The AL Basis for the solution of elliptic problems in heterogeneous media.

L. Grasedyck* I. Greff† S. Sauter‡

January 17, 2011

In this paper, we will show that, for elliptic problems in heterogeneous media, there exists a local (generalized) finite element basis (AL basis) consisting of $O\left(\left(\log \frac{1}{h}\right)^{d+1}\right)$ basis functions per nodal point such that the convergence rates of the classical finite element method for Poisson-type problems are preserved. We provide several numerical examples beyond our theory, where even $O(1)$ basis functions per nodal point are sufficient to preserve the convergence rates.

1 Introduction

The efficient numerical modelling of elliptic problems in heterogeneous media is of fundamental and practical importance and arises in applications such as composite materials, porous media and turbulent transport. If the geometric details, e.g., inclusions in the material, have complicated structure and/or are tiny, then the resolution of all details by conventional finite elements becomes too costly — especially for three-dimensional problems.

In the recent years, many types of *generalized* finite element methods have been developed where the characteristic physical behavior of the solution is incorporated in the *shape* of the trial functions so that the geometric details may not be resolved by the finite element mesh while the goal is to preserve the asymptotic convergence rates also for these coarse-scale discretizations. Early papers on this topic are [1], [2]. We omit an extensive list of references on the construction of generalized finite element methods because our goal here is *not* to introduce a new method but to discuss the general theoretical question: What is the minimal dimension of a generalized finite element space for elliptic problems with rough coefficients such that the "textbook"-convergence of polynomial finite elements is preserved.

*Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Templergraben 55, D-52056 Aachen, Germany, e-mail: lgr@igpm.rwth-aachen.de

†Laboratoire de Mathématiques Appliquées, Université de Pau et des pays de l'Adour, F-64013 Pau Cedex, France, e-mail: isabelle.greff@univ-pau.fr

‡Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, e-mail: stas@math.uzh.ch

In this paper we shall deal with the following problem. Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain and let the diffusion matrix $A \in L^\infty\left(\Omega, \mathbb{R}^{d\times d}_{\mathrm{sym}}\right)$ be uniformly elliptic:

$$0 < \alpha = \operatorname*{essinf}_{x\in\Omega} \inf_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\langle Av, v\rangle}{\langle v, v\rangle} \;\leq\; \operatorname*{esssup}_{x\in\Omega} \sup_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\langle Av, v\rangle}{\langle v, v\rangle} = \beta < \infty. \tag{1}$$

For $m \in \mathbb{N}_0$, let $H^m(\Omega)$ denote the usual Sobolev space with norm $\|\cdot\|_{H^m(\Omega)}$ and let $H^m_0(\Omega)$ be the closure of $C^\infty_0(\Omega)$ with respect to the norm $\|\cdot\|_{H^m(\Omega)}$. The dual space of $H^m_0(\Omega)$ is denoted by $H^{-m}(\Omega)$.

For given $f \in L^2(\Omega)$, we are seeking $u \in H^1_0(\Omega)$ such that

$$a(u,v) := \int_\Omega \langle A\nabla u, \nabla v\rangle = \int_\Omega fv =: F(v) \qquad \forall v \in H^1_0(\Omega). \tag{2}$$

The abstract conforming Galerkin method to this problem is given by specifying a finite-dimensional subspace $S \subset H^1_0(\Omega)$ and seeking $u_S \in S$ such that

$$a(u_S, v) = F(v) \qquad \forall v \in S.$$

For conventional finite elements the space S is based on a regular (in the sense of Ciarlet [5]) finite element mesh $\mathcal{G} = \{\tau_i : 1 \leq i \leq N\}$ consisting of shape-regular simplices τ_i. The mesh width is given by $H = \max\{\operatorname{diam}\tau : \tau \in \mathcal{G}\}$.

We assume that Ω is polygonal (polyhedral in 3-d) and $\mathcal{G}$ defines a disjoint covering of Ω. The space of continuous, piecewise linear finite elements is given by

$$S = \left\{u \in H^1_0(\Omega) \mid \forall \tau \in \mathcal{G} : u_\tau \in \mathbb{P}_1\right\}. \tag{3}$$

Let b_i denote the usual local nodal basis of S ("hat functions") and their support is denoted by

$$\omega_i := \operatorname{supp} b_i. \tag{4}$$

It is well known that — as long as the mesh $\mathcal{G}$ does not resolve the (possible) discontinuities and oscillations of A — the "textbook"-convergence rates of linear finite elements are substantially reduced.

In this paper we will address the following question: Is there a set of basis functions $b_{i,j} \in H^1_0(\Omega)$, $1 \leq j \leq p$, $1 \leq i \leq n := \dim(S)$, such that

$$\operatorname{supp} b_{i,j} \subset \omega_i$$

and the *linear convergence property* (cf. Definition 1) holds?

