An a posteriori error estimate for the generalized finite element method for transient heat diffusion problems
Iqbal, Muhammad; Gimperlein, Heiko; Mohamed, M. Shadi; Laghrouche, Omar

Published in:
International Journal for Numerical Methods in Engineering

DOI:
10.1002/nme.5440

Publication date:
2017

Document Version
Peer reviewed version

Link to publication in Heriot-Watt University Research Portal

Citation for published version (APA):
An a posteriori error estimate for the generalized finite element method for transient heat diffusion problems

Muhammad Iqbal* Heiko Gimperlein†‡ M. Shadi Mohamed* Omar Laghrouche*

Abstract

We propose the study of a posteriori error estimates for time–dependent generalized finite element simulations of heat transfer problems. A residual estimate is shown to provide reliable and practically useful upper bounds for the numerical errors, independent of the heuristically chosen enrichment functions. Two sets of numerical experiments are presented. First, the error estimate is shown to capture the decrease in the error as the number of enrichment functions is increased or the time discretization refined. Second, the estimate is used to predict the behaviour of the error where no exact solution is available. It also reflects the errors incurred in the poorly conditioned systems typically encountered in generalised finite element methods. Finally we study local error indicators in individual time steps and elements of the mesh. This creates a basis towards the adaptive selection and refinement of the enrichment functions.

1 Introduction

Time–dependent finite element methods provide an accurate and stable approximation to the solution of time–dependent diffusion equations in engineering applications. For complex geometries or complex source/sink terms unstructured meshes can be highly advantageous on the basis of their ability to provide local mesh refinement near important diffusion features and structures. However, classical finite elements based on linear or quadratic polynomial ansatz functions on such meshes still require considerable computational efforts. This is particularly the case in the presence of steep fronts or boundary layers in the solution, which need to be resolved accurately in applications [17, 39, 40]. Applications which suffer from these complications include heat transfer in the presence of thermal radiation [9], optical tomography [18] or the cooling down of molten glass [46, 48].

In analogy to the partition of unity finite element methods (PUFEM) for steady–state elliptic equations [24], the generalized finite element methods (GFEM) whose approximation space is enriched by non–polynomial enrichment functions have been proposed to overcome these difficulties also for transient problems [14, 26–30]. As it is based on a variational formulation of the underlying PDE, GFEM inherits the stability and accuracy of finite elements, but allows to adapt the trial and test functions to reflect a priori information about the physical properties of the considered problem. Similar to the time–independent case, the method shows a reduction in computation time and memory requirements for complex engineering problems, given reasonable enriched approximation spaces.

Enriched finite element methods have found applications in a wide variety of contexts, based on enrichment functions tailored to the problem at hand. The PUFEM has been thoroughly investigated...
for time-harmonic wave problems from acoustics or elasticity [8, 19, 20] and has been adapted to boundary element methods [21, 34], ultraweak formulations [4, 15, 16] and discontinuous Galerkin methods [44, 45, 50]. It has also been presented in the context of the GFEM [6, 41, 42]. A prominent example in a different direction, fracture mechanics, is given by the extended finite element method. Here the enrichment functions reproduce the singular displacements or stresses locally around a crack tip [5].

The current work explores rigorous computable (a posteriori) error estimates for the approximate solutions provided by the time-dependent GFEM. In particular, we address enrichments in the whole spatial domain and for time-dependent problems. While suitable enrichment functions are easily proposed, such estimates provide rigorous insight into the accuracy of the resulting numerical scheme and the relative merits of different numbers and kinds of enrichment functions, as well as possible mesh refinements.

Similar estimates in time-independent or time-harmonic settings go back to the very beginning of enriched finite element methods [25], with later works especially by Strouboulis, Babuska and coauthors [41, 43]. More recently a variety of posteriori error estimates for the extended finite element method has been explored, i.e. for enrichments localised in a small part of the mesh, with a particular emphasis on crack problems. We mention representative works of Bordas and Duflot [2, 3, 7], Prange, Loehnert and Wriggers [23, 35], Pannachet, Sluys and coauthors [32, 33], Barros and coauthors [1], resp. Rodenas, Estrada and coauthors [36–38]. Specifically, the role of statically admissible recovery has been addressed in [12] and [49].

Here we explore the relevance of a residual a posteriori estimate for time-dependent simulations of heat transfer problems, in the context of [26]. While for standard $h$-method finite elements such residual error estimates have a long tradition and are known to give sharp upper bounds for the error as $h \to 0$ [47], for GFEM one is particularly interested in a fixed coarse mesh. Also, unlike for piecewise polynomial ansatz functions or the plane-wave enrichment for wave problems, little is known about the approximation provided by the heuristically chosen enrichment functions. Nevertheless, our simple error estimates are rigorously shown to provide reliable and practically useful upper bounds for the numerical error. Numerical experiments indicate that they efficiently capture the decrease of the error as the number of enrichment functions is increased or the time discretization refined. Both the global error in the whole space-time domain and local error indicators in the individual time steps and elements of the mesh are studied, with a view towards the adaptive selection and refinements of the enrichment functions.

For applications the error in particular observables may sometimes be of more interest than the error of the solution. Goal-oriented a posteriori error estimates are currently being studied for the extended finite element methods [10, 11, 13, 22, 31]; they are clearly relevant also in the context of this article and will be pursued in future work.

As for the structure of this article: Section 2 states the precise formulation of the time-dependent diffusion problem which we consider. It recalls the weak form of the equation and its numerical approximation by the generalized finite element method. The residual a posteriori error estimate is formulated and proven in Section 3. Section 4 contains some algorithmic considerations, before numerical experiments study the error estimate in model problems in Section 5. Some concluding remarks are the content of Section 6.

