Adaptive Multiresolution Discontinuous Galerkin Schemes for Conservation Laws

Nune Hovhannisyan*† Siegfried Müller‡ Roland Schäfer‡

September 16, 2010

A multiresolution-based adaptation concept is proposed that aims at accelerating Discontinuous Galerkin schemes applied to nonlinear hyperbolic conservation laws. Opposite to standard adaptation concepts no error estimates are needed to tag mesh elements for refinement. Instead of this, a multiresolution analysis is performed on a hierarchy of nested grids for the data given on a uniformly refined mesh. This provides difference information between successive refinement levels, that may become negligibly small in regions where the solution is locally smooth. Applying hard thresholding the data are highly compressed and local grid adaptation is triggered by the remaining significant coefficients. A central mathematical problem addressed in this work is then to show that the adaptive solution is of the same accuracy as the reference solution on a uniformly refined mesh. Numerical comparisons demonstrate the efficiency of the concept and provide reliable estimates of the actual error in the numerical solution.

Conservation laws, Discontinuous Galerkin schemes, grid adaptation, multiwavelets.
35L65, 65M12, 65M60, 65T60, 74S05

*Faculty of Informatics and Applied Mathematics, Alex Manoogian 1, Yerevan 0044, Armenia (alnune03@yahoo.com)

†This work has been performed with funding by the Deutsche Forschungsgemeinschaft in the Collaborative Research Center SFB 401 "Flow Modulation and Fluid-Structure Interaction at Airplane Wings" of RWTH Aachen University, Germany.

‡Institut für Geometrie und Praktische Mathematik, RWTH Aachen University, Templergraben 55, 52056 Aachen, Germany (mueller,schaefer@igpm.rwth-aachen.de)

Contents

1 Introduction

The solution of a nonlinear hyperbolic conservation law typically develops discontinuities. In order to avoid instabilities in the numerical discretization, Finite Volume (FV) schemes are frequently used. These have been developed and investigated for many decades. An overview can be found in the textbooks of Godlewski and Raviart [20], Kröner [30] and Leveque [32]. Although these schemes have proven to be very robust in practice, they suffer from discretization stencils that are growing with increasing order. In particular, on unstructured meshes this turns out to be a severe drawback in particular on unstructured meshes. This problem can be overcome by using Discontinuous Galerkin (DG) schemes which are locally conservative, stable and high-order accurate. These methods can easily handle complex geometries and meshes with hanging nodes as well as approximations with locally changing polynomial degree in different elements. In particular, they are easy to implement and allow for an efficient parallelization. After their introduction in 1973 by Reed and Hill [37] a major development was carried out by Cockburn et al. [11, 10, 8, 12, 13]. A review on Runge-Kutta DG schemes for convection-dominated problems can be found in [14]. By now DG methods have been applied to partial differential equations and ordinary differential equations [9].

Due to the formation of discontinuities the solution of a nonlinear hyperbolic conservation law often exhibits locally steep gradients and large regions where it is smooth. To account for the highly nonuniform spatial behavior, we need numerical schemes that adequately resolve the

different scales, i.e., use a high resolution only near sharp transition regions and singularities but a moderate resolution in regions with smooth, slowly varying behavior of the solution. For this purpose, strategies for DG discretizations have been discussed or are under current investigation that aim at adapting the internal degrees of freedom, such as the spatial and temporal discretization or the discretization order, to the local behavior of the flow field. So far numerous refinement indicators have been developed that are based on interpolation error estimates of some key quantity using some a priori knowledge of the solution. Although these concepts turn out to be very efficient in practice, they offer no reliable error control. For this purpose, error estimates are needed which provide a control of the actual error in the approximate solution. Bey and Oden [4] obtained both a priori and a posteriori error estimates for adaptive strategies. By now many types of a posteriori error estimates have been designed to control the adaptive process. A lot of work has been done by Flaherty et al. [1, 38], Houston et al. [25, 24, 26], Dedner et al. [19], and recently Mavriplis et al. [43].

All these approaches have in common that they aim at estimating the error of the solution. However, in general there are no mathematical rigorous error estimates available for nonlinear systems of conservation laws arising, for instance, from the balance equations of continuum mechanics. Therefore we propose an alternative adaptation strategy that does *not* rely on the existence of some error estimator. The rationale behind its design is to *accelerate* a given DG scheme or a FV scheme (reference scheme) on a uniformly refined mesh (reference mesh) through computing actually only on a locally refined adapted subgrid, while *preserving* (up to a fixed constant multiple) the accuracy of the discretization on the full uniform grid. For this purpose, a multiresolution analysis (MRA) is performed, where the data corresponding to the current solution are represented as data on some coarse level and the fine scale information is encoded in arrays of detail coefficients of ascending resolution. The new data format reveals insight into the local behaviour of the solution. It can be shown that the details become small with increasing refinement level when the underlying function is smooth. As suggested by this so-called cancellation property, we may determine a locally refined grid performing data compression on the array of detail coefficients using hard thresholding. This significantly reduces the complexity of the data. Based on the thresholded array local grid adaptation is performed, where we refine an element whenever there exists a significant detail. Of course, the crux in this context is to arrange this procedure in such a way that at no stage of the computation there is ever made use of the fully refined mesh. A central mathematical problem is then to show that the solution on the adapted mesh is of the same accuracy as the solution on the reference mesh.

Such multiresolution-based mesh adaptation methods using biorthogonal wavelets [15] have been introduced in [33, 16] and have been quite successful with FV solvers for compressible fluid flow, see Bramkamp et al. [5]. A comprehensive review on their development can be found in [34] and [17] provides some overview on recent trends. Note, that the rate of decay of the details fastens with increasing number of vanishing moments of the wavelets, i.e., more details may be discarded in smooth regions and the adaptive mesh becomes coarser. However, for biorthogonal wavelets to realize more vanishing moments requires to extend the support of the wavelet functions. Its construction, in particular, becomes even more complicate on unstructured grid hierarchies. Therefore it is natural to extend the multiresolution-based mesh adaptation concept to higher order DG discretizations using so-called multiwavelets [41, 28]. Multiwavelets allow for higher order vanishing moments, while being defined on one mesh element.

In the present work, we develop the concept of multiresolution-based DG (MR-DG) schemes,

where for the sake of analysis we focus on scalar one-dimensional conservation laws. We emphasize that none of the conceptual ingredients are restricted to this frame but can be extended straight-forwardly to multidimensional systems of conservation laws.

For this purpose, we first summarize in Section 2 the basic ingredients of Runge-Kutta DG schemes and the main convergence results. These schemes will be used as the reference scheme that will provide us with data on a reference mesh. In Section 3 we perform a MRA of these data. For this purpose we embed the reference mesh in a hierarchy of nested grids. In Section 4, we then derive evolution equations for these detail coefficients by means of the MRA. This set of equations will be reduced significantly by data compression performing hard thresholding. In order to control the threshold error that might accumulate in each time step, we have to make sure that significant information is well-resolved on both the old and the new time level. For this purpose we need to predict significant detail coefficients from the data of the old time level, where we have to analyse the evolution of the detail coefficients. This evolution process is strongly intertwined with the limiting process for the higher order coefficients. This is investigated in detail in Section 5, where we prove that the accumulative threshold error is uniformly bounded. Since for scalar problems there is an estimate available for the discretization error of the reference DG scheme [11], we may then choose a priorily a threshold value that allows to balance the discretization error with the threshold error. Finally we have implemented the MR-DG scheme for scalar one-dimensional conservation laws. Numerical comparisons in Section 6 demonstrate the efficiency of our concept by verifying the reliability of the estimates of the actual error of numerical solutions.

2 The reference DG scheme

In order to simplify the notation, we confine ourselves to one space dimension. We like to point out once more that the concepts extend to higher dimensional problems as well. In the framework of adaptive MR-FV schemes this has been realized for the multidimensional case and has been successfully applied to complex configurations in fluid dynamics, cf. [5]. Here we consider the initial value problem for a scalar *inhomogeneous* conservation law

$$\begin{aligned} &u_t(t,x) + (f(u(t,x)))_x = s(u(t,x)), \quad t>0,\, x\in\mathbb{R}, \\ &u(0,x) = u_0(x), \quad x\in\mathbb{R}. \end{aligned} \tag{2.1}$$

If the initial data $u_0 \in BV(\mathbb{R}) \cap L^\infty(\mathbb{R}) \cap L^1(\mathbb{R})$, the flux $f \in C^1(\mathbb{R})$ and the source term $s : \mathbb{R} \to \mathbb{R}$ satisfying a global Lipschitz condition with Lipschitz constant L_s and $s(0) = 0$, then there exists a unique entropy solution $u \in L^\infty(0,T,L^1_{loc}(\mathbb{R}))$ for all $T > 0$, see [35, 31, 42]. We will assume additionally that u_0 is compactly supported.

Weak formulation The entropy solution is approximated by a DG scheme. Since computations are only performed on bounded domains and convergence analysis is only performed for compact sets, we confine ourselves to uniform finite discretizations of $\Omega := [a,b] = \bigcup_{k=0}^{N-1} V_k$, where the cells $V_k := [x_k, x_{k+1}]$, $k \in \mathcal{I}_h := \{0,\dots,N-1\}$, are determined by the discretization points $x_k := a + k\,h$ with uniform spatial discretization $h = (b-a)/N$.

On this discretization we introduce the space $S_h^p := \{w \in L^2([a,b]) :\ w|_{V_k} \in \Pi_{p-1},\ \forall\ k \in \mathcal{I}_h\}$ of piecewise polynomial functions of degree less than p. For this finite-dimensional space we introduce two sets of basis functions, $\Phi_h := \{\varphi_{k,i}\}_{k\in\mathcal{I}_h, i\in\mathcal{P}}$ and $\tilde{\Phi}_h := \{\tilde{\varphi}_{k,i}\}_{k\in\mathcal{I}_h, i\in\mathcal{P}}$ with

$\mathcal{P} := \{0, \dots, p-1\}$ such that $S_h^p = \operatorname{span} \Phi_h = \operatorname{span} \tilde{\Phi}_h$. We assume that these functions fulfil the biorthogonality relation

$$\langle \varphi_{k',i'}, \tilde{\varphi}_{k,i} \rangle_\Omega = \delta_{i,i'} \, \delta_{k,k'} \tag{2.2}$$

and are compactly supported, i.e., supp $\varphi_{k,i} =$ supp $\tilde{\varphi}_{k,i} = V_k$. Here $\langle f, g \rangle_\Omega := \int_\Omega f(x) \, g(x) \, dx$ denotes the standard L^2-inner product. A typical example are the shifts and translates of the Legendre polynomials normalized with respect to L^∞ and L^1, respectively. In order to derive a DG discretization of the initial value problem (2.1), we now assume that the approximate solution can be written as an expansion of the basis Φ_h, i.e.,

$$u_h(t, \cdot) = \sum_{k \in \mathcal{I}_h} \sum_{i \in \mathcal{P}} v_{k,i}(t) \varphi_{k,i}(\cdot) \in S_h^p, \tag{2.3}$$

where the coefficients $v_{k,i}$ are determined by the biorthogonality relation (2.2) as

$$v_{k,i}(t) = \langle u_h(t, \cdot), \tilde{\varphi}_{k,i} \rangle_\Omega.$$

Multiplying (2.1) by $\tilde{\varphi}_{k,i}$ and integrating over its support V_k we obtain

$$\langle (u_h(t, \cdot))_t, \tilde{\varphi}_{k,i} \rangle_{V_k} + \langle (f(u_h(t, \cdot)))_x, \tilde{\varphi}_{k,i} \rangle_{V_k} = \langle s(u_h(t, \cdot)), \tilde{\varphi}_{k,i} \rangle_{V_k} \tag{2.4}$$

Inserting (2.3) into (2.4) and performing integration by parts we end up with

$$\frac{d}{dt} v_{k,i}(t) + f(u_h(t, x)) \tilde{\varphi}_{k,i}(x) \big|_{x_k^+}^{x_{k+1}^-} - \langle f(u_h), \tilde{\varphi}'_{k,i} \rangle_{V_k} = \langle s(u_h), \tilde{\varphi}_{k,i} \rangle_{V_k}, \tag{2.5}$$

where we employ biorthogonality (2.2) and the local support of the basis functions. Here $x^\pm$ always denote the right-sided $(+)$ and left-sided $(-)$ limit, respectively. Since the basis functions and, hence, the approximate solution u_h suffer from jumps at the interval points x_k, the flux evaluations $f(u_h(t, x^\pm))$ are replaced by numerical fluxes $F(u_h(t, x_k^-), u_h(t, x_k^+))$, where $F : \mathbb{R}^2 \to \mathbb{R}$ is a two-point flux function such that

1. F is consistent with the flux f, i.e., $F(v, v) = f(v)$ for all $v \in \mathbb{R}$,
2. F is Lipschitz continuous, i.e., $|F(u_1, v_1) - F(u_2, v_2)| \leq L_1 |u_1 - u_2| + L_2 |v_1 - v_2|$ for all $|u_i|, \ |v_i| \leq C$, $i = 1, 2$, with C arbitrary but fixed;
3. F is monotone, i.e., nondecreasing in the first argument and nonincreasing in the second argument.

Typical examples are the flux functions due to Engquist-Osher, (local) Lax-Friedrichs, Godunov, Roe with entropy fix, cf. [20]. In general, monotonicity only holds if a Courant-Friedrichs-Levy (CFL) condition is satisfied for the time and space discretization. Finally, we obtain the semi-discrete DG method

$$\frac{d}{dt} v_{k,i}(t) = - F(u_h(t, x), u_h(t, x)) \tilde{\varphi}_{k,i}(x) \big|_{x_k^+}^{x_{k+1}^-} + \langle f(u_h), \tilde{\varphi}'_{k,i} \rangle_{V_k} + \langle s(u_h), \tilde{\varphi}_{k,i} \rangle_{V_k}. \tag{2.6}$$

Convergence of DG scheme in the mean The building block for a higher order time discretization is a forward Euler step applied to (2.6). For this purpose we introduce the time discretization $t_{n+1} := t_n + \tau$, where for reasons of simplicity the time step $\tau > 0$ is assumed to be constant. Then the forward Euler step reads

$$v_{k,i}^{n+1} = v_{k,i}^n - \tau \left(B_{k,i}^n - G_{k,i}^n - S_{k,i}^n \right), \tag{2.7}$$

where the numerical flux F_k^n, the numerical flux balance $B_{k,i}^n$, the flux quadrature $G_{k,i}^n$ and the source quadrature $S_{k,i}^n$ are defined by

$$\begin{aligned} F_k^n &:= F(u_h(t^n, x_k^-), u_h(t^n, x_k^+)), && (2.8)\\ B_{k,i}^n &:= F_{k+1}^n \cdot \tilde{\varphi}_{k,i}(x_{k+1}^-) - F_k^n \cdot \tilde{\varphi}_{k,i}(x_k^+), && (2.9)\\ G_{k,i}^n &:= \langle f(u_h(t^n, \cdot)), \tilde{\varphi}'_{k,i} \rangle_{V_k}, \quad S_{k,i}^n := \langle s(u_h(t^n, \cdot)), \tilde{\varphi}_{k,i} \rangle_{V_k}. && (2.10) \end{aligned}$$

Typically, convergence of (2.7) is investigated for the averaged function of piecewise constant data

$$\hat{u}_h(t, x) := \langle u_h(t, \cdot), \tilde{\varphi}_{k,0} \rangle_{V_k} = v_{k,0}^n, \qquad x \in V_k,\ t_n \le t < t_{n+1}, \tag{2.11}$$

where the right-hand side follows by the biorthogonality of the basis functions. The convergence proof follows from the framework for finite volume schemes. Therefore, it is naturally to identify the zero order coefficient $v_{k,0}^n$ with the cell average. For this purpose, we assume in the following that $\tilde{\varphi}_{k,0} = h^{-1} \chi_{V_k}$. Details of the convergence proof can be found in [11]. Here we briefly summarize the main ingredients that are needed later on in the analysis of the adaptive MR-DG scheme. Following the roadmap of convergence to the entropy solution of the initial value problem (2.1) as can be found, for instance, in [39], we have to verify that the family of grid functions $\{\hat{u}_h\}_h$ is uniformly bounded in L^1, the total variation of the sequence is uniformly bounded and each of the grid functions is L^1-continuous in time. Then by compactness arguments and Cantor's diagonalization argument there exists a subsequence of $\{\hat{u}_h\}_h$ uniformly convergent on any bounded interval $[0, T]$. Furthermore, by definition $\hat{u}_h(0, \cdot)$ converges to the initial function u_0 in $L^1(\mathbb{R})$ as h tends to 0, cf. [39]. If in addition $\{\hat{u}_h\}_h$ is uniformly bounded in L^∞, then using the same arguments as in the proof of the Lax-Wendroff theorem, cf. [20, pp. 100–102], the limit of the subsequence is converging in $L^\infty(L^1_{loc}(\mathbb{R}); [0, T])$ to a weak solution of (2.1). Note that for initial data $u_0 \in L^1(\mathbb{R})$ we may infer uniform boundedness in L^∞ as well as L^1-continuity in time provided that the sequence $\{\hat{u}_h\}_h$ is total variation diminishing (TVD), i.e.,

$$TV(\hat{u}_h(t, \cdot)) := \sum_{k \in \mathbb{Z}} |v_{k+1,0}^n - v_{k,0}^n| \le TV(\hat{u}_h(0, \cdot), \quad t_n \le t < t_{n+1} \tag{2.12}$$

or at least total variation bounded (TVB), i.e.,

$$TV(\hat{u}_h(t, \cdot)) \le C\, TV(\hat{u}_h(0, \cdot), \quad t \in [0, T], \tag{2.13}$$

where C is a constant independent of the discretization h. The proof can be found in [20, p. 131], in the context of FV schemes for homogeneous problems. Since we assume that $s(0) = 0$, the proof applies also to inhomogeneous problems and the DG context as well. Since by definition the total variation of $\hat{u}_h(0, \cdot)$ is bounded by the total variation of u_0, cf. [39], the right-hand sides in (2.12) and (2.13) can be bounded independently of the discretization only depending

on the initial data $u_0 \in BV(\mathbb{R})$ and the interval $[0,T]$. In order to prove convergence of the sequence $\{\hat{u}_h\}_h$, it is sufficient to verify the TVD or TVB property. For homogeneous problems, Cockburn and Shu [11] give the sufficient conditions

$$-\theta \le \frac{\overline{v}_{k+1}^+ - \overline{v}_k^+}{v_{k+1,0} - v_{k,0}} \le 1, \quad -\theta \le -\frac{\overline{v}_{k+1}^- - \overline{v}_k^-}{v_{k+1,0} - v_{k,0}} \le 1, \tag{2.14}$$

that ensure the TVD condition under the CFL condition $(L_1 + L_2)\,\tau/h \le 1/(1+\theta)$, where

$$\overline{v}_k^- := \sum_{i \in \mathcal{P}^*} v_{k,i}\,\varphi_{k,i}(x_{k+1}^-), \quad \overline{v}_k^+ := -\sum_{i \in \mathcal{P}^*} v_{k,i}\,\varphi_{k,i}(x_k^+), \quad \mathcal{P}^* := \mathcal{P}\backslash\{0\} \tag{2.15}$$

denote the contribution of the higher order coefficients to the values of u_h at the cell interfaces. In the inhomogeneous case, we have to relax TVD by TVB, but in principle the proof works also in this case under the above assumptions on the source function s. In order to ensure the conditions (2.14), the values $\overline{v}_k^\pm$ are modified typically using local projection limiters. For instance, in [11] the following two examples are given:
Example 1: (van Leer limiter)

$$\tilde{\overline{v}}_k^\pm := m\left(\overline{v}_k^\pm, v_{k+1,0} - v_{k,0}, v_{k,0} - v_{k-1,0}\right) \tag{2.16}$$

Example 2: (Shu limiter)

$$\tilde{\overline{v}}_k^\pm := \begin{cases} \overline{v}_k^\pm & , \quad |\overline{v}_k^\pm| \le M\,h^2 \\ m(\overline{v}_k^\pm, v_{k+1,0} - v_{k,0}, v_{k,0} - v_{k-1,0}) & , \quad \text{else.} \end{cases} \tag{2.17}$$

Here m denotes the standard minmod function defined as

$$m(a_1, \ldots, a_n) := \begin{cases} s\,\min_{1 \le i \le n} |a_i| & , \quad \operatorname{sgn}(a_1) = \ldots = \operatorname{sgn}(a_n) = s \\ 0 & , \quad \text{else.} \end{cases} \tag{2.18}$$

The constant M in the Shu limiter has to be chosen appropriately such that the loss of accuracy at critical points as in case of the van Leer limiter is avoided. It typically depends on the bound M_2 to be chosen such that $|u_{xx}| \le M_2$ in any region where $u \in C^2$. If we now replace the numerical flux in the discrete evolution equation (2.7) by

$$F_k := F(v_{k-1,0} + \tilde{\overline{v}}_{k-1}^-, v_{k,0} - \tilde{\overline{v}}_k^+) \tag{2.19}$$

using either the van Leer or the Shu limiter, then the conditions (2.14) hold for $\theta = 1$ and, hence, the sequence $\{\hat{u}_h\}_h$ is TVD (van Leer) and TVB (Shu) in case of a homogeneous problem. In the general inhomogeneous case, it is TVB for both limiters. Note that $\tilde{\overline{v}}_k^\pm$ is only a modification on some combinations of higher order coefficients $v_{k,i}$, $i \in \mathcal{P}^*$. Since the limiting process is performed locally on each cell V_k, one might split the limiting step from the evolution step, where we first compute modified higher order coefficients $\tilde{v}_{k,i}$ that are determined such that

$$\sum_{i \in \mathcal{P}^*} \tilde{v}_{k,i}\,\varphi_{k,i}(x_{k+1}^-) = \tilde{\overline{v}}_k^-, \quad \sum_{i \in \mathcal{P}^*} \tilde{v}_{k,i}\,\varphi_{k,i}(x_k^+) = -\tilde{\overline{v}}_k^+. \tag{2.20}$$

For $p = 2, 3$ this leads to uniquely determined coefficients, for $p > 3$ the coefficients are no longer unique, cf. [11]. Note that by the above limiting process the zero-order coefficients remain unchanged, i.e., the limiting process is conservative.

Finally, if the minmod function (2.18) in the van Leer limiter (2.16) and the Shu limiter (2.17) is replaced by $m(a_1, a_2, a_3, \text{sign}(a_1) C_\alpha\, h^\alpha)$ for some fixed $C_\alpha > 0$ and $\alpha > 0$, i.e.,

$$|\tilde{\overline{v}}_k^{\pm}| \le C_\alpha\, h^\alpha \tag{2.21}$$

holds, then the sequence $\{\hat{u}_h\}_h$ converges to the unique entropy solution and the error on any bounded domain $\Omega \subset \mathbb{R}$ can be estimated by

$$\|\hat{u} - \hat{u}_h\|_{L^\infty(0,T;L^1(\Omega))} \le C(T,\Omega)\, h^{\min(1,\alpha)/2}. \tag{2.22}$$

This follows from the convergence of quasi-monotone finite volume schemes, cf. [7], where the numerical flux in the evolution equation (2.7) for the zero-order coefficients $i = 0$ is rewritten as a quasi-monotone flux, i.e., a perturbation of the two-point monotone flux F where the perturbation is uniformly bounded by h^α. Then the entropy production rate vanishes when h tends to zero.

Runge-Kutta time discretization For a higher order time discretization we have to apply an appropriate time discretization to the semi-discrete DG scheme (2.6) that can be considered a system of ordinary differential equations (ODE). In view of convergence, the ODE solver has to be chosen such that the TVD property (homogeneous case) and the TVB property (inhomogeneous case), respectively, are maintained. For this purpose, Shu and Osher [40] introduced a new class of m-stage Runge-Kutta (RK) methods that can be written as

$$\begin{aligned} \mathbf{u}^{(0)} &= \mathbf{v}_h^n \equiv \{v_{k,i}^n\}_{k\in\mathcal{I}_h,\, i\in\mathcal{P}} \\ \mathbf{u}^{(s)} &= \sum_{j=0}^{s-1} \Big(\alpha_{s,j}\, \mathbf{u}^{(j)} + \tau\, \beta_{s,j}\, \mathbf{L}_h(\mathbf{u}^{(j)})\Big), \quad s = 1,\dots,m \\ \mathbf{v}_h^{n+1} &= \mathbf{u}^{(m)}, \end{aligned} \tag{2.23}$$

where we introduce the notation $\mathbf{u} := \{u_{k,i}\}_{k\in\mathcal{I}_h, i\in\mathcal{P}}$. Then the operator $\mathbf{L}_h(\mathbf{u}) := \{L_{k,i}(\mathbf{u})\}_{k\in\mathcal{I}_h,i\in\mathcal{P}}$ is defined as the right-hand side in (2.6), where the piecewise polynomial function $u_h(t,\cdot)$ is determined by (2.3) with coefficients $\mathbf{u}$. In order to be consistent in the trivial case where the right-hand side in (2.6) vanishes, i.e., $\mathbf{L} = \mathbf{0}$, the coefficients have to satisfy the consistency condition

$$\sum_{j=0}^{s-1} \alpha_{s,j} = 1, \quad s = 1,\dots,m. \tag{2.24}$$

This representation allows to rewrite the evolution on each intermediate stage in (2.23) as a forward Euler step

$$\mathbf{u}^{(s)} = \sum_{j=0}^{s-1} \alpha_{s,j} \Big(\mathbf{u}^{(j)} + \tau_{s,j}\, \mathbf{L}_h(\mathbf{u}^{(j)})\Big) \equiv \sum_{j=0}^{s-1} \alpha_{s,j}\, \mathcal{E}(\mathbf{u}^{(j)}, \tau_{s,j}), \quad s = 1,\dots,m \tag{2.25}$$

with time step size $\tau_{s,j} := \tau\, \beta_{s,j}/\alpha_{s,j}$. Here we assume that $\beta_{s,j}$ vanishes whenever $\alpha_{s,j}$ is zero. In addition, if the coefficients $\alpha_{s,j}$ are non-negative, then the RK method (2.6) can be interpreted as a convex combination of forward Euler steps. This allows to extend any stability condition for the forward Euler step to the RK method. Formally a stability condition can be written as

$$\rho(\mathcal{E}(\mathbf{u}_h, \tau)) \le \rho(\mathbf{u}_h)\,(1 + c\,\tau), \quad \tau \le \tau_{FE}, \tag{2.26}$$

where ρ denotes a sublinear functional and $c \geq 0$ some constant independent of the discretization. Assuming that the coefficients $\beta_{s,j}$ are also non-negative, then by induction on the RK stages the stability result follows for the RK step, i.e.,

$$\rho(\mathbf{v}_h^{n+1}) \leq \rho(\mathbf{v}_h^n)\,(1 + \bar{c}\,\tau), \quad \tau \max_{s,j} \beta_{s,j}/\alpha_{s,j} \leq \tau_{FE}, \tag{2.27}$$

This result also holds for negative coefficients $\beta_{s,j}$, where the operator $\mathbf{L}_h$ is replaced by $\tilde{\mathbf{L}}_h$ and $-\tilde{\mathbf{L}}_h$ is an operator approximating the time inverse problem of (2.1), i.e., $\tilde{t} = -t$.

Originally, the Shu-Osher RK schemes have been designed to maintain the TVD property, i.e., the sublinear functional ρ is chosen as the TV semi-norm, and therefore these methods were introduced as TVD-RK methods, cf. [40]. Nowadays the notion has changed to nonlinear strongly stability preserving (SSP) RK methods, cf. [21, 29]. Note that "strongly" refers to $c = \bar{c} = 0$ in (2.26) and (2.27). For the homogeneous problem the strong stability result can be verified, but for the inhomogeneous case we have to relax the stability condition in the above sense corresponding to TVB instead of TVD.

In summary, if we apply a TVD-RK method to the semi-discrete DG scheme (2.6), then the higher order discretization inherits the TVB property from the DG scheme in the mean (2.11) determined by the forward Euler step (2.7). According to the above discussion on the convergence, the DG scheme in the mean with this higher order discretization is converging.

Convergence of DG scheme So far the convergence of the DG scheme in the mean $\hat{u}_h$, has been discussed. To infer now convergence of the DG scheme u_h, a perturbation argument is used. A sufficient condition is to assume that the difference of the DG scheme and its mean value can be bounded by the total variation of the mean function, i.e.,

$$\|u_h(t,\cdot) - \hat{u}_h(t,\cdot)\|_{L^1} \leq c\,h\,TV(\hat{u}_h(t,\cdot)) \tag{2.28}$$

holds for any $t \in [0,T]$, where $c > 0$ is a constant independent of the discretization. Then convergence can be concluded from

$$\|u_h(t,\cdot) - u(t,\cdot)\|_{L^1} \leq \|u_h(t,\cdot) - \hat{u}_h(t,\cdot)\|_{L^1} + \|\hat{u}_h(t,\cdot) - u(t,\cdot)\|_{L^1} \tag{2.29}$$

and the TVD/TVB property (2.12), (2.13) as well as the convergence of the DG scheme in the mean. Note that the condition (2.28) holds provided that the limited higher order coefficients $\tilde{v}_{k,i}$ can be estimated by

$$|\tilde{v}_{k,i}| \leq c_1\,\min(|v_{k+1,0} - v_{k,0}|, |v_{k,0} - v_{k-1,0}|) + c_2\,h^2, \quad i \in \mathcal{P}^*. \tag{2.30}$$

In case of $p = 2, 3$, this sufficient condition can be verified for the coefficients $\tilde{v}_{k,i}$ that are uniquely determined by (2.20).

3 Multiresolution analysis

The DG discretization typically works on an array of coefficients. In order to realize a certain target accuracy at the expense of a possibly low number of degrees of freedom, viz. a possibly low computational effort, one should keep the size of the cells large wherever the data exhibit little variation, reflecting a high regularity of the searched solution components. Our analysis of the local regularity behavior of the data is based on the concept of multiwavelets, cf. [28].

This can be considered a natural extension of the MRA for cell averages corresponding to a FV discretization, i.e., $p = 1$, where biorthogonal wavelets have been used to construct an appropriate MRA, cf. [18, 33]. Here we will briefly summarize the basic ideas of the MRA concept. Although we will confine ourselves to the one-dimensional case the framework is applicable to arbitrary discretizations in higher dimensions. The core ingredients are (i) a hierarchy of nested grids, (ii) (biorthogonal) multiwavelets and (iii) the multiscale decomposition. Note that the *multiwavelet* MRA generalizes the discrete MRA in [3, 44] motivated by Harten's original discrete framework [23] for cell averages.

