On a Fictitious Domain Method for Unilateral Problems

Jaroslav Haslinger, Tomáš Kozubek and Radek Kučera

1 Introduction

This contribution deals with numerical realization of elliptic boundary value problems with unilateral boundary conditions using a fictitious domain method. Any fictitious domain formulation [2] extends the original problem defined in a domain ω to a new (fictitious) domain Ω with a simple geometry (e.g. a box) which contains $\overline{\omega}$. The main advantage consists in possibility to use a uniform mesh in Ω leading to a structured stiffness matrix. This enables us to apply highly efficient multiplying procedures [6].

Fictitious domain formulations of problems with the classical Dirichlet or Neumann boundary conditions lead after a finite element discretization typically to algebraic saddle-point systems. For their solution one can use the algorithm studied in [4] that combines the Schur complement reduction with the null-space method. The situation is not so easy for unilateral problems since their weak formulation contains a non-differentiable projection operator. Fortunately, a resulting algebraic representation is described by a system that is semi-smooth in the sense of [1] so that a generalized Newton method can be applied. This method has been already used in [5] for solving complementarity problems. In our case each Newton step relates to a mixed Dirichlet-Neumann problem and therefore the algorithm from [4] can be used for solving inner linear systems. Due to the *superlinear* convergence rate of the Newton iteration [1], the computations are only slightly more expensive than the solution of pure Dirichlet or Neumann problems.

In this paper we compare two variants of the fictitious domain method. The first one enforces unilateral conditions by Langrange multipliers defined on the boundary γ of the original domain ω. Therefore the fictitious domain solution has a singularity

Jaroslav Haslinger
Charles University, Prague, Czech Republic, e-mail: hasling@karlin.mff.cuni.cz

Tomáš Kozubek and Radek Kučera
VŠB-TU Ostrava, Czech Republic e-mail: tomas.kozubek@vsb.cz, radek.kucera@vsb.cz

on γ that can result in an intrinsic error of the computed solution. The second one uses an auxiliary boundary Γ located outside of $\overline{\omega}$ on which we introduce a new control variable in order to satisfy the conditions on γ. In the second approach the singularity is moved away from $\overline{\omega}$ so that the computed solution is smoother in ω. We shall experimentally show that the discretization error is significantly smaller in this case. For more details we refere to [3, 4].

2 Setting of the problem

We shall consider the following unilateral problem in a bounded domain $\omega \subset \mathbb{R}^2$ with the Lipschitz boundary γ:

$$\left.\begin{aligned} &-\Delta u + u = f \quad \text{in } \omega, \\ &u \geq g,\ \frac{\partial u}{\partial n_\gamma} \geq 0,\ \frac{\partial u}{\partial n_\gamma}(u-g) = 0 \quad \text{on } \gamma, \end{aligned}\right\} \tag{1}$$

where $f \in L^2_{loc}(\mathbb{R}^2)$, $g \in H^{1/2}(\gamma)$ are given functions and $\frac{\partial}{\partial n_\gamma}$ stands for the normal derivative of a function on γ. We denote by $(\cdot,\cdot)_{k,S}$ the scalar product in $H^k(S)$, $k \geq 0$ integer ($H^0(S) := L^2(S)$).

The weak form of (1) reads as follows:

$$\left.\begin{aligned} &\textit{Find } u \in H^1(\omega) \textit{ such that} \\ &(u,v)_{1,\omega} = (f,v)_{0,\omega} + \langle \frac{\partial u}{\partial n_\gamma}, v\rangle_\gamma \quad \forall v \in H^1(\omega), \\ &\frac{\partial u}{\partial n_\gamma} \in H^{-1/2}_+(\gamma), \\ &\langle \mu - \frac{\partial u}{\partial n_\gamma}, u-g\rangle_\gamma \geq 0 \quad \forall \mu \in H^{-1/2}_+(\gamma), \end{aligned}\right\} \tag{2}$$

where $\langle\cdot,\cdot\rangle_\gamma$ denotes the duality pairing between $H^{-1/2}(\gamma)$ and $H^{1/2}(\gamma)$. It is well-known that this problem has a unique solution.