## 2 Weak formulation and numerical approximation

Given an open bounded domain $\Omega \subset \mathbb{R}^2$ with polygonal boundary $\Gamma$ and a given time interval $[0, T]$, we consider the diffusion equation

$$\frac{\partial u}{\partial t} - \lambda \Delta u = f, \quad \text{in } [0, T] \times \Omega,$$

(1)
where $\lambda > 0$ is the diffusion coefficient and the right hand side $f$ represents the effects of internal sources/sinks. An initial condition
\[ u(t = 0, x) = U_0(x), \quad x \in \Omega , \] (2)
and a Robin–type boundary condition
\[ \frac{\partial u}{\partial n} + hu = g, \quad \text{in } [0, T] \times \Gamma , \] (3)
are imposed. Here $n$ denotes the outward unit normal on the boundary $\Gamma$, and $h \geq 0$ is the heat convection coefficient on $\Gamma$, and $g$ represents boundary sources.

We will numerically solve a weak formulation of the initial–boundary problem (1) - (3): Multiplying by a smooth test function $W$, then integrating over $\Omega$ we obtain
\[ \int_\Omega \left( W \frac{\partial u}{\partial t} - \lambda W \Delta u \right) \, d\Omega = \int_\Omega W f \, d\Omega . \] (4)
The divergence theorem allows to write this as
\[ \int_\Omega \left( W \frac{\partial u}{\partial t} + \lambda \nabla W \nabla u \right) \, d\Omega - \int_\Gamma \lambda W \nabla u \cdot n \, d\Gamma = \int_\Omega W f \, d\Omega , \] (5)
and with the boundary condition (3) we conclude the weak formulation of the heat diffusion problem: Find a solution $u$ on $[0, T] \times \Omega$ such that $u(0, x) = U_0(x)$ and for all test functions $W$ on $\Omega$ and all $t \in [0, T]$:
\[ \int_\Omega \left( W \frac{\partial u}{\partial t} + \lambda \nabla W \nabla u \right) \, d\Omega + \int_\Gamma \lambda W_{\,hu} \, d\Gamma = \int_\Omega W f \, d\Omega + \int_\Gamma \lambda W_{\,g} \, d\Gamma . \] (6)

This article will be concerned with the numerical approximation of (6). As time–discretization, we choose an implicit Euler method: The time interval is divided into $N$ subintervals $[t_n, t_{n+1}]$ of size $\Delta t = t_{n+1} - t_n$ for $n = 0, 1, \ldots, N$ and approximate the time derivative in (6) by a difference quotient:
\[ \int_\Omega \left( W \frac{u^{n+1} - u^n}{\Delta t} + \lambda \nabla W \nabla u^{n+1} \right) \, d\Omega + \int_\Gamma \lambda W_{\,hu} \, d\Gamma = \int_\Omega W F^{n+1} \, d\Omega + \int_\Gamma \lambda W_{\,g} \, d\Gamma . \] (7)
Rearranging the terms we conclude the time–discretized variant of the weak formulation (6): Find a solution $u^{n+1}$ on $\Omega$ such that $u^0 = u_0$ and for all test functions $W$ on $\Omega$ and all $n \in \mathbb{N}$:
\[ \int_\Omega (\nabla W \cdot \nabla u^{n+1} + W_{\,ku} u^{n+1}) \, d\Omega + \int_\Gamma W_{\,hu} u^{n+1} \, d\Gamma = \int_\Omega W F^{n+1} \, d\Omega + \int_\Gamma W_{\,g}^{n+1} \, d\Gamma , \] (8)
where $u_0$ is a discretization of $U_0$, and $F^{n+1}$ and $k$ are defined as
\[ F^{n+1} = k \left( \delta t f(t_{n+1}, x) + u^n \right), \quad k = \frac{1}{\lambda \Delta t} . \]

Our aim is to find an approximate solution $u^{n+1}$ of the weak form (8) using the generalized finite element method. To do so, we assume that $\Omega$ is a polygon and fix a regular mesh $\Omega = \cup K$. The edges of the elements $K$ are denoted by $E$. We look for $u^{n+1}$ of the form
\[ u^{n+1}(x) = \sum_{j=1}^{\mathcal{M}} \sum_{q=1}^{\mathcal{Q}} A^q_j N_j(x) G_q(x) . \] (9)
Here $A^q_j \in \mathbb{R},$ and $N_j$ are the piecewise polynomial shape functions associated to the mesh. $Q$ denotes the number of enrichment functions, and $M$ the number of nodes in the mesh. As in [26], we choose the global enrichment functions $G_q$ to be
\[ G_q(x) = e^{-\left( \frac{r^2}{\delta^2} \right)^q} - e^{-\left( \frac{r^2}{\delta^2} \right)^q}, \quad q = 1, 2, \ldots, Q . \] (10)
From the numerical solution \(u_3\) a residual a posteriori error estimate can become a more relevant choice. For developing the proposed residual error estimate in Theorem 1 below. For specific applications, their choice can be optimized by accurately assessing the error; this is a main motivation. The choice of enrichment functions is mainly motivated by the physical behaviour of the solution.

The choice of enrichment functions is mainly motivated by the physical behaviour of the solution. As a main point, the estimate does not depend on the choice of our enrichment functions – little can be said about their approximation properties in general, and this will allow to choose the enrichment adaptively. Also, for GFEM we are concerned with a fixed, coarse mesh.

\[
\int_\Omega (\nabla P_r \cdot \nabla u^{n+1} + P_r h u^{n+1}) d\Omega + \int_\Gamma P_r h u^{n+1} d\Gamma = \int_\Omega P_r F^{n+1} d\Omega + \int_\Gamma P_r g^{n+1} d\Gamma. \tag{11}
\]

The functions \(P_r\) are written in terms of global coordinates but modulated locally as they also include the local shape functions \(N_j\). On the other hand the global nature of the enrichment functions makes them highly efficient in modelling the behaviour of the solution in time as well as in space. The choice of enrichment functions is mainly motivated by the physical behaviour of the solution. However, their choice can be optimized by accurately assessing the error; this is a main motivation for developing the proposed residual error estimate in Theorem 1 below. For specific applications goal oriented error estimates can become a more relevant choice.

