No. 628 March 2020

On the threshold condition for Dörfler marking

L. Diening, C. Kreuzer

ISSN: 2190-1767

ON THE THRESHOLD CONDITION FOR DÖRFLER MARKING

LARS DIENING AND CHRISTIAN KREUZER

Abstract. It is an open question if the threshold condition $\theta < \theta_\star$ for the Dörfler marking parameter is necessary to obtain optimal algebraic rates of adaptive finite element methods. We present a (non-PDE) example fitting into the common abstract convergence framework (axioms of adaptivity) and which is potentially converging with exponential rates. However, for Dörfler marking $\theta > \theta_\star$ the algebraic converges rate can be made arbitrarily small.

1. Introduction

In the pioneering work [Ste07] of Stevenson proved rate optimality of the standard adaptive finite element method for the Poisson problem. In contrast to the prior result [BDD04], which used optimal coarsening based on fast tree approximation [BD04], Stevenson proved optimality of the refinement based on Dörfler marking totally avoiding coarsening as is standard for stationary problems. One main ingredient of the proof is a local estimate, which bounds the distance of discrete solutions on nested meshes relative to the refined elements only. As a consequence each refinement that ensures some error reduction must satisfy a Dörfler marking condition. This is the key in the optimality proof of [Ste07], since it allows to compare the adaptive refinement strategy with optimal refinements when the adaptive marking parameter is below a certain threshold depending on the ratio of the efficiency and reliability constants of the a posteriori estimator.

During the last decade, the approach of Stevenson became very popular and was further developed e.g. in [CKNS08, DK08, BDK12, CN11, KS11, FFP14, FKMP13, CPR13, BN10, Gan13]. For and overview of the topic see e.g. the monographs [NV12, NSV09] and for a more exhaustive list of related works, we refer the reader to [CFPP14].

2010 *Mathematics Subject Classification.* 65N50, 65N15, 41A25.

Key words and phrases. Adaptive finite element methods, optimal complexity, convergence, Dörfler marking.

The research of Lars Diening was partially supported by the DFG through the CRC 1283.

It is, however, still open whether or not the threshold for the marking is sharp in general or if it is only a technical artefact. In [CFPP14], Carstensen, Feischl, Page, and Praetorius presented an unifying axiomatic approach for proving rate optimality of adaptive finite element methods. Therein, the threshold is obtained with a slightly different technique (compare with Section 2.4) and it is stated that

> *...the upper bound for adaptivity parameters which guarantee quasi-optimal convergence rates, is independent of the efficiency constant. Such an observation might be a first step to the mathematical understanding of the empirical observation that each adaptivity parameter $0 < \theta \leq 0.5$ yields optimal convergence rates in the asymptotic regime.*

In this work, we present an example, which satisfies the abstract axioms of adaptivity from [CFPP14, CR17] but also efficiency, whichs allows to apply the original techniques from [Ste07]. However, for any $\theta \in (0, 1)$, we can adjust the parameters such that, although exponential convergence is possible, the adaptive loop with Dörfler marking fails to have arbitrary bad algebraic rates. It turns out that, the threshold parameter of [CFPP14, CR17] not involving efficiency suggests for our example values, which are slightly too conservative. The threshold parameter of [Ste07] however, which is the ratio of the reliability and efficiency constant, is sharp.

We emphasise that the example is not within the context of finite element discretisations of partial differential equations (compare also with Remark 9). Therefore, technically speaking, we cannot claim any conclusions for the relevant practical cases, however, we clarify that the threshold condition can neither be avoided nor significantly improved within the axiomatic framework of [CFPP14, CR17] even when relying on efficiency as in [Ste07]. Notice that our example also confirms the result of [DKS16, KS16] that the maximum marking strategy is more robust in the sense that it provides optimal convergence rates without restriction on the maximum marking parameter; compare with Section 3.4.

2. Axioms of adaptivity

In this section, we recall the axiomatic approach in [CFPP14, CR17] of the proof of optimal convergence rates for adaptive finite element methods. The presentation mainly follows [CR17] neglecting the additional refinement indicators but is simplified tailored to our needs, i.e., we do not cover the full generality of [CFPP14, CR17]. In particular, we consider the most simple case of dimension $d = 1$ and let $\Omega \subset \mathbb{R}$ be a non-empty open interval. Before we state the axioms, we first verify the refinement conditions.

2.1. **Refinement by bisection.** Let $\mathcal{T}_0$ be an initial partition of Ω into closed intervals (called macro elements) and denote by $\mathbb{T}$ the set of its possible refinements. To be more precise, we introduce bisection of a closed interval $[a,b]$, $a<b$ by

$$\mathrm{BISECT}([a,b]) = \{[a, \tfrac{a+b}{2}],\, [\tfrac{a+b}{2}, b]\}.$$

We say that $\mathcal{T}_\star$ is a refinement of $\mathcal{T}$ (or $\mathcal{T}_\star \geq \mathcal{T}$) iff there exist a finite sequence of partitions $\{\mathcal{T}_n\}_{n=1}^N$ and $T_n \in \mathcal{T}_n$, $n=1,\ldots,N-1$, such that $\mathcal{T}_\star = \mathcal{T}_N$ and $\mathcal{T}_1 = \mathcal{T}$ as well as

$$\mathcal{T}_{n+1} = (\mathcal{T}_n \setminus \{T_n\}) \cup \mathrm{BISECT}(T_n), \quad n = 1,\ldots,N-1.$$

With this definition $(\mathbb{T}, \leq)$ becomes a lattice and we can define for $\mathcal{T}_a, \mathcal{T}_b \in \mathbb{T}$

$$\mathcal{T}_a \wedge \mathcal{T}_b := \arg\max\{\mathcal{T}' \in \mathbb{T} \colon \mathcal{T}' \leq \mathcal{T}_a \text{ and } \mathcal{T}' \leq \mathcal{T}_b\}$$

and

$$\mathcal{T}_a \vee \mathcal{T}_b := \arg\min\{\mathcal{T}' \in \mathbb{T} \colon \mathcal{T}_a \leq \mathcal{T}' \text{ and } \mathcal{T}_b \leq \mathcal{T}'\};$$

i.e. the finest common coarsening respective the coarsest common refinement. Moreover, we have

$$\#(\mathcal{T}_a \vee \mathcal{T}_b) = \#\mathcal{T}_a + \#\mathcal{T}_b - \#(\mathcal{T}_a \wedge \mathcal{T}_b) \leq \#\mathcal{T}_a + \#\mathcal{T}_b - \#\mathcal{T}_0.$$

Thanks to the bisection rule, we can also recursively assign to each $T \in \mathcal{T}$, $\mathcal{T} \in \mathbb{T}$, a generation by

$$g(T) = 0 \text{ if } T \in \mathcal{T}_0 \qquad \text{and} \qquad g(T) = g(T') + 1 \text{ if } T \in \mathrm{BISECT}(T').$$

Defining for $\mathcal{T} \in \mathbb{T}$ and $\mathcal{M} \subset \mathcal{T}$ the refinement procedure

$$\mathrm{REFINE}(\mathcal{T};\mathcal{M}) := \arg\min\{\mathcal{T}' \in \mathbb{T} \colon \mathcal{T} \leq \mathcal{T}' \text{ and } \mathcal{M} \cap \mathcal{T}' = \emptyset\},$$

we obviously have

$$\mathrm{REFINE}(\mathcal{T};\mathcal{M}) = (\mathcal{T} \setminus \mathcal{M}) \cup \mathrm{BISECT}(\mathcal{M}) \tag{2.1}$$

with $\mathrm{BISECT}(\mathcal{M}) := \bigcup\{\mathrm{BISECT}(T) : T \in \mathcal{M}\}$. Obviously, we thus have $\mathcal{T}_\star = \mathrm{REFINE}(\mathcal{T};\mathcal{M}) \in \mathbb{T}$ and

$$\#\mathcal{T}_\star - \#\mathcal{T} = \#\mathcal{M}.$$

We conclude that our refinement framework satisfies the requirements in [CFPP14, Section 2.4].

