

# Semi-smooth Newton method for Solving Unilateral Problems in Fictitious Domain Formulations

Jaroslav Haslinger, Charles University, Prague

Tomáš Kozubek, VŠB–TU Ostrava

Radek Kučera, VŠB–TU Ostrava

SNA'08, Liberec

# OUTLINE

Motivation

Fictitious domain formulation

Algorithm – Semi-Smooth Newton method

Inner solvers

Numerical experiments

# OUTLINE

Motivation

Fictitious domain formulation

Algorithm – Semi-Smooth Newton method

Inner solvers

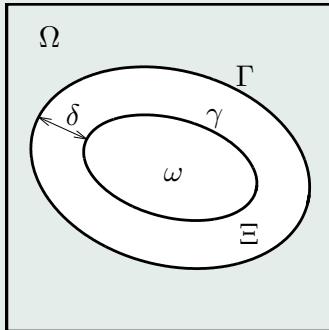
Numerical experiments

## FDM for Dirichlet problem:

$$\left. \begin{array}{l} -\Delta u + u = f \quad \text{on } \omega \\ u = g \quad \text{in } \gamma \equiv \partial\omega \end{array} \right\}$$

### Main idea of FDM:

PDE is solved on the fictitious domain  $\Omega$ ,  $\bar{\omega} \subset \Omega$ , with a simple geometry. The corresponding stiffness matrix  $\mathbf{A}$  is structured. The original boundary conditions on  $\gamma$  are enforced by Lagrange multipliers or control variables.



Classical FDM with  $\Gamma \equiv \gamma$

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}_\gamma^\top \\ \mathbf{B}_\gamma & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \boldsymbol{\lambda}_\gamma \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix}$$

Smooth FDM with  $\Gamma \not\equiv \gamma$

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}_\Gamma^\top \\ \mathbf{B}_\gamma & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \boldsymbol{\lambda}_\Gamma \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix}$$

## FDM for unilateral problem:

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

Equivalent form of BC:  $\frac{\partial u}{\partial n_\gamma} = \max \left\{ 0, \frac{\partial u}{\partial n_\gamma} - \rho(u - g) \right\}, \quad \rho > 0$

### Discretized FDM

$$\left. \begin{array}{l} \mathbf{A}\mathbf{u} + \mathbf{B}_\Gamma \boldsymbol{\lambda}_\Gamma - \mathbf{f} = \mathbf{0} \\ \mathbf{C}_\gamma \mathbf{u} - \max \left\{ 0, \mathbf{C}_\gamma \mathbf{u} - \rho(\mathbf{B}_\gamma \mathbf{u} - \mathbf{g}) \right\} = \mathbf{0} \end{array} \right\} \quad F(\mathbf{u}, \boldsymbol{\lambda}_\Gamma) = \mathbf{0}$$

where  $\mathbf{B}_\gamma$ ,  $\mathbf{B}_\Gamma$  and  $\mathbf{C}_\gamma$  are Dirichlet and Neumann trace matrices, respectively.

This system may be solved by a semi-smooth Newton method with:

$$\textbf{Jacobian} \sim \begin{pmatrix} \mathbf{A} & \mathbf{B}_\Gamma^\top \\ \partial G(\mathbf{u}) & \mathbf{0} \end{pmatrix}$$

# **OUTLINE**

Motivation

Fictitious domain formulation

Algorithm – Semi-Smooth Newton method

Inner solvers

Numerical experiments

## Formulation of the problem:

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

where  $f \in L^2_{loc}(\mathbb{R}^2)$ ,  $g \in H^{1/2}(\gamma)$ .

## Weak formulation:

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

It is equivalent to a convex minimization on a convex set. Here,  $\lambda := \frac{\partial u}{\partial n_\gamma}$  plays the role of the Lagrange multiplier.

**Reformulation of the inequality:** for  $\frac{\partial u}{\partial n_\gamma} \in L_+^2(\gamma)$ ,  $\rho > 0$

$$\langle \mu - \frac{\partial u}{\partial n_\gamma}, u - g \rangle_\gamma \geq 0 \quad \forall \mu \in H_+^{-1/2}(\gamma)$$

$\Updownarrow$

$$(\mu - \frac{\partial u}{\partial n_\gamma}, u - g)_{0,\gamma} \geq 0 \quad \forall \mu \in L_+^2(\gamma)$$

$\Updownarrow$

$$(\mu - \frac{\partial u}{\partial n_\gamma}, \frac{\partial u}{\partial n_\gamma} - \rho(u - g) - \frac{\partial u}{\partial n_\gamma})_{0,\gamma} \leq 0 \quad \forall \mu \in L_+^2(\gamma)$$

