(k)}} e^{-2\sqrt{\eta + k^2} L},
\]

where \( k \) is the Fourier variable in the \( y \) direction and \( \sigma_v \) and \( \sigma_w \) denote the symbols of \( S_v \) and \( S_w \). The optimal transmission operators \( S_v \) and \( S_w \) have thus the symbols \( \sigma_v = \sqrt{\eta + k^2} \) and \( \sigma_w = -\sqrt{\eta + k^2} \) because then the convergence rate vanishes and hence the algorithm converges in 2 steps, independent of the size of the overlap \( L \). Unfortunately these operators are non-local, they require the evaluation of a convolution and hence polynomial approximations have been introduced for various types of partial differential equations, see [3, 13, 4, 7, 10, 11, 8, 12]. The simplest approximation is to use a constant, which leads to the Robin transmission conditions \( S_v = -S_w = p \) for some constant \( p > 0 \). The sign of \( p \) is needed for well-posedness, but it also guarantees convergence of the algorithm, since then the convergence rate becomes

\[
\rho(k, p) = \left( \frac{\sqrt{\eta + k^2 - \sigma_v(k)}}{\sqrt{\eta + k^2 + \sigma_v(k)}} \cdot \frac{\sqrt{\eta + k^2 + \sigma_w(k)}}{\sqrt{\eta + k^2 - \sigma_w(k)}} e^{-2\sqrt{\eta + k^2} L} \right)^2
\]

which is less than one for all \( k < \infty \), even if the overlap is zero, i.e. \( L = 0 \). To find the best Robin parameter, one minimizes the convergence rate over all the frequencies relevant to a given discretization, \( k_{\min} < |k| < k_{\max} \), which leads to the min-max problem

\[
\min_{p \geq 0} \left( \max_{k_{\min} < |k| < k_{\max}} \rho(k, p) \right).
\]

The solution of this problem, with or without overlap, can be found in [9] and the convergence rate depends mildly on the mesh parameter \( h \); for \( L = 0 \) one finds \( \rho = 1 - O(\sqrt{h}) \) and for \( L = O(h) \) the result is \( \rho = 1 - O(h^{1/3}) \). In the following section we will make the convergence rate as weakly dependent on \( h \) as desired for the case \( L = 0 \).

4. Arbitrarily Weak Dependence on \( h \). The idea is to use different parameters \( p_j \) for different steps of the iteration. Suppose we want to use \( m \) different values \( p_j \), \( j = 1, \ldots, m \) in the Robin transmission condition. We then cycle through these different parameters in the optimized Schwarz algorithm

\[
\begin{align*}
(\eta - \Delta)v^n &= f, & \text{in } \Omega_1, \\
(\eta - \Delta)w^n &= f, & \text{in } \Omega_2, \\
(\partial_x + p_n \text{ mod } m+1)v^n &= (\partial_x + p_n \text{ mod } m+1)w^{n-1} & \text{on } \Gamma_L, \\
(\partial_x - p_n \text{ mod } m+1)w^n &= (\partial_x - p_n \text{ mod } m+1)v^{n-1} & \text{on } \Gamma_0.
\end{align*}
\]


Performing again a Fourier analysis in $y$ with the parameter $k$ of this algorithm, we obtain the convergence rate depending on $\mathbf{p} = (p_1, p_2, \ldots, p_m)$

$$\rho(m, \mathbf{p}, \eta, k) = e^{-2\sqrt{k^2 + \eta^2} L} \left( \prod_{j=1}^{m} \left( \frac{\sqrt{\eta^2 + k^2} - p_j}{\sqrt{\eta^2 + k^2} + p_j} \right)^{\frac{1}{2}} \right)^{\frac{1}{2}}.$$ 

To optimize the performance of this new algorithm, the parameters $p_j$, $j = 1, \ldots, m$ in the vector $\mathbf{p}$ have to be the solution of the min-max problem

$$\min_{\mathbf{p} \geq 0} \left( \max_{k_{\min} < k < k_{\max}} \rho(m, \mathbf{p}, \eta, k) \right).$$

This optimization problem has to be solved numerically in general, but for $L = 0$ and $m = 2^l$ it has an elegant solution in closed form for the ADI method in [19].