2.2. **Adaptive algorithm with Dörfler marking.** In the following we formulate the basic conditions from [CFPP14, CR17] sufficient for optimal convergence rates of the adaptive Dörfler marking strategy. The precise algorithm and the optimality result is stated in section 2.3 below.

We assume that for any $\mathcal{T} \in \mathbb{T}$, and any element $T \in \mathcal{T}$, we have nonnegative indicators $\eta_{\mathcal{T}}(T)$ available and set

$$\eta_{\mathcal{T}}^2(\mathcal{M}) = \sum_{T \in \mathcal{M}} \eta_{\mathcal{T}}^2(T) \quad \text{for any } \mathcal{M} \subset \mathcal{T}.$$

Moreover, we assume that there is a nonnegative distance measure on $\mathbb{T}$ denoted by $\delta(\mathcal{T}, \mathcal{T}_\star)$ for $\mathcal{T}, \mathcal{T}_\star \in \mathbb{T}$. This distance measures in the application the error between to discrete solutions.

Based on the above indicators, we can formulate the adaptive algorithm.

Algorithm 1 (AFEM with Dörfler marking)**.** *Let $\mathcal{T}_0$ be an initial triangulation of Ω and $\theta \in (0,1)$ a given marking parameter. Set $k := 0$ and iterate*

- *Compute the indicators $\{\eta_{\mathcal{T}_k}(T) : T \in \mathcal{T}_k\}$.*
- *Choose $\mathcal{M}_k \subset \mathcal{T}_k$ such that*

$$\theta \eta_{\mathcal{T}_k}^2(\mathcal{T}_k) \leq \eta_{\mathcal{T}_k}^2(\mathcal{M}_k) \tag{2.2}$$

  with quasi-minimal cardinality, i.e., $\#\mathcal{M}_k \leq C_{DM} \#\mathcal{M}$ for some fixed constant $C_{DM} \geq 1$ and all $\mathcal{M} \subset \mathcal{T}_k$ with $\theta \eta_{\mathcal{T}_k}(\mathcal{T}_k) \leq \eta_{\mathcal{T}_k}(\mathcal{M})$.

  If the set $\mathcal{M}_k$ has minimal cardinality (i.e. $C_{DM} = 1$) then (2.2) is called optimal Dörfler marking*.*
- *Construct the refinement*

$$\mathcal{T}_{k+1} = \text{REFINE}(\mathcal{T}_k; \mathcal{M}_k)$$

  and set $k := k + 1$.

2.3. **The axioms.** In this section, we present the axioms of adaptivity from [CFPP14, CR17] in a simplified version tailored to our needs. We assume that the indicators and the distance measure from the previous section 2.2 satisfy the following conditions:

(A1) **Stability.** For all $\mathcal{T}, \mathcal{T}_\star \in \mathbb{T}$ with $\mathcal{T}_\star \geq \mathcal{T}$, we have

$$|\eta_{\mathcal{T}}(\mathcal{T} \cap \mathcal{T}_\star) - \eta_{\mathcal{T}_\star}(\mathcal{T} \cap \mathcal{T}_\star)| \leq \delta(\mathcal{T}, \mathcal{T}_\star)$$

(A2) **Reduction.** There exists $\rho \in [0,1)$ such that for all $\mathcal{T}, \mathcal{T}_\star \in \mathbb{T}$ with $\mathcal{T}_\star \geq \mathcal{T}$, we have

$$\eta_{\mathcal{T}_\star}(\mathcal{T}_\star \setminus \mathcal{T}) \leq \rho \eta_{\mathcal{T}}(\mathcal{T}_\star \setminus \mathcal{T}).$$

(A3) **Discrete reliability.** There exists $C_3 > 0$, such that for all $\mathcal{T}, \mathcal{T}_\star \in \mathbb{T}$ with $\mathcal{T}_\star \geq \mathcal{T}$, we have
$$\delta(\mathcal{T}, \mathcal{T}_\star)^2 \leq C_3 \eta^2_{\mathcal{T}}(\mathcal{T} \setminus \mathcal{T}_\star).$$

(A4) **Quasi-orthogonality.** There exists $C_4 > 0$, such that for any sequence $\{\mathcal{T}_k\}_k \subset \mathbb{T}$ of nested partitions (i.e. $\mathcal{T}_1 \leq \mathcal{T}_2 \leq \ldots$), we have for all $\ell \in \mathbb{N}$ that
$$\sum_{k=\ell}^{\infty} \delta(\mathcal{T}_{k+1}, \mathcal{T}_k)^2 \leq C_4 \eta^2_{\mathcal{T}_\ell}(\mathcal{T}_\ell).$$

Remark 2. *We note that (A1)–(A4) correspond to the respective conditions in* [CR17] *with*

(A1) $$\Lambda_1 = 1$$

(A2) $$\rho_2 = \rho, \qquad \Lambda_2 = 0$$

(A3) $$\Lambda_{ref} = 1, \quad \Lambda_3 = C_3, \quad \textit{and} \quad \hat{\Lambda}_3 = 0$$

(A4) $$\Lambda_4 = C_4.$$

The conditions (B1) and (B2) in [CR17] *do not apply since we assume $\mu_\ell \equiv 0$ for the second indicator in* [CR17]*. Note that therefore also the quasi-monotonicity (QM) condition in* [CR17] *is satisfied automatically since we may chose $\hat{\Lambda}_3 = 0$ in* [CR17, Theorem 3.2].

We recall the following main theorem from [CFPP14, CR17].

Theorem 3. *Suppose that (A1)–(A4) hold and define the threshold*
$$\theta_\star := \frac{1}{1 + C_3}.$$
Then Algorithm 1 is rate optimal if $\theta < \theta_\star$, i.e., in this case we have for all $s > 0$ there exists $C > 0$ with
$$\begin{aligned}\sup_{k \in \mathbb{N}} \big((\#\mathcal{T}_k - \#\mathcal{T}_0)^s \eta_{\mathcal{T}_k}(\mathcal{T}_k) \big) \\ \leq C \sup_{N \in \mathbb{N}} \big(N^s \min\{\eta_{\mathcal{T}}(\mathcal{T}) \colon \#\mathcal{T} - \#\mathcal{T}_0 \leq N\} \big).\end{aligned}$$

The original approach of Stevenson is slightly different in that it utilizes also the efficiency of the estimator. This allows to show convergence rates for the error rather than for the estimator; compare with Remark 5. To work in the framework of Stevenson we modify/sharpen two of the axioms.

