

# ON SOLVING NON-SYMMETRIC SADDLE-POINT SYSTEMS ARISING FROM FICTITIOUS DOMAIN APPROACHES\*

Radek Kučera, Tomáš Kozubek, Jaroslav Haslinger

## 1. Introduction

We propose a fast method for finding a pair  $(u, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$  that solves a linear system of algebraic equations called the *(generalized) saddle-point system*:

$$\begin{pmatrix} A & B_1^\top \\ B_2 & 0 \end{pmatrix} \begin{pmatrix} u \\ \lambda \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}, \quad (1)$$

where the diagonal block  $A$  is an  $(n \times n)$  matrix, the off-diagonal blocks  $B_1$  and  $B_2$  are  $(m \times n)$  matrices with full row-rank and vectors  $f, g$  are of order  $n, m$ , respectively.

Our contribution is inspired by a class of saddle-point systems arising from fictitious domain formulations of PDEs [3, 4]. Therefore we will be interested especially in systems (1) with  $n$  large,  $A$  singular and  $B_1, B_2$  sparse. Moreover, we will assume that  $m$  is much smaller than  $n$  and that the defect  $l$  of  $A$ , i.e.  $l = n - \text{rank } A$ , is much smaller than  $m$ .

There are several basic approaches used for solving (1); see e.g. [1]. Due to the structure of our matrices, we pay our attention to the class of methods that are based on the Schur complement reduction. Their key idea consists in eliminating the first unknown  $u$ . This leads, in the case of non-singular  $A$ , to the reduced system for the second unknown  $\lambda$ . The matrix of this system is the (negative) Schur complement  $-S = B_2 A^{-1} B_1^\top$ . If this system is solved by an iterative method, we do not need to form  $S$  explicitly since only the matrix-vector products with  $S$  are needed.

The situation is not so easy if  $A$  is singular. In this case, the first unknown  $u$  can not be completely eliminated from (1). The Schur complement reduction leads now to another saddle-point system for  $\lambda$  and a new unknown, say  $\alpha$ , that corresponds to the null-space of  $A$ . Fortunately after applying orthogonal projectors, we obtain an equation only in terms of  $\lambda$ . As our original saddle-point system (1) is non-symmetric, this equation can be solved by a *projected* Krylov method for non-symmetric matrices. In our numerical tests, we will use the projected variant of the BiCGSTAB algorithm.

The presented method generalizes ideas used in the algebraic description of FETI domain decomposition methods [2], in which  $A$  is symmetric, positive semidefinite and  $B_1 = B_2$ .

---

\*Supported by the National Program of Research "Information Society" under project 1ET400300415 and by the grant IAA1075402 of the Grant Agency of the Czech Academy of Sciences.

## 2. A new variant of the fictitious domain method

Let  $\Omega$  be a bounded domain in  $\mathbb{R}^d$ ,  $d = 2, 3$  with the Lipschitz boundary  $\partial\Omega$ , which is split into three non-overlapping parts  $\Gamma_D$ ,  $\Gamma_N$  and  $\Gamma_G$  (see Figure 2.1). We will be concerned with the following abstract class of mixed boundary value problems:

$$\left. \begin{aligned} \mathcal{L}u &= f && \text{in } \Omega, \\ u &= g_D && \text{on } \Gamma_D, \\ \frac{\partial u}{\partial \nu_{\mathcal{L}}} &= g_N && \text{on } \Gamma_N, \\ \frac{\partial u}{\partial \nu_{\mathcal{L}}} + \beta u &= g_G && \text{on } \Gamma_G, \end{aligned} \right\} \quad (\mathcal{P})$$

where  $\mathcal{L}$  is an elliptic operator of the second order,  $f \in L^2(\Omega)$ ,  $g_D \in H^{1/2}(\Gamma_D)$ ,  $g_N \in L^2(\Gamma_N)$ ,  $g_G \in L^2(\Gamma_G)$ ,  $\beta$  is a constant and  $\frac{\partial}{\partial \nu_{\mathcal{L}}}$  denotes the normal derivative on  $\partial\Omega$ . We assume that  $(\mathcal{P})$  has a unique solution  $u$ .

