On the derivation of optimized transmission conditions for the Stokes-Darcy coupling

Martin J. Gander[1] and Tommaso Vanzan[1]

1 Introduction

Recently a lot of attention has been devoted to the Stokes-Darcy coupling which is a system of equations used to model the flow of fluids in porous media. In [2, 1] a non standard behaviour of the optimized Schwarz method (OSM) has been observed: the optimized parameters obtained solving the classical min-max problems do not lead to an optimized convergence. The authors in [2, 1] proposed to consider a different optimization problem and they claim that the unexpected behaviour is due to the Krylov acceleration. In this manuscript, we study OSM as an iterative method and as a preconditioner for GMRES and we show that the discrepancy is not due to the Krylov acceleration but to a limitation in the derived convergence factor.

2 The Stokes-Darcy model

We consider a domain Ω divided by an interface Γ into two subdomains, Ω_1 and Ω_2. In Ω_1, a Newtonian fluid is present described by the Stokes equations whose unknowns are the velocity field $\mathbf{u}_f = (u, v)^\top$ and the pressure field p_f,

$$\begin{aligned} -\nabla \cdot \mathbb{T} &= f \quad \text{in} \quad \Omega_1, \\ \nabla \cdot \mathbf{u}_f &= 0 \quad \text{in} \quad \Omega_1, \end{aligned} \tag{1}$$

where $\mathbb{T} = 2\mu_f(\nabla^s \mathbf{u}_f) - p_f \mathbb{I}$ is the stress tensor, with $\nabla^s \mathbf{u}_f$ the symmetrized gradient, and μ_f is the fluid viscosity. The motion of the fluid in the porous media is modelled through the Darcy equations whose unknowns are the

[1] Section de mathématiques, Université de Genève, 2-4 rue du Lièvre, Genève, e-mail: martin.gander,tommaso.vanzan@unige.ch.

velocity and pressure fields in the porous media domain $\mathbf{u}_d$, p_d,

$$\mathbf{u}_d = -\mathbb{K}\nabla p_d + \mathbf{g}, \quad \nabla \cdot \mathbf{u}_d = 0 \quad \text{in} \quad \Omega_2, \tag{2}$$

where $\mathbb{K}$ is the permeability tensor and $\mathbf{g}$ is a body force vector. Equation (2) can be simplified taking the divergence of the first equation to obtain a second order elliptic PDE only for the pressure field,

$$-\nabla \cdot \mathbb{K}\nabla p_d = -\nabla \cdot \mathbf{g} \quad \text{in} \quad \Omega_2. \tag{3}$$

Both (1) and (3) are closed by Dirichlet boundary conditions on the external boundary $\partial\Omega \setminus \Gamma$, i.e. $\mathbf{u}_f = \mathbf{h}_f$, $p_d = h_d$ on $\partial\Omega \setminus \Gamma$. However the Stokes and Darcy equations still need to be coupled along the common interface Γ and there are many possible choices, see Paragraph 3 of [3]. In the following we prescribe the continuity of the normal velocities and of the normal stresses and the so called Beaver-Joseph-Saffman (BJS) condition,

$$\begin{aligned} \mathbf{u}_f \cdot \mathbf{n} &= -(\mathbb{K}\nabla p_d) \cdot \mathbf{n} + \mathbf{g} \cdot \mathbf{n}, \\ -\mathbf{n} \cdot (2\mu_f \nabla^s \mathbf{u}_f - p_f \mathbb{I}) \cdot \mathbf{n} &= p_d, \\ -\tau \cdot (2\mu_f \nabla^s \mathbf{u}_f - p_f \mathbb{I}) \cdot \mathbf{n} &= \chi_s (\mathbf{u}_f)_\tau. \end{aligned} \tag{4}$$

We remark that the BJS condition $(4)_3$ is not a coupling condition but only a closure condition for the Stokes equations. OSMs use enhanced transmission conditions on the interface, thus we take a linear combination of the coupling conditions $(4)_{1,2}$ introducing the real parameters s_1 and s_2 which are chosen to optimize the convergence. The OSM for the Stokes-Darcy system (1)-(3)-(4) then computes for iterations $n = 1, 2 \ldots$

