Science & Technology Facilities Council

Technical Report

Discrete fractional Sobolev norms for domain decomposition preconditioning

Mario Arioli, D. Kourounis, and Daniel Loghin

December 23, 2008

RAL-TR-2008-031

Library and Information Services
SFTC Rutherford Appleton Laboratory
Harwell Science and Innovation Campus
Didcot
OX11 0QX
UK
Tel: +44 (0)1235 445384
Fax: +44(0)1235 446403
Email: library@rl.ac.uk

ISSN 1358-6254

RAL-TR-2008-031

Discrete fractional Sobolev norms for domain decomposition preconditioning

M. Arioli[1], D. Kourounis[2] , and D. Loghin[3]

ABSTRACT

Is is well-known that the Steklov-Poincaré operators arising in non-overlapping domain decomposition algorithms are coercive and continuous with respect to Sobolev norms of index 1/2 defined on the skeleton of the domain subdivision. This suggests that discrete fractional Sobolev norms are candidates for interface preconditioning in domain decomposition methods. Our approach is based on recent work by Arioli and Loghin, 2008 which provides a description of finite element projections of fractional Sobolev norms. We show that these norms are optimal preconditioners with respect to the mesh-size for domain decomposition methods for standard elliptic problems. Numerical results indicate that the dependence on the number of domains is logarithmic.

Keywords: Generalized Laplace operators, Fractional Sobolev spaces, Krylov methods, Generalised Lanczos method, Domain decomposition, Matrix square root.

AMS(MOS) subject classifications: 35J15, 35J50, 47G30, 65F10, 65F15, 65F30, 65N30.

Current reports available by anonymous ftp to ftp.numerical.rl.ac.uk in directory pub/reports.

[1] m.arioli@rl.ac.uk, Rutherford Appleton Laboratory.

2 Ioannina University, Ioannina, Greece.

3 d.loghin@bham.ac.ukm.arioli@rl.ac.uk, School of Mathematics
University of Birmingham, UK.

The work was supported by EPSRC grant GR/S42170/01.

Computational Science and Engineering Department
Atlas Centre
Rutherford Appleton Laboratory
Oxon OX11 0QX

December 23, 2008

Contents

1 Introduction

The usefulness of any domain decomposition method (DD) rests on the ability to solve a problem involving a pseudo-differential operator: the Steklov-Poincaré operator. Since under discretisation this gives rise to a system with a dense matrix, for large problems this needs to be solved approximately via a procedure which computes the action of the inverse of the discrete operator on a given vector. To this aim, a great number of iterative approaches have been suggested in the literature; classical algorithms include Dirichlet-Neumann, Neumann-Neumann, FETI methods, Schwarz methods, together with two-level and overlapping variants. For descriptions and analyses see Toselli and Widlund, 2005 and Quarteroni and Valli, 1999.
An alternative that has not been considered to date and which can be shown to be competitive is based on a well-known property of the discrete Steklov-Poincaré operator: it is norm-equivalent to a Sobolev norm-matrix of index 1/2 Quarteroni and Valli, 1999, the discrete representation of which can be written in terms of the square-root of a discrete Laplacian defined on the union of the boundaries of each subdomain Peisker, 1988. This discrete norm has a non-sparse representation; however, since only the action of its inverse on a vector is required, we can achieve this using a standard approach based on a Krylov subspace approximation. The resulting algorithm is a generalised Lanczos procedure and the ensuing preconditioning procedure is independent of the size of the problem.

2 Problem description

We review below the standard formulation of non-overlapping domain decomposition problems for a general scalar elliptic problem.

2.1 Notation and definitions

Throughout the paper we will use the following notation and standard results. Given an open simply-connected domain U in $\mathbb{R}^d$ its boundary will be denoted by ∂U. We denote by $C_0^\infty(U)$ the space of infinitely differentiable functions defined on U with compact support in U. We will also denote by $L^2(U)$ the Lebesgue space of square-integrable functions defined on U endowed with inner-product $(\cdot,\cdot)$ and by $H^m(U)$ the Sobolev space of order m equipped with norm $\|\cdot\|_{m,U}$ and semi-norm $|\cdot|_{m,U}$ with the convention $H^0(U) = L^2(U)$. The Sobolev spaces of real index $0 \leq s \leq m$ are defined as interpolation spaces of index $\theta = 1 - s/m$ for the pair $[H^m(U), L^2(U)]$

$$H^s(\Omega) := [H^m(U), L^2(U)]_\theta \qquad \theta = 1 - s/m.$$

For any s, the space $H_0^s(U)$ denotes the completion of $C_0^\infty(U)$ in $H^s(U)$ (see e.g. Lions and Magenes, 1968, p 60). In particular, we shall be interested in the interpolation space

$$H^{1/2}(U) = [H^1(U), L^2(U)]_{1/2}.$$

for which there holds $H_0^{1/2}(U) \equiv H^{1/2}(U)$. Another space of interest is $H_{00}^{1/2}(U)$ which is a subspace of $H_0^{1/2}(U)$ and is defined as the interpolation space of index 1/2 for the pair $[H_0^1(U), L^2(U)]$

$$H_{00}^{1/2}(U) = [H_0^1(U), H^0(U)]_{1/2}.$$

Norms on $H^{1/2}(U), H_{00}^{1/2}(U)$ will be denoted by the same notation $|\cdot|_{1/2,U}$ or $\|\cdot\|_{1/2,U}$, with the assumption that it is evident from the context which space is under consideration. We will return to the definition of these norms in Section 3. The dual of $H_{00}^{1/2}(U)$ is denoted by

$(H_{00}^{1/2}(U))' \subset H^{-1/2}(U)$ where $H^{-1/2}(U) := (H^{1/2}(U))' \equiv (H_0^{1/2}(U))'$. The duality between $H_{00}^{1/2}(U)$ and its dual will be denoted by $\langle \cdot, \cdot \rangle$.
Finally, we will make use of the trace operator $\gamma_0 : H^1(U) \to H^{1/2}(\partial U)$ which is known to be surjective and continuous, i.e., there exists a constant $c_\gamma(U)$ such that

$$\|\gamma_0 v\|_{1/2,(\partial U)} \leq c_\gamma(U)\|v\|_{1,U} \quad \forall v \in H^1(U). \tag{1}$$

A similar inequality holds if we take $\gamma_0 : H_0^1(U) \to H_{00}^{1/2}(\partial U)$:

$$\|\gamma_0 v\|_{1/2,(\partial U)} \leq c_\gamma(U)\|v\|_{1,U} \quad \forall v \in H_0^1(U). \tag{2}$$

We will also assume that the following Poincaré inequality holds

$$\|v\|_{0,U} \leq C_P(U)|v|_{1,U}. \tag{3}$$

2.2 Domain decomposition for scalar elliptic PDE

Let now Ω denote an open subset of $\mathbb{R}^d$ with boundary $\partial\Omega$ and consider the problem

$$\begin{cases} \mathcal{L}u = -\mathrm{div}(a\nabla u) + \vec{b} \cdot \nabla u + cu &= f \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \partial\Omega. \end{cases} \tag{4}$$

where $f \in L^2(\Omega)$, $c \in L^\infty(\Omega)$, $\vec{b}$ is a vector function whose entries are Lipschitz continuous real-valued functions on $\bar{\Omega}$, and a is a symmetric $d \times d$ matrix whose entries are bounded, piecewise continuous real-valued functions defined on $\bar{\Omega}$, with

$$0 < a_{\min} \leq \zeta^T a(\mathbf{x})\zeta \leq a_{\max} \quad \forall \zeta \in \mathbb{R}^d, \quad \text{a.e. } \mathbf{x} \in \bar{\Omega}. \tag{5}$$

We will also assume the following standard condition holds

$$c - \frac{1}{2}\nabla \cdot \vec{b} \geq c_{\min} \ \text{ a.e. } \mathbf{x} \in \Omega. \tag{6}$$

The weak formulation of problem (4) reads

$$\begin{cases} \text{Find } u \in H_0^1(\Omega) \text{ such that for all } v \in H_0^1(\Omega), \\ B(u, v) = (f, v). \end{cases} \tag{7}$$

where the bilinear form $B(\cdot, \cdot) : H_0^1(\Omega) \times H_0^1(\Omega) \to \mathbb{R}$ is defined via

$$B(v, w) = (a \cdot \nabla v, \nabla w) + \left(\vec{b} \cdot \nabla v + cv, w\right).$$

Let

$$\bar{\Omega} = \bigcup_{i=1}^N \bar{\Omega}_i, \quad \Omega_i \cap \Omega_j \equiv \emptyset \ \ (i \neq j),$$

and let $\Gamma \subset \mathbb{R}^{d-1}$ denote the set of internal boundaries associated with the above partition of Ω

$$\Gamma = \bigcup_{i=1}^N \Gamma_i \quad (\Gamma_i := \partial\Omega_i \setminus \partial\Omega).$$

Given a function v defined on Ω we will denote by v_i the restriction of v to Ω_i: $v_i = v\,|_{\Omega_i}$. With this notation, we define the bilinear forms $B_i(\cdot, \cdot) : H_0^1(\Omega_i) \times H_0^1(\Omega_i) \to \mathbb{R}$ similarly to $B(\cdot, \cdot)$:

$$B_i(v_i, w_i) = (a_i \cdot \nabla v_i, \nabla w_i) + \left(\vec{b}_i \cdot \nabla v_i + c_i v_i, w_i\right).$$

Let now

$$H^1_D(\Omega_i) = \left\{ w \in H^1(\Omega_i) : w \mid_{\partial\Omega \cap \partial\Omega_i} = 0 \right\}$$

and let $v \in H^1_0(\Omega)$. Then $v_i = v \mid_{\Omega_i} \in H^1_D(\Omega_i)$ and there holds

$$B(u,v) = \sum_{i=1}^{N} B_i(u_i, v_i), \quad (f,v) = \sum_{i=1}^{N} (f, v_i). \tag{8}$$