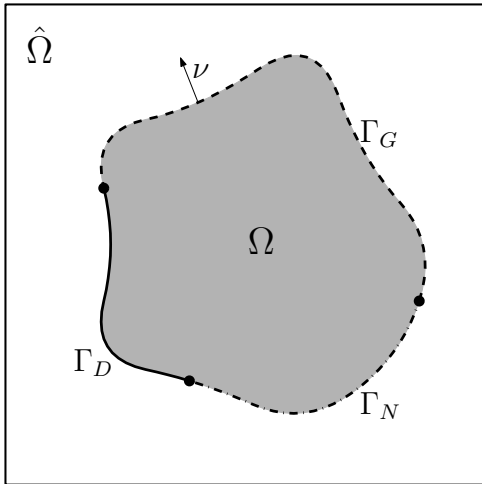


Figure 2.1: Geometry.

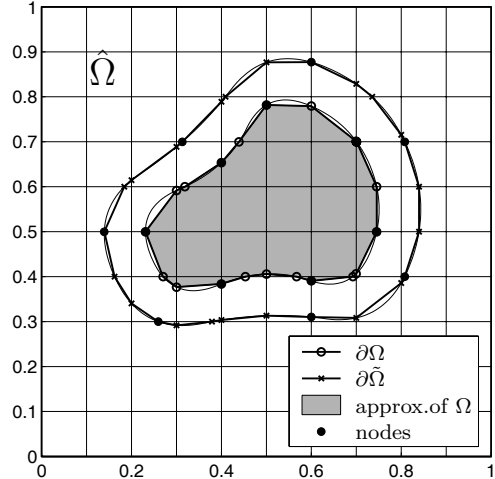


Figure 2.2: Auxiliary boundary  $\partial\tilde{\Omega}$ .

Any fictitious domain (FD) formulation of PDEs transforms the original problem defined in a domain  $\Omega$  to a new one solved in a simple shaped domain  $\hat{\Omega}$  (e.g. a box), which contains  $\bar{\Omega}$ . Its solution will be denoted by  $\hat{u}$ . The standard boundary Lagrange multiplier FD approach (see [3]) gives rise to a singularity of  $\hat{u}$  located on the boundary  $\partial\Omega$ . This fact can result in an intrinsic error of the computed solution. Therefore we recommend to move this singularity further of  $\partial\Omega$ , i.e. to enforce the prescribed boundary conditions by new control variables defined not on  $\partial\Omega$  but on an auxiliary boundary  $\partial\tilde{\Omega} = \tilde{\Gamma}_D \cup \tilde{\Gamma}_N \cup \tilde{\Gamma}_G$  obtained by shifting the Bezièr approximation of  $\partial\Omega = \Gamma_D \cup \Gamma_N \cup \Gamma_G$  in the outer normal direction with a step  $\delta$  (see Fig. 2.2). *This approach improves significantly the error of the computed FD solution and the rates of convergence.*

Let us introduce boundary control variables  $\tilde{\lambda}_D \in \tilde{\Lambda}_D := H^{-1/2}(\tilde{\Gamma}_D)$ ,  $\tilde{\lambda}_N \in \tilde{\Lambda}_N := H^{-1/2}(\tilde{\Gamma}_N)$  and  $\tilde{\lambda}_G \in \tilde{\Lambda}_G := H^{-1/2}(\tilde{\Gamma}_G)$  defined on  $\tilde{\Gamma}_D$ ,  $\tilde{\Gamma}_N$  and  $\tilde{\Gamma}_G$ , respectively. Instead of  $(\mathcal{P})$ , we will solve the following problem:

