LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-209688

# Overlapping Schwarz for Nonlinear Problems. An Element Agglomeration Nonlinear Additive Schwarz Preconditioned Newton Method for Unstructured Finite Element Problems

X. C. Cai, L. Marcinkowski, P. S. Vassilevski

February 14, 2005

Application of Mathematics

# Overlapping Schwarz for Nonlinear Problems. An Element Agglomeration Nonlinear Additive Schwarz Preconditioned Newton Method for Unstructured Finite Element Problems

Xiao-Chuan Cai[1], Leszek Marcinkowski[2], and Panayot S. Vassilevski[3] *

February 10, 2005

[1]*Department of Computer Science*
*University of Colorado at Boulder,*
*Boulder, CO 80309, USA.*
*e-mail:cai@cs.colorado.edu*

[2]*Department of Mathematics, Warsaw University,*
*Banacha 2, 02-097 Warsaw, Poland.*
*e-mail:lmarcin@mimuw.edu.pl*

[3]*Center for Applied Scientific Computing,*
*UC Lawrence Livermore National Laboratory,*
*P.O. Box 808, L-560, Livermore, CA 94551, USA.*
*e-mail:vassilevski1@llnl.gov*

**Abstract**

This paper extends previous results on nonlinear Schwarz preconditioning ([4]) to unstructured finite element elliptic problems exploiting now nonlocal (but small) subspaces. The nonlocal finite element subspaces are associated with subdomains obtained from a non-overlapping element partitioning of the original set of elements and are coarse outside the prescribed element subdomain. The coarsening is based on a modification of the agglomeration based AMGe method proposed in [8]. Then, the algebraic construction from [9] of the corresponding non–linear finite element subproblems is applied to generate the subspace based nonlinear preconditioner. The overall nonlinearly preconditioned problem is solved by an inexact Newton method. Numerical illustration is also provided.

Keywords: *algebraic multigrid, agglomeration, non–linear elliptic problems, nonlinear preconditioning, Newton method, finite elements*

AMS Classification *65F10, 65N20, 65N30*

## 1 Introduction

In this paper we introduce a parallel iterative method for solving system of nonlinear algebraic equations arising in the discretization of nonlinear elliptic partial differential equations using finite element methods. We follow the general framework of nonlinear additive Schwarz preconditioned inexact Newton methods, as outlined in [4]. In the classical Schwarz type algorithms, the subproblems are "local" problems in the sense that they only cover a small portion of the computational domain. Such an approach is good for distributed parallel computing since the localized subproblems can be solved independently on different processors, however, due to the lack of communication among the

*This work was performed under the auspices of the U.S. Department of Energy by the University of California Lawrence Livermore National Laboratory: contract/grant number: W-7405-Eng-48. The contribution of the second author was also partially supported by Polish Scientific Grant 2/P03A/005/24.

subproblems, the convergence degenerates as the number of processors grows without using a coarse problem that connects all the subproblems. A new approach to the linear Schwarz algorithm was introduced in [1], in which all "local" subproblems are made "global", in the sense that the mesh for the subproblem is dense (or fine) in only a small subdomain and coarse outside the small subdomain, and the subproblem mesh effectively covers the whole domain. In this paper, we study a parallel and optimal method for nonlinear problems by combining the ideas of agglomeration, nonlinear Schwarz, and preconditioned Newton methods. Under certain assumption, we prove the optimal convergence of the proposed algorithms, and some numerical experiments are presented to support the theory.

To be specific, we consider a model second order elliptic problem,

$$-\nabla \cdot (a(u)\nabla u) + g(u)u = f, \tag{1}$$

defined on a polygonal domain $\Omega$ with Dirichlet boundary conditions, $u = 0$ on $\partial\Omega$. The coefficient functions $a = a(\xi) > 0$, $g = g(\xi) \geq 0$, for any $\xi \in R$, and the right–hand side $f = f(x)$ are given. In what follows, we assume that they can be analytically evaluated for any value of their argument. Otherwise, in practice, one has to utilize interpolation. Let $\mathcal{T}_h$ be a given triangulation of $\Omega$ of triangular elements and let $V = V_h$ be a conforming finite element space of piecewise linear continuous functions associated with $\mathcal{T}_h$ and vanishing on $\partial\Omega$. Also, for any element $T \in \mathcal{T}_h$, we define the averaged coefficient

$$a_T(u) = a\left(\frac{1}{3}\sum_{x_i \in \text{verticesof}T} u(x_i)\right).$$

Similarly, let

$$g_T(u) = g\left(\frac{1}{3}\sum_{x_i \in \text{verticesof}T} u(x_i)\right).$$

A general element–based procedure for averaging functions and their derivatives was outlined in [9], see also Section 3 of the present paper.

The finite element discretization of (1) under consideration reads: Find $v = u_h \in V$ such that,

$$a(v,w) \equiv \sum_{T\in\mathcal{T}_h}\left(a_T(v)\int_T \nabla v\cdot\nabla w\,dx + g_T(v)\int_T vw dx\right) = (f,w), \quad \forall\, w \in V.$$

It is clear that the discrete nonlinear problem takes the form,

$$F(\mathbf{u}^*)\mathbf{u}^* = \mathbf{f}, \tag{2}$$

where $\mathbf{u}^*$ denotes the exact solution of the nonlinear system, and $F(\mathbf{v})$ is a linear operator (matrix) assembled from the local element matrices

$$a_T(v)A_T + g_T(v)M_T.$$

Here $A_T$ corresponds to the element matrices coming from the Laplace operator, and $M_T$ stands for the element mass matrix; i.e.,

$$A_T = \left\{\int_T \nabla\varphi_j\cdot\nabla\varphi_i\,dx\right\} \text{ and } M_T = \left\{\int_T \varphi_j\varphi_i\,dx\right\}, \tag{3}$$

where $\{\varphi_i\}$ span the fine grid finite element space. Hence the element matrices corresponding to $F(\mathbf{v})$, for any given $v \in V$, are the linear combinations $a_T(v)A_T + g_T(v)M_T$.

The objective of the present paper is based on an algebraic construction of a coarser version of the original fine–grid nonlinear problem discretized on generally unstructured grids (as described in

[9]) to study the behavior of the nonlinearly preconditioned inexact Newton method proposed in [4]. In addition we provide some model analysis of the method that applies to a simplified version of the coarsened away meshes (similarly to the case of linear problems studied in [2]).

For a given subdomain $G$, which is a union of elements $T$ from $\mathcal{T}_h$, we define a set of agglomerated elements $E$, where $E = T$ for all $T$ outside $G$ and $E = \{T\}_{T \subset G}$. Then we run the agglomeration based algorithm from [8] labeling all faces in $G$ as unacceptable for agglomeration. We run the agglomeration algorithm until the number of final agglomerated elements is acceptably small. Thus, we end up with a sequence of coarser triangulations, and at every coarsening level the original fine–grid elements in $G$ are still present (see Fig. 2, p. 18, for an illustration). Then one is able to define the non–local coarse nonlinear problems.

# 2 Generating agglomerated meshes that are coarsened away from a given mesh domain

Given a finite element mesh (triangulation) $\{T\}$ and a domain $G$ which is union of finite elements one can exploit certain topological relations of the mesh to create agglomerated elements $\{E\}$, where each $E$ is a list of fine–grid elements and each fine–grid element belongs to exactly one agglomerate $E$. More specifically, we assume that one has access to the relations "**element_face**", the adjoint one "**face_element**" and the "**face_face**" connectivity (on the fine mesh). These relations "**obj1_obj2**" can be viewed as boolean sparse matrices, where the rows correspond to the "**obj1**" and the columns to "**obj2**" and non–zero entries at position $(i, j)$ of the respective table indicate that $i$th "**obj1**" is "related" to the $j$th entry of "**obj2**". For example, if "**obj1**" stands for "**element**" and "**obj2**" stands for "**face**', "related" means that element $i$ has (geometrical) face $j$, whereas if **obj1**" stands for "**face**" and "**obj2**" stands for "**face**', "related" now means that face $i$ intersects face $j$. Based on the three relations ("**element_face**", "**face_element**" and "**face_face**") in [8] an element agglomeration algorithm was proposed. We can easily modify it here in order to have the elements in $G$ stay at all recursive applications of the agglomeration algorithm.

At step $l \geq 0$ of the algorithm we set $G_0 = G$ and for $l > 0$, we define $G_l = G_{l-1} \cup \{T \in \mathcal{T}_{l-1}$ is a neighbor of $G_{l-1}\}$. Label all faces of $T \in G_l$ as unacceptable in the following agglomeration step. That is, every face of the elements $T \in \mathcal{T}_l$ come with an integer weight $w = w(f)$ which is initially 0 or $-1$ if $f$ is unacceptable.

**Algorithm 2.1 (Building coarsened away agglomerated elements)**

1. *global search: find a face $f$ with a maximal weight and put the elements $T^+$ and $T^-$ which share that face on the list of the current agglomerate $E$; then update the weights of the neighboring non-eliminated faces $g$ (i.e. for which $w(g) > -1$) $w(g) := w(g) + 1$. Here we use the relation "***face_face***". Also, if $g$ and $f$ belong to a common element (here we use the relations "***face_element***" and "***element_face***") we increase the weight of $g$ once more, that is, we set $w(g) := w(g) + 1$ and finally we label $f$ as eliminated (or unacceptable) by setting $w(f) = -1$. If all faces have been eliminated (or unacceptable) go to step (3).*

2. *local search: loop over the faces of all elements already agglomerated in $E$ and find a face $f$ with maximal weight ($> -1$); if the weight of that face is less than the weight of the last eliminated face, or all local faces have weight less than zero, the agglomerate $E$ is completed and we go to step (1) for a global search. Otherwise, we eliminate $f$ and add the elements which share $f$ on the list $E$. Then we perform the weight increase of the neighboring (non-eliminated) faces $g$ of $f$ as described before. Then we repeat the loop (2) again.*

3. *final step: label all hanging elements which have not been agglomerated as new agglomerates (each such agglomerate consists of one fine element).*

The above agglomeration step can be efficiently implemented using double linked lists. We also note, that the original elements in $G$ will stay on all levels as (hanging) agglomerated elements since their faces are labeled as unacceptable to begin with and they are isolated from the rest.