Next, we shall suppose that $\frac{\partial u}{\partial n_\gamma} \in L^2_+(\gamma)$. Thus the duality pairing in (2) is represented by the $L^2(\gamma)$-scalar product and the inequality in (2) is equivalent to

$$\frac{\partial u}{\partial n_\gamma} = P(\frac{\partial u}{\partial n_\gamma} - \rho(u-g)), \tag{3}$$

where P denotes the projection of $L^2(\gamma)$ onto $L^2_+(\gamma)$ and $\rho > 0$ is arbitrary but fixed.

We shall present two variants of a fictitious domain formulation. To this end we choose a bounded domain Ω having a simple shape such that $\overline{\omega} \subset \Omega$ and construct a closed curve $\Gamma \subset \Omega$ surrounding ω. We shall distinguish two cases concerning the mutual positions of γ and Γ:

(i) $\gamma \equiv \Gamma$; (non-smooth variant)
(ii) $\mathrm{dist}(\gamma, \Gamma) > 0$; (smooth variant).

Instead of (2), we propose to solve the extended problem in Ω called the *fictitious domain formulation* of (1):

$$\left.\begin{array}{l} \textit{Find } (\hat{u}, \lambda) \in H_0^1(\Omega) \times H^{-1/2}(\Gamma) \textit{ such that} \\ (\hat{u}, v)_{1,\Omega} = (f, v)_{0,\Omega} + \langle \lambda, v \rangle_\Gamma \quad \forall v \in H_0^1(\Omega), \\ \dfrac{\partial \hat{u}_{|\omega}}{\partial n_\gamma} \in L^2(\gamma), \\ \dfrac{\partial \hat{u}_{|\omega}}{\partial n_\gamma} = P\left(\dfrac{\partial \hat{u}_{|\omega}}{\partial n_\gamma} - \rho(\hat{u}_{|\omega} - g)\right), \end{array}\right\} \tag{4}$$

where $\langle \cdot, \cdot \rangle_\Gamma$ stands for the duality pairing between $H^{-1/2}(\Gamma)$ and $H^{1/2}(\Gamma)$. It is readily seen that $\hat{u}_{|\omega}$ solves (2), where $\hat{u}$ is the first component of the solution to (4). An existence analysis for this problem is discussed in [3].

3 Discretization

Let us consider finite dimensional subspaces $V_h \subset H_0^1(\Omega)$, $\Lambda_H(\gamma) \subset L^2(\gamma)$, $\Lambda_H(\Gamma) \subset L^2(\Gamma)$ such that $\dim V_h = n$, $\dim \Lambda_H(\gamma) = \dim \Lambda_H(\Gamma) = m$. By a discretization of (4) we mean the following problem:

$$\left.\begin{array}{l} \textit{Find } (\hat{u}_h, \lambda_H) \in V_h \times \Lambda_H(\Gamma) \textit{ such that} \\ (\hat{u}_h, v_h)_{1,\Omega} = (f, v_h)_{0,\Omega} + (\lambda_H, v_h)_{0,\Gamma} \quad \forall v_h \in V_h, \\ \delta_H \hat{u}_h = P(\delta_H \hat{u}_h - \rho(\tau_H \hat{u}_h - g_H)), \end{array}\right\} \tag{5}$$

where $\delta_H \hat{u}_h$, $\tau_H \hat{u}_h$ and g_H are appropriate approximations of $\frac{\partial \hat{u}_{h|\omega}}{\partial n_\gamma}$, $\hat{u}_{h|\gamma}$ and g, respectively, in $\Lambda_H(\gamma)$ [3].