$$\begin{aligned} -\nabla \cdot (2\mu_f \nabla^s \mathbf{u}_f^n - p_f^n \mathbb{I}) &= \mathbf{f}, \quad \text{in} \quad \Omega_1, \\ \nabla \cdot \mathbf{u}_f^n &= 0, \quad \text{in} \quad \Omega_1 \\ -\nabla \cdot \mathbb{K}\nabla p_d^n &= -\nabla \cdot \mathbf{g}, \quad \text{in} \quad \Omega_2, \\ p_d^n - s_1 \left(\mathbb{K}\nabla p_d^n \cdot \mathbf{n} - \mathbf{g} \cdot \mathbf{n}\right) &= -\mathbf{n} \cdot (2\mu_f \nabla^s \mathbf{u}_f^{n-1} - p_f^{n-1} \mathbb{I}) \cdot \mathbf{n} + s_1 \mathbf{u}_f^{n-1} \cdot \mathbf{n} \quad \text{on } \Gamma, \\ -\mathbf{n} \cdot (2\mu_f \nabla^s \mathbf{u}_f^n - p_f^n \mathbb{I}) \cdot \mathbf{n} - s_2 \mathbf{u}_f^n \cdot \mathbf{n} &= p_d^{n-1} + s_2 \left(\mathbb{K}\nabla p_d^{n-1} \cdot \mathbf{n} - \mathbf{g} \cdot \mathbf{n}\right) \quad \text{on } \Gamma, \\ -\tau \cdot (2\mu_f \nabla^s \mathbf{u}_f^n - p_f^n \mathbb{I}) \cdot \mathbf{n} &= \chi_s (\mathbf{u}_f^n)_\tau \quad \text{on } \Gamma. \end{aligned} \tag{5}$$

In [2], the authors perform a Fourier analysis of the OSM (5). Their analysis follows one of the standard approaches in the literature, i.e. the problem of interest is posed in a simplified setting where one can exploit the Fourier transform for unbounded domains or separation of variables for bounded domains. Unfortunately this last approach is not possible here since no analytical expression is available for the eigenvectors of the Stokes operator in bounded domains with Dirichlet boundary conditions. Furthermore, to simplify the calculations they assume that $\mathbb{K} = \text{diag}(\eta_1, \eta_2)$ with $\eta_j > 0, j = 1, 2$. They finally obtain that the convergence factor of algorithm (5) for all the

Fourier frequencies $k \in \mathbb{R}$ is

$$\rho(k, s_1, s_2) = \left| \frac{2\mu_f|k| - s_1}{2\mu_f|k| + s_2} \cdot \frac{1 - s_2\sqrt{\eta_1\eta_2}|k|}{1 + s_1\sqrt{\eta_1\eta_2}|k|} \right|. \tag{6}$$

The optimal choice $s_1 = 2\mu_f|k|$ and $s_2 = \frac{1}{\sqrt{\eta_1\eta_2|k|}}$ would lead to a direct method which converges in just two iterations, however this choice corresponds to non-local operators once backtransformed. Therefore a more practical choice is to set $s_1 = 2\mu_f p$ and $s_2 = \frac{1}{\sqrt{\eta_1\eta_2 p}}$ for some $p \in \mathbb{R}$. An equivalent choice of optimized parameters has been treated in [2] where the authors obtain the following result:

Theorem 1 (Proposition 3.3 in [2]). *The unique solution of the min-max problem*

$$\min_{p} \max_{k\in[k_{\min},k_{\max}]} \rho(k, p), \tag{7}$$

is given by the unique root of the non linear equation $\rho(k_{\min}, p) = \rho(k_{\max}, p)$.

A possible improvement consists in considering two free parameters, choosing $s_1 = 2\mu_f p$ and $s_2 = \frac{1}{\sqrt{\eta_1\eta_2 q}}$ with $p, q \in \mathbb{R}$. In [1], the authors propose to choose the couple p, q such that $\rho(k_{\min}, p, q) = \rho(\hat{k}, p, q) = \rho(k_{\max}, p, q)$, i.e. they impose equioscillation to obtain the optimized parameters. Even though often the solution of such min-max problems is indeed given by equioscillation, a priori there is no reason why this should be the case also for the Stokes-Darcy coupling. In fact for heterogenous problems, it has been observed that there can exist a couple of parameters which satisfies the equioscillation property, but leads to a non optimized convergence or even to a divergence method, see [6, 4, 7]. In Theorem 2 we refine Proposition 1 of [1].