3 A residual a posteriori error estimate

From the numerical solution \(u^{n+1}\) at time \(t_{n+1}\) we define, by piecewise constant resp. piecewise linear interpolation, numerical solutions for all positive \(t\):

\[
u(t, x) = \frac{t - t_n}{t_{n+1} - t_n} u^{n+1}(x) + \frac{t_{n+1} - t}{t_{n+1} - t_n} u^n(x),
\]

\[
\hat{u}(t, x) = u(t_{n+1}, x), \quad \hat{f}(t, x) = f(t_{n+1}, x), \quad \hat{g}(t, x) = g(t_{n+1}, x) \quad \text{for} \quad t \in [t_n, t_{n+1}].
\]

Note from (7) that the GFEM discretization (11) is equivalent to finding \(u^{n+1}\) of the form (9) such that \(u^0 = u_0\) and for all \(r = 1, \ldots, MQ\):

\[
\int_\Omega (P_r \partial_t u + \lambda \nabla P_r \nabla \hat{u}) d\Omega + \int_\Gamma \lambda P_r h \hat{u} d\Gamma = \int_\Omega P_r \hat{f} d\Omega + \int_\Gamma \lambda P_r \hat{g} d\Gamma. \tag{12}
\]

In this notation, we obtain a classical a posteriori estimate of residual type for the error of the GFEM solution, similar to extensively used estimates for adaptive \(h\)- and \(hp\)-finite element methods. It is given by computable error indicators \(\eta_1, \ldots, \eta_6\), and the following theorem shows that it is reliable, in the sense that \(\eta_1, \ldots, \eta_6\) rigorously bound the error.

As a main point, the estimate does not depend on the choice of our enrichment functions – little can be said about their approximation properties in general, and this will allow to choose the enrichment adaptively. Also, for GFEM we are concerned with a fixed, coarse mesh.

**Theorem 1.** Let \(U\) be the solution of the exact weak formulation (6) and \(u\) the solution of the GFEM discretization (11). Then there exists a constant \(c > 0\) such that:

\[
\int_\Omega |U(T, x) - u(T, x)|^2 d\Omega + \lambda \int_0^T \int_\Omega |\nabla (U - \hat{u})|^2 d\Omega dt \leq c\{\eta_1^2 + \eta_2^2 + \eta_3^2 + \eta_4^2 + \eta_5^2 + \eta_6^2\}, \tag{13}
\]
where

\[
\eta_1^2 = \|U_0 - u_0\|_{L^2(\Omega)}^2 \quad (14)
\]

\[
\eta_2^2 = \sum_k \int_0^T \left\| \dot{f} - \partial_t u + \lambda \Delta \hat{u} \right\|_{H^{-1}(K)}^2 \, dt \quad (15)
\]

\[
\eta_3^2 = \int_0^T \left\| f - \hat{f} \right\|_{H^{-1}(\Omega)}^2 \, dt + \lambda \int_0^T \left\| \hat{g} - g \right\|_{H^{-1/2}(\Gamma)}^2 \, dt \quad (16)
\]

\[
\eta_4^2 = \lambda \int_0^T \| \nabla (u - \hat{u}) \|_{L^2(\Omega)}^2 \, dt \quad (17)
\]

\[
\eta_5^2 = \sum_{E: \Gamma \cap E \neq \emptyset} \int_0^T \left\| \left[ \frac{\partial \hat{u}}{\partial n} \right] \left( L^2(E) \right) \right\|^2 \, dt \quad (18)
\]

\[
\eta_6^2 = \lambda \int_0^T \| \hat{g} - \frac{\partial \hat{u}}{\partial n} - h \hat{u} \|_{H^{-1/2}(\Gamma)}^2 \, dt .
\]

The left hand side of expression (13) measures the (typically unknown) size of the actual error between the GFEM and exact solutions. The error indicators on the right hand side, \( \eta_1^2 \) to \( \eta_6^2 \), however can be computed. The theorem proves that the error is at most \( c \{ \eta_1^2 + \cdots + \eta_6^2 \} \), in particular the \( \eta_j \) never underestimate the error. According to the proof, the constant \( c \) only depends on the geometry, \( \lambda \), and the boundary condition (3). More precisely, it can be obtained from the lowest eigenvalue of \( \lambda \Delta \) in \( \Omega \) with boundary condition (3) and the optimal constant \( C \) in the trace theorem \( \| V \|_{H^{1/2}(\Gamma)} \leq C \| V \|_{H^1(\Omega)} \). In the numerical examples below \( c \sim 10^{-1} \).

For the \( h \)-method one can often show that a residual error estimate is efficient, in the sense that it does not overestimate the error by more than a fixed multiplicative constant \( [47] \). Proving such a result would require a detailed analysis of the particular enrichment functions; it is less relevant for key applications such as adaptive refinements, where Theorem 1 is crucial. The numerical experiments in Section 5 will investigate the relation between the error indicators and the actual error numerically.