$\Updownarrow$

$$\frac{\partial u}{\partial n_\gamma} = P\left(\frac{\partial u}{\partial n_\gamma} - \rho(u - g)\right)$$

**Projection:**  $P : L^2(\gamma) \mapsto L_+^2(\gamma) : P\varphi = \max\{0, \varphi(x)\}$ ,  $x \in \gamma$

**Reformulation of the weak formulation:** if  $\frac{\partial u}{\partial n_\gamma} \in L_+^2(\gamma)$

$$\left. \begin{aligned} (u, v)_{1,\omega} &= (f, v)_{0,\omega} + (\frac{\partial u}{\partial n_\gamma}, v)_{0,\gamma} \quad \forall v \in H^1(\omega) \\ \frac{\partial u}{\partial n_\gamma} &= P(\frac{\partial u}{\partial n_\gamma} - \rho(u - g)) \end{aligned} \right\}$$

**Fictitious domain formulation:**

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

Here,  $\lambda$  represents the jump of normal derivative on  $\Gamma$ .

## Discretization:

Let  $V_h \subset H_{per}^1(\Omega)$ ,  $\Lambda_H(\gamma) \subset L^2(\gamma)$ ,  $\Lambda_H(\Gamma) \subset L^2(\Gamma)$  be such that

$$\dim V_h = n, \quad \dim \Lambda_H(\gamma) = \dim \Lambda_H(\Gamma) = m.$$

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

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  $\Lambda_H(\gamma)$ .

## Algebraic form:

$$\left. \begin{array}{l} \mathbf{A}\mathbf{u} - \mathbf{B}_\Gamma \boldsymbol{\lambda}_\Gamma - \mathbf{f} = \mathbf{0} \\ \mathbf{C}_\gamma \mathbf{u} - \max \{0, \mathbf{C}_\gamma \mathbf{u} - \rho(\mathbf{B}_\gamma \mathbf{u} - \mathbf{g})\} = \mathbf{0} \end{array} \right\} \quad F(\mathbf{u}, \boldsymbol{\lambda}_\Gamma) = \mathbf{0}$$

# **OUTLINE**

Motivation

Fictitious domain formulation

**Algorithm – Semi-Smooth Newton method**

Inner solvers

Numerical experiments

## **Definition of slanting functions** (X. Chen, Z. Nashed, L. Qi – 2000)

A function  $F : Y \mapsto Z$  is said to be slantly differentiable at  $y$  if

- $\exists F^o : Y \mapsto L(Y, Z)$ :

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

- $F^o$  is uniformly bounded in an open neighbourhood of  $y$ .

$F^o$  is said to be a slanting function for  $F$  at  $y$ .

**Example:**  $\psi : \mathbb{R} \mapsto \mathbb{R}, y \mapsto \max\{0, y\}$

$$\psi^o(y) = \begin{cases} 0 & \text{if } y < 0 \\ 1 & \text{if } y > 0 \\ \sigma \in \mathbb{R} & \text{if } y = 0 \end{cases}$$

## Theorem (X. Chen, Z. Nashed, L. Qi – 2000)

Let  $F$  be slantly differentiable in an open subset  $U \subset Y$  with a slanting function  $F^o : U \mapsto L(Y, Z)$ , and  $y^* \in U$  be 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 by a constant  $M \geq 0$ , then the Newton iteration

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

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

### Super-linear convergence:

$$\lim_{k \rightarrow \infty} \frac{\|y^{k+1} - y^*\|}{\|y^k - y^*\|} = 0$$

**Linear convergence:**  $\|y^{k+1} - y^*\| \leq \eta \|y^k - y^*\|, \quad 0 \leq \eta < 1$

**Quadratic convergence:**  $\|y^{k+1} - y^*\| \leq C \|y^k - y^*\|^2, \quad 0 \leq C$

## Idea of the proof:

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

$$y^{k+1} - y^* = y^k - y^* - F^o(y^k)^{-1}(F(y^k) - F(y^*)), \quad h = y^k - y^*$$

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

$$\|y^{k+1} - y^*\| \leq M \|F^o(y^* + h)h - F(y^* + h) + F(y^*)\|$$

$$\frac{\|y^{k+1} - y^*\|}{\|y^k - y^*\|} \leq M \frac{\|F(y^* + h) - F(y^*) - F^o(y^* + h)h\|}{\|h\|}$$

- if it converges, then it converges super-linearly
- it converges, provided that  $\|y^0 - y^*\|$  is sufficiently small (by induction)