**Theorem 4.1 (Wachspress (1962))** If $m = 2^l$ then the optimal choice for the parameters $p_j$, $j = 1, 2, \ldots, m$ is given by

$$p_j = \alpha_{0,j}, \quad j = 1, 2, \ldots, m$$

(4.2)

where the $\alpha_{0,j}$ are recursively defined using the forward recursion

$$\begin{align*}
x_0 &= \sqrt{\eta + k_{\min}^2}, & x_{i+1} &= \frac{\sqrt{x_i y_i}}{x_i + y_i} \quad i = 0, 1, \ldots, l \\
y_0 &= \sqrt{\eta + k_{\max}^2}, & y_{i+1} &= \frac{x_i y_i}{x_i + y_i}
\end{align*}$$

(4.3)

and the backward recursion

$$\begin{align*}
\alpha_{i,2j-1} &= \alpha_{i+1,j} - \sqrt{\alpha_{i+1,j}^2 - x_i y_i} \\
\alpha_{i,2j} &= \alpha_{i+1,j} + \sqrt{\alpha_{i+1,j}^2 - x_i y_i}
\end{align*}$$

(4.4)

where $i = l-1, l-2, \ldots, 0$ and $j = 1, 2, \ldots, 2^{l-i-1}$ for each $i$. The convergence rate obtained with these parameters is given by

$$\max_{k_{\min} < k < k_{\max}} |\rho(k, m)| = \left( \frac{\sqrt{y_i} - \sqrt{x_i}}{\sqrt{y_i} + \sqrt{x_i}} \right)^{\frac{1}{2}}.$$ (4.5)

**Proof.** The proof uses the equioscillation property of the optimum similar to the case of the Chebyshev polynomials and is due to Wachspress in [19]. An elegant version of the proof can be found in Varga [18].

In Figure 4.1 we show how the optimal choice of an increasing number of parameters $p_j$ affects the convergence rate of the optimized Schwarz method. From Figure 4.1 on the right we see that the more optimization parameters we use, the weaker the dependence on $h$ of the convergence rate becomes. This indicates that we can define a sequence of non-overlapping optimized Schwarz methods with an arbitrarily weak dependence of the convergence rate on the mesh parameter $h$ using $m$ different constants in the Robin transmission conditions. To prove this result, we first need the following
Figure 4.1: On the left dependence of $|\rho_{opt}|$ on the frequency $k$ when 1, 2, 4, 8 and 16 optimization parameters are used on a fixed range of frequencies, $k_{\text{max}} = 100\pi$, and on the right dependence of $1-\rho_{\text{max}}$ on $h$ as $h$ goes to zero for 1, 2, 4, 8 and 16 optimization parameters.

**Lemma 4.1** For $k_{\text{max}} = \pi/h$ the recursively defined $x_i$ and $y_i$ in equation (4.3) have for $h$ small the asymptotic expansion

$$
x_i = 2^{2-(i-1)\frac{2}{\pi h}} \left( \eta + k_{\text{min}}^2 \right)^{\frac{1}{2\pi h}} \left( \frac{\pi}{h} \right)^{1-\frac{1}{2h}} + O\left( \left( \frac{1}{h} \right)^{1-\frac{1}{2h}} \right) \quad i = 0, 1, \ldots \quad (4.6)
$$

**Proof.** The proof is done by induction. For $i = 0$, we have

$$
x_0 = \sqrt{\eta + k_{\text{min}}^2}, \quad y_0 = \sqrt{\eta + \left( \frac{\pi}{h} \right)^2} = \frac{\pi}{h} + O(h).
$$

Now we assume that (4.6) holds for $i$ and compute for $i+1$, using the recursive definition (4.3) first for $x_{i+1}$

$$
x_{i+1} = \sqrt{x_i y_i}
$$

$$
= \sqrt{2^{2-(i-1)\frac{2}{\pi h}} \left( \eta + k_{\text{min}}^2 \right)^{\frac{1}{2\pi h}} \left( \frac{\pi}{h} \right)^{1-\frac{1}{2h}} + O\left( \left( \frac{1}{h} \right)^{1-\frac{1}{2h}} \right)}
$$

$$
= \sqrt{2^{2-(i+1)\frac{2}{\pi h}} \left( \eta + k_{\text{min}}^2 \right)^{\frac{1}{2\pi h}} \left( \frac{\pi}{h} \right)^{1-\frac{1}{2h}} + O\left( \left( \frac{1}{h} \right)^{1-\frac{1}{2h}} \right)}
$$