The algebraic representation of (5) can be written in the form

$$F(\overline{y}) = 0 \tag{6}$$

with $F : \mathbb{R}^{n+m} \mapsto \mathbb{R}^{n+m}$ defined by

$$F(\overline{y}) := \begin{pmatrix} A\,\overline{u} - B_\Gamma^\top \overline{\lambda} - \overline{f} \\ G(\overline{u}) \end{pmatrix}, \quad \overline{y} := \begin{pmatrix} \overline{u} \\ \overline{\lambda} \end{pmatrix}, \tag{7}$$

where

$$G(\overline{u}) := C_\gamma\,\overline{u} - \max\{0, C_\gamma\,\overline{u} - \rho(B_\gamma\,\overline{u} - \overline{g})\}$$

and the max-function is understood componentwisely. Here, $A \in \mathbb{R}^{n\times n}$ denotes the standard stiffness matrix, $B_\gamma, B_\Gamma \in \mathbb{R}^{m\times n}$ are the Dirichlet trace matrices related to γ, Γ, respectively, $C_\gamma \in \mathbb{R}^{m\times n}$ is the Neumann trace matrix on γ and $f \in \mathbb{R}^n$, $\overline{g} \in \mathbb{R}^m$.

The equation (6) is nonsmooth due to the presence of the max-function. Fortunately, it is semismooth in the sense of [1] so that a semismooth variant of the Newton method can be used.

4 Algorithm

The concept of semismoothness uses slant differentiability of a function. Here, we recall basic results of [1] related to our problem.

Let Y, Z be Banach spaces and $\mathscr{L}(Y,Z)$ denote the set of all bounded linear mappings of Y into Z. Let $U \subseteq Y$ be an open subset and $F : U \mapsto Z$ a function.

Definition 1. (i) The function F is called slantly differentiable at $y \in U$ if there exists a mapping $F^o : U \mapsto \mathscr{L}(Y,Z)$ such that $\{F^o(y+h)\}$ are uniformly bounded for sufficiently small $h \in Y$ and

$$\lim_{h\to 0} \frac{1}{\|h\|} \|F(y+h) - F(y) - F^o(y+h)h\| = 0.$$

The function F^o is called a slanting function for F at y.
(ii) The function F is called slantly differentiable in U if there exists $F^o : U \mapsto \mathscr{L}(Y,Z)$ such that F^o is a slanting function for F at every point $y \in U$. The function F^o is called a slanting function for F in U.

Theorem 1. *Let F be slantly differentiable in U with a slanting function F^o. Suppose that $y^* \in U$ is a solution to the nonlinear equation $F(y) = 0$. If $F^o(y)$ is non-singular for all $y \in U$ and $\{\|F^o(y)^{-1}\| : y \in U\}$ is bounded, then the Newton method*

$$y^{k+1} = y^k - F^o(y^k)^{-1} F(y^k)$$

converges superlinearly to y^*, provided that $\|y^0 - y^*\|$ is sufficiently small.

Let us focus on the max-function $\psi(y) = \max\{0, y\}$ with $Y = Z = \mathbb{R}$. This function is slantly differentiable and

$$\psi^o(y) = \begin{cases} 1, & y > 0, \\ \sigma, & y = 0, \\ 0, & y < 0, \end{cases}$$

is the slanting function in $\mathbb{R}$ for an arbitrary (but fixed) real number σ. Since the convergence rate of the Newton method does not depend on the choice of a slanting function, we shall use $\psi^o(0) = 0$ below.

The function F defined by (7) is slantly differentiable in $\mathbb{R}^{n+m}$ with

$$F^o(\overline{y}) = \begin{pmatrix} A & -B_\Gamma^\top \\ G^o(\overline{u}) & 0 \end{pmatrix},$$

where

$$\begin{aligned} G^o(\overline{u}) &= (G_1^o(\overline{u}), \ldots, G_m^o(\overline{u}))^\top, \\ G_i^o(\overline{u}) &= C_{\gamma,i} - \psi^o(C_{\gamma,i}\overline{u} - \rho(B_{\gamma,i}\overline{u} - g_i))(C_{\gamma,i} - \rho B_{\gamma,i}), \quad i = 1, \ldots, m, \end{aligned}$$

and the subscript i denotes the i-th row of the corresponding matrix. A more convenient setting of this slanting function F^o uses an active set terminology.