The error indicators \( \eta_1^2 \) to \( \eta_6^2 \) in Theorem 1 have clear physical meanings: \( \eta_1 \) and \( \eta_3 \) describe the error in the approximation of the initial condition, resp. the source term. We will usually consider situations in which the exact initial condition and source are used in the computation, so that \( \eta_1 \) and \( \eta_3 \) vanish. The violation of the original PDE (1) is measured by \( \eta_2 \), the violation of the boundary condition (3) by \( \eta_5 \). Finally, \( \eta_6 \) measures that the numerical heat flux is not conserved across element edges, while \( \eta_4 \) involves the error in the time discretization.

**Proof of Theorem 1.** Let \( \hat{e} = U - u, \pi \) some stable projection onto the space of basis functions and \( \hat{e}_I = \pi \hat{e} \). We start by applying the Fundamental Theorem of Calculus to the first term in expression (13):

\[
\int_\Omega |U(T, x) - u(T, x)|^2 \, d\Omega = \int_\Omega |U(0, x) - u(0, x)|^2 \, d\Omega \\
+ 2 \int_0^T \int_\Omega (\partial_t U(t, x) - \partial_t u(t, x))(U(t, x) - u(t, x)) \, d\Omega \, dt \\
= \eta_1^2 + 2 \int_0^T \int_\Omega (\partial_t U(t, x) - \partial_t u(t, x))\hat{e}(t, x) \, d\Omega \, dt .
\]
Now adding and subtracting the same quantities

\[
\int_{0}^{T} \int_{\Omega} (\partial_t U - \partial_t u) \dot{e} \, d\Omega \, dt + \lambda \int_{0}^{T} \int_{\Omega} |\nabla (U - \dot{u})|^2 \, d\Omega \, dt
\]

\[
= \int_{0}^{T} \int_{\Omega} (\partial_t U - \partial_t u)(\dot{e} - \dot{e}_I) \, d\Omega \, dt + \int_{0}^{T} \int_{\Omega} (\partial_t U - \partial_t u) \dot{e}_I \, d\Omega \, dt
\]

\[
+ \lambda \int_{0}^{T} \int_{\Omega} \nabla (U - \dot{u}) \nabla (\dot{e} - \dot{e}_I) \, d\Omega \, dt + \lambda \int_{0}^{T} \int_{\Omega} \nabla (U - \dot{u}) \nabla \dot{e}_I \, d\Omega \, dt
\]

\[
+ \lambda \int_{0}^{T} \int_{\Omega} \nabla (U - \dot{u}) \nabla (u - \dot{u}) \, d\Omega \, dt .
\]

Using the exact weak formulation (6), this equals

\[
- \int_{0}^{T} \int_{\Omega} \partial_t u (\dot{e} - \dot{e}_I) \, d\Omega \, dt - \int_{0}^{T} \int_{\Omega} \partial_t u \dot{e}_I \, d\Omega \, dt
\]

\[
- \lambda \int_{0}^{T} \int_{\Omega} \nabla \dot{u} \nabla (\dot{e} - \dot{e}_I) \, d\Omega \, dt - \lambda \int_{0}^{T} \int_{\Omega} \nabla \dot{u} \nabla \dot{e}_I \, d\Omega \, dt
\]

\[
+ \lambda \int_{0}^{T} \int_{\Omega} \nabla (U - \dot{u}) \nabla (u - \dot{u}) \, d\Omega \, dt
\]

\[
+ \int_{0}^{T} \int_{\Omega} f(\dot{e} - \dot{e}_I) \, d\Omega \, dt + \int_{0}^{T} \int_{\Gamma} \lambda g(\dot{e} - \dot{e}_I) \, d\Gamma \, dt - \int_{0}^{T} \int_{\Gamma} \lambda h (\dot{u} - \dot{e}_I) \, d\Gamma \, dt
\]

\[
+ \int_{0}^{T} \int_{\Omega} f\dot{e}_I \, d\Omega \, dt + \int_{0}^{T} \int_{\Gamma} \lambda g\dot{e}_I \, d\Gamma \, dt - \int_{0}^{T} \int_{\Gamma} \lambda h \dot{e}_I \, d\Gamma \, dt .
\]

Further, using the GFEM equation (12), the sum of the second and fourth terms becomes

\[
\int_{0}^{T} \int_{\Gamma} \lambda h \dot{u} \dot{e}_I \, d\Gamma \, dt - \int_{0}^{T} \int_{\Omega} \dot{f} \dot{e}_I \, d\Omega \, dt - \int_{0}^{T} \int_{\Gamma} \lambda g \dot{e}_I \, d\Gamma \, dt .
\]

On each of the elements \( K \) of the mesh we integrate by parts in the third term,

\[
- \lambda \int_{0}^{T} \int_{\Omega} \nabla \dot{u} \nabla (\dot{e} - \dot{e}_I) \, d\Omega \, dt = \lambda \sum_{K} \int_{0}^{T} \int_{K} \Delta \dot{u} (\dot{e} - \dot{e}_I) \, d\Omega \, dt
\]

\[
- \lambda \sum_{K} \int_{0}^{T} \int_{\partial K} \frac{\partial \dot{u}}{\partial n} (\dot{e} - \dot{e}_I) \, d(\partial K) \, dt .
\]

