DISCONTINUOUS GALERKIN FINITE ELEMENT METHODS FOR INTERFACE PROBLEMS: A PRIORI AND A POSTERIORI ERROR ESTIMATIONS

ZHIQIANG CAI†, XIU YE‡, AND SHUN ZHANG§

Abstract. Discontinuous Galerkin (DG) finite element methods were studied by many researchers for second-order elliptic partial differential equations, and a priori error estimates were established when the solution of the underlying problem is piecewise \( H^{3/2+\epsilon} \) smooth with \( \epsilon > 0 \). However, elliptic interface problems with intersecting interfaces do not possess such a smoothness. In this paper, we establish a quasi-optimal a priori error estimate for interface problems whose solutions are only \( H^{1+\alpha} \) smooth with \( \alpha \in (0, 1) \) and, hence, fill a theoretical gap of the DG method for elliptic problems with low regularity. The second part of the paper deals with the design and analysis of robust residual- and recovery-based a posteriori error estimators. Theoretically, we show that the residual and recovery estimators studied in this paper are robust with respect to the DG norm, i.e., their reliability and efficiency bounds do not depend on the jump, provided that the distribution of coefficients is locally quasi-monotone.

Key words. a priori error estimation, a posteriori error estimator, discontinuous Galerkin methods, interface problems

AMS subject classifications. 65N30, 65N15

DOI. 10.1137/100805133

1. Introduction. Consider the following interface problem:

\begin{equation}
-\nabla \cdot (k(x)\nabla u) = f \quad \text{in } \Omega
\end{equation}

with boundary conditions

\begin{equation}
u = g_D \quad \text{on } \Gamma_D \quad \text{and} \quad n \cdot (k\nabla u) = g_N \quad \text{on } \Gamma_N,
\end{equation}

where \( f, g_D, \) and \( g_N \) are given scalar-valued functions; \( \Omega \) is a bounded polygonal domain in \( \mathbb{R}^2 \) with boundary \( \partial \Omega = \Gamma_D \cup \Gamma_N \) and \( \Gamma_D \cap \Gamma_N = \emptyset; \) \( n = (n_1, n_2) \) is the outward unit vector normal to the boundary; and diffusion coefficient \( k(x) \) is positive and piecewise constant on polygonal subdomains of \( \Omega \) with possible large jumps across subdomain boundaries (interfaces):

\begin{equation}
k(x) = k_i > 0 \quad \text{in } \Omega_i
\end{equation}

for \( i = 1, \ldots, n \). Here, \( \{\Omega_i\}_{i=1}^n \) is a partition of the domain \( \Omega \) with \( \Omega_i \) being an open polygonal domain. Define
\[ k_{\text{min}} = \min_{1 \leq i \leq n} k_i \quad \text{and} \quad k_{\text{max}} = \max_{1 \leq i \leq n} k_i. \]

Assume that \( f \) is square integrable over \( \Omega \). For simplicity, we assume that \( g_D \) and \( g_N \) are piecewise linear and constant, respectively, and that \( \Gamma_D \) is not empty (i.e., \( \text{mes}(\Gamma_D) \neq 0 \)).

Discontinuous Galerkin (DG) finite element methods for elliptic boundary value problems have been studied since the late 1970s and now constitute an active research area (see, e.g., [6, 7] and recent books [25, 33]). For problems with discontinuous coefficients such as interface problems, the stabilization (edge jump) term in the DG finite element method needs special treatments in order to be robust. Robustness in this paper means that constants in a priori error estimates or in the reliability and efficiency bounds of a posteriori error estimators are independent of the jump of the coefficients. Recently, Ern, Stephansen, and Zunino [23] developed a DG method using general weighted averages and proper weights for the stabilization term for advection-diffusion equations and established a robust a priori error estimate, provided that the solution is piecewise \( H^2 \) smooth.

A posteriori error estimation for continuous Galerkin finite element methods has been extensively studied for the past three decades (see, e.g., books by Verfürth [37], Ainsworth and Oden [4], and Babuška and Strouboulis [8] and the references therein). Recently, there has been increasing interest in a posteriori error estimation for the DG finite element method (see, e.g., [27, 9, 26, 2, 20, 34, 35, 36, 18]).

For elliptic interface problems considered in this paper, robust a posteriori error estimators have been investigated. For the continuous Galerkin method, Bernardi and Verfürth [11] and Petzoldt [31] studied a residual-based estimator, and we in [14] studied recovery-based estimators that were further extended to mixed and non-conforming finite element methods in [15]. For a DG finite element method without proper weights, Ainsworth [3] developed an a posteriori error estimator based on the so-called numerical flux. This result is further extended to meshes with hanging nodes in [5]. However, the error estimator and its analysis in [3, 5] depend on the jumps of coefficients. Recently, Ern, Nicaise, and Vohralik [19] and Ern, Stephansen, and Vohralik [21] investigated an equilibrated error estimator using the numerical flux for the DG method with the harmonic average weight; and Ern and Stephansen [22] studied a residual-based error estimator for DG approximations to diffusion and to advection-diffusion-reaction equations. Both the equilibrated and the residual-based estimators studied in [21, 22] contain a so-called nonconforming error term which is the energy norm of difference between the DG approximation and its recovery through the Oswald interpolation. Theoretically, they showed that their reliability constants are independent of the jump of the diffusion coefficients, but their efficiency constants do depend on the jump. This dependency could be removed for the interface problem by using a modified Oswald interpolate (see, e.g., section 4) as remarked in [22], provided that the diffusion coefficient is locally quasi-monotone. Such a hypothesis is assumed in order to prove the robustness of estimators in [11, 31, 14, 15] and in this paper.

For elliptic interface problems with intersecting interfaces, it is well known [28, 24] that their solutions may have only \( H^{1+\epsilon} \) regularity with small \( \epsilon > 0 \) and are not piecewise \( H^s \), \( s > 3/2 \), smooth. Since the standard a priori error estimate of the DG method requires that the exact solution be piecewise \( H^{3/2+\epsilon} \) smooth, there are no known results for elliptic interface problems. One of the purposes of this paper is to fill this theoretical gap. To this end, we introduce a nonstandard variational formulation (see (2.14)) that uses general weighted averages. The formulation is
defined in an appropriate solution space $V$ (see (2.9)) that permits discontinuity across interior edges of a triangulation $\mathcal{T}$ and that does not require piecewise $H^{s}$, $s > 3/2$, smoothness. The corresponding DG (or standard Galerkin) finite element approximation is then the solution of this variational problem in a discontinuous (or continuous) finite dimensional subspace. In this setting, the error equation is then obtained in a straightforward manner and, most importantly, unnecessary smoothness of the solution is not assumed.

As usual, this formulation involves a parameter $\theta \in \{-1, 0, 1\}$ and a stabilization parameter $\gamma_\theta$. For $\theta = 1$, the formulation is stable for all $\gamma_\theta \geq 0$. For $\theta = -1$ or $0$, we show that there exists a positive constant $\gamma_\theta$ such that the variational problem in the discontinuous piecewise polynomial space is stable, provided that $\gamma_\theta \geq \gamma_\theta$. The $\gamma_\theta$ is computable and depends only on the shape of elements but not on the mesh size and the jump of the diffusion coefficient. With this discrete stability and the error equation, we are then able to obtain a robust a priori error estimate of the DG method for the interface problem with low regularity. Note that the DG method corresponding to $\theta = -1$ was first introduced and analyzed in [23] for smooth solution.

The second part of the paper deals with the development and analysis of various robust a posteriori error estimators including residual- and recovery-based a posteriori error estimators. The residual-based estimator studied in the paper is standard and may be regarded as an extension of that by Bernardi and Verfürth [11] and Petzold [31] to the DG method. The recovery-based a posteriori error estimators follow ideas of our previous work in [14, 15] for conforming, nonconforming, and mixed finite element methods. Theoretically, we show that the residual and recovery estimators are robust with respect to the DG norm; i.e., their reliability and efficiency bounds do not depend on the jump, provided that the distribution of coefficients is locally quasi-monotone. Finally, we remark that there is numerical evidence showing that both the residual and the recovery estimators are not subject to the locally quasi-monotone assumption (see, e.g., [14, 15]).

This paper is organized as follows. DG finite element methods are introduced and their well-posedness is established in section 2. In section 3, we obtain a robust a priori error estimate in a norm which is stronger than the broken energy norm. Modified Oswald and Clément types of interpolations are described in section 4. A residual-based a posteriori error estimator is introduced and analyzed in section 5. We introduce and analyze flux recoveries and the resulting recovery-based estimators in sections 6 and 7, respectively. Numerical results for a test problem are reported in section 8.

1.1. Notation. For a subdomain $G \subset \Omega$, we use the standard notation and definitions for Sobolev spaces (see, e.g., Lions and Magenes [29], Adams [1], or Grisvard [24]). For $G$ being an open region, denote the Sobolev space by $W^{s,r}(B)$ on $B = G$ or $\partial G$ equipped with the standard Sobolev norm $\| \cdot \|_{s,r,B}$, where $s$ is a real number and $1 \leq r \leq \infty$. When $r = 2$, $W^{s,2}(B)$ is a Hilbert space and is denoted by $H^s(B)$ with the norm $\| \cdot \|_{s,B}$. When $s = 0$, $W^{0,r}(B)$ is the standard $L^r(B)$ space. By the trace theorem [24], the trace of any function in $W^{s,r}(G)$ lies in $W^{s-1/r, r} (\partial G)$, provided that $s - 1/r > 0$ is not an integer.

In two dimensions, for a vector-valued function $\tau = (\tau_1, \tau_2)^t$, define the divergence by

$$\nabla \cdot \tau = \frac{\partial \tau_1}{\partial x_1} + \frac{\partial \tau_2}{\partial x_2}.$$
For a scalar-valued function \( v \), define the operator \( \nabla^\perp \) by

\[
\nabla^\perp v = Q \nabla v = \left( -\frac{\partial v}{\partial x_2}, \frac{\partial v}{\partial x_1} \right)^T \quad \text{with} \quad Q = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.
\]

We shall use the Hilbert space

\[
H(\text{div}; G) = \{ \tau \in L^2(G)^2 : \nabla \cdot \tau \in L^2(G) \}
\]

equipped with the norm

\[
\|\tau\|_{H(\text{div}; G)} = \left( \|\tau\|_{0,G}^2 + \|\nabla \cdot \tau\|_{0,G}^2 \right)^{1/2}.
\]

Finally, let

\[
H^1_{g,D}(\Omega) = \{ v \in H^1(\Omega) : v = g \text{ on } \Gamma_D \}
\]

and

\[
H_{g,N}(\text{div}; \Omega) = \{ \tau \in H(\text{div}; \Omega) : \tau \cdot n = g_N \text{ on } \Gamma_N \}.
\]

2. Discontinuous finite element approximation.

2.1. Finite element spaces. For simplicity of presentation, consider only triangular elements. Let \( \mathcal{T} = \{ K \} \) be a finite element partition of the domain \( \Omega \). Assume that the triangulation \( \mathcal{T} \) is regular (see [17]); i.e., for all \( K \in \mathcal{T} \), there exists a positive constant \( \kappa \) such that

\[
h_K \leq \kappa \rho_K,
\]

where \( h_K \) denotes the diameter of the element \( K \) and \( \rho_K \) the diameter of the largest circle that may be inscribed in \( K \). Note that the assumption of the regularity does not exclude highly locally refined meshes. Furthermore, assume that interfaces

\[
F = \{ \partial \Omega_i \cap \partial \Omega_j : i, j = 1, \ldots, n \}
\]
do not cut through any element \( K \in \mathcal{T} \).

