THE ROLE OF NUMERICAL INTEGRATION IN NUMERICAL HOMOGENIZATION [*]

Assyr Abdulle[1]

Abstract. Finite elements methods (FEMs) with numerical integration play a central role in numerical homogenization methods for partial differential equations with multiple scales, as the effective data in a homogenization problem can only be recovered from a microscopic solver at a finite number of points in the computational domain. In a multiscale framework the convergence of a FEM with numerical integration applied to the effective (homogenized) problem guarantees that the so-called macroscopic solver is consistent and convergent. Convergence results for FEM with numerical integration are however scarce in the literature and need often to be derived as a first step to analyze a numerical homogenization method for a given problem. In this paper we review and explain the main ideas in deriving convergence results for FEM with numerical integration for linear and nonlinear elliptic problems and explain the role of these methods in numerical homogenization.

Résumé. Les méthodes d'éléments finis avec intégration numérique par quadrature jouent un rôle central dans l'homogénéisation numérique des équations aux dérivées partielles multi-échelles. En effet, les coefficients de l'équation homogénéisée ne peuvent être déterminés que pour un nombre fini de points du domaine considéré. Dans le cadre des méthodes multi-échelles basées sur un schéma macroscopique avec intégration numérique en la variable macroscopique et couplées à des schémas microscopiques autour des points de quadratures, la convergence d'une méthode d'élément fini avec intégration numérique pour le problème homogénéisé garantit que la méthode macroscopique est consistante et convergente. Comme les résultats de convergence pour les méthodes d'éléments finis avec intégration numérique ne sont connus que pour un nombre limité de problèmes, des nouveaux résultats de ce type sont alors un premier pas indispensable pour établir la convergence d'une méthode d'homogénéisation numérique pour un problème donné. Dans ce papier, nous effectuons un survol des idées clés pour l'analyse des méthodes d'éléments finis avec intégration numérique pour des problèmes linéaires et non linéaires et expliquons le rôle de ces méthodes dans l'homogénéisation numérique.

Introduction

When partial differential equations (PDEs) have coefficients that vary over multiple scales (highly oscillatory), classical numerical methods such as the finite element method (FEM), the finite volume method (FVM) or the finite difference method (FDM) are inefficient, as the mesh involved in any of these methods needs to resolve the smallest scale in the PDE to recover usual convergence rates. Such scale resolution is often prohibitive in many applications and there is a need for other type of numerical methods, called multiscale methods.

[1] ANMC, Section de Mathématiques, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland, (assyr.abdulle@epfl.ch)

Among the large variety of multiscale problems, we concentrate on numerical methods for homogenization problems. Homogenization is a mathematical theory that study the convergence of PDEs with rapidly oscillating coefficients towards an averaged PDE describing the macroscopic or effective behavior of the physical process modeled by the differential equation [18, 31, 32].

Numerical homogenization methods are numerical methods able to approximate the effective solution of a highly oscillatory PDE without resolving the full oscillatory equation by direct discretization (see [3, 29] for recent reviews). Such methods are also called multiscale methods as they typically couple different solvers at different scales. The numerical analysis of numerical homogenization methods has been studied by many authors and we refer to [15] for a contribution of a pioneer in the field and to [3] for a recent review with numerous bibliographical entries.

In this paper, we first discuss in Section 1 the typical structure of a numerical homogenization method. We then discuss the central role of FEM with numerical integration for such methods. We then review numerical homogenization methods for linear elliptic homogenization problems, nonlinear monotone elliptic homogenization problems and nonlinear nonmonotone elliptic homogenization problems in Sections 2,3,4, respectively. As our focus is on numerical integration we explain the various steps needed to control this error in a numerical homogenization method. We will show that for numerical homogenization, in order to obtain quantitative error bounds, FEM with numerical integration cannot be avoided as by its nature the macroscopic solver can only be defined at a finite number of points, which should be quadrature nodes of a macroscopic mesh. This raises the question of quantitative error bounds results for FEM with numerical integration for the various aforementioned problems. While the literature for the analysis of the FEM is abundant, works concerned with the analysis of the FEM with numerical integration are sparse even though for practical problems with non-constant parameters numerical integration is unavoidable. For linear elliptic problems, this problem has been thoroughly analyzed by Strang & Fix [37] and Ciarlet & Raviart [23]. For linear parabolic and hyperbolic problems results by Raviart [35] and Baker & Dougalis [17] are available. However for nonlinear problems quantitative error bounds have only recently be obtained for elliptic problems (of monotone and nonmonotone type) motivated by numerical homogenization [7, 11, 12].

We close this introduction by mentioning that while we concentrate on a specific numerical homogenization method, namely the finite element heterogeneous multiscale method (FE-HMM) built in the framework of the heterogeneous multiscale method (HMM) [25], the link between numerical homogenization method and FEM with numerical integration is of general nature and not restricted to the specific framework described here.

Notations. Let $\Omega \subset \mathbb{R}^d$ be open and denote by $W^{s,p}(\Omega)$ the standard Sobolev spaces. We use the standard Sobolev norms $\|\cdot\|_{W^{s,p}(\Omega)}$ and semi-norms $|\cdot|_{W^{s,p}(\Omega)}$. For $p = 2$ we will use the notations $\|\cdot\|_{H^s(\Omega)}$ and $|\cdot|_{H^s(\Omega)}$, respectively. For $k \in \mathbb{N}$ and $1 \le p \le \infty$, the broken norms $\|\cdot\|_{\bar{W}^{k,p}(\Omega)}$ are given by

$$\|v\|_{\bar{W}^{k,p}(\Omega)} = \left(\sum_{K \in \mathcal{T}_H} \|v\|_{W^{k,p}(K)}^p \right)^{1/p}, \text{ if } p < \infty, \quad \|v\|_{\bar{W}^{k,\infty}(\Omega)} = \max_{K \in \mathcal{T}_H} \|v\|_{W^{k,\infty}(K)},$$

where $\mathcal{T}_H$ is a partition of the closure of Ω, i.e., $\overline{\Omega} = \bigcup_{K \in \mathcal{T}_H} K$, $W^{k,p}(K)$ and $\|\cdot\|_{W^{k,p}(K)}$ denote the usual Sobolev spaces and norms on a closed subset K of Ω, respectively, and $\bar{W}^{k,2}(\Omega)$ is written as $\bar{H}^k(\Omega)$. The notations $|v|_{\bar{W}^{k,p}(\Omega)}, |v|_{\bar{W}^{k,\infty}(\Omega)}$ will be used for the corresponding semi-norm.

1. Numerical homogenization methods

The general methodology of a numerical homogenization method such as the FE-HMM can be described as follows. Let Ω be an open bounded polygonal domain in $\mathbb{R}^d$, V be a Hilbert space and consider the following (multiscale) problem: find $u^\varepsilon \in V$ such that

$$L^\varepsilon(u^\varepsilon, a^\varepsilon) = f \quad \text{in } \Omega, \tag{1}$$

with appropriate boundary conditions. Here L^ε denotes a differential operator, a^ε highly oscillatory data and f a right-hand side [1]. The multiscale nature of the data, the operator and the solution is emphasized by the superscript ε (representing the typical size of a small scale in the considered problem). Assume now that u^ε converges (weakly) in V to u as $\varepsilon \to 0$, where the function u (a homogenized solution) solves a homogenized problem of the form

$$L(u, a) = f \quad \text{in } \Omega. \tag{2}$$

In this latter problem, the small scales have been averaged out in the homogenization process.

First, when one is not interested in the fine scale details involved in the problem (1) it is attractive to solve (2) numerically instead of (1), as (2) can in principle be solved with a standard FEM. Second, when fine scale features are required (often only in part of the computational domain) one can add corrector functions on the homogenized solution u to recover small scale information [34] or couple the solution of (1) in a part ω of the domain with the homogenized solution u in $\Omega \backslash \omega$ (in this direction see the recent results [9, 16]). However, even when the primary goal is to solve numerically (2), this cannot be done straightforwardly as the effective data $a(x)$, $x \in \Omega$, are not known explicitly (except in very special cases) and must be computed with a micro FEM for a problem involving the original operator (1) on patches $K_{\delta_j} = x_j + \delta Y$ (sampling domains) around a given point $x_j \in \Omega$ (here $\delta \geq \varepsilon$ is a parameter usually of the size of ε and $Y = (-1/2, 1/2)^d$). Hence $a(x)$ can only be approximated at a finite number of points x_j, $j = 1, \ldots, N$, in the computational domain.

Let us describe a general framework for numerical homogenization methods such as the FE-HMM. Such methods rely on

- a macroscopic domain Ω with a family of macroscopic triangulations $\mathcal{T}_H$ such that $\bigcup_{K \in \mathcal{T}_H} K = \overline{\Omega}$, where H is the maximum diameter of the element $K \in \mathcal{T}_H$ and $\overline{\Omega}$ is the closure of Ω;
- a family of microscopic domains (sampling domains) $K_{\delta_j} = x_j + \delta Y$, $j = 1, \ldots, N$, and microscopic triangulation $\mathcal{T}_h$ such that $\bigcup_{T \in \mathcal{T}_h} T = \overline{K}_{\delta_j}$, where h is the maximum diameter of the element $T \in \mathcal{T}_h$.

A numerical homogenization method (such as the FE-HMM) relies on (at least) two solvers, and on a data recovery process

(1) a macroscopic solver L_{HMM} for the effective problem L *with a priori unknown data* $\{a^h(x_j)\}_{j=1}^N$ defined on a macroscopic finite dimensional subspace $V_H(\Omega)$ of V based on piecewise polynomials on each element K of the macroscopic triangulation $\mathcal{T}_H$ of Ω;

(2) a microscopic solver involving the operator L^ε constrained by the macroscopic state (usually through boundary conditions involving the macro solution) defined on microscopic finite element spaces $V_h(K_{\delta_j})$ based on piecewise polynomials on each element T of the microscopic triangulation $\mathcal{T}_{h_j}$;

(3) a data recovery process in which the effective data $a^h(x_j)$ at the point x_j are computed using a suitable average involving the fine scale data a^ε and the microscopic finite element solutions in each K_{δ_j}.

We then consider the following problem: find $u_{HMM} \in V_H$ such that

$$L_{HMM}(u_{HMM}) = f \quad \text{in } \Omega. \tag{3}$$

Fundamental questions are now the following. How should we choose the nodes x_j in Ω to obtain the optimal accuracy of the numerical homogenization method with a minimal cost ? Recall that the computation of $a(x_j)$ for each x_j is expensive as it involves a (micro) boundary value problem in a patch around x_j. How should we choose the macro and micro meshsizes and the polynomial degree in order to obtain a given accuracy with the minimal computational cost ? We note that the macroscopic meshsize H is usually both larger and independent of ε and its size is solely dictated by the required accuracy for the effective problem. The microscopic meshsize is smaller than ε *but only* on patches K_{δ_j} that are usually also of size $\delta \simeq \varepsilon$. A fully discrete a priori error estimates [1] then reveals that both meshes have to be refined simultaneously and that the computational cost, independent of ε, is proportional to the macroscopic degrees of freedom.

---

[1]For simplicity we do not consider a highly oscillatory right-hand side (e.g., f^ε) but we note that this situation can usually be handled with minor modifications of the method described below (see e.g., [10]).

For the analysis of such a method, we usually introduce an intermediate problem, namely a macroscopic solver L_H for the effective problem L with data $\{a(x_j)\}_{j=1}^N$ assumed to be exact and computed at the same location $x_j \in \Omega$ as for the FE-HMM. We thus have three macroscopic solutions: u the solution of (2), u_{HMM} the solution of (3) and u_H the solution of $L_H(u_H) = f$ in Ω.