$$
= 2^{2-(i+1)\frac{2}{\pi h}} \left( \eta + k_{\text{min}}^2 \right)^{\frac{1}{2\pi h}} \left( \frac{\pi}{h} \right)^{1-\frac{1}{2h}} \sqrt{1 + O(h^{2-\frac{1}{2h}})}
$$

and then for $y_{i+1}$

$$
y_{i+1} = \frac{x_{i+1}}{2^{2-(i+1)\frac{2}{\pi h}} \left( \eta + k_{\text{min}}^2 \right)^{\frac{1}{2\pi h}} \left( \frac{\pi}{h} \right)^{1-\frac{1}{2h}} + O\left( \left( \frac{1}{h} \right)^{1-\frac{1}{2h}} \right) + O\left( \left( \frac{1}{h} \right)^{1-\frac{1}{2h}} \right)}
$$

$$
= \frac{1}{2^{2-(i+1)\frac{2}{\pi h}} \left( \eta + k_{\text{min}}^2 \right)^{\frac{1}{2\pi h}} \left( \frac{\pi}{h} \right)^{1-\frac{1}{2h}} + O\left( \left( \frac{1}{h} \right)^{1-\frac{1}{2h}} \right)} + O\left( \left( \frac{1}{h} \right)^{1-\frac{1}{2h}} \right)
$$

which completes the induction. 

The asymptotic convergence rates for small mesh parameter $h$ are given in the following
Theorem 4.2 The non-overlapping optimized Schwarz method (4.1) with \( m = 2^l \) optimally chosen parameters \( p_j, j = 1, 2, \ldots, m \) in the Robin transmission conditions according to (4.2) has for small mesh parameter \( h \) the asymptotic convergence rate

\[
\rho_{opt} = 1 - 2^{-1 - \frac{1}{2} (\eta + k_{\text{max}}^2)} \frac{m^{-\eta}}{\eta^{\frac{1}{m}}} h^{\frac{1}{2m}} + O(h^{\frac{1}{m}}). \tag{4.7}
\]

Proof. We first need the asymptotic expansions of the square roots of \( x_l \) and \( y_l \) given in Lemma 4.1,

\[
\sqrt{x_l} = 2^{-1 - \frac{1}{2} (\eta + k_{\text{max}}^2)} \left( \frac{\pi}{2} \right)^{\frac{1}{2}} h^{\frac{1}{2m}} + O(h^{\frac{1}{2m}})
\]

\[
\sqrt{y_l} = \left( \frac{\pi}{2} \right)^{\frac{1}{2}} h^{\frac{1}{2m}} + O(h^{\frac{1}{2m}}).
\]

Inserting these expansions into the expression for the optimized convergence rate (4.5) of Theorem 4.1 we obtain

\[
\rho_{opt} = \left( \frac{\sqrt{x_l} - \sqrt{y_l}}{\sqrt{x_l} + \sqrt{y_l}} \right)^{\frac{1}{2m}} = \left( \frac{1 - 2^{-1 - \frac{1}{2} (\eta + k_{\text{max}}^2)} \left( \frac{\pi}{2} \right)^{\frac{1}{2}} h^{\frac{1}{2m}} + O(h^{\frac{1}{2m}})}{1 + 2^{-1 - \frac{1}{2} (\eta + k_{\text{max}}^2)} \left( \frac{\pi}{2} \right)^{\frac{1}{2}} h^{\frac{1}{2m}} + O(h^{\frac{1}{2m}})} \right)^{\frac{1}{2m}}
\]

\[
= \left( 1 - 2^{-1 - \frac{1}{2} (\eta + k_{\text{max}}^2)} \left( \frac{\pi}{2} \right)^{\frac{1}{2}} h^{\frac{1}{2m}} + O(h^{\frac{1}{2m}}) \right)^{\frac{1}{2m}} = 1 - 2^{-1 - \frac{1}{2} (\eta + k_{\text{max}}^2)} \left( \frac{\pi}{2} \right)^{\frac{1}{2}} h^{\frac{1}{2m}} + O(h^{\frac{1}{2m}})
\]

and the result follows by noting that \( m = 2^l \).

