UNIVERSIDAD DE CONCEPCIÓN

Centro de Investigación en Ingeniería Matemática (CI^{2}MA)

Weno reconstructions of unconditionally optimal high order

Antonio Baeza, Raimund Bürger,
Pep Mulet, David Zorío

PREPRINT 2018-14

SERIE DE PRE-PUBLICACIONES

WENO RECONSTRUCTIONS OF UNCONDITIONALLY OPTIMAL HIGH ORDER

ANTONIO BAEZA[A], RAIMUND BÜRGER[B], PEP MULET[C], AND DAVID ZORÍO[D]

Abstract. A modified Weighted Essentially Non-Oscillatory (WENO) technique preventing accuracy loss near smooth extrema, regardless of their order, is presented. This approach uses only local data from around the reconstruction stencil. The resulting weights to account for discontinuities are non-dimensional and scale-independent. Two different ways to define the weights and prove that both give the sought accuracy order are provided. Several validation tests are also presented, in which the method is applied to algebraic equations, scalar conservation laws and systems of conservation laws.

1. Introduction

1.1. **Scope.** Weighted Essentially Non-Oscillatory (WENO) schemes, initially proposed in [16] and later improved in [13], have become a very useful tool to solve hyperbolic conservation laws, that is, initial-value problems of the type

$$\boldsymbol{u}_t + \sum_{i=1}^{d} \boldsymbol{f}_i(\boldsymbol{u})_{x_i} = \boldsymbol{0}, \tag{1.1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, $t > 0$, $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}, t) = (u_1, \ldots, u_N)^{\mathrm{T}}$ is the sought vector of unknowns, and $\boldsymbol{f}_i(\boldsymbol{u}) = (f_{i,1}(\boldsymbol{u}), \ldots, f_{i,N}(\boldsymbol{u}))^{\mathrm{T}}$ for $i = 1, \ldots, d$, supplied with the initial condition

$$\boldsymbol{u}(\boldsymbol{x}, 0) = \boldsymbol{u}_0(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^d. \tag{1.2}$$

WENO schemes present a high order of accuracy in smooth zones and avoid the oscillatory behaviour typical of the reconstructions from discontinuous data through a sophisticated construction of non-linear weights [13]. However, such weights are sensitive not only to discontinuities, but also, and in general, to abrupt changes in any higher derivative of the function that generates the data, which leads to an undesired loss of accuracy near smooth extrema. A variety of solutions to handle this problem have been proposed in the literature; see for instance [1,2,11,21]. However,

Date: March 27, 2018.

Key words and phrases. finite-difference schemes, WENO reconstructions, optimal order, smooth extrema.

Corresponding author: David Zorío Ventura. E-mail: dzorio@ci2ma.udec.cl.

[A]Departament de Matemàtiques, Universitat de València, Av. Vicent Andrés Estellés, E-46100 Burjassot, Spain. E-Mail: antonio.baeza@uv.es.

[B]CI^{2}MA and Departamento de Ingeniería Matemática, Universidad de Concepción, Casilla 160-C, Concepción, Chile. E-Mail: rburger@ing-mat.udec.cl.

[C]Departament de Matemàtiques, Universitat de València, Av. Vicent Andrés Estellés, E-46100 Burjassot, Spain. E-Mail: pep.mulet@uv.es.

[D]CI^{2}MA, Universidad de Concepción, Casilla 160-C, Concepción, Chile. E-Mail: dzorio@ci2ma.udec.cl.

none of the proposed solutions allows to unconditionally attain the optimal order of accuracy (that is, regardless of the order of the smooth extrema) depending only on the local data without ending up with dimensional or scale-dependent weights. In other words, the proposed solutions either use some dimensional (namely, depending on the grid size) or not properly scaled parameter, or use data from the global numerical solution to define a non-dimensional and scale-independent parameter to prevent such loss of accuracy.

It is the purpose of this paper to perform an extensive accuracy analysis of the elements involved in the definition of weights in WENO schemes, mainly smoothness indicators and undivided differences, and to find solid upper and lower bounds on these quantities in order to allow a fine theoretical analysis. Such analysis is also used to find a suitable weight design whose associated reconstruction algorithm does not loose accuracy in smooth zones, even in presence of smooth extrema of any order, through non-dimensional and scale-independent weights that only use the information from the local data of the stencil.

1.2. **Related work.** Many works in the literature tackle the problem of achieving optimal order of accuracy near smooth extrema. In [11] the authors obtain optimal order convergence near critical points for the case of fifth order through a simple modification of Jiang-Shu weights [13], which involves mapping the weights to values that verify an optimality condition. The approach was further extended up to order 17 in [9] and further enhanced in [8] by means of a different mapping.

A different weight design was followed in the fifth order WENO-Z method of [3], which attains fourth order accuracy even at critical points. In [4] the WENO-Z scheme was extended to any odd order of accuracy, getting optimal order at smooth extrema by proper parameter tuning.

Yamaleev and Carpenter introduced in [21] a new method, named ESWENO, based on the third-order case previously introduced in [20] that ensures energy stability in an L_2 norm. Even though it was not their primary goal to enhance order at critical points, it turns out that the resulting scheme achieves optimal order in the presence of smooth extrema provided that the number of zero derivatives is at most the order of the scheme minus three.

Another way to handle the problem of the order loss at critical points is the modification of the smoothness indicators. In [10] a new smoothness measurement provides optimal order for functions with critical points, but in which the second derivative is not zero.

The design of weights in WENO schemes typically involves a quantity that avoids division by zero whenever an smoothness indicator becomes zero. It was noted in [1] that the choice of this parameter is crucial for the achievement of optimal order at critical points and that, for the case of the original weights of Jiang and Shu, the choice of the parameter proportional to the square of the mesh size provides the desired accuracy even at critical points. A similar analysis was later performed in [6], regarding WENO-Z schemes, and in [2] with respect to the ESWENO weights of Yamaleev and Carpenter. See also [14].