The analysis of the FE-HMM relies on the following decomposition [4]

$$\|u - u_{HMM}\| \leq \|u - u_H\| + \|u_H - u_{HMM}\|. \tag{4}$$

The first term in the right-hand side of (4) is usually called the macroscopic error. It depends on the type of discretization used at the macroscopic level, the macroscopic meshsize and on the quadrature points x_j chosen in the computational domain to recover $\{a(x_j)\}$. The second term in the right-hand side of (4) is usually called the macroscopic error comprise the so-called microscopic and modeling errors and depends on the type of discretization used at the microscopic level, the macroscopic meshsize, the sampling domain size and the type of boundary conditions used for the micro solver.

We note that it is the second term that involves the mathematical homogenization theory for its analysis, while the control of the first term involves a classical question in the numerical analysis of the FEM as described in the introduction, which goes back to Strang & Fix [37] and Ciarlet & Raviart [23] for linear elliptic problem. For numerical homogenization methods, as we will show in this paper, FEM with numerical integration cannot be avoided as by its nature the macroscopic solver can only be defined at quadrature nodes. Thus any numerical homogenization method for a new class of problems must come with its companion analysis of FEM with numerical integration. These results are however not trivial already in the linear case (see [23] for a very general analysis). In particular for nonlinear problems, quantitative error bounds have only recently be obtained for elliptic problems motivated by numerical homogenization [7, 11, 12]. We will highlight in the following sections the main ideas to derive error estimates for FEM with numerical integration for various linear and nonlinear problems and show how such solvers naturally arise in numerical homogenization. We close this introduction by noting that while we only describe here multiscale methods coupling elliptic equations at both the macro and the micro scales, one could also couple different physics at different scales. This has been pursued for example in [6], where the macroscopic solver discretizes an effective Darcy equation, while the microscopic solver discretizes Stokes problems around quadrature points of the macromesh. There again, results for FEM with numerical integration for the Darcy problem are important.

2. NUMERICAL HOMOGENIZATION AND NUMERICAL INTEGRATION FOR LINEAR ELLIPTIC PROBLEMS

In this section we first detail the abstract method described in the introduction for elliptic linear multiscale problems of the form [2]

$$L^\varepsilon(u^\varepsilon, a^\varepsilon) = -\nabla \cdot (a^\varepsilon \nabla u^\varepsilon) = f \text{ in } \Omega, \quad u^\varepsilon = 0 \text{ on } \partial\Omega. \tag{5}$$

We assume that $f \in L^2(\Omega)$ and that the family of tensors $a^\varepsilon(x) \in (L^\infty(\Omega))^{d \times d}$ indexed by $\varepsilon > 0$ is uniformly elliptic and bounded, thus for any ε there exist a unique solution $u^\varepsilon \in H_0^1(\Omega)$ of Problem (5) and the family of solution $\{u^\varepsilon\}$ is bounded independently of ε. For such problems the homogenization theory [18, 31, 32] ensures the existence of a subsequence of $\{u^\varepsilon\}$ that weakly converges in $H_0^1(\Omega)$ to function u, solution of the homogenized problem $L(u, a) = -\nabla \cdot (a\nabla u) = f$ in Ω with $u = 0$ on $\partial\Omega$ that reads in weak form: find $u \in H_0^1(\Omega)$ such that

$$B(u, v) = \int_\Omega a(x) \nabla u \cdot \nabla v \, dx = \int_\Omega f v \, dx, \quad \forall v \in H_0^1(\Omega). \tag{6}$$

---

[2]Here we take zero Dirichlet boundary conditions for simplicity, but stress that the FE-HMM method can be analyzed for other boundary conditions.

The tensor $a(x)$ is again uniformly elliptic and bounded

$$\exists \lambda, \Lambda > 0 \text{ such that } \lambda|\xi|^2 \le a(x)\xi \cdot \xi, \ |a(x)\xi| \le \Lambda|\xi|, \ \forall \xi \in \mathbb{R}^d, \text{ a.e. } x \in \Omega. \tag{7}$$

We note that for locally periodic coefficients, i.e., $a^\varepsilon(x) = a(x, x/\varepsilon) = a(x, y)$ Y-periodic in y, the whole sequence $\{u^\varepsilon\}$ weakly converges to the unique solution u of (6).

2.1. Numerical homogenization method.

For the numerical homogenization method, based on a family of triangulations $\mathcal{T}_H$ (the triangulation defined in the introduction), we first define the macroscopic finite element space by

$$V_H^\ell(\Omega, \mathcal{T}_H) = \{v_H \in H_0^1(\Omega); \ v_H|_K \in \mathcal{R}^\ell(K), \ \forall K \in \mathcal{T}_H\}, \tag{8}$$

where $\mathcal{R}^\ell(K)$ is the space $\mathcal{P}^\ell(K)$ of polynomials on K of total degree at most ℓ if K is a simplicial FE, or the space $\mathcal{Q}^\ell(K)$ of polynomials on K of degree at most ℓ in each variable if K is a parallelogram (rectangular) FE. Then, for each $K \in \mathcal{T}_H$ we choose quadrature nodes and weights $\{x_{K_j}, \omega_{K_j}\}_{j=1}^J$ and define the macroscopic solver for $v_H, w_H \in V_H^\ell(\Omega, \mathcal{T}_H)$ by

$$B_{HMM}(v_H, w_H) = \sum_{K \in \mathcal{T}_H} \sum_{j=1}^J \omega_{K_j} a_{K_j}^h \nabla v_H(x_{K_j}) \cdot \nabla w_H(x_{K_j}), \tag{9}$$

where $a_{K_j}^h$ is an approximation of the homogenized tensor $a(x)$ at the node $x = x_{K_j}$, determined by a micro method solving a micro problem on a patch K_{δ_j} around x_{K_j}. The numerical homogenization solution is then defined as follows: find $u_{HMM} \in V_H^\ell(\Omega, \mathcal{T}_H)$ such that

$$B_{HMM}(u_H, v_H) = \int_\Omega f v_H dx, \ \ \forall v_H \in V_H^\ell(\Omega, \mathcal{T}_H). \tag{10}$$

We see from the bilinear form (9) that numerical integration is mandatory in numerical homogenization already to define the numerical method as the effective data $a_{K_j}^h$ can only be accessed at a finite number of points $x_{K_j} \in \Omega$ (unless $a(x)$ is constant). For simplicity we do not consider numerical integration for the right-hand side of (10) but emphasize that the generalization for that case would not introduce additional difficulties. As mentioned in the introduction for the analysis of the method (10) we introduce an intermediate problem to decompose the error into macroscopic error and "numerical homogenization errors". This intermediate problem is based here on the bilinear form

$$B_H(v_H, w_H) = \sum_{K \in \mathcal{T}_H} \sum_{j=1}^J \omega_{K_j} a(x_{K_j}) \nabla v_H(x_{K_j}) \cdot \nabla w_H(x_{K_j}), \tag{11}$$

where $a(x_{K_j})$ is the value of the homogenized tensor at the node $x = x_{K_j}$. We then define $u_H \in V_H^\ell(\Omega, \mathcal{T}_H)$ such that

$$B_H(u_H, v_H) = \int_\Omega f v_H dx, \ \ \forall v_H \in V_H^\ell(\Omega, \mathcal{T}_H). \tag{12}$$

Of course some assumptions on the quadrature formula are needed to ensure that Problem (10) is well-posed, in particular we need to ensure that the bilinear form B_H is elliptic. This follows from the assumption (**Q1**) below. Going back to the decomposition (4), we see that the error $\|u - u_H\|$ quantifies indeed the error of FEM with numerical integration for a single scale problem. When the problem is linear the error analysis goes back

to [23, 37]. These results that are classical are reviewed in the next subsection. We close this presentation of the FE-HMM for linear elliptic problem by noting that the term $\|u_H - u_{HMM}\|$ has been analyzed in various publications [1, 26] for a variety of macro and micro FEM (see [3, 4] for reviews).

Micro solver. For completeness we give the definition of the micro solver that can be used to compute $a_{K_j}^h$. For each quadrature node x_{K_j} we consider the sampling domain K_{δ_j}, its micro triangulation defined in Section 1 and the micro FE space $V_h^q(K_{\delta_j}, \mathcal{T}_h) = \{z_h \in W(K_{\delta_j}); \ z_h|_T \in \mathcal{R}^q(T), \ T \in \mathcal{T}_h\}$, where $W(K_{\delta_j})$ is a given Sobolev space. The choice of $W(K_{\delta_j})$ sets the coupling condition between macro and micro functions, typical examples are $H_{per}^1(K_{\delta_j})/\mathbb{R}$ (where $H_{per}^1(K_{\delta_j})$ is the closure of smooth periodic functions on K_{δ_j} for the H^1 norm) or $H_0^1(K_{\delta_j})$. Next for each $\xi \in \mathbb{R}^d$ and $x_{K_j} \in \Omega$, $a_{K_j}^h \in \mathbb{R}^{d \times d}$ is given by

$$a_{K_j}^h \xi = \frac{1}{|K_{\delta_j}|} \int_{K_{\delta_j}} a^\varepsilon(x)(\xi + \nabla \chi_{h,K_j}^\xi) dx$$

and χ_{h,K_j}^ξ solves the micro problem: find $\chi_{h,K_j}^\xi \in V_h^q(K_{\delta_j}, \mathcal{T}_h)$ such that

$$\int_{K_{\delta_j}} a^\varepsilon(x)(\xi + \nabla \chi_{h,K_j}^\xi) \cdot \nabla z_h \, dx = 0, \qquad \forall z_h \in V_h^q(K_{\delta_j}, \mathcal{T}_h).$$

Alternatively for $v_H \in V_H^\ell(\Omega, \mathcal{T}_H)$ we can define a micro function v_{h,K_j}, satisfying $(v_{h,K_j} - v_{H,lin}) \in V_h^q(K_{\delta_j}, \mathcal{T}_h)$ and solution of

$$\int_{K_{\delta_j}} a^\varepsilon(x) \nabla v_{h,K_j} \cdot \nabla z_h dx = 0, \quad \forall z_h \in V_h^q(K_{\delta_j}, \mathcal{T}_h), \tag{13}$$

where $v_{H,lin}|_{K_{\delta_j}} = v_H(x_{K_j}) + (x - x_{K_j}) \cdot \nabla v_H(x_{K_j})$ is the linearization of v_H at the quadrature point x_{K_j}. We then have the relation

$$a_{K_j}^h \nabla v_H(x_{K_j}) \cdot \nabla w_H(x_{K_j}) = \frac{1}{|K_\delta|} \int_{K_{\delta_j}} a^\varepsilon(x) \nabla v_{h,K_j} \cdot \nabla w_{h,K_j} dx$$

(see [1–4] for details).

2.2. FEM with numerical integration for linear problems

We review here results for FEM with numerical integration for linear problems for simplicial or parallelogram elements. We will assume (throughout the paper) that the family of triangulation $\mathcal{T}_H$ is regular, i.e.

$$\frac{H_K}{\rho_K} \leq \sigma \ \text{ for all } K \in \bigcup_H \mathcal{T}_H, \tag{14}$$

where H_K is the diameter of K and ρ_K the diameter of the largest ball contained in K.

For each element $K \in \mathcal{T}_H$ we consider a $\mathcal{C}^1$-diffeomorphism F_K such that $K = F_K(\hat{K})$, where $\hat{K}$ is the simplicial or the quadrilateral reference element. Let $J \in \mathbb{N}$ and $\{\hat{x}_j, \hat{\omega}_j\}_{j=1}^J$ be a given quadrature formula on $\hat{K}$ with positive weights $\hat{\omega}_j$ and $\hat{x}_j \in \hat{K}$. The transformation F_K induces a quadrature formula on K with integration points $x_{K_j} = F_K(\hat{x}_j)$ and weights $\omega_{K_j} = \hat{\omega}_j |\det(\partial F_K(\hat{x}_j))|$. We will consider for simplicity affine maps, i.e., the case of simplicial straight elements or parallelogram, [3] hence $\det(\partial F_K)$ is constant for each K. We make the following classical assumptions on the quadrature formula $\{\hat{x}_j, \hat{\omega}_j\}_{j=1}^J$ on the reference element $\hat{K}$ (see [23])

(Q1) $\hat{\omega}_j > 0, \ j = 1, \ldots, J, \quad \sum_{j=1}^J \hat{\omega}_j |\nabla \hat{p}(\hat{x}_j)|^2 \geq \hat{\lambda} \|\nabla \hat{p}\|_{L^2(\hat{K})}^2, \ \forall \hat{p}(\hat{x}) \in \mathcal{R}^\ell(\hat{K}), \ \hat{\lambda} > 0;$

---

[3] In what follows we will refer to simplicial straight elements as simplicial elements.