Let \( P_k(K) \) be the space of polynomials of degree \( k \) on element \( K \). Denote the conforming continuous piecewise linear finite element space and the discontinuous Galerkin linear finite element space associated with the triangulation \( \mathcal{T} \) by

\[
\mathcal{U} = \{ v \in H^1(\Omega) : v|_K \in P_1(K) \quad \forall \quad K \in \mathcal{T} \}
\]

and

\[
\mathcal{U}^{DG} = \{ v \in L^2(\Omega) : v|_K \in P_1(K) \quad \forall \quad K \in \mathcal{T} \},
\]

and denote the subspaces of \( \mathcal{U} \) by

\[
\mathcal{U}_g = \{ v \in \mathcal{U} : v = g_D \text{ on } \Gamma_D \} \quad \text{and} \quad \mathcal{U}_0 = \{ v \in \mathcal{U} : v = 0 \text{ on } \Gamma_D \}.
\]

Denote the local lowest-order Raviart–Thomas (RT) [32, 13] and Brezzi–Douglas–Marini (BDM) spaces [12, 13] on element \( K \in \mathcal{T} \) by

\[
RT_0(K) = P_0(K)^2 + x P_0(K) \quad \text{and} \quad BDM_1(K) = P_1(K)^2,
\]
respectively, where \( \mathbf{x} = (x_1, x_2) \). Then the standard \( H(\text{div}; \Omega) \) conforming RT and BDM spaces are defined by

\[
RT_0 = \{ \tau \in H(\text{div}; \Omega) : \tau|_K \in RT_0(K) \quad \forall \ K \in \mathcal{T} \}
\]

and

\[
BDM_1 = \{ \tau \in H(\text{div}; \Omega) : \tau|_K \in BDM_1(K) \quad \forall \ K \in \mathcal{T} \},
\]

respectively. For convenience, denote \( RT_0 \) or \( BDM_1 \) by \( V \).

2.2. Jumps and averages. Denote by \( E_K \) the set of three edges of element \( K \in \mathcal{T} \). Denote the set of all edges of the triangulation \( \mathcal{T} \) by

\[
E := E_I \cup E_D \cup E_N,
\]

where \( E_I \) is the set of all interior element edges, and \( E_D \) and \( E_N \) are the sets of all boundary edges belonging to the respective \( \Gamma_D \) and \( \Gamma_N \). For each \( e \in E \), denote by \( h_e \) the length of the edge \( e \); denote by \( \mathbf{n}_e \) a unit vector normal to \( e \). When \( e \in E_D \cup E_N \), denote by \( K_{+}^e \) the element with the edge \( e \), and assume that \( \mathbf{n}_e \) is the unit outward normal vector. For each interior edge \( e \in E_I \), let \( K_{+}^e \) and \( K_{-}^e \) be the two elements sharing the common edge \( e \) such that the unit outward normal vector of \( K_{+}^e \) coincides with \( \mathbf{n}_e \). For any \( e \in E \), denote by \( v_{+}^e \) and \( v_{-}^e \), respectively, the traces of a function \( v \) over \( e \).

Define jumps over edges by

\[
[v]_e := \begin{cases} v_{+}^e - v_{-}^e, & e \in E_I, \\ v_{+}^e, & e \in E_D \cup E_N. \end{cases}
\]

Let \( w_{+}^e \in [0, 1] \) and \( w_{-}^e \in [0, 1] \) be weights defined on \( e \) satisfying

\[
(2.1) \quad w_{+}^e(x) + w_{-}^e(x) = 1,
\]

and define the weighted averages

\[
\{v(x)\}_w^e = \begin{cases} w_{+}^e v_{+}^e + w_{-}^e v_{-}^e, & e \in E_I, \\ v_{+}^e, & e \in E_D \cup E_N, \end{cases}
\]

and

\[
\{v(x)\}_w^e = \begin{cases} w_{+}^e v_{+}^e + w_{-}^e v_{-}^e, & e \in E_I, \\ 0, & e \in E_D \cup E_N, \end{cases}
\]

for all \( e \in E \). When there is no ambiguity, the subscript or superscript \( e \) in the designation of the jump and the weighted averages will be dropped. A simple calculation leads to the following identity:

\[
(2.2) \quad [wv]_e = \{v\}_w^e [u]_e + \{u\}_w^e [v]_e.
\]

Let \( e \) be the interface of elements \( K_{+}^e \) and \( K_{-}^e \),

\[
e = \partial K_{+}^e \cap \partial K_{-}^e,
\]
and denote by $k^e_+$ and $k^e_-$ the diffusion coefficients on $K^e_+$ and $K^e_-$, respectively. There are several possible choices of the weights:

$$w^e_{+,1} = \frac{1}{2}, \quad w^e_{+,2} = \frac{k^e_+}{k^e_+ + k^e_-}, \quad \text{and} \quad w^e_{+,3} = \frac{\sqrt{k^e_+}}{\sqrt{k^e_+} + \sqrt{k^e_-}}.$$  

Denote by

$$W^e = \{k\}^e_w$$

the weighted average of $k$ on edge $e$. For the above choices of the weights, $W^e$ is then the arithmetic, the harmonic, and the geometric averages:

$$W^e_{e,1} = \frac{k^e_+ + k^e_-}{2}, \quad W^e_{e,2} = \frac{2k^e_+ k^e_-}{k^e_+ + k^e_-}, \quad \text{and} \quad W^e_{e,3} = \sqrt{k^e_- k^e_+},$$

respectively. It is well known that these averages have the following relations:

$$W^e_{e,2} \leq W^e_{e,3} \leq W^e_{e,1}.$$  

Since

$$\frac{(w^e)^2 k^e_+}{W^e} = \frac{(w^e)^2 k^e_+}{w^e_+ k^e_+ + w^e_- k^e_-} \leq \frac{(w^e)^2 k^e_+}{w^e_+ k^e_+} \leq w^e_+ \leq 1,$$

the same argument shows that

$$\frac{(w^e)^2 k^e_+}{W^e} \leq 1 \quad \text{and} \quad \frac{(w^e)^2 k^e_-}{W^e} \leq 1$$

for any weights satisfying (2.1). Letting

$$k^e_{\min} = \min\{k^e_+, k^e_-.\} \quad \text{and} \quad k^e_{\max} = \max\{k^e_+, k^e_-\},$$

it is easy to check that

$$\frac{1}{2} k^e_{\max} \leq W^e_{e,1} \leq k^e_{\max}, \quad k^e_{\min} \leq W^e_{e,2} \leq 2k^e_{\min}, \quad \text{and} \quad k^e_{\min} \leq W^e_{e,3} \leq k^e_{\max},$$

which implies that the arithmetic and harmonic averages are equivalent to the maximum and the minimum, respectively. For boundary edge $e \in \partial K^e_+ \cap \partial \Omega$, set

$$w^e_+ = 1 \quad \text{and} \quad W^e = k^e_+.$$

### 2.3. DG finite element approximation.

To describe the DG finite element method, we introduce a nonstandard variational formulation for (1.1)–(1.2) defined in a proper solution space which is a subspace of piecewise $H^{1+\epsilon}$ functions for $0 < \epsilon \ll 1$. The corresponding DG finite element approximation is the solution of this variational problem in a discontinuous finite dimensional subspace. Also, the standard Galerkin method is the variational problem restricted in a continuous finite dimensional subspace.

To define a proper solution space, let us start with the following Green’s formula:

$$\int_{\partial K} (\nabla w \cdot \mathbf{n}) v \, ds := (\nabla w \cdot \mathbf{n}, v)_{\partial K} = (\Delta w, v)_K + (\nabla w, \nabla v)_K \quad \forall \ K \in \mathcal{T}$$
holds for all \( w \in H^{1+\epsilon}(K) \) with \( \Delta w \in L^2(K) \) and for all \( v \in H^{1-\epsilon}(K) \) with \( 0 < \epsilon < 1/2 \). Let \( H^{-\epsilon}(K) \) be the dual of \( H^{\epsilon}_0(K) \) which is the closure of \( C_0^\infty(K) \) in the \( H^s(K) \) norm. Since \( H^s(K) \) is the same space as \( H^s_0(K) \) for \( s \in (0, 1/2) \) (see, e.g., Theorem 1.4.2.4 in [24]), \( \nabla v \) is then in \( H^{-\epsilon}(K) \). That is, the term \( (\nabla w, \nabla v)_K \) in (2.7) can be viewed as a duality pairing between \( H^s(K)^2 \) and \( H^{-s}(K)^2 \). The validity of (2.7) follows from the standard density argument and the fact that (2.7) holds for \( C^\infty(K) \) functions.

By the trace theorem [24], \( v|_{\partial K} \) is in \( H^{1/2-\epsilon}(\partial K) \). Hence, the formal boundary integral in the left-hand side of (2.7) may be regarded as the duality pairing between \( H^{\epsilon-1/2}(\partial K) \) and \( H^{1/2-\epsilon}(\partial K) \), which is defined by the right-hand side of (2.7). Since, for each edge \( e \subset \partial K \), the trivial extension of functions in \( H^{1/2-\epsilon}(e) \) by zero to all of \( \partial K \) belongs to \( H^{1/2-\epsilon}(\partial K) \) (see, e.g., Theorem 1.5.2.3 in [24]), this interpretation enables us to define the duality pairing on each edge \( e \) of \( \partial K \),

\[
\int_e (\nabla w \cdot n) v \; ds := (\nabla w \cdot n, v)_e,
\]

where \( (\nabla w \cdot n)|_e \in H^{\epsilon-1/2}(e) \) and \( v|_e \in H^{1/2-\epsilon}(e) \).

