Acceleration of the Convergence of the Asynchronous RAS Method

D. Tromeur-Dervout

1 Introduction

Nowadays high performance computers have several thousand cores and more and more complex hierarchical communication networks. For these architectures, the use of a global reduction operation such as the dot product involved in the GMRES acceleration can be a bottleneck for the performance. In this context domain decomposition’s solvers with local communications are becoming particularly interesting. Nevertheless, the probability of temporarily failures/unavailability of a set of processors/clusters is non-zero, which leads to the need for fault tolerant algorithms such as asynchronous Schwarz type’s methods. With the asynchronism the transmission conditions (TC) at artificial interfaces generated by the domain decomposition may not have been updated for some subdomains and for some iterations. The message passing interface MPI-3 standard provides one-sided communication protocol where a process can directly write on the local memory of another process without synchronizing. This can also occur in the OpenMP implementation. For asynchronous methods, it is very difficult to know if the update has been performed and most papers fail to give the level of asynchronism in their implementation results.

From the numerical point of view, this asynchronism affects the linear operator of the interface problem. In this context Aitken’s acceleration of the convergence should not be applicable as it is based on the pure linear convergence of the DDM [6] [10] [11], i.e. there exists a linear operator $P$ independent of the iteration that connects the error at the artificial interfaces of two consecutive iterations. This paper focuses on Aitken’s acceleration of the convergence of the asynchronous Restricted Additive Schwarz (RAS) iterations. We develop a mathematical model of the Asynchronous RAS allowing us to set the percentage of the number of randomly chosen local artificial interfaces where transmission conditions are not updated. Then we show how this ratio deteriorates the convergence of the Asynchronous RAS and how some

1 University of Lyon, UMR5208 U.Lyon1-CNRS, Institut Camille Jordan, e-mail: Damien.tromeur-derout@univ-lyon1.fr
regularization techniques on the traces of the iterative solutions at artificial interfaces allow us to accelerate the convergence to the true solution.

The plan of the paper is the following. Section 2 gives the notation and the principles of the Aitken-Schwarz method using some low-rank approximation of the interface error operator. Section 3 presents the modeling of the asynchronous RAS on a 2D Poisson problem allowing us to define the level of asynchronism. Section 4 present the results of the acceleration with respect to the level of asynchronism and the enhancement of this acceleration with regularization techniques before concluding in section 5.

2 Aitken-Schwarz method principles

By adapting the notations of [3], we consider a non-singular matrix $A \in \mathbb{R}^{n \times n}$ having a non-zero pattern and the associated graph $G = (W, F)$, where the set of vertices $W = \{1, \ldots, n\}$ represents the $n$ unknowns and the edge set $F = \{(i, j)|a_{i,j} \neq 0\}$ represents the pairs of vertices that are coupled by a nonzero element in $A$. Then we assume that a graph partitioning has been applied and has resulted in $N$ non-overlapping subsets $W_i^p$ whose union is $W$. Let $W_i^p$ be the $p$-overlap partition of $W$, obtained by including all the immediate neighboring vertices of the vertices from $W_i^{p-1}$. Let $W_i^{p} = W_i^{p+1} \setminus W_i^{p}$. Then let $R_i^p \in \mathbb{R}^{n_i \times n}$ ($R_i^{p,e} \in \mathbb{R}^{n_i \times n_i}$ and $R_0^0 \in \mathbb{R}^{n_i \times n_i}$ respectively) be the operator that restricts $x \in \mathbb{R}^n$ to the components of $x$ belonging to $W_i^p$ ($W_i^{p,e}$ and $W_0^0$ respectively, and the operator $R_0^0 \in \mathbb{R}^{n_i \times n}$ puts 0 to those unknowns belonging to $W_i^{p} \setminus W_0^0$). We define the operators $A_i = R_i^0 A R_i^{pT}$ and $E_i = R_i^0 A R_i^{pT}$, the vectors $x_i = R_i^p x$, $b_i = R_i^p b$, and $x_{i,e} = R_i^{p,e} x$, then the RAS iteration $k + 1$ writes locally for the partition $W_i^p$:

$$A_i^{-1} b_i - E_i x_{i,e}.$$  

By defining $M_{RAS}^{-1} \overset{\text{def}}{=} \sum_{i=0}^{N-1} R_i^{0T} A_i^{-1} R_i^p$ and adding the contribution of each partition $W_i^p$, RAS can be viewed as a Richardson’s process:

$$\sum_{i=0}^{N-1} R_i^{0T} R_i^p x^{k+1} = \sum_{i=0}^{N-1} R_i^{0T} A_i^{-1} R_i^p b - \sum_{i=0}^{N-1} R_i^{0T} A_i^{-1} R_i^p A R_i^{pT} x^k,$$

$$x^{k+1} = M_{RAS}^{-1} b - A x^k + x^k = x^k + M_{RAS}^{-1} (b - A x^k).$$

The Richardson’s process (3) is deduced from (2) (see [5, Theorem 3.7]) with using the property $R_i^p A = R_i^0 A (R_i^{pT} R_i^p + R_i^{pT} R_i^{p,e})$. It can be reduced to a problem with the unknowns on the interface (see [12, eq. (2.12) and (2.13)]).