$$\left. \begin{aligned} &\text{Find } (\hat{u}, \tilde{\lambda}_D, \tilde{\lambda}_N, \tilde{\lambda}_G) \in V \times \tilde{\Lambda}_D \times \tilde{\Lambda}_N \times \tilde{\Lambda}_G \text{ such that} \\ &a(\hat{u}, \hat{v}) + \tilde{b}_D(\tilde{\lambda}_D, \tilde{\tau}_D \hat{v}) + \tilde{b}_N(\tilde{\lambda}_N, \tilde{\tau}_N \hat{v}) + \tilde{b}_G(\tilde{\lambda}_G, \tilde{\tau}_G \hat{v}) = (\hat{f}, \hat{v})_{0, \hat{\Omega}} \quad \forall \hat{v} \in V, \\ &b_D(\mu_D, \tau_D \hat{u}) = b_D(\mu_D, g_D) \quad \forall \mu_D \in \Lambda_D, \\ &b_N(\mu_N, \frac{\partial \hat{u}}{\partial \nu_C}) = b_N(\mu_N, g_N) \quad \forall \mu_N \in \Lambda_N, \\ &b_G(\mu_G, \frac{\partial \hat{u}}{\partial \nu_C} + \beta \tau_G \hat{u}) = b_G(\mu_G, g_G) \quad \forall \mu_G \in \Lambda_G, \end{aligned} \right\} \quad (\hat{\mathcal{P}})$$

where  $a : V \times V \rightarrow \mathbb{R}^1$  is a continuous, coercive bilinear form resulting from the weak formulation of the first equation in  $(\mathcal{P})$ ,  $\hat{f}$  is an extension of  $f$  from  $\Omega$  to  $\hat{\Omega}$ ,  $\tau_D : V \mapsto H^{1/2}(\Gamma_D)$ ,  $\tau_G : V \mapsto H^{1/2}(\Gamma_G)$ ,  $\tilde{\tau}_D : V \mapsto H^{1/2}(\tilde{\Gamma}_D)$ ,  $\tilde{\tau}_N : V \mapsto H^{1/2}(\tilde{\Gamma}_N)$  and  $\tilde{\tau}_G : V \mapsto H^{1/2}(\tilde{\Gamma}_G)$  stand for the trace mappings, respectively, and the bilinear forms  $b_D$ ,  $b_N$ ,  $b_G$  and  $\tilde{b}_D$ ,  $\tilde{b}_N$ ,  $\tilde{b}_G$  denote the corresponding duality pairings. Finally,  $\Lambda_D := H^{-1/2}(\Gamma_D)$ ,  $\Lambda_N := H^{-1/2}(\Gamma_N)$ ,  $\Lambda_G := H^{-1/2}(\Gamma_G)$  and  $V$  is a closed subspace of  $H^1(\hat{\Omega})$ . Typical choices for  $V$  are:  $H^1(\hat{\Omega})$ ,  $H_0^1(\hat{\Omega})$ , or  $H_P^1(\hat{\Omega}) = \{v | v \in H^1(\hat{\Omega}), v \text{ is periodic on } \partial \hat{\Omega}\}$  if  $\hat{\Omega}$  is a cartesian product of intervals.

A discretization of  $(\hat{\mathcal{P}})$  based on a mixed finite element method leads to a saddle-point system (1). One can use fairly structured meshes in  $\hat{\Omega}$  ensuring favorable properties of the stiffness matrix  $A$ . Therefore actions of a generalized inverse  $A^\dagger$  (or inverse  $A^{-1}$ ) are cheaply computable and, in addition, the null-space of  $A$  and  $A^\top$  can be easily identified [6]. The geometry of  $\partial \Omega$  together with the type of boundary conditions are characterized by  $B_1$ ,  $B_2$ , which are highly sparse.

