Contemporary Mathematics

An Adaptive Algorithm for Ordinary, Stochastic and Partial Differential Equations

Kyoung-Sook Moon, Erik von Schwerin, Anders Szepessy, and Raúl Tempone

ABSTRACT. The theory of a posteriori error estimates suitable for adaptive refinement is well established. This work focuses on the fundamental, but less studied, issue of convergence rates of adaptive algorithms. In particular, this work describes a simple and general adaptive algorithm applied to ordinary, stochastic and partial differential equations with proven convergence rates. The presentation has three parts: The error approximations used to build error indicators for the adaptive algorithm are based on error expansions with computable leading order terms. It is explained how to measure optimal convergence rates for approximation of functionals of the solution, and why convergence of the error density is always useful and subtle in the case of stochastic and partial differential equations. The adaptive algorithm, performing successive mesh refinements, either reduces the maximal error indicator by a factor or stops with the error asymptotically bounded by the prescribed accuracy requirement. Furthermore, the algorithm stops using the optimal number of degrees of freedom, up to a problem independent factor.

1. Introduction to the Adaptive Algorithm

This work presents an overview of the authors work on the convergence rate of an adaptive algorithm to compute functionals of solutions to ordinary, stochastic and partial differential equations. The main ingredient of the adaptive algorithm is an error expansion of the form

$$\text{Global error} = \sum \text{local error} \cdot \text{weight} + \text{higher order error}, \tag{1.1}$$

with computable leading order terms. The weight is the sensitivity of the functional of the solution with respect to perturbations in the differential equation. For an ordinary differential equation, the error expansion (1.1) can be derived by the variational principle in [23] and for weak approximation of stochastic differential

1991 *Mathematics Subject Classification.* 65L50,65C30, 65N50.

Key words and phrases. adaptive methods, mesh refinement algorithm, a posteriori error estimate, computational complexity.

Support from the Swedish Research Council grants 2002-6285 and 2002-4961, UdelaR in Uruguay, a Swedish Foundation for Strategic Research grant and the European network HYKE, funded by the EC as contract HPRN-CT-2002-00282, is acknowledged.

equations the error expansion (1.1) can be derived based on computable stochastic flows and discrete dual backward problems in [1]. For partial differential equations, [24] derives an asymptotic expansion of the error using the dual weighted residual method.

The goal of the adaptive algorithm for differential equations based on these error expansions, is to approximate the desired quantity of interest with an adapted mesh using as few mesh elements as possible, for a given level of accuracy using a given approximation method, for instance the Euler method or piecewise bilinear finite elements with varying mesh size. Based on the a posteriori error expansions of the form (1.1) the global error can be asymptotically approximated by

$$\text{Global error} \approx \sum_{\mathcal{K}} \text{error indicator}, \tag{1.2}$$

where $\mathcal{K}$ is a set of time steps or elements. A typical adaptive algorithm does two things iteratively :

- **(i)** if the error indicators satisfy an accuracy condition, then it stops; otherwise
- **(ii)** the algorithm chooses where to refine the mesh, recomputes the error indicators and then makes an iterative step to **(i)**.

Therefore the indicators not only estimate the localization of the global error but also give information on how to refine the mesh in order to achieve optimal efficiency.

Despite the established use of adaptive algorithms and the well developed theory of a posteriori error estimates, e.g. [2–4, 6–8, 16–20, 26] the theoretical study of adaptive mesh refinement algorithms is more recent, cf. [5,9,12,13,21,22,24,25,27]. To introduce the main ingredients, let us consider a simple integration problem, namely for a given function $X : [0,T] \to \mathbb{R}$ approximate the integral $g(X) = \int_0^T X(t)dt$. Let us first discretize the time interval $[0,T]$ into N subintervals $0 = t_0 < t_1 < \cdots < t_N = T$ with corresponding time steps $h_n := t_{n+1} - t_n$. Now, if we approximate $g(X)$ using the left point rule (forward Euler) denoted by $\bar{g}$, our global discretization error becomes

$$\text{Global Error } = g(X) - \bar{g} = \sum_{n=0}^{N-1} (h_n)^2 \rho_n + \text{ higher order terms}, \tag{1.3}$$

with the error density function ρ defined by $\rho_n := \frac{dX}{dt}(t_n)/2$. From the definition of the number of time steps

$$N(h) := \int_0^T \frac{1}{h(\tau)} d\tau, \tag{1.4}$$

the number N_u of uniform steps to reach a given level of accuracy TOL turns out to be asymptotically proportional to the L^1-norm of the function ρ, i.e.

$$N_u \simeq \frac{T}{\text{TOL}} \|\rho\|_{L^1(0,T)}.$$

On the other hand, by minimizing the number of steps in (1.4) subject to an accuracy constraint, i.e. imposing the leading order of (1.3) to be TOL, a standard Lagrangian relaxation yields the optimal choice

$$h_n^2 \, \rho_n = \text{ constant for all} \, n$$

and as a consequence the number N_a of adaptive steps becomes proportional to the smaller $L^{\frac{1}{2}}$ quasi-norm of the error density, i.e.

$$N_a \simeq \frac{1}{\mathrm{TOL}} \|\rho\|_{L^{\frac{1}{2}}(0,T)}.$$

For instance, if we have a singular case, $X(t) = 1/\sqrt{t}$, we can instead compute with $X_\epsilon(t) = 1/\sqrt{t+\epsilon}$, choosing the positive parameter ϵ such that

$$\left| \int_0^T (X(t) - X_\epsilon(t))\, dt \right| = o(\mathrm{TOL})$$

i.e. $\epsilon^{1/2} \lesssim o(\mathrm{TOL})$. Therefore, the number of uniform time steps becomes

$$N_u \simeq \frac{T/4}{\mathrm{TOL}} \int_0^T \frac{dt}{(t+\epsilon)^{3/2}} \simeq \frac{T/4}{\mathrm{TOL}} \frac{1}{\epsilon^{1/2}} \gtrsim \mathcal{O}(\mathrm{TOL}^{-2})$$

while the number of adaptive time steps is the smaller

$$N_a \simeq \frac{1/4}{\mathrm{TOL}} \left(\int_0^T \frac{dt}{(t+\epsilon)^{3/4}} \right)^2 \approx \mathcal{O}(\mathrm{TOL}^{-1})$$

which clearly shows the advantage of an adaptive approach. In the sequel we will consider multiscale problems, i.e. problems that have very different scales so that the error density, although being bounded, may be very large, e.g. $\frac{1}{\epsilon^{3/2}}$.

Thus, having motivated the need for adaptivity in the previous example, we now state the main questions to answer, namely

What is the notion of error density for ordinary, stochastic and partial differential equations?

What is a suitable approximation for such an error density?

What can be concluded about the convergence rate of the adaptive algorithm?

In this paper, Section 2 describes an adaptive algorithm based on previously derived a posteriori error expansions for ODEs [23], SDEs [21], and PDEs [24]. Then, Section 3 presents results on the convergence rates of the adaptive algorithm and Section 4 illustrates the behavior of the adaptive algorithm using a numerical example.

2. Convergence of the Error Density and the Adaptive Algorithm

2.1. An Error Expansion for ODEs. Let us consider an ordinary differential equation (ODE) of the form

$$\frac{dX(t)}{dt} = a(t, X(t)),\ 0 < t \le T, \tag{2.1}$$

with an initial value $X(0) = X_0 \in \mathbb{R}^d$ and a given flux $a : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$. First discretize the time interval $[0,T]$ into N subintervals $0 = t_0 < t_1 < \ldots < t_N = T$ and let $\overline{X}$ be an approximation of X in (2.1) by any p-th order accurate numerical method, satisfying the same initial condition $\overline{X}(0) = X(0) = X_0$.

