

ERROR CONTROL FOR THE LOCALIZED REDUCED BASIS MULTISCALE METHOD WITH ADAPTIVE ON-LINE ENRICHMENT*

M. OHLBERGER[†] AND F. SCHINDLER[†]

Abstract. In this contribution we consider localized, robust, and efficient a posteriori error estimation of the localized reduced basis multiscale (LRBMS) method for parametric elliptic problems with possibly heterogeneous diffusion coefficient. The numerical treatment of such parametric multiscale problems is characterized by a high computational complexity, arising from the multiscale character of the underlying differential equation and the additional parameter dependence. The LRBMS method can be seen as a combination of numerical multiscale methods and model reduction using reduced basis (RB) methods to efficiently reduce the computational complexity with respect to the multiscale as well as the parametric aspect of the problem, simultaneously. In contrast to the classical residual based error estimators currently used in RB methods, we are considering error estimators that are based on conservative flux reconstruction and provide an efficient and rigorous bound on the full error with respect to the weak solution. In addition, the resulting error estimator is localized and can thus be used in the on-line phase to adaptively enrich the solution space locally where needed. The resulting certified LRBMS method with adaptive on-line enrichment thus guarantees the quality of the reduced solution during the on-line phase, given any (possibly insufficient) reduced basis that was generated during the off-line phase. Numerical experiments are given to demonstrate the applicability of the resulting algorithm with on-line enrichment to single phase flow in heterogeneous media.

Key words. model reduction, multiscale, error analysis, flux reconstruction

AMS subject classifications. 65G99, 65N55, 65N15, 35J20, 65N30, 76S05

DOI. 10.1137/151003660

1. Introduction. We are interested in efficient and reliable numerical approximations of elliptic parametric multiscale problems. Such problems consist of finding $p(\boldsymbol{\mu}) \in Q$, such that

$$(1.1) \quad b(p(\boldsymbol{\mu}), q; \boldsymbol{\mu}) = l(q) \quad \text{for all } q \in Q,$$

in a suitable space Q for parameters $\boldsymbol{\mu} \in \mathcal{P} \subset \mathbb{R}^p$ for $p \in \mathbb{N}$, where the data functions involved may depend on an a priori given multiscale parameter $\varepsilon > 0$ (described in detail in section 2). An approximation $p_h(\boldsymbol{\mu}) \in Q_h$ of $p(\boldsymbol{\mu})$ is usually obtained by a discretization of (1.1) resulting from a Galerkin projection onto a discrete space Q_h , associated with a triangulation τ_h of Ω . Since τ_h should resolve all features of (1.1) associated with the fine scale ε , solving parametric heterogeneous multiscale problems accurately can be challenging and computationally costly, in particular for strongly varying scales and parameter ranges (see the references in [44]). Two traditional approaches exist to reduce the computational complexity of the discrete problem: numerical multiscale methods and model order reduction techniques. Numerical multiscale methods reduce the complexity of multiscale problems with respect to ε for

*Submitted to the journal's Methods and Algorithms for Scientific Computing section January 20, 2015; accepted for publication (in revised form) September 23, 2015; published electronically December 3, 2015.

<http://www.siam.org/journals/sisc/37-6/100366.html>

[†]Applied Mathematics, University of Münster, Einsteinstr. 62, D-48149 Münster, Germany (mario.ohlberger@uni-muenster.de, felix.schindler@uni-muenster.de). The second author's work was supported by the German Federal Ministry of Education and Research (BMBF) under contract 05M13PMA.

a fixed $\boldsymbol{\mu}$, while model order reduction techniques reduce the complexity of parametric problems with respect to $\boldsymbol{\mu}$ for moderate scales ε (see [44] for an overview). In general, numerical multiscale methods capture the macroscopic behavior of the solution in a coarse approximation space $Q_H \subset Q_h$, e.g., associated with a coarse triangulation \mathcal{T}_H of Ω , and recover the microscopic behavior of the solution by local fine-scale corrections. Inserting this additive decomposition into (1.1) yields a coupled system of a fine- and a coarse-scale variational problem. By appropriately selecting trial and test spaces and defining the localization operators to decouple this system, a variety of numerical multiscale methods can be recovered, e.g., the multiscale finite element method (MsFEM) [22], the variational multiscale method [43, 37], the multiscale finite volume method [30], and the heterogeneous multiscale method (HMM) [4, 31], just to name a few. Model order reduction using reduced basis (RB) methods, on the other hand, is based on the idea of introducing a reduced space $Q_{\text{red}} \subset Q_h$, spanned by discrete solutions for a limited number of parameters $\boldsymbol{\mu}$. These training parameters are iteratively selected by an adaptive greedy procedure (see, e.g., [56] and the references therein). Depending on the choice of the training parameters and the nature of the problem, Q_{red} is expected to be of a significantly smaller dimension than Q_h . Additionally, if b allows for an affine decomposition with respect to $\boldsymbol{\mu}$, its components can be projected onto Q_{red} , which can then be used to effectively split the computation into an off-line and an on-line part. In the off-line phase all parameter-independent quantities are precomputed, such that the on-line phase's complexity only depends on the dimension of Q_{red} . The idea of the recently presented localized reduced basis multiscale (LRBMS) approach (see [36, 7]) is to combine numerical multiscale and RB methods and to generate a local reduced space $Q_{\text{red}}^T \subset Q_h^T$ for each coarse element T of \mathcal{T}_H , given a tensor-product-type decomposition of the fine approximation space, $Q_h = \bigoplus_{T \in \mathcal{T}_H} Q_h^T$. The coarse reduced space is then given as $Q_{\text{red}}(\mathcal{T}_H) := \bigoplus_{T \in \mathcal{T}_H} Q_{\text{red}}^T \subset Q_h$, resulting in a multiplicative decomposition of the solution into $p_{\text{red}}(x; \boldsymbol{\mu}) = \sum_{n=1}^{\dim Q_{\text{red}}(\mathcal{T}_H)} p_n(x; \boldsymbol{\mu}) \varphi_n(x)$, where the reduced basis functions φ_n capture the microscopic behavior of the solution associated with the fine scale ε and the coefficient functions p_n only vary on the coarse partition \mathcal{T}_H .

Other model reduction approaches for parametric (but not multi-scale) problems that incorporate localization ideas and concepts from domain decomposition (DD) methods are the RB element method [39, 38], the RB hybrid method [33, 34] and the port reduced static condensation RB element method [52]. While the idea of the former is to share one reduced basis on all subdomains the idea of the latter two is to generate one reduced basis for each class of subdomains which are then coupled appropriately. There also exist several approaches to use model reduction techniques for homogenization problems (see [13]) and problems with multiple scales, such as the RB finite element HMM [2, 3]. For the case of no scale-separation there exist the generalized MsFEM method [20], which incorporates model reduction ideas, and most recently a work combining the RB method with localized orthogonal decomposition (see [5]).

It is vital for an efficient and reliable use of RB as well as LRBMS methods to have access to an estimate on the model reduction error. Such an estimate is used to drive the adaptive greedy basis generation during the off-line phase of the computation and to ensure the quality of the reduced solution during the on-line phase. It is usually given by a residual based estimator involving the stability constant and the residual in a dual norm. It was shown in [7] that such an estimator can be successfully applied in the context of the LRBMS, but it was also pointed out that an estimator relying on global information might not be computationally feasible since too much work is required in the off-line part of the computation.

The novelty of this contribution lies in a completely different approach to error estimation—at least in the context of RB methods. We make use of the ansatz of local error estimation presented in [26] which measures the error by a conforming reconstruction of the physical quantities involved, specifically the diffusive flux $-\lambda(\boldsymbol{\mu})\kappa_\varepsilon\nabla p(\boldsymbol{\mu})$. This kind of local error estimation was proven to be very successful in the context of multiscale problems and very robust with respect to ε . We show in this work how we can transfer those ideas to the framework of the LRBMS to obtain an estimate of the error $\|p(\boldsymbol{\mu}) - p_{\text{red}}(\boldsymbol{\mu})\|_{\boldsymbol{\mu}}$. We would like to point out that we are able to estimate the error against the weak solution $p(\boldsymbol{\mu})$ in a parameter dependent energy norm while traditional RB approaches only allow us to estimate the model reduction error in a parameter-independent norm and only against the discrete solution. In principal, this approach is able to turn the LRBMS method into a full multiscale approximation scheme, while traditional RB methods can only be seen as a model reduction technique. We would also like to point out that, to the best of our knowledge, this approach (first published in [45]) is the first one to make use of local error information in the context of RB methods.

2. Problem formulation and discretization. We consider linear elliptic problems in a bounded connected domain $\Omega \subset \mathbb{R}^d$ for $d = 2, 3$, with polygonal boundary $\partial\Omega$ and a multiplicative splitting of the influences of the multiscale parameter $\varepsilon > 0$ and the parameter $\boldsymbol{\mu}$. An example is given by the problem of finding a global pressure $p(\boldsymbol{\mu}) \in H_0^1(\Omega)$ for a set of admissible parameters $\boldsymbol{\mu} \in \mathcal{P}$, such that

$$(2.1) \quad -\nabla \cdot (\lambda(\boldsymbol{\mu})\kappa_\varepsilon\nabla p(\boldsymbol{\mu})) = f \quad \text{in } \Omega$$

in a weak sense in $H_0^1(\Omega)$, where H^1 denotes the usual Sobolev space of weakly differentiable functions and $H_0^1(\Omega) \subset H^1(\Omega)$ its elements which vanish on the boundary in the sense of traces. Problems of this kind arise for instance in the context of the global pressure formulation of two-phase flow in porous media. Using an implicit pressure explicit saturation (IMPES) discretization scheme of the pressure/saturation system (see [16] and the references therein), an equation of the form (2.1) has to be solved in each time step for a different parameter $\boldsymbol{\mu} \in \mathcal{P}$. In that context $\lambda(\boldsymbol{\mu})$ denotes the scalar total mobility and κ_ε denotes a possibly complex heterogeneous permeability tensor; external forces are collected in f and the parameter $\boldsymbol{\mu}$ models the influence of the global saturation (see Definition 2.1 for details on λ , κ_ε , and f). Note that more complex boundary conditions and additional parameter dependencies of the boundary values as well as the right-hand side (modeling parameter dependent wells, for instance) are possible, but do not lie within the scope of this work.

Triangulation. We require two nested partitions of Ω , a coarse one, \mathcal{T}_H , and a fine one, τ_h . Let τ_h be a simplicial triangulation of Ω with elements $t \in \tau_h$. In the context of multiscale problems we call τ_h a *fine triangulation* if it resolves all features of the quantities involved in (2.1), specifically if $\kappa_\varepsilon|_t \in [L^\infty(t)]^{d \times d}$ is constant for all $t \in \tau_h$. This assumption is a natural one in the context of two-phase flow problems where the permeability field is usually given by piecewise constant measurement data. For simplicity we require τ_h to fulfill the requirements stated in [26, section 2.1], namely, shape regularity and the absence of hanging nodes; an extension to more general triangulations is possible analogously to [26, A.1]. We only require the coarse elements $T \in \mathcal{T}_H$ to be shaped such that a local Poincaré inequality in $H^1(T)$ is fulfilled (see the requirements of Theorem 4.2). We collect all fine faces in \mathcal{F}_h , all coarse faces in \mathcal{F}_H , and denote by $\mathcal{N}(t) \subset \tau_h$ and $\mathcal{N}(T) \subset \mathcal{T}_H$ the neighbors of $t \in \tau_h$ and $T \in \mathcal{T}_H$, respectively, and by h_* the diameter of any element $*$ of the sets τ_h ,

\mathcal{T}_H , \mathcal{F}_h , or \mathcal{F}_H . We collect in $\tau_h^T \subset \tau_h$ the fine elements of τ_h that cover the coarse element T and in $\mathcal{F}_h^* \subset \mathcal{F}_h$ all faces that cover the set $*$, e.g., by \mathcal{F}_h^t the faces of a fine element $t \in \tau_h$, by \mathcal{F}_h^E the faces that cover a coarse face $E \in \mathcal{F}_H$, and so forth; the same notation is used for coarse faces $\mathcal{F}_H^* \subset \mathcal{F}_H$. In addition we denote the set of all boundary faces by $\overline{\mathcal{F}}_h \subset \mathcal{F}_h$ and the set of all inner faces, that share two elements, by $\overset{\circ}{\mathcal{F}}_h \subset \mathcal{F}_h$, such that $\overline{\mathcal{F}}_h \cup \overset{\circ}{\mathcal{F}}_h = \mathcal{F}_h$ and $\overline{\mathcal{F}}_h \cap \overset{\circ}{\mathcal{F}}_h = \emptyset$. We also denote the set of fine faces which lie on the boundary of any coarse element $T \in \mathcal{T}_H$ by $\overline{\mathcal{F}}_h^T := \bigcup_{E \in \mathcal{F}_H^T} \mathcal{F}_h^E$ and by $\overset{\circ}{\mathcal{F}}_h^T := \mathcal{F}_h^T \setminus \overline{\mathcal{F}}_h^T$ the set of fine faces which lie in the interior of the coarse element. Finally, we assign a unit normal $n_e \in \mathbb{R}^d$ to each inner face $\partial t^- \cap \partial t^+ = e \in \overset{\circ}{\mathcal{F}}_h$, pointing from t^- to t^+ , and also denote the unit outward normal to $\partial\Omega$ by n_e for a boundary face $e = \partial t^- \cap \partial\Omega$, for $t^-, t^+ \in \tau_h$.

2.1. The continuous problem. For our analysis we define the *broken Sobolev space* $H^1(\tau_h) \subset L^2(\Omega)$ by $H^1(\tau_h) := \{q \in L^2(\Omega) \mid q|_t \in H^1(t) \ \forall t \in \tau_h\}$ as the most general space for the weak, the discrete, and the reduced solution defined below. In the same manner we denote the local broken Sobolev spaces $H^1(\tau_h^T) \subset L^2(T)$ for all coarse elements $T \in \mathcal{T}_H$ and denote by $\nabla_h : H^1(\tau_h) \rightarrow [L^2(\Omega)]^d$ the *broken gradient operator* which is locally defined by $(\nabla_h q)|_t := \nabla(q|_t)$ for all $t \in \tau_h$ and $q \in H^1(\tau_h)$. Now given $\lambda(\boldsymbol{\mu})$, κ_ε , and f as stated in Definition 2.1 we define the parametric bilinear form $b : \mathcal{P} \rightarrow [H^1(\tau_h) \times H^1(\tau_h) \rightarrow \mathbb{R}]$, $\boldsymbol{\mu} \mapsto [(p, q) \mapsto b(p, q; \boldsymbol{\mu})]$, and the linear form $l : H^1(\tau_h) \rightarrow \mathbb{R}$ by $b(p, q; \boldsymbol{\mu}) := \sum_{T \in \mathcal{T}_H} b^T(p, q; \boldsymbol{\mu})$ and $l(q) := \sum_{T \in \mathcal{T}_H} l^T(q)$, respectively, and their local counterparts $b^T : \mathcal{P} \rightarrow [H^1(\tau_h^T) \times H^1(\tau_h^T) \rightarrow \mathbb{R}]$, $\boldsymbol{\mu} \mapsto [(p, q) \mapsto b^T(p, q; \boldsymbol{\mu})]$, and $l^T : H^1(\tau_h^T) \rightarrow \mathbb{R}$ for all $T \in \mathcal{T}_H$, $\boldsymbol{\mu} \in \mathcal{P}$, and $p, q \in H^1(\tau_h)$ by

$$b^T(p, q; \boldsymbol{\mu}) := \int_T (\lambda(\boldsymbol{\mu})\kappa_\varepsilon \nabla_h p) \cdot \nabla_h q \, dx \quad \text{and} \quad l^T(q) := \int_T f q \, dx,$$

respectively.

DEFINITION 2.1 (weak solution). *Let $f \in L^2(\Omega)$ be bounded and $\kappa_\varepsilon \in [L^\infty(\Omega)]^{d \times d}$ be symmetric and uniformly positive definite. For $\boldsymbol{\mu} \in \mathcal{P}$, let $\lambda(\boldsymbol{\mu}) \in L^\infty(\Omega)$ be strictly positive, such that $\lambda(\boldsymbol{\mu})\kappa_\varepsilon \in [L^\infty(\Omega)]^{d \times d}$ is bounded from below (away from 0) and above. We define the weak solution $p : \mathcal{P} \rightarrow H_0^1(\Omega)$ for a parameter $\boldsymbol{\mu} \in \mathcal{P}$, such that*

$$(2.2) \quad b(p(\boldsymbol{\mu}), q; \boldsymbol{\mu}) = l(q) \quad \text{for all } q \in H_0^1(\Omega).$$

Note that, since $b(\cdot, \cdot; \boldsymbol{\mu})$ is continuous and coercive for all $\boldsymbol{\mu} \in \mathcal{P}$ (due to the assumptions on $\lambda(\boldsymbol{\mu})\kappa_\varepsilon$) and since l is bounded, there exists a unique solution of (2.2) due to the Lax–Milgram theorem.

A note on parameters. In addition to the assumptions we posed on λ above we also demand it to be *affinely decomposable* with respect to $\boldsymbol{\mu} \in \mathcal{P}$, i.e., there exist $\Xi \in \mathbb{N}$ strictly positive *coefficients* $\theta_\xi : \mathcal{P} \rightarrow \mathbb{R}$ and Ξ nonparametric *components* $\lambda_\xi \in L^\infty(\Omega)$ for $1 \leq \xi \leq \Xi$, such that $\lambda(x; \boldsymbol{\mu}) = \sum_{\xi=1}^\Xi \theta_\xi(\boldsymbol{\mu})\lambda_\xi(x)$. We can then compare λ for two parameters $\boldsymbol{\mu}, \overline{\boldsymbol{\mu}} \in \mathcal{P}$ by $\alpha(\boldsymbol{\mu}, \overline{\boldsymbol{\mu}})\lambda(\overline{\boldsymbol{\mu}}) \leq \lambda(\boldsymbol{\mu}) \leq \gamma(\boldsymbol{\mu}, \overline{\boldsymbol{\mu}})\lambda(\overline{\boldsymbol{\mu}})$, where $\alpha(\boldsymbol{\mu}, \overline{\boldsymbol{\mu}}) := \min_{\xi=1}^\Xi \theta_\xi(\boldsymbol{\mu})\theta_\xi(\overline{\boldsymbol{\mu}})^{-1}$ and $\gamma(\boldsymbol{\mu}, \overline{\boldsymbol{\mu}}) := \max_{\xi=1}^\Xi \theta_\xi(\boldsymbol{\mu})\theta_\xi(\overline{\boldsymbol{\mu}})^{-1}$ denote the positive equivalence constants. This assumption on the data function λ is a common one in the context of RB methods and covers a wide range of physical problems. If λ does not exhibit such a decomposition one can replace λ by an arbitrary close approximation using empirical interpolation techniques (see [9, 19]) which does not impact our analysis. All quantities that linearly depend on λ inherit the above affine decomposition in a straightforward way. In particular there exist component bilinear

forms such that b and b^T (and their discrete counterparts introduced further below) are also affinely decomposable.