### 3. Algorithms

Denote  $\mathbb{N}(B|\mathbb{V})$  the null-space and  $\mathbb{R}(B|\mathbb{V})$  the range-space of an  $(m \times n)$  matrix  $B$  in a subspace  $\mathbb{V} \subset \mathbb{R}^n$ . If  $\mathbb{V} = \mathbb{R}^n$ , we simply write  $\mathbb{N}(B) := \mathbb{N}(B|\mathbb{R}^n)$  and  $\mathbb{R}(B) := \mathbb{R}(B|\mathbb{R}^n)$ . The system (1) has a unique solution iff [5]

$$\mathbb{N}(A) \cap \mathbb{N}(B_2) = \{0\}, \quad (2)$$

$$\mathbb{R}(A|\mathbb{N}(B_2)) \cap \mathbb{R}(B_1^\top) = \{0\}. \quad (3)$$

Suppose that  $A$  is singular with the defect  $l = \dim \mathbb{N}(A)$ ,  $l \geq 1$  and consider  $(n \times l)$  matrices  $N$  and  $M$  whose columns span the null-space  $\mathbb{N}(A)$  and  $\mathbb{N}(A^\top)$ , respectively. Finally, denote by  $A^\dagger$  a generalized inverse to  $A$ . In what follows we will consider an arbitrary but fixed selections of  $A^\dagger$ ,  $N$  and  $M$ .

The *generalized Schur complement* of  $A$  in (1) is defined by

$$\mathcal{S} = \begin{pmatrix} -B_2 A^\dagger B_1^\top & B_2 N \\ M^\top B_1^\top & 0 \end{pmatrix}.$$

Notice that  $\mathcal{S}$  is invertible provided that (2), (3) are satisfied. The following theorem describes the Schur complement reduction.

**Theorem 3.1** [5] *Assume that both  $B_1, B_2$  have full row-ranks and that (2), (3) are satisfied. Then the second component  $\lambda$  of a solution to (1) is the first component of a solution to*

$$\begin{pmatrix} F & G_1^\top \\ G_2 & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \alpha \end{pmatrix} = \begin{pmatrix} d \\ e \end{pmatrix}, \quad (4)$$

where  $F := B_2 A^\dagger B_1^\top, G_1 := -N^\top B_2^\top, G_2 := -M^\top B_1^\top, d := B_2 A^\dagger f - g$  and  $e := -M^\top f$ . The first component  $u$  of a solution to (1) is given by the formulae

$$u = A^\dagger(f - B_1^\top \lambda) + N\alpha.$$

Let us point out that (4) is formally the same saddle-point system as (1), but its size is considerably smaller. We will modify the new system (4) by two orthogonal projectors

$$P_1 := I - G_1^\top (G_1 G_1^\top)^{-1} G_1, \quad P_2 := I - G_2^\top (G_2 G_2^\top)^{-1} G_2,$$

on  $\mathbb{N}(G_1), \mathbb{N}(G_2)$ , respectively. Our method is based on the following results.

**Lemma 3.1** [5] *The linear operator  $P_1 F : \mathbb{N}(G_2) \mapsto \mathbb{N}(G_1)$  is invertible.*

**Theorem 3.2** [5] *Let  $\lambda_{\mathbb{N}} \in \mathbb{N}(G_2), \lambda_{\mathbb{R}} \in \mathbb{R}(G_2^\top)$ . Then  $\lambda = \lambda_{\mathbb{N}} + \lambda_{\mathbb{R}}$  is the first component of a solution to (4) iff*

$$\lambda_{\mathbb{R}} = G_2^\top (G_2 G_2^\top)^{-1} e$$

and

$$P_1 F \lambda_{\mathbb{N}} = P_1 (d - F \lambda_{\mathbb{R}}).$$

The second component  $\alpha$  is given by

$$\alpha = (G_1 G_1^\top)^{-1} G_1 (d - F \lambda).$$

Let us summarize the previous results in the algorithm scheme. It turns out to be reasonable to form and store the  $(l \times m)$  matrices  $G_1, G_2$  and the  $(l \times l)$  matrices  $H_1 := (G_1 G_1^\top)^{-1}, H_2 := (G_2 G_2^\top)^{-1}$  because  $l$  is small. On the other hand, the  $(m \times m)$  matrices  $F, P_1$  and  $P_2$  are not assembled explicitly.

