Numer. Math. Theor. Meth. Appl. Vol. **9**, No. 1, pp. 51-86
doi: 10.4208/nmtma.2015.m1412 February 2016

Convergence and Quasi-Optimality of an Adaptive Multi-Penalty Discontinuous Galerkin Method

Zhenhua Zhou and Haijun Wu*

Department of Mathematics, Nanjing University, Jiangsu, 210093, P. R. China.

Received 28 April 2014; Accepted (in revised version) 31 January 2015

Abstract. An adaptive multi-penalty discontinuous Galerkin method (AMPDG) for the diffusion problem is considered. Convergence and quasi-optimality of the AMPDG are proved. Compared with the analyses for the adaptive finite element method or the adaptive interior penalty discontinuous Galerkin method, extra works are done to overcome the difficulties caused by the additional penalty terms.

AMS subject classifications: 65N30, 65N12

Key words: Multi-penalty discontinuous Galerkin method, adaptive algorithm, convergence, quasi-optimality.

1. Introduction

Let $\Omega \subset \mathbb{R}^d, d = 2, 3$ be a bounded, polyhedral domain. We study the convergence of an *adaptive multi-penalty discontinuous Galerkin* (AMPDG) method for the diffusion problem

$$L(u) := -\operatorname{div}(A\nabla u) = f \quad \text{in } \Omega, \tag{1.1a}$$
$$lu = 0 \quad \text{on } \partial\Omega. \tag{1.1b}$$

Precise conditions on $A : \Omega \to \mathbb{R}^{d\times d}$ and $f : \Omega \to \mathbb{R}$ are specified later. We would like to point out that, in the present work we restrict ourselves to homogeneous Dirichlet boundary conditions to simplify the already technical presentation. Similarly, a reaction term of the type cu with $0 \le c \in L^\infty(\Omega)$ could have been added to the development as in [10] without changing the essence.

The adaptivity has been a fundamental technique for about four decades in finite element methods (FEM), interior penalty discontinuous Galerkin (IPDG) methods, and many other methods to deal with various singularities. There have been many works

*Corresponding author. *Email addresses:* `njuzzh@gmail.com` (Z.-H. Zhou Zhou), `hjw@nju.edu.cn` (H.-J. Wu)

on the convergence analysis of the adaptive FEMs (AFEM). It started with Babuška and Vogelius [2], who gave a detailed treatment for the one-dimensional boundary value problems. In 1996 Dörfler [14] introduced a crucial marking, from now on called Dörfler's marking, and proved the first convergent result for the two-dimensional case. He proved strict energy error reduction for the Poisson's equation provided the initial mesh satisfies a fineness assumption. For the results after that, we refer the reader to [10,13,17,18,20–23, etc.], and the references therein. Here we note that in [21,22], the MARK procedure of the adaptive algorithm marks not only for the estimator, but also for the oscillation, and the interior node property should be involved. However, in [10,23] the impractical ingredients, that is, the mark of oscillation and the interior node property, were removed. This is a great improvement in a practical point of view.

Although there have been many works for the AFEM, the convergence results of the adaptive IPDG methods (AIPDG) are rather recent. The first convergence result on the AIPDG was given by Karakashian and Pascal [19], then Hoppe, Kanschat and Warburton [15] improved the results upon [19]. In 2010, Bonito and Nochetto [3] proved the convergence of the AIPDG for the diffusion problem with general data on nonconforming partitions. They also gave a quasi-optimal asymptotic rate of convergence for the AIPDG, which was the first result of this type in the literature for DG methods.

The multi-penalty discontinuous Galerkin (MPDG) method considered in this paper, which was first introduced by Arnold [1], penalizes not only the jump of discrete solution, but also the jump of the (higher) derivatives of the discrete solution at mesh interfaces. We point out that the latter one has successfully applied to convection-dominated problems as a stabilization technique [5–9], and more recently, it has shown great potential for simulating Helmholtz scattering problems with high wave number [25,26].

The purpose of this paper is to prove convergence and quasi-optimality for the AMPDG based on an a posteriori error estimator of residual type and the Dörfler's marking strategy. The basic idea of the analysis is to mimic that of the AIPDG (cf. [3]), but some essential difficulties caused by the additional penalty terms need to be treated specially. Compared with [3], we introduce the extra penalty terms without any other restrictions for the data in model problem (1.1a) or the regularity of weak solution. Note that when the penalty parameters of the extra penalty terms equal to zero, the AMPDG reduces to the AIPDG, our results extend those of AIPDG [3].

The rest of this paper is organized as follows. In Section 2 we introduce the MPDG method and its adaptive algorithm, and give the preliminaries used to derive the main results of this paper. In Section 3 we give the upper and lower error estimates for the MPDG. Section 4 is devote to prove the contraction property of the adaptive algorithm. The quasi-optimality of the AMPDG is proved in Section 5.

In order to simplify the notation, we write $a \lesssim b$ whenever $a \leq Cb$ with a constant C independent of parameters which a and b may depend on. We also write $a \eqsim b$ for $a \lesssim b$ and $b \lesssim a$.

2. The multi-penalty discontinuous Galerkin method

In this section, we state the AMPDG on nonconforming meshes. Notation and preliminaries are also introduced. Throughout this paper, the standard space, norm and inner product notation are adopted, their definitions can be found in [4, 11]. In particular, $(\cdot,\cdot)_\omega$ and $\|\cdot\|_\omega$ denote the L^2-inner product and L^2-norm on $L^2(\omega)$ space, respectively. Denote by $(\cdot,\cdot) := (\cdot,\cdot)_\Omega$.

2.1. Data assumptions and the weak formulation

Let $p \geq 1$ be a given polynomial degree. Given an initial conforming partition $\mathcal{M}_0$ of Ω, we assume that $f \in L^2(\Omega)$, and $A \in \prod_{K\in\mathcal{M}_0} W^{p,\infty}(K)^{d\times d}$ is symmetric positive definite with eigenvalues in $0 < a_m \leq a_M < \infty$, i.e.,

$$a_m|y|^2 \leq Ay\cdot y \leq a_M|y|^2 \quad \forall y \in \mathbb{R}^d \text{ a.e. in } \Omega.$$

Invoking the Lax-Milgram lemma, the above assumptions ensure that the weak formulation of (1.1a), namely

$$u \in H_0^1(\Omega): \quad (A\nabla u, \nabla v) = (f, v) \quad \forall v \in H_0^1(\Omega), \tag{2.1}$$

possesses a unique solution.

2.2. Partitions of Ω

In the theoretical analyses hereafter, nonconforming partitions made of triangles (tetrahedrons when $d = 3$) or quadrilaterals (hexahedra when $d = 3$) are considered. They can be generated from the initial mesh $\mathcal{M}_0$ by using some typical refine strategies, such as *quad refinement* for quadrilaterals (hexahedra when $d = 3$) and *red refinement* or *bisection* for triangles (tetrahedrons when $d = 3$), see Section 2.1.2 in [3] for more details.

From now on, we use the two dimensional denomination for elements of the partitions even when discussing the three dimensional case. For example, we say an element K is a quadrilateral, meaning K is a quadrilateral when $d = 2$, and K is a hexahedra when $d = 3$.

We call an element K' is an ancestor of element K, if K is generated by a finite number of refinements from K'. Given a partition $\mathcal{M}_H$ of Ω, we say that $\mathcal{M}_h$ is a refinement of $\mathcal{M}_H$ if every element in $\mathcal{M}_h$ is either an element in $\mathcal{M}_H$, or has an ancestor element in $\mathcal{M}_H$. To this end, we write $\mathcal{M}_H \leq \mathcal{M}_h$ or $\mathcal{M}_h \geq \mathcal{M}_H$, and denote by $\mathcal{R}_{\mathcal{M}_H\to\mathcal{M}_h}$ the set of elements in $\mathcal{M}_H$ refined to obtain $\mathcal{M}_h$. We remark here that every partition appearing below is a refinement of the initial partition $\mathcal{M}_0$.

2.3. Definitions and Notation

Given a partition $\mathcal{M}_h \geq \mathcal{M}_0$, we define the *energy space* to be the broken H^1 space

$$H^1(\mathcal{M}_h) := \prod_{K\in\mathcal{M}_h} H^1(K).$$

For any element $K \in \mathcal{M}_h$, let $\mathcal{P}_p(K)$ denote the set of all polynomials whose degrees in all variables (total degree) $\leq p$ if K is a simplex, and the set of all polynomials whose degrees in each variable (separate degrees) $\leq p$ if K is a quadrilateral. Define the discontinuous finite element space

$$V_h := \{v_h : v_h|_K \in \mathcal{P}_p(K),\ \forall K \in \mathcal{M}_h\},$$

and define $V_h^0 := V_h \cap H_0^1(\Omega)$ to be the underlying conforming finite element space. Note that V_h^0 is defined on the nonconforming mesh. Let $\mathcal{E}_h^I$ and $\mathcal{E}_h^D$ be the set of all interior and boundary edges of $\mathcal{M}_h$, respectively. Write $\mathcal{E}_h = \mathcal{E}_h^I \cup \mathcal{E}_h^D$. Define the mesh size h to be

$$h_K = |K|^{\frac{1}{d}},\ \forall K \in \mathcal{M}_h, \qquad h_e = |e|^{\frac{1}{d-1}},\ \forall e \in \mathcal{E}_h.$$

To shorten notation, we define for any set of elements $\mathcal{M}_h^* \subseteq \mathcal{M}_h$ and for any set of edges $\mathcal{E}_h^* \subseteq \mathcal{E}_h$

$$(v,w)_{\mathcal{M}_h^*} := \sum_{K\in\mathcal{M}_h^*} (v,w)_K \qquad \forall v,w \in L^2(\mathcal{M}_h^*),$$
$$\langle\phi,\varphi\rangle_{\mathcal{E}_h^*} := \sum_{e\in\mathcal{E}_h^*} \langle\phi,\varphi\rangle_e \qquad \forall \phi,\varphi \in L^2(\mathcal{E}_h^*).$$

The same notation is used for functions in $L^2(\mathcal{M}_h^*)^d$ or $L^2(\mathcal{E}_h^*)^d$. We also define the corresponding *broken norms*

$$\|v\|_{\mathcal{M}_h^*} := (v,v)_{\mathcal{M}_h^*}^{\frac{1}{2}}, \quad \|\phi\|_{\mathcal{E}_h^*} := \langle\phi,\phi\rangle_{\mathcal{E}_h^*}^{\frac{1}{2}}.$$

Let $e \in \mathcal{E}_h^I$ be an interior edge shared by the elements $K_1^e, K_2^e \in \mathcal{M}_h$, where the global index of K_2^e is greater than that of K_1^e. Define $n|_e$ to be the unit normal vector associate to edge e pointing to K_1^e which has smaller global index. We also define the jump $[v]$ and average $\{v\}$ of v on e to be

$$[v]|_e := v|_{K_2^e} - v|_{K_1^e}, \quad \{v\}|_e := \frac{v|_{K_1^e} + v|_{K_2^e}}{2}.$$

If $e \in \mathcal{E}_h^D$, that is, $e = \partial K^e \cap \partial\Omega$ for some element $K^e \in \mathcal{M}_h$, we define $n|_e$ to be the unit normal vector associated to e pointing to the outer of Ω, and let

$$[v]|_e = \{v\}|_e = v|_{K^e}.$$

For an element $K\in\mathcal{M}_h$, let $\omega(K)\subset\mathcal{M}_h$ be a point set including K such that

$$\operatorname{diam}(\omega(K))\le Ch_K, \tag{2.2}$$

with $C>0$ a constant depending only on the shape regularity of $\mathcal{M}_h$. For precise definition of $\omega(K)$, see Section 6.1 in [3]. This set is helpful in describing the local properties of a Clément-type interpolation operator onto V_h^0. See Condition 2.2 in Section 2.7 or Lemma 6.6 in [3]. Let $\sigma(K)=\cup_{e\in\mathcal{E}_h}e\cap\omega(K)$ be the skeleton of $\mathcal{M}_h$ within $\omega(K)$, and for a set of elements $\mathcal{M}_h^*\subseteq\mathcal{M}_h$, let

$$\omega(\mathcal{M}_h^*):=\bigcup_{K\in\mathcal{M}_h^*}\omega(K),\qquad \sigma(\mathcal{M}_h^*):=\cup_{e\in\mathcal{E}_h}e\cap\omega(\mathcal{M}_h^*). \tag{2.3}$$

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two refinements of $\mathcal{M}_0$, we define $\mathcal{M}_1\oplus\mathcal{M}_2$ to be the overlay of $\mathcal{M}_1$ and $\mathcal{M}_2$, see Section 2.1.2 in [3] for more details. Therefore, $\mathcal{M}_1\oplus\mathcal{M}_2$ is a refinement of both $\mathcal{M}_1$ and $\mathcal{M}_2$, and under Condition 2.1 (see Section 2.7), we have [3]

$$\#(\mathcal{M}_1\oplus\mathcal{M}_2)\le\#\mathcal{M}_1+\#\mathcal{M}_2-\#\mathcal{M}_0. \tag{2.4}$$

2.4. The discrete formulation

In order to simplify the notation, we write

$$a(v,w):=(A\nabla v,\nabla w)_{\mathcal{M}_h}\quad\forall v,w\in H^1(\mathcal{M}_h).$$

Let $\gamma_0>0$. The usual symmetric IPDG formulation reads as follows: given a triangulation $\mathcal{M}_h$ of Ω, find $u_h\in V_h$ such that

$$a_h^{\mathrm{IP}}(u_h,v_h)=(f,v_h)\quad\forall v_h\in V_h,$$

where

$$\begin{aligned}
&a_h^{\mathrm{IP}}(u_h,v_h):=a(u_h,v_h)-\langle\{A\nabla u_h\cdot n\},[v_h]\rangle_{\mathcal{E}_h}-\langle\{A\nabla v_h\cdot n\},[u_h]\rangle_{\mathcal{E}_h}+J_h^0(u_h,v_h),\\
&J_h^0(u_h,v_h):=\gamma_0\left\langle h^{-1}[u_h],[v_h]\right\rangle_{\mathcal{E}_h}.
\end{aligned}$$

We note that $a_h^{\mathrm{IP}}(\cdot,\cdot)$ is not well defined on $H^1(\mathcal{M}_h)\times H^1(\mathcal{M}_h)$. However, this can be remedied by means of the lifting operator.