Definition 1 (Linear Convergence Property) *Let $a(\cdot,\cdot)$ be as in (2) and S be as in (3) with supports ω_i of basis functions as in (4). Let $\tilde{S} \subset H^1_0(\Omega)$ be a finite dimensional subspace which satisfies*

$$\tilde{S} = \operatorname{span}\left\{b_{i,j} \mid 1 \leq j \leq p, 1 \leq i \leq n \text{ and } \operatorname{supp} b_{i,j} \subset \omega_i\right\}. \tag{5}$$

*$\tilde{S}$ has the **linear convergence property** (LCP) if, for any $f \in L^2(\Omega)$ the solution to the problem of finding $u_{\tilde{S}} \in \tilde{S}$ such that*

$$a\left(u_{\tilde{S}}, v\right) = \int_\Omega fv \qquad \forall v \in \tilde{S}$$

satisfies the error estimate

$$\left\|u - u_{\tilde{S}}\right\|_{H^1(\Omega)} \leq CH\left\|f\right\|_{L^2(\Omega)},$$

where C only depends on α and β (cf. (1)).

Remark 2 *Note that the linear convergence property is defined for a* given *set of supports ω_i. More generally, one could also include an optimal choice of these supports in the definition. This would be appropriate if the (possibly low) regularity of the solution is* not *distributed uniformly over the domain. Our simplified definition is suitable for problems where the diffusion coefficient is rough/oscillating over the whole domain and a quasi-uniform mesh $\mathcal{G}$ is an adequate choice.*

We will prove a) theoretically that, by choosing the number p in (5) proportionally to $O\left(\log^{d+1}\frac{1}{H}\right)$, such a set of basis functions exists and b) by numerical experiments that there exist problems of the form (2) with very complicated diffusion matrix where, still, the choice $p = O(1)$ is sufficient.

We emphasize that our construction of $b_{i,j}$ is very costly and we consider our results rather as a theoretical insight than a practical method. Forthcoming papers will address the question how to construct the basis $b_{i,j}$ efficiently.

For problems with periodic coefficient $A = A_0\left(\frac{\cdot}{\varepsilon}\right)$ or locally periodic coefficient $A(\cdot) = A_0\left(\cdot, \frac{\cdot}{\varepsilon}\right)$ (with slowly varying functions $A_0(\cdot)$ resp. $A_0(\cdot,\cdot)$) the convergence of finite elements, generalized finite elements as well as methods based on homogenization-type techniques has been analyzed in the literature (see, e.g., [11], [8] – for a more general, non-periodic setting see [12]). In contrast, we do not impose any assumption on the structure of A but only assume uniform ellipticity and continuity for the corresponding bilinear form. Our results rely strongly on the approximability of the Green's function for general elliptic problems (see [3]).

2 The AL Basis

Let $L : H_0^1(\Omega) \to H^{-1}(\Omega)$ denote the operator associated to the bilinear form $a(\cdot,\cdot)$. Let the standard finite element basis be denoted by b_i (cf. (4)) and set

$$B_i := L^{-1} b_i, \qquad i \in \mathcal{I} := \{1, 2, \ldots, n\}.$$

Recall the definition of the supports ω_i as in (4).

Remark 3 *The condition* $\operatorname{supp} b_{i,j} \subset \omega_i$ *in (5) on the localness of the basis functions implies the sparsity of the arising stiffness matrix and is crucial for the computational/storage complexity of the discretization. Without this condition, the basis B_i would be a very good choice preserving the optimal error estimates. However, the functions B_i, in general, are non-local and the generation of the system matrix would be prohibitively expensive.*

Next, we define simplex layers around ω_i by the following recursion. Let $\omega_{i,0} := \omega_i$ and define, for $j = 0, 1, 2, \ldots$,

$$\omega_{i,j+1} := \bigcup \{\bar{\tau} \mid \tau \in \mathcal{G} \text{ and } \omega_{i,j} \cap \bar{\tau} \neq \emptyset\}.$$

We set $\omega_i^\star := \omega_{i,m}$, where m is chosen such that

$$\eta \operatorname{diam} \omega_i \leq \operatorname{dist}(\omega_i, \partial\omega_i^\star) \tag{6}$$

for some sufficiently small $0 < \eta = O(1)$. As an assumption on the local quasi-uniformity of the mesh we assume that $m = m(\eta) \leq C$.

To reduce technicalities we assume that, for all $1 \leq i \leq m$, the sets ω_i and $\omega_i^\star$ are convex.

For an index $i \in \mathcal{I}$, we define a nearfield and a farfield by setting[1]

$$\mathcal{I}_i^{\text{near}} := \{j \in \mathcal{I} \mid \ 0 < |\omega_i^\star \cap \operatorname{supp} b_j|\} \quad \text{and} \quad \mathcal{I}_i^{\text{far}} := \mathcal{I} \backslash \mathcal{I}_i^{\text{near}}.$$

Then, we set

$$X_i^{\text{far}} := \operatorname{span}\left\{ B_j|_{\omega_i^\star} \mid \ j \in \mathcal{I}_i^{\text{far}} \right\}$$

and

$$V_i^{\text{near}} := \operatorname{span}\left\{ b_i B_j \mid \ j \in \mathcal{I}_i^{\text{near}} \right\}.$$

Note that the functions in X_i^{far} are L-harmonic in $\omega_i^\star$, i.e., any $v \in X_i^{\text{far}}$ satisfies

$$\int_{\omega_i^\star} \langle A \nabla v, \nabla w \rangle = 0 \qquad \forall w \in H_0^1\left(\omega_i^\star\right).$$

It turns out that the space X_i^{far} can be approximated by a low dimensional space and we employ the construction which has been proposed in [3].