Let u denote the solution of (4) and let $u_i = u \mid_{\Omega_i}$. Assuming the value of the exact solution is known on each Γ_i, say $u_i \mid_{\Gamma_i} = \lambda_i$, problem (4) can be equivalently be written as a set of problems defined on Ω_i for all i

$$\left\{ \begin{array}{rll} \mathcal{L}u_i &= f & \text{in } \Omega_i, \\ u_i &= 0 & \text{on } \partial\Omega \cap \partial\Omega_i, \\ u_i &= \lambda_i & \text{on } \Gamma_i. \end{array} \right. \tag{9}$$

Under the same assumption that λ_i are known, problems (9) can be decoupled into two sets of problems

$$\left\{ \begin{array}{rll} \mathcal{L}u_i^{\{1\}} &= f & \text{in } \Omega_i, \\ u_i^{\{1\}} &= 0 & \text{on } \partial\Omega \cap \partial\Omega_i, \\ u_i^{\{1\}} &= 0 & \text{on } \Gamma_i. \end{array} \right. \quad \left\{ \begin{array}{rll} \mathcal{L}u_i^{\{2\}} &= 0 & \text{in } \Omega_i, \\ u_i^{\{1\}} &= 0 & \text{on } \partial\Omega \cap \partial\Omega_i, \\ u_i^{\{2\}} &= \lambda_i & \text{on } \Gamma_i. \end{array} \right. \tag{10}$$

with the solution $u \mid_{\Omega_i} = u_i = u_i^{\{1\}} + u_i^{\{2\}}$.
To find an equation for λ_i we integrate the two sets of problems in (10) against $v_i \in H^1_D(\Omega_i)$ to get the identities

$$B(u_i^{\{1\}}, v_i) = (f, v_i) + \int_{\Gamma_i} \mathbf{n}_i \cdot a \cdot \nabla u_i^{\{1\}} v_i \mathrm{d}s(\Gamma_i)$$

and

$$B(u_i^{\{2\}}, v_i) = \int_{\Gamma_i} \mathbf{n}_i \cdot a \cdot \nabla u_i^{\{2\}} v_i \mathrm{d}s(\Gamma_i).$$

Adding them up and then summing over i we find (using (8))

$$B(u,v) = (f,v) + \sum_{i=1}^{N} \int_{\Gamma_i} \mathbf{n}_i \cdot a \cdot \nabla u_i^{\{1\}} v_i \mathrm{d}s(\Gamma_i) + \sum_{i=1}^{N} \int_{\Gamma_i} \mathbf{n}_i \cdot a \cdot \nabla u_i^{\{2\}} v_i \mathrm{d}s(\Gamma_i)$$

which yields, using (7), the *Steklov-Poincaré equation* for the decomposition (9)

$$\sum_{i=1}^{N} \int_{\Gamma_i} \mathbf{n}_i \cdot a \cdot \nabla u_i^{\{2\}} v_i \mathrm{d}s(\Gamma_i) = -\sum_{i=1}^{N} \int_{\Gamma_i} \mathbf{n}_i \cdot a \cdot \nabla u_i^{\{1\}} v_i \mathrm{d}s(\Gamma_i).$$

Let now $\eta \in H^{1/2}_{00}(\Gamma)$ with $\eta_i = \eta \mid_{\Gamma_i}$ and let v_i be the solution to the problem

$$\left\{ \begin{array}{rll} \mathcal{L}v_i &= 0 & \text{in } \Omega_i, \\ v_i &= 0 & \text{on } \partial\Omega_i \setminus \Gamma_i, \\ v_i &= \eta_i & \text{on } \Gamma_i. \end{array} \right. \tag{11}$$

Note that in the case of the Poisson problem where $\mathcal{L} = -\Delta$ the functions v_i are harmonic extensions of η_i to Ω_i. In general, one can view v_i as $\mathcal{L}$-extensions of the corresponding data η_i to Ω_i. Furthermore, the function v defined via $v \mid_{\Omega_i} = v_i$ can be viewed as a generalised $\mathcal{L}$-extension of the function $\eta \in H^{1/2}_{00}(\Gamma)$ to the domain Ω. Henceforth, such generalised $\mathcal{L}$-extensions of

functions η or η_i will be denoted by $E\eta$ and $E_i\eta_i$, respectively. Any other extensions will be denoted by $F\eta$ and $F_i\eta_i$.
We will also need the following elliptic regularity result which is known to hold for the weak solution of (11) (see for example Agmon, Douglis and Nirenberg, 1959)

$$\|v_i\|_{1,\Omega_i} = \|E_i\eta_i\|_{1,\Omega_i} \leq C_e\|\eta_i\|_{1/2,\Gamma_i}. \tag{12}$$

We define the *Steklov-Poincaré operator* $\mathcal{S} : H_{00}^{1/2}(\Gamma) \to (H_{00}^{1/2}(\Gamma))'$ as follows. Let $\eta, \mu \in H_{00}^{1/2}(\Gamma)$ with $\eta\,|_{\Gamma_i} =: \eta_i, \mu\,|_{\Gamma_i} =: \mu_i$. We define $\mathcal{S}$ via

$$\langle \mathcal{S}\eta, \mu\rangle = \sum_{i=1}^{N} \int_{\Gamma_i} \mathbf{n}_i \cdot a \cdot \nabla(E\eta_i)\ \mu_i \mathrm{d}s(\Gamma_i) =: \sum_{i=1}^{N} \langle \mathcal{S}_i\eta_i, \mu_i\rangle\,. \tag{13}$$

Using integration by parts the operator $\mathcal{S}$ can be given the following alternative representation

$$\langle \mathcal{S}\eta, \mu\rangle = B(E\eta, F\mu) = \sum_{i=1}^{N} B_i(E_i\eta_i, F_i\mu_i) \quad \forall \eta, \mu \in H_{00}^{1/2}(\Gamma). \tag{14}$$

Note that these definitions amend those given in Quarteroni and Valli, 1999, pp 142–143.
With this definition of $\mathcal{S}$ our model problem can be recast as an ordered sequence of three decoupled sets of problems involving the same operator $\mathcal{L}$ with essential boundary conditions on each subdomain together with a problem set on the interface Γ.

$$\begin{aligned}
&\text{(i)} \begin{cases} \mathcal{L}u_i^{\{1\}} &= f \ \text{ in } \Omega_i, \\ u_i^{\{1\}} &= 0 \ \text{ on } \partial\Omega_i, \end{cases} \\
&\text{(ii)} \begin{cases} \mathcal{S}\lambda = -\displaystyle\sum_{i=1}^{N} \mathbf{n}_i \cdot a \cdot \nabla u_i^{\{1\}} \ \text{ on } \Gamma, \end{cases} \\
&\text{(iii)} \begin{cases} \mathcal{L}u_i^{\{2\}} &= 0 \ \text{ in } \Omega_i, \\ u_i^{\{2\}} &= \lambda_i \ \text{ on } \Gamma_i. \\ u_i^{\{2\}} &= 0 \ \text{ on } \partial\Omega_i \setminus \Gamma_i. \end{cases}
\end{aligned} \tag{15}$$

The resulting solution is

$$u|_{\Omega_i} = u_i^{\{1\}} + u_i^{\{2\}}.$$

We now turn to the properties of the interface operator $\mathcal{S}$. Given representation (13) we can immediately see that $\mathcal{S}$ is non-symmetric unless $\vec{b} = \mathbf{0}$. One can show further that $\mathcal{S}$ is a bounded positive operator on $H_{00}^{1/2}(\Gamma)$.

Lemma 2.1 *Let $\mathcal{S}$ be defined by* (13) *and let* (6) *hold. Then there exist constants α_1, α_2 such that for all $\eta, \mu \in H_{00}^{1/2}(\Gamma)$*

$$\alpha_1\|\eta\|_{1/2,\Gamma}^2 \leq \langle S\eta, \eta\rangle\,, \qquad \langle S\eta, \mu\rangle \leq \alpha_2\|\eta\|_{1/2,\Gamma}\|\mu\|_{1/2,\Gamma}.$$

Proof: Let $v_i = E_i\eta_i, w_i = E_i\mu_i$ satisfy (11). We have, using (6),

$$\begin{aligned}
\langle \mathcal{S}_i\eta_i, \eta_i\rangle &= B_i(v_i, v_i) \\
&= (a \cdot \nabla v_i, v_i) + \left(\vec{b} \cdot \nabla v_i, v_i\right) + (cv_i, v_i) \\
&= (a \cdot \nabla v_i, v_i) + \left((c - \frac{1}{2}\nabla \cdot \vec{b})v_i, v_i\right) \\
&\geq a_{\min}|v_i|_{1,\Omega_i}^2 + c_{\min}\|v_i\|_{0,\Omega_i}^2 \\
&\geq \min\{a_{\min}, c_{\min}\}\,\|v_i\|_{1,\Omega_i}^2
\end{aligned}$$

Moreover, using the Poincaré inequality (3) we get

$$\begin{aligned}\langle \mathcal{S}\rangle\eta_i,\mu_i\rangle &= B_i(v_i,w_i)\\ &\le a_{\max}|v_i|_{1,\Omega_i}|w_i|_{1,\Omega_i}+\|\vec{b}\|_{L^\infty(\Omega_i)}|v_i|_{1,\Omega_i}\|w_i\|_{0,\Omega_i}+\|c\|_{L^\infty(\Omega_i)}\|v_i\|_{0,\Omega_i}\|w_i\|_{0,\Omega_i}\\ &\le \max\left\{a_{\max}+\|\vec{b}\|_{L^\infty(\Omega_i)}C_P(\Omega_i),\|c\|_{L^\infty(\Omega_i)}\right\}\|v_i\|_{1,\Omega_i}\|w_i\|_{1,\Omega_i}.\end{aligned}$$