The restriction of (3) to the interface $\Gamma = \{W_{0,e}^0, \ldots, W_{N-1,e}^0\}$ of size $n_\Gamma = \sum_{i=0}^{N-1} n_{i,e}$, by defining $R_\Gamma = (R_{0,e}^0, \ldots, R_{N-1,e}^0)^T \in \mathbb{R}^{n_\Gamma \times n}$ and by using the
property $R_{i,e}^T R_{i,e} R_{i,e}^T R_{i,e} = R_{i,e}^T R_{i,e}$, writes:

$$R_{i,e} x^{k+1} = R_{i,e} \left( I - M_{RAS}^{-1} A \right) R_{i,e}^T R_{i,e} x^k + R_{i,e} M_{RAS}^{-1} b.$$  \hspace{1cm} (4)

The pure linear convergence of the RAS at the interface given by $\gamma = y^k - y^\infty$ (the error operator $P$ does not depend of the iteration $k$) allows to apply the Aitken’s acceleration of the convergence technique to obtain the true solution $y^\infty$ on the interface $\Gamma$: $y^\infty = (I - P)^{-1}(y^k - P y^{k-1})$, and thus after another local resolving, the true solution $x^\infty$. Let us note that we can accelerate the convergence to the solution for a convergent or a divergent iterative method. The only need is that 1 is not one of the eigen values of $P$. Considering $e^k = y^k - y^{k-1}$, $k = 1, \ldots,$ the operator $P \in \mathbb{R}^{m \times m}$ can be computed algebraically after $n_\gamma + 1$ iterations as $P = [e_1, \ldots, e_{n_\gamma}] [e_1, \ldots, e_{n_\gamma}]^{-1}$. Nevertheless, for 2D or 3D problems, the value $n_\gamma$ may be too large to have an efficient method. So a low-rank approximation of $P$ is computed using the iterated interface solutions and the Aitken’s acceleration is performed on the low-rank space of dimension $n_\gamma \ll n_\Gamma$. As we search the converged interface solution $y^\infty$, we build from the singular value decomposition \cite{9} of the matrix $Y = [y^0, \ldots, y^\gamma] = U \Sigma V^T$ a low-rank space with selecting the $n_\gamma$ singular vectors associated to the most significant singular values.

**Algorithm 1** Approximated Aitken’s acceleration

**Require:** $x^0$ an arbitrary initial condition, $\epsilon > 0$ a given tolerance, $y^0 = R_{i,e} x^0$.

1. **repeat**
2.  **for** $k = 1, \ldots, q$ **do**
3.   $x^k = x^{k-1} + M_{RAS} (b - A x^{k-1})$, $y^k = R_{i,e} x^k$ // RAS iteration
4.  **end for**
5.  Compute SVD of $[y^0, y^1, \ldots, y^q] = U \Sigma V^T$, keep the $n_\gamma$ singular vectors $U_{1:n_\gamma}$ such that $\sigma_{n_\gamma+1} < \epsilon$
6.  Compute $[\hat{\gamma}^{n_\gamma+2}, \ldots, \hat{\gamma}^q] = U_{1:n_\gamma} [\hat{\gamma}^{n_\gamma+2}, \ldots, \hat{\gamma}^q]$, and $\hat{\gamma}^k = \hat{\gamma}^k - \hat{\gamma}^{k-1}$
7.  Compute $\hat{P} = [\hat{\gamma}^{n_\gamma+1}, \ldots, \hat{\gamma}^q] [\hat{\gamma}^{n_\gamma+1}, \ldots, \hat{\gamma}^q]^{-1}$
8.  $y^0 \leftarrow U_{1:n_\gamma} \left( I - \hat{P} \right)^{-1} \left( \hat{\gamma}^q - \hat{P} \hat{\gamma}^{q-1} \right)$
9.  **until** convergence

This low-rank approximation of the acceleration has been very efficient to solve 3D Darcy flow with highly heterogeneous and randomly generated permeability field \cite{1}. Step 7 of the algorithm may be subject to bad conditioning and matrix inversion can be replaced by pseudo inverse. Other techniques developed in \cite{1} avoid the matrix inversion. For 1D partitioning (i.e $\forall j \in \{0, \ldots, N - 1\}$, $W_{i,e}^P \cap W_{j}^Q = \emptyset$, $\forall j \neq \{i - 1, i + 1\}$), we can use the sparsity of $P$ to define a Sparse-Aitken acceleration, numerically more efficient by using local SVD for each subdomain \cite{2}. 