Definition 2.1 (Lifting operator). *Let $L_h:H^1(\mathcal{M}_h)\to V_h^d$ be defined by* [3, 16]

$$\left(L_h(v),Aw_h^d\right)_{\mathcal{M}_h}:=\left\langle[v],\{Aw_h^d\cdot n\}\right\rangle_{\mathcal{E}_h}\quad\forall w_h^d\in V_h^d. \tag{2.5}$$

Remark 2.1. By invoking the Lax-Milgram lemma, the lifting operator L_h is well defined on the whole broken energy space $H^1(\mathcal{M}_h)$ and $L_h(v)=0$ for all $v\in H_0^1(\Omega)$. In addition, as has been shown in [3, Lemma 2.1], the lifting operator is stable in the sense that there exists a constant $C_L>0$ depending only on A, p and the shape regularity of $\mathcal{M}_h$ such that

$$\|L_h(v)\|_{\mathcal{M}_h}\le C_L\left\|h^{-\frac12}[v]\right\|_{\mathcal{E}_h}\quad\forall v\in H^1(\mathcal{M}_h). \tag{2.6}$$

Using the lifting operator mentioned above, let $a_h^{\rm L}(\cdot,\cdot): H^1(\mathcal{M}_h)\times H^1(\mathcal{M}_h)\to\mathbb{R}$ be defined by [3]

$$a_h^{\rm L}(v,w) := a(v,w) - (L_h(w), A\nabla v)_{\mathcal{M}_h} - (L_h(v), A\nabla w)_{\mathcal{M}_h} + J_h^0(v,w). \tag{2.7}$$

Note that the definition of the lifting operator (2.5) directly implies

$$a_h^{\rm L}(v_h,w_h) = a_h^{\rm IP}(v_h,w_h) \quad \forall v_h,w_h\in V_h.$$

Since $L_h(v)=0$ for all $v\in H_0^1(\Omega)$, the solution $u\in H_0^1(\Omega)$ of (2.1) satisfies

$$a_h^{\rm L}(u,v) = (f,v) \quad \forall v\in H_0^1(\Omega).$$

Let $0\le \gamma_1,\cdots,\gamma_p\lesssim 1$ be the penalty parameters. The MPDG considered in this paper reads: find $u_h\in V_h$ such that

$$a_h(u_h,v_h) := a_h^{\rm L}(u_h,v_h) + \sum_{j=1}^p J_h^j(u_h,v_h) = (f,v_h) \quad \forall v_h\in V_h, \tag{2.8}$$

where

$$J_h^j(u_h,v_h) := \sum_{e\in\mathcal{E}_h^I} \gamma_j h_e^{2j-1}\left\langle\left[\frac{\partial^{j-1}(A\nabla u_h\cdot n)}{\partial n^{j-1}}\right],\left[\frac{\partial^{j-1}(A\nabla v_h\cdot n)}{\partial n^{j-1}}\right]\right\rangle_e,\ 1\le j\le p.$$

For any $v_h,w_h\in V_h$, $\mathcal{M}_h^*\subseteq\mathcal{M}_h$, $\mathcal{E}_h^*\subseteq\mathcal{E}_h^I$, and $1\le j\le p$, we write

$$\begin{aligned}
J_h^j(v_h,w_h)_e &:= \gamma_j h_e^{2j-1}\left\langle\left[\frac{\partial^{j-1}(A\nabla v_h\cdot n)}{\partial n^{j-1}}\right],\left[\frac{\partial^{j-1}(A\nabla w_h\cdot n)}{\partial n^{j-1}}\right]\right\rangle_e,\\
\widetilde{J}_h^j(v_h,w_h)_K &:= \gamma_j h_K^{2j-1}\left\langle\left[\frac{\partial^{j-1}(A\nabla v_h\cdot n)}{\partial n^{j-1}}\right],\left[\frac{\partial^{j-1}(A\nabla w_h\cdot n)}{\partial n^{j-1}}\right]\right\rangle_{\partial K\cap\Omega},\\
J_h^j(v_h,w_h)_{\mathcal{E}_h^*} &:= \sum_{e\in\mathcal{E}_h^*} J_h^j(v_h,w_h)_e,\quad \widetilde{J}_h^j(v_h,w_h)_{\mathcal{M}_h^*} := \sum_{K\in\mathcal{M}_h^*}\widetilde{J}_h^j(v_h,w_h)_K,
\end{aligned}$$

and write $\widetilde{J}_h^j(v_h,w_h)=\widetilde{J}_h^j(v_h,w_h)_{\mathcal{M}_h}$ for simplicity. We introduce the following mesh dependent norms:

$$\begin{aligned}
&|||v||| := a(v,v)^{\frac12} \quad\text{and}\quad |||v|||_{0,h} := \left\{|||v|||^2 + J_h^0(v,v)\right\}^{\frac12} && \forall v\in H^1(\mathcal{M}_h),\\
&|||v_h|||_{p,h} := \left\{|||v_h|||_{0,h}^2 + \sum_{j=1}^p J_h^j(v_h,v_h)\right\}^{\frac12} && \forall v_h\in V_h.
\end{aligned}$$

The bilinear forms $a_h^{\rm L}(\cdot,\cdot)$ and $a_h(\cdot,\cdot)$ satisfy the following properties.

Lemma 2.1 (Continuity and coercivity). *Let $\mathcal{M}_h\ge\mathcal{M}_0$ and $\gamma_j\ge 0, j=1,\cdots,p$. There exist positive constants $\gamma_0^{(c)}$, C_n and C_e such that for $\gamma_0\ge\gamma_0^{(c)}$ there holds*

$$\begin{aligned}
a_h^{\rm L}(v,w) &\le C_n|||v|||_{0,h}|||w|||_{0,h} \quad\text{and}\quad a_h^{\rm L}(v,v)\ge C_e|||v|||_{0,h}^2 && \forall v,w\in H^1(\mathcal{M}_h),\\
a_h(v_h,w_h) &\le C_n|||v_h|||_{p,h}|||w_h|||_{p,h} \quad\text{and}\quad a_h(v_h,v_h)\ge C_e|||v_h|||_{p,h}^2 && \forall v_h,w_h\in V_h.
\end{aligned}$$

Proof. We prove only for $a_h(\cdot,\cdot)$; the properties for $a_h^{\rm L}(\cdot,\cdot)$ can be found in [3, Lemma 2.2].

By using the Cauchy-Schwarz inequality for each term in $a_h(\cdot,\cdot)$ and the stability of lifting operator (2.6), we have

$$
\begin{aligned}
& a_h(v_h,w_h) \\
\le & \Big\{(1+C_L a_M^{\frac{1}{2}})|||v_h|||^2+(\gamma_0+C_L a_M^{\frac{1}{2}})\left\|h^{-\frac{1}{2}}[v_h]\right\|_{\mathcal{E}_h}^2+\sum_{j=1}^p J_h^j(v_h,v_h)\Big\}^{\frac{1}{2}} \\
& \times\Big\{(1+C_L a_M^{\frac{1}{2}})|||w_h|||^2+(\gamma_0+C_L a_M^{\frac{1}{2}})\left\|h^{-\frac{1}{2}}[w_h]\right\|_{\mathcal{E}_h}^2+\sum_{j=1}^p J_h^j(w_h,w_h)\Big\}^{\frac{1}{2}}.
\end{aligned}
$$

Then the continuity of $a_h(\cdot,\cdot)$ follows by setting $C_n=1+C_L a_M^{\frac{1}{2}}$ and $\gamma_0\ge 1$.

Using (2.6) again, together with Young's inequality with parameter $\varepsilon>0$, yields

$$
a_h(v_h,v_h)\ge(1-\frac{\varepsilon}{2})|||v_h|||^2+(\gamma_0-\frac{2}{\varepsilon}C_L^2a_M)\left\|h^{-\frac{1}{2}}[v_h]\right\|_{\mathcal{E}_h}^2+\sum_{j=1}^p J_h^j(v_h,v_h).
$$

Let $\varepsilon=1$, $\gamma_0\ge 4C_L^2a_M$, we obtain the coercivity of $a_h(\cdot,\cdot)$ with $C_e=\frac{1}{2}$. The lemma follows by setting $\gamma_0^{(c)}:=\max\{1,4C_L^2a_M\}$. □

2.5. Estimator and oscillation

Let $v_h\in V_h$, we define the *element residual* on $K\in\mathcal{M}_h$ and *jump residual* on $e\in\mathcal{E}_h^I$ of v_h by

$$
R(v_h)|_K:=\big(f-L(v_h)\big)|_K \quad\text{and}\quad J(v_h)|_e:=\big([A\nabla v_h\cdot n]\big)|_e.
$$

Let $m\in\mathbb{N}_0$, we denote by $P_m^q, q\in[2,\infty]$, the L^q-best approximation operator onto the set of piecewise polynomials of total degree $\le m$ on d or $(d-1)$-simplices, or piecewise polynomials of separate degrees $\le m$ on d or $(d-1)$-quadrilaterals. That is, for any $K\in\mathcal{M}_h$ or $e\in\mathcal{E}_h^I$, $P_m^q v$ satisfies

$$
\|v-P_m^qv\|_{L^q(\omega)}=\inf_{z\in\mathcal{P}_m(\omega)}\|v-z\|_{L^q(\omega)}\,,\quad \omega=K \text{ or } \omega=e.
$$

Let $Q_m^q=I-P_m^q$, where I stands for the identity operator. Define the error indicators $E_h(v_h,K)$, $\eta_h(v_h,K)$ and oscillation $\mathrm{osc}_h(v_h,K)$ by

$$
E_h^2(v_h,K):=\eta_h^2(v_h,K)+\sum_{j=2}^p\widetilde{J}_h^j(v_h,v_h)_K, \tag{2.9}
$$

$$
\eta_h^2(v_h,K):=h_K^2\,\|R(v_h)\|_K^2+h_K\,\|J(v_h)\|_{\partial K\cap\Omega}^2\,, \tag{2.10}
$$

$$
\mathrm{osc}_h^2(v_h,K):=h_K^2\,\left\|Q_m^2R(v_h)\right\|_K^2+h_K\,\left\|Q_{m'}^2J(v_h)\right\|_{\partial K\cap\Omega}^2\,. \tag{2.11}
$$

The specific choice of m and m' depends on the element type:

$$m = 2p-2, m' = 2p-1 \quad \text{for simplices;}$$
$$m = 2p, \quad m' = 2p \quad \text{for quadrilaterals.}$$

Remark 2.2. If $A \in \prod_{K\in\mathcal{M}_0} \mathcal{P}_p(K)^{d\times d}$, the choice of m and m' directly implies that $\operatorname{osc}_h(v_h, K) = h_K \left\|Q_m^2 f\right\|_K$, which is independent of v_h. This is because $\nabla\mathcal{P}_p(K) \subset \mathcal{P}_p(K)^d$ for quadrilaterals while $\nabla\mathcal{P}_p(K) \subset \mathcal{P}_{p-1}(K)^d$ for simplices. Moreover, both choices yield Lemma 2.3 and Lemma 2.4 (below).

Finally, for any subset $\mathcal{M}_h^* \subseteq \mathcal{M}_h$ we set

$$E_h^2(v_h, \mathcal{M}_h^*) := \sum_{K\in\mathcal{M}_h^*} E_h^2(v_h, K),$$
$$\eta_h^2(v_h, \mathcal{M}_h^*) := \sum_{K\in\mathcal{M}_h^*} \eta_h^2(v_h, K),$$
$$\operatorname{osc}_h^2(v_h, \mathcal{M}_h^*) := \sum_{K\in\mathcal{M}_h^*} \operatorname{osc}_h^2(v_h, K).$$

We also define the similar notions for the matrix A:

$$E_h(A, K) := \max_{0\le j\le p} \left\{h_K^j |A|_{W^{j,\infty}(K)}\right\},$$
$$\operatorname{osc}_h(A, K) := \max\left\{h_K \left\|Q_{m''}^\infty \operatorname{div} A\right\|_{L^\infty(K)}, \left\|Q_p^\infty A\right\|_{L^\infty(K)}\right\},$$

where $|A|_{W^{j,\infty}(K)} := \max_{1\le s,t\le d} |A_{st}|_{W^{j,\infty}(K)}$, and $m'' = p-1$ if K is a simplex, $m'' = p$ if K is a quadrilateral. For any subset $\mathcal{M}_h^* \subseteq \mathcal{M}_h$ we set

$$E_h(A, \mathcal{M}_h^*) := \max_{K\in\mathcal{M}_h^*} E_h(A, K) \quad \text{and} \quad \operatorname{osc}_h(A, \mathcal{M}_h^*) := \max_{K\in\mathcal{M}_h^*} \operatorname{osc}_h(A, K).$$

Remark 2.3 (Monotonicity)**.** The definitions above imply the following monotonicity property: let $\mathcal{M}_0 \le \mathcal{M}_H \le \mathcal{M}_h$, for any $K' \in \mathcal{M}_H$ and $v_H \in V_H$, there holds

$$\eta_h(v_H, \mathcal{M}_h) \le \eta_H(v_H, \mathcal{M}_H), \quad \operatorname{osc}_h(v_H, \mathcal{M}_h) \le \operatorname{osc}_H(v_H, \mathcal{M}_H),$$
$$\operatorname{osc}_H(v_H, K') \le \eta_H(v_H, K'),$$
$$E_h(A, \mathcal{M}_h) \le E_H(A, \mathcal{M}_H) \le E_0(A, \mathcal{M}_0),$$
$$\operatorname{osc}_h(A, \mathcal{M}_h) \le \operatorname{osc}_H(A, \mathcal{M}_H) \le \operatorname{osc}_0(A, \mathcal{M}_0).$$

From now on, we shall not specify the dependency on $E_h(A, \mathcal{M}_h)$ and $\operatorname{osc}_h(A, \mathcal{M}_h)$ for a mesh $\mathcal{M}_h \ge \mathcal{M}_0$, since they can be bounded by $E_0(A, \mathcal{M}_0)$ and $\operatorname{osc}_0(A, \mathcal{M}_0)$, which are fixed constants depending on A and the initial mesh $\mathcal{M}_0$.

Lemma 2.2. *For any $v_h \in V_h$ and $1 \le j \le p$, we have*

$$J_h^j(v_h, v_h) \eqsim \widetilde{J}_h^j(v_h, v_h) \lesssim \gamma_j \left\|\nabla v_h\right\|_{\mathcal{M}_h}^2,$$

the constant hiding in "$\lesssim$" depends on A, p and the shape regularity of $\mathcal{M}_h$.

Proof. For any $v \in W^{k,\infty}(K)$, if $n = (n_1, \cdots, n_d)$ is a fixed unit vector, then

$$\begin{aligned}\left\|\frac{\partial^k v}{\partial n^k}\right\|_{L^\infty(K)} &= \left\|\sum_{1\le i_1,\cdots,i_k\le d} n_{i_1}\cdots n_{i_k}\frac{\partial^k v}{\partial x_{i_1}\cdots\partial x_{i_k}}\right\|_{L^\infty(K)} \\ &\le |v|_{W^{k,\infty}(K)}\left(\sum_{i=1}^d |n_i|\right)^k \le d^{\frac{k}{2}}|v|_{W^{k,\infty}(K)}.\end{aligned}$$

For any $w_h \in V_h$, the inverse inequality says that $\|\nabla w_h\|_K \le C h_K^{-1}\|w_h\|_K$, where C is a constant depending only on p and the shape regularity of K. Therefore,

$$\begin{aligned}&\left\|\frac{\partial^{j-1}(A\nabla v_h\cdot n)}{\partial n^{j-1}}\right\|_K \\ =&\left\|\sum_{i=0}^{j-1} C_{j-1}^i \sum_{s=1}^d\sum_{t=1}^d n_s \frac{\partial^{j-1-i}A_{st}}{\partial n^{j-1-i}}\cdot\frac{\partial^i}{\partial n^i}(\frac{\partial v_h}{\partial x_t})\right\|_K \\ \le& \sum_{i=0}^{j-1} C_{j-1}^i \sum_{s=1}^d\sum_{t=1}^d n_s \left\|\frac{\partial^{j-1-i}A_{st}}{\partial n^{j-1-i}}\right\|_{L^\infty(K)}\left\|\frac{\partial^i}{\partial n^i}(\frac{\partial v_h}{\partial x_t})\right\|_K \\ \le& \sum_{i=0}^{j-1} C_{j-1}^i \sqrt{d}^{j-1-i}|A|_{W^{j-1-i,\infty}(K)}(\frac{C}{h_K})^i \sum_{s=1}^d\sum_{t=1}^d n_s\left\|\frac{\partial v_h}{\partial x_t}\right\|_K \\ \le& d(C+\sqrt{d})^{j-1}\max_{0\le i\le j-1}\left\{h_K^i|A|_{W^{i,\infty}(K)}\right\}\cdot h_K^{-j+1}\|\nabla v_h\|_K \\ \lesssim& E_h(A,K)h_K^{-j+1}\|\nabla v_h\|_K.\end{aligned}$$

Similarity, we have

$$\left\|\nabla\left(\frac{\partial^{j-1}(A\nabla v_h\cdot n)}{\partial n^{j-1}}\right)\right\|_K \lesssim E_h(A,K)h_K^{-j}\|\nabla v_h\|_K.$$

Since the mesh $\mathcal{M}_h$ is local quasi-uniformly, the trace inequality yields

$$\begin{aligned}\widetilde{J}_h^j(v_h,v_h)_K \lesssim& \sum_{T\in\omega_K}\gamma_j h_T^{2j-1}\sum_{e\subset\partial T\cap\Omega}\left\|\frac{\partial^{j-1}(A\nabla v_h\cdot n)}{\partial n^{j-1}}\right\|_e^2 \\ \lesssim& \sum_{T\in\omega_K}\gamma_j h_T^{2j-1}\sum_{e\subset\partial T\cap\Omega}\left\{h_T^{-1}\left\|\frac{\partial^{j-1}(A\nabla v_h\cdot n)}{\partial n^{j-1}}\right\|_T^2\right. \\ &\left.+\left\|\frac{\partial^{j-1}(A\nabla v_h\cdot n)}{\partial n^{j-1}}\right\|_T\left\|\nabla\left(\frac{\partial^{j-1}(A\nabla v_h\cdot n)}{\partial n^{j-1}}\right)\right\|_T\right\} \\ \lesssim& \gamma_j E_h^2(A,\omega_K)\|\nabla v_h\|_{\omega_K}^2.\end{aligned}$$

Then

$$\widetilde{J}_h^j(v_h,v_h)=\sum_{K\in\mathcal{M}_h}\widetilde{J}_h^j(v_h,v_h)_K$$
$$\lesssim \gamma_j E_h^2(A,\mathcal{M}_h)\,\|\nabla v_h\|^2_{\mathcal{M}_h}\lesssim \gamma_j E_0^2(A,\mathcal{M}_0)\,\|\nabla v_h\|^2_{\mathcal{M}_h}.$$

This completes the proof of the lemma. □

We have the following two lemmas for the error indicator and the oscillation (cf. Lemma 2.4 and Lemma 2.5 in [3]). These two lemmas are instrumental in deriving the main results of this paper.