Since $\gamma_0 v_i=\eta_i,\gamma_0 w_i=\mu_i$, the trace inequalities (1,2) read for all $i=1,\dots,N$

$$\|\eta_i\|_{1/2,\Gamma_i}\le C_\gamma(\Omega_i)\|v_i\|_{1,\Omega_i},\qquad \|\mu_i\|_{1/2,\Gamma_i}\le C_\gamma(\Omega_i)\|w_i\|_{1,\Omega_i}$$

and the result follows from the regularity estimate (12) and the definition (13) of the operator $\mathcal{S}$. ■

3 Finite element discretisations

In order to write down the weak formulation of problems (15) we re-write the set of equations (15,(iii)) as

$$\text{(iii)}\begin{cases}\mathcal{L}\tilde{u}_i^{\{2\}} &= -\mathcal{L}z_i \quad \text{in } \Omega_i,\\ \tilde{u}_i^{\{2\}} &= 0 \qquad \text{on } \Gamma_i.\\ \tilde{u}_i^{\{2\}} &= 0 \qquad \text{on } \partial\Omega_i\setminus\Gamma_i.\end{cases}$$

where $\tilde{u}_i^{\{2\}}=u_i^{\{2\}}-z$. With this new notation, the weak formulations of problems (15) are

$$\begin{aligned}&\text{(i)}\begin{cases}\text{Find } u_i^{\{1\}}\in H_0^1(\Omega_i) \text{ such that for all } v_i\in H_0^1(\Omega_i),\\ B_i(u_i^{\{1\}},v_i)=(f_i,v_i).\end{cases}\\ &\text{(ii)}\begin{cases}\text{Find } \lambda\in H_{00}^{1/2}(\Gamma) \text{ such that for all } \eta\in H_{00}^{1/2}(\Gamma),\\ s(\lambda,\eta):=\langle\mathcal{S}\lambda,\eta\rangle=\sum_{i=1}^N\left[(f_i,F_i\eta_i)-B_i(u_i^{\{1\}},F_i\eta_i)\right].\end{cases}\\ &\text{(iii)}\begin{cases}\text{Find } \tilde{u}_i^{\{2\}}=u_i^{\{2\}}-z_i\in H_0^1(\Omega_i) \text{ such that for all } v_i\in H_0^1(\Omega_i),\\ B_i(\tilde{u}_i^{\{2\}},v_i)=-B_i(z_i,v_i).\end{cases}\end{aligned}\tag{16}$$

Note that $z_i\in H^1(\Omega_i)$ and $z\in H^1(\Omega)$ with $\gamma_0(\Gamma_i)z_i=\lambda_i$.
Let $P_r(\mathfrak{t})$ denote the space of polynomials in d variables of degree r defined on a set $\mathfrak{t}\subset\mathbb{R}^d$. Let

$$V_i^h=V_i^{h,r}:=\left\{w\in C^0(\Omega_i): w|_{\mathfrak{t}}\in P_k\ \ \forall\mathfrak{t}\in\mathfrak{T}_h,\ w\,|_{\partial\Omega\cap\partial\Omega_i}=0\right\}\subset H_D^1(\Omega_i)\tag{17}$$

be a finite-dimensional space of piecewise polynomial functions defined on some subdivision $\mathfrak{T}_h$ of Ω into simplices $\mathfrak{t}$ of maximum diameter h. Let further $V_{iI}^h,V_{iB}^h\subset V_i^h$ satisfy $V_{iI}^h\oplus V_{iB}^h\equiv V_i^h$. Let also

$$V_{iI}^h=\text{span}\left\{\phi_k^i,k=1\dots n_i^I\right\}$$

and

$$V_{iB}^h=\text{span}\left\{\psi_k^i,k=1\dots n_i^B\right\}$$

and set $n_I=\sum_i n_i^I$. Let further

$$V_B^h=\bigcup_{i=1}^N V_{iB}^h$$

and let $\{\psi_k, k = 1, \dots n_B\}$ denote a basis for V_B^h. Let $S_i^h = \text{span}\left\{\gamma_0(\Gamma_i)\psi_k, k = 1 \dots n_i^B\right\}$ with $S^h = \cup_{i=1}^N S_i^h$. Finally, let

$$V^h = \bigcup_{i=1}^N V_i^h \subset H_0^1(\Omega).$$

The finite element discretisation of the weak formulation (7) reads

$$\left\{ \begin{array}{l} \text{Find } u_h \in V^h \text{ such that for all } v_h \in V^h, \\ B(u_h, v_h) = (f, v_h). \end{array} \right. \tag{18}$$

The finite element discretisations of the weak formulations (16) are as follows.

$$\begin{array}{ll} \text{(i)} & \left\{ \begin{array}{l} \text{Find } u_{hi}^{\{1\}} \in V_{iI}^h \text{ such that for all } v_{hi} \in V_{iI}^h, \\ B_i(u_{hi}^{\{1\}}, v_{hi}) = (f_i, v_{hi}). \end{array} \right. \\ \text{(ii)} & \left\{ \begin{array}{l} \text{Find } \lambda_h \in S^h \text{ such that for all } \eta_h \in S^h, \\ s(\lambda_h, \eta_h) = \sum_{i=1}^N \left[(f_i, F_i\eta_{hi}) - B_i(u_{hi}^{\{1\}}, F_i\eta_{hi})) \right]. \end{array} \right. \\ \text{(iii)} & \left\{ \begin{array}{l} \text{Find } \tilde{u}_{hi}^{\{2\}} = u_{hi}^{\{2\}} - z_{hi} \in V_{iI}^h \text{ such that for all } v_{hi} \in V_{iI}^h, \\ B_i(\tilde{u}_{hi}^{\{2\}}, v_{hi}) = -B_i(z_{hi}, v_{hi}). \end{array} \right. \end{array} \tag{19}$$

3.1 Matrix formulation

Formulation (19) amounts to a Schur complement approach for the solution of the discrete weak formulation (18). For completeness of exposition we include this characterisation below.
Let

$$A\mathbf{u} = \begin{pmatrix} A_{II} & A_{IB} \\ A_{BI} & A_{BB} \end{pmatrix} \begin{pmatrix} \mathbf{u}_I \\ \mathbf{u}_B \end{pmatrix} = \begin{pmatrix} \mathbf{f}_I \\ \mathbf{f}_B \end{pmatrix} = \mathbf{f} \tag{20}$$

represent the linear system associated with the discrete formulation (18) with $A \in \mathbb{R}^{n\times n}, \mathbf{u} \in \mathbb{R}^n$ where $n = n_I + n_B$ and $A_{II} \in \mathbb{R}^{n_I \times n_I}, A_{IB}, A_{BI}^T \in \mathbb{R}^{n_I \times n_B}, A_{BB} \in \mathbb{R}^{n_B \times n_B}$ are given by

$$A_{II} = \begin{pmatrix} A_{II}^1 & & & & \\ & \ddots & & & \\ & & A_{II}^i & & \\ & & & \ddots & \\ & & & & A_{II}^N \end{pmatrix}, \quad A_{IB} = \begin{pmatrix} A_{IB}^1 \\ \vdots \\ A_{IB}^i \\ \ddots \\ A_{IB}^N \end{pmatrix}, \quad A_{IB} = \begin{pmatrix} A_{BI}^1 & \cdots & A_{BI}^i & \cdots & A_{BI}^N \end{pmatrix}$$

with

$$\begin{array}{rcl} (A_{II}^i)_{kk} & = & B_i(\phi_k^i, \phi_k^i), \\ (A_{IB}^i)_{kj} & = & B_i(\phi_k^i, \psi_j^i), \\ (A_{BI}^i)_{jk} & = & B_i(\psi_l^i, \phi_l^i), \\ (A_{BB})_{ll} & = & B(\psi_l, \psi_l), \end{array}$$

for all $k = 1, \dots, n_i^I, j = 1, \dots, n_i^B, l = 1, \dots, n_B$.

Lemma 3.1 *With the above notation, the solution of problems* (11) *has finite element coefficients*

$$\mathbf{v} = \begin{pmatrix} \mathbf{v}_I \\ \mathbf{v}_B \end{pmatrix} = \begin{pmatrix} -A_{II}^{-1} A_{IB} \mathbf{v}_B \\ \mathbf{v}_B \end{pmatrix},$$

where $\mathbf{v}_B$ *are the coefficients of* η *with respect to the basis of* V_B^h.

Proof: We start by considering the weak formulation of problems (11). If we let $w_i \mid_{\Gamma_i} = \eta_i$ we can re-write our problems as

$$\begin{cases} \mathcal{L}\tilde{v}_i &= -\mathcal{L}z_i \quad \text{in } \Omega_i, \\ \tilde{v}_i &= 0 \quad \text{on } \partial\Omega_i \setminus \Gamma_i, \\ \tilde{v}_i &= 0 \quad \text{on } \Gamma_i. \end{cases} \tag{21}$$

where we set $\tilde{v}_i = v_i - z_i$. This problem has the following discrete weak formulation

$$\begin{cases} \text{Find } \tilde{v}_{hi} = v_{hi} - z_{hi} \in V_{iI}^h \text{ such that for all } w_{hi} \in V_{iI}^h, \\ B_i(\tilde{v}_{hi}, w_{hi}) = -B_i(z_{hi}, w_{hi}). \end{cases}$$

so that the global matrix representation resulting after summing over i is

$$\begin{pmatrix} \mathbf{w}_I^T & 0 \end{pmatrix} \begin{pmatrix} A_{II} & A_{IB} \\ A_{BI} & A_{BB} \end{pmatrix} \begin{pmatrix} \mathbf{v}_I - \mathbf{z}_I \\ 0 \end{pmatrix} = - \begin{pmatrix} \mathbf{w}_I^T & 0 \end{pmatrix} \begin{pmatrix} A_{II} & A_{IB} \\ A_{BI} & A_{BB} \end{pmatrix} \begin{pmatrix} \mathbf{z}_I \\ \mathbf{v}_B \end{pmatrix}$$

and the statement of the lemma follows. ∎

Consider now the matrix representation of operator $\mathcal{S}$ in the basis $\{\psi_k, k = 1, \ldots, n_B\}$. The discrete form of definition (14) is

$$s(\eta_h, \mu_h) = B_i(E_i \eta_{ih}, F_i \mu_{ih}) \tag{22}$$

where we recall that F_i is an arbitrary extension operator to Ω_i while E_i is an $\mathcal{L}$-extension to the same domain. Setting $v \mid_{\Gamma_i} = \eta_{ih}, w \mid_{\Gamma_i} = \mu_{ih}$, the corresponding discrete representations of these extensions will have the form (using the Lemma 3.1)

$$E_i \eta_{ih} = \begin{pmatrix} -A_{II}^{-1} A_{IB} \mathbf{v}_B \\ \mathbf{v}_B \end{pmatrix}, F_i \mu_{ih} = \begin{pmatrix} \mathbf{w}_I \\ \mathbf{w}_B \end{pmatrix}.$$