2.2. Notation. Throughout this work, we denote scalar functions by p, q, s, φ , their discrete counterparts by $p_h, q_h \in Q_h^k$ (compare section 2.3), and their reduced counterparts by $p_{\text{red}}, q_{\text{red}} \in Q_{\text{red}}(\mathcal{T}_H)$ (compare section 3). We denote the space of vector valued functions, the divergence of which lies in $L^2(\Omega)$, by $H_{\text{div}}(\Omega) \subset [L^2(\Omega)]^d$, its elements by $u, v \in H_{\text{div}}(\Omega)$, and their discrete counterparts by $u_h \in V_h^l(\tau_h)$ (compare (4.5)). For any scalar- or vector-valued space $\mathbb{V}(\omega)$, for $\omega \subseteq \Omega$, we denote the associated inner product by $(\cdot, \cdot)_{\mathbb{V}, \omega}$ and omit ω if $\omega = \Omega$; the same holds for the induced norm $\|\cdot\|_{\mathbb{V}, \omega}$. In addition, we use $(\cdot, \cdot)_{L^2, \omega}$ for the L^2 -inner product of scalar- and vector-valued functions.

Additionally, since we would like to estimate the error in a problem dependent norm we introduce the *parameter-dependent energy (semi)norm* $\|\cdot\|_{\mu} : \mathcal{P} \rightarrow [H^1(\tau_h) \rightarrow \mathbb{R}]$, $\mu \rightarrow [q \rightarrow \|q\|_{\mu}]$, defined by $\|q\|_{\mu} := (\sum_{T \in \mathcal{T}_H} \|q\|_{\mu, T}^2)^{1/2}$ with the local seminorm defined by $\|q\|_{\mu, T} := b^T(q, q; \mu)^{1/2}$ for all $T \in \mathcal{T}_H$, $\mu \in \mathcal{P}$, and $q \in H^1(\tau_h)$; the local seminorm on the fine triangulation is defined by $\|q\|_{\mu, t} := b^t(q, q; \mu)^{1/2}$ for any $t \in \tau_h$, with b^t defined analogously to b^T . Note that $\|\cdot\|_{\mu}$ is a norm only on $H_0^1(\Omega)$. We can compare these norms for any two parameters $\mu, \bar{\mu} \in \mathcal{P}$ using the decomposition of λ from above (the same holds true for the local seminorms):

$$(2.3) \quad \sqrt{\alpha(\mu, \bar{\mu})} \|\cdot\|_{\bar{\mu}} \leq \|\cdot\|_{\mu} \leq \sqrt{\gamma(\mu, \bar{\mu})} \|\cdot\|_{\bar{\mu}}.$$

Given the requirements on λ and κ from above, we denote by $0 < c_{\varepsilon}^t(\mu)$ and $c_{\varepsilon}^t(\mu) \leq C_{\varepsilon}^t(\mu)$ the smallest and largest eigenvalue of $(\lambda(\mu)\kappa_{\varepsilon})|_t$, respectively, for any $\mu \in \mathcal{P}$ and $t \in \tau_h$ and additionally define $0 < c_{\varepsilon}^t := \min_{\mu \in \mathcal{P}} c_{\varepsilon}^t(\mu)$ and $c_{\varepsilon}^t < C_{\varepsilon}^t := \max_{\mu \in \mathcal{P}} C_{\varepsilon}^t(\mu)$.

2.3. The discretization. We discretize (2.2) by allowing for a suitable discretization of at least first order inside each coarse element $T \in \mathcal{T}_H$ and by coupling those with a symmetric weighted interior-penalty discontinuous Galerkin (SWIPDG) discretization along the coarse faces of \mathcal{T}_H . This ansatz can be either interpreted as an extension of the SWIPDG discretization on the coarse partition \mathcal{T}_H , where we further refine each coarse element and introduce an additional local discretization, or it can be interpreted as a DD approach, where we use local discretizations, defined on subdomains given by the coarse partition, which are then coupled by the SWIPDG fluxes. In view of the latter, this ansatz shows some similarities to [15] but allows for a wider range of local discretizations. A similar ansatz for a multinumerics discretization using a different coupling strategy was independently developed and recently introduced in [46]. We will present two particular choices for the local discretizations and continue with the definition of the overall discontinuous Galerkin (DG) discretization.

Local discretizations. The main idea is to approximate the local bilinear forms b^T , which are defined on the local subdomain triangulations τ_h^T , by local discrete bilinear forms b_h^T discretizing (2.1) on T with homogeneous Neumann boundary values. We will additionally choose local discrete ansatz spaces $Q_h^{k, T} \subset H^1(\tau_h^T)$, with polynomial order $k \in \mathbb{N}$, to complete the definition of the local discretizations. The first natural choice for the local discretization is to use a standard continuous Galerkin (CG) discretization, which we obtain by setting $b_h^T = b^T$ and $Q_h^{k, T} = S_h^k(\tau_h^T)$ with

$$S_h^k(\tau_h^T) := \{q \in C^0(T) \mid q|_t \in \mathbb{P}_k(t) \quad \forall t \in \tau_h^T\} \subset H^1(T),$$

where $\mathbb{P}_k(\omega)$ denotes the set of polynomials on $\omega \subseteq \Omega$ with total degree at most $k \in \mathbb{N}$. Another choice is to use a discontinuous space $Q_h^{k,T} = Q_h^k(\tau_h^T)$, given by

$$Q_h^k(\tau_h^T) := \{q \in L^2(T) \mid q|_t \in \mathbb{P}_k(t) \quad \forall t \in \tau_h^T\} \subset H^1(\tau_h^T),$$

and to choose b_h^T from a family of DG discretizations. Therefore we introduce the technicalities needed to state a common framework for the SWIPDG and the nonsymmetric, the incomplete, and the symmetric interior-penalty DG discretization (henceforth denoted by NIPDG, IIPDG, and SIPDG, respectively; see [27] and the references therein), following [26, section 2.3].

For a function $q \in H^1(\tau_h)$ that is double valued on interior faces we denote its jump on an inner face $e \in \mathring{\mathcal{F}}_h$ by $\llbracket q \rrbracket_e := q^- - q^+$ with $q^\pm := q|_{t^\pm}$, recalling that $e = t^- \cap t^+$ for $t^\pm \in \tau_h$. We also assign weights $\omega_e^-, \omega_e^+ > 0$ to each inner face, such that $\omega_e^- + \omega_e^+ = 1$, and denote the weighted average of q by $\{\{q\}\}_e := \omega_e^- q^- + \omega_e^+ q^+$. On a boundary face $e \in \mathcal{F}_h$ we set $\omega_e^- = 1, \omega_e^+ = 0, \llbracket q \rrbracket_e := q$, and $\{\{q\}\}_e := q$. With these definitions we define the local bilinear form $b_h^T : \mathcal{P} \rightarrow [H^1(\tau_h^T) \times H^1(\tau_h^T) \rightarrow \mathbb{R}], \boldsymbol{\mu} \mapsto [(p, q) \mapsto b_h^T(p, q; \boldsymbol{\mu})]$ for $\vartheta \in \{-1, 0, 1\}$ as specified below, by

$$(2.4) \quad b_h^T(p, q; \boldsymbol{\mu}) := b^T(p, q; \boldsymbol{\mu}) + \sum_{e \in \mathring{\mathcal{F}}_h^T} \left(\vartheta b_c^e(q, p; \boldsymbol{\mu}) + b_c^e(p, q; \boldsymbol{\mu}) + b_p^e(q, p; \boldsymbol{\mu}) \right)$$

on $T \in \mathcal{T}_H$, with its coupling and penalty parts b_c^e and b_p^e , respectively, defined by

$$b_c^e(p, q; \boldsymbol{\mu}) := \int_e - \{ \{ (\lambda(\boldsymbol{\mu}) \kappa_\varepsilon \nabla p) \cdot n_e \} \}_e \llbracket q \rrbracket_e \, ds \quad \text{and} \quad b_p^e(p, q; \boldsymbol{\mu}) := \int_e \sigma_e(\boldsymbol{\mu}) \llbracket p \rrbracket_e \llbracket q \rrbracket_e \, ds$$

for all $\boldsymbol{\mu} \in \mathcal{P}$ all $p, q \in H^1(\tau_h)$, and all $e \in \mathcal{F}_h$. The parametric positive penalty function $\sigma_e(\boldsymbol{\mu}) : \mathcal{P} \rightarrow \mathbb{R}$ is given by $\sigma_e(\boldsymbol{\mu}) := \sigma h_e^{-1} \{ \{ \lambda(\boldsymbol{\mu}) \} \}_e \sigma_\varepsilon^e$, where $\sigma \geq 1$ denotes a user-dependent parameter and the locally adaptive weight is given by $\sigma_\varepsilon^e := \delta_e^+ \delta_e^- (\delta_e^+ + \delta_e^-)^{-1}$ for an interior face $e \in \mathring{\mathcal{F}}_h$ and by $\sigma_\varepsilon^e := \delta_e^-$ on a boundary face $e \in \mathcal{F}_h$, respectively, with $\delta_e^\pm := n_e \cdot \kappa_\varepsilon^\pm \cdot n_e$. Using the weights $\omega_e^\pm := 1/2$ we obtain the NIPDG bilinear form for $\vartheta = -1$, the IIPDG bilinear form for $\vartheta = 0$, and the SIPDG bilinear form for $\vartheta = 1$. We obtain the SWIPDG bilinear form for $\vartheta = 1$ by using locally adaptive weights $\omega_e^- := \delta_e^+ (\delta_e^+ + \delta_e^-)^{-1}$ and $\omega_e^+ := \delta_e^- (\delta_e^+ + \delta_e^-)^{-1}$. From now on we assume that b_h^T is of the form (2.4), since this is the most general case and also covers a CG discretization, where all face terms vanish due to the nature of $S_h^k(\tau_h^T)$.

Global coupling. Now given suitable local discretizations on the coarse elements $T \in \mathcal{T}_H$ we couple those along the coarse faces $E \in \mathcal{F}_H$ using a SWIPDG discretization again and define the bilinear form $b_h : \mathcal{P} \rightarrow [H^1(\tau_h) \times H^1(\tau_h) \rightarrow \mathbb{R}], \boldsymbol{\mu} \mapsto [(p, q) \mapsto b_h(p, q; \boldsymbol{\mu})]$, by

$$(2.5) \quad b_h(p, q; \boldsymbol{\mu}) := \sum_{T \in \mathcal{T}_H} b_h^T(p, q; \boldsymbol{\mu}) + \sum_{E \in \mathcal{F}_H} b_h^E(p, q; \boldsymbol{\mu}),$$

where we use the SWIPDG variants of ω_e^\pm to define the coupling bilinear form $b_h^E : \mathcal{P} \rightarrow [H^1(\tau_h) \times H^1(\tau_h) \rightarrow \mathbb{R}], \boldsymbol{\mu} \mapsto [(p, q) \mapsto b_h^E(p, q; \boldsymbol{\mu})]$, by

$$(2.6) \quad b_h^E(p, q; \boldsymbol{\mu}) := \sum_{e \in \mathcal{F}_h^E} \left(b_c^e(q, p; \boldsymbol{\mu}) + b_c^e(p, q; \boldsymbol{\mu}) + b_p^e(q, p; \boldsymbol{\mu}) \right)$$

for all $E \in \mathcal{F}_H$, all $\boldsymbol{\mu} \in \mathcal{P}$, and all $p, q \in H^1(\tau_h)$. Accordingly, we define the DG space $Q_h^k \subset H^1(\tau_h)$ for $k \geq 1$ by

$$Q_h^k := \{q \in H^1(\tau_h) \mid q|_T \in Q_h^{k,T} \quad \forall T \in \mathcal{T}_H\}$$

with $Q_h^{k,T}$ either being the local CG space $S_h^k(\tau_h^T)$ or the local DG space $Q_h^k(\tau_h^T)$.

DEFINITION 2.2 (DG solution). *Let σ be large enough, such that $b_h(\cdot, \cdot; \boldsymbol{\mu})$ is continuous and coercive for all $\boldsymbol{\mu} \in \mathcal{P}$. We define the DG solution $p_h : \mathcal{P} \rightarrow Q_h^1$ for a parameter $\boldsymbol{\mu} \in \mathcal{P}$, such that*

$$(2.7) \quad b_h(p_h(\boldsymbol{\mu}), q_h; \boldsymbol{\mu}) = l(q_h) \quad \text{for all } q_h \in Q_h^1.$$

As always with DG methods, coercivity is to be understood with respect to a DG norm on Q_h^1 (i.e., given by the semi H^1 norm combined with a DG jump norm) and there exists a unique solution of (2.7) due to the Lax–Milgram theorem, if σ is chosen accordingly.

Remark 2.3 (properties of the discretization). Depending on the choice of the coarse partition and the local discretization we can recover several discretizations from (2.5). Choosing $\mathcal{T}_H = \Omega$ and $Q_h^{k,T} = S_h^k(\tau_h^T)$ yields a standard CG discretization (except for the treatment of the boundary values), for instance. Choosing $\mathcal{T}_H = \tau_h$, on the other hand, results in the standard SWIPDG discretization proposed in [27, 26]. Note that the local discretizations as well as the polynomial degree k do not have to coincide on each coarse element (or even inside a coarse element when using a local DG space). It is thus possible to balance the computational effort by choosing local CG or k -adaptive DG discretizations. This puts our discretization close to the multinumerics discretization proposed in [46], where the latter allows for an even wider range of local discretizations while coupling along the coarse faces using mortar methods. Concerning the choice of the user-dependent penalty factor, we found an automated choice of σ depending on the polynomial degree k , as proposed in [23], to work very well.

3. Model reduction. Disregarding the multiscale parameter ε for the moment, model reduction using RB methods is a well-established technique to reduce the computational complexity of (2.7) with respect to the parameter $\boldsymbol{\mu}$. It has been successfully applied in multiquery contexts, where (2.7) has to be solved for many parameters $\boldsymbol{\mu}$ (e.g., in the context of optimization), and in real-time contexts, where a solution of (2.7) (or a quantity depending on it) has to be available for some $\boldsymbol{\mu}$ as fast as possible (see [7, 19, 44] and the references therein). The general idea of RB methods is to span a reduced space $Q_{\text{red}} \subset Q_h^k$ by postprocessed solutions of (2.7) for a selection of parameters and to apply a Galerkin projection of (2.7) onto this reduced space. Together with the assumption we posed on the data function in section 2 this allows for the well-known off-line/on-line decomposition of the RB method, where all quantities depending on Q_h^k (in particular high-dimensional evaluations of the bilinear form) can be precomputed and stored in so called Gramian-matrices during a possibly computationally expensive off-line phase. In the on-line phase of the simulation only low-dimensional quantities depending on the parameter dependency of the data functions and the dimension of Q_{red} need to be computed which can usually be done in real time and even on low end devices such as smartphones or embedded devices.

It has been shown in [7], however, that the computational complexity of the off-line phase can become unbearable if the RB method is applied to parametric *multiscale* problems, such as (2.7) for small scales $\varepsilon \ll 1$. There are two main drawbacks

of classical RB methods in this context: (i) all high-dimensional quantities (solution snapshots, operator matrices, ...) usually depend on the fine triangulation τ_h , the size of which scales with $\mathcal{O}(1/\varepsilon)$. Correspondingly all evaluations involving these quantities (scalar products, operator inversions, ...) become increasingly expensive for small scales ε ; (ii) the generation of the reduced space Q_{red} usually requires the discrete problem (2.7) to be solved for several parameters $\boldsymbol{\mu}$. In the context of multiscale problems, however, it is usually only feasible to carry out very few global solutions of the original problem on the fine triangulation, if at all.

The first of these shortcomings has been successfully addressed by the LRBMS method introduced in [36, 7], which takes advantage of the coarse partition and the underlying discretization to form local reduced spaces $Q_{\text{red}}^T \subset Q_h^{k,T}$ on each coarse element instead of the usual global RB space. This allows us to balance the computational effort of the off-line versus the on-line phase by selecting an appropriate size of the coarse partition \mathcal{T}_H depending on the multiscale parameter ε and the real-time or multiquery context of the application. But it was also shown in [7], that the standard residual based estimator which is usually used in the context of RB methods does not scale well in the context of multiscale problems. We thus finalize the definition of the LRBMS method in this contribution with an efficient off-line/on-line decomposable a posteriori error estimator (see the next section). The second shortcoming of classical RB methods in the context of multiscale problems is addressed in section 5, where we propose an on-line enrichment extension for the LRBMS method based on the presented error estimator, which finally turns the LRBMS method into a fully capable multiscale method.

The reduced problem. Let us assume that we are given a local reduced space $Q_{\text{red}}^T \subset Q_h^{k,T}$ on each coarse element $T \in \mathcal{T}_H$. Those are usually, but not necessarily, spanned by solutions of (2.7) and of low dimension, i.e., $\dim Q_{\text{red}}^T \ll \dim Q_h^{k,T}$. We then define the *coarse reduced space* $Q_{\text{red}}(\mathcal{T}_H) \subset Q_h^k$ by

$$Q_{\text{red}}(\mathcal{T}_H) := \{q \in Q_h^k \mid q|_T \in Q_{\text{red}}^T \quad \forall T \in \mathcal{T}_H\} \subset Q_h^k,$$

and obtain the reduced solution by a Galerkin projection of (2.7) onto $Q_{\text{red}}(\mathcal{T}_H)$.