**Lemma 2.1.** Letting \( K \in \mathcal{T} \), \( e \in \partial K \), and \( 0 < \epsilon < 1/2 \), for any \( \phi \in H^{1+\epsilon}(K) \) with \( \Delta \phi \in L^2(K) \), there exists a positive constant \( C \) independent of \( \phi \) such that

\[
||\nabla \phi \cdot n||_{e-1/2, e} \leq C (||\nabla \phi||_{e, K} + h_{K}^{1-\epsilon} ||\Delta \phi||_{0, K}).
\]

**Proof.** Inequality (2.8) is contained in the proof of Corollary 3.3 on page 1384 of [10]. For the convenience of readers, we provide a proof here. For any \( g \in H^{1/2-\epsilon}(e) \), there exists a lifting \( v_g \) of \( g \) such that \( v_g \in H^{1-\epsilon}(K) \), \( v_g|_e = g \), \( v_g|_{\partial K \setminus e} = 0 \), and

\[
||\nabla v_g||_{-\epsilon, K} + h_{K}^{1-\epsilon} ||v_g||_{0, K} \leq c ||g||_{1/2-\epsilon, e}.
\]

It then follows from the Green’s formula in (2.7), the Cauchy–Schwarz inequality, and the definition of the dual norm that

\[
(\nabla \phi \cdot n, g)_e = (\nabla \phi \cdot n, v_g)_{\partial K} = (\Delta \phi, v_g)_K + (\nabla \phi, \nabla v_g)_K
\]

\[
\leq ||\Delta \phi||_{0, K} ||v_g||_{0, K} + ||\nabla \phi||_{e, K} ||\nabla v_g||_{-\epsilon, K}
\]

\[
\leq C (||\nabla \phi||_{e, K} + h_{K}^{1-\epsilon} ||\Delta \phi||_{0, K}) ||g||_{1/2-\epsilon, e},
\]

which, combining with the definition of the dual norm

\[
||\nabla \phi||_{e-1/2, e} = \sup_{g \in H^{1/2-\epsilon}(e)} \frac{(\nabla \phi \cdot n, g)_e}{||g||_{1/2-\epsilon, e}},
\]

implies (2.8). This completes the proof of the lemma. \( \square \)

Denote by \( H^s(\mathcal{T}) \) the broken Sobolev space of degree \( s > 0 \) with respect to \( \mathcal{T} \),

\[
H^s(\mathcal{T}) = \{ v \in L^2(\Omega) : v|_K \in H^s(K) \quad \forall K \in \mathcal{T} \},
\]

and denote its subspace by

\[
V^s(\mathcal{T}) = \{ v \in H^s(\mathcal{T}) : \nabla \cdot (k \nabla v) \in L^2(K) \quad \forall K \in \mathcal{T} \}.
\]

Let \( u \) be the solution of problem (1.1)–(1.2); then it is well known from the regularity estimate [28, 24] that \( u \in H^{1+\alpha}(\Omega) \) for some positive \( \alpha \) which could be very small.
Since \( f \in L^2(\Omega) \), it is then easy to see that \( u \in V^{1+\epsilon}(\mathcal{T}) \) for any \( 0 < \epsilon < \alpha \), and the flux \( \sigma = -k \nabla u \) is in \( H(\text{div}; \Omega) \cap H^\epsilon(\mathcal{T})^2 \).

Denote the discrete gradient and divergence operators by
\[
(\nabla_h v)|_K = \nabla(v|_K) \quad \text{and} \quad (\nabla_h \cdot \tau)|_K = \nabla \cdot (\tau|_K)
\]
for all \( K \in \mathcal{T} \), respectively. Multiplying (1.1) by a test function \( v \in V^{1+\epsilon}(\mathcal{T}) \), integrating by parts, and using boundary conditions (1.2), we have
\[
(f, v) = \sum_{K \in \mathcal{T}} (k \nabla u, \nabla v)_K - \sum_{K \in \mathcal{T}} \int_{\partial K} (k \nabla u \cdot n_e) v \, ds
\]
\[
= (k \nabla_h u, \nabla_h v) - \sum_{e \in \mathcal{E}_I} \int_e \left[ (k \nabla u \cdot n_e) v \right] ds - \sum_{e \in \mathcal{E}_D} \int_e (k \nabla u \cdot n_e) v \, ds - \int_{\Gamma_N} g_N \nabla v \, ds.
\]
Together with (2.2) and the continuity of the flux, for all \( v \in V^{1+\epsilon}(\mathcal{T}) \),
\[
(2.10) \quad \int_e [k \nabla u \cdot n_e] \{v\} \, ds = 0 \quad \forall \, e \in \mathcal{E}_I
\]
implies that
\[
(2.11) \quad (k \nabla_h u, \nabla_h v) - \sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} \int_e [k \nabla u \cdot n_e] w[v] \, ds = (f, v) + \int_{\Gamma_N} g_N \nabla v \, ds
\]
for all \( v \in V^{1+\epsilon}(\mathcal{T}) \). By the continuity of \( u \in H^{1+\alpha}(\Omega) \) and the Dirichlet boundary condition, we have that
\[
(2.12) \quad \sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} \int_e \{k \nabla \cdot n_e\} w[u] \, ds = \sum_{e \in \mathcal{E}_D} g_N (k \nabla v \cdot n_e) \, ds
\]
and
\[
(2.13) \quad \sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} \int_e \gamma_\theta h_e^{-1} W_e[u] [v] \, ds = \sum_{e \in \mathcal{E}_D} \gamma_\theta h_e^{-1} W_e \int_e g_D v \, ds
\]
for all \( v \in V^{1+\epsilon}(\mathcal{T}) \), where \( \gamma_\theta > 0 \) is a parameter to be determined.

For \( \theta \in \{-1, 0, 1\} \), define a bilinear form for \( u, v \in V^{1+\epsilon}(\mathcal{T}) \) by
\[
a_\theta(u, v) = (k \nabla_h u, \nabla_h v) + \sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} \int_e \gamma_\theta h_e^{-1} W_e[u] [v] \, ds
\]
\[
- \sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} \int_e \{k \nabla u \cdot n_e\} w[v] \, ds + \theta \sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} \int_e \{k \nabla v \cdot n_e\} w[u] \, ds
\]
and a linear form for \( v \in V^{1+\epsilon}(\mathcal{T}) \) by
\[
f_\theta(v) = \sum_{K \in \mathcal{T}} (f, v)_K + \sum_{e \in \mathcal{E}_D} \gamma_\theta h_e^{-1} W_e \int_e g_D v \, ds
\]
\[
+ \sum_{e \in \mathcal{E}_N} \int_e g_N v \, ds + \theta \sum_{e \in \mathcal{E}_D} \int_e g_D (k \nabla v \cdot n_e) \, ds.
\]
Note that the bilinear form is symmetric if $\theta = -1$ and nonsymmetric otherwise. Using (2.11), (2.12), and (2.13), the weak solution of (1.1)–(1.2) satisfies the following variational equation:

$$a_\theta(u, v) = f_\theta(v) \quad \forall \ v \in V^{1+\epsilon}(T).$$

Note that (2.13) is used in (2.14) for stabilizing the formulation or, equivalently, enforcing weakly the continuity of the solution. For $\theta = 1$, (2.12) is used for symmetrizing and stabilizing the bilinear form. Note also that the Dirichlet boundary condition is enforced weakly in (2.14). One could enforce it strongly in the solution space by removing all boundary integrals over edge $e \in \mathcal{E}_D$.

**Remark 2.2.** It is obvious that the bilinear form $a_1(\cdot, \cdot)$ is coercive in $V^{1+\epsilon}(T)$ with respect to the energy/DG norm if $\gamma_1 > 0$. But it is difficult, if not impossible, to show that the bilinear form $a_\theta(\cdot, \cdot)$ is coercive in $V^{1+\epsilon}(T)$ for $\theta = -1, 0$. Nonetheless, the fact that the weak solution of (1.1)–(1.2) satisfies problem (2.14) implies existence of solutions of problem (2.14). Uniqueness of problem (2.14) follows from that of (1.1)–(1.2) and the fact that solution of problem (2.14) satisfies (1.1)–(1.2) in the weak sense.

Now, the corresponding DG finite element method is to find $u_\tau \in \mathcal{U}^{DG} \subset V^{1+\epsilon}(T)$ such that

$$a_\theta(u_\tau, v) = f_\theta(v) \quad \forall \ v \in \mathcal{U}^{DG}.$$

Methods corresponding to $\theta = -1, 0, \text{ or } 1$ are the so-called symmetric interior penalty Galerkin (SIPG), incomplete interior penalty Galerkin (IIPG), or nonsymmetric interior penalty Galerkin (NIPG) methods, respectively. A special and interesting case of the NIPG method is $\gamma_1 = 0$, which was studied in [30] by Oden, Babuška, and Baumann. With the special choices of weights in (2.3), the corresponding DG methods in (2.15) are called the arithmetic, the harmonic, and the geometric weighted DG methods, respectively. The SIPG method defined in (2.15) with general weights but a slightly different stabilization term was introduced and analyzed recently in [23] and a robust a priori error bound was obtained, provided that the solution is piecewise $H^2$ smooth and that $\gamma_\theta$ is large enough.

### 2.4. Well-posedness of the DG finite element formulation.

For any element $K \in \mathcal{T}$, let $\{\lambda_i(x)\}_{i=1}^3$ be the standard barycentric coordinates of $K$. Following the idea of Lemma 1 in [3], we introduce an element stiffness matrix of the Laplace operator (instead of the diffusion operator in [3]):

$$S_K = (S_{ij}^K)_{3 \times 3} \quad \text{and} \quad S_{ij}^K = (\nabla \lambda_i, \nabla \lambda_j)_K.$$

The matrix $S_K$ is positive semidefinite, and its largest eigenvalue, $\rho(S_K)$, depends only on the shape of the element $K$ but not on its size $h_K$. Obviously, $\rho(S_K)$ is independent of the coefficient $k$. If shapes of the elements in $\mathcal{T}$ are reasonably regular, $\rho(S_K)$ is of similar size for all $K \in \mathcal{T}$. Denote the maximum of $\rho(S_K)$ over $\mathcal{T}$ by $\rho_T = \max_{K \in \mathcal{T}} \rho(S_K)$.

For any $v \in \mathcal{U}^{DG}$, let

$$\|v\|_{DG,K} = \left(\|k^{\frac{1}{2}} \nabla v\|_{0,K}^2 + \sum_{e \in \mathcal{E}_K \setminus \mathcal{E}_N} h_e^{-1} W_e \|v\|_{0,e}^2\right)^{1/2}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
for any $K \in T$. The so-called DG norm is defined as follows:

$$\|v\|_{DG} = \left( \|k^2 \nabla_h v\|_{0,\Omega}^2 + \sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} h_e^{-1} W_e \|v\|_{0,e}^2 \right)^{1/2}.$$  

**Lemma 2.3** (uniqueness and coercivity).

(i) The bilinear form $a_1(\cdot, \cdot)$ is coercive in $\mathcal{U}_{DG}$ with the coercivity constant $\min\{1, \gamma_1\}$, provided that $\gamma_1 > 0$, i.e.,

$$a_1(v, v) \geq \min\{1, \gamma_1\} \|v\|_{DG}^2 \quad \forall v \in \mathcal{U}_{DG}.$$

(ii) Let $w_+^e$ and $w_-^e$ be weights satisfying (2.1). Then SIPG and IIPG problems (2.15) have a unique solution, provided that $\gamma_0 > 2(1 + \theta)^2 \rho_\gamma$. Moreover, the symmetric/incomplete bilinear form $a_\theta(\cdot, \cdot)$ for $\theta = -1$ or $0$ is coercive in $\mathcal{U}_{DG}$ with a coercivity constant $\alpha_0 \in (0, 1)$ independent of the mesh size and the ratio $k_{max}/k_{min}$, i.e.,

$$a_\theta(v, v) \geq \alpha_0 \|v\|_{DG}^2 \quad \forall v \in \mathcal{U}_{DG},$$

for $\theta = -1$ and $0$, provided that $\gamma_0 > \frac{2(1 + \theta)^2 \rho_\gamma}{1 - \alpha_0} + \alpha_0$.

**Proof.** Let $\delta$ be a positive constant to be determined. For any $v \in \mathcal{U}_{DG}$ and for $e \in \mathcal{E}_I \cup \mathcal{E}_D$, the Cauchy–Schwarz inequality and the inequality of arithmetic and geometric means give

$$2 \sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} \int_e \{k \nabla v \cdot n_e\}_w[v] \, ds \leq \sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} \frac{\delta h_e}{W_e} \|k \nabla v \cdot n_e\}_w \|_{0,e}^2 + \sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} \frac{W_e}{\delta h_e} \|v\|_{0,e}^2.$$  

For $e \in \mathcal{E}_I$, let $e = \partial K_+^e \cap \partial K_-$.