1.3. **Outline of this paper.** To properly support the grounds of this work, the required theoretical background is presented in Section 2. Section 3 is devoted to the definition of a modified WENO scheme which attains optimal order of accuracy

regardless of the number of consecutive zero derivatives of the function to be reconstructed, and without using any scaling parameter. We prove that the new scheme has unconditionally optimal order of accuracy under those conditions. The section is divided into two subsections in which we propose two alternative definitions of the weights. Both methods attain unconditionally the optimal order for smooth data: on one hand, Subsection 3.1 introduces a new weight design based on adding an additional node to the stencil, together with the theoretical results which guarantee that this scheme entails the optimal order; on the other hand, in Subsection 3.2 a second alternative is presented in which no additional node is required to attain the optimal order, whose accuracy is proven using previous results from Subsection 3.1 and some additional results provided in Appendix A. In section 4 we present some numerical experiments, both for algebraic problems in Subsection 4.1 and problems involving hyperbolic conservation laws in Subsection 4.2. Finally, in Section 5 some conclusions are drawn. Some additional results required to prove the accuracy of the novel schemes are collected in Appendix A.

2. Regularity properties of functions

2.1. **Preliminaries.** In this work, given a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ the notation $f(h) = \mathcal{O}(h^\alpha)$ for $\alpha \in \mathbb{Z}$ always corresponds to the behaviour of a function f as $h \to 0$ in the standard sense, that is,

$$f(h) = \mathcal{O}(h^\alpha) \Leftrightarrow \limsup_{h\to 0} \left| \frac{f(h)}{h^\alpha} \right| < \infty.$$

Furthermore, we write $f(h) = \bar{\mathcal{O}}(h^\alpha)$ to express the more restrictive property

$$f(h) = \bar{\mathcal{O}}(h^\alpha) \Leftrightarrow \limsup_{h\to 0} \left| \frac{f(h)}{h^\alpha} \right| < \infty \quad \text{and} \quad \liminf_{h\to 0} \left| \frac{f(h)}{h^\alpha} \right| > 0.$$

Since for positive functions f and g,

$$\limsup_{h\to 0} f(h)g(h) \leq \limsup_{h\to 0} f(h) \limsup_{h\to 0} g(h),$$
$$\liminf_{h\to 0} f(h)g(h) \geq \liminf_{h\to 0} f(h) \liminf_{h\to 0} g(h),$$

it follows for $\alpha, \beta \in \mathbb{Z}$ that $\mathcal{O}(h^\alpha)\mathcal{O}(h^\beta) = \mathcal{O}(h^{\alpha+\beta})$ and $\bar{\mathcal{O}}(h^\alpha)\bar{\mathcal{O}}(h^\beta) = \bar{\mathcal{O}}(h^{\alpha+\beta})$.

Since, for $A_i \subseteq (0, \infty)$, $\sup_i \inf A_i = \inf_i \sup A_i^{-1}$, $A_i^{-1} = \{1/x \colon x \in A_i\}$, it follows that for positive f,

$$\liminf_{h\to 0} f(h) = \left(\limsup_{h\to 0} f(h)^{-1} \right)^{-1},$$

therefore, if f is positive, $f(h) = \bar{\mathcal{O}}(h^\alpha)$ implies $f(h)^{-1} = \mathcal{O}(h^{-\alpha})$.

Definition 2.1. *We say that a function f has a smooth critical point of maximal order $k \geq 0$ at x if $f^{(l)}(x) = 0$, $l = 1, \ldots, k$ and $f^{(k+1)}(x) \neq 0$.*

2.2. **WENO reconstructions.** For a $(2r-1)$-point stencil $S = \{x_{-r+1}, \ldots, x_{r-1}\}$ and a scalar function f we assume that the data $\{f_{-r+1}, \ldots, f_{r-1}\}$ is defined either as the point values

$$f_j = f(x_j) \quad -r+1 \leq j \leq r-1 \tag{2.1}$$

or the cell averages

$$f_j = \frac{1}{h}\int_{x_{j-1/2}}^{x_{j+1/2}} f(x)\,\mathrm{d}x \quad -r+1 \le j \le r-1, \tag{2.2}$$

where in both cases $x_j = x_{j,h}$, $x_{j+1} - x_j = h$ for $-r+1 \le j \le r-1$, and assuming that we wish to approximate the point value $f(x_{1/2})$. In what follows we state some results that will be helpful for the analysis of the accuracy of the WENO reconstructions for both cases.

We denote by $p_{r,i} \in \Pi_{r-1}$ the reconstruction polynomials of the corresponding substencils $S_{r,i} = \{x_{-r+1+i}, \dots, x_i\}$, $0 \le i \le r-1$, where the corresponding interpolation property on $S_{r,i}$ is assumed to be satisfied, i.e., $p_{r,i}(x_j) = f_j$ for all $x_j \in S_{r,i}$. In what follows we omit the subindex r when no confusion may arise.

The WENO strategy consists in defining a reconstruction q as a convex combination of the individual reconstructions p_i with appropriately designed weights $\omega_0, \dots, \omega_{r-1} \ge 0$, where $\omega_0 + \cdots + \omega_{r-1} = 1$:

$$q(x) = \sum_{i=0}^{r-1} \omega_i p_i(x). \tag{2.3}$$

The weights ω_i are functions of some smoothness indicators, which we take according to the proposal by Jiang and Shu [13]:

$$I_{r,i} = \sum_{l=1}^{r-1} \int_{x_{-1/2}}^{x_{1/2}} h^{2l-1}\big(p_{r,i}^{(l)}(x)\big)^2\,\mathrm{d}x. \tag{2.4}$$

In the remainder of the paper, we denote by Π_k, $k \in \mathbb{N}_0$, the space of polynomials of *maximal* degree k, and by $\bar{\Pi}_k$ the space of polynomials of *exact* degree k, $k \in \mathbb{N}_0$.

Lemma 2.1. *Let $c_0 < c_1 < \cdots < c_n$ and z be fixed real numbers. Let $S = \{x_{0,h}, \dots, x_{n,h}\}$ be an $(n+1)$-point stencil with $x_{i,h} = z + c_i h$ for $h > 0$. For any real function f, let $p_h = p_h[f] \in \Pi_n$ be the reconstruction polynomial that satisfies either $p_h(x_{i,h}) = f(x_{i,h})$ for $i = 0, \dots, n$ or*

$$\int_{x_{i,h}-h/2}^{x_{i,h}+h/2} p_h(x)\,\mathrm{d}x = \int_{x_{i,h}-h/2}^{x_{i,h}+h/2} f(x)\,\mathrm{d}x \quad \text{for } i = 0, \dots, n,$$

depending on whether the underlying data are point values (2.1) *or cell averages* (2.2).