Lemma 2.3 (Estimator reduction). *Assume that the refinement strategy satisfies Condition 2.1 (below). Let $\mathcal{M}_0\le\mathcal{M}_H\le\mathcal{M}_h$, and let $\widehat{\mathcal{M}}_H\subset\mathcal{M}_H$ be the set of elements marked for refinement to obtain $\mathcal{M}_h$. Write $\lambda=1-\beta_M$, where $0<\beta_M<1$ is the constant appearing in Condition 2.1. Then there exists a constant C_E depending on A, p and the shape regularity of $\mathcal{M}_0$ such that, for any $v_H\in V_H$, $v_h\in V_h$, and any $\delta>0$, there holds that*

$$E_h^2(v_h,\mathcal{M}_h)\le(1+\delta)\big(E_H^2(v_H,\mathcal{M}_H)-\lambda E_H^2(v_H,\widehat{\mathcal{M}}_H)\big)$$
$$+(1+\delta^{-1})C_E|||v_h-v_H|||^2.$$

Proof. Let $z_h=v_h-v_H$. The triangle inequality and Young's inequality imply

$$\begin{aligned}E_h^2(v_h,\mathcal{M}_h)\le(1+\delta)E_h^2(v_H,\mathcal{M}_h)+(1+\delta^{-1})\sum_{K\in\mathcal{M}_h}\Big\{h_K^2\,\|\mathrm{div}(A\nabla z_h)\|_K^2\\ +h_K\,\|J(z_h)\|^2_{\partial K\cap\Omega}+\sum_{j=2}^p\widetilde{J}_h^j(z_h,z_h)_K\Big\}.\end{aligned}\tag{2.12}$$

Since $\big[\frac{\partial^{j-1}(A\nabla v_H\cdot n)}{\partial n^{j-1}}\big]\big|_e=0$ for any side $e\in\mathcal{E}_h^I$ in the interior of some element $K'\in\mathcal{M}_H$, and $h_K\le\beta_M H_{K'}$ if K, K' satisfies $K\subset K'\in\widehat{\mathcal{M}}_H$, we have

$$\begin{aligned}E_h^2(v_H,\mathcal{M}_h)=&\sum_{K\subset K'\in\mathcal{M}_H\backslash\widehat{\mathcal{M}}_H}\Big(\eta_h^2(v_H,K)+\sum_{j=2}^p\widetilde{J}_h^j(v_H,v_H)_K\Big)\\ &+\sum_{K\subset K'\in\widehat{\mathcal{M}}_H}\Big(\eta_h^2(v_H,K)+\sum_{j=2}^p\widetilde{J}_h^j(v_H,v_H)_K\Big)\\ \le&\sum_{K'\in\mathcal{M}_H\backslash\widehat{\mathcal{M}}_H}\Big(\eta_H^2(v_H,K')+\sum_{j=2}^p\widetilde{J}_H^j(v_H,v_H)_{K'}\Big)\\ &+\beta_M\sum_{K'\in\widehat{\mathcal{M}}_H}\Big(\eta_H^2(v_H,K')+\sum_{j=2}^p\widetilde{J}_H^j(v_H,v_H)_{K'}\Big)\\ =&E_H^2(v_H,\mathcal{M}_H)-\lambda E_H^2(v_H,\widehat{\mathcal{M}}_H).\end{aligned}$$

Using the similar tricks as in Lemma 2.2, we have

$$\begin{aligned}\|\mathrm{div}(A\nabla z_h)\|_K &= \left\|\sum_{i=1}^d\sum_{j=1}^d\Big(\frac{\partial A_{ij}}{\partial x_i}\cdot\frac{\partial z_h}{\partial x_j}+A_{ij}\cdot\frac{\partial^2 z_h}{\partial x_i\partial x_j}\Big)\right\|_K\\ &\le |A|_{W^{1,\infty}(K)}\sum_{i=1}^d\sqrt{d}\,\|\nabla z_h\|_K+|A|_{L^\infty(K)}\sum_{i=1}^d\sqrt{d}\left\|\nabla\Big(\frac{\partial z_h}{\partial x_i}\Big)\right\|_K\\ &\lesssim E_h(A,K)\cdot h_K^{-1}\,\|\nabla z_h\|_K\,.\end{aligned}$$

Noting that $0\le\gamma_j\lesssim 1, j=1,\cdots,p$, and the minimal eigenvalue of A is $\ge a_m$, using Lemma 2.2 for the other terms in (2.12), and combining the above estimates, we can get the desired conclusion. □

Lemma 2.4 (Perturbation of oscillation). *Let $\mathcal{M}_0\le\mathcal{M}_H\le\mathcal{M}_h$. Then there exists a constant C_{osc} depending on A, p and the shape regularity of $\mathcal{M}_0$ such that, for all $v_H\in V_H$, $v_h\in V_h$, and any $\mathcal{M}^*\subseteq\mathcal{M}_H\cap\mathcal{M}_h$, there holds*

$$\mathrm{osc}_H^2(v_H,\mathcal{M}^*)\le 2\mathrm{osc}_h^2(v_h,\mathcal{M}^*)+C_{osc}|||v_h-v_H|||^2.$$

Remark 2.4. Lemma 2.4 is a little different from [3, Lemma 2.5], their proofs are almost the same (see e.g., [10, Corollary 3.5]). We only need to sum the local perturbation of oscillation (see [10, Proposition 3.3]) over elements only in $\mathcal{M}^*$ instead of all elements in $\mathcal{M}_H\cap\mathcal{M}_h$. Moreover, if $\mathcal{M}_H=\mathcal{M}_h$, Lemma 2.4 is still valid.

2.6. The adaptive algorithm

As usual, the AMPDG may be described as loops of the form with counter $k\ge 0$:

$$\text{SOLVE}\to\text{ESTIMATE}\to\text{MARK}\to\text{REFINE}.$$

Let $\{\mathcal{M}_k\}_{k\ge0}$, $\{V_k\}_{k\ge0}$, $\{u_k\}_{k\ge0}$ etc. be the sequence of meshes, discontinuous finite element spaces, discrete solutions etc. generated by the above iteration.

In the MARK procedure we use the Dörfler strategy [14]: given $\theta\in(0,1]$, mark elements in a subset $\widehat{\mathcal{M}}_k$ of $\mathcal{M}_k$ such that

$$E_{h_k}(u_k,\widehat{\mathcal{M}}_k)\ge\theta\,E_{h_k}(u_k,\mathcal{M}_k). \tag{2.13}$$

The pseudocode of the AMPDG is then given by the following iteration which is almost the same as that of AFEM (cf. [10, Subsection 2.7]):

AMPDG

Given the initial mesh $\mathcal{M}_0$ and marking parameter $0 < \theta \leq 1$
set $k := 0$ and iterate
1: Solve (2.8) on $\mathcal{M}_k$ for u_k;
2: Compute estimators $E_{h_k}(u_k, K), K \in \mathcal{M}_k$ via (2.9);
3: Find the set of marked elements $\widehat{\mathcal{M}}_k$ via (2.13);
4: Refine $\widehat{\mathcal{M}}_k$ in $\mathcal{M}_k$ to obtain the mesh $\mathcal{M}_{k+1}$, set $k = k + 1$.

2.7. Some conditions

To unify the treatment of the various refinement strategies, we impose some conditions below, as have been shown in [3]. They are important in deriving the main results of this paper. Their correctness are verified in [3, Section 6] for quad refinement of quadrilateral meshes and bisection of triangle meshes, while red refinement of triangle meshes, which is somewhat in between the two, is not discussed in detail.

Condition 2.1 (Atomic refinement)**.** *The subdivision of an element into children is called atomic refinement and is dictated exclusively by the initial partition $\mathcal{M}_0$ and the refinement rules. If K is a child of $K' \in \mathcal{M}_H$ and $h_K = |K|^{\frac{1}{d}}$, $H_{K'} = |K'|^{\frac{1}{d}}$ are the mesh sizes, then there exists a constant $\beta_M < 1$ depending only on the dimension d such that*

$$h_K \leq \beta_M H_{K'}.$$

Moreover, the shape regularity of any mesh $\mathcal{M}_H \geq \mathcal{M}_0$ is determined by that of $\mathcal{M}_0$.

Condition 2.2 (Interpolation operator onto V_h^0)**.** *There exists an interpolation operator $I_h : H^1(\mathcal{M}_h) \to V_h^0$ and a constant C_I depending only on the shape regularity of $\mathcal{M}_h$ such that for all $K \in \mathcal{M}_h$ the following inequalities hold:*

$$\|v - I_h v\|_K \leq C_I \|h\nabla v\|_{\omega(K)} \quad \forall v \in H_0^1(\Omega), \tag{2.14}$$

and for $m = 0, 1$,

$$\|\nabla^m (v_h - I_h v_h)\|_K \leq C_I \left\| h^{\frac{1-2m}{2}} [v_h] \right\|_{\sigma(K)} \quad \forall v_h \in V_h. \tag{2.15}$$

Here $\omega(K)$ satisfying (2.2) *is defined in Section 6.1 of* [3] *and $\sigma(K) = \cup_{e\in\mathcal{E}_h} e \cap \omega(K)$. The operator I_h is defined locally and preserves V_h^0 locally.*

Remark 2.5. I_h is a Clément-type interpolation operator, see [3, Lemma 6.6] for more details. Moreover, I_h is H^1-stable: $\forall v \in H_0^1(\Omega)$, let $v_K = \frac{1}{|K|}\int_K v\,\mathrm{d}x$, then the inverse inequality, triangle inequality, (2.14) and Poincaré inequality yield

$$\begin{aligned}\|\nabla I_h v\|_K =& \|\nabla(I_h v - v_K)\|_K \\ \lesssim& h_K^{-1}\big(\|I_h v - v\|_K + \|v - v_K\|_K \big) \lesssim \|\nabla v\|_{\omega(K)}.\end{aligned}$$

Condition 2.3 (Properties of REFINE). *$\mathcal{M}_{k+1}$ is generated from $\mathcal{M}_k$ by successive application of atomic refinements such that the following hold:*

(a) the atomic refinements satisfy Condition 2.1;

(b) all elements of $\widehat{\mathcal{M}}_k$ are refined at least b times, with $b \geq 1$ a given integer;

(c) no element of $\mathcal{M}_k$ can undergo more than B atomic refinements, with $B \geq 1$ a universal integer, to give rise to elements of $\mathcal{M}_{k+1}$;

(d) Condition 2.2 is valid on V_{k+1}.

Condition 2.4 (Complexity of REFINE). *Let $\mathcal{M}_0$ be an initial conforming subdivision. The REFINE procedure produces the sequence of subdivisions $\{\mathcal{M}_k\}_{k\geq 0}$ such that*

$$\#\mathcal{M}_k - \#\mathcal{M}_0 \leq C_0 \sum_{j=0}^{k-1} \#\widehat{\mathcal{M}}_j,$$

where $C_0 > 0$ is a universal constant depending on $\mathcal{M}_0$, the dimension d, and the number of refinements $b \geq 1$.

Condition 2.5 (Assumptions about MARK). *The set of marked elements $\widehat{\mathcal{M}}_k$ and marking parameter θ satisfy*

$$\widehat{\mathcal{M}}_k \text{ has minimal cardinality and } \theta \in (0, \theta_*),$$

where θ_* is defined by (5.10).

Condition 2.6 (Mesh overlay). *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be such that Condition 2.2 holds. Then, Condition 2.2 holds on $\mathcal{M}_1 \oplus \mathcal{M}_2$.*

3. The a posteriori error estimates

In this part, we derive the a posteriori error estimates of the MPDG, which are crucial ingredients leading to the convergence and quasi-optimality of the AMPDG.

3.1. Space decomposition and continuous approximation

It has been shown in [3] that the decomposition of V_h into its continuous and discontinuous components is useful. We introduce the following orthogonal decomposition of V_h:

$$V_h := V_h^0 \oplus V_h^{\perp}, \tag{3.1}$$

where $V_h^{\perp}$ is the orthogonal complement of V_h^0 in V_h with respect to the discrete scalar product $a_h(\cdot,\cdot)$.

Remark 3.1 (CG solution)**.** Let $u_h \in V_h$ be the DG solution of (2.8). Write $u_h = u_h^0 + u_h^\perp$, where $u_h^0 \in V_h^0$, $u_h^\perp \in V_h^\perp$. Since V_h^0 and $V_h^\perp$ are orthogonal with respect to $a_h(\cdot,\cdot)$, we obtain for all $v_h^0 \in V_h^0$,

$$(f, v_h^0) = a_h(u_h, v_h^0) = a_h(u_h^0, v_h^0), \tag{3.2}$$

that is, u_h^0 is the corresponding CG solution of (2.8).

Lemma 3.1 (Estimate on the nonconforming component)**.** *Let* $\gamma_0 \geq \gamma_0^{(c)}$*,* $0 \leq \gamma_1, \cdots, \gamma_p \lesssim 1$*. If* $v_h = v_h^0 + v_h^\perp \in V_h$ *according to* (3.1)*, and if Condition 2.2 holds, then*

$$|||v_h^\perp|||_{p,h} \lesssim \gamma_0^{\frac{1}{2}} \left\| h^{-\frac{1}{2}} [v_h] \right\|_{\mathcal{E}_h}. \tag{3.3}$$

The constant hiding in "$\lesssim$" depends on A, p and the shape regularity of $\mathcal{M}_h$.

Proof. This proof is similar to that of [3, Lemma 2.9], we write down below for completeness.

From the definition (3.1) of $V_h^\perp$, we obtain

$$a_h(v_h^\perp, v_h^\perp) = \inf_{w_h^0 \in V_h^0} a_h(v_h - w_h^0, v_h - w_h^0). \tag{3.4}$$

If $\gamma_0 \geq \gamma_0^{(c)}$, by invoking Lemma 2.1, we have

$$|||v_h^\perp|||_{p,h}^2 \lesssim a_h(v_h^\perp, v_h^\perp) \leq a_h(v_h - I_h v_h, v_h - I_h v_h) \lesssim |||v_h - I_h v_h|||_{p,h}^2.$$

We will estimate each term in the right-most side of the above inequality. From (2.15),

$$|||v_h - I_h v_h|||^2 \lesssim \sum_{K \in \mathcal{M}_h} \left\| h^{-\frac{1}{2}} [v_h] \right\|_{\sigma(K)}^2 \lesssim \left\| h^{-\frac{1}{2}} [v_h] \right\|_{\mathcal{E}_h}^2.$$

Since $I_h v_h \in H_0^1(\Omega)$, we can easily see that

$$J_h^0(v_h - I_h v_h, v_h - I_h v_h) = \gamma_0 \left\| h^{-\frac{1}{2}} [v_h] \right\|_{\mathcal{E}_h}^2.$$

When $0 \leq \gamma_1, \cdots, \gamma_p \lesssim 1$, Lemma 2.2 and (2.15) yield

$$\sum_{j=1}^{p} J_h^j(v_h - I_h v_h, v_h - I_h v_h) \lesssim \sum_{K \in \mathcal{M}_h} \|\nabla(v_h - I_h v_h)\|_K^2 \lesssim \left\| h^{-\frac{1}{2}} [v_h] \right\|_{\mathcal{E}_h}^2.$$

The conclusion follows by collecting all above estimates. □

3.2. The a posteriori error estimates

Given a discrete function $v_h \in V_h$, we define its *discrete energy error* to the exact solution u as follows:

$$e_h(v_h) := \left\{ |||u - v_h|||_{0,h}^2 + \sum_{j=1}^{p} J_h^j(v_h, v_h) \right\}^{\frac{1}{2}}.$$

Remark 3.2. Lemma 4.1 (below), which compares the errors of two approximate solutions on two successive meshes, may suggest using the discrete energy error $e_h(v_h)$ to measure the error $u - v_h$ for non-smooth solution $u \in H^1(\Omega)$.

The crucial ingredient of any adaptive algorithm is the control of the error by the estimator. The following lemma is similar to [3, Lemma 3.1], but contains something different. Our goal is to control the error $e_h(u_h)$.