Grid hierarchy Starting point for the construction of a MRA is a sequence of nested grids. Here we confine ourselves to 1D dyadic grid refinements on the interval $\Omega = [a, b]$, see Figure 1 for an illustration. Let $\mathcal{G}_l := \{V_{l,k}\}_{k\in\mathcal{I}_l}$, $\mathcal{I}_l = \{0, \dots, N_l - 1\}$, $l \in \mathbb{N}_0$, be a sequence of grids with increasing resolution. These meshes are composed of the intervals $V_{l,k} = [x_{l,k}, x_{l,k+1}]$ determined by the discretization points $x_{l,k} := a + k\, h_l$ with uniform spatial discretization $h_l = (b - a)/N_l$. The cell center is given by $\hat{x}_{l,k} = (x_{l,k} + x_{l,k+1})/2$. Hence, the resulting grid hierarchy is *nested* because of the subdivision

$$V_{l,k} = V_{l+1,2k} \cup V_{l+1,2k+1}, \qquad \forall\, k \in \mathcal{I}_l,\ l \in \mathbb{N}_0. \tag{3.1}$$

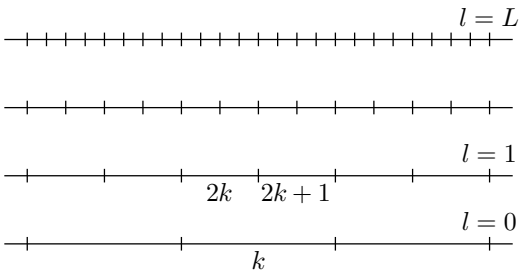

Figure 1: Sequence of nested dyadic grids

Multiresolution sequence On the above hierarchy of nested grids we introduce the spaces

$$S_l^p = \{f \in L^2([a, b]) :\ f|_{V_{l,k}} \in \Pi_{p-1},\ \forall\ k \in \mathcal{I}_l\} \tag{3.2}$$

of piecewise polynomial functions of degree less than p. Obviously, these spaces are nested, i.e.,

$$S_0^p \subset S_1^p \subset ... \subset S_l^p \subset ... \subset L^2([a, b]). \tag{3.3}$$

Hence, there exist complement spaces W_l^p such that

$$S_{l+1}^p = S_l^p \oplus W_l^p, \quad S_l^p = S_0^p \oplus W_0^p \oplus W_1^p \oplus ... \oplus W_{l-1}^p \tag{3.4}$$

These spaces are spanned by the single-scale basis and the multiwavelet basis

$$\Phi_l^p := \{\varphi_{l,k,i} \,:\, k \in \mathcal{I}_l,\ i \in \mathcal{P}\}, \quad \Psi_l^p := \{\psi_{l,k,i} \,:\, k \in \mathcal{I}_l,\ i \in \mathcal{P}\}, \tag{3.5}$$

respectively, i.e., $S_l^p = \operatorname{span} \Phi_l^p$ and $W_l^p = \operatorname{span} \Psi_l^p$. The multiresolution sequence is assumed to be dense in $L_2([a,b])$, i.e.,

$$L^2([a,b]) = \operatorname{clos}_{L^2} \left\{ \operatorname{span} \left(\Phi^p \right) \right\} = \operatorname{clos}_{L^2} \left\{ \operatorname{span} \left(\Psi^p \right) \right\}, \tag{3.6}$$

where $\Phi^p := \bigcup_{l \in \mathbb{N}_0} \Phi_l^p$ and $\Psi^p := \Phi_0^p \cup \bigcup_{l \in \mathbb{N}_0} \Psi_l^p$ are the *single-scale basis* and the *multiscale basis* of order p for $L^2([a,b])$, respectively. Then any function $u \in L_2([a,b])$ can be approximated by the sequence of projections

$$u_L = \sum_{k \in \mathcal{I}_L} \sum_{i \in \mathcal{P}} u_{L,k,i} \, \varphi_{L,k,i} = \sum_{k \in \mathcal{I}_0} \sum_{i \in \mathcal{P}} u_{0,k,i} \, \varphi_{0,k,i} + \sum_{l=0}^{L-1} \sum_{k \in \mathcal{I}_l} \sum_{i \in \mathcal{P}} d_{l,k,i} \, \psi_{l,k,i} \tag{3.7}$$

that is uniformly converging to u in L^2. If there exists a dual system

$$\tilde{\Phi}_l^p := \{ \tilde{\varphi}_{l,k,i} \, : \, k \in \mathcal{I}_l, \; i \in \mathcal{P} \}, \quad \tilde{\Psi}_l^p := \{ \tilde{\psi}_{l,k,i} \, : \, k \in \mathcal{I}_l, \; i \in \mathcal{P} \}, \tag{3.8}$$

of biorthogonal functions, i.e., $\Phi_l^p \cup \Psi_l^p$ is orthogonal to $\tilde{\Phi}_l^p \cup \tilde{\Psi}_l^p$, then the single-scale coefficients and wavelet coefficients (details) in (3.7) are determined by

$$u_{l,k,i} := \langle u, \tilde{\varphi}_{l,k,i} \rangle_{[a,b]}, \quad d_{l,k,i} := \langle u, \tilde{\psi}_{l,k,i} \rangle_{[a,b]}. \tag{3.9}$$

In particular, if $\Phi_l^p \cup \Psi_l^p$ is orthogonal, then it coincides with the dual system.

Construction of bases Since the space S_l^p contains piecewise polynomial functions that are not necessarily continuous at the cell edges, the design of appropriate basis functions can be performed independently for each cell $V_{l,k}$, where we take into account the nesting (3.1). In a first step we therefore need to construct an appropriate basis for the space Π_{p-1} on $V_{l,k}$, i.e., $\Phi_{l,k}^p := \{ \varphi_{l,k,i} \, : \, i \in \mathcal{P} \}$. In a second step, we then have to find a completion $\Psi_{l,k}^p := \{ \psi_{l,k,i} \, : \, i \in \mathcal{P} \}$ of $\Phi_{l,k}^p$ such that $\operatorname{span}(\Phi_{l,k}^p \cup \Psi_{l,k}^p) = \operatorname{span}(\Phi_{l+1,2k}^p \cup \Phi_{l+1,2k+1}^p)$. The construction of appropriate wavelet functions is subject to certain constraints:
(i) they are locally supported, i.e.,

$$\operatorname{supp} \varphi_{l,k,i} = \operatorname{supp} \psi_{l,k,i} = \operatorname{supp} \tilde{\varphi}_{l,k,i} = \operatorname{supp} \tilde{\psi}_{l,k,i} = V_{l,k} \tag{3.10}$$

(ii) provide vanishing moments of order $M_{p,i} \geq 1$, i.e.,

$$\langle P, \tilde{\psi}_{l,k,i} \rangle_{[a,b]} = 0, \quad \forall P \in \Pi_{M_{p,i}-1} \tag{3.11}$$

(iii) there exists a biorthogonal system of dual functions, i.e.,

$$\langle \varphi_{l,k,i}, \tilde{\varphi}_{l,k',i'} \rangle_{[a,b]} = \delta_{i,i'} \, \delta_{k,k'}, \quad \langle \varphi_{l,k,i}, \tilde{\psi}_{l,k',i'} \rangle_{[a,b]} = 0, \tag{3.12}$$

$$\langle \psi_{l,k,i}, \tilde{\psi}_{l,k',i'} \rangle_{[a,b]} = \delta_{i,i'} \, \delta_{k,k'}, \quad \langle \psi_{l,k,i}, \tilde{\varphi}_{l,k',i'} \rangle_{[a,b]} = 0. \tag{3.13}$$

The existence of a dual basis is closely related to the Riesz basis property of the infinite collection $\Phi_0^p \cup \bigcup_{l=0}^{\infty} \Psi_l^p$ of $L_2([a,b])$. This is typically hard to satisfy. For details we refer to the *concept of stable completions*, see [6].

With regard to the convergence analysis in Section 2, we assume that the primal and dual functions are normalized with respect to L^∞ and L^1, respectively, i.e.,

$$\| \varphi_{l,k,i} \|_{L^\infty([a,b])} \lesssim 1, \quad \| \psi_{l,k,i} \|_{L^\infty([a,b])} \lesssim 1, \tag{3.14}$$

$$\| \tilde{\varphi}_{l,k,i} \|_{L^1([a,b])} \lesssim 1, \quad \| \tilde{\psi}_{l,k,i} \|_{L^1([a,b])} \lesssim 1, \tag{3.15}$$

where "$\lesssim$" denotes a bound up to some constant independent of the level l and the position k. Furthermore, we assume that $\tilde{\varphi}_{l,k,0}$ coincides with the L^1-normalized characteristic function, i.e.,

$$\tilde{\varphi}_{l,k,0} = h_l^{-1}\, \chi_{V_{l,k}}. \tag{3.16}$$

Hence, the zero order coefficients in (2.3) coincide with cell averages, i.e., the DG scheme in the mean (2.11) can be considered as a FV discretization.

Typically the local basis functions with respect to a cell $V_{l,k}$ are deduced from a set of basis functions on some reference element $[0,1]$, i.e.,

$$\varphi_{l,k,i} = \varphi_i((x-\hat{x}_{l,k})/h_l), \quad \psi_{l,k,i} = \psi_i((x-\hat{x}_{l,k})/h_l), \tag{3.17}$$

$$\tilde{\varphi}_{l,k,i} = \tilde{\varphi}_i((x-\hat{x}_{l,k})/h_l), \quad \tilde{\psi}_{l,k,i} = \tilde{\psi}_i((x-\hat{x}_{l,k})/h_l), \tag{3.18}$$

where the primal and dual mother functions φ_i, ψ_i and $\tilde{\varphi}_i$, $\tilde{\psi}_i$ are normalized with respect to $L^\infty([0,1])$ and $L^1([0,1])$, respectively, satisfying biorthogonal relations similar to (3.12) and (3.13). Later on we need to estimate the derivatives of the primal scaling functions. Obviously, these are bounded by

$$\left\| \frac{d^r}{d\,x^r}\, \varphi_{l,k,i} \right\|_{L^\infty(V_{l,k})} \le h_l^{-r} \left\| \frac{d^r}{d\,x^r}\, \varphi_i \right\|_{L^\infty([0,1])} =: h_l^{-r}\, D_{r,i}. \tag{3.19}$$

An important example that fits into the above general frame is Alpert's orthogonal multiwavelet basis, cf. [2]. It is summarized in Appendix A.1.

Cancellation property The vanishing moments (3.11) and the normalization (3.15) imply that the details become small with increasing refinement level when the underlying function is locally smooth, i.e.,

$$|\langle u, \tilde{\psi}_{l,k,i}\rangle_{[a,b]}| \le \inf_{P\in\Pi_{M-1}} |\langle u-P, \tilde{\psi}_{l,k,i}\rangle_{[a,b]}| \lesssim 2^{-l\,M} \|u\|_{W^{1,M}(V_{l,k})}. \tag{3.20}$$

Here we assume that the grid hierarchy is quasi-uniform in the sense that the diameters of the cells on each level l are proportional to 2^{-l}. More precisely, the details decay at a rate of at least $2^{-l\,M}$, provided that the function u has sufficient regularity on the support of the multiwavelet, i.e., $u \in W^{1,M}(V_{l,k})$, and the multiwavelet has vanishing moments of order M. In fact, the higher M the more details may be discarded in smooth regions. Typically, the minimal number of vanishing moments is determined by $M = M_{p,i} \ge p$. Therefore, opposite to biorthogonal wavelets, the number of vanishing moments can be easily improved by increasing the order p *without* enlarging the support.

Multiscale transformation In order to exploit the above compression potential, the idea is to transform the array of single-scale coefficients corresponding to a finest uniform discretization level

$$\mathbf{u}_L := (\mathbf{u}_{L,k})_{k\in\mathcal{I}_L}, \quad \mathbf{u}_{L,k} := (u_{L,k,i})_{i\in\mathcal{P}} \tag{3.21}$$

into a sequence of coarse grid data and details

$$\mathbf{u}_0 := (\mathbf{u}_{0,k})_{k\in\mathcal{I}_0}, \quad \mathbf{d}_l := (\mathbf{d}_{l,k})_{k\in\mathcal{I}_l}, \quad \mathbf{d}_{l,k} := (d_{l,k,i})_{i\in\mathcal{P}}, \quad l = 0,\dots,L-1, \tag{3.22}$$

representing the successive update from a coarser resolution to a finer resolution. Here the coefficients $u_{l,k,i}$ and $d_{l,k,i}$ are defined according to (3.9). In fact, the single-scale and the multiscale representation realize the change of basis in (3.7).

For this purpose, we first note that the nesting of the spaces (3.4) as well as the biorthogonality relations (3.12) and (3.13) ensure the existence of the two-scale decomposition

$$\tilde{\varphi}_{l,k,i} = \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{0,s}_{i,j} \tilde{\varphi}_{l+1,2k+s,j},\ i \in \mathcal{P} \tag{3.23}$$

$$\tilde{\psi}_{l,k,i} = \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{1,s}_{i,j} \tilde{\varphi}_{l+1,2k+s,j},\ i \in \mathcal{P} \tag{3.24}$$

and its inverse

$$\tilde{\varphi}_{l+1,2k+s,i} = \sum_{j\in\mathcal{P}} \left(\tilde{g}^{0,s}_{i,j} \tilde{\varphi}_{l,k,j} + \tilde{g}^{1,s}_{i,j} \tilde{\psi}_{l,k,j} \right),\ s \in \{0,1\},\ i \in \mathcal{P}. \tag{3.25}$$

Here the mask coefficients are determined by the biorthogonality conditions as

$$\tilde{m}^{0,s}_{i,j} = \langle \tilde{\varphi}_{l,k,i}, \varphi_{l+1,2k+s,j} \rangle_{[a,b]}, \quad \tilde{m}^{1,s}_{i,j} = \langle \tilde{\psi}_{l,k,i}, \varphi_{l+1,2k+s,j} \rangle_{[a,b]}, \tag{3.26}$$

$$\tilde{g}^{0,s}_{i,j} = \langle \varphi_{l+1,2k+s,i}, \tilde{\varphi}_{l,k,j} \rangle_{[a,b]}, \quad \tilde{g}^{1,s}_{i,j} = \langle \varphi_{l+1,2k+s,i}, \tilde{\psi}_{l,k,j} \rangle_{[a,b]}. \tag{3.27}$$

In summary, according to (3.7), the change of basis provides two–scale relations for the coefficients inherited from the two–scale relations of the single-scale functions and the multiwavelet functions,

$$u_{l,k,i} = \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{0,s}_{i,j} u_{l+1,2k+s,j},\ i \in \mathcal{P}, \tag{3.28}$$

$$d_{l,k,i} = \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{1,s}_{i,j} u_{l+1,2k+s,j},\ i \in \mathcal{P}, \tag{3.29}$$

and, conversely,

$$u_{l+1,2k+s,i} = \sum_{j\in\mathcal{P}} \left(\tilde{g}^{0,s}_{i,j} u_{l,k,j} + \tilde{g}^{1,s}_{i,j} d_{l,k,j} \right),\ s \in \{0,1\},\ i \in \mathcal{P}, \tag{3.30}$$

which reflects the typical cascadic format of a wavelet transform. The two-scale relations are illustrated for the 1D case in Figure 2. A successive application of the relations (3.28) and (3.29), see Figure 3, decomposes the array $\mathbf{u}_L$ into coarse scale coefficients $\mathbf{d}_0$ and higher level fluctuations $\mathbf{d}_0, \ldots, \mathbf{d}_{L-1}$. We refer to this transformation as the *multiscale transformation* determined by the *multiscale operator* $\mathcal{M}_L : \mathbf{u}_L \longrightarrow (\mathbf{u}_0, \mathbf{d}_0, \ldots, \mathbf{d}_{L-1})$ with

$$\mathbf{u}_l = \mathbf{M}^T_{l,0} \mathbf{u}_{l+1}, \quad \mathbf{d}_l = \mathbf{M}^T_{l,1} \hat{\mathbf{u}}_{l+1}.$$

Here the mask matrices $\mathbf{M}_{l,0}$ and $\mathbf{M}_{l,1}$ are defined by the mask coefficients $\tilde{m}^{0,s}_{i,j}$ and $\tilde{m}^{1,s}_{i,j}$, respectively. It is reversed by recursively applying the two-scale relation (3.30). The resulting *inverse multiscale transformation* is described by the *inverse multiscale operator* $\mathcal{M}_L^{-1} : (\mathbf{u}_0, \mathbf{d}_0, \ldots, \mathbf{d}_{L-1}) \longrightarrow \mathbf{u}_L$ with

$$\mathbf{u}_{l+1} = \mathbf{G}^T_{l,0} \mathbf{u}_l + \mathbf{G}^T_{l,1} \mathbf{d}_l.$$

Here again the mask matrices $\mathbf{G}_{l,0}$ and $\mathbf{G}_{l,1}$ are determined by the mask coefficients $\tilde{g}^{0,s}_{i,j}$ and $\tilde{g}^{1,s}_{i,j}$, respectively.

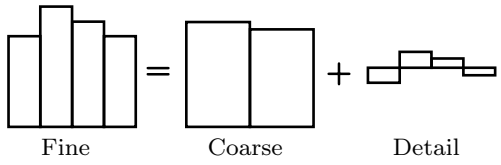

Figure 2: Two-scale Transformation ($p = 1$)

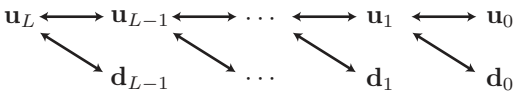

Figure 3: Multiscale transformation

Thresholding and approximation Due to the cancellation property, details might become negligibly small whenever the underlying function is locally smooth. This gives rise to *hard thresholding* characterized by the index set

$$\mathcal{D}_\varepsilon := \{(l,k,i) \,:\, |d_{l,k,i}| > \varepsilon_l\}.$$

Here ε denotes the threshold value. The level-dependent threshold values ε_l are determined by an appropriate scaling of ε. Later on we will choose $\varepsilon_l := 2^{l-L}\,\varepsilon$. Then the threshold operator $\mathcal{T}_{\mathcal{D}_\varepsilon} : (\mathbf{u}_0, \mathbf{d}_0, \ldots, \mathbf{d}_{L-1}) \longrightarrow (\tilde{\mathbf{u}}_0, \tilde{\mathbf{d}}_0, \ldots, \tilde{\mathbf{d}}_{L-1})$ is defined element-wise by

$$\tilde{d}_{l,k,i} := \begin{cases} d_{l,k,i} & , \quad (l,k,i) \in \mathcal{D}_\varepsilon \\ 0 & , \quad \text{else} \end{cases}, \qquad \tilde{u}_{0,k,i} := u_{0,k,i}, \quad k \in \mathcal{I}_0.$$

Note that thresholding is not performed on the coarse-scale averages in order to maintain the conservation property of the adaptive DG scheme.

Later on we will perform thresholding also for an arbitrary index set $\mathcal{D}$ instead of $\mathcal{D}_\varepsilon$. Let u_L and $u_{L,\tilde{\mathcal{D}}_\varepsilon}$ be the projections according to (3.7) corresponding to the coefficients $\mathbf{u}_L$ and $\mathcal{A}_\mathcal{D}\,\mathbf{u}_L$, where

$$\mathcal{A}_\mathcal{D} := \mathcal{M}_L^{-1}\,\mathcal{T}_\mathcal{D}\,\mathcal{M}_L, \tag{3.31}$$

and, in particular for $\mathcal{D} = \mathcal{D}_\varepsilon$,

$$\mathcal{A}_\varepsilon := \mathcal{M}_L^{-1}\,\mathcal{T}_{\mathcal{D}_\varepsilon}\,\mathcal{M}_L. \tag{3.32}$$

In order to investigate the approximation error introduced by thresholding we first note that by (3.7) we may write for any $x \in V_{L,r}$, $r \in \mathcal{I}_L$,

$$u_L(x) = \sum_{l=-1}^{L-1} \sum_{k \in \mathcal{I}_l} \sum_{i \in \mathcal{P}} d_{l,k,i}\psi_{l,k,i}(x) = \sum_{(l,k,i) \in \overline{\Sigma}_{L,r}} d^n_{l,k,i}\psi_{l,k,i}(x), \tag{3.33}$$

where we define $d_{-1,0,i} := v_{0,0,i}$, $\psi_{-1,0,i} := \varphi_{0,0,i}$ and $\mathcal{I}_{-1} := \mathcal{I}_0$. Note that for $x \in V_{L,r}$ the summation in (3.33) is performed only for those indices (l,k,i) with $x \in \operatorname{supp}\psi_{l,k,i} = V_{l,k}$ according to (3.10). These indices are determined by the support $\overline{\Sigma}_{L,r}$

$$\begin{aligned} \overline{\Sigma}_{L,r} &:= \{(l,k,i) : V_{L,r} \cap V_{l,k} \neq \emptyset,\, l = -1, ..., L-1,\, i \in \mathcal{P}\} \\ &= \{(l, \lfloor r/2^{L-l_+} \rfloor, i) : l = -1, ..., L-1,\, i \in \mathcal{P}\} \end{aligned} \tag{3.34}$$

corresponding to $V_{L,r}$. The last identity holds because of the dyadic grid hierarchy (3.1), where $l_+ := \max(0,l)$. Then we may rewrite u_L locally as

$$u_L(x) = \sum_{(l,k,i) \in \overline{\Sigma}_{L,r}} d^n_{l,k,i}\psi_{l,k,i}(x) = \sum_{i \in \mathcal{P}} \sum_{l=-1}^{L-1} d^n_{l,\lfloor r/2^{L-l_+} \rfloor,i}\psi_{l,\lfloor r/2^{L-l_+} \rfloor,i}(x) \tag{3.35}$$

and the approximation error due to thresholding is locally determined by

$$u_L(x) - u_{L,\mathcal{D}_\varepsilon}(x) = \sum_{(l,k,i)\notin\mathcal{D}_\varepsilon} d_{l,k,i}\,\psi_{l,k,i} = \sum_{(l,k,i)\in\overline{\Sigma}_{L,r}\setminus\mathcal{D}_\varepsilon} d^n_{l,k,i}\psi_{l,k,i}(x). \tag{3.36}$$

Then we can estimate the approximation error in the L^q-metric, $q\in[1,\infty)$, by

$$\|u_L - u_{L,\mathcal{D}_\varepsilon}\|^q_{L_q(\Omega)} \le \sum_{(l,k,i)\notin\mathcal{D}_\varepsilon} |d_{l,k,i}|^q \int_\Omega |\psi_{l,k,i}(x)|^q\,dx. \tag{3.37}$$

Since $\psi_{l,k,i}$ has compact support and is normalized with respect to L^∞ according to (3.10) and (3.14), we infer

$$\|\psi_{l,k,i}\|^q_{L^q(\Omega)} \le (\text{supp }\psi_{l,k,i})\,\|\psi_{l,k,i}\|^q_{L_\infty(\Omega)} \lesssim 2^{-l_+}.$$

The non-significant multiscale coefficients can be estimated by $|d_{l,k,i}|\le\varepsilon_{l_+}$ for all $(l,k,i)\notin\mathcal{D}_\varepsilon$. There are at most $p\,N_0\,2^{l_+}$ coefficients on level l_+. Hence, the approximation error is bounded by

$$\|u_L - u_{L,\mathcal{D}_\varepsilon}\|^q_{L^q(\Omega)} \lesssim \sum_{l=-1}^{L-1}(\varepsilon_{l_+})^q = \varepsilon^q\overline{a}^{-Lq}\left(1+\sum_{l=0}^{L-1}\overline{a}^{lq}\right) \le \varepsilon^q\Big(1+\frac{1}{\overline{a}^q-1}\Big) \lesssim \varepsilon^q, \tag{3.38}$$

where we choose $\varepsilon_l = \overline{a}^{l_+-L}\varepsilon$, $\overline{a}>1$.

In case of $q=\infty$ we have to proceed slightly different. Since $\psi_{l,k,i}$ is normalized with respect to L^∞, we conclude from (3.36) and (3.34)

$$|u_L(x) - u_{L,\mathcal{D}_\varepsilon}(x)| \le \sum_{(l,k,i)\in\bar{\Sigma}_{L,r}\setminus\mathcal{D}_\varepsilon} |d_{l,k,i}|\cdot|\psi_{l,k,i}(x)| \lesssim \sum_{(l,k,i)\in\bar{\Sigma}_{L,r}} \varepsilon_{l_+} = \sum_{l=-1}^{L-1}\overline{a}^{l_+-L}\varepsilon \lesssim \varepsilon.$$

A similar result holds for the approximation error of the mean function (2.11). According to (3.36) this error is determined by

$$\hat{u}_L(x) - \hat{u}_{L,\mathcal{D}_\varepsilon}(x) = \sum_{(l,k,i)\in\overline{\Sigma}_{L,r}\setminus\mathcal{D}_\varepsilon} d^n_{l,k,i}\langle\psi_{l,k,i},\tilde{\varphi}_{L,r,0}\rangle_{[a,b]}, \qquad x\in V_{L,r}. \tag{3.39}$$

By the normalization (3.14) and (3.15) of the basis functions the inner products are uniformly bounded, i.e.,

$$|\langle\psi_{l,k,i},\tilde{\varphi}_{L,r,0}\rangle_{[a,b]}| \lesssim 1, \qquad (l,k,i)\in\overline{\Sigma}_{L,r}.$$

Hence we conclude in a similar way as above

$$|\hat{u}_L(x) - \hat{u}_{L,\mathcal{D}_\varepsilon}(x)| \le \sum_{(l,k,i)\in\bar{\Sigma}_{L,r}} \varepsilon_{l_+} = \sum_{l=-1}^{L-1}\overline{a}^{l_+-L}\varepsilon \lesssim \varepsilon, \qquad x\in V_{L,r}. \tag{3.40}$$

Then we obtain for the approximation error of the mean function in the L^q-metric, $q\in[1,\infty)$

$$\|\hat{u}_L - \hat{u}_{L,\mathcal{D}_\varepsilon}\|^q_{L^q(\Omega)} = \sum_{r\in\mathcal{I}_L}\int_{V_{L,r}} |\hat{u}_L(x) - \hat{u}_{L,\mathcal{D}_\varepsilon}(x)|^q dx \lesssim \varepsilon^q,$$

where we use $\sum_{r\in\mathcal{I}_L} h_L = (b-a) < \infty$, i.e., the boundedness of the domain $\Omega=[a,b]$. For $q=\infty$ the result directly follows from (3.40). Thus we have proven the following proposition.

Proposition 1 *(Approximation error)*
Let $\Omega = [a,b]$ be bounded and $\varepsilon_l = \overline{a}^{l-L}\varepsilon$ with $\overline{a} > 1$. Then the approximation error with respect to the set of significant details $\mathcal{D}_\varepsilon$ is uniformly bounded with respect to $L^q(\Omega)$, $q \in [1,\infty]$, i.e.,

$$\|u_L - u_{L,\mathcal{D}_\varepsilon}\|_{L^q(\Omega)} \lesssim \varepsilon \qquad \textit{and} \qquad \|\hat{u}_L - \hat{u}_{L,\mathcal{D}_\varepsilon}\|_{L^q(\Omega)} \lesssim \varepsilon. \tag{3.41}$$

4 Adaptive multiresolution DG scheme

We now combine the MRA in Section 3 with the DG scheme presented in Section 2. The idea is first to derive evolution equations for the single-scale and the multiscale coefficients, respectively, and then to apply thresholding to discard non-significant contributions and, hence, to reduce the complexity of the evolution system. The procedure is similar to the derivation of adaptive MR-FV schemes, cf. [33]. Starting point is a given DG discretization that is defined on the uniform highest resolution level L, i.e., $v_L \equiv u_h$ with $v_{L,k,i} \equiv v_{k,i}$ in (2.3). This will be referred to as the *reference scheme* on the *reference mesh*. Since for the time discretization a TVD-Runge-Kutta method (2.23) is used, where each stage can be interpreted as a linear combination of forward Euler steps, we confine ourselves to the forward Euler step (2.7) with (2.8), (2.9) and (2.10). Furthermore, the limiting process is split from the evolution process as is justified by (2.20) and (2.30).

In a first step we now apply the two-scale relation (3.28) to the evolution equation (2.7), where we replace $u_{l+1,2k+s,j}$ by $v^{\mu}_{l+1,2k+s,j}$ for $\mu = n$ and $\mu = n+1$. Then we recursively obtain for any $k \in \mathcal{I}_l$, $i \in \mathcal{P}$, the two-scale evolution equations for the single-scale coefficients

$$v^{n+1}_{l,k,i} = v^n_{l,k,i} - \tau\left(B^n_{l,k,i} - G^n_{l,k,i} - S^n_{l,k,i}\right), \tag{4.1}$$

where we proceed levelwise from fine to coarse. It turns out that the numerical flux $F^n_{l,k}$, the numerical flux balance $B^n_{l,k,i}$, the flux quadrature $G^n_{l,k,i}$ and the source quadrature $S^n_{l,k,i}$ are recursively determined by

$$F^n_{l,k} := F^n_{l+1,2k} = \ldots = F^n_{L,2^{L-l}k} = F(v_L(t^n, x^-_{L,2^{L-l}k}), v_L(t^n, x^+_{L,2^{L-l}k})), \tag{4.2}$$

$$B^n_{l,k,i} := F^n_{l,k+1} \cdot \tilde{\varphi}_{l,k,i}(x^-_{l,k+1}) - F^n_{l,k} \cdot \tilde{\varphi}_{l,k,i}(x^+_{l,k}), \tag{4.3}$$

$$G^n_{l,k,i} := \sum_{s\in\{0,1\}} \sum_{j\in\mathcal{P}} \tilde{m}^{0,s}_{ij} G^n_{l+1,2k+s,j} = \int_{V_{l,k}} f(v_L(t^n,x))\tilde{\varphi}'_{l,k,i}(x)\,dx, \tag{4.4}$$

$$S^n_{l,k,i} := \sum_{s\in\{0,1\}} \sum_{j\in\mathcal{P}} \tilde{m}^{0,s}_{ij} S^n_{l+1,2k+s,j} = \int_{V_{l,k}} s(v_L(t^n,x)) \cdot \tilde{\varphi}_{l,k,i}(x)\,dx. \tag{4.5}$$

Details of the derivation can be found in Appendix A.2. Note that due to the nestedness of the grid hierarchy and the consistency property of the numerical fluxes, the flux balances are only computed by the fine-scale fluxes corresponding to the edges of the coarse cell, see (4.3). These, in particular, have to be determined by the fine scale data. However, the internal fluxes cancel and, hence, the overall complexity is reduced, because we only have to compute two fluxes for $V_{l,k}$ instead of $2^{L-l}+1$ fluxes corresponding to the cells $V_{L,2^{L-l}k+r}$, $r = 0,\ldots,2^{L-l}-1$. In higher dimensions, the subdivision of the cell faces has to be taken into account. For instance, for a d-dimensional Cartesian grid hierarchy we would have to compute $2d\,2^{(L-l)(d-1)}$ fluxes that correspond to all interfaces on level L that form a partition of an interface on a coarser scale l. On the other hand, the volume integrals of the flux and source quadrature, (4.4) and (4.5), are

determined by the sum of all integrals on the finest scale L according to (2.10). Hence there is no complexity reduction, i.e., we still have the complexity of the reference grid. However, if the integrals can be computed *exactly*, then this sum can be replaced by one integral on the local scale l.