Since $k \nabla v$ is constant on each element,

$$\frac{h_e}{W_e} \|k \nabla v \cdot n_e\}_w \|_{0,e}^2 \leq 2 h_e^2 \left\{ \left( \frac{w_+^e}{W_e} \right)^2 k_+ (\nabla v_+ \cdot n_e)^2 + \left( \frac{w_-^e}{W_e} \right)^2 k_- (\nabla v_- \cdot n_e)^2 \right\}$$

$$\leq 2 h_e^2 \max \left\{ \left( \frac{w_+^e}{W_e} \right)^2, \left( \frac{w_-^e}{W_e} \right)^2 \right\} (k_+ (\nabla v_+ \cdot n_e)^2 + k_- (\nabla v_- \cdot n_e)^2).$$

Similarly, for $e \in \mathcal{E}_D$ and $e \subset \partial K$, we have

$$\frac{h_e}{W_e} \|k \nabla v \cdot n_e\}_w \|_{0,e}^2 = k_+^{-1} h_e^2 (k_K \nabla v_K \cdot n_e)^2 = h_e^2 k_K (\nabla v_K \cdot n_e)^2.$$  

Summing up over all edges in $\mathcal{E}_I \cup \mathcal{E}_D$ and using (2.5) imply that

$$(2.19) \quad \sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} \frac{h_e}{W_e} \|k \nabla v \cdot n_e\}_w \|_{0,e}^2 \leq 2 \sum_{K \in T} \sum_{e \in \mathcal{E}_K} h_e^2 k_K (\nabla v_K \cdot n_e)^2.$$  

It is proved in [3] that

$$\sum_{e \in \mathcal{E}_K} h_e^2 (\nabla v \cdot n_K)^2 = 4 v_K^T S_K^2 \tilde{v}_K,$$
where $\vec{v}_K$ is the vector of values of $v$ at vertices of $K$. Since $k_K\vec{v}_K^T \mathbf{S}_K \vec{v}_K = (k \nabla v, \nabla v)_K$, thus
\[
\sum_{e \in \mathcal{E}_K} h_e^2 k_e (\nabla v_K \cdot \mathbf{n}_e)^2 \leq 4 k_K \vec{v}_K \mathbf{S}_K^2 \vec{v}_K \leq 4 \rho(\mathbf{S}_K) k_K \vec{v}_K^T \mathbf{S}_K \vec{v}_K = 4 \rho(\mathbf{S}_K) (k \nabla v, \nabla v)_K,
\]
which, together with (2.19), leads to
\[
\sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} \frac{h_e}{W_e} \| \{ k \nabla v \cdot \mathbf{n} \}_w \|_e^2 \leq 8 \rho_\tau (k \nabla h_v, \nabla h_v).
\]
Using (2.18), we now have
\[
\sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} \int_e \{ k \nabla v \cdot \mathbf{n}_e \}_w[v] \, ds \leq 4 \delta \rho_\tau (k \nabla h_v, \nabla h_v) + \sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} \frac{W_e}{2 \delta h_e} \| [v] \|_0,e^2.
\]
Hence,
\[
a_s(v, v) \geq (1 - 4(1 + \theta) \delta \rho_\tau)(k \nabla h_v, \nabla h_v) + \sum_{e \in \mathcal{E}_I \cup \mathcal{E}_D} \left( \gamma_\theta - \frac{1 + \theta}{2 \delta} \right) \frac{W_e}{h_e} \| [v] \|_0,e^2.
\]
For any constant $\alpha_0 \in [0, 1)$, assume that $\gamma_\theta > \frac{2(1+\theta)^2 \rho_\tau}{1-\alpha_0} + \alpha_0$; then there exists $\delta > 0$ such that $\frac{2(1-\alpha_0)}{1+\theta} > \delta^{-1} > \frac{2(1+\theta) \rho_\tau}{1-\alpha_0}$, which is equivalent to
\[
1 - 4(1 + \theta) \delta \rho_\tau > \alpha_0 \quad \text{and} \quad \gamma_\theta - \frac{1 + \theta}{2 \delta} > \alpha_0.
\]
This implies the coercivity of $a_s(v, v)$ in (2.17) for any $\alpha_0 \in (0, 1)$. When $\alpha_0 = 0$, it yields that $a_s(v, v)$ is positive and definite in $\mathcal{U}^{DG}$ and, hence, problem (2.15) has a unique solution. This completes the proof of the lemma. $\Box$

**Remark 2.4.** The constant $\gamma_\theta$ that appears in [5] is chosen to be greater than $(1 + \theta)^2 \max_{K \in \mathcal{T}} k_K \rho(\mathbf{S}_K)$, which depends on $k$ for $\theta \neq -1$, and, hence, it is not optimal.

**3. A priori error estimate.** Let $e = u - u_\tau$, where $u$ and $u_\tau$ are the solutions of (2.14) and (2.15), respectively. The difference of (2.14) and (2.15) yields the following error equation:
\[
(3.1) \quad a_s(e, v) = 0 \quad \forall \ v \in \mathcal{U}^{DG}.
\]
Let $\epsilon > 0$ be a very small constant, and define
\[
\| v \|_{k_\epsilon, \Omega} = \| k^{1/2} \nabla v \|_{0, \Omega}.
\]
Let $P_\tau : H^{1+s}_g(D)(\Omega) \to \mathcal{U}_g$ be the orthogonal projection operator from $H^{1+s}_g(D)(\Omega)$ onto $\mathcal{U}_g$ with respect to the inner product associated with the norm $\| \cdot \|_{k_\epsilon, \Omega}$. Then the standard interpolation argument and an analysis similar to that for Proposition 2.4 in [11] give that for $\phi \in H^{1+s}_g(D) \cap H^{1+s}(\mathcal{T})$ with $\epsilon \leq s \leq 1$,
\[
(3.2) \quad \| k^{1/2} \nabla (\phi - P_\tau \phi) \|_{0, \Omega} \leq C \left( \sum_{K \in \mathcal{T}} h_K^{2(s-\epsilon)} \| k^{1/2} \nabla \phi \|_0^2 \right)^{1/2},
\]
where $C$ is a positive constant independent of the mesh size and the ratio $k_{\text{max}}/k_{\text{min}}$. For any $v \in H^{1+s}(\mathcal{T})$, $0 < s \leq 1$, denote

$$B_s(h, v) = \left( \sum_{K \in \mathcal{T}} h_K^{2(s-\varepsilon)} \| k^{1/2} \nabla v \|^2_{s, K} \right)^{1/2} + \left( \sum_{K \in \mathcal{T}} h_K^{-1} \| f \|^2_{0, K} \right)^{1/2}.$$ 

**Lemma 3.1.** Assume that the solution $u \in V^{1+s}(\mathcal{T})$ of problem (2.14) belongs to $H^{1+s}(\mathcal{T})$ with $0 < \varepsilon \leq s \leq 1$. Then

$$\sum_{e \in E_{1} \cup E_{D}} \int_{e} \{ k \nabla (P_{T} u - u) \cdot n_{e} \}_{w} \| P_{T} u - u_{T} \|_{e} ds \leq C B_s(h, u) \| P_{T} u - u_{T} \|_{DG},$$

where $C$ is a positive constant independent of the mesh size and the ratio $k_{\text{max}}/k_{\text{min}}$.

**Proof.** Let $z = P_{T} u - u$ and $z_{T} = P_{T} u - u_{T}$. By using the definition of the dual norm, the triangle inequality, the inverse inequality, (2.5), Lemma 2.1, and (3.2), we have

$$\sum_{e \in E_{1} \cup E_{D}} \int_{e} \{ k \nabla (P_{T} u - u) \cdot n_{e} \}_{w} \| P_{T} u - u_{T} \|_{e} ds = \sum_{e \in E_{1} \cup E_{D}} \int_{e} \{ k \nabla z \cdot n_{e} \}_{w} \| z_{T} \|_{e} ds$$

$$\leq \sum_{e \in E_{1} \cup E_{D}} \| \{ k \nabla z \cdot n_{e} \}_{w} \|_{e-1/2, e} \| z_{T} \|_{1/2-\varepsilon, e}$$

$$\leq \sum_{e \in E_{1} \cup E_{D}} \left( \| k^{-1/2} \nabla z \|_{K_{e}} \| n_{e} \|_{e-1/2, e} + \| k^{1/2} \nabla z \|_{K_{e}} \| n_{e} \|_{e-1/2, e} \right) h_{e}^{-1/2} \| z_{T} \|_{0, e}$$

$$\leq C \sum_{e \in E_{1} \cup E_{D}} \left( \sum_{K_{e} \in u_{e}} h_{e}^{2} \| k^{1/2} \nabla z \|_{e, K} + h_{K} \| k^{1/2} \Delta z \|_{0, K} \right) W_{e}^{1/2} h_{e}^{-1/2} \| z_{T} \|_{0, e}$$

$$\leq C B_s(h, u) \| P_{T} u - u_{T} \|_{DG}.$$

This completes the proof of the lemma. \qed

**Theorem 3.2.** Assume that the solution $u \in V^{1+s}(\mathcal{T})$ of problem (2.14) belongs to $H^{1+s}(\mathcal{T}) \cap H^{1+s}(\Omega)$ with $0 < \varepsilon \leq s \leq 1$ and that $\gamma_{0} > \frac{2(1+\theta)^{2} \rho_{e}}{1-\alpha_{0}} + \alpha_{0}$ for $\theta = +1$, $0$, and $-1$. Then we have the following a priori error bound:

$$\| u - u_{T} \|_{DG} \leq C B_s(h, u),$$

where $C$ is a positive constant independent of the mesh size and the ratio $k_{\text{max}}/k_{\text{min}}$.

**Proof.** The triangle inequality gives

$$\| u \|_{DG} \leq \| u - P_{T} u \|_{DG} + \| P_{T} u - u_{T} \|_{DG}.$$ 

Since $u - P_{T} u$ is continuous and vanishes on $\Gamma_{D}$, thus

$$\| u - P_{T} u \|_{DG} = \| k^{1/2} \nabla (u - P_{T} u) \|_{0, \Omega} \leq \| k^{1/2} \nabla (u - P_{T} u) \|_{e, \Omega}.$$ 

Now, by (3.2) with $\phi = u$ it suffices to show that

$$\| P_{T} u - u_{T} \|_{DG} \leq C (\| u - P_{T} u \|_{DG} + B_s(h, u)).$$
To this end, using the coercivity of $a_\omega(\cdot, \cdot)$ in (2.17), the error equation in (3.1), the Cauchy–Schwarz inequality, and the fact that $\|P_T u - u_T\|_{DG} = 0$, we have
\[
\alpha_0\|P_T u - u_T\|^2_{DG} \leq a_\omega(P_T u - u_T, P_T u - u_T) = a_\omega(P_T u - u, P_T u - u_T)
\]
\[
= (k\nabla_h(P_T u - u), \nabla_h(P_T u - u_T)) + \sum_{e \in E \cup T} \int_e \gamma_\omega h_x^{-1}W_e |P_T u - u||P_T u - u_T| ds
\]
\[
- \sum_{e \in E \cup T} \int_e (\{k\nabla(P_T u - u) \cdot n_e\}_w [P_T u - u_T] - \theta [k\nabla(P_T u - u_T) \cdot n_e]_w [P_T u - u_T]) ds
\]
\[
\leq C \|P_T u - u\|_{DG} \|P_T u - u_T\|_{DG} + \sum_{e \in E \cup T} \int_e \{k \nabla(P_T u - u) \cdot n_e\}_w [P_T u - u_T] ds,
\]
which, together with Lemma 3.1, implies (3.5) and, hence, (3.4). This completes the proof of the theorem. \(\Box\)

4. Oswald- and Clément-type interpolations. Denote by $\mathcal{N}$, $\mathcal{N}_D$, and $\mathcal{N}_K$ the sets of all vertices of the triangulation $T$, on the $\Gamma_D$, and of element $K \in T$, respectively. For any $z \in N$, denote by $\phi_z$ the nodal basis function of $\mathcal{U}$, and let
\[
\omega_z = \{K \in T : K \subset \text{supp}(\phi_z)\} \quad \text{and} \quad \hat{\omega}_z = \left\{K \in \omega_z : k_K = \max_{K' \in \omega_z} k_{K'}\right\}.
\]
The number of elements in $\hat{\omega}_z$ is denoted by $cd(z)$. Also, denote by $\hat{E}_K$ the set of edges that share at least a vertex with $K$.