DEFINITION 3.1 (reduced solution). *We define the reduced solution $p_{\text{red}} : \mathcal{P} \rightarrow Q_{\text{red}}(\mathcal{T}_H)$ for a parameter $\boldsymbol{\mu} \in \mathcal{P}$, such that*

$$(3.1) \quad b_h(p_{\text{red}}(\boldsymbol{\mu}), q_{\text{red}}; \boldsymbol{\mu}) = l(q_{\text{red}}) \quad \text{for all } q_{\text{red}} \in Q_{\text{red}}(\mathcal{T}_H).$$

Off-line/on-line decomposition. The well-known off-line/on-line decomposition of RB methods relies on an affine parameter dependence of the data functions (see section 2) that carries over to the discrete bilinear form, i.e., there exist Ξ non-parametric component bilinear forms $b_{h,\xi} : H^1(\tau_h) \times H^1(\tau_h) \rightarrow \mathbb{R}$, for $1 \leq \xi \leq \Xi$, such that $b_h(p, q; \boldsymbol{\mu}) = \sum_{\xi=1}^{\Xi} \theta_{\xi}(\boldsymbol{\mu}) b_{h,\xi}(p, q)$ with θ_{ξ} being the coefficients of the corresponding decomposition of λ . In the standard RB framework, given a basis $\{q_{\text{red},1}, \dots, q_{\text{red},N}\}$ of Q_{red} with $N := \dim Q_{\text{red}}$, one would precompute dense Gramian matrices $\underline{b}_{h,\xi} \in \mathbb{R}^{N \times N}$, given by $(\underline{b}_{h,\xi})_{i,j} := b_{h,\xi}(q_{\text{red},i}, q_{\text{red},j})$ during the off-line phase. In the on-line phase, given a parameter $\boldsymbol{\mu} \in \mathcal{P}$, one would obtain the reduced system matrix by a low-dimensional summation $\underline{b}_h(\boldsymbol{\mu}) := \sum_{\xi=1}^{\Xi} \theta_{\xi}(\boldsymbol{\mu}) \underline{b}_{\xi,h} \in \mathbb{R}^{N \times N}$, which only involves evaluations of the scalar coefficient functions θ_{ξ} for the current parameter. The resulting dense reduced system is of size $N \times N$ and does not depend on the dimension of Q_h^k . It can thus be solved with a complexity of $\mathcal{O}(N^3)$ in the on-line phase. Denoting by $N_h := \dim Q_h^k$ and $N_h^T := \dim Q_h^k(\tau_h^T)$ (with $N_h = \sum_{T \in \mathcal{T}_H} N_h^T$)

the number of degrees of freedom of the global and the local spaces, respectively, the reduction of b_h during the off-line phase would be of a computational complexity of $\mathcal{O}(\Xi N^2 N_h)$, which can become too expensive in the context of small scales ε .

The flexible framework of the LRBMS method, however, allows for a more local approach. Since the affine decomposition of the data function λ also carries over to the local bilinear forms b_h^T and the coupling bilinear forms b_h^E , we define local Gramian matrices $\underline{b}_{h,\xi}^T \in \mathbb{R}^{N^T \times N^T}$ on each coarse element $T \in \mathcal{T}_H$, given by

$$\left(\underline{b}_{h,\xi}^T\right)_{i,j} := b_{h,\xi}^T(q_{\text{red},i}^T, q_{\text{red},j}^T) + \sum_{E \in \mathcal{F}_H^T} b_{h,\xi}^E(q_{\text{red},i}^T, q_{\text{red},j}^T),$$

where $\{q_{\text{red},1}^T, \dots, q_{\text{red},N^T}^T\}$ denotes a basis of Q_{red}^T with $N^T := \dim Q_{\text{red}}^T$ (where a basis function $q_{\text{red},i}^T$ is implicitly extended by 0 outside of T , if required). In addition, we define coupling Gramian matrices $\underline{b}_{h,\xi}^{T,S} \in \mathbb{R}^{N^T \times N^S}$ for all coarse elements $T \in \mathcal{T}_H$ and their neighbors $S \in \mathcal{N}(T)$ by

$$\left(\underline{b}_{h,\xi}^{T,S}\right)_{i,j} := \sum_{E \in \mathcal{F}_H^T \cap \mathcal{F}_H^S} b_{h,\xi}^E(q_{\text{red},i}^T, q_{\text{red},j}^S).$$

In the same manner, we define local vectors $\underline{l}^T \in \mathbb{R}^{N^T}$, given by $(\underline{l}^T)_j := l^T(q_{\text{red},j}^T)$, for all $T \in \mathcal{T}_H$. The global Gramian matrices $\underline{b}_{h,\xi} \in \mathbb{R}^{N \times N}$ and the global vector $\underline{l} \in \mathbb{R}^N$, with $N := \sum_{T \in \mathcal{T}_H} N^T$, are then given by a standard DG mapping (with respect to the coarse triangulation) of their local counterparts. The complexity to directly solve the reduced system in the on-line phase, $\mathcal{O}((\sum_{T \in \mathcal{T}_H} N^T)^3)$, is by definition larger than for standard RB methods. The reduction of b_h can be carried out with a complexity of $\mathcal{O}(\Xi \sum_{T \in \mathcal{T}_H} N^{T^2} N_h^T)$, which, however, can be computed significantly faster than in the standard RB setting (depending on the choice of \mathcal{T}_H), since $\Xi |\mathcal{T}_H|$ local computations can be carried out in parallel.

Remark 3.2 (properties of the reduced system). The reduced system matrix of the LRBMS method is obviously larger than the one we would obtain for standard RB methods (given that it scales with $|\mathcal{T}_H|$). On the other hand, we obtain a sparse matrix of dense blocks (stemming from the local and coupling blocks and the DG mapping) that is usually much smaller than the system matrix of the high-dimensional problem. Thus, if $|\mathcal{T}_H|$ is large, the reduced system can also be solved using sparse iterative solvers with a complexity of $\mathcal{O}((\sum_{T \in \mathcal{T}_H} N^T)^2)$.

4. Error analysis. Our error analysis is a generalization of the ansatz presented in [26] to provide an estimator for our DG approximation (2.7) as well as for our LRBMS approximation (3.1). The main idea of the error estimator presented in [57, 26] is to observe that the approximate DG diffusive flux $-\lambda(\boldsymbol{\mu}) \kappa_\varepsilon \nabla_h p_h(\boldsymbol{\mu})$ is non-conforming while its exact counterpart belongs to $H_{\text{div}}(\Omega)$. The idea of [57, 26] is to reconstruct the discrete diffusive flux in a conforming Raviart–Thomas–Nédélec space $V_h^l(\tau_h) \subset H_{\text{div}}(\Omega)$ and compare it to the nonconforming one. Their error analysis relies on a local conservation property of the reconstructed flux on the fine triangulation τ_h to prove estimates local to the fine triangulation.

We transfer this concept to the DG discretization defined in section 2.3 and prove estimates local to the coarse triangulation that are valid for the DG approximation

as well as for our LRBMS approximation. We obtain mild requirements for the coarse triangulation and the local approximation spaces, namely, that a local Poincaré inequality holds on each coarse element and that the constant function $\mathbb{1}$ is present in the local approximation spaces. The latter is obvious for traditional discretizations and can be easily achieved for the LRBMS approximation. The estimates are fully off-line/on-line decomposable and can thus be efficiently used for model reduction in the context of the LRBMS. Preliminary results were published in [45].

We begin by stating an *abstract energy norm estimate* (see [26, Lemma 4.1]) that splits the difference between the weak solution $p \in H_0^1(\Omega)$ of (2.2) and any function $p_h \in H^1(\tau_h)$ into two contributions. This abstract estimate does not depend on the discretization and thus leaves the choice of s and v open. Note that we formulate the following lemma with different parameters for the energy norm and the weak solution. The price we have to pay for this flexibility is the additional constants involving $\alpha(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})$ and $\gamma(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})$, that are equal to 1 if $\bar{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}$ coincide.

LEMMA 4.1 (abstract energy norm estimate). *Let $p(\boldsymbol{\mu}) \in H_0^1(\Omega)$ be the weak solution of (2.2) for $\boldsymbol{\mu} \in \mathcal{P}$ and let $p_h \in H^1(\tau_h)$ and $\bar{\boldsymbol{\mu}} \in \mathcal{P}$ be arbitrary. Then*

$$\begin{aligned}
 \|p(\boldsymbol{\mu}) - p_h\|_{\bar{\boldsymbol{\mu}}} &\leq \frac{1}{\sqrt{\alpha(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})}} \left(\inf_{s \in H_0^1(\Omega)} \sqrt{\gamma(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})} \|p_h - s\|_{\bar{\boldsymbol{\mu}}} \right. \\
 &\quad \left. + \inf_{v \in H_{div}(\Omega)} \left\{ \sup_{\substack{\varphi \in H_0^1(\Omega) \\ \|\varphi\|_{\boldsymbol{\mu}}=1}} \left\{ (f - \nabla \cdot v, \varphi)_{L^2} \right. \right. \right. \\
 &\quad \left. \left. \left. - (\lambda(\boldsymbol{\mu})\kappa_\varepsilon \cdot \nabla_h p_h + v, \nabla \varphi)_{L^2} \right\} \right\} \right) \\
 (4.1) \qquad &\leq \frac{\sqrt{\gamma(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})}}{\sqrt{\alpha(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})}} 2 \|p(\boldsymbol{\mu}) - p_h\|_{\bar{\boldsymbol{\mu}}}.
 \end{aligned}$$

Proof. We mainly follow the proof of [26, Lemma 4.1] while accounting for the parameter dependency of the energy norm and the weak solution. It holds for arbitrary $\boldsymbol{\mu} \in \mathcal{P}$, $p \in H_0^1(\Omega)$, and $p_h \in H^1(\tau_h)$, that

$$(4.2) \qquad \|p - p_h\|_{\boldsymbol{\mu}} \leq \inf_{s \in H_0^1(\Omega)} \|p_h - s\|_{\boldsymbol{\mu}} + \sup_{\substack{\varphi \in H_0^1(\Omega) \\ \|\varphi\|_{\boldsymbol{\mu}}=1}} b(p - p_h, \varphi, \boldsymbol{\mu})$$

(see [57, Lemma 7.1]) and for the weak solution $p(\boldsymbol{\mu}) \in H_0^1(\Omega)$ of (2.2), that

$$\begin{aligned}
 b(p(\boldsymbol{\mu}) - p_h, \varphi; \boldsymbol{\mu}) &= (f, \varphi)_{L^2} - (\lambda(\boldsymbol{\mu})\kappa_\varepsilon \nabla_h p_h, \nabla \varphi)_{L^2}, \\
 (4.3) \qquad &= (f - \nabla \cdot v, \varphi)_{L^2} - (\lambda(\boldsymbol{\mu})\kappa_\varepsilon \cdot \nabla_h p_h + v, \nabla \varphi)_{L^2}
 \end{aligned}$$

for all $\varphi \in H_0^1(\Omega)$ and all $v \in H_{div}(\Omega)$, where we used the definition of b in the first equality and the fact that $(v, \nabla \varphi)_{L^2} = -(\nabla \cdot v, \varphi)_{L^2}$ due to Green's theorem and $\varphi \in H_0^1(\Omega)$ in the second one. Inserting (4.3) into (4.2) with $p = p(\boldsymbol{\mu})$ and using the norm equivalence (2.3) then yields the first inequality in (4.1).

To obtain the second inequality we choose $s = p(\boldsymbol{\mu})$ and $v = -\lambda(\boldsymbol{\mu})\kappa_\varepsilon \nabla p(\boldsymbol{\mu})$ in the right-hand side of (4.1) which eliminates the two infimums and leaves us with two

terms yet to be estimated arising inside the supremum. Using Green’s theorem and the definition of b we observe the vanishing of the first term. We estimate the second term as

$$\begin{aligned} & |(\lambda(\boldsymbol{\mu})\kappa_\varepsilon \nabla_h p_h - \lambda(\boldsymbol{\mu})\kappa_\varepsilon \nabla p(\boldsymbol{\mu}), \nabla \varphi)_{L^2}| \\ &= |((\lambda(\boldsymbol{\mu})\kappa_\varepsilon)^{1/2} \nabla_h(p_h - p(\boldsymbol{\mu})), (\lambda(\boldsymbol{\mu})\kappa_\varepsilon)^{1/2} \nabla \varphi)_{L^2}| \\ &\leq \left\| (\lambda(\boldsymbol{\mu})\kappa_\varepsilon)^{1/2} \nabla_h(p_h - p(\boldsymbol{\mu})) \right\|_{L^2} \left\| (\lambda(\boldsymbol{\mu})\kappa_\varepsilon)^{1/2} \nabla \varphi \right\|_{L^2} \\ &= \|p_h - p(\boldsymbol{\mu})\|_{\boldsymbol{\mu}} \|\varphi\|_{\boldsymbol{\mu}}, \end{aligned}$$

where we used the Cauchy–Schwarz inequality and the definition of the energy norm. We finally obtain the second inequality of (4.1) from the bound above by observing that the supremum vanishes (due to $\|\varphi\|_{\boldsymbol{\mu}} = 1$) and by using the norm equivalence (2.3) again. \square

The following theorem states the main localization result and gives an indication on how to proceed with the choice of v : it allows us to localize the estimate of the above lemma, if v fulfills a local conservation property. It is still an abstract estimate in the sense that it does not use any information of the discretization and does not yet fully prescribe s and v . We also introduce the additional flexibility to choose a fixed parameter $\hat{\boldsymbol{\mu}}$, which is required for an off-line/on-line decomposition of the diffusive flux estimate (not shown).

THEOREM 4.2 (locally computable abstract energy norm estimate). *Let $p(\boldsymbol{\mu}) \in H_0^1(\Omega)$ be the weak solution of (2.2) for $\boldsymbol{\mu} \in \mathcal{P}$, let $s \in H_0^1(\Omega)$ and $p_h \in H^1(\tau_h)$ be arbitrary, let $v \in H_{div}(\Omega)$ fulfill the local conservation property $(\nabla \cdot v, \mathbb{1})_{L^2, T} = (f, \mathbb{1})_{L^2, T}$ and let $C_P^T > 0$ denote the constant from the Poincaré inequality $\|\varphi - \Pi_0^T \varphi\|_{L^2, T}^2 \leq C_P^T h_T^2 \|\nabla \varphi\|_{L^2, T}^2$ for all $\varphi \in H^1(T)$ on all $T \in \mathcal{T}_H$, where Π_l^ω denotes the L^2 -orthogonal projection onto $\mathbb{P}_l(\omega)$ for $l \in \mathbb{N}$ and $\omega \subseteq \Omega$. It then holds for arbitrary $\bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}} \in \mathcal{P}$, that*

$$\|p(\boldsymbol{\mu}) - p_h\|_{\bar{\boldsymbol{\mu}}} \leq \tilde{\eta}(p_h, s, v; \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}})$$

with the abstract global estimator $\tilde{\eta}(p_h, s, v; \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}})$ defined as

$$\begin{aligned} \tilde{\eta}(p_h, s, v; \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}}) := & \frac{1}{\sqrt{\alpha(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})}} \left[\sqrt{\gamma(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})} \left(\sum_{T \in \mathcal{T}_H} \tilde{\eta}_{\text{nc}}^T(p_h, s; \bar{\boldsymbol{\mu}})^2 \right)^{1/2} + \left(\sum_{T \in \mathcal{T}_H} \tilde{\eta}_r^T(v)^2 \right)^{1/2} \right. \\ & \left. + \frac{1}{\sqrt{\alpha(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})}} \left(\sum_{T \in \mathcal{T}_H} \tilde{\eta}_{\text{df}}^T(p_h, v; \hat{\boldsymbol{\mu}})^2 \right)^{1/2} \right] \end{aligned}$$

and the local nonconformity estimator defined as $\tilde{\eta}_{\text{nc}}^T(p_h, s; \bar{\boldsymbol{\mu}}) := \|p_h - s\|_{\bar{\boldsymbol{\mu}}, T}$, the local residual estimator defined as $\tilde{\eta}_r^T(v) := (C_P^T/c_\varepsilon^T)^{1/2} h_T \|f - \nabla \cdot v\|_{L^2, T}$, and the local diffusive flux estimator defined as $\tilde{\eta}_{\text{df}}^T(p_h, v; \hat{\boldsymbol{\mu}}) := \|(\lambda(\hat{\boldsymbol{\mu}})\kappa_\varepsilon)^{-1/2} (\lambda(\boldsymbol{\mu})\kappa_\varepsilon \nabla_h p_h + v)\|_{L^2, T}$ for all coarse elements $T \in \mathcal{T}_H$, where $c_\varepsilon^T := \min_{t \in \tau_h^T} c_\varepsilon^t$.

Proof. We loosely follow the proof of [26, Theorem 3.1] while accounting for the parameter dependency and the coarse triangulation. Fixing an arbitrary $s \in H_0^1(\Omega)$

in (4.1) and localizing with respect to the coarse triangulation yields

$$(4.4) \quad \|p(\boldsymbol{\mu}) - p_h\|_{\overline{\boldsymbol{\mu}}} \leq \frac{1}{\sqrt{\alpha(\boldsymbol{\mu}, \overline{\boldsymbol{\mu}})}} \left(\sqrt{\gamma(\boldsymbol{\mu}, \overline{\boldsymbol{\mu}})} \sqrt{\sum_{T \in \mathcal{T}_H} \|p_h - s\|_{\boldsymbol{\mu}, T}^2} \right. \\ \left. + \sup_{\substack{\varphi \in H_0^1(\Omega) \\ \|\varphi\|_{\boldsymbol{\mu}} = 1}} \left\{ \sum_{T \in \mathcal{T}_H} \underbrace{\left((f - \nabla \cdot v, \varphi)_{L^2, T} \right)}_{:= (i)} \right. \right. \\ \left. \left. - \underbrace{\left(\lambda(\boldsymbol{\mu}) \kappa_\varepsilon \cdot \nabla_h p_h + v, \nabla \varphi \right)_{L^2, T}}_{:= (ii)} \right\} \right)$$

which leaves us with two local terms we will estimate separately.

- (i) Since $(f - \nabla \cdot v, \Pi_0^T \varphi)_{L^2, T} = 0$ due to the local conservation property of v we can estimate the first term as

$$|(f - \nabla \cdot v, \varphi)_{L^2, T}| \leq \|f - \nabla \cdot v\|_{L^2, T} \|\varphi - \Pi_0^T \varphi\|_{L^2, T} \\ \leq \sqrt{C_P^T} h_T \left(\max_{t \in \tau_h^T} \frac{1}{c_\varepsilon^t} \right)^{1/2} \|f - \nabla \cdot v\|_{L^2, T} \|\varphi\|_{\boldsymbol{\mu}, T},$$

where we used the Cauchy–Schwarz inequality, the Poincaré inequality, and an eigenvalue estimate of $\lambda(\boldsymbol{\mu}) \kappa_\varepsilon|_t$ on all $t \in \tau_h^T$.