We are interested in computing a function value $g(X(T))$ for a given general function $g : \mathbb{R}^d \to \mathbb{R}$, which may represent the quantities of physical interest. Using the variational principle, we show in [23] that the global error has the expansion

$$g(X(T)) - g(\overline{X}(T)) = \sum_{n=1}^{N} \left(\bar{e}(t_n), \bar{\Psi}(t_n) \right) + \int_0^T o(h^p(t))dt, \tag{2.2}$$

where $(\cdot,\cdot)$ is the standard scalar product on $\mathbb{R}^d$. Here the approximate local error is defined by $\bar{e}(t_n) := \gamma(\bar{\overline{X}}(t_n) - \overline{X}(t_n))$, where γ is a Richardson extrapolation constant and the approximate local exact solution $\bar{\overline{X}}$ is computed with smaller time steps or a higher order method than $\overline{X}$. The weight $\bar{\Psi}$ is an approximation of Ψ, which solves the dual equation

$$
\begin{aligned}
-\frac{d\Psi(s)}{ds} &= (a')^*(s,X(s))\Psi(s),\ s<T,\\
\Psi(T) &= g'(X(T)),
\end{aligned}
\tag{2.3}
$$

where $(a')^*(s,x)$ is the transpose of the Jacobian matrix.

Therefore the leading order term in (2.2) has the approximate error density

$$\bar{\rho}_n := \frac{\big(\bar{e}(t_n), \bar{\Psi}(t_n)\big)}{h_n^{p+1}},$$

which is then used in the adaptive algorithm, see Section 2.4.

2.2. An Error Expansion for SDEs. Let us consider an Itô Stochastic differential equation (SDE) of the form

$$dX_k(t) = a_k(t,X(t))dt + \sum_{\ell=1}^{\ell_0} b_k^\ell(t,X(t))dW^\ell(t),\ t>0, \tag{2.4}$$

where $k=1,\ldots,d$ and $(X(t;\omega))$ is a stochastic process in $\mathbb{R}^d$, with randomness generated by the independent one dimensional Wiener processes $W^\ell(t;\omega)$, $\ell=1,\ldots,\ell_0$, on the probability space $(\Omega,\mathcal{F},P)$. The functions $a(t,x)\in\mathbb{R}^d$ and $b^\ell(t,x)\in\mathbb{R}^d$, $\ell=1,\ldots,\ell_0$, are given drift and diffusion fluxes.

The goal is to construct approximations to the expected value $E[g(X(T))]$ by a Monte Carlo method, for a given function $g:\mathbb{R}^d\to\mathbb{R}$. Examples of such an expected value are the computation of option prices in mathematical finance, stochastic climate prediction, wave propagation in random media, etc. The Monte Carlo Euler method approximates the unknown process X by the Euler method $\overline{X}(t_n)$ which is a time discretization based on the nodes $0=t_0<t_1<\cdots<t_N=T$ where

$$\overline{X}(t_{n+1}) - \overline{X}(t_n) = h_n a(t_n,\overline{X}(t_n)) + \sum_{\ell=1}^{\ell_0} \Delta W_n^\ell b^\ell(t_n,\overline{X}(t_n)), \tag{2.5}$$

and $h_n \equiv t_{n+1}-t_n$, $\Delta W_n^\ell \equiv W^\ell(t_{n+1}) - W^\ell(t_n)$, $n=0,1,2,\ldots,N-1$. The aim of the adaptive algorithm is to choose the size of the time steps, h_n, and the number of independent identically distributed samples $\overline{X}(\cdot,\omega_j)$, $j=1,2,\ldots,M$, such that the computational work, $N\cdot M$, is minimal while the approximation error is bounded by a given error tolerance, TOL, i.e. the event

$$\left| E[g(X(T))] - \frac{1}{M}\sum_{j=1}^{M} g(\overline{X}(T;\omega_j)) \right| \le \mathrm{TOL} \tag{2.6}$$

has a probability close to one. A priori error estimates of the computational error in (2.6) was first derived by Talay and Tubaro in [28]. The work [1] modified Talay's and Tubaro's error expansion to an expansion with computable leading order term in a posteriori form, based on computable stochastic flows and discrete dual backward problems. Stopped diffusion, including for example the barrier option, is an example where adaptive time steps improve the convergence rate, see [14].

Assume that the process X satisfies (2.4) and its approximation, $\overline{X}$, is given by (2.5), we have, see [1,21]

THEOREM 2.1 (Error expansion for SDEs). *Suppose there are positive constants k and C and an integer m_0 with the bounds*

$$\begin{aligned} &g \in \mathcal{C}^{m_0}_{loc}(\mathbb{R}^d),\ |\partial_\alpha g(x)| \le C(1+|x|^k), \quad \textit{for all} \quad |\alpha| \le m_0,\\ &E\left[|X(0)|^{2k+d+1} + |\overline{X}(0)|^{2k+d+1}\right] \le C, \end{aligned}$$

and

$$a \textit{ and } b \textit{ are bounded in } \mathcal{C}^{m_0}([0,T]\times\mathbb{R}^d).$$

Assume that $\overline{X}$ is constructed by the forward Euler method with step sizes h_n produced by the stochastic time step version of the adaptive algorithm in Section 2.4 and the corresponding $\Delta W_n \equiv W(t_{n+1}) - W(t_n)$ are generated by Brownian bridges. Assume also that $\overline{X}(0) = X(0)$ and $E[|X(0)|^{k_0}] \le C$ for some $k_0 \ge 16$. Then the time discretization error has the expansion

$$\begin{aligned} E[g(X(T)) \quad - \quad g(\overline{X}(T))] &= E\left[\sum_{n=1}^{N} \bar\rho_n (h_n)^2\right] \\ &+\mathcal{O}\left(\sqrt{\frac{\mathrm{TOL}}{c(\mathrm{TOL})}}\left(\frac{C(\mathrm{TOL})}{c(\mathrm{TOL})}\right)^{8/k_0}\right) E\left[\sum_{n=1}^{N}(h_n)^2\right] \end{aligned} \tag{2.7}$$

with computable leading order terms, where

$$\begin{aligned} \bar\rho_n(t_n,\overline{X}) \quad &\equiv \quad \frac{1}{2}\left(\frac{\partial}{\partial t}a_k + \partial_j a_k a_j + \partial_{ij} a_k d_{ij}\right)\varphi_k(t_{n+1}) \\ &+\frac{1}{2}\left(\frac{\partial}{\partial t}d_{km} + \partial_j d_{km} a_j + \partial_{ij} d_{km} d_{ij} + 2\partial_j a_k d_{jm}\right)\varphi'_{km}(t_{n+1}) \\ &+\partial_j d_{km} d_{jr} \varphi''_{kmr}(t_{n+1}), \end{aligned} \tag{2.8}$$

with $d_{ij} = \frac{1}{2}\sum_{l=1}^{l_0} b_i^l b_j^l$. The terms in the sum of (2.8) are evaluated at the a posteriori known points $(t_n, \overline{X}(t_n))$, i.e.

$$\partial_\alpha a \equiv \partial_\alpha a(t_n,\overline{X}(t_n)),\ \partial_\alpha b \equiv \partial_\alpha b(t_n,\overline{X}(t_n)),\ \partial_\alpha d \equiv \partial_\alpha d(t_n,\overline{X}(t_n)).$$

Here $\varphi \in \mathbb{R}^d$ is the solution of the discrete dual backward problem

$$\begin{aligned} \varphi_i(t_n) \quad &= \quad \partial_i c_j(t_n,\overline{X}(t_n))\varphi_j(t_{n+1}),\ t_n < T, \\ \varphi_i(T) \quad &= \quad \partial_i g(\overline{X}(T)), \end{aligned} \tag{2.9}$$

with

$$c_i(t_n,x) \equiv x_i + h_n a_i(t_n,x) + \Delta W_n^\ell b_i^\ell(t_n,x) \tag{2.10}$$

and its first and second variation

$$\varphi'_{ij} \quad \equiv \quad \partial_{x_j(t_n)}\varphi_i(t_n) \equiv \frac{\partial \varphi_i(t_n;\overline{X}(t_n)=x)}{\partial x_j}, \tag{2.11}$$

$$\varphi''_{ikm}(t_n) \quad \equiv \quad \partial_{x_m(t_n)}\varphi'_{ik}(t_n) \equiv \frac{\partial \varphi'_{ik}(t_n;\overline{X}(t_n)=x)}{\partial x_m}, \tag{2.12}$$