Lemma 3.2 (First upper bound). *Let $\mathcal{M}_h \geq \mathcal{M}_0$, and let Condition* 2.2 *be valid. Let $u \in H_0^1(\Omega)$, $u_h \in V_h$ be the corresponding solutions of* (2.1)*,* (2.8)*, respectively. Then for $\gamma_0 \geq \gamma_0^{(c)}$ and $0 \leq \gamma_1, \cdots, \gamma_p \lesssim 1$, there holds that*

$$e_h^2(u_h) \lesssim E_h^2(u_h, \mathcal{M}_h) + \gamma_0 \left\| h^{-\frac{1}{2}} [u_h] \right\|_{\mathcal{E}_h}^2. \tag{3.5}$$

Proof. From (2.1) and (2.8), we have

$$a(u - u_h, v_h^0) = -\big(L_h(u_h), A\nabla v_h^0\big)_{\mathcal{M}_h} + \sum_{j=1}^{p} J_h^j(u_h, v_h^0) \quad \forall v_h^0 \in V_h^0. \tag{3.6}$$

Write $u_h = u_h^0 + u_h^{\perp}$ according to (3.1). Denote by $\xi = u - u_h = v - u_h^{\perp}$, where $v = u - u_h^0 \in H_0^1(\Omega)$, then

$$\begin{aligned} a(\xi, \xi) &= a(\xi, v - I_h v) - a(\xi, u_h^{\perp}) + a(\xi, I_h v) \\ &= a(\xi, v - I_h v) - a(\xi, u_h^{\perp}) - (L_h(u_h), A\nabla I_h v)_{\mathcal{M}_h} + \sum_{j=1}^{p} J_h^j(u_h, I_h v). \end{aligned} \tag{3.7}$$

Using the Cauchy-Schwarz inequality, (2.14) and Remark 2.5, we get

$$\begin{aligned} a(\xi, v - I_h v) &= \big(R(u_h), v - I_h v\big)_{\mathcal{M}_h} - \langle J(u_h), v - I_h v \rangle_{\mathcal{E}_h^I} \\ &\leq \left(\|hR(u_h)\|_{\mathcal{M}_h} + \left\| h^{\frac{1}{2}} J(u_h) \right\|_{\mathcal{E}_h^I} \right) \\ &\qquad \times \left(\left\| h^{-1}(v - I_h v) \right\|_{\mathcal{M}_h} + \left\| h^{-\frac{1}{2}}(v - I_h v) \right\|_{\mathcal{E}_h^I} \right) \\ &\lesssim \eta_h(u_h, \mathcal{M}_h) \, \|\nabla v\|_{\mathcal{M}_h}. \end{aligned}$$

Concerning the last term of (3.7), if $\gamma_1,\cdots,\gamma_p \lesssim 1$, then Lemma 2.2 and Remark 2.5 yield

$$\begin{aligned}\sum_{j=1}^{p} J_h^j(u_h, I_h v) &\le \Big(\sum_{j=1}^{p} J_h^j(u_h,u_h)\Big)^{\frac{1}{2}} \Big(\sum_{j=1}^{p} J_h^j(I_h v, I_h v)\Big)^{\frac{1}{2}} \\ &\lesssim E_h(u_h,\mathcal{M}_h)\,\|\nabla v\|_{\mathcal{M}_h}.\end{aligned}$$

Insert the above two estimates into (3.7), using (2.6), the triangle inequality, and Young's inequality, we obtain

$$\begin{aligned}a(\xi,\xi) &\lesssim \Big(E_h(u_h,\mathcal{M}_h) + \Big\|h^{-\frac{1}{2}}[u_h]\Big\|_{\mathcal{E}_h}\Big)\Big(|||\xi||| + |||u_h^{\perp}|||\Big) + |||\xi|||\cdot|||u_h^{\perp}||| \\ &\le \frac{1}{2}|||\xi|||^2 + C\Big(E_h^2(u_h,\mathcal{M}_h) + \Big\|h^{-\frac{1}{2}}[u_h]\Big\|_{\mathcal{E}_h}^2 + |||u_h^{\perp}|||^2\Big),\end{aligned}$$

which implies

$$|||\xi|||^2 \lesssim E_h^2(u_h,\mathcal{M}_h) + \Big\|h^{-\frac{1}{2}}[u_h]\Big\|_{\mathcal{E}_h}^2 + |||u_h^{\perp}|||^2.$$

Therefore

$$\begin{aligned}e_h^2(u_h) &= |||\xi|||^2 + \gamma_0\Big\|h^{-\frac{1}{2}}[u_h]\Big\|_{\mathcal{E}_h}^2 + \sum_{j=1}^{p} J_h^j(u_h,u_h) \\ &\lesssim E_h^2(u_h,\mathcal{M}_h) + \gamma_0\Big\|h^{-\frac{1}{2}}[u_h]\Big\|_{\mathcal{E}_h}^2 + |||u_h^{\perp}|||^2.\end{aligned}$$

The assertion follows from (3.3) and the above inequality. □

The upper bound we obtained is almost the same as the AIPDG [3]. The residual estimator $E_h(u_h,\mathcal{M}_h)$ is just what we want, while the jump term $\gamma_0^{\frac{1}{2}}\Big\|h^{-\frac{1}{2}}[u_h]\Big\|_{\mathcal{E}_h}$ is not. This is because adding the latter to $E_h(u_h,\mathcal{M}_h)$ would destroy the monotonicity property of the estimator (see Remark 2.3) with respect to the mesh size, which is instrumental for Lemma 2.3 and Lemma 2.4 and thus for the proof of the main result. However, we can control this jump term in terms of $E_h(u_h,\mathcal{M}_h)$ provided γ_0 sufficiently large, as stated in the following lemma. See [3, Lemma 3.3] for a similar result of the AIPDG.

Lemma 3.3 (Jump control)**.** *Let Condition* 2.2 *be valid on any partition* $\mathcal{M}_h$ *of* Ω*. Let* $u_h \in V_h$ *be the solution of* (2.8)*. There exists a constant* $\gamma_0^{(E)} \ge \gamma_0^{(c)}$ *such that for* $\gamma_0 \ge \gamma_0^{(E)}$ *and* $0 \le \gamma_1,\cdots,\gamma_p \lesssim 1$ *there holds that*

$$\gamma_0\Big\|h^{-\frac{1}{2}}[u_h]\Big\|_{\mathcal{E}_h} \lesssim E_h(u_h,\mathcal{M}_h). \tag{3.8}$$

Proof. This proof is similar to that of [3, Lemma 3.3], we write down below for completeness.

Let $\gamma_0 \ge \gamma_0^{(c)}$. From Lemma 2.1 and (2.8), for any $v_h^0 \in V_h^0$,

$$\begin{aligned}\gamma_0 \left\| h^{-\frac{1}{2}}[u_h] \right\|_{\mathcal{E}_h}^2 &= J_h^0(u_h - v_h^0, u_h - v_h^0) \\ &\lesssim a_h(u_h - v_h^0, u_h - v_h^0) = (f, u_h - v_h^0) - a_h(v_h^0, u_h - v_h^0). \end{aligned} \tag{3.9}$$

Moreover, (2.8), (2.7), (2.5), and integration by parts yield

$$\begin{aligned}&a_h(v_h^0, u_h - v_h^0) \\ =& a(u_h, u_h - v_h^0) - |||u_h - v_h^0|||^2 - \big(L_h(u_h), A\nabla v_h^0\big)_{\mathcal{M}_h} + \sum_{j=1}^p J_h^j(v_h^0, u_h - v_h^0) \\ =& \big(-\operatorname{div}(A\nabla u_h), u_h - v_h^0\big)_{\mathcal{M}_h} + \big\langle J(u_h), \{u_h - v_h^0\}\big\rangle_{\mathcal{E}_h^I} + \big(L_h(u_h), A\nabla(u_h - v_h^0)\big)_{\mathcal{M}_h} \\ &- |||u_h - v_h^0|||^2 + \sum_{j=1}^p J_h^j(u_h, u_h - v_h^0) - \sum_{j=1}^p J_h^j(u_h - v_h^0, u_h - v_h^0).\end{aligned}$$

Insert the above equality into (3.9), using the Cauchy-Schwarz inequality, (2.6), and Lemma 2.2, we have

$$\begin{aligned}\gamma_0 \left\| h^{-\frac{1}{2}}[u_h] \right\|_{\mathcal{E}_h}^2 \lesssim& \big(R(u_h), u_h - v_h^0\big)_{\mathcal{M}_h} - \big\langle J(u_h), \{u_h - v_h^0\}\big\rangle_{\mathcal{E}_h^I} \\ &- \big(L_h(u_h), A\nabla(u_h - v_h^0)\big)_{\mathcal{M}_h} + |||u_h - v_h^0|||^2 \\ &- \sum_{j=1}^p J_h^j(u_h, u_h - v_h^0) + \sum_{j=1}^p J_h^j(u_h - v_h^0, u_h - v_h^0) \\ \lesssim& \eta_h(u_h, \mathcal{M}_h)\left(\left\| h^{-1}(u_h - v_h^0)\right\|_{\mathcal{M}_h} + \left\| h^{-\frac{1}{2}}\{u_h - v_h^0\}\right\|_{\mathcal{E}_h^I}\right) \\ &+ \left\| h^{-\frac{1}{2}}[u_h] \right\|_{\mathcal{E}_h} |||u_h - v_h^0||| + |||u_h - v_h^0|||^2 \\ &+ E_h(u_h, \mathcal{M}_h)\left\| \nabla(u_h - v_h^0)\right\|_{\mathcal{M}_h} + \left\| \nabla(u_h - v_h^0)\right\|_{\mathcal{M}_h}^2.\end{aligned}$$

Let $v_h^0 = I_h u_h$, and denote C a constant independent of γ_0. From Condition 2.2 and the trace and inverse inequalities, we obtain

$$\gamma_0 \left\| h^{-\frac{1}{2}}[u_h] \right\|_{\mathcal{E}_h}^2 \le C E_h(u_h, \mathcal{M}_h)\left\| h^{-\frac{1}{2}}[u_h] \right\|_{\mathcal{E}_h} + C\left\| h^{-\frac{1}{2}}[u_h] \right\|_{\mathcal{E}_h}^2.$$

The proof finishes by setting $\gamma_0 \ge \gamma_0^{(E)} := \max(\gamma_0^{(c)}, 2C)$. □

As a direct consequence of the previous lemma, we obtain the following upper bound.

Theorem 3.1 (Second upper bound). *Let Condition* 2.2 *be valid on any partition* $\mathcal{M}_h$ *of* Ω. *Let* $u \in H_0^1(\Omega)$, $u_h \in V_h$ *be the corresponding solutions of* (2.1), (2.8)*, respectively. For* $\gamma_0 \geq \gamma_0^{(E)}$*, and* $0 \leq \gamma_1, \cdots, \gamma_p \lesssim 1$*, there exists a constant* $C_{GU} > 0$ *depending on* A*,* p *and the shape regularity of* $\mathcal{M}_h$ *such that*

$$e_h^2(u_h) \leq C_{GU} E_h^2(u_h, \mathcal{M}_h). \tag{3.10}$$

Proof. This is a directly consequence of (3.5) and (3.8). □

Lemma 3.4 (Global lower bound). *Let* $\mathcal{M}_h \geq \mathcal{M}_0$*, and let* $u \in H_0^1(\Omega)$*,* $u_h \in V_h$ *be the solutions of* (2.1), (2.8)*, respectively. Then there exists a constant* C_{GL} *depending on* A*,* p *and the shape regularity of* $\mathcal{M}_h$ *such that*

$$C_{GL} E_h^2(u_h, \mathcal{M}_h) \leq e_h^2(u_h) + \operatorname{osc}_h^2(u_h, \mathcal{M}_h). \tag{3.11}$$

Proof. From [3, Lemma 3.6],

$$\eta_h^2(u_h, \mathcal{M}_h) \lesssim \operatorname{osc}_h^2(u_h, \mathcal{M}_h) + |||u - u_h|||^2.$$

The proof completes by combing the above inequality and the truth that $\widetilde{J}_h^j(u_h, u_h) \eqsim J_h^j(u_h, u_h), j = 1, \cdots, p$. □

4. Convergence of the adaptive algorithm

In this part, we prove the convergence of the AMPDG method by showing that the AMPDG is a contraction with respect to the sum of the discrete energy error and a scaled error estimator. We first state the following lemma which compares the errors of two approximate solutions.

Lemma 4.1 (Quasi orthogonality). *Let* $\mathcal{M}_H \leq \mathcal{M}_h$ *be two successive refinements created by REFINE, and let* $u_H \in V_H$*,* $u_h \in V_h$ *be the corresponding solutions of* (2.8)*. Let* $u \in H_0^1(\Omega)$ *be the solution of* (2.1)*. If Condition* 2.3 *holds, then there exists a constant* C_m*, independent of* γ_0*,* $\gamma_1, \cdots, \gamma_p$ *and the mesh size, such that for all* $\gamma_0 \geq \gamma_0^{(E)}$ ($\gamma_0^{(E)}$ *is from Lemma* 3.3) *and* $0 < \varepsilon \leq 1$ *there holds that*

$$e_h^2(u_h) \leq (1+\varepsilon) e_H^2(u_H) - \frac{1}{4}|||u_h - u_H|||^2 + \frac{C_m}{\varepsilon\gamma_0}\big(E_h^2(u_h, \mathcal{M}_h) + E_H^2(u_H, \mathcal{M}_H)\big).$$

Proof. Write $u_H = u_H^0 + u_H^\perp$ and $u_h = u_h^0 + u_h^\perp$ according to (3.1). Notice that $u - u_h + u_h^0 - u_H^0 = u - u_H + u_H^\perp - u_h^\perp$, and $u_h^0 - u_H^0 \in V_h^0$, by invoking (3.6) we have

$$\begin{aligned} a(u-u_h, u-u_h) &= a(u-u_h+u_h^0-u_H^0, u-u_h+u_h^0-u_H^0) \\ &\quad - a(u_h^0-u_H^0, u_h^0-u_H^0) - 2a(u-u_h, u_h^0-u_H^0) \\ &= a(u-u_H+u_H^\perp-u_h^\perp, u-u_H+u_H^\perp-u_h^\perp) \\ &\quad - a(u_h^0-u_H^0, u_h^0-u_H^0) + 2\big(L_h(u_h), A\nabla(u_h^0-u_H^0)\big)_{\mathcal{M}_h} \\ &\quad - 2\sum_{j=1}^p J_h^j(u_h^\perp, u_h^0-u_H^0) - 2\sum_{j=1}^p J_h^j(u_h^0, u_h^0-u_H^0). \end{aligned}$$

The definition (3.1) of $V_h^\perp$ implies $a_h(u_h^\perp, u_h^0 - u_H^0) = 0$, that is

$$a(u_h^\perp, u_h^0-u_H^0) = \big(L_h(u_h^\perp), A\nabla(u_h^0-u_H^0)\big)_{\mathcal{M}_h} - \sum_{j=1}^p J_h^j(u_h^\perp, u_h^0-u_H^0).$$

Since $L_h(u_h) = L_h(u_h^\perp)$, we obtain

$$\begin{aligned} |||u-u_h|||^2 &= |||u-u_H+u_H^\perp-u_h^\perp|||^2 - |||u_h^0-u_H^0|||^2 \\ &\quad + 2a(u_h^\perp, u_h^0-u_H^0) - 2\sum_{j=1}^p J_h^j(u_h^0,u_h^0) + 2\sum_{j=1}^p J_h^j(u_h^0,u_H^0). \end{aligned} \tag{4.1}$$

Noting that $[A\nabla u_H^0 \cdot n]|_e = 0$ for any $e \in \mathcal{E}_h^I$ in the interior of some element $K' \in \mathcal{M}_H$, therefore

$$\begin{aligned} 2\sum_{j=1}^p J_h^j(u_h^0,u_H^0) &\le \sum_{j=1}^p J_h^j(u_h^0,u_h^0) + \sum_{j=1}^p J_h^j(u_H^0,u_H^0) \\ &\le \sum_{j=1}^p J_h^j(u_h^0,u_h^0) + \sum_{j=1}^p J_H^j(u_H^0,u_H^0). \end{aligned}$$

Inserting the above inequality into (4.1), and using Young's inequality lead to

$$\begin{aligned} |||u-u_h|||^2 &\le |||u-u_H+u_H^\perp-u_h^\perp|||^2 - |||u_h^0-u_H^0|||^2 \\ &\quad + 2|||u_h^\perp|||\cdot|||u_h^0-u_H^0||| - \sum_{j=1}^p J_h^j(u_h^0,u_h^0) + \sum_{j=1}^p J_H^j(u_H^0,u_H^0) \\ &\le \Big(1+\frac{\varepsilon}{3}\Big)|||u-u_H|||^2 + \Big(1+\frac{3}{\varepsilon}\Big)|||u_h^\perp - u_H^\perp|||^2 - \frac{1}{2}|||u_h^0-u_H^0|||^2 \\ &\quad + 2|||u_h^\perp|||^2 - \sum_{j=1}^p J_h^j(u_h^0,u_h^0) + \sum_{j=1}^p J_H^j(u_H^0,u_H^0). \end{aligned}$$

The triangle inequality implies

$$-|||u_h^0 - u_H^0|||^2 \leq |||u_h^\perp - u_H^\perp|||^2 - \frac{1}{2}|||u_h - u_H|||^2.$$