We replace the stability (A1) by the following efficiency condition.

(A1') **Efficiency.** For all $\mathcal{T} \in \mathbb{T}$, we have
$$C_1 \eta_{\mathcal{T}}^2(\mathcal{T}) \le \delta(\mathcal{T})^2$$
for some constant $C_1 > 0$.

Typically, $\delta(\mathcal{T})$ measures the error, e.g. in applications the distance between the discrete solution and the exact solution of the PDE. In our example, $\delta(\mathcal{T})$ is the distance to the 'finest' partition

$$\delta(\mathcal{T}) := \inf\{\delta(\mathcal{T}_\star, \mathcal{T}) : \mathcal{T}_\star \in \mathbb{T} \text{ with } \mathcal{T} \ge \mathcal{T}\}. \tag{2.3}$$

As a consequence of the lattice structure of $(\mathbb{T}, \ge)$, we have that the definition (2.3) is unique when $\delta(\cdot, \mathcal{T}) : \mathbb{T} \to \mathbb{R}$ is non-increasing under refinement. This property is immediate when replacing the quasi-orthogonality condition (A4) by the following orthogonality property; compare also with Remark 5 below.

(A4') **Orthogonality.** For all $\mathcal{T}, \mathcal{T}_\star, \mathcal{T}_\circ \in \mathbb{T}$ with $\mathcal{T} \le \mathcal{T}_\star \le \mathcal{T}_\circ$, we have
$$\delta(\mathcal{T}_\circ, \mathcal{T}_\star)^2 + \delta(\mathcal{T}_\star, \mathcal{T})^2 = \delta(\mathcal{T}_\circ, \mathcal{T})^2.$$

Then Stevenson proved in [Ste07] the following version of Theorem 3 with a different threshold.

Theorem 4. *Suppose that (A2)–(A3) and assume in addition (A1') and (A4'). Define the threshold*
$$\tilde{\theta}_\star := \frac{C_1}{C_3}.$$
Then Algorithm 1 is rate optimal if $\theta < \tilde{\theta}_\star$, i.e., in this case we have for all $s > 0$ there exists $C > 0$ with
$$\sup_{k \in \mathbb{N}} (\#\mathcal{T}_k - \#\mathcal{T}_0)^s \delta(\mathcal{T}_k) \le C \sup_{N \in \mathbb{N}} N^s \min\{\delta(\mathcal{T}) \colon \#\mathcal{T} - \#\mathcal{T}_0 \le N\}.$$

Remark 5 ($\delta(\mathcal{T})$ vs. $\eta_{\mathcal{T}}(\mathcal{T})$)**.** *We remark that in [Ste07] optimal convergence rates are proved for $\delta(\mathcal{T})$ in contrast to [CFPP14, CR17], which focus on $\eta_{\mathcal{T}}(\mathcal{T})$. Let us compare these two approaches.*

Since $\delta(\mathcal{T}_\star, \mathcal{T}) \ge 0$ for $\mathcal{T}_\star, \mathcal{T} \in \mathbb{T}$ with $\mathcal{T}_\star \ge \mathcal{T}$, we conclude from (A4') that $\delta : \mathbb{T} \to \mathbb{R}_\ge$ is monotone decreasing under refinement. Moreover, recalling (2.3), it follows from (A3) that
$$\delta(\mathcal{T}) \le C_3 \eta_{\mathcal{T}}(\mathcal{T}).$$
This is an upper bound or equivalently (A4) with $C_4 = C_3$.

Combining this with the efficiency (A1'), we have equivalence of the error and the estimator, i.e. $C_1 \eta_{\mathcal{T}}^2(\mathcal{T}) \le \delta(\mathcal{T})^2 \le C_3 \eta_{\mathcal{T}}^2(\mathcal{T})$.

As a consequence, we have that $\delta(\mathcal{T})$ converges iff $\eta_{\mathcal{T}}(\mathcal{T})$ converges and both converge then with the same rates. Taking the setting (A1)–(A4) of [CFPP14, CR17]*, however, the convergence behavior of δ and η may differ. In particular, in view of (A3), the convergence rate of δ may be better than the one of η.*

2.4. **The origin of the thresholds.** We start with discussing the threshold $\theta_\star$ from Theorem 3. Let $\mathcal{T}_k \in \mathbb{T}$ be from Algorithm 1 for some $k \in \mathbb{N}$ and assume $\mathcal{T}_\star \geq \mathcal{T}_k$ with

$$\eta_{\mathcal{T}_\star}(\mathcal{T}_\star) \leq \kappa \eta_{\mathcal{T}_k}(\mathcal{T}_k) \tag{2.4}$$

for some arbitrarily fixed $\kappa \in (0,1)$. Then from $\eta_{\mathcal{T}_\star}(\mathcal{T}_\star \cap \mathcal{T}_k) \leq \eta_{\mathcal{T}_\star}(\mathcal{T}_\star)$ and (A1) for $0 < \gamma < \frac{1}{\kappa^2} - 1$ in Youngs inequality we conclude that

$$\begin{aligned}
(1-(1+\gamma)\kappa^2)\eta_{\mathcal{T}}^2(\mathcal{T}_k) &\leq \eta_{\mathcal{T}_k}^2(\mathcal{T}_k) - (1+\gamma)\eta_{\mathcal{T}_\star}^2(\mathcal{T}_\star) \\
&\leq \eta_{\mathcal{T}_k}^2(\mathcal{T}_k) - (1+\gamma)\eta_{\mathcal{T}_\star}^2(\mathcal{T}_\star \cap \mathcal{T}_k) \\
&\leq \eta_{\mathcal{T}_k}^2(\mathcal{T}_k \setminus \mathcal{T}_\star) + \frac{1}{1+\gamma^{-1}}\delta(\mathcal{T}_\star, \mathcal{T}_k)^2.
\end{aligned}$$

Now applying (A3), we obtain

$$\frac{1-(1+\gamma)\kappa^2}{1+\frac{C_3}{1+\gamma^{-1}}}\,\eta_{\mathcal{T}_k}^2(\mathcal{T}) \leq \eta_{\mathcal{T}_k}^2(\mathcal{T}_k \setminus \mathcal{T}_\star). \tag{2.5}$$