Similarly, we derive evolution equations for the details where we recursively apply the multiscale transformation (3.29) to the evolution equations (4.1) of the single-scale coefficients for $l = L-1 \searrow 0$. Then the evolution process (2.7) on the uniform reference mesh is equivalent to the evolution of the multiscale coefficients, i.e., coarse-scale coefficients and details,

$$v_{0,k,i}^{n+1} = v_{0,k,i}^{n} - \tau \left(B_{0,k,i}^{n} - G_{0,k,i}^{n} - S_{0,k,i}^{n} \right), \tag{4.6}$$

$$d_{l,k,i}^{n+1} = d_{l,k,i}^{n} - \tau \left(\overline{B}_{l,k,i}^{n} - \overline{G}_{l,k,i}^{n} - \overline{S}_{l,k,i}^{n} \right). \tag{4.7}$$

Here the numerical flux balance, the flux quadrature and the source quadrature differ from those in (4.1):

$$\begin{aligned} \overline{B}_{l,k,i}^{n} := & \sum_{j \in \mathcal{P}} \sum_{s \in \{0,1\}} \tilde{m}_{i,j}^{1,s} B_{l+1,2k+s,j}^{n} = \\ & F_{l,k+1}^{n} \cdot \tilde{\psi}_{l,k,i}(x_{l,k+1}^{-}) - F_{l,k}^{n} \cdot \tilde{\psi}_{l,k,i}(x_{l,k}^{+}) + \\ & F_{l+1,2k+1}^{n} \cdot (\tilde{\psi}_{l,k,i}(x_{l+1,2k+1}^{-}) - \tilde{\psi}_{l,k,i}(x_{l+1,2k+1}^{+})), \end{aligned} \tag{4.8}$$

$$\overline{G}_{l,k,i}^{n} := \sum_{j \in \mathcal{P}} \sum_{s \in \{0,1\}} \tilde{m}_{i,j}^{1,s} G_{l+1,2k+s,j}^{n} = \sum_{s \in \{0,1\}} \int_{V_{l+1,2k+s}} f(v_L(t^n, x)) \tilde{\psi}'_{l,k,i}(x)\, dx, \tag{4.9}$$

$$\overline{S}_{l,k,i}^{n} := \sum_{j \in \mathcal{P}} \sum_{s \in \{0,1\}} \tilde{m}_{i,j}^{1,s} S_{l+1,2k+s,j}^{n} = \int_{V_{l,k}} s(v_L(t^n, x)) \cdot \tilde{\psi}_{l,k,i}(x)\, dx. \tag{4.10}$$

In particular, we note that the multiwavelet $\psi_{l,k,i}$ typically jumps at the cell center $x_{l,k+\frac{1}{2}} = x_{l+1,2k+1}$. Therefore the internal numerical flux $F_{l+1,2k+1}^{n}$ does *not* cancel in (4.8) and we have to split the integral in (4.9) into two subintervals. These equations are employed later on in the analysis of the adaptive scheme. Details on their derivation can be found again in Appendix A.2. However, in practice, it is more convenient to use (4.1) for the cells of the locally refined grid.

According to the change of basis (3.7) applied to (2.3), the evolution step (2.7) of the reference scheme can be rewritten as

$$v_L^{n+1} = \sum_{k \in \mathcal{I}_0} \sum_{i \in \mathcal{P}} v_{0,k,i}^{n+1} \varphi_{0,k,i} + \sum_{l=0}^{L-1} \sum_{k \in \mathcal{I}_l} \sum_{i \in \mathcal{P}} d_{l,k,i}^{n+1} \psi_{l,k,i}, \tag{4.11}$$

where the multiscale coefficients are determined by (4.6) and (4.7). As is motivated by the cancellation property (3.20), we may reduce the degrees of freedom by applying first *data compression* to the multiscale coefficients without loosing significantly in accuracy. Then the compressed set of equations is solved on a locally refined *adaptive grid*.

Data compression The idea of the adaptive DG scheme is to perform the evolution step only for *significant* details

$$\mathcal{D}^{n+1} := \left\{ (l,k,i) \ ; \ |d_{l,k,i}^{n+1}| > \varepsilon_l, \ i \in \mathcal{P}, \ k \in \mathcal{I}_l, \ l \in \{0, \ldots, L-1\} \right\}$$

and to discard all other equations. Since this set cannot be computed before the data at time level t^{n+1} are known, a *prediction set* $\tilde{\mathcal{D}}^{n+1}$ has to be computed from $\mathcal{D}^n$ such that the *reliability condition*

$$\mathcal{D}^n \cup \mathcal{D}^{n+1} \subset \tilde{\mathcal{D}}^{n+1} \tag{4.12}$$

holds. In particular, this condition must be satisfied after having performed the whole update consisting of the evolution step *and* the limiting process. In Section 5.2 we will present a prediction strategy satisfying the reliability condition (4.12).

Adaptive grid According to the change of basis (3.7), the reduced system of evolution equations for the *significant* multiscale coefficients corresponding to the prediction set $\tilde{\mathcal{D}}^{n+1}$ can be replaced by a system of evolution equations for single-scale coefficients corresponding to a *locally refined grid* $\tilde{\mathcal{G}}^{n+1}$. The adaptive grid is characterized by the index set $\mathcal{G} \subset \{(l,k)\,;\, k \in \mathcal{I}_l,\ l = 0,\dots,L\}$, i.e., $\mathbb{R} = \bigcup_{(l,k)\in\mathcal{G}} V_{l,k}$ with $|V_{l,k} \cap V_{l',k'}| = 0$ for $(l,k) \neq (l',k')$, which is computed from $\mathcal{D} = \tilde{\mathcal{D}}^{n+1}$. For this purpose we have to assume that $\mathcal{D}$ is a tree, i.e., the relation

$$(l,k,i) \in \mathcal{D} \Rightarrow (l,k,j) \in \mathcal{D} \text{ and } (l-1,\lfloor k/2 \rfloor, j) \in \mathcal{D} \quad \forall j \in \mathcal{P}, \tag{4.13}$$

holds for any $l \in \{1,\dots,L-1\}$. Note that in case of adaptive MR-FV schemes [33] the tree has to be graded of degree $q \geq 1$. This leads to larger trees and, hence, degrades the efficiency. Then $\mathcal{G}$ can be determined recursively. For this purpose, the index set $\mathcal{G}$ is initialized by all indices of the coarsest discretization. Then, traversing through the levels from coarse to fine we proceed as follows: if $(l,k,i) \in \mathcal{D}$ for some $i \in \mathcal{P}$, then the cell $V_{l,k}$ is locally refined, i.e., the index (l,k) is removed from $\mathcal{G}$ and the indices of the subcells on the finer level are added to $\mathcal{G}$. Finally, we obtain a locally adapted grid which naturally corresponds to the outer leaves of the tree of significant details. Note that, if $\mathcal{D}$ is a tree, the system of evolution equations for the single-scale coefficients on grid $\mathcal{G}$ is equivalent to the system of evolution equations for the multiscale coefficients.

Change of basis By the sets $\mathcal{D}$ and $\mathcal{G}$ a local change of basis is performed that, in analogy to (3.7), reads:

$$v_L = \sum_{(l,k)\in\mathcal{G}} \sum_{i\in\mathcal{P}} v_{l,k,i}\,\varphi_{l,k,i} = \sum_{k\in\mathcal{I}_0} \sum_{i\in\mathcal{P}} v_{0,k,i}\,\varphi_{0,k,i} + \sum_{(l,k,i)\in\mathcal{D}} d_{l,k,i}\,\psi_{l,k,i}. \tag{4.14}$$

For analytical purposes it is convenient to write the evolution process in terms of the multiscale coefficients. However, from a practical point of view, it is preferable to evolve the cell averages according to (4.1) on a locally refined grid, because the local single-scale coefficients are needed in the limiting process. Moreover, we need to evaluate v_L when computing the numerical flux (4.2) as well as the flux quadrature (4.4) and source quadrature (4.5). For this purpose, the evaluation of the single-scale representation is more efficient.

Finally, we summarize the adaptive MR-DG scheme in the following algorithm. It basically consists of three steps corresponding to (i) refinement, (ii) limiting and evolution as well as (iii) coarsening.

Algorithm 1 *(Adaptive MR-DG scheme)*
Step 1. *(Refinement) Determine the prediction set* $\tilde{\mathcal{D}}^{n+1}$ *and apply the approximation operator* $\mathcal{A}_{\tilde{\mathcal{D}}^{n+1}}$ *to the given data, i.e., compute*

$$\mathbf{v}^n_{L,\tilde{\mathcal{D}}^{n+1}} := \mathcal{A}_{\tilde{\mathcal{D}}^{n+1}} \mathbf{v}^n_{L,\mathcal{D}^n}. \tag{4.15}$$

Step 2a. *(Limiting) We apply a limiter to the data on the adaptive grid $\tilde{\mathcal{G}}^{n+1}$, where only the higher order coefficients on the highest resolution level $v_{L,k,i}^{n+1}$, $i \in \mathcal{P}^*$, $(L,k) \in \tilde{\mathcal{G}}^{n+1}$, are modified, i.e.,*

$$\overline{\mathbf{v}}_{L,\mathcal{D}^{n+1}}^n := \Pi_{L,\tilde{\mathcal{D}}^{n+1}} \mathbf{v}_{L,\tilde{\mathcal{D}}^{n+1}}^n \quad \text{with} \quad \overline{v}_{L,k,0}^n = v_{L,k,0}^n,\ (L,k) \in \tilde{\mathcal{G}}^{n+1}. \tag{4.16}$$

Step 2b. *(Evolution) Perform time evolution of the multiscale coefficients corresponding to the prediction set $\tilde{\mathcal{D}}^{n+1}$ that are determined by the limited data $\overline{\mathbf{v}}_{L,\mathcal{D}^{n+1}}^n$ according to (4.6), (4.7), i.e.,*

$$\tilde{\mathbf{v}}_{L,\tilde{\mathcal{D}}^{n+1}}^{n+1} := \mathcal{E}_{L,\tilde{\mathcal{D}}^{n+1}} \overline{\mathbf{v}}_{L,\tilde{\mathcal{D}}^{n+1}}^n. \tag{4.17}$$

Since $\tilde{\mathcal{D}}^{n+1}$ is supposed to be a tree, we may equivalently evolve the single-scale coefficients on the adaptive grid corresponding to the set $\tilde{\mathcal{G}}^{n+1}$ in time according to (4.1).

Step 3. *(Coarsening) Threshold the data on the new time level by applying the approximation operator $\mathcal{A}_\varepsilon$, i.e.,*

$$\mathbf{v}_{L,\mathcal{D}^{n+1}}^{n+1} := \mathcal{A}_\varepsilon \tilde{\mathbf{v}}_{L,\tilde{\mathcal{D}}^{n+1}}^{n+1}. \tag{4.18}$$

We emphasize that all operators are applied locally, i.e., the multiscale operators $\mathcal{M}_L$, $\mathcal{M}_L^{-1}$, the threshold operator $\mathcal{T}_\mathcal{D}$ and the approximation operator $\mathcal{A}_\mathcal{D}$ only work on the set $\mathcal{D}$ of *significant* coefficients. Therefore, if the integrals in the flux quadrature (4.4) and the source quadrature (4.5) can be computed *exactly*, then the complexity of the resulting algorithm might be significantly reduced from the cardinality of the reference mesh to the cardinality of the set $\mathcal{D}$. Otherwise we have to approximate the local integrals such that the error can be uniformly bounded by the local threshold value ε_l. This has been investigated in detail for adaptive MR-FV schemes in [27].

5 Error analysis

The objective of the proposed adaptive MR-DG scheme is to reduce for a given DG scheme performed on the uniform reference mesh computational cost and memory requirements while preserving the accuracy of the reference scheme. In order to achieve this, we proceed similarly as in case of the adaptive MR-FV scheme, cf. [33], using a perturbation argument. For this purpose we first introduce the entropy solution u of the initial value problem (2.1) and its projection u_L to the space S_L^p according to (2.3). Let v_L and $v_{L,\varepsilon}$ denote the approximate solution functions determined by the reference DG scheme and the adaptive MR-DG scheme, respectively. An ideal strategy would be to prescribe an error tolerance *tol*. Then the number of refinement levels L should be determined during the computation such that the error meets the tolerance, i.e.,

$$\|\hat{u}_L(t,\cdot) - \hat{v}_{L,\varepsilon}(t,\cdot)\|_{L^1(\Omega)} \leq tol$$

for possibly small L. Here $\hat{w}_L$ denotes the mean of a function $w_L \in S_L^p$ according to (2.11). Since in general no error estimator is available for the adaptive scheme, we split the error into two parts corresponding to the *discretization error* $\hat{\eta}_L := \hat{u}_L - \hat{v}_L$ of the reference DG scheme and the *perturbation error* $\hat{e}_L := \hat{v}_L - \hat{v}_{L,\varepsilon}$, i.e.,

$$\|\hat{u}_L(t,\cdot) - \hat{v}_{L,\varepsilon}(t,\cdot)\|_{L^1(\Omega)} \leq \|\hat{\eta}_L(t,\cdot)\|_{L^1(\Omega)} + \|\hat{e}_L(t,\cdot)\|_{L^1(\Omega)} \leq tol. \tag{5.1}$$

We now assume that there is an a priori error estimate (2.22) of the discretization error, i.e., $\|\hat{\eta}_L(t,\cdot)\|_{L^1(\Omega)} \sim h_L^\beta$ where h_L denotes the spatial step size and β the convergence order. Then,

ideally we would determine the number of refinement levels L such that $h_L^\beta \sim tol$. In order to preserve the accuracy of the reference DG scheme we may now admit a perturbation error which is proportional to the discretization error, i.e., $\|\hat{e}_L(t,\cdot)\|_{L^1(\Omega)} \sim \|\hat{\eta}_L(t,\cdot)\|_{L^1(\Omega)}$. From this, we conclude

$$L = L(tol, \beta) \quad \text{and} \quad \varepsilon = \varepsilon(L). \tag{5.2}$$

Therefore it remains to verify that the perturbation error can be controlled. Note that in each time step we introduce an error due to the thresholding procedure. Obviously, this error accumulates in each step, i.e., the best we can hope for is an estimate of the form

$$\|\hat{e}_L(t_n,\cdot)\|_{L^1(\Omega)} \leq C\, n\, \varepsilon.$$

However, the threshold error may be amplified in addition by the evolution step. In order to control the cumulative perturbation error, we have to prove that the constant C is independent of L, n, τ and ε. For this purpose, we will consider the following two issues in more detail, namely, (i) the uniform boundedness of the perturbation error and (ii) the reliability of the prediction procedure. Again, the analysis is only performed for a forward Euler time step. Due to the stability property (2.27) it also holds for the general case of a TVD-RK time discretization, where we apply Algorithm 1 to each Runge-Kutta stage in (2.25).

5.1 Perturbation error

In a first step we verify the uniform boundedness of the perturbation error between the reference DG scheme and the adaptive MR-DG scheme. For this purpose, we rewrite both schemes at some time t_n in single-scale representation as

$$v_L^n(x) = \sum_{k\in\mathcal{I}_L}\sum_{i\in\mathcal{P}} v_{L,k,i}^n \varphi_{L,k,i}(x), \quad v_{L,\varepsilon}^n(x) = \sum_{k\in\mathcal{I}_L}\sum_{i\in\mathcal{P}} \overline{v}_{L,k,i}^n \varphi_{L,k,i}(x), \quad x\in\Omega \tag{5.3}$$

according to (3.7). Note that for the adaptive scheme we prolongate the coefficients from the adaptive grid to the reference grid performing a change of basis according to (3.7), where we put all missing detail coefficients to zero.

According to (3.21) the single-scale representations are characterized by the sequence of coefficients $\mathbf{v}_L$ and $\overline{\mathbf{v}}_L$, respectively. Since we assume that the initial data u_0 are compactly supported, then, by the finite speed of propagation the solution is compactly supported as well. Therefore it is sufficient to consider an arbitrary but fixed compact set $\Omega = [a,b]$ containing the support of u for $t\in[0,T]$.

In order to investigate the perturbation error, we introduce the following discrete operators defined on the coefficients of the single-scale representation:

1. the limiting operator Π_L satisfying the sufficient TVB condition (2.30) on the reference mesh and acting only on the higher order coefficients in order to maintain the conservation property,

2. the evolution operator $\mathcal{E}_L$ of the reference DG scheme determined by (2.7), where the numerical flux, the numerical flux balance, the flux quadrature and the source quadrature are determined by (2.8), (2.9) and (2.10) on the reference mesh,

3. the prediction operator $\mathcal{A}_{\tilde{\mathcal{D}}^n}$ determined by (3.31) realizing the refinement step (4.15) in Algorithm 1; the prediction set $\tilde{\mathcal{D}}^n$ is assumed to (i) satisfy the reliability condition (4.12), (ii) to be a tree, i.e., (4.13) holds, and (iii) not to change the data, i.e.,

$$\mathcal{A}_{\tilde{\mathcal{D}}(\mathbf{v})}\mathbf{v} = \mathbf{v} \quad \text{with } \mathbf{v} = \mathcal{A}_\varepsilon \mathbf{w} \text{ for all } \mathbf{w} \in (\mathbb{R}^p)^{N_L}. \tag{5.4}$$

   This, in particular, implies the inclusion $\tilde{\mathcal{D}}(\mathbf{v}) \supset \mathcal{D}_\varepsilon(\mathbf{v})$.

4. the limiting operator $\Pi_{L,\tilde{\mathcal{D}}^n}$ realizing the limiting step (4.16) in Algorithm 1, where limiting is performed only for the *higher* order coefficients on the *highest* resolution level L of the *adaptive* grid,

5. the evolution operator $\mathcal{E}_{L,\tilde{\mathcal{D}}^n}$ realizing the evolution step (4.17) in Algorithm 1, where the data corresponding to the adaptive grid are evolved, and

6. the approximation operator $\mathcal{A}_\varepsilon$ determined by (3.32) realizing the coarsening step (4.18) in Algorithm 1 with respect to the set $\mathcal{D}^n$ of significant details.

Then *one* step of the reference and the adaptive scheme can be represented by these discrete operators as

$$\mathbf{v}_L^n = \mathcal{E}_L\, \Pi_L\, \mathbf{v}_L^{n-1}, \quad \overline{\mathbf{v}}_L^n = \mathcal{A}_\varepsilon\, \mathcal{E}_{L,\tilde{\mathcal{D}}^n}\, \Pi_{L,\tilde{\mathcal{D}}^n}\, \mathcal{A}_{\tilde{\mathcal{D}}^n}\, \overline{\mathbf{v}}_L^{n-1}. \tag{5.5}$$

In the following this step will be referred to as *update*.

Finally, for any mean function $\hat{w}_L$ of a function $w_L \in S_L^p$ that is defined by (2.11), we conclude from its single-scale representation (3.7)

$$\|\hat{w}_L\|_{L^q(\Omega)}^q = h_L \sum_{k\in\mathcal{I}_L} |w_{L,k,0}|^q =: \|\mathcal{R}\mathbf{w}_L\|_q^q, \qquad q \in [1,\infty) \tag{5.6}$$

$$\|\hat{w}_L\|_{L^\infty(\Omega)} = \max_{k\in\mathcal{I}_L} |w_{L,k,0}| =: \|\mathcal{R}\mathbf{w}_L\|_\infty. \tag{5.7}$$

Here $\mathcal{R} : (\mathbb{R}^p)^{N_L} \to \mathbb{R}^{N_L}$ is the restriction of a vector of higher order coefficients $\mathbf{w}_L = (w_{L,k,i})_{k\in\mathcal{I}_L, i\in\mathcal{P}}$ to the vector of mean values $\mathcal{R}\mathbf{w}_L := (w_{L,k,0})_{k\in\mathcal{I}_L}$. Note that we may estimate the L^q-norm, $q \in [1,\infty)$, by the L^∞-norm

$$\|\hat{w}_L\|_{L^q(\Omega)} \le (b-a)^{1/q}\, \|\hat{w}_L\|_{L^\infty(\Omega)} \quad \text{resp.} \quad \|\mathcal{R}\mathbf{w}_L\|_q \le (b-a)^{1/q}\, \|\mathcal{R}\mathbf{w}_L\|_\infty, \tag{5.8}$$

since the domain $\Omega = [a,b]$ is bounded. Then we can estimate the perturbation error.

Theorem 1 *(Uniform boundedness of perturbation error)*
Let the following assumptions hold:

(A1) The initial data are consistent, i.e.,
$\|\mathcal{R}(\overline{\mathbf{v}}_L^0 - \mathbf{v}_L^0)\|_1 \le C_1\, \varepsilon$ *with fixed* C_1.

(A2) The approximation error is uniformly bounded, i.e.,
$\|\mathcal{R}(\mathcal{A}_\varepsilon\, \mathbf{w} - \mathbf{w})\|_1 \le C_2\, \varepsilon$ *for all* $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$ *with fixed* C_2.

(A3) Both the reference and the adaptive scheme have uniformly bounded solutions, i.e.,
$\|\mathcal{R}\overline{\mathbf{v}}_L^\mu\|_\infty, \|\mathcal{R}\mathbf{v}_L^\mu\|_\infty \le C(T,u_0)$ *for all* $\mu = 0,\dots,n$.

(A4) The perturbation by the adaptive scheme in each time step is uniformly bounded, i.e.,
$\|\mathcal{R}(\mathcal{E}_{L,\tilde{\mathcal{D}}(\mathbf{w})}\,\Pi_{L,\tilde{\mathcal{D}}(\mathbf{w})}\,\mathcal{A}_{\tilde{\mathcal{D}}(\mathbf{w})}\,\mathbf{w} - \mathcal{E}_L\,\Pi_L\,\mathcal{A}_{\tilde{\mathcal{D}}(\mathbf{w})}\,\mathbf{w})\|_1 \le C_4\,\varepsilon$
holds for all $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$ *with* $\|\mathcal{R}\mathbf{w}\|_\infty \le C(T,u_0)$.

(A5) The reference scheme is locally l_1*-stable with uniform constants, i.e.,*
$\|\mathcal{R}(\mathcal{E}_L\,\Pi_L\,\mathbf{w} - \mathcal{E}_L\,\Pi_L\,\overline{\mathbf{w}})\|_1 \le (1 + C_s\tau)\|\mathcal{R}(\mathbf{w} - \overline{\mathbf{w}})\|_1 + C_5\,h_L^\alpha$
holds for $\mathbf{w}, \overline{\mathbf{w}} \in (\mathbb{R}^p)^{N_L}$ *with* $\|\mathcal{R}\mathbf{w}\|_\infty, \|\mathcal{R}\overline{\mathbf{w}}\|_\infty \le C(T,u_0)$.

Then the perturbation error is uniformly bounded for $\tau \le \tau_{max}$*, i.e.,*

$$\|\hat{v}_{L,\varepsilon}^n - \hat{v}_L^n\|_{L_1(\Omega)} = \|\mathcal{R}(\overline{\mathbf{v}}_L^n - \mathbf{v}_L^n)\|_1 \le C(h_L^\alpha + \varepsilon)/\tau \qquad \forall n\,\tau \le T. \tag{5.9}$$

where the constant C *is given by* $C = \bar{C}\,(T + \tau_{max})$ *if* $C_s = 0$ *and* $C = \bar{C}e^{C_s(T+\tau_{max})}/C_s$ *if* $C_s \neq 0$ *with* $\bar{C} = \max\big(C_1, C_2 + C_4, C_5\big)$.

This theorem can be considered a roadmap to verify the boundedness of the perturbation error which is needed to balance the discretization error and the perturbation error in (5.1). We have not yet specified a particular choice for the MRA, the prediction strategy, the limiter and the reference DG scheme. The main objective and also the most original part of the present work is the appropriate choice of these key ingredients such that the assumptions of Theorem 1 hold. For this purpose, we first comment on these assumptions before proving the theorem:

Remark 1 *(Remarks on Assumptions (A1) – (A5))*

(A1) The consistency of the initial values can be ensured by first computing the coefficients $v_{L,k,i}^0 := \langle u_0, \tilde{\varphi}_{L,k,i}\rangle$ *and then applying hard thresholding to* $\mathbf{v}_L$*, i.e.,* $\overline{\mathbf{v}}_L^0 := \mathcal{A}_\varepsilon \mathbf{v}_L^0$*. Then Assumption (A1) can be concluded from Proposition 1 for the approximation error.*

(A2) The approximation error immediately follows from Proposition 1. To see this, we note that $\|\mathcal{R}(\mathbf{w}_\varepsilon - \mathbf{w})\|_q = \|\hat{w}_\varepsilon - \hat{w}\|_{L^q(\Omega)}$*, where* $\hat{w}$ *and* $\hat{w}_\varepsilon$ *are the mean functions of the functions* w *and* w_ε *defined by the single-scale coefficients* $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$ *and* $\mathbf{w}_\varepsilon := \mathcal{A}_\varepsilon\,\mathbf{w} \in (\mathbb{R}^p)^{N_L}$ *according to (3.7).*

(A3) Note that the local *Lipschitz constant* L_F *of the numerical flux* F *enters the constant* C_5 *in (A5). To render this constant independent of the discretization parameters, the data of both the reference and the adaptive DG scheme in the mean have to be uniformly bounded. For the reference DG scheme this holds true for sufficiently small grid size* $h_L \le h_{max}$ *and a sufficiently small constant ratio* τ/h_L*, see Appendix A.3. For the adaptive scheme this will be verified in Section 5.2.4 under a similar bound for* h_L *and a bound on* ε *of the form* $\varepsilon \le C\,h_L$ *with constant* C*.*

(A4) In order to verify this assumption, we need to design a prediction strategy and an adaptive limiter, that do not introduce an error larger than ε*. By setting* $\Pi_{L,\tilde{\mathcal{D}}^n} = \Pi_L$ *and* $\mathcal{E}_{L,\tilde{\mathcal{D}}^n} = \mathcal{E}_L$ *this trivially holds, but we seek a much more efficient solution using less degrees of freedom.*

(A5) The local l^1*-stability of the reference scheme can be proven to hold for a sufficiently small CFL condition. It is strongly related to the convergence of the DG scheme. The basic idea is to interpret the DG scheme in the mean (2.7) with numerical flux (2.19) as a perturbation of a monotone finite volume scheme. Then it is crucial to assume that the*

limiter Π_L does not modify the zero order coefficients (mean values) and (2.21) holds with uniform constant C_α. For convenience of the reader we detail the proof in Appendix A.3. Note that there seems to be no result available verifying convergence to the entropy solution of the initial value problem (2.1) for general fluxes that avoids the uniform bound (2.21) of the point evaluations at the cell interfaces. Therefore we confine ourselves to this class of converging DG schemes. Furthermore the constant C_s is related to the source function s. It will vanish for a homogeneous problem, i.e., $s \equiv 0$. Finally, the constant C_5 depends on the local Lipschitz bound for the numerical flux. It can be estimated by a uniform bound due to the boundedness of the reference scheme, see the remark on (A3).

Proof: (Theorem 1) In a first step we split the perturbation error into three parts

$$\|\mathcal{R}\mathbf{e}_L^n\|_1 = \|\mathcal{R}(\overline{\mathbf{v}}_L^n - \mathbf{v}_L^n)\|_1 \le a_{n-1} + b_{n-1} + c_{n-1}$$

with

$$\begin{aligned} a_{n-1} &:= \|\mathcal{R}(\mathcal{A}_\varepsilon\, \mathcal{E}_{L,\tilde{\mathcal{D}}^n}\, \Pi_{L,\tilde{\mathcal{D}}^n}\, \mathcal{A}_{\tilde{\mathcal{D}}^n}\, \overline{\mathbf{v}}_L^{n-1} - \mathcal{E}_{L,\tilde{\mathcal{D}}^n}\, \Pi_{L,\tilde{\mathcal{D}}^n}\, \mathcal{A}_{\tilde{\mathcal{D}}^n}\, \overline{\mathbf{v}}_L^{n-1})\|_1, \\ b_{n-1} &:= \|\mathcal{R}(\mathcal{E}_{L,\tilde{\mathcal{D}}^n}\, \Pi_{L,\tilde{\mathcal{D}}^n}\, \mathcal{A}_{\tilde{\mathcal{D}}^n}\, \overline{\mathbf{v}}_L^{n-1} - \mathcal{E}_L\, \Pi_L\, \mathcal{A}_{\tilde{\mathcal{D}}^n}\, \overline{\mathbf{v}}_L^{n-1})\|_1, \\ c_{n-1} &:= \|\mathcal{R}(\mathcal{E}_L\, \Pi_L\, \mathcal{A}_{\tilde{\mathcal{D}}^n}\, \overline{\mathbf{v}}_L^{n-1} - \mathcal{E}_L\, \Pi_L\, \mathbf{v}_L^{n-1})\|_1. \end{aligned}$$

Then the first term can be estimated by the approximation property (A2), the second term by the boundedness of the perturbation (A4) and the third term by the stability property (A5). Note that by (5.4) and (5.5) we have

$$\|\mathcal{R}(\mathcal{A}_{\tilde{\mathcal{D}}^n}\, \overline{\mathbf{v}}_L^{n-1} - \mathbf{v}_L^{n-1})\|_1 = \|\mathcal{R}(\overline{\mathbf{v}}_L^{n-1} - \mathbf{v}_L^{n-1})\|_1.$$

Hence the perturbation error can be bounded by

$$\|\mathcal{R}\mathbf{e}_L^n\|_1 \le \|\mathcal{R}\mathbf{e}_L^{n-1}\|_1\,(1 + C_s\,\tau) + C_5\, h_L^\alpha + (C_2 + C_4)\,\varepsilon.$$

Repeating this argument we obtain further

$$\|\mathcal{R}\mathbf{e}_L^n\|_1 \le \|\mathcal{R}\mathbf{e}_L^0\|_1\,(1 + C_s\,\tau)^n + (C_5\, h_L^\alpha + \varepsilon\,(C_2 + C_4))\,\sum_{i=0}^{n-1}(1 + C_s\,\tau)^i.$$

We now put $\bar{C} = \max(C_1, C_2 + C_4, C_5)$. In case of $C_s \neq 0$ we conclude with Assumption (A1)

$$\|\mathcal{R}\mathbf{e}_L^n\|_1 \le (h_L^\alpha + \varepsilon)\bar{C}\frac{(1 + C_s\tau)^{n+1} - 1}{C_s\tau} \le (h_L^\alpha + \varepsilon)\bar{C}\frac{e^{C_s(n+1)\tau} - 1}{C_s\tau},$$

otherwise we obtain

$$\|\mathcal{R}\mathbf{e}_L^n\|_1 \le (h_L^\alpha + \varepsilon)\bar{C}\frac{n\tau + \tau}{\tau}.$$

Since the maximal number of time steps is bounded by $n \le T/\tau$ for a bounded time interval $[0, T]$, $T < \infty$, and the time step is bounded by $\tau_{\max}$, the assertion holds. □

From Theorem 1 and (5.1) we immediately conclude that the accuracy of the reference DG scheme is preserved, provided that ε and α are appropriately chosen.

Corollary 1 *(Choice of the threshold parameter) If the discretization error of the reference DG scheme is bounded by $\|\hat{\eta}_L(t,\cdot)\|_{L^1(\Omega)} \le C\, 2^{-\beta\, L}$ for some $\beta > 0$ and $\alpha \ge 1 + \beta$ in the entropy bound (2.21), then the accuracy is preserved by the adaptive MR-DG scheme provided that $\varepsilon \sim 2^{-\gamma\, L}$ with $\gamma = 1 + \beta$ and the time step τ is limited by a CFL condition.*

In order to apply the above perturbation analysis, we now need to design a reliable prediction strategy and an appropriate limiting process such that the reliability condition (4.12) holds true and the adaptive MR-DG scheme is uniformly bounded, i.e., the assumptions (A3) and (A4) can be verified.