Let $\gamma_0 \geq \gamma_0^{(E)}$, then (3.3) and (3.8) yield

$$|||u_h^\perp - u_H^\perp|||^2 \lesssim \gamma_0^{-1}\big(E_h^2(u_h,\mathcal{M}_h) + E_H^2(u_H,\mathcal{M}_H)\big), \quad |||u_h^\perp|||^2 \lesssim \gamma_0^{-1}E_h^2(u_h,\mathcal{M}_h).$$

Collecting the above estimates, setting $\varepsilon \leq 1$, and denoting C a general constant independent of γ_0 and ε, we obtain

$$\begin{aligned}|||u-u_h|||^2 + \sum_{j=1}^p J_h^j(u_h^0,u_h^0) \leq \Big(1+\frac{\varepsilon}{3}\Big)|||u-u_H|||^2 + \sum_{j=1}^p J_H^j(u_H^0,u_H^0)\\ -\frac{1}{4}|||u_h-u_H|||^2 + \frac{C}{\varepsilon\gamma_0}\big(E_h^2(u_h,\mathcal{M}_h) + E_H^2(u_H,\mathcal{M}_H)\big).\end{aligned}$$

Therefore, from Young's inequality, Lemma 2.2, (3.3) and (3.8), we have

$$\begin{aligned}e_h^2(u_h) &= |||u-u_h|||^2 + \gamma_0\Big\|h^{-\frac{1}{2}}[u_h]\Big\|_{\mathcal{E}_h}^2 + \sum_{j=1}^p J_h^j(u_h,u_h)\\ &\leq \Big(1+\frac{\varepsilon}{3}\Big)\Big(|||u-u_h|||^2 + \sum_{j=1}^p J_h^j(u_h^0,u_h^0)\Big)\\ &\qquad + \gamma_0\Big\|h^{-\frac{1}{2}}[u_h]\Big\|_{\mathcal{E}_h}^2 + (1+\frac{3}{\varepsilon})\sum_{j=1}^p J_h^j(u_h^\perp,u_h^\perp)\\ &\leq \Big(1+\frac{\varepsilon}{3}\Big)^2|||u-u_H|||^2 + (1+\frac{\varepsilon}{3})\Big((1+\frac{\varepsilon}{3})\sum_{j=1}^p J_H^j(u_H,u_H)\\ &\qquad + \Big(1+\frac{3}{\varepsilon}\Big)\sum_{j=1}^p J_H^j(u_H^\perp,u_H^\perp)\Big) - \frac{1}{4}|||u_h-u_H|||^2\\ &\qquad + \frac{C}{\varepsilon\gamma_0}\big(E_h^2(u_h,\mathcal{M}_h) + E_H^2(u_H,\mathcal{M}_H)\big)\\ &\leq (1+\varepsilon)\Big(|||u-u_H|||^2 + \sum_{j=1}^p J_H^j(u_H,u_H)\Big) - \frac{1}{4}|||u_h-u_H|||^2\\ &\qquad + \frac{C}{\varepsilon\gamma_0}\big(E_h^2(u_h,\mathcal{M}_h) + E_H^2(u_H,\mathcal{M}_H)\big),\end{aligned}$$

which completes the proof of this lemma. □

Finally we prove the following contraction property of AMPDG method.

Theorem 4.1 (Contraction property). *Let Condition* 2.3 *be valid. Let $\mathcal{M}_0$ be the initial conforming partition of Ω and $\mathcal{M}_k \le \mathcal{M}_{k+1}$ be two consecutive meshes obtained from $\mathcal{M}_0$ by the adaptive algorithm. Let $\theta \in (0,1]$ be the Dörfler marking parameter. Let $u \in H_0^1(\Omega)$ be the solution of* (2.1)*, and let u_k, u_{k+1} be the corresponding DG solutions of* (2.8)*, respectively. Then there exist constants $\gamma_0^{(m)} \ge \gamma_0^{(E)}$ ($\gamma_0^{(E)}$ is from Lemma* 3.3*), $\beta > 0$ and $0 < \alpha < 1$, depending only on the shape regularity of $\mathcal{M}_0$, the matrix A, the polynomial degree p and the marking parameter θ, such that for $\gamma_0 \ge \gamma_0^{(m)}$ and $0 \le \gamma_1, \cdots, \gamma_p \lesssim 1$, there holds that*

$$e^2_{h_{k+1}}(u_{k+1}) + \beta E^2_{h_{k+1}}(u_{k+1}, \mathcal{M}_{k+1}) \le \alpha^2 \Big(e^2_{h_k}(u_k) + \beta E^2_{h_k}(u_k, \mathcal{M}_k) \Big).$$

Proof. For brevity, denote by

$$e_k := e_{h_k}(u_k), \quad E_k := E_{h_k}(u_k, \mathcal{M}_k), \quad \widehat{E}_k := E_{h_k}(u_k, \widehat{\mathcal{M}}_k).$$

Let $\beta' > 0$, Lemma 2.3 and Lemma 4.1 lead to

$$e^2_{k+1} + \left(\beta' - \frac{C_m}{\varepsilon \gamma_0} \right) E^2_{k+1} \le (1+\varepsilon) e^2_k - \frac{1}{4} |||u_{k+1} - u_k|||^2 + \beta'(1+\delta)(E^2_k - \lambda \widehat{E}^2_k) + \beta'(1+\delta^{-1}) C_E |||u_{k+1} - u_k|||^2 + \frac{C_m}{\varepsilon \gamma_0} E^2_k.$$

Choose β' dependent on δ to eliminate the term $|||u_{k+1} - u_k|||$, that is

$$\beta' = \frac{1}{4(1+\delta^{-1}) C_E} \iff (1+\delta)\beta' = \frac{\delta}{4 C_E},$$

to obtain

$$e^2_{k+1} + \left(\beta' - \frac{C_m}{\varepsilon \gamma_0} \right) E^2_{k+1} \le (1+\varepsilon) e^2_k + \beta'(1+\delta) \left(E^2_k - \lambda \widehat{E}^2_k \right) + \frac{C_m}{\varepsilon \gamma_0} E^2_k.$$

From the Dörfler marking strategy (2.13), we have

$$e^2_{k+1} + \left(\beta' - \frac{C_m}{\varepsilon \gamma_0} \right) E^2_{k+1} \le (1+\varepsilon) e^2_k + \left(\beta'(1+\delta)(1-\lambda\theta^2) + \frac{C_m}{\varepsilon \gamma_0} \right) E^2_k.$$

Let $\beta = \beta' - \frac{C_m}{\varepsilon\gamma_0}$. In order to prove

$$e^2_{k+1} + \beta E^2_{k+1} \le \alpha^2 \big(e^2_k + \beta E^2_k \big) \text{ for some } \beta > 0, 0 < \alpha < 1,$$

we only need to prove that

$$(1+\varepsilon) e^2_k + \left(\beta'(1+\delta)(1-\lambda\theta^2) + \frac{C_m}{\varepsilon \gamma_0} \right) E^2_k \le \alpha^2 \big(e^2_k + \beta E^2_k \big),$$

or, equivalently,

$$(1-\alpha^2+\varepsilon)e_k^2 \le \left(\alpha^2\big(\beta' - \frac{C_m}{\varepsilon\gamma_0}\big) - \beta'(1+\delta)(1-\lambda\theta^2) - \frac{C_m}{\varepsilon\gamma_0}\right) E_k^2.$$

Noting form (3.10) that $e_k^2 \le C_{GU} E_k^2$, we set

$$\alpha^2 = 1 - \frac{\beta'\big(1-(1+\delta)(1-\lambda\theta^2)\big) - \varepsilon C_{GU} - \frac{2C_m}{\varepsilon\gamma_0}}{C_{GU} + \beta' - \frac{C_m}{\varepsilon\gamma_0}}.$$

Choose δ such that $(1+\delta)(1-\lambda\theta^2) = 1 - \frac{1}{2}\lambda\theta^2$, that is, $\delta = \frac{\frac{1}{2}\lambda\theta^2}{1-\lambda\theta^2}$, then

$$\alpha^2 = 1 - \frac{\frac{1}{2}\lambda\theta^2\beta' - \varepsilon C_{GU} - \frac{2C_m}{\varepsilon\gamma_0}}{C_{GU} + \beta' - \frac{C_m}{\varepsilon\gamma_0}}.$$

Let $\varepsilon = \frac{\lambda\theta^2\beta'}{8C_{GU}}$ and choose $\gamma_0^{(m)} \ge \gamma_0^{(E)}$ such that $\frac{2C_m}{\varepsilon\gamma_0^{(m)}} \le \frac{1}{8}\lambda\theta^2\beta'$, then for $\gamma_0 \ge \gamma_0^{(m)}$,

$$\begin{aligned}
\alpha^2 &\le 1 - \frac{\frac{1}{2}\lambda\theta^2\beta' - \varepsilon C_{GU} - \frac{2C_m}{\varepsilon\gamma_0^{(m)}}}{C_{GU} + \beta'} \le 1 - \frac{\frac{1}{4}\lambda\theta^2\beta'}{C_{GU}+\beta'} < 1,\\
\beta &= \beta' - \frac{C_m}{\varepsilon\gamma_0} \ge \beta' - \frac{C_m}{\varepsilon\gamma_0^{(m)}} \ge \left(1 - \frac{1}{16}\lambda\theta^2\right)\beta' > 0,
\end{aligned}$$

which finishes the proof of this theorem. □

Remark 4.1. (i) As a direct consequence of the theorem we obtain

$$|||u - u_k||| \lesssim \alpha^k, \quad k = 0,1,2,\cdots,$$

which implies the convergence of the adaptive algorithm.

(ii) This proof and the resulting conclusion are similar as those of the FEM (cf. [10, Theorem 4.1]) and the IPDG (cf. [3, Theorem 4.4]), where different measure of errors and different orthogonality relation are employed.

(iii) The choice of δ implies that

$$\beta' = \frac{1}{4(1+\delta^{-1})C_E} = \frac{\lambda\theta^2}{8C_E(1-\frac{1}{2}\lambda\theta^2)} \Longrightarrow \frac{\lambda\theta^2}{8C_E} \le \beta' \le \frac{\lambda\theta^2}{8C_E(1-\frac{1}{2}\lambda)}.$$

Since $(1-\frac{1}{16}\lambda)\beta' \le (1-\frac{1}{16}\lambda\theta^2)\beta' \le \beta < \beta'$, we have $\beta \eqsim \theta^2$. Then we estimate α^2,

$$\begin{aligned}
\alpha^2 &\le 1 - \frac{\frac{1}{4}\lambda\theta^2\beta'}{C_{GU}+\beta'} = 1 - \frac{\lambda^2\theta^4}{32C_EC_{GU}(1-\frac{1}{2}\lambda\theta^2) + 4\lambda\theta^2}\\
&\le 1 - \frac{\lambda^2\theta^4}{32C_EC_{GU}+4},
\end{aligned}$$

that is, $\alpha^2 \le 1 - C\theta^4$ for some constant $C \in (0,1)$ independent of θ.

5. Quasi optimality of the AMPDG

The purpose of this section is to prove the following asymptotic estimate for the quasi-error

$$\left(e_{h_k}^2(u_k) + \beta E_{h_k}^2(u_k, \mathcal{M}_k)\right)^{\frac{1}{2}} \lesssim \left(\#\mathcal{M}_k - \#\mathcal{M}_0\right)^{-s}, \tag{5.1}$$

provided that $(u, f, D) \in \widetilde{\mathbb{A}}_s$, while $\widetilde{\mathbb{A}}_s$ is an approximation class defined in Section 5.3. Noting that when the penalty parameter $\gamma_1, \cdots, \gamma_p = 0$, the AMPDG reduces to the AIPDG [3], our results extend those of AIPDG.

5.1. Quasi-localized upper bound

To prove the optimality of the AMPDG, we need a localized upper bound for the distance between two nested solutions as stated in the next lemma, which is similar as the one in [3], but in a sense weaker than [10, 24].

Lemma 5.1 (Quasi-localized upper bound)**.** *Let Condition* 2.2 *hold. Let* $\mathcal{M}_0 \leq \mathcal{M}_H \leq \mathcal{M}_h$ *and write* $\mathcal{R}_H = \mathcal{R}_{\mathcal{M}_H \to \mathcal{M}_h}$*. Let* $u_H \in V_H$*,* $u_h \in V_h$ *be the corresponding solutions of* (2.8)*. Then there exists a constant* C_{LU} *depending on* A*,* p *and the shape regularity of* $\mathcal{M}_H$ *such that for* $\gamma_0 \geq \gamma_0^{(E)}$ *and* $\gamma_1, \cdots, \gamma_p \lesssim 1$*, there holds*

$$|||u_h^0 - u_H|||_{0,H}^2 \leq C_{LU}\Big(E_H^2\big(u_H, \omega(\mathcal{R}_H)\big) + \gamma_0^{-1} E_H^2(u_H, \mathcal{M}_H)\Big),$$

where $u_h = u_h^0 + u_h^{\perp}$ *is the orthogonal decomposition according to* (3.1)*, and* $\gamma_0^{(E)}$ *is from Lemma 3.3.*

Proof. Write $u_H = u_H^0 + u_H^{\perp}$ with $u_H^0 \in V_H^0$, $u_H^{\perp} \in V_H^{\perp}$ according to (3.1). From (2.7), (2.8) and (3.2), for any $v_H^0 \in V_H^0$, we have

$$\begin{aligned} a_H(u_h^0, v_H^0) &= a_h(u_h^0, v_H^0) - \sum_{j=1}^{p} J_h^j(u_h^0, v_H^0) + \sum_{j=1}^{p} J_H^j(u_h^0, v_H^0) \\ &= a_H(u_H, v_H^0) - \sum_{j=1}^{p} J_h^j(u_h^0, v_H^0) + \sum_{j=1}^{p} J_H^j(u_h^0, v_H^0). \end{aligned}$$

Therefore,

$$\begin{aligned} a_H^{\mathrm{L}}(u_h^0 - u_H, v_H^0) &= a_H(u_h^0 - u_H, v_H^0) - \sum_{j=1}^{p} J_H^j(u_h^0 - u_H, v_H^0) \\ &= \sum_{j=1}^{p} J_H^j(u_H, v_H^0) - \sum_{j=1}^{p} J_h^j(u_h^0, v_H^0). \end{aligned}$$

Write $u_h^0-u_H=z_h^0-u_H^\perp+v_H^0$ where $z_h^0=u_h^0-u_H^0-v_H^0\in V_h^0$. Using (2.8) and (3.2) yield

$$\begin{aligned}
&a_H^{\rm L}(u_h^0-u_H,u_h^0-u_H)\\
&=a_H^{\rm L}(u_h^0-u_H,z_h^0)-a_H^{\rm L}(u_h^0-u_H,u_H^\perp)+a_H^{\rm L}(u_h^0-u_H,v_H^0)\\
&=\Big(a_h(u_h^0,z_h^0)-\sum_{j=1}^p J_h^j(u_h^0,z_h^0)-a_H^{\rm L}(u_H,z_h^0)\Big)\\
&\qquad -a_H^{\rm L}(u_h^0-u_H,u_H^\perp)+\Big(\sum_{j=1}^p J_H^j(u_H,v_H^0)-\sum_{j=1}^p J_h^j(u_h^0,v_H^0)\Big)\\
&=(f,z_h^0)-a_H^{\rm L}(u_H,z_h^0)-a_H^{\rm L}(u_h^0-u_H,u_H^\perp)+\sum_{j=1}^p J_H^j(u_H,v_H^0)\\
&\qquad -\sum_{j=1}^p J_h^j(u_H^0,u_h^0-u_H^0)-\sum_{j=1}^p J_h^j(u_h^0-u_H^0,u_h^0-u_H^0).
\end{aligned}\tag{5.2}$$

Let $v_H^0=I_H(u_h^0-u_H^0)$, where I_H is given in Condition 2.2. Since I_H is locally a projection [3, Lemma 6.6], the error z_h^0 vanishes outside the set $\omega(\mathcal{R}_H)$ (see (2.3)), then (2.6) and the Cauchy-Schwarz inequalities imply

$$\begin{aligned}
&(f,z_h^0)-a_H^{\rm L}(u_H,z_h^0)\\
&=(f,z_h^0)-a(u_H,z_h^0)+\big(L_H(u_H),A\nabla z_h^0\big)_{\mathcal{M}_H}\\
&=\big(R(u_H),z_h^0\big)_{\omega(\mathcal{R}_H)}-\big\langle J(u_H),z_h^0\big\rangle_{\sigma(\mathcal{R}_H)}+\big(L_H(u_H),A\nabla z_h^0\big)_{\mathcal{M}_H}\\
&\lesssim\Big(\|HR(u_H)\|_{\omega(\mathcal{R}_H)}+\Big\|H^{\frac12}J(u_H)\Big\|_{\sigma(\mathcal{R}_H)}+\Big\|H^{-\frac12}[u_H]\Big\|_{\mathcal{E}_H}\Big)\\
&\qquad\times\Big(\|H^{-1}z_h^0\|_{\omega(\mathcal{R}_H)}+\Big\|H^{-\frac12}z_h^0\Big\|_{\sigma(\mathcal{R}_H)}+\|\nabla z_h^0\|_{\mathcal{M}_H}\Big).
\end{aligned}$$