which satisfy

$$\begin{aligned}\varphi'_{ik}(t_n) &= \partial_i c_j(t_n,\overline{X}(t_n))\partial_k c_p(t_n,\overline{X}(t_n))\varphi'_{jp}(t_{n+1}) \\ &\quad +\partial_{ik} c_j(t_n,\overline{X}(t_n))\varphi_j(t_{n+1}),\ t_n<T, \\ \varphi'_{ik}(T) &= \partial_{ik} g(\overline{X}(T)),\end{aligned} \tag{2.13}$$

and

$$\begin{aligned}\varphi''_{ikm}(t_n) &= \partial_i c_j(t_n,\overline{X}(t_n))\partial_k c_p(t_n,\overline{X}(t_n))\partial_m c_r(t_n,\overline{X}(t_n))\varphi''_{jpr}(t_{n+1}) \\ &+ \partial_{im} c_j(t_n,\overline{X}(t_n))\partial_k c_p(t_n,\overline{X}(t_n))\varphi'_{jp}(t_{n+1}) \\ &+ \partial_i c_j(t_n,\overline{X}(t_n))\partial_{km} c_p(t_n,\overline{X}(t_n))\varphi'_{jp}(t_{n+1}) \\ &+ \partial_{ik} c_j(t_n,\overline{X}(t_n))\partial_m c_p(t_n,\overline{X}(t_n))\varphi'_{jp}(t_{n+1}) \\ &+ \partial_{ikm} c_j(t_n,\overline{X}(t_n))\varphi_j(t_{n+1}),\ t_n<T, \\ \varphi''_{ikm}(T) &= \partial_{ikm} g(\overline{X}(T)),\end{aligned} \tag{2.14}$$

respectively.

The previous result can also be directly applied to the particular case of deterministic time steps. Observe that the error expansion in Theorem 2.1 has the form

$$E[g(X(T))-g(\overline{X}(T))] = E\left[\sum_{n=1}^{N}\bar\rho_n h_n^2\right] + \text{ higher order terms} \tag{2.15}$$

and due to the almost sure convergence of the density $\bar\rho_n$ as we refine the discretization, see [21], it is suitable for use in the adaptive algorithm.

The computational error in (2.6) naturally separates into the time discretization error and the statistical error

$$\begin{aligned} &E[g(X(T))] - \frac{1}{M}\sum_{j=1}^{M} g(\overline{X}(T;\omega_j)) \\ &= \big(E[g(X(T))-g(\overline{X}(T))]\big) + \left(E[g(\overline{X}(T))] - \frac{1}{M}\sum_{j=1}^{M} g(\overline{X}(T;\omega_j))\right) \\ &\equiv \mathcal{E}_T + \mathcal{E}_S.\end{aligned} \tag{2.16}$$

The time steps for the realizations of the approximate solution $\overline{X}$ are determined from statistical approximations of the time discretization error, $\mathcal{E}_T$, and the number, M, of realizations of $\overline{X}$ is determined from the statistical error, $\mathcal{E}_S$. The statistical error and the time discretization error are combined in order to bound the computational error (2.16). Therefore we split a given error tolerance TOL into a statistical tolerance, TOL_S, and a time discretization tolerance, TOL_T. The computational work is roughly $\mathcal{O}(N\cdot M) = \mathcal{O}(\mathrm{TOL}_T^{-1}\ \mathrm{TOL}_S^{-2})$, therefore we use

$$\mathrm{TOL}_T = \frac{1}{3}\mathrm{TOL} \ \text{ and } \ \mathrm{TOL}_S = \frac{2}{3}\mathrm{TOL}, \tag{2.17}$$

by minimizing $\mathrm{TOL}_T^{-1}\ \mathrm{TOL}_S^{-2}$ under the constraint $\mathrm{TOL}_T + \mathrm{TOL}_S = \mathrm{TOL}$.

From the central limit theorem, the statistical error is bounded by the following quantity, i.e. the event

$$|\mathcal{E}_S(\overline{X};M)| \leq \mathtt{E}_S(\overline{X};M) \equiv c_0\ \frac{\mathcal{S}(\overline{X};M)}{\sqrt{M}} \tag{2.18}$$

has probability close to one, where $\mathcal{S}(\overline{X};M)$ is the sample standard deviation of $\overline{X}$ and c_0 is a constant related to the confidence interval.

The time discretization error is approximated by (2.15) and its contribution from each of the realizations controlled according to the adaptive algorithm described in Section 2.4.

2.3. An Error Expansion for PDEs. Consider a problem to compute a linear functional

$$g(u) := \int_D u\, G dx$$

for a given function $G \in L^2(D)$ and u is the solution of a second order elliptic partial differential equation of the form

$$-\mathrm{div}(a\nabla u) = f \tag{2.19}$$

in a given open bounded domain $D \subset \mathbb{R}^d$ with Dirichlet boundary data $u|_{\partial D} = 0$.

The finite element approximation u_h, of u in (2.19), is based on the standard variational formulation in the function set V_h of continuous piecewise isoparametric bilinear quadrilateral functions in $H_0^1(D)$, using an adaptive quadrilateral mesh with hanging nodes cf. [8]. The Sobolev space $H_0^1(D)$ is the usual Hilbert space of functions on D, vanishing on ∂D, with bounded first derivatives in $L^2(D)$. Let $\mathcal{T}$ denote the set of convex quadrilaterals K and let h_K be the local mesh size, i.e. the length of the longest edge of K. Then the variational problems for $u \in H_0^1(D)$ and $u_h \in V_h$ are

$$\begin{aligned}
\int_D a\nabla u \cdot \nabla v\, dx &= \int_D f\, v\, dx, \quad \forall v \in H_0^1(D),\\
\int_D a\nabla u_h \cdot \nabla v\, dx &= \int_D f\, v\, dx, \quad \forall v \in V_h.
\end{aligned} \tag{2.20}$$

A central role in the dual weighted error representation for $g(u) - g(u_h)$ is played by the dual function $\varphi \in H_0^1(D)$ which satisfies

$$\int_D a\nabla \varphi \cdot \nabla v\, dx = \int_D G\, v\, dx, \quad \forall v \in H_0^1(D). \tag{2.21}$$

Besides, its finite element approximation $\varphi_h \in V_h$, defined by

$$\int_D a\nabla \varphi_h \cdot \nabla v\, dx = \int_D G\, v\, dx, \quad \forall v \in V_h \tag{2.22}$$

is used to construct the error density $\bar\rho$ in Theorem 2.2.

For general meshes the convergence of the error density does not hold, since the orientation of the elements varies. Thus, here the analysis considers the asymptotic behavior of the error density $\bar\rho$ for adaptive refinements, with general quadrilateral initial meshes: successive division of reference square elements into four similar squares generates hanging node meshes consisting of unions of structured adapted meshes, where each structured mesh has the domain of an initial element; viewed in the initial reference element the structured adaptive mesh is an adaptive hanging node mesh with square elements. We restrict the study to such unions of structured adaptive hanging node meshes. The use of quadrilaterals can directly be extended to higher space dimension using tensor reference elements. Other refinements using e.g. subdivision of a simplex, in three and higher dimensions cf. [15], generate new edges which are not parallel to the old and would require additional analysis.

There is a smooth mapping of each initial element to a square, so that the refined initial element is mapped to a square hanging node mesh. Let $\mathcal{T}_I$ denote the subset of elements with an edge on the initial mesh. Theorem 2.2 states that the error density has a precise expansion using that the isoparametric bilinear coordinate transformation $X^{-1} : [0,1]^2 \to K_I$ maps the square and the square hanging node mesh to the initial element K_I and its refined hanging node quadrilateral mesh.

Let us now study the transformation of the variational formulation under such a mapping $X : K_I \to [0,1]^2$

$$\sum_{ij} \int_{K_I} (a_{ij} \frac{\partial u_h}{\partial x_j} \frac{\partial v}{\partial x_i} - fv) dx = \int_{[0,1]^2} (aX' u_h' \cdot X' v' - fv) J dx'$$

where X' is the Jacobian of X and J is the Jacobian determinant. Here we abuse the notation by writing v instead of $(v \circ X^{-1})$ and similarly for a, u_h, and f, for $x \in K_I$. Besides, we write $v' = \frac{\partial v}{\partial x_i'}$ instead of $\frac{\partial (v \circ X^{-1})}{\partial x_i'}$.