**Slanting function for our problem:**  $F : \mathbb{R}^{n+m} \mapsto \mathbb{R}^{n+m}$ ,  $\mathbf{y} := (\mathbf{u}^\top, \boldsymbol{\lambda}^\top)^\top$

$$F(\mathbf{y}) := \begin{pmatrix} \mathbf{A}\mathbf{u} - \mathbf{B}_\Gamma^\top \boldsymbol{\lambda} - \mathbf{f} \\ G(\mathbf{u}) \end{pmatrix}, \quad G(\mathbf{u}) := (G_1(\mathbf{u}), \dots, G_m(\mathbf{u}))^\top$$

$$G_i(\mathbf{u}) := \mathbf{C}_{\gamma,i}\mathbf{u} - \max \{0, \mathbf{C}_{\gamma,i}\mathbf{u} - \rho(\mathbf{B}_{\gamma,i}\mathbf{u} - \mathbf{g}_i)\}$$

$$F^o(\mathbf{y}) = \begin{pmatrix} \mathbf{A} & -\mathbf{B}_\Gamma^\top \\ G^o(\mathbf{u}) & 0 \end{pmatrix}, \quad G^o(\mathbf{u}) = (G_1^o(\mathbf{u}), \dots, G_m^o(\mathbf{u}))^\top$$

$$G_i^o(\mathbf{u}) = \mathbf{C}_{\gamma,i} - \psi^o(\mathbf{C}_{\gamma,i}\mathbf{u} - \rho(\mathbf{B}_{\gamma,i}\mathbf{u} - \mathbf{g}_i))(\mathbf{C}_{\gamma,i} - \rho\mathbf{B}_{\gamma,i})$$

**Active/Inactive set terminology:**  $\mathcal{M} := \{1, 2, \dots, m\}$

$$\mathcal{A} := \{i \in \mathcal{M} : \mathbf{C}_{\gamma,i}\mathbf{u} - \rho(\mathbf{B}_{\gamma,i}\mathbf{u} - \mathbf{g}_i) > 0\}$$

$$\mathcal{I} := \{i \in \mathcal{M} : \mathbf{C}_{\gamma,i}\mathbf{u} - \rho(\mathbf{B}_{\gamma,i}\mathbf{u} - \mathbf{g}_i) \leq 0\}$$

$$G_i^o(\mathbf{u}) = \begin{cases} \rho\mathbf{B}_{\gamma,i} & \text{for } i \in \mathcal{A} \\ \mathbf{C}_{\gamma,i} & \text{for } i \in \mathcal{I} \end{cases}$$

## Active set algorithm:

(0) Set  $k := 0$ ,  $\rho > 0$ ,  $\varepsilon_u > 0$ ,  $\mathbf{u}^0 \in \mathbb{R}^n$ ,  $\boldsymbol{\lambda}^0 \in \mathbb{R}^m$ .

(1) Define the inactive and active sets by:

$$\begin{aligned}\mathcal{I}^k &:= \{i : \mathbf{C}_{\gamma,i}\mathbf{u}^k - \rho(\mathbf{B}_{\gamma,i}\mathbf{u}^k - \mathbf{g}_i) \leq 0\} \\ \mathcal{A}^k &:= \{i : \mathbf{C}_{\gamma,i}\mathbf{u}^k - \rho(\mathbf{B}_{\gamma,i}\mathbf{u}^k - \mathbf{g}_i) > 0\}\end{aligned}$$

(2) Solve:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}_\Gamma^\top \\ \rho\mathbf{B}_{\gamma,\mathcal{A}^k} & \mathbf{0} \\ \mathbf{C}_{\gamma,\mathcal{I}^k} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u}^{k+1} \\ \boldsymbol{\lambda}_\Gamma^{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \rho\mathbf{g}_{\mathcal{A}^k} \\ \mathbf{0} \end{pmatrix}$$

(3) If  $\|\mathbf{u}^{k+1} - \mathbf{u}^k\| / \|\mathbf{u}^{k+1}\| \leq \varepsilon_u$ , return  $\mathbf{u} := \mathbf{u}^k$ .

(4) Set  $k := k + 1$ , and go to step (1).

Remark: The mixed Dirichlet-Neumann problem is solved in each Newton step.

# OUTLINE

Motivation

Fictitious domain formulation

Algorithm – Semi-Smooth Newton method

Inner solvers: Schur complement + Null-space method

Projected BiCGSTAB

Circulant matrices

Numerical experiments

## Non-symmetric saddle-point system

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}_1^\top \\ \mathbf{B}_2 & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix}$$

### General assumptions

**A** ... **non-symmetric** ( $n \times n$ )–matrix  
... **singular** with  $p = \dim \text{Ker } \mathbf{A}$

**B**<sub>1</sub>, **B**<sub>2</sub> ... **full rank** ( $m \times n$ )–matrices  
...  **$\mathbf{B}_1 \neq \mathbf{B}_2$**