By using the trace and inverse inequalities, (2.14) and Remark 2.5, we have

$$\|H^{-1}z_h^0\|_{\omega(\mathcal{R}_H)}+\Big\|H^{-\frac12}z_h^0\Big\|_{\sigma(\mathcal{R}_H)}+\|\nabla z_h^0\|_{\mathcal{M}_H}\lesssim\big\|\nabla(u_h^0-u_H^0)\big\|_{\mathcal{M}_h}.$$

The triangle inequality and (3.3) yield

$$\big\|\nabla(u_h^0-u_H^0)\big\|_{\mathcal{M}_h}\lesssim|||u_h^0-u_H|||+\gamma_0^{\frac12}\Big\|H^{-\frac12}[u_h^0-u_H]\Big\|_{\mathcal{E}_H}\lesssim|||u_h^0-u_H|||_{0,H}.$$

Collecting the above three estimates, together with (3.8) leads to

$$(f,z_h^0)-a_H^{\rm L}(u_H,z_h^0)\lesssim\big(\eta_H(u_H,\omega(\mathcal{R}_H))+\gamma_0^{-1}E_H(u_H,\mathcal{M}_H)\big)|||u_h^0-u_H|||_{0,H}.\tag{5.3}$$

By invoking Lemma 2.1, (3.3) and (3.8), the third term in (5.2) can be estimated as follows:

$$\begin{aligned}\left|a_H^{\rm L}(u_h^0-u_H,u_H^\perp)\right| &\lesssim |||u_H^\perp|||_{0,H}\cdot|||u_h^0-u_H|||_{0,H}\\ &\lesssim \gamma_0^{\frac{1}{2}}\left\|H^{-\frac{1}{2}}[u_H]\right\|_{\mathcal{E}_H}|||u_h^0-u_H|||_{0,H}\\ &\lesssim \gamma_0^{-\frac{1}{2}}E_H(u_H,\mathcal{M}_H)|||u_h^0-u_H|||_{0,H}.\end{aligned} \tag{5.4}$$

It remains to estimate the fourth and fifth terms in (5.2). Noting that $v_H^0=u_h^0-u_H^0$ outside the set $\omega(\mathcal{R}_H)$, let $\mathcal{E}_h^I(\omega(\mathcal{R}_H))$ be the restriction of $\mathcal{E}_h^I$ in $\omega(\mathcal{R}_H)$, then we can easily see that $\mathcal{E}_H^I\setminus\mathcal{E}_h^I(\sigma(\mathcal{R}_H))=\mathcal{E}_h^I\setminus\mathcal{E}_h^I(\omega(\mathcal{R}_H))$. From Lemma 2.2, (3.3), and (3.8), we have

$$\begin{aligned}&\sum_{j=1}^p J_H^j(u_H,v_H^0)-\sum_{j=1}^p J_h^j(u_H^0,u_h^0-u_H^0)\\ =&\sum_{j=1}^p J_H^j(u_H,v_H^0)-\sum_{j=1}^p J_h^j(u_H,u_h^0-u_H^0)+\sum_{j=1}^p J_h^j(u_H^\perp,u_h^0-u_H^0)\\ =&\sum_{j=1}^p J_H^j(u_H,v_H^0)_{\mathcal{E}_h^I(\sigma(\mathcal{R}_H))}-\sum_{j=1}^p J_h^j(u_H,u_h^0-u_H^0)_{\mathcal{E}_h^I(\omega(\mathcal{R}_H))}+\sum_{j=1}^p J_h^j(u_H^\perp,u_h^0-u_H^0)\\ \lesssim& E_H(u_H,\omega(\mathcal{R}_H))\left\|\nabla v_H^0\right\|_{\mathcal{M}_h}+E_H(u_H,\omega(\mathcal{R}_H))\left\|\nabla(u_h^0-u_H^0)\right\|_{\mathcal{M}_h}\\ &\quad+|||u_H^\perp|||\cdot\left\|\nabla(u_h^0-u_H^0)\right\|_{\mathcal{M}_h}\\ \lesssim&\left(E_H(u_H,\omega(\mathcal{R}_H))+\gamma_0^{-\frac{1}{2}}E_H(u_H,\mathcal{M}_H)\right)|||u_h^0-u_H|||_{0,H}.\end{aligned} \tag{5.5}$$

Insert (5.3), (5.4) and (5.5) into (5.2), and using Lemma 2.1 for $a_H^{\rm L}(u_h^0-u_H,u_h^0-u_H)$, we get the desired conclusion. □

Remark 5.1. Here we have estimated the jump terms (see (5.5)) carefully and obtained the quasi-localized upper bound for the AMPDG, which is almost the same as that for the AIPDG (cf. [3, Lemma 3.5]).

5.2. Céa's lemma on discrete total error

According to the upper bound (3.10) and the truth that $\mathrm{osc}_h^2(u_h,\mathcal{M}_h)\le\eta_h^2(u_h,\mathcal{M}_h)$, we have

$$e_h^2(u_h)+\mathrm{osc}_h^2(u_h,\mathcal{M}_h)\lesssim E_h^2(u_h,\mathcal{M}_h).$$

Combine the above inequality with the global lower bound (3.11) we realize that

$$E_h(u_h,\mathcal{M}_h)\approx\left(e_h^2(u_h)+\mathrm{osc}_h^2(u_h,\mathcal{M}_h)\right)^{\frac{1}{2}}, \tag{5.6}$$

and we call the right-hand side the *discrete total error*. We can also see that the discrete total error is equivalent to the *discrete quasi-error*

$$\big(e_h^2(u_h)+\mathrm{osc}_h^2(u_h,\mathcal{M}_h)\big)^{\frac12}\eqsim\big(e_h^2(u_h)+\beta E_h^2(u_h,\mathcal{M}_h)\big)^{\frac12}, \tag{5.7}$$

which is strictly reduced by the AMPDG method (see Theorem 4.1). Finally, the discrete total error satisfies the following Céa's lemma.

Lemma 5.2 (Quasi optimality of the discrete total error). *Let $\mathcal{M}_h\ge\mathcal{M}_0$ and Condition 2.2 hold. Let $u\in H_0^1(\Omega)$, $u_h\in V_h$ be the solutions of* (2.1), (2.8), *respectively. Then there exist constants C_{opt} and $\gamma_0^{(opt)}\ge\gamma_0^{(E)}$ depending on A, p and the shape regularity of $\mathcal{M}_0$, such that for $\gamma_0\ge\gamma_0^{(opt)}$ there holds*

$$e_h^2(u_h)+\mathrm{osc}_h^2(u_h,\mathcal{M}_h)\le C_{opt}\inf_{v_h\in V_h}\Big(e_h^2(v_h)+\mathrm{osc}_h^2(v_h,\mathcal{M}_h)\Big),$$

where $\gamma_0^{(E)}$ is from Lemma 3.3.

Proof. Write $u_h=u_h^0+u_h^\perp$, $v_h=v_h^0+v_h^\perp$ according to (3.1). Following the proof of Lemma 4.1, we have

$$\begin{aligned}
|||u-u_h|||^2&=|||u-v_h+v_h^\perp-u_h^\perp|||^2-|||u_h^0-v_h^0|||^2\\
&\quad+2a(u_h^\perp,u_h^0-v_h^0)-2\sum_{j=1}^p J_h^j(u_h^0,u_h^0)+2\sum_{j=1}^p J_h^j(u_h^0,v_h^0)\\
&\le 2|||u-v_h|||^2+2|||u_h^\perp-v_h^\perp|||^2-\frac12|||u_h^0-v_h^0|||^2\\
&\quad+2|||u_h^\perp|||^2-\sum_{j=1}^p J_h^j(u_h^0,u_h^0)+\sum_{j=1}^p J_h^j(v_h^0,v_h^0)\\
&\le 2|||u-v_h|||^2+\frac52|||u_h^\perp-v_h^\perp|||^2-\frac14|||u_h-v_h|||^2\\
&\quad+2|||u_h^\perp|||^2-\sum_{j=1}^p J_h^j(u_h^0,u_h^0)+\sum_{j=1}^p J_h^j(v_h^0,v_h^0).
\end{aligned}\tag{5.8}$$

Let $\gamma_0\ge\gamma_0^{(E)}$. Using Young's inequality, Lemma 2.2, (3.3), (3.8) and (3.11) yield

$$\begin{aligned}
e_h^2(u_h)&\le 2\Big(|||u-u_h|||^2+\sum_{j=1}^p J_h^j(u_h^0,u_h^0)\Big)+2\sum_{j=1}^p J_h^j(u_h^\perp,u_h^\perp)+\gamma_0\Big\|h^{-\frac12}[u_h]\Big\|_{\mathcal{E}_h}^2\\
&\le 4|||u-v_h|||^2+5|||u_h^\perp-v_h^\perp|||^2-\frac12|||u_h-v_h|||^2+4|||u_h^\perp|||^2\\
&\quad+4\sum_{j=1}^p J_h^j(v_h,v_h)+4\sum_{j=1}^p J_h^j(v_h^\perp,v_h^\perp)+2\sum_{j=1}^p J_h^j(u_h^\perp,u_h^\perp)+\gamma_0\Big\|h^{-\frac12}[u_h]\Big\|_{\mathcal{E}_h}^2\\
&\le Ce_h^2(v_h)-\frac12|||u_h-v_h|||^2+C\gamma_0^{-1}\big(e_h^2(u_h)+\mathrm{osc}_h^2(u_h,\mathcal{M}_h)\big).
\end{aligned}\tag{5.9}$$

Here $C \geq 4$ is a constant independent of γ_0. Let $\mathcal{M}^* = \mathcal{M}_H = \mathcal{M}_h$ and $v_H = u_h$ in Lemma 2.4, we get

$$\mathrm{osc}_h^2(u_h, \mathcal{M}_h) \leq 2\mathrm{osc}_h^2(v_h, \mathcal{M}_h) + C_{osc}|||u_h - v_h|||^2,$$

which combined with (5.9) leads to

$$\begin{aligned} & e_h^2(u_h) + \mathrm{osc}_h^2(u_h, \mathcal{M}_h) \\ \leq\ & 2\mathrm{osc}_h^2(v_h, \mathcal{M}_h) + C_{osc}|||u_h - v_h|||^2 + Ce_h^2(v_h) + C\gamma_0^{-1}\big(e_h^2(u_h) + \mathrm{osc}_h^2(u_h, \mathcal{M}_h)\big) \\ \leq\ & 2\mathrm{osc}_h^2(v_h, \mathcal{M}_h) + C\max(2C_{osc}, 1)\Big(e_h^2(v_h) + \gamma_0^{-1}\big(e_h^2(u_h) + \mathrm{osc}_h^2(u_h, \mathcal{M}_h)\big)\Big). \end{aligned}$$

Therefore

$$\begin{aligned} & \big(1 - C\max(2C_{osc}, 1)\gamma_0^{-1}\big)\Big(e_h^2(u_h) + \mathrm{osc}_h^2(u_h, \mathcal{M}_h)\Big) \\ \leq\ & C\max(2C_{osc}, 1)\Big(e_h^2(v_h) + \mathrm{osc}_h^2(v_h, \mathcal{M}_h)\Big). \end{aligned}$$

Let

$$\gamma_0^{(opt)} := \max\big(\gamma_0^{(E)}, 2C\max(2C_{osc}, 1)\big) \quad \text{and} \quad C_{opt} := 2C\max(2C_{osc}, 1),$$

which completes the proof of this lemma. □

5.3. Approximation class

Now we are going to define the nonlinear approximation class $\widetilde{\mathbb{A}}_s$ which is suitable for the AMPDG. The analysis leading to the definition of $\widetilde{\mathbb{A}}_s$ is similar as those of AFEM (see [10, Section 5.1]) and AIPDG (see [3, Section 5.2]), excepting that we employ the discrete energy error instead of the energy norm, so we only keep the important points.

Let N be an integer and $\mathbb{M}_N$ be the set of all possible subdivisions generated from $\mathcal{M}_0$ with at most N elements more than that of $\mathcal{M}_0$, that is,

$$\mathbb{M}_N := \big\{\mathcal{M}_h \,|\, \#\mathcal{M}_h - \#\mathcal{M}_0 \leq N\big\}.$$

The quality of the best approximation to the discrete total error in the set $\mathbb{M}_N$ is given by

$$\widetilde{\sigma}(N; u, f, A) := \inf_{\mathcal{M}_h \in \mathbb{M}_N} \inf_{v_h \in V_h} \Big(e_h^2(v_h) + \mathrm{osc}_h^2(v_h, \mathcal{M}_h)\Big)^{\frac{1}{2}},$$

and for $s > 0$ we define the nonlinear approximation class $\widetilde{\mathbb{A}}_s$ to be

$$\widetilde{\mathbb{A}}_s := \Big\{(u, f, A) \;\big|\; |u, f, A|_s := \sup_{N>0} \big(N^s \widetilde{\sigma}(N; u, f, A)\big) < \infty\Big\}.$$

We also define the counterparts for continuous finite element approximations

$$\widetilde{\sigma}^0(N;u,f,A) := \inf_{\mathcal{M}_h\in\mathbb{M}_N}\inf_{v_h^0\in V_h^0}\Big(e_h^2(v_h^0)+\mathrm{osc}_h^2(v_h^0,\mathcal{M}_h)\Big)^{\frac{1}{2}},$$

and

$$\widetilde{\mathbb{A}}_s^0 := \Big\{(u,f,A)\mid |u,f,A|_s^0 := \sup_{N>0}\big(N^s\widetilde{\sigma}^0(N;u,f,A)\big)<\infty\Big\}.$$

Remark 5.2. The approximation classes $\widetilde{\mathbb{A}}_s$ and $\widetilde{\mathbb{A}}_s^0$ defined here are different from those for IPDG in [3], which are defined as follows:

$$\mathbb{A}_s := \Big\{(u,f,A)\mid \sup_{N>0}\Big(N^s\inf_{\mathcal{M}_h\in\mathbb{M}_N}\inf_{v_h\in V_h}\big(|||u-v_h|||_{0,h}^2+\mathrm{osc}_h^2(v_h,\mathcal{M}_h)\big)^{\frac{1}{2}}\Big)<\infty\Big\},$$

$$\mathbb{A}_s^0 := \Big\{(u,f,A)\mid \sup_{N>0}\Big(N^s\inf_{\mathcal{M}_h\in\mathbb{M}_N}\inf_{v_h^0\in V_h^0}\big(|||u-v_h^0|||_{0,h}^2+\mathrm{osc}_h^2(v_h^0,\mathcal{M}_h)\big)^{\frac{1}{2}}\Big)<\infty\Big\}.$$

We remark that $\mathbb{A}_s^0\equiv\mathbb{A}_s$ if Condition 2.2 holds and $0<s\le\frac{p}{d}$ [3, Proposition 5.2].

It is obviously that $\widetilde{\mathbb{A}}_s\subset\mathbb{A}_s$ and $\widetilde{\mathbb{A}}_s^0\subset\mathbb{A}_s^0$. But we will show in Lemma 5.4 more relations among these approximation classes. To do so, we first state a perturbation result on the error estimator η.

Lemma 5.3 (Perturbation of estimator). *Let $\mathcal{M}_h$ be a refinement of $\mathcal{M}_0$. For any pair of discrete functions $v_h,w_h\in V_h$, we have*

$$\eta_h(v_h,\mathcal{M}_h)\lesssim\eta_h(w_h,\mathcal{M}_h)+|||v_h-w_h|||,$$

the constant hiding in "$\lesssim$" depends on A, p and the shape regularity of $\mathcal{M}_0$.

Proof. Although this lemma is stated for general mesh made of triangles or quadrilaterals, its original proof for conforming simplices (see [10]) is still valid, so we omit the details.

Lemma 5.4 (Equivalence classes). *Let Condition* 2.2 *be valid. Suppose that u, f and A satisfy* (2.1). *Then $\widetilde{\mathbb{A}}_s\equiv\widetilde{\mathbb{A}}_s^0$. Moreover, if $\gamma_2=\cdots=\gamma_p=0$ and $(u,f,A)\in\mathbb{A}_s$, then $(u,f,A)\in\widetilde{\mathbb{A}}_s$ and $(u,f,A)\in\widetilde{\mathbb{A}}_s^0$.*

Proof. If $(u,f,A)\in\widetilde{\mathbb{A}}_s^0$, then the relation $|u,f,A|_s\le|u,f,A|_s^0$ directly implies that $(u,f,A)\in\widetilde{\mathbb{A}}_s$, which leads to $\widetilde{\mathbb{A}}_s^0\subseteq\widetilde{\mathbb{A}}_s$.