In other words, the set of elements $\mathcal{T}_k \setminus \mathcal{T}_\star$ from $\mathcal{T}_k$ which are refined in $\mathcal{T}_\star$ satisfies a Dörfler marking property. When

$$\theta \leq \frac{1-(1+\gamma)\kappa^2}{1+\frac{C_3}{1+\gamma^{-1}}}, \tag{2.6}$$

then the quasi minimal cardinality of $\mathcal{M}_k$ in Algorithm 1 implies

$$\#(\mathcal{T}_k \setminus \mathcal{T}_\star) \leq C \# \mathcal{M}_k,$$

which is the key in the proof of the rate optimality Theorem 3; compare e.g. with [Ste07, CKNS08, CFPP14]. By choosing $\kappa > 0$ small, we observe that (2.6) can only hold if

$$\theta < \theta_\star = \frac{1}{1+C_3}. \tag{2.7}$$

In order to discuss the threshold $\tilde{\theta}_\star$ from Theorem 4, instead of satisfying the estimator reduction (2.4), we assume that $\mathcal{T}_\star \in \mathbb{T}$, $\mathcal{T}_\star \geq \mathcal{T}_k$ reduces the distance

$$\delta(\mathcal{T}_\star) \leq \kappa \delta(\mathcal{T})$$

for some $\kappa \in (0,1)$ arbitrarily fixed. We then have

$$\begin{aligned}(1-\kappa^2)C_1\,\eta^2_{\mathcal{T}}(\mathcal{T}) &\le (1-\kappa^2)\delta(\mathcal{T})^2\\ &\le \delta(\mathcal{T})^2-\delta(\mathcal{T}_\star)^2=\delta(\mathcal{T}_\star,\mathcal{T})^2\\ &\le C_3\eta^2_{\mathcal{T}}(\mathcal{T}\setminus\mathcal{T}_\star).\end{aligned}$$

Arguing as before, there exists $\kappa>0$ such that the above computation implies a Dörfler condition $\theta\eta^2_{\mathcal{T}_k}(\mathcal{T}_k)\le\eta^2_{\mathcal{T}_k}(\mathcal{T}_k\setminus\mathcal{T}_\star)$ only if $\theta\le\tilde{\theta}_\star=\frac{C_1}{C_3}$.

3. Dörfler marking with suboptimal convergence rates

For a given marking parameter $\theta\in(0,1)$, and $s_0>0$, we construct an example with an exponential optimal convergence rate that satisfies the axioms of adaptivity (A1)–(A4) and also (A1')+(A4') with $\delta(\mathcal{T})=\eta_{\mathcal{T}}(\mathcal{T})$, $\mathcal{T}\in\mathbb{T}$ (i.e. $C_1=1$) and local reliability constant $C_3=K>0$. Thus, in this situation $\frac{1}{K}=\tilde{\theta}_\star\ge\theta_\star=\frac{1}{1+K}$, i.e. the threshold of Theorem 4 is less conservative than the one in Theorem 3. In particular, thanks to the possible exponential convergence, if $\theta\le\tilde{\theta}_\star$ then Algorithm 1 converges with any possible algebraic rate $s>0$.

However, we will see that for any $\theta\in(0,1)$, the example can be adjusted with arbitrary close $\tilde{\theta}_\star<\theta$, such that the adaptive Algorithm 1 will not converge with rate s_0, more precisely

$$\sup_{k\in\mathbb{N}}(\#\mathcal{T}_k-\#\mathcal{T}_0)^{s_0}\eta_{\mathcal{T}_k}(\mathcal{T}_k)=\sup_{k\in\mathbb{N}}(\#\mathcal{T}_k-\#\mathcal{T}_0)^{s_0}\delta(\mathcal{T}_k)=\infty.$$

This shows that a threshold conditions as in Theorems 3 and 4 cannot be avoided in the axiomatic framework of [CFPP14, CR17] and, moreover, can be arbitrarily restrictive.

3.1. **The setup.** For $\Omega=(0,M+1)$, $M\in\mathbb{N}$, consider the initial partition

$$\mathcal{T}_0=\{[0,1],[1,2],\dots,[M,M+1]\} \tag{3.1}$$

and denote the set of admissible refinements according to Section 2.1 by $\mathbb{T}$. For $A\subset\Omega$ and $\mathcal{T}\in\mathbb{T}$, we use the notation

$$\mathcal{T}|_A:=\{T\in\mathcal{T}\,:\,T\subset\overline{A}\}.$$

We denote by $T_0(\mathcal{T})$ the element of $\mathcal{T}$ that contains zero and by $g_0(\mathcal{T}):=g(T_0(\mathcal{T}))$ its generation.

For fixed $\alpha,\beta>0$ and $K>1$, we define

$$\eta^2_{\mathcal{T}}(T):=\begin{cases}2^{-\alpha g_0(\mathcal{T})-\beta(g(T)+\frac{m-1}{M})}\,|T| & \text{if } T\subset[m,m+1],\ m\ge 1\\ \frac{1}{K-1}\,\eta^2_{\mathcal{T}}(\mathcal{T}|_{[1,M+1]}) & \text{if } T=T_0(\mathcal{T}),\\ 0 & \text{else.}\end{cases} \tag{3.2}$$

The constant K will be the reliability constant, i.e. $C_3 = K$. The constants $\alpha > 0$ and $M \in \mathbb{N}$, will be chosen later depending on the convergence rate s_0 and the marking parameter θ.

This yields the immediate relation

$$\eta_{\mathcal{T}}^2(T_0(\mathcal{T})) = \frac{1}{K}\,\eta_{\mathcal{T}}^2(\mathcal{T}). \tag{3.3}$$

In particular, the estimator of the element $T_0(\mathcal{T})$ is comparable to the estimator on all of $\mathcal{T}$.

For $\mathcal{T}_* \geq \mathcal{T}$ we define

$$\delta(\mathcal{T}_\star, \mathcal{T})^2 := \eta_{\mathcal{T}}^2(\mathcal{T}) - \eta_{\mathcal{T}_\star}^2(\mathcal{T}_\star) \geq 0. \tag{3.4}$$

Thus, $\delta(\mathcal{T})^2 := \eta_{\mathcal{T}}^2(\mathcal{T})$.

Note that a refinement of $T_0(\mathcal{T})$ decreases all estimators by a factor of $2^{-\alpha}$. Thus, for suitable refinement, the estimator η defined in (3.2) converges exponentially.

Lemma 6 (Exponential Convergence)**.** *Assume that $\{\mathcal{T}_k\}_{k\in\mathbb{N}_0}$ is generated by a repeatedly refinement of $T_0(\mathcal{T}_k)$, i.e*

$$\mathcal{T}_i = \text{REFINE}(\{T_0(\mathcal{T}_{i-1})\}; \mathcal{T}_{i-1}).$$

Then the estimator and the distance converge exponentially, i.e.

$$\delta(\mathcal{T}_k)^2 = \eta_{\mathcal{T}_k}^2(\mathcal{T}_k) = 2^{-\alpha k}\eta_{\mathcal{T}_0}^2(\mathcal{T}_0) \quad \textit{and} \quad \#\mathcal{T}_k - \#\mathcal{T}_0 = k.$$

In particular, we have for all rates $s > 0$ that

$$\sup_{N\in\mathbb{N}} N^s \min\{\eta_{\mathcal{T}}(\mathcal{T})\colon \#\mathcal{T} - \#\mathcal{T}_0 \leq N\} < \infty.$$

Proof. Observing that

$$\#\mathcal{T}_k - \#\mathcal{T}_0 = k = g(T_0(\mathcal{T}_k)),$$

the assertion is an immediate consequence of (3.2) and (2.1). $\square$

Remark 7. *We are using in our setup the bisection method without conforming closure. This is just for the sake of a clear presentation. All observations remain valid if a conforming closure step is included.*

3.2. **Verifying the axioms.** In order to verify the conditions (A1)–(A4) as well as (A1') and (A4'), we first observe the estimator defined in (3.2) is locally non-increasing under refinement.