### Special FDM assumptions

- $n$  is large ( $n = 4198401$ )
- $m \ll n$  ( $m = 360$ )
- $p \ll m$  ( $p = 1$ , periodic BC on the fictitious boundary  $\partial\Omega$  for  $-\Delta$ )
- **A** is **structured** (circulant) so that actions of  $\mathbf{A}^\dagger$  (or  $\mathbf{A}^{-1}$ ) are "cheap"
- **B**<sub>1</sub>, **B**<sub>2</sub> are highly **sparse** so that their actions are "cheap"

## Schur complement reduction

- a generalized inverse  $\mathbf{A}^\dagger$
- columns of  $(n \times p)$ -matrices  $\mathbf{N}, \mathbf{M}$  span  $\text{Ker } \mathbf{A}, \text{Ker } \mathbf{A}^\top$ , respectively

$$\mathbf{A}\mathbf{u} + \mathbf{B}_1^\top \boldsymbol{\lambda} = \mathbf{f} \iff \mathbf{f} - \mathbf{B}_1^\top \boldsymbol{\lambda} \in \text{Im } \mathbf{A} \perp \text{Ker } \mathbf{A}^\top$$

$\Updownarrow$

$\Updownarrow$

$$\mathbf{u} = \mathbf{A}^\dagger(\mathbf{f} - \mathbf{B}_1^\top \boldsymbol{\lambda}) + \mathbf{N}\boldsymbol{\alpha} \quad \mathbf{M}^\top(\mathbf{f} - \mathbf{B}_1^\top \boldsymbol{\lambda}) = \mathbf{0}$$

&

The reduced system:

$$\mathbf{B}_2\mathbf{u} = \mathbf{g}$$

$$\begin{pmatrix} \mathbf{B}_2\mathbf{A}^\dagger\mathbf{B}_1^\top & -\mathbf{B}_2\mathbf{N} \\ -\mathbf{M}^\top\mathbf{B}_1^\top & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_2\mathbf{A}^\dagger\mathbf{f} - \mathbf{g} \\ -\mathbf{M}^\top\mathbf{f} \end{pmatrix}$$

$\Downarrow$

$$\mathbf{B}_2\mathbf{A}^\dagger\mathbf{B}_1^\top \boldsymbol{\lambda} - \mathbf{B}_2\mathbf{N}\boldsymbol{\alpha} = \mathbf{B}_2\mathbf{A}^\dagger\mathbf{f} - \mathbf{g}$$

## Null-space method for the reduced system

$$\begin{pmatrix} \mathbf{F} & \mathbf{G}_1^\top \\ \mathbf{G}_2 & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{d} \\ \mathbf{e} \end{pmatrix}$$

Two orthogonal projectors  $\mathbf{P}_1$  and  $\mathbf{P}_2$  onto  $\text{Ker } \mathbf{G}_1$  and  $\text{Ker } \mathbf{G}_2$ :

$$\mathbf{P}_k : \mathbb{R}^m \mapsto \text{Ker } \mathbf{G}_k, \quad \mathbf{P}_k := \mathbf{I} - \mathbf{G}_k^\top (\mathbf{G}_k \mathbf{G}_k^\top)^{-1} \mathbf{G}_k, \quad k = 1, 2$$

Property:  $\text{Ker } \mathbf{P}_k = \text{Im } \mathbf{G}_k^\top \iff \mathbf{P}_k \mathbf{G}_k^\top = \mathbf{0}$

- $\mathbf{P}_1$  splits the saddle-point structure:  $\mathbf{P}_1 \mathbf{F} \boldsymbol{\lambda} + \mathbf{P}_1 \mathbf{G}_1^\top \boldsymbol{\alpha} = \mathbf{P}_1 \mathbf{d}$

$$\mathbf{P}_1 \mathbf{F} \boldsymbol{\lambda} = \mathbf{P}_1 \mathbf{d}, \quad \mathbf{G}_2 \boldsymbol{\lambda} = \mathbf{e}, \quad \boldsymbol{\alpha} := (\mathbf{G}_1 \mathbf{G}_1^\top)^{-1} (\mathbf{G}_1 \mathbf{d} - \mathbf{G}_1 \mathbf{F} \boldsymbol{\lambda})$$

- $\mathbf{P}_2$  decomposes  $\boldsymbol{\lambda} = \boldsymbol{\lambda}_{\text{Im}} + \boldsymbol{\lambda}_{\text{Ker}}$ ,  $\boldsymbol{\lambda}_{\text{Im}} \in \text{Im } \mathbf{G}_2^\top$ ,  $\boldsymbol{\lambda}_{\text{Ker}} \in \text{Ker } \mathbf{G}_2$