5.2 Prediction and limiting

The efficiency of the adaptive MR-DG scheme crucially relies on the compression rate of the threshold process, since time evolution is only performed for the significant multiscale coefficients or, alternatively, the single-scale coefficients corresponding to the adaptive grid. On the other hand, if we miss a significant detail in the prediction step, then it might contribute significantly to the threshold error and, hence, dominate the discretization error of the reference DG scheme. Therefore the prediction set has to be designed such that all detail coefficients *not* contained in the prediction set remain non-significant after the update (5.5) corresponding to Algorithm 1.

Since the update contains both the limiting process *and* the evolution step, see Step 2 in Algorithm 1, the prediction strategy must account for both nonlinearities.

In the context of the adaptive MR-FV scheme, a prediction strategy has been developed for which the reliability condition (4.12) could be analytically verified, see [16, 27]. Unfortunately, we can not apply the same strategy to the adaptive MR-DG scheme, because of the limiting process and missing regularity of the primal multiwavelet functions.

In order to verify (A2) in Theorem 1 and later on (B2) in Theorem 4, we split the evolution error into two parts:

$$\|\mathcal{R}(\mathcal{E}_{L,\tilde{\mathcal{D}}(\mathbf{w})}\,\Pi_{L,\tilde{\mathcal{D}}(\mathbf{w})}\,\mathbf{v}-\mathcal{E}_L\,\Pi_L\,\mathbf{v})\|_q \le \|\mathcal{R}(\mathcal{E}_{L,\tilde{\mathcal{D}}(\mathbf{w})}\,\mathbf{u}-\mathcal{E}_L\,\mathbf{u})\|_q + \|\mathcal{R}(\mathcal{E}_L\Pi_{L,\tilde{\mathcal{D}}(\mathbf{w})}\mathbf{v}-\mathcal{E}_L\Pi_L\,\mathbf{v})\|_q,$$

where $\mathbf{v} := \mathcal{A}_{\tilde{\mathcal{D}}(\mathbf{w})}\,\mathbf{w}$ and $\mathbf{u} := \Pi_{L,\tilde{\mathcal{D}}(\mathbf{w})}\,\mathbf{v}$ are defined by data $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$ and $q \in [1,\infty]$. Thus, we need to bound the two terms by ε, i.e., we have to construct the prediction set in a way such that

$$\text{(E)}\quad \|\mathcal{R}(\mathcal{E}_{L,\tilde{\mathcal{D}}(\mathbf{w})}\mathbf{u}-\mathcal{E}_L\mathbf{u})\|_q \lesssim \varepsilon \qquad \text{and} \qquad \text{(L)}\quad \|\mathcal{R}(\mathcal{E}_L\Pi_{L,\tilde{\mathcal{D}}(\mathbf{w})}\mathbf{v}-\mathcal{E}_L\Pi_L\,\mathbf{v})\|_q \lesssim \varepsilon$$

hold. This will ensure reliability of the evolution step and the limiting process: For the first estimate (E), we thus have to verify that the detail coefficients $d_{l,k,i}$, $(l,k,i) \notin \tilde{\mathcal{D}}$, that are not updated by the adaptive MR-DG scheme stay bounded by ε_l. For the second estimate (L), we have to make sure that the difference of the two different limiters is bounded, and then employ the Lipschitz-continuity of $\mathcal{R}\mathcal{E}_L$ that follows from the Lipschitz-continuity of the numerical flux function. Due to conditions (L) and (E) the evolution error is uniformly bounded and the following theorem holds.

Theorem 2 *(Uniform boundedness of evolution error) Let the prediction fulfil the estimates in Lemma 1 stated below. Let $\Omega = [a,b]$ be bounded and $\varepsilon_l = 2^{l-L}\varepsilon$. Then the evolution error is uniformly bounded with respect to $L^q(\Omega)$, $q \in [1,\infty]$, i.e., there exist constants C_q independent of L and ε such that*

$$\|\mathcal{R}(\mathcal{E}_{L,\tilde{\mathcal{D}}(\mathbf{w})}\,\Pi_{L,\tilde{\mathcal{D}}(\mathbf{w})}\,\mathcal{A}_{\tilde{\mathcal{D}}(\mathbf{w})}\,\mathbf{w}-\mathcal{E}_L\,\Pi_L\,\mathcal{A}_{\tilde{\mathcal{D}}(\mathbf{w})}\,\mathbf{w})\|_q \le C_q\,\varepsilon \tag{5.10}$$

holds for all $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$ with $C_q = C_q(\mathbf{w})$.

From the construction of the prediction to be presented in Section 5.2.1 we will conclude the estimates (L) and (E) in Sections 5.2.2 and 5.2.3.

5.2.1 Prediction strategy

The reliability condition (4.12) implies that for any index $(l,k,i) \notin \tilde{\mathcal{D}}^{n+1}$ the detail coefficient remains bounded by the order of the threshold value ε_l after having performed the update, i.e., $|d_{l,k,i}^{n+1}| \lesssim \varepsilon_l$. We therefore have to ensure that the terms on the right-hand side of the discrete evolution equation (4.7) can be bounded by

$$|\overline{B}_{l,k,i}^n| \leq 2^L \varepsilon_l, \quad |\overline{G}_{l,k,i}^n| \leq 2^L \varepsilon_l, \quad |\overline{S}_{l,k,i}^n| \leq 2^L \varepsilon_l.$$

Note that the factor 2^L cancels out because the time step size τ is proportional to the space discretization, i.e., $\tau \sim 2^{-L}$ due to a CFL condition. As we will see in Section 5.2.3, these conditions are satisfied, if the prediction strategy is designed such that the following Lemma holds true. For this purpose, we have to consider the update composed of the evolution step and the limiting process. In Section 5.2.2 we will state properties of the reference limiter. This will help to enforce estimates in the prediction that will be used to control the action of the limiter on the data.

Lemma 1 *(Properties of the prediction set): Let $\tilde{\mathcal{D}}(\mathbf{w})$ be the prediction set determined by the data $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$ and $\mathbf{v} = \mathcal{A}_{\tilde{\mathcal{D}}(\mathbf{w})}\mathbf{w}$. Then the following estimates hold for the function v_L corresponding to the coefficients $\mathbf{v}$ on any cell $V_{l,k}$ with $(l,k,i) \notin \tilde{\mathcal{D}}(\mathbf{w})$ and $l < L$ for all $i \in \mathcal{P}$:*

- *The function v_L is a polynomial on $V_{l,k}$, i.e.,*

$$v_L|_{V_{l,k}} \in \Pi_{p-1} \tag{P0}$$

- *Reliability of evolution step:*

$$|v_L(x^+) - v_L(x^-)| \leq \varepsilon, \quad x = x_{l,k} \text{ or } x = x_{l,k+1} \tag{E1}$$

$$\frac{(h_l/2)^p}{p!} \left| \frac{d^{p+1}}{dx^{p+1}} f(v_L(x)) \right| \leq 2^l \varepsilon \quad \forall x \in V_{l,k} \tag{E2}$$

$$\frac{(h_l/2)^p}{p!} \left| \frac{d^{p}}{dx^{p}} s(v_L(x)) \right| \leq 2^l \varepsilon \quad \forall x \in V_{l,k} \tag{E3}$$

- *Reliability of limiter process:*
  - *Monotonicity at cell interfaces:*

$$\begin{aligned}
&(i) \quad \begin{aligned} &\operatorname{sign}(v_{L,k^+-1,0} - v_L(x_{l,k}^+)) = \operatorname{sign}(v_L(x_{l,k}^+) - v_{L,k^+,0}) \quad \text{or} \\ &|v_L(x_{l,k}^+) - v_{L,k^+,0}| \leq \varepsilon \text{ and } |v_L(x_{L,k^++1}^-) - v_{L,k^+,0}| \leq \varepsilon \end{aligned} \\
&(ii) \quad \begin{aligned} &\operatorname{sign}(v_{L,k^-,0} - v_L(x_{l,k+1}^-)) = \operatorname{sign}(v_L(x_{l,k+1}^-) - v_{L,k^-+1,0}) \quad \text{or} \\ &|v_L(x_{l,k+1}^-) - v_{L,k^-,0}| \leq \varepsilon \text{ and } |v_L(x_{L,k^-}^+) - v_{L,k^-,0}| \leq \varepsilon \end{aligned} \\
&\text{with } k^+ = 2^{(L-l)}k \text{ and } k^- = 2^{(L-l)}(k+1) - 1
\end{aligned} \tag{L4}$$

  - *Monotonicity or small curvature on cell:*

$$v_L \text{ is either monotone on } V_{l,k} \text{ or } |v_L''(x)| \leq \frac{1}{4h_L^2}\varepsilon \text{ holds for all } x \in V_{l,k} \tag{L5}$$

– *Entropy correction:*

$$|v_L'(x)| \leq C_\alpha h^{\alpha-1} \quad \forall x \in V_{l,k} \tag{L6}$$

- *Buffer zone to cells on lower levels:*

$$\begin{aligned} &\textit{If for the grid } \tilde{\mathcal{G}}(\mathbf{w}) \textit{ corresponding to } \tilde{\mathcal{D}}(\mathbf{w}) \textit{ the} \\ &\textit{index } (L,k) \in \tilde{\mathcal{G}}(\mathbf{w}) \textit{ with } (L,k-1) \notin \tilde{\mathcal{G}}(\mathbf{w}) \textit{ or } (L,k+1) \notin \tilde{\mathcal{G}}(\mathbf{w}), \\ &\textit{then (L4)–(L6) must hold for the index } (L,k) \textit{ as well.} \end{aligned} \tag{B7}$$

Property (P0) will directly follow from the construction of the prediction set, for which we enforce the tree property (4.13). Then there are no detail coefficients corresponding to cells on higher levels contained in the cell $V_{l,k}$, and, hence, v_L is a polynomial on that cell. Estimate (E1) bounds the jumps of the function v_L at the cell interfaces. Thus the contribution of the numerical fluxes in the evolution of the detail coefficients will be bounded as well. This we will infer later on from the Lipschitz-continuity of the numerical flux and the definition (4.8) of the numerical flux balance $\overline{B}^n_{l,k,i}$. Similarly, the contributions by the flux quadrature $\overline{G}^n_{l,k,i}$ and the source quadrature $\overline{S}^n_{l,k,i}$ defined by (4.9) and (4.10) are also bounded. For this purpose, we will perform a Taylor expansion and employ regularity of the flux f and the source function s, respectively, as well as the vanishing moments of the dual wavelets.

The estimates (L4) and (L5) ensure, that in regions where the reference limiter modifies significantly the data on level L, we resolve these modifications such that the error introduced by the adaptive limiter is bounded. Furthermore, the estimate (L6) is needed to enforce convergence to the entropy solution, see the discussion in Section 5.2.2. Note that the conditions (L4)–(L6) are adjusted to the reference limiter that is chosen as the Shu limiter (2.17) with entropy correction (2.21). Then the reference DG scheme is known to converge to the entropy solution.

In order to proof Lemma 1, we will now define the prediction set judisciously, i.e., we locally refine the adaptive grid until either the estimates hold for some $l' < L$ or we arrive at the highest level $l' = L$. Moreover, to estimate the evolution error, see condition (E), we need later on that the estimates for the evolution still hold after limiting, i.e., (E1)–(E3) remain true with $\mathbf{v}$ replaced by $\Pi_{L,\tilde{\mathcal{D}}^n}\mathbf{v}$. For this reason we have to inflate the prediction set in the neighborhood of cells located on the highest refinement level L, otherwise the limiter could alter the point value of the data from outside in (E1), this is enforced by (B7).

Algorithm 2 *(Construction of prediction set) Let* $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$*. To compute the prediction set* $\tilde{\mathcal{D}} = \tilde{\mathcal{D}}(\mathbf{w})$ *we perform the following nine steps:*

Step 0. (Initial inclusion of $\mathcal{D}_0$) *The prediction set has to include all indices corresponding to non-vanishing detail coefficients of the data* $\mathbf{w}$ *contained in the support* $\mathcal{D}_0 = \{(l,k,i) : |d_{l,k,i}| > 0\}$*, i.e.,* $\tilde{\mathcal{D}} \leftarrow \mathcal{D}_0$*. In particular, if* $\mathbf{w}$ *is the result of a thresholding process, e.g., the thresholded data of the last time step, then* $\mathbf{w} = \mathcal{A}_\varepsilon \mathbf{w}$ *and* $\mathcal{D}_0(\mathbf{w}) = \mathcal{D}_\varepsilon(\mathbf{w})$*.*

Then, the Steps E1–E3 and L4–L6 are performed for any index $(l,k) \in \mathcal{G}_0$ *corresponding to all cells of the current grid:*

Step E1. (Jump at cell interfaces) *The jump at the left and right cell interface is determined, i.e.,* $J_L := |v_L(x^+_{l,k}) - v_L(x^-_{l,k})|$ *and* $J_R := |v_L(x^+_{l,k+1}) - v_L(x^-_{l,k+1})|$*. If the jump at* $x_{l,k}$ *is larger than* ε*, we add the left and right cell neighbors attached to the grid point* $x_{l,k}$ *on level* L *to the*

adaptive grid. This is enforced by the inclusion of the corresponding details on level $L-1$ to the prediction set, i.e.,

$$\tilde{\mathcal{D}} \leftarrow \tilde{\mathcal{D}} \cup \{(L, 2^{L-l-1}\,k-1, i)\,;\, i \in \mathcal{P}\} \cup \{(L, 2^{L-l-1}\,k, i)\,;\, i \in \mathcal{P}\}.$$

Similarly we proceed with J_R and $x_{l,k+1}$. Note that by refining we create new cell interfaces. However, the function v_L is continuous at these internal grid points and, hence, (E1) also holds for the new cells at these points.

Step E2. (Flux quadrature bound) *For each cell $(l,k) \in \mathcal{G}_0$ we calculate bounds D_v^r for the r-th derivative of v_L in $V_{l,k}$, i.e.,*

$$\left|\frac{d^r}{dx^r} v_L(x)\right| \le h_l^{-r} \sum_{i\in\mathcal{P}} |v_{l,k,i}| D_{r,i} =: D_v^r, \qquad \forall\, x \in V_{l,k},$$

where the constants $D_{r,i}$ are determined according to (3.19). Then we may bound the $(p+1)$st derivative of the flux f via chain rule by

$$\left\| \frac{d^{p+1}}{dx^{p+1}} f(v_L(t^n,\cdot)) \right\|_{L^\infty(V_{l,k})} \le \sum_{\mathbf{k}\in T_{p+1}} \frac{(p+1)!}{k_1!\cdots k_{p+1}!} D_f^{|\mathbf{k}|} \prod_{j=1,k_j\ge 1}^{p+1} \left(\frac{D_v^j}{j!}\right)^{k_j},$$

where the set T_{p+1} contains all $(p+1)$-tupels $\mathbf{k} = (k_1,\dots,k_{p+1})$ of non-negative integers such that $\sum_{i=1}^{p+1} i\,k_i = p+1$ and $|\mathbf{k}| = \sum_{i=1}^{p+1} k_i$. The constant D_f^r is a problem-dependent bound for the rth derivative of the flux, i.e.,

$$D_f^r := \max_{|u|\le C(T,u_0)} |D^r f(u)|.$$

Finally, we ensure (E2) by requiring

$$(h_l/2)^p \sum_{\mathbf{k}\in T_{p+1}} \frac{p+1}{k_1!\cdots k_{p+1}!} D_f^{|\mathbf{k}|} \prod_{j=1,k_j\ge 1}^{p+1} \left(\frac{D_v^j}{j!}\right)^{k_j} \le 2^l \varepsilon.$$

In order to enforce this inequality, we determine a level

$$l' \ge \log_{2^{p+1}} \left(\frac{1}{\varepsilon} (h_0/2)^p \sum\nolimits_{\mathbf{k}\in T_{p+1}} \frac{p+1}{k_1!\cdots k_{p+1}!} D_f^{|\mathbf{k}|} \prod_{j=1,k_j\ge 1}^{p+1} \left(\frac{D_v^j}{j!}\right)^{k_j} \right)$$

depending on the polynomial degree p of the basis functions. Note that due to dyadic grid refinement $h_l = 2^{-l}\,h_0$. Then we include all subcells $V_{l',k'} \subset V_{l,k}$ of $V_{l,k}$ on higher refinement levels, i.e.,

$$\tilde{\mathcal{D}} \leftarrow \tilde{\mathcal{D}} \cup \{(l',k',i)\,;\, V_{l',k'} \subset V_{l,k},\ i \in \mathcal{P}\}. \tag{5.11}$$

Step E3. (Source quadrature bound) *We proceed similarly for derivatives of the source function s, but only up to the order p, i.e.,*

$$l' \ge \log_{2^{p+1}} \left(\frac{1}{\varepsilon} (h_0/2)^p \sum\nolimits_{\mathbf{k}\in T_p} \frac{1}{k_1!\cdots k_p!} D_s^{|\mathbf{k}|} \prod_{j=1,k_j\ge 1}^{p} \left(\frac{D_v^j}{j!}\right)^{k_j} \right)$$

with $D_s^{|\mathbf{k}|}$ instead of $D_f^{|\mathbf{k}|}$.

Step L4. (Limiting at cell interfaces) *Let be $k^+ = 2^{(L-l)}k$ and $k^- = 2^{(L-l)}(k+1)-1$. We calculate the mean values $v_{L,k^+-1,0}$, $v_{L,k^+,0}$ and $v_{L,k^-,0}$, $v_{L,k^-+1,0}$ of the neighbors on level L to the cell interfaces at $x_{l,k}$ and $x_{l,k+1}$, respectively. An efficient procedure for this is described in Appendix A.4. Then we evaluate v_L at the cell interfaces. First of all we consider the left cell interface $x_{l,k}^+$: if we do not have monotonicity, i.e.,*

$$\operatorname{sign}(v_{L,k^+-1,0} - v_L(x_{l,k}^+)) \neq \operatorname{sign}(v_L(x_{l,k}^+) - v_{L,k^+,0}),$$

then we check whether

$$|v_L(x_{l,k}^+) - v_{L,k^+,0}| \geq \varepsilon \text{ or } |v_L(x_{L,k^++1}) - v_{L,k^+,0}| \geq \varepsilon.$$

If this holds true, then limiting of the data on the cell V_{L,k^+} might introduce an error greater than ε. Therefore, we enforce refinement up to level L, i.e.,

$$\tilde{\mathcal{D}} \leftarrow \tilde{\mathcal{D}} \cup \{(L, k^+, i)\,;\ i \in \mathcal{P}\}. \tag{5.12}$$

The same check is performed at the right cell interface $x_{l,k+1}^-$: if we do not have monotonicity, i.e.,

$$\operatorname{sign}(v_{L,k^-,0} - v_L(x_{l,k+1}^-)) \neq \operatorname{sign}(v_L(x_{l,k+1}^-) - v_{L,k^-+1,0}),$$

and

$$|v_L(x_{l,k+1}^-) - v_{L,k^-,0}| \geq \varepsilon \text{ or } |v_L(x_{L,k^-}) - v_{L,k^-,0}| \geq \varepsilon,$$

then we refine up to level L, i.e.,

$$\tilde{\mathcal{D}} \leftarrow \tilde{\mathcal{D}} \cup \{(L, k^-, i)\,;\ i \in \mathcal{P}\}.$$

Step L5. (Limiting inside of cells) *If v_L is not monotone on $V_{l,k}$, the limiting might introduce an error. We first find an upper bound for the second derivative of v_L (see Step E2)*

$$|v_L''(x)| \leq D_v^2, \qquad x \in V_{l,k}.$$

Then we must have the bound

$$D_v^2 \leq \frac{1}{4h_L^2}\varepsilon,$$

otherwise we refine up to level L, i.e.,

$$\tilde{\mathcal{D}} \leftarrow \tilde{\mathcal{D}} \cup \{(L, k', i)\,;\, V_{L,k'} \subset V_{l,k},\ i \in \mathcal{P}\}. \tag{5.13}$$

Step L6. (Entropy bound on the gradient) *We find an upper bound for the first derivative of v_L (see Step E2)*

$$|v_L'(x)| \leq D_v^1, \qquad x \in V_{l,k}.$$

Then we check, if

$$D_v^1 \leq C_\alpha h_L^{\alpha-1},$$

where C_α and α stem from the entropy correction of the limiter, see Section 5.2.2. If the estimate does not hold, we refine up to level L, i.e.,

$$\tilde{\mathcal{D}} \leftarrow \tilde{\mathcal{D}} \cup \{(L, k', i)\,;\, V_{L,k'} \subset V_{l,k},\ i \in \mathcal{P}\}. \tag{5.14}$$

Step B7. (Additional refinement of neighbors to level L) *The set $\tilde{\mathcal{D}}$ determined by the previous steps characterizes an adaptive grid $\tilde{\mathcal{G}}$ applying the procedure described in Section 4. The cells in $\tilde{\mathcal{G}}$ on the highest refinement level are given by*

$$\Sigma(\tilde{\mathcal{D}}, L) := \{(L, k)\,;\, \exists\,(L-1, \lfloor k/2 \rfloor, i) \in \tilde{\mathcal{D}}\} = \{(L, k) \in \tilde{\mathcal{G}}\}.$$

In order to verify Lemma 4 we enlarge the prediction set $\tilde{\mathcal{D}}$ such that all neighbors of cells on level L characterized by

$$\Sigma^\partial := \{(L, k) \notin \Sigma(\tilde{\mathcal{D}}, L)\,;\, (L, k-1) \in \Sigma(\tilde{\mathcal{D}}, L) \text{ or } (L, k+1) \in \Sigma(\tilde{\mathcal{D}}, L)\}$$

are included in the adaptive grid, i.e.,

$$\tilde{\mathcal{D}} \leftarrow \tilde{\mathcal{D}} \cup \{(L-1, \lfloor k/2 \rfloor, i)\,;\, (L, k) \in \Sigma^\partial,\ i \in \mathcal{P}\}.$$

Step 8. (Ensure tree property) *Enlarge the set $\tilde{\mathcal{D}}$, such that the tree property* (4.13) *holds.*

Obviously, Lemma 1 holds true by construction of the prediction set according to the above algorithm, where (P0) holds due to Step 8.

5.2.2 Limiting strategy

We now have to define the limiter $\Pi_{L,\tilde{\mathcal{D}}}$ of the adaptive MR-DG scheme. For this purpose, we first have to specify the limiter Π_L of the reference DG scheme: Let $v \in S_L^p$ be determined by the coefficients $\mathbf{v} \in (\mathbb{R}^p)^{N_L}$ and let $u \in S_L^p$ be characterized by $\mathbf{u} = \Pi_L \mathbf{v}$. Then we assume the following properties to hold for Π_L:

(RefL1) The mean values remain unchanged, i.e., $u_{L,k,0} = v_{L,k,0}$.

(RefL2) The effect of the limiter on the data can be described as a scaling of the difference of the point value at the boundary and the mean value of the cell, see Figure 4, i.e.,

$$u(x_{L,k}^+) - u_{L,k,0} = c_{L,k}^+ \left(v(x_{L,k}^+) - v_{L,k,0}\right)$$

where $c_{L,k}^+ \in [0, 1]$, and analogously for $u(x_{L,k+1}^-)$. The coefficients $c_{L,k}^+$, $c_{L,k+1}^-$ only depend on the mean values $v_{L,k,0}$ of the cell, its neighbors $v_{L,k\pm1,0}$ and the point values at the cell interfaces $v(x_{L,k}^+)$, $v(x_{L,k+1}^-)$.

(RefL3) If the gradient is bounded, i.e.,

$$|v'(x)| \le C_\alpha h_L^{\alpha-1}, \tag{5.15a}$$

and the data are monotone in the following sense

$$\operatorname{sign}(v_{L,k-1,0} - v(x_{L,k}^+)) = \operatorname{sign}(v(x_{L,k}^+) - v_{L,k,0}) \tag{5.15b}$$

$$\operatorname{sign}(v_{L,k,0} - v(x_{L,k+1}^-)) = \operatorname{sign}(v(x_{L,k+1}^-) - v_{L,k+1,0}) \tag{5.15c}$$

then the limiter does not change the data, i.e., $c_{L,k}^+ = c_{L,k+1}^- = 1$.

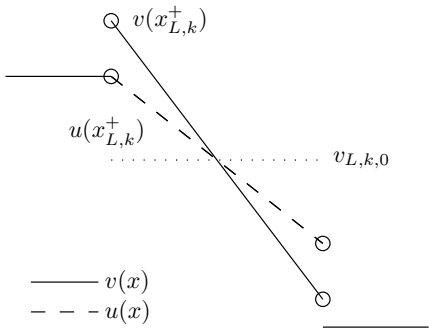

Figure 4: Effect of limiter on point value at cell interface

These properties are fulfilled, for example by the Shu limiter with entropy stabilization, i.e.,

$$\tilde{\overline{v}}^{\pm}_{L,k} := \begin{cases} \overline{v}^{\pm}_{L,k} & , \ |\overline{v}^{\pm}_{L,k}| \le M\, h_L^2 \\ m(\overline{v}^{\pm}_{L,k}, v_{L,k+1,0} - v_{L,k,0}, v_{L,k,0} - v_{L,k-1,0}, \operatorname{sign}(\overline{v}^{\pm}_{L,k}) C_\alpha\, h_L^\alpha) & , \text{ else} \end{cases} \tag{5.16}$$

where (RefL3) holds, because the minmod limiter selects the first argument due to the estimate

$$|\overline{v}^{+}_{L,k}| = \left| \int_{x_k}^{\bar{x}} v'(x)dx \right| \le h_L C_\alpha h_L^{\alpha-1} = C_\alpha h_L^\alpha.$$

If the entropy stabilization term $C_\alpha h_L^\alpha$ is not needed, for example for convex flux functions, where convergence to the entropy solution can be shown without this term, we can in practice relax the condition (5.15a) by setting $C_\alpha = \infty$, as observed in the numerical experiments, see Section 6. Then Step L6 of the prediction strategy can be omitted.

Then we define the adaptive limiter $\Pi_{L,\tilde{\mathcal{D}}}$ by the following algorithm:

Algorithm 3 *(Adaptive limiter) Let $\tilde{\mathcal{D}} = \tilde{\mathcal{D}}(\mathbf{w})$ be the prediction set determined by the data $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$ according to Algorithm 2. The set $\tilde{\mathcal{D}}$ characterizes an adaptive grid $\tilde{\mathcal{G}}$ applying the procedure described in Section 4. Then the limited values $\overline{\mathbf{w}} := \Pi_{L,\tilde{\mathcal{D}}}\, \mathbf{w}$ are locally determined for $(l,k) \in \tilde{\mathcal{G}}$ by the following procedure:*
Case 1. (Highest level: $l = L$)
If the cell $V_{L,k}$ has a neighboring cell that is not on the same level L, it stems from Step B7 in the prediction strategy. In that case, we do nothing. Otherwise we apply the reference limiter.
Case 2. (Lower level: $l < L$)
No limiting is performed at all, i.e., $\overline{w}_{l,k,i} = w_{l,k,i}$, $i \in \mathcal{P}$.

Then by construction of the adaptive limiter and the prediction set we make the following observations:

Lemma 2 *(Point-wise error of limiter) Let $\tilde{\mathcal{D}} = \tilde{\mathcal{D}}(\mathbf{w})$ be the prediction set determined by the data $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$ and w the corresponding function. Let $\tilde{u}$ and u be the functions corresponding to $\tilde{\mathbf{u}} := \Pi_{L,\tilde{\mathcal{D}}}\mathbf{w}$ and $\mathbf{u} := \Pi_L \mathbf{w}$, respectively. Then we can estimate the difference of point evaluations at cell boundaries on the reference grid by*

$$|\tilde{u}(x^{\pm}_{L,k}) - u(x^{\pm}_{L,k})| \le \varepsilon \quad \forall\, k \in \{0, \dots, N_L\}. \tag{5.17}$$

Proof: We confine ourselfs to the case $x^+_{L,k}$, since for $x^-_{L,k}$ we may proceed in complete analogy. We now consider the cell $V_{L,k}$ and distinguish three cases:

(1) If $(L,k) \notin \tilde{\mathcal{G}}$, then $V_{L,k}$ is contained in $V_{l',k'}$ for some $(l',k') \in \tilde{\mathcal{G}}$ with $l' < L$ and estimates (L4)–(L6) hold on $V_{l',k'}$. If $(L,k) \in \tilde{\mathcal{G}}$, then, either

(2) $(L,k+1) \in \tilde{\mathcal{G}}$ and $(L,k-1) \in \tilde{\mathcal{G}}$ and thus $\tilde{u}(x^+_{L,k}) = u(x^+_{L,k})$ holds by Case 1 of Algorithm 3, or

(3) by (B7) the estimates (L4)–(L6) hold, where we set for notational convenience $(l',k') = (L,k)$.

It remains to consider a cell $V_{L,k} \subset V_{l',k'}$, $(l',k') \in \tilde{\mathcal{G}}$, where (L4)–(L6) hold on $V_{l',k'}$ for cases (1) and (3). By Case 2 of Algorithm 3 we note, that

$$\tilde{u}(x^+_{L,k}) - \tilde{u}_{L,k,0} = w(x^+_{L,k}) - w_{L,k,0} \tag{5.18}$$

and by (RefL2) we have

$$u(x^+_{L,k}) - u_{L,k,0} = c^+_{L,k}(w(x^+_{L,k}) - w_{L,k,0}). \tag{5.19}$$

Subtracting (5.18) and (5.19) and using the fact that the limiters do not change the mean values, i.e., $u_{L,k,0} = \tilde{u}_{L,k,0} = w_{L,k,0}$, we obtain

$$|\tilde{u}(x^+_{L,k}) - u(x^+_{L,k})| = (1 - c^+_{L,k})|w(x^+_{L,k}) - w_{L,k,0}|.$$

We will now verify (5.17) by concluding that either $c^+_{L,k} = 1$ holds by using (RefL3), i.e., the reference limiter does not change the data, or

$$|w(x^+_{L,k}) - w_{L,k,0}| \leq \varepsilon, \tag{5.20}$$

where we use (L4)–(L6).

For this purpose, we first note that for polynomials we have the following property: let be $q \in \Pi_p$ any polynomial and $(q_{L,\bar{k},i})_{\bar{k}\in\mathcal{I}_L, i\in\mathcal{P}} \in (\mathbb{R}^p)^{N_L}$ the corresponding coefficients determined by $q_{L,\bar{k},i} = \langle q, \tilde{\varphi}_{L,\bar{k},i}\rangle_\Omega$. If q is monotone on $V_{L,\bar{k}-1} \cup V_{L,\bar{k}}$, then

$$\operatorname{sign}(q_{L,\bar{k}-1,0} - q(x_{L,\bar{k}})) = \operatorname{sign}(q(x_{L,\bar{k}}) - q_{L,\bar{k},0}). \tag{5.21}$$

This is true, due to the fact that the mean value $q_{L,\bar{k},0}$ of q on a cell $V_{L,\bar{k}}$ is a point evaluation $q_{L,\bar{k},0} = q(\bar{x})$ for some $\bar{x} \in V_{L,\bar{k}}$.