In this section and sections 5 and 7, assume that the distribution of the coefficients $k_K$ for all $K \in T$ is locally quasi-monotone [31], which is slightly weaker than Hypothesis 2.7 in [11]. For the convenience of the readers, we restate it here.

DEFINITION 4.1. Given a vertex $z \in \mathcal{N}$, the distribution of coefficients $k_K$, $K \in \omega_z$, is said to be quasi-monotone with respect to the vertex $z$ if there exists a subset $\hat{\omega}_{K,z,\text{qm}}$ of $\omega_z$ such that the union of elements in $\hat{\omega}_{K,z,\text{qm}}$ is a Lipschitz domain and that the following hold:

- if $z \in \mathcal{N} \setminus \mathcal{N}_D$, then $\{K\} \cup \hat{\omega}_z \subset \hat{\omega}_{K,z,\text{qm}}$ and $k_K \leq \max_{K' \in \omega_z} k_{K'}$;
- if $z \in \mathcal{N}_D$, then $K \in \hat{\omega}_{K,z,\text{qm}}$, $\partial \hat{\omega}_{K,z,\text{qm}} \cap \Gamma_D \neq \emptyset$, and $k_K \leq \max_{K' \in \omega_z} k_{K'}$.

The distribution of coefficients $k_K$, $K \in T$, is said to be locally quasi-monotone if it is quasi-monotone with respect to every vertex $z \in \mathcal{N}$.

For a given function $v \in \mathcal{U}^{DG}$, define the Oswald interpolation operator $\mathcal{I} : \mathcal{U}^{DG} \to \mathcal{U}_g$ by
\[
\mathcal{I}v = \sum_{z \in \mathcal{N}} \mathcal{I}v(z) \phi_z(x),
\]
where the nodal value of the interpolant $\mathcal{I}v$ at $z$ is defined by
\[
\mathcal{I}v(z) = \begin{cases}
    g_p(z) & \text{if } z \in \mathcal{N}_D, \\
    \frac{1}{cd(z)} \sum_{K \in \omega_z} v_K(z) & \text{if } z \in \mathcal{N} \setminus \mathcal{N}_D
\end{cases}
\]
with \( v_K = v|_K \) the restriction of \( v \) on \( K \).

**Lemma 4.2.** Assume that the triangulation \( T \) is quasi-uniform. Then for any \( v \in \mathcal{U}^{DG} \), there exists a positive constant \( C \) independent of the ratio \( k_{\max}/k_{\min} \) such that

\[
(4.1) \quad k_K \| v - \mathcal{I}v \|_{0,K}^2 \leq C \left( \sum_{e \in \partial K \setminus \partial D} h_e W_e \| [v] \|_{0,e}^2 + \sum_{e \in \partial K \cap \partial D} h_e W_e \| v - g_D \|_{0,e}^2 \right)
\]

and

\[
(4.2) \quad \| k_K^{1/2} \nabla (v - \mathcal{I}v) \|_{0,K}^2 \leq C \left( \sum_{e \in \partial K \setminus \partial D} W_e h_e \| [v] \|_{0,e}^2 + \sum_{e \in \partial K \cap \partial D} W_e h_e \| v - g_D \|_{0,e}^2 \right)
\]

for all \( K \in T \).

**Proof.** For any \( v \in \mathcal{U}^{DG} \) and any \( K \in T \), the inverse inequality implies that

\[
h_K^2 \| k_K^{1/2} \nabla (v - \mathcal{I}v) \|_{0,K}^2 \leq C k_K \| v - \mathcal{I}v \|_{0,K}^2.
\]

Hence, it suffices to establish the validity of (4.1).

To this end, for any \( K \in T \) and any \( z \in \mathcal{N}_K \), denote by \( \phi_z(K) = \phi_z(x)|_K \) the restriction of \( \phi_z \) in \( K \). Then

\[
v_K(x) = v|_K(x) = \sum_{z \in \mathcal{N}_K} v_K(z) \phi_z, K(x) \quad \forall v \in \mathcal{U}^{DG}.
\]

Since \( \| \phi_z, K \|_{0,K} \leq C h_K \) and

\[
v_K(z) - \mathcal{I}v(z) = v_K(z) - \frac{1}{cd(z)} \sum_{K' \in \hat{\omega}_z} v_{K'}(z) = \frac{1}{cd(z)} \sum_{K' \in \hat{\omega}_z} (v_K(z) - v_{K'}(z))
\]

for all \( z \in \mathcal{N} \setminus \mathcal{N}_D \), we then have

\[
\| v - \mathcal{I}v \|_{0,K} = \| \sum_{z \in \mathcal{N}_K} (v - \mathcal{I}v)(z) \phi_z, K \|_{0,K}
\]

\[
\leq \sum_{z \in \mathcal{N}_K} \| (v - \mathcal{I}v)(z) \phi_z, K \|_{0,K} \leq C \sum_{z \in \mathcal{N}_K} h_K v_K(z) - \mathcal{I}v(z)\|
\]

(4.3) \[
\leq C \left( \sum_{z \in \mathcal{N}_K \setminus \mathcal{N}_D} \sum_{K' \in \hat{\omega}_z} h_K |v_K(z) - v_{K'}(z)| + \sum_{z \in \mathcal{N}_K \setminus \mathcal{N}_D} h_K |g_D(z) - v_K(z)| \right).
\]

For \( z \in \mathcal{N}_K \setminus \mathcal{N}_D \) and any \( K' \in \hat{\omega}_z \), by the fact that the distribution of the coefficients of \( K \) is quasi-monotone with respect to \( z \), there is a connected path,

\[
\{ K = K_0, K_1, \ldots, K_l = K' \} \quad \text{with} \quad K_i \in \omega_z,
\]

from \( K \) to \( K' \) such that the diffusion coefficient is monotonically increasing. Denote by \( e_i \) the common edge of \( K_{i-1} \) and \( K_i \); then we have

\[
k_K \leq W_{e_1} \leq \cdots \leq W_{e_l}.
\]
Now, it follows from the triangle inequality and the inverse inequality that for any $K' \in \hat{\omega}_z$,

$$k_K^{1/2}|v_K(z) - v_{K'}(z)| \leq k_K^{1/2} \sum_{i=0}^l |v_{K_i}(z) - v_{K_{i+1}}(z)|$$

(4.4)

$$\leq \sum_{i=0}^l \sqrt{W_{e_i}} \|v\|_{\infty,e_i} \leq C \sum_{i=0}^l \sqrt{W_{e_i}} \|v\|_{0,e_i}.$$  

For $z \in N_K \cap N_D$, by Definition 4.1, there exists a $K' \in \hat{\omega}_z$ such that $E_K \cap E_D \neq \emptyset$. Let $e_D = E_K \cap E_D \neq \emptyset$; then the triangle inequality and an argument similar to that above yield

$$k_K^{1/2}|g_D - v_K(z)| \leq k_K^{1/2}|g_D - v_{K'}(z)| + k_K^{1/2}|v_{K'}(z) - v_K(z)|$$

$$\leq \sqrt{W_{e_D}} \|g_D - v\|_{\infty,e_D} + \sum_{i=0}^{l'} \sqrt{W_{e_i}} \|v\|_{\infty,e_i}$$

(4.5)

$$\leq C \sqrt{W_{e_D}} \|g_D - v\|_{0,e_D} + C \sum_{i=0}^{l'} \sqrt{W_{e_i}} \|v\|_{0,e_i}.$$  

Now, (4.1) is a direct consequence of (4.3), (4.4), (4.5), and the Cauchy–Schwarz inequality. This completes the proof of the lemma.  

Clément-type interpolation operators (see, e.g., [11, 31]) are often used for establishing the reliability bound of a posteriori error estimators. We define a weighted Clément-type interpolation operator and state its approximation and stability properties. For more details, see [14].

For a given function $v$, define its weighted average over $\hat{\omega}_z$ by

$$\int_{\hat{\omega}_z} v \, dx = \frac{\int_{\hat{\omega}_z} v \phi_z \, dx}{\int_{\hat{\omega}_z} \phi_z \, dx}.$$  

(6.6)

Following [11, 31], define the Clément-type interpolation operator $J : L^2(\Omega) \rightarrow U_0$ by

$$Jv = \sum_{z \in N} (\pi_z v) \phi_z(x),$$  

where the nodal value at $z$ is defined by

$$(Jv)(z) = \pi_z v = \begin{cases} f_{\hat{\omega}_z} v \, dx, & z \in N \setminus \Gamma_D, \\ 0, & z \in N \cap \Gamma_D. \end{cases}$$

**Lemma 4.3** (see [14]). For any $K \in T$ and $v \in H^1_{0,D}(\Omega)$, the estimates

$$\|v - Jv\|_{0,K} \leq C h_K k_K^{-1/2} \|k^{1/2} \nabla v\|_{0,\Delta_K}$$

(4.7)

and

$$\|\nabla(v - Jv)\|_{0,K} \leq C k_K^{-1/2} \|k^{1/2} \nabla v\|_{0,\Delta_K}$$

(4.8)
hold, where $\Delta_K$ is the union of all elements that share at least one vertex with $K$. For any $e \in E_I \cup E_N$ and $v \in H^{1,0}_D(\Omega)$, the estimate

$$
\|v - Jv\|_{0,e} \leq C h_{e}^{1/2}(W_{e,1})^{-1/2}\|k^{1/2}\nabla v\|_{0,\Delta_e}
$$

holds, where $\Delta_e$ is the union of all elements that share at least one vertex with edge $e$.