In order to be able to recursively apply the above algorithm one needs to define faces (AEfaces) of the agglomerated elements (AEs) and build the respective (coarse) relations "**AE_AEface**", "**AEface_AE**" and "**AEface_AEface**". Algorithms for this are found in [8] and [13].

# 3 The Schwarz method

## 3.1 The construction of Schwarz subspaces

Let $\{G\}$ be a partitioning of the original set of elements $\mathcal{T}_h = \mathcal{T}_0$. Each $G$ is a union of elements and each element $T$ is contained in exactly one $G$. The partitioning can be carried out for example, by the graph partitioning software METIS, available at [10]. Then for each $G$ we perform the algorithm from the preceding section thus ending up with a sequence of nested triangulations $\mathcal{T}_l$ (which depend on $G$). We assume that we have access to the element matrices $A_T$ and $M_T$ corresponding to the Laplacian and identity operator at the fine level (for $T \in \mathcal{T}_0$), as defined in (3). Based on the agglomeration AMGe method from [8] one can compute corresponding element matrices $A_T$ and $M_T$ for agglomerated elements $T \in \mathcal{T}_l$ for every coarse level $l > 0$. The same algorithm selects coarse degrees of freedom $\mathcal{N}_l$ as vertices of the agglomerated elements at every coarsening level and builds interpolation matrices $P_l : V(\mathcal{N}_{l+1}) \mapsto V(\mathcal{N}_l)$ which maps the space of coarse vectors to fine grid vectors. Let $\ell$ be the coarsest level produced by the agglomeration **Algorithm** 2.1 (starting with the subdomain $G$). Then we define

$$P_G = P_0 P_1 \dots P_{\ell-1}$$

as the resulting interpolation from the coarsest level grid $\mathcal{T}_G \equiv \mathcal{T}_\ell$ to the finest one. We denote the coarse grid (the set of degrees of freedom– vertices) at level $\ell$ by $\mathcal{N}_G$ and the respective vector space by $V_G = V(\mathcal{N}_G)$.

Assume now that we are dealing with the model nonlinear operator (1) for a given positive function $a(u)$ and a non–negative function $g(u)$. A coarse discretization on the subspace $V_G$ can be defined as: Find $\mathbf{u} := \mathbf{u}_G \in V_G$ such that

$$\sum_{T \in \mathcal{T}_G} \left( a_T(\mathbf{u}) \mathbf{w}_T^T A_T \mathbf{u}_T + g_T(\mathbf{u}) \mathbf{w}_T^T M_T \mathbf{u}_T \right) = \mathbf{w}^T (P_G^T \mathbf{f}), \quad \text{for any } \mathbf{w} \in V_G. \tag{4}$$

Here, on each agglomerated element $T$, we use the following averages:

$$a_T(\mathbf{u}) = a \left( \frac{1}{\text{number of vertices of } T} \sum_{x_i \in \{\text{vertices of } T\}} u(x_i) \right),$$

and similarly

$$g_T(\mathbf{u}) = g \left( \frac{1}{\text{number of vertices of } T} \sum_{x_i \in \{\text{vertices of } T\}} u(x_i) \right).$$

Throughout the present paper, we use the dual notation $\mathbf{u}$ (a vector) and $u = u(x_i)$ (a discrete function). That is, $i$th entry of $\mathbf{u}$ corresponds to the $i$th coarse degree of freedom – the vertex $x_i$. Also, $\mathbf{v}_T$ stands for the restriction of a vector $\mathbf{v}$ on $T$ ($T$ as a set of coarse degrees of freedom). In matrix–vector notation (4) reads:

$$F_G(\mathbf{u})\mathbf{u} = P_G^T \mathbf{f}. \tag{5}$$

Here, the matrix $F_G(\mathbf{u})$ is assembled from the element matrices:

$$\{a_T(\mathbf{u}_T) A_T + g_T(\mathbf{u}_T) M_T\}.$$

### 3.2 Setting up the nonlinearly preconditioned problem

Having the small dimensional nonlinear problems (4) defined, we are ready to introduce a nonlinearly preconditioning method, in a similar way to the one in [5], cf. also [4]. In the present context it is formulated as follows:

$$0 = \mathcal{F}(\mathbf{u}) \quad \equiv \sum_G P_G \mathbf{g}_G. \tag{6}$$

Here, $\mathbf{g}_G$ solves the nonlinear problem (for any given $\mathbf{u}$)

$$F_G(\mathbf{u}^*_G + \mathbf{g}_G)(\mathbf{u}^*_G + \mathbf{g}_G) = P_G^T F(\mathbf{u})\mathbf{u}, \tag{7}$$

where $\mathbf{u}^*_G$ is a solution of (5), which, in practice must be precomputed. It is clear that if $\mathbf{u}$ is the exact solution of $F(\mathbf{u})\mathbf{u} = \mathbf{f}$ then $\mathbf{g}_G = 0$ and hence $\mathcal{F}(\mathbf{u}) = 0$.

We will further show that under certain assumptions the converse statement is also true.

**Lemma 3.1** *The Jacobian $J$ at $v_0$ ($\mathbf{v}_0$) of the fine grid nonlinear operator $F(\mathbf{v})\mathbf{v}$ is given by the expression:*

$$\begin{aligned}\mathbf{w}^T J(\mathbf{v}_0)\mathbf{v} \quad &= \sum_{T\in\mathcal{T}_h} \big(a_T(v_0)\mathbf{w}_T^T A_T \mathbf{v}_T + g_T(v_0)\mathbf{w}_T^T M_T \mathbf{v}_T \\ &\quad + (va'(v_0))_T \mathbf{w}_T^T A_T \mathbf{v}_{0,\,T} + (vg'(v_0))_T \mathbf{w}_T^T M_T \mathbf{v}_{0,\,T}\big).\end{aligned}$$

*Here $\mathbf{v}_{0,\,T}$ is the restriction of $\mathbf{v}_0$ (or equivalently of $v_0$) on $T$ and*

$$\begin{aligned}(va'(v_0))_T \quad &\equiv a'\left(\tfrac{1}{3}\sum_{x_i-\text{vertices of } T} v_0(x_i)\right)\left(\tfrac{1}{3}\sum_{x_i-\text{vertices of } T} v(x_i)\right) \\ &= v_T a'_T(v_0).\end{aligned}$$

*The expression for $(vg'(v_0))_T$ is the same ($a'$ replaced with $g'$). It is clear that*

$$J(\mathbf{v}_0) = F(\mathbf{v}_0) + \text{ a convective term},$$

*which is assembled from the low–rank element matrices $\mathbf{b}_T\mathbf{e}_T^T$, where*

$$\mathbf{b}_T = \mathbf{b}_T(\mathbf{v}_0) \equiv \frac{1}{3}a'_T(\mathbf{v}_0)\mathbf{a}_T + \frac{1}{3}g'_T(\mathbf{v}_0)\mathbf{m}_T,$$

*$\mathbf{a}_T = A_T\mathbf{v}_{0,\,T}$, $\mathbf{m}_T = M_T\mathbf{v}_{0,\,T}$ and $\mathbf{e}_T = (1)$ is the constant vector of $1$s restricted to $T$.*

*The Jacobian actions of the nonlocal coarse nonlinear operators $F_G(\mathbf{u}_G)\mathbf{u}_G$ in (5) are computed similarly*

$$\begin{aligned}\mathbf{w}^T J_G(\mathbf{v}_0)\mathbf{v} \quad &= \sum_{T\in\mathcal{T}_G} \big(a_T(v_0)\mathbf{w}_T^T A_T \mathbf{v}_T + g_T(v_0)\mathbf{w}_T^T M_T \mathbf{v}_T \\ &\quad + (va'(v_0))_T \mathbf{w}_T^T A_T \mathbf{v}_{0,\,T} + (vg'(v_0))_T \mathbf{w}_T^T M_T \mathbf{v}_{0,\,T}\big).\end{aligned}$$

*Here $\mathbf{v}_{0,\,T}$ is the restriction of $\mathbf{v}_0$ (or equivalently of $v_0$) on $T$ and*

$$\begin{aligned}(va'(v_0))_T \quad &\equiv a'\left(\tfrac{1}{\text{number of vertices of } T}\sum_{x_i-\text{vertices of } T} v_0(x_i)\right) \\ &\quad \times \left(\tfrac{1}{\text{number of vertices of } T}\sum_{x_i-\text{vertices of } T} v(x_i)\right) \\ &= v_T a'_T(v_0).\end{aligned}$$

*Similarly, $(vg'(v_0))_T = v_T g'_T(v_0)$.*

*The Jacobian $\mathcal{J}$ of the nonlinearly preconditioned problem (6) is simply equal to*

$$\mathcal{J}(\mathbf{v}^0) = \sum_G P_G \left(J_G(\mathbf{v}^0_G)\right)^{-1} (P_G)^T J(\mathbf{v}^0), \tag{8}$$

*where* $\mathbf{v}_G^0$ *solves the equation*

$$F_G(\mathbf{v}_G^0)\mathbf{v}_G^0 = P_G^T F(\mathbf{v}^0)\mathbf{v}^0.$$

*That is, computing the Jacobian requires solution of the above linear subdomain problems.*

**Proof.** To verify the above formula for the Jacobian of (6) we need to expand

$$\mathcal{F}(\mathbf{v}^0 + \mathbf{g}) \simeq \mathcal{F}(\mathbf{v}^0) + \mathcal{J}(\mathbf{v}^0)\mathbf{g}.$$

