Acceleration of the Convergence of the Asynchronous RAS Method

D. Tromeur-Dervout[1]

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/unavailabilty 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 an 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 consecutives 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-dervout@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 regularisation 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 $W = \{1, \dots, n\}$ represents the n unknowns and the edge set $F = \left\{(i, j)|a_{i,j} \neq 0\right\}$ 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 nonoverlapping subsets W_i^0 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 be $W_{i,e}^p = W_i^{p+1}\backslash W_i^p$. Then let $R_i^p \in \mathbb{R}^{n_i\times n}$ ($R_{i,e}^p \in \mathbb{R}^{n_{i,e}\times n}$ and $\tilde{R}_i^0 \in \mathbb{R}^{n_i\times n}$ respectively) be the operator that restricts $x \in \mathbb{R}^n$ to the components of x belonging to W_i^p ($W_{i,e}^p$ and W_i^0 respectively, and the operator $\tilde{R}_i^0 \in \mathbb{R}^{n_i\times n}$ puts 0 to those unknowns belonging to $W_i^p\backslash W_i^0$). We define the operators $A_i = R_i^p A R_i^{pT}$ and $E_i = R_i^p A R_{i,e}^{pT}$, the vectors $x_i = R_i^p x$, $b_i = R_i^p b$, and $x_{i,e} = R_{i,e}^p x$, then the RAS iteration $k+1$ writes locally for the partition W_i^p:

$$x_i^{k+1} = A_i^{-1}(b_i - E_i x_{i,e}^k). \tag{1}$$

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

$$\sum_{i=0}^{N-1} \tilde{R}_i^{0T} R_i^p x^{k+1} = \sum_{i=0}^{N-1} \tilde{R}_i^{0T} A_i^{-1} R_i^p b - \sum_{i=0}^{N-1} \tilde{R}_i^{0T} A_i^{-1} R_i^p A R_{i,e}^{pT} x^k, \tag{2}$$

$$x^{k+1} = M_{RAS}^{-1} b - M_{RAS}^{-1} A x^k + x^k = x^k + M_{RAS}^{-1}(b - Ax^k). \tag{3}$$

The Richardson's process (3) is deduced from (2) (see [5, Theorem 3.7]) with using the property $R_i^p A = R_i^p A(R_i^{pT} R_i^p + R_{i,e}^{pT} R_{i,e}^p)$. 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 $\Gamma = \left\{W_{0,e}^p, \dots, W_{N-1,e}^p\right\}$ of size $n_\Gamma = \sum_{i=0}^{N-1} n_{i,e}$, by defining $R_\Gamma = (R_{0,e}^p, \dots, R_{N-1,e}^p)^T \in \mathbb{R}^{n_\Gamma\times n}$ and by using the property $R_{i,e}^{pT} R_{i,e}^{p} R_{\Gamma}^{T} R_{\Gamma} = R_{i,e}^{pT} R_{i,e}^{p}$, writes:

$$\underbrace{R_\Gamma x^{k+1}}_{y^{k+1}} = \underbrace{R_\Gamma \left(I - M_{RAS}^{-1} A\right) R_\Gamma^T}_{P} \underbrace{R_\Gamma x^k}_{y^k} + \underbrace{R_\Gamma M_{RAS}^{-1} b}_{c} . \tag{4}$$