Therefore the variational equation in the transformed coordinates, x', takes the same form with a and f replaced by $a^* \equiv J(X')^t a X'$ and $f^* \equiv Jf$, respectively. Note that a^* and f^* are as smooth on $X(K_I)$ as the functions a and f are on K_I. To avoid messy notation, we will not always use the prime notation for coordinates obviously in the reference elements; we will also avoid notation for the dependence of X on the initial element K_I and assume that we for a point $x \in D$ choose the mapping X that corresponds to the initial element K_I which contains x. We will use the set of transformed elements $\mathcal{T}' \equiv \{X(K) : K \in \mathcal{T}\}$.

To define the approximate error density, $\bar{\rho}$, we will use averages of second difference quotients as follows. Consider a function w which is defined on a discretization of an interval $[0, L]$ with nodes $\{x_j : j = 0, \ldots, \bar{N}+1\} =: \bar{\mathcal{N}}$, where $x_0 = 0$ and $x_{\bar{N}+1} = L$. Let $h_+ \equiv x_{j+1} - x_j$ and $h_- \equiv x_j - x_{j-1}$ denote two consecutive edge sizes. Then define the average mesh size $\bar{h}$ and the difference quotients

$$\begin{aligned} \bar{h}_j &\equiv \frac{h_+ + h_-}{2} = \frac{x_{j+1} - x_{j-1}}{2}, \\ Dw(x_j) &\equiv \frac{w(x_j + h_+) - w(x_j)}{h_+} \\ D^2 w(x_j) &\equiv \frac{1}{\bar{h}_j} \left(\frac{w(x_j + h_+) - w(x_j)}{h_+} - \frac{w(x_j) - w(x_j - h_-)}{h_-} \right). \end{aligned} \tag{2.23}$$

Define $\overline{D^2 w} \in \mathbb{R}^{\bar{N}}$, implicitly as the solution $Y \in \mathbb{R}^{\bar{N}+2}$ of an auxiliary equation, i.e.

$$\begin{aligned} \overline{D^2 w_n} \equiv& Y_n, \ n = 1, \ldots, \bar{N}, \text{ where} \\ Y_n - \alpha^2 D^2 Y_n =& D^2 w_n, \ \ n = 1, \ldots, \bar{N}, \end{aligned} \tag{2.24}$$

with homogeneous Neumann boundary conditions, $Y_0 = Y_1$, $Y_{\bar{N}} = Y_{\bar{N}+1}$. The work [24] reports numerical results of different alternative averages, including the fast nearest neighbor variant. The convergence proof requires α to be sufficiently large compared to the mesh size.

Let us define $\bar{h}D_i^2 w$ as the difference quotients $\bar{h}D^2 w$, in (2.23), with respect to the x_i' reference directions $i = 1, 2$, respectively, and analogously for $D_i w$. The approximate error density, $\bar{\rho}$, in the transformed coordinates is now defined by

$$\overline{\rho}_K \equiv \frac{1}{48}\sum_{j=1}^{4}\Big(a^*_{11}\overline{D_1^2 u_h}\ \overline{D_1^2\varphi_h} + a^*_{22}\overline{D_2^2 u_h}\ \overline{D_2^2\varphi_h}\Big)(x_j^K) \tag{2.25}$$

where $x_1^K, x_2^K, x_3^K, x_4^K$ are the four corners of the square $K \in \mathcal{T}'$ illustrated in Figure 1.

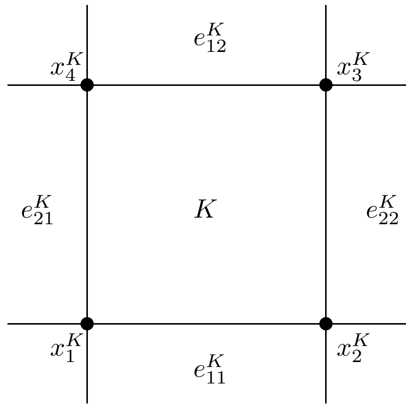


FIGURE 1. Corners x_j^K and edges e_{ij}^K of a square $K \in \mathcal{T}'$

Let $\mathcal{T}_H$ denote the subset of elements with hanging nodes in neighbors and let $\bar{\mathcal{T}}_H \equiv \bigcup_{K\in\mathcal{T}_H} K$. Let $W^{1,\infty}(D)$ denote the usual Sobolev space of functions with bounded first order derivatives in $L^\infty(D)$ and let h_{max} and h_{min} be the maximal and the minimal edge length in the mesh of V_h.

The proof of convergence of the error density, $\bar{\rho}$, uses the assumption that for some $\gamma \in (0, 1]$

$$\begin{aligned} \|u - u_h\|_{W^{1,\infty}(D)} + \|\varphi - \varphi_h\|_{W^{1,\infty}(D)} &= \mathcal{O}(C_\delta h_{max}),\\ \|u - u_h\|_{L^\infty(D)} + \|\varphi - \varphi_h\|_{L^\infty(D)} &= \mathcal{O}(h_{max}^{2\gamma}). \end{aligned} \tag{2.26}$$

In [10] such estimates are proved for finite element approximations of the coercive linear problems (2.19) and (2.21), with piecewise isoparametric bilinear quadrilateral elements and quasi uniform meshes, provided $u, \varphi \in \mathcal{C}^2(\bar{D})$; see [11] for nonlinear problems. If $|\bar{\rho}|$ is bounded away from zero, then the algorithm described in Section 2.4 produces quasi uniform meshes, h_{max}/h_{min} bounded by a constant that is possibly large depending on u and φ but independent of TOL. To include the case with $|\bar{\rho}|$ close to zero the algorithm uses a positive approximate error density $\hat{\rho} \geq \delta > 0$, defined in (2.38), with a parameter δ which tends to zero as TOL tends to zero; see Remark 2.3. With N denoting the number of elements in the final mesh, it is proved in [24] that

$$c\frac{\text{TOL}}{N} \leq \hat{\rho}_K h_K^4 \leq S_1\frac{\text{TOL}}{N}, \tag{2.27}$$

for problem independent constants c and S_1. With u and φ in $\mathcal{C}^3$ the quotient of the maximal and minimal mesh sizes becomes

$$\frac{h_{max}}{h_{min}} \le C\Big(\frac{\hat\rho_{max}}{\hat\rho_{min}}\Big)^{1/4} \le \frac{C_u}{\delta^{1/4}} \equiv \sqrt{C_\delta}, \tag{2.28}$$

where $C = (S_1/c)^{1/4}$ is independent of TOL, u and φ; while the constant C_u is independent of TOL but depends on u and φ. The δ-dependence of the quotient between h_{max} and h_{min} changes an $\mathcal{O}(h_{max})$ estimate of the right hand side in (2.26) to $\mathcal{O}(C_\delta h_{max})$.

The main result in [24] is

THEOREM 2.2. *Assume that $a \in \mathcal{C}^1(\bar D)$ and that the solutions $u \in \mathcal{C}^3(\bar D)$, $\varphi \in \mathcal{C}^3(\bar D)$ of* (2.19) *and* (2.21)*, respectively, are for some $\gamma \in (0,1]$ approximated uniformly with error satisfying* (2.26) *using piecewise isoparametric bilinear quadrilateral elements and a refined mesh, with at most one hanging node per edge, obtained by successively dividing the reference square elements into four similar squares. Assume also that the total area of the elements with a hanging node on a neighbor or with an edge on the initial mesh is asymptotically zero:*

$$\int_{\bar{\mathcal{T}}_H \cup \bar{\mathcal{T}}_I} dx = o(1), \qquad \text{as } h_{max} \to 0+\,. \tag{2.29}$$

Then the global error has the expansion

$$g(u) - g(u_h) = \sum_{K\in\mathcal{T}'} \Big(\bar\rho_K + \mathcal{O}(h^\gamma_{max}/\alpha + \alpha)\Big) h_K^4 + \mathcal{O}(C_\delta h_{max}) \int_{\bar{\mathcal{T}}_H \cup \bar{\mathcal{T}}_I} h_K dx \tag{2.30}$$

with uniformly convergent computable error density $\bar\rho$, defined by (2.25) *and* (2.23)*-*(2.24) *for $\alpha^{-1} = o(h^{-\gamma}_{max})$, satisfying*

$$\bar\rho = \widetilde\rho + \mathcal{O}(h^\gamma_{max}/\alpha + \alpha) \tag{2.31}$$

where

$$\widetilde\rho \equiv \frac{1}{12}\left(a^*_{11}\frac{\partial^2 u}{\partial x_1^2}\frac{\partial^2\varphi}{\partial x_1^2} + a^*_{22}\frac{\partial^2 u}{\partial x_2^2}\frac{\partial^2\varphi}{\partial x_2^2}\right)$$

is evaluated in the transformed coordinates on $[0,1]^2$.