Let $\mathscr{M} := \{1, 2, \ldots, m\}$. We define the sets of *inactive* and *active* indices at $\overline{y} = (\overline{u}^\top, \overline{\lambda}^\top)^\top \in \mathbb{R}^{n+m}$ by

$$\begin{aligned} \mathscr{I} &:= \{i \in \mathscr{M} : C_{\gamma,i}\overline{u} - \rho(B_{\gamma,i}\overline{u} - g_i) \le 0\}, \\ \mathscr{A} &:= \{i \in \mathscr{M} : C_{\gamma,i}\overline{u} - \rho(B_{\gamma,i}\overline{u} - g_i) > 0\}. \end{aligned}$$

It is easily seen that

$$G_i^o(\overline{u}) = \begin{cases} C_{\gamma,i}, & i \in \mathscr{I}, \\ \rho B_{\gamma,i}, & i \in \mathscr{A}, \end{cases}$$

therefore

$$G^o(\overline{u}) = D(\mathscr{I})C_\gamma + \rho D(\mathscr{A})B_\gamma,$$

where $D(\mathscr{S})$ denotes the diagonal matrix for $\mathscr{S} \subseteq \mathscr{M}$ defined by

$$D(\mathscr{S}) = \operatorname{diag}(s_1, \ldots, s_m) \quad \text{with} \quad s_i = \begin{cases} 1, & i \in \mathscr{S}, \\ 0, & i \notin \mathscr{S}. \end{cases}$$

Finally, we obtain

$$F^o(\overline{y}) = \begin{pmatrix} A & -B_\Gamma^\top \\ D(\mathscr{I})C_\gamma + \rho D(\mathscr{A})B_\gamma & 0 \end{pmatrix}.$$

The Newton method leads to the following active-set type algorithm.

Algorithm ASM (Active-Set Method)

(0) Set $k := 0$ and choose $\underline{\rho} > 0$, $\varepsilon_u > 0$ ($\varepsilon_u = 10^{-5}$).
Initialize $\overline{u}^0 \in \mathbb{R}^n$ and $\lambda^0 \in \mathbb{R}^m$.

(1) Define the inactive and active sets by:

$$\begin{aligned} \mathscr{I}^k &:= \{i \in \mathscr{M} : C_{\gamma,i}\overline{u}^k - \rho(B_{\gamma,i}\overline{u}^k - g_i) \le 0\}, \\ \mathscr{A}^k &:= \{i \in \mathscr{M} : C_{\gamma,i}\overline{u}^k - \rho(B_{\gamma,i}\overline{u}^k - g_i) > 0\}. \end{aligned}$$

(2) Solve:

$$\begin{pmatrix} A & -B_\Gamma^\top \\ D(\mathscr{I}^k)C_\gamma + \rho D(\mathscr{A}^k)B_\gamma & 0 \end{pmatrix} \begin{pmatrix} \overline{u}^{k+1} \\ \overline{\lambda}^{k+1} \end{pmatrix} = \begin{pmatrix} \overline{f} \\ \rho D(\mathscr{A}^k)\,\overline{g} \end{pmatrix}.$$

(3) Set $err(k) := \|\overline{u}^{k+1} - \overline{u}^k\|/\|\overline{u}^{k+1}\|$. If $err(k) \le \varepsilon_u$, return $\overline{u} := \overline{u}^{k+1}$.
(4) Set $k := k+1$ and go to step (1).

Remark 1. Let us point out that the algorithm has the finite terminating property provided that all possible matrices in the step (2) are non-singular. This follows directly from the fact that the number of the active sets is finite so that the active set corresponding to the solution, and hence the solution itself, must be found in a finite number of the Newton iterations.

Remark 2. It is readily seen that ρ can be omitted from the linear systems in the step (2). Indeed, if $k \ge 1$ then ρ does not play any role in the definitions of $\mathscr{I}^k$ and $\mathscr{A}^k$ since always either $C_{\gamma,i}\,\overline{u}^k = 0$ or $B_{\gamma,i}\,\overline{u}^k - g_i = 0$. Moreover, an appropriate choice of the initial iterate $\overline{u}^0$ (e.g related to the Dirichlet problem) make possible to omit ρ completely from the algorithm.