**Lemma 4.4** (see [14]). For any $v \in H^{1,0}_D(\Omega)$, there exists a positive constant $C$ independent of the mesh size and the ratio $k_{\text{max}}/k_{\text{min}}$ such that

$$
\|(f, v - Jv)\| \leq C H_f \|k^{1/2}\nabla v\|_{0,\Omega}
$$

with

$$
H_f = \left( \sum_{e \in \mathcal{N}(\partial \Omega)} \sum_{K \subset \omega_e} k^{-1}_K h_K^2 \|f\|^2_{0,K} + \sum_{e \in \mathcal{N}(\partial \Omega)} \sum_{K \subset \omega_e} k^{-1}_K h_K^2 \|f - \int_{\omega_e} f dx\|^2_{0,K} \right)^{1/2}.
$$

**Remark 4.5** (see [16]). If $f \in L^2(\Omega)$, the second term in $H_f$ is $o(\max_{K \in \mathcal{T}} h_K)$. If $f \in L^p(\Omega)$ with $p > 2$, the same holds for the first term.

5. Residual-based a posteriori error estimators. In this section, we study the following residual-based a posteriori error estimator:

$$
\eta_R = \left( \sum_{K \in \mathcal{T}} \eta^2_{R,K} \right)^{1/2},
$$

where the local indicator on $K \in \mathcal{T}$ is defined by

$$
\eta_{R,K} = \left( \eta^2_{R_f,K} + \eta^2_{J_u,K} + \eta^2_{J_v,K} + \eta^2_{R_D,K} + \eta^2_{R_N,K} \right)^{1/2}
$$

with

$$
\eta^2_{R_f,K} = \frac{h_K^2\|f\|^2_{0,K}}{k_K}, \quad \eta^2_{R_N,K} = \sum_{e \in \mathcal{E}_K \cap \mathcal{E}_N} \frac{h_e}{k_e} \|g_N - k\nabla u_T \cdot n\|^2_{0,e},
$$

$$
\eta^2_{R_D,K} = \sum_{e \in \mathcal{E}_K \cap \mathcal{E}_D} \frac{W_e}{k_e} \|g_D - u_T\|^2_{0,e},
$$

$$
\eta^2_{J_u,K} = \frac{1}{2} \sum_{e \in \mathcal{E}_K \cap \mathcal{E}_I} \frac{h_e}{k_e} \|[k \nabla u_T \cdot n]\|^2_{0,e}, \quad \text{and} \quad \eta^2_{J_v,K} = \sum_{e \in \mathcal{E}_K \cap \mathcal{E}_I} \frac{W_e}{h_e} \|[u_T]\|^2_{0,e}.
$$

The $\eta_{R_f,K}$ is the element residual, the $\eta_{R_D,K}$ and $\eta_{R_N,K}$ are the boundary residuals, and the $\eta_{J_u,K}$ and $\eta_{J_v,K}$ are associated with edge jumps of the flux and the solution, respectively. For $k = 1$, i.e., the Poisson equation, this residual-based estimator is identical to that of [27]. For $k$ being a tensor, Ern and Stephansen in [22] recently studied a residual-based error estimator, which can be made robust for scalar $k$ and under the assumption of local monotonicity. Their estimator differs from the estimator $\eta_{R,K}$ in both the flux jump term and the solution jump term. Instead of $\eta_{J_v,K}$, they...
use the so-called nonconforming error which is the energy norm of difference between
the DG approximation and its continuous recovery through the Oswald interpolation.

To analyze the estimator \( \eta_h \) and recovery-based estimators to be introduced in
section 7, we employ a standard technique that uses the Helmholtz decomposition. To this end, we cite the following decomposition (see, e.g., [3]): for any given vector-valued function \( \tau \in L^2(\Omega)^2 \), there exist \( p \in H^1_D(\Omega) \) and \( q \in H \) such that

\[
\tau = k(x) \nabla p + \nabla^\perp q,
\]

where \( H \) is a subspace of \( H^1(\Omega) \) having zero mean value and homogeneous tangential derivatives on \( \Gamma_N \):

\[
H = \left\{ v \in H^1(\Omega) : \int_{\Omega} v \, dx = 0 \text{ and } \frac{\partial v}{\partial n} = 0 \text{ on } \Gamma_N \right\}.
\]

Integrating by parts gives

\[
(\nabla p, \nabla^\perp q) = 0
\]

for all \( p \in H^1_D(\Omega) \) and all \( q \in H \), which, in turn, implies that the decomposition is orthogonal with respect to the weighted \( L^2 \) inner product \( (k^{-1}, \cdot, \cdot) \):

\[
(k^{-1} \tau, \tau) = (k \nabla p, \nabla p) + (k^{-1} \nabla^\perp q, \nabla^\perp q).
\]

Let \( e = u - u_r \); then there exist \( p \in H^1_D(\Omega) \) and \( q \in H \) such that

\[
k \nabla Ke = k \nabla p + \nabla^\perp q
\]

and

\[
\|k^{1/2} \nabla Ke\|_{0,\Omega}^2 = \|k^{1/2} \nabla p\|_{0,\Omega}^2 + \|k^{-1/2} \nabla^\perp q\|_{0,\Omega}^2.
\]

Denote the weighted oscillations of the data \( f \) over the collection \( T' \) of elements by

\[
\text{osc}(f, T')^2 = \sum_{K \in T'} \frac{h_K^2}{k_K} \|f - \bar{f}_K\|_{0,K}^2,
\]

where \( \bar{f}_K \) is the average of \( f \) over \( K \).

**Theorem 5.1** (reliability). Assume that \( u \in V^{1+\epsilon}(T) \) is the solution of problem
(2.14) and that the triangulation is quasi-uniform. Then the residual-based a posteriori
error estimator \( \eta_h \) satisfies the following global reliability bound:

\[
\|u - u_r\|_{DG} \leq C \eta_h,
\]

where \( C \) is a positive constant independent of the mesh size and the ratio \( k_{\max}/k_{\min} \).

**Proof.** It follows from (5.2), the error equation in (3.1) with \( v = Jp \), and the continuity of \( Jp \in U_0 \) that

\[
\|k^{1/2} \nabla p\|_{0,\Omega}^2 = (k \nabla e, \nabla p)
\]

\[
= (k \nabla e, \nabla (p - Jp)) + \sum_{e \in E_D} \int_e \left( \{k \nabla e \cdot n\}_w [Jp] - \theta (k \nabla Jp \cdot n)_w [\epsilon] \right.
\]

\[
- \gamma_e \frac{W_e}{h_e} [Jp] [\epsilon] \right) \, ds
\]

\[
(5.7) = (k \nabla e, \nabla (p - Jp)) - \theta \sum_{e \in E_D} \int_e \{k \nabla Jp \cdot n\}_w [\epsilon] \, ds.
\]
For any \( e \in \mathcal{E}_I \), using the Cauchy–Schwarz inequality, (2.5), the trace theorem, and the inverse inequality, we have

\[
\int_e \{k \nabla \mathcal{J} p \cdot \mathbf{n}\}_w \|e\| ds = \int_e w_+(k_+ \nabla \mathcal{J} p \cdot \mathbf{n}) \|e\| ds + \int_e w_-(k_- \nabla \mathcal{J} p \cdot \mathbf{n}) \|e\| ds
\]

\[
\leq \left( \frac{h_e^{1/2} w_+ k_+^{1/2}}{W_e^{1/2}} \|k_+^{1/2} \nabla \mathcal{J} p_+ \cdot \mathbf{n}\|_{0,e} + \frac{h_e^{1/2} w_- k_-^{1/2}}{W_e^{1/2}} \|k_-^{1/2} \nabla \mathcal{J} p_- \cdot \mathbf{n}\|_{0,e} \right) \frac{W_e^{1/2}}{h_e^{1/2}} \|[u_r]\|_{0,e}
\]

\[
\leq \left( \|k_+^{1/2} \nabla \mathcal{J} p\|_{0,K_e} + \|k_-^{1/2} \nabla \mathcal{J} p\|_{0,K_e} \right) \frac{W_e^{1/2}}{h_e^{1/2}} \|[u_r]\|_{0,e}.
\]

Thus, summing over all \( e \in \mathcal{E}_I \) and using the Cauchy–Schwarz inequality lead to

\[
(5.8) \quad \sum_{e \in \mathcal{E}_I} \int_e \{k \nabla \mathcal{J} p \cdot \mathbf{n}\}_w \|e\| ds \leq C \left( \sum_{e \in \mathcal{E}_I} \frac{W_e}{h_e} \|\|u_r\|\|_{0,e}^2 \right)^{1/2} \|k^{1/2} \nabla p\|_{0,\Omega}.
\]

Similarly, we have

\[
(5.9) \quad \sum_{e \in \mathcal{E}_D} \int_e \{k \nabla \mathcal{J} p \cdot \mathbf{n}\}_w \|e\| ds \leq C \left( \sum_{e \in \mathcal{E}_D} \frac{W_e}{h_e} \|g_D - u_r\|_{0,e}^2 \right)^{1/2} \|k^{1/2} \nabla p\|_{0,\Omega}.
\]

Denote \( e_J = p - \mathcal{J} p \). Using integration by parts, (1.1), (2.10), the facts that

\[
\nabla \cdot (k \nabla u_r) = 0 \quad \text{on} \ K \in T \quad \text{and} \ e_J \in H^1_{0,D}(\Omega),
\]

(1.2), the Cauchy–Schwarz inequality, and Lemma 4.3, we have

\[
(5.10) \quad (k \nabla h_e, \nabla (p - \mathcal{J} p)) = -\sum_{K \in T} (\nabla \cdot (k \nabla h_e), e_J) + \sum_{e \in \mathcal{E}_T} \int_{\partial K} (k \nabla e \cdot \mathbf{n}) e_J ds
\]

\[
= \sum_{K \in T} (f, e_J) - \sum_{e \in \mathcal{E}_I} \int_e \|k \nabla u_r \cdot \mathbf{n}\|_{0,e} e_J ds + \sum_{e \in \mathcal{E}_N} \int_e (g_N - k \nabla u_r \cdot \mathbf{n}) e_J ds
\]

\[
\leq \sum_{K \in T} \|f\|_{0,K} \|e_J\|_{0,K} + \sum_{e \in \mathcal{E}_I} \|k \nabla u_r \cdot \mathbf{n}\|_{0,e} \|e_J\|_{0,e}
\]

\[
+ \sum_{e \in \mathcal{E}_N} \|g_N - k \nabla u_r \cdot \mathbf{n}\|_{0,e} \|e_J\|_{0,e}
\]

\[
\leq C \left( \sum_{K \in T} \frac{h_K}{k_K^{1/2}} \|f\|_{0,K} \|k^{1/2} \nabla p\|_{0,\Delta_K} + \sum_{e \in \mathcal{E}_I} \sqrt{\frac{h_e}{W_e^{1/2}}} \|k \nabla u_r \cdot \mathbf{n}\|_{0,e} \|k^{1/2} \nabla p\|_{0,\Delta_e}
\]