For $\mathbf{g} = 0$ we obtain that

$$\mathcal{F}(\mathbf{v}^0) = \sum_G P_G(\mathbf{v}_G^0 - \mathbf{u}_G^0),$$

where $\mathbf{v}_G^0$ solves the nonlinear problem

$$F_G(\mathbf{v}_G^0)\mathbf{v}_G^0 = P_G^T F(\mathbf{v}^0)\mathbf{v}^0$$

and $\mathbf{u}_G^{(0)}$ is a solution of (5). Then,

$$\mathcal{F}(\mathbf{v}^0 + \mathbf{g}) \simeq \sum_G P_G(\mathbf{v}_G^0 + \mathbf{g}_G)$$

for a vector $\mathbf{g}_G$ which depends linearly on $\mathbf{g}$. Using the definition of the Jacobian $J_G(\mathbf{v}_G^0)$ we have that

$$F_G(\mathbf{v}_G^0 + \mathbf{g}_G)(\mathbf{v}_G^0 + \mathbf{g}_G) \simeq F_G(\mathbf{v}_G^0)\mathbf{v}_G^0 + J_G(\mathbf{v}_G^0)\mathbf{g}_G, \tag{9}$$

and, on the other hand, $\mathbf{v}_G^0 + \mathbf{g}_G$ solves (up to higher order terms) the nonlinear problem

$$\begin{aligned} F_G(\mathbf{v}_G^0 + \mathbf{g}_G)(\mathbf{v}_G^0 + \mathbf{g}_G) &\simeq P_G^T F(\mathbf{v}^0 + \mathbf{g})(\mathbf{v}^0 + \mathbf{g}) \\ &\simeq P_G^T F(\mathbf{v}^0)\mathbf{v}^0 + P_G^T J(\mathbf{v}^0)\mathbf{g} \\ &= F_G(\mathbf{v}_G^0)\mathbf{v}_G^0 + P_G^T J(\mathbf{v}^0)\mathbf{g}. \end{aligned} \tag{10}$$

Comparing the linear parts of both expressions for $F_G(\mathbf{v}_G^0 + \mathbf{g}_G)(\mathbf{v}_G^0 + \mathbf{g}_G)$, (9) and (10), we end up with the equality,

$$F_G(\mathbf{v}_G^0)\mathbf{v}_G^0 + J_G(\mathbf{v}_G^0)\mathbf{g}_G = F_G(\mathbf{v}_G^0)\mathbf{v}_G^0 + P_G^T J(\mathbf{v}^0)\mathbf{g}.$$

That is,

$$J_G(\mathbf{v}_G^0)\mathbf{g}_G = P_G^T J(\mathbf{v}^0)\mathbf{g}.$$

As a result we arrive at the desired formula (8) for $\mathcal{J}$,

$$\mathcal{J}(\mathbf{v}^0)\mathbf{g} = \sum_G P_G \mathbf{g}_G = \sum_G P_G \left(J_G(\mathbf{v}_G^0)\right)^{-1} (P_G)^T J(\mathbf{v}^0)\mathbf{g}.$$

□

### 3.3 An ASPIN algorithm

Then the additive Schwarz preconditioned inexact Newton (sometimes referred to as ASPIN) algorithm, for computing the next iterate $\mathbf{u}^{(k+1)}$ from a given iterate $\mathbf{u}^{(k)}$, in the present context takes the form:

**Algorithm 3.1 (ASPIN)** *Given a current iterate* $\mathbf{u}^{(k)}$*, and let* $\mathbf{u}_G^{(k)}$ *be coarse approximations to the respective solutions of (5). Then for* $k = 0, 1, \ldots$*, one computes:*

1. *Compute the nonlinear residual* $\mathbf{g}^{(k)} = \mathcal{F}(\mathbf{u}^{(k)})$ *through the following steps:*

*a) Find* $\mathbf{g}_G^{(k)}$ *for every subdomain* $G$ *by solving the coarse nonlinear problems*

$$F_G(\mathbf{u}_G^{(0)} + \mathbf{g}_G^{(k)})(\mathbf{u}_G^{(0)} + \mathbf{g}_G^{(k)}) = P_G^T F(\mathbf{u}^{(k)})\mathbf{u}^{(k)}$$

*with a starting iterate* $\mathbf{g}_G^{(k)} = 0$.

*b) Form the global residual*

$$\mathbf{g}^{(k)} = \mathcal{F}(\mathbf{u}^{(k)}) = \sum_G P_G \mathbf{g}_G^{(k)}.$$

*c) Check for convergence based on a norm* $\|.\|$ *of* $\mathbf{g}^{(k)}$.

2. *Find an inexact Newton direction* $\mathbf{p}^{(k)}$ *by solving the Jacobian system approximately,*

$$\left( \sum_G P_G \left( J_G(\mathbf{u}_G^{(0)} + \mathbf{g}_G^{(k)}) \right)^{-1} (P_G)^T J(\mathbf{u}^{(k)}) \right) \mathbf{p}^{(k)} = \mathbf{g}^{(k)}$$

*in the sense that*

$$\left\| \mathbf{g}^{(k)} - \sum_G P_G \left( J_G(\mathbf{u}_G^{(0)} + \mathbf{g}_G^{(k)}) \right)^{-1} (P_G)^T J(\mathbf{u}^{(k)}) \mathbf{p}^{(k)} \right\| \leq \eta_k \|\mathbf{g}^{(k)}\|$$

*for some given forcing sequence* $\eta_k \in [0, 1)$.

3. *Compute the new iterate*

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} - \lambda^{(k)} \mathbf{p}^{(k)}.$$

*where* $\lambda^{(k)}$ *is a damping parameter.*

### 3.4 The general coefficient case

In practice, one may be interested in the general case of non–linear coefficients, $a$ and $g$, that is,

$$a = a(x, u, \nabla u) \text{ and} g = g(x, u, \nabla u).$$

The first thing is to notice that in order to be able to perform various averages one needs to be able to compute (approximate) derivatives of the finite element functions on coarse discretization levels. In order to do this (as already utilized in [9]), we assume that one has access to the coordinates of the vertices of the elements on all grids. It is more convenient to state this, that one has access to the coordinate vectors $\mathbf{X} = (x_i)$ and $\mathbf{Y} = (y_i)$ where $(x_i,\ y_i)$ are the geometric coordinates of the $i$th vertex at a given discretization level. Then, since

$$\frac{\partial u}{\partial x} = \nabla u \cdot \nabla x \simeq \frac{\int_T \nabla u \cdot \nabla x}{\int_T 1} = \frac{\mathbf{X}_T^T A_T \mathbf{u}_T}{1_T^T M_T 1_T},$$

where $\simeq$ is actually equality if $u$ is linear over the element $T$. Here, $1_T$ is the vector of ones restricted to $T$. Note, that here we again need access to both $A_T$ and $M_T$, the element matrices corresponding to the Laplacian and the identity (mass) operators.

According to the above formula, one is able to compute derivatives of functions on every level, assuming access to the element matrices $A_T$ and $M_T$ as well as access to the coordinate vectors $\mathbf{X}$ and $\mathbf{Y}$. Then, first the formula for the non–linear operator at a given grid reads as before, $F(\mathbf{u})\mathbf{u}$, where $F(\mathbf{u})$ is assembled from the following weighted combination of the element matrices $A_T$ and $M_T$,

$$a\left((u)_T,\ \left(\frac{\partial u}{\partial x}\right)_T,\ \left(\frac{\partial u}{\partial y}\right)_T\right) A_T + g\left((u)_T,\ \left(\frac{\partial u}{\partial x}\right)_T,\ \left(\frac{\partial u}{\partial y}\right)_T\right) M_T.$$

The derivative averages read as above; for example,

$$\left(\frac{\partial u}{\partial y}\right)_T = \frac{\mathbf{Y}_T^T A_T \mathbf{u}_T}{1_T^T M_T 1_T}.$$

In order to compute the Jacobian of $F(\mathbf{v})\mathbf{v}$ at $\mathbf{v}_0$ ($v_0$), $J(\mathbf{v}_0)$, one can use the following formulas. Let, $a = a(v, v_x, v_y)$ and $g = g(v, v_x, v_y)$ and assume that one can analytically compute the partial derivatives

$$a' = \frac{\partial a}{\partial v},\ a'_x = \frac{\partial a}{\partial v_x},\ a'_y = \frac{\partial a}{\partial v_y},\ \text{and} g' = \frac{\partial g}{\partial v},\ g'_x = \frac{\partial g}{\partial v_x},\ g'_y = \frac{\partial g}{\partial v_y}.$$

The corresponding formula for $J(\mathbf{v}_0)$, for any $\mathbf{w}$ and $\mathbf{v}$, then reads,

$$\begin{aligned}
\mathbf{w}^T J(\mathbf{v}_0)\mathbf{v} \quad &= \mathbf{w}^T F(\mathbf{v}_0)\mathbf{v} \\
&+ \Bigg\{ \sum_{T\in\mathcal{T}_h} \Big[ (v)_T\ a'\left((v_0)_T,\ (\tfrac{\partial v_0}{\partial x})_T,\ \left(\tfrac{\partial v_0}{\partial y}\right)_T\right) \\
&\qquad + (\tfrac{\partial v}{\partial x})_T\ a'_x\left((v_0)_T,\ (\tfrac{\partial v_0}{\partial x})_T,\ \left(\tfrac{\partial v_0}{\partial y}\right)_T\right) \\
&\qquad + \left(\tfrac{\partial v}{\partial y}\right)_T\ a'_y\left((v_0)_T,\ (\tfrac{\partial v_0}{\partial x})_T,\ \left(\tfrac{\partial v_0}{\partial y}\right)_T\right)\Big]\ \mathbf{w}_T^T A_T \mathbf{v}_{0,\,T} \\
&+ \sum_{T\in\mathcal{T}_h} \Big[ (v)_T\ g'\left((v_0)_T,\ (\tfrac{\partial v_0}{\partial x})_T,\ \left(\tfrac{\partial v_0}{\partial y}\right)_T\right) \\
&\qquad + (\tfrac{\partial v}{\partial x})_T\ g'_x\left((v_0)_T,\ (\tfrac{\partial v_0}{\partial x})_T,\ \left(\tfrac{\partial v_0}{\partial y}\right)_T\right) \\
&\qquad + \left(\tfrac{\partial v}{\partial y}\right)_T\ g'_y\left((v_0)_T,\ (\tfrac{\partial v_0}{\partial x})_T,\ \left(\tfrac{\partial v_0}{\partial y}\right)_T\right)\Big]\ \mathbf{w}_T^T M_T \mathbf{v}_{0,\,T} \Bigg\}.
\end{aligned}$$

# 4 An abstract convergence theory for ASPIN

In this section we develop a simple general convergence theory for the nonlinear additive Schwarz preconditioned Newton method outlined in the previous section. The abstract theory will be based on several assumptions that will be verified in the next section for a class of specific nonlinear problems.