Next we prove $\widetilde{\mathbb{A}}_s\subseteq\widetilde{\mathbb{A}}_s^0$. Let $(u,f,A)\in\widetilde{\mathbb{A}}_s$, for $N>0$, let $\mathcal{M}_{h_*}\in\mathbb{M}_N$, $v_{h_*}\in V_{h_*}$ be such that

$$e_{h_*}^2(v_{h_*})+\mathrm{osc}_{h_*}^2(v_{h_*},\mathcal{M}_{h_*})=\inf_{\mathcal{M}_h\in\mathbb{M}_N}\inf_{v_h\in V_h}\Big(e_h^2(v_h)+\mathrm{osc}_h^2(v_h,\mathcal{M}_h)\Big).$$

Let I_{h_*} be given by Condition 2.2. The triangle inequality, Lemma 2.2, (3.3) and the error estimates for I_{h_*} show that

$$\begin{aligned} e^2_{h_*}(I_{h_*}v_{h_*}) &= |||u - I_{h_*}v_{h_*}|||^2 + \sum_{j=1}^{p} J^j_{h_*}(I_{h_*}v_{h_*}, I_{h_*}v_{h_*}) \\ &\leq 2|||u - v_{h_*}|||^2 + 2\sum_{j=1}^{p} J^j_{h_*}(v_{h_*}, v_{h_*}) + 2|||v_{h_*} - I_{h_*}v_{h_*}|||^2 \\ &\quad + 2\sum_{j=1}^{p} J^j_{h_*}(v_{h_*} - I_{h_*}v_{h_*}, v_{h_*} - I_{h_*}v_{h_*}) \\ &\lesssim e^2_{h_*}(v_{h_*}). \end{aligned}$$

On the other hand, from Lemma 2.4,

$$\mathrm{osc}^2_{h_*}(I_{h_*}v_{h_*}, \mathcal{M}_{h_*}) \lesssim \mathrm{osc}^2_{h_*}(v_{h_*}, \mathcal{M}_{h_*}) + \left\|h^{-\frac{1}{2}}[v_{h_*}]\right\|^2_{\mathcal{E}_{h_*}}.$$

Therefore

$$e^2_{h_*}(I_{h_*}v_{h_*}) + \mathrm{osc}^2_{h_*}(I_{h_*}v_{h_*}, \mathcal{M}_{h_*}) \lesssim e^2_{h_*}(v_{h_*}) + \mathrm{osc}^2_{h_*}(v_{h_*}, \mathcal{M}_{h_*}) \lesssim N^{-s},$$

which implies $(u, f, A) \in \widetilde{\mathbb{A}}^0_s$, therefore $\widetilde{\mathbb{A}}_s \subseteq \widetilde{\mathbb{A}}^0_s$. Thus $\widetilde{\mathbb{A}}_s \equiv \widetilde{\mathbb{A}}^0_s$.

Suppose that $\gamma_2 = \cdots = \gamma_p = 0$, for any $\mathcal{M}_H \geq \mathcal{M}_0$ and $v_H \in V_H$, we have $E_H(v_H, \mathcal{M}_H) = \eta_H(v_H, \mathcal{M}_H)$. If $(u, f, A) \in \mathbb{A}_s$, then for any $N > 0$ there exists a subdivision $\mathcal{M}_h \geq \mathcal{M}_0$ such that

$$\#\mathcal{M}_h - \#\mathcal{M}_0 \leq N \quad \text{and} \quad \inf_{v_h \in V_h} \left(|||u - v_h|||^2_{0,h} + \mathrm{osc}^2_h(v_h, \mathcal{M}_h)\right)^{\frac{1}{2}} \lesssim N^{-s}.$$

Let $u^{\mathrm{L}}_h \in V_h$ be the discrete solution of IPDG method, that is, u^{L}_h satisfies

$$a^{\mathrm{L}}_h(u^{\mathrm{L}}_h, v_h) = (f, v_h) \quad \forall v_h \in V_h.$$

From (5.6) and Lemma 5.2 with $\gamma_1 = 0$, we know that

$$\begin{aligned} \eta_h(u^{\mathrm{L}}_h, \mathcal{M}_h) &\eqsim \left(|||u - u^{\mathrm{L}}_h|||^2_{0,h} + \mathrm{osc}^2_h(u^{\mathrm{L}}_h, \mathcal{M}_h)\right)^{\frac{1}{2}} \\ &\lesssim \inf_{v_h \in V_h} \left(|||u - v_h|||^2_{0,h} + \mathrm{osc}^2_h(v_h, \mathcal{M}_h)\right)^{\frac{1}{2}} \lesssim N^{-s}. \end{aligned}$$

By setting $v_h = u_h$ and $w_h = u^{\mathrm{L}}_h$ in Lemma 5.3 we deduce that

$$\eta_h(u_h, \mathcal{M}_h) \lesssim \eta_h(u^{\mathrm{L}}_h, \mathcal{M}_h) + |||u_h - u^{\mathrm{L}}_h|||,$$

where $|||u_h - u^{\mathrm{L}}_h|||$ can be estimated as follows:

$$\begin{aligned} |||u_h - u^{\mathrm{L}}_h|||^2 &\lesssim a^{\mathrm{L}}_h(u^{\mathrm{L}}_h - u_h, u^{\mathrm{L}}_h - u_h) + J^1_h(u^{\mathrm{L}}_h - u_h, u^{\mathrm{L}}_h - u_h) \\ &= a^{\mathrm{L}}_h(u^{\mathrm{L}}_h, u^{\mathrm{L}}_h - u_h) - a_h(u_h, u^{\mathrm{L}}_h - u_h) + J^1_h(u^{\mathrm{L}}_h, u^{\mathrm{L}}_h - u_h) \\ &= J^1_h(u^{\mathrm{L}}_h, u^{\mathrm{L}}_h - u_h) \lesssim \eta_h(u^{\mathrm{L}}_h, \mathcal{M}_h)|||u^{\mathrm{L}}_h - u_h|||. \end{aligned}$$

By invoking (5.6) and Lemma 5.2 we have

$$\begin{aligned}\widetilde{\sigma}(N;u,f,A) &\eqsim \inf_{\mathcal{M}_H\in\mathbb{M}_N}\eta_H(u_H,\mathcal{M}_H)\\ &\lesssim \eta_h(u_h,\mathcal{M}_h)\lesssim \eta_h(u_h^{\rm L},\mathcal{M}_h)\lesssim N^{-s},\end{aligned}$$

which implies $(u,f,A)\in\widetilde{\mathbb{A}}_s$. Since $\widetilde{\mathbb{A}}_s\equiv\widetilde{\mathbb{A}}_s^0$, we have $(u,f,A)\in\widetilde{\mathbb{A}}_s^0$. This completes the proof of the lemma. □

Remark 5.3. (i) The characterization of $\widetilde{\mathbb{A}}_s$ (or $\mathbb{A}_s$) and $\widetilde{\mathbb{A}}_s^0$ (or $\mathbb{A}_s^0$) is a pending issue. When the matrix A is piecewise polynomial of degree at most p over $\mathcal{M}_0$, some analysis is given in [10], while we won't talk about it in this paper.

(ii) We would like to point out that if $s\ge p/d$, then $\mathbb{A}_s$ (or $\mathbb{A}_s^0$) contains only trivial functions, see [10,12].

5.4. Cardinality of $\widehat{\mathcal{M}}_k$

Assume that $(u,f,A)\in\widetilde{\mathbb{A}}_s$ for some $s>0$. Similar as in [3], we are now going to prove that the approximation u_k generated by AMPDG converges to u with the same rate $(\#\mathcal{M}_k-\#\mathcal{M}_0)^{-s}$ as the best approximation described by $\widetilde{\mathbb{A}}_s$. We assume the following conditions regarding the parameters (θ,γ) and the marking procedure MARK.

$$\begin{aligned}&\gamma_0>\gamma_0^*:=\max\left(\frac{C_{om}+C_{LU}(2+C_{osc}+C_{om})}{C_{GL}},\gamma_0^{(m)},\gamma_0^{(opt)},\gamma_0^{(\mu)}\right),\\ &0<\theta<\theta_*:=\left(\frac{C_{GL}-\big(C_{om}+C_{LU}(2+C_{osc}+C_{om})\big)\gamma_0^{-1}}{1+C_{om}+C_{LU}(2+C_{osc}+C_{om})}\right)^{1/2}.\end{aligned}\tag{5.10}$$

Here $\gamma_0^{(m)}$, $\gamma_0^{(opt)}$, $\gamma_0^{(\mu)}$, C_{osc}, C_{GL}, C_{LU}, C_{om} are constants from Theorem 4.1, Lemma 5.2, Lemma 5.5 (below), (2.4), (3.11), Lemma 5.1 and (5.15) (below). We remark that $\theta_*>0$ if $\gamma_0>\gamma_0^*$.

In order to simplify the notation, let $0<\mu<\frac{1}{2}$ be defined by

$$\mu:=\frac{C_{GL}-\big(C_{om}+C_{LU}(2+C_{osc}+C_{om})\big)\gamma_0^{-1}}{2C_{GL}}\Big(1-\frac{\theta^2}{\theta_*^2}\Big),\quad\forall\gamma_0>\gamma_0^*,0<\theta<\theta_*.$$

Lemma 5.5 (Optimal marking)**.** *Let condition 2.3 hold. Let $\mathcal{M}_H$ be a refinement of $\mathcal{M}_0$ and let $\mathcal{M}_h$ be an admissible refinement of $\mathcal{M}_H$ such that*

$$e_h^2(u_h^0)+\operatorname{osc}_h^2(u_h^0,\mathcal{M}_h)\le\mu\Big(e_H^2(u_H)+\operatorname{osc}_H^2(u_H,\mathcal{M}_H)\Big),\tag{5.11}$$

where $u_H\in V_H$, $u_h\in V_h$ are the corresponding solutions of (2.8), respectively, and $u_h=u_h^0+u_h^\perp$ is the orthogonal decomposition according to (3.1). Then for $\gamma_0>\gamma_0^*$ and $\theta\in(0,\theta_*)$, the set of elements $\mathcal{R}_H:=\mathcal{R}_{\mathcal{M}_H\to\mathcal{M}_h}$ (see Section 2.2) satisfies a Dörfler marking property

$$E_H(u_H,\omega(\mathcal{R}_H))\ge\theta E_H(u_H,\mathcal{M}_H).$$

Proof. From the global lower bound (3.11) and (5.11) we have

$$\begin{aligned}&(1-2\mu)C_{GL}E_H^2(u_H,\mathcal{M}_H)\\ \leq\ &(1-2\mu)\Big(\mathrm{osc}_H^2(u_H,\mathcal{M}_H)+e_H^2(u_H)\Big)\\ \leq\ &\mathrm{osc}_H^2(u_H,\mathcal{M}_H)+e_H^2(u_H)-2\Big(\mathrm{osc}_h^2(u_h^0,\mathcal{M}_h)+e_h^2(u_h^0)\Big).\end{aligned}\tag{5.12}$$

Noting that $\mathrm{osc}_H^2(u_H,K')\leq\eta_H^2(u_H,K')$ for $K'\in\mathcal{M}_H$, and $\mathcal{M}_H\setminus\omega(\mathcal{R}_H)\subseteq\mathcal{M}_H\cap\mathcal{M}_h$, then Lemma 2.4 with $\mathcal{M}^*=\mathcal{M}_H\setminus\omega(\mathcal{R}_H)$ imply

$$\begin{aligned}&\mathrm{osc}_H^2(u_H,\mathcal{M}_H)-2\mathrm{osc}_h^2(u_h^0,\mathcal{M}_h)\\ \leq\ &\eta_H^2(u_H,\omega(\mathcal{R}_H))+\mathrm{osc}_H^2(u_H,\mathcal{M}_H\setminus\omega(\mathcal{R}_H))-2\mathrm{osc}_h^2(u_h^0,\mathcal{M}_H\setminus\omega(\mathcal{R}_H))\\ \leq\ &\eta_H^2(u_H,\omega(\mathcal{R}_H))+C_{osc}|||u_h^0-u_H|||_{0,H}^2.\end{aligned}\tag{5.13}$$

Next we estimate

$$\begin{aligned}e_H^2(u_H)-2e_h^2(u_h^0)=\ &|||u-u_H|||_{0,H}^2-2|||u-u_h^0|||_{0,h}^2\\ &+\sum_{j=1}^p J_H^j(u_H,u_H)-2\sum_{j=1}^p J_h^j(u_h,u_h).\end{aligned}$$

Since $u-u_h^0\in H_0^1(\Omega)$, the triangle inequality yields

$$|||u-u_H|||_{0,H}^2-2|||u-u_h^0|||_{0,h}^2\leq 2|||u_h^0-u_H|||_{0,H}^2.\tag{5.14}$$

Again using the triangle inequality, together with (3.3), (3.8) and Lemma 2.3 with $\delta=1$, leads to

$$\begin{aligned}|||u_h-u_H|||^2&\leq 2|||u_h^0-u_H|||^2+2|||u_h^\perp|||^2\\ &\leq 2|||u_h^0-u_H|||^2+C\gamma_0^{-1}\Big(E_H^2(u_H,\mathcal{M}_H)+C_E|||u_h-u_H|||^2\Big).\end{aligned}$$

Let $\gamma_0^{(\mu)}=\max(\gamma_0^{(E)},2CC_E)$. If $\gamma_0\geq\gamma_0^{(\mu)}$, then

$$|||u_h-u_H|||^2\leq 4|||u_h^0-u_H|||^2+2C\gamma_0^{-1}E_H^2(u_H,\mathcal{M}_H).$$

Considering the fact that any edge in $\mathcal{E}_H^I\setminus\sigma(\mathcal{R}_H)$ is also an edge in $\mathcal{E}_h^I$, by using Lemma 2.2, together with the above inequality, we obtain

$$\begin{aligned}&\sum_{j=1}^p J_H^j(u_H,u_H)-2\sum_{j=1}^p J_h^j(u_h,u_h)\\ \leq\ &\sum_{j=1}^p J_H^j(u_H,u_H)_{\sigma(\mathcal{R}_H)}+\sum_{j=1}^p J_H^j(u_H,u_H)_{\mathcal{E}_H^I\setminus\sigma(\mathcal{R}_H)}-2\sum_{j=1}^p J_h^j(u_h,u_h)_{\mathcal{E}_H^I\setminus\sigma(\mathcal{R}_H)}\\ \leq\ &\sum_{j=1}^p J_H^j(u_H,u_H)_{\sigma(\mathcal{R}_H)}+2\sum_{j=1}^p J_h^j(u_h-u_H,u_h-u_H)_{\mathcal{E}_H^I\setminus\sigma(\mathcal{R}_H)}\\ \leq\ &C_{om}\Big(E_H^2(u_H,\omega(\mathcal{R}_H))+\gamma_0^{-1}E_H^2(u_H,\mathcal{M}_H)+|||u_h^0-u_H|||_{0,H}^2\Big),\end{aligned}\tag{5.15}$$

where C_{om} is a constant depending on A, p and the shape regularity of $\mathcal{M}_0$. Insert (5.14), (5.13) and (5.15) into (5.12), we arrive at

$$\begin{aligned}&(1-2\mu)C_{GL}E_H^2(u_H,\mathcal{M}_H)\\ \le\ &(1+C_{om})E_H^2(u_H,\omega(\mathcal{R}_H))+C_{om}\gamma_0^{-1}E_H^2(u_H,\mathcal{M}_H)\\ &+(2+C_{osc}+C_{om})|||u_h^0-u_H|||_{0,H}^2,\end{aligned}$$

which together with the quasi localized upper bound Lemma 5.1 yield

$$\begin{aligned}&(1-2\mu)C_{GL}E_H^2(u_H,\mathcal{M}_H)\\ \le\ &\big(1+C_{om}+C_{LU}(2+C_{osc}+C_{om})\big)E_H^2(u_H,\omega(\mathcal{R}_H))\\ &+\big(C_{om}+C_{LU}(2+C_{osc}+C_{om})\big)\gamma_0^{-1}E_H^2(u_H,\mathcal{M}_H).\end{aligned}$$

If $\gamma_0>\gamma_0^*$, then, employing the definitions of θ_* and u results in

$$\begin{aligned}&E_H^2(u_H,\omega(\mathcal{R}_H))\\ \ge\ &\frac{(1-2\mu)C_{GL}-\big(C_{om}+C_{LU}(2+C_{osc}+C_{om})\big)\gamma_0^{-1}}{1+C_{om}+C_{LU}(2+C_{osc}+C_{om})}E_H^2(u_H,\mathcal{M}_H)\\ =\ &\theta^2E_H^2(u_H,\mathcal{M}_H).\end{aligned}$$

This completes the proof of the lemma.