Hence (22) has the representation

$$\mathbf{w}_B^T S \mathbf{v}_B = \begin{pmatrix} \mathbf{w}_I^T & \mathbf{w}_B^T \end{pmatrix} \begin{pmatrix} A_{II} & A_{IB} \\ A_{BI} & A_{BB} \end{pmatrix} \begin{pmatrix} -A_{II}^{-1} A_{IB} \mathbf{v}_B \\ \mathbf{v}_B \end{pmatrix}$$

which is equivalent to

$$\mathbf{w}_B^T S \mathbf{v}_B = \mathbf{w}_B^T (A_{BB} - A_{BI} A_{II}^{-1} A_{IB}) \mathbf{v}_B$$

for all $\mathbf{v}_B, \mathbf{w}_B \in \mathbb{R}^{n_B \times n_B}$ so that S is the Schur complement of A_{BB} in the global matrix A. With this notation, (19 (i)) has the following matrix formulation

$$\text{(i) } A_{II} \mathbf{u}_I^{\{1\}} = \mathbf{f}_I,$$

while (19 (ii)) becomes

$$\text{(ii) } \mathbf{w}_B^T S \mathbf{u}_B = \begin{pmatrix} \mathbf{w}_I^T & \mathbf{w}_B^T \end{pmatrix} \begin{pmatrix} \mathbf{f}_I \\ \mathbf{f}_B \end{pmatrix} - \begin{pmatrix} \mathbf{w}_I^T & \mathbf{w}_B^T \end{pmatrix} \begin{pmatrix} A_{II} & A_{IB} \\ A_{BI} & A_{BB} \end{pmatrix} \begin{pmatrix} \mathbf{u}_I^{\{1\}} \\ 0 \end{pmatrix},$$

for any $\mathbf{w}_i \in \mathbb{R}^{n_I}, \mathbf{w}_B \in \mathbb{R}^{n_B}$; this equation simplifies to

$$\text{(ii) } S\mathbf{u}_B = \mathbf{f}_B - A_{BI}\mathbf{u}_I^{\{1\}}.$$

Finally, using Lemma 3.1, the discrete form of (19 (iii)) is seen to be

$$\text{(iii) } \mathbf{u}_I^{\{2\}} = -A_{II}^{-1}A_{IB}\mathbf{u}_B.$$

These equations are easily seen to represent a Schur complement approach for the original linear system (20) with global solution $\mathbf{u}$ given by

$$\mathbf{u} = \begin{pmatrix} \mathbf{u}_I \\ \mathbf{u}_B \end{pmatrix} = \begin{pmatrix} \mathbf{u}_I^{\{1\}} \\ \mathbf{u}_B^{\{1\}} \end{pmatrix} + \begin{pmatrix} \mathbf{u}_I^{\{1\}} \\ \mathbf{u}_B^{\{1\}} \end{pmatrix}.$$

4 Preconditioners for the Steklov-Poincaré operator

The result of Lemma 2.1 holds also in the discrete case for the choice of space S_h introduced in the previous section. In particular it translates into the following coercivity and continuity bounds for $s(\cdot,\cdot) : S_h \times S_h$.

Lemma 4.1 *Let $s(\cdot,\cdot)$ be defined as in* (16) *and let* (6) *hold. Then there exist constants α_1, α_2 such that for all $\eta_h, \mu_h \in S_h \subset H_{00}^{1/2}(\Gamma)$*

$$\alpha_1\|\eta_h\|_{1/2,\Gamma}^2 \leq s(\eta_h,\eta_h), \qquad s(\eta_h,\mu_h) \leq \alpha_2\|\eta_h\|_{1/2,\Gamma}\|\mu_h\|_{1/2,\Gamma}.$$

In order to derive the corresponding matrix formulation of the above result we need to recall the results in Arioli and Loghin, 2008.

4.1 Discrete fractional Sobolev norms

Let X, Y denote two Hilbert spaces with $X \subset Y$, X dense and continuously embedded in Y. Let $\langle\cdot,\cdot\rangle_X, \langle\cdot,\cdot\rangle_Y$ denote the corresponding inner products, and $\|\cdot\|_X, \|\cdot\|_Y$ the respective norms. By the Riesz representation theory (see for example Riesz and Sz-Nagy, 1956) there exists an operator $\mathcal{J} : X \to Y$ which is positive and self-adjoint with respect to $\langle\cdot,\cdot\rangle_Y$ such that

$$\langle u, v\rangle_X = \langle u, \mathcal{J}v\rangle_Y . \tag{23}$$

Using the spectral decomposition of $\mathcal{J}$ we define the operator $\mathcal{E} = \mathcal{J}^{1/2} : X \to Y$, which in turn is positive self-adjoint. Moreover, the spectral decomposition of $\mathcal{E}$ can be used to define any real power of $\mathcal{E}$. Let $\theta \in [0,1]$ and let $\|\cdot\|_\theta$ denote the scale of graph norms

$$\|u\|_\theta := \left(\|u\|_Y^2 + \|\mathcal{E}^{1-\theta}u\|_Y^2\right)^{1/2}. \tag{24}$$

One can then show that the domain of $\mathcal{E}^{1-\theta}$ endowed with the inner-product

$$\langle u, v\rangle_\theta = \langle u, v\rangle_Y + \left\langle \mathcal{E}^{1-\theta}u, \mathcal{E}^{1-\theta}v \right\rangle_Y$$

is a Hilbert space (Lions and Magenes, 1968). This is an *interpolation space of index θ for the pair* $[X, Y]$ and is denoted by $[X, Y]_\theta$

$$[X, Y]_\theta := D(\mathcal{E}^{1-\theta}), \qquad 0 \leq \theta \leq 1.$$

Let now $X_h \subset X, Y_h \subset Y$ denote two finite-dimensional subspaces of X, Y respectively, with $n = \dim X_h = \dim Y_h$. They are Hilbert spaces when endowed with the inner-products $\langle \cdot, \cdot \rangle_X, \langle \cdot, \cdot \rangle_Y$. We can similarly define corresponding positive, self-adjoint operators $\mathcal{J}_h, \mathcal{E}_h : X_h \to Y_h$

$$\langle u_h, v_h \rangle_X = \langle u_h, \mathcal{J}_h v_h \rangle_Y \qquad u_h, v_h \in X_h \tag{25}$$

where $\mathcal{J}_h$ is positive self-adjoint and $\mathcal{E}_h = \mathcal{J}_h^{1/2}$. We define the discrete interpolation spaces

$$[X_h, Y_h]_\theta := D(\mathcal{E}_h^{1-\theta}).$$

Furthermore, we define the scale of discrete norms

$$\|u_h\|_{\theta,h} := \left(\|u_h\|_Y^2 + \|\mathcal{E}_h^{1-\theta} u_h\|_Y^2 \right)^{1/2}. \tag{26}$$

The following result can be found in Arioli and Loghin, 2008.

Lemma 4.2 *Let $X_h \subset Y_h, X \subset Y$ be Hilbert spaces with inner-products $\langle \cdot, \cdot \rangle_X, \langle \cdot, \cdot \rangle_Y$ and let $\| \cdot \|_\theta, \| \cdot \|_{\theta,h}$ be defined by* (24), (26)*, respectively. Let us assume that there exists an operator I_h such that $I_h : \mathcal{L}(X; X_h) \cap \mathcal{L}(Y; Y_h)$ and $I_h u = u_h$ for all $u_h \in X_h$. Then the norms $\| \cdot \|_\theta, \| \cdot \|_{\theta,h}$ are equivalent on $[X_h, Y_h]_\theta$ for all $\theta \in (0, 1)$.*

In the following we assume that Γ is the union of planar (straight) faces (segments) Γ_i. We let Y be the space of square-integrable functions defined on Γ and let ∇_Γ denote the tangential gradient of a scalar function $v(\mathbf{x}) : \Omega$

$$\nabla_\Gamma v(\mathbf{x}) := \nabla v(\mathbf{x}) - \mathbf{n}(\mathbf{n} \cdot \nabla v(\mathbf{x})) \tag{27}$$

i.e., the projection of the gradient of v onto the plane tangent to Γ at $\mathbf{x} \in \Gamma$. Let the Sobolev space of index 1 be defined as

$$H^1(\Gamma) := \left\{ v \in L^2(\Gamma) : \int_\Gamma |\nabla_\Gamma v|^2 \, \mathrm{d}s(\Gamma) < \infty, \right\}.$$

Let $\partial\Gamma \subset \partial\Omega$ denote the set of points (faces) on the boundary of our domain which represent the skeleton boundary. We define X as follows

$$X = H_0^1(\Gamma) := \left\{ v \in H_0^1(\Gamma) : v \mid_{\partial\Gamma} = 0 \right\}.$$