Assume that we are given a finite dimensional space $V$ and a set of spaces $V_G$ where $G$ runs over a finite set. The spaces $V$ and $V_G$ are related through a set of extension and restriction operators $E_G : V_G \to V$ and $R_G : V \to V_G$.

Both $V$ and $V_G$ are equipped with different pairs of norms. The norms defined over $V$ are denoted by $\|\cdot\|_X$ and $\|\cdot\|_Z$, whereas the respective norms defined over $V_G$ are denoted by $\|\cdot\|_{X,G}$ and $\|\cdot\|_{Z,G}$. We will drop the subscript $G$ whenever it will not cause any confusion.

Assume that we originally want to solve the problem:

$$F(u^*) = 0, \tag{11}$$

where $F : D(F) \subset V \to V$ is in $C^1$, and $u^*$ is the exact solution to be computed. To define a nonlinearly preconditioned problem we introduce a mapping $u \to u_G \in V_G$ as a solution of the following nonlinear system of equations defined on $G$.

$$F_G(u_G) = R_G F(u), \tag{12}$$

where

$$u_G = g_G + u_G^*$$

as defined in (7) and $F_G : D(F_G) \subset V_G \to V_G$ is a $C^1$ function.

The abstract ASPIN algorithm then reads:

Find the solution $u^*$ of (11) by solving the following nonlinearly preconditioned system of equations using an inexact Newton method

$$\mathcal{F}(u^*) \equiv \sum_G E_G\,(u_G - u_G^*) = 0, \tag{13}$$

where $u_G^* \in V_G$ is a pre-computed solution of a sub nonlinear system of equations

$$F_G\left(u_G^*\right) = 0. \tag{14}$$

It is straightforward to show that a solution of (11) is also a solution of (13). As it will be shown, the converse is also true under certain assumptions. We next state the assumptions for $F$,$F_G$, $E_G$, and $R_G$.

**Basic Assumptions** on $F$, $F_G$, $E_G$, and $R_G$.

(A1) The problem (11) has a solution; i.e., there exists $u^* \in V$ such that $F(u^*) = 0$.

(A2) In the neighborhood of $u^*$ there exists a nonsingular derivative of $F$, denoted by $DF$, with

$$\|DF(u^*)\|_{X\to Z} \le C_*.$$

(A3) The derivative $DF$ is Lipschitz in the neighborhood of $u^*$, i.e. for any $u, v$ in this neighborhood we have

$$\|DF(u) - DF(v)||_{X\to Z} \le L\|u - v\|_X$$

(A4) There is a ball about the origin (in the sense of $\|\cdot\|_{Z,G}$) such that for any $f_G$ in this ball, there exists a unique $u_G \in V_G$ such that $F_G(u_G) = f_G$ and $u_G$ is in a fixed neighborhood of $u_G^*$ (the solution of (14)).

(A5) For any $v$ in a neighborhood of $u_G^*$ there exists a derivative $DF_G(v)$ of $F_G$.

(A6) The derivative $DF_G(v)$ is invertible and $(DF_G(v))^{-1}$ is a uniformly bounded linear operator for any $v$ in a neighborhood of $u_G^*$ , that is, with a constant $\mu_G$ (independent of $v$) one has,

$$\|DF_G(v)^{-1}\|_{Z\to X} \le \mu_G,$$

(A7) The inverse of the derivative $DF_G^{-1}$ is Lipschitz in the neighborhood of $u_G^*$, i.e., for all $w, v$ in this neighborhood we have

$$\left\|(DF_G(w))^{-1} - (DF_G(v))^{-1}\right\|_{Z\to X} \le L_G\|w - v\|_X$$

(A8) boundedness of the extension and restriction mapping, i.e.,

$$\|E_G\|_{X\to X} \le C \ \text{ and } \ \|R_G\|_{Z\to Z} \le C.$$

As mentioned before, we drop the subscript $G$ whenever it does not introduce confusion.
The following result follows directly from the above assumptions.

**Theorem 4.1** *Under our assumptions the function $\mathcal{F}$ is well defined in a neighborhood of $u^*$ and has properly defined continuous derivative:*

$$D\mathcal{F}(u) = \mathcal{J}(u) = \sum_G E_G(DF_G(u_G))^{-1}R_G DF(u).$$

Finally, we also assume that

(A9) the inverse of the derivative of $\mathcal{F}$ is locally bounded, i.e. there is a positive constant $\beta_{\text{ASM}}$, such that for any $u$ in a neighborhood of $u^*$, one has,

$$\left\|\sum_G E_G(DF_G(u_G))^{-1}R_G DF(u)\right\|_{X\to X} \ge \beta_{ASM}.$$

By using the inverse theorem in Calculus and assumption (A9) one easily proves,

**Corollary 4.1** *The solution $u^*$ of the original nonlinear system of equations (11) is also a locally unique solution of the nonlinearly preconditioned system of equations (13).*

The next lemma represents a Lipschitz continuity property of $\mathcal{J}$ that ensures quadratic convergence of Newton method.

**Lemma 4.1** *Under our assumptions, there is a positive constant $L_{ASM}$ such that*

$$\|D\mathcal{F}(u) - D\mathcal{F}(v)\|_{X\to X} \le L_{ASM}\|u - v\|_X.$$

*for any $u$ and $v$ in a neighborhood of $u^*$.*

**Proof.** Since the function $\mathcal{F}$ is defined as a sum over all $G$, and we have assumed that their number is finite, it is sufficient to show, that for each $G$, the function $u \mapsto E_G(DF_G(u_G))^{-1}R_G DF(u)$ satisfies the statement of the lemma; i.e. the function is Lipschitz. By (A8) and the triangle inequality, we have

$$\begin{aligned}
&\|E_G(DF_G(u_G))^{-1}R_G DF(u) - E_G(DF_G(v_G))^{-1}R_G DF(v)\|_{X\to X} \le \\
&\le \|E_G\|_{X\to X}\|(DF_G(u_G))^{-1}R_G DF(u) - (DF_G(v_G))^{-1}R_G DF(v)\|_{X\to X} \\
&\le C\|(DF_G(u_G))^{-1}R_G DF(u) - (DF_G(v_G))^{-1}R_G DF(v)\|_{X\to X} \\
&\le \|(DF_G(u_G))^{-1}R_G\,[DF(u) - DF(v)]\|_{X\to X} + \\
&\quad \|[(DF_G(u_G))^{-1}R_G - (DF_G(v_G))^{-1}R_G]\,DF(v)\|_{X\to X}
\end{aligned}$$

We estimate the first term by (A3), (A6) and (A8) and obtain:

$$\begin{aligned}
&\|(DF_G(u_G))^{-1}R_G\,[DF(u) - DF(v)]\|_{X\to X} \\
&\le \|(DF_G(u_G))^{-1}\|_{Z\to X}\|R_G\|_{Z\to Z}\|DF(u) - DF(v)\|_{X\to Z} \\
&\le \beta_G\, L\, \|u - v\|_X.
\end{aligned}$$

For the second term by (A7) and (A8) we obtain

$$\begin{aligned}
&\|[(DF_G(u_G))^{-1} - (DF_G(v_G))^{-1})]R_G\,DF(v)\|_{X\to X} \\
&\le \|(DF_G(u_G))^{-1} - (DF_G(v_G))^{-1}\|_{Z\to X}\|R_G\|_{Z\to Z}\|DF(v)\|_{X\to Z} \\
&\le C\|\,u_G - v_G\|_{X,G}\|DF(v)\|_{X\to Z} \\
&\le C\|u_G - v_G\|_{X,G}\,\{\|DF(v) - DF(u^*)\|_{X\to Z} + \|DF(u^*)\|_{X\to Z}\}
\end{aligned}$$

In the last inequality, the second term is bounded by $\|DF(u^*)\|_{X\to Z} + L\delta$, due to (A2). Here $\delta$ is the diameter of the neighborhood of $u^*$. The Lipschitz continuity of the mapping $u \mapsto u_G$ (which follows from (A3),(A6) and (A8)) completes the proof. □

The main results of this section are summarized in the following theorem.

**Theorem 4.2** *For a properly chosen initial guess $u_0$, the Newton iteration defined by*

$$u_{n+1} = u_n - [D\mathcal{F}(u_n)]^{-1}\mathcal{F}(u_n)$$

*converges, and the convergence rate is quadratic; i.e.:*

$$\|u_{n+1} - u^*\|_X \le C\|u_n - u^*\|_X^2.$$

*where $C = L_{ASM}/\beta_{ASM}$ is a positive constant.*

**Proof.** The proof is standard and it follows directly from the previous lemma and assumption (A9), e.g. cf. Theorem 5.2.1, p. 90 in [6]. □

# 5 Application to second order semi–linear elliptic problems

In this section we apply the abstract convergence theory developed in the previous section to a class of semi-linear elliptic problems, and we show that this class of equations satisfy the assumptions proposed in the previous section, and therefore our ASPIN algorithm converges locally and the convergence rate is quadratic.