(**Q2**) $\int_{\hat{K}} \hat{p}(\hat{x}) d\hat{x} = \sum_{j=1}^{J} \hat{\omega}_j \hat{p}(\hat{x}_j)$, $\forall \hat{p}(\hat{x}) \in \mathcal{R}^{\sigma}(\hat{K})$, where $\sigma = \max(2\ell - 2, \ell)$ if $\hat{K}$ is a simplicial FE, or $\sigma = \max(2\ell - 1, \ell + 1)$ if $\hat{K}$ is a rectangular FE.

We observe that for simplicial elements (**Q2**) implies (**Q1**) and (**Q1**) holds with $\hat{\lambda} = 1$. The following lemma ensures that the problem (12) is well-posed (see for example [22, Thm. 27.1] for a proof). We note that a similar result holds for Problem (10) if in addition to the hypothesis of Lemma 2.1, $a^{\varepsilon}(x)$ is uniformly elliptic and bounded (see [1, 3]).

Lemma 2.1. *Assume that (***Q1***) holds. Then the bilinear form B_H defined in (11) is uniformly elliptic and bounded, hence Problem (12) is well-posed.*

Theorem 2.2 (Ciarlet & Raviart [23])**.** *Consider u the solution of problem (6) and u_H the solution of problem (12). Let $\ell \geq 1$ and $\mu = 0$ or 1. Assume (***Q1***), (***Q2***), (7) and*

$$u \in H^{\ell+1}(\Omega), \tag{15}$$

$$a_{mn} \in W^{\ell+\mu,\infty}(\Omega), \quad \forall m, n = 1, \ldots, d. \tag{16}$$

Then we have the a priori error estimates

$$\|u - u_H\|_{H^1(\Omega)} \leq CH^{\ell} \quad (\mu = 0), \tag{17}$$

$$\|u - u_H\|_{L^2(\Omega)} \leq CH^{\ell+1} \quad (\mu = 1), \tag{18}$$

where C is independent of H.

Let us explain the main ingredients for the proof of the above theorem. It is important to understand these steps as they are also useful in more complicated situation, namely for nonlinear problems.

Step 1. We start with the H^1 estimate. The first step consists in a generalization of Cea's Lemma due to Strang [37]. Using Lemma 2.1 it is easily seen that

$$\|u - u_H\|_{H^1(\Omega)} \leq C \inf_{v_H \in V_H^{\ell}(\Omega, \mathcal{T}_H)} \left(\|u - v_H\|_{H^1(\Omega)} + \sup_{w_H \in V_H^{\ell}(\Omega, \mathcal{T}_H)} \frac{|B(v_H, w_H) - B_H(v_H, w_H)|}{\|w_H\|_{H^1(\Omega)}} \right). \tag{19}$$

For the L^2 estimate, we need to consider an adjoint problem following Aubin [14] and Nitsche [33]. For $g \in L^2(\Omega)$, let $z \in H_0^1(\Omega)$ be the solution of the adjoint problem

$$B(v, z) = \int_{\Omega} gv dx, \quad \forall v \in H_0^1(\Omega), \tag{20}$$

where B is defined in (6). Using then

$$\|u - u_H\|_{L^2(\Omega)} = \sup_{g \in L^2(\Omega)} \frac{|\int_{\Omega} (u - u_H) g dx|}{\|g\|_{L^2(\Omega)}}$$

together with the H^2-regularity $\|z\|_{H^2(\Omega)} \leq C\|g\|_{L^2(\Omega)}$ it is also seen that

$$\|u - u_H\|_{L^2(\Omega)} \leq C \sup_{z \in H_0^1(\Omega) \cap H^2(\Omega)} \frac{1}{\|z\|_{H^2(\Omega)}} \left(\inf_{z_H \in V_H^{\ell}} \left(\|u - u_H\|_{H^1(\Omega)} \|z - z_H\|_{H^1(\Omega)} + |B(u_H, z_H) - B_H(u_H, z_H)| \right) \right). \tag{21}$$

Step 2. In the inequality (19) the first term of the right-hand side of the inequality can be dealt with by using an appropriate interpolant for u. The first term of the right-hand side in the inequality (21) can be bounded by using the H^1 estimate and an appropriate interpolant for z. For the second term in both inequalities (19) and (21), an appropriate consistency error estimate of $B(\cdot, \cdot) - B_H(\cdot, \cdot)$ is needed. This relies on quadrature

error estimates and on the Bramble-Hilbert lemma. Consider for the reference element $\hat{K}$ the quadrature error functional defined for all continuous functions $\hat{\varphi}$ on $\hat{K}$

$$E_{\hat{K}}(\hat{\varphi}) := \int_{\hat{K}} \hat{\varphi}(\hat{x})d\hat{x} - \sum_{j=1}^{J} \hat{\omega}_j \hat{\varphi}(\hat{x}_j),$$

Then for any element $K \in \mathcal{T}_H$ the quadrature scheme over $\hat{K}$ induces a quadrature scheme over K by

$$E_K(\varphi) := \int_K \varphi(x)dx - \sum_{j=1}^{J} \omega_{K_j} \varphi(x_{K_j}),$$

where $x_{K_j} = F_K(\hat{x}_j)$ and $\omega_{K_j} = |\det \partial F_K|\hat{\omega}_j$ for $1 \le j \le J$ (recall that F_K is the affine transformation such that $K = F_K(\hat{K})$) and we observe that

$$E_K(\varphi) = |\det \partial F_K|E_{\hat{K}}(\hat{\varphi}), \tag{22}$$

where $\hat{\varphi}(\hat{x}) = \varphi(F_K(\hat{x}))$. We thus have

$$B(v_H, w_H) - B_H(v_H, w_H) = \sum_{K \in \mathcal{T}_H} E_K(a\nabla v_H \cdot \nabla w_H). \tag{23}$$

Consider next $E_K(a_{m,n}w_{(n)}v_{(m)})$, where $v_{(m)}, w_{(n)}$ denote the components of $\nabla v_H|_K$ and $\nabla w_H|_K$, i.e., $\partial v_H/\partial x_m$, $\partial w_H/\partial x_n$. In view of (22) it is enough to estimate

$$E_{\hat{K}}(\hat{a}_{m,n}\hat{v}_{(n)}\hat{w}_{(m)}),$$

where we use the correspondence $\hat{\varphi}(\hat{x}) = \varphi(F_K(\hat{x}))$ for each function in the above expression. Let $\ell \ge 1$ and consider for a given polynomial $\hat{\psi}$ on $\hat{K}$ the linear form

$$\hat{\varphi} \mapsto E_{\hat{K}}(\hat{\psi}\hat{\varphi}).$$

This form is continuous on $W^{\ell,\infty}(\hat{K})$ with norm $\le C\|\hat{\psi}\|_{L^2(\hat{K})}$, i.e.,

$$|E_{\hat{K}}(\hat{\psi}\hat{\varphi})| \le C\|\hat{\psi}\|_{L^2(\hat{K})}\|\hat{\varphi}\|_{W^{\ell,\infty}(\hat{K})}.$$

Due to the assumption (**Q2**) this form vanishes on $\mathcal{P}^{\ell-1}(\hat{K})$ if

- $\hat{\psi} \in \mathcal{P}^{\ell-1}(\hat{K})$ (simplicial elements);
- $\hat{\psi} \in \mathcal{Q}^{\ell}(\hat{K}) \cap \mathcal{P}^{d\ell-1}(\hat{K})$, the space of all derivatives of polynomials belonging to $\mathcal{Q}^{\ell}(\hat{K})$, (parallelogram elements).[4]

Thus, in both cases, the Bramble-Hilbert lemma [20] allows to replace the norm $W^{\ell,\infty}$ of $\hat{\varphi}$ by the semi-norm involving only derivatives of order ℓ, i.e.,

$$|E_{\hat{K}}(\hat{\psi}\hat{\varphi})| \le C\|\hat{\psi}\|_{L^2(\hat{K})}|\hat{\varphi}|_{W^{\ell,\infty}(\hat{K})}, \quad \forall \hat{\varphi} \in W^{\ell,\infty}(\hat{K}). \tag{24}$$

---

[4]Observe that if $\hat{w} \in \mathcal{Q}^{\ell}(\hat{K})$ then $\hat{\psi} = \hat{w}_{(n)} \in \mathcal{Q}^{\ell}(\hat{K}) \cap \mathcal{P}^{d\ell-1}(\hat{K})$.

We now take $\hat{\varphi} = \hat{a}_{m,n}\hat{v}_{(n)}$ and $\hat{\psi} = \hat{w}_{(m)}$, use the equivalence of norms over the finite dimensional spaces of polynomial to obtain

$$|\hat{a}_{m,n}\hat{v}_{(n)}|_{W^{\ell,\infty}(\hat{K})} \quad \leq \quad C\sum_{j=0}^{\ell}|\hat{a}|_{W^{j,\infty}(\hat{K})}|\hat{v}_{(n)}|_{H^{\ell-j}(\hat{K})}, \tag{25}$$

where the sum starts at 1 for simplicial elements. Finally using the following inequalities for all $0 \leq j \leq \ell$ (see [22, Theorems 15.1 and 15.2])

$$|\hat{a}|_{W^{j,\infty}(\hat{K})} \quad \leq \quad CH_K^j|a|_{W^{j,\infty}(K)}, \tag{26}$$

$$|\hat{v}_{(n)}|_{H^{\ell-j}(\hat{K})} \quad \leq \quad CH_K^{\ell-j}|\det\partial F_K|^{-1/2}|v_{(n)}|_{H^{\ell-j}(K)}, \tag{27}$$

and $|\hat{w}_{(m)}|_{L^2(\hat{K})} \leq C|\det\partial F_K|^{-1/2}|w_{(m)}|_{L^2(K)}$, we get in view of (23) and (22)

$$|B(v_H, w_H) - B_H(v_H, w_H)| \leq CH^\ell\|a\|_{W^{\ell,\infty}(\Omega)}\|v_H\|_{\bar{H}^{\ell+1}(\Omega)}\|\nabla w_H\|_{L^2(\Omega)} \tag{28}$$

For the estimates in the L^2 norm we need to extract an additional power of H (see [23, Theorem 8]). Define the linear operator $\hat{\Pi}_0 : L^1(\hat{K}) \to \mathcal{P}^0(\hat{K})$ by

$$\hat{\Pi}_0(\hat{\psi}) = \frac{1}{|\hat{K}|}\int_{\hat{K}}\hat{\psi}(\hat{x})d\hat{x}.$$

Assume $\hat{a}_{m,n} \in W^{\ell+1,\infty}(\hat{K})$ and consider

$$E_{\hat{K}}(\hat{a}_{m,n}\hat{v}_{(n)}\hat{w}_{(m)}) \quad = \quad \underbrace{E_{\hat{K}}\left(\hat{a}_{m,n}\hat{v}_{(n)}\big(\hat{w}_{(m)} - \hat{\Pi}_0(\hat{w}_{(m)})\big)\right)}_{E_1} + \underbrace{E_{\hat{K}}\left(\hat{a}_{m,n}\big(\hat{v}_{(n)} - \hat{\Pi}_0(\hat{v}_{(n)})\big)\hat{\Pi}_0(\hat{w}_{(m)})\right)}_{E_2}$$

$$+ \quad \underbrace{E_{\hat{K}}\left(\hat{a}_{m,n}\hat{\Pi}_0(\hat{v}_{(n)})\hat{\Pi}_0(\hat{w}_{(m)})\right)}_{E_3}.$$