\[
+ \sum_{e \in \mathcal{E}_N} \sqrt{\frac{h_e}{k_e}} \|g_N - k \nabla u_r \cdot \mathbf{n}\|_{0,e} \|k^{1/2} \nabla p\|_{0,\Delta_e} \right) \leq C \left( \sum_{e \in \mathcal{E}_I} \left( \eta_{n,e,K}^2 + \eta_{n,e,K}^2 + \eta_{n,e,K}^2 \right) \right)^{1/2} \|k^{1/2} \nabla p\|_{0,\Omega}.
\]
Combining (5.7), (5.8), (5.9), and (5.11) yields
\[ \| k^{1/2} \nabla p \|_{0, \Omega} \leq C \eta_R. \]
Together with (5.5), then it suffices to show
\[ (5.11) \quad \| k^{-1/2} \nabla^\perp q \|_{0, \Omega} \leq C \eta_R. \]
To this end, since \( u \in H^1(\Omega), Iu_\tau \in H^1(\Omega), u - Iu_\tau = 0 \) on \( \Gamma_D \), and \( q \in \mathcal{H} \), integrating by parts gives
\[ \langle \nabla (u - Iu_\tau), \nabla^\perp q \rangle = 0. \]
This, together with the Cauchy–Schwarz inequality and (4.2), implies that
\[ \| k^{-1/2} \nabla^\perp q \|_{0, \Omega} = \sum_{K \in \mathcal{T}} (\nabla^h e, \nabla^\perp q)_K = \sum_{K \in \mathcal{T}} (\nabla(Iu_\tau - u_\tau), \nabla^\perp q)_K \]
\[ \leq \sum_{K \in \mathcal{T}} \| k^{1/2} \nabla(Iu_\tau - u_\tau) \|_{0,K} \| k^{-1/2} \nabla^\perp q \|_{0,K} \]
\[ \leq \left( \sum_{K \in \mathcal{T}} \| k^{1/2} \nabla(Iu_\tau - u_\tau) \|_{0,K}^2 \right)^{1/2} \| k^{-1/2} \nabla^\perp q \|_{0,\Omega} \]
\[ \leq C \sum_{K \in \mathcal{T}} \left( \eta^2_{Iu_\tau,K} + \eta^2_{Iu_\tau,K} \right)^{1/2} \| k^{-1/2} \nabla^\perp q \|_{0,\Omega}. \]
Thus
\[ (5.12) \quad \| k^{-1/2} \nabla^\perp q \|_{0, \Omega} \leq C \sum_{K \in \mathcal{T}} \left( \eta^2_{Iu_\tau,K} + \eta^2_{Iu_\tau,K} \right)^{1/2}, \]
which proves the validity of (5.11) and, hence, the theorem. \( \square \)

Note that (5.4) indicates that the true error \( e = u - u_\tau \) comes from two kinds of sources: discontinuous approximations of the normal component of the flux and the tangential component of the gradient. As shown in the following lemma, the element residual, the Neumann boundary residual, and the edge jump of the flux may be bounded by the energy norm of \( p \) plus higher-order terms.

**Lemma 5.2.** There exists a positive constant \( C \) independent of the mesh size and the ratio \( k_{\text{max}}/k_{\text{min}} \) such that
\[ (5.13) \quad \eta_{u_\tau,K} \leq C \left( \| k^{1/2} \nabla p \|_{0,K} + \text{osc}(f,K) \right) \quad \forall K \in \mathcal{T}, \]
\[ (5.14) \quad \sqrt{\frac{h_e}{W_{e,1}}} \| k \nabla u_\tau \cdot n \|_{0,e} \leq C \left( \| k^{1/2} \nabla p \|_{0,\omega_e} + \text{osc}(f,\omega_e) \right) \quad \forall e \in \mathcal{E}_1, \]
\[ (5.15) \quad \sqrt{\frac{h_e}{k_e}} \| g_N - k \nabla u_\tau \|_{0,e} \leq C \left( \| k^{1/2} \nabla p \|_{0,\omega_e} + \text{osc}(f,\omega_e) \right) \quad \forall e \in \mathcal{E}_N, \]

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
where $\omega_e$ is the collection of all elements that share the common edge $e$.

**Proof.** For any element $K \in T$, let $b_K$ be the standard cubic bubble function whose support is $K$. Then (see, e.g., [37])

(5.16) \[ \frac{C h_K^{1/2}}{k} \leq \|b_K^{1/2}\|_{0,K}, \quad \|b_K\|_{\infty,K} \leq C, \quad \text{and} \quad \|\nabla b_K\|_{\infty,K} \leq C \frac{h_K^{-1}}{k}. \]

Using (5.4) and (5.2), we have

(5.17) \[ (k\nabla p, \nabla v) = (k\nabla h_e, \nabla v) \quad \forall \; v \in H^1_{0,B}(\Omega). \]

Choosing $v = \tilde{f}_K b_K$ in (5.17) and integrating by parts lead to

\[ (k\nabla p, \nabla (\tilde{f}_K b_K))_K = (k\nabla h_e, \nabla (\tilde{f}_K b_K))_K = (f - \tilde{f}_K, \tilde{f}_K b_K)_K + (\tilde{f}_K, \tilde{f}_K b_K)_K. \]

It follows from the Cauchy–Schwarz inequality and (5.16) that

\[ C\|\tilde{f}_K\|^2_{0,K} \leq (\tilde{f}_K, \tilde{f}_K b_K)_K = (k\nabla p, \nabla (\tilde{f}_K b_K))_K - (f - \tilde{f}_K, \tilde{f}_K b_K)_K \leq \|k^{1/2}\nabla p\|_{0,K} \|k^{1/2}\tilde{f}_K \nabla b_K\|_{0,K} + \|f - \tilde{f}_K\|_{0,K} \|\tilde{f}_K b_K\|_{0,K} \leq C \left( \frac{k^{1/2}}{h_K}\|k^{1/2}\nabla p\|_{0,K} + \frac{h_K}{k^{1/2}}\|f - \tilde{f}_K\|_{0,K} \right). \]

Hence,

\[ \frac{h_K}{k^{1/2}}\|\tilde{f}_K\|_{0,K} \leq C \left( \|k^{1/2}\nabla p\|_{0,K} + \frac{h_K}{k^{1/2}}\|f - \tilde{f}_K\|_{0,K} \right). \]

Now, (5.13) is a direct consequence of the triangle inequality.

For any edge $e \in E$, let $b_e$ be the standard piecewise quadratic edge bubble function corresponding to the edge $e$ whose support is $\omega_e$. Then (see, e.g., [37])

(5.18) \[ C h_e^{1/2} \leq \|b_e^{1/2}\|_{0,e}, \quad \|b_e\|_{\infty,\omega_e} \leq C, \quad \text{and} \quad \|\nabla b_e\|_{\infty,\omega_e} \leq C h_e^{-1}. \]

Choosing $v = \|k\nabla u_T \cdot n\|_{0,e} b_e$ in (5.17) and integrating by parts lead to

\[ (k\nabla p, \nabla ((k\nabla u_T \cdot n)_{e} b_e))_{\omega_e} = (k\nabla h_e, \nabla ((k\nabla u_T \cdot n)_{e} b_e))_{\omega_e} = (f, (k\nabla u_T \cdot n)_{e} b_e)_{\omega_e} + \|k\nabla u_T \cdot n\|_{0,e}^2. \]

It follows from (5.18), the Cauchy–Schwarz inequality, and (2.6) that

\[ C\|k\nabla u_T \cdot n\|^2_{0,e} \leq \|k\nabla u_T \cdot n\|_{0,e}^2 b_e, \leq \|k\nabla p\|_{0,\omega_e} \|k\nabla u_T \cdot n\|_{0,\omega_e} \|\nabla b_e\|_{0,\omega_e} + \|f\|_{0,\omega_e} \|k\nabla u_T \cdot n\|_{0,\omega_e} \|b_e\|_{0,\omega_e} \leq C h_e^{-1/2} \|k\nabla u_T \cdot n\|_{0,e} \sum_{K \in \omega_e} k_K^{1/2} \left( \|k^{1/2}\nabla p\|_{0,K} + h_K k_K^{-1/2} \|f\|_{0,K} \right), \]

which, together with (2.6) and the triangle inequality, implies that

\[ \sqrt{\frac{h_e}{W_{e,1}}} \|k\nabla u_T \cdot n\|_{0,e} \leq C \left( \|k^{1/2}\nabla p\|_{0,\omega_e} + \sum_{K \in \omega_e} \frac{h_K(\|\tilde{f}_K\|_{0,K} + \|f - \tilde{f}_K\|_{0,K})}{k_K^{1/2}} \right). \]
Now, (5.14) follows from (5.13). For $e \in E_N$, (5.15) may be proved in a similar fashion by choosing $v = (g_N - k \nabla u_T \cdot n) b_e$. This completes the proof of the lemma.

**Theorem 5.3 (efficiency).** Assume that the diffusion coefficient is locally quasi-monotone. Then there exists a positive constant $C$ independent of the mesh size and the ratio $k_{\max}/k_{\min}$ such that

$$\eta_{h,K} \leq C (\|e\|_{DG,K} + \text{osc}(f, \omega_K)) \quad \forall K \in T,$$

where $\omega_K$ is the collection of elements in $T$ that share a common edge with $K$.

**Proof.** The local efficiency bound is a straightforward consequence of Lemma 5.2 and the definition of the DG norm.

6. **Flux recovery.** The flux defined by

$$\sigma = -k(x) \nabla u \quad \text{in } \Omega$$

is an important physical quantity which is often the primary concern in practice. In this section, we describe both implicit and explicit recoveries of the flux that are used to design robust a posteriori error estimators in the subsequent section. In addition, we show that the implicitly recovered flux is a good approximation to the flux.

6.1. **Implicit approximation.** The implicitly recovered flux is defined as follows: find $\sigma_T \in V_N$ such that

$$(k^{-1} \sigma_T, \tau) = - (\nabla h u_T, \tau) \quad \forall \tau \in V_N,$$

where $V_N$ is the subspace of $V (RT_0$ or $BDM_1)$ satisfying the Neumann boundary conditions

$$V_N = \{ \tau \in V : \tau \cdot n = g_N \text{ on } \Gamma_N \}.$$

**Theorem 6.1.** There exists a positive constant $C$ independent of the ratio $k_{\max}/k_{\min}$ such that the a priori error bound

$$(6.3) \quad \|k^{-1/2}(\sigma - \sigma_T)\|_{0,\Omega} \leq C \left( \inf_{\tau \in V_N} \|k^{-1/2}(\sigma - \tau)\|_{0,\Omega} + \|k^{1/2}(\nabla u - \nabla h u_T)\|_{0,\Omega} \right)$$

holds.

**Proof.** By using the error equation

$$(k^{-1}(\sigma - \sigma_T), \tau) = (\nabla h u_T - \nabla u, \tau) \quad \forall \tau \in V_N,$$

(6.3) may be proved in a fashion similar to that of Theorem 3.1 in [14].