---

**Acceleration of the Convergence of the Asynchronous RAS Method**

727
3 Modeling the Asynchronous RAS

If the Schwarz DDM converges then the asynchronous Schwarz does the same [8, Theorem 5 with assumption 2], under the additional hypothesis that the TC have been generated before their use, no subdomain stop updating its components and no subdomain have a TC that is never updated.

We consider the 2D Poisson problem:

$$\begin{cases}
-\left(\frac{\partial^2}{\partial z_1^2} + \frac{\partial^2}{\partial z_2^2}\right)x(z_1, z_2) = b(z_1, z_2), \quad (z_1, z_2) \in ]0, 1[ \times ]0, 1[, \\
\text{with homogeneous Dirichlet B.C.}
\end{cases}$$

(5)

We discretize (5) with second order centered finite differences on a regular Cartesian mesh of \( n_{z_1}^g \times n_{z_2}^g = n \) points.

Given a non-prime number \( N \in \mathbb{N} \), we split the domain \([0, 1]^2\) in \( N = n_{z_1} \times n_{z_2} \) overlapping partitions \( W_i^P \). For the sake of simplicity, we consider that each partition \( W_i^P \) has \( n_i = n_{z_1}^i \times n_{z_2}^i \) points of discretizing and we define \( n_{z_1}^i \) and \( n_{z_2}^i \) accordingly. Due to the Cartesian mesh discretizing, the set \( W_i^P \), for each \( i \), can be split in a maximum of four parts corresponding to the four local artificial interfaces generated by the partitioning. Two: \( W_i^{O_i^p} \) and \( W_i^{E_i^p} \) (respectively \( W_i^{S_i^p} \) and \( W_i^{N_i^p} \)) are in the \( z_1 \) (respectively \( z_2 \)) direction.

The asynchronous RAS algorithm does not wait that the updates of the transmission conditions (TC) (the term \( E_i^p x^k \) in (1)) are done before starting the next iteration. Consequently, the TC of one partition could have not been totally or partially updated. As there is not control on the restraining of the communication network, it is difficult to evaluate the number of update of the local TC that are missing.

In order to modelize the asynchronous RAS, we propose a model where each of the four TC of each subdomains are totally update or not, following a random draw of four numbers \((l_{O_i}^p, l_{W_i}^p, l_{S_i}^p, l_{N_i}^p)\) per \( W_i^P \). Only if a draw associated to a local TC is greater than a fixed limit \( l \) then this local TC is updated. The value \( l \) gives the percentage of missing TC updates. The synchronous RAS algorithm is obtained setting \( l = 0 \) and we note \( l\text{-RAS} \) the asynchronous RAS with a \( l \) level of asynchronism. The \( l\text{-RAS} \) iterates until \( R_t x^k \) does not evolve anymore. Figure 1 (left) shows that the level of asynchronism deteriorates the convergence of the RAS. The error between two consecutive iterations oscillates quite strongly with \( l \). These oscillations are smoother for the error with the true solution. Table 1 shows the log10 of the error with the true solution of the asynchronous \( l\text{-RAS} \) for 240 iterations and the associated Aitken’s acceleration of the convergence. The results for \( l\text{-RAS} \), with respect to the asynchronism level \( l \), have an increasing variance but the min, max and mean values of the error are close. The Aitken’s acceleration of the convergence, using the set of 240 \( l\text{-RAS} \) iterations, still accelerates even at a high level \( l \) of asynchronism, even though the acceleration deteriorates with increasing \( l \). Those results have a more stable variance and the mean value is closer to the max value than to the min value. We limited \( n_T \) to be 40 for \( l \neq 0 \) and to be 20 for \( l = 0 \).
Acceleration of the Convergence of the Asynchronous RAS Method

Fig. 1: $l$-RAS convergence with respect to the level of asynchronism $l$: for two consecutive iterations (continuous line) and (left) with the true solution (+), (right) two consecutive iterations after Cesaro’s summation (+). $(n_{z_1} = n_{z_2} = 10, N_{z_1} = N_{z_2} = 5, n_y = 40)$

due to the strong decreasing of the firsts singular values. Let us notice for this test case $n_l = 544$ and the low-rank space is of size $n_y = 40$.