Note that the convergence of $\bar\rho$ is uniform while the convergence of $\check\rho$, defined by $g(u)-g(u_h) \equiv \sum_{K\in\mathcal{T}'} \check\rho h_K^4$, is in $L^1(D)$ by assumption (2.29). It is important to notice that our restriction of the data, required by $u,\ \varphi \in \mathcal{C}^3(\bar D)$, includes examples with substantial adaptive gain. Section 3 shows that the optimal number of adaptive elements is $N^{opt} \simeq \mathrm{TOL}^{-1}\|\bar\rho\|^{d/2}_{L^{\frac{d}{d+2}}}$, while the number of uniform elements becomes $N^{uni} \simeq \mathrm{TOL}^{-1}\|\bar\rho\|^{d/2}_{L^1}$ to achieve the same error TOL. Although $u,\ \varphi \in \mathcal{C}^3(\bar D)$ their norms in these spaces may be large so that $\|\bar\rho\|_{L^{\frac{d}{d+2}}} \ll \|\bar\rho\|_{L^1}$.

In general, second order difference quotients of the interpolant on meshes with hanging nodes do not converge uniformly on D and this is why the averages are needed [24].

Theorem 2.2 proves that the error expansion

$$g(u) - g(u_h) = \sum_{K\in\mathcal{T}'} (\bar\rho_K + \mathcal{O}(\frac{h^\gamma_{max}}{\alpha} + \alpha)) h_K^{2+d} + \mathcal{O}(C_\delta h_{max}) \sum_{K\in\mathcal{T}'_H\cup\mathcal{T}'_I} h_K^{1+d} \tag{2.32}$$

has a well defined leading order error density $\overline{\rho}$ which converges uniformly as $h_{max} \to 0+$. We now assume that α has been chosen such that

$$\frac{h^{\gamma}_{max}}{\alpha} + \alpha = \mathcal{O}(h^{\widehat{\gamma}}_{\max}), \tag{2.33}$$

where $\widehat{\gamma} > 0$.

2.4. The Adaptive Algorithm. To motivate the approximate equidistribution of the error indicators in an adaptive algorithm, consider an asymptotic error expansion

$$\text{error} \simeq \sum_n \rho_n h_n^{p+d},$$

where h is the local isotropic mesh size and ρ is independent of h. The number of elements that corresponds to a mesh with size h can be determined by

$$N(h) \equiv \int_D \frac{dx}{h^d(x)}. \tag{2.34}$$

It seems hard to use the sign of the error indicators for constructing the mesh. Instead, we minimize the number of elements N in (2.34) under the more stringent constraint

$$\sum_{n=1}^{\bar{N}} |\rho_n| h_n^{d+p} = \int_D |\rho(x)| h^p dx = \text{TOL}, \tag{2.35}$$

with $D = [0, T]$ and $d = 1$ for ODEs and SDEs. The global order of convergence satisfies $p = 1$ for the Euler-Maruyama SDE approximation and $p = 2$ for the d-linear finite elements approximations used here. A standard application of a Lagrange multiplier yields the optimum

$$|\rho| (h^*)^{d+p} = \ \textit{constant} \tag{2.36}$$

and

$$h^* \equiv \frac{\text{TOL}^{\frac{1}{p}}}{|\rho|^{\frac{1}{d+p}}} \left(\int_D |\rho(x)|^{\frac{d}{d+p}} dx \right)^{-\frac{1}{p}}. \tag{2.37}$$

This condition is optimal only for density functions ρ with one sign, and in the PDE case, for meshes with shape regular elements, i.e. non stretched elements. To use the sign of the density or orientation of stretched elements in an optimal way is not considered here.

In the adaptive algorithm below we will use the positive approximate error density $\hat{\rho}_K$ defined by

$$\hat{\rho}|_K \equiv \hat{\rho}_K \equiv \min\big(\max\left(|\overline{\rho}_K|,\ \delta\right), \text{TOL}^{-r}\big) \tag{2.38}$$

with $r > 0$ and where the lower bound, $\delta > 0$, is chosen according to

Remark 2.3 (Lower bound for the error density).

$$\delta \equiv \text{TOL}^{\bar{\gamma}} \tag{2.39}$$

where the parameter $\bar{\gamma}$ is $0 < \bar{\gamma} < 1/(p+1)$ for ODEs, $\bar{\gamma} = 1/9$ for SDEs, and for PDEs it is chosen such that satisfies the two lower bounds

$$\bar{\gamma} < \frac{\widehat{\gamma}}{\widehat{\gamma} + 2} \ \text{ and } \int_{\bar{\mathcal{T}}_H \cup \bar{\mathcal{T}}_I} dx/\delta = o(1) \ \text{ as TOL} \to 0+, \tag{2.40}$$

and the upper bound $\delta = o(1)$ as TOL $\to 0+$. The lower bounds on $\delta > 0$ are motivated by the requirements that $h_{max} \to 0$ as TOL $\to 0$, that the bounds for the error density in (3.4) hold and that the error from hanging node elements becomes asymptotically negligible, see Theorem 3.2. The convergence of $\hat\rho$ towards the exact density requires the upper bound $\delta \to 0$.

The goal of the adaptive algorithm described below is to construct a mesh such that

$$\hat\rho_n h_n^{d+p} \approx \frac{\text{TOL}}{N}, \; n = 1, \dots, N, \tag{2.41}$$

which is an approximation of the optimal (2.36). Let the index $[k]$ refer to the refinement level in the sequence of adaptively refined meshes. For a mesh with elements $\{K_1, K_2, K_3, \dots, K_N\}$, we consider the piecewise constant error density and mesh functions $\rho|_{K_n} \equiv \rho_n \equiv \rho_{K_n}$, $\hat\rho|_{K_n} \equiv \hat\rho_n \equiv \hat\rho_{K_n}$ and $h|_{K_n} \equiv h_n \equiv h_{K_n}$. To achieve (2.41) let $s_1 \approx 1$ be a given constant, start with an initial mesh of size $h[1]$ and then specify iteratively a new mesh $h[k+1]$, from $h[k]$, using the following dividing strategy:

(2.42)

for all intervals (elements) $n = 1, 2, \dots, N[k]$

$\bar r_n[k] \equiv \hat\rho_n[k](h_n[k])^{d+p}$

if $\bar r_n[k] > s_1 \dfrac{\text{TOL}}{N[k]}$ **then**

mark interval (element) n for division.

(In addition, for the PDE case mark recursively all neighbors

that need division due to the hanging node constraint:

at most one hanging node per edge.)

endif

endfor

divide every marked interval (element) into 2^d uniform sub intervals (elements).

With this dividing strategy, it is natural to use the stopping criterion:

$$\textbf{if} \; \left(\max_{1\le n\le N[k]} \bar r_n[k] \le S_1 \frac{\text{TOL}}{N[k]} \right) \; \textbf{then} \; \text{stop.} \tag{2.43}$$

Here S_1 is a given constant, with $S_1 > s_1 \approx 1$, determined more precisely as follows: we want that the maximal error indicator decays quickly to the stopping level $S_1\text{TOL}/N$, but when almost all error indicators $\bar r_n$ satisfy $\bar r_n < s_1 \frac{\text{TOL}}{N}$ the reduction of the error may be slow. Theorem 3.1 shows that a slow reduction is avoided if S_1 satisfies (3.5).

REMARK 2.4 (SDE case: Stochastic Time Steps). Let $\overline{g} \equiv \sum_{j=1}^M g(\bar X(T);\omega_j)$ be the sample average approximation of the expected value $E[g(X(T))]$ and let $\overline{N}[m]$ be the sample average of the final number of time steps in the m-th batch of $M[m]$ realizations. In this case (2.42), is used iteratively for each of the realizations, $j = 1, \dots, M[m]$, with TOL_T instead of TOL and with $\overline{N}$ instead of N, see [21]. Replacing the integrals $\int_D \dots\, dx$ by $\int\int \dots\, dt\, dP = E\int \dots\, dt$ formally motivates the equidistribution of the error indicators for each realization of the Brownian motion.