- (ii) We estimate the second term as

$$|(\lambda(\boldsymbol{\mu}) \kappa_\varepsilon \nabla_h p_h + v, \nabla \varphi)_{L^2, T}| \\ \leq \|(\lambda(\boldsymbol{\mu}) \kappa_\varepsilon)^{-1/2} (\lambda(\boldsymbol{\mu}) \kappa_\varepsilon \nabla_h p_h + v)\|_{L^2, T} \|\varphi\|_{\boldsymbol{\mu}, T} \\ \leq \sqrt{\alpha(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})}^{-1} \|(\lambda(\hat{\boldsymbol{\mu}}) \kappa_\varepsilon)^{-1/2} (\lambda(\boldsymbol{\mu}) \kappa_\varepsilon \nabla_h p_h + v)\|_{L^2, T} \|\varphi\|_{\boldsymbol{\mu}, T}$$

using the Cauchy–Schwarz inequality, the definition of the local energy seminorms, and the parameter equivalency from section 2.

Inserting the last two inequalities in (4.4) and using the Cauchy–Schwarz inequality and the definition of the local estimators and of c_ε^T yields

$$\|p(\boldsymbol{\mu}) - p_h\|_{\overline{\boldsymbol{\mu}}} \\ \leq \frac{1}{\sqrt{\alpha(\boldsymbol{\mu}, \overline{\boldsymbol{\mu}})}} \left(\sqrt{\gamma(\boldsymbol{\mu}, \overline{\boldsymbol{\mu}})} \left(\sum_{T \in \mathcal{T}_H} \tilde{\eta}_{\text{nc}}^T(p_h, s; \overline{\boldsymbol{\mu}})^2 \right)^{1/2} \right. \\ \left. + \sup_{\substack{\varphi \in H_0^1(\Omega) \\ \|\varphi\|_{\boldsymbol{\mu}} = 1}} \left[\underbrace{\sum_{T \in \mathcal{T}_H} \left(\tilde{\eta}_r^T(v) + \frac{1}{\sqrt{\alpha(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})}} \tilde{\eta}_{\text{df}}^T(p_h, v; \hat{\boldsymbol{\mu}}) \right) \|\varphi\|_{\boldsymbol{\mu}, T}}_{(iii)} \right]^2 \right).$$

Using the Cauchy–Schwarz inequality again we can further estimate (iii) as

$$(iii) \leq \left[\left(\sum_{T \in \mathcal{T}_H} \tilde{\eta}_r^T(v)^2 \right)^{1/2} + \frac{1}{\sqrt{\alpha(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})}} \left(\sum_{T \in \mathcal{T}_H} \tilde{\eta}_{\text{df}}^T(p_h, v; \hat{\boldsymbol{\mu}})^2 \right)^{1/2} \right] \|\varphi\|_{\boldsymbol{\mu}}.$$

The previous two inequalities combined give the final result, since the supremum vanishes due to $\|\varphi\|_{\mu} = 1$. \square

Remark 4.3 (properties of the locally computable abstract energy norm estimate). In contrast to the estimator proposed in [26] the above estimate is local with respect to \mathcal{T}_H , not τ_h . Choosing $\mathcal{T}_H = \tau_h$ we obtain nearly the same estimate as the one in [26] for the pure diffusion case (apart from a slightly less favorable summation). In general, however, we can only expect $\tilde{\eta}_r$ to be superconvergent if we refine \mathcal{T}_H along with τ_h (see section 6.1), thus keeping the ratio H/h fixed.

What is left now in order to turn the abstract estimate of Theorem 4.2 into a fully computable one is to specify s and v , given a DG solution p_h . We will do so in the following paragraphs, finally using the knowledge that p_h was computed using our DG discretization.

Oswald interpolation. The form of the nonconformity indicator in Theorem 4.2 already indicates how to choose s : it should be close to p_h , in order to minimize $\tilde{\eta}_{nc}$, and it should be computable with reasonable effort. Both requirements are met by the Oswald interpolation operator, that goes back to [35] (in the context of a posteriori error estimates; see also [26, section 2.5] and the references therein). Given any nonconforming approximation $p_h \in Q_h^k(\mathcal{T}_H) \not\subset H_0^1(\Omega)$ we choose $s \in H_0^1(\Omega)$ as a conforming reconstruction of p_h by the *Oswald interpolation operator* $I_{os} : Q_h^k(\mathcal{T}_H) \rightarrow Q_h^k(\mathcal{T}_H) \cap H_0^1(\Omega)$ which is defined by prescribing its values on the Lagrange nodes ν of the triangulation: we set $I_{os}[p_h](\nu) := p_h^t(\nu)$ inside any $t \in \tau_h$ and

$$I_{os}[p_h](\nu) := \frac{1}{|\tau_h^\nu|} \sum_{t \in \tau_h^\nu} p_h^t(\nu) \quad \text{for all inner nodes of } \tau_h \text{ and} \quad I_{os}[p_h](\nu) := 0$$

for all boundary nodes of τ_h , where $\tau_h^\nu \subset \tau_h$ denotes the set of all simplices of the fine triangulation which share ν as a node.

We continue with the specification of v , which is a bit more involved. The only formal requirement we have is for v to fulfill the local conservation property on each coarse element, although the diffusive flux estimator already gives a good hint on the specific form of v (namely, to be close to $-\lambda(\boldsymbol{\mu})\kappa_\varepsilon \nabla_h p_h(\boldsymbol{\mu})$). A particular choice is given by the elementwise diffusive flux reconstruction (with respect to the fine triangulation) that was proposed in [26], which fulfills the local conservation property on the coarse elements if properly defined with respect to our DG discretization.

Diffusive flux reconstruction. We will reconstruct a conforming diffusive flux approximation $u_h(\boldsymbol{\mu}) \in H_{\text{div}}(\Omega)$ of the nonconforming discrete diffusive flux $-\lambda(\boldsymbol{\mu})\kappa_\varepsilon \nabla_h p_h(\boldsymbol{\mu}) \notin H_{\text{div}}(\Omega)$ in a conforming discrete subspace $V_h^l(\tau_h) \subset H_{\text{div}}(\Omega)$, namely, the *Raviart–Thomas–Nédélec* space of vector functions (see [26] and the references therein), defined for $k - 1 \leq l \leq k$ by

$$(4.5) \quad V_h^l(\tau_h) := \{v \in H_{\text{div}}(\Omega) \mid v|_t \in [\mathbb{P}_l(t)]^d + \boldsymbol{x}\mathbb{P}_l(t) \quad \forall t \in \tau_h\}.$$

See [26, section 2.4] and the references therein for a detailed discussion of the role of the polynomial degree l , the properties of elements of $V_h^l(\tau_h)$, and the origin of the use of diffusive flux reconstructions in the context of error estimation in general. We define the *diffusive flux reconstruction* operator $R_h^l : \mathcal{P} \rightarrow [Q_h^k(\mathcal{T}_H) \rightarrow V_h^l(\tau_h)]$, $\boldsymbol{\mu} \mapsto [q_h \mapsto R_h^l[q_h; \boldsymbol{\mu}]]$, by locally specifying $R_h^l[q_h; \boldsymbol{\mu}] \in V_h^l(\tau_h)$, such that

$$(4.6) \quad (R_h^l[q_h; \boldsymbol{\mu}] \cdot \boldsymbol{n}_e, q)_{L^2, e} = b_c^e(q_h, q; \boldsymbol{\mu}) + b_p^e(q_h, q; \boldsymbol{\mu}) \quad \text{for all } q \in \mathbb{P}_l(e)$$

and all $e \in \mathcal{F}_h^t$ and

$$(4.7) \quad (R_h^l[q_h; \boldsymbol{\mu}], \nabla q)_{L^2,t} = -b^t(q_h, q; \boldsymbol{\mu}) - \vartheta \sum_{e \in \mathcal{F}_h^t} b_c^e(q, q_h; \boldsymbol{\mu}) \quad \text{for all } \nabla q \in [\mathbb{P}_{l-1}(t)]^d$$

with $q \in \mathbb{P}_l(t)$ for all $t \in \tau_h$, where ϑ is given by the local discretization inside each coarse element and by $\vartheta = 1$ on all fine faces that lie on a coarse face. The next lemma shows that this reconstruction of the diffusive flux is sensible for the DG solution as well as the reduced solution, since the reconstructions of both fulfill the requirements of Theorem 4.2.

LEMMA 4.4 (local conservativity). *Let $\mathbf{1} \in Q_{\text{red}}^T \subset Q_h^{k,T}$ for all $T \in \mathcal{T}_H$ and let $p_h(\boldsymbol{\mu}) \in Q_h^k(\mathcal{T}_H)$ be the DG solution of (2.7) and $p_{\text{red}}(\boldsymbol{\mu}) \in Q_{\text{red}}(\mathcal{T}_H)$ the reduced solution of (3.1) for a parameter $\boldsymbol{\mu} \in \mathcal{P}$ and let $u_h(\boldsymbol{\mu}) := R_h^l[p_h(\boldsymbol{\mu}); \boldsymbol{\mu}] \in V_h^l(\tau_h)$ and $u_{\text{red}}(\boldsymbol{\mu}) := R_h^l[p_{\text{red}}(\boldsymbol{\mu}); \boldsymbol{\mu}] \in V_h^l(\tau_h)$ denote their respective diffusive flux reconstructions. It then holds that $u_h(\boldsymbol{\mu})$ and $u_{\text{red}}(\boldsymbol{\mu})$ fulfill the local conservation property of Theorem 4.2, i.e.,*

$$(\nabla \cdot u_h(\boldsymbol{\mu}), \mathbf{1})_{L^2,T} = (f, \mathbf{1})_{L^2,T} = (\nabla \cdot u_{\text{red}}(\boldsymbol{\mu}), \mathbf{1})_{L^2,T} \quad \text{for all } T \in \mathcal{T}_H.$$

Proof. We follow the ideas of [26, Lemma 2.1] while accounting for the coarse triangulation. Let $\mathbf{1}^T \in Q_h^k(\mathcal{T}_H)$ be an indicator for $T \in \mathcal{T}_H$, such that $\mathbf{1}^T|_T = \mathbf{1} \in Q_h^{k,T}$ and zero everywhere else. It then holds with $* = h, \text{red}$, that

$$\begin{aligned} (\nabla \cdot u_*(\boldsymbol{\mu}), \mathbf{1})_{L^2,T} &= \sum_{t \in \tau_h^T} [(u_*(\boldsymbol{\mu}) \cdot n, \mathbf{1})_{L^2,\partial t} - (u_*(\boldsymbol{\mu}), \nabla \mathbf{1})_{L^2,t}] \\ &= b_h(p_*(\boldsymbol{\mu}), \mathbf{1}^T; \boldsymbol{\mu}) = (f, \mathbf{1})_{L^2,T} \end{aligned}$$

for all $T \in \mathcal{T}_H$, where we used Green's theorem in the first equality, the definition of the diffusive flux reconstruction, (4.6), (4.7), and the definition of $\mathbf{1}^T$ and b_h in the second, and the fact that $\mathbf{1} \in Q_h^{k,T}$ and p_* solves (2.7) or (3.1), respectively, in the third. \square

Inserting the Oswald interpolation for s and the diffusive flux reconstruction for v in Theorem 4.2 then yields a locally computable energy estimate for the DG as well as the reduced solution.

COROLLARY 4.5 (locally computable energy norm estimate). *Let $p(\boldsymbol{\mu}) \in H_0^1(\Omega)$ be the weak solution of (2.2), let $p_h(\boldsymbol{\mu}) \in Q_h^1(\mathcal{T}_H)$ be the DG solution of (2.7), let $p_{\text{red}}(\boldsymbol{\mu}) \in Q_{\text{red}}(\mathcal{T}_H)$ be the reduced solution of (3.1), and let R_h^l denote the diffusive flux reconstruction operator. Let the assumptions of Theorem 4.2 and Lemma 4.4 be fulfilled and let $\bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}} \in \mathcal{P}$ be arbitrary. It then holds that*

$$\begin{aligned} \|p(\boldsymbol{\mu}) - p_h(\boldsymbol{\mu})\|_{\bar{\boldsymbol{\mu}}} &\leq \eta(p_h(\boldsymbol{\mu}); \boldsymbol{\mu}, \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}}), \\ \|p(\boldsymbol{\mu}) - p_{\text{red}}(\boldsymbol{\mu})\|_{\bar{\boldsymbol{\mu}}} &\leq \eta(p_{\text{red}}(\boldsymbol{\mu}); \boldsymbol{\mu}, \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}}) \end{aligned}$$

with

$$\begin{aligned} \eta(\cdot; \boldsymbol{\mu}, \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}}) &:= \frac{1}{\sqrt{\alpha(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})}} \left[\sqrt{\gamma(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})} \left(\sum_{T \in \mathcal{T}_H} \eta_{\text{mc}}^T(\cdot; \bar{\boldsymbol{\mu}})^2 \right)^{1/2} + \left(\sum_{T \in \mathcal{T}_H} \eta_r^T(\cdot; \boldsymbol{\mu})^2 \right)^{1/2} \right. \\ &\quad \left. + \frac{1}{\sqrt{\alpha(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})}} \left(\sum_{T \in \mathcal{T}_H} \eta_{\text{df}}^T(\cdot; \boldsymbol{\mu}, \hat{\boldsymbol{\mu}})^2 \right)^{1/2} \right], \end{aligned}$$

and

$$\eta_{\text{nc}}^T(\cdot; \bar{\boldsymbol{\mu}}) := \tilde{\eta}_{\text{nc}}^T(\cdot, I_{\text{os}}[\cdot]; \bar{\boldsymbol{\mu}}), \quad \eta_{\text{r}}^T(\cdot; \boldsymbol{\mu}) := \tilde{\eta}_{\text{r}}^T(R_h^l[\cdot; \boldsymbol{\mu}]), \quad \eta_{\text{df}}^T(\cdot; \boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) := \tilde{\eta}_{\text{df}}^T(\cdot, R_h^l[\cdot; \boldsymbol{\mu}]; \hat{\boldsymbol{\mu}}).$$

Local efficiency. The global efficiency of the abstract estimate was already shown in Lemma 4.1 (again note that $\gamma(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}}) = \alpha(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}}) = 1$ if $\boldsymbol{\mu}$ and $\bar{\boldsymbol{\mu}}$ coincide); see [25, Remarks 4.2 and 4.3] for a discussion of the global efficiency of the abstract estimate. We also state a local efficiency of the local estimates using the knowledge of the discretization, the Oswald interpolation, and the diffusive flux reconstruction. To do so we further localize our estimates with respect to the fine triangulation and apply the ideas of [24, 25, 26]. We denote by \lesssim a proportionality relation between two quantities a and b in the sense that $a \lesssim b \iff a \leq Cb$, where the positive constant C depends on the space dimension, the polynomial degree k , the polynomial degree of f , the shape regularity of τ_h , and the DG parameters σ and ϑ , only. We additionally denote the set of all fine elements that touch $T \in \mathcal{T}_H$ by $\tilde{\tau}_h^T := \{t \in \tau_h \mid t \cap T \neq \emptyset\}$, the set of all fine faces that touch T by $\tilde{\mathcal{F}}_h^T := \{e \in \mathcal{F}_h \mid \exists t \in \tau_h^T : e \cap t \neq \emptyset\}$, and the weighted jump seminorms $\llbracket \cdot \rrbracket_{\boldsymbol{\mu}, \mathcal{F}} : \mathcal{P} \rightarrow [H^1(\tau_h) \rightarrow \mathbb{R}]$, $\boldsymbol{\mu} \mapsto [q \mapsto \llbracket [q] \rrbracket_{\boldsymbol{\mu}, \mathcal{F}}]$ and $\llbracket \cdot \rrbracket_{p, \boldsymbol{\mu}, \mathcal{F}} : \mathcal{P} \rightarrow [H^1(\tau_h) \rightarrow \mathbb{R}]$, $\boldsymbol{\mu} \mapsto [q \mapsto \llbracket [q] \rrbracket_{p, \boldsymbol{\mu}, \mathcal{F}}]$ for any subset $\mathcal{F} \subset \mathcal{F}_h$, all $\boldsymbol{\mu} \in \mathcal{P}$, and $q \in H^1(\tau_h)$ by

$$\begin{aligned} \llbracket [q] \rrbracket_{\boldsymbol{\mu}, \mathcal{F}} &:= \left(\sum_{e \in \mathcal{F}} \left\| \llbracket [(\lambda(\boldsymbol{\mu}) \kappa_\varepsilon \nabla q) \cdot n_e] \rrbracket_e \right\|_{L^2, e} \right)^{1/2} \quad \text{and} \\ \llbracket [q] \rrbracket_{p, \boldsymbol{\mu}, \mathcal{F}} &:= \left(\sum_{e \in \mathcal{F}} b_p^e(q, q; \boldsymbol{\mu}) \right)^{1/2}, \end{aligned}$$

respectively. We analogously denote the set of all fine elements that touch $t \in \tau_h$ by $\tilde{\tau}_h^t$ and the set of all fine faces that touch t by $\tilde{\mathcal{F}}_h^t$ and define

$$\begin{aligned} \tilde{c}_\varepsilon^T &:= \min_{t \in \tilde{\tau}_h^T} c_\varepsilon^t, & \overline{h}_T &:= \max_{t \in \tau_h^T} h_t, & \overline{\omega}^T &:= \left(\max_{e \in \mathcal{F}_h^T} \omega_e^{+2} \right)^{1/2}, \\ C_\varepsilon^T &:= \max_{t \in \tau_h^T} C_\varepsilon^t, & \underline{h}_T &:= \min_{t \in \tau_h^T} h_t, & \underline{C}_\varepsilon^T &:= \max_{t \in \tau_h^T} \left(\left(\max_{s \in t \cup \mathcal{N}(t)} \frac{C_\varepsilon^s}{c_\varepsilon^s} \right)^2 \right) \end{aligned}$$

for all $T \in \mathcal{T}_H$.

THEOREM 4.6 (local efficiency of the locally computable energy norm estimate).