The finite terminating property mentioned in Remark 1 assumes the exact solution of the linear systems in the step (2). Numerical experiments however show that the inexact implementation is more efficient. In order to maintain the finite terminating property, we drive the precision control in solving inner linear systems adaptively. Our main idea consists in respecting $err(k-1)$ achieved in the previous Newton iteration and, if the progress is not sufficiently large then the precision of the inner loop is increased independently of $err(k-1)$. Denoting $\delta(k)$ the upper bound for the relative residual terminating iterations of the inner solver [4] in the k-th Newton step, we can express our strategy by

$$\delta(k) := \min\{\varepsilon_{min} \times err(k-1), c_{fact} \times \delta(k-1)\}$$

with $0 < \varepsilon_{min} < 1$, $0 < c_{fact} < 1$, $err(0) = 2$ and $\delta(0) = \varepsilon_{min}/c_{fact}$ (typically $\varepsilon_{min} = 10^{-2}$ and $c_{fact} = 0.2$).

5 Numerical experiments

We illustrate the efficiency of the presented method on the model problem (1), in which $\omega = \{(x,y) \in \mathbb{R}^2 | (x-0.5)^2/0.4^2 + (y-0.5)^2/0.2^2 < 1\}$, $f \equiv -10$ and $g(x,y) = 5\sin(2\varphi)(r^2 + r(\cos\varphi + \sin\varphi) + 0.5)^{1/2} - 1.5$ on γ, where (φ, r) is the polar coordinate of $(x-0.5, y-0.5)$. In the fictitious domain formulation (4) we take $\Omega = (0,1)\times(0,1)$. Moreover we replace $H_0^1(\Omega)$ by $H^1_{per}(\Omega)$ containing periodic functions on Ω that enables us to apply multiplying procedures based on circulant matrices [6]. In the discretized problem (5) we consider V_h formed by piecewise

bilinear functions on a uniform rectangulation of Ω with a stepsize h and $\Lambda_H(\gamma)$, $\Lambda_H(\Gamma)$ defined by piecewise constant functions on partitions of polygonal approximations of γ, Γ, respectively. The curve Γ is constructed by shifting γ three h units in the direction of the outward normal vector n_γ and $H/h = 5$; see Figure 1. The definition of $\delta_H \hat{u}_h$ in (5) uses averaging of gradients.

Fig. 1 Geometry of the problem.

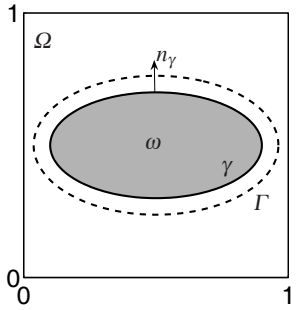

Fig. 2 Solution $\hat{u}_h$ and obstacle g.

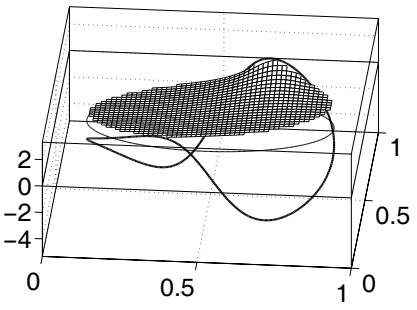

Table 1 Non-smooth fictitious domain formulation ($\gamma \equiv \Gamma$).

Step h	$n/m_{\mathscr{A}}/m_{\mathscr{I}}$	out./$\sum$inn.its.	C.time[s]	$\mathrm{Err}_{L^2(\omega)}$	$\mathrm{Err}_{H^1(\omega)}$	$\mathrm{Err}_{L^2(\gamma)}$
1/128	16641/6/44	11/48	0.81	4.0280e-003	1.7229e-001	1.5350e-002
1/256	66049/13/87	11/62	3.50	2.3784e-003	1.0700e-001	6.3671e-003
1/512	263169/23/175	12/76	31.31	1.9782e-003	1.1129e-001	4.5262e-003
1/1024	1050625/45/351	11/118	185.32	1.0554e-003	8.3205e-002	2.3919e-003
Convergence rates:				0.6063	0.3094	0.8538

Table 2 Smooth fictitious domain formulation ($\gamma \not\equiv \Gamma$).