At first:  $\mathbf{G}_2 \boldsymbol{\lambda} = \mathbf{G}_2 \boldsymbol{\lambda}_{\text{Im}} = \mathbf{e} \implies \boldsymbol{\lambda}_{\text{Im}} := \mathbf{G}_2^\top (\mathbf{G}_2 \mathbf{G}_2^\top)^{-1} \mathbf{e}$

At second:  $\mathbf{P}_1 \mathbf{F} \boldsymbol{\lambda}_{\text{Ker}} = \mathbf{P}_1 (\mathbf{d} - \mathbf{F} \boldsymbol{\lambda}_{\text{Im}}) \text{ on } \text{Ker } \mathbf{G}_2$

## Algorithm PSCM

Step 1.a: Assemble  $\mathbf{G}_1 := -\mathbf{N}^\top \mathbf{B}_2^\top$ ,  $\mathbf{G}_2 := -\mathbf{M}^\top \mathbf{B}_1^\top$ .

Step 1.b: Assemble  $\mathbf{d} := \mathbf{B}_2 \mathbf{A}^\dagger \mathbf{f} - \mathbf{g}$ ,  $\mathbf{e} := -\mathbf{M}^\top \mathbf{f}$ .

Step 1.c: Assemble  $\mathbf{H}_1 := (\mathbf{G}_1 \mathbf{G}_1^\top)^{-1}$ ,  $\mathbf{H}_2 := (\mathbf{G}_2 \mathbf{G}_2^\top)^{-1}$ .

Step 1.d: Assemble  $\boldsymbol{\lambda}_{Im} := \mathbf{G}_2^\top \mathbf{H}_2 \mathbf{e}$ ,  $\tilde{\mathbf{d}} := \mathbf{P}_1(\mathbf{d} - \mathbf{F}\boldsymbol{\lambda}_{Im})$ .

**Step 1.e:** Solve  $\mathbf{P}_1 \mathbf{F} \boldsymbol{\lambda}_{Ker} = \tilde{\mathbf{d}}$  on  $Ker \mathbf{G}_2$ .

Step 1.f: Assemble  $\boldsymbol{\lambda} := \boldsymbol{\lambda}_{Im} + \boldsymbol{\lambda}_{Ker}$ .

Step 2: Assemble  $\boldsymbol{\alpha} := \mathbf{H}_1 \mathbf{G}_1 (\mathbf{d} - \mathbf{F} \boldsymbol{\lambda})$ .

Step 3: Assemble  $\mathbf{u} := \mathbf{A}^\dagger (\mathbf{f} - \mathbf{B}_1^\top \boldsymbol{\lambda}) + \mathbf{N} \boldsymbol{\alpha}$ .

- an iterative **projected** Krylov subspace method for non-symmetric operators can be used in Step 1.e
- matrix-vector products

$$\mathbf{F} \boldsymbol{\mu} := \left( \mathbf{B}_2 \left( \mathbf{A}^\dagger \left( \mathbf{B}_1^\top \boldsymbol{\mu} \right) \right) \right), \quad \mathbf{P}_k \boldsymbol{\mu} := \boldsymbol{\mu} - \left( \mathbf{G}_k^\top \left( \mathbf{H}_k \left( \mathbf{G}_k \boldsymbol{\mu} \right) \right) \right)$$

Idea: keep the iterative process for solving  $\mathbf{P}_1 \mathbf{F} \boldsymbol{\lambda}_{Ker} = \tilde{\mathbf{d}}$  on  $Ker \mathbf{G}_2$ .

**Algorithm ProjBiCGSTAB**  $[\epsilon, \boldsymbol{\lambda}^0, \mathbf{F}, \mathbf{P}_1, \mathbf{P}_2, \tilde{\mathbf{d}}] \rightarrow \boldsymbol{\lambda}$

Initialize:  $\boldsymbol{\lambda}^0 \in Ker \mathbf{G}_2$ ,  $\mathbf{r}^0 := \mathbf{P}_2 \tilde{\mathbf{d}} - \mathbf{P}_2 \mathbf{P}_1 \mathbf{F} \boldsymbol{\lambda}^0$ ,  $\mathbf{p}^0 := \mathbf{r}^0$ ,  $\tilde{\mathbf{r}}^0$ ,  $k := 0$