The second sum over \( K \) can be written as a sum over interior and boundary edges,

\[
- \lambda \sum_{K} \int_{0}^{T} \int_{\partial K} \frac{\partial \dot{u}}{\partial n} (\dot{e} - \dot{e}_I) \, d(\partial K) \, dt
\]

\[
= - \lambda \sum_{E \cap \Gamma = 0} \int_{0}^{T} \int_{E} \left[ \frac{\partial \dot{u}}{\partial n} \right] (\dot{e} - \dot{e}_I) \, dE \, dt - \lambda \int_{0}^{T} \int_{\Gamma} \frac{\partial \dot{u}}{\partial n} (\dot{e} - \dot{e}_I) \, d\Gamma \, dt .
\]
We conclude

\[
\int_{0}^{T} \int_{\Omega} (\partial_t U - \partial_t u) \hat{\epsilon} \, d\Omega \, dt + \lambda \int_{0}^{T} \int_{\Omega} |\nabla (U - \hat{u})|^2 \, d\Omega \, dt
\]
\[= \sum_{K} \int_{0}^{T} \int_{K} (\hat{f} - \partial_t u + \lambda \Delta \hat{u}) \, d\Omega \, dt \]
\[- \lambda \sum_{E \cap I = \emptyset} \int_{0}^{T} \int_{E} \left[ \frac{\partial \hat{u}}{\partial n} \right] (\hat{\epsilon} - \hat{e}_I) \, dE \, dt + \lambda \int_{0}^{T} \int_{\Omega} \nabla (U - \hat{u}) \nabla (u - \hat{u}) \, d\Omega \, dt \]
\[+ \int_{0}^{T} \int_{\Gamma} \lambda \left( \hat{g} - \frac{\partial \hat{u}}{\partial n} - h\hat{u} \right) (\hat{\epsilon} - \hat{e}_I) \, d\Gamma \, dt - \int_{0}^{T} \int_{\Gamma} \lambda h (U - \hat{u}) (\hat{\epsilon} - \hat{e}_I) \, d\Gamma \, dt \]
\[+ \int_{0}^{T} \int_{\Omega} (f - \hat{f}) \hat{\epsilon} \, d\Omega \, dt + \int_{0}^{T} \int_{\Gamma} \lambda (g - \hat{g}) \hat{\epsilon} \, d\Gamma \, dt - \int_{0}^{T} \int_{\Gamma} \lambda h (U - \hat{u}) \hat{e}_I \, d\Gamma \, dt . \]  

(19)

Note that

\[- \int_{0}^{T} \int_{\Gamma} \lambda h (U - \hat{u}) (\hat{\epsilon} - \hat{e}_I) \, d\Gamma \, dt - \int_{0}^{T} \int_{\Gamma} \lambda h (U - \hat{u}) \hat{e}_I \, d\Gamma \, dt \]
\[= \int_{0}^{T} \int_{\Gamma} \lambda h (-\hat{\epsilon} + \hat{u} - u) \hat{\epsilon} \, d\Gamma \, dt . \]

As we do not have further information about the approximation properties of the basis functions, we do not lose much by choosing \( \epsilon_I = 0 \). With the Cauchy–Schwarz inequality \( ab \leq \frac{a^2}{2} + \frac{b^2}{2} \) for any \( a, b \in \mathbb{R}, \epsilon > 0 \), the third term on the right hand side of (19) is smaller than \( \frac{\epsilon \lambda}{2} \int_{0}^{T} \int_{\Omega} |\nabla (U - \hat{u})|^2 \, d\Omega \, dt + \frac{1}{2} \eta_4^2 \), and for small \( \epsilon > 0 \) we may move the first of these summands to the left hand side.

The continuity of the bilinear pairing between \( H^1 \) and \( H^1_0 \) and the Sobolev inequality similarly result in \( \eta_2^2 \) from the first term resp. \( \eta_3^2 \) from the term which involves \( f - \hat{f} \). Finally, the trace theorem \( \| V \|_{H^{1/2}(\Gamma)} \leq C \| V \|_{H^1(\Omega)} \) in the same manner leads to \( \eta_5^2 \) from the second term on the right hand side, to \( \eta_6^2 \) from the fourth term, and \( \eta_3^2 \) from the penultimate term involving \( g - \hat{g} \).

For the remaining terms, the trace theorem and the Sobolev inequality allow to estimate them by \( \eta_4^2 \).

\[ \square \]

4 Algorithmic considerations

This section discusses the detailed implementation of the error indicator in Theorem 1.

We compute \( \eta_2^2 \) as \( \eta_2^2 = \sum_{n=0}^{N_t} \sum_{K} \eta_2^2(n, K) \), with

\[
\eta_2^2(n, K) = \int_{t_n}^{t_{n+1}} \| \hat{f} - \partial_t u + \lambda \Delta \hat{u} \|_{H^{-1}(K)}^2 \, dt \]
\[\leq \int_{t_n}^{t_{n+1}} dt \int_{K} (\hat{f} - \partial_t u + \lambda \Delta \hat{u})^2 \, d\Omega \]
\[= \delta t \int_{K} \left( u^{n+1} - \frac{u^{n+1} - u^n}{\delta t} + \lambda \left( \frac{\partial^2 u^{n+1}}{\partial x^2} + \frac{\partial^2 u^{n+1}}{\partial y^2} \right) \right)^2 \, d\Omega . \]

Here, the final line is efficiently computable and only a slight overestimate of \( \eta_2 \). The values of \( f \) and \( u^n \) are calculated at each integration point and then accumulated over the whole domain. They are updated at every time step. Values from the present and previous time step are used in the calculation of \( \eta_2^2(n, K) \).