Step h	$n/m_{\mathscr{A}}/m_{\mathscr{I}}$	out./$\sum$inn.its.	C.time[s]	$\mathrm{Err}_{L^2(\omega)}$	$\mathrm{Err}_{H^1(\omega)}$	$\mathrm{Err}_{L^2(\gamma)}$
1/128	16641/6/44	10/41	0.6875	5.6320e-003	2.6868e-001	2.2502e-002
1/256	66049/13/87	9/73	3.891	1.9606e-003	1.2138e-001	7.3177e-003
1/512	263169/23/175	9/90	34.11	2.8203e-004	2.4997e-002	1.2019e-003
1/1024	1050625/45/351	9/104	161	2.2655e-005	5.5767e-003	1.4466e-004
Convergence rates:				2.6670	1.9051	2.4450

Figure 2 shows the solution $\hat{u}_h$ for $h = 1/256$. In Tables 1, 2 we report the number of primal variables (n), the number of active ($m_{\mathscr{A}} = |\mathscr{A}|$) and inactive ($m_{\mathscr{I}} = |\mathscr{I}|$) control variables, the number of outer (Newton) iterations, the total number of inner (BiCGSTAB) iterations, the computational time and the errors of approximate

solutions in the indicated norms (the comparisons are done with respect to the reference solution computed on the fine mesh with $h = 1/2048$). From the errors, we determine the convergence rate of fictitious domain approaches.

The last experiment in Table 3 compares efficiency of the inexact and exact (with $\delta(k) \equiv 10^{-12}$) implementations of the semi-smooth Newton method. When the active and inactive sets corresponding to the solution are recognized then the exact implementation finds immediately the solution. The inexact implementation divides computations of the solution into several Newton iterations, the total cost is however considerably smaller.

Table 3 Iteration history for $h = 1/256$ and various implementations.

	Exact		Inexact, $c_{fact} = 0.01$		Inexact, $c_{fact} = 0.5$	
k	inn.its.	$m_{\mathscr{A}^k}/m_{\mathscr{I}^k}$	inn.its.	$m_{\mathscr{A}^k}/m_{\mathscr{I}^k}$	inn.its.	$m_{\mathscr{A}^k}/m_{\mathscr{I}^k}$
0	65	33/67	1	33/67	2	33/67
1	84	26/74	2	25/75	1	21/79
2	70	20/80	13	16/84	5	16/84
3	69	16/84	23	14/86	8	19/81
4	54	14/86	34	13/87	10	16/84
5	51	13/87	13	13/87	9	13/87
6	0	13/87			4	13/87
7					6	13/87
8					6	13/87
9					6	13/87
10					3	13/87
11					6	13/87
12					4	13/87
$\sum$inn.its.	393		86		70	

Acknowledgments: This research is supported by the grant GAČR 201/07/0294 and by the research project MSM0021620839. The second author also acknowledges the support of the grant IAA1075402.

References

1. Chen, X., Nashed, Z., Qi, L.: Smoothing methods and semismooth methods for nondifferentiable operator equation. SIAM J. Numer. Anal. **38**, 1200–1216 (2000)
2. Glowinski, R., Pan, T., Periaux, J.: A fictitious domain method for Dirichlet problem and applications. Comput. Meth. Appl. Mech. Eng. **111**, 283–303 (1994)
3. Haslinger, J., Kozubek, T., Kučera: Fictitious domain formulation of unilateral problems: analysis and algorithms. In preparing (2008)
4. Haslinger, J., Kozubek, T., Kučera, R., Peichel, G.: Projected Schur complement method for solving non-symmetric systems arising from a smooth fictitious domain approach. Num. Lin. Algebra Appl. **14**, 713–739 (2007)
5. Hintermüller, M., Ito, K., Kunisch, K.: The primal-dual active set strategy as a semismooth Newton method. SIAM J. Optim. **13**, 865–888 (2002)
6. Kučera, R.: Complexity of an algorithm for solving saddle-point systems with singular blocks arising in wavelet-Galerkin discretizations. Appl. Math. **50**(3), 291–308 (2005)