We introduce intermediate layers between ω_i and $\omega_i^\star$, by setting $r_{i,0} := \operatorname{dist}\left(\omega_i, \partial\omega_i^\star\right)$ and

$$r_{i,j} := \left(1 - \frac{j}{\ell}\right) r_{i,0} \qquad 1 \leq j \leq \ell,$$

where ℓ will be fixed later. The intermediate layers are given by

$$D_{i,j} := \left\{x \in \omega_i^\star \mid \operatorname{dist}\left(x, \omega_i\right) \leq r_{i,j}\right\}, \qquad 0 \leq j \leq \ell,$$

and satisfy $\omega_i = D_{i,\ell} \subset D_{i,\ell-1} \subset \ldots \subset D_{i,0} \subset \omega_i^\star$. For $\rho > 0$, let $\mathcal{G}_\rho$ denote a Cartesian tensor mesh on $\mathbb{R}^d$ consisting of d-dimensional elements (hypercubes) with side lengths ρ and let

$$\mathcal{G}_{i,j} := \left\{ D_{i,j} \cap \tau \;\middle|\; \tau \in \mathcal{G}_\rho,\ \rho := \frac{\operatorname{diam} D_{i,j}}{k} \right\},$$

where $k \in \mathbb{N}_{\geq 1}$ will be fixed later. For $t \in \mathcal{G}_{i,j}$, we denote the characteristic function for t by $\chi_t : \Omega \to \mathbb{R}$. We define

$$\widetilde{V}_{i,j}^{\text{far}} := \operatorname{span}\left\{ (P\chi_t)|_{\omega_i} \mid t \in \mathcal{G}_{i,j} \right\},$$

where $P : L^2(\Omega) \to X_i^{\text{far}}$ is the L^2-orthogonal projection. Then,

$$\widetilde{V}_i^{\text{far}} := \widetilde{V}_{i,0}^{\text{far}} + \widetilde{V}_{i,1}^{\text{far}} + \ldots + \widetilde{V}_{i,\ell}^{\text{far}} \tag{7}$$

and, finally,

$$V_i^{\text{far}} := \left\{ b_i v \mid v \in \widetilde{V}_i^{\text{far}} \right\}.$$

Obviously, we have

$$V_i^{\text{far}} \subset H_0^1(\Omega), \qquad \dim V_i^{\text{far}} \leq \sum_{j=0}^{\ell} \#\mathcal{G}_{i,j} \leq \sum_{j=0}^{\ell} k^d = (\ell+1)\, k^d.$$

[1] For a measurable subset $M \subset \mathbb{R}^d$, we set $|M| := \int_M 1$.

Definition 4 (AL basis) *For any support ω_i (cf. (4)) the set of AL basis functions consists of the functions $b_i B_j$, $j \in \mathcal{I}_i^{\text{near}}$, and of the functions*

$$b_i P \chi_t \qquad \forall t \in \mathcal{G}_{i,q} \qquad 0 \leq q \leq \ell.$$

The general notation is $b_{i,j}$, $1 \leq j \leq p$, $1 \leq i \leq n$, where $p := \dim\left(V_i^{\text{far}} + V_i^{\text{near}}\right)$. The corresponding generalized finite element space is given by

$$V_{\text{AL}} := \left(V_1^{\text{near}} + V_1^{\text{far}}\right) + \left(V_2^{\text{near}} + V_2^{\text{far}}\right) + \ldots + \left(V_n^{\text{near}} + V_n^{\text{far}}\right).$$

Remark 5 *Since the index m in the definition of $\omega_i^\star$ is independent of H, we have $\dim V_i^{\text{near}} = O(1)$. As a consequence of the error analysis it will turn out that $\dim V_i^{\text{far}} = O\left(\log^{d+1} \frac{1}{H}\right)$.*

The Galerkin discretization for the generalized finite element space V_{AL} is given by seeking $u_{\text{AL}} \in V_{\text{AL}}$ such that

$$a(u_{\text{AL}}, v) = F(v) \qquad \forall v \in V_{\text{AL}}. \tag{8}$$

3 Error Analysis

The error analysis is based on the results in [3]. The constants in the error estimates of this section depend on α and $\beta \in \mathbb{R}_{>0}$ without writing this dependence explicitly. Our emphasis is to prove that the estimates are uniform for all diffusion matrices $A \in L^\infty\left(\Omega, \mathbb{R}^{d\times d}_{\text{sym}}\right)$ which satisfy (1). Note that the assumptions on A imply

$$\left\|L^{-1}\right\|_{H_0^1(\Omega) \leftarrow H^{-1}(\Omega)} \leq C.$$

Assumption 6 *The domains ω_i, $\omega_i^\star$ (cf. (4) and (6)) are convex and satisfy (6) for some $\eta \gtrsim 1$.*