The pure linear convergence of the RAS at the interface given by : $y^k - y^\infty = P(y^{k-1} - y^\infty)$ (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^∞ on the interface Γ: $y^\infty = (I - P)^{-1}(y^k - Py^{k-1})$, and thus after another local resolving, the true solution x^∞. 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 $P \in \mathbb{R}^{n_\Gamma \times n_\Gamma}$ can be computed algebraically after $n_\Gamma + 1$ iterations as $P = [e^{n_\Gamma+1}, \ldots, e^2][e^{n_\Gamma}, \ldots, e^1]^{-1}$. Nevertheless, for 2D or 3D problems, the value n_Γ 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^∞, we build from the singular value decomposition [9] of the matrix $Y = [y^0, \ldots y^q] = U\Sigma V^T$ a low-rank space with selecting the n_γ 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, $y^0 = R_\Gamma x^0$,
1: **repeat**
2: **for** $k = 1 \ldots q$ **do**
3: $x^k = x^{k-1} + M_{RAS}^{-1}\left(b - Ax^{k-1}\right)$, $y^k = R_\Gamma x^k$ // RAS iteration
4: **end for**
5: Compute SVD of $[y^0, y^1, \ldots, y^q] = U\Sigma V'$,
keep the n_γ singular vectors $U_{1:n_\gamma}$ such that $\sigma_{n_\gamma+1} < \epsilon$
6: Compute $[\hat{y}^{q-n_\gamma-2}, \ldots, \hat{y}^q] = U_{1:n_\gamma}^T [y^{q-n_\gamma-2}, \ldots, y^q]$, and $\hat{e}^k = \hat{y}^k - \hat{y}^{k-1}$
7: Compute $\widehat{P} = [\hat{e}^{q-n_\gamma} \ldots, \hat{e}^q][\hat{e}^{q-n_\gamma-1}, \ldots, \hat{e}^{q-1}]^{-1}$
8: $y^0 \leftarrow U_{1:n_\gamma}\left(I - \widehat{P}\right)^{-1}\left(\hat{y}^q - \widehat{P}\hat{y}^{q-1}\right)$
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 [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 [1] avoid the matrix inversion. For 1D partitioning (i.e $\forall i = \{0, \ldots N-1\}$, $W_{i,e}^p \cap W_j^0 = \emptyset, \forall j \neq \{i-1, i+1\}$), we can use the sparsity of P to define a Sparse-Aitken acceleration, numerically more efficient by using local SVD for each subdomain [2].

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 hyppothesis 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:

$$\begin{cases} -(\dfrac{\partial^2}{\partial z_1^2}+\dfrac{\partial^2}{\partial z_2^2})\, x(z_1,z_2)=b(z_1,z_2),\ (z_1,z_2)\in]0,1[\times]0,1[, \\ \text{with homogeneous Dirichlet B.C.} \end{cases} \tag{5}$$

We discretize (5) with second order centered finite differences on a regular Cartesian mesh of $n^g_{z_1}\times n^g_{z_2}=n$ 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 $n_i=n^l_{z_1}\times n^l_{z_2}$ points of discretizing and we define $n^g_{z_1}$ and $n^g_{z_2}$ accordingly. Due to the Cartesian mesh discretizing, the set $W^p_{i,e}$, for each i, can be split in a maximum of four parts corresponding to the four local artificial interfaces generated by the partitioning. Two: $W^{O,p}_{i,e}$ and $W^{E,p}_{i,e}$ (respectively $W^{S,p}_{i,e}$ and $W^{N,p}_{i,e}$) 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 $E_i x^k$ 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 $(l_i^O, l_i^W, l_i^S, l_i^N)$ per W_i^p. Only if the 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-RAS the asynchronous RAS with a l level of asynchronism. The l-RAS iterates until $R_\Gamma x^k$ 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-RAS for 240 iterations and the associated Aitken's acceleration of the convergence. The results for l-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-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_γ to be 40 for $l\neq 0$ and to be 20 for $l=0$

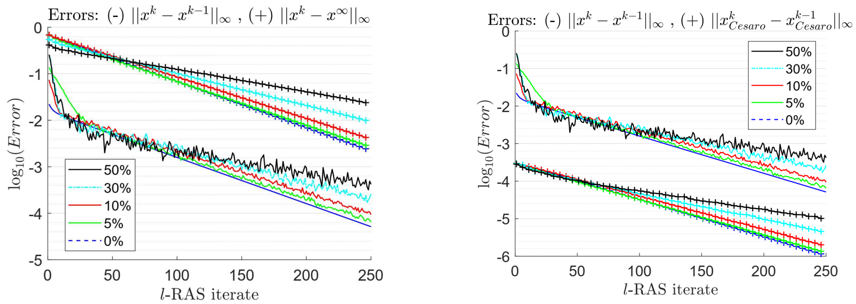

Fig. 1: l-RAS convergence with respect to the level of asynchronism l: for two consecutive iterations (continuous line) and (left) with the true solution (+) , (right) two consecutive iterations after Césaro's summation (+). ($n^l_{z_1} = n^l_{z_2} = 10, N_{z_1} = N_{z_2} = 5, n_\gamma = 40$)