Then, for $1 \le j \le n$ and $s \ge j$, there exists polynomials $b_{s,j} \in \Pi_{n-j}$, depending uniquely on the type of reconstruction and parameters $c_0, \dots, c_n$, such that for any $f \in \mathcal{C}^{m+1}$

$$p_h^{(j)}(z + wh) = \sum_{s=j}^{m} b_{s,j}(w) h^{s-j} f^{(s)}(z) + \mathcal{O}(h^{m+1-j}) \tag{2.5}$$

for sufficiently small wh. The functions $b_{s,j}$ have the following properties:

(1) *For $j \le s \le n$,*

$$b_{s,j}(w) = s!\binom{s}{j} w^{s-j}.$$

(2) *$b_{s,1} \equiv 0$ if and only if $n = 1$, s is even and $c_0 = -c_1$, and $b_{s,1} \not\equiv 0$ otherwise.*

The proof relies on applying appropriate operators to Taylor expansions, based on the following result.

Lemma 2.2. *Let $\mathcal{L}\colon \mathcal{C}^{m+1}[a,b] \to \Pi_n$ be a linear and continuous operator with respect to the norm $\|\cdot\| = \|\cdot\|_\infty$. Then there exists $K > 0$ such that for any $\zeta \in [a,b]$ and $w \in [a,b]$,*

$$\mathcal{L}[f](w) = \sum_{s=0}^{m} \frac{f^{(s)}(\zeta)}{s!}\mathcal{L}[(w-\zeta)^s] + \Delta_{m+1,\zeta}\mathcal{L}[f],$$

where $\Delta_{m+1,\zeta}\mathcal{L}[f]$ is a remainder term that satisfies

$$\big\|\Delta_{m+1,\zeta}\mathcal{L}[f]\big\| \le K\big\|f^{(m+1)}\big\|.$$

Proof of Lemma 2.2. The result follows by applying $\mathcal{L}$ to the Taylor expansion of f about ζ

$$f(w) = \sum_{s=0}^{m} \frac{f^{(s)}(\zeta)}{s!}(w-\zeta)^s + R_{m+1,\zeta}f(w),$$

taking $\Delta_{m+1,\zeta}\mathcal{L}[f] = \mathcal{L}[R_{m+1,\zeta}f]$ and using the bound $\|R_{m+1,\zeta}f\| \le K_1\|f^{(m+1)}\|$ to set $K = K_1\|\mathcal{L}\|$. □

Proof of Lemma 2.1. We let $a = c_0 - 1/2$ and $b = c_n + 1/2$ and define the operators $\tilde{\mathcal{L}}_\nu, \mathcal{L}_{\nu,j}\colon \mathcal{C}^{m+1}[a,b] \to \Pi_n$, $\nu = 1, 2$, $j \ge 1$ through the following conditions, where $i = 0, \dots, n$ and $j \le n$:

$$\tilde{\mathcal{L}}_1[f](c_i) = f(c_i), \qquad \mathcal{L}_{1,j}[f] = \big(\tilde{\mathcal{L}}_1[f]\big)^{(j)}, \tag{2.6}$$

$$\int_{c_i-1/2}^{c_i+1/2} \tilde{\mathcal{L}}_2[f](x)\,\mathrm{d}x = \int_{c_i-1/2}^{c_i+1/2} f(x)\,\mathrm{d}x, \qquad \mathcal{L}_{2,j}[f] = \big(\tilde{\mathcal{L}}_2[f]\big)^{(j)}. \tag{2.7}$$

The linearity of $\tilde{\mathcal{L}}_\nu$ and $\mathcal{L}_{\nu,j}$ is clear and the continuity can be proven by exploiting the finite set of conditions (2.6), (2.7), e.g., by using Lagrange basis polynomials φ_i (standard ones for point evaluation); namely, if we define

$$\tilde{\mathcal{L}}_1[f] := \sum_{i=0}^{n} f(c_i)\varphi_i,$$

this implies

$$\mathcal{L}_{1,j}[f] = \sum_{i=0}^{n} f(c_i)\varphi_i^{(j)}, \quad \big\|\mathcal{L}_{1,j}[f]\big\| \le \max_{0\le i\le n}(f(c_i)) \sum_{i=0}^{n}\big\|\varphi_i^{(j)}\big\| \le \|f\| \sum_{i=0}^{n}\big\|\varphi_i^{(j)}\big\|.$$

Similar arguments apply to the cell-average case ($\nu = 2$).

With the notation $S_{z,h}(w) := z + wh$, the polynomials (2.5) can be expressed as $p_h = \tilde{\mathcal{L}}[f \circ S_{z,h}] \circ S_{z,h}^{-1}$, which means that

$$p_h(x) = \tilde{\mathcal{L}}[f \circ S_{z,h}]\left(\frac{x-z}{h}\right)$$

and where $\tilde{\mathcal{L}}$ denotes either $\tilde{\mathcal{L}}_1$ or $\tilde{\mathcal{L}}_2$, and $\mathcal{L}_j$ the corresponding operator $\mathcal{L}_{1,j}$ or $\mathcal{L}_{2,j}$. Since $(f \circ S_{z,h})^{(s)}(w) = h^s f^{(s)}(z + wh)$, Lemma 2.2 for $\zeta = 0$ yields

$$p_h^{(j)}(x) = h^{-j}\tilde{\mathcal{L}}[f \circ S_{z,h}]^{(j)}\left(\frac{x-z}{h}\right) = h^{-j}\mathcal{L}_j[f \circ S_{z,h}]\left(\frac{x-z}{h}\right),$$

$$p_h^{(j)}(z + wh) = h^{-j}\mathcal{L}_j[f \circ S_{z,h}](w)$$

$$= h^{-j} \sum_{s=0}^{m} \frac{(f \circ S_{z,h})^{(s)}(0)}{s!} \mathcal{L}_j[w^s] + h^{-j} \Delta_{m+1,0} \mathcal{L}_j[f \circ S_{z,h}]$$
$$= \sum_{s=j}^{m} h^{s-j} \frac{f^{(s)}(z)}{s!} \mathcal{L}_j[w^s] + \mathcal{O}(h^{m+1-j}),$$

since $\tilde{\mathcal{L}}[w^s] = w^s$ for $s \leq n$, therefore $\mathcal{L}_j[w^s] = (\tilde{\mathcal{L}}[w^s])^{(j)} = 0$ for $s < j$, and

$$\Delta_{m+1,0} \mathcal{L}_j[f \circ S_{z,h}] \leq K \left\| (f \circ S_{z,h})^{(m+1)} \right\|_{[a,b]} = K h^{m+1} \left\| f^{(m+1)} \right\|_{S_{z,h}([a,b])}.$$