By (L6) we see that (5.15a) holds on $V_{L,k}$. Consider the grid point $x_{L,k}$. If $x_{L,k}$ is not at the interface of $V_{l',k'}$, then w is a polynomial on $I := V_{L,k-1} \cup V_{L,k}$. Then, w is either monoton on I and (5.15b) holds by (5.21), or w is not monotone on I and, by (L5), $|w''| \leq \frac{1}{4h_L^2}\varepsilon$. In the latter case let $\bar{x}$ be the local extremum in I. Expanding $w(x)$ around $\bar{x}$ and using $w'(\bar{x}) = 0$ gives $w(x) = w(\bar{x}) + \frac{(\bar{x}-x)^2}{2}w''(\xi)$. Together with $w_{L,k,0} = w(\tilde{x})$ for some $\tilde{x} \in V_{L,k}$ and using the bound $|\bar{x} - x| \leq 2h_L$ we obtain (5.20) by

$$|w(x^+_{L,k}) - w_{L,k,0}| \leq |w(x^+_{L,k}) - w(\bar{x})| + |w(\bar{x}) - w(\tilde{x})| \leq 2\frac{(2h_L)^2}{2}\frac{1}{4h_L^2}\varepsilon = \varepsilon. \tag{5.22}$$

If $x_{L,k}$ coincides with an interface of $V_{l',k'}$, then we see by (L4) that (5.20) is fulfilled, or (5.15b) holds. Analogously we conclude for $x_{L,k+1}$ that (5.20) or (5.15c) holds. Therefore, either (5.20) holds directly, or we can conclude $c^+_{L,k} = 1$ from (RefL3). □

Since by the previous lemma, the difference of the adaptive and the reference scheme at the cell interfaces is bounded by the threshold value, this holds true for the mean values when performing one evolution step with the reference DG scheme.

Lemma 3 *(Bound on evolution of means) Consider two functions u, $\tilde{u}$ defined by the coefficients $\mathbf{u}$, $\tilde{\mathbf{u}} \in (\mathbb{R}^p)^{N_L}$ with identical mean values, i.e., $\tilde{u}_{L,k,0} = u_{L,k,0}$. If the point values at the interfaces of the reference mesh are bounded by the threshold value, i.e., $|u(x^{\pm}_{L,k}) - \tilde{u}(x^{\pm}_{L,k})| \lesssim \varepsilon$, then*

$$\|\mathcal{R}\mathcal{E}_L\mathbf{u} - \mathcal{R}\mathcal{E}_L\tilde{\mathbf{u}}\|_q \lesssim \varepsilon$$

holds for $q \in [1,\infty]$, provided that some CFL condition is satisfied.

Proof: For the time evolution of the mean values we conclude by the Lipschitz-continuity of the numerical fluxes

$$\begin{aligned}
&\|\mathcal{R}\mathcal{E}_L\,\mathbf{u} - \mathcal{R}\mathcal{E}_L\,\tilde{\mathbf{u}}\|_\infty \\
&= \max_{k\in\mathcal{I}_L} |\, u_{L,k,0} - \lambda(F(u(x^-_{L,k+1}), u(x^+_{L,k+1})) - F(u(x^-_{L,k}), u(x^+_{L,k}))) - \\
&\qquad \tilde{u}_{L,k,0} + \lambda(F(\tilde{u}(x^-_{L,k+1}), \tilde{u}(x^+_{L,k+1})) - F(\tilde{u}(x^-_{L,k}), \tilde{u}(x^+_{L,k})))| \\
&\le 2\lambda \max_{k\in\mathcal{I}_L} |F(u(x^-_{L,k}), u(x^+_{L,k})) - F(\tilde{u}(x^-_{L,k}), \tilde{u}(x^+_{L,k}))| \le 2\lambda(L_1\varepsilon + L_2\varepsilon),
\end{aligned}$$

where $L_{1,2}$ are the Lipschitz-bounds of the numerical flux function F. For $q \in [1,\infty)$, the assertion follows from the boundedness of the domain. □

Using the estimates in Lemma 2 and Lemma 3 we then conclude

Conclusion 1 *(Condition (L))*
Let $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$, $q \in [1,\infty]$ and $\mathbf{v} := \mathcal{A}_{\tilde{\mathcal{D}}(\mathbf{w})}\,\mathbf{w}$. Then Condition (L) holds, i.e.,

$$\|\mathcal{R}(\mathcal{E}_L\Pi_{L,\tilde{\mathcal{D}}(\mathbf{w})}\mathbf{v} - \mathcal{E}_L\Pi_L\mathbf{v})\|_q \lesssim \varepsilon.$$

Furthermore, from the construction of the limiter we note that the estimates for the evolution step in Lemma 1 being enforced by the prediction strategy still hold after applying the adaptive limiter.

Lemma 4 *(Invariance of estimates after limiting) Let $\tilde{\mathcal{D}}(\mathbf{w})$ be the prediction set determined by the data $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$. Then the estimates (E1), (E2) and (E3) hold for the function v_L corresponding to the* limited *coefficients $\mathbf{v} := \Pi_{L,\tilde{\mathcal{D}}}\mathbf{w}$.*

Proof: Let $(l',k') \notin \tilde{\mathcal{G}}$. Then there exists a coarse-scale cell $V_{l,k} \supset V_{l',k'}$ such that $(l,k) \in \tilde{\mathcal{G}}$ with $l < l'$. According to Case 2 of Algorithm 3 the adaptive limiter does not change the coefficients of this cell. This holds true also for the neighbor cells, because of Step B7 of Algorithm 2. Therefore, neither the point values at the cell interfaces nor the coefficients in the interior are modified by the limiter. Hence the estimates (E1), (E2) and (E3) hold. □

5.2.3 Reliability of evolution

We now are ready to verify the reliability condition of the update step, as is needed in Theorem 2. For this purpose, we investigate the evolution process (4.7) of the detail coefficients.

Theorem 3 *(Growth of non-significant details) We assume that*

1. *the flux $f : \mathbb{R} \to \mathbb{R}$ is sufficiently smooth;*
2. *the source function $s : \mathbb{R} \to \mathbb{R}$ is sufficiently smooth;*
3. *the numerical flux $F : \mathbb{R}^2 \to \mathbb{R}$ is consistent with the flux f and locally Lipschitz-continuous with local Lipschitz constant L_F;*
4. *the temporal discretization τ is chosen such that a CFL condition holds, i.e., $L_F\,\tau/h_L \leq$ CFL holds for sufficiently small CFL number.*

Let $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$ and $\tilde{\mathcal{D}}(\mathbf{w})$ the prediction set according to Algorithm 2. Then, according to Lemma 4, the estimates (E1)–(E3) *hold for $\overline{\mathbf{w}} := \Pi_{L,\tilde{\mathcal{D}}(\mathbf{w})}\mathcal{A}_{\tilde{\mathcal{D}}(\mathbf{w})}\mathbf{w}$. Consider the evolved data $\mathbf{v} := \mathcal{E}_L\overline{\mathbf{w}}$. Then the detail coefficients $d_{l,k,i}$ of $\mathbf{v}$* not *contained in the prediction set $\tilde{\mathcal{D}}(\mathbf{w})$ are* not *significant, i.e.,*

$$|d_{l,k,i}| \leq C\varepsilon_l \qquad \forall\ (l,k,i) \notin \tilde{\mathcal{D}}(\mathbf{w}), \tag{5.23}$$

where $C > 0$ is a constant depending on the local Lipschitz constant $L_F = L_F(\mathbf{w})$ of the numerical flux function.

Proof: Let $(l,k,i) \notin \tilde{\mathcal{D}}(\mathbf{w})$. Furthermore, $\bar{w} \in S_L^p$ and $w \in S_L^p$ denote the functions corresponding to $\overline{\mathbf{w}}$ and $\mathbf{w}$, respectively. Analogously, $d^{\overline{\mathbf{w}}}_{l,k,i}$ and $d^{\mathbf{w}}_{l,k,i}$ denote the detail coefficients of $\overline{\mathbf{w}}$ and $\mathbf{w}$. Then we note that $\bar{w} = w \in S_l^p$ on $V_{l,k}$ and $d^{\overline{\mathbf{w}}}_{l,k,i} = d^{\mathbf{w}}_{l,k,i} = 0$, because of (P0) and the definition of the adaptive limiter. Hence, w is differentiable on $V_{l+1,2k+r}$, $r = 0, 1$, and we may apply integration by parts on each subdomain, i.e.,

$$\int_{V_{l+1,2k+r}} f(w(x))\tilde{\psi}'_{l,k,i}(x)\,dx = -\int_{V_{l+1,2k+r}} f(w(x))_x\tilde{\psi}_{l,k,i}(x)\,dx + f(w(y))\tilde{\psi}_{l,k,i}(y)\Big|_{x^+_{l+1,2k+r}}^{x^-_{l+1,2k+r+1}}.$$

Since $F^n_{l,k} = F^n_{l+1,2k}$ and $F^n_{l,k+1} = F^n_{l+1,2k+2}$, see (4.2), as well as $x_{l,k} = x_{l+1,2k}$ and $x_{l,k+1} = x_{l+1,2k+2}$ we obtain by rearranging the corresponding terms in (4.7)

$$|d_{l,k,i}| \leq |d^{\overline{\mathbf{w}}}_{l,k,i}| + \tau(|I| + |II| + |III| + |IV|)$$

with

$$I := \sum_{r=1}^{2} \left(f(w(x^-_{l+1,2k+r})) - F^n_{l+1,2k+r}\right)\tilde{\psi}_{l,k,i}(x^-_{l+1,2k+r}), \tag{5.24}$$

$$II := \sum_{r=0}^{1} \left(F^n_{l+1,2k+r} - f(w(x^+_{l+1,2k+r}))\right)\tilde{\psi}_{l,k,i}(x^+_{l+1,2k+r}), \tag{5.25}$$

$$III := \int_{V_{l,k}} f(w(x))_x\,\tilde{\psi}_{l,k,i}(x)\,dx, \tag{5.26}$$

$$IV := \int_{V_{l,k}} s(w(x))\,\tilde{\psi}_{l,k,i}(x)\,dx. \tag{5.27}$$

We already noted that $d^{\overline{\mathbf{w}}}_{l,k,i} = 0$, the other terms can be estimated with the help of (E1)–(E3). First of all we consider the terms I and II. Since the numerical flux is assumed to be consistent with the flux f, i.e., $F(u,u) = f(u)$ for all $u \in \mathbb{R}$, we first obtain by the local Lipschitz-continuity of the numerical flux

$$|f(w(x^{\pm}_{l+1,2k+r^{\pm}})) - F^n_{l+1,2k+r^{\pm}}| \leq L_F |w(x^-_{l+1,2k+r^{\pm}}) - w(x^+_{l+1,2k+r^{\pm}})|$$

for $r^+ = 0, 1$ and $r^- = 1, 2$. Note that the Lipschitz constant depends on the data, i.e., $L_F = L_F(\mathbf{w})$.

Since the dual functions $\tilde{\psi}_{l,k,i}$ are piecewise polynomials and L^1-normalized (3.15) we may bound them by

$$|\tilde{\psi}_{l,k,i}(x)| \lesssim 2^l, \quad x \in \mathbb{R}. \tag{5.28}$$

Moreover, w is continuous inside the cell $V_{l,k}$. Therefore, only the differences at the cell interfaces, i.e., $x = x_{l,k}$ and $x = x_{l,k+1}$, give a non-trivial contribution. Hence we can estimate the sum $I + II$ by

$$|I| + |II| \lesssim 2^l \left(|w(x^+_{l,k}) - w(x^-_{l,k})| + |w(x^+_{l,k+1}) - w(x^-_{l,k+1})| \right).$$

Together with (E1) we finally obtain

$$|I| + |II| \lesssim 2^l \varepsilon. \tag{5.29}$$

In order to estimate the third term III we will use the vanishing moments of the wavelets. First of all, we write the derivative of the flux in a Taylor expansion

$$\frac{d}{dx} f(w(x)) = \sum_{r=0}^{p-1} \frac{(x - \hat{x})^r}{r!} \frac{d^{r+1}}{dx^{r+1}} f(w(\hat{x})) + R_{p-1}(x),$$

where the reminder is determined by

$$R_{p-1}(x) = \frac{1}{p!} (x - \hat{x})^p \frac{d^{p+1}}{dx^{p+1}} f(w(\xi)), \quad \xi \in [x, \hat{x}].$$

Then we estimate

$$\begin{aligned} |III| &= \left| \int_{V_{l,k}} (f(w(x)))_x \tilde{\psi}_{l,k,i}(x)\, dx \right| = \left| \int_{V_{l,k}} R_{p-1}(x) \tilde{\psi}_{l,k,i}(x)\, dx \right| \\ &\leq \|R_{p-1}\|_{L^\infty(V_{l,k})} \|\tilde{\psi}_{l,k,i}\|_{L^1(V_{l,k})} \lesssim \frac{(h_l/2)^p}{p!} \left\| \frac{d^{p+1}}{dx^{p+1}} f(w) \right\|_{L^\infty(V_{l,k})} \lesssim 2^l \varepsilon, \end{aligned} \tag{5.30}$$

where we choose $\hat{x} = \hat{x}_{l,k}$ Note that the dual multiwavelets $\tilde{\psi}_{l,k,i}$ have at least p vanishing moments, i.e., $M_{p,i} \geq p$ in (3.11), and are L^1-normalized according to (3.15). The last inequality is enforced by (E2).

Finally we have to estimate the fourth term. This is done as before for the third term, using the smoothness of the source function s, the vanishing moments and estimate (E3).

$$|IV| = \left| \int_{V_{l,k}} s(w(x)) \tilde{\psi}_{l,k,i}(x)\, dx \right| \lesssim \frac{(h_l/2)^p}{p!} \left\| \frac{d^p}{dx^p} s(w(\cdot)) \right\|_{L^\infty(V_{l,k})} \lesssim 2^l \varepsilon \tag{5.31}$$

Since a CFL condition is assumed to hold with respect to level L, i.e., $\tau \sim 2^{-L}$, the assertion follows by (5.29), (5.30) and (5.31). □

Using the above theorem, we now can verify:

Conclusion 2 *(Condition (E))*
Let $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$, $\tilde{\mathcal{D}} = \tilde{\mathcal{D}}(\mathbf{w})$ *be the prediction set and* $\mathbf{u} = \Pi_{L,\tilde{\mathcal{D}}(\mathbf{w})} \mathcal{A}_{\tilde{\mathcal{D}}(\mathbf{w})} \mathbf{w}$. *If the flux and source quadrature are performed exactly, then Condition (E) holds, i.e.,*

$$\|\mathcal{R}(\mathcal{E}_{L,\tilde{\mathcal{D}}(\mathbf{w})} \mathbf{u} - \mathcal{E}_L \mathbf{u})\|_q \lesssim \varepsilon.$$

Proof: We note that by Lemma 4 the estimates (E1)–(E3) hold for $\mathbf{u}$. In order to show (E), we have to investigate the difference of $\overline{\mathbf{v}} := \mathcal{E}_{L,\tilde{\mathcal{D}}(\mathbf{w})} \mathbf{u}$ and $\mathbf{v} := \mathcal{E}_L \mathbf{u}$. For this purpose we apply the multiscale transformation (3.28) and (3.29) to $\overline{\mathbf{v}}$ and $\mathbf{v}$, respectively. The resulting multiscale coefficients are denoted by $\overline{v}_{0,k,i}$, $\overline{d}_{l,k,i}$ and $v_{0,k,i}$, $d_{l,k,i}$, respectively. Since the evolution operators $\mathcal{E}_{L,\tilde{\mathcal{D}}(\mathbf{w})}$ and $\mathcal{E}_L$ act on the same data and the integrations are performed exactly, the evolution equations (4.6) and (4.7) give the same update for the coarse-scale coefficients and the significant detail coefficients, i.e., $\overline{v}_{0,k,i} = v_{0,k,i}$ and $\overline{d}_{l,k,i} = d_{l,k,i}$, $(l,k,i) \in \tilde{\mathcal{D}}(\mathbf{w})$. There is only a difference for the details not in the prediction set. According to the adaptive MR-DG scheme no evolution is performed for these coefficients, i.e., $\overline{d}_{l,k,i} = 0$, $(l,k,i) \notin \tilde{\mathcal{D}}(\mathbf{w})$, whereas these are updated for the reference DG scheme. By Theorem 3 applied to $\mathbf{u}$ we have $|d_{l,k,i}| \leq C\,\varepsilon_l$. To estimate the difference $\overline{\mathbf{v}} - \mathbf{v}$ we thus may proceed as in the proof of the approximation error, see Proposition 1. □

5.2.4 Uniform boundedness of adaptive MR-DG scheme

We now can use the local perturbation estimate of the previous section to show the uniform boundedness of the adaptive scheme. This is one of the assumptions in Theorem 1. We first choose a constant bound and then choose h and ε small enough such that the adaptive solution is bounded by that constant. In doing so we can replace the local Lipschitz constants with a uniform constant, depending only on the global constant bound of the solution.

Theorem 4 *(Uniform boundedness of adaptive MR-DG scheme) Let* $M > 0$ *be arbitrary but fixed and* $C_\infty := e^{C_s T}(\|u_0\|_{L^\infty} + M)$. *The constant* C_s *depends only on the global Lipschitz bound of the source function* s. *The ratio* $\lambda = \tau/h$ *is assumed to be constant and chosen small enough such that the CFL condition*

$$\lambda L_F \leq CFL_{max}$$

holds. Here $L_F = L_F(C_\infty)$ *denotes the Lipschitz constant of the numerical flux function for all data* $\mathbf{w}$ *with* $\|\mathcal{R}\mathbf{w}\|_\infty \leq C_\infty$. *Furthermore let the following assumptions hold:*

(B1) The error of the initial data approximation is bounded by
$\|\mathcal{R}(\overline{\mathbf{v}}^0 - \mathbf{u}^0)\|_\infty \leq C_1\,\varepsilon$,
where $\mathbf{u}^0$ *denotes the coefficients on discretization level* L *corresponding to the initial function* u_0.

(B2) The approximation error is uniformly bounded, i.e.,
$\|\mathcal{R}(\mathcal{A}_\varepsilon\,\mathbf{w} - \mathbf{w})\|_\infty \leq C_2\,\varepsilon$ *for all* $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$.

(B3) The evolution error is uniformly bounded, i.e.,
$\|\mathcal{R}(\mathcal{E}_{L,\tilde{\mathcal{D}}(\mathbf{w})}\,\Pi_{L,\tilde{\mathcal{D}}(\mathbf{w})}\,\mathcal{A}_{\tilde{\mathcal{D}}(\mathbf{w})}\,\mathbf{w} - \mathcal{E}_L\,\Pi_L\,\mathcal{A}_{\tilde{\mathcal{D}}(\mathbf{w})}\,\mathbf{w})\|_\infty \leq C_3\,\varepsilon$
holds for all $\mathbf{w} \in (\mathbb{R}^p)^{N_L}$ *with* $\|\mathcal{R}\mathbf{w}\|_\infty \leq C_\infty$.

(B4) The reference DG scheme in the mean is uniformly bounded, i.e.,
$\|\mathcal{R}\,\mathcal{E}_L\,\Pi_L\,\mathbf{w}\|_\infty \le (1+C_s\tau)\|\mathcal{R}\mathbf{w}\|_\infty + C_4\,h_L^\alpha$
holds for all $\mathbf{w}\in(\mathbb{R}^p)^{N_L}$ *with* $\|\mathcal{R}\mathbf{w}\|_\infty \le C_\infty$.

If $h_L \le \min\{1,h_1\}$ *and* $\varepsilon \le \min\{1, M/(3C_1), \tilde{C}h_L\}$ *with*

$$h_1 = \begin{cases} \left(\frac{\lambda M}{3TC_4}\right)^{\frac{1}{\alpha-1}} & \text{if } C_s = 0 \\ \left(\frac{\lambda M C_s}{3C_4}\right)^{\frac{1}{\alpha-1}} & \text{if } C_s \neq 0 \end{cases} \qquad \text{and} \qquad \tilde{C} = \begin{cases} \frac{\lambda M}{3T(C_2+C_3)} & \text{if } C_s = 0 \\ \frac{\lambda C_s M}{3(C_2+C_3)} & \text{if } C_s \neq 0 \end{cases}$$

for $\alpha > 1$, *then we have for all* $n \le T/\tau$

$$\|\mathcal{R}\overline{\mathbf{v}}^n\|_\infty \le C_\infty. \tag{5.32}$$

Remark 2 *(Remarks on the Assumptions)*

- *The constants in (B3) and (B4) depend on the local Lipschitz constant* L_F *of the numerical flux function, acting on limited data. If we consider arbitrary data* $\mathbf{w}\in(\mathbb{R}^p)^{N_L}$, *then we can estimate the limited data* $\Pi\mathbf{w}$ *by use of* (2.21) *and Lemma 2, i.e.,*

  $$\|\Pi\mathbf{w}\|_\infty \le \|\mathcal{R}\mathbf{w}\|_\infty + C_\alpha h^\alpha + \varepsilon$$

  for $\Pi = \Pi_L$ *and* $\Pi = \Pi_{\tilde{\mathcal{D}},L}$. *Since we only consider data bounded in the mean by* C_∞, *we have an uniform Lipschitz constant* $L_F = L_F(C_\infty + C_\alpha + 1)$ *since* $h_L \le 1$ *and* $\varepsilon \le 1$. *Hence we can use the uniform constants* $C_3 = C_3(C_\infty)$ *and* $C_4 = C_4(C_\infty)$. *Then we have to check the assumption* $\|\mathcal{R}\mathbf{w}\|_\infty \le C_\infty$ *in the course of the proof to make use of (B3) and (B4).*

- *We note that* ε *is bounded by* h_L, *up to a constant. That imposes no additional constraint due to the choice of* ε *in Corollary 1.*

Proof: We show (5.32) by induction. First of all, we estimate for $n=0$

$$\|\mathcal{R}\overline{\mathbf{v}}^0\|_\infty \le \|\mathcal{R}(\overline{\mathbf{v}}^0 - \mathbf{u}^0)\|_\infty + \|\mathcal{R}\mathbf{u}^0\|_\infty \le C_1\varepsilon + \|u_0\|_{L^\infty} \le \frac{M}{3} + \|u_0\|_{L^\infty} \le C_\infty, \tag{5.33}$$

where we use the bound on ε. Now we assume that $\|\mathcal{R}\overline{\mathbf{v}}^j\|_\infty \le C_\infty$ for $j = 0,\dots,n-1$.

Let be $\mathbf{w}^{n-1} := \mathcal{A}_{\tilde{\mathcal{D}}^n}\overline{\mathbf{v}}^{n-1}$ and $\tilde{\mathcal{D}}^n := \tilde{\mathcal{D}}(\overline{\mathbf{v}}^{n-1})$. Then we estimate the mean values of the adaptive MR-DG scheme by

$$\|\mathcal{R}\overline{\mathbf{v}}^n\|_\infty \le a_{n-1} + b_{n-1} + c_{n-1}$$

with

$$a_{n-1} := \|\mathcal{R}(\mathcal{A}_\varepsilon\,\mathcal{E}_{L,\tilde{\mathcal{D}}^n}\,\Pi_{L,\tilde{\mathcal{D}}^n}\,\mathbf{w}^{n-1} - \mathcal{E}_{L,\tilde{\mathcal{D}}^n}\,\Pi_{L,\tilde{\mathcal{D}}^n}\,\mathbf{w}^{n-1})\|_\infty,$$
$$b_{n-1} := \|\mathcal{R}(\mathcal{E}_{L,\tilde{\mathcal{D}}^n}\,\Pi_{L,\tilde{\mathcal{D}}^n}\,\mathbf{w}^{n-1} - \mathcal{E}_L\,\Pi_L\,\mathbf{w}^{n-1})\|_\infty, \qquad c_{n-1} := \|\mathcal{R}\,\mathcal{E}_L\,\Pi_L\,\mathbf{w}^{n-1}\|_\infty.$$

The terms of the right-hand side can be estimated by the Assumptions (B2), (B3) and (B4). Note that by (5.4) we have $\mathcal{A}_{\tilde{\mathcal{D}}^n}\overline{\mathbf{v}}^{n-1} = \overline{\mathbf{v}}^{n-1}$ and therefore

$$\|\mathcal{R}\mathcal{A}_{\tilde{\mathcal{D}}^n}\,\overline{\mathbf{v}}^{n-1}\|_\infty = \|\mathcal{R}\,\overline{\mathbf{v}}^{n-1}\|_\infty.$$

Therefore we may estimate the mean values of the adaptive MR-DG scheme by

$$\|\mathcal{R}\overline{\mathbf{v}}^n\|_\infty \leq \|\mathcal{R}\overline{\mathbf{v}}^{n-1}\|_\infty (1 + C_s\,\tau) + C_4\,h_L^\alpha + (C_2 + C_3)\,\varepsilon$$

and then infer by repeating this estimate

$$\|\mathcal{R}\overline{\mathbf{v}}^n\|_\infty \leq \|\mathcal{R}\overline{\mathbf{v}}^0\|_\infty (1 + C_s\,\tau)^n + (C_4\,h_L^\alpha + \varepsilon\,(C_2 + C_3)) \sum_{i=0}^{n-1} (1 + C_s\,\tau)^i. \tag{5.34}$$

We now use (5.33). By the bounds on h_L and ε we then further estimate (5.34), where we distinguish two cases. If $C_s = 0$, then we obtain

$$\begin{aligned}\|\mathcal{R}\overline{\mathbf{v}}^n\|_\infty &\leq \|u_0\|_{L^\infty} + \frac{M}{3} + \frac{T}{\lambda}\big((C_2 + C_3)\frac{\varepsilon}{h_L} + C_4 h_L^{\alpha-1}\big) \\ &\leq \|u_0\|_{L^\infty} + M/3 + M/3 + M/3 = C_\infty.\end{aligned}$$

If $C_s \neq 0$, we bound the geometric sum by $e^{C_s T}/(C_s\tau)$ with $\tau = \lambda h_L \leq T/n$ and then estimate

$$\begin{aligned}\|\mathcal{R}\overline{\mathbf{v}}^n\|_\infty &\leq e^{C_s T}\big(\|u_0\|_{L^\infty} + M/3 + +((C_2 + C_3)\varepsilon + C_4 h_L^\alpha)/(C_s\tau)\big) \\ &\leq e^{C_s T}(\|u_0\|_{L^\infty} + M/3 + M/3 + M/3) = C_\infty.\end{aligned}$$

□

6 Numerical results

In order to verify the analytical results we consider the inviscid Burgers' equation with flux $f(u) = 0.5\,u^2$. Numerical computations are performed for two configurations characterized by the source term and the initial conditions

C1 (homogeneous problem):
$s(u) = 0$ with $u_0(x) = 1$ in $[0, 0.5]$ and $u_0(x) = 0$ in $[0.5, 1]$,

C2 (inhomogeneous problem):
$s(u) = u(u - 0.5)(u - 1)$ with $u_0(x) = \sin(2\pi x)$ in $[0, 1]$.

For both configurations the computational domain is $\Omega = [0, 1]$, where we impose periodic boundary conditions. Thus, for configuration C1 the entropy solution exhibits a shock wave emanating at $x = 0.5$ and moving with speed 0.5. Due to the periodic boundary conditions a rarefaction wave develops at the left boundary with minimal and maximal characteristic speed $c_{min} = 0$ and $c_{max} = 1$, respectively. In Figure 5 we present the entropy solution at $T = 0.5$. For configuration C2, a shock is developing at some time $t > 0$ that is moving at negative speed. Note that at the boundaries the characteristic speeds are zero, i.e., the solution does not change there. Since there is no explicit representation of the entropy solution available, an approximation of the entropy solution is shown in Figure 6 at time $T = 0.24$.

The computational domain is discretized by $N_0 = 5$ cells on the coarsest level, i.e., $h_0 = 0.2$. Hence the resolution for higher refinement levels is $N_l = 2^l\,N_0$ and $h_l = 2^{-l}\,h_0$. For both configurations we choose $L = 9$ refinement levels. For the time discretization, we use the SSP-RK(3,3) scheme from [40]. We have to respect a CFL condition of $CFL = 0.1$ to ensure both linear L^2 and nonlinear TVD stability, therefore we choose $\tau_0 = 0.02$ (C1) and $\tau_0 = 0.005$

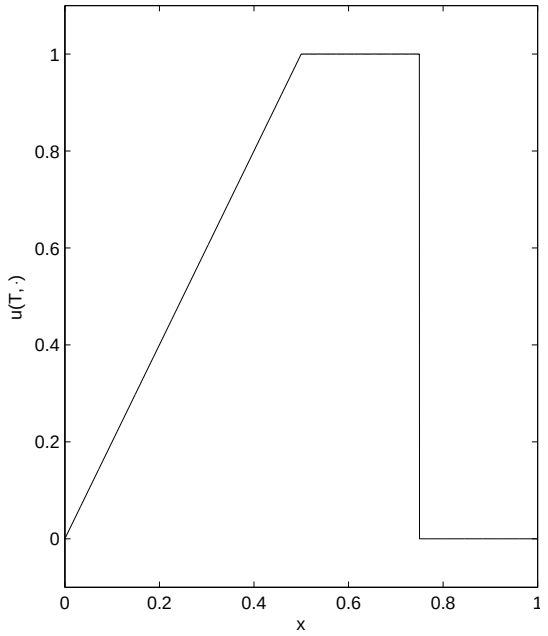

Figure 5: Entropy solution in case of C1 at $T = 0.5$.

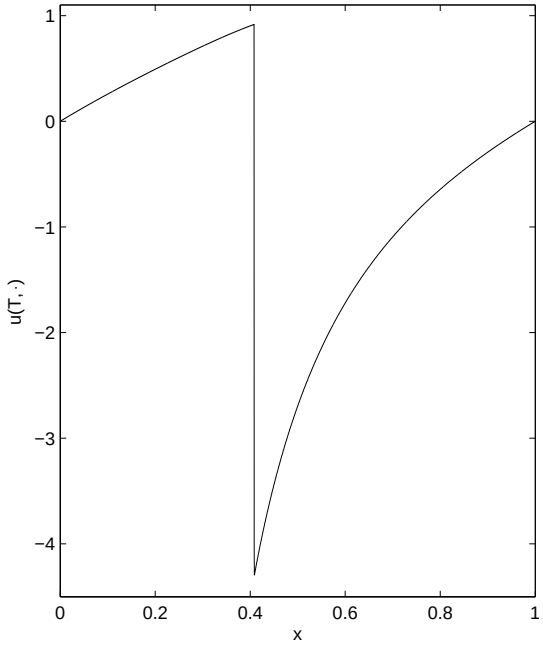

Figure 6: Approximate entropy solution in case of C2 at $T = 0.24$.