Lemma 8 (Monotonicity)**.** *Let $\mathcal{T}_\star \geq \mathcal{T}$ such that $T \in \mathcal{T}$ is bisected into $\{T_1, T_2\} = \text{BISECT}(T) \subset \mathcal{T}_\star$, then*

$$\eta_{\mathcal{T}_\star}^2(T_1) + \eta_{\mathcal{T}_\star}^2(T_2) \leq \eta_{\mathcal{T}}^2(T).$$

In particular, for all $\mathcal{T}_\star \geq \mathcal{T}$, we have $\eta_{\mathcal{T}_\star}(\mathcal{T}_\star) \leq \eta_{\mathcal{T}}(\mathcal{T})$.

Proof. We consider first the case $0 \notin T$, then

$$\eta^2_{\mathcal{T}_\star}(T_i) \le 2^{-1-\beta}\eta^2_{\mathcal{T}}(T), \qquad i = 1,2. \tag{3.5}$$

Therefore, we conclude from $\alpha \ge 0$ that

$$\eta^2_{\mathcal{T}_\star}(T_1) + \eta^2_{\mathcal{T}_\star}(T_2) \le 2\,2^{-1-\beta}\eta^2_{\mathcal{T}}(T) = 2^{-\beta}\eta^2_{\mathcal{T}}(T) \le \eta^2_{\mathcal{T}}(T).$$

Assume now that $0 \in T$ (i.e. $T = T_0(\mathcal{T})$). We first observe for unrefined elements

$$\tilde{T} \in \mathcal{T} \cap \mathcal{T}_\star \qquad \Rightarrow \qquad \eta^2_{\mathcal{T}_\star}(\tilde{T}) \le \eta^2_{\mathcal{T}}(\tilde{T}). \tag{3.6}$$

W.l.o.g. let $0 \in T_1$, then we have from (3.6) and (3.5), that $\eta^2_{\mathcal{T}_\star}(T_2) = 0$ and thus

$$\begin{aligned}
\eta^2_{\mathcal{T}_\star}(T_1) + \eta^2_{\mathcal{T}_\star}(T_2) = \eta^2_{\mathcal{T}_\star}(T_1) &= \frac{1}{K-1}\,\eta^2_{\mathcal{T}_\star}(\mathcal{T}_\star|_{[1,M+1]}) \\
&\le 2^{-\alpha}\frac{1}{K-1}\,\eta^2_{\mathcal{T}}(\mathcal{T}|_{[1,M+1]}) \\
&\le \frac{1}{K-1}\,\eta^2_{\mathcal{T}}(\mathcal{T}|_{[1,M+1]}) = \eta^2_{\mathcal{T}}(T_0(\mathcal{T}))
\end{aligned} \tag{3.7}$$

This finishes the proof. ◻

We are now in the position to verify the axioms of adaptivity from Section 2.3.

(A1) **Stability.** We recall from (3.6) that $\eta^2_{\mathcal{T}_\star}(T) \ge \eta^2_{\mathcal{T}}(T)$ for each unrefined $T \in \mathcal{T} \cap \mathcal{T}_\star$. Moreover, by the local monotonicity (Lemma 8), we have $\eta^2_{\mathcal{T}_\star}(\mathcal{T}_\star \setminus \mathcal{T}) \le \eta^2_{\mathcal{T}}(\mathcal{T} \setminus \mathcal{T}_\star)$ and therefore

$$\begin{aligned}
&\big|\eta^2_{\mathcal{T}}(\mathcal{T} \cap \mathcal{T}_\star) - \eta^2_{\mathcal{T}_\star}(\mathcal{T} \cap \mathcal{T}_\star)\big| \\
&\quad = \eta^2_{\mathcal{T}}(\mathcal{T} \cap \mathcal{T}_\star) - \eta^2_{\mathcal{T}_\star}(\mathcal{T} \cap \mathcal{T}_\star) \\
&\quad \le \eta^2_{\mathcal{T}}(\mathcal{T} \cap \mathcal{T}_\star) - \eta^2_{\mathcal{T}_\star}(\mathcal{T} \cap \mathcal{T}_\star) + \eta^2_{\mathcal{T}}(\mathcal{T} \setminus \mathcal{T}_\star) - \eta^2_{\mathcal{T}_\star}(\mathcal{T}_\star \setminus \mathcal{T}) \\
&\quad = \eta^2_{\mathcal{T}}(\mathcal{T}) - \eta^2_{\mathcal{T}_\star}(\mathcal{T}_\star) \\
&\quad = \delta^2(\mathcal{T}, \mathcal{T}_\star).
\end{aligned}$$

This and $|a-b| \le \sqrt{|a^2-b^2|}$ for $a \ge b \ge 0$ imply (A1).

(A2) **Reduction.** Assume first, that $T_0(\mathcal{T})$ is not refined in $\mathcal{T}_\star \ge \mathcal{T}$. Then we have

$$\eta^2_{\mathcal{T}_\star}(\mathcal{T}_\star \setminus \mathcal{T}) \le 2^{-\beta}\eta^2_{\mathcal{T}}(\mathcal{T} \setminus \mathcal{T}_\star).$$

If on the other hand $T_0(\mathcal{T})$ is refined in $\mathcal{T}_\star$, then each estimator is at least reduced by the factor $2^{-\alpha}$, and thus similar to (3.7), we obtain

$$\eta^2_{\mathcal{T}_\star}(\mathcal{T}_\star \setminus \mathcal{T}) \le 2^{-\alpha}\eta^2_{\mathcal{T}}(\mathcal{T} \setminus \mathcal{T}_\star).$$

Thus in both cases we conclude (A2) with $\rho = 2^{-\frac{\min\{\alpha,\beta\}}{2}}$.