The indicator \( \eta_4^2 = \sum_{n=0}^{N_t} \sum_{K} \eta_4^2(n, K) \) calculates the change in the derivative of \( u \) in every time step. We calculate the derivative of the temperature as above at times \( t_n \) and \( t_{n+1} \) at every
integration point and then accumulate over the whole domain:

$$
\eta_3^2(n, K) = \lambda \int_{t_n}^{t_{n+1}} \| \nabla (u - \hat{u}) \|^2_{L^2(K)} dt
$$

$$
= \lambda \int_{t_n}^{t_{n+1}} \left( \frac{t_{n+1} - t}{t_{n+1} - t_n} \right)^2 dt \int_K \left[ \left( \frac{\partial u^{n+1}}{\partial x} - \frac{\partial \hat{u}}{\partial x} \right)^2 + \left( \frac{\partial u^{n+1}}{\partial y} - \frac{\partial \hat{u}}{\partial y} \right)^2 \right] d\Omega
$$

$$
= \frac{\lambda \delta T}{3} \left( \int_K \left( \frac{\partial u^{n+1}}{\partial x} - \frac{\partial u^n}{\partial x} \right)^2 d\Omega + \int_K \left( \frac{\partial u^{n+1}}{\partial y} - \frac{\partial u^n}{\partial y} \right)^2 d\Omega \right).
$$

Here we use that, by definition,

$$
u(t, x) - \hat{u}(t, x) = - \frac{t_{n+1} - t}{t_{n+1} - t_n} \left( u^{n+1}(x) - u^n(x) \right).$$

For $$\eta_5^2 = \sum_{n=0}^{N_t} \sum_E \eta_5^2(n, E)$$ error indicator calculates the jump of $$\frac{\partial \hat{u}}{\partial n}$$ across the interior edges $$E$$ of the mesh. At each integration point these values are calculated for adjacent edges of the elements. The difference between these values at the adjacent elements is then calculated and integrated over the whole domain. This means,

$$
\eta_5^2(n, E) = \int_{t_n}^{t_{n+1}} \left\| \frac{\partial \hat{u}}{\partial n} \right\|^2_{L^2(E)} dt
$$

$$
= \int_{t_n}^{t_{n+1}} \left\| \nabla u^{n+1} \cdot n \right\|^2_{L^2(E)} dt
$$

$$
= \int_{t_n}^{t_{n+1}} \int_E \left( \nabla u^{n+1}_{E_1} n_1 + \nabla u^{n+1}_{E_2} n_2 \right)^2 dE dt
$$

$$
= \int_{t_n}^{t_{n+1}} \int_E \left( \nabla u^{n+1}_{E_1} n_1 - \nabla u^{n+1}_{E_2} n_2 \right)^2 dE dt
$$

$$
= \delta \int_E \left( \left( \frac{\partial u^{n+1}}{\partial x} n_{1x} + \frac{\partial u^{n+1}}{\partial y} n_{1y} \right)_{E_1} - \left( \frac{\partial u^{n+1}}{\partial x} n_{1x} + \frac{\partial u^{n+1}}{\partial y} n_{1y} \right)_{E_2} \right)^2 dE,
$$

where $$E_1$$ refers to the boundary value of the function on the edge taken from the first element and $$E_2$$ refers to the boundary value of the function on the same edge, but taken from the second element. We denote by $$n_1 = (n_{1x}, n_{1y})$$ and $$n_2 = (n_{2x}, n_{2y})$$ the unit normals for elements 1 and 2, respectively.

## 5 Numerical experiments

This section investigates the relevance and sharpness of the a posteriori error estimate in Theorem 1 in numerical experiments. We find numerical approximations to the heat transfer problem (1) - (3) by computing the solution to the GFEM discretization (9).

For the experiments we choose a quadrilateral mesh with piecewise bilinear shape functions. The parameters in the GFEM basis functions $$G_q$$ are taken to be $$R_c = \sqrt{\frac{2800}{239}}$$ and $$C = \frac{200}{239}$$, and $$R_0 = |x - x_c|$$ is the distance from the point $$x_c = (1, 1)$$. All integrals over $$\Omega$$ are evaluated numerically, using a Gauss–Legendre quadrature with 22 integration points. In the numerical examples we consider, the error estimators $$\eta_1$$ and $$\eta_3$$, which arise from the approximation of the initial condition and the source terms, are either identically 0 or negligible. The same holds for the violation of the boundary condition, $$\eta_6$$. Both the numerical and exact solutions themselves are either 0 or very close to 0 at the boundary.

To compare the results for different model problems, we focus on the relative error between the exact solution $$U$$ and its GFEM approximation $$\hat{u}$$, defined as the square root of

$$
\frac{\int_{\Omega} |U - \hat{u}|^2 d\Omega + \lambda \int_0^T \int_{\Omega} \| \nabla (U - \hat{u}) \|^2 d\Omega dt}{\int_{\Omega} |U|^2 d\Omega + \lambda \int_0^T \int_{\Omega} \| \nabla U \|^2 d\Omega dt}.
$$

(20)
The corresponding relative error indicator is given as the square root of

$$\frac{\eta_2^2 + \eta_3^2 + \eta_5^2}{\int_{\Omega} |U|^2 d\Omega + \lambda \int_0^T \int_{\Omega} |\nabla U|^2 d\Omega \ dt}.$$  \hfill (21)

If the exact solution $U$ is not known, we replace it by an FEM approximation on a fine mesh.

### 5.1 Example Problem 1

In this example we test the error indicator by comparing it to the computed error for a problem with a given exact solution that is proposed in [27]. We consider a square domain $\Omega = \{(x, y) \in \mathbb{R}^2 : 0 \leq x, y \leq 2\}$ with heat diffusion coefficient $\lambda = 0.1 kg m/s^2$ and convection heat transfer coefficient $h = 1 kg/\degree Cs^2$. The data $f$, $g$ and $U_0$ are chosen such that

$$U(t, x) = x^{20} y^{20} (2-x)^{20} (2-y)^{20} (1-e^{-\lambda t})$$  \hfill (22)

is an exact solution to (1) - (3).

For a fixed coarse spatial mesh of 25 elements we vary the number of enrichment functions $Q = 2, \ldots, 6$. We compute both the actual error of the GFEM solution and the error indicators and
Figure 2: Comparison of relative error (■) and relative error indicator (▲) for δt = 0.05, 0.5 and 1s at t = 5s (a), t = 10s (b) and t = 100s (c).