**While**  $\|\mathbf{r}^k\| > \epsilon$

$$1^\circ \quad \tilde{\mathbf{p}}^k := \mathbf{P}_2 \mathbf{P}_1 \mathbf{F} \mathbf{p}^k$$

$$2^\circ \quad \alpha_k := (\mathbf{r}^k)^\top \tilde{\mathbf{r}}^0 / (\tilde{\mathbf{p}}^k)^\top \tilde{\mathbf{r}}^0$$

$$3^\circ \quad \mathbf{s}^k := \mathbf{r}^k - \alpha_k \tilde{\mathbf{p}}^k$$

$$4^\circ \quad \tilde{\mathbf{s}}^k := \mathbf{P}_2 \mathbf{P}_1 \mathbf{F} \mathbf{s}^k$$

$$5^\circ \quad \omega_k := (\tilde{\mathbf{s}}^k)^\top \mathbf{s}^k / (\tilde{\mathbf{s}}^k)^\top \tilde{\mathbf{s}}^k$$

$$6^\circ \quad \boldsymbol{\lambda}^{k+1} := \boldsymbol{\lambda}^k + \alpha_k \mathbf{p}^k + \omega_k \mathbf{s}^k$$

$$7^\circ \quad \mathbf{r}^{k+1} := \mathbf{s}^k - \omega_k \tilde{\mathbf{s}}^k$$

$$8^\circ \quad \beta_{k+1} := (\alpha_k / \omega_k) (\mathbf{r}^{k+1})^\top \tilde{\mathbf{r}}^0 / (\mathbf{r}^k)^\top \tilde{\mathbf{r}}^0$$

$$9^\circ \quad \mathbf{p}^{k+1} := \mathbf{r}^{k+1} + \beta_{k+1} (\mathbf{p}^k - \omega_k \tilde{\mathbf{p}}^k)$$

$$10^\circ \quad k := k + 1$$

**end**

**Theorem 2** The saddle-point matrix  $\mathcal{A} := \begin{pmatrix} \mathbf{A} & \mathbf{B}_1^\top \\ \mathbf{B}_2 & \mathbf{0} \end{pmatrix}$  is invertible iff

$$\left. \begin{array}{l} \mathbf{B}_1 \text{ has full row-rank} \\ \mathbf{Ker} \mathbf{A} \cap \mathbf{Ker} \mathbf{B}_2 = \{\mathbf{0}\} \\ \mathbf{A} \mathbf{Ker} \mathbf{B}_2 \cap \mathbf{Im} \mathbf{B}_1^\top = \{\mathbf{0}\} \end{array} \right\} \text{(NSC)}$$

The generalized Schur complement: the matrix of the reduced system

$$\mathcal{S} := \begin{pmatrix} -\mathbf{B}_2 \mathbf{A}^\dagger \mathbf{B}_1^\top & \mathbf{B}_2 \mathbf{N} \\ \mathbf{M}^\top \mathbf{B}_1^\top & \mathbf{0} \end{pmatrix}$$

**Theorem 3** The following three statements are equivalent:

- The necessary and sufficient condition (NSC) holds.
- $\mathcal{A}$  is invertible.
- $\mathcal{S}$  is invertible.

Remark: The generalized Schur complement  $\mathcal{S}$  is not defined uniquely.

**Theorem 4** Let  $\mathcal{A}$  be invertible. The linear operator  $\mathbf{P}_1\mathbf{F}: \text{Ker } \mathbf{G}_2 \mapsto \text{Ker } \mathbf{G}_1$  is invertible.

**Proof:** Notice  $\dim \text{Ker } \mathbf{G}_1 = \dim \text{Ker } \mathbf{G}_2$ .

Let  $\mu \in \text{Ker } \mathbf{G}_2$  be such that  $\mathbf{P}_1\mathbf{F}\mu = \mathbf{0}$ . Then  $\mathbf{F}\mu \in \text{Ker } \mathbf{P}_1 = \text{Im } \mathbf{G}_1^\top$  and, therefore, there is  $\beta \in \mathbb{R}^p$  so that

$$\mathbf{F}\mu = \mathbf{G}_1^\top \beta \quad \text{and} \quad \mathbf{G}_2\mu = \mathbf{0}.$$

We obtain

$$\begin{pmatrix} \mathbf{F} & \mathbf{G}_1^\top \\ \mathbf{G}_2 & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mu \\ -\beta \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

where the matrix is  $-\mathcal{S}$  (i.e. the generalized Schur complement) that is invertible iff  $\mathcal{A}$  is invertible. Therefore  $\mu = \mathbf{0}$ .

## Circulant matrices and Fourier transform

$$\mathbf{A} = \begin{pmatrix} a_1 & a_n & \dots & a_2 \\ a_2 & a_1 & \dots & a_3 \\ a_3 & a_2 & \dots & a_4 \\ \vdots & \vdots & \ddots & \vdots \\ a_n & a_{n-1} & \dots & a_1 \end{pmatrix} = (\mathbf{a}, \mathbf{T}\mathbf{a}, \mathbf{T}^2\mathbf{a}, \dots, \mathbf{T}^{n-1}\mathbf{a})$$

$$\widehat{\mathcal{T}_k f}(\omega) = \int_R f(x - k) e^{-ix\omega} dx = e^{-ik\omega} \widehat{f}(\omega)$$

$$\mathbf{X}\mathbf{A} = (\mathbf{Dx}_0, \mathbf{Dx}_1, \mathbf{Dx}_2, \dots, \mathbf{Dx}_{n-1}) = \mathbf{DX}$$

*Lemma:* Let  $\mathbf{A}$  be circulant. Then

$$\mathbf{A} = \mathbf{X}^{-1} \mathbf{D} \mathbf{X},$$

where  $\mathbf{X}$  is the DFT matrix and  $\mathbf{D} = \text{diag}(\widehat{\mathbf{a}})$ ,  $\widehat{\mathbf{a}} = \mathbf{X}\mathbf{a}$ ,  $\mathbf{a} = \mathbf{A}(:, 1)$ .