#### ALGORITHM: PROJECTED SCHUR COMPLEMENT METHOD (PSCM)

Step 1.a: Assemble  $G_1 = -N^\top B_2^\top, G_2 = -M^\top B_1^\top, d = B_2 A^\dagger f - g$  and  $e = -M^\top f$ .

Step 1.b: Assemble  $H_1 = (G_1 G_1^\top)^{-1}$  and  $H_2 = (G_2 G_2^\top)^{-1}$ .

Step 1.c: Assemble  $\lambda_{\mathbb{R}} = G_2^\top H_2 e$ .

Step 1.d: Assemble  $\tilde{d} = P_1 (d - F \lambda_{\mathbb{R}})$ .

Step 1.e: Solve the equation  $P_1 F \lambda_{\mathbb{N}} = \tilde{d}$  on  $\mathbb{N}(G_2)$ .

Step 1.f: Compute  $\lambda = \lambda_{\mathbb{N}} + \lambda_{\mathbb{R}}$ .

Step 2: Compute  $\alpha = H_1 G_1 (d - F \lambda)$ .

Step 3: Compute  $u = A^\dagger (f - B_1^\top \lambda) + N\alpha$ .

The heart of the algorithm consists in Step 1.e. Its solution can be computed by a *projected* Krylov subspace method. The projected BiCGSTAB algorithm [5] can be derived from the non-projected one [7] by choosing an initial iterate  $\lambda_N^0$  in  $\mathbb{N}(G_2)$ , projecting the initial residual in  $\mathbb{N}(G_2)$  and replacing the operator  $P_1F$  by its projected version  $P_2P_1F$ . Finally, let us point out that convergence of the projected BiCGSTAB algorithm can be accelerated by a reorthogonalization procedure or by a multigrid technique.

#### 4. Numerical experiments

We illustrate the efficiency of the presented method on a model problem ( $\mathcal{P}$ ). Let  $\mathcal{L} = -\Delta$ ,  $\Omega = \{(x, y) \in \mathbb{R}^2 \mid (x - 0.5)^2/0.4^2 + (y - 0.5)^2/0.2^2 < 1\}$  and consider the mixed Dirichlet-Neumann boundary conditions with  $\Gamma_D$  and  $\Gamma_N$  corresponding to the upper and lower half-part of the ellipse  $\partial\Omega$ , respectively. Let us choose the right hand-sides  $f$ ,  $g_D$  and  $g_N$  in ( $\mathcal{P}$ ) appropriately to the exact solution  $u_{ex}(x, y) = 100((x - 0.5)^3 - (y - 0.5)^3)$ . In the FD formulation ( $\hat{\mathcal{P}}$ ), we take  $\hat{\Omega} \equiv (0, 1) \times (0, 1)$  and  $V = H_P^1(\hat{\Omega})$ . This space is approximated by piecewise bilinear functions defined on a rectangulation of  $\hat{\Omega}$  with a stepsize  $h$ . The spaces  $\Lambda_D$ ,  $\Lambda_N$  and their tilded counterparts are approximated by piecewise constant functions defined on partitions of polygonal approximations of  $\partial\Omega$  and  $\partial\tilde{\Omega}$ .

In tables below, we report the errors of the approximate solution  $u_h$  with respect to the stepsize  $h$  in the indicated norms together with the number of primal ( $n$ ) and control ( $m$ ) variables, the number of BiCGSTAB iterations and the computational time.

Tables 4.1 and 4.2 summarize results obtained by a classical FD method with boundary Lagrange multipliers on  $\partial\Omega$ . The BiCGSTAB iterations are accelerated by biorthogonalization, when  $B_2$  in (1) is replaced by  $(B_2B_1)^{-1}B_2$ .

Table 4.1: Convergence **without**  $\partial\tilde{\Omega}$ .

Step $h$	$n/m$	Iters.	C.time[s]	$\delta_{L^2(\Omega)}$	$\delta_{H^1(\Omega)}$	$\delta_{L^2(\partial\Omega)}$
1/128	16641/40	15	0.188	2.3637e-002	2.1633e+000	9.0989e-002
1/256	66049/70	24	1.36	1.2831e-002	1.4736e+000	4.9341e-002
1/512	263169/124	32	14.24	7.1820e-003	9.9318e-001	2.7571e-002
1/1024	1050625/220	46	93.11	3.9157e-003	7.1732e-001	1.5345e-002

Table 4.2: Convergence **without**  $\partial\tilde{\Omega}$ , biorthogonalization.