The three linear forms on the left-hand side are of the form $\hat{\varphi} \in W^{\ell+1,\infty}(\hat{K}) \mapsto E_{\hat{K}}(\hat{\psi}\hat{\varphi})$. Using (**Q2**) we see that the first linear form (we set $\hat{\varphi} = \hat{a}_{m,n}\hat{v}_{(m)}$ $\hat{\psi} = (\hat{w}_{(n)} - \hat{\Pi}_0(\hat{w}_{(n)}))$) vanishes on $\mathcal{P}^{\ell-1}(\hat{K})$ for $\hat{\psi} \in \mathcal{P}^{\ell-1}(\hat{K})$ or for $\hat{\psi} \in \mathcal{Q}^\ell(\hat{K}) \cap \mathcal{P}^{2\ell-1}(\hat{K})$). Thus by the Bramble-Hilbert lemma we obtain

$$|E_1| \leq C|\hat{a}_{m,n}\hat{v}_{(n)}|_{W^{\ell,\infty}(\hat{K})}\|\hat{w}_{(m)} - \hat{\Pi}_0(\hat{w}_{(m)})\|_{L^2(\hat{K})} \leq C|\hat{a}_{m,n}\hat{v}_{(n)}|_{W^{\ell,\infty}(\hat{K})}|\hat{w}_{(m)}|_{H^1(\hat{K})},$$

where the second inequality is obtained by noting that since $\hat{\psi} = \hat{\Pi}_0(\hat{\psi})$ on $\mathcal{P}^0(\hat{K})$, we have $\|\hat{\psi} - \hat{\Pi}_0(\hat{\psi})\|_{L^2(\hat{K})} \leq C|\hat{\psi}|_{H^1(\hat{K})}$ (using a result on polynomial preserving operator see [22, Theorems 15.3]). For the second linear form (we set $\hat{\varphi} = \hat{a}_{m,n}$, $\hat{\psi} = \hat{v}_{(n)} - \hat{\Pi}_0(\hat{v}_{(n)})\hat{\Pi}_0(\hat{w}_{(m)})$) a similar arguments yields the bound

$$|E_2| \leq C|\hat{a}_{m,n}|_{W^{\ell,\infty}(\hat{K})}|\hat{v}_{(n)}|_{H^1(\hat{K})}|\hat{w}_{(m)}|_{L^2(\hat{K})}.$$

For the third linear form we set $\hat{\varphi} = \hat{a}_{m,n}$, $\hat{\psi} = \hat{\Pi}_0(\hat{v}_{(n)})\hat{\Pi}_0(\hat{w}_{(m)})$. Then using (**Q2**) we see that this linear form vanishes on $\mathcal{P}^\ell(\hat{K})$ and applying the Bramble-Hilbert lemma yields the bound

$$|E_3| \leq C|\hat{a}_{m,n}|_{W^{\ell+1,\infty}(\hat{K})}|\hat{v}_{(n)}|_{L^2(\hat{K})}|\hat{w}_{(m)}|_{L^2(\hat{K})}.$$

Using (25),(26),(27) similarly as for the H^1 estimate gives

$$|B(v_H, w_H) - B_H(v_H, w_H)| \leq CH^{\ell+1} \|a\|_{W^{\ell+1,\infty}(\Omega)} \|v_H\|_{\bar{H}^{\ell+1}(\Omega)} \|w_H\|_{\bar{H}^2(\Omega)}. \tag{29}$$

Finally we also need for the L^2 bound an error estimate of the form

$$|B(v_H, w_H) - B_H(v_H, w_H)| \leq CH \|a\|_{W^{1,\infty}(\Omega)} \|\nabla v_H\|_{L^2(\Omega)} \|w_H\|_{\bar{H}^2(\Omega)}, \tag{30}$$

for all $v_H, w_H \in V_H^\ell(\Omega, \mathcal{T}_H)$ (observe that this does not follow from (28) when $\ell > 1$). Consider $\hat{\varphi} \mapsto E_{\hat{K}}(\hat{\psi}\hat{\varphi})$, a continuous linear form on $W^{1,\infty}(\hat{K})$, with $\hat{\psi} \in \mathcal{P}^{\ell-1}(\hat{K})$ or $\hat{\psi} \in \mathcal{Q}^\ell(\hat{K}) \cap \mathcal{P}^{d\ell-1}(\hat{K})$. Due to (**Q2**) this linear form vanishes on $\mathcal{P}^0(\hat{K})$ and the Bramble-Hilbert lemma yields $|E_{\hat{K}}(\hat{\psi}\hat{\varphi})| \leq C\|\hat{\psi}\|_{L^2(\hat{K})} |\hat{\varphi}|_{W^{1,\infty}(\hat{K})}$ and (30) is obtained using similar arguments as in the proof of (30).

Step 3. In order to have a constant C independent of H in the estimates (17),(18), appropriate test functions have to be used in (19),(28) and (21),(29). As $u \in H^{\ell+1}(\Omega)$ we can take the nodal interpolant $v_H = \mathcal{I}_H u$ (recall that the dimension $d \leq 3$) in (19)-(28) and use the standard estimates

$$\|u - \mathcal{I}_H u\|_{H^1(\Omega)} \leq CH^\ell |u|_{H^{\ell+1}(\Omega)}, \qquad \|\mathcal{I}_H u\|_{\bar{H}^{\ell+1}(\Omega)} \leq C\|u\|_{H^{\ell+1}(\Omega)}. \tag{31}$$

For the L^2 estimates we take the nodal interpolant $z_H = \mathcal{I}_H z$ and use and standard estimates

$$\|z - \mathcal{I}_H z\|_{H^1(\Omega)} \leq CH |z|_{H^2(\Omega)}, \qquad \|\mathcal{I}_H z\|_{\bar{H}^2(\Omega)} \leq C\|z\|_{H^2(\Omega)}. \tag{32}$$

In order to bound the last term in (21) we consider

$$B(u_H, \mathcal{I}_H z) - B_H(u_H, \mathcal{I}_H z) = \underbrace{B(u_H - \mathcal{I}_H u, \mathcal{I}_H z) - B_H(u_H - \mathcal{I}_H u, \mathcal{I}_H z)}_{B_1} + \underbrace{B(\mathcal{I}_H u, \mathcal{I}_H z) - B_H(\mathcal{I}_H u, \mathcal{I}_H z)}_{B_2}.$$

The term B_1 is estimated using (30),(31) and (32) and the H^1 estimate $\|u - u_H\|_{H^1(\Omega)} \leq CH^\ell |u|_{H^{\ell+1}(\Omega)}$, while the term B_2 is estimated using (29),(31) and (32). Finally, both terms can be bounded by

$$CH^{\ell+1} |u|_{H^{\ell+1}(\Omega)} |z|_{H^2(\Omega)}.$$

2.3. Numerical homogenization and numerical integration for $N+1$ scale problems

Consider now (5) assuming that a^ε varies over $N+1$ scales. Precisely, assume $a^\varepsilon = a(x, \frac{x}{\varepsilon_1}, \cdots, \frac{x}{\varepsilon_N})$, and $\varepsilon_1, \ldots, \varepsilon_N$ are N positive functions $\varepsilon_i(\varepsilon)$ that converge to 0 when $\varepsilon \to 0$ and that are well-separated in the sense that $\lim_{\varepsilon \to 0} \frac{\varepsilon_{i+1}(\varepsilon)}{\varepsilon_i(\varepsilon)} = 0$ for $i = 1, \ldots N-1$. In this situation, we have one macroscopic scale and N microscopic scales. For $N = 1$ we have the numerical method described in Section 2.1.

We observe that the micro problems were assumed to be solved with a FEM *without* numerical quadrature in Section 2.1. Already in this situation one can ask: what are the regularity conditions needed on the micro problems when using FEM with numerical quadrature in order to guarantee (optimal) convergence of the macro FEM with numerical quadrature ? Next for problems with more than two scales, one is *obliged* to use FEM with numerical quadrature at the meso and the macro scales as the data in these FEM methods rely on boundary value problems at smaller scales and are only accessible at a finite number of points. We have thus a cascade of interdepending FEM with numerical quadrature leading to a cascade of variational crimes. A generalization of error estimates for FEM with numerical quadrature for this situation has recently be obtained in [5]. As for two scale problems, such results are crucial to analyze the numerical homogenization method with multiple scales in order to determine the optimal mesh refinement at each scale to obtain a given convergence rate at the macro scale with minimal computational complexity.

3. Numerical homogenization and numerical integration for nonlinear monotone elliptic problems

In this section we review FEMs with numerical integration and numerical homogenization for nonlinear monotone elliptic problems of the form

$$L^\varepsilon(u^\varepsilon, a^\varepsilon) = -\nabla \cdot (a^\varepsilon(x, \nabla u^\varepsilon)) = f \text{ in } \Omega, \quad u^\varepsilon = 0 \text{ on } \partial\Omega. \tag{33}$$

We assume that $f \in L^2(\Omega)$ and that the maps $a^\varepsilon \colon \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ (indexed by ε) have the property that $a^\varepsilon(\cdot, \xi) \colon \Omega \to \mathbb{R}^d$ is Lebesgue measurable for every $\xi \in \mathbb{R}^d$. In addition we assume the following conditions uniformly in $\varepsilon > 0$

$(\mathcal{A}_0)$ there is some $C_0 > 0$ such that $|a^\varepsilon(x, 0)| \leq C_0$ for almost every (a.e.) $x \in \Omega$;

$(\mathcal{A}_1)$ the map $a^\varepsilon(x, \cdot) \colon \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz continuous, i.e., there exists $L > 0$ such that

$$|a^\varepsilon(x, \xi_1) - a^\varepsilon(x, \xi_2)| \leq L |\xi_1 - \xi_2|, \quad \forall \xi_1, \xi_2 \in \mathbb{R}^d, \text{ a.e. } x \in \Omega;$$

$(\mathcal{A}_2)$ the map $a^\varepsilon(x, \cdot) \colon \mathbb{R}^d \to \mathbb{R}^d$ is strongly monotone, i.e., there exists $\lambda > 0$ such that

$$(a^\varepsilon(x, \xi_1) - a^\varepsilon(x, \xi_2)) \cdot (\xi_1 - \xi_2) \geq \lambda |\xi_1 - \xi_2|^2, \quad \forall \xi_1, \xi_2 \in \mathbb{R}^d, \text{ a.e. } x \in \Omega.$$

Under these assumptions it can be seen that Problem (33) has a unique weak solution $u^\varepsilon \in H_0^1(\Omega)$ for each $\varepsilon > 0$. Homogenization theory (see [38]) ensures that a subsequence of $\{u^\varepsilon\}$ converges weakly to a function $u \in H_0^1(\Omega)$, solution of the homogenized problem that reads in weak form

$$B(u, v) = \int_\Omega a(x, \nabla u) \cdot \nabla v \, dx = \int_\Omega f v \, dx, \quad \forall v \in H_0^1(\Omega), \tag{34}$$

where $a \colon \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ is the homogenized map that satisfies again $(\mathcal{A}_{0-2})$.