Multiplying procedure:  $\mathbf{A}^\dagger \mathbf{v} := \mathbf{X}^{-1} (\mathbf{D}^\dagger (\mathbf{X}\mathbf{v})) \dots$  Moore-Penrose

$$\left. \begin{array}{ll} 0^\circ & \mathbf{d} := \text{fft}(\mathbf{a}) \\ 1^\circ & \mathbf{v} := \text{fft}(\mathbf{v}) \\ 2^\circ & \mathbf{v} := \mathbf{v} \cdot * \mathbf{d}^{-1} \\ 3^\circ & \mathbf{A}^\dagger \mathbf{v} := \text{ifft}(\mathbf{v}) \end{array} \right\} \mathcal{O}(2n \log_2 n)$$

Multiplying procedures:  $\mathbf{N}\boldsymbol{\alpha}$ ,  $\mathbf{N}^\top \mathbf{v}$  (and  $\mathbf{M}\boldsymbol{\alpha}$ ,  $\mathbf{M}^\top \mathbf{v}$ )

As  $\mathbf{A}\mathbf{N} = \mathbf{0}$ , the matrix  $\mathbf{N}$  may be formed by eigenvectors corresponding to zero eigenvalues.

$$\mathbf{I} - \mathbf{D}\mathbf{D}^\dagger = \text{diag}(1, 1, 1, 0, \dots, 0) \implies \mathbf{X}^{-1} = (\mathbf{N}, \mathbf{Y}), \quad \mathbf{X}^{-1} = \begin{pmatrix} \mathbf{N}^\top \\ \mathbf{Y} \end{pmatrix}$$

Therefore we can define the operation:  $\text{ind}(\boldsymbol{\alpha}) = \begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{0} \end{pmatrix} \in \mathbb{R}^n$

$$\left. \begin{array}{ll} 1^\circ & \mathbf{v}_\boldsymbol{\alpha} := \text{ind}(\boldsymbol{\alpha}) \\ 2^\circ & \mathbf{N}\boldsymbol{\alpha} := \text{ifft}(\mathbf{v}_\boldsymbol{\alpha}) \end{array} \right. \left. \begin{array}{ll} 1^\circ & \mathbf{v} := \text{ifft}(\mathbf{v}) \\ 2^\circ & \mathbf{N}^\top \mathbf{v} := \text{ind}^{-1}(\mathbf{v}) \end{array} \right\} \mathcal{O}(n \log_2 n)$$

Kronecker product of matrices:  $\mathbf{A}_x \in \mathbb{R}^{n_x \times n_x}$ ,  $\mathbf{A}_y \in \mathbb{R}^{n_y \times n_y}$

$$\mathbf{A}_x \otimes \mathbf{A}_y = \begin{pmatrix} a_{11}^y \mathbf{A}_x & \dots & a_{1n_y}^y \mathbf{A}_x \\ \vdots & \ddots & \vdots \\ a_{n_y 1}^y \mathbf{A}_x & \dots & a_{n_y n_y}^y \mathbf{A}_x \end{pmatrix}$$

*Lemma 1:*  $(\mathbf{A}_x \otimes \mathbf{A}_y)(\mathbf{B}_x \otimes \mathbf{B}_y) = \mathbf{A}_x \mathbf{B}_x \otimes \mathbf{A}_y \mathbf{B}_y$

$$(\mathbf{A}_x \otimes \mathbf{A}_y)^\dagger = \mathbf{A}_x^\dagger \otimes \mathbf{A}_y^\dagger$$

$$\mathbf{N} = \mathbf{N}_x \otimes \mathbf{N}_y$$

*Lemma 2:*  $(\mathbf{A}_x \otimes \mathbf{A}_y)\mathbf{v} = \text{vec}(\mathbf{A}_x \mathbf{V} \mathbf{A}_y^\top)$ , where  $\mathbf{V} = \text{vec}^{-1}(\mathbf{v})$ .

$$\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_{n_y}) \in \mathbb{R}^{n_x \times n_y} \iff \text{vec}(\mathbf{V}) = \begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_{n_y} \end{pmatrix} \in \mathbb{R}^{n_x n_y}$$