3. Convergence Rates for the Adaptive Mesh Algorithm

This section presents results on the stopping, accuracy and efficiency properties of adaptive algorithm introduced in Section 2.

3.1. Adaptive Refinements and Stopping. To analyze the decay of the maximal error indicator, it is useful to understand the variation of the density $\hat\rho$ at different refinement levels, in particular we will consider an element or time step $K[k]$ and its parent on a previous refinement level, $p(K,k)$, with the corresponding error density $\hat\rho(K)[p(K,k)]$. It is possible to verify that the choice (2.39) of δ implies that $h_{max} \to 0$ as TOL $\to 0+$, see [24], [22], [21]. Hence Theorem 2.2 shows, for the PDE, that there is a limit error density $\widetilde\rho$ such that

$$\check\rho \xrightarrow{L^1} \widetilde\rho,\ \overline\rho \to \widetilde\rho \text{ and } \hat\rho \to |\widetilde\rho|,\ \text{ as TOL} \to 0+\,. \tag{3.1}$$

Similarly, the choice (2.39) of δ is used to show in [22] for ODEs that

$$\hat\rho \to |\widetilde\rho|,\ \text{ as TOL} \to 0+, \tag{3.2}$$

and in [21] that for each realization of the SDE, with $0 < \alpha < 1/2$,

$$\lim_{\text{TOL}\to 0+} h_{max}^{-\alpha}(\hat\rho - |\widetilde\rho|) = 0, \qquad \text{almost surely.} \tag{3.3}$$

A consequence of the uniform convergence $\hat\rho \to |\widetilde\rho|$, as TOL $\to 0+$, and (2.38) is that for all elements K and all refinement levels k there exists positive functions $\hat c$ and $\hat C$ close to 1 for sufficiently refined meshes, such that the error density satisfies

$$\begin{aligned} \hat c(K) &\le \frac{\hat\rho(K)[p(K,k)]}{\hat\rho(K)[k]} \le \hat C(K), \\ \hat c(K) &\le \frac{\hat\rho(K)[k-1]}{\hat\rho(K)[k]} \le \hat C(K), \end{aligned} \tag{3.4}$$

provided $\max_{K,k} h_K[k]$ is sufficiently small. In other words, (3.4) holds with e.g. $\hat c = 2^{-1} = \hat C^{-1}$ for sufficiently small $\max_{K,k} h_K[k]$. For SDEs the functions $\hat c$ and $\hat C$ are close to 1, almost surely.

THEOREM 3.1 (Stopping). *Suppose the adaptive algorithm uses the strategy* (2.42)-(2.43). *Assume that $\hat c$ satisfies* (3.4)*, for the elements or time steps corresponding to the maximal error indicator on each refinement level, and that*

$$S_1 \ge \frac{2^d}{\hat c} s_1, \quad 1 > \frac{\hat c^{-1}}{2^{d+p}}. \tag{3.5}$$

Then each refinement level either decreases the maximal error indicator with the factor

$$\max_{1\le n\le N[k+1]} \bar r_n[k+1] \le \frac{\hat c^{-1}}{2^{d+p}} \max_{1\le n\le N[k]} \bar r_n[k], \tag{3.6}$$

or stops the algorithm.

Here, the global order of convergence is $p = 1$ for the Euler-Maruyama SDE approximation and $p = 2$ for the d-linear finite elements approximations.

3.2. Accuracy of the Adaptive Algorithm. The adaptive algorithm guarantees that the estimate of the global error is bounded by a given error tolerance, TOL. An important question is whether the true global error is bounded by TOL asymptotically. Using the upper bound (2.43) of the error indicators and the convergence of ρ and $\bar\rho$ in Theorem 2.2, or the convergence (3.2), (3.3) respectively, the global error has the following estimate.

THEOREM 3.2 (Accuracy). *Suppose* (2.38)–(2.39) *hold and that, for PDEs* (2.33) *and the assumptions of Theorem 2.2 hold, or, for ODEs and SDEs,* (3.2) *and* (3.3) *holds, respectively. Then the adaptive algorithm* (2.42)–(2.43) *satisfies*

$$\limsup_{\mathrm{TOL}\to 0+} \Big(\mathrm{TOL}^{-1}\big|g(X(T)) - g(\overline{X}(T))\big| \Big) \le S_1, \quad \textit{for the ODE,}$$

$$\limsup_{\mathrm{TOL}\to 0+} \Big(\mathrm{TOL}^{-1}\big|g(u) - g(u_h)\big| \Big) \le S_1, \quad \textit{for the PDE,}$$

and, for the SDE, with the number of realizations M and any $c_0 > 0$ determined by (2.18),

$$\liminf_{\mathrm{TOL}\to 0+} P\left(\frac{1}{\mathrm{TOL}}\left|E[g(X(T))] - \frac{1}{M}\sum_{j=1}^{M} g(\overline{X}(T;\omega_j))\right| \le \frac{S_1+2}{3}\right) \ge \int_{-c_0}^{c_0} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx.$$

3.3. Efficiency of the Adaptive Algorithm. An important issue for the adaptive method is its efficiency; we want to determine a mesh with as few elements or time steps as possible providing the desired accuracy. From the definition (2.34) and the optimality condition (2.37), the number of optimal adaptive elements, N^{opt}, satisfies

$$N^{\mathrm{opt}} = \int_D \frac{dx}{(h^*(x))^d} = \frac{1}{\mathrm{TOL}^{\frac{d}{p}}}\left(\int_D |\rho[k](x)|^{\frac{d}{d+p}}\,dx\right)^{\frac{d+p}{p}} = \frac{1}{\mathrm{TOL}^{\frac{d}{p}}}\|\rho\|^{\frac{d}{p}}_{L^{\frac{d}{d+p}}}. \tag{3.7}$$

On the other hand, for the uniform mesh with elements $h = \textit{constant}$, the number of elements, N^{uni}, to achieve $\sum_{i=1}^{N}|\rho_i|h^{d+p} = \mathrm{TOL}$ becomes

$$N^{\mathrm{uni}} = \int_D \frac{dx}{h^d(x)} = \frac{\int_D dx}{\mathrm{TOL}^{\frac{d}{p}}}\left(\int_D |\rho[k](x)|dx\right)^{\frac{d}{p}} = \frac{\int_D dx}{\mathrm{TOL}^{\frac{d}{p}}}\|\rho\|^{\frac{d}{p}}_{L^1}. \tag{3.8}$$

Hence, the number of uniform elements is measured in the L^1-norm while the optimal number of elements is measured in the $L^{\frac{d}{d+p}}$ quasi-norm. Jensen's inequality implies $\|f\|_{L^{\frac{d}{d+p}}} \le (\int_D dx)^{\frac{p}{d}}\|f\|_{L^1}$, therefore an adaptive method may use fewer elements than the uniform element size method. For the SDE we get the optimal expected number of adaptive steps $E[N^{\mathrm{opt}}] = \frac{1}{\mathrm{TOL}}\left(E\int_0^T\sqrt{|\rho|}\,dt\right)^2 = \frac{1}{\mathrm{TOL}}\|\rho\|_{L^{\frac{1}{2}}(dtdP)}$ while with uniform time steps $E[N^{\mathrm{uni}}] = \frac{T}{\mathrm{TOL}}\int_0^T E|\rho|\,dt$.

The following theorem uses a lower bound of the error indicators, obtained from the refinement criterion (2.42) for the refined parent error indicator and the ratio of the error density (3.4), to show that the algorithm (2.42)-(2.43) generates a mesh which is optimal, up to a multiplicative constant independent of the data. In order to guarantee that, for sufficiently small TOL, all elements on the initial mesh are refined, the initial mesh size is assumed to obey

$$h_K[1] \ge \mathrm{TOL}^s, \tag{3.9}$$

where the parameter s has the upper bound $s < \frac{1-\bar{\gamma}}{p}$ and the lower bounds $0 < s$, for ODEs and SDEs, and $\frac{\bar{\gamma}}{\hat{\gamma}} < s$, for PDEs.