The constant

$$C_\sharp := \max_{i\in\mathcal{I}} \#I_i^{\text{near}},$$

*depends only on the shape-regularity of the finite element mesh $\mathcal{G}$ and the number $m = O(1)$ (depending on the **local** quasi-uniformity of $\mathcal{G}$) in the definition of $\omega_i^\star$.*

Finally, there exists a constant C_q such that

$$\#\mathcal{I} \leq C_q H^{-d}.$$

Theorem 7 *Let u denote the solution of (2). Let the parameters ℓ and k in the definition of the farfield part of V_{AL} be chosen according to*

$$\ell := \max\left\{2, \left\lceil \frac{2+d}{2\log 2} \log \frac{1}{H} \right\rceil\right\} \quad \text{and} \quad k := \left\lceil \frac{2c_0\ell^2}{(\ell-1)} \right\rceil$$

for some $c_0 = O(1)$. Let u_{AL} be the corresponding Galerkin solution (cf. (8)). Then, the error estimate

$$\|u - u_{\text{AL}}\|_{H^1(\Omega)} \leq CH \|f\|_{L^2(\Omega)}$$

holds while

$$\dim V_{\text{AL}} \leq C_d N \ell^{d+1} \leq \tilde{C}_d H^{-d} \log^{d+1} \frac{1}{H}.$$

Proof. Let $P_S : L^2(\Omega) \to S$ denote the L^2-orthogonal projection onto S. For $f \in L^2(\Omega)$, let $u = L^{-1}f$. Then, the substitution of f by $P_S f$ leads to a consistent perturbation

$$\left\|u - L^{-1}P_S f\right\|_{H^1(\Omega)} \leq C \left\|f - P_S f\right\|_{H^{-1}(\Omega)} \leq CH \left\|f\right\|_{L^2(\Omega)}. \tag{9}$$

We introduce the nearfield and the farfield parts of f with respect to some $i \in \mathcal{I}$ by

$$f_i^{\text{near}} := \sum_{j \in \mathcal{I}_i^{\text{near}}} (P_S f)_j \, b_j \quad \text{and} \quad f_i^{\text{far}} := \sum_{j \in \mathcal{I}_i^{\text{far}}} (P_S f)_j \, b_j,$$

where $(P_S f)_j := (P_S f)(x_j)$ and x_j is the nodal point corresponding to b_j. Then,

$$L^{-1}P_S f = \sum_{i=1}^{n} \underbrace{b_i L^{-1} f_i^{\text{near}}}_{u_i^{\text{near}}} + \sum_{i=1}^{n} b_i \underbrace{L^{-1} f_i^{\text{far}}}_{u_i^{\text{far}}}.$$

Since $u_i^{\text{near}} \in V_i^{\text{near}}$, the approximation problem is reduced to the approximation of u_i^{far}. Note that the function $u_i^{\text{far}}\big|_{\omega_i^\star} \in X_i^{\text{far}}$. As a consequence of the approximation results in [3] (for the details see [4, Lemma 4, Proof of Theorem 5 by choosing $q \leftarrow 1/2$ and $p \leftarrow \ell$ therein.]) there exists $\tilde{u}_i^{\text{far}} \in \widetilde{V}_i^{\text{far}}$ (cf. (7)) such that

$$\left\|u_i^{\text{far}} - \tilde{u}_i^{\text{far}}\right\|_{H^m(\omega_i)} \leq CH^{s-m} \left\|\nabla L^{-1} f_i^{\text{far}}\right\|_{L^2(\omega_i)} \qquad m = 0, 1 \tag{10}$$

with $s = 2 + d/2$. The approximation of u finally is given by

$$\tilde{u} := \sum_{i=1}^{n} u_i^{\text{near}} + \sum_{i=1}^{n} b_i \tilde{u}_i^{\text{far}} \in V_{\text{AL}}.$$

By using (9) and a triangle inequality we obtain

$$\left\|u - \tilde{u}\right\|_{H^1(\Omega)} \leq CH \left\|f\right\|_{L^2(\Omega)} + \left\|\sum_{i=1}^{n} b_i \left(u_i^{\text{far}} - \tilde{u}_i^{\text{far}}\right)\right\|_{H^1(\Omega)}.$$