Theorem 2. *The solutions of the min-max problem*

$$\min_{p,q\in\mathbb{R}} \max_{k\in[k_{\min},k_{\max}]} \rho(k,p,q) = \min_{p,q\in\mathbb{R}} \max_{k\in[k_{\min},k_{\max}]} 2\mu_f\sqrt{\eta_1\eta_2}\left| \frac{k-p}{1+2\mu_f\sqrt{\eta_1\eta_2}kp} \cdot \frac{k-q}{1+2\mu_f\sqrt{\eta_1\eta_2}kq} \right|, \tag{8}$$

are given by two pairs (p_i^*, q_i^*), $i = 1, 2$ *which satisfy the non linear equations* $|\rho(k_{\min}, p_i^*, q_i^*)| = |\rho(\widehat{k}, p_i^*, q_i^*)| = |\rho(k_{\max}, p_i^*, q_i^*)|$, $\widehat{k}$ *being an interior maximum. Moreover* $p_2^* = q_1^*$ *and* $q_2^* = p_1^*$.

Proof. The proof is based on arguments presented in [4, 8, 7] and we outline the main steps. We first observe that $\rho(k, p, q)$ is invariant under $p \leftrightarrow q$, hence we consider only $p < q$ and moreover $\rho(k, p, q) = 0$ for $k = q$ and $k = p$. The partial derivatives with respect to the parameters satisfy $\text{sign}(\partial_p\rho) = \text{sign}(p - k)$ and $\text{sign}(\partial_q\rho) = \text{sign}(q - k)$, therefore at optimality we conclude that p, q lie in $[k_{\min}, k_{\max}]$, see the proof of Theorem 1 in [8]. Solving $\partial_k\rho = 0$, we get that there exists a unique interior maximum $\widehat{k}$, with $p < \widehat{k} < q$, so that we can restrict $\max_{k\in[k_{\min},k_{\max}]} \rho(k, p, q) =$

$\max\{\rho(k_{\min},p,q),\rho(\widehat{k},p,q),\rho(k_{\max},p,q)\}$. Repeating the same arguments of Lemma 2.9 in [7], we obtain that at the optimum we must have $\rho(k_{\min},p,q)=\rho(k_{\max},p,q)$, so that we can express q as function of p and we can restrict the study to $\min_p \max\{\rho(k_{\min},p,q(p)),\rho(\widehat{k},p,q(p))\}$. Defining $\delta:=2\mu_f\sqrt{\eta_1\eta_2}$, the equioscillation constraint is equivalent to

$$l(p):=\frac{k_{\min}-p}{1+\delta k_{\min}p}\frac{1+\delta k_{\max}p}{k_{\max}-p}=\frac{k_{\max}-q(p)}{1+\delta q(p)k_{\max}}\frac{1+\delta q(p)k_{\min}}{k_{\min}-q(p)}=:g(p). \quad (9)$$

Since $\partial_p l(p)<0$ and $\partial_p g(p)>0$, $q(p)$ must be a decreasing function of p so that eq (9) is satisfied. Then using the sign of the derivatives of ρ with respect to p and q and the explicit expression of $q(p)$, we have $\frac{d\rho(k_{\min},p)}{dp}>0$ and $\frac{d\rho(\widehat{k},p)}{dp}<0$ for $k_{\min}<p<q(p)$. These observations are sufficient to conclude, see Theorem 1 in [8], that the solution of $\min_p \max\{\rho(k_{\min},p,q(p)),\rho(\widehat{k},p,q(p))\}$ is given by the unique p_1^*, such that $\rho(k_{\min},p_1^*,q(p_1^*))=\rho(\widehat{k},p_1^*,q(p_1^*))$ and q_1^* given by $q_1^*=q(p_1^*)$. Due to the invariance $p\leftrightarrow q$, we get the same results in the case $q<p$ and we conclude that the other couple satisfies $p_2^*=q_1^*$ and $q_2^*=p_1^*$.

In [2, 1], the authors studied extensively the methods obtained from Theorems 1-2 as preconditioners for GMRES. They observed that these optimized parameters do not lead to an optimized convergence and they proposed to minimize the L^1 norm instead of the maximum of the convergence factor,

$$\min_p \frac{1}{k_{\max}-k_{\min}}\int_{k_{\min}}^{k_{\max}}\rho(k,p)dk. \quad (10)$$

The reason behind this choice lies in the assumption that the Krylov method can take care of isolated slow frequencies, and therefore it would be better to have a convergence factor that is very small for a large set of frequencies with possibly high peaks. This approach was first discussed in [5] for the Helmholtz problem, with the significant difference that the OSM does not converge for the Helmholtz frequency ω, and thus the authors proposed to minimize $\min_p \max_{k\in[k_{\min},\omega^-]\cup[\omega^+,k_{\max}]}\rho(k,p)$. Since such a bad performance of the optimized parameters obtained from a min-max problem in combination with a Krylov method does not have comparison in the literature, we investigate it in details in the next Section.