<table>
<thead>
<tr>
<th>$l$</th>
<th>Aitken $l$-RAS</th>
<th>$l$-RAS</th>
<th>Update failures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>min</td>
<td>max</td>
<td>mean</td>
</tr>
<tr>
<td>0.0%</td>
<td>-11.12</td>
<td>-11.12</td>
<td>-11.12</td>
</tr>
<tr>
<td>0.5%</td>
<td>-3.666</td>
<td>-5.839</td>
<td>-4.969</td>
</tr>
<tr>
<td>1.0%</td>
<td>-2.814</td>
<td>-5.440</td>
<td>-4.751</td>
</tr>
<tr>
<td>5.0%</td>
<td>-2.521</td>
<td>-5.023</td>
<td>-4.284</td>
</tr>
<tr>
<td>10%</td>
<td>-1.729</td>
<td>-4.707</td>
<td>-3.956</td>
</tr>
<tr>
<td>30%</td>
<td>-1.037</td>
<td>-4.005</td>
<td>-3.280</td>
</tr>
<tr>
<td>50%</td>
<td>0.548</td>
<td>-3.613</td>
<td>-2.643</td>
</tr>
</tbody>
</table>

Table 1: Statistics (min,max,mean and variance $\sigma$), based on 100 runs, of $log_{10}(||x^{240} - x^\infty||_\infty)$, with respect to $l$, for the asynchronous $l$-RAS and its Aitken’s acceleration of the convergence (with the same data). $(n_{z_1} = n_{z_2} = 10, N_{z_1} = N_{z_2} = 5, n_y = 40)$

4 Regularization of the Aitken acceleration of the convergence of the Asynchronous RAS

At first glance, previous results on Aitken’s acceleration of the convergence of the $l$-RAS are surprising as the pure linear convergence of the RAS is destroyed with the asynchronism, i.e. the error operator depends of the iteration: $y^{k+1} - y^k = P_k(y^k - y^{k-1})$. The explanation comes from the low-rank space built with the SVD. Let $Y_l = [y_0^l, \ldots, y_t^l]$ be the matrix of the iterated $l$-RAS interface solutions. As the asynchronous $l$-RAS converges, we can write $Y_l = Y_0 + E_l$ where $E_l$ is a perturbation matrix with smaller and smaller entries with respect to the iterations. Then using the Fan inequality [4, Theorem 2, p.764] of the SVD of a perturbation matrix, we have:
Table 2: Statistics (min, max, mean and variance \( \sigma \)) for 100 runs of \( \log_{10} \) of the error with the true solution of the Aitken acceleration of the convergence of l-RAS with Cesaro’s mean with respect to the asynchronism level \( l \). \( (n_{c1} = n_{c2} = 10, N_x = N_y = 5, n_y = 40, m = 200) \)

<table>
<thead>
<tr>
<th>( l )</th>
<th>( \text{min} )</th>
<th>( \text{max} )</th>
<th>( \sigma )</th>
<th>( \text{mean} )</th>
<th>( \text{bound} )</th>
<th>( \text{mean} )</th>
<th>( \text{mean} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0%</td>
<td>-12.42</td>
<td>-12.42</td>
<td>8.3e-15</td>
<td>-12.27</td>
<td>-11.111</td>
<td>-2.543</td>
<td></td>
</tr>
<tr>
<td>0.5%</td>
<td>-4.059</td>
<td>-6.968</td>
<td>4.5e-1</td>
<td>-6.120</td>
<td>-4.969</td>
<td>-2.533</td>
<td></td>
</tr>
<tr>
<td>1.0%</td>
<td>-4.667</td>
<td>-6.636</td>
<td>3.9e-1</td>
<td>-5.902</td>
<td>-4.751</td>
<td>-2.524</td>
<td></td>
</tr>
<tr>
<td>5.0%</td>
<td>-4.184</td>
<td>-6.383</td>
<td>4.9e-1</td>
<td>-5.434</td>
<td>-4.284</td>
<td>-2.443</td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>-3.344</td>
<td>-6.047</td>
<td>4.5e-1</td>
<td>-5.106</td>
<td>-3.956</td>
<td>-2.347</td>
<td></td>
</tr>
<tr>
<td>30%</td>
<td>-3.457</td>
<td>-5.261</td>
<td>3.9e-1</td>
<td>-4.431</td>
<td>-3.280</td>
<td>-1.974</td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>-2.505</td>
<td>-4.553</td>
<td>4.7e-1</td>
<td>-3.794</td>
<td>-2.643</td>
<td>-1.678</td>
<td></td>
</tr>
</tbody>
</table>

\( \sigma_{r+s+1}(Y_0 + E_l) \leq \sigma_{r+s+1}(Y_0) + \sigma_{s+1}(E_l) \) with \( r, s \geq 0, r + s + 1 \leq q + 1 \).