The second sum can be estimated by combining the Leibniz rule for products with a triangle inequality, a Hölder's inequality, an inverse inequality for b_i, and (10):

$$\begin{aligned}
\left\|\sum_{i=1}^{n} b_i \left(u_i^{\text{far}} - \tilde{u}_i^{\text{far}}\right)\right\|_{H^1(\Omega)} &\leq \sum_{i=1}^{n} \Bigg(\left\|b_i\right\|_{L^\infty(\omega_i)} \left\|u_i^{\text{far}} - \tilde{u}_i^{\text{far}}\right\|_{H^1(\omega_i)} + \\
&\qquad \left\|\nabla b_i\right\|_{L^\infty(\omega_i)} \left\|u_i^{\text{far}} - \tilde{u}_i^{\text{far}}\right\|_{L^2(\omega_i)} \Bigg) \\
&\leq CH^{s-1} \sum_{i=1}^{n} \left\|\nabla L^{-1} f_i^{\text{far}}\right\|_{L^2(\omega_i)} \\
&\leq CH^{s-1} \sum_{i=1}^{n} \left(\left\|\nabla L^{-1} P_S f\right\|_{L^2(\omega_i)} + \left\|\nabla L^{-1} f_i^{\text{near}}\right\|_{L^2(\omega_i)} \right) \\
&\leq CH^{s-1}\sqrt{n} \left(\left\|f\right\|_{L^2(\Omega)} + \sqrt{\sum_{i=1}^{n} \left\|\nabla L^{-1} f_i^{\text{near}}\right\|_{L^2(\omega_i)}^2} \right).
\end{aligned}$$

In order to estimate the last sum we use the representation of L^{-1} via the Green's function

$$L^{-1} f_i^{\text{near}} = \int_\Omega G(x,y)\, f_i^{\text{near}}(y)\, dy,$$

where the estimate

$$\sup_{x\in\Omega} \|\nabla G(x,y)\|_{L^\rho(\Omega)} \le C_{d,\alpha,\beta,\varepsilon} \quad \text{with} \quad \alpha,\beta \text{ as in (1)}, \rho := \frac{d}{d-1} - \varepsilon$$

for any $0 < \varepsilon \le \frac{1}{d-1}$ follows from [9, Theorem 1.1 and (1.12)] for $d \ge 3$ and from [7, Remark 2.19] for $d = 2$. For $d = 1$ the estimate

$$\sup_{x\in\Omega} \left\|G'(x,y)\right\|_{L^\infty(\Omega)} \le C_{\alpha,\beta}$$

follows from [10, (10.14)]. In the following we work out only the case $d \ge 2$ while the case $d = 1$ can be derived analogously. Hence,

$$\begin{aligned}\left\|\nabla L^{-1} f_i^{\text{near}}\right\|^2_{L^2(\omega_i)} &\le \int_{\omega_i} \left| \int_\Omega \nabla_x G(x,y)\, f_i^{\text{near}}(y)\, dy \right|^2 dx \\ &\le C^2_{d,\alpha,\beta,\varepsilon} |\omega_i| \, \|f_i^{\text{near}}\|^2_{L^p(\Omega)} \le C^2_{d,\alpha,\beta,\varepsilon} H^d \, \|f_i^{\text{near}}\|^2_{L^p(\Omega)}\end{aligned}$$

for $p = \frac{d+\varepsilon(1-d)}{1+\varepsilon(1-d)} \ge 2$. From [6, Proposition 3.10 (choosing $p' \leftarrow p$, $p \leftarrow 2$, $\alpha \leftarrow 0$ therein)] we conclude that

$$\|f_i^{\text{near}}\|^2_{L^p(\Omega)} \le H^{-\zeta} \|f_i^{\text{near}}\|^2_{L^2(\Omega)}, \qquad \zeta := \frac{d(d-\varepsilon(1-d)-2)}{d+\varepsilon(1-d)}$$

so that

$$\left\|\nabla L^{-1} f_i^{\text{near}}\right\|^2_{L^2(\omega_i)} \le C^2_{d,\alpha,\beta,\varepsilon} H^{2-2q} \|f_i^{\text{near}}\|^2_{L^2(\omega_i^\star)} \le C^2_{d,\alpha,\beta,\varepsilon} H^{2-2q} \|P_S f\|^2_{L^2(\omega_i^\star)},$$

where $q := \varepsilon \frac{(d-1)^2}{d-\varepsilon(d-1)}$. Hence,

$$\sum_{i=1}^n \left\|\nabla L^{-1} f_i^{\text{near}}\right\|^2_{L^2(\omega_i)} \le C^2_{d,\alpha,\beta,\varepsilon} H^{2-2q} \sum_{i=1}^n \|P_S f\|^2_{L^2(\omega_i^\star)} \le C^2_{d,\alpha,\beta,\varepsilon} H^{2-2q} \|f\|^2_{L^2(\Omega)}.$$

In summary, we have proved (by using $q \le 1$ for all $\varepsilon \in [0, 1/(d-1)]$)

$$\begin{aligned}\|u - \tilde{u}\|_{H^1(\Omega)} &\le CH \|f\|_{L^2(\Omega)} + CH^{s-1}\sqrt{n}\left(\|f\|_{L^2(\Omega)} + H^{1-q}\|f\|_{L^2(\Omega)}\right) \\ &\overset{\text{Assumpt. 6}}{\le} C\left(H + \sqrt{C_q} H^{s-1-d/2}\right)\|f\|_{L^2(\Omega)}\end{aligned}$$

and the choice of s yields the assertion. ■

4 Numerical Experiment

We have shown theoretically that $O\left(\log^{d+1}\frac{1}{H}\right)$ degrees of freedom per nodal point are sufficient for problems with general L^∞ diffusion coefficients in order to obtain linear convergence rates with respect to the $H^1(\Omega)$-norm.