3 Numerical study of the optimized Schwarz method

We consider the domains $\Omega_1=(0,1)\times(0,1)$, $\Omega_2=(0,1)\times(-1,0)$ and a uniform structured mesh with mesh size $h=0.02$, so that $k_{\min}=\pi$ and $k_{\max}=\pi/h$. We discretize the corresponding error equations of (5) with Taylor-Hood finite elements $\mathbb{P}_2^2-\mathbb{P}_1$ for the Stokes unknowns and $\mathbb{P}_2$ el-

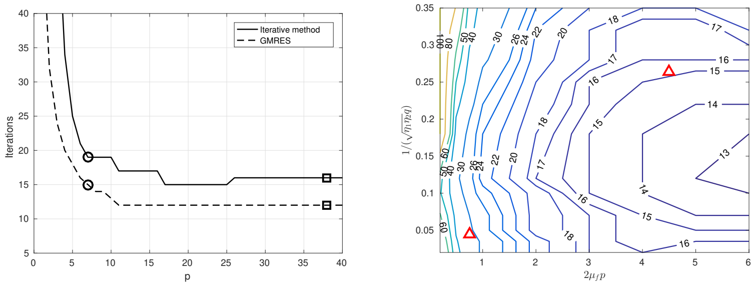

Fig. 1 Number of iterations to reach the tolerance 10^{-9} for different optimized parameters. On the left, the circle represents the solution of Theorem 1, the square corresponds to the solution of (10). On the right the triangles correspond to the double solutions of Theorem 2 and the contour plot refers to the iterative method.

ements for the Darcy pressure. The physical parameters are set equal to $\mu_f = 0.1$, $\eta_1 = \eta_2 = 1$. The stopping criterion for the iterative method is $\|u^n\|_{H^1} + \|v^n\|_{H^1} + \|p_f^n\|_{L^2} + \|p_d^n\|_{H^1} < 10^{-9}$ and similarly for GMRES the tolerance is 10^{-9}. Figure 1 shows the number of iterations to reach convergence. On the left panel we show with a circle the optimized parameter p obtained from Theorem 1 and with a square the optimized p obtained solving (10). We observe that indeed the solution of (10) leads to a faster convergence than the classical approach of Theorem 1 for the preconditioned GMRES. This is in accordance with the results proposed in [2, 1], where it has been shown that the solution of (10) leads to an equivalent or faster convergence than Theorem 1 for a wide range of parameters. However, we remark that (10) leads to a faster method than (7) also for the iterative method and not only under Krylov acceleration! On the right panel of Fig. 1 we observe that also Theorem 2 does not lead to an optimized convergence and the symmetry of the parameters has disappeared. To understand better the behaviour of the method, we initialize it setting as initial condition one by one the sine functions which correspond to the restriction of the Fourier basis $\{e^{-ikx}\}_k$ on bounded domains with Dirichlet boundary conditions. We then compute numerically an approximation of the convergence factor defining $\rho_v(k,p) = \left(\frac{\|v^3\|_{H^1}}{\|v^1\|_{H^1}}\right)$, $\rho_{p_d}(k,p) = \left(\frac{\|p_d^3\|_{H^1}}{\|p_d^1\|_{H^1}}\right)$, where v^n is the Stokes velocity in the y direction at iteration n and p_d^n is the Darcy pressure at iteration n. From the results presented in Figure 2, we observe two major issues: the first one is a very poor approximation of high frequencies. This is due to the fact that the chosen finite element spaces $\mathbb{P}_2^2 - \mathbb{P}_1 - \mathbb{P}_2$ are not capable of representing properly the exponential boundary layer of the high frequencies near the interface. We propose two remedies which can also be combined. We could first raise the order of the approximation of the finite element spaces to $\mathbb{P}_3^2 - \mathbb{P}_2 - \mathbb{P}_3$ and/or refine the mesh in the normal direction to the interface. Both remedies improve the representation of the high frequencies and

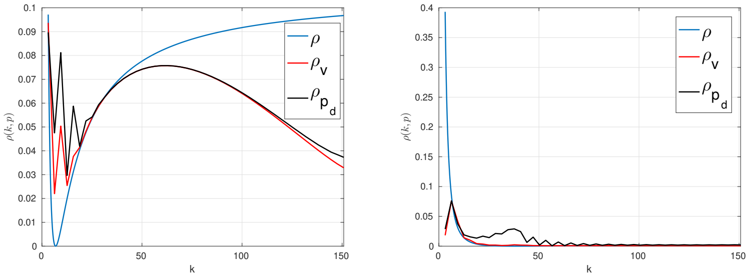

Fig. 2 Comparison of the theoretical and numerical convergence factors. On the left, optimized parameter from Theorem 1 and on the right, optimized parameter from (10).