Setting \( s = 0 \), we have \( |\sigma_{r+s+1}(Y_0 + E_l) - \sigma_{r+s+1}(Y_0)| \leq \sigma_1(E_l) = ||E_l||_2, \forall r \leq q \). By using the Schmidt’s Theorem [7, Theorem 2.5.3] on the SVD approximation, we can write:

\[
\min_{X, \text{rank} X = k} (||Y_l - X||_2) = \sigma_{k+1}(Y_l) = \min_{X, \text{rank} X = k} (||Y_l - Y_0 + Y_0 - X||_2) \\
\leq ||Y_l - Y_0||_2 + \min_{X, \text{rank} X = k} ||Y_0 - X||_2 \\
\leq \sigma_1(E_l) + \sigma_{k+1}(Y_0)
\]

This result implies that:

- the low-rank space \( U_l \) built from \( Y_l \) is an approximation of \( U_0 \) with a small perturbation \( ||E_l||_2 = \sigma_1(E_l) \).
- As \( \lim_{k \to \infty} y_t^k \to y^{\infty} \), the perturbation matrix \( E_l \) has its columns with a decreasing 2-norm. Thus, a better acceleration is obtained with considering the last \( q \) iterations to build \( U_l \).

This last result suggests an improvement of the Aitken’s acceleration of the convergence with the Cesaro’s mean of the iterated interface solutions. We transform the sequence \( (y_t) \) in another sequence \( (\tilde{y}_t) \) defined as \( \tilde{y}_t^i = \frac{1}{m} \sum_{j=0}^{m-1} y_{t+j}^{i+j} \). The summation still preserves the pure linear convergence of the synchronous 0%-RAS: \( \tilde{y}_0^{k+1} - y^{\infty} = P(y_0^{k+1} - y^{\infty}) \) and will smooth the perturbation \( E_l \). Figure 1 (right) shows the log10 of the error with the true solution of the iterated interface solution with the Cesaro’s mean with \( m = 200 \). This last allows to smooth the error oscillations on the convergence of l-RAS. The difference between two consecutive iterations of the sequence \( (\tilde{y}_t) \) has a smaller amplitude than for the original sequence \( (y_t) \). This leads to have a low-rank space \( U_l \) built from this \( (\tilde{y}_t) \) more representative of the space where the true solution lives.

Table 2 gives the statistics for 100 runs of the Aitken’s acceleration of the convergence for the l-RAS using the Cesaro’s mean with respect to \( l \). The acceleration of the convergence is enhanced using \( (\tilde{y}_t) \) than \( (y_t) \). The variance and the amplitude between the min and the max values of the results are smaller. Even the 0%-RAS is
better accelerated. Moreover, it shows a upper bound for the mean acceleration of 
the $l$-RAS with the Césaro’s mean to be $\frac{1}{\sqrt{m}}$ the mean acceleration of the $l$-RAS. 
Figure 2 gives the singular values $(\sigma_i)$ of the SVD of $Y_l$ obtained with $l$-RAS with 
respect to the level $l$ of asynchronism. It shows that the fast decreasing of $(\sigma_i)$ is 
lost with the asynchronism. It still exhibits some decreasing of $(\sigma_i)$ that allows the 
Aitken’s acceleration of the convergence. The right figure shows that even with a 
very small level $l$ of asynchronism, the decreasing of $\sigma_l$ is deteriorated even with 
few TC update failures (the total number of update for 300 0%-RAS iterations is 
$300 \times (4 \times 2 + 12 \times 3 + 9 \times 4) = 24000$).

5 Conclusion

We have succeed to accelerate the asynchronous RAS with the Aitken’s acceleration 
of the convergence technique based on the low-rank approximation of the error 
operator with the SVD of the matrix of interface iterated solutions. The SVD allows 
to smooth the asynchronous effect over the iterations. We proposed a modeling for 
setting the level of asynchronism. It can be used to estimate the asynchronism in 
real application. Knowing the observed convergence rate of the real application, 
we can extrapolate the level of asynchronism of the implementation. The model 
proposed here considers a uniform probability for TC update failure (the worst case) 
but we also can consider that only certain parts of the domain decomposition may 
be temporarily at fault. Finally, we proposed a regularisation technique based on the
Césaro’s mean of the $l$-RAS iterated interface solutions that improves the Aitken’s acceleration of the convergence even on the synchronous RAS.

References