For the numerical homogenization method, we consider $V_H^\ell(\Omega, \mathcal{T}_H)$ defined in (8), a family of triangulations $\mathcal{T}_H$ and an associated quadrature formula $\{x_{K_j}, \omega_{K_j}\}_{j=1}^J$ for each $K \in \mathcal{T}_H$. We define a macroscopic solver for $v_H, w_H \in V_H^\ell(\Omega, \mathcal{T}_H)$ by

$$B_{HMM}(v_H, w_H) = \sum_{K \in \mathcal{T}_H} \sum_{j=1}^J \omega_{K_j} a_{K_j}^h(\nabla v_H(x_{K_j})) \cdot \nabla w_H(x_{K_j}), \tag{35}$$

where for $\xi \in \mathbb{R}^d$ and $x_{K_j} \in \Omega$, $a_{K_j}^h$ is given by

$$a_{K_j}^h(\xi) = \frac{1}{\left|K_{\delta_j}\right|} \int_{K_{\delta_j}} a^\varepsilon(x, \xi + \nabla \chi_{h, K_j}^\xi) dx$$

and χ_{h, K_j}^ξ solves the nonlinear micro problem: find $\chi_{h, K_j}^\xi \in V_h^q(K_{\delta_j}, \mathcal{T}_h)$ such that

$$\int_{K_{\delta_j}} a^\varepsilon(x, \xi + \nabla \chi_{h, K_j}^\xi) \cdot \nabla z_h \, dx = 0, \quad \forall z_h \in V_h^q(K_{\delta_j}, \mathcal{T}_h).$$

The numerical homogenization method is then defined as follows: find $u_{HMM} \in V_H^\ell(\Omega, \mathcal{T}_H)$ such that

$$B_{HMM}(u_{HMM}, v_H) = \int_\Omega f v_H \, dx, \quad \forall v_H \in V_H^\ell(\Omega, \mathcal{T}_H). \tag{36}$$

Here also we introduce an intermediate problem to decompose the error into macroscopic error and "numerical homogenization errors". This intermediate problem is based on the form

$$B_H(v_H, w_H) = \sum_{K \in \mathcal{T}_H} \sum_{j=1}^{J} \omega_{K_j} \, a(x_{K_j}, \nabla v_H(x_{K_j})) \cdot \nabla w_H(x_{K_j}), \tag{37}$$

where $a(x_{K_j}, \xi)$ is the value of the nonlinear homogenized map at the nodes $x = x_{K_j}$ for $\xi \in \mathbb{R}^d$ and reads: find $u_H \in V_H^\ell(\Omega, \mathcal{T}_H)$ such that

$$B_H(u_H, v_H) = \int_\Omega f v_H dx, \quad \forall v_H \in V_H^\ell(\Omega, \mathcal{T}_H). \tag{38}$$

The existence of a solution of Problems (36) and (38) can be shown by establishing that the forms B_{HMM} and B_H are continuous in the second variable and Lipschitz continuous as well as strongly monotone in the first variable. This follows from the condition (**Q1**) and the hypotheses ($\mathcal{A}_{0-2}$). Hence Problems (36) and (38) have a unique solution (see [8] for details). Optimale convergence rates for the FE-HMM (in the H^1 and L^2 norms) for arbitrary simplicial elements for (33) have been obtained in [7]. We also mention the work [30], where convergence rates (in the H^1 norm) for P^1 macro elements have been derived for the FE-HMM for a class of elliptic monotone PDEs (associated to minimization problems)

3.1. **FEM with numerical integration for monotone problems**

The effect of numerical integration for a FEM applied to (34) has been studied in [28] for the H^1 norm in dimension $d = 2$ and for polynomial degree $\ell = 1$. Recently in [7] we proved optimal convergence rates in the H^1 and L^2 norms for FEM with numerical integration with arbitrary polynomial degree ℓ in dimension $d \leq 3$ and for simplicial FEs. For simplicity we discuss only the case of simplicial FEs in this section but emphasize that error estimates for parallelogram could be derived following the same arguments with additional technicalities. The setting (family of triangulations and quadrature formula) is the same as in Section 2.2 and we assume as usual that (14) holds. We need however a slightly more restrictive assumption (**Q2**) for the L^2 error estimate, namely

(**Q2'**) $\int_{\hat{K}} \hat{p}(\hat{x}) d\hat{x} = \sum_{j=1}^{J} \hat{\omega}_j \hat{p}(\hat{x}_j), \; \forall \hat{p}(\hat{x}) \in \mathcal{R}^\sigma(\hat{K})$, where $\sigma = 2\ell - 1$.

We discuss now the a priori error estimates for FEM with numerical integration for nonlinear monotone problems obtained in [7]. We first state the following regularity assumptions that will be used for the H^1 error estimate for $\mu = 0$ and for the L^2 error estimate for $\mu = 1$

$$\begin{aligned} &a(\cdot, \xi) \in W^{1+\mu,\infty}(\Omega), \;\; \|a(\cdot, \xi)\|_{W^{1+\mu,\infty}(\Omega)} \leq C(L_0 + |\xi|), \; \forall \xi \in \mathbb{R}^d, && \text{if } \ell = 1, \\ &a \in W^{\ell+\mu,\infty}(\Omega \times B_R(0); \mathbb{R}^d), \;\; \|a\|_{W^{\ell+\mu,\infty}(\Omega \times B_R(0); \mathbb{R}^d)} \leq C(L_0 + R), \; \forall R > 0, && \text{if } \ell \geq 2, \end{aligned} \tag{39}$$

for some $L_0 \geq 0$. The main result for FEM with numerical integration for Problem (37) has been proved in [7] and is stated in the next theorem.

Theorem 3.1. *Consider u the solution of problem* (34) *and u_H the solution of problem* (37). *Assume that a satisfies* ($\mathcal{A}_{0-2}$), *that $u \in H^{\ell+1}(\Omega)$ and that the quadrature formula satisfies* (**Q2**) *and* (39) *for $\mu = 0$. Then, we have the a priori error estimate*

$$\|u - u_H\|_{H^1(\Omega)} \leq C H^\ell, \tag{40}$$

where C is independent of H. Assume that the quadrature formula satisfies (**Q2'**) *and* (39) *for* $\mu = 1$ *and in addition*

$$a(x,\cdot) \in W^{2,\infty}(\mathbb{R}^d; \mathbb{R}^d) \text{ with } \|D_\xi a(x,\cdot)\|_{W^{1,\infty}(\mathbb{R}^d; \mathbb{R}^{d\times d})} \leq L_a \text{ for a.e. } x \in \Omega,$$

$$u \in H^{l+1}(\Omega) \cap W^{2,\infty}(\Omega), \qquad \mathfrak{a}_{ij} \in W^{1,\infty}(\Omega), \text{ for } 1 \leq i, j \leq d, \tag{41}$$

where $\mathfrak{a}(x) = D_\xi a(x, \nabla u(x))$ *and* $L_a > 0$. *Then, assuming quasi-uniform meshes, there exists* H_0 *such that for all* $H < H_0$ *the following estimate holds*

$$\|u - u_H\|_{L^2(\Omega)} \leq CH^{\ell+1}. \tag{42}$$

We explain the main steps of the proof.

Step 1. For both the L^2 and the H^1 error estimate we consider the decomposition

$$\|u - u_H\|_{H^{1-\mu}(\Omega)} \leq \|u - \mathcal{U}_H\|_{H^{1-\mu}(\Omega)} + C\|\nabla\theta_H\|_{L^2(\Omega)}, \tag{43}$$

where C depends on the Poincaré constant. We set $\theta_H = \mathcal{U}_H - u_H$, where $\mathcal{U}_H \in V_H^\ell(\Omega, \mathcal{T}_H)$ will be either the nodal interpolant $\mathcal{I}_H u$ for the error in the H^1 norm, or an elliptic projection $\pi_H(u)$ described below for the error in the L^2 norm. Using the strong monotonicity of B_H

$$\lambda\|\nabla\theta_H\|_{L^2(\Omega)}^2 \leq B_H(u_H, \theta_H) - B_H(\mathcal{U}_H, \theta_H)$$
$$= B(u, \theta_H) - B(\mathcal{U}_H, \theta_H) + B(\mathcal{U}_H, \theta_H) - B_H(\mathcal{U}_H, \theta_H). \tag{44}$$

Step 2. We need an error estimate for $|B(\mathcal{U}_H, \theta_H) - B_H(\mathcal{U}_H, \theta_H)|$ in (44). This can be estimated as follows. Let $\mu \in \{0, 1\}$ and assume (**Q2**) or (**Q2'**) and the regularity (39) of a for $\mu \in \{0, 1\}$. Then, we have

$$|B(v_H, w_H) - B_H(v_H, w_H)| \leq CH^{l+\mu}\mathcal{D}(v_H)\|\nabla w_H\|_{L^2(\Omega)}, \qquad \forall v_H, w_H \in V_H^\ell(\Omega, \mathcal{T}_H), \tag{45}$$

where C is independent of H. The term $\mathcal{D}(v_H)$ is given by

$$\mathcal{D}(v_H) = L_0 + \|\nabla v_H\|_{L^2(\Omega)}, \qquad\qquad\qquad \text{if } l = 1,$$

$$\mathcal{D}(v_H) = \left(L_0 + |\nabla v_H|_{L^\infty(\Omega)}\right)\left(1 + \|v_H\|_{\bar{W}^{l-1,\infty}(\Omega)}^{l+\mu}\right)\left(1 + \sum_{\kappa=1}^{\kappa_l} |v_H|_{\bar{W}^{l,2\kappa}(\Omega)}^\kappa\right), \qquad \text{if } l \geq 2,$$

where L_0 is the constant from (39) and $\kappa_2 = 2 + \mu$, $\kappa_3 = 1 + \mu$ and $\kappa_l = 1$ for $l \geq 4$.

For the proof of (45) we consider the local (componentwise) quadrature error. As $\nabla v_H, \nabla w_H \in (\mathcal{P}^{\ell-1}(K))^d$, we consider for $\mathbf{p}(x) = (p_1(x), \ldots, p_d(x))^T \in (\mathcal{P}^{\ell-1}(K))^d$ and $q \in \mathcal{P}^{\ell-1}(K)$ and the quadrature error functional

$$E_K^i(a, \mathbf{p}, q) = \int_K a_i(x, \mathbf{p}(x))q(x)dx - \sum_{j=1}^J \omega_{K_j} a_i(x_{K_j}, \mathbf{p}(x_{K_j}))q(x_{K_j}).$$

We next transfer this error on the reference element $\hat{K}$ similarly to (22), use (**Q2**) and the Bramble-Hilbert lemma to obtain

$$\left|\hat{E}(\hat{a}, \hat{\mathbf{p}}, \hat{q})\right| \leq C|\hat{a}(\cdot, \hat{\mathbf{p}}(\cdot))|_{W^{l+\mu,\infty}(\hat{K})}\|\hat{q}\|_{L^2(\hat{K})}. \tag{46}$$

In contrast to the linear case, some care is needed to bound the term $|\hat{a}(\cdot, \hat{\mathbf{p}})|_{W^{l+\mu,\infty}(\hat{K})}$ in terms of derivatives of $\hat{\mathbf{p}}$ and terms of the type $\|a\|_{W^{l+\mu,\infty}(K \times B_R(0))}\left(1 + \|\mathbf{p}\|_{W^{l-2,\infty}(K)}^{l+\mu}\right)$. For that we use the Faà-di-Bruno formula,

the equivalence of norm on the finite dimensional space $(\mathcal{P}^{l-1}(\hat{K}))^d$ and estimates of the type (26)-(27).

Step 3 (H^1 error estimate). Using the Lipschitz continuity of B in (44) we obtain

$$\lambda\|\nabla\theta_H\|_{L^2(\Omega)}^2 \le L\|\nabla u - \nabla\mathcal{I}_H u\|_{L^2(\Omega)}\|\nabla\theta_H\|_{L^2(\Omega)} + |B(\mathcal{I}_H u, \theta_H) - B_H(\mathcal{I}_H u, \theta_H)|.$$

We use (45) with $v_H = \mathcal{I}_H$ (for $\mu = 0$) and the standard interpolation estimate $\|\nabla u - \nabla\mathcal{I}_H u\|_{L^2(\Omega)} \le CH^\ell\|u\|_{H^{l+1}(\Omega)}$. It remains to bound $\mathcal{D}(\mathcal{I}_H u)$. Since $u \in H^{l+1}(\Omega)$, it follows from the Sobolev embeddings $H^{l+1}(\Omega) \hookrightarrow W^{l,4}(\Omega), W^{l-1,\infty}(\Omega)$, which hold for $d \le 3$, and from the interpolation estimate [22, Theorems 15.3] that $\mathcal{D}(\mathcal{I}_H u) \le C\|u\|_{H^{l+1}(\Omega)}$ for $l \ge 1$. Finally (43) and (44) yield (40).