In this section, we will investigate whether our complexity estimates are sharp for certain (complicated) choices of the diffusion coefficient $A(x)$ (which in some cases violate

the assumptions on uniform boundedness on $A(\cdot)$). Interestingly, for all considered examples which satisfy assumption (1) uniformly in α and β, only 1 degree of freedom per coarse grid nodal point is sufficient for optimal convergence rates if the basis functions are properly designed. We have tested many choices of $A(\cdot)$ – however we will describe only three characteristic examples here for sake of brevity and explain the (heuristic) reasons for these choices in the following.

Choice of Right-Hand Sides

As a subset of all possible right-hand sides we consider real valued plane waves of the form $f_j := \sin(2\pi \langle \xi_j, \cdot \rangle)$, $\xi_j = \left(\sin \frac{\pi j}{20}, \cos \frac{\pi j}{20}\right)^\intercal$, $j = 1, \ldots, 20$.

We expect (and observe) that these choices of right-hand sides generate solutions which exhibit very different directions of oscillations so that the use of only four shape functions per element is a critical test for approximability.

Generalized Finite Element Space and Error Measure

Recall that for the Poisson problem – discretized by continuous, piecewise bilinear elements – the *textbook* convergence order for the relative $L^2(\Omega)$-error is given by $O\left(h^2\right)$ and for the relative $H^1(\Omega)$-error by $O(h)$. For simplicity we consider the $L^2(\Omega)$-error rather than the $H^1(\Omega)$-error. Although we have not considered a periodic example (in order to avoid special effects which, possibly, only arise in a periodic setting), the oscillatory behavior of the diffusion coefficient for all the numerical examples is (relatively) uniformly distributed over the domain – that is the reason why we employed a *uniform* Cartesian grid of mesh width H for setting up the (generalized) finite element spaces.

To make the construction of the generalized finite element space $\tilde{S}$ computable we choose a sufficiently small mesh width $h = H^2$ and denote by S_h the usual linear $\mathbb{P}_1$ finite element space on the fine mesh. The subspace $\tilde{S}$ is chosen as a subspace of S_h. Let u denote the Galerkin solution for the fine mesh S_h which we use as the "nearly exact" reference solution and let $u_{\tilde{S}}$ be the Galerkin solution for the space $\tilde{S}$. The relative error is given by $\left\|u - u_{\tilde{S}}\right\|_{L^2(\Omega)} / \|u\|_{L^2(\Omega)}$. For our numerical experiments, it turns out that the size of the local norms are distributed uniformly over the domain, i.e., they satisfy $\|u\|_{L^2(\tau)} \sim H \|u\|_{L^2(\Omega)} \sim H$ and $\|u - u_S\|_{L^2(\tau)} \sim H \left\|u - u_{\tilde{S}}\right\|_{L^2(\Omega)}$ for all $\tau \in \mathcal{G}_H$. This justifies to investigate the convergence of the *local* relative L^2-error $\|u - u_S\|_{L^2(\tau)} / \|u\|_{L^2(\tau)}$ for all fine grid Galerkin solutions $u \in S_h$ corresponding to the right-hand sides f_j. From the numerical point of view this is a much simpler task. Hence, we construct only the local basis functions for a single cell

$$\tau := [\frac{1}{2} - H, \frac{1}{2}] \times [\frac{1}{2} - H, \frac{1}{2}] \in \mathcal{G}_H.$$

The local space $\tilde{S}|_\tau$ is obtained by an SVD of the fine grid solutions as follows.

Let the vector c_j be the coefficient vector of the Galerkin solution $u_j \in S_h$ with respect to the standard bilinear "hat" basis on the fine mesh for the right-hand side f_j. We denote by $c_{j,\tau}$ the restriction of c_j to the values at vertices of the cell τ. We define the matrix A columnwise

$$A := [c_{1,\tau} \mid \cdots \mid c_{20,\tau}]$$

and compute the left and right singular vectors v_j, w_j as well as the corresponding

singular values $\sigma_{\tau,j}$ of A. By the inequality

$$\|A - \sum_{r=1}^{k} v_r \sigma_{\tau,r} w_r^\intercal\|_2 \leq \sigma_{\tau,k+1}.$$

We conclude that

$$\|c_{j,\tau} - \tilde{c}_{j,\tau}\|_2 \leq \sigma_{\tau,k+1}, \quad \text{where} \quad \tilde{c}_{j,\tau} := \sum_{r=1}^{k} v_r \sigma_{\tau,r} (w_r)_j$$

holds. This says that the first k left singular vectors v_r of A define a k-dimensional space $S_\tau := \text{span}\{v_r \mid r = 1, \ldots, k\}$ for the cell τ such that we can approximate the τ-parts of the coefficients of all solutions u_j in S_τ up to an error of size $\sigma_{\tau,k+1}$. Note that quadratic convergence with respect to the relative $L^2(\tau)$-norm is equivalent to quadratic convergence of the ratio $\sigma_{\tau,5}/\sigma_{\tau,1}$ as a function of H.