in the following we only consider the first one. The second issue lies in a unusual oscillatory behaviour of the low, odd frequencies. This is due to the fact that the unbounded analysis used to obtain the convergence factor is not transferable to the bounded case, since the sines do not form a separated variable solution for the Stokes operator with Dirichlet boundary conditions. Hence, for instance in the right panel of Figure 2, the first frequency $\sin(\pi x)$ is transformed after one iteration into a complicated combination of higher frequencies so that actually the parameter p makes the method much faster than the theory predicts. Therefore it is not possible to diagonalize the iteration as the formula of the convergence factor (6) assumes. This phenomeon was first discussed in [8, 7] where the authors show that for the coupling of the Laplace equation with an advection-diffusion equation with tangential advection, the unbounded analysis leads to inefficient optimized parameters since the two equations lack a common eigenbasis. We consider now the Stokes-Darcy system (5) with periodic boundary conditions on the vertical edges in order to make the bounded problem as similar as possible to the unbounded case. In this setting there exists a separated variable solution for the Stokes problem involving the Fourier basis $\{e^{-ikx}\}_k$, see [9]. In Figure 3 we show both the numerical and theoretical convergence factors computed for even frequencies $\{\sin(2k\pi x)\}_k$. The same results are obtained using the other periodic frequencies $\{\cos(2k\pi x)\}_k$. Comparing with Figure 2, we observe that now we have an excellent agreement between the numerical and theoretical convergence factors and thus we would expect that the optimized parameters from the min-max theorems provide optimized convergence. We thus start the OSM method (5) with initial guesses given by a linear combination of periodic sine and cosine functions multiplied by random coefficients. Figure 4 shows that both Theorem 1 and 2 now lead to optimized convergence for the iterative method (5) and we also observe the symmetry of the optimized parameters in the right panel as Theorem 2 predicts. However concerning GMRES, we note that the optimized parameter from Theorem 1 is still a bit too small. This can be understood studying the eigenvalues of the preconditioned matrix system which are shown in Fig 5. Analyzing the large real

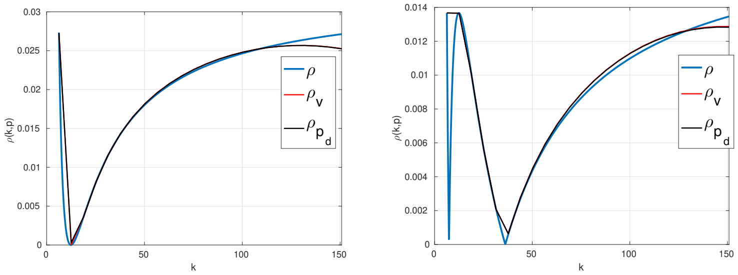

Fig. 3 Comparison of the theoretical and numerical convergence factors. On the left for the single sided optimized parameter from Theorem 1 and on the right one for the double sided parameters of Theorem 2. The minimum frequency is now $k_{\min} = 2\pi$.

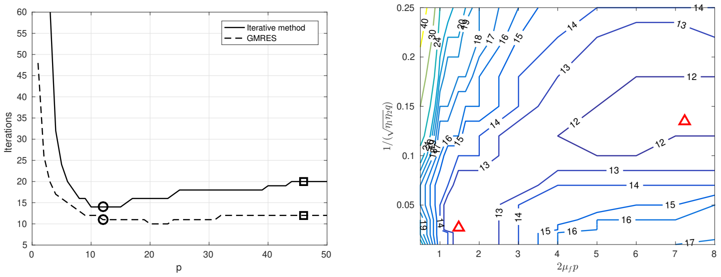

Fig. 4 Number of iterations to reach the tolerance 10^{-9} for different optimized parameters. On the left, the circle represents the solution of Theorem 1, the square corresponds to the approach of (10). On the right the triangles correspond to the double solutions of Theorem 2 and the contour plot refers to the iterative method.