Step 3 (L^2 error estimate). We want to obtain an error $\lambda\|\nabla\theta_H\|_{L^2(\Omega)}^2 \le CH^{\ell+1}\|\nabla\theta^H\|_{L^2(\Omega)}$, where C depends on some Sobolev norm of the exact solution u. The key ingredient is the following linear elliptic projection defined by: find $\pi_H(u) \in V_H^\ell(\Omega, \mathcal{T}_H)$ solution of

$$B_\pi(\pi_H(u), w_H) = B_\pi(u, w_H), \quad \forall\, w_H \in V_H^\ell(\Omega, \mathcal{T}_H), \tag{47}$$

where the bilinear form B_π, for $v, w \in H_0^1(\Omega)$, is given by

$$B_\pi(v, w) = \int_\Omega \mathfrak{a}(x)\nabla v \cdot \nabla w \, dx, \quad \text{with } \mathfrak{a}(x) = D_\xi a(x, \nabla u(x)) \text{ for a.e. } x \in \Omega. \tag{48}$$

The above variational problem is a linear elliptic problem. If $a\colon \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ satisfies $(\mathcal{A}_{1-2})$ and $a(x, \cdot) \in \mathcal{C}^1(\mathbb{R}^d; \mathbb{R}^d)$ for a.e. $x \in \Omega$, then, $D_\xi a(x, \xi)$ is uniformly elliptic and bounded (for details see [8, Sect. 5.1]). Hence we have existence and uniqueness of $\pi_H(u)$. We have also for quasi-uniform meshes the following bounds

$$\|\pi_H(u)\|_{\bar{H}^{l+1}(\Omega)} + \|\pi_H(u)\|_{\bar{W}^{l,6}(\Omega)} + \|\pi_H(u)\|_{\bar{W}^{l-1,\infty}(\Omega)} \le C\|u\|_{H^{l+1}(\Omega)}, \tag{49}$$

where C is independent of H (see [12, Lemma 1] for a proof). A Taylor expansion of $a(x, \cdot)$, using the definition (47) of the elliptic projection $\pi_H(u)$ yields

$$B(\pi_H(u); \theta_H) - B(u, \theta_H) = P(u, \pi_H(u), w_H), \tag{50}$$

where $P(u, \pi_H(u); w_H) = \int_\Omega \int_0^1 D_\xi a(x, \nabla u + \tau(\nabla\pi_H(u) - \nabla u)) - D_\xi a(x, \nabla u)d\tau(\nabla\pi_H(u) - \nabla u) \cdot \nabla w_H dx$. This term can be estimated using the Lipschitz continuity of $D_\xi a(x, \cdot)$ and we obtain

$$P(u, \pi_H(u); w^H) \le L_a\|u - \pi_H(u)\|_{W^{1,\infty}(\Omega)}\|\nabla u - \nabla\pi_H(u)\|_{L^2(\Omega)}\|\nabla w^H\|_{L^2(\Omega)}. \tag{51}$$

The maximum norm error estimate for linear FEM from [21, Sect. 8] yields $\|u - \pi_H(u)\|_{W^{1,\infty}(\Omega)} \le CH\|u\|_{W^{2,\infty}(\Omega)}$, a standard estimate for linear elliptic problem gives $\|\nabla u - \nabla\pi_H(u)\|_{L^2(\Omega)} \le CH^l|u|_{H^{l+1}(\Omega)}$. Using (49) allows to obtain the bound $\mathcal{D}(\pi_H(u)) \le C\|u\|_{H^{l+1}(\Omega)}$. Finally using (43) and (44) allows to obtain the error bound (42) in the L^2 norm.

4. Numerical homogenization and numerical integration for nonlinear nonmonotone elliptic problems

In this section we review FEMs with numerical integration and numerical homogenization for nonlinear nonmonotone elliptic problems of the form

$$L^\varepsilon(u^\varepsilon, a^\varepsilon) = -\nabla \cdot (a^\varepsilon(x, u^\varepsilon(x))\nabla u^\varepsilon(x)) = f(x) \text{ in } \Omega, \qquad u^\varepsilon(x) = 0 \text{ on } \partial\Omega. \tag{52}$$

We assume that the tensors $a^\varepsilon(x, s) = (a_{mn}^\varepsilon(x, s))_{1 \le m,n \le d}$ satisfy the following conditions uniformly in $\varepsilon > 0$

($\mathcal{A}_1$) there exists $\Lambda_1 > 0$ such that $|a_{mn}^\varepsilon(x, s_1) - a_{mn}^\varepsilon(x, s_2)| \leq \Lambda_1 |s_1 - s_2|$, $\forall x \in \overline{\Omega}, \forall s_1, s_2 \in \mathbb{R}$,
$\forall\, 1 \leq m, n \leq d$;

($\mathcal{A}_2$) there exist $\lambda, \Lambda_0 > 0$ such that $\lambda \|\xi\|^2 \leq a^\varepsilon(x, s)\xi \cdot \xi$, $\|a^\varepsilon(x, s)\xi\| \leq \Lambda_0 \|\xi\|$, $\forall \xi \in \mathbb{R}^d, \forall x \in \overline{\Omega}, \forall s \in \mathbb{R}$.

Under these assumptions there exists a unique solution $u^\varepsilon \in H_0^1(\Omega)$ of problem (52). Then, homogenization theory [19, Theorem 3.6] ensures the existence of a subsequence $\{u_\varepsilon\}$ that converges weakly to u in $H_0^1(\Omega)$ solution of the homogenized problem that reads in weak form

$$B(u; u, v) = \int_\Omega a(x, u(x)) \nabla u \cdot \nabla v\, dx = \int_\Omega f v\, dx, \quad \forall v \in H_0^1(\Omega), \tag{53}$$

where $a(x, s)$ is the homogenized tensor that satisfies again ($\mathcal{A}_{1-2}$). For the numerical homogenization method, we consider $V_H^\ell(\Omega, \mathcal{T}_H)$ as previously defined (see (8)), a family of triangulations $\mathcal{T}_H$ and an associated quadrature formula $\{x_{K_j}, \omega_{K_j}\}_{j=1}^J$ for each $K \in \mathcal{T}_H$. We define a macroscopic solver for $z_H, v_H, w_H \in V_H^\ell(\Omega, \mathcal{T}_H)$ by

$$B_{HMM}(z_H; v_H, w_H) = \sum_{K \in \mathcal{T}_H} \sum_{j=1}^J \omega_{K_j} a_{K_j}(z_H(x_{K_j})) \nabla v_H(x_{K_j}) \cdot \nabla w_H(x_{K_j}), \tag{54}$$

where for $s \in \mathbb{R}$, $\xi \in \mathbb{R}^d$ and $x_{K_j} \in \Omega$, $a_{K_j}^h(s) \in \mathbb{R}^{d \times d}$ is given by

$$a_{K_j}^h(s)\xi = \frac{1}{|K_{\delta_j}|} \int_{K_{\delta_j}} a^\varepsilon(x, s)(\xi + \nabla \chi_{h, K_j}^{\xi, s})dx$$

and $\chi_{h, K_j}^{\xi, s}$ solves the following micro problem: find $\chi_{h, K_j}^{\xi, s} \in V_h^q(K_{\delta_j}, \mathcal{T}_h)$ such that

$$\int_{K_{\delta_j}} a^\varepsilon(x, s)(\xi + \nabla \chi_{h, K_j}^{\xi, s})) \cdot \nabla z_h\, dx = 0, \quad \forall z_h \in V_h^q(K_{\delta_j}, \mathcal{T}_h).$$

The numerical homogenization method is then defined as follows: find $u_{HMM} \in V_H^\ell(\Omega, \mathcal{T}_H)$ such that

$$B_{HMM}(u_{HMM}; u_{HMM}, v_H) = \int_\Omega f v_H\, dx, \quad \forall v_H \in V_H^\ell(\Omega, \mathcal{T}_H). \tag{55}$$

As usual we introduce an intermediate problem to decompose the error into macroscopic error and "numerical homogenization errors". This intermediate problem is based on the form

$$B_H(z_H; v_H, w_H) = \sum_{K \in \mathcal{T}_H} \sum_{j=1}^J \omega_{K_j} a(x_{K_j}, z_H(x_{K_j})) \nabla v_H(x_{K_j}) \cdot \nabla w_H(x_{K_j}), \tag{56}$$

where $a(x_{K_j}, s)$ is the value of the nonlinear homogenized map at the nodes $x = x_{K_j}$ for $s \in \mathbb{R}$ and reads: find $u_H \in V_H^\ell(\Omega, \mathcal{T}_H)$ such that

$$B_H(u_H; u_H, v_H) = \int_\Omega f v_H\, dx, \quad \forall v_H \in V_H^\ell(\Omega, \mathcal{T}_H). \tag{57}$$

Using (**Q1**), ($\mathcal{A}_1$) and ($\mathcal{A}_2$), it can be shown that the bilinear form $B_{HMM}(z_H; \cdot, \cdot), B_H(z_H; \cdot, \cdot)$ are elliptic and bounded, i.e., there exist two constants $c_1, c_2 > 0$ such that

$$\|v_H\|_{H^1(\Omega)}^2 \leq c_1 B_*(z_H; v_H, v_H), \quad |B_*(z_H; v_H, w_H)| \leq c_2 \|v_H\|_{H^1(\Omega)} \|w_H\|_{H^1(\Omega)}, \tag{58}$$

for all $z_H, v_H, w_H \in V_H^\ell(\Omega, \mathcal{T}_H)$, where $B_*(\cdot;\cdot,\cdot) = B_{HMM}(\cdot;\cdot,\cdot)$ or $B_H(\cdot;\cdot,\cdot)$. The existence of a solution of Problems (55) and (57), respectively, can can be shown by using the Brouwer fixed point theorem (see [12,13]). The uniqueness of the solution is however more involved (see below).

4.1. FEM with numerical integration for nonmonotone problems

The effect of numerical integration for a FEM applied to (53) has first been studied in [27] for piecewise linear FEs. Convergence of the FE method *without convergence rates* has been obtained in the $H^1(\Omega)$ norm. Optimal convergence rates in the H^1 and L^2 norms for FEM with numerical integration with arbitrary polynomial degree ℓ in dimension $d \leq 3$ for simplicial and parallelogram FEs have been obtained in [12].

We consider a family of triangulations of simplicial or parallelogram FEs. In addition to the shape regularity hypothesis (14), we assume that the family of triangulations satisfies the inverse assumption

$$\frac{H}{H_K} \leq C \ \text{ for all } K \in \mathcal{T}_H \text{ and all } \mathcal{T}_H \text{ of the family of triangulations,} \qquad (59)$$

where $H = \max_{K \in \mathcal{T}_H} H_K$. The main result for FEM with numerical integration for Problem (53) has been proved in [12] and is stated in the next theorem.