(A3) Discrete reliability. Assume first that $T_0(\mathcal{T})$ is not refined in $\mathcal{T}_\star \geq \mathcal{T}$, i.e. $T_0(\mathcal{T}) \in \mathcal{T}_\star$. Then $\eta^2_{\mathcal{T}}(\mathcal{T} \cap \mathcal{T}_\star) = \eta^2_{\mathcal{T}_\star}(\mathcal{T} \cap \mathcal{T}_\star)$ and

$$\begin{aligned}\delta^2(\mathcal{T},\mathcal{T}_\star) &= \eta^2_{\mathcal{T}}(\mathcal{T}) - \eta^2_{\mathcal{T}_\star}(\mathcal{T}_\star)\\ &= \eta^2_{\mathcal{T}}(\mathcal{T}\setminus\mathcal{T}_\star) + \eta^2_{\mathcal{T}}(\mathcal{T}\cap\mathcal{T}_\star) - \eta^2_{\mathcal{T}_\star}(\mathcal{T}_\star\setminus\mathcal{T}) - \eta^2_{\mathcal{T}_\star}(\mathcal{T}\cap\mathcal{T}_\star)\\ &= \eta^2_{\mathcal{T}}(\mathcal{T}\setminus\mathcal{T}_\star) - \eta^2_{\mathcal{T}_\star}(\mathcal{T}_\star\setminus\mathcal{T})\\ &\leq \eta^2_{\mathcal{T}}(\mathcal{T}\setminus\mathcal{T}_\star)\end{aligned}$$

If otherwise $T_0(\mathcal{T}) \in \mathcal{T}\setminus\mathcal{T}_\star$, then we obtain with (3.2) that

$$\begin{aligned}\delta^2(\mathcal{T},\mathcal{T}_\star) &\leq \eta^2_{\mathcal{T}}(\mathcal{T})\\ &= K\,\eta^2_{\mathcal{T}}(T_0(\mathcal{T}))\\ &\leq K\,\eta^2_{\mathcal{T}}(\mathcal{T}\setminus\mathcal{T}_\star).\end{aligned}$$

In other words, we have (A3) with $C_3 = \max\{1, K\} = K$.

(A4) Quasi-orthogonality. For a sequence of nested meshes $\mathcal{T}_1 \leq \mathcal{T}_2 \leq \cdots$, in $\mathbb{T}$, we have

$$\begin{aligned}\sum_{k=1}^{N} \delta(\mathcal{T}_{k+1},\mathcal{T}_k)^2 &= \sum_{k=1}^{N} \eta^2_{\mathcal{T}_k}(\mathcal{T}_k) - \eta^2_{\mathcal{T}_{k+1}}(\mathcal{T}_{k+1})\\ &= \eta^2_{\mathcal{T}_1}(\mathcal{T}_1) - \eta^2_{\mathcal{T}_{N+1}}(\mathcal{T}_{N+1})\\ &\leq \eta^2_{\mathcal{T}_1}(\mathcal{T}_1).\end{aligned}$$

Taking the limit $N \to \infty$ and observing from Lemma 8 that $\delta(\mathcal{T}_{k+1},\mathcal{T}_k)^2 \geq 0$, we conclude (A4) with $C_4 = 1$.

Also (A1') and (A4') are satisfied by the error indicators.

(A1') Efficiency. For $\mathcal{T} \in \mathbb{T}$, we have from Lemma 6 that

$$\delta(\mathcal{T}) = \eta_{\mathcal{T}}(\mathcal{T}), \qquad \text{i.e.,} \quad C_1 = 1.$$

(A4') Orthogonality. Indeed, from (3.4), we have for $\mathcal{T}, \mathcal{T}_\star, \mathcal{T}_\circ \in \mathbb{T}$ with $\mathcal{T} \leq \mathcal{T}_\star \leq \mathcal{T}_\circ$ that

$$\begin{aligned}\delta(\mathcal{T}_\circ,\mathcal{T}_\star)^2 + \delta(\mathcal{T}_\star,\mathcal{T})^2 &= \eta^2_{\mathcal{T}_\star}(\mathcal{T}_\star) - \eta^2_{\mathcal{T}_\circ}(\mathcal{T}_\circ) + \eta^2_{\mathcal{T}}(\mathcal{T}) - \eta^2_{\mathcal{T}_\star}(\mathcal{T}_\star)\\ &= \eta^2_{\mathcal{T}}(\mathcal{T}) - \eta^2_{\mathcal{T}_\circ}(\mathcal{T}_\circ) = \delta(\mathcal{T}_\circ,\mathcal{T})^2.\end{aligned}$$

Concluding, we have that Theorem 3 and Theorem 4 apply with the thresholds

$$\theta_\star = \frac{1}{K+1} \qquad \text{and} \qquad \tilde{\theta}_\star = \frac{1}{K}, \tag{3.8}$$

respectively.

Remark 9. *We have verified that our indicators η and the distance function δ satisfies all stated axioms of adaptivity. Nevertheless, we suspect that our example (3.2) can be realised within the context of finite elements for differential equations, as is suggested by the following example.*

Let $a : \Omega = (0, M) \to \mathbb{R}_{>}$ piecewise constant with respect to $\mathcal{T}_0$. We consider the following one dimensional problem: For $f \in H^{-1}(\Omega)$, find $u \in H_0^1(\Omega)$ such that

$$\forall v \in H_0^1(\Omega) \qquad \int_0^M a\, u'v' \mathrm{d}x = \langle f, v\rangle.$$

For $\mathcal{T} \in \mathbb{T}$, we chose $\mathbb{V}(\mathcal{T}) := \{v \in H_0^1(\Omega) \colon v|_T \in \mathbb{P}_k, T \in \mathcal{T}\}$ and define $u \in \mathcal{T} \in \mathbb{V}(\mathcal{T})$ to be the Galerkin approximation of u in $\mathbb{V}(\mathcal{T})$. Recalling $H_0^1(\Omega) \hookrightarrow C_0(\bar{\Omega})$ since $d = 1$, we have that the Lagrange interpolant is stable. Using this, standard a posteriori techniques readily show that

$$\int_T a(u' - u'_{\mathcal{T}})^2 \mathrm{d}x = \frac{1}{a_{|T}} \|f + (au''_{\mathcal{T}})\|^2_{H^{-1}(T)} \quad \forall T \in \mathcal{T}.$$

An error indicator is then typically obtained by estimating the local residuals on the right hand side in a computable way. However, their relation to the error is purely local and therefore a dependence of the local indicators on the generation $g(T_0(\mathcal{T}))$ as in (3.2) is not possible.

3.3. **Dörfler marking.** We recall that $K > 1$ is just our reliability constant, i.e. $C_3 = K$, which is related to the threshold by

$$\tilde{\theta}_\star = \frac{1}{C_3} = \frac{1}{K}.$$

Theorem 10. *Let $\theta \in (0, 1)$ (the Dörfler parameter) and $s_0 > 0$ (the rate) be given. Then there exist $\alpha, \beta > 0$, $M \in \mathbb{N}$, and*

$$\frac{1}{K} = \tilde{\theta}_\star < \theta \quad \textit{arbitrary close,}$$

such that Algorithm 1 with optimal Dörfler marking fails to converge with rate s_0, i.e.

$$\sup_{k\in\mathbb{N}} (\#\mathcal{T}_k - \#\mathcal{T}_0)^{s_0} \eta_{\mathcal{T}_k}(\mathcal{T}_k) = \sup_{k\in\mathbb{N}} (\#\mathcal{T}_k - \#\mathcal{T}_0)^{s_0} \delta(\mathcal{T}_k) = \infty.$$

Proof. For an arbitrary fixed $\varepsilon > 0$ we will determine parameters K, α, β and M such that $\tilde{\theta}_\star = \frac{1}{K}$ satisfies

$$\tilde{\theta}_\star < \theta < \tilde{\theta}_\star + \varepsilon,$$

i.e. for some $\gamma \in (0, \varepsilon)$

$$\theta = \tilde{\theta}_\star + \gamma = \frac{1}{K} + \gamma \qquad \text{or equivalently} \qquad K = \frac{1}{\theta - \gamma}. \tag{3.9}$$

The constants $\alpha, \beta > 0$ are related to the rate s_0. We fix $\beta = s_0 > 0$ and determined α at the end of the proof.