We endow X with the norm

$$|v|_{H_0^1(\Gamma)} = \|\nabla_\Gamma v\|_{L^2(\Gamma)}.$$

Let now $X_h = (S_h, \| \cdot \|_{H_0^1(\Gamma)}) \subset X, Y_h = (S_h, \| \cdot \|_{L^2(\Gamma)}) \subset Y$. A norm for the discrete interpolation space $[X_h, Y_h]_{1/2}$ has the matrix representation Arioli and Loghin, 2008

$$H_{1/2} = M + M(M^{-1}L)^{1/2}, \tag{28}$$

which can be shown to be spectrally equivalent to (Arioli and Loghin 2008)

$$\widehat{H}_{1/2} = M(M^{-1}L)^{1/2}, \tag{29}$$

where

$$M_{ij} = (\psi_i, \psi_j)_{L^2(\Gamma)}, \quad L_{ij} = (\nabla_\Gamma \psi_i, \nabla_\Gamma \psi_j)_{H_0^1(\Gamma)}$$

where $\psi_i \in S^h$ for $i = 1, \ldots, n_B$. With this notation, Lemma 4.2 applies with I_h the finite element projector onto S^h. Thus, for all $\lambda_h \in S^h$ with $\lambda_h = \sum_{i=1}^{n_B} \boldsymbol{\lambda}_i \psi_i$ there exist constants κ_1, κ_2 such that

$$\kappa_1 \|\lambda_h\|_{1/2,\Gamma} \leq \|\boldsymbol{\lambda}\|_{H_{1/2}} \leq \kappa_2 \|\lambda_h\|_{1/2,\Gamma}.$$

We immediately derive the following result.

Proposition 4.3 *Let $s(\cdot,\cdot)$ be defined as in* (16) *and let* (6) *hold. Let $\boldsymbol{\beta},\boldsymbol{\mu}$ denote the coefficients of η_h,μ_h with respect to the basis $\{\psi_i, i=1,\ldots,n_B\}$ of S^h. Let S denote the matrix representation of $s(\cdot,\cdot)$ with respect to the same basis. Then there exist constants $\tilde{\alpha}_1,\tilde{\alpha}_2,\widehat{\alpha}_1,\widehat{\alpha}_2$ such that*

$$\tilde{\alpha}_1\|\boldsymbol{\beta}\|^2_{H_{1/2}} \leq \boldsymbol{\beta}^T S\boldsymbol{\beta}, \qquad \boldsymbol{\mu}^T S\boldsymbol{\beta} \leq \tilde{\alpha}_2\|\boldsymbol{\beta}\|_{H_{1/2}}\|\boldsymbol{\mu}\|_{H_{1/2}}$$

and

$$\widehat{\alpha}_1\|\boldsymbol{\beta}\|^2_{\widehat{H}_{1/2}} \leq \boldsymbol{\beta}^T S\boldsymbol{\beta}, \qquad \boldsymbol{\mu}^T S\boldsymbol{\beta} \leq \widehat{\alpha}_2\|\boldsymbol{\beta}\|_{\widehat{H}_{1/2}}\|\boldsymbol{\mu}\|_{\widehat{H}_{1/2}}$$

for all $\eta_h,\mu_h \in S_h \subset H^{1/2}_{00}(\Gamma)$.

We will see below that the above equivalence indicates that the norm-matrices $H_{1/2},\widehat{H}_{1/2}$ can be used as preconditioners for the Schur complement and that they are optimal in some sense to be described.

4.2 Mesh-independent preconditioners

The solution of linear system (20) requires an iterative approach in the case of large scale problems. A useful approach is to consider an iterative solver such as GMRES together with a suitable preconditioning strategy. In our case, one could for example employ a right preconditioner which will incorporate the solution of problems posed on the interior of each domain (achieved in parallel) and the (approximate) solution of a problem involving the discrete Steklov-Poincaré operator. Given the equivalence in Proposition 4.3, a candidate as right preconditioner can be taken to be

$$P_R = \begin{pmatrix} A_{II} & A_{IB} \\ 0 & H_{1/2} \end{pmatrix}.$$

With this choice, the preconditioned system is

$$AP_R^{-1} = \begin{pmatrix} I & 0 \\ A_{BI}A_{II}^{-1} & SH_{1/2}^{-1} \end{pmatrix}.$$

This block structure indicates that the convergence of an iterative algorithm such as GMRES will depend on the ability of $H_{1/2}$ to approximate S. In particular, the eigenvalues of the above preconditioned matrix are either equal to one or coincide with one of the eigenvalues of $SH_{1/2}^{-1}$. Note that the spectrum of S depends on n_B and also on the subdomain decomposition: number of subdomains, partitioning configuration, subdomain regularity etc. We show below that the eigenvalues of $SH_{1/2}^{-1}$ lie in a region of the complex plane which is independent of the size of the problem n_B and which also lies in the right half-plane. We will investigate numerically the dependence on the type of decomposition employed.
We start by recalling the definition of the H-field of values of a matrix A, given a symmetric and positive-definite matrix H.

Definition 1 *Let $R,H \in \mathbb{R}^{n\times n}$, with H symmetric and positive definite. The H-field of values of the matrix R, denoted by $\mathcal{W}_H(R)$, is a set in the complex plane given by*

$$\mathcal{W}_H(R) = \left\{ z \in \mathbb{C} : z = \frac{\mathbf{x}^* HR\mathbf{x}}{\mathbf{x}^* H\mathbf{x}} = \frac{\langle \mathbf{x}, R\mathbf{x}\rangle_H}{\langle \mathbf{x},\mathbf{x}\rangle_H}, \quad \mathbf{x}\in\mathbb{C}^n \setminus \{0\} \right\}.$$

When $H=I$, the set is called the field of values and is denoted by $\mathcal{W}(R)$.

We also need to recall a related result concerning the convergence of GMRES (see Elman, 1982 and Saad, 2003).

Lemma 4.4 *Let $H \in \mathbb{R}^{n\times n}$ be a positive-definite matrix. Let $R, P \in \mathbb{R}^{n\times n}$ be nonsingular matrices such that the following bounds hold:*

$$\xi_1 \leq \frac{\left\langle \mathbf{x}, RP^{-1}\mathbf{x}\right\rangle_H}{\langle \mathbf{x},\mathbf{x}\rangle_H}, \quad \frac{\|RP^{-1}\mathbf{x}\|_H}{\|\mathbf{x}\|_H} \leq \xi_2 \tag{30}$$

for some positive constants ξ_1 and ξ_2. Then the GMRES algorithm in the H-inner product yields a residual $\mathbf{r}^k$ after k iterations which satisfies

$$\frac{\|\mathbf{r}^k\|_H}{\|\mathbf{r}^0\|_H} \leq \left(1 - \frac{\xi_1^2}{\xi_2^2}\right)^{k/2}. \tag{31}$$

The following result provides bounds on the $H_{1/2}^{-1}$-field of values.

Proposition 4.5 *Let the hypothesis of Proposition* 4.3 *hold. Then the $H_{1/2}^{-1}$-field of values of $SH_{1/2}^{-1}$ is in the right half-plane and is bounded independently of n_B.*

Proof: The projection on the real line of the $H_{1/2}^{-1}$-field of values is bounded from below by

$$\min_{z\in\mathcal{W}_{H_{1/2}}(SH_{1/2}^{-1})} |z| = \min_{\boldsymbol{\beta}\in\mathbb{R}^{n_B}\setminus\{\mathbf{0}\}} \frac{\left\langle \boldsymbol{\beta}, SH_{1/2}^{-1}\boldsymbol{\beta}\right\rangle_{H_{1/2}^{-1}}}{\langle \boldsymbol{\beta},\boldsymbol{\beta}\rangle_{H_{1/2}^{-1}}} = \min_{\boldsymbol{\beta}\in\mathbb{R}^{n_B}\setminus\{\mathbf{0}\}} \frac{\boldsymbol{\beta}^T S\boldsymbol{\beta}}{\boldsymbol{\beta}^T H_{1/2}\boldsymbol{\beta}} \geq \tilde{\alpha}_1 > 0.$$

An upper bound for the field of values is provided by the numerical radius which in turn is bounded by the maximum $H_{1/2}^{-1}$-singular value. The resulting bound on the $H_{1/2}^{-1}$-field of values of $SH_{1/2}^{-1}$ is

$$|z| \leq \max_{\boldsymbol{\beta}\in\mathbb{R}^{n_B}\setminus\{\mathbf{0}\}} \frac{\|S\boldsymbol{\beta}\|_{H_{1/2}^{-1}}}{\|\boldsymbol{\beta}\|_{H_{1/2}}} = \max_{\boldsymbol{\beta}\in\mathbb{R}^{n_B}\setminus\{\mathbf{0}\}} \max_{\boldsymbol{\mu}\in\mathbb{R}^{n_B}\setminus\{\mathbf{0}\}} \frac{\boldsymbol{\beta}S\boldsymbol{\mu}}{\|\boldsymbol{\beta}\|_{H_{1/2}}\|\boldsymbol{\mu}\|_{H_{1/2}}} \leq \tilde{\alpha}_2.$$

■

Note that the above bounds also imply the following bounds independent of n_B on the eigenvalues of the preconditioned discrete Steklov-Poincaré operator:

$$\tilde{\alpha}_1 \leq \left|\lambda(SH_{1/2}^{-1})\right| \leq \tilde{\alpha}_2.$$

Given the result of Proposition 4.3, a convergence bound can be immediately derived for a system of equations with $SH_{1/2}^{-1}$ as a coefficient matrix.