Theorem 4.1. *Consider u the solution of problem* (53). *Let $\ell \geq 1$. Let $\mu = 0$ or 1. Assume* (**Q1**), (**Q2**), (59), *and*

$$u \in H^{\ell+1}(\Omega) \cap W^{1,\infty}(\Omega),$$
$$a_{mn} \in W^{\ell+\mu,\infty}(\Omega \times \mathbb{R}), \qquad\qquad \forall m, n = 1, \ldots, d.$$

In addition to $(\mathcal{A}_1),(\mathcal{A}_2)$, *assume that $\partial_s a_{mn} \in W^{1,\infty}(\Omega \times \mathbb{R})$, and that the coefficients $a_{mn}(x,s)$ are twice differentiable with respect to s, with the first and second order derivatives continuous and bounded on $\overline{\Omega} \times \mathbb{R}$, for all $m, n = 1, \ldots, d$.*

Then there exists $H_0 > 0$ such that for all $H \leq H_0$, the solution u_H of (57) *is unique, and the following H^1 and L^2 error estimates hold,*

$$if\ \mu = 0, \qquad \|u - u_H\|_{H^1(\Omega)} \leq CH^\ell \qquad for\ all\ H \leq H_0, \qquad (60)$$

$$if\ \mu = 1, \qquad \|u - u_H\|_{L^2(\Omega)} \leq CH^{\ell+1} \qquad for\ all\ H \leq H_0, \qquad (61)$$

where the constants C are independent of H.

Remark 4.2. We note that if u satisfies a smallness assumption, i.e., $C\lambda^{-1}\Lambda_1\|u\|_{H^2(\Omega)} < 1$ (where C depends only on Ω and $(V_H^\ell(\Omega, \mathcal{T}_H))_{H>0}$), then the uniqueness of a numerical solution u_H of (57) and the H^1 estimates can be obtained by assuming only $u \in H^{\ell+1}(\Omega)$, $a_{mn} \in W^{\ell,\infty}(\Omega \times \mathbb{R})$, $\forall m, n = 1, \ldots, d$ and $(\mathcal{A}_1),(\mathcal{A}_2)$. The additional regularity assumption on a_{mn}, u and the assumption (59) are not needed. In view of Theorem 2.2, we see that under this smallness assumption we obtain the same convergence results as for linear problems with the same hypothesis.

One main difficulty to obtain convergence rates for the FE method for this class of nonlinear problems is that the nonlinear form $(v, w) \longrightarrow B(v; v, w)$ is not monotone in general, i.e.

$$B(v; v, v - w) - B(w; w, v - w) \geq C\|v - w\|_{H^1(\Omega)}^2,$$

does *not* hold in general even with $C = 0$ (take for example $a(x, u) = b(u)I$, where I is the $d \times d$ identity matrix and b a scalar function satisfying $s_0 b'(s_0) + b(s_0) < 0$ for some real s_0). Such an estimate is however the first step to obtain a Strang like lemma in the convergence results for linear or monotone problems (see Sections 2 and 3). Other techniques are needed here and the convergence results rely on a compactness argument together with simultaneous $H^1 - L^2$ estimates. This was first suggested in [24] for FEM without numerical quadrature.

Step 1. We first need to show that

$$\|u - u_H\|_{L^2(\Omega)} \to 0 \text{ for } H \to 0. \tag{62}$$

For that we need an estimate of the form

$$|B_H(z_H; v_H, w_H) - B(z_H; v_H, w_H)| \leq CH\|\nabla v_H\|_{L^2(\Omega)}(\|\nabla w_H\|_{\bar{H}^1(\Omega)} + \|\nabla z_H\|_{L^\alpha(\Omega)}\|\nabla w_H\|_{L^\beta(\Omega)}), \tag{63}$$

where C is independent of H and $1 \leq \alpha, \beta \leq \infty$ and $1/\alpha + 1/\beta = 1/2$. Next using the boundedness of a numerical solution in $H_0^1(\Omega)$, the uniqueness in $H_0^1(\Omega)$ of the exact solution of (53) and the compact injection $H^1(\Omega) \subset L^2(\Omega)$, estimate (63) with $\alpha = 2, \beta = \infty$ one can show (62).

Step 2. We want to derive the following $H^1 - L^2$ estimate

$$\|u - u_H\|_{H^1(\Omega)} \leq C(H^\ell + \|u - u_H\|_{L^2(\Omega)}), \text{ for all } H > 0. \tag{64}$$

We note an unusual additional term $\|u - u_H\|_{L^2(\Omega)}$ in the right-hand side of (64), which is due to the non-monotonicity of the differential operator (53). Hence already for the H^1 error estimate we need an error estimate in the L^2 norm.

We start by deriving the following estimates

$$|B_H(\mathcal{I}_H u; \mathcal{I}_H u, w_H) - B(\mathcal{I}_H u; \mathcal{I}_H u, w_H)| \leq CH^\ell\|w_H\|_{H^1(\Omega)}, \tag{65}$$

where C depends on $\|a\|_{(W^{\ell,\infty}(\Omega\times\mathbb{R}))^{d\times d}}$ and $\|u\|_{H^{\ell+1}(\Omega)}$ but is independent of H. Here $\mathcal{I}_H u$ denotes the usual nodal interpolant of u in $V_H^\ell(\Omega, \mathcal{T}_H)$. To prove (65), we need to pull back the error estimates on each K to the reference element $\hat{K}$, use the Bramble-Hilbert lemma, the Faà-di-Bruno formula and the equivalence of norm on finite dimensional spaces to obtain estimates on K via the affine transformation $K = F_K(\hat{K})$. Next we use the ellipticity of $B_H(z_H; \cdot, \cdot)$ to obtain

$$
\begin{aligned}
C\|\xi_H\|_{H^1(\Omega)}^2 \leq B_H(u_H; u_H - \mathcal{I}_H u, \xi_H) &= B(u; u - \mathcal{I}_H u, \xi_H) \\
&+ B(u; \mathcal{I}_H u, \xi_H) - B(\mathcal{I}_H u; \mathcal{I}_H u, \xi_H) \\
&+ B(\mathcal{I}_H u; \mathcal{I}_H u, \xi_H) - B_H(\mathcal{I}_H u; \mathcal{I}_H u, \xi_H) \\
&+ B_H(\mathcal{I}_H u; \mathcal{I}_H u, \xi_H) - B_H(u_H; \mathcal{I}_H u, \xi_H),
\end{aligned} \tag{66}
$$

where $\xi_H = u_H - \mathcal{I}_H u$. For the first term we use the boundedness of B and an interpolation estimate to obtain the bound $CH^\ell\|u\|_{H^{\ell+1}(\Omega)}\|\nabla \xi_H\|_{L^2(\Omega)}$. For the second term, we use $(\mathcal{A}_1)$ and observe that $(u - \mathcal{I}_H u) \in L^3(\Omega)$ thanks to the continuous embedding $H^1(\Omega) \subset L^3(\Omega)$ (recall that $d \leq 3$). We then obtain

$$\|a(\cdot, u) - a(\cdot, \mathcal{I}_H u))\nabla \mathcal{I}_H u\|_{L^2(\Omega)} \leq C\|u - \mathcal{I}_H u\|_{L^3(\Omega)}\|\mathcal{I}_H u\|_{W^{1,6}(\Omega)},$$

where we used Hölder inequality and the continuous embedding $H^1(\Omega) \subset L^6(\Omega)$ for $\nabla \mathcal{I}_H u \in (H^1(\Omega))^d$. We finally find

$$|B(u; \mathcal{I}_H u, \xi_H) - B(\mathcal{I}_H u; \mathcal{I}_H u, \xi_H)| \leq C\|u - \mathcal{I}_{\mathcal{H}}\|_{L^3(\Omega)}\|\mathcal{I}_H u\|_{W^{1,6}(\Omega)}\|\nabla \xi_H\|_{L^2(\Omega)},$$

that can be bounded by $CH^\ell\|u\|_{H^{\ell+1}(\Omega)}\|\nabla \xi_H\|_{L^2(\Omega)}$, using again the above continuous Sobolev embedding and classical interpolation estimates [22, Theorem 16.2]. Using similar arguments for the fourth term (via mesh dependent semi-norms) we obtain

$$|B_H(\mathcal{I}_H u; \mathcal{I}_H u, \xi_H) - B_H(u_H; \mathcal{I}_H u, \xi_H)| \leq C\|\xi_H\|_{L^3(\Omega)}\|\mathcal{I}_H u\|_{W^{1,6}(\Omega)}\|\nabla \xi_H\|_{L^2(\Omega)}.$$

For the third term we use (65) and obtain the bound $CH^\ell\|u\|_{H^{\ell+1}(\Omega)}\|\nabla\xi_H\|_{L^2(\Omega)}$. Using all these estimates in (66) yield

$$\|\xi_H\|_{H^1(\Omega)} \leq C(H^\ell + \|\xi_H\|_{L^3(\Omega)}).$$

Finally combining the Gagliardo-Nirenberg inequality $\|v\|_{L^3(\Omega)} \leq C\|v\|_{L^2(\Omega)}^{1/2}\|v\|_{H^1(\Omega)}^{1/2}$ (valid for functions in H^1 for $d \leq 3$) and Young inequality gives $\|\xi_H\|_{H^1(\Omega)} \leq C(H^\ell + \|\xi_H\|_{L^2(\Omega)})$. Finally the triangular inequalities $\|u - u_H\| \leq \|u - \mathcal{I}_H u\| + \|\xi_H\|$, $\|\xi_H\| \leq \|u - u_H\| + \|u - \mathcal{I}_H u\|$ gives the H^1 a priori error bound (64).

Step 3. For the L^2 estimate we first derive similarly as for (65) the estimate

$$|B_H(\pi_H(u); \pi_H(u), w_H) - B(\pi_H(u); \pi_H(u), w_H) \quad \leq \quad CH^{\ell+1}(\|w_H\|_{\bar{H}^2(\Omega)} + \|w_H\|_{W^{1,6}(\Omega)}) \qquad (67)$$

where $\pi_H(u)$ is the L^2 projection and where C depends on $\|a\|_{(W^{\ell+1,\infty}(\Omega\times\mathbb{R}))^{d\times d}}$ and $\|u\|_{H^{\ell+1}(\Omega)}$ but is independent of H. The other key ingredient is an Aubin-Nitsche duality argument, where we consider the adjoint $\mathcal{L}^*$ of the linearized operator associated to (53) given

$$\mathcal{L}\varphi = -\nabla \cdot (a(\cdot, u)\nabla\varphi + \varphi\partial_s a(\cdot, u)\nabla u).$$

We also consider its discrete counterpart denoted by f $\mathcal{L}_H$ (linearized at $\mathcal{I}_H u$). Let g be an arbitrary function in $L^2(\Omega)$. A crucial step is then an estimate between the solution $\varphi \in H_0^1(\Omega)$ of the problem $(\mathcal{L}\varphi, w) = (g, w)$, for all $w \in H_0^1(\Omega)$ and the FEM solution $\varphi_H \in V_H^\ell(\Omega, \mathcal{T}_H)$ of the problem $(\mathcal{L}_H\varphi_H, w_H) = (g, w_H)$, for all $w_H \in V_H^\ell(\Omega, \mathcal{T}_H)$. This is obtained using a compactness argument of Schatz [36]. Under the assumptions of Theorem 4.1 and for H small enough, the above problems have a unique solution $\varphi \in H^2(\Omega) \cap H_0^1(\Omega)$, $\varphi_H \in V_H^\ell(\Omega, \mathcal{T}_H)$ and the following a priori error estimates are valid

$$\|\varphi - \varphi_H\|_{H^1(\Omega)} \leq CH\|g\|_{L^2(\Omega)}, \qquad (68)$$

$$\|\varphi_H\|_{\bar{H}^2(\Omega)} + \|\varphi_H\|_{W^{1,6}(\Omega)} \leq C\|g\|_{L^2(\Omega)}, \qquad (69)$$

where C is independent of H. It is then possible to show that there exists $H_1 > 0$ such that for $\mu = 0, 1$ we have

$$\|u - u_H\|_{L^2(\Omega)} \leq C(H^{\ell+\mu} + \|u - u_H\|_{H^1(\Omega)}^2), \text{ for all } H \leq H_1. \qquad (70)$$

Step 4. To prove the estimates (60),(61), we combine the H^1 and L^2 estimates. Substituting (64) into (70) (with $\mu = 0$) yields

$$\|u - u_H\|_{H^1(\Omega)} \leq C(H^\ell + \|u - u_H\|_{H^1(\Omega)}^2), \text{ for all } H \leq H_1.$$

Substituting (62) into (64), we obtain $\|u - u_H\|_{H^1(\Omega)} \to 0$ for $H \to 0$. There exists thus H_2 small enough (but independent of the particular solution u_H) such that for all $H \leq H_2$ we have $0 < 1/2 \leq 1 - C\|u - u_H\|_{H^1(\Omega)}$, hence (60) is established for all $H \leq H_0 = \min\{H_1, H_2\}$. The estimate (61) is deduced by substituting (60) into (70) with $\mu = 1$.