THEOREM 3.3 (Efficiency). *Assume that $\hat{C} = \hat{C}(t)$, $\hat{C}(t,\omega)$ or $\hat{C}(x)$ satisfies* (3.4) *for all elements at the final refinement level, that the assumptions of Theorem 3.2 hold, and that the initial mesh satisfies* (3.9) *for all elements K. Then there exists a constant $C > 0$, bounded by $(\frac{2^{d+p}}{s_1})^{\frac{d}{p}}$, such that, for sufficiently small* TOL, *the final number of adaptive time steps or elements N, of the algorithm* (2.42)-(2.43) *for ODEs or PDEs, satisfies*

$$(\mathrm{TOL}^{\frac{d}{p}} N) \leq C \, \|\hat{C}\hat{\rho}\|^{\frac{d}{p}}_{L^{\frac{d}{d+p}}} \leq C \left(\max_{x\in D} \hat{C}(x)^{\frac{d}{p}} \right) \|\hat{\rho}\|^{\frac{d}{p}}_{L^{\frac{d}{d+p}}}, \tag{3.10}$$

and

$$\lim_{\mathrm{TOL}\to 0+} \|\hat{\rho}\|_{L^{\frac{d}{d+p}}} = \|\widetilde{\rho}\|_{L^{\frac{d}{d+p}}},$$

$$\lim_{\mathrm{TOL}\to 0+} \max_{x\in D} \hat{C}(x)^{\frac{d}{p}} = 1,$$

i.e. the number of elements is asymptotically optimal up to the problem independent factor $C \leq (\frac{2^{d+p}}{s_1})^{\frac{d}{p}}$. For the SDE case the final sample average $\bar{N}[m] = \frac{1}{M[m]}\sum_{j=1}^{M[m]} N(\omega_j)$ of the number of adaptive steps of the algorithm (2.42)-(2.43) *satisfies*

$$\frac{\mathrm{TOL}_T \bar{N}[m]^2}{\bar{N}[m-1]} < C^2 \left(\int_0^T \frac{1}{M[m]} \sum_{j=1}^{M[m]} \sqrt{\hat{\rho}\hat{C}} \, dt \right)^2$$

and asymptotically

$$\limsup_{\mathrm{TOL}_T \to 0+} \mathrm{TOL}_T E[N] \leq C^2 \|\widetilde{\rho}\|_{L^{\frac{1}{2}}(dtdP)}$$

4. A Numerical Example

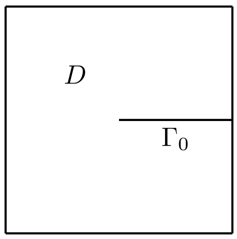
This section presents numerical results obtained with an implementation of the adaptive algorithm in Section 2.4 for a simplified elasticity problem using the error expansion (2.30) and the approximate error density (2.25). A more detailed description of the problem and the numerical results is given in [24].

EXAMPLE 4.1. The problem is to compute the functional

$$g(u) = \int_D u \, dx,$$

where the function u solves the Laplace equation in a slit domain

$$\begin{aligned} -\Delta u &= 0, && \text{in } D = (-1,1)^2 \setminus \Gamma_0, \\ u &= 0, && \text{on } \Gamma_0 = [0,1) \times \{0\}, \\ u &= u_b, && \text{on } \partial D \setminus \Gamma_0. \end{aligned}$$



With the function u_b given by $(r,\theta) \mapsto r^{1/2}\sin(\theta/2)$, in the polar coordinates (r,θ), the exact solution, u, is $r^{1/2}\sin(\theta/2)$. This singular problem is related to a problem with a rounded tip, leading to a smooth solution with multiple scales, see [24]. Numerical results show that the adaptive algorithm needs far less elements than the number of uniform elements needed to get comparable accuracy for this problem; see Table 1. Figure 2 shows that the number of elements in the accepted adaptively refined mesh is close to the estimated optimal number.

refinements	N	error	error estimate	TOL
uniform	32768	$4.8 \cdot 10^{-4}$	$2.0 \cdot 10^{-4}$	
adaptive	725	$3.2 \cdot 10^{-4}$	$3.1 \cdot 10^{-4}$	$2.0 \cdot 10^{-3}$
	3464	$3.8 \cdot 10^{-5}$	$4.5 \cdot 10^{-5}$	$2.4 \cdot 10^{-4}$
	20288	$4.4 \cdot 10^{-6}$	$6.9 \cdot 10^{-6}$	$3.1 \cdot 10^{-5}$

TABLE 1. The adaptive algorithm (2.42)–(2.43) uses far less elements than the number of uniform elements needed to get comparable accuracy in Example 4.1. The error $g(u) - g(u_h)$ is estimated by $\sum_K \bar{\rho}_K h_K^4$ using the signed error density (2.25), spatially varying averaging $\alpha(x) = \sqrt{r(x)h(x)}$, and lower bound $\delta = \sqrt{\mathrm{TOL}}$. The adaptive refinement and stopping used the parameter values $s_1 = 1$ and $S_1 = 10$.

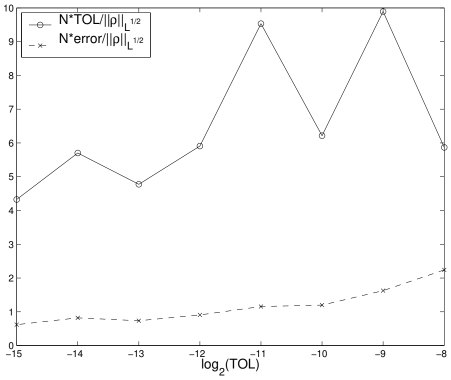


FIGURE 2. Efficiency estimate of the accepted mesh with the parameters in Table 1 for a sequence of tolerances. The number of elements, N, on the accepted mesh is compared to the estimated optimal $N^{\mathrm{opt}} = ||\hat{\rho}||_{L^{1/2}}/\mathrm{TOL}$, where the quasi-norm of the error density is computed on the finest mesh, corresponding to $\mathrm{TOL} = 2^{-15}$. Note that by the construction of the algorithm the final stopping error is expected to be only within a multiplicative factor $[1/4, 4]$ from the accuracy requirement TOL (without fine tuning of s_1).

The singular mode $r^{1/2}\sin(\theta/2)$ in u is present also in the solution to the dual problem

$$-\Delta\varphi = 1 \quad \text{in } D, \quad \varphi|_{\partial D} = 0.$$

Thus the conditions $u, \phi \in C^3(\bar{D})$ in Theorem 2.2 are violated and uniform convergence of the error density does not hold. Instead the error density grows like

$$\hat{\rho}(x) = \frac{\mathcal{O}(1)}{r^3}. \tag{4.1}$$

Still the numerical results in Figure 4 show that the minimal requirement on $\hat{c}$ and $\hat{C}$ in (3.4) to prove Theorem 3.1, 3.2 and 3.3 behave well.

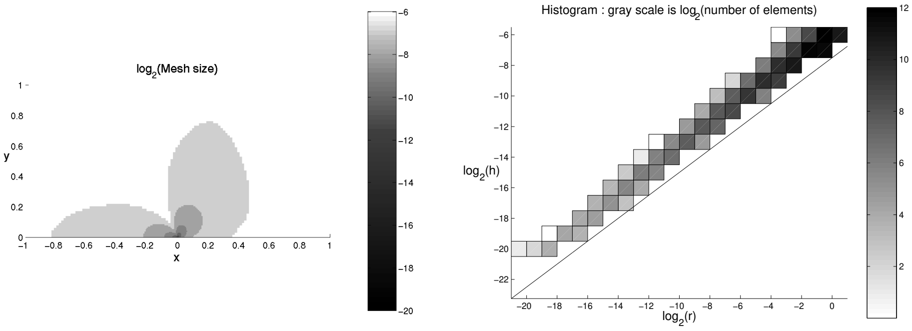


FIGURE 3. This is the mesh of the accepted solution corresponding to the smallest tolerance in Table 1. The element sizes h vary with the distance from the origin approximately like the optimal grading $r^{3/4}$. The reference line is the function $\sqrt{\mathrm{TOL}}\, r^{3/4}$.