With the notation and assumptions from Corollary 4.5, let f be polynomial and $\max_{t \in \tau_h} h_t \leq 1$. It then holds with $*$ = h, red , respectively, that

$$\begin{aligned} \eta_{\text{nc}}^T(p_*(\boldsymbol{\mu}); \bar{\boldsymbol{\mu}}) &\lesssim (C_\varepsilon^T / \tilde{c}_\varepsilon^T)^{1/2} \llbracket [p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu})] \rrbracket_{p, \bar{\boldsymbol{\mu}}, \tilde{\mathcal{F}}_h^T}, \\ \eta_{\text{r}}^T(p_*(\boldsymbol{\mu}); \boldsymbol{\mu}) &\lesssim \sqrt{\gamma(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})} (C_p^T / c_\varepsilon^T)^{1/2} h_T \left[\begin{aligned} &C_\varepsilon^T \underline{h}_T^{-1} \llbracket [p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu})] \rrbracket_{\bar{\boldsymbol{\mu}}, T} \\ &+ \overline{\omega}^T \overline{h}_T \llbracket [p_*(\boldsymbol{\mu})] \rrbracket_{\bar{\boldsymbol{\mu}}, \tilde{\mathcal{F}}_h^T} \\ &+ \sigma^{1/2} \underline{h}_T^{-1} \llbracket [p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu})] \rrbracket_{p, \bar{\boldsymbol{\mu}}, \tilde{\mathcal{F}}_h^T} \end{aligned} \right], \\ \eta_{\text{df}}^T(p_*(\boldsymbol{\mu}); \boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) &\lesssim \sqrt{\gamma(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})} \sqrt{\gamma(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})} \underline{C}_\varepsilon^T{}^{1/2} \left(\begin{aligned} &\llbracket [p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu})] \rrbracket_{\bar{\boldsymbol{\mu}}, T} \\ &+ \llbracket [p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu})] \rrbracket_{p, \bar{\boldsymbol{\mu}}, \tilde{\mathcal{F}}_h^T} \end{aligned} \right) \end{aligned}$$

for all coarse elements $T \in \mathcal{T}_H$.

Proof. We estimate each local estimator separately. It holds for the nonconformity estimator that

$$\begin{aligned} \eta_{\text{nc}}^T(p_*(\boldsymbol{\mu}); \bar{\boldsymbol{\mu}}) &= \left(\sum_{t \in \tau_h^T} \|p_*(\boldsymbol{\mu}) - I_{\text{os}}[p_*(\boldsymbol{\mu})]\|_{\bar{\boldsymbol{\mu}}, t}^2 \right)^{1/2} \\ &\lesssim \left(\sum_{t \in \tau_h^T} C_\varepsilon^t \min_{\tilde{t} \in \tilde{\tau}_h^t} (c_\varepsilon^{\tilde{t}})^{-1} \llbracket p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu}) \rrbracket_{p, \bar{\boldsymbol{\mu}}, \tilde{\mathcal{F}}_h^t}^2 \right)^{1/2} \\ &\leq (C_\varepsilon^T / \tilde{c}_\varepsilon^T)^{1/2} \llbracket p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu}) \rrbracket_{p, \bar{\boldsymbol{\mu}}, \tilde{\mathcal{F}}_h^T}, \end{aligned}$$

where we use the definition of $\eta_{\text{nc}}^T[p_*(\boldsymbol{\mu})]$ and τ_h^T in the equality and the arguments of [26, Proof of Theorem 3.2] in the first and the definition of C_ε^T , \tilde{c}_ε^T and $\tilde{\mathcal{F}}_h^T$ in the second inequality.

It holds for the residual estimator that

$$\begin{aligned} \eta_r^T(p_*(\boldsymbol{\mu}); \boldsymbol{\mu}) &\leq (C_p^T / c_\varepsilon^T)^{1/2} h_T \left(\underbrace{\|f + \nabla \cdot (\lambda(\boldsymbol{\mu}) \kappa_\varepsilon \nabla p_*(\boldsymbol{\mu}))\|_{L^2, T}}_{:= (i)} \right. \\ &\quad \left. + \underbrace{\|\nabla \cdot (\lambda(\boldsymbol{\mu}) \kappa_\varepsilon \nabla p_*(\boldsymbol{\mu}) + u_*(\boldsymbol{\mu}))\|_{L^2, T}}_{:= (ii)} \right) \end{aligned}$$

with u_* as in Lemma 4.4, where we used the definition of $\eta_r^T[u_*(\boldsymbol{\mu})]$ and the triangle inequality, which leaves us with two terms we will estimate separately.

- (i) The first term can be estimated as follows, using the definition of τ_h^T and the arguments of [24, Proposition 3.3] in the first and the definition of C_ε^T and τ_h^T and the fact that $\max_{t \in \tau_h} h_t \leq 1$ in the second inequality:

$$(i) \lesssim \left(\sum_{t \in \tau_h^T} C_\varepsilon^t h_t^{-2} \llbracket p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu}) \rrbracket_{\boldsymbol{\mu}, t}^2 \right)^{-1/2} \leq C_\varepsilon^T \underline{h}_T^{-1} \llbracket p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu}) \rrbracket_{\boldsymbol{\mu}, T}.$$

- (ii) The second term can be estimated as

$$\begin{aligned} (ii) &\lesssim \left[\sum_{t \in \tau_h} \left(h_t^{1/2} \sum_{e \in \mathcal{F}_h^t} \omega_e^+ \left\| [(\lambda(\boldsymbol{\mu}) \kappa_\varepsilon \nabla p_*(\boldsymbol{\mu})) \cdot n_e] \right\|_{L^2, e} \right. \right. \\ &\quad \left. \left. + \sigma^{1/2} h_t^{-1} \llbracket p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu}) \rrbracket_{p, \boldsymbol{\mu}, \mathcal{F}_h^t} \right)^2 \right]^{1/2} \\ &\lesssim \overline{\omega}^T \overline{h}_T \llbracket p_*(\boldsymbol{\mu}) \rrbracket_{\boldsymbol{\mu}, \mathcal{F}_h^T} + \underline{h}_T^{-1} \sigma^{1/2} \llbracket p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu}) \rrbracket_{p, \boldsymbol{\mu}, \mathcal{F}_h^T}, \end{aligned}$$

using the definition of τ_h^T and the arguments of [26, Proof of Theorem 3.2] in the first and Young's inequality and the definition of $\overline{\omega}^T$, \overline{h}_T , \underline{h}_T , and τ_h^T in the second inequality.

Applying the norm equivalence (2.3) yields the desired result for the residual estimator.

Finally, it holds for the diffusive flux estimator that

$$\begin{aligned} \eta_{\text{df}}^T(p_*(\boldsymbol{\mu}); \boldsymbol{\mu}, \bar{\boldsymbol{\mu}}) &\leq \sqrt{\gamma(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})} \left(\sum_{t \in \tau_h^T} \left\| (\lambda(\boldsymbol{\mu})\kappa_\varepsilon)^{-1/2} (\lambda(\boldsymbol{\mu})\kappa_\varepsilon \nabla p_*(\boldsymbol{\mu}) + u_*(\boldsymbol{\mu})) \right\|_{L^2, t}^2 \right)^{1/2} \\ &\lesssim \sqrt{\gamma(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})} \left[\sum_{t \in \tau_h^T} \left(\max_{s \in t \cup \mathcal{N}(t)} \frac{C_\varepsilon^s}{c_\varepsilon^s} \right)^2 \left(\|p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu})\|_{\boldsymbol{\mu}, t} \right. \right. \\ &\qquad \qquad \qquad \left. \left. + \llbracket p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu}) \rrbracket_{p, \boldsymbol{\mu}, \mathcal{F}_h^t} \right)^2 \right]^{1/2} \\ &\lesssim \sqrt{\gamma(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})} \underline{C}_\varepsilon^T{}^{1/2} \left(\|p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu})\|_{\boldsymbol{\mu}, T} + \llbracket p(\boldsymbol{\mu}) - p_*(\boldsymbol{\mu}) \rrbracket_{p, \boldsymbol{\mu}, \mathcal{F}_h^T} \right), \end{aligned}$$

where we used the definition of $\eta_{\text{df}}^T[p_*(\boldsymbol{\mu})]$ and τ_h^T and the parameter equivalence from (2.3) in the first, [25, Lemma 4.12] in the second, and the definition of $\underline{C}_\varepsilon^T$ and \mathcal{F}_h^T in the third inequality. Applying the norm equivalence (2.3) again finally yields the desired result for the diffusive flux estimator. \square

5. On-line enrichment. As mentioned in section 3 there are two drawbacks of classical RB methods in the context of parametric multiscale problems, stemming from the fact that $\dim Q_h^k$ roughly scales with $\mathcal{O}(\varepsilon^{-1})$: (i) expensive high-dimensional evaluations of global quantities during reduction and orthogonalization and (ii) expensive high-dimensional inversions of (2.7) during the basis generation. The first shortcoming was partially addressed by the LRBMS method introduced in [36, 7] by localization and by the local error estimator introduced in the previous section, that can be off-line/on-line decomposed with a computational complexity depending only linearly on $|\tau_h|$ (not shown) while the original estimator presented in [7] required the computation of $|\mathcal{T}_H|$ global Riesz representatives for each snapshot (and thus inversions of high-dimensional product matrices). But also the LRBMS method suffers from the second shortcoming, namely, that the basis generation (for instance using an adaptive greedy procedure) requires the global high-dimensional problem (2.7) to be solved for a possibly large number of parameters out of a finite training set $\mathcal{P}_{\text{train}} \subset \mathcal{P}$. In the context of parametric multiscale problems, however, solving (2.7) may be prohibitively expensive and one may only have the resources to do so for a very limited number of parameters, if at all. Such an RB space constructed out of only a very few solution snapshots is usually insufficient for nearly all parameters $\boldsymbol{\mu} \in \mathcal{P}$ that were not included in the basis generation.

To address this shortcoming we relax the notion of a strict off-line/on-line splitting of the computation in the classical sense. While the computational complexity of the on-line phase must not depend on any global high-dimensional quantities (namely, $\dim Q_h^k$) for RB methods, we allow for high-dimensional but local computations (of order $\dim Q_h^{k, T}$ with $T \in \mathcal{T}_H$) in the context of the LRBMS method. The idea is as follows: during the off-line phase we initialize the local reduced bases Q_{red}^T with a DG basis of order up to $k_H \in \mathbb{N}$ with respect to the coarse elements $T \in \mathcal{T}_H$, thus ensuring that any reduced solution is at least as good as a DG solution on the coarse triangulation. We then carry out a discrete weak greedy algorithm based on the error estimator defined in Corollary 4.5 while allowing only for a limited amount of global solution snapshots, $N_{\text{greedy}} \in \mathbb{N}$, and extend the local bases with these snapshots using an orthonormalization algorithm ONB locally on each $T \in \mathcal{T}_H$. This procedure is summarized in Algorithm 5.1.

ALGORITHM 5.1. DISCRETE WEAK GREEDY ALGORITHM USED IN THE LRBMS METHOD.

Input: ONB, $k_H \in \mathbb{N}$, $\mathcal{P}_{\text{train}} \subset \mathcal{P}$, $\Delta_{\text{greedy}} > 0$, $N_{\text{greedy}} \in \mathbb{N}$, $\bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}} \in \mathcal{P}$

Output: A local reduced basis Φ^T for each coarse element $T \in \mathcal{T}_H$.

Initialize the local reduced bases with the coarse DG basis:

for all $T \in \mathcal{T}_H$ **do**

$$\Phi^{T(0)} \leftarrow \text{ONB}(\{\text{DG shape functions of order up to } k_H \text{ w.r.t. } T\})$$

end for

$$Q_{\text{red}}(\mathcal{T}_H)^{(0)} \leftarrow \bigoplus_{T \in \mathcal{T}_H} \text{span}(\Phi^{T(0)})$$

$n \leftarrow 0$

while $\max_{\boldsymbol{\mu} \in \mathcal{P}_{\text{train}}} \eta(p_{\text{red}}(\boldsymbol{\mu}); \boldsymbol{\mu}, \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}}) > \Delta_{\text{greedy}}$, with $p_{\text{red}}(\boldsymbol{\mu}) \in Q_{\text{red}}(\mathcal{T}_H)^{(n)}$

solving (3.1) **and** $n < N_{\text{greedy}}$ **do**

Compute all reduced quantities w.r.t. $Q_{\text{red}}(\mathcal{T}_H)^{(n)}$.

Find the worst approximated parameter, with $p_{\text{red}}(\boldsymbol{\mu}) \in Q_{\text{red}}(\mathcal{T}_H)^{(n)}$

solving (3.1):

$$\boldsymbol{\mu}_{\text{max}} \leftarrow \arg \max_{\boldsymbol{\mu} \in \mathcal{P}_{\text{train}}} \eta(p_{\text{red}}(\boldsymbol{\mu}); \boldsymbol{\mu}, \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}})$$

Extend the local reduced bases, with $p_h(\boldsymbol{\mu}_{\text{max}}) \in Q_h^k$ solving (2.7):

for all $T \in \mathcal{T}_H$ **do**

$$\Phi^{T(n+1)} \leftarrow \text{ONB}(\{\Phi^{T(n)}, p_h(\boldsymbol{\mu}_{\text{max}})|_T\})$$

end for

$$Q_{\text{red}}(\mathcal{T}_H)^{(n+1)} \leftarrow \bigoplus_{T \in \mathcal{T}_H} \text{span}(\Phi^{T(n+1)})$$

$n \leftarrow n + 1$

end while

return $\{\Phi^{T(n)}\}_{T \in \mathcal{T}_H}$

During the on-line phase, given any $\boldsymbol{\mu} \in \mathcal{P}$, we compute a reduced solution $p_{\text{red}}(\boldsymbol{\mu}) \in Q_{\text{red}}(\mathcal{T}_H)$ and efficiently assess its quality using the error estimator. If the estimated error is above a prescribed tolerance, $\Delta_{\text{online}} > 0$, we start an intermediate local enrichment phase to enrich the reduced bases in the SEMR (solve \rightarrow estimate \rightarrow mark \rightarrow refine) spirit of adaptive mesh refinement (the procedure is summarized in Algorithm 5.2): we first compute local error indicators $\eta^T(p_{\text{red}}(\boldsymbol{\mu}); \boldsymbol{\mu}, \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}})$ for all $T \in \mathcal{T}_H$, such that $\eta(\cdot; \boldsymbol{\mu}, \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}})^2 \leq \sum_{T \in \mathcal{T}_H} \eta^T(\cdot; \boldsymbol{\mu}, \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}})^2$, defined as¹

(5.1)

$$\eta^T(\cdot; \boldsymbol{\mu}, \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}})^2 := \frac{3}{\sqrt{\alpha(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})}} \left[\sqrt{\gamma(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})} \eta_{\text{nc}}^T(\cdot; \bar{\boldsymbol{\mu}})^2 + \eta_r^T(\cdot; \boldsymbol{\mu})^2 + \frac{1}{\sqrt{\alpha(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})}} \eta_{\text{df}}^T(\cdot; \boldsymbol{\mu}, \hat{\boldsymbol{\mu}})^2 \right]$$

and mark coarse elements $\tilde{\mathcal{T}}_H \subseteq \mathcal{T}_H$ for enrichment, given a marking strategy MARK. For each marked $T \in \tilde{\mathcal{T}}_H$ we solve

$$(5.2) \quad b_h^{T_\delta}(p_h^{T_\delta}(\boldsymbol{\mu}), q_h; \boldsymbol{\mu}) = l_h^{T_\delta}(q_h) \quad \text{for all } q_h \in Q_h^k(\tau_h^{T_\delta})$$

on an overlapping domain $T_\delta \supset T$ with the insufficient reduced solution $p_{\text{red}}(\boldsymbol{\mu})$ as Dirichlet boundary values on ∂T_δ to obtain an updated detailed solution $p_h^{T_\delta}(\boldsymbol{\mu}) \in Q_h^k(\tau_h^{T_\delta})$. Here $Q_h^k(\tau_h^{T_\delta})$, $b_h^{T_\delta}$, and $l_h^{T_\delta}$ are extensions of $Q_h^k(\tau_h^T)$, b_h^T , and l^T , respec-

¹The chosen indicator is an upper bound for the error estimator which can be seen by applying Young's inequality.

tively, to the oversampled domain T_δ while additionally encoding $p_{\text{red}}(\boldsymbol{\mu})$ as Dirichlet boundary values. We then extend each marked local reduced basis by performing an orthonormalization procedure on $p_h^{T_\delta}(\boldsymbol{\mu})|_T$ with respect to the existing local basis and update all reduced quantities with respect to the newly added basis vector. We finally compute an updated reduced solution in the updated coarse reduced space and estimate the error again. We repeat this procedure until the estimated error falls below the prescribed tolerance Δ_{online} or until the prescribed maximum number of iterations, $N_{\text{online}} \in \mathbb{N}$, is reached. Possible choices for ONB and MARK are given in section 6.2. Note that first steps in the direction of on-line enrichment were published in [8].

Remark 5.1. It is also possible to use other methods to compute (or approximate) $p_h(\boldsymbol{\mu})$ during the greedy procedure, in particular other DD or multiscale methods. If the resulting approximation does not fulfill the local conservation property of Lemma 4.4, however, the estimator would have to be replaced during the greedy algorithm, for instance in the spirit of [46]. But during the on-line phase the estimator would be valid for any reduced solution, as long as the basis is initialized with at least a constant function. In particular one could use variants of the MsFEM (see [22]), the HMM (see [4]), or in particular the DG-HMM (see [1, 6]) during the greedy procedure to generate a coarse reduced basis with approximation properties of order H . Fine scale features of the solution would then be adaptively recovered during the on-line enrichment phase, if and where needed.

Remark 5.2. The computation of the local error indicators in Algorithm 5.2 can be efficiently off-line/on-line decomposed (not shown here). Once a set of subdomains has been marked, the enrichment can be carried out in parallel without any communication. For the update of the reduced quantities only local information and the information on one layer of neighboring fine grid cells is needed. The on-line phase,

ALGORITHM 5.2. ADAPTIVE BASIS ENRICHMENT IN THE INTERMEDIATE LOCAL ENRICHMENT PHASE.

Input: MARK, ONB, $\{\Phi^T\}_{T \in \mathcal{T}_H}$, $p_{\text{red}}(\boldsymbol{\mu})$, $\boldsymbol{\mu} \in \mathcal{P}$, $\Delta_{\text{online}} > 0$, $N_{\text{online}} \in \mathbb{N}$, $\bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}} \in \mathcal{P}$

Output: Updated reduced solution and local reduced bases.