Kronecker product and circulant matrices: Let  $\mathbf{A}_x, \mathbf{A}_y$  be circulant then:

$$\begin{aligned}\mathbf{A} &= \mathbf{A}_x \otimes \mathbf{I}_y + \mathbf{I}_x \otimes \mathbf{A}_y \\&= \mathbf{X}_x^{-1} \mathbf{D}_x \mathbf{X}_x \otimes \mathbf{X}_y^{-1} \mathbf{X}_y + \mathbf{X}_x^{-1} \mathbf{X}_x \otimes \mathbf{X}_y^{-1} \mathbf{D}_y \mathbf{X}_y \\&= (\mathbf{X}_x^{-1} \otimes \mathbf{X}_y^{-1})(\mathbf{D}_x \otimes \mathbf{I}_y + \mathbf{I}_x \otimes \mathbf{D}_y)(\mathbf{X}_x \otimes \mathbf{X}_y) \\&= \mathbf{X}^{-1} \mathbf{D} \mathbf{X}\end{aligned}$$

with

$$\mathbf{X} = \mathbf{X}_x \otimes \mathbf{X}_y \quad (\text{DFT matrix in 2D})$$

$$\mathbf{D} = \mathbf{D}_x \otimes \mathbf{I}_y + \mathbf{I}_x \otimes \mathbf{D}_y \quad (\text{diagonal matrix})$$

where  $\mathbf{X}_x, \mathbf{X}_y$  are the DFT matrices,  $\mathbf{D}_x = \text{diag}(\mathbf{X}_x \mathbf{a}_x)$ ,  $\mathbf{D}_y = \text{diag}(\mathbf{X}_y \mathbf{a}_y)$  and  $\mathbf{a}_x = \mathbf{A}_x(:, 1)$ ,  $\mathbf{a}_y = \mathbf{A}_y(:, 1)$ , respectively.

Multiplying procedure:  $\mathbf{A}^\dagger \mathbf{v} := \mathbf{X}^{-1} (\mathbf{D}^\dagger (\mathbf{X} \mathbf{v}))$

- 0°  $\mathbf{d}_x := \text{fft}(\mathbf{a}_x), \mathbf{d}_y := \text{fft}(\mathbf{a}_y)$
- $\mathbf{V} := \text{vec}^{-1}(\mathbf{v})$
- 1°  $\mathbf{V} := \text{fft}(\mathbf{V})$
- 2°  $\mathbf{V} := \text{fft}(\mathbf{V}^\top)^\top$
- 3°  $\mathbf{V} := \text{vec}^{-1}(\mathbf{D}^\dagger \text{vec}(\mathbf{V}))$
- 4°  $\mathbf{V} := \text{ifft}(\mathbf{V})$
- 5°  $\mathbf{V} := \text{ifft}(\mathbf{V}^\top)^\top$
- $\mathbf{A}^\dagger \mathbf{v} := \text{vec}(\mathbf{V})$

Number of arithmetic operations:

$$\mathcal{O}(2n(\log_2 n_x + \log_2 n_y) + n) \approx \mathcal{O}(n \log_2 n), \quad n = n_x n_y$$

Multiplying procedures:  $\mathbf{N}\boldsymbol{\alpha}, \mathbf{N}^\top \mathbf{v}, \mathbf{M}\boldsymbol{\alpha}, \mathbf{M}^\top \mathbf{v}$  . . . analogous

## REFERENCES

- J. Haslinger, T. Kozubek, R.K. (2008): *Fictitious domain formulation of unilateral problems: analysis and algorithms*, in preparing.
- J. Haslinger, T. Kozubek, R.K., G. Peichel (2007): *Projected Schur complement method for solving non-symmetric systems arising from a smooth fictitious domain approach*, Num. Lin. Algebra Appl., 14, 713–739.
- Farhat, C., Mandel, J., Roux, F., X. (1994): *Optimal convergence properties of the FETI domain decomposition method*, Comput. Meth. Appl. Mech. Engrg., 115, 365–385.
- R. Glowinski, T. Pan, J. Periaux (1994): *A fictitious domain method for Dirichlet problem and applications*. Comput. Meth. Appl. Mech. Engrg. 111, 283–303.
- R. K. (2005): *Complexity of an algorithm for solving saddle-point systems with singular blocks arising in wavelet-Galerkin discretizations*, Appl. Math. 50, 291–308.
- H. A. Van der Vorst (1992): *BiCGSTAB: a fast and smoothly converging variant of BiCG for solution of nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 13, 631–644.
- M. Benzi, G. H. Golub, J. Liesen(2005): *Numerical solution of saddle point systems*, Acta Numerica, pp. 1–137.