## 5.1 A model problem and its discretization.

We consider the following boundary value problem: Find $u^* \in H_0^1(\Omega)$ such that

$$\mathcal{L}u = Au - b(x,u) = f \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega, \tag{15}$$

where $\Omega$ is a convex polygonal domain in $\mathbf{R}^2$ and $f \in L^2(\Omega)$ and $A$ is an $H_0^1$ elliptic operator. We define $\mathcal{L}: \ \mathcal{V} \equiv H_0^1(\Omega) \mapsto \mathcal{V}' = H^{-1}(\Omega)$. Let $V_h \subset \mathcal{V}$ be a sequence of finite element spaces of continuous piecewise linear polynomials such that $\cup\{V_h, \ h \le h_0\} = \mathcal{V}$. In the present setting, for a given mesh $h$, $V_h$ corresponds to the finest level mesh from Section 2 and $V_G \subset V_h$ is one of the subspaces based on a coarsened away from $G$ mesh (i.e., $V_G = P_G V(N_G) \subset V_h$, cf. Section 2). We also assume that the original fine mesh is quasi-uniform, which implies certain inverse inequalities for the finite element functions. The coarsened away from $G$ mesh will be specified in the next subsection (in a model situation). The Galerkin operators induced by $\mathcal{L}$ on $V_h$, and similarly on $V_G$, denoted by $Av - b(v)$, are defined as follows

$$(Av - b(v), \ \varphi) = (\mathcal{A}\nabla v, \ \nabla\varphi) - (b(x,v), \ \varphi), \quad \forall v, \ \varphi \in V_h \ (\text{or} V_G),$$

where $\mathcal{A}(x)$ is uniformly positive definite in $\Omega$.

In the same way, the Jacobian of $Av - b(v)$ at $v$, which is a linear operator denoted by $A - b_u(v)$ ($b_u(v) \equiv \frac{\partial b(.,\, v)}{\partial u}$), is defined by

$$(A\xi - b_u(v)\xi, \ \varphi) = (\mathcal{A}\nabla\xi, \nabla\varphi) - \left(\frac{\partial b(x,v)}{\partial u}\xi, \ \varphi\right), \quad \forall \xi, \ \varphi \in V_h \ (\text{or} V_G).$$

The discrete counterpart of (15) reads: Find $u_h \in V_h$ such that

$$\begin{aligned} (L_h u_h, \ \varphi) &\equiv (Au_h - b(u_h), \ \varphi) \\ &\equiv (\mathcal{A}\nabla u_h, \nabla\varphi) - (b(x,u_h), \ \varphi) = (f, \ \varphi), \quad \forall \varphi \in V_h. \end{aligned} \tag{16}$$

Similarly, one defines $L_G$ by replacing in (16) $V_h$ with $V_G$. In the notation of the previous section, the discrete nonlinear problem (16) and its subdomain counter-part are formulated as

$$F_h(u_h) \equiv L_h(u_h) - Q_h f = 0 \tag{17}$$

$$F_G(u_h) \equiv L_G(u_h) - Q_G f = 0, \tag{18}$$

where $Q_h(\text{or} Q_G): \ L^2(\Omega) \mapsto V_h \ (\text{or} V_G)$ is the corresponding $L^2$–projection onto $V_h$ (or$V_G$). In what follows we denote the solution of (17) by $u_h^*$ and similarly, the solution of (18) by $u_G^*$. The derivative $F_h'(v)$, $v \in V_h$, is defined variationally as

$$(DF_h(v)\psi, \ \varphi) \equiv (A\psi - b_u(v) \ \psi, \ \varphi), \quad \forall \psi, \ \varphi \in V_h.$$

The derivative $DF_G$ is defined analogously by replacing the subscript $h$ by $G$ in the previous line.

**Definition 5.1 (Discrete Banach spaces)**

- $X = V_h$ *with a norm* $\|\cdot\|$, *such that* $\|\psi\| \le C\|F_h'(u_h^*)\psi\|_0$ *for any* $\psi \in V_h$.
- $Z = V_h$ *equipped with* $\|\cdot\| = \|\cdot\|_0$.

*Above, $\|\cdot\|_s$ stands for the Sobolev space $H^s$–norms, and $L^\infty$ on $V_h$ is actually the maximum norm (since the functions in $V_h$ are continuous).*

It is known that under certain regularity assumptions, one has with a mesh–independent constant $C$,

$$\|\psi\|_1 \le \max\{\|\psi\|_1,\ \|\psi\|_{L_\infty}\} \le C\|F_h^{'}(u_h^*)\psi\|_0.$$

In what follows, we will consider the two (mesh–independent) norms in $X$

$$\|\psi\|_X \equiv \max\{\|\psi\|_1,\ \|\psi\|_\infty\}, \quad \text{and} \quad \|\psi\|_{X,1} \equiv \|\psi\|_1.$$

The estimates to be obtained in the first norm will be a bit weaker, i.e., they will exhibit a very weak (logarithmic) dependence on the mesh size.

## 5.2 Subdomain spaces and norms.

Recall that to build $V_G$ in Section 2 we first obtain a sequence of subdomains $\{G_k\}$ with $G_0 = G$ and $G_{k-1} \subset G_k$ for $k = 0, \ldots, l$ with a preset (fixed) value of $\ell$. We then introduce a finite sequence of disjoint subdomains defined by $S_1 = G_1$ and $S_{k+1} = G_{2k+1} \setminus G_{2k-1}$, $k = 1, 2, \ldots$, cf. Fig. 1, and consider for each $S_k$ a triangulation $\mathcal{T}_h(S_k)$ inherited from $\mathcal{T}_h(G_{2k-1})$, i.e., formed from the (agglomerated) elements of $\mathcal{T}_h(G_{2k-1})$ restricted to $S_k$.

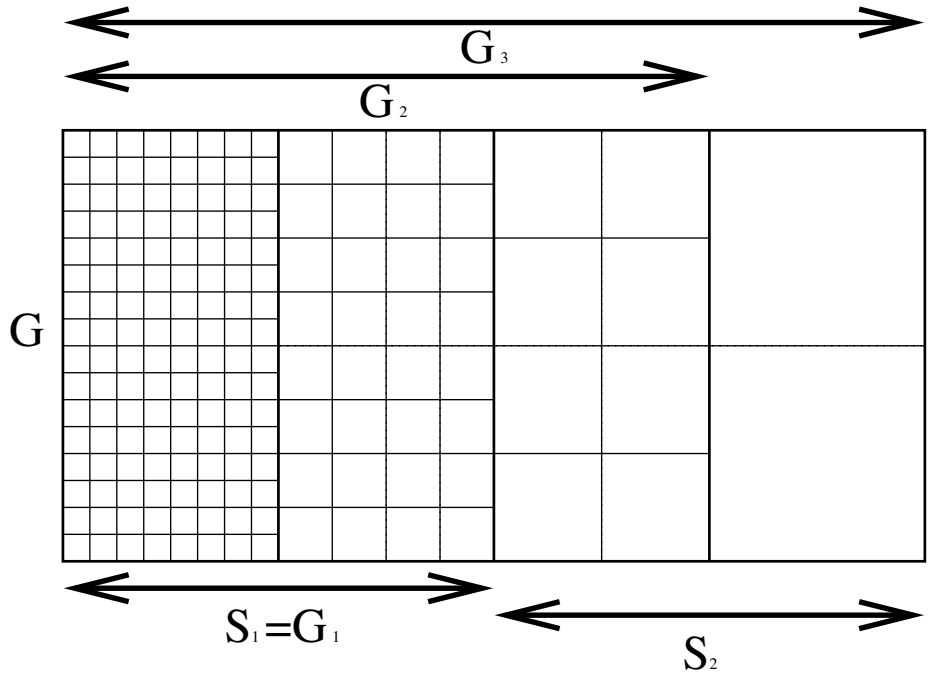


Figure 1: Subdomains $G_k$ and $S_k$.

We assume that the number of $S_k$ is bounded by some constant $(\ell/2)$. Because of the graduate coarsening $\mathcal{T}_h(S_k)$ contains elements only of $\mathcal{T}_{2k-1}$ and $\mathcal{T}_{2k-2}$; i.e., of the size between $h_{2k-1}$ and $h_{2k-2}$, thus it is reasonable to assume:

**Assumption 5.1 (Inverse Assumption)** *The mesh restricted to a subdomain $S_k$, i.e. $\mathcal{T}_h(S_k)$ is quasi-uniform. That is, the maximal diameter of an element in $S_k$ is of the same order as the minimal one.*

**Remark 5.1** *We note that inverse assumption will be needed only for convergence results in* $\|\cdot\|_X$.

We also want to consider different norms associated with subdomains $S_k$:

$$\|u\|_{\infty,\,G} = \sum_k \|u\|_{\infty,\,S_k} \quad \text{and} \quad \|u\|_{s,\,G} = \sum_k \|u\|_{s,\,S_k},$$

where $\|u\|_{\infty,\,S_k} = \max_{x\in S_k} |u(x)|$ and $\|u\|_{s,\,S_k}$ stands for a standard Sobolev norm $\|u\|_{H^s(S_k)}$. Then from the classical inverse inequality (cf. [12]), and a standard argument the following inverse inequality follows:

$$\|u\|_{\infty,\,G} \le C\sum_k (1+|\log(h(S_k))|^{1/2})\|u\|_{1,\,S_k} \le C(1+\max_k\{|\log(h(S_k))|^{1/2}\})\|u\|_{1,G}. \tag{19}$$

We stress upon the fact that the constants in the analysis to follow in generally depend on the subdomains $G$ and their number.

**Definition 5.2 (Subdomain spaces and norms)**
*For any subdomain* $G$ *we introduce the following spaces equipped with norms:*

- $\mathcal{X}_G = V_G$ *with a norm* $\|\psi\|_{X,G} = \max\{\|\psi\|_{\infty,G}, \|\psi\|_{1,G}\}$ *in the case of the first norm, i.e. for* $(X, \|\cdot\|_X)$*, and* $\|\psi\|_{X,1,G} \equiv \|\psi\|_{1,G}$*, in the other case, for any* $\psi \in V_G$*,*
- $\mathcal{Z}_G = V_G$ *equipped in a norm* $\|\psi\|_{Z,G} = \|\psi\|_{0,G}$*, in both cases for any* $\psi \in V_G$*.*

The subscript $G$ is dropped in what follows, whenever this does not cause any ambiguity.