Step 5. The uniqueness of the numerical solution is proved via convergence of the Newton method. This generalizes [24] to the case of FEMs with numerical integration. Given an initial guess $z_{H,0} \in V_H^\ell(\Omega, \mathcal{T}_H)$, the sequence $(z_{H,k})$ of a Newton method is defined by

$$N_H(z_{H,k}; z_{H,k+1} - z_{H,k}, v_H) = \int_\Omega f v_H dx - B_H(z_{H,k}; z_{H,k}, v_H), \quad \forall v_H \in V_H^\ell(\Omega, \mathcal{T}_H), \qquad (71)$$

with

$$N_H(z_H; v_H, w_H) = B_H(z_H; v_H, w_H) + (v_H\partial_s a(\cdot, z_H)\nabla z_H, \nabla w_H)_H,$$

where we use the notation $(v, w)_H = \sum_{K\in\mathcal{T}_H}\sum_{j=1}^J \omega_{K_j} v(x_{K_j}) \cdot w(x_{K_j})$ (defined for piecewise continuous functions v, w). Consider for all H the quantity $\sigma_H = \sup_{v_H\in V_H^\ell(\Omega, \mathcal{T}_H)} \|v_H\|_{L^\infty(\Omega)}/\|v_H\|_{H^1(\Omega)}$. Using (59), an inverse

inequality and the continuous injection $H^1(\Omega) \subset L^p(\Omega)$ (with $p = 6$ for $d = 3$ and $1 \leq p < \infty$ for $d = 2$) one can show the estimates

$$\sigma_H \leq C(1 + |\ln H|)^{1/2} \text{ for } d = 2, \quad \sigma_H \leq CH^{-1/2} \text{ for } d = 3,$$

where C is independent of H. Consider u_H solution of (57) and define $e_k := \|z_{H,k} - u_H\|_{H^1(\Omega)}$ for the sequence $\{z_{H,k}\}$ from (71). Then under the assumptions of Theorem 4.1, there exist $H_0, \rho > 0$ such that if $H \leq H_0$ and $\sigma_H \|z_{H,0} - u_H\|_{H^1(\Omega)} \leq \rho$ the sequence $\{z_{H,k}\}$ for the Newton method (71) is well-defined and e_k is a decreasing sequence that converges quadratically to 0 for $k \to \infty$,

$$e_{k+1} \leq C\sigma_H e_k^2, \tag{72}$$

where C is a constant independent of H, k. The convergence (72) allows to show the uniqueness of the numerical solution of Problem (57). Indeed, given two solutions $u_H, \widetilde{u}_H$ of (57), we consider the Newton method with initial value $z_{H,0} = \widetilde{u}_H$. Then, on one hand, $z_{H,k} = z_{H,0}$ for all k (as $\widetilde{u}_H$ solves (57)). On the other hand, $\sigma_H \|\widetilde{u}_H - u_H\|_{H^1(\Omega)} \leq C\sigma_H H^\ell \to 0$ (as both $\widetilde{u}_H, u_H$ satisfy (60)). Hence (72) shows $u_H = \widetilde{u}_H$ for all $H \leq H_3$ where H_3 is small enough (observe that $H_3 \leq H_0$, where H_0 is defined in Theorem 4.1).

5. Conclusion

In this paper we have explained the role of FEM with numerical integration for numerical homogenization methods. We have seen that variational crimes cannot be avoided and are built in the very definition of the averaging process as effective data can only be recovered at finite locations of the computational domain. We have reviewed numerical homogenization methods for monotone elliptic problems (both linear and nonlinear) and nonmonotone problems and highlighted the various steps needed to derive a priori estimates in the H^1 and L^2 norm. For nonlinear problems (in particular for nonmonotone problems) the analysis requires additional tools to the interpolation results, Strang and Bramble-Hilbert lemmas that are the cornerstones for the error estimates in the linear case.

References

[1] A. Abdulle. On a priori error analysis of fully discrete heterogeneous multiscale FEM. *Multiscale Model. Simul.*, 4(2):447–459, 2005.

[2] A. Abdulle. Heterogeneous multiscale methods with quadrilateral finite elements. In A. B. de Castro, D. Gómez, P. Quintela, and P. Salgado, editors, *Numerical Mathematics and Advanced Applications*, pages 743–751. Springer Berlin Heidelberg, 2006.

[3] A. Abdulle. The finite element heterogeneous multiscale method: a computational strategy for multiscale PDEs. In *Multiple scales problems in biomathematics, mechanics, physics and numerics*, volume 31 of *GAKUTO Internat. Ser. Math. Sci. Appl.*, pages 133–181. Gakkōtosho, Tokyo, 2009.

[4] A. Abdulle. A priori and a posteriori error analysis for numerical homogenization: a unified framework. *Ser. Contemp. Appl. Math. CAM*, 16:280–305, 2011.

[5] A. Abdulle and Y. Bai. Fully discrete analysis of the heterogeneous multiscale method for elliptic problems with multiple scales. *IMA Journal of Numerical Analysis*, 2014.

[6] A. Abdulle and O. Budáč. An adaptive finite element heterogeneous multiscale method for the Stokes problem in porous media. to appear in SIAM MMS.

[7] A. Abdulle and M. E. Huber. Error estimates for finite element approximations of nonlinear monotone elliptic problems with application to numerical homogenization. MATHICSE Technical Report 36.2014, École Polytechnique Fédérale de Lausanne.

[8] A. Abdulle and M. E. Huber. Finite element heterogeneous multiscale method for nonlinear monotone parabolic homogenization problems. MATHICSE Technical Report 31.2014, École Polytechnique Fédérale de Lausanne.

[9] A. Abdulle and O. Jecker. An optimization-based multiscale coupling method. MATHICSE Technical Report 27.2014, École Polytechnique Fédérale de Lausanne.

[10] A. Abdulle and C. Schwab. Heterogeneous multiscale FEM for diffusion problems on rough surfaces. *Multiscale Model. Simul.*, 3(1):195–220, 2005.

[11] A. Abdulle and G. Vilmart. The effect of numerical integration in the finite element method for nonmonotone nonlinear elliptic problems with application to numerical homogenization methods. *C. R. Acad. Sci. Paris, Ser. I*, 349(19–20):1041–1046, 2011.

[12] A. Abdulle and G. Vilmart. A priori error estimates for finite element methods with numerical quadrature for nonmonotone nonlinear elliptic problems. *Numer. Math.*, 121(3):397–431, 2012.

[13] A. Abdulle and G. Vilmart. Analysis of the finite element heterogeneous multiscale method for quasilinear elliptic homogenization problems. *Math. Comp.*, 83(286):513–536, 2014.

[14] J.-P. Aubin. Behavior of the error of the approximate solutions of boundary value problems for linear elliptic operators by Galerkin's and finite difference methods. *Ann. Scuola Norm. Sup. Pisa (3)*, 21:599–637, 1967.

[15] I. Babuška. Homogenization and its application. Mathematical and computational problems. In *Numerical solution of partial differential equations, III (Proc. Third Sympos. (SYNSPADE), Univ. Maryland, College Park, Md., 1975)*, pages 89–116, New York, 1976. Academic Press.

[16] I. Babuška and R. Lipton. Optimal local approximation spaces for generalized finite element methods with application to multiscale problems. *Multiscale Model. Simul.*, 9:373–406, 2011.

[17] G. Baker and V. Dougalis. The effect of quadrature errors on finite element approximations for second order hyperbolic equations. *SIAM J. Numer. Anal.*, 13(4), 1976.

[18] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic analysis for periodic structures*. North-Holland Publishing Co., Amsterdam, 1978.

[19] L. Boccardo and F. Murat. Homogénéisation de problèmes quasi-linéaires. *Publ. IRMA, Lille*, 3(7):13–51, 1981.

[20] J. H. Bramble and S. R. Hilbert. Estimation of linear functionals on Sobolev spaces with application to Fourier transforms and spline interpolation. *SIAM J. Numer. Anal.*, 7:112–124, 1970.

[21] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[22] P. G. Ciarlet. Basic error estimates for elliptic problems. In *Handbook of numerical analysis, Vol. II*, Handb. Numer. Anal., II, pages 17–351. North-Holland, Amsterdam, 1991.

[23] P. G. Ciarlet and P. A. Raviart. The combined effect of curved boundaries and numerical integration in isoparametric finite element methods. In *The mathematical foundations of the finite element method with applications to partial differential equations*, pages 409–474, 1972.

[24] J. Douglas, Jr. and T. Dupont. A Galerkin method for a nonlinear Dirichlet problem. *Math. Comp.*, 29(131):689–696, 1975.

[25] W. E and B. Engquist. The heterogeneous multiscale methods. *Commun. Math. Sci.*, 1(1):87–132, 2003.

[26] W. E, P. Ming, and P. Zhang. Analysis of the heterogeneous multiscale method for elliptic homogenization problems. *J. Amer. Math. Soc.*, 18(1):121–156, 2005.

[27] M. Feistauer, M. Křížek, and V. Sobotíková. An analysis of finite element variational crimes for a nonlinear elliptic problem of a nonmonotone type. *East-West J. Numer. Math.*, 1(4):267–285, 1993.

[28] M. Feistauer and A. Ženíšek. Finite element solution of nonlinear elliptic problems. *Numer. Math.*, 50(4):451–475, 1987.

[29] M. G. D. Geers, V. G. Kouznetsova, and W. A. M. Brekelmans. Multi-scale computational homogenization: Trends and challenges. *J. Comput. Appl. Math.*, 234:2175–2182, 2010.

[30] A. Gloria. An analytical framework for the numerical homogenization of monotone elliptic operators and quasiconvex energies. *Multiscale Model. Simul.*, 5(3):996–1043 (electronic), 2006.

[31] V. V. Jikov, S. M. Kozlov, and O. A. Oleinik. *Homogenization of differential operators and integral functionals*. Springer-Verlag, Berlin, Heidelberg, 1994.

[32] F. Murat and L. Tartar. *H*-convergence. In *Topics in the mathematical modelling of composite materials*, volume 31 of *Progr. Nonlinear Differential Equations Appl.*, pages 21–43. Birkhäuser Boston, Boston, MA, 1997.

[33] J. A. Nitsche. Verfahren von Ritz und Spline-Interpolation bei Sturm-Liouville-Randwertproblemen. *Numer. Math.*, 13(3):260–265, 1969.

[34] J. T. Oden and K. S. Vemaganti. Estimation of local modeling error and goal-oriented adaptive modeling of heterogeneous materials. I. Error estimates and adaptive algorithms. *J. Comput. Phys.*, 164(1):22–47, 2000.

[35] P. A. Raviart. The use of numerical integration in finite element methods for solving parabolic equations. In J. J. H. Miller, editor, *Topics in numerical analysis. Proceedings of the Royal Irish Academy, Conference on Numerical Analysis, 1972*, pages 233–264. Academic Press, 1973.

[36] A. H. Schatz. An observation concerning Ritz-Galerkin methods with indefinite bilinear forms. *Math. Comp.*, 28:959–962, 1974.

[37] G. Strang and G. J. Fix. *An analysis of the finite element method*. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1973. Prentice-Hall Series in Automatic Computation.

[38] L. Tartar. Cours Peccot. Collège de France, 1977.