Step $h$	$n/m$	Iters.	C.time[s]	$\delta_{L^2(\Omega)}$	$\delta_{H^1(\Omega)}$	$\delta_{L^2(\partial\Omega)}$
1/128	16641/40	9	0.11	2.3386e-002	2.1550e+000	8.9462e-002
1/256	66049/70	12	0.735	1.2808e-002	1.4734e+000	4.9238e-002
1/512	263169/124	22	10.03	7.1183e-003	9.9261e-001	2.7336e-002
1/1024	1050625/220	30	60.23	3.8315e-003	7.1694e-001	1.5064e-002

From Tables 4.3 and 4.4 one can see that the errors are significantly smaller, when the auxiliary boundary  $\partial\tilde{\Omega}$  (with  $\delta = 8h$ ) is used. Here the BiCGSTAB iterations are accelerated by a multigrid strategy.

Table 4.3: Convergence **with**  $\partial\tilde{\Omega}$ .

Step $h$	$n/m$	Iters.	C.time[s]	$\delta_{L^2(\Omega)}$	$\delta_{H^1(\Omega)}$	$\delta_{L^2(\partial\Omega)}$
1/128	16641/40	25	0.281	5.3431e-004	2.4639e-002	1.8577e-003
1/256	66049/70	39	2.218	1.4133e-004	1.2407e-002	5.7929e-004
1/512	263169/124	99	42.22	4.3848e-005	7.0675e-003	2.2314e-004
1/1024	1050625/220	200	371.5	1.2541e-005	3.6767e-003	6.9726e-005

Table 4.4: Convergence **with**  $\partial\tilde{\Omega}$ , multigrid.

Step $h$	$n/m$	Iters.	C.time[s]	$\delta_{L^2(\Omega)}$	$\delta_{H^1(\Omega)}$	$\delta_{L^2(\partial\Omega)}$
1/128	16641/40	16	0.266	7.3218e-004	2.8843e-002	2.3947e-003
1/256	66049/68	20	1.39	1.3533e-004	1.1927e-002	5.0063e-004
1/512	263169/124	33	16.37	3.3349e-005	5.9480e-003	1.4539e-004
1/1024	1050625/220	38	94.25	1.3469e-005	3.7054e-003	5.2209e-005

## References

- [1] M. Benzi, G.H. Golub and J. Liesen: Numerical solution of saddle point systems. Acta Numerica (2005), 1–137.
- [2] C. Farhat, J. Mandel and F.X. Roux: Optimal convergence properties of the FETI domain decomposition method. Comput. Methods Appl. Mech. Engrg. **115** (1994), 365–385.
- [3] R. Glowinski, T. Pan and J. Periaux: A fictitious domain method for Dirichlet problem and applications. Comput. Meth. Appl. Mech. Engrg. **111** (1994), 283–303.
- [4] J. Haslinger, T. Kozubek, R. Kučera, K. Kunisch and G. Peichl: Fictitious domain approach for solving boundary value problems with the mixed Dirichlet-Neumann boundary conditions. In preparation (2006).
- [5] R. Kučera, T. Kozubek and J. Haslinger: Projected Schur complement method for solving nonsymmetric saddle point systems arising in fictitious domain approach. Submitted to Appl. Math. (2006).
- [6] R. Kučera: Complexity of an algorithm for solving saddle-point systems with singular blocks arising in wavelet-Galerkin discretizations. Appl. Math. **50** (2005), 3, 291–308.
- [7] H. A. Van der Vorst: BiCGSTAB: a fast and smoothly converging variant of BiCG for solution of nonsymmetric linear systems. SIAM J. Sci. Statist. Comput. **13** (1992), 631–644.