Figure 3: Domain configuration for Example Problem 2 with a heat source in the centre.

compare these values as in Theorem 1. In Figure 1, we show the relative error (20) of the GFEM solution and the relative error indicator (21), for different numbers Q of enrichment functions at times t = 0.05, 0.1 and 1s. In each of the cases we do so for different time steps of size δt = 0.01, 0.001 and 0.0001s. In the graphs logarithmic scale is used.

In all cases the actual error of the GFEM solution and the error indicator show a similar decrease as we increase the number of enrichment functions. The ratio of the estimator and the error is close to 10, consistently in all cases. This corresponds to a constant c ~ 10^{-1} in Theorem 1. In this sense the error indicator efficiently captures the behaviour of the real error.

Figure 2 investigates the influence of the time step δt when the spatial discretization error is small. We use a uniform mesh with 100 elements, Q = 6, and compare the relative error and error indicator for δt = 0.05, 0.5 and 1s. The results are plotted at t = 5, 10 and 100s. At 5s the error decreases significantly for smaller time steps. As time progresses the error becomes insensitive to the time step which can be seen at 100s. This is due to the large temperature differences between the central part and rest of the domain. As time passes, heat propagates from the central part to other parts, hence the temperature gradient is reduced in size. Figure 2 shows remarkable consistency between the behaviour of the error indicator and the actual error. The behaviour described above is accurately recovered by the error indicator. For a residual a posteriori estimate, the multiplicative constant between the indicator and the actual error is essentially the constant c.

5.2 Example Problem 2

The second example considers a square domain with a square heat source as shown in Figure 3. The source is switched on from t = 0s to t = 0.02s, and then it is switched off. The source diffuses heat at two rates. For x, y ∈ [0.8,1.2], the central part of the source, f is constant and given by f = 200°C/s. Outside this part f decreases linearly to f = 0 on the external boundaries of the
source, where either $x$ or $y$ is one of $\in \{0.4, 1.6\}$. In this example the thermal conductivity is taken as $\lambda = 0.1 kg/m^2Cs^2$ and the convective heat transfer coefficient as $h = 1 kg/\circ Cs^2$. The initial temperature $U_0$ and the boundary sources $g$ are both chosen to be 0. This example tests the error estimates for a more realistic heat problem and makes a first step towards an adaptive selection of enrichment functions. Similar sharp gradients would also appear as boundary layers.

Here we only compute the error indicator, because the exact solution is not known. The magnitude of the error indicator is considered relative to a reference value defined by Equation (20), where the exact solution $U$ is replaced with a reference FEM solution on a fine uniform mesh of 12800 triangular elements with piecewise linear basis functions. Both the FEM and the GFEM meshes are shown in Figure 4.

For the GFEM solution we fix a coarse mesh of 25 elements and vary the number of enrichment functions $Q = 2, \ldots, 6$. Figure 5 compares the temperature distribution of the FEM solution to the GFEM solution. Figure 6 shows the relative error indicators as a function of the number $Q$ of enrichment functions at times $t = 0.01, 0.05, 0.1, 0.15$ and $0.2s$. In each of the cases different time steps are considered: $\delta t = 0.001, 0.0001$ and $0.00001s$. The results show a decrease in the error indicator as we increase the number of enrichment functions. For all the values of $\delta t$ and at early times the results show similar trends (as in Figure 6(a)). However, at later times and for the smallest considered time step $\delta t = 0.00001$ the error indicator with higher $Q$ starts to increase rapidly. This can be seen clearly for $Q = 5$ and $6$ in Figure 6(d) and (e). For example in the case with $Q = 6$ the error indicator increases from less than 0.3 in Figure 6(c) to more than 0.6 in Figure 6(e).

The rapid increase can be attributed to two facts. First, at higher number of enrichment function the conditioning of the system matrix from Equation (11) becomes much worse as can be seen in Table 1. The size of the condition number is mostly determined by $Q$, with negligible dependence on $\delta t$. For example the condition number increases from about $1.5E+5$ with $Q = 2$ and $\delta t = 0.001s$ to about $2.0E+13$ with $Q = 6$ and $\delta t = 0.00001s$. The increase in the condition number causes an increase in the error. This behaviour seems to be reflected accurately by the error indicator as can be seen in Figure 6. Second, smaller values of $\delta t$ require more time steps to cover the same time span. This accumulates the computational errors, in particular floating point errors which increase with higher condition numbers. To verify this, in Figure 5 we compare the temperature distribution obtained with the reference solution to the GFEM solution with 4 and 6 enrichment functions with

<table>
<thead>
<tr>
<th>$Q$</th>
<th>$0.001$</th>
<th>$0.0001$</th>
<th>$0.00001$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.461E+05</td>
<td>1.509E+05</td>
<td>1.514E+05</td>
</tr>
<tr>
<td>3</td>
<td>2.534E+07</td>
<td>2.727E+07</td>
<td>2.748E+07</td>
</tr>
<tr>
<td>4</td>
<td>4.302E+09</td>
<td>4.805E+09</td>
<td>4.862E+09</td>
</tr>
<tr>
<td>5</td>
<td>5.118E+11</td>
<td>6.270E+11</td>
<td>6.416E+11</td>
</tr>
<tr>
<td>6</td>
<td>1.637E+13</td>
<td>1.992E+13</td>
<td>2.037E+13</td>
</tr>
</tbody>
</table>
Figure 5: Temperature distribution for the reference solution (FEM) (left column), GFEM solution with $Q = 4$ (middle column) and GFEM solution with $Q = 6$ (right column) at $t = 0.05s$ (top row), $t = 0.1s$ (middle row) and $t = 0.2s$ (bottom row). The distribution is obtained with $\delta t = 0.00001s$.