Decay of singular values

In this section we consider three numerical examples, all of them for the domain $\Omega = [-1,1]\times[-1,1]$ and the bilinear form a as defined in (2). We compute the singular values for the cell τ and investigate their decay behavior. In the first experiment (**Problem 1**) the diffusion coefficient is oscillatory:

$$A_1(x,y) = \nu(x,y)I \quad \text{with} \quad \nu(x,y) := 2 - \frac{\cos(2\pi x^2/H) + \cos(2\pi y/H)}{11\left(\frac{1}{10} + (x+y)^2\right)}$$

and I denoting the 2×2 identity matrix. Note that this coefficient satisfies assumption (1) with $\alpha = \frac{2}{11}$ and $\beta = \frac{42}{11}$.

For comparison we have used standard P1 finite elements and a sufficient quadrature (regular refinement to fine-scale) for the setup of the stiffness matrix in order to compute the approximate solution u_{P1}.

The results in Table 1 show that for Problem 1 with fine scale oscillations the optimal shape functions preserve the quadratic convergence rate (cf. Figures 1,2 for a plot of the four shape functions), whereas the $P1$ finite elements are not sufficient for quadratic or even linear convergence rates.

Problem 1

$H =$	**AL**	**Ratio**	**P1**	**Ratio**
0.25	1.59_{e-2}		4.65_{e-2}	
0.125	7.76_{e-3}	2.04	2.75_{e-2}	2.91
0.0625	1.82_{e-3}	4.26	1.83_{e-2}	2.18
0.03125	4.44_{e-4}	4.10	1.59_{e-2}	1.15

Table 1: Convergence rates for the optimal shape functions (AL) and standard P1 shape functions (P1) for a non-periodic oscillating coefficient.

In the second experiment (**Problem 2**) we consider a singularly perturbed diffusion coefficient:

$$A_2(x,y) = \text{diag}(\delta(x,y), H^2), \quad \delta(x,y) := \frac{3}{2} + \sin(\frac{2\pi x^2}{H^{3/2}}),$$

where now α becomes small as $H \to 0$.

Problem 2

$H =$	**AL**	**Ratio**	**P1**	**Ratio**
0.25	5.75_{e-3}		1.93_{e-1}	
0.125	1.97_{e-3}	2.92	2.06_{e-1}	0.94
0.0625	9.04_{e-4}	2.18	2.23_{e-1}	0.92
0.03125	2.93_{e-4}	3.09	2.36_{e-1}	0.94

Table 2: Convergence rates for the optimal shape functions (AL) and standard P1 shape functions (P1) for a singularly perturbed Problem.

In the case of the singularly perturbed Problem 2 we observe in Table 2 at least linear convergence. Note that the coefficient A in this case is not uniformly elliptic as $H \to 0$ and, hence, the assumptions for the theory are violated. This was the only example, where the convergence rates are found to be clearly less than quadratic.

In the last experiment (**Problem 3**) we choose for each $\tau_i, 1 \leq i \leq N$, a random coefficient

$$A_3|_{\tau_i} \equiv \gamma I \quad \text{with } \gamma \text{ drawn randomly (uniform) from } \{0.01, 0.1, 1, 10, 100\}.$$

The relative approximation errors are averaged over 10 random samples of A_3.

The results in Table 3 show that even for this medium-contrast random coefficient the rate of convergence is quadratic.

Problem 3

$H =$	**Avg**	**Ratio**	**Min**	**Max**
0.25	3.10_{e-1}		1.80_{e-1}	4.06_{e-1}
0.125	7.80_{e-2}	3.97	6.20_{e-2}	9.17_{e-2}
0.0625	1.84_{e-2}	4.24	1.63_{e-2}	2.04_{e-2}
0.03125	4.33_{e-3}	4.25	4.22_{e-3}	4.48_{e-3}

Table 3: Convergence rates for the optimal shape functions for a random diffusion coefficient (average, minimal and maximal error from 10 samples).

The numerical experiments were conducted with the HLib (`http://www.hlib.org`) software library. We conclude that for all test examples only four (properly selected)