$\Phi^{T(0)} \leftarrow \Phi^T, \forall T \in \mathcal{T}_H$

$n \leftarrow 0$

while $\eta(p_{\text{red}}(\boldsymbol{\mu}); \boldsymbol{\mu}, \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}}) > \Delta_{\text{online}}$, with $p_{\text{red}}(\boldsymbol{\mu}) \in Q_{\text{red}}(\mathcal{T}_H)^{(n)}$ solving (3.1)

and $n < N_{\text{online}}$ **do**

for all $T \in \mathcal{T}_H$ **do**

Compute local error indicator $\eta^T(p_{\text{red}}(\boldsymbol{\mu}); \boldsymbol{\mu}, \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}})$ according to (5.1).

end for

$\tilde{\mathcal{T}}_H \leftarrow \text{MARK}(\mathcal{T}_H, \{\eta^T(p_{\text{red}}(\boldsymbol{\mu}); \boldsymbol{\mu}, \bar{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}})\}_{T \in \mathcal{T}_H})$

for all $T \in \tilde{\mathcal{T}}_H$ **do**

Solve (5.2) for $p_h^{T_\delta}(\boldsymbol{\mu}) \in Q_h^k(\tau_h^{T_\delta})$.

$\Phi^{T(n+1)} \leftarrow \text{ONB}(\{\Phi^{T(n)}, p_h^{T_\delta}(\boldsymbol{\mu})|_T\})$

end for

$Q_{\text{red}}(\mathcal{T}_H)^{(n+1)} \leftarrow \bigoplus_{T \in \tilde{\mathcal{T}}_H} \text{span}(\Phi^{T(n+1)}) \oplus \bigoplus_{T \in \mathcal{T}_H \setminus \tilde{\mathcal{T}}_H} \text{span}(\Phi^{T(n)})$

Update all reduced quantities w.r.t $Q_{\text{red}}(\mathcal{T}_H)^{(n+1)}$.

Solve (3.1) for $p_{\text{red}}(\boldsymbol{\mu}) \in Q_{\text{red}}(\mathcal{T}_H)^{(n+1)}$.

$n \leftarrow n + 1$

end while

return $p_{\text{red}}(\boldsymbol{\mu}), \{\Phi^{T(n)}\}_{T \in \mathcal{T}_H}$

however, requires information on \mathcal{T}_H (in particular the number of coarse elements and neighboring information), in contrast to traditional RB methods.

Remark 5.3. Our choice of the greedy algorithm and the adaptive on-line enrichment covers a wide range of scenarios. Disabling the on-line enrichment (by setting $N_{\text{online}} = 0$) and choosing any suitable Δ_{greedy} and N_{greedy} yields the standard greedy basis generation, well known in the RB context. Setting $N_{\text{greedy}} = 0$ and $k_H = 1$, on the other hand, disables the greedy procedure and merely initializes the reduced bases with the coarse DG basis of order one. This is of particular interest in situations where the computation of solutions of the detailed problem during the greedy procedure might be too costly. In that setup nearly all work is done in the adaptive on-line enrichment phase. Many other variants are possible, e.g., other local boundary conditions, several marking strategies MARK or orthonormalization algorithms ONB or other stopping criteria; one could also limit the number of intermediate snapshots added to the local bases. Depending on these choices the resulting method is then close to existing DD methods (i.e., a DD method with overlapping subdomains; see [47]) or multiscale methods (i.e., the adaptive iterative multiscale finite volume method [30]).

The LRBMS method with the proposed adaptive on-line enrichment strategy is now suitable for a far wider range of circumstances than standard RB methods or the previously published variant of the LRBMS method. As mentioned before it can now be applied if the computational power available for the off-line phase is limited by time or resource constraints. It can also be applied if the set of training parameters given to the greedy algorithm was insufficiently chosen or even if on-line a solution to a parameter is requested that is outside of the original bounds of the parameter space. In general, the on-line adaptive LRBMS method can be applied whenever the basis that was generated during the off-line phase turns out to not be sufficient for what is required during the on-line phase.

6. Numerical experiments. In this section we investigate the performance of the error estimator in the context of the DG discretization defined in section 2.3 as well as in the context of the the LRBMS method defined in section 3. We also investigate the performance of the on-line enrichment procedure we proposed in the previous section.

We used several software packages for the implementation: everything concerning the discretization was implemented within the high performance C++ software framework DUNE [11, 10] while everything related to model reduction was implemented based on the pyMOR package [40]. Data functions, container, and linear solvers were implemented within `dune-stuff` [41], discrete function spaces (based on the discretization modules `dune-fem` [18] and `dune-pdelab` [21]), operators and products for the discretization as well as the error estimator were implemented within `dune-gdt` [49]. The coarse triangulation was implemented in `dune-grid-multiscale` [50] while the Python bindings forming the bridge between DUNE and pyMOR were implemented in `dune-pymor` [48]. Finally a high-level solver for linear elliptic (and possibly parametric) PDEs was implemented in `dune-hdd` [51] and exposed to pyMOR, where everything related to model reduction was implemented.

The fine triangulations τ_h used are conforming refinements of triangular grids, represented by instances of `ALUGrid< 2, 2, simplex, conforming >` (see [17]). In the following experiments we choose a DG space $Q_h^{1,T}(\tau_h^T)$ locally on all coarse elements $T \in \mathcal{T}_H$. The resulting discretization thus coincides with the one defined in [27, 26] (though in general other choices are possible). All coarse triangulations \mathcal{T}_H used consist of squared elements (though arbitrary shapes are possible).

Most figures in this publication were created or arranged using TikZ [55, 54] and pgfplots [28], the colorblind safe colors in Figures 4 and 6 were selected with ColorBrewer [14].

6.1. Study of the a posteriori error estimator. To study the convergence properties of our estimator we consider two experiments. The first one serves as an academic example and as a comparison to the work of [25, 26]. The second experiment demonstrates the efficiency of the estimator in realistic circumstances. In both experiments we compute estimator components $\eta_*^2 := \sum_{T \in \mathcal{T}_H} \eta_*^{T^2}$ for $*$ =, nc, r, df, and the estimator η as defined in Corollary 4.5, using a 0th order diffusive flux reconstruction ($l = 0$ in (4.6), (4.7)). If no analytical solution $p(\boldsymbol{\mu})$ is available we approximate the discretization error $\|p(\boldsymbol{\mu}) - p_h(\boldsymbol{\mu})\|_{\overline{\boldsymbol{\mu}}}$ by substituting $p(\boldsymbol{\mu})$ for a discrete solution on a finer grid and by computing all integrals using a high order quadrature on the finer grid. We denote the efficiency of the estimator, $\eta(p_h(\boldsymbol{\mu}); \boldsymbol{\mu}, \overline{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}}) / \|p(\boldsymbol{\mu}) - p_h(\boldsymbol{\mu})\|_{\overline{\boldsymbol{\mu}}} \geq 1$, by “eff.” and the average (over all refinement steps) experimental order of convergence of a quantity by “order.”

Academic example. We consider (2.1) on $\Omega = [-1, 1]^2$ with a parameter space $\mathcal{P} = [0.1, 1]$, $\kappa_\varepsilon \equiv \text{id}$, $f(x, y) = \frac{1}{2}\pi^2 \cos(\frac{1}{2}\pi x) \cos(\frac{1}{2}\pi y)$, $\lambda(x, y; \boldsymbol{\mu}) = 1 + (1 - \boldsymbol{\mu}) \cos(\frac{1}{2}\pi x) \cos(\frac{1}{2}\pi y)$, and homogeneous Dirichlet boundary values. We study the components of the estimator as well as its efficiency in several circumstances, i.e., for different parameters $\boldsymbol{\mu}, \overline{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}} \in \mathcal{P}$ and triangulations τ_h and \mathcal{T}_H . We choose τ_h just as in [25, section 8.1] and begin with $\boldsymbol{\mu} = \overline{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}} = 1$, thus reproducing the nonparametric example studied in [25, section 8.1] (since $\lambda \equiv 1$ and all constants involving α and γ are equal to 1). For this specific choice of parameters an exact solution is available (see [25, section 8.1]). In this configuration, η_{nc} and η_{df} coincide with their respective counterparts defined in [25, 26] while η_r is directly influenced by the choice of the coarse triangulation and the parametric nature of λ (entering c_ε^T). Choosing $\mathcal{T}_H = \Omega$ (the coarse grid configuration with the worst efficiency), we observe results similar to [25, Table 1] for η_{df} and η_{nc} in Table 1. In contrast, η_r shows only linear convergence while the residual estimator in [25, Table 1] converges with second order. Overall, the efficiency of the estimator η is around 3.5 (for fixed $|\mathcal{T}_H| = 1$) while the efficiency of the estimator in [25, Table 1] is around 1.2. We can recover the superconvergence of η_r , however, by refining \mathcal{T}_H along with τ_h (thus keeping the ratio H/h fixed); see the left columns of Table 2. In order to make the estimator off-line/on-line decomposable in the parametric setting, $\hat{\boldsymbol{\mu}}$ has to be fixed throughout the experiment. Choosing $\hat{\boldsymbol{\mu}} = 0.1$ has no negative impact on the efficiency of the estimator, as we observe in the right columns of Table 2. It is often desirable to additionally fix the error norm throughout the experiments. Choosing $\overline{\boldsymbol{\mu}} = 0.1$ we still observe a very reasonable efficiency in Table 3.

Multiscale example. We consider (2.1) on $\Omega = [0, 5] \times [0, 1]$ with $f(x, y) = 2 \cdot 10^3$ if $(x, y) \in [0.95, 1.10] \times [0.30, 0.45]$, $f(x, y) = -1 \cdot 10^3$ if $(x, y) \in [3.00, 3.15] \times [0.75, 0.90]$

TABLE 1

Discretization error, estimator components, and efficiency of the error estimator for the academic example in section 6.1 with $|\mathcal{T}_H| = 1$ and $\boldsymbol{\mu} = \overline{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}} = 1$.

$ \tau_h $	$\ p(\boldsymbol{\mu}) - p_h(\boldsymbol{\mu})\ _{\overline{\boldsymbol{\mu}}}$	$\eta_{nc}(\cdot; \overline{\boldsymbol{\mu}})$	$\eta_r(\cdot; \boldsymbol{\mu})$	$\eta_{df}(\cdot; \boldsymbol{\mu}, \hat{\boldsymbol{\mu}})$	eff.
128	$3.28 \cdot 10^{-1}$	$1.66 \cdot 10^{-1}$	$5.79 \cdot 10^{-1}$	$3.55 \cdot 10^{-1}$	3.36
512	$1.60 \cdot 10^{-1}$	$7.89 \cdot 10^{-2}$	$2.90 \cdot 10^{-1}$	$1.76 \cdot 10^{-1}$	3.40
2,048	$7.78 \cdot 10^{-2}$	$3.91 \cdot 10^{-2}$	$1.45 \cdot 10^{-1}$	$8.73 \cdot 10^{-2}$	3.49
8,192	$3.47 \cdot 10^{-2}$	$1.95 \cdot 10^{-2}$	$7.27 \cdot 10^{-2}$	$4.35 \cdot 10^{-2}$	3.91
order	1.08	1.03	1.00	1.01	–

TABLE 2

Selected estimator components, estimated error, and efficiency of the error estimator for the academic example in section 6.1 with τ_h and \mathcal{T}_H simultaneously refined, $\mu = 1$, and two choices of $\hat{\mu}$. Note that the estimator components η_{nc} and η_{df} are not influenced by \mathcal{T}_H , the estimator components η_{nc} and η_r are not influenced by $\hat{\mu}$, and the discretization error is not influenced by either. Thus only η_r , η , and its efficiency are given for $\hat{\mu} = 1$ and only η_{df} , η , and its efficiency are given for $\hat{\mu} = 0.1$ (the other quantities coincide with the ones in Table 1).

$ \tau_h / \mathcal{T}_H $	$\eta_r(\cdot; \mu)$	$\hat{\mu} = 1$		$\hat{\mu} = 0.1$		
		$\eta(\cdot; \mu, \bar{\mu}, \hat{\mu})$	eff.	$\eta_{df}(\cdot; \mu, \hat{\mu})$	$\eta(\cdot; \mu, \bar{\mu}, \hat{\mu})$	eff.
128 / 2×2	$2.89 \cdot 10^{-1}$	$8.10 \cdot 10^{-1}$	2.47	$3.16 \cdot 10^{-1}$	$7.71 \cdot 10^{-1}$	2.35
512 / 4×4	$7.26 \cdot 10^{-2}$	$3.27 \cdot 10^{-1}$	2.04	$1.56 \cdot 10^{-1}$	$3.08 \cdot 10^{-1}$	1.92
2,048 / 8×8	$1.82 \cdot 10^{-2}$	$1.45 \cdot 10^{-1}$	1.86	$7.74 \cdot 10^{-2}$	$1.35 \cdot 10^{-1}$	1.73
8,192 / 16×16	$4.54 \cdot 10^{-3}$	$6.76 \cdot 10^{-2}$	1.95	$3.85 \cdot 10^{-2}$	$6.26 \cdot 10^{-2}$	1.80
order	2.00	1.20	–	1.01	1.21	–

TABLE 3

Discretization error, selected estimator component, estimated error, and efficiency of the error estimator for the academic example in section 6.1 with τ_h and \mathcal{T}_H simultaneously refined, $\mu = 1$, and $\bar{\mu} = \hat{\mu} = 0.1$. Note that η_r and η_{df} are not influenced by $\bar{\mu}$ and coincide with Table 2.

$ \tau_h / \mathcal{T}_H $	$\ p(\mu) - p_h(\mu)\ _{\bar{\mu}}$	$\eta_{nc}(\cdot; \bar{\mu})$	$\eta(\cdot; \mu, \bar{\mu}, \hat{\mu})$	eff.
128 / 2×2	$3.81 \cdot 10^{-1}$	$1.82 \cdot 10^{-1}$	$1.18 \cdot 10^0$	3.10
512 / 4×4	$1.87 \cdot 10^{-1}$	$8.57 \cdot 10^{-2}$	$5.00 \cdot 10^{-1}$	2.67
2,048 / 8×8	$9.08 \cdot 10^{-2}$	$4.22 \cdot 10^{-2}$	$2.29 \cdot 10^{-1}$	2.52
8,192 / 16×16	$4.05 \cdot 10^{-2}$	$2.11 \cdot 10^{-2}$	$1.10 \cdot 10^{-1}$	2.71
order	1.08	1.03	1.14	–



FIG. 1. Data functions of the multiscale example in section 6.1 on a triangulation with $|\tau_h| = 16,000$ elements: location of the channel function λ_c (left) and plot of the force f (right) modeling one source (black: $2 \cdot 10^3$) and two sinks (dark gray: $-1 \cdot 10^3$, zero elsewhere).

or $(x, y) \in [4.25, 4.40] \times [0.25, 0.40]$, and 0 everywhere else, $\lambda(x, y; \mu) = 1 + (1 - \mu)\lambda_c(x, y)$, $\kappa_\varepsilon = \kappa \text{id}$, homogeneous Dirichlet boundary values, and a parameter space $\mathcal{P} = [0.1, 1]$. On each $t \in \tau_h$, $\kappa|_t$ is the corresponding 0th entry of the permeability tensor used in the first model of the 10th SPE Comparative Solution Project (which is given by 100×20 constant tensors; see [53]) and λ_c models a channel, as depicted in Figure 1. The right-hand side f models a strong source in the middle left of the domain and two sinks in the top and right middle of the domain, as is visible in the structure of the solutions (see Figure 1). The role of the parameter μ is to toggle the existence of the channel λ_c . Thus $\lambda(\mu)\kappa$ reduces to the abovementioned permeability tensor for $\mu = 1$ while $\mu = 0.1$ models the removal of a large high-conductivity region near the center of the domain (see the first row in Figure 2). This missing channel has a visible impact on the structure of the pressure distribution as well as the reconstructed velocities, as we observe in the left column of Figure 2. With a contrast of 10^6 in the diffusion tensor and an ε of about $|\Omega|/2,000$ this setup is a challenging heterogeneous multiscale problem.

While we observe in Table 4 that the convergence rates of the estimator components are not as good as in the experiment studied before, the estimator shows an average efficiency of 5.5, which is quite remarkable considering the contrast of the

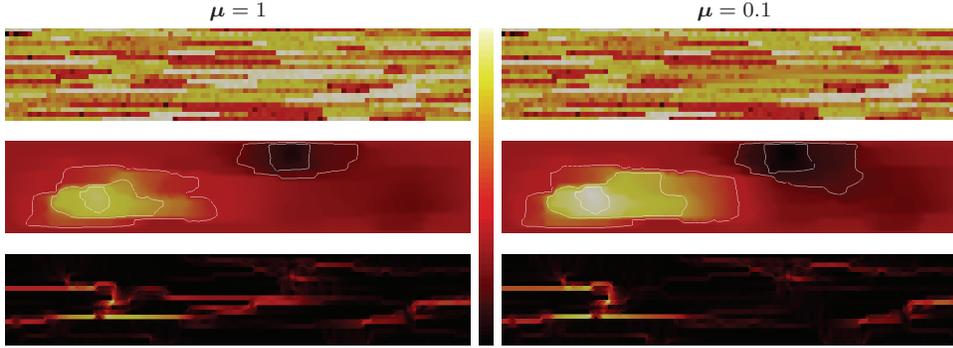


FIG. 2. Data functions and sample solutions of the multiscale example in section 6.1 on a triangulation with $|\tau_h| = 16,000$ elements for parameters $\mu = 1$ (left column) and $\mu = 0.1$ (right column). In each row both plots share the same color map (middle) with different ranges per row. From top to bottom: logarithmic plot of $\lambda(\mu)\kappa$ (dark: $1.41 \cdot 10^{-3}$, light: $1.41 \cdot 10^3$), plot of the pressure $p_h(\mu)$ (solution of (2.7), dark: $-3.92 \cdot 10^{-1}$, light: $7.61 \cdot 10^{-1}$, isolines at 10%, 20%, 45%, 75%, and 95%), and plot of the magnitude of the reconstructed diffusive flux $R_h^0[p_h(\mu); \mu]$ (defined in (4.6) and (4.7), dark: $3.10 \cdot 10^{-6}$, light: $3.01 \cdot 10^2$). Note the presence of high-conductivity channels in the permeability (top left, light regions) throughout large parts of the domain. The parameter dependency models a removal of one such channel in the middle right of the domain (top right), well visible in the reconstructed Darcy velocity fields (bottom).