Lemma 5.6 (Cardinality of $\widehat{\mathcal{M}}_k$)**.** *Let Condition* 2.3, 2.4, 2.5, 2.6 *be valid. Let* $u\in H_0^1(\Omega)$ *be the solution of* (2.1)*, and* $u_k\in V_k$ *be the* k^{th} *solution of* (2.8) *generated by the adaptive algorithm. If* $(u,f,A)\in\widetilde{\mathbb{A}}_s$*, then, for* $\gamma_0>\gamma_0^*$ *and* $0<\theta<\theta_*$*, the following estimate holds:*

$$\#\widehat{\mathcal{M}}_k\lesssim|u,f,A|_s^{\frac{1}{s}}\mu^{-\frac{1}{2s}}C_{opt}^{\frac{1}{2s}}\Big(e_{h_k}^2(u_k)+\mathrm{osc}_k^2(u_k,\mathcal{M}_k)\Big)^{-\frac{1}{2s}}.$$

The constant hiding in "$\lesssim$" is independent of k.

Proof. This proof is similar to that of [3, Lemma 5.5], we write down below for completeness. Let

$$\varepsilon^2:=\mu C_{opt}^{-1}\Big(e_{h_k}^2(u_k)+\mathrm{osc}_{h_k}^2(u_k,\mathcal{M}_k)\Big).$$

Since $(u,f,A)\in\widetilde{\mathbb{A}}_s$, from Lemma 5.4 we have $(u,f,A)\in\widetilde{\mathbb{A}}_s^0$, then from the definition of $\widetilde{\mathbb{A}}_s^0$ with $N\eqsim|u,f,A|_s^{\frac{1}{s}}\varepsilon^{-\frac{1}{s}}$, there exists $(\mathcal{M}_{\widehat{h}},v_{\widehat{h}}^0)$ with $\mathcal{M}_{\widehat{h}}\ge\mathcal{M}_0, v_{\widehat{h}}^0\in V_{\widehat{h}}^0$ such that

$$\#\mathcal{M}_{\widehat{h}}-\#\mathcal{M}_0\lesssim|u,f,A|_s^{\frac{1}{s}}\varepsilon^{-\frac{1}{s}}\quad\text{and}\quad e_{\widehat{h}}^2(v_{\widehat{h}}^0)+\mathrm{osc}_{\widehat{h}}^2(v_{\widehat{h}}^0,\mathcal{M}_{\widehat{h}})\le\varepsilon^2. \tag{5.16}$$

Let $\mathcal{M}_{\widehat{k}} = \mathcal{M}_k \oplus \mathcal{M}_{\widehat{h}}$ be the overlay of $\mathcal{M}_k$ and $\mathcal{M}_{\widehat{h}}$, and let $u_{\widehat{k}}$ be the corresponding solution of (2.8). Write $u_{\widehat{k}} = u^0_{\widehat{k}} + u^{\perp}_{\widehat{k}}$ according to (3.1). Then from Lemma 5.2, Remark 2.3, and noting that $e^2_{h_{\widehat{k}}}(v^0_{\widehat{h}}) \le e^2_{\widehat{h}}(v^0_{\widehat{h}})$, we have

$$\begin{aligned} e^2_{h_{\widehat{k}}}(u^0_{\widehat{k}}) + \operatorname{osc}^2_{h_{\widehat{k}}}(u^0_{\widehat{k}}, \mathcal{M}_{\widehat{k}}) &\le C_{opt}\Big(e^2_{h_{\widehat{k}}}(v^0_{\widehat{h}}) + \operatorname{osc}^2_{h_{\widehat{k}}}(v^0_{\widehat{h}}, \mathcal{M}_{\widehat{k}})\Big) \\ &\le C_{opt}\Big(e^2_{\widehat{h}}(v^0_{\widehat{h}}) + \operatorname{osc}^2_{\widehat{h}}(v^0_{\widehat{h}}, \mathcal{M}_{\widehat{h}})\Big) \\ &\le C_{opt}\varepsilon^2 = \mu\Big(e^2_{h_k}(u_k) + \operatorname{osc}^2_{h_k}(u_k, \mathcal{M}_k)\Big). \end{aligned}$$

By using Lemma 5.5, we deduce that the subset $\mathcal{R}_k := \mathcal{R}_{\mathcal{M}_k \to \mathcal{M}_{k\varepsilon}}$ verifies the following Dörfler marking property for $\theta < \theta_*$:

$$E_{h_k}(u_k, \omega(\mathcal{R}_k)) \ge \theta E_{h_k}(u_k, \mathcal{M}_k).$$

According to Condition 2.5, the procedure MARK selects a subset $\widehat{\mathcal{M}}_k \subset \mathcal{M}_k$ with minimal cardinality, then

$$\#\widehat{\mathcal{M}}_k \le \#\omega(\mathcal{R}_k) \lesssim \#\mathcal{R}_k,$$

where we have used (2.2) and the shape regularity of $\mathcal{M}_k$ to get the last estimate. Then from Condition 2.3 and (2.4) we obtain

$$\#\widehat{\mathcal{M}}_k \lesssim \#\mathcal{R}_k \le \#\mathcal{M}_{\widehat{k}} - \#\mathcal{M}_k \le \#\mathcal{M}_{\widehat{h}} - \#\mathcal{M}_0.$$

The proof finishes by combining this with (5.16). □

5.5. Quasi optimality

Assume that $(u, f, A) \in \widetilde{\mathbb{A}}_s$ for some $s > 0$. We are going to prove that the approximation u_k generated by the AMPDG converges to u with the same rate $(\#\mathcal{M}_k - \#\mathcal{M}_0)^{-s}$ as the best approximation described by $\widetilde{\mathbb{A}}_s$.

As a consequence of the previous estimates and the fact that the AMPDG is a contraction for the discrete quasi-error, we obtain quasi optimality of the discrete quasi-error.

Theorem 5.1 (Quasi optimality). *Let Condition* 2.3, 2.4, 2.5, 2.6 *be valid. Let f and A satisfy the assumptions in Section 2.1, and let $u \in H^1_0(\Omega)$ be the solution of* (2.1). *Let $\{\mathcal{M}_k, V_k, u_k\}_{k\ge 0}$ be the sequence of meshes, approximate spaces, and discrete solutions generated by the AMPDG algorithm. Let γ_0^* and θ_* be given as in* (5.10). *If $(u, f, A) \in \widetilde{\mathbb{A}}_s$ and $\gamma_0 > \gamma_0^*$, $0 < \theta < \theta_*$, then there holds that*

$$\Big(e^2_{h_k}(u_k) + \beta E^2_{h_k}(u_k, \mathcal{M}_k)\Big)^{\frac{1}{2}} \lesssim \left(1 - \frac{\theta^2}{\theta^2_*}\right)^{-\frac{1}{2}} \theta^{-4s} |u, f, A|_s \left(\#\mathcal{M}_k - \#\mathcal{M}_0\right)^{-s}.$$

Proof. The global lower bound (3.11) implies

$$e_{h_j}^2(u_j) + \beta E_{h_j}^2(u_j,\mathcal{M}_j) \le \left(1+\frac{\beta}{C_{GL}}\right)\left(e_{h_j}^2(u_j) + \mathrm{osc}_{h_j}^2(u_j,\mathcal{M}_j)\right).$$

By invoking Condition 2.4, Lemma 5.6 and Theorem 4.1, we can deduce that

$$\begin{aligned}
&\#\mathcal{M}_k - \#\mathcal{M}_0 \lesssim \sum_{j=0}^{k-1} \#\widehat{\mathcal{M}}_j \\
&\lesssim |u,f,A|_s^{\frac{1}{s}} \mu^{-\frac{1}{2s}} \sum_{j=0}^{k-1}\left\{e_{h_j}^2(u_j) + \mathrm{osc}_{h_j}^2(u_j,\mathcal{M}_j)\right\}^{-\frac{1}{2s}} \\
&\lesssim \left(1+\frac{\beta}{C_{GL}}\right)^{\frac{1}{2s}} |u,f,A|_s^{\frac{1}{s}} \mu^{-\frac{1}{2s}} \sum_{j=0}^{k-1}\left\{e_{h_j}^2(u_j) + \beta E_{h_j}^2(u_j,\mathcal{M}_j)\right\}^{-\frac{1}{2s}} \\
&\lesssim \left(1+\frac{\beta}{C_{GL}}\right)^{\frac{1}{2s}} |u,f,A|_s^{\frac{1}{s}} \mu^{-\frac{1}{2s}} \left\{e_{h_k}^2(u_k) + \beta E_{h_k}^2(u_k,\mathcal{M}_k)\right\}^{-\frac{1}{2s}} \sum_{j=0}^{k-1}\alpha^{\frac{k-j}{s}} \\
&\lesssim \left(1+\frac{\beta}{C_{GL}}\right)^{\frac{1}{2s}} |u,f,A|_s^{\frac{1}{s}} \mu^{-\frac{1}{2s}} \left\{e_{h_k}^2(u_k) + \beta E_{h_k}^2(u_k,\mathcal{M}_k)\right\}^{-\frac{1}{2s}} \frac{\alpha^{\frac{1}{s}}}{1-\alpha^{\frac{1}{s}}}.
\end{aligned}$$

By raising to the sth power and reordering, we arrive at

$$\begin{aligned}
&\left(e_{h_k}^2(u_k) + \beta E_{h_k}^2(u_k,\mathcal{M}_k)\right)^{\frac{1}{2}} \\
&\lesssim \left(1+\frac{\beta}{C_{GL}}\right)^{\frac{1}{2}} \mu^{-\frac{1}{2}} \left(\frac{\alpha^{\frac{1}{s}}}{1-\alpha^{\frac{1}{s}}}\right)^s |u,f,A|_s (\#\mathcal{M}_k - \#\mathcal{M}_0)^{-s}.
\end{aligned}$$

From Remark 4.1(iii), we have $\beta \lesssim \theta^2$ and

$$\left(\frac{\alpha^{\frac{1}{s}}}{1-\alpha^{\frac{1}{s}}}\right)^s \le \frac{(1-C\theta^4)^{\frac{1}{2}}}{\left(1-(1-C\theta^4)^{\frac{1}{2s}}\right)^s}$$

for some constant $C\in(0,1)$. Simple calculations show that, given $s>0$, the function $g(t) := \frac{1-t^{\frac{1}{2s}}}{1-t}$ is monotone in $t\in(0,1)$. Since the limiting values $g(0+) = 1$ and $g(1-) = \frac{1}{2s}$, we have $g(t) \ge \min(1,\frac{1}{2s}) \gtrsim 1$. Apply this inequality with $t = 1-C\theta^4$ to obtain

$$\left(\frac{\alpha^{\frac{1}{s}}}{1-\alpha^{\frac{1}{s}}}\right)^s \lesssim \frac{(1-C\theta^4)^{\frac{1}{2}}}{(C\theta^4)^s} \lesssim \theta^{-4s}.$$

This completes the proof of the theorem. □

Remark 5.4. (i) The same decay rate is obtained for the discrete total error due to its equivalence to the discrete quasi-error (see (5.7)).

(ii) The quasi-optimal result is almost the same as that for the AIPDG (cf. [3, Theorem 5.7]) except we use the discrete quasi-error instead of the total error.

(iii) $|||u - u_k|||$ satisfies the same quasi-optimal convergence rate.

6. Concluding remarks

In this paper we have analyzed an adaptive multi-penalty discontinuous Galerkin method. Convergence and quasi-optimality of the AMPDG method are proved for the diffusion problem. Extra works have been done to deal with the additional penalty terms. We will investigate the convergence properties of the AMPDG for Helmholtz scattering problems with high wave numbers in a future work.

Acknowledgments The author would like to thank the referees for their helpful suggestions. This research was partially the National Natural Science Foundation of China under grants 11525103 and 91130004.

References

[1] D. Arnold, *An interior penalty finite element method with discontinuous elements*, SIAM J. Numer. Anal., 19 (1982), pp. 742–760.

[2] I. Babuška and M. Vogelius, *Feedback and adaptive finite element solution of one-dimensional boundary value problems*, Numer. Math., 44 (1984), pp. 75–102.

[3] A. Bonito and R. H. Nochetto, *Quasi-optimal convergence rate of an adaptive discontinuous Galerkin method*, SIAM J. Numer. Anal., 48 (2010), pp. 734–771.

[4] S.C. Brenner and L.R. Scott, *The mathematical theory of finite element methods*, Springer-Verlag, third ed., 2008.

[5] E. Burman, *A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty*, SIAM J. Numer. Anal., 43 (2005), pp. 2012–2033.

[6] E. Burman and A. Ern, *Stabilized Galerkin approximation of convection-diffusion-reaction equations: discrete maximum principle and convergence*, Math. Comp., 74 (2005), pp. 1637–1652.

[7] E. Burman and A. Ern, *Continuous interior penalty hp-finite element methods for advection and advection-diffusion equations*, Math. Comp., 259 (2007), pp. 1119–1140.

[8] E. Burman, M.A. Fernández, and P. Hansbo, *Continuous interior penalty finite element method for Oseen's equations*, SIAM J. Numer. Anal., 44 (2006), pp. 1248–1274.

[9] E. Burman and P. Hansbo, *Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems*, Comput. Meth. Appl. Mech. Engrg., 193 (2004), pp. 1437–1453.

[10] J. M. Cascon, C. Kreuzer, R. H. Nochetto, and K. G. Siebert, *Quasi-optimal convergence rate for an adaptive finite element method*, SIAM J. Numer. Anal., 46 (2008), pp. 2524–2550.

[11] P. G. Ciarlet, *The finite element method for elliptic problems*, North-Holland, 1978.

[12] R. A. DeVore, *Nonlinear approximation*, Acta numerica, 7 (1998), pp. 51–150.

[13] L. Diening and C. Kreuzer, *Linear convergence of an adaptive finite element method for the p-Laplacian equation*,SIAM J. Numer. Anal., 46 (2008), pp. 614–638.
[14] W. Dörfler, *A convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal., 33 (1996), pp. 1106–1124.
[15] R. H. W. Hoppe, G. Kanschat, and T. Warburton, *Convergence analysis of an adaptive interior penalty discontinuous Galerkin method*, SIAM J. Numer. Anal., 47 (2008), pp. 534–550.
[16] P. Houston, D. Schötzau, and T. P. Wihler, *Mixed hp-discontinuous Galerkin finite element methods for the Stokes problem in polygons*, in Numerical mathematics and advanced applications, Springer, 2004, pp. 493–501.
[17] J. Hu, Z.-C. Shi, and J.-C. Xu, *Convergence and optimality of the adaptive Morley element method*, Numer. Math., 121 (2012), pp. 731–752.
[18] J.-G. Huang and Y.-F. Xu, *Convergence and complexity of arbitrary order adaptive mixed element methods for the Poisson equation*, Science China Mathematics, 55 (2012), pp. 1083–1098.
[19] O. A. Karakashian and F. Pascal, *Convergence of adaptive discontinuous Galerkin approximations of second-order elliptic problems*, SIAM J. Numer. Anal., 45 (2007), pp. 641–665.
[20] K. Mekchay and R. H. Nochetto, *Convergence of adaptive finite element methods for general second order linear elliptic PDEs*, SIAM J. Numer. Anal., 43 (2005), pp. 1803–1827.
[21] P. Morin, R. H. Nochetto, and K. G. Siebert, *Data oscillation and convergence of adaptive FEM*, SIAM J. Numer. Anal., 38 (2000), pp. 466–488.
[22] P. Morin, R. H. Nochetto, and K. G. Siebert, *Convergence of adaptive finite element methods*, SIAM review, 44 (2002), pp. 631–658.
[23] P. Morin, K. G. Siebert, and A. Veeser, *A basic convergence result for conforming adaptive finite elements*, Math. Model. Meth. Appl. Sci., 18 (2008), pp. 707–737.
[24] R. Stevenson, *Optimality of a standard adaptive finite element method*, Found. Comput. Math., 7 (2007), pp. 245–269.
[25] H. J. Wu, *Pre-asymptotic error analysis of CIP-FEM and FEM for the Helmholtz equation with high wave number. part I: linear version*, IMA J. Numer. Anal. 34 (2014), pp. 1266–1288.
[26] L. X. Zhu and H. J. Wu, *Preasymptotic error analysis of CIP-FEM and FEM for Helmholtz equation with high wave number. part II: hp version*, SIAM J. Numer. Anal., 51 (2013), pp. 1828–1852.