(C2). Since we perform global time stepping, the CFL condition has to hold for the smallest cells corresponding to the highest refinement level L, i.e., $\tau = 2^{-L}\tau_0 = 4 \times 10^{-5}$ (C1), and $\tau = 1 \times 10^{-5}$ (C2). The final integration time is $T = 0.5$ (C1) and $T = 0.24$ (C2). We choose cubic polynomials for the discretization space S_h^p, i.e. $p = 4$, where the multiscale analysis is performed by Alpert's multiwavelets, see Appendix A.1. For the prediction we apply Algorithm 2.

The reference DG scheme (2.7) is determined by the Engquist-Osher flux

$$F^{EO}(u, v) = \int_0^v \min(f'(s), 0)\mathrm{d}s + \int_0^u \max(f'(s), 0)\mathrm{d}s + f(0)$$

and the integrals in the quadrature terms (2.10) are computed by a 5-point Gaussian quadrature rule of order 9. Limiting is performed using the Shu limiter (2.17), where we do not incorporate the entropy bound (2.21), because we consider a convex flux, see [36].

Note that the analysis of the adaptive MR-DG scheme in Section 5 has only been performed for a forward Euler step. Since each stage of the TVD-RK method (2.23) can be written as a linear combination of forward Euler steps (2.25), we may apply Algorithm 1 to each of these steps. However, the CFL condition ensures that an information cannot move more than one cell in one time step. In order to investigate the need of these intermediate grid adaptations, we perform all computations twice, where (i) the refinement step and the coarsening step are only performed *once* before and after the TVD-RK scheme, respectively, and omit the prediction step B7, and (ii) intermediate grid adaptations, including B7, are performed before each stage of the Runge-Kutta scheme; this requires to synchronize the intermediate adaptive grids. In the figures the results are distinguished by "with RKpred" and "without".

To conclude on the efficiency and reliability of the adaptive MR-DG scheme, we investigate the computational effort (memory and CPU time) and the accuracy (discretization and

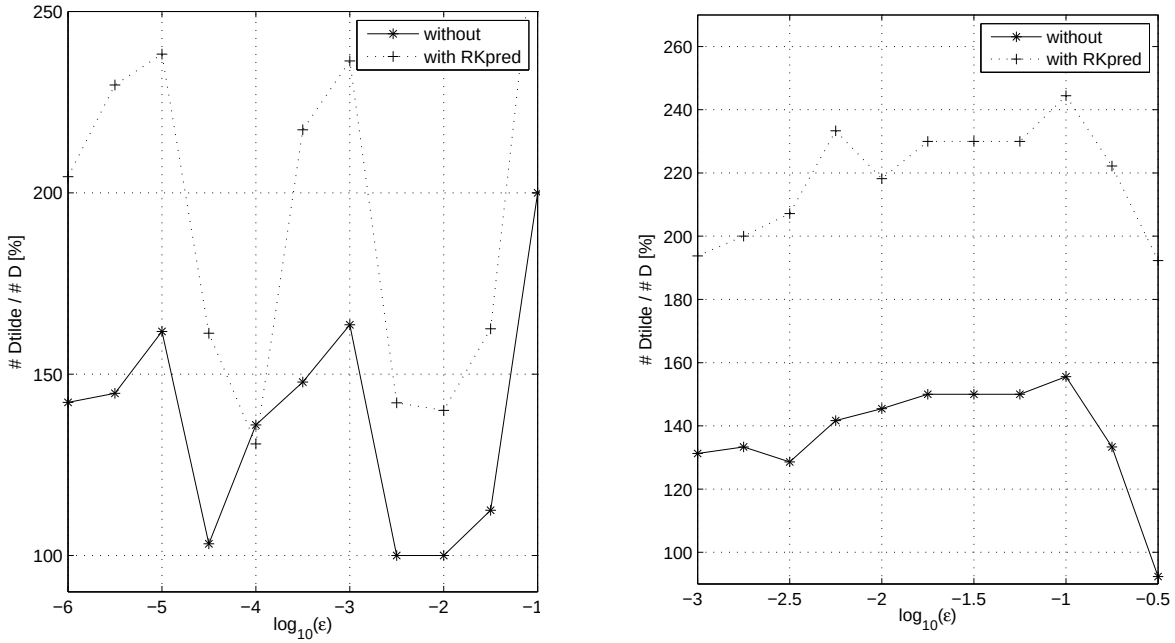

Figure 7: Variation of $\#\tilde{\mathcal{D}}/\#\mathcal{D}$ for different threshold values ε for configuration C1 (left) and C2 (right).

perturbation error) for varying threshold values ε.

Grid refinement The results of the adaptive DG scheme without intermediate Runge-Kutta prediction are presented in Figures 12 and 13 for varying threshold values $\varepsilon = 10^{-1}, \ldots, 10^{-6}$ (C1) and $\varepsilon = 10^{-1/2}, \ldots, 10^{-3}$ (C2). For each computation we plot the adaptive solution (top), the adaptive grid after hard thresholding (middle) characterized by the set $\mathcal{D}$ corresponding to the significant details and the predicted grid corresponding to the prediction set $\tilde{\mathcal{D}}$ (bottom), that was used to calculate that solution. There is hardly any difference visible between the solution of the adaptive DG scheme and the reference DG scheme performed on the uniform mesh corresponding to the highest refinement level $L = 9$. However, there are differences in the locally refined grids. These are represented by the position of the local cells in the grid hierarchy. To each cell pair $V_{l+1,2k}$, $V_{l+1,2k+1}$ four grey scales are shown for the absolute value of the multiscale coefficients $|d_{l,k,i}|$, $i = 0, \ldots, 3$, where white corresponds to 0 and black to the maximal value on level l. The magnitude of the scales is given as labels on the right axis. Typically the absolute value of these coefficients in smooth regions decreases with increasing order i, due to the $i+p$ vanishing moments, see (A.9). We note that the largest values correspond to the discontinuity. Therefore the grid is locally refined near to its position at $x = 0.75$ (C1) and $x \approx 0.41$ (C2). With smaller threshold values, the kinks at $x = 0$ and $x = 0.5$ in case of (C1) are also detected by the multiscale analysis and refinement is locally triggered near to these points as well. Away from these isolated points, the discretization is coarse because of the local smoothness of the solution. We note that the adaptive grid is inflated due to the prediction.

In Figure 7 the ratio $\#\tilde{\mathcal{D}}/\#\mathcal{D}$ of the prediction set $\tilde{\mathcal{D}}$ and the set $\mathcal{D}$ of significant details is plotted versus the threshold value. The ratio indicates a reasonable efficient prediction

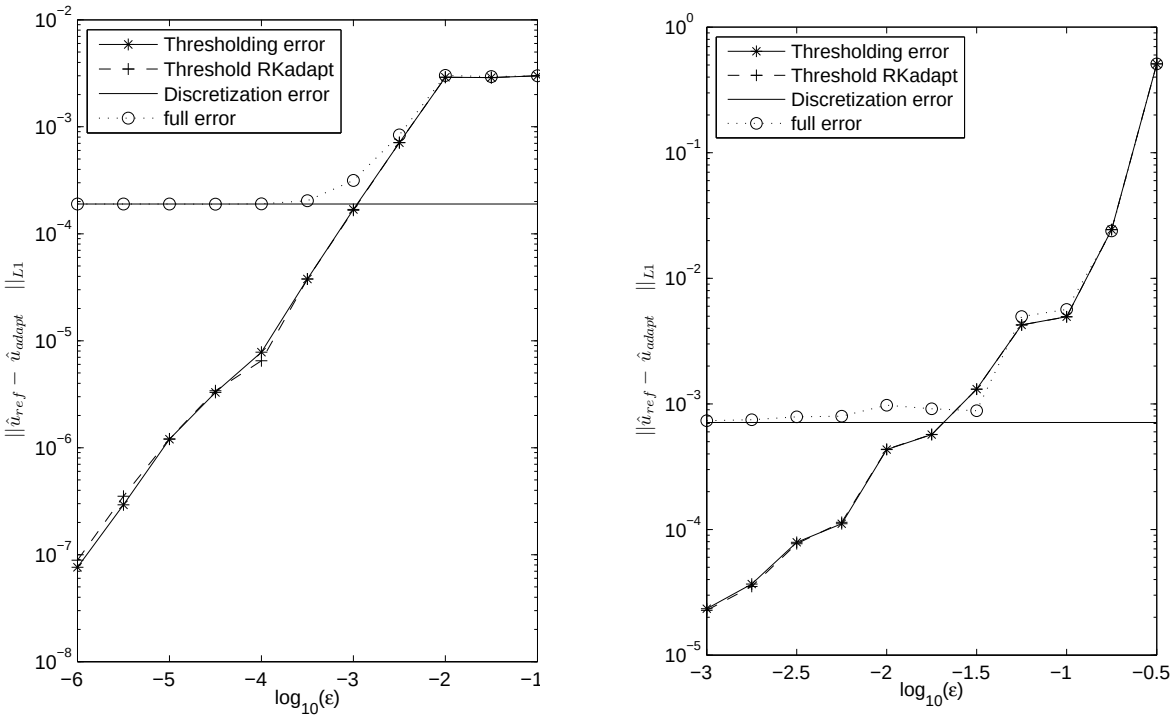

Figure 8: Error of adaptive solution with $L = 9$ and varying threshold value ε for configuration C1 (left) and C2 (right).

strategy. Additionally, the absolute numbers $\#\tilde{\mathcal{D}}$ and $\#\mathcal{D}$ are negligibly small in comparison to the uniform reference mesh. In particular, we note that the intermediate grid adaptations trigger additional grid refinement, that will result in a higher computational load but with no effect on the overall accuracy as is discussed below.

Optimal choice of threshold value The ideal strategy presented in Section 5 suggests that the optimal threshold value ε_{opt} should be chosen such that the discretization error $\hat{\eta}_L = \hat{u} - \hat{v}_L$ of the reference scheme and the perturbation error $\hat{e}_L = \hat{v}_L - \hat{v}_{L,\varepsilon}$ are balanced. For $L = 9$ we obtain $\|\hat{\eta}_L\|_{L^1([0,1])} = 1.9 \times 10^{-4}$ (C1) and $\|\hat{\eta}_L\|_{L^1([0,1])} = 7.1 \times 10^{-4}$ (C2), respectively. Note that for an inhomogeneous problem, the constant in the error estimates is typically much larger than for a homogeneous problem. This is caused by a small perturbation of order τ in each time step that is accumulating to e^{cT}, see for instance the estimates in Appendix A.3. Note that for (C2) the entropy solution is not explicitly available. Therefore we approximate the entropy solution by a computation with the reference DG scheme on a uniformly refined grid corresponding to $L = 13$ refinement levels. In order to determine ε_{opt} we compare in Figure 8 the perturbation error and the discretization error on the uniform reference mesh. For this purpose, we project the data of the adaptive grid to the reference mesh performing a local change of basis according to (3.7), where we put the non-significant detail coefficients to zero and apply the inverse multiscale transformation (3.30). Obviously, the perturbation error is decreasing with smaller threshold values. In particular, it tends to zero for $\varepsilon \to 0^+$, i.e., the adaptive solution approaches the reference solution obtained on the reference grid with L refinement levels. Of course, we do not gain in accuracy when choosing a very small threshold value,

because the discretization error is fixed by the number of refinement levels. Note that there is almost no difference visible, when performing the computation with intermediate Runge-Kutta prediction.

As we can depict from Figure 8 an optimal choice would be $\varepsilon_{opt} \sim 10^{-3}$ (C1) and $\varepsilon_{opt} \sim 10^{-1.75}$ (C2), respectively. Obviously, the error of the adaptive scheme is decreasing with decreasing threshold value ε as long as $\varepsilon > \varepsilon_{opt}$ whereas it stalls for $\varepsilon < \varepsilon_{opt}$, i.e., for $\varepsilon > \varepsilon_{opt}$ the perturbation error due to thresholding dominates whereas for $\varepsilon < \varepsilon_{opt}$ the discretization error dominates.

Efficiency versus accuracy To conclude on the efficiency of the adaptive scheme, we consider the computational costs. First we discuss the size of the adaptive grids that determine the memory requirements, see Figures 9 and 10. We observe that with decreasing threshold value the number of cells is increasing and, thus, resulting in higher CPU times. We note that the intermediate Runge-Kutta prediction triggers additional grid refinement that results in higher computational times that are almost twice as high. Since the ultimate goal is to save computational costs *and* to preserve the accuracy of the reference solution, we plot the perturbation error for varying threshold values versus the CPU time in Figure 11. However, we want to point out that in practice the optimal threshold value ε_{opt} can only be roughly estimated, i.e., we either loose accuracy if $\varepsilon \gg \varepsilon_{opt}$, see Figure 8, or the computational cost is significantly higher if $\varepsilon \ll \varepsilon_{opt}$, see Figures 9 and 10.

The choice $\varepsilon \sim h_L^{1+\beta}$ suggested in Corollary 1 is much too pessimistic: according to [11] the discretization error is of order h_L^{β} where $\beta = \min\{1, \alpha\}/2$ and α corresponds to the entropy bound (2.21) in the Shu limiter. For instance, for our configuration we compute $h_L^{1+\beta} \sim 7.7 \times 10^{-6}$, where we put $\alpha \geq 1$ and $C_\alpha = \infty$ and $h_L = 0.2 \times 2^{-9}$. However, the discretization error is approximately $\hat{\eta}_L \sim 1.9 \times 10^{-4}$ (C1) and $\hat{\eta}_L \sim 7.1 \times 10^{-4}$ (C2), respectively, that fits much better to ε_{opt}. Since the threshold error accumulates when performing all time steps, we have to compensate the accumulation process by the choice of the threshold value, i.e., we are loosing one order of magnitude. Since Corollary 1 relies on a-priori error estimates of the discretization error, it is a worst case estimate. Therefore we suggest to use $\varepsilon_{opt} \sim h_L^{\beta}$ in practice.

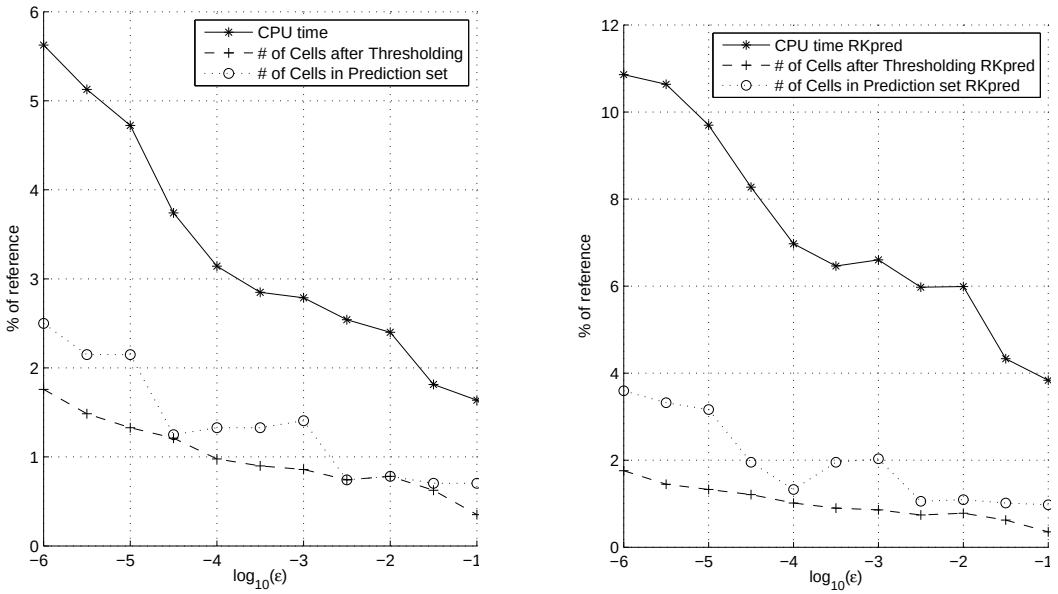

Figure 9: Number of cells and CPU time w.r.t. reference solution ($L = 9$): Adaptive computations with $L = 9$ and varying threshold value ε in case of configuration C1.

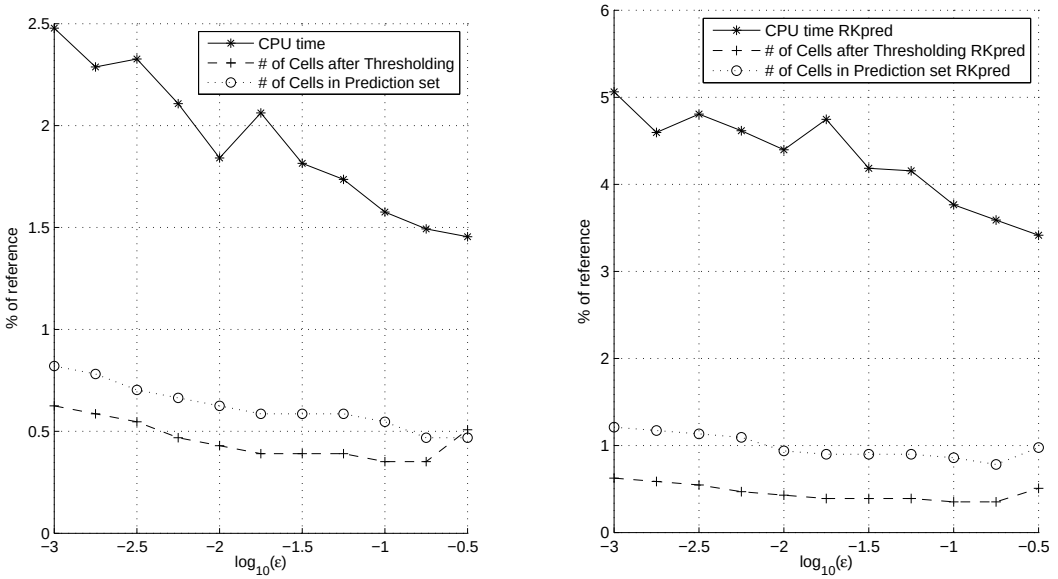

Figure 10: Number of cells and CPU time w.r.t. reference solution ($L = 9$): Adaptive computations with $L = 9$ and varying threshold value ε in case of configuration C2.

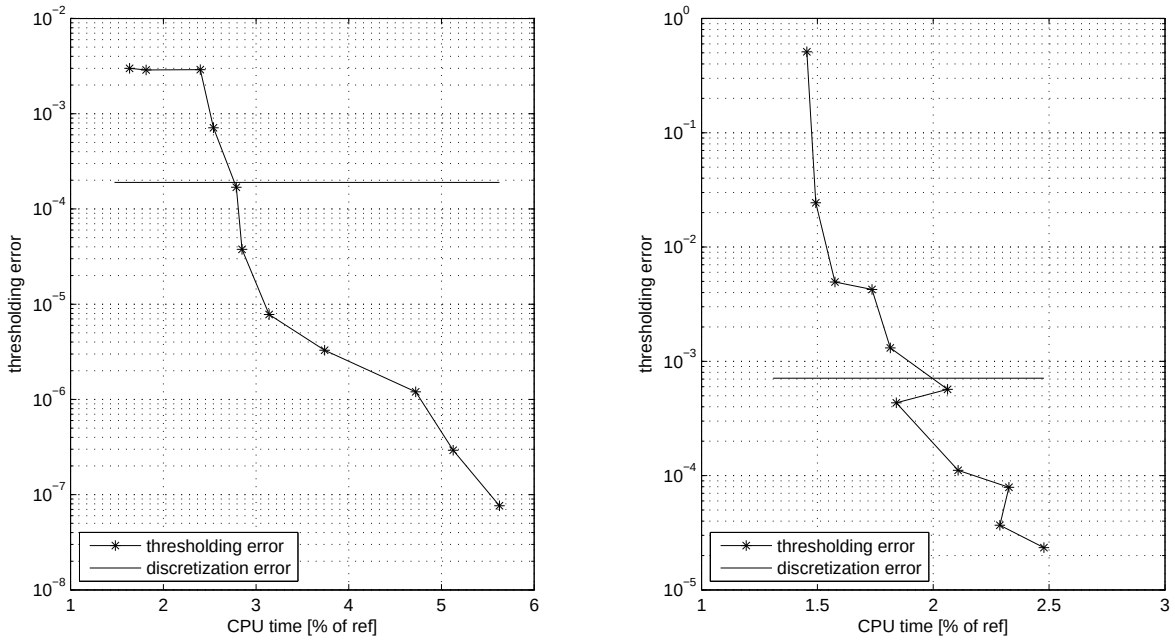

Figure 11: Comparison of CPU time and error of adaptive scheme for different threshold values for configuration C1 (left) and C2 (right) without intermediate Runge-Kutta prediction.

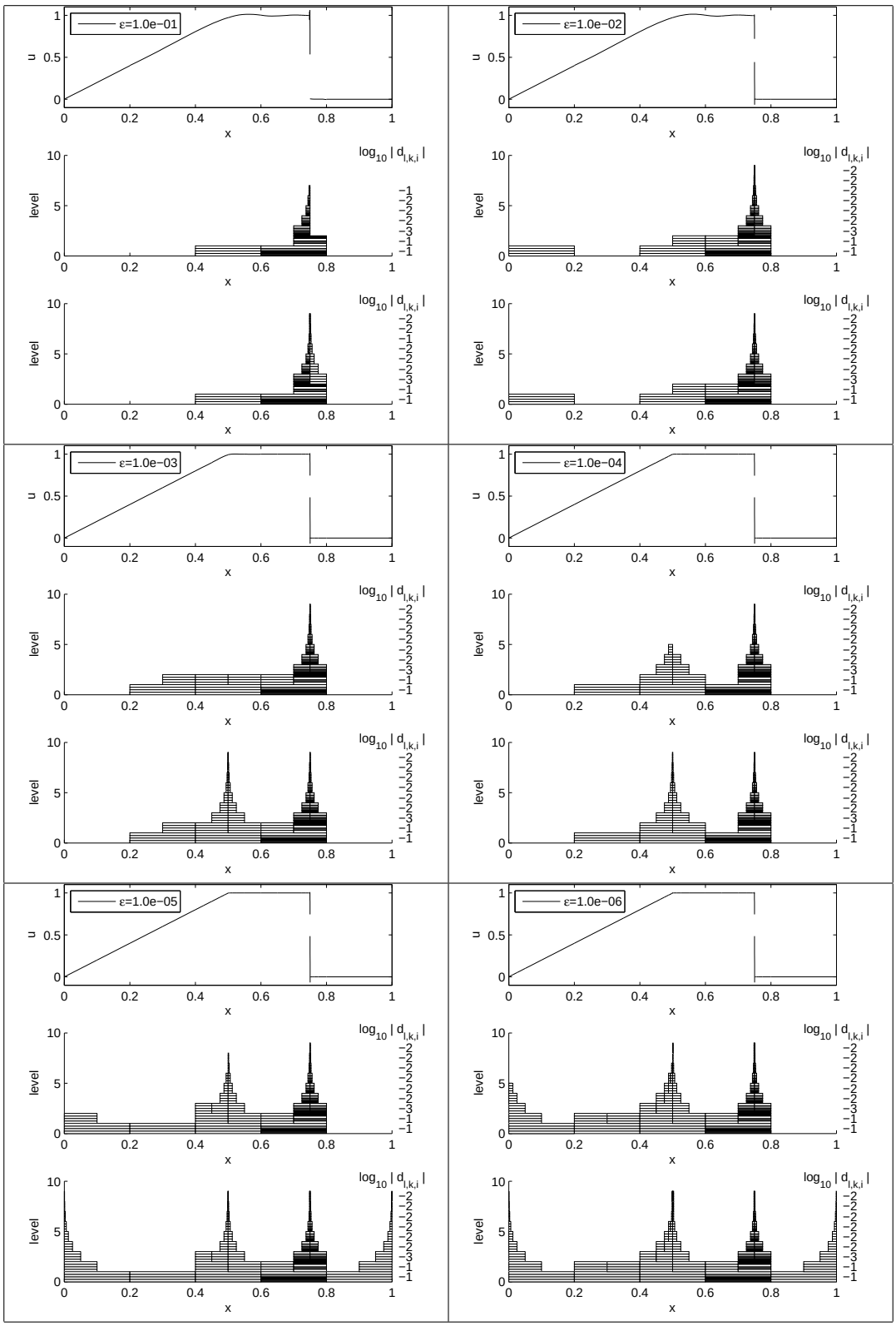

Figure 12: Results of adaptive DG scheme for configuration C1 without intermediate Runge-Kutta prediction at time $T = 0.5$ using different threshold values ε. For each value the adaptive solution (top), the adaptive grid after hard thresholding (middle) and the predicted grid corresponding to the prediction set (bottom) are shown in each frame.

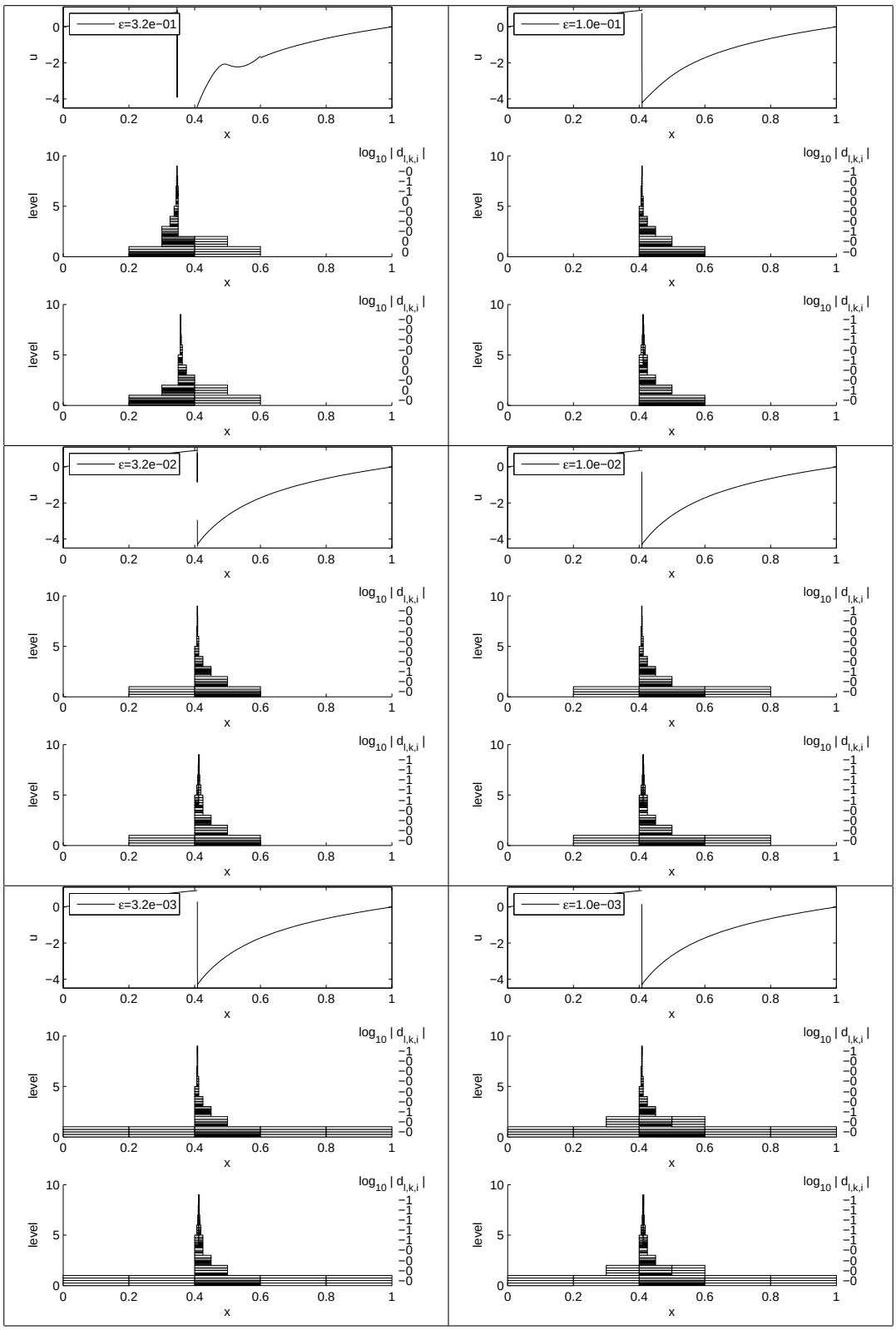

Figure 13: Results of adaptive DG scheme for configuration C2 without intermediate Runge-Kutta prediction at time $T = 0.24$ using different threshold values ε. For each value the adaptive solution (top), the adaptive grid after hard thresholding (middle) and the predicted grid corresponding to the prediction set (bottom) are shown in each frame.

References

[1] S. Adjerid, K.D. Devine, J.E. Flaherty, and L. Krivodonova. A posteriori error estimation for discontinuous Galerkin solutions of hyperbolic problems. *Comput. Methods Appl. Mech. Eng.*, 191:1097–1112, 2002.

[2] B. Alpert, G. Beylkin, D. Gines, and L. Vozovoi. Adaptive solution of partial differential equation in multiwavelet bases. *J. Comp. Phys.*, 182:149 –190, 2002.

[3] R.M. Beam and R.F. Warming. Multiresolution analysis and supercompact multiwavelets. *SISC*, 22(4):1238–1268, 2000.

[4] K.S. Bey and J.T. Oden. *hp*-version discontinuous Galerkin methods for hyperbolic conservation laws. *Comput. Methods Appl. Mech. Eng.*, 133(3-4):259–286, 1996.

[5] F. Bramkamp, Ph. Lamby, and S. Müller. An adaptive multiscale finite volume solver for unsteady an steady state flow computations. *J. Comp. Phys.*, 197(2):460 –490, 2004.

[6] J.M. Carnicer, W. Dahmen, and J.M. Peña. Local decomposition of refinable spaces and wavelets. *Appl. Comput. Harmon. Anal.*, 3:127–153, 1996.

[7] B. Cockburn. Quasimonotone schemes for scalar conservation laws: Part I. *SIAM J. Numer. Anal.*, 26(6):1325–1341, 1989.

[8] B. Cockburn, S. Hou, and C.-W. Shu. The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. IV: The multidimensional case. *Math. Comp.*, 54(190):545–581, 1990.

[9] B. Cockburn, G.E. Karniadakis, and C.-W. Shu. The development of discontinuous Galerkin methods. Cockburn, Bernardo (ed.) et al., Discontinuous Galerkin methods. Theory, computation and applications. 1st international symposium on DGM, Newport, RI, USA, May 24-26, 1999. Berlin: Springer. Lect. Notes Comput. Sci. Eng. 11, 3-50 (2000)., 2000.

[10] B. Cockburn, S.-Y. Lin, and C.-W. Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. III: One-dimensional systems. *J. Comput. Phys.*, 84(1):90–113, 1989.

[11] B. Cockburn and C.-W. Shu. TVB Runge-Kutta Local Projection Discontinuous Galerkin Finite Element Method for conservation laws II: General framework. *Math. Comp.*, 52(186):411–435, 1989.

[12] B. Cockburn and C.-W. Shu. The Runge-Kutta local projection P^1-discontinuous-Galerkin finite element method for scalar conservation laws. *RAIRO Modélisation Math. Anal. Numér.*, 25(3):337–361, 1991.