TABLE 4

Discretization error, estimator components, and efficiency of the error estimator for the multiscale example in section 6.1 with τ_h and \mathcal{T}_H simultaneously refined and $\mu = \bar{\mu} = \hat{\mu} = 1$. Note that η_r should be close to zero (since f is piecewise constant, compare Figure 1) and suffers from numerical inaccuracies (ignoring the last refinement would yield an average order of 2.33 for η_r).

$ \tau_h $ / $ \mathcal{T}_H $	$\ p(\mu) - p_h(\mu)\ _{\bar{\mu}}$	$\eta_{nc}(\cdot; \bar{\mu})$	$\eta_r(\cdot; \mu)$	$\eta_{df}(\cdot; \mu, \hat{\mu})$	eff.
16,000 / 25×5	$7.49 \cdot 10^{-1}$	$2.13 \cdot 10^0$	$1.88 \cdot 10^{-9}$	$9.66 \cdot 10^{-1}$	4.14
64,000 / 50×10	$4.52 \cdot 10^{-1}$	$1.46 \cdot 10^0$	$7.05 \cdot 10^{-10}$	$6.05 \cdot 10^{-1}$	4.58
256,000 / 100×20	$2.58 \cdot 10^{-1}$	$1.02 \cdot 10^0$	$7.44 \cdot 10^{-11}$	$3.85 \cdot 10^{-1}$	5.44
1,024,000 / 200×40	$1.26 \cdot 10^{-1}$	$7.20 \cdot 10^{-1}$	$2.00 \cdot 10^{-10}$	$2.49 \cdot 10^{-1}$	7.70
order	0.86	0.52	1.07	0.65	-

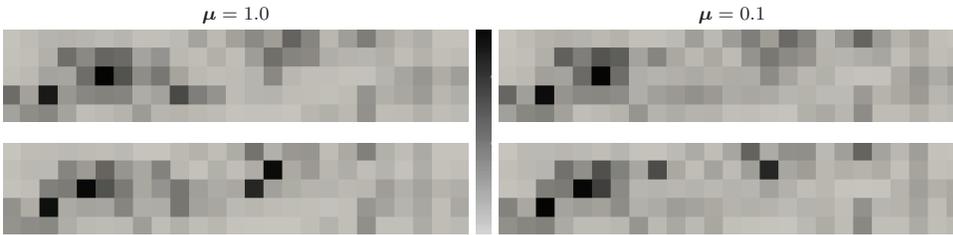


FIG. 3. Spatial distribution of the relative error contribution (top), $\|p(\mu) - p_h(\mu)\|_{\bar{\mu}, T} / \|p(\mu) - p_h(\mu)\|_{\bar{\mu}}$, and the relative estimated error contribution (bottom), $\eta^T(p_h(\mu); \mu, \bar{\mu}, \hat{\mu}) / (\sum_{T \in \mathcal{T}_H} \eta^T(p_h(\mu); \mu, \bar{\mu}, \hat{\mu})^2)^{-1/2}$, for all $T \in \mathcal{T}_H$, for the multiscale example in section 6.1 with $|\mathcal{T}_H| = 25 \times 5$ and $\bar{\mu} = \hat{\mu} = 0.1$ for parameters $\mu = 1$ (left column, light: $2.26 \cdot 10^{-3}$, dark: $3.78 \cdot 10^{-1}$) and $\mu = 0.1$ (right column, light: $4.02 \cdot 10^{-3}$, dark: $3.73 \cdot 10^{-1}$).

data functions. Fixing $\bar{\mu} = \hat{\mu} = 0.1$ (to obtain a fully off-line/on-line decomposable configuration with a fixed error norm) yields an average efficiency of 10.2 which is still very reasonable. In addition we observe a good agreement between the spatial distribution of the error and the estimator indicators in Figure 3.

6.2. Adaptive on-line enrichment. To demonstrate the proposed adaptive on-line enrichment Algorithm 5.2 and the flexibility of the LRBMS we study two distinct circumstances. We first consider a smooth academic example where we disable the greedy procedure and build the reduced bases only by local enrichment. The second example is again a multiscale one with global channels in the permeability and we allow for very few global solution snapshots and adaptively enrich afterwards.

For the orthonormalization algorithm ONB in the greedy Algorithm 5.1 we use the stabilized Gram–Schmidt procedure implemented² in `pyMOR` with respect to the scalar product given by $(p, q) \mapsto b^T(p, q; \bar{\mu}) + \sum_{e \in \mathcal{F}_h^T} b_p^T(p, q; \bar{\mu})$ on each $T \in \mathcal{T}_H$. In contrast to other possible basis extension algorithms (e.g., using a proper orthogonal decomposition) the Gram–Schmidt basis extension yields hierarchical local reduced bases. This is of particular interest in the context of on-line enrichment, since we do not need to update reduced quantities w.r.t existing basis vectors after enrichment. We always initialize the local reduced bases with the coarse DG basis of order up to one by setting $k_H = 1$ in the greedy Algorithm 5.1. We choose the same orthonormalization algorithm in the adaptive basis enrichment Algorithm 5.2 and use several marking strategies for MARK, depending on the circumstances (detailed below). Regarding the overlap for the local enrichment we always choose the overlapping subdomains $T_\delta \supset T$ to include T and all subdomains that touch it, thus choosing an overlap of $\mathcal{O}(H)$ as motivated in [32].

Since the error of any reduced solution $\|p(\mu) - p_{\text{red}}(\mu)\|_{\bar{\mu}}$ cannot be lower than the error of the corresponding detailed solution $\|p(\mu) - p_h(\mu)\|_{\bar{\mu}}$ we always choose Δ_{online} in Algorithm 5.2 to be slightly larger than $\max_{\mu \in \mathcal{P}_{\text{online}}} \eta(p_h(\mu); \mu, \bar{\mu}, \hat{\mu})$ in our experiments (see below), where $\mathcal{P}_{\text{online}} \subset \mathcal{P}$ is the set of all parameters we consider during the on-line phase. This is only necessary since we do not allow for an adaptation of τ_h ; combining our on-line adaptive LRBMS with the ideas of [58] would overcome this restriction.

Academic example. We again consider the academic example detailed in section 6.1 on fixed triangulations with $|\tau_h| = 131,072$ and $|\mathcal{T}_H| = 8 \times 8$ and choose the set of on-line parameters $\mathcal{P}_{\text{online}}$ to consist of 10 randomly chosen parameters $\mu_0, \dots, \mu_9 \in \mathcal{P}$. We set $N_{\text{greedy}} = 0$ and $k_H = 1$, thus disabling any Greedy search and initializing the local bases with the coarse DG basis of order up to one (consisting of 4 shape functions). In this setup $\max_{\mu \in \mathcal{P}_{\text{online}}} \eta(p_h(\mu); \mu, \bar{\mu}, \hat{\mu}) = 2.79 \cdot 10^{-2}$ and we choose $\Delta_{\text{online}} = 5 \cdot 10^{-2}$ in Algorithm 5.2.

We begin by choosing MARK such that $\tilde{\mathcal{T}}_H = \mathcal{T}_H$ (all coarse elements are marked; denoted by `uniform` in the following). For each parameter $\mu \in \mathcal{P}_{\text{online}}$ this results in a method that is similar to DD methods with overlapping subdomains. In contrast to traditional DD methods, however, we start with an initial coarse basis and perform a reduced solve before each iteration which helps to spatially spread information. As we observe in Figure 4, top left, it takes four enrichment steps to lower the estimated error for the first on-line parameter μ_0 below the desired tolerance and another two enrichment steps for the next on-line parameter μ_1 (which is $\max_{\mu \in \mathcal{P}_{\text{online}}} \mu$). With uniform marking this increases the local basis sizes from four to ten on each coarse element. The resulting coarse reduced space of dimension 640 is then sufficient to solve for the next four on-line parameters μ_2, \dots, μ_5 without enrichment. One additional enrichment phase is needed for μ_6 (which is $\min_{\mu \in \mathcal{P}_{\text{online}}} \mu$) and none for the remaining three on-line parameters. Note that while this uniform marking strategy may be

²http://docs.pymor.org/en/0.2.x/_modules/pymor/la/gram_schmidt.html#gram_schmidt

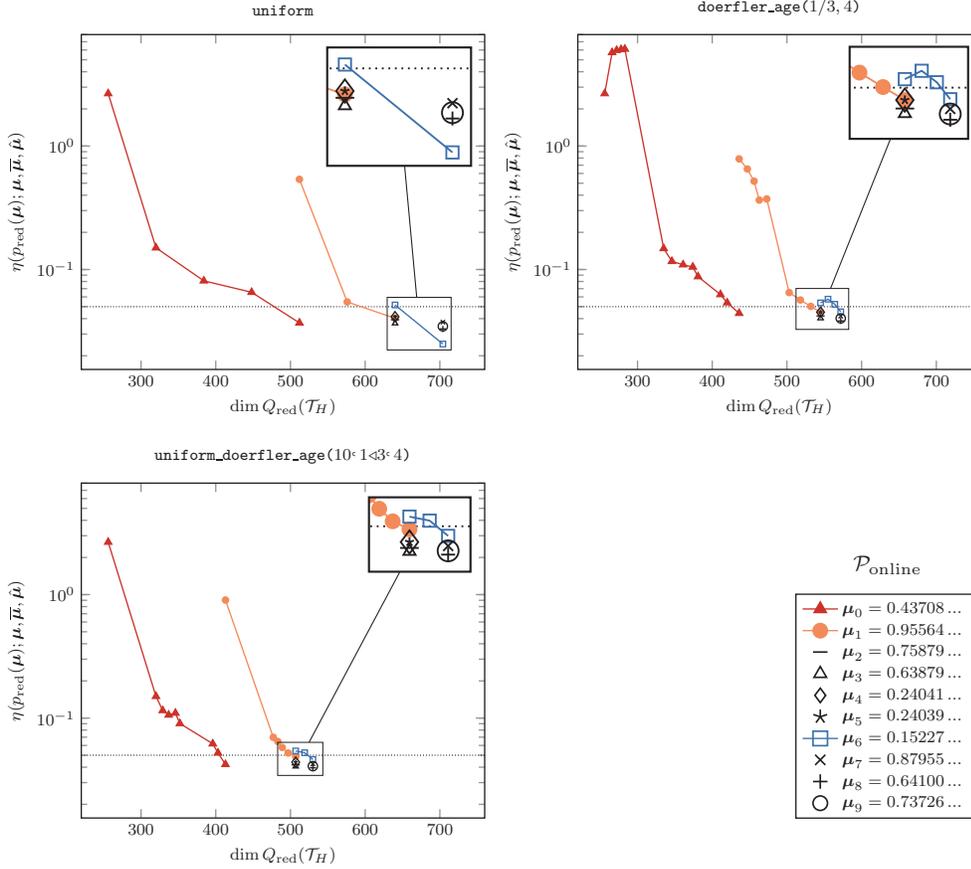
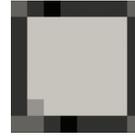


FIG. 4. Estimated error evolution during the adaptive on-line phase for the academic example in section 6.2 with $|\mathcal{T}_H| = 64$, $N_{\text{greedy}} = 0$, $k_H = 1$, $\Delta_{\text{online}} = 5 \cdot 10^{-2}$ (dotted line), and $\bar{\mu} = \hat{\mu} = 0.1$ for several marking strategies: uniform marking of all subdomains (top left), combined Dörfler marking with $\theta_{\text{doerf}} = 1/3$ and age-based marking with $N_{\text{age}} = 4$ (top right), and additional uniform marking while $\eta(p_{\text{red}}(\mu); \mu, \bar{\mu}, \hat{\mu}) > \theta_{\text{uni}} \Delta_{\text{online}}$ with $\theta_{\text{uni}} = 10$ (bottom left). With each strategy the local reduced bases are enriched according to Algorithm 5.2 for each subsequently processed on-line parameter μ_0, \dots, μ_9 (bottom right).

optimal in the number of enrichment steps it takes to reach the desired error for all on-line parameters it also leads to an unnecessarily high-dimensional coarse reduced space (of $\dim Q_{\text{red}}(\mathcal{T}_H) = 704$) and a high workload in each enrichment step.

Another popular choice in the context of adaptive mesh refinement is a Dörfler marking strategy (see [42] and the references therein), where we collect those coarse elements that contribute most to $\theta_{\text{doerf}} \sum_{T \in \mathcal{T}_H} \eta^T(\cdot; \mu, \bar{\mu}, \hat{\mu})^2$ in $\tilde{\mathcal{T}}_H \subseteq \mathcal{T}_H$, for a given user-dependent parameter $0 < \theta_{\text{doerf}} \leq 1$. In addition, similarly to [12, 29], we compute an “age” count of each $T \in \mathcal{T}_H$ (by counting how often an element was not marked for refinement) and mark those elements the age of which is larger than a prescribed $N_{\text{age}} \in \mathbb{N}$ (resetting the age count of each selected element). We denote this marking strategy by `doerfler_age`($\theta_{\text{doerf}}, N_{\text{age}}$). We found that a combination of $\theta_{\text{doerf}} = 1/3$ and $N_{\text{age}} = 4$ yielded the smallest overall basis size (of 572), compared to other combinations of θ_{doerf} and N_{age} and the `uniform` marking strategy. The number of elements marked per step range between five and 52 (over all on-line parameters and all enrichment steps; 23 steps in total) with a mean of 14 and a median of ten.

FIG. 5. Spatial distribution of the final sizes of the local reduced bases, $|\Phi^T|$ (light: 7, dark: 11), for all $T \in \mathcal{T}_H$ after the adaptive on-line phase for the academic example in section 6.2 with $\Omega = [-1, 1]^2$, $|\mathcal{T}_H| = 8 \times 8$, and the `uniform_dorfler_age`(10, 1/3, 4) marking strategy (see Figure 4, bottom left).



Out of these marked elements, between one and 44 have been marked due to their age in 12 of these 23 steps (with an average of 12 and a median of eight, taken over only those 12 steps where elements have been marked due to their age). We observe in Figure 4, top right, that the general behavior of the method with this marking strategy is similar to the one with `uniform` marking, with some commonalities and differences worth noting. First of all it also takes three enrichment phases to reach the prescribed error tolerance, and for the same parameters μ_0 , μ_1 , and μ_6 as above. But each of these enrichment phases naturally need more steps and large improvements can usually be observed after a lot of elements have been marked due to their age count (see for instance the fifth enrichment step for μ_0 or μ_1). In addition we observe that the estimated error for a parameter sometimes increases, in particular in the very beginning (see the first four steps for μ_0 , the fourth step for μ_1 , or the first step for μ_6). This is not troublesome since we can only expect a strict improvement in the energy norm induced by the bilinear form that is used in the enrichment. This shows, never the less, that there is still room for improvement, although using the `dorfler_age` marking we could reach a significantly lower overall basis size than using the `uniform` marking (572 vs. 704).

We propose a combination of the two strategies, namely, a uniform marking while the estimated error is far away from the desired tolerance, i.e., $\eta(p_{\text{red}}(\mu); \mu, \bar{\mu}, \hat{\mu}) > \theta_{\text{uni}} \Delta_{\text{online}}$ for some $\theta_{\text{uni}} > 0$, followed by a Dörfler and age-based marking as detailed above. We denote this marking strategy by `uniform_dorfler_age`($\theta_{\text{uni}}, \theta_{\text{dorfler}}, N_{\text{age}}$). As we observe in Figure 4, bottom left, this marking strategy combines advantages of both previous approaches, recovering the rapid error decrease of the `uniform` marking strategy far away from the desired tolerance (see the first step for μ_0 and μ_1) while yielding the smallest overall basis size of 530 (using a factor of $\theta_{\text{uni}} = 10$) due to the `dorfler_age` marking strategy. The smoothness and symmetry of the problem is reflected in the spatial distribution of the final local basis sizes (see Figure 5).

Multiscale example. We again consider the multiscale example detailed in section 6.1 on fixed triangulations with $|\mathcal{T}_h| = 1,014,000$ and $|\mathcal{T}_H| = 25 \times 5$ and choose the set of on-line parameters $\mathcal{P}_{\text{online}}$ to consist of the same 10 randomly chosen parameters $\mu_0, \dots, \mu_9 \in \mathcal{P}$ as in the previous example. In this setup $\max_{\mu \in \mathcal{P}_{\text{online}}} \eta(p_h(\mu); \mu, \bar{\mu}, \hat{\mu}) = 2.66$ and we choose $\Delta_{\text{online}} = 2$ in Algorithm 5.2.