In order to introduce the general idea of the proof, we define

$$I_k := \big[\big((k-1) \bmod M\big) + 1, \big((k-1) \bmod M\big) + 2\big].$$

Therefore, for any $j \in \mathbb{N}$, we have that $I_{0+j}, I_{1+j}, \ldots, I_{M-1+j}$ represent the intervals $[1, 2], \ldots, [M, M+1]$ with order shifted by j. Below, we will adjust the parameters such that in each iteration $k = 0, 1, 2, \ldots$, the set

$$\mathcal{M}_k = \{T_0(\mathcal{T}_k)\} \cup \{T \in \mathcal{T}_k \,:\, T \in I_k\} \tag{3.10}$$

satisfies *optimal Dörfler marking*. In fact, we will have $\mathcal{M}_k \subset \mathcal{T}_k$ such that $\#\mathcal{M}_k$ is minimal with the property

$$\eta^2_{\mathcal{T}_k}(\mathcal{M}_k) = \theta \eta^2_{\mathcal{T}_k}(\mathcal{T}_k). \tag{3.11}$$

Consider first $k = 0$. It follows from (3.3) that

$$\eta^2_{\mathcal{T}_0}(T_0(\mathcal{T}_0)) = \frac{1}{K}\eta^2_{\mathcal{T}_0}(\mathcal{T}_0),$$
$$\eta^2_{\mathcal{T}_0}(\mathcal{T}_0|_{[1,M+1]}) = \left(1 - \frac{1}{K}\right)\eta^2_{\mathcal{T}_0}(\mathcal{T}_0).$$

Moreover, we have from the definition of our indicators (3.2) for $M \in \mathbb{N}$, that

$$\eta^2_{\mathcal{T}_0}(\mathcal{T}_0|_{I_j}) = 2^{-\beta\frac{j-1}{M}}\eta^2_{\mathcal{T}_0}(\mathcal{T}_0|_{[1,2]}), \qquad \text{for } j = 1, \ldots, M. \tag{3.12}$$

Consequently, it follows from (3.3) that

$$\begin{aligned}
\eta^2_{\mathcal{T}_0}(\mathcal{T}_0) = K\eta^2_{\mathcal{T}_0}(T_0(\mathcal{T}_0)) &= \frac{K}{K-1}\eta^2_{\mathcal{T}_0}(\mathcal{T}_0|_{[1,M+1]}) \\
&= \frac{K}{K-1}\sum_{j=1}^{M} 2^{-\beta\frac{j-1}{M}}\eta^2_{\mathcal{T}_0}(\mathcal{T}_0|_{[1,2]}) = \frac{K}{K-1}S(\beta, M)\eta^2_{\mathcal{T}_0}(\mathcal{T}_0|_{[1,2]}),
\end{aligned}$$

where

$$S(\beta, M) := \sum_{j=1}^{M} 2^{-\beta\frac{j-1}{M}} = \frac{1 - 2^{-\beta}}{1 - 2^{-\frac{\beta}{M}}}.$$

In other words

$$\eta^2_{\mathcal{T}_0}(\mathcal{T}_0|_{[1,2]}) = \frac{\eta^2_{\mathcal{T}_0}(\mathcal{T}_0|_{[1,M+1]})}{S(\beta, M)} = \frac{1}{S(\beta, M)}\left(1 - \frac{1}{K}\right)\eta^2_{\mathcal{T}_0}(\mathcal{T}_0), \tag{3.13}$$

and thus the Dörfler marking condition (3.11) reduces to finding $K > 1$ and $M \in \mathbb{N}$ with

$$\frac{1}{K} + \frac{1}{S(\beta, M)}\left(1 - \frac{1}{K}\right) = \theta$$

or equivalently (recall (3.9))

$$S(\beta, M) = \frac{1 - \frac{1}{K}}{\theta - \frac{1}{K}} = \frac{1 - \theta + \gamma}{\gamma} = \frac{1}{\gamma}(1 - \theta) + 1. \tag{3.14}$$

Since $\beta > 0$, we have $S(\beta, M) = 1$ and $\lim_{M\to\infty} S(\beta, M) = \infty$ and thus there exist $M \in \mathbb{N}$ and $\gamma \in (0, \varepsilon)$ satisfying (3.14).

Overall, thanks to (3.12) and the fact that $\#\mathcal{T}_0|_{I_0} = \#\mathcal{T}_0|_{I_1} = \cdots = \#\mathcal{T}_0|_{I_{M-1}} = 1$, for $\beta = s_0 > 0$, we have fixed the parameters $M \in \mathbb{N}$ and $K > 1$, such that (3.9) and (3.14) hold. This implies in particular optimal Dörfler marking (2.2) for $k = 0$.

We shall now deal with the case $k > 0$ and let $k = \ell M + m \in \mathbb{N}$ with $\ell \in \mathbb{N}_0$ and $m \in \{0, \ldots, M - 1\}$. It is easy to see from (3.10) that

$$g(T) = \begin{cases} \ell + 1, & \text{if } T \subset I_j \text{ for some } j \in \{0, \ldots, m - 1\} \\ \ell, & \text{if } T \subset I_j \text{ for some } j \in \{m, \ldots, M - 1\}. \end{cases} \tag{3.15}$$

Consequently, we have

$$\eta^2_{\mathcal{T}_k}(\mathcal{T}_k|_{I_{m+j}}) = 2^{-\beta\frac{j-1}{M}}\eta^2_{\mathcal{T}_k}(\mathcal{T}_k|_{I_m}), \qquad \text{for } j = 1, \ldots, M.$$

Therefore, the relative sizes of the indicators on the intervals I_{m+j} correspond to a cyclic permutation of the initial situation in (3.12). In other words we have (3.11). Note that $I_m = I_k$ by construction and thus

$$\eta^2_{\mathcal{T}_k}(\mathcal{T}_k|_{I_k}) \geq \eta^2_{\mathcal{T}_k}(\mathcal{T}_k|_{I_j}), \qquad j \in \{0, \ldots, M - 1\}.$$

Moreover, it follows from (3.15) that

$$\#\mathcal{T}_k|_{I_k} \leq \#\mathcal{T}_k|_{I_j}, \qquad j \in \{0, \ldots, M - 1\}$$

and thus the Dörfler marking is again minimal.