Therefore, the result follows with

$$b_{s,j}(w) = \frac{\mathcal{L}_j[w^s]}{s!}.$$

Finally, if $n \geq 1$ and $b_{s,1}(w) = 0$, then for the first operator we have the equivalences

$$\tilde{\mathcal{L}}_1[w^s] = \alpha \Leftrightarrow c_i^s = \alpha,\ i = 0, \ldots, n \Leftrightarrow n = 1,\ s \text{ is even and } c_0 = -\alpha^{1/s},\ c_1 = \alpha^{1/s}.$$

For the second operator, we have

$$b_{s,1}(w) = 0 \Leftrightarrow \tilde{\mathcal{L}}_2[w^s] = \alpha$$
$$\Leftrightarrow q_s(c_i) = \left(c_i + \frac{1}{2}\right)^{s+1} - \left(c_i - \frac{1}{2}\right)^{s+1} = (s+1)\alpha, \quad i = 0, \ldots, n,$$

which means that

$$q_s(x) = \left(x + \frac{1}{2}\right)^{s+1} - \left(x - \frac{1}{2}\right)^{s+1} = \sum_{l=0}^{s+1} \binom{s+1}{l} x^{s+1-l} \frac{1}{2^l} (1 - (-1)^l)$$
$$= \sum_{l=0}^{\lfloor s/2 \rfloor} \binom{s+1}{2l+1} \frac{1}{2^{2l}} x^{s-2l}.$$

This implies, by Rolle's theorem, that there exist numbers $\tilde{c}_i \in (c_{i-1}, c_i)$, $i = 1, \ldots, n$ such that $q_s'(\tilde{c}_i) = 0$. But

$$q_s'(x) = \sum_{l=0}^{\lfloor s/2 \rfloor} \binom{s+1}{2l+1} \frac{1}{2^{2l}} (s - 2l) x^{s-2l-1}$$

has only even-degree terms, with strictly positive coefficients, when s is odd (and therefore no roots) and only odd-degree terms, with strictly positive coefficients, when s is even (and therefore 0 as only root). This implies that s is even, $n = 1$ and $\tilde{c}_1 = 0$, which yields $c_0 < \tilde{c}_1 = 0 < c_1$. Since q_s is an even function and strictly increasing in $(0, \infty)$, for even s, $q_s(c_0) = q_s(-c_0) = q_s(c_1)$ implies $c_1 = -c_0$. The converse is clear, since $n = 1$, $c_1 = -c_0$ and even s implies that $q_s(c_1) = q_s(c_0) = \alpha$ and therefore $\tilde{\mathcal{L}}_2[w^s] = \alpha$ and $b_{s,1}(w) = \frac{1}{s!} \tilde{\mathcal{L}}_2[w^s]' = 0$. □

Lemma 2.3. *If $f \in \mathcal{C}^\infty(z - \gamma, z + \gamma)$ for some $\gamma > 0$ and*

$$f^{(s')}(z) = 0 \quad \text{for all } s' < s, \tag{2.8}$$

then

$$e_i(h) := f(z + h/2) - p_i(z + h/2) = \mathcal{O}(h^{\max(r,s)}), \tag{2.9}$$
$$e(h) := f(z + h/2) - q(z + h/2) = \mathcal{O}(h^{\max(r,s)}). \tag{2.10}$$

Proof. We prove the result for the interpolatory case, the cell-average case is similar. Without loss of generality assume $z = 0$. Using the representation of the interpolation error by divided differences (see e.g. [5]), we get

$$\begin{aligned} e_i(h) &= f(x_{1/2}) - p_i(x_{1/2}) \\ &= f[x_{i-r+1}, \dots, x_i, x_{1/2}] \prod_{l=0}^{r-1} (x_{1/2} - x_{i-l}) = \frac{f^{(r)}(\xi)}{r!} h^r \prod_{l=0}^{r-1} \left(\frac{1}{2} - i + l \right), \end{aligned}$$

where $|\xi - z| < \max(r - 1 - i, i)h < rh$. The result follows for $s \le r$. For $s > r$, by (2.8) and Taylor's remainder theorem, we get

$$f^{(r)}(\xi) = \frac{f^{(s)}(\xi_{s,r})}{(s-r)!} (\xi - z)^{s-r} |\xi_{s,r} - z| < |\xi - z|.$$

It follows that

$$\left| e_i(h) \right| \le \max_{|\xi - z| < r h_0} \left| f^{(s)}(\xi) \right| \frac{r^s}{r!(s-r)!} h^s$$

for $0 < h < h_0$ so that $rh_0 < \gamma$. This concludes the proof of (2.9), and (2.10) follows from $\sum_j \omega_j = 1$. □

In order to use the previous results, we consider $x_{i,h} = z + (\alpha + i)h$, with $\alpha \in \mathbb{R}$ fixed and $i \in \mathbb{Q}$, so that, for instance $x_{1/2,h} = z + (\alpha + 1/2)h$. The reconstruction polynomial $p_{r,i}$ associated to the substencil $S_{r,i} = \{x_{-r+1+i}, \dots, x_i\}$ corresponds to p_h in Lemma 2.1 for $n = r - 1$ and

$$c_j = c_{j,i} := \alpha - r + i + 1 + j, \quad j = 0, \dots, r-1. \tag{2.11}$$

Theorem 2.1. *If f has a smooth extremum of maximal order k at z, then the Jiang-Shu smoothness indicator $I_{r,i}$ (see (2.4)) satisfies $I_{r,i} = \bar{\mathcal{O}}(h^{2q})$, where*

$$q = \begin{cases} \min\{l \in \mathbb{N} \colon 2|l, l \ge k, f^{(l+1)}(z) \ne 0\} & \text{for } r = 2 \text{ and } \alpha + i = 1/2, \\ k + 1 & \text{otherwise.} \end{cases}$$