Proposition 4.6 *Let the hypothesis of Proposition* 4.3 *hold. Then GMRES algorithm applied to the linear system*

$$SH_{1/2}^{-1}\tilde{\mathbf{y}} = \mathbf{z}, \qquad (\tilde{\mathbf{y}} = H_{1/2}\mathbf{y})$$

in the $H_{1/2}^{-1}$-inner product yields a residual $\mathbf{r}^k$ after k iterations which satisfies

$$\frac{\|\mathbf{r}^k\|_{H_{1/2}}}{\|\mathbf{r}^0\|_{H_{1/2}}} \leq \left(1 - \frac{\tilde{\alpha}_1^2}{\tilde{\alpha}_2^2}\right)^{k/2}. \tag{32}$$

The following result is adapted from Loghin and Wathen, 2004, Thm 3.7

Proposition 4.7 *Let the hypothesis of Proposition* 4.3 *hold and let* P_R *be given by*

$$P_R = \begin{pmatrix} A_{II} & A_{IB} \\ 0 & \rho H_{1/2} \end{pmatrix}.$$

Then there exists $\rho_0 > 0$ *such that for all* $\rho > \rho_0$ *conditions* (30) *hold with* R, P *replaced by* A, P_R *and for the choice*

$$H = \begin{pmatrix} A_{II} & 0 \\ 0 & H_{1/2} \end{pmatrix}^{-1}.$$

As before, this result indicates that block triangular preconditioners $P_R(\rho)$ are optimal preconditioners when we use a suitable GMRES iteration to obtain the solution of the *global linear system*.

Remark 4.1 *All results in this section apply with* $H_{1/2}$ *replaced by* $\widehat{H}_{1/2}$ *as defined in* (29).

4.3 Algorithms for the matrix square-root

The discrete fractional Sobolev norms introduced in the previous section require the evaluation of the square root function of a matrix. According to application this task can be achieved in different ways. Direct approaches are based on a generalised eigenvalue decomposition which is known to have complexity $O(n^3)$; for problems with structure we can also employ fast algorithms such as the FFT (Peisker 1988) with reduced complexity ($O(n \log n)$). The alternative is to use iterative techniques. Standard iterations, such as Newton's method, may not have better complexity than a direct method. Hale, Higham and Trefethen, 2008 could claim reduced complexity . In our implementation we used a Krylov subspace approximation which takes advantage of the sparsity properties of the matrices involved in the definition of our discrete fractional Sobolev norms. In particular, we employed a generalised Lanczos algorithm which we describe below together with some related approximations required inside a preconditioning procedure.

4.4 A generalised Lanczos algorithm

Given a pair of symmetric and positive-definite matrices (M, L), the generalised Lanczos algorithm constructs a set of M-orthogonal vectors $\mathbf{v}_i$ such that

$$LV_k = MV_kT_k + \beta_{k+1}M\mathbf{v}_{k+1}\mathbf{e}_k^T, \quad V_k^TMV_k = I_k$$

where the columns $\mathbf{v}_i$ of $V_k = [\mathbf{v}_1, \mathbf{v}_2 \ldots, \mathbf{v}_k]$ are known as the Lanczos vectors and $I_k \in \mathbb{R}^{k\times k}$ is the identity matrix with kth column denoted by $\mathbf{e}_k$, while the matrix $T_k \in \mathbb{R}^{k\times k}$ is a symmetric and tridiagonal matrix (Parlett 1998). The standard algorithm corresponds to the case $M = I$. Note that T_k can be seen as a projection of L onto the space spanned by the M-orthogonal columns of V_k

$$V_k^TLV_k = T_k, \quad V_k^TMV_k = I_k. \tag{33}$$

In exact arithmetic, when $k = n$, the algorithm can be seen as providing simultaneous factorisations of the matrix pair (M, L) as

$$L = V_n^{-T}T_nV_n^{-1}, \quad M = V_n^{-T}V_n^{-1}.$$

We recall the algorithm below (Parlett 1998).

Algorithm 1. Generalised Lanczos Algorithm

Input: $L, M \in \mathbb{R}^{n\times n}(\text{spd}), \mathbf{v} \in \mathbb{R}^n$
Output: $V_k \in \mathbb{R}^{n\times k}, T_k \in \mathbb{R}^{k\times k}$

Set $\beta_1 = 0, \mathbf{v}_0 = 0, \mathbf{v}_1 = \mathbf{v}/\|\mathbf{v}\|_M$
for $i = 1 : k$
  $\mathbf{w}_i = M^{-1}L\mathbf{v}_i - \beta_i\mathbf{v}_{i-1}$
  $\alpha_i = (\mathbf{w}_i, \mathbf{v}_i)_M$
  $\mathbf{w}_i = \mathbf{w}_i - \alpha_i\mathbf{v}_i$
  $\beta_{i+1} = \|\mathbf{w}_i\|_M$
  if $\beta_{i+1} = 0$ stop
  $\mathbf{v}_{i+1} = \mathbf{w}_i/\beta_{i+1}$
end
$T_k = \text{tridiag}[\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\beta}]$

The explicit form of T_k is given below

$$T_k = \text{tridiag}[\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\beta}] = \begin{pmatrix} \alpha_1 & \beta_2 & & 0 \\ \beta_2 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_k \\ 0 & & \beta_k & \alpha_k \end{pmatrix}.$$

Consider now the generalised Lanczos factorisation for the same matrix-pair (M, L) written as:

$$LV = MVT, \quad V^T MV = I. \tag{34}$$

where we used the notation $V = V_n, T = T_n$. We can immediately derive the following result.

Lemma 4.8 *Let* (34) *hold and let* $H_{1/2} = M + M(M^{-1}L)^{1/2}$ *and* $\widehat{H}_{1/2} = M(M^{-1}L)^{1/2}$. *Then*

$$H_{1/2} = \left(V(I + T^{1/2})^{-1}V^T\right)^{-1} = MV(I + T^{1/2})V^T M \tag{35}$$

and

$$\widehat{H}_{1/2} = \left(VT^{-1/2}V^T\right)^{-1} = MVT^{1/2}V^T M. \tag{36}$$

The complexity of the full ($k = n$) generalised Lanczos algorithm is in general $O(n^3)$. However, in many applications of interest we do not need to compute $H_{1/2}$, but simply apply it (or its inverse) to a given vector $\mathbf{z} \in \mathbb{R}^n$. In such cases, a truncated version of the algorithm is used in practice with only k Lanczos vectors being constructed. As we are interested in approximations of $H_{1/2}\mathbf{z}$ we note first that if we start the Lanczos process with $\mathbf{v} = \mathbf{z}$ then

$$V_k^T M\mathbf{z} = \mathbf{e}_1\|\mathbf{z}\|_M$$

where $\mathbf{e}_1 \in \mathbb{R}^k$ is the first column of the identity I_k. This leads us to consider the following approximations of the matrix-vector products:

$$H_{1/2}\mathbf{z} \approx MV_k(I_k + T_k^{1/2})\mathbf{e}_1\|\mathbf{z}\|_M.$$

and

$$\widehat{H}_{1/2}\mathbf{z} \approx MV_k T_k^{1/2}\mathbf{e}_1\|\mathbf{z}\|_M$$

Similarly, if we wish to apply the inverse of $H_{1/2}$ to a given vector $\mathbf{z}$ we first note that if we start the iteration with $\mathbf{v} = M^{-1}\mathbf{z}$ then

$$V_k^T \mathbf{z} = V_k^T M (M)^{-1} \mathbf{z} = \mathbf{e}_1 \|M^{-1}\mathbf{z}\|_M = \mathbf{e}_1 \|\mathbf{z}\|_{M^{-1}}.$$

This leads us to consider the following approximations (cf. Lemma 4.8)

$$H_{1/2}^{-1}\mathbf{z} \approx V_k \left(I_k + T_k^{1/2}\right)^{-1} V_k^T \mathbf{z} = V_k \left(I_k + T_k^{1/2}\right)^{-1} \mathbf{e}_1 \|\mathbf{z}\|_{M^{-1}}. \tag{37}$$

and

$$\widehat{H}_{1/2}^{-1}\mathbf{z} \approx V_k T_k^{-1/2} V_k^T \mathbf{z} = V_k T_k^{-1/2} \mathbf{e}_1 \|\mathbf{z}\|_{M^{-1}} \tag{38}$$

The complexity of the above operations depends on the complexity corresponding to the application of the inverse of M. If this operation can be achieved in $O(n)$ operations, then the overall complexity of computing $H_{1/2}\mathbf{z}, H_{1/2}^{-1}\mathbf{z}$ is of order $O(kn)$ for $k \ll n$, with storage requirements of the same order.

5 Numerical experiments

We performed a series of experiments on some standard elliptic problems both for 2D and 3D domains. The domains were subdivided *a priori* (prior to triangulation) so that the resulting subdomain boundaries were linear (planar) faces. This allows for tangential gradient (27) to be implemented in a natural way. In both cases we used a number of levels of refinement to investigate performance. Mesh information is included for each test problem.
The iterative method employed in all cases is the GMRES method with right preconditioners

$$P_R = \begin{pmatrix} A_{II} & A_{IB} \\ 0 & H_{1/2} \end{pmatrix}, \quad \widehat{P}_R = \begin{pmatrix} A_{II} & A_{IB} \\ 0 & \widehat{H}_{1/2} \end{pmatrix}.$$

and we recall here that

$$H_{1/2} = M + M(M^{-1}L)^{1/2}, \quad \widehat{H}_{1/2} = M(M^{-1}L)^{1/2}.$$

On a uniform mesh with mesh-size h the mass matrix is known to be spectrally equivalent to a h^2-scaling of the identity: $M \approx \widetilde{M} := h^2 I_{BB}$. This allows us to reduce $\widehat{H}_{1/2}$ to

$$\widetilde{H}_{1/2} = \widetilde{M}(\widetilde{M}^{-1}L)^{1/2} = hL^{1/2}.$$

We denote the corresponding block-triangular preconditioner by $\widetilde{P}_R$.
The preconditioners were implemented using the following decomposition of the inverse preconditioner

$$P_R^{-1} = \begin{pmatrix} A_{II}^{-1} & 0 \\ 0 & I_{BB} \end{pmatrix} \begin{pmatrix} I_{II} & -A_{IB} \\ 0 & I_{BB} \end{pmatrix} \begin{pmatrix} I_{II} & 0 \\ 0 & H_{1/2}^{-1} \end{pmatrix}.$$

which highlights the fact that this choice of block-triangular preconditioner is parallelizable. In particular, we need to invert a discrete operator separately on each subdomain (A_{II} being block-diagonal), perform in parallel some boundary-to-domain updates (A_{IB} being a block-rectangular matrix) and also apply the action of a discrete $H^{1/2}$-norm on the internal boundary (or skeleton). The latter was achieved both exactly using a direct method to compute a generalised eigenvalue decomposition and iteratively using the Lanczos approximations (37), (38), in which case we employed the flexible GMRES method (Saad 1993) to take into account the changing nature of

		n_B			
level	n	$N = 4$	$N = 16$	$N = 64$	$N = 256$
1	16,641	253	753	1,729	3,585
2	66,049	509	1,521	3,521	7,425
3	263,169	1,021	3,057	7,105	15,105