## 5.3 Problem specific assumptions and their verifications

**Assumption 5.2 (cf. [3])** *We assume that the nonlinear boundary value problem has a solution* $u^* \in H_0^1(\Omega) \cap H^2(\Omega)$*. The latter implies that* $u^* \in L^\infty(\Omega)$ *as well.*
*In addition, we assume:*

*(i)* $\Omega$ *is a bounded convex polygon and the principal linear elliptic part* $A$ *of* $\mathcal{L}$ *is* $H^2$*–regular; that is, for any* $g \in L^2(\Omega)$ *the solution of the linear boundary value problem*

$$(Aw,\ v) = (g,\ v) \quad \forall v \in H_0^1(\Omega),$$

*satisfies the* a priori *estimate for a constant* $C_R > 0$ *(independent of the r.h.s.* $g$*),*

$$\|w\|_2 \le C_R\|g\|_0.$$

*Of course this is a well-known result for* $A = -\triangle$*, i.e., for identity diffusion coefficient matrix* $\mathcal{A} = I|_{\mathbf{R}^2}$*.*

*(ii) The function* $b(x,\ u)$ *is continuously differentiable; that is,* $\frac{\partial b(x,\ u)}{\partial u}$ *exists in a ball with a center at* $u^*$ *- the solution of (15) and is Lipschitz as a function of* $u \in \mathbf{R}$*, uniformly in* $x \in \Omega$

*(iii) The function* $|\frac{\partial b(x,\ u)}{\partial u}|$ *is bounded in* $\Omega$ *for* $u$ *close to* $u^*$*.*

*(iv) Finally, we assume that* $\frac{\partial b(x,\ u)}{\partial u} \le 0$*.*

For the analysis in the norm $\|\cdot\|_{X,1}$ we will need somewhat stronger assumptions on the function $b(.,.)$:

**Assumption 5.3** *(The case of $\|\cdot\|_{X,1}$ norm.) Let (i) and (iv) of Assumption 5.2 hold and we assume that $\frac{\partial b(x,\,u)}{\partial u}$ (and $b(x,u)$) exists on $\Omega \times \mathbf{R}$ and*

$$\begin{aligned} |b_u(x,\xi_1) - b_u(x,\xi_2)| &\leq L(1+|\xi_1|^r+|\xi_2|^r)|\xi_1-\xi_2| \quad \forall \xi_1,\xi_2 \in \mathbf{R}, \forall x \in \Omega, \\ |b_u(x,\xi)| &\leq C(1+|\xi|^s) \quad \forall \xi \in \mathbf{R}, \forall x \in \Omega, \end{aligned} \tag{20}$$

*where $L, C$ are positive constants and $r, s$ are non–negative constants.*

It is clear that (20) implies $(ii)$ and $(iii)$ of Assumption 5.2.

**Remark 5.2** *We note that (20) is a type of growth assumption, and such are commonly used in the theory of nonlinear PDEs, e.g. cf. [14].*

**Remark 5.3** *A simple example which satisfies Assumption 5.3 is $b(x,u) = -u^3$ which leads to the model problem: $Au + u^3 = f$.*

The main result of the present section is the following convergence theorem:

**Theorem 5.1** *For an initial guess $u_0$ in a neighborhood of $u_h^*$ the ASPIN algorithm is quadratically convergent with*

$$\begin{aligned} \|u_{n+1} - u_h^*\|_X &\leq C(1+|\log(h)|)^{1/2}\|u_n - u_h^*\|_X^2 \\ \|u_{n+1} - u_h^*\|_1 &\leq C\|u_n - u_h^*\|_1^2. \end{aligned}$$

**Proof.** We apply Theorem 4.2, for which we have to verify the assumptions (A1)-(A9) of Section 4 in both norms. The next section provides the necessary lemmas where these assumptions are verified. More specifically, assumptions (A2) and (A5) follow from Lemmas 3.3 and 4.1 in [3] and because $R_G \equiv Q_G$ is an $L^2$ projection onto $V_G$ then (A8) holds. It follows from Lemma 6.2, see below, that assumptions (A1) and (A4) are satisfied in the norms $\|\cdot\|_X$ and $\|\cdot\|_{X,1}$. Lemma 6.4 verifies (A3) and Lemma 6.3 implies (A6) and (A7). Finally (A9) is validated (in both norms) by Lemma 6.5. Thus the proof is complete. □

# 6 Technical lemmas

This section contains proofs of a few technical lemmas, which show that the assumptions from Section 4 are fulfilled in both norms for our problem (already used in the proof of Theorem 5.1). Some of the proofs here closely follow Sections 3 and 4 of [3].

**Lemma 6.1** *Let $w \in V_G$ solve the linear problem*

$$DF_G(u)w = g$$

*for a given $u \in V_G$ in a neighborhood of $u^*$ and $g \in V_G$. Then under Assumption. 5.2, we have*

$$\|w\|_{X,\,G} \leq C_F\|g\|_{0,G}.$$

*with a uniform constant $C_F$.*

**Proof.** We follow Lemma 3.3 in [3]. First we rewrite the problem into a variational form:

$$(DF_G(u)w,\psi) = (A - b_u(u)w,\psi) = (g,\psi) \;\; \forall \psi \in V_G.$$

The stability in $H^1$ norm (and thus in $\|\cdot\|_{X,\,G}$) is obtained in the standard way by taking $\psi = w$:

$$\begin{aligned} (Aw,w) &\leq (Aw,w) - (b_u(u)w,w) \\ &\leq \|g\|_0\|w\|_0 \leq \|g\|_0\|w\|_1 \\ &\leq \|g\|_0\|w\|_{X,G}. \end{aligned}$$

Note that due to Assumption 5.2 (iv), the following coercivity estimate holds $\|w\|_1^2 \le C(Aw, w) \le C\|w\|_1^2$, which shows $\|w\|_{1,G} \le C\ \|g\|_0$. It remains to show that $\|w\|_\infty \le C\|g\|_0$. For this purpose, we consider the differential problem:

$$((A - b_u(u))\Psi, \phi) = (g, \phi)\ \ \forall \phi \in H_0^1(\Omega).$$

From Lemma 3.2 in [3] it follows that $\|\Psi\|_2 \le C\|g\|_0$. Based on the approximation property of the nodal interpolation in $V_G$ there exists $\Psi_G \in V_G$ such that

$$\|\Psi - \Psi_G\|_{H^1(S_k)} \le Ch(S_k)|\Psi|_{H^2(S_k)}, \quad \|\Psi_G\|_{\infty,G} \le \|\Psi\|_{\infty,G}$$

and by the standard argument:

$$\|w - \Psi\|_{1,G} \le C\sum_k \|\Psi - \Psi_G\|_{H^1(S_k)} \le C\sum_k h(S_k)|\Psi|_{H^2(S_k)}.$$

Now by (19) we get

$$\|w - \Psi_G\|_{\infty,\,G} \le \sum_k (1 + |\log(h(S_k))|^{1/2})\|w - \Psi_G\|_{H^1(S_k)}.$$

Then we conclude that

$$\begin{aligned}
\|w\|_{\infty,G} &\le \|w - \Psi_G\|_{\infty,\,G} + \|\Psi_G\|_{\infty,\,G} \\
&\le C(\sum_k (1 + |\log(h(S_k))|^{1/2})\|w - \Psi_G\|_{H^1(S_k)} + \|\Psi\|_{\infty,\,G}) \\
&\le C(\sum_k (1 + |\log(h(S_k))|^{1/2})h(S_k)|\Psi|_{H^2(S_k)} + \|\Psi\|_{2,\,G}) \\
&\le C\|\Psi\|_{2,\,G} \le C\|g\|_{0,\,G}.
\end{aligned}$$

Thus the proof is complete. □

**Lemma 6.2**

1. *Under Assumptions 5.2 the discrete problem (17) has a unique solution $u_h^*$. Moreover the following error estimates hold:*

$$\|u^* - u_h^*\|_1 \le Ch\|u^*\|_2,\ \ \|u^* - u_h^*\|_\infty \le Ch^\alpha\|u^*\|_2,$$

*for some positive $\alpha$ $(< 1)$.*

2. *The problem (16) has a unique solution $u_G^*$ and the following error estimates hold:*

$$\|u^* - u_G^*\|_{1,G} \le CH\|u\|_{2.G},\ \ \|u^* - u_G^*\|_{\infty,G} \le CH^\alpha\|u^*\|_{2,G},$$

*for a positive $\alpha$ $(< 1)$ and $H = \max_k h(S_k)$. If we take $Q_G f + \delta f$ in (18) with $\delta f:\ \|\delta f\|_{Z,G}$ is sufficiently small then these estimates still hold.*

**Proof.** The first statement of the present lemma is Lemma 3.1 in [3], and the second one can be proved along the same lines there, the only difference is that we have to use our Inverse Assumption and Lemma 6.1 for the second statement of the lemma. □

The next results deal with the inverse of $DF_G$ for a local problem associated with a subdomain $G$.

**Lemma 6.3** *For any $u_1,\ u_2 \in V_G$ in a neighborhood of $u^*$ it holds that*

$$\begin{aligned}
\|(DF_G(u_1))^{-1}\|_{Z\to X} &\le C, \\
\|(DF_G(u_1))^{-1} - (DF_G(u_2))^{-1}\|_{Z\to X} &\le C\|u_1 - u_2\|_{X,G}.
\end{aligned}$$

*Under Assumptions 5.3 the same results hold in the $H^1$ type norms $\|\cdot\|_{X,1,G}$.*

**Proof.** For any $w \in L^2(\Omega)$ with $\|w\|_0 = 1$ we have the following identities

$$(DF_G(u_i)p_i, \varphi) = (Ap_i - b_u(u_i)p_i, \varphi) = (w, \varphi) \ \ \forall \varphi \in V_G, \ \ i = 1, 2.$$

and

$$(DF_G(u_1)p_2, \varphi) = (Ap_2 - b_u(u_1)p_2, \varphi) \ \ \forall \varphi \in V_G.$$

The first statement of the lemma now follows from Lemma 6.1, cf. Lemma 3.3 in [3], but we give here a proof of the stability in the $H^1$ norm. Taking $\varphi = p_2$ we have

$$(Ap_2, p_2) - (b_u(u_2)p_2, p_2) = (w, p_2).$$

Thus by Assumption (iv) and Poincare inequality we get

$$\|p_2\|_1^2 \leq \|w\|_0 \|p_2\|_0 \leq C\|w\|_0 \|p_2\|_1,$$

which for $\|w\|_0 = 1$ implies $\|p_2\|_1 \leq C$.