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

	Aitken l-RAS				l-RAS				Update failures			
l	min	max	mean	σ	min	max	mean	σ	min	max	mean	σ
0.0%	-11.12	-11.12	-11.12	2e-14	-2.543	-2.543	-2.543	3e-15	0	0	0	0
0.5%	-3.666	-5.839	-4.969	4.0e-1	-2.527	-2.556	-2.533	4.8e-3	99	145	120.7	9.9
1.0%	-2.814	-5.440	-4.751	4.7e-1	-2.513	-2.544	-2.524	7.1e-3	202	277	239.48	15.6
5.0%	-2.521	-5.023	-4.284	4.2e-1	-2.415	-2.479	-2.443	1.4e-2	1121	1286	1197.3	34.3
10.%	-1.729	-4.707	-3.956	5.3e-1	-2.303	-2.406	-2.347	2.1e-2	2267	2502	2397.9	43.6
30.%	-1.037	-4.005	-3.280	4.6e-1	-1.868	-2.089	-1.974	4.7e-2	7044	7349	7203.3	66.5
50.%	0.548	-3.613	-2.643	6.1e-1	-1.472	-1.961	-1.678	9.3e-2	11860	12199	12013	66.1

Table 1: Statistics (min,max,mean and variance σ), based on 100 runs, of $log10(||x^{240} - x^\infty||_\infty)$, with respect to l, for the asynchronous l-RAS and its Aitken's acceleration of the convergence (with the same data). ($n^l_{z_1} = n^l_{z_2} = 10, N_{z_1} = N_{z_2} = 5, n_\gamma = 40$)

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 $Y_l = [y^0_l, \ldots, y^q_l]$ 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:

	Aitken Césaro l-RAS				Upper	Aitken l-RAS	l-RAS
l	min	max	σ	mean	bound	mean	mean
0.0%	-12.42	-12.42	8.e-15	-12.42	-12.27	-11.111	-2.543
0.5%	-4.059	-6.968	4.5e-1	-6.284	-6.120	-4.969	-2.533
1.0%	-4.667	-6.856	3.9e-1	-6.096	-5.902	-4.751	-2.524
5.0%	-4.184	-6.383	4.9e-1	-5.546	-5.434	-4.284	-2.443
10.%	-3.844	-6.047	4.5e-1	-5.294	-5.106	-3.956	-2.347
30.%	-3.457	-5.261	3.9e-1	-4.500	-4.431	-3.280	-1.974
50.%	-2.505	-4.553	4.7e-1	-3.841	-3.794	-2.643	-1.678

Table 2: Statistics (min,max,mean and variance σ) for 100 runs of $log10$ 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. ($n^l_{z_1} = n^l_{z_2} = 10$, $N_x = N_y = 5$, $n_\gamma = 40$, $m = 200$)

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

$$\begin{aligned}\min_{X,rankX=k}(||Y_l - X||_2) = \sigma_{k+1}(Y_l) &= \min_{X,rankX=k}(||Y_l - Y_0 + Y_0 - X||_2 \\ &\leq ||Y_l - Y_0||_2 + \min_{X,rankX=k} ||Y_0 - X||_2 \\ &\leq \sigma_1(E_l) + \sigma_{k+1}(Y_0) \end{aligned} \quad (6)$$

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 $\lim_{k\to\infty} y^k_l \to y^\infty$, 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 Césaro's mean of the iterated interface solutions. We transform the sequence (y_l) in an another sequence $(\tilde{y}_l)$ defined as $\tilde{y}^i_l = \frac{1}{m}\sum_{j=0}^{m-1} y^{i+j}_l$. The summation still preserves the pure linear convergence of the synchronous 0%-RAS: $\tilde{y}^{k+1}_0 - y^\infty = P(\tilde{y}^k_0 - y^\infty)$ 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 Césaro'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 $(\tilde{y}_l)$ has a smaller amplitude than for the original sequence (y_l). This leads to have a low-rank space U_l built from this $(\tilde{y}_l)$ 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 Césaro's mean with respect to l. The acceleration of the convergence is enhanced using $(\tilde{y}_l)$ than (y_l). 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 (σ_i) of the SVD of Y_l obtained with l-RAS with