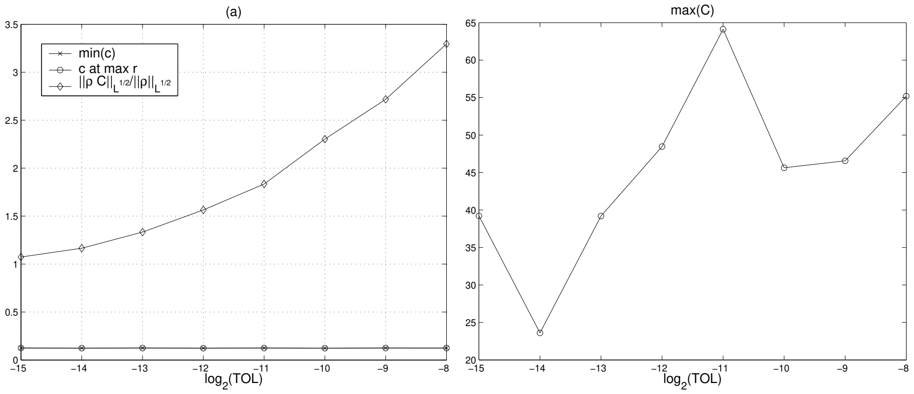


FIGURE 4. The quotients $\hat{c}(K) = \min\{\frac{\hat{\rho}(K)[p(K,k)]}{\hat{\rho}(K)[k]}, \frac{\hat{\rho}(K)[k-1]}{\hat{\rho}(K)[k]}\}$ and $\hat{C}(K) = \frac{\hat{\rho}(K)[p(K,k)]}{\hat{\rho}(K)[k]}$ have been computed for a sequence of tolerances, where K is an element on the accepted mesh, k. The minimal requirements for Theorems 3.1 and 3.2 use bounds on $\hat{c}^{-1}$ for the maximal error indicator, which is here approximately 8, because of (4.1). For Theorem 3.3 the minimal requirements use bounds on $||\hat{\rho}\hat{C}||_{L^{1/2}}$, which is close to $||\hat{\rho}||_{L^{1/2}}$ computed on the accepted mesh using $\mathrm{TOL} = 2^{-15}$; the maximal $\hat{C}$ may be significantly larger as seen on the right plot.

References

1. R. Tempone A. Szepessy and G. E. Zouraris, *Adaptive weak approximation of stochastic differential equations*, Comm. Pure Appl. Math. **54** (2001), 1169–1214.
2. M. Ainsworth and J. T. Oden, *A posteriori error estimation in finite element analysis*, Comput. Methods Appl. Mech. Engrg. **142** (1997), 1–88.
3. I. Babuska and W. C. Rheinboldt, *Error estimates for adaptive finite element computations*, SIAM J. Numer. Anal. **15** (1978), 736–754.
4. I. Babuška, A. Miller, and M. Vogelius, *Adaptive methods and error estimation for elliptic problems of structural mechanics*, Adaptive computational methods for partial differential equations, SIAM, Philadelphia, Pa., 1983, pp. 57–73.
5. I. Babuška and M. Vogelius, *Feedback and adaptive finite element solution of one-dimensional boundary value problems*, Numer. Math. **44** (1984), no. 1, 75–102.
6. R. E. Bank and A. Weiser, *Some a posteriori error estimators for elliptic partial differential equations*, Math. Comp. **44** (1985), 283–301.
7. R. Becker and R. Rannacher, *A feed-back approach to error control in finite element methods: basic analysis and examples*, East-West J. Numer. Math. **4** (1996), no. 4, 237–264.
8. ——, *An optimal control approach to a posteriori error estimation in finite element methods*, Acta Numerica **1** (2001), 102.
9. P. Binev, W. Dahmen, and R. DeVore, *Adaptive finite element methods with convergence rates*, Numer. Math. **97** (2004), no. 2, 219–268.
10. S. C. Brenner and L. R. Scott, *The mathematical theory of finite element methods*, Texts in Applied Mathematics 15, Springer–Verlag, New York, 1994.
11. F. Christian and G. Santos, *A posteriori estimators for nonlinear elliptic partial differential equations*, J. Comput. Appl. Math. **103** (1999), 99–114.
12. A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods for elliptic operator equations: convergence rates*, Math. Comp. **70** (2001), no. 233, 25–75.
13. W. Dörfler, *A convergent adaptive algorithm for poisson's equation*, SIAM J. Numer. Anal. **33** (1996), no. 3, 1106–1124.
14. A. Dzougoutov, K.-S. Moon, E. von Schwerin A. Szepessy, and R. Tempone, *Adaptive monte carlo algorithms for stopped diffusion*.
15. H. Edelsbrunner and D. R. Grayson, *Edgewise subdivision of a simplex*, Discrete Comput. Geom. **24** (2000), 707–719, ACM Symposium on Computational Geometry (Miami, FL, 1999).
16. K. Eriksson, D. Estep, P. Hansbo, and C. Johnson, *Introduction to adaptive methods for differential equations*, Acta Numerica, 1995, pp. 105–158.
17. T. Strouboulis I. Babuska and S. K. Gangaraj, *Guaranteed computable bounds for the exact error in the finite element solution. I. one dimensional model problem*, Comput. Methods Appl. Mech. Engrg. **176** (1999), 51–79.
18. C. Johnson and A. Szepessy, *Adaptive finite element methods for conservation laws based on a posteriori error estimates*, Comm. Pure Appl. Math. **48** (1995), 199–234.
19. L. Machiels, A. T. Patera, J. Peraire, and Y. Maday, *A general framework for finite element a posteriori error control: application to linear and nonlinear convection-dominated problems*, preprint (presentation at ICFD Conference on Numerical Methods for Fluid Dynamics, Oxford, England, March 31-April 3, 1998).
20. Y. Maday, A. T. Patera, and J. Peraire, *A general formulation for a posteriori bounds for output functionals of partial differential equations; applications to the eigenvalue problem*, C. R. Acad. Sci. Paris Sér. I Math. **328** (1999), 823–828.
21. K.-S. Moon, A. Szepessy, R. Tempone, and G. E. Zouraris, *Convergence rates for adaptive weak approximation of stochastic differential equations*, preprint, 2002.
22. ——, *Convergence rates for adaptive approximation of ordinary differential equations*, Numer. Math. **96** (2003), 99–129.
23. ——, *A variational principle for adaptive approximation of ordinary differential equations*, Numer. Math. **96** (2003), 131–152.
24. K.-S. Moon, A. Szepessy, E. von Schwerin, and R. Tempone, *Convergence rates for an adaptive dual weighted residual finite element algorithm*, http://www.nada.kth.se/∼szepessy.
25. P. Morin, R. Nochetto, and K. G. Siebert, *Convergence of adaptive finite element methods*, SIAM Rev. **44** (2002), no. 4, 631–658, (electronic) (2003), Revised reprint of *Data oscillation and convergence of adaptive FEM*, SIAM J. Numer. Anal., **38** (2000), no. 2, 466–488.

26. I. Babuška, J. Hugger, T. Strouboulis, K Copps, and S. K. Gangaraj, *The asymptotically optimal meshsize function for bi-p degree interpolation over rectangular elements*, J. Comput. Appl. Math. **90** (1998), no. 2, 185–221.
27. R. Stevenson, *An optimal adaptive finite element method*, preprint, to appear in SIAM J. Numer. Anal., 2004.
28. D. Talay and L. Tubaro, *Expansion of the global error for numerical schemes solving stochastic differential equations*, Stochastic Anal. Appl. **8** (1990), 483–509.

Department of Mathematics, University of Maryland, College Park MD 20742-4015, USA

E-mail address: moon@math.umd.edu

Institutionen för Numerisk Analys och Datalogi, Kungl. Tekniska Högskolan, S–100 44 Stockholm, Sweden.

E-mail address: schwerin@nada.kth.se

Matematiska Institutionen, Kungl. Tekniska Högskolan, S–100 44 Stockholm, Sweden.

E-mail address: szepessy@nada.kth.se

Institute for Computational and Engineering Sciences (ICES), The University of Texas at Austin, 1 University Station C0200, Austin, Texas 78712, USA.

E-mail address: rtempone@ices.utexas.edu