[13] B. Cockburn and C.-W. Shu. The Runge-Kutta discontinuous Galerkin method for conservation laws. V: Multidimensional systems. *J. Comput. Phys.*, 141(2):199–224, 1998.

[14] B. Cockburn and C.-W. Shu. Runge-Kutta Discontinuous Galerkin Methods for Convection-Dominated Problems. *J. Sci. Computing*, 16(3):173–261, 2001.

[15] A. Cohen, I. Daubechies, and J. Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45:485–560, 1992.

[16] A. Cohen, S.M. Kaber, S. Müller, and M. Postel. Fully Adaptive Multiresolution Finite Volume Schemes for Conservation Laws. *Math. Comp.*, 72(241):183–225, 2003.

[17] F. Coquel, Y. Maday, S. Müller, M. Postel, and Q.H. Tran. New trends in multiresolution and adaptive methods for convection-dominated problems. *ESAIM, Proc.*, 29:1–7, 2009. `http://www.esaim-proc.org/index.php?option=article&access=standard&Itemid=129&url=/articles/proc/abs/2009/04/proc092900f/proc092900f.html`.

[18] W. Dahmen, B. Gottschlich–Müller, and S. Müller. Multiresolution schemes for conservation laws. *Numer. Math.*, 88(3):399–443, 2000.

[19] A. Dedner, C. Makridakis, and M. Ohlberger. Error control for a class of runge kutta discontinuous Galerkin methods for nonlinear conservation laws. *SIAM J. Numer. Anal.*, 45:514–538, 2007.

[20] E. Godlewski and P.-A. Raviart. *Hyperbolic systems of conservation laws*. Mathématiques et Applications (Paris). 3-4. Paris: Ellipses, 1991.

[21] S. Gottlieb, C.-W. Shu, and E. Tadmor. Strong Stability Preserving High-Order Time Discretization Methods. *SIAM Review*, 43(1):89–112, 2001.

[22] A. Haar. Zur Theorie der orthogonalen Funktionen-Systeme. *Math. Ann.*, 69:331 –371, 1910.

[23] A. Harten. Multiresolution representation of data: A general framework. *SIAM J. Numer. Anal.*, 33(3):1205–1256, 1996.

[24] R. Hartmann and P. Houston. Adaptive discontinuous Galerkin finite element methods for nonlinear hyperbolic conservation laws. *SISC*, 24:979–1004, 2002.

[25] R. Hartmann and P. Houston. Adaptive discontinuous Galerkin finite element methods for the compressible Euler equations. *J. Comp. Phys.*, 183:508–532, 2002.

[26] P. Houston, B. Senior, and E. Süli. *hp*-discontinuous Galerkin finite element methods for hyperbolic problems: Error analysis and adaptivity. *Int. J. Numer. Methods Fluids*, 40(1-2):153–169, 2002.

[27] N. Hovhannisyan and S. Müller. On the stability of fully adaptive multiscale schemes for conservation laws using approximate flux and source reconstruction strategies. *IMA Journal of Numerical Analysis*, 2009. DOI 10.1093/imanum/drp010.

[28] F. Keinert. *Wavelets and Multiwavelets*. Studies in Advanced Mathematics. Chapman & Hall/CRC, 2003.

[29] D. Ketcheson, C.B. Macdonald, and S. Gottlieb. Optimal implicit strong stability preserving Runge-Kutta methods. *Applied Numerical Mathematics*, 59:373–392, 2009.

[30] D. Kröner. *Numerical Schemes for Conservation Laws*. Advances in Numerical Mathematics. Wiley–Teubner, 1997.

[31] S.N. Kruzhkov. First order quasilinear equations in several independent variables. *Math. USSR Sbornik*, 10:217–243, 1970.

[32] R.J. LeVeque. *Finite volume methods for hyperbolic problems*. Cambridge Texts in Applied Mathematics. Cambridge: Cambridge University Press., 2002.

[33] S. Müller. *Adaptive Multiscale Schemes for Conservation Laws*, volume 27 of *Lecture Notes on Computational Science and Engineering*. Springer, 2003.

[34] S. Müller. Multiresolution schemes for conservation laws. DeVore, Ronald (ed.) et al., Multiscale, nonlinear and adaptive approximation. Dedicated to Wolfgang Dahmen on the occasion of his 60th birthday. Berlin: Springer. 379-4080, 2009.

[35] O.A. Oleinik. Discontinuous solutions of nonlinear differential equations. *Amer. Math. Soc. Transl. Ser. 2*, 26:95–172, 1963.

[36] S. Osher. Convergence of generalized MUSCL schemes. *SIAM J. Numer. Anal.*, 22(5):947–961, 1985.

[37] W.H. Reed and T.R. Hill. Triangular mesh methods for neutron transport equation. Technical Report LA-UR-73-479, Los Alamos Scientific Laboratory, Los Alamos, NA, 1973.

[38] J.-F. Remacle, J.E. Flaherty, and M.S. Shephard. An adaptive discontinuous Galerkin technique with an orthogonal basis applied to compressible flow problems. *SIAM Rev.*, 45(1):53–72, 2003.

[39] R. Sanders. On convergence of monotone finite-difference schemes with variable spatial differencing. *Math. Comp.*, 40:91–106, 1983.

[40] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing scheme. *J. Comp. Physics*, 77:439–471, 1988.

[41] V. Strela. *Multiwavelets: Theory and Applications*. PhD thesis, Massachusetts Institute of Technology, 1996.

[42] T. Tang and Z.-H. Teng. Error bounds for fractional step methods for conservation laws with source terms. *SIAM J. Numer. Anal.*, 32(1):110–127, 1995.

[43] L. Wang and D.J. Mavriplis. Adjoint-based *h-p* adaptive discontinuous Galerkin methods for the 2d compressible Euler equations. *J. Comp. Phys.*, 228(20):7643–7661, 2009.

[44] R.F. Warming and R.M. Beam. Discrete multiresolution analysis using Hermite interpolation biorthogonal multiwavelets. *SISC*, 22(4):1269–1317, 2000.

A Appendix

A.1 Example: Alpert's orthogonal multiwavelet basis

An important example that fits into the general framework of Section 3 is Alpert's orthogonal multiwavelet basis, cf. [2]. The main ideas are briefly summarized. In order to construct an appropriate MRA on $L_2([a,b])$ we define the scaling functions

$$\Phi_i(x) = \begin{cases} \sqrt{(2i+1)/2}\, P_i(x), & x \in [-1,1], \\ 0, & \text{otherwise}, \end{cases} \tag{A.1}$$

by means of the Legendre polynomials $P_i \in \Pi_i$, $i \in \mathcal{P}$, on the interval $[-1,1]$. These form an orthonormal basis for the space $S_0^p([-1,1])$. In Table 1, the Legendre polynomials are explictly given up to degree 5 and plotted in Figure 14.

$p_0(x)$	$=$	1
$p_1(x)$	$=$	x
$p_2(x)$	$=$	$\frac{1}{2}(3x^2-1)$
$p_3(x)$	$=$	$\frac{1}{2}(5x^3-3x)$
$p_4(x)$	$=$	$\frac{1}{8}(35x^4-30x^2+3)$
$p_5(x)$	$=$	$\frac{1}{8}(63x^5-70x^3+15x)$

Table 1: Legendre polynomials on $[-1,1]$.

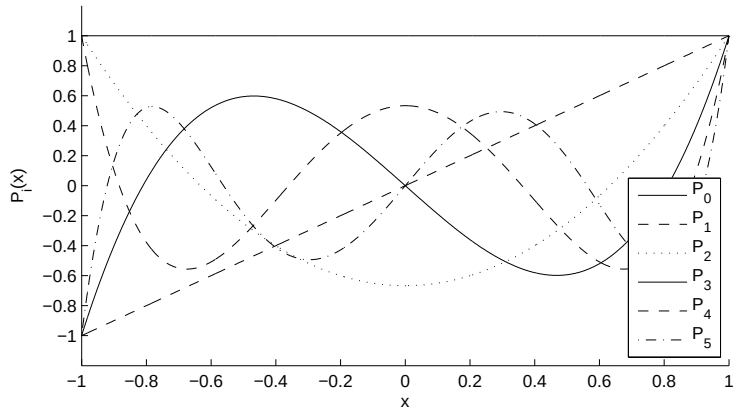

Figure 14: Graphs of Legendre Polynomials for $i = 0, \ldots, 5$

Alpert et al. [2] constructed a system of orthonormal multiwavelets $\{f_{p,i} : [-1,1] \to \mathbb{R}\}_{i \in \mathcal{P}}$, that satisfy the following properties:

1. The restriction of f_i to the interval $(0,1)$ is a polynomial of degree $p-1$.
2. The function f_i is extended to the interval $(-1,0)$ as an even or odd function according to the parity of $i+p$.

3. The functions $f_0, ..., f_{p-1}$ are orthogonal, i.e., $\langle f_i, f_j \rangle_{[-1,1]} = \delta_{ij}$, $i, j = 0, ..., p-1$.

4. The function f_i has $i+p$ vanishing moments, i.e., $\langle f_i, P \rangle_{[-1,1]} = 0$, $P \in \Pi_{i+p-1}$

Note that these functions are discontinuous at $x = 0$. Up to degree 5 they are explicitly given in Table 2 and plotted in Figure 15.

$p = 1$		
$f_0(x)$	$=$	$\sqrt{\frac{1}{2}}$
$p = 2$		
$f_0(x)$	$=$	$\sqrt{\frac{3}{2}}(-1 + 2x)$
$f_1(x)$	$=$	$\sqrt{\frac{1}{2}}(-2 + 3x)$
$p = 3$		
$f_0(x)$	$=$	$\frac{1}{3}\sqrt{\frac{1}{2}}(1 - 24x + 30x^2)$
$f_1(x)$	$=$	$\frac{1}{2}\sqrt{\frac{3}{2}}(3 - 16x + 15x^2)$
$f_2(x)$	$=$	$\frac{1}{3}\sqrt{\frac{5}{2}}(4 - 15x + 12x^2)$
$p = 4$		
$f_0(x)$	$=$	$\sqrt{\frac{15}{34}}(1 + 4x - 30x^2 + 28x^3)$
$f_1(x)$	$=$	$\sqrt{\frac{1}{42}}(-4 + 105x - 300x^2 + 210x^3)$
$f_2(x)$	$=$	$\frac{1}{2}\sqrt{\frac{35}{34}}(-5 + 48x - 105x^2 + 64x^3)$
$f_3(x)$	$=$	$\frac{1}{2}\sqrt{\frac{5}{42}}(-16 + 105x - 192x^2 + 105x^3)$
$p = 5$		
$f_0(x)$	$=$	$\sqrt{\frac{1}{186}}(1 + 30x + 210x^2 - 840x^3 + 630x^4)$
$f_1(x)$	$=$	$\frac{1}{2}\sqrt{\frac{1}{38}}(-5 - 144x + 1155x^2 - 2240x^3 + 1260x^4)$
$f_2(x)$	$=$	$\sqrt{\frac{35}{14694}}(22 - 735x + 3504x^2 - 5460x^3 + 2700x^4)$
$f_3(x)$	$=$	$\frac{1}{8}\sqrt{\frac{21}{38}}(35 - 512x + 1890x^2 - 2560x^3 + 1155x^4)$
$f_4(x)$	$=$	$\frac{1}{2}\sqrt{\frac{7}{158}}(32 - 315x + 960x^2 - 1155x^3 + 480x^4)$

Table 2: Alpert's multiwavelet construction: Expressions for the orthonormal, vanishing-moment functions $f_0, ..., f_{p-1}$, for various p and $x \in (0, 1)$. The function f_i is extended to the interval $(-1, 0)$ as an odd or even function, according to the formula $f_i(x) = (-1)^{i+p} f_i(-x)$ for $x \in (-1, 0)$ and is zero outside the interval $(-1, 1)$.

Then the system of functions defined by

$$\Psi_i(x) = \begin{cases} f_{p,i}(x), & x \in [-1, 1], \\ 0, & \text{otherwise}, \end{cases} \tag{A.2}$$

form an orthonormal basis for the complement space $W_0^p([-1, 1])$. Note that in the case of $p = 1$ this coincides with the well-known Haar basis [22]. Finally, by the affine transformation

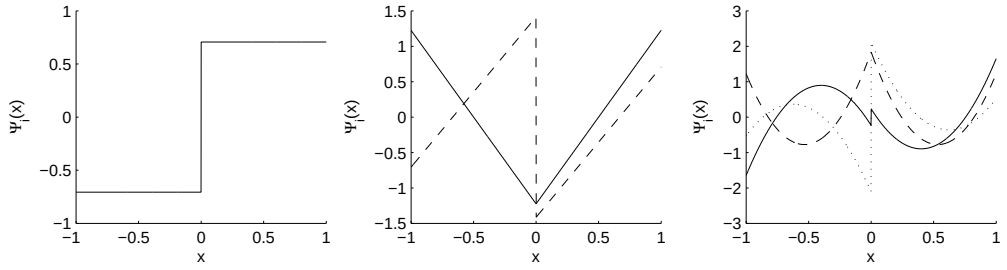

Figure 15: Multiwavelet system of order p: Left ($p = 1$): Haar wavelet; Middle ($p = 2$): "—" - ψ_0, "- - -" - ψ_1; Right ($p = 3$): "—" - ψ_0, "- - -" - ψ_1, "$\cdots$" - ψ_2.

$x \to 2\,(x - \hat{x}_{l,k})/h_l$ with the cell center $\hat{x}_{l,k} := (x_{l,k} + x_{l,k+1})/2$ we obtain a MRA on $L_2([a,b])$ where the basis functions are determined by

$$\bar{\varphi}_{l,k,i}(x) := \sqrt{\frac{2}{h_l}}\Phi_i\left(2\,\frac{x-\hat{x}_{l,k}}{h_l}\right), \quad \bar{\psi}_{l,k,i}(x) := \sqrt{\frac{2}{h_l}}\Psi_i\left(2\,\frac{x-\hat{x}_{l,k}}{h_l}\right). \tag{A.3}$$

Then we conclude by the orthogonality relation of the Legendre polynomials and Alpert's multiwavelets, respectively, the following properties of the MRA on $L^2([a,b])$:

1. the basis functions are locally supported, i.e.,

$$\text{supp } \bar{\varphi}_{l,k,i} = \text{supp } \bar{\psi}_{l,k,i} = V_{l,k}, \tag{A.4}$$

2. the basis function are normalized with respect to $L^2([a,b])$, i.e.,

$$\|\bar{\varphi}_{l,k,i}\|_{L^2([a,b])} = \|\bar{\psi}_{l,k,i}\|_{L^2([a,b])} = 1, \tag{A.5}$$

3. the basis functions satisfy the following orthogonality relations, i.e.,

$$\begin{aligned}
\langle \bar{\varphi}_{l,k,i}, \bar{\varphi}_{l,k',i'}\rangle_{[a,b]} &= \delta_{i,i'}\,\delta_{k,k'} \quad (\{\bar{\varphi}_{l,k,i}\}_{k,i} \text{ orthon. basis of } S_l^p), && \text{(A.6)}\\
\langle \bar{\varphi}_{l,k,i}, \bar{\psi}_{l,k',i'}\rangle_{[a,b]} &= 0 \quad (W_l^p \perp S_l^p), && \text{(A.7)}\\
\langle \bar{\psi}_{l,k,i}, \bar{\psi}_{l',k',i'}\rangle_{[a,b]} &= \delta_{i,i'}\,\delta_{k,k'}\,\delta_{l,l'} \quad (W_l^p \perp W_{l'}^p), && \text{(A.8)}
\end{aligned}$$

4. the multiwavelet functions have $M_{p,i} := i + p$ vanishing moments, i.e.,

$$\langle P, \bar{\psi}_{l,k,i}\rangle_{[a,b]} = 0, \quad \forall P \in \Pi_{M_{p,i}-1}. \tag{A.9}$$

With regard to (3.14) and (3.15) the L^2-normalized basis functions have to be appropriately scaled. For this purpose we normalize the bases with respect to $L^q([a,b])$ and $L^{\tilde{q}}([a,b])$ by the scaling factors $C_q^l := h_l^{1/2-1/q}$ and $C_{\tilde{q}}^l := h_l^{1/2-1/\tilde{q}} = 1/C_q^l$, respectively, i.e.,

$$\varphi_{l,k,i} := C_q^l\,\bar{\varphi}_{l,k,i}, \quad \psi_{l,k,i} := C_q^l\,\bar{\psi}_{l,k,i}, \tag{A.10}$$

$$\tilde{\varphi}_{l,k,i} := C_{\tilde{q}}^l\bar{\varphi}_{l,k,i}, \quad \tilde{\psi}_{l,k,i} := C_{\tilde{q}}^l\bar{\psi}_{l,k,i}, \tag{A.11}$$

where we choose $q = \infty$ and $\tilde{q} = 1$. Obviously, these functions inherit the above properties from their L^2-counterparts, where the functions are uniformly bounded by $\|\Phi_i\|_{L_q([a,b])}$, $\|\Psi_i\|_{L_q([a,b])}$

and $\|\Phi_i\|_{L_{\tilde{q}}([a,b])}$, $\|\Psi_i\|_{L_{\tilde{q}}([a,b])}$, respectively, instead of 1. Note, in particular, that condition (3.16) holds.

The vanishing moments (A.9) and the normalization (A.11) imply that the details decay at a rate of at least $2^{-l\,M}$ when the underlying function is locally smooth, i.e.,

$$|\langle u, \tilde{\psi}_{l,k,i}\rangle_{[a,b]}| \leq \inf_{P\in\Pi_{M-1}} |\langle f-P, \tilde{\psi}_{l,k,i}\rangle_{[a,b]}| \lesssim 2^{-l\,M}\|f\|_{H^M(V_{l,k})}. \tag{A.12}$$

Here, the number of vanishing moments is at least $M = M_{p,i} \geq p$.

Furthermore, the orthogonality relations (A.6), (A.7) and (A.8) together with the scaling (A.10) and (A.11) ensure the existence of the two-scale decomposition (3.23) and (3.24) and its inverse (3.25). The corresponding mask coefficients are determined as

$$\begin{aligned}
\tilde{m}^{0,s}_{i,j} &= \langle \tilde{\varphi}_{l,k,i}, \varphi_{l+1,2k+s,j}\rangle_{[a,b])} = \bar{h}^s_{ij}/\sqrt{2}, && \text{(A.13)}\\
\tilde{m}^{1,s}_{i,j} &= \langle \tilde{\psi}_{l,k,i}, \varphi_{l+1,2k+s,j}\rangle_{[a,b])} = \bar{g}^s_{ij}/\sqrt{2}, && \text{(A.14)}\\
\tilde{g}^{0,s}_{i,j} &= \langle \varphi_{l+1,2k+s,i}, \tilde{\varphi}_{l,k,j}\rangle_{[a,b])} = \bar{h}^s_{j,i}/\sqrt{2}, && \text{(A.15)}\\
\tilde{g}^{1,s}_{i,j} &= \langle \varphi_{l+1,2k+s,i}, \tilde{\psi}_{l,k,j}\rangle_{[a,b])} = \bar{g}^s_{j,i}/\sqrt{2} && \text{(A.16)}
\end{aligned}$$

for $s = \{0,1\}$. Note that by construction the inner products simplify to

$$\begin{aligned}
\bar{h}^s_{ij} &= \langle \bar{\varphi}_{l,k,i}, \bar{\varphi}_{l+1,2k+s,j}\rangle = \sqrt{2}\langle \varphi_i(\cdot), \varphi_j(2\cdot+2s-1)\rangle, && \text{(A.17)}\\
\bar{g}^s_{ij} &= \langle \bar{\psi}_{l,k,i}, \bar{\varphi}_{l+1,2k+s,j}\rangle = \sqrt{2}\langle \psi_i(\cdot), \psi_j(2\cdot+2s-1)\rangle, && \text{(A.18)}
\end{aligned}$$

where we use the supports of the scaling functions and multiwavelets determined by (A.4) and their shifts and translates (A.3). Since for any orthogonal polynomial sequence $\{P_i\}_{i\geq 0}$ with $P_i \in \Pi_i$, any polynomial of degree i can be expanded in terms of $P_0, ..., P_i$, the matrices $\{\bar{h}^s_{ij}\}_{i,j=0,1}$, $s = 0, 1$, are lower triangular.

In particular, for the Legendre scaling functions and Alpert's multiwavelets we obtain

$$\bar{h}^1_{ij} = (-1)^{i+j}\bar{h}^0_{ij}, \quad \bar{g}^1_{ij} = (-1)^{i+j+p}\bar{g}^0_{ij} \tag{A.19}$$

using their symmetry properties.

A.2 Two-scale relations of evolution equations for single-scale and multiscale coefficients

Here we will discuss in detail how to derive the two-scale evolution equations (4.1) and (4.7) for the single-scale coefficients and the multiscale coefficients, respectively. First of all, we recursively derive the evolution equations for the single-scale coefficients. For this purpose we assume that on level $l+1$ the evolution equations for any $r \in I_{l+1}$ can be written as

$$v^{n+1}_{l+1,r,i} = v^n_{l+1,r,i} - \tau\left(B^n_{l+1,r,i} - G^n_{l+1,r,i} - S^n_{l+1,r,i}\right), \tag{A.20}$$

where the numerical flux, the numerical flux balance, the flux quadrature and the source quadrature are determined by (4.2), (4.3), (4.4) and (4.5) with $l+1$ instead of l. For $l = L-1$ these equations obviously coincide with the reference scheme (2.7) and (2.8), (2.9), (2.10). In order to determine the recursive formulae for the numerical flux $F^n_{l,k}$, the numerical flux balance $B^n_{l,k,i}$,

the flux quadrature $G^n_{l,k,i}$ and the source quadrature $S^n_{l,k,i}$, we apply the two-scale transformation (3.28) to the evolution equation (A.20) and define the single-scale coefficients on level l as

$$v^{\mu}_{l,k,i} := \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{0,s}_{ij} v^{\mu}_{l+1,2k+s,j}, \quad \mu = n, n+1, \tag{A.21}$$

where $k \in I_l$. Then we obtain

$$v^{n+1}_{l,k,i} = \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{0,s}_{ij} v^{n+1}_{l+1,2k+s,j} = T_v - \tau\left(T_B - T_G - T_S\right) \tag{A.22}$$

with

$$\begin{aligned} T_v &:= \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{0,s}_{ij} v^{n}_{l+1,2k+s,j},\ T_B := \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{0,s}_{ij} B^{n}_{l+1,2k+s,j}, \\ T_G &:= \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{0,s}_{ij} G^{n}_{l+1,2k+s,j},\ T_S := \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{0,s}_{ij} S^{n}_{l+1,2k+1,j}. \end{aligned}$$

From the two-scale relation (A.21) we directly conclude

$$T_v = v^n_{l,k,i}. \tag{A.23}$$

Since (4.3) holds by assumption for $l+1$, the second term can be split into two parts, i.e., $T_B = T_{B1} + T_{B2}$ with

$$\begin{aligned} T_{B1} &:= \sum_{j\in\mathcal{P}} \left(\tilde{m}^{0,1}_{ij} F^n_{l+1,2k+2} \tilde{\varphi}_{l+1,2k+1,j}(x^-_{l+1,2k+2}) - \tilde{m}^{0,0}_{ij} F^n_{l+1,2k} \tilde{\varphi}_{l+1,2k,j}(x^+_{l+1,2k}) \right), \\ T_{B2} &:= F^n_{l+1,2k+1} \sum_{j\in\mathcal{P}} \left(\tilde{m}^{0,0}_{ij} \tilde{\varphi}_{l+1,2k,j}(x^-_{l+1,2k+1}) - \tilde{m}^{0,1}_{ij} \tilde{\varphi}_{l+1,2k+1,j}(x^+_{l+1,2k+1}) \right). \end{aligned}$$

The nestedness (3.1) of the grid hierarchy implies that the following grid points coincide: $x^+_{l,k} = x^+_{l+1,2k}$, $x^{\pm}_{l,k+\frac{1}{2}} = x^{\pm}_{l+1,2k+1}$ and $x^-_{l,k+1} = x^-_{l+1,2k+2}$. From the two-scale relation (3.23) we then conclude

$$\tilde{\varphi}_{l,k,i}(x^+_{l,k}) = \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{0,s}_{ij} \tilde{\varphi}_{l+1,2k+s,j}(x^+_{l+1,2k}), \tag{A.24}$$

$$\tilde{\varphi}_{l,k,i}(x^{\pm}_{l,k+\frac{1}{2}}) = \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{0,s}_{ij} \tilde{\varphi}_{l+1,2k+s,j}(x^{\pm}_{l+1,2k+1}), \tag{A.25}$$

$$\tilde{\varphi}_{l,k,i}(x^-_{l,k+1}) = \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{0,s}_{ij} \tilde{\varphi}_{l+1,2k+s,j}(x^-_{l+1,2k+2}). \tag{A.26}$$

Furthermore, we obtain by the supports (3.10) of the basis functions

$$\tilde{\varphi}_{l+1,2k,j}(x^{\pm}_{l+1,2k+2}) = 0,\ \tilde{\varphi}_{l+1,2k+1,j}(x^-_{l+1,2k+1}) = 0, \tag{A.27}$$

$$\tilde{\varphi}_{l+1,2k+1,j}(x^{\pm}_{l+1,2k}) = 0,\ \tilde{\varphi}_{l+1,2k,j}(x^+_{l+1,2k+1}) = 0. \tag{A.28}$$

Inflating the term T_{B1} by the zeros in (A.27) and using the identities (A.24) and (A.26) we end up with

$$\begin{aligned} T_{B1} &= F^n_{l+1,2k+2} \sum_{j\in\mathcal{P}} \tilde{m}^{0,1}_{ij} \tilde{\varphi}_{l+1,2k+1,j}(x^-_{l+1,2k+2}) - F^n_{l+1,2k} \sum_{j\in\mathcal{P}} \tilde{m}^{0,0}_{ij} \tilde{\varphi}_{l+1,2k,j}(x^+_{l+1,2k}) \\ &= F^n_{l+1,2k+2} \tilde{\varphi}_{l,k,i}(x^-_{l,k+1}) - F^n_{l+1,2k} \tilde{\varphi}_{l,k,i}(x^+_{l,k}). \end{aligned} \tag{A.29}$$

Similarly we inflate the term T_{B2} by the zeros in (A.28) and employ the identity (A.25). Then the term

$$T_{B2} = F^n_{l+1,2k+1}\left(\tilde{\varphi}_{l,k,i}(x^-_{l+1,2k+1}) - \tilde{\varphi}_{l,k,i}(x^+_{l+1,2k+1})\right) = 0 \tag{A.30}$$

vanishes, because $\tilde{\varphi}_{l,k,i}$ is continuous inside its support $V_{l,k}$ and $x_{l+1,2k+1} = x_{l,k+\frac{1}{2}} \in V_{l,k}$.

For the term T_G we proceed in a similar manner. By assumption that (2.10) holds for level $l+1$, we may rewrite this term as

$$T_G = \sum_{j\in\mathcal{P}} \sum_{r\in\{0,1\}} \tilde{m}^{0,r}_{ij} \int_{V_{l+1,2k+r}} f(v_L(t^n,x)) \cdot \tilde{\varphi}'_{l+1,2k+r,j}(x)\,dx. \tag{A.31}$$

Since the two-scale relation (3.23) also holds for the derivative of the scaling function in $V_{l,k}$ and the supports of the scaling functions are determined by (3.10), then we conclude from the nesting (3.1)

$$\begin{aligned} T_G &= \sum_{r\in\{0,1\}} \int_{V_{l+1,2k+r}} f(v_L(t^n,x)) \cdot \sum_{j=0}^{p-1} \sum_{s\in\{0,1\}} \tilde{m}^{0,s}_{ij} \tilde{\varphi}'_{l+1,2k+s,j}(x)\,dx \\ &= \sum_{r\in\{0,1\}} \int_{V_{l+1,2k+r}} f(v_L(t^n,x)) \cdot \tilde{\varphi}'_{l,k,i}(x)\,dx = \int_{V_{l,k}} f(v_L(t^n,x)) \cdot \tilde{\varphi}'_{l,k,i}(x)\,dx. \end{aligned} \tag{A.32}$$

Analogously, it can be shown that

$$T_S = \int_{V_{l,k}} s(v_L(t^n,x)) \cdot \tilde{\varphi}_{l,k,i}(x)\,dx.$$

Thus we have derived the evolution equations of the single-scale coefficients on the coarse scale l from those on the fine scale $l+1$.

Finally, the two-scale evolution equations (4.7) for the multiscale coefficients are determined by definition. Motivated by the two-scale transformation (3.29) we first define the multiscale coefficients as

$$d^\mu_{l,k,i} := \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{1,s}_{ij} v^\mu_{l+1,2k+s,j}, \quad \mu = n, n+1. \tag{A.33}$$

Then incorporating the evolution equation (A.20) we directly end up with (4.7). In order to derive the right-hand sides in (4.8), (4.9) and (4.10) we have to proceed in a similar way as in the above case of the evolution equations for the single-scale coefficients. In principle we have to replace the mask coefficients $\tilde{m}^{0,s}_{ij}$ by $\tilde{m}^{1,s}_{ij}$ and the single-scale functions $\varphi_{l,k,i}$ by the multiwavelet $\psi_{l,k,i}$.