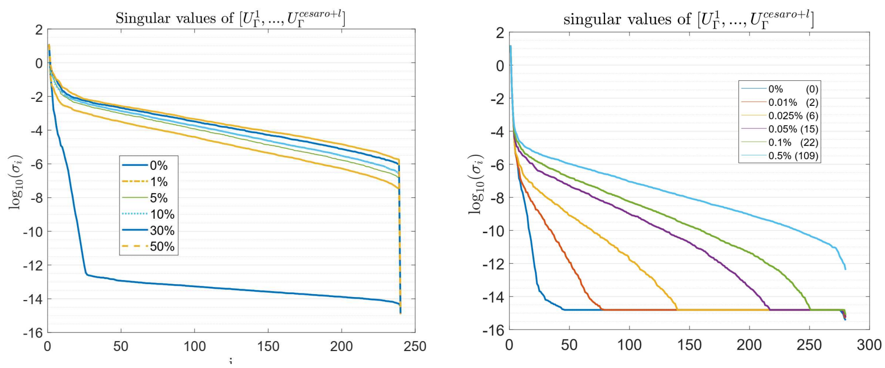

Fig. 2: Singular values of one sample of 250 l-RAS iterations for $l = \{0\%, 1\%, 5\%, 10\%, 30\%, 50\%\}$ (left) and for $l = \{0\%, 0.01\%, 0.025\%, 0.05\%, 0.1\%, 0.5\%\}$ and 300 iterations with the number of transmission condition update failures in brackets (right). ($N_x^l = N_y^l = 10$, $P_x = P_y = 5$, $p = 40$)

respect to the level l of asynchronism. It shows that the fast decreasing of (σ_i) is lost with the asynchronism. It still exhibits some decreasing of (σ_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 σ_i 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

1. L. Berenguer, T. Dufaud, and D. Tromeur-Dervout. Aitken's acceleration of the Schwarz process using singular value decomposition for heterogeneous 3D groundwater flow problems. *Computers & Fluids*, 80:320–326, 2013.
2. L. Berenguer and D. Tromeur-Dervout. Sparse Aitken-Schwarz with application to Darcy flow. https://hal.archives-ouvertes.fr/hal-03090521, 2020.
3. X.-C. Cai and M. Sarkis. A restricted additive Schwarz preconditioner for general sparse linear systems. *SIAM J. Sci. Comput.*, 21(2):792–797 (electronic), 1999.
4. K Fan. Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proc. of the National Academy of Sciences*, 37(11):760–766, 1951.
5. M. J. Gander. Schwarz methods over the course of time. *Electronic Transactions on Numerical Analysis*, pages 228–255, 2008.
6. M. Garbey and D. Tromeur-Dervout. On some Aitken-like acceleration of the Schwarz method. *Internat. J. Numer. Methods Fluids*, 40(12):1493–1513, 2002. LMS Workshop on Domain Decomposition Methods in Fluid Mechanics (London, 2001).
7. G. H. Golub and Ch. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, third edition, 1996.
8. F. Magoules, D. B. Szyld, and C. Venet. Asynchronous optimized Schwarz methods with and without overlap. *Numerische Mathematik*, 137(1):199–227, SEP 2017.
9. G. W. Stewart. On the early history of the singular value decomposition. *SIAM Rev.*, 35(4):551–566, 1993.
10. D. Tromeur-Dervout. Meshfree Adaptative Aitken-Schwarz Domain Decomposition with application to Darcy Flow. In Topping, BHV and Ivanyi, P, editor, *Parallel, Distributed and Grid Computing for Engineering*, volume 21 of *CSET Series*, pages 217–250. Saxe-Coburg Publications, 2009.
11. D. Tromeur-Dervout. Approximating the trace of iterative solutions at the interfaces with nonuniform Fourier transform and singular value decomposition for cost-effectively accelerating the convergence of Schwarz domain decomposition. *ESAIM: PROCEEDINGS*, 42:34–60, 2013.
12. P Wilders and E Brakkee. Schwarz and Schur: An algebraical note on equivalence properties. *SIAM J. Sci Comput.*, 20(6):2297–2303, JUL 22 1999.