eigenvalue, we have observed that the corresponding eigenvector is given by a zero velocity field $\mathbf{u}_f$, a constant pressure p_f and a linear Darcy pressure p_d. This constant mode is actually not treated by the unbounded Fourier analysis and it is not present in our initial guess for the iterative method. Defining the functions $p_d^n = D^n(y + L)$ and $p_f = P^n$ with $P, D \in \mathbb{R}$ and L is the vertical length of the subdomains, and inserting them into the OSM algorithm (5), we obtain a convergence factor $\rho(k = 0, p) := \frac{1-s_2}{1+s_1}$. Solving numerically the min-max problem $\min_p \max_{k\in\{0\}\cup[k_{\min},k_{\max}]} \rho(k, p)$ we obtain the equioscillation between $\rho(0, p)$ and $\rho(k_{\min}, p)$ and a numerical value of $p \approx 48$. In the right panel of Fig. 5 we start the method with a totally random initial guess and this shows that taking into account the constant mode actually makes our analysis exact.

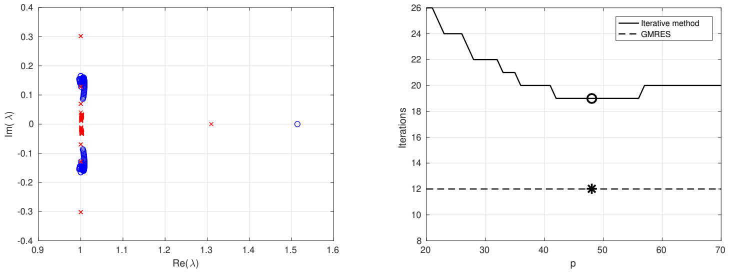

Fig. 5 On the left panel, the blue circles correspond to first 100 eigenvalues of the preconditioned volume matrix in the case with the optimized parameter of Theorem 1 and the red crosses in the case using the solution of (10). On the right panel we show the number of iterations to reach convergence with periodic boundary conditions and with a random initial guess. The circle corresponds to the solution of Theorem 1 and the star to the value of p such that we have the minimal residual of GMRES.

4 Conclusions

In this manuscript we showed that the bad performance of the optimized parameters of the min-max problems for the Stokes-Darcy coupling is not due to the Krylov acceleration but to the difficulty of transferring the unbounded Fourier analysis to the bounded case. For Dirichlet boundary conditions, the problem lies in the odd frequencies which mix among them during the iterations and therefore the convergence factor (6) loses its accuracy. For periodic boundary conditions, we recover a perfect agreement between the unbounded analysis and the numerical simulations for periodic frequencies, however the Fourier analysis does not deal with the constant mode which is present in the bounded case. Including the constant mode in the analysis we recover the optimality of the min-max optimized parameters for periodic boundary conditions.

References

1. Discacciati, M., Gerardo-Giorda, L.: Is minimising the convergence rate a good choice for efficient optimized Schwarz preconditioning in heterogeneous coupling? the Stokes-Darcy case. In: Domain Decomposition Methods in Science and Engineering XXIV, pp. 233–241. Springer International Publishing, Cham (2018)
2. Discacciati, M., Gerardo-Giorda, L.: Optimized Schwarz methods for the Stokes-Darcy coupling. IMA Journal of Numerical Analysis (2018)
3. Discacciati, M., Quarteroni, A.: Navier-Stokes/Darcy coupling: modeling, analysis, and numerical approximation. Revista Matematica Complutense (2009)
4. Gander, M.J., Dubois, O.: Optimized Schwarz methods for a diffusion problem with discontinuous coefficient. Numerical Algorithms **69**(1), 109–144 (2015)

5. Gander, M.J., Magoules, F., Nataf, F.: Optimized Schwarz methods without overlap for the Helmholtz equation. SIAM Journal on Scientific Computing **24**(1), 38–60 (2002)
6. Gander, M.J., Vanzan, T.: Heterogeneous optimized Schwarz methods for coupling Helmholtz and Laplace equations. In: Domain Decomposition Methods in Science and Engineering XXIV, pp. 311–320. Springer International Publishing, Cham (2018)
7. Gander, M.J., Vanzan, T.: Heterogeneous optimized Schwarz methods for second order elliptic PDEs. to appear in SIAM Journal of Scientific Computing (2019)
8. Gander, M.J., Vanzan, T.: Optimized Schwarz methods for advection diffusion equations in bounded domains. In: Numerical Mathematics and Advanced Applications ENUMATH 2017, pp. 921–929. Springer International Publishing, Cham (2019)
9. Rummler, B.: The eigenfunctions of the Stokes operator in special domains. ii. ZAMM - Journal of Applied Mathematics and Mechanics **77**(9), 669–675 (1997)