Table 2: Details of the $q$-bands.

<table>
<thead>
<tr>
<th>Case</th>
<th>Band</th>
<th>Case</th>
<th>Band</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$q = 1, 2, 3$</td>
<td>11</td>
<td>$q = 2, 3, 6$</td>
</tr>
<tr>
<td>2</td>
<td>$q = 2, 3, 4$</td>
<td>12</td>
<td>$q = 2, 3, 5$</td>
</tr>
<tr>
<td>3</td>
<td>$q = 3, 4, 5$</td>
<td>13</td>
<td>$q = 3, 5, 6$</td>
</tr>
<tr>
<td>4</td>
<td>$q = 4, 5, 6$</td>
<td>14</td>
<td>$q = 3, 4, 6$</td>
</tr>
<tr>
<td>5</td>
<td>$q = 1, 2, 4$</td>
<td>15</td>
<td>$q = 1, 3, 4$</td>
</tr>
<tr>
<td>6</td>
<td>$q = 1, 2, 5$</td>
<td>16</td>
<td>$q = 1, 3, 5$</td>
</tr>
<tr>
<td>7</td>
<td>$q = 1, 2, 6$</td>
<td>17</td>
<td>$q = 1, 3, 6$</td>
</tr>
<tr>
<td>8</td>
<td>$q = 2, 5, 6$</td>
<td>18</td>
<td>$q = 1, 4, 5$</td>
</tr>
<tr>
<td>9</td>
<td>$q = 2, 4, 6$</td>
<td>19</td>
<td>$q = 1, 4, 6$</td>
</tr>
<tr>
<td>10</td>
<td>$q = 2, 4, 5$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6: Relative error indicator for $\delta t = 0.01s$, $\delta t = 0.001s$ and $\delta t = 0.0001s$ at $t = 0.01s$ (a), $t = 0.05s$ (b), $t = 0.1s$ (c), $t = 0.15s$ (d) and $t = 0.2s$ (e).

Figure 7: Relative error indicator for different $q$-bands with $\delta t = 0.001s$ at $t = 0.01s$ (a), $t = 0.05s$ (b), $t = 0.1s$ (c), $t = 0.15s$ (d) and $t = 0.2s$ (e).
\[ \delta t = 0.00001. \] The solution at the early time steps shown in the figure, namely \( t = 0.05s \) and \( 0.01s \), seems very similar in all the three cases. However, at later times \( (t = 0.2s) \) the distribution is still similar for FEM and GFEM with 4 enrichment functions, but it deteriorates with 6 enrichment functions.

Two final experiments investigate the possible use of a posteriori error estimates as in Theorem 1 for adaptive selection or refinement procedures, when little is known about the approximation properties of the enrichment functions.

Figure 7 calculates the relative time–integrated error indicator over the whole domain \( \Omega \) for different choices of enrichment functions and different time intervals \( t = 0.01, 0.05, 0.1, 0.15 \) and \( 0.2s \), with fixed \( \delta t = 0.001s \). For the enrichments stated in Table 2, we observe that the three enrichment functions \( q = 1, 2, 3 \) (Case 1) consistently give rise to the lowest error indicator. Especially for long times, enrichment with steep functions such as \( q = 4, 5, 6 \) (Case 4) is seen to yield worse numerical approximations. More generally, in each instance the enrichment function with \( q = 1 \) leads to a substantial improvement of the error indicator.

In order to decide about locally refined enrichments, the spatial distribution of the error indicators proves relevant. In Figure 8 we depict the time–integrated error indicators for each element of the mesh for the optimal choice of enrichment functions \( q = 1, 2, 3 \) (Case 1) from Figure 7. Short \( (t = 0.05s) \), intermediate \( (t = 0.1s) \) and longer time intervals \( (t = 0.2s) \) are considered, with \( \delta t = 0.001s \). The contribution from the interior edges, \( \eta_5 \), is assigned in equal parts to the adjacent elements. It should be noted that the enrichment functions have a radial symmetry, which is not present in the exact solution for short times. Correspondingly, for short times the error indicator exhibits large contributions along both diagonals of the square domain. It also concentrates in the hot central element of the mesh, while the sharp gradient seems to be captured by the enrichment functions for short and intermediate times. For larger times, \( t = 0.2s \), both the solution and the error indicators include contributions near the boundary, but we also observe increasing contributions from the large gradients.

### 6 Conclusions

We have studied an a posteriori error estimate for time–dependent generalized finite element simulations of heat transfer problems. The easily implemented residual estimate we propose does not depend on the choice of enrichment functions and is seen to efficiently and reliably reflect the behaviour of the numerical error for the coarse meshes of an enriched method. It also reflects the errors incurred in the poorly conditioned systems typically encountered in generalised finite element methods. The estimate thereby allows to adaptively choose a suitable global enrichment.

We have investigated the estimate both for problems with a known exact solution, for which we compare the indicator with the known error, and in cases where the exact solution is not known.
The results indicate the potential of the a posteriori error estimates to be used also for adaptive local choices of the number and type of enrichment functions. A rigorous exploration of adaptivity and applications to both strong boundary layers as in [28] and solutions with singularities will be pursued in future work. Furthermore, goal oriented error estimates for GFEM solutions to heat transfer problems would also be relevant; developing such estimates becomes an obvious choice when local instead of global enrichment is used as in [13, 31].

Our results provide the basis for space–time adaptive generalized finite element methods for the heat equation. Fast methods will hinge on a better understanding of the adaptive refinement procedures and on the development of more efficiently computable indicators, as in [7]. Also error indicators without unknown constants will be of interest.

Acknowledgements

H. G. is supported by ERC Advanced Grant HARG 268105.

References