We turn now to investigate the rate of the algorithm so to fix α. After each M iterations in Algorithm 1 each element of $[1, M]$ is refined once. Thus, for all $\ell \in \mathbb{N}$

$$\#\mathcal{T}_{\ell M} - \#\mathcal{T}_0 \geq 2^\ell M.$$

Moreover, after M algorithm cycles the element containing zero is M times refined and all elements in $[1, M]$ are refined once. Thus, the error indicator of the whole partition decreases after M cycles by $2^{-\alpha M-\beta}$, i.e.

$$\eta^2_{\mathcal{T}_{\ell M}}(\mathcal{T}_{\ell M}) = 2^{(-\alpha M-\beta)\ell}\eta^2_{\mathcal{T}_0}(\mathcal{T}_0).$$

Therefore, we have with $\beta = s_0$ that

$$(\#\mathcal{T}_{\ell M} - \#\mathcal{T}_0)^{s_0}\eta_{\mathcal{T}_{\ell M}}(\mathcal{T}_{\ell M}) \geq M^{s_0}2^{(s_0-\frac{\alpha}{2}M-\frac{\beta}{2})\ell} = M^{s_0}2^{(\frac{s_0}{2}-\frac{\alpha}{2}M)\ell}.$$

Choosing $\alpha \in (0, \frac{s_0}{M})$, we have $\frac{s_0}{2} - \frac{\alpha}{2}M > 0$ and thus

$$\sup_{k\in\mathbb{N}}(\#\mathcal{T}_k - \#\mathcal{T}_0)^{s_0}\eta_{\mathcal{T}_k}(\mathcal{T}_k) = \infty.$$

This finishes the proof. □

3.4. **Maximums Strategy.** Another popular refinement strategy is the *maximum strategy.* For this the Dörfler marking (2.2) in Algorithm 1 is replaced by

$$\mathcal{M}_k := \big\{T \in \mathcal{T}_k \colon \eta^2_{\mathcal{T}_k}(T) \geq \mu \max\{\eta^2_{\mathcal{T}_k}(T')\colon T' \in \mathcal{T}_k\}\big\}, \tag{3.16}$$

for some marking parameter $\mu \in (0, 1]$. The strategy requires to determine the maximal local indicator. Then all elements with indicators that are up to the factor μ maximal are refined. Obviously, the strategy is getting more selective as closer μ is to one.

The maximum strategy has been analyzed in [DKS16] and it has been shown that for any $\mu \in (0, 1]$ the algorithm is *instance optimal.* The term *instance optimality* means that the algorithm produces meshes with up to a fixed constant optimal cardinality relativ to the achieved energy error. Different from the Dörfler marking strategy there is no restriction on the marking parameter μ, i.e., all $\mu \in (0, 1]$ are admissible for *instance optimality.*

Let us briefly analyze how the maximum strategy will perform for the setup of Subection 3.1. It may actually happen in the first iterations that elements in $[1, M + 1]$ are refined. However, these elements are getting then smaller relative to $\eta^2_{\mathcal{T}_k}(T_0(\mathcal{T}_k))$ due to bisection, thanks to the fact that $|T| = 2^{-g(T)}$. Therefore, eventually all elements in $[1, M+1]$ are smaller than $\mu\,\eta^2_{\mathcal{T}_k}(T_0(\mathcal{T}_k))$. From that point on only $T_0(\mathcal{T}_k)$ will be refined and we obtain exponential convergence similar as in Lemma 6.

This confirms the expected perfomance of the maximum strategy.

References

[BD04] P. Binev and R. DeVore, *Fast computation in adaptive tree approximation*, Numer. Math. **97** (2004), no. 2, 193–217.

[BDD04] P. Binev, W. Dahmen, and R. DeVore, *Adaptive finite element methods with convergence rates*, Numer. Math **97** (2004), 219–268.

[BDK12] L. Belenki, L. Diening, and C. Kreuzer, *Optimality of an adaptive finite element method for the p-Laplacian equation*, IMA J. Numer. Anal. **32** (2012), no. 2, 484–510.

[BN10] A. Bonito and R. H. Nochetto, *Quasi-optimal convergence rate of an adaptive discontinuous Galerkin method*, SIAM J. Numer. Anal. **48** (2010), no. 2, 734–771.

[CFPP14] C. Carstensen, M. Feischl, M. Page, and D. Praetorius, *Axioms of adaptivity*, Computers & Mathematics with Applications **67** (2014), no. 6, 1195–1253.

[CKNS08] J. M. Cascón, C. Kreuzer, R. H. Nochetto, and K. G. Siebert, *Quasi-optimal convergence rate for an adaptive finite element method*, SIAM J. Numer. Anal. **46** (2008), no. 5, 2524–2550.

[CN11] J. M. Cascón and R. H. Nochetto, *Quasioptimal cardinality of AFEM driven by nonresidual estimators*, IMA Journal of Numerical Analysis **32** (2011), no. 1, 1–29.

[CPR13] C. Carstensen, D. Peterseim, and H. Rabus, *Optimal adaptive nonconforming FEM for the Stokes problem*, Numer. Math. **123** (2013), no. 2, 291–308.

[CR17] C. Carstensen and H. Rabus, *Axioms of adaptivity with separate marking for data resolution*, SIAM J. Numer. Anal. **55** (2017), no. 6, 2644–2665.

[DK08] L. Diening and C. Kreuzer, *Linear convergence of an adaptive finite element method for the p-Laplacian equation*, SIAM J. Numer. Anal. **46** (2008), no. 2, 614–638.

[DKS16] L. Diening, C. Kreuzer, and R. Stevenson, *Instance optimality of the adaptive maximum strategy*, Found. Comput. Math. **16** (2016), no. 1, 33–68.

[FFP14] M. Feischl, T. Führer, and D. Praetorius, *Adaptive FEM with optimal convergence rates for a certain class of nonsymmetric and possibly nonlinear problems*, SIAM J. Numer. Anal. **52** (2014), no. 2, 601–625.

[FKMP13] M. Feischl, M. Karkulik, J. M. Melenk, and D. Praetorius, *Quasi-optimal convergence rate for an adaptive boundary element method*, SIAM J. Numer. Anal. **51** (2013), no. 2, 1327–1348.

[Gan13] T. Gantumur, *Adaptive boundary element methods with convergence rates*, Numer. Math. **124** (2013), no. 3, 471–516.

[KS11] C. Kreuzer and K. G. Siebert, *Decay rates of adaptive finite elements with Dörfler Marking*, Numer. Math. **117** (2011), no. 4, 679–716.

[KS16] C. Kreuzer and M. Schedensack, *Instance optimal Crouzeix-Raviart adaptive finite element methods for the Poisson and Stokes problems*, IMA J. Numer. Anal. **36** (2016), no. 2, 593–617.

[NSV09] R. H. Nochetto, K. G. Siebert, and A. Veeser, *Theory of adaptive finite element methods: an introduction*, Multiscale, nonlinear and adaptive approximation, Springer, Berlin, 2009, pp. 409–542.

[NV12] R. H. Nochetto and A. Veeser, *Primer of adaptive finite element methods*, Multiscale and adaptivity: modeling, numerics and applications, Lecture Notes in Math., vol. 2040, Springer, Heidelberg, 2012, pp. 125–225.

[Ste07] R. Stevenson, *Optimality of a standard adaptive finite element method*, Found. Comput. Math. **7** (2007), no. 2, 245–269.

Universität Bielefeld, Fakultät für Mathematik, D-33615 Bielefeld, Germany
Email address: `lars.diening@uni-bielefeld.de`

TU Dortmund, Fakultät für Mathematik, D-44221 Dortmund, Germany
Email address: `christian.kreuzer@tu-dortmund.de`