Proof. From Lemma 2.1 applied to $n = r - 1$ and (2.11), we obtain

$$p_{r,i}^{(j)}(z + wh) = \sum_{s=j}^{m} b_{i,s,j}(w) h^{s-j} f^{(s)}(z) + \mathcal{O}(h^{m+1-j}), \tag{2.12}$$

where $b_{i,s,j}$ denotes the function $b_{s,j}$ given by Lemma 2.1 corresponding to the stencil $S_{r,i}$. Notice that the condition $\alpha + i = 1/2$ is equivalent to $c_{0,i} = -c_{1,i}$. We apply (2.12) for $m = q$, where

$$q = \min\big\{\nu \in \mathbb{N} \colon b_{i,\nu,1}(w) f^{(\nu)}(z) \ne 0\big\}.$$

Then by the definition of k and Lemma 2.1 we get for $j \le \min(r - 1, q)$:

$$p_{r,i}^{(j)}(z + wh) = b_{i,q,j}(w) h^{q-j} f^{(q)}(z) + \mathcal{O}(h^{q+1-j}). \tag{2.13}$$

We use the change of variables $x = z + wh$ to get from (2.13) for $j = 1$:

$$\begin{aligned} \int_{x_{-1/2}}^{x_{1/2}} \big(p_{r,i}^{(1)}(x)\big)^2 \mathrm{d}x &= h \int_{\alpha}^{\alpha+1} \big(b_{i,q,1}(w) h^{q-1} f^{(q)}(z) + \mathcal{O}(h^q)\big)^2 \mathrm{d}w \\ &= h^{2q-1} \mu_{i,1} + \mathcal{O}(h^{2q}), \quad \mu_{i,1} := f^{(q)}(z)^2 \int_{\alpha}^{\alpha+1} b_{i,q,1}(w)^2 \mathrm{d}w > 0. \end{aligned}$$

For $1 < j \leq q$ (and, a fortiori, $r > 2$, therefore $q = k + 1$) we obtain

$$\int_{x_{-1/2}}^{x_{1/2}} \big(p_{r,i}^{(j)}(x)\big)^2 \,\mathrm{d}x = h \int_{x_{-1/2}}^{x_{1/2}} \big(b_{i,q,j}(w) h^{q-j} f^{(q)}(z) + \mathcal{O}(h^{q+1-j})\big)^2 \mathrm{d}w$$

$$= \mu_{i,j} h^{2(q-j)+1} + \mathcal{O}(h^{2(q-j)+2}), \quad \mu_{i,j} := f^{(q)}(z)^2 \int_{\alpha}^{\alpha+1} \big(b_{i,q,j}(w)\big)^2 \mathrm{d}w \geq 0.$$

For $j > q$ we get

$$\int_{x_{-1/2}}^{x_{1/2}} \big(p_{r,i}^{(j)}(x)\big)^2 \,\mathrm{d}x = h \int_{x_{-1/2}}^{x_{1/2}} \big(b_{i,j,j}(w) f^{(j)}(z) + \mathcal{O}(h)\big)^2 \mathrm{d}w = \mathcal{O}(h).$$

The proof is complete after substitution of these terms into (2.4):

$$\begin{aligned} I_{r,i} &= \sum_{j=1}^{r-1} h^{2j-1} \int_{x_{-1/2}}^{x_{1/2}} \big(p_{r,i}^{(j)}(x)\big)^2 \,\mathrm{d}x \\ &= \sum_{j=1}^{\min(q,r-1)} h^{2j-1} \big(\mu_{i,j} h^{2(q-j)+1} + \mathcal{O}(h^{2(q-j)+2})\big) + \sum_{j=\min(q,r-1)+1}^{r-1} h^{2j-1} \mathcal{O}(h) \\ &= h^{2q} \sum_{j=1}^{\min(q,r-1)} \mu_{i,j} + \mathcal{O}(h^{2q+1}), \end{aligned}$$

taking into account that $\sum_{j=1}^{\min(q,r-1)} \mu_{i,j} > 0$. □

3. Design of WENO weights

We next proceed to the definition of our modified scheme. The goal is to define weights in a way such that the resulting scheme has the maximal order of accuracy $2r-1$, for $r > 2$. For simplicity, assume a right-biased reconstruction (the procedure for left-biased reconstructions is analogous, nevertheless we will state the differences along the construction of the scheme where required).

From this point on, two alternative approaches are presented in which the optimal accuracy is unconditionally attained in presence of smooth data. Approach 1 relies on the inclusion of an additional node to the stencil. Approach 2 is based on a modification of a parameter of Approach 1, in a way so that no additional node is required and the optimal accuracy is still attained.

In both approaches we will work under the assumption $r > 2$, namely, considering WENO reconstructions of order greater than 3.

3.1. Approach 1: Optimal WENO reconstructions with an additional node.

Weights ω_i in WENO schemes are defined through a relation of the type

$$\omega_i = \frac{\alpha_i}{\alpha_0 + \cdots + \alpha_{r-1}}, \quad 0 \leq i \leq r-1, \tag{3.1}$$

so that the sum of the weights is 1.

In this section the quantities $\alpha_0, \ldots, \alpha_{r-1}$ are given by

$$\alpha_i = c_i \left(1 + \frac{d(\varepsilon, h)^{s_1}}{I_i^{s_1} + \varepsilon}\right)^{s_2}, \quad 0 \leq i \leq r-1, \tag{3.2}$$

for some $s_1, s_2 > 0$, $c_i > 0$ with $\sum_{i=0}^{r-1} c_i = 1$ and $d(\varepsilon, h)$ to be defined below. This approach is related to the one of Yamaleev and Carpenter [21]. The ultimate goal

is to obtain the maximal order of convergence $2r-1$, regardless of the presence of neighboring extrema [1, 2, 11, 21] and without using the small number $\varepsilon > 0$ that ensures the strict positivity of the denominators. Contrarily to other approaches [1, 2], our design does *not* rely on a functional relation between ε and h. Although this parameter is necessary if conditionals are to be avoided (which in turn may be necessary to avoid divisions by zero), our arguments will show that ε can be neglected in the asymptotical analysis of the order with respect to h.