Hence increasing \( m \) we can achieve an as weak dependence of the convergence rate on the mesh parameter \( h \) as we like. The numerical experiments in the next section show that this result also holds for the discretized algorithm.

5. Numerical Experiments. We perform all our computations on a bounded domain for the model problem

\[
\mathcal{L}(u) := (\eta - \Delta)u = f \quad \text{in } \Omega = [0, 1]^2
\]

with homogeneous Dirichlet boundary conditions. We decompose the domain into two non-overlapping subdomains \( \Omega_1 = [0, \frac{1}{2}] \times [0, 1] \) and \( \Omega_2 = [\frac{1}{2}, 1] \times [0, 1] \) and apply the optimized non-overlapping Schwarz method for various values of the parameter \( m \). We simulate directly the error equations, i.e. \( f = 0 \), and we show the results for \( \eta = 1 \). To solve the subdomain problems, we use the standard five point finite difference discretization with uniform mesh spacing \( h \) in both the \( x \) and \( y \) directions.

We start the iteration with a random initial guess so that it contains all the frequencies on the given mesh and we iterate until the relative residual is smaller than \( 10^{-6} \). Table 5.1 shows the number of iterations required as one refines the mesh parameter \( h \).

There are two important things to notice: first one can see that the dependence of the number of iterations gets weaker as \( m \) becomes larger, as predicted by the analysis. Second for small \( m \), using Krylov acceleration leads to significant improvement in the performance, whereas for bigger \( m \), the improvement is almost negligible. Schwarz by itself is already such a good solver that Krylov acceleration is not needed. This is a property also observed for multi-grid methods applied to this problem. To see the dependence of the convergence rate on \( h \) more clearly, we plotted in Figure 5.1 the
Table 5.1: Dependence on $h$ and $m$ of the number of iterations when the optimized Schwarz method is used as a solver or as a preconditioner for a Krylov method.

<table>
<thead>
<tr>
<th></th>
<th>Schwarz as a solver</th>
<th>Schwarz as a preconditioner</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h$</td>
<td>$m = 1$</td>
<td>$m = 2$</td>
</tr>
<tr>
<td>$1/50$</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>$1/100$</td>
<td>34</td>
<td>8</td>
</tr>
<tr>
<td>$1/200$</td>
<td>48</td>
<td>10</td>
</tr>
<tr>
<td>$1/400$</td>
<td>68</td>
<td>12</td>
</tr>
<tr>
<td>$1/800$</td>
<td>95</td>
<td>14</td>
</tr>
</tbody>
</table>

Figure 5.1: Asymptotic behavior of the optimized Schwarz method, on the left used as an iterative solver and on the right as a preconditioner.

number of iterations together with the asymptotic rates expected from our analysis. One can see that the asymptotic analysis predicts very well the numerically observed results. One even gains almost the additional square-root from the Krylov method when Schwarz is used as a preconditioner.

Finally we emphasize that the number of iterations given in Table 5.1 is the number of times we cycled through all parameter values. In the current implementation therefore the cost of one iteration with $m = 4$ is four times the cost of one iteration with $m = 1$. But note that not each iteration of the Schwarz method needs the same resolution now, since it only needs to be effective in the frequency range around the corresponding $p_j$. The values of $p_j$ for $m = 4$ with $h = 1/400$ are for example $p_1 = 4.78$, $p_2 = 25.85$, $p_3 = 160.26$ and $p_4 = 866.71$. Hence the solve with $p_1$ in the transmission condition can be done on a very coarse grid, the one with $p_2$ on quite a coarse grid, the one with $p_2$ on an intermediate grid and only the solve with $p_4$ needs to be on a fine grid. In addition we do not need to solve exactly; it is only required to reduce the error in the corresponding frequency range. Using a relaxation iteration, which leads to an algorithm with natural inner and outer iterations. Doing this, the cost for arbitrary $m$ will be only a constant times the cost for $m = 1$ and hence the number of iterations we gave become the relevant ones to compare. Furthermore in that case, the factor $1/m$ in the asymptotic convergence rate (4.7) disappears because
now the relevant quantities to compare are $\rho_{\text{opt}}^m$, and hence one can obtain a convergence rate independent of $h$ by choosing the number $m$ like the logarithm of $1/h$ as $h$ is refined. Such an algorithm then has the key properties of multigrid, but is naturally parallel like the Schwarz algorithm.

REFERENCES