First of all, we again apply the two-scale transformation (3.29) to the evolution equation (A.20) and obtain

$$d^{n+1}_{l,k,i} = \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{1,s}_{ij} v^{n+1}_{l+1,2k+s,j} = \overline{T}_v - \tau\left(\overline{T}_B - \overline{T}_G - \overline{T}_S\right) \tag{A.34}$$

with

$$\begin{aligned} \overline{T}_v &:= \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{1,s}_{ij} v^n_{l+1,2k+s,j},\ \overline{T}_B := \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{1,s}_{ij} B^n_{l+1,2k+s,j}, \\ \overline{T}_G &:= \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{1,s}_{ij} G^n_{l+1,2k+s,j},\ \overline{T}_S := \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{1,s}_{ij} S^n_{l+1,2k+1,j}. \end{aligned}$$

From the two-scale relation ($A.33$) we directly conclude

$$\overline{T}_v = d^n_{l,k,i}. \tag{A.35}$$

Since (4.3) holds by assumption for $l+1$, the second term can be split into two parts, i.e., $\overline{T}_B = \overline{T}_{B1} + \overline{T}_{B2}$ with

$$\begin{aligned}
\overline{T}_{B1} &:= \sum_{j\in\mathcal{P}} \left(\tilde{m}^{1,1}_{ij} F^n_{l+1,2k+2} \tilde{\varphi}_{l+1,2k+1,j}(x^-_{l+1,2k+2}) - \tilde{m}^{1,0}_{ij} F^n_{l+1,2k} \tilde{\varphi}_{l+1,2k,j}(x^+_{l+1,2k}) \right), \\
\overline{T}_{B2} &:= F^n_{l+1,2k+1} \sum_{j\in\mathcal{P}} \left(\tilde{m}^{1,0}_{ij} \, \tilde{\varphi}_{l+1,2k,j}(x^-_{l+1,2k+1}) - \tilde{m}^{1,1}_{ij} \, \tilde{\varphi}_{l+1,2k+1,j}(x^+_{l+1,2k+1}) \right).
\end{aligned}$$

Due to the nestedness (3.1) of the grid hierarchy we infer by the two-scale relation (3.24)

$$\tilde{\psi}_{l,k,i}(x^+_{l,k}) = \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{1,s}_{ij} \tilde{\varphi}_{l+1,2k+s,j}(x^+_{l+1,2k}), \tag{A.36}$$

$$\tilde{\psi}_{l,k,i}(x^\pm_{l,k+\frac{1}{2}}) = \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{1,s}_{ij} \tilde{\varphi}_{l+1,2k+s,j}(x^\pm_{l+1,2k+1}), \tag{A.37}$$

$$\tilde{\psi}_{l,k,i}(x^-_{l,k+1}) = \sum_{j\in\mathcal{P}} \sum_{s\in\{0,1\}} \tilde{m}^{1,s}_{ij} \tilde{\varphi}_{l+1,2k+s,j}(x^-_{l+1,2k+2}). \tag{A.38}$$

Inflating the term $\overline{T}_{B1}$ by the zeros in (A.27) and using the identities (A.36) and (A.38) we end up with

$$\begin{aligned}
\overline{T}_{B1} &= F^n_{l+1,2k+2} \sum_{j\in\mathcal{P}} \tilde{m}^{1,1}_{ij} \, \tilde{\varphi}_{l+1,2k+1,j}(x^-_{l+1,2k+2}) - F^n_{l+1,2k} \sum_{j\in\mathcal{P}} \tilde{m}^{1,0}_{ij} \, \tilde{\varphi}_{l+1,2k,j}(x^+_{l+1,2k}) \\
&= F^n_{l+1,2k+2} \, \tilde{\psi}_{l,k,i}(x^-_{l,k+1}) - F^n_{l+1,2k} \, \tilde{\psi}_{l,k,i}(x^+_{l,k}).
\end{aligned} \tag{A.39}$$

Similarly we inflate the term $\overline{T}_{B2}$ by the zeros in (A.28) and employ the identity (A.37). However, opposite to (A.30), the term $\overline{T}_{B2}$ does *not* vanish

$$\overline{T}_{B2} = F^n_{l+1,2k+1} \left(\tilde{\psi}_{l,k,i}(x^-_{l+1,2k+1}) - \tilde{\psi}_{l,k,i}(x^+_{l+1,2k+1}) \right) = 0, \tag{A.40}$$

because $\tilde{\psi}_{l,k,i}$ is discontinuous at the center $x_{l+1,2k+1} = x_{l,k+\frac{1}{2}}$ of its support $V_{l,k}$.

For the term $\overline{T}_G$ we proceed in a similar manner. By assumption that (2.10) holds for level $l+1$, we may rewrite this term

$$\overline{T}_G = \sum_{j\in\mathcal{P}} \sum_{r\in\{0,1\}} \tilde{m}^{1,r}_{ij} \int_{V_{l+1,2k+r}} f(v_L(t^n,x)) \cdot \tilde{\psi}'_{l+1,2k+r,j}(x) \, dx. \tag{A.41}$$

Since the two-scale relation (3.24) also holds for the derivative of the multiwavelet in $V_{l,k}$ and the supports of the multiwavelets are determined by (3.10), then we conclude from the nesting (3.1)

$$\begin{aligned}
\overline{T}_G &= \sum_{r\in\{0,1\}} \int_{V_{l+1,2k+r}} f(v_L(t^n,x)) \cdot \sum_{j=0}^{p-1} \sum_{s\in\{0,1\}} \tilde{m}^{1,s}_{ij} \tilde{\varphi}'_{l+1,2k+s,j}(x) \, dx \\
&= \sum_{r\in\{0,1\}} \int_{V_{l+1,2k+r}} f(v_L(t^n,x)) \cdot \tilde{\psi}'_{l,k,i}(x) \, dx.
\end{aligned} \tag{A.42}$$

Analogously, it can be shown that

$$\overline{T}_S = \int_{V_{l,k}} s(v_L(t^n, x)) \cdot \tilde{\psi}_{l,k,i}(x)\, dx.$$

Thus we have derived the evolution equations of the details on the coarse scale l from the evolution equations of the single-scale coefficients on the fine scale $l+1$.

A.3 Some properties of the DG scheme in the mean

The analytical investigation of the adaptive MR-DG scheme in Section 5 uses the l^1-stability and the l^∞-boundedness of the underlying reference DG scheme in the mean. For completeness of this manuscript and for convenience of the reader we give here the full proofs. Here again we confine ourselves to the forward Euler time discretization.

In order to verify these properties, the DG scheme in the mean (2.7) with numerical flux (2.19) is interpreted as a perturbation of a monotone finite volume scheme. For this purpose, we first rewrite the numerical flux (2.19) as a perturbation of a monotone flux evaluated with respect to the mean values, i.e.,

$$\begin{aligned} F_k &= F(v_{k-1,0}, v_{k,0}) + \Delta F_k \quad \text{with} \\ \Delta F_k &:= F(v_{k-1,0} + \tilde{\overline{v}}^-_{k-1}, v_{k,0} - \tilde{\overline{v}}^+_k) - F(v_{k-1,0}, v_{k,0}). \end{aligned}$$

Hence the evolution equation for the mean values reads

$$v^{n+1}_{k,0} = H^n_k - \lambda\,(\Delta F^n_{k+1} - \Delta F^n_k) + \tau\, s^n_{k,0}, \tag{A.43}$$

where the evolution operator

$$H^n_k := v^n_{k,0} - \lambda(F(v^n_{k,0}, v^n_{k+1,0}) - F(v^n_{k-1,0}, v^n_{k,0})), \quad \lambda := \tau/h, \tag{A.44}$$

is only defined by the mean values, i.e., $H^n_k = H(v^n_{k-1,0}, v^n_{k,0}, v^n_{k+1,0})$. Here we use that $\varphi_{k,0} = h^{-1}\,\chi_{V_k}$ holds according to (3.16). Note that for $k = 0$ and $k = N-1$ we access to data outside the set $\mathcal{I}_h$. For this purpose we extend the vectors by $v_{-1,i} = v_{0,i}$ and $v_{N,i} = v_{N-1,i}$, $i \in \mathcal{P}$. This is justified, because we assume that for the time interval $[0, T]$, the solution of the initial value problem (2.1) vanishes outside the interval $[a, b]$.

The numerical flux F is assumed to be monotone. This typically holds, if a CFL condition is satisfied, i.e.,

$$\lambda\, L_F < CFL_{\max} < 1, \tag{A.45}$$

where the Lipschitz constant

$$L_F := \max_{|u|,|v|,|w| \le C} \frac{|F(u,w) - F(v,w)|}{|u-v|} + \max_{|u|,|v|,|w| \le C} \frac{|F(w,u) - F(w,v)|}{|u-v|} \tag{A.46}$$

depends on the initial data, i.e., $C = \|u_0\|_{L^\infty(\mathbb{R})}$. Hence, the discrete evolution operator H defined by (A.44) corresponds to a *monotone* finite volume scheme for sufficiently small time step size τ. It is well-known that a monotone FV scheme satisfies a min-max condition and is l^1-stable, cf. [39], i.e.,

$$\min(v^n_{k-1,0}, v^n_{k,0}, v^n_{k+1,0}) \le H^n_k \le \max(v^n_{k-1,0}, v^n_{k,0}, v^n_{k+1,0}) \tag{A.47}$$

$$\sum_k |H^n_k| \le \sum_k |v^n_{k,0}| \tag{A.48}$$

Furthermore, if the limiter satisfies the condition (2.21), then we may estimate the flux perturbations by the Lipschitz continuity of the numerical flux F

$$|\Delta F_k| \le L^-_{1,k}\,|\tilde{\overline{v}}^-_{k-1}| + L^+_{2,k}\,|\tilde{\overline{v}}^+_k| \le L_F\,C_\alpha\,h^\alpha, \tag{A.49}$$

where the local Lipschitz constants are determined by

$$\begin{aligned} L^-_{1,k} &:= \frac{F(v_{k-1,0}+\tilde{\overline{v}}^-_{k-1}, v_{k,0}-\tilde{\overline{v}}^+_k) - F(v_{k-1,0}, v_{k,0}-\tilde{\overline{v}}^+_k)}{|\tilde{\overline{v}}^-_{k-1}|}, \\ L^+_{2,k} &:= \frac{F(v_{k-1,0}, v_{k,0}-\tilde{\overline{v}}^+_k) - F(v_{k-1,0}, v_{k,0})}{|\tilde{\overline{v}}^+_k|}. \end{aligned}$$

Here the Lipschitz constant L_F depends on the bound

$$\|\mathcal{R}\mathbf{v}\|_\infty + C_\alpha\,h^\alpha. \tag{A.50}$$

In the following we show that the DG scheme in the mean inherits the above properties of the underlying monotone FV scheme, where we basically have to assume that the time step is sufficiently small and the higher order coefficients are bounded by an appropriate limiting process. For this purpose we first verify that one time step is l^∞-stable.

Lemma 5 *(Local l^∞-stability of reference DG scheme in the mean) Let $\Omega = [a,b]$ be bounded. Furthermore we assume that*

1. *the numerical flux $F : \mathbb{R}^2 \to \mathbb{R}$ is consistent with the flux f, locally Lipschitz continuous and monotone under the CFL condition CFL_{max},*
2. *the source function $s : \mathbb{R} \to \mathbb{R}$ satisfies a global Lipschitz condition with Lipschitz constant L_s and $s(0) = 0$; in particular, $L_s = 0$ if $s \equiv 0$.*
3. *the limiter Π_L does not modify the zero order coefficients (mean values) and (2.21) holds at the cell interfaces and in case of an inhomogeneous problem ($s \not\equiv 0$) also inside the cell, i.e.,*

$$|\sum_{i\in\mathcal{P}^*} (\Pi_L\mathbf{v})_{k,i}\varphi_{k,i}(x)| \le C_\alpha h^{\alpha-1}, \quad x \in V_k. \tag{A.51}$$

For any $\mathbf{v} \in (\mathbb{R}^p)^{N_L}$ let $L_F = L_F(\|\mathcal{R}\mathbf{v}\|_\infty + C_\alpha h^\alpha)$ be the Lipschitz constant in (A.46) *and let $\lambda = \tau/h$ be small enough such that the CFL condition* (A.45) *holds, thus, the numerical flux F is monotone. Then the forward Euler step of the reference DG scheme is l^∞-stable, i.e.,*

$$\|\mathcal{R}\,\mathcal{E}_L\,\Pi_L\,\mathbf{v}\|_\infty \le (1 + L_s\tau)\|\mathcal{R}\,\mathbf{v}\|_\infty + \lambda\,(2\,L_F + L_s)C_\alpha\,h^\alpha. \tag{A.52}$$

Proof: Let be $\mathbf{v} \in (\mathbb{R}^p)^{N_L}$. Then the higher order coefficients are limited such that (2.20) holds, i.e., $\mathbf{w} := \Pi_L\mathbf{v}$, before applying the evolution operator $\mathcal{E}_L$. According to (A.43) the updated data may be estimated by

$$\|\mathcal{R}\,\mathcal{E}_L\,\mathbf{w}\|_\infty \le \|\mathbf{H}(\mathbf{w}_0)\|_\infty + \lambda\,\|\mathbf{\Delta}(\mathbf{w})\|_\infty + \tau\,\|\mathbf{S}(\mathbf{w})\|_\infty, \tag{A.53}$$

where the vectors $\mathbf{H}$, $\mathbf{\Delta}$, $\mathbf{S} \in \mathbb{R}^{N_L}$ are determined by the components H_k, $|\Delta F_k| + |\Delta F_{k+1}|$ and $s_{k,0}$ that are computed by the data $\mathbf{w}_0 := \mathcal{R}\,\mathbf{w}$ and $\mathbf{w}$, respectively.

Since the numerical flux F is monotone, we infer by the monotonicity of the corresponding FV scheme

$$\|\mathbf{H}(\mathbf{w}_0)\|_\infty \leq \|\mathbf{w}_0\|_\infty = \|\mathcal{R}\mathbf{v}\|_\infty. \tag{A.54}$$

Here we use that the mean values are not modified by the limiter, i.e.,

$$\mathbf{w}_0 = \mathcal{R}\mathbf{w} = \mathcal{R}\Pi_L\mathbf{v} = \mathcal{R}\mathbf{v}. \tag{A.55}$$

Since we perform limiting before the time evolution, condition (2.21) holds and we may estimate the flux perturbations by (A.49) and obtain

$$\lambda \, \|\mathbf{\Delta}(\mathbf{w})\|_\infty \leq 2 \, \lambda \, L_F \, C_\alpha \, h^\alpha. \tag{A.56}$$

Furthermore, we may estimate the source quadrature by

$$|S_{k,0}| \leq \int_{V_k} |s(w(x)) - s(0)| \cdot |\tilde{\varphi}_{k,0}(x)| \, dx \leq L_s \, \|w\|_{L^\infty(V_k)},$$

where the function w is defined according to (2.3) with single-scale coefficients $\mathbf{w}$. Here we use that the source function s is assumed to be globally Lipschitz continuous and $s(0) = 0$, and $\tilde{\varphi}_{k,0} = h^{-1}\chi_{V_k}$ is determined by (3.16). Since condition (A.51) holds inside the cell, we infer

$$|w(x)| \leq |w_{k,0} \, \varphi_{k,0}(x) + \sum_{i \in \mathcal{P}^*} w_{k,i} \, \varphi_{k,i}(x)| \leq |w_{k,0}| + C_\alpha \, h^{\alpha-1} \leq \|\mathbf{w}_0\|_\infty + C_\alpha h^{\alpha-1},$$

where we use $\varphi_{k,0} = h\tilde{\varphi}_{k,0}$. Then we obtain with (A.55) for the source quadrature

$$\|\mathbf{S}(\mathbf{w})\|_\infty \leq L_s(\|\mathcal{R}\,\mathbf{v}\|_\infty + C_\alpha \, h^{\alpha-1}). \tag{A.57}$$

Finally, we conclude the assertion (A.52) from (A.53), where we use the estimates (A.54), (A.56), (A.57), i.e.,

$$\|\mathcal{R}\,\mathcal{E}_L\,\Pi_L\,\mathbf{v}\|_\infty \leq \|\mathcal{R}\mathbf{v}\|_\infty + 2\lambda L_F C_\alpha h^\alpha + \tau \, L_s \, \|\mathcal{R}\,\mathbf{v}\|_\infty + \lambda \, L_s \, C_\alpha \, h^\alpha.$$

From the local l^∞-stability we may now infer the uniform stability result. □

Proposition 2 *(l^∞-stability of reference DG scheme in the mean) Let the assumptions of Lemma 5 hold true and*

$$\|\mathcal{R}\mathcal{E}_L\Pi_L\mathbf{w}\|_\infty \leq (1 + L_s\tau)\|\mathcal{R}\mathbf{w}\|_\infty + C_p h^\alpha \quad \textit{for all } \mathbf{w} \in (\mathbb{R}^p)^{N_L} \tag{A.58}$$

with $C_p = C_p(\|\mathcal{R}\mathbf{w}\|_\infty + C_\alpha h^\alpha)$ and $\alpha > 1$. Put $C_\infty := e^{L_s T}(\|u_0\|_{L^\infty} + M)$, where $M > 0$ is an arbitrary but fixed constant. Let $\bar{C}_p$ and $\bar{L}_F$ be upper bounds for $C_p(\mathbf{w})$ and $L_F(\mathbf{w})$ such that $\bar{C}_p \geq C_p(\mathbf{w})$ and $\bar{L}_F \geq L_F(\mathbf{w})$ hold for all $\|\mathcal{R}\mathbf{w}\|_\infty \leq C_\infty$. If $\lambda = \tau/h$ is chosen constant and sufficiently small, such that a CFL condition holds with $\bar{L}_F$, implying the monotonicity of F, and

$$h \leq \begin{cases} \left(\frac{\lambda M L_s}{\bar{C}_p}\right)^{\frac{1}{\alpha-1}} & \textit{if } L_s \neq 0 \\ \left(\frac{\lambda M}{T\bar{C}_p}\right)^{\frac{1}{\alpha-1}} & \textit{if } L_s = 0 \end{cases}$$

then the forward Euler step of the reference DG scheme is uniformly l^∞-stable, i.e., for all time levels n with $n\tau \leq T$ we have

$$\|\mathcal{R}\,\mathbf{v}_L^n\|_\infty \leq C_\infty \qquad \textit{for} \quad n\,\tau \leq T, \tag{A.59}$$

where in particular the constant C_∞ is independent of the discretization parameters h, τ and n.

Proof: We will show the proposition by induction. First we consider $n = 0$. Then by the definition (3.16) of $\tilde{\varphi}_{k,0}$ and the choice of C_∞ we may estimate the discrete norm by the function norm, i.e.,

$$\|\mathcal{R}\mathbf{v}_L^0\|_\infty \le \|u_0\|_{L^\infty} \le C_\infty,$$

where we use $1 \le e^{L_s T}$ and $M > 0$.

Now we assume that $\|\mathcal{R}\mathbf{v}_L^j\|_\infty \le C_\infty$ holds for $j < n$. Then by assumption (A.58) holds with $\bar{C}_p$ for all $\mathbf{v}_L^j$, $j < n$. Therefore we may apply (A.58) repeatedly and obtain

$$\|\mathcal{R}\,\mathbf{v}_L^n\|_\infty = \|\mathcal{R}\,\mathcal{E}_L\,\Pi_L\,\mathbf{v}_L^{n-1}\|_\infty \le (1 + L_s\,\tau)^n \|\mathcal{R}\,\mathbf{v}_L^0\|_\infty + \bar{C}_p\,h^\alpha \sum_{i=0}^{n-1} (1 + L_s\,\tau)^i.$$

Now we use $n\tau \le T$ and the bound on h. If $L_s = 0$, this implies

$$\|\mathcal{R}\mathbf{v}_L^n\|_\infty \le \|\mathcal{R}\mathbf{v}_L^0\|_\infty + n\bar{C}_p h^\alpha \le \|\mathcal{R}\mathbf{v}_L^0\|_\infty + T\bar{C}_p h^{\alpha-1}/\lambda \le \|\mathcal{R}\mathbf{v}_L^0\|_\infty + M = C_\infty.$$

Otherwise, if $L_s \neq 0$, we estimate the geometric sum by $(1 + L_s\tau)^n/(L_s\tau) \le e^{C_s T}/(C_s\tau)$ and obtain

$$\|\mathcal{R}\mathbf{v}_L^n\| \le e^{L_s T}\big(\|\mathcal{R}\mathbf{v}_L^0\|_\infty + \bar{C}_p h^\alpha/(L_s\tau)\big) \le e^{L_s T}\big(\|\mathcal{R}\mathbf{v}_L^0\|_\infty + M) = C_\infty.$$

□

The result on the boundedness of the DG scheme in the mean is now used to verify l^1-stability.

Lemma 6 *(Local l^1-stability of reference DG scheme in the mean) Let the assumptions of Lemma 5 hold true. For $\mathbf{v}, \overline{\mathbf{v}} \in (\mathbb{R}^p)^{N_L}$ let $L_F \ge \max(L_F(\mathbf{v}), L_F(\overline{\mathbf{v}}))$. Let λ be sufficiently small such that the CFL condition* (A.45) *holds with L_F, then the forward Euler step of the reference DG scheme in the mean is l^1-stable, i.e.,*

$$\|\mathcal{R}(\mathcal{E}_L\,\Pi_L\,\mathbf{v} - \mathcal{E}_L\,\Pi_L\,\overline{\mathbf{v}})\|_1 \le (1 + L_s\tau)\|\mathcal{R}(\mathbf{v} - \overline{\mathbf{v}})\|_1 + 2\,(b-a)\,\lambda\,(2L_F + L_s)\,C_\alpha\,h^\alpha. \quad \text{(A.60)}$$

Proof: Let be $\mathbf{v}, \overline{\mathbf{v}} \in (\mathbb{R}^p)^{N_L}$. Then the higher order coefficients are limited such that (2.20) holds, i.e., $\mathbf{w} := \Pi_L\,\mathbf{v}$, $\overline{\mathbf{w}} := \Pi_L\,\overline{\mathbf{v}}$, before applying the evolution operator $\mathcal{E}_L$. According to (A.43) the difference of the updated data may be estimated by

$$\begin{aligned}&\|\mathcal{R}(\mathcal{E}_L\,\mathbf{w} - \mathcal{E}_L\,\overline{\mathbf{w}})\|_1 \le \\ &\|\mathbf{H}(\mathbf{w}_0) - \mathbf{H}(\overline{\mathbf{w}}_0)\|_1 + \lambda\,(\|\boldsymbol{\Delta}(\mathbf{w})\|_1 + \|\boldsymbol{\Delta}(\overline{\mathbf{w}})\|_1) + \tau\,\|\mathbf{S}(\mathbf{w}) - \mathbf{S}(\overline{\mathbf{w}})\|_1, \quad \text{(A.61)}\end{aligned}$$

where the vectors $\mathbf{H}$, $\boldsymbol{\Delta}$, $\mathbf{S} \in \mathbb{R}^{N_L}$ are determined by the components H_k, $|\Delta\,F_k| + |\Delta\,F_{k+1}|$ and $s_{k,0}$ that are computed by the data $\mathbf{w}_0 := \mathcal{R}\,\mathbf{w}$, $\mathbf{w}$ and $\overline{\mathbf{w}}_0 := \mathcal{R}\,\overline{\mathbf{w}}$, $\overline{\mathbf{w}}$, respectively. Since the numerical flux F is monotone due to the CFL condition, we infer by the l^1-contractivity of the corresponding FV scheme

$$\|\mathbf{H}(\mathbf{w}_0) - \mathbf{H}(\overline{\mathbf{w}}_0)\|_1 \le \|\mathbf{w}_0 - \overline{\mathbf{w}}_0\|_1 = \|\mathcal{R}(\mathbf{v} - \overline{\mathbf{v}})\|_1. \quad \text{(A.62)}$$

Here we use that the mean values are not modified by the limiter, i.e.,

$$\mathbf{w}_0 = \mathcal{R}\mathbf{w} = \mathcal{R}\Pi_L\mathbf{v} = \mathcal{R}\mathbf{v}, \quad \overline{\mathbf{w}}_0 = \mathcal{R}\overline{\mathbf{w}} = \mathcal{R}\Pi_L\overline{\mathbf{v}} = \mathcal{R}\overline{\mathbf{v}}. \quad \text{(A.63)}$$

Since we perform limiting before the time evolution, condition (2.21) holds and we may estimate the flux perturbations by (A.49) and obtain by the choice of L_F

$$\lambda\,(\|\boldsymbol{\Delta}(\mathbf{w})\|_1 + \|\boldsymbol{\Delta}(\overline{\mathbf{w}})\|_1) \le 2\,(b-a)\,\lambda\,2L_F\,C_\alpha\,h^\alpha. \quad \text{(A.64)}$$

We may estimate the source quadrature by

$$|S_{k,0} - \overline{S}_{k,0}| \leq \int_{V_k} |s(w(x)) - s(\overline{w}(x))| \cdot |\tilde{\varphi}_{k,0}(x)|\, dx \leq L_s \|w - \overline{w}\|_{L_\infty(V_k)},$$

where the functions w and $\overline{w}$ are defined according to (2.3) with single-scale coefficients $\mathbf{w}$ and $\overline{\mathbf{w}}$, respectively. Since condition (A.51) is assumed to hold, we infer from (A.45)

$$\|w - \overline{w}\|_{L_\infty(V_k)} \leq |w_{k,0} - \overline{w}_{k,0}| + \| \sum_{i \in \mathcal{P}^*} (w_{k,i} - \overline{w}_{k,i})\, \varphi_{k,i} \|_{L_\infty(V_k)} \leq |w_{k,0} - \overline{w}_{k,0}| + 2C_\alpha\, h^{\alpha-1},$$

where we use $\varphi_{k,0} = \chi_{V_k}$. Then we obtain with (A.63) for the source quadrature

$$\|\mathbf{S}(\mathbf{w}) - \mathbf{S}(\overline{\mathbf{w}})\|_1 \leq L_s\, (\|\mathcal{R}\,(\mathbf{v} - \overline{\mathbf{v}})\|_1 + 2(b-a)\, C_\alpha\, h^{\alpha-1}). \tag{A.65}$$

Finally, we conclude the assertion (A.60) from (A.61), where we use the estimates (A.62), (A.64), (A.65), i.e.,

$$\|\mathcal{R}(\mathcal{E}_L\, \mathbf{w} - \mathcal{E}_L\, \overline{\mathbf{w}})\|_1 \leq \|\mathcal{R}\,(\mathbf{v} - \overline{\mathbf{v}})\|_1 + 2(b-a)\lambda 2 L_F C_\alpha h^\alpha + \tau\, L_s\, \|\mathcal{R}\,(\mathbf{v} - \overline{\mathbf{v}})\|_1 + \lambda\, L_s\, 2(b-a) C_\alpha\, h^\alpha.$$

□

In order to control the perturbation error of the adaptive MR-DG scheme and the reference DG scheme in the mean we employ Lemma 6 in Theorem 1. For this purpose we need that the constants on the right-hand side of (A.60) do not depend on the discretization. In Proposition 2 this is proven to hold for the reference DG scheme provided that the discretization is chosen sufficiently small. For the adaptive MR-DG scheme this follows by Theorem 4.

Finally we conclude uniform l^1-stability using Lemma 6.

Proposition 3 *(l^1-stability of reference DG scheme in the mean) Define by $\mathbf{v}_L^n := \mathcal{E}_L\, \Pi_L\, \mathbf{v}_L^{n-1}$ and $\overline{\mathbf{v}}_L^n := \mathcal{E}_L\, \Pi_L\, \overline{\mathbf{v}}_L^{n-1}$ two approximations of the DG scheme with initial data $u_0 \in L^1(\mathbb{R})$ and $\overline{u}_0 \in L^1(\mathbb{R})$, respectively. For both sequences let the assumptions of Proposition 2 hold true, i.e., both sequences are bounded in the l^∞-norm and there exists a uniform Lipschitz bound L_F. Let $C_1 = e^{L_s T}$ and $C_2 = 2(b-a)(2L_F + L_s)C_\alpha / L_s$ if $L_s \neq 0$ and $C_2 = 4T(b-a)L_F C_\alpha$ if $L_s = 0$. Then the forward Euler step of the reference DG scheme in the mean is uniformly l^1-stable, i.e.,*

$$\|\mathcal{R}\,(\mathbf{v}_L^n - \overline{\mathbf{v}}_L^n)\|_1 \leq C_1(1 + C_2) \max(\|u_0 - \overline{u}_0\|_{L^1}, h^{\alpha-1}) \qquad \text{for} \quad n\,\tau \leq T, \tag{A.66}$$

where in particular the constants C_1, C_2 are independent of the discretization parameters h, τ and n.

Proof: Without loss of generality, the CFL condition is chosen such that the numerical flux is monotone and the discrete evolution operator H defined by (A.44) corresponds to a *monotone* finite volume scheme. Then we may apply (A.60) repeatedly and obtain

$$\|\mathcal{R}\,(\mathbf{v}_L^n - \overline{\mathbf{v}}_L^n)\|_1 \leq (1 + L_s\,\tau)^n \|\mathcal{R}\,(\mathbf{v}_L^0 - \overline{\mathbf{v}}_L^0)\|_1 + 2\,(b-a)\,\lambda\,(2L_F + L_s)\, C_\alpha\, h^\alpha \sum_{i=0}^{n-1} (1 + L_s\,\tau)^i.$$

By definition (3.16) of $\tilde{\varphi}_{k,0}$ we furthermore may estimate the discrete norm by the function norm, i.e.,

$$\|\mathcal{R}\,(\mathbf{v}_L^0 - \overline{\mathbf{v}}_L^0)\|_1 \leq \|u_0 - \overline{u}_0\|_{L^1}.$$

Similar to the proof of Proposition 2 we now distinguish the two cases $L_s = 0$ and $L_s \neq 0$. For both cases we deduce

$$\|\mathcal{R}\,(\mathbf{v}_L^n - \tilde{\mathbf{v}}_L^n)\|_1 \leq C_1 \|\mathcal{R}\,(\mathbf{v}_L^0 - \tilde{\mathbf{v}}_L^0)\|_1 + C_1 C_2 h^{\alpha-1},$$

where we use $n\tau \leq T$ and the bound on h according to Proposition 2. This verifies (A.66). □

It is worthwhile mentioning, that the above stability result implies the TVB property of the DG scheme in the mean. For this purpose we have to choose $\overline{\mathbf{v}}_L^n$ as a shift of $\mathbf{v}_L^n$ by one index. Hence the assumption (2.21) of uniformly bounded higher order coefficients does not only imply that the weak limit satisfies an entropy condition, but ensures convergence at all. Hence (2.21) is a strong assumption.

A.4 Projection of mean values from coarse to fine scales

In order to perform Step L4 in Algorithm 2 we have to compute mean values on level L from data on coarser levels. This can be done in an efficient way, where we have to pre-compute some mask coefficients. For this purpose, we consider the function $v \in S_L^p$ characterized by the coefficients on an adaptive grid $\mathcal{G}$. Consider a cell $V_{l,k}$ with $(l,k) \in \mathcal{G}$. For any cell $V_{L,k'} \subset V_{l,k}$, we want to calculate the mean values $v_{L,k',0}$. The single scale-representation gives

$$v_{L,k',0} = \int_{V_{L,k'}} v(x) \cdot \tilde{\varphi}_{L,k',0}(x)dx = \sum_{i\in\mathcal{P}} v_{l,k,i} \int_{V_{L,k'}} \varphi_{l,k,i}(x) \cdot \tilde{\varphi}_{L,k',0}(x)dx.$$

Note that there are no detail coefficients due to the assumptions. Since $\varphi_{l,k,i}$ is defined by shifts and translates of the mother-scaling function, the value of

$$c_{d,s,i} := \int_{V_{L,k'}} \varphi_{l,k,i}(x) \cdot \tilde{\varphi}_{L,k',0}(x)dx$$

only depends on the difference in levels $d = L - l$ and the relative position $s = 2^d k' - k$.

For instance, for the basis constructed in Appendix A.1, one can easily compute

$$c_{d,s,i} = \begin{cases} 1 & i = 0 \\ \sqrt{3}\left(2^{1-d}s + 2^{-d} - 1\right) & i = 1 \\ \sqrt{5}\left(6\,s^2 4^{-d} + 6\,4^{-d}s + 2\,4^{-d} - 6\,s\,2^{-d} - 3\,2^{-d} + 1\right) & i = 2 \\ \sqrt{7}\left(20\,s^3 8^{-d} + 30\,8^{-d}s^2 + 20\,8^{-d}s + 5\,8^{-d} - 30\,s^2 4^{-d} - 30\,4^{-d}s \right. & \\ \quad \left. -10\,4^{-d} + 12\,s\,2^{-d} + 6\,2^{-d} - 1\right) & i = 3 \\ \ldots & i > 3 \end{cases}$$

With these coefficients we obtain the mean value by

$$v_{L,k',0} = \sum_{i\in P} c_{L-l,2^{L-l}k'-k,i} v_{l,k,i}.$$