Table 1: Mesh information for two-dimensional experiments

the preconditioner. Note that while in some cases the problem is symmetric, our preconditioner is non-symmetric - we found that the block-triangular preconditioner introduced in the previous section out-performed standard symmetric Krylov solvers. Not least, our aim was to monitor methodically the performance of a single (non-symmetric) iterative method as the problems become more and more non-symmetric (e.g., convection-diffusion problem with diminishing diffusion). Finally, we remark here that we did not use a two-level approach in order to highlight the raw performance of our preconditioners. Clearly, a multi-level approach can also be considered in this case in a standard fashion.

5.1 2D results

In this section we present the numerical experiments obtained by solving some standard elliptic problems in two dimensions. The problems were solved on the same domain $\Omega = (-1, 1)^2$. We decomposed Ω into $N = N_x \times N_y$ subdomains of size $2/N_x \times 2/N_y$ each, with $N_x = N_y \in \{2, 4, 8, 16\}$. Each subdomain was triangulated uniformly so that we work with a sequence of nested grids as well as nested subdomain partitions. The mesh/subdomain information is shown in Table 1 together with the number of nodes n_B on the internal boundary (skeleton). We also used the same domain decomposition with non-uniform mesh refinements with parameters n, n_B of the same order as those corresponding to the uniform case. We chose not to include here the mesh information for the non-uniform case in order to keep the presentation of numerical results succinct.

Problem 1 Consider the following standard model problem

$$\begin{cases} -\Delta u &= 1, \quad \text{in } \Omega, \\ u &= 0, \quad \text{on } \Omega. \end{cases} \tag{39}$$

The results for the choice of preconditioner $\widehat{P}_R$ implemented exactly are shown in Table 2. We find indeed that the number of iterations is independent of the size of the problem n, but increases logarithmically as the size of the subdomains is reduced. We also show in Table 2 the results for the simplified choice $\widetilde{P}_R$ computed exactly. As in the first case, the number of iterations is independent of the size of the problem n, but exhibits still a logarithmic dependence on the number of subdomains, though the number of iterations was reduced.

We also solved Problem 1 using a sequence of non-nested, non-uniform meshes. The results are shown in Table 3 and they indicate similar convergence behaviour: independence of level (size of problem) and logarithmic dependence on the number of subdomains.

Problem 2 We consider now the following model for reaction-diffusion problems posed on the

same domain and using the discretisations detailed in Table 1.

$$\begin{cases} -\Delta u + \alpha u &= 1, \quad \text{in } \Omega, \\ u &= 0, \quad \text{on } \partial\Omega. \end{cases} \tag{40}$$

The iteration counts for $\alpha = 1$, $\alpha = 10$, and $\alpha = 100$ are shown in Table 4. It is known that increasing α makes the problem 'easier' to solve iteratively, due to a increasingly dominant mass matrix. Our preconditioning strategy reflects this behaviour. Aside from independence of problem size, we see that a larger α leads to independence of number of domains as well.

Problem 3 Consider the following convection-diffusion problem

$$\begin{cases} -\nu\Delta u + \vec{b}\cdot\nabla u &= 0, \quad \text{in } \Omega, \\ u &= u_0, \quad \text{on } \partial\Omega. \end{cases} \tag{41}$$

with 'diagonal wind' $\vec{b} = (-1,-1)$ and also with 'rotating wind' $\vec{b} = (2y(1-x^2), -2x(1-y^2))$. The boundary data u_0 was chosen to be constant on some part of the boundary and zero elsewhere. This is a non-symmetric problem with a non-symmetric Steklov-Poincaré operator which results in a non-symmetric boundary Schur complement. While our preconditioners are non-symmetric, the matrices $H_{1/2}$, $\widehat{H}_{1/2}$, $\widetilde{H}_{1/2}$ are all symmetric and positive-definite. We would therefore expect performance to deteriorate as the degree of non-symmetry increases, which is the case for decreasing ν. We solved the problem using a standard stabilisation technique: streamline upwinding Petrov-Galerkin (SUPG). Our implementation adds an amount of diffusion in the direction of $\vec{b}$ which decreases with the Peclet number $Pe = h\|\vec{b}\|/\nu$.

The results are shown in Table 5 for diagonal wind and Table 6 for 'rotating wind'. In both cases we see that while independence of the size of the problem still holds for this problem, as predicted by theory, the number of iterations remains approximately constant with decreasing ν on finer meshes. Furthermore, the number of iterations decreases with decreasing h - this is due to the fact that the discrete operator is better resolved on finer meshes, with the amount of SUPG diffusion reduced considerably.

Finally, the logarithmic dependence on the number of subdomains remains unchanged for coarser meshes, but vanishes when the problem is well-resolved. Domain independence is neither predicted nor infirmed by our analysis; it appears that for the convection-diffusion problems under consideration this property holds rather robustly. Further analysis is required to explain the numerics.

5.2 3D results

We now present the numerical experiments obtained by solving some of the above elliptic problems in three dimensions. The problems were solved on the same domain $\Omega = (0,1)^3$. We decomposed Ω into $N = N_x \times N_y \times N_z$ subdomains of size $1/N_x \times 1/N_y \times 1/N_z$ each, with

preconditioner=	$\widehat{P}_R$				$\widetilde{P}_R$			
domains =	4	16	64	256	4	16	64	256
level = 1	13	16	19	27	8	11	14	19
2	14	16	20	27	8	11	15	20
3	14	16	21	28	9	12	15	20

Table 2: GMRES iterations for Problem 1 using a uniform mesh.

preconditioner=	$\widehat{P}_R$				$\widetilde{P}_R$			
domains =	4	16	64	256	4	16	64	256
level = 1	12	15	19	27	9	12	16	23
2	12	15	20	27	9	12	16	23
3	13	16	20	26	10	12	16	23

Table 3: GMRES iterations for Problem 1 using a non-uniform mesh.

$N_x = N_y = N_z \in \{2, 4, 8\}$. We used a non-uniform refinement of each subdomain. The mesh/subdomain information is shown in Table 7 together with the number of nodes n_B on the skeleton. The boundary preconditioner used in our block-triangular preconditioners was approximated using the Lanczos approximations of section 4 with $k = 20$. Note that this choice is not optimal for all mesh configurations – see section 5.3 for details.

Problem 1 We solved the three-dimensional version of problem (39) posed in the unit cube. The results are shown in Table 8.
As before, we see a domain dependence which appears to be logarithmic and virtually no dependence on the size of the problem. However, the latter property is somewhat affected by the poor Lanczos approximation (only 20 vectors).

Problem 2 We also solved the three-dimensional version of problem (40) for the same values of α. The results are shown in Table 9.
We note that while the independence of n is still evident, there is a dependence on the number of domains for all values of α. This was not the case for the 2D version of the problem; however, this behaviour is due to the coarse Lanczos approximation of our boundary preconditioner.

5.3 Lanczos approximation of $H^{1/2}$-norms

We end our experiments with a numerical study of the Lanczos approximation $L^{1/2}$. We chose to run the 2D experiment corresponding to problem (39) on a uniform mesh using a Lanczos approximation with several choices of k for all the mesh configurations employed in that experiment. It is expected that the number of iterations deteriorates with decreasing k. On the other hand, in the limit $k = n_B$, we should be able to recover the results shown in Table 2 corresponding to $\widetilde{P}_R$. Let $it(k)$ denote the number of GMRES iterations corresponding to an approximation of $\widetilde{P}_R$ using k Lanczos vectors. To describe the dependence on k we chose to plot the additional number of FGMRES iterations $it(k) - it(n_B)$ against k. The results are shown in Fig. 1 for all levels and all domains. A comparison between domains for a fixed level is shown in Fig. 2 for levels 1 and 2.

α=	1				10				100			
domains =	4	16	64	256	4	16	64	256	4	16	64	256
level = 1	12	14	16	22	11	13	13	16	10	10	12	13
2	12	14	17	23	11	13	12	15	10	11	12	13
3	12	15	17	23	11	13	12	15	10	11	12	11

Table 4: GMRES iterations for Problem 2 using P_R and a uniform mesh.