Next, note that for any $\varphi \in V_G$ we have

$$(Ap_2, \varphi) \quad = \quad (w + b_u(u_2)p_2, \varphi)$$

thus

$$\begin{aligned} (DF_G(u_1)(p_2), \varphi) &= (Ap_2, \varphi) - (b_u(u_1)p_2, \varphi) \\ &= (w + (b_u(u_2) - b_u(u_1))p_2, \varphi), \end{aligned}$$

hence $p_1 - p_2$ solves,

$$(DF_G(u_1)(p_1 - p_2), \varphi) = ((b_u(u_1) - b_u(u_2))p_2, \varphi). \tag{21}$$

In the case of $\|\cdot\|_X$ norm, Lemma 6.1 yields (based on the Lipschitz continuity of $b_u(.)$):

$$\begin{aligned} \max\{\|p_1 - p_2\|_1, \|p_1 - p_2\|_\infty\} &\leq C\|(b_u(u_2) - b_u(u_1))p_2\|_0 \\ &\leq C\|u_2 - u_1\|_\infty \|p_2\|_0 \\ &\leq C\|u_2 - u_1\|_X \|p_2\|_0 \leq C\|u_2 - u_1\|_X. \end{aligned}$$

Equivalently, we have

$$\|((DF_G(u_1))^{-1} - (DF_G(u_2))^{-1})w\|_X \leq C\|u_2 - u_1\|_X,$$

which completes the proof in the case of $\|\cdot\|_X$ norm.

The case of the other norm is a bit different. By $(ii)$ and $(iv)$ in Assumption 5.3, letting $\phi = p_1 - p_2$ in (21), Hölder inequality and Sobolev embeddings ($H^1(\Omega) \subset L^q(\Omega)$, $q \geq 1$ in 2D), we get

$$\begin{aligned} \|p_1 - p_2\|_1^2 &\leq C((b_u(u_2) - b_u(u_1))p_2, p_1 - p_2) \\ &\leq C(1 + \sum_{k=1,2} \|u_k\|^r_{L^{q_1 r}(\Omega)}) \|u_1 - u_2\|_{L^{q_2}(\Omega)} \|p_2\|_{L^{q_3}(\Omega)} \|p_1 - p_2\|_{L^{q_4}(\Omega)} \\ &\leq C(1 + \sum_{k=1,2} \|u_k\|_1^r) \|u_1 - u_2\|_1 \|p_2\|_1 \|p_1 - p_2\|_1 \\ &\leq C\|u_1 - u_2\|_1 \|p_1 - p_2\|_1, \end{aligned}$$

for appropriate coefficients $q_i \geq 1$ such that $\sum_{i=1}^4 \frac{1}{q_i} = 1$, and also $q_1 r \geq 1$ if $r$ from Assumption 5.3 is positive (if $r = 0$ the latter is not needed). □

In the next lemma we check if the derivative of $F$ is Lipschitz.

**Lemma 6.4** *For $u_1, u_2 \in V_h$ in a neighborhood of $u^*$ it holds*

$$\|DF(u_1) - DF(u_2)\|_{X\to Z} \quad \leq \quad C\|u_1 - u_2\|_X.$$

*Under Assumptions 5.3 the same results holds in the $H^1$ norm $\|\cdot\|_{X,1}$.*

**Proof.** For any $w, \psi \in V_h$ we have

$$\begin{aligned}((DF(u_1)w - DF(u_2))w, \psi) &= ((b_u(u_2) - b_u(u_1))w, \psi)\\ &\leq \|b_u(u_1) - b_u(u_2)\|_\infty \|w\|_0 \|\psi\|_0\\ &\leq C\|u_1 - u_2\|_\infty \|w\|_0 \|\psi\|_0.\end{aligned}$$

We utilized Assumption 5.2 (ii). Thus we can verify the statement of the lemma by taking the maximum over all $w \in V_h$ with $\|w\|_X \leq 1$:

$$\|(DF(u_1)w - DF(u_2))w\|_0 \leq C\|u_1 - u_2\|_\infty \|w\|_0 \leq C\|u_1 - u_2\|_X \|w\|_X.$$

The case of $\|\cdot\|_{X,1}$ can be proven with the help of Hölder inequality and Sobolev embedding, in a similar way to the one utilized in the proof of Lemma 6.3. □

The next lemma verifies the boundedness of the Jacobian of $\mathcal{F}$ and its coercivity in $H^1$ norm, as well as the boundedness of its inverse in $\|.\|_X$.

**Lemma 6.5** *For all $u \in V_h$ in the neighborhood of $u^*$, and all $v \in V_h$ we have*

$$\begin{aligned}&c\|v\|_1 \leq \|D\mathcal{F}(u)v\|_1 \leq C\|v\|_1\\ &\beta \leq \|D\mathcal{F}(u)\|_{X\to X}\end{aligned}$$

*where $\beta = c(1 + |\log(h)|)^{-1/2}$ for positive constants $c$, $C$ which are independent of the mesh size. Under Assumptions 5.3 the first statement of the lemma holds in the $H^1$ norm $\|\cdot\|_{X,1}$.*

**Proof.** In this proof the following notation is used:

$$\|u\|_C := (Cu, u)$$

for any symmetric and positive definite operator $C : V_h \to V_h$. Let

$$M^{-1} := \sum_G E_G (DF_G(u_G))^{-1} R_G,$$

then

$$D\mathcal{F}(u) = M^{-1}\, DF(u) = \sum_G E_g (DF_G(u_G))^{-1} R_G\, DF(u)$$

We first consider the case of $\|\cdot\|_X$ norm. Note that for $u \in V_h$ in a neighborhood of $u_h^*$ we have that $u_G = E_G F_G^{-1}(RF(u))$ is close to $u_G^*$, cf. Lemma 6.3, and thus $\|b_u(u_G)\|_\infty$ (and $\|b_u(u)\|_\infty$) is bounded by Assumption 5.2 (iii). It follows that

$$(DF_G(u_G)\psi, \psi) \quad \leq \quad C(A\psi, \psi) \;\; \forall \psi \in V_G, \tag{22}$$

$$(DF(u)\;\varphi,\; \varphi) \quad \leq \quad C(A\varphi, \varphi) \;\; \forall \varphi \in V_h. \tag{23}$$

Next note that for $u$ and $u_G$ in the neighborhood of $u^*$ we have by Assumption 5.2 (iv) $-b_u(u) \geq 0$ and $-b_u(u_G) \geq 0$, and since $DF(u) = A - b_u(u)$, the following estimates are straightforward

$$\begin{array}{rcccl}(A\varphi, \varphi) & \leq & (DF(u)\;\varphi,\; \varphi) & \leq & C(A\varphi, \varphi) \quad \forall \varphi \in V_h,\\ (A\psi, \psi) & \leq & (DF_G(u_G)\;\psi,\; \psi) & \leq & C(A\psi, \psi) \quad \forall \psi \in V_G.\end{array} \tag{24}$$

Introduce $A_G : V_G \to V_G$ defined by

$$(A_G\phi, \psi) = (\mathcal{A}\nabla\phi, \nabla\psi) \;\; \forall \phi, \psi \in V_G.$$

Then $B^{-1} := \sum_G E_G A_G^{-1} R_G$ is invertible over $V_h$ and we have

$$c(A\varphi, \varphi) \leq (B\varphi, \varphi) \leq C(A\varphi, \varphi), \quad \forall \varphi \in V_h,$$

see [2]. From the standard additive Schwarz method (ASM) theory (e.g., [11]) and (24) we get that if we replace $A_G$ by $DF_G(u_G)$, (i.e. $B^{-1}$ by $M^{-1}$) then it holds

$$(A\varphi, \varphi) \leq C(M\varphi, \varphi) \leq C(DF(u)\varphi, \varphi) \leq C(A\varphi, \varphi), \quad \forall \varphi \in V_h. \tag{25}$$

and by a standard argument, cf. e.g. [7], we get

$$c(A^{-1}\varphi, \varphi) \leq (M^{-1}\varphi, \varphi) \leq C(DF(u)^{-1}\varphi, \varphi) \leq C(A^{-1}\varphi, \varphi), \quad \forall \varphi \in V_h. \tag{26}$$

This is equivalent to say that the norms $\|u\|_M$, $\|u\|_{DF(u)}$ are equivalent to $\|u\|_A$ over $V_h$. The latter one (based on the Poincare's inequality) is equivalent to $\|u\|_1$.