3.1.1. *Motivation.* In the classical WENO order-enhancing argument in case of sufficient smoothness (see the proof of Lemma 3.2 below), for a function with an extremum of maximal order k, the order of the reconstruction is $\mathrm{ord}_{\max} = \min(\max(2r-1, k+1), s+\max(r, k+1))$, where $\max(2r-1, k+1)$, resp. $\max(r, k+1)$, are the orders of the reconstructions with $p_{2r-1,r-1}$, resp. $p_{r,i}$ (see Lemma 2.3) and $s \geq 0$ satisfies $\omega_i = c_i + \mathcal{O}(h^s)$.

A natural way of defining α_i in (3.2) could be to define $d(\varepsilon, h)$ independently of ε through the following squared undivided difference related to the $2r-1$ consecutive values $(f_{-r+1}, \ldots, f_{r-1})$:

$$d(\varepsilon, h) = d_{r,2r-2}(h) := \left(\sum_{j=-r+1}^{r-1} (-1)^{j+r-1} \binom{2r-2}{j+r-1} f_j \right)^2. \tag{3.3}$$

With this definition of $d(\varepsilon, h)$ it holds that $d(0, h) = \mathcal{O}(h^{4r-4})$ which, together with $I_j = \bar{\mathcal{O}}(h^{2k+2})$, gives from Theorem 2.1:

$$1 + \frac{d(0,h)^{s_1}}{I_i^{s_1}} = 1 + \mathcal{O}(h^{4r-2k-6}).$$

The order-enhancing argument in this context requires $\frac{d(0,h)^{s_1}}{I_i^{s_1}} \to 0$, which is not met if $k = 2r-3$. On the other hand, if $k \geq 2r-2$, then $\mathrm{ord}_{\max} \geq 2r-1$. So there remains an order loss gap at $k = 2r-3$.

3.1.2. *Theoretical analysis.* Based on the previous analysis, in order to design a reconstruction procedure that attains the optimal order of accuracy, we include an additional node to the stencil, namely, we sample an additional value of f. In our case, the most convenient value for that purpose is $f_r = f(x_r)$ with $x_r = x_{r-1}+h$, since in that case, the resulting extended stencil, $\bar{S} = \{f_{-r+1}, \ldots, f_{r-1}, f_r\}$ is centered with respect to $x_{1/2}$. For the case of a left-biased reconstruction at $x_{-1/2}$ the added value would be $f_{-r} = f(x_{-r})$.

After the above preliminaries, which for now are all the elements used by Jiang-Shu in their original WENO scheme, we can proceed to define our additional items and modified weights. Now, we also consider the square of the undivided difference of degree $2r-1$, associated to the whole extended stencil $\bar{S}$. We use the notation $d_1 := d_{r,2r-2}$ as given by (3.3), and define

$$d_2 := d_{r,2r-1} := \left(\sum_{j=-r+1}^{r} (-1)^{j+r} \binom{2r-1}{j+r-1} f_j \right)^2, \tag{3.4}$$

$$\bar{d}(\varepsilon, h) := \frac{d_1^{s_1} d_2^{s_1}}{d_1^{s_1} + d_2^{s_1} + \varepsilon}. \tag{3.5}$$

The function $\bar{d}(\varepsilon,h)$ is related to the harmonic mean of $d_1^{s_1}$ and $d_2^{s_1}$ (for $\varepsilon=0$ is actually half the harmonic mean) and its definition is motivated by the following fact:

$$\bar{d}(0,h)=\begin{cases}\mathcal{O}(h^{4r-2}) & \text{if } \bar{S} \text{ is smooth,}\\ \mathcal{O}(h^{4r-4}) & \text{if } S \text{ is smooth, but a discontinuity crosses } \bar{S},\\ \bar{\mathcal{O}}(1) & \text{if a discontinuity crosses } S.\end{cases} \tag{3.6}$$

We perform now an asymptotical analysis of the weights for $\varepsilon\to 0$. To be precise, for $0\le i\le r-1$, $\varepsilon>0$, we denote:

$$\begin{aligned}\bar{\beta}_i(\varepsilon,h) &:= 1+\frac{\bar{d}(\varepsilon,h)}{I_i^{s_1}(h)+\varepsilon}=\frac{(d_1^{s_1}+d_2^{s_1}+\varepsilon)\prod_k(I_k^{s_1}+\varepsilon)+d_1^{s_1}d_2^{s_1}\prod_{k\neq i}(I_k^{s_1}+\varepsilon)}{(d_1^{s_1}+d_2^{s_1}+\varepsilon)\prod_k(I_k^{s_1}+\varepsilon)}\\ &=\frac{\mu_i(\varepsilon,h)}{(d_1^{s_1}+d_2^{s_1}+\varepsilon)\prod_k(I_k^{s_1}+\varepsilon)},\\ \bar{\omega}_i(\varepsilon,h) &:= \frac{c_i\bar{\beta}_i(\varepsilon,h)^{s_2}}{\sum_{j=0}^{r-1}c_j\bar{\beta}_j(\varepsilon,h)^{s_2}}=\frac{c_i\mu_i(\varepsilon,h)^{s_2}}{\sum_{j=0}^{r-1}c_j\mu_j(\varepsilon,h)^{s_2}},\end{aligned}$$

with $\Pi_k\equiv\Pi_{k=0}^{r-1}$, $\Pi_{k\neq i}\equiv\Pi_{k=0,k\neq i}^{r-1}$ and where we have denoted

$$\mu_i(\varepsilon,h)=(d_1^{s_1}+d_2^{s_1}+\varepsilon)\prod_k(I_k^{s_1}+\varepsilon)+d_1^{s_1}d_2^{s_1}\prod_{k\neq i}(I_k^{s_1}+\varepsilon).$$