$\nu=$	1				10^{-1}				10^{-2}			
domains =	4	16	64	256	4	16	64	256	4	16	64	256
level = 1	8	10	14	18	7	9	12	15	9	12	14	15
2	7	9	13	16	6	8	9	9	8	10	11	10
3	6	8	10	13	5	6	6	5	7	8	8	6

Table 5: GMRES iterations for Problem 3 with 'diagonal wind' using $\widetilde{P}_R$ and a uniform mesh

$\nu=$	1				10^{-1}				10^{-2}			
domains =	4	16	64	256	4	16	64	256	4	16	64	256
level = 1	8	10	13	17	6	8	9	10	9	14	13	10
2	7	9	12	15	6	6	6	6	8	11	9	7
3	6	8	10	12	5	5	4	5	7	7	5	4

Table 6: GMRES iterations for Problem 3 with 'rotating wind' using $\widetilde{P}_R$ and a uniform mesh

The experiments suggest that the size of the Lanczos basis depends on both n_B and N. In particular, note that in order to maintain a fixed number of GMRES iterations with varying level we need to double the size of the Lanczos basis with each regular mesh refinement, i.e., $k \approx n_B \approx h^{-1}$. This result holds for any N. We can also deduce that $k \approx N^{1/2}$ for fixed n, which is what Fig. 2 indicates. Finally, we note that while the above asymptotic behaviour indicates a dependence on n_B, the actual number of basis vectors needed in practice is remarkably low ($k \approx 100$ for $n = 10^6$) with the subdomain complexity dominating by far the overhead in computing the action of the inverse of our $H^{1/2}$-preconditioners.

6 Conclusions

We proved that our preconditioner is independent of the mesh size and we illustrated this in our numerical experiments. However, we still have a logaritmic dependence on the number of subdomains used. This dependence is more evident in the solution of Lapace equations, but for the reaction-diffusion problems it seems to weaken for large values of α. The same behaviour has been found for the convection-diffusion problems where for fine meshes we almost achieve independence from both mesh size and number of subdomains.

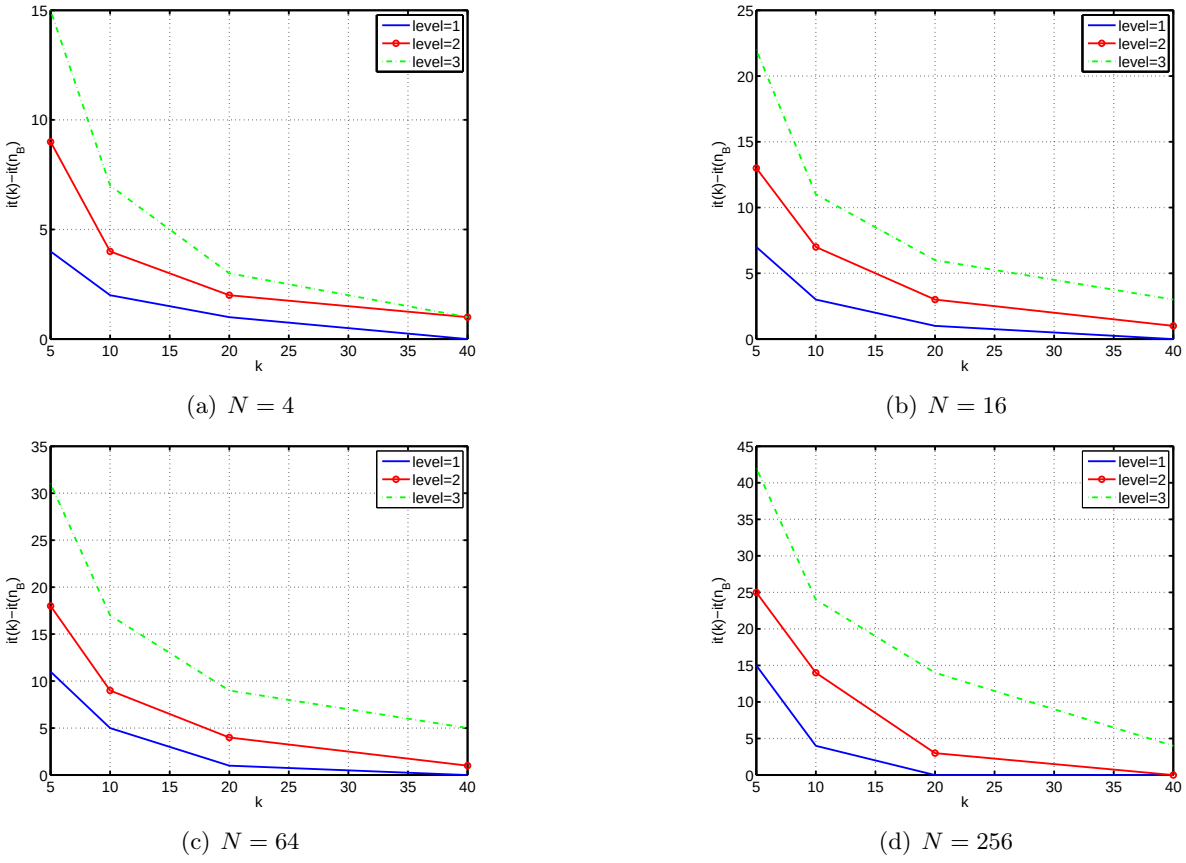
The numerical cost of our preconditioner scales with n_B and N, and we analysed experimentally the effect of a reduced number of Lanczos steps on the overall algorithm. Lanczos subspaces

	n			n_B		
domains=	8	64	512	8	64	512
level=1	28,603	29,943	34,821	3,214	9,231	20,579
2	229,041	238,839	255,606	12,880	38,980	89,677
3	1,884,996	1,902,206	1,939,420	53,460	158,733	364,470

Table 7: Mesh information for three-dimensional experiments

preconditioner=	$\widehat{P}_R(k)$			$\widetilde{P}_R(k)$		
domains =	8	64	512	8	64	512
level = 1	23	25	34	21	25	32
2	22	27	33	19	26	31
3	24	29	34	23	26	33

Table 8: FGMRES iterations for Problem 1 (3D): Lanczos with $k = 20$.



(a) $N = 4$ (b) $N = 16$

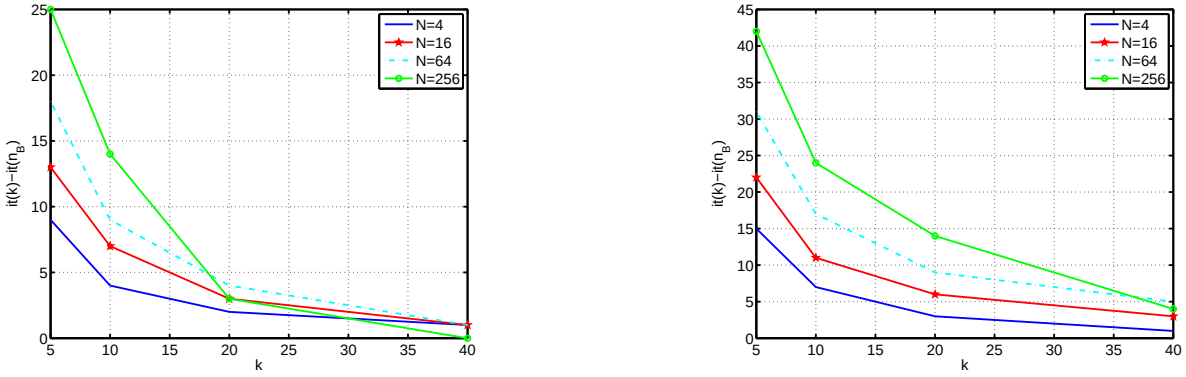
(c) $N = 64$ (d) $N = 256$

Figure 1: Additional FGMRES iterations as a function of k for fixed N.

of very limited dimensions produce an effective preconditioner for all our test problems. In particular for 3D problems, the practical cost of computing the fractional norms is less than the cost

α=	1			10			100		
domains =	8	64	512	8	64	512	8	64	512
level = 1	21	25	32	21	24	31	20	23	28
2	19	26	31	19	25	31	18	25	28
3	23	26	32	23	26	32	22	25	28

Table 9: FGMRES iterations for Problem 2 (3D) using $\widetilde{P}_R$: Lanczos with $k = 20$.



(a) level= 2 (b) level= 3

Figure 2: Additional FGMRES iterations as a function of k for fixed n.

of solving the 3D problem in one of the subdomain.

Further study will be devoted to the analysis of the approximation error and of the computational complexity of the Lanczos process for the computation of the $H^{1/2}$-norms.

References

Agmon, S., Douglis, A. and Nirenberg, L. (1959), 'Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions. I', *Comm. Pure Appl. Math.* **12**, 623 – 727.

Arioli, M. and Loghin, D. (2008), Discrete interpolation norms with applications, Technical Report Technical Report RAL-TR-2008-012, Rutherford Appleton Laboratory, Didcot, UK,.

Elman, H. C. (1982), Iterative methods for large sparse non-symmetric systems of linear equations, Phd thesis, Yale University, New Haven.

Hale, N., Higham, N. J. and Trefethen, L. N. (2008), 'Computing A^α , $\log A$ and related matrix functions by contour integrals', *SIAM J. Numer. Anal.* To appear.

Lions, J. and Magenes, E. (1968), *Problèmes aux limites non homogènes et applications*, Dunod, Paris, France.

Loghin, D. and Wathen, A. J. (2004), 'Analysis of preconditioners for saddle-point problems', *SIAM J. Sci. Comp.* **25**(6), 2029 – 2049.

Parlett, B. N. (1998), *The Symmetric Eigenvalue Problem*, Classics in Applied Mathematics 20, SIAM, Philadelphia, USA.

Peisker, P. (1988), 'On the numerical solution of the first biharmonic equation', *Mathematical Modelling and Numerical Analysis* **22**(4), 655–676.

Quarteroni, A. and Valli, A. (1999), *Domain Decomposition Methods for Partial Differential Equations*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, UK.

Riesz, F. and Sz-Nagy, B. (1956), *Functional Analysis*, Blackie and Son Limited.

Saad, Y. (1993), 'A flexible inner-outer preconditioned GMRES algorithm', *SIAM J. Sci. Comp.* **14**(2), 461 – 469.

Saad, Y. (2003), *Iterative Methods for Sparse Linear Systems*, second edition edn, SIAM, Philadelphia, USA.

Toselli, A. and Widlund, O. (2005), *Domain Decomposition Methods – Algorithms and Theory*, Springer, Berlin, Heidelberg.