We first set $N_{\text{greedy}} = 0$ and $k_H = 1$ (thus disabling any greedy search in the off-line phase and initializing the local bases with the coarse DG basis of order up to one); in the on-line phase we use the `uniform` marking strategy (see above). As we observe in Figure 6, top, it takes 129 enrichment steps to lower the estimated error below the desired tolerance for the first on-line parameter μ_0 . This is not surprising since the data functions exhibit strong multiscale features and nonlocal high-conductivity channels connecting domain boundaries (see Figure 2). After this extensive enrichment it takes 12 steps for μ_1 and none or one enrichment step to reach the desired tolerance for the other on-line parameters. The resulting coarse reduced space is of size 10,749 (with an average of 86 basis functions per subdomain), which is clearly not optimal. Although each subdomain was marked for enrichment, the sizes of the final local reduced bases may differ since the local Gram–Schmidt

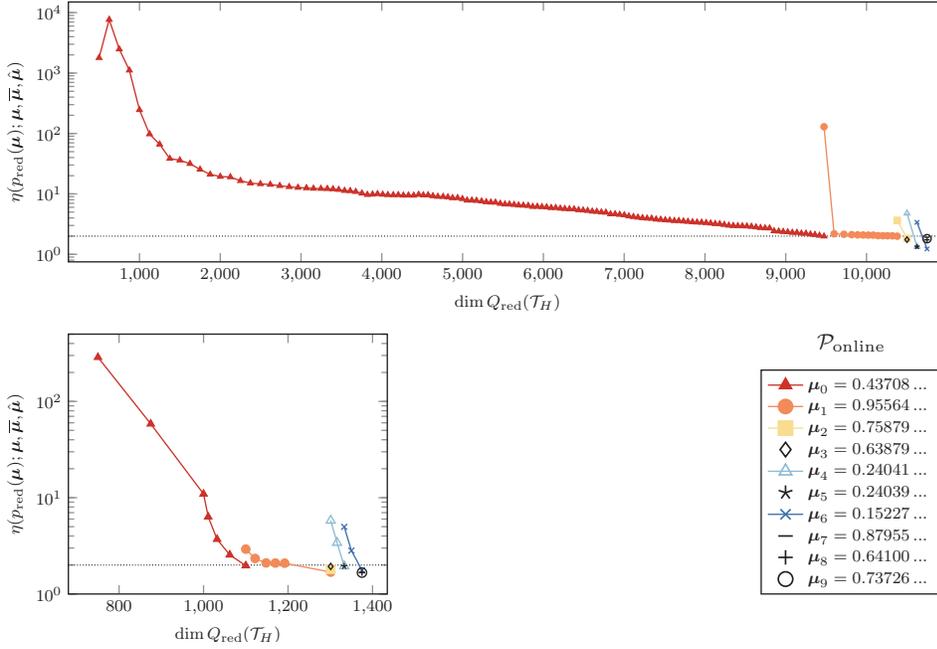


FIG. 6. *Estimated error evolution during the adaptive on-line phase for the multiscale example in section 6.2 with $|\mathcal{T}_H| = 125$, $k_H = 1$, $\Delta_{online} = 2$ (dotted line), $\bar{\mu} = \hat{\mu} = 0.1$ for different on-line and off-line strategies: no global snapshot (greedy search disabled, $N_{greedy} = 0$) during the off-line phase, uniform marking during the on-line phase (top) and two global snapshots (greedy search on $\mathcal{P}_{train} = \{0.1, 1\}$, $N_{greedy} = 2$) and combined uniform marking while $\eta(p_{red}(\mu); \mu, \bar{\mu}, \hat{\mu}) > \theta_{uni} \Delta_{online}$ with $\theta_{uni} = 10$, Dörfler marking with $\theta_{doerf} = 0.85$, and age-based marking with $N_{age} = 4$ (bottom left); note the different scales. With each strategy the local reduced bases are enriched according to Algorithm 5.2 while subsequently processing the on-line parameters μ_0, \dots, μ_9 (bottom right).*

basis extension may reject updates (if the added basis function is locally not linearly independent). As we observe in Figure 7, left, this is indeed the case with local basis sizes ranging between 24 and 148.

To remedy the situation we allow for two global snapshots during the off-line phase (setting $N_{greedy} = 2$, $\mathcal{P}_{train} = \{0.1, 1\}$) and use the adaptive `uniform.dörfler.age` marking strategy (see above) in the on-line phase. With two global solution snapshots incorporated in the basis the situation improves significantly, as we observe in Figure 6, bottom left, and there is no qualitative difference of the evolution of the estimated error during the adaptive on-line phase between the academic example studied above and this highly heterogeneous multiscale example (compare Figure 4, bottom left). In total we observe only two enrichment steps with uniform marking (see the first two of the six steps for μ_0). The number of elements marked range between 11 and 110 (over all on-line parameters and all but the first two enrichment steps) with a mean of 29 and a median of 22. Of these marked elements only once have 87 out of 110 elements been marked due to their age (see the last step for μ_1). Overall we could reach a significantly lower overall basis size than in the previous setup (1,375 versus 10,749) and the sizes of the final local bases range between only nine and 2 (compared to 24 to 148 above). We also observe in Figure 7, right, that the spatial distribution of the basis sizes follows the spatial structure of the data functions (compare Figures 1, 2), which nicely shows the localization qualities of our error estimator.



FIG. 7. *Spatial distribution of the final sizes of the local reduced bases, $|\Phi^T|$ for all $T \in \mathcal{T}_H$, after the adaptive online phase for the multiscale example in section 6.2 with $\Omega = [0, 5] \times [0, 1]$, $|\mathcal{T}_H| = 25 \times 5$ for the two strategies shown in Figure 6: no global snapshot with uniform enrichment (left, light: 24, dark: 148) and two global snapshots with adaptive enrichment (right, light: 9, dark: 20). Note the pronounced structure (right) reflecting the spatial structure of the data functions (compare Figures 1 and 2).*

7. Conclusion. In this contribution we equipped the LRBMS method with a rigorous and efficient localized a posteriori error estimator and a new adaptive on-line enrichment procedure. The LRBMS method was originally introduced in [36, 7] with a standard residual based a posteriori error estimator for the reduced solution against the high-dimensional solution that could not give local error information and was computationally very costly during the off-line phase of the computation. In addition the method was restricted to the SWIPDG discretization. We extended the original LRBMS method to locally allow for arbitrary discretizations of at least first order within each subdomain. In addition, we proposed a new error estimator for the high-dimensional as well as the reduced solution against the weak solution in the spirit of [25, 26], based on a local conservative flux reconstruction. The proposed estimator yields a guaranteed upper bound (involving no unknown constants), it is locally efficient, and can be computed using local information only. The estimator is efficiently off-line/on-line decomposable and allows us to estimate the error in parameter dependent norms.

Using the local information of the estimator we proposed a new adaptive on-line enrichment strategy, where we extend the local reduced bases by solutions of local corrector problems posed on overlapping subdomains. This is done during the on-line phase, thus deviating from the strict off-line/on-line separation of traditional RB methods, but only on those subdomains that have been selected by the estimator. This strategy allows us to guarantee the quality of the reduced solution during the on-line phase, even if an insufficient reduced basis had been prepared in the off-line phase (in contrast to traditional RB methods where the reduced basis is fixed for the whole on-line phase).

We provide numerical experiments to demonstrate the performance of the proposed estimator in the parametric setting for the high-dimensional discretization for an academic as well as a highly heterogeneous multiscale example, where the estimator proves to be very efficient (see section 6.1). We also demonstrate the performance of the newly proposed adaptive on-line enrichment strategy for both examples. It is particular noteworthy that only very few global solution snapshots were needed during the off-line phase for the multiscale example to sufficiently prepare a reduced basis that was then adaptively enriched during the off-line phase (see section 6.2). For the academic example of smooth data functions without any multi-scale features no global solution snapshots were required at all.

First findings regarding the new estimator have been published in [45] while first steps in the direction of on-line enrichment have been discussed in [8].

REFERENCES

- [1] A. ABDULLE, *Discontinuous Galerkin finite element heterogeneous multiscale method for elliptic problems with multiple scales*, Math. Comp., 81 (2012), pp. 687–713.
- [2] A. ABDULLE AND Y. BAI, *Adaptive reduced basis finite element heterogeneous multiscale method*, Comput. Methods Appl. Mech. Engrg., 257 (2013), pp. 203–220.
- [3] A. ABDULLE AND Y. BAI, *Reduced-order modelling numerical homogenization*, R. Soc. Lond. Ser. Philos. Trans. A Math. Phys. Eng. Sci., 372 (2014), 20130388.
- [4] A. ABDULLE, W. E. B. ENQUIST, AND E. VANDEN-EIJNDEN, *The heterogeneous multiscale method*, Acta Numer., 21 (2012), pp. 1–87.
- [5] A. ABDULLE AND P. HENNING, *A Reduced Basis Localized Orthogonal Decomposition*, J. Comput. Phys., 295 (2015), pp. 379–401.
- [6] A. ABDULLE AND M. E. HUBER, *Discontinuous Galerkin finite element heterogeneous multiscale method for advection-diffusion problems with multiple scales*, Numer. Math., 126 (2014), pp. 589–633.
- [7] F. ALBRECHT, B. HAASDONK, S. KAULMANN, AND M. OHLBERGER, *The localized reduced basis multiscale method*, in Proceedings of Algorithmy 2012, Conference on Scientific Computing, Vysoke Tatry, Podbanske, 2012, Slovak University of Technology, Bratislava, 2012, pp. 393–403.
- [8] F. ALBRECHT AND M. OHLBERGER, *The localized reduced basis multi-scale method with online enrichment*, Oberwolfach Rep., 7 (2013), pp. 12–15.
- [9] M. BARRAULT, Y. MADAY, N. C. NGUYEN, AND A. T. PATERA, *An ‘empirical interpolation’ method: Application to efficient reduced-basis discretization of partial differential equations*, C. R. Math. Acad. Sci. Paris, 339 (2004), pp. 667–672.
- [10] P. BASTIAN, M. BLATT, A. DEDNER, C. ENGWER, R. KLÖFKORN, R. KORNUBER, M. OHLBERGER, AND O. SANDER, *A generic grid interface for parallel and adaptive scientific computing. II. Implementation and tests in DUNE*, Computing, 82 (2008), pp. 121–138.
- [11] P. BASTIAN, M. BLATT, A. DEDNER, C. ENGWER, R. KLÖFKORN, M. OHLBERGER, AND O. SANDER, *A generic grid interface for parallel and adaptive scientific computing. I. Abstract framework*, Computing, 82 (2008), pp. 103–119.
- [12] P. BINEV, W. DAHMEN, AND R. DEVORE, *Adaptive finite element methods with convergence rates*, Numer. Math., 97 (2004), pp. 219–268.
- [13] S. BOYAVAL, *Reduced-basis approach for homogenization beyond the periodic setting*, Multiscale Model. Simul., 7 (2008), pp. 466–494.
- [14] C. A. BREWER, *Colorbrewer*, <http://colorbrewer2.org/> (2015).
- [15] E. BURMAN AND P. ZUNINO, *A domain decomposition method based on weighted interior penalties for advection-diffusion-reaction problems*, SIAM J. Numer. Anal., 44 (2006), pp. 1612–1638.
- [16] Z. CHEN, G. HUAN, AND Y. MA, *Computational Methods for Multiphase Flows in Porous Media*, Comput. Sci. Eng. 2, SIAM, Philadelphia, 2006.
- [17] A. DEDNER, R. KLÖFKORN, AND M. NOLTE, *The DUNE-ALUGrid Module*, preprint, arXiv:1407.6954 [cs.MS], 2014.
- [18] A. DEDNER, R. KLÖFKORN, M. NOLTE, AND M. OHLBERGER, *A generic interface for parallel and adaptive discretization schemes: Abstraction principles and the DUNE-FEM module*, Computing, 90 (2010), pp. 165–196.
- [19] M. DROHMANN, B. HAASDONK, AND M. OHLBERGER, *Reduced basis approximation for nonlinear parametrized evolution equations based on empirical operator interpolation*, SIAM J. Sci. Comput., 34 (2012), pp. A937–A969.
- [20] Y. EFENDIEV, J. GALVIS, AND T. Y. HOU, *Generalized multiscale finite element methods (GMS-FEM)*, J. Comput. Phys., 251 (2013), pp. 116–135.
- [21] DUNE-PDELAB, <http://www.dune-project.org/pdelab/index.html> (2014).
- [22] Y. EFENDIEV AND T. Y. HOU, *Multiscale Finite Element Methods*, Surv. Tutor. Appl. Math. Sci. 4, Springer, New York, 2009.
- [23] Y. EPSHTEYN AND B. RIVIÈRE, *Estimation of penalty parameters for symmetric interior penalty Galerkin methods*, J. Comput. Appl. Math., 206 (2007), pp. 843–872.
- [24] A. ERN AND A. F. STEPHANSEN, *A posteriori energy-norm error estimates for advection-diffusion equations approximated by weighted interior penalty methods*, J. Comput. Math., 26 (2008), pp. 488–510.
- [25] A. ERN, A. F. STEPHANSEN, AND M. VOHRALÍK, *Improved Energy Norm A Posteriori Error Estimation Based on Flux Reconstruction for Discontinuous Galerkin Methods*, preprint, HAL archives-ouvertes, hal-0019354UV1, 2007.

- [26] A. ERN, A. F. STEPHANSEN, AND M. VOHRALÍK, *Guaranteed and robust discontinuous Galerkin a posteriori error estimates for convection–diffusion–reaction problems*, J. Comput. Appl. Math., 234 (2010), pp. 114–130.
- [27] A. ERN, A. F. STEPHANSEN, AND P. ZUNINO, *A discontinuous Galerkin method with weighted averages for advection–diffusion equations with locally small and anisotropic diffusivity*, IMA J. Numer. Anal., 29 (2009), pp. 235–256.
- [28] C. FEUERSÄNGER, *PGF and TikZ–Graphic Systems for TeX*, <http://sourceforge.net/projects/pgf/> (version 1.11).
- [29] B. HAASDONK, M. DIHLMANN, AND M. OHLBERGER, *A training set and multiple bases generation approach for parameterized model reduction based on adaptive grids in parameter space*, Math. Comput. Model. Dyn. Syst., 17 (2011), pp. 423–442.
- [30] H. HAJIBEYGI AND P. JENNY, *Adaptive iterative multiscale finite volume method*, J. Comput. Phys., 230 (2011), pp. 628–643.
- [31] P. HENNING AND M. OHLBERGER, *The heterogeneous multiscale finite element method for elliptic homogenization problems in perforated domains*, Numer. Math., 113 (2009), pp. 601–629.
- [32] P. HENNING AND D. PETERSEIM, *Oversampling for the multiscale finite element method*, Multiscale Model. Simul., 11 (2013), pp. 1149–1175.
- [33] L. IAPICHINO, A. QUARTERONI, AND G. ROZZA, *A reduced basis hybrid method for the coupling of parametrized domains represented by fluidic networks*, Comput. Methods Appl. Mech. Engrg., 221/222 (2012), pp. 63–82.
- [34] L. IAPICHINO, A. QUARTERONI, G. ROZZA, AND S. VOLKWEIN, *Reduced basis method for the Stokes equations in decomposable parametrized domains using greedy optimization*, in Progress in Industrial Mathematics ECMI 2014, Springer, Heidelberg, 2014, pp. 1–7.
- [35] O. A. KARAKASHIAN AND F. PASCAL, *A posteriori error estimates for a discontinuous Galerkin approximation of second-order elliptic problems*, SIAM J. Numer. Anal., 41 (2003), pp. 2374–2399.
- [36] S. KAULMANN, M. OHLBERGER, AND B. HAASDONK, *A new local reduced basis discontinuous Galerkin approach for heterogeneous multiscale problems*, C. R. Math. Acad. Sci. Paris, 349 (2011), pp. 1233–1238.
- [37] M. G. LARSON AND A. MÁLQVIST, *Adaptive variational multiscale methods based on a posteriori error estimation: Energy norm estimates for elliptic problems*, Comput. Methods Appl. Mech. Engrg., 196 (2007), pp. 2313–2324.
- [38] A. E. LØVGREN, Y. MADAY, AND E. M. RØNQUIST, *The reduced basis element method for fluid flows*, in Analysis and Simulation of Fluid Dynamics, Adv. Math. Fluid Mech., Birkhäuser, Basel, 2007, pp. 129–154.
- [39] Y. MADAY AND E. M. RØNQUIST, *A reduced-basis element method*, C. R. Math. Acad. Sci. Paris, 335 (2002), pp. 195–200.
- [40] R. MILK, S. RAVE, AND F. SCHINDLER, *pyMOR: Companion Software to OS2015*, <http://dx.doi.org/10.5281/zenodo.13897>.
- [41] R. MILK AND F. SCHINDLER, *dune-stuff: Companion Software to OS2015*, <http://dx.doi.org/10.5281/zenodo.13891>.
- [42] P. MORIN, R. H. NOCHETTO, AND K. G. SIEBERT, *Convergence of adaptive finite element methods*, SIAM Rev., 44 (2002), pp. 631–658 (2003).
- [43] J. M. NORDBOTTEN, *Adaptive variational multiscale methods for multiphase flow in porous media*, Multiscale Model. Simul., 7 (2009), pp. 1455–1473.
- [44] M. OHLBERGER, *Error control based model reduction for multiscale problems*, in Proceedings of Algoritmy 2012, Conference on Scientific Computing, Vysoke Tatry, Podbanske, 2012, Slovak University of Technology, Bratislava, 2012, pp. 1–10.
- [45] M. OHLBERGER AND F. SCHINDLER, *A-posteriori error estimates for the localized reduced basis multi-scale method*, in Finite Volumes for Complex Applications VII-Methods and Theoretical Aspects, J. Fuhrmann, M. Ohlberger, and C. Rohde, eds., Springer Proc. Math. Stat. 77, Springer, Berlin, 2014, pp. 421–429.
- [46] G. V. PENCHEVA, M. VOHRALÍK, M. F. WHEELER, AND T. WILDEY, *Robust a posteriori error control and adaptivity for multiscale, multinumercs, and mortar coupling*, SIAM J. Numer. Anal., 51 (2013), pp. 526–554.
- [47] A. QUARTERONI AND A. VALLI, *Domain Decomposition Methods for Partial Differential Equations*, Oxford University Press, Oxford, 1999.
- [48] S. RAVE AND F. SCHINDLER, *dune-pymor: Companion Software to OS2015*, <http://dx.doi.org/10.5281/zenodo.13896>.
- [49] F. SCHINDLER, *dune-gdt: Companion Software to OS2015*, <http://dx.doi.org/10.5281/zenodo.13893>.

- [50] F. SCHINDLER, *dune-grid-multiscale: Companion Software to OS2015*, <http://dx.doi.org/10.5281/zenodo.13894>.
- [51] F. SCHINDLER, *dune-hdd: Comparison Software to OS2015*, <http://dx.doi.org/10.5281/zenodo.13895>.
- [52] K. SMETANA, *A new certification framework for the port reduced static condensation reduced basis element method*, *Comput. Methods Appl. Mech. Engrg.*, 283 (2015), pp. 352–383.
- [53] SOCIETY OF PETROLEUM ENGINEERS, *SPE Comparative Solution Project*, <http://www.spe.org/web/csp/index.html> (2001).
- [54] T. TANTAU, *PGF and TikZ–Graphic Systems for TeX*, <http://sourceforge.net/projects/pgf/> (version 3.0.0).
- [55] T. TANTAU, *Graph drawing in TikZ*, *J. Graph Algorithms Appl.*, 17 (2013), pp. 495–513.
- [56] K. VEROY, C. PRUD’HOMME, AND A. T. PATERA, *Reduced-basis approximation of the viscous Burgers equation: Rigorous a posteriori error bounds*, *C. R. Math. Acad. Sci. Paris*, 337 (2003), pp. 619–624.
- [57] M. VOHRALÍK, *A posteriori error estimates for lowest-order mixed finite element discretizations of convection-diffusion-reaction equations*, *SIAM J. Numer. Anal.*, 45 (2007), pp. 1570–1599.
- [58] M. YANO, *A minimum-residual mixed reduced basis method: Exact residual certification and simultaneous finite-element reduced-basis refinement*, *ESAIM Math. Model. Numer. Anal.*, to appear.