We define the index set $\mathcal{P}_s:=\{J\subseteq\{0,\dots,r-1\}\colon |J|=s\}$. Then, since

$$\prod_{k=0}^{r-1}(I_k^{s_1}+\varepsilon)=\sum_{l=0}^{r}\varepsilon^l\sum_{J\in\mathcal{P}_{r-l}}\prod_{k\in J}I_k^{s_1},\quad \prod_{\substack{k=0\\k\neq i}}^{r-1}(I_k^{s_1}+\varepsilon)\quad=\sum_{l=0}^{r-1}\varepsilon^l\sum_{\substack{J\in P_{r-l}\\ i\notin J}}\prod_{k\in J}I_k^{s_1},$$

it follows that

$$\mu_i(\varepsilon,h)=\varepsilon^{r+1}+\sum_{l=0}^{r}\mu_{i,l}(h)\varepsilon^l,$$

where

$$\begin{aligned}\mu_{i,r} &= \sum_{k=0}^{r-1}I_k^{s_1}+d_1^{s_1}+d_2^{s_1},\\ \mu_{i,l} &= \sum_{J\in P_{r-l+1}}\prod_{k\in J}I_k^{s_1}+(d_1^{s_1}+d_2^{s_1})\sum_{J\in P_{r-l}}\prod_{k\in J}I_k^{s_1}+d_1^{s_1}d_2^{s_1}\sum_{\substack{J\in P_{r-l}\\ i\notin J}}\prod_{k\in J}I_k^{s_1},\\ &\quad l=1,\dots,r-1,\\ \mu_{i,0} &= (d_1^{s_1}+d_2^{s_1})\prod_{k=0}^{r-1}I_k^{s_1}+d_1^{s_1}d_2^{s_1}\prod_{\substack{k=0\\k\neq i}}^{r-1}I_k^{s_1}.\end{aligned}$$

For a fixed h, it follows:

(1) If $\bar{d}(\varepsilon,h)=0$, then $\bar{\beta}_i(\varepsilon,h)=1$ for all i and ε, and $\bar{\omega}_i(\varepsilon,h)=c_i$ for any $\varepsilon>0$.
(2) Assume that $\bar{d}(\varepsilon,h)\neq 0$) and assume that there exists some k with $I_k=0$. (This excludes that $I_k=0$ for *all* k, since in that case we would have $\bar{d}(\varepsilon,h)=0$). Then take $J_*:=\{k\colon I_k\neq 0\}$ with $|J_*|=r-s$, $r>s\ge 1$. If

$l < s-1$ then $\mu_{i,l} = 0$ for all i, since any J with $|J| = r-l > r-s+1$ intersects $\{0,\dots,r-1\} \setminus J$, so that the products are 0. By a similar argument, if $i \notin J_*$ (i.e., $I_i = 0$),

$$\mu_{i,s-1}(h) = d_1^{s_1} d_2^{s_1} \prod_{k\in J_*} I_k^{s_1} \neq 0, \tag{3.7}$$

and if $i \in J_*$, then $\mu_{i,s-1} = 0$. Therefore, for $i \notin J_*$,

$$\begin{aligned}
\bar\omega_i(\varepsilon,h) &= \frac{\big(c_i d_1^{s_1} d_2^{s_1} \prod_{k\in J_*} I_k^{s_1}\varepsilon^{s-1} + \mathcal{O}(\varepsilon^s)\big)^{s_2}}{(\sum_{j\notin J_*} c_j d_1^{s_1} d_2^{s_1} \prod_{k\in J_*} I_k^{s_1}\varepsilon^{s-1} + \mathcal{O}(\varepsilon^s))^{s_2}} \\
&= \frac{(c_i + \mathcal{O}(\varepsilon))^{s_2}}{\big(\sum_{j\notin J_*} c_j + \mathcal{O}(\varepsilon)\big)^{s_2}} \\
&= \frac{c_i^{s_2}}{\big(\sum_{j\notin J_*} c_j\big)^{s_2}} + \mathcal{O}(\varepsilon^{s_2}) \overset{\varepsilon\to 0}{\longrightarrow} \frac{c_i^{s_2}}{\big(\sum_{j\notin J_*} c_j\big)^{s_2}}
\end{aligned}$$

On the other hand, if $i \in J_*$,

$$\begin{aligned}
\bar\omega_i(\varepsilon,h) &= \frac{\mathcal{O}(\varepsilon^s)^{s_2}}{(\sum_{j\notin J_*} c_j d_1^{s_1} d_2^{s_1} \prod_{k\in J_*} I_k^{s_1}\varepsilon^{s-1} + \mathcal{O}(\varepsilon^s))^{s_2}} \\
&= \mathcal{O}(\varepsilon^{s_2}) \overset{\varepsilon\to 0}{\longrightarrow} 0
\end{aligned}$$

(3) If $d \neq 0$ and $I_k \neq 0$ for all k, then

$$\begin{aligned}
&\bar\omega_i(\varepsilon,h) \\
&= \frac{c_i\big(\mu_{i,0} + \mathcal{O}(\varepsilon)\big)^{s_2}}{\sum_j c_j\big(\mu_{j,0} + \mathcal{O}(\varepsilon)\big)^{s_2}} = \frac{c_i \mu_{i,0}^{s_2}}{\sum_j c_j \mu_{j,0}^{s_2}} + \mathcal{O}(\varepsilon^{s_2}) \\
&= \frac{c_i\big((d_1^{s_1} + d_2^{s_1}) \prod_{k=0}^{r-1} I_k^{s_1} + d_1^{s_1} d_2^{s_1} \prod_{k=0, i\notin J}^{r-1} I_k^{s_1}\big)^{s_2}}{\sum_j c_j\big((d_1^{s_1} + d_2^{s_1}) \prod_{k=0}^{r-1} I_k^{s_1} + d_1^{s_1} d_2^{s_1} \prod_{k=0, j\notin J}^{r-1} I_k^{s_1}\big)^{s_2}} + \mathcal{O}(\varepsilon^{s_2})
\end{aligned}$$

Dividing both terms of this fraction by $(d_1^{s_1} + d_2^{s_1}) \prod_{k=0}^{r-1} I_k^{s_1} \neq 0$, we get

$$\bar\omega_i(\varepsilon,h) = \frac{c_i\big(1 + \bar d(0,h)/I_i^{s_1}\big)^{s_2}}{\sum_j c_j\big(1 + \bar d(0,h)/I_j^{s_1}\big)^{s_2}} + \mathcal{O}(\varepsilon^{s_2}). \tag{3.8}$$

Defining $\omega_i(h) = \lim_{\varepsilon\to 0} \bar\omega_i(\varepsilon,h) := \bar\omega_i(0,h)$, we have just proved that

$$\bar\omega_i(\varepsilon,h) = \omega_i(h) + \mathcal{O}(\varepsilon^{s_2}). \tag{3.9}$$

We also define

$$\bar q[\varepsilon,h](x) := \sum_{i=0}^{r-1} \bar\omega_i(\varepsilon,h) p_i(x), \quad q[h](x) := \sum_{i=0}^{r-1} \omega_i(h) p_i(x).$$