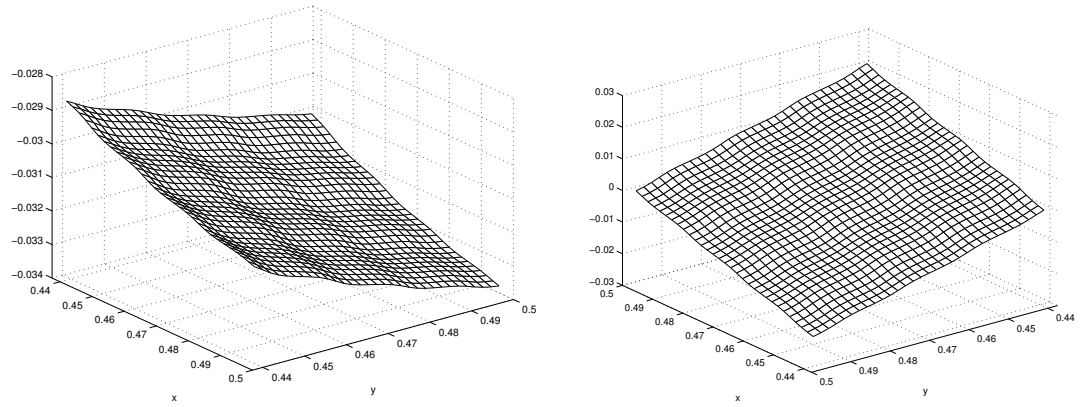

Figure 1: The first and second shape function for the oscillatory coeff. of Problem 1.

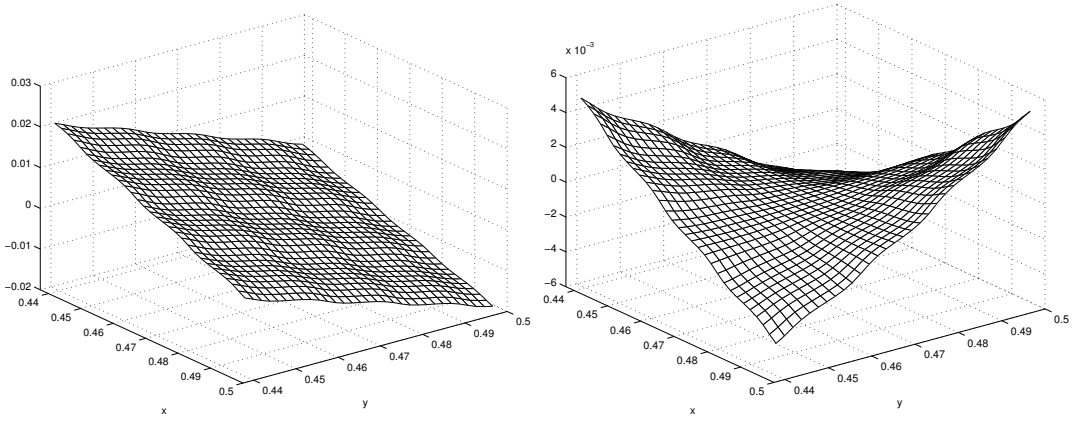

Figure 2: The third and fourth shape function for the oscillatory coeff. of Problem 1.

basis functions per cell preserve the convergence with respect to the relative L^2 norm even for cases where the diffusion coefficient is rather complicated.

References

[1] I. Babuška, G. Caloz, and J. E. Osborn. Special finite element methods for a class of second order elliptic problems with rough coefficients. *SIAM J. Numer. Anal.*, 31(4):945–981, 1994.

[2] I. Babuška and J. E. Osborn. Generalized finite element methods: their performance and their relation to mixed methods. *SIAM J. Numer. Anal.*, 20(3):510–536, 1983.

[3] M. Bebendorf and W. Hackbusch. Existence of $\mathcal{H}$-matrix approximants to the inverse FE-matrix of elliptic operators with L^∞-coefficients. *Numer. Math.*, 95(1):1–28, 2003.

[4] S. Börm. Approximation of solution operators of elliptic partial differential equations by $\mathcal{H}$- and $\mathcal{H}^2$-matrices. *Numer. Math.*, published online 2010, DOI 10.1007/s00211-009-0278-7.

[5] P. Ciarlet. *The finite element method for elliptic problems.* North-Holland, 1987.

[6] W. Dahmen, B. Faermann, I. Graham, W. Hackbusch, and S. Sauter. Inverse Inequalities on Non-Qasiuniform Meshes and Applications to the Mortar Element Method. *Math. Comp.*, 73(247):1107–1138, 2003.

[7] H. Dong and S. Kim. Green's Matrices of Second Order Elliptic Systems with Measurable Coefficients in Two Dimensional Domains. *Trans. AMS*, 361(6):3303–3323, 2009.

[8] R. Du and P. Ming. Convergence of heterogeneous multiscale method for elliptic problem with nonsmooth coefficients. *SIAM Multiscale Model. Simul.*, to appear.

[9] M. Grüter and K.-O. Widman. The Green Function for Uniformly Elliptic Equations. *Manuscripta Mathematica*, 37:303–342, 1982.

[10] A. Henrot. *Extremum Problems for Eigenvalues of Elliptic Operators.* Birkhäuser, 2006.

[11] A. Matache and C. Schwab. Two-Scale FEM for Homogenization Problems. *M2AN. Mathematical Modelling and Numerical Analysis*, 36(4):537–572, 2002.

[12] D. Peterseim and S. Sauter. Error Estimates for Finite Element Discretizations of Elliptic Problems with Oscillatory Coefficients. *Preprint 13-2010 Institut für Mathematik, Universität Zürich.*