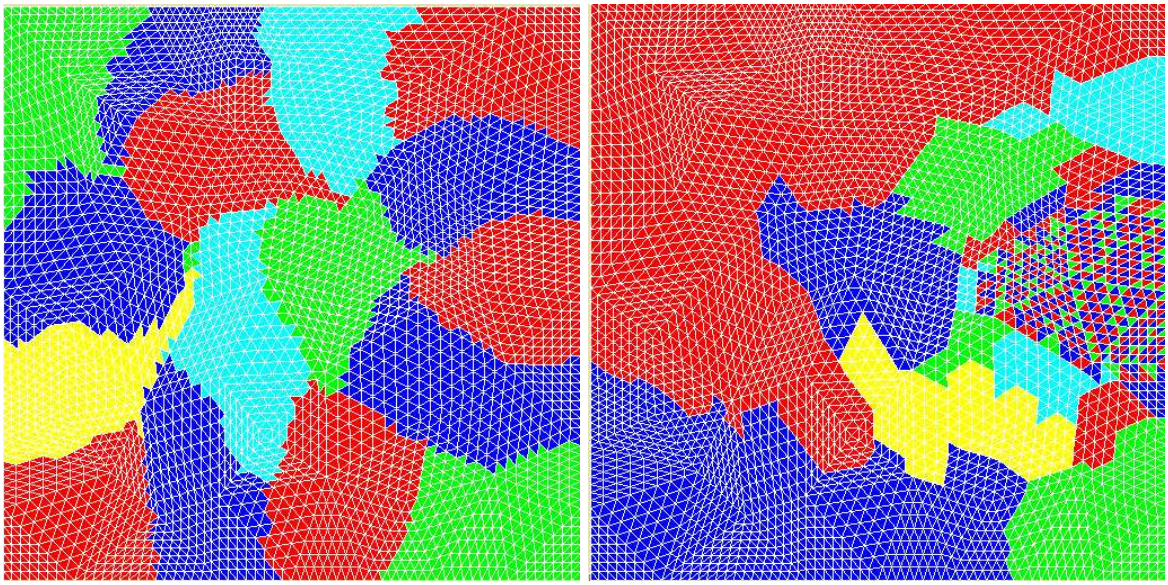

Figure 2: The left figure shows a mesh of 6,400 fine elements partitioned into 16 subdomains. Each color represents a subdomain of fine mesh elements. The right figure shows an agglomeration based coarsened away mesh containing 6,400 fine elements, 456 agglomerated elements and 400 subdomain elements.

Define $p = \mathcal{F}(u)\, v = M^{-1}DF(u)v$, for any $u$ in the neighborhood of $u^*$ and any $v \in V_h$, i.e., let $DF(u)\, v = Mp$. Then, (25) and (26) yield

$$\begin{aligned} \|v\|_1 &\leq C(Av, v) \leq C(DF(u)v, v) = C\|DF(u)v\|_{DF(u)^{-1}} \\ &\leq C\|Mp\|_{DF(u)^{-1}} \leq C\|Mp\|_{M^{-1}} = C(Mp, p) \leq C(Ap, p) \leq C\|p\|_1. \end{aligned} \tag{27}$$

The converse is also true, i.e., $\|p\|_1 \leq C\|v\|_1$. By an inverse inequality, e.g. cf. [12], and (27) we get

$$\|v\|_\infty \leq C(1 + |\log(h)|)^{1/2}\|v\|_1 \leq C(1 + |\log(h)|)^{1/2}\|p\|_X.$$

To handle the case of $H^1$ norm, we have to use Assumption 5.3, then apply Hölder inequality, use Sobolev embedding and the Poincare's inequality. Thus, we get

$$\begin{aligned}(DF(u)\phi,\phi) &\leq (A\phi,\phi)+(b_u(u)\phi,\phi)\\ &\leq (A\phi,\phi)+C(1+\|u\|^r_{L^{rp}(\Omega)})\|\phi\|^2_{L^{2q}(\Omega)}\\ &\leq (A\phi,\phi)+C(1+\|u\|^r_1)\|\phi\|^2_1\\ &\leq (A\phi,\phi)+C\|\phi\|^2_1 \leq C(A\phi,\phi)\end{aligned}$$

for any coefficients $p,\ q>0$ such that $\frac{1}{p}+\frac{1}{q}=1$, $rp\geq 1$ (if $r=0$ the latter is not needed) and any $\phi\in V$, i.e., (22) and (23) hold . The rest of the proof follows the lines of the case of $\|\cdot\|_X$. □

# 7 Numerical experiments

In this section, we report a series of numerical experiments for the model problem (1) with the functions $a, g:\ \mathbf{R}\to\mathbf{R}$ defined by

$$a(u)=\sqrt{1+u^2}\ \text{ and }\ g(u)=u^2.$$

Table 1: 400 elements, 231 degrees of freedom; *Nsub* - number of subdomains, *Nonlin iter* - number of nonlinear iterations, *Lin iter* - number of linear iterations, *Average lin* - average number of linear iterations per Newton iteration.

| Nsub | Nonliniter | Lineariter | Averagelin |
|---|---|---|---|
| 4 | 3 | 35 | 11 |
| 9 | 3 | 30 | 10 |
| 16 | 3 | 35 | 11 |
| 64 | 3 | 45 | 15 |

We are interested in the number of linear and nonlinear iterations, and how they change with respect to the number of subdomains, as well as the fine–grid mesh size (or number of fine–grid elements) .

Table 2: 1600 elements, 861 degrees of freedom; *Nsub* - number of subdomains, *Nonlin ite* - number of nonlinear iterations, *Lin ite* - number of linear iterations, *Average lin* - average number of linear iterations per Newton iteration.

| Nsub | Nonliniter | Lineariter | Averagelin |
|---|---|---|---|
| 4 | 3 | 37 | 12 |
| 9 | 3 | 42 | 14 |
| 16 | 3 | 52 | 17 |
| 64 | 4 | 61 | 15 |

The stopping criterion is to reduce the $L^2$–norm of the initial nonlinear residual by a factor of $10^{-4}$. In the following three tables we show the global number of nonlinear Newton iterations, the global number of linear iterations and the average number of linear iterations necessary for one Newton iteration for three meshes and different number of subdomains. In Table 1 we present the results for a mesh with 400 elements. Table 2 refers to a mesh with 1600 elements, whereas Table 3 stands for a mesh with 6400 elements, cf. Fig. 2, p. 18. Lastly, Table 4 presents the distribution of

Table 3: 6400 elements, 3321 degrees of freedom; *Nsub* - number of subdomains, *Nonlin iter* - number of nonlinear iterations, *Lin iter* - number of linear iterations, *Average lin* - average number of linear iterations per Newton iteration.

| Nsub | Nonliniter | Lineariter | Averagelin |
|---|---|---|---|
| 4 | 3 | 34 | 11 |
| 9 | 4 | 71 | 17 |
| 16 | 4 | 62 | 15 |
| 64 | 4 | 71 | 17 |

the degrees of freedom and number of elements corresponding to various partitioning of the original mesh with 6400 elements which gives an indication about the complexity of the subdomain solvers (shown in Table 3).

Table 4: Number of degrees of freedom and of elements of a few subdomains for mesh with 6400 elements and 3321 degrees of freedom. *NU SUB* - number of subdomains, *SN* - subdomain number, *NE* - number of elements, *NDoF* - number of degrees of freedom

| NU SUB | SN | NE | NDoF |
|---|---|---|---|
| 4 | 0 | 1699 | 952 |
| 4 | 1 | 1705 | 980 |
| 4 | 2 | 1680 | 944 |
| 4 | 3 | 1663 | 943 |
| 9 | 0 | 781 | 469 |
| 9 | 4 | 764 | 456 |
| 9 | 5 | 793 | 479 |
| 9 | 8 | 792 | 504 |
| 16 | 0 | 456 | 287 |
| 16 | 4 | 459 | 291 |
| 16 | 10 | 484 | 320 |
| 16 | 14 | 438 | 279 |
| 64 | 3 | 152 | 119 |
| 64 | 14 | 147 | 122 |
| 64 | 36 | 142 | 116 |
| 64 | 59 | 145 | 121 |

The numerical results show that overall the number of nonlinear iterations and the average number of linear iterations per nonlinear step are stable and nearly independent of the mesh size and the number of subdomains. Note that our discrete subdomain problems are obtained by the non–linear element agglomeration AMGe method summarized in Section 2 and Section 3 (originally described in [9] which is an extension of the linear one from [8]) which should explain the somewhat increase of the number of the linear iterations with the number of subdomains.

# References

[1] R. E. Bank and M. Holst, *A new paradigm for parallel adaptive meshing algorithms*, SIAM J. Sci. Comput., 22(2000), 1411-1443.

[2] R. E. Bank, P. K. Jimack, S. A. Nadeem, and S. V. Nepomnyaschikh, *A weakly overlapping domain decomposition preconditioner for the finite element solution of elliptic partial differential equations*, SIAM J. Sci. Comput., 23(2002), 1817-1841.

[3] P. N. Brown, P. S. Vassilevski, and C. S. Woodward, *On mesh-independent convergence of an inexact Newton-multigrid algorithm*, SIAM J. Scientific Computing, 25(2003) pp. 570–590.

[4] X.-C. Cai and D. E. Keyes, *Nonlinearly preconditioned inexact Newton algorithms*, SIAM J. Sci. Comput., 24 (2002), 183-200.

[5] X.-C. Cai, D. E. Keyes, and L. Marcinkowski, *Nonlinear additive Schwarz preconditioners and applications in computational fluid dynamics,* Int. J. Numer. Meth. Fluid Mech., 40 (2002), 1463-1470.

[6] J. E. Dennis Jr. and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Classics in Applied Mathematics, Vol. 16, SIAM, Philadelphia, 1996.

[7] E. G. D'yakonov, *Optimization in Solving Elliptic Problems*, CRC Press, Boca Raton, FL, 1996. Translated from the 1989 Russian original, Translation edited and with a preface by Steve McCormick.

[8] J. E. Jones and P. S. Vassilevski, *AMGe based on element agglomeration*, SIAM J. Sci. Comput., 23(2001), 109-133.

[9] J. E. Jones, P. S. Vassilevski, and C. S. Woodward, *Nonlinear Schwarz-FAS methods for unstructured finite element problems,* in the Proceedings of the Second M.I.T. Conference on Computational Fluid and Solid Mechanics, Cambridge, MA, June 17-20, 2003. Elsevier, vol. 2, pp. 2008-2011.

[10] G. Karypis and V. Kumar, *METIS: A family of multilevel partitioning algorithms*, available at: *http://www-users.cs.umn.edu/karypis/metis/.*

[11] B. Smith, P. Bjørstad, and W. Gropp, *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*, Cambridge University Press, 1996.

[12] V. Thomée, *Galerkin finite element methods for parabolic problems*, vol. 1054 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1984.

[13] P. S. Vassilevski, *Sparse matrix element topology with application to AMG and preconditioning*, Numer. Lin. Alg. Applic., 9(2002), 429-444.

[14] A. Ženíšek, *Nonlinear elliptic and evolution problems and their finite element approximations.* Computational Mathematics and Applications. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1990. With a foreword by P.-A. Raviart.