6.2. **Explicit approximations.** Let $\delta_{e,e'}$ denote the Kronecker delta:

$$\delta_{e,e'} = \begin{cases} 1 & \text{if } e = e', \\ 0 & \text{if } e \neq e'. \end{cases}$$

For $RT_0$, its nodal basis function, $\phi_e$, corresponding to the edge $e \in E$ is uniquely determined by

$$\phi_e \cdot n_{e'} = \delta_{e,e'} \quad \forall e' \in E.$$
Define the explicit approximation \( \hat{\sigma}_T \) in \( RT_0 = \text{span}\{\phi_e : e \in E\} \) by

\[
\hat{\sigma}_T = \sum_{e \in E} \hat{\sigma}_e \phi_e,
\]

where \( \hat{\sigma}_e \) is the normal component of \( \hat{\sigma}_T \) on the edge \( e \in E \) defined as a weighted average of the normal components of the approximated flux \(-k \nabla_h u_T\); i.e.,

\[
\hat{\sigma}_e := \{-k \nabla_h u_T \cdot n_e\}_{w_i}^e
\]

with weights \( w_i \) defined in (2.3). To ensure the efficiency bound independent of the size of jumps, we choose \( i = 2 \) or \( i = 3 \), i.e., the harmonic or the geometric weights. Note that these weights satisfy the following inequality:

\[
\max\left\{\left(\frac{(w_{+i})^2}{k_-^e}, \frac{(w_{-i})^2}{k_+^e}\right), \frac{1}{k_-^e + k_+^e}\right\} \leq \frac{1}{k_-^e + k_+^e}, \quad i = 2, 3.
\]

### 7. Recovery-based a posteriori error estimators.

For any element \( K \in \mathcal{T} \), based on the implicitly and explicitly recovered fluxes, define

\[
\eta_{\sigma,K} = \|k^{-1/2} \sigma_T + k^{1/2} \nabla_h u_T\|_{0,K} \quad \text{and} \quad \hat{\eta}_{\sigma,K} = \|k^{-1/2} \hat{\sigma}_T + k^{1/2} \nabla_h u_T\|_{0,K}.
\]

Obviously,

\[
\eta_{\sigma,K} \leq \hat{\eta}_{\sigma,K}.
\]

Let

\[
\eta_{\sigma}^2 = \sum_{K \in \mathcal{T}} \eta_{\sigma,K}^2, \quad \eta_{J_u}^2 = \sum_{K \in \mathcal{T}} \eta_{J_u,K}^2, \quad \text{and} \quad \eta_{R_D}^2 = \sum_{K \in \mathcal{T}} \eta_{R_D,K}^2.
\]

It is easy to see that

\[
\eta_{\sigma} = \|k^{-1/2} \sigma_T + k^{1/2} \nabla_h u_T\|_{0,\Omega} = \min_{\tau \in \mathcal{V}_N} \|k^{-1/2} \tau + k^{1/2} \nabla_h u_T\|_{0,\Omega}.
\]

Now, we define the recovery-based a posteriori error estimators as follows:

\[
\xi = \left(\eta_{\sigma}^2 + \eta_{J_u}^2 + \eta_{R_D}^2\right)^{1/2} \quad \text{and} \quad \hat{\xi} = \left(\hat{\eta}_{\sigma}^2 + \hat{\eta}_{J_u}^2 + \hat{\eta}_{R_D}^2\right)^{1/2}.
\]

**Theorem 7.1** (reliability). There exists a positive constant \( C \) independent of the mesh size and the ratio \( k_{\max}/k_{\min} \) such that

\[
\|\epsilon\|_{DG} \leq C \left(\xi + H_f\right) \leq C \left(\hat{\xi} + H_f\right).
\]

**Proof.** The second inequality in (7.2) follows directly from (7.1). To establish the validity of the first inequality in (7.2), by (5.5), (5.12), and the definition of the DG norm, it suffices to show that

\[
\|k^{1/2} \nabla p\|_{0,\Omega} \leq C \left(\xi + H_f\right).
\]

Letting \( \epsilon_J = p - J p \), where \( J \) is the Clément-type interpolation operator defined in section 4, it then follows from integration by parts, homogeneous boundary conditions...
of \( e_J \) on \( \Gamma_D \) and \((k\nabla u + \sigma_T) \cdot n \) on \( \Gamma_N \), the Cauchy–Schwarz inequality, (4.8), the fact that \( \nabla_h \cdot (k\nabla_h u_T) = 0 \), the inverse inequality, (4.7), and (4.10) that
\[
(k\nabla_h e, \nabla e_J) = (k\nabla u + \sigma_T, \nabla e_J) - (\sigma_T + k\nabla_h u_T, \nabla e_J)
\leq (f, e_J) + C \eta_p \|k^{1/2}\nabla p\|_{0,\Omega}
= (f, e_J) - (\nabla_h \cdot (\sigma_T + k\nabla_h u_T), e_J) + C \eta_p \|k^{1/2}\nabla p\|_{0,\Omega}
\leq C (\eta_p + H_f) \|k^{1/2}\nabla p\|_{0,\Omega}.
\]
Combining with (5.7), (5.8), and (5.9) yields the first inequality in (7.3). This completes the proof of the theorem. \( \Box \)

Remark 7.2. A different explicit recovery-based estimator is derived in [21]. Its flux recovery achieves a tighter local connection between \( \nabla \cdot \sigma \) and \( f \). Its reliability bound similar to (7.2) is established with \( C = 1 \) and \( H_f \) being replaced by a superconvergent term if \( f \in H^1(\Omega) \).

Lemma 7.3. For any element \( K \in \mathcal{T} \), let \( \omega_K \) be the union of elements sharing a common edge with \( K \). There exists a positive constant \( C \) independent of the mesh size and the ratio \( k_{\max}/k_{\min} \) such that
\[
\hat{\eta}_{e,K} \leq C \left( \|k^{1/2}\nabla p\|_{0,\omega_K}^2 + \text{osc}(f, \omega_K)^2 \right)^{1/2} \quad \forall K \in \mathcal{T}.
\]

Proof. To show the validity of (7.4), by (5.14) and (5.15), it suffices to prove that for any element \( K \in \mathcal{T} \),
\[
\hat{\eta}_{e,K}^2 \leq C \left( \sum_{e \in e_K \cap \mathcal{E}_I} \frac{h_e}{W_{e,1}} \|k\nabla u_T \cdot n_e\|^2_{0,e} \right) + \sum_{e \in e_K \cap \mathcal{E}_N} \frac{h_e}{k} \|g_n - k\nabla u_T \cdot n_e\|^2_{0,e}.
\]
We provide the proof of (7.5) only in the case that \( \mathcal{E}_K \cap \mathcal{E}_N = \emptyset \) because it can be proved in a similar fashion in the case that \( \mathcal{E}_K \cap \mathcal{E}_N \neq \emptyset \). To this end, for any edge \( e \in \mathcal{E}_K \), without loss of generality, let \( n_e \) be the outward unit vector normal to \( \partial K \), and denote by \( K_e \) the adjacent element with the common edge \( e \). Since \( \tau = -k\nabla_h u_T \) is piecewise constant, \( \tau|_K \) may be represented in terms of the nodal basis function of \( R T_0 \), \( \{\phi_e\}_{e \in \partial K} \), as follows:
\[
\tau|_K = \sum_{e \in e_K} \tau_{e,K} \phi_e.
\]
For any \( x \in K \), (6.4) and (6.5) give
\[
\hat{\sigma}_T - \tau = \sum_{e \in \mathcal{E}_K} (\hat{\sigma}_e - \tau_{e,K}) \phi_e = \sum_{e \in \mathcal{E}_K} (w_{+,e} - 1) (\tau_{e,K} - \tau_{e,K}) \phi_e
= -\sum_{e \in \mathcal{E}_K} w_{-,e} [\tau \cdot n_e] \phi_e.
\]
Now, it follows from the triangle inequality, the fact that \( \int_K |\phi_e|^2 \, dx \leq C \), and (6.6) that
\[
\hat{\eta}_{e,K}^2 = \|k^{-1/2} (\hat{\sigma}_T - \tau) \|^2_{0,K} \leq C k_{K}^{-1} \|\hat{\sigma}_T - \tau\|^2_{0,K}
\leq C \sum_{e \in \mathcal{E}_K} k_{K}^{-1} w_{+,e} h_e \int_e [\tau \cdot n_e]^2 \, ds \leq C \left( \sum_{e \in \mathcal{E}_K} \frac{h_e}{W_{e,1}} \|k\nabla u_T \cdot n_e\|^2_{0,e} \right),
\]

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
which proves (7.5) and, hence, (7.4). This completes the proof of the lemma.

**Theorem 7.4 (efficiency).** Assume that the diffusion coefficient is locally quasi-
monotone. Then there exists a positive constant $C$ independent of the mesh size and
the ratio $k_{\text{max}}/k_{\text{min}}$ such that

$$
\xi \leq \hat{\xi} \leq C \left( \|e\|_{\text{DG}} + \text{osc}(f, T) \right).
$$

**Proof.** Inequality (7.6) is a direct consequence of (7.1), (7.4), and the definition
of the DG norm.

8. Numerical experiment. In this section, we report some numerical results
for an interface problem with intersecting interfaces used by many authors, which
is considered as a benchmark test problem. For this test problem, we numerically
illustrate the discretization error of the DG method and demonstrate the robustness
of our error estimators.

To this end, let $\Omega = (-1,1)^2$ and

$$
u(r, \theta) = r^\beta \mu(\theta)
$$
in the polar coordinates at the origin with $\mu(\theta)$ being a smooth function of $\theta$ (see,
e.g., [14]). The function $u(r, \theta)$ satisfies the interface equation with $A = kI$, $\Gamma_N = \emptyset$,
$f = 0$, and

$$
k(x) = \begin{cases} 
R & \text{in } (0,1)^2 \cup (-1,0)^2, \\
1 & \text{in } \Omega \setminus ([0,1]^2 \cup [-1,0]^2).
\end{cases}
$$

Note that the solution $u(r, \theta)$ is only in $H^{1+\beta-\epsilon}(\Omega)$ for any $\epsilon > 0$ and, hence, it is very
singular for small $\beta$ at the origin. This suggests that refinement is centered around
the origin. The $\beta$ depends on the size of the jump. For the test problem, we choose
$\beta = 0.1$ which is corresponding to $R \approx 161$.

For simplicity of presentation, the harmonic weighted incomplete interior penalty
Galerkin (IIPG) method is used. Let $u_T \in \mathcal{U}^{\text{DG}}$ be the discontinuous finite element
approximation of the solution. Denote by $N$ the number of unknowns. We start
with the coarsest triangulation $T_0$ obtained from halving 16 congruent squares by
connecting the bottom left and upper right corners.

Numerical results on uniform meshes are reported in Figure 1. The a priori error
estimate in (3.4) is of order $h^{0.1}$. This indicates that the slope of the log(dof)-log(error)
should be $-0.05$. Figure 1 shows that the asymptotic convergence rate for this test
problem is slightly better than the theoretical prediction.

Starting with the coarse triangulation $T_0$, a sequence of meshes is generated by
using a standard adaptive meshing algorithm that adopts the maximum marking
strategy: (1) mark those elements such that $\eta_K \geq 0.5 \max_{K' \in T} \eta_{K'}$ and (2) refine the
marked triangles by bisection. The stopping criterion

$$
\text{rel-err} := \frac{\|u - u_T\|_{\text{DG}}}{\|k^{1/2} \nabla u\|_{0, \Omega}} \leq \text{tol}
$$
is used, and numerical results with $\text{tol} = 0.1$ are reported.

Meshes generated by $\eta_\xi$ and $\xi$ are depicted in Figures 2 and 4, respectively.
Refinements are centered at origin as expected for efficient estimators. Meshes generated
by various a posteriori error estimators for various types of discretizations of
this test problem can be found in [14, 15]. As shown in Figures 3 and 5, the slopes of the log(dof)-log(relative error) for the estimators are close to $-1/2$ when there are enough grid points (about several hundreds of unknowns). This implies the optimal decay of the error with respect to the number of unknowns. While the effectivity index, $\text{eff-index} := \frac{\eta}{|u-u_T|_{DG}}$, of $\eta_R$ is about 4, that of $\xi$ is close to 1. This means that the estimator $\xi$ is very accurate.
REFERENCES


Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.