Theorem 3.1. *If f is $2r-1$ times continuously differentiable in a neighborhood of z then*

$$e(\varepsilon,h) = f(x_{1/2}) - \bar q(x_{1/2}) = \mathcal{O}(h^{2r-1}) + \mathcal{O}(\varepsilon^{s_2}). \tag{3.10}$$

The first step to prove Theorem 3.1 is to use the previous results to bound

$$\begin{aligned} e(\varepsilon,h) &:= f(x_{1/2}) - \bar{q}(x_{1/2}) = f(x_{1/2}) - q(x_{1/2}) + q(x_{1/2}) - \bar{q}(x_{1/2}) \\ &= e(h) + \sum_{i=0}^{r-1} \big(\omega_i(h) - \bar{\omega}_i(\varepsilon,h)\big) p_i(x_{1/2}) = e(h) + \sum_{i=0}^{r-1} \mathcal{O}(\varepsilon^{s_2})\mathcal{O}(1) \\ &= e(h) + \mathcal{O}(\varepsilon^{s_2}). \end{aligned} \tag{3.11}$$

The proof of Theorem 3.1 will be complete when we prove that

$$e(h) = \mathcal{O}(h^{2r-1}), \tag{3.12}$$

which we will achieve through several intermediate results.

To analyze the behavior of $\omega_i(h)$ we distinguish the two mutually exclusive cases:

(1) There exist $\{h_n\} = \{h_n\}_{n\in\mathbb{N}} > 0$ and $\{j_n\} = \{j_n\}_{n\in\mathbb{N}} \subset \{0,\dots,r-1\}$ with $\lim_{n\to\infty} h_n = 0$ and $I_{j_n}(h_n) = 0$.
(2) There exists $h_0 > 0$ such that $I_j(h) \neq 0$ for all j and $0 < h < h_0$.

Lemma 3.1. *If $f \in \mathcal{C}^s(z-\gamma, z+\gamma)$, $\gamma > 0$, and Case (1) is in effect, then condition* (2.8) *of Lemma 2.3 is satisfied.*

Proof of Lemma 3.1. From the definition of $I_j(h)$ it follows that p_j' is zero on an interval of positive length, so that p_j' is overall zero, i.e., f is constant at the points $\{z + (\alpha + j)h_n\}$, $j = -r+1+j_n, \dots, j_n$. Therefore there exists $\{z_n\}_{n\in\mathbb{N}}$ with $z_n \to z$ with $f'(z_n) = 0$. A recursive use of Rolle's theorem and continuity yields (2.8). □

If we define

$$k := \inf\big\{l \in \mathbb{N}\colon f^{(l)}(z) \neq 0\big\} - 1, \tag{3.13}$$

then Lemma 2.3 yields that $e(h) = \mathcal{O}(h^{k+1})$, so, in order to prove (3.12) we may assume

$$k < 2r - 2. \tag{3.14}$$

In this case Lemma 3.1 implies the existence of $h_0 > 0$ such that $I_k(h) > 0$ for $k \in \{0,\dots,r-1\}$, $0 < h < h_0$ (i.e., Case (2) is in effect), therefore (3.8) yields

$$\omega_i(h) = \frac{c_i}{\sum_j c_j (\beta_j(0,h)/\beta_i(0,h))^{s_2}}. \tag{3.15}$$

From now on we use the simplified notations $d = \bar{d}(0,h)$ and $\beta_i = \bar{\beta}_i(0,h) = 1 + \frac{d}{I_i^{s_1}}$

Lemma 3.2. *Assume $f \in \mathcal{C}^{2r-1}(z-\varepsilon, z+\varepsilon)$, $r \geq 3$, and that k defined by* (3.13) *satisfies* (3.14). *Then* (3.12) *holds for any $s_1, s_2 \in \mathbb{N}$.*

Proof of Lemma 3.2. As mentioned above, Case (2) is in effect, i.e., we may assume that there exists $h_0 > 0$ such that $I_k(h) > 0$ for $k \in \{0,\dots,r-1\}, 0 < h < h_0$. We have by Theorem 2.1 and (3.6) $I_j = \bar{\mathcal{O}}(h^{2(k+1)})$ and $d = \mathcal{O}(h^{4r-2})$. We analyze (3.15) with these estimates:

$$\left|\frac{\beta_j}{\beta_i} - 1\right| = \frac{d}{1 + d/I_i^{s_1}} \frac{|I_i^{s_1} - I_j^{s_1}|}{I_i^{s_1} I_j^{s_1}} \leq \frac{d(I_i^{s_1} + I_j^{s_1})}{I_i^{s_1} I_j^{s_1}} = \frac{O(h^{s_1(4r-2)})\mathcal{O}(h^{(2(k+1))s_1})}{\bar{\mathcal{O}}(h^{4s_1(k+1)})},$$

which means that

$$\beta_j/\beta_i = 1 + \mathcal{O}(h^{\zeta}), \quad \zeta := 2s_1(2r-2-k). \tag{3.16}$$

It follows from (3.15) that

$$\omega_i(h) = \frac{c_i}{\sum_j c_j \left(1 + \mathcal{O}(h^\zeta)\right)^{s_2}} = \frac{c_i}{\sum_j c_j (1 + \mathcal{O}(h^\zeta))} = c_i + \mathcal{O}(h^\zeta). \tag{3.17}$$

By using that $\sum_i \omega_i = \sum_i c_i$, $f(z + h/2) - p_{2r-1,r-1}(z + h/2) = \mathcal{O}(h^{2r-1})$, and Lemma 2.3, we obtain from (3.17)

$$\begin{aligned} e(h) = \sum_{i=0}^{r-1} \omega_i e_i(h) &= \sum_{i=0}^{r-1} \big(c_i + \mathcal{O}(h^\zeta)\big) \left(f\left(z + \frac{h}{2}\right) - p_i\left(z + \frac{h}{2}\right)\right) \\ &= \sum_{i=0}^{r-1} c_i \left(f\left(z + \frac{h}{2}\right) - p_i\left(z + \frac{h}{2}\right)\right) + \sum_{i=0}^{r-1} \mathcal{O}(h^\zeta)\mathcal{O}(h^{\max(r,k+1)}) \\ &= f(z + h/2) - p(z + h/2) + \mathcal{O}(h^{\zeta+\max(r,k+1)}) \\ &= \mathcal{O}(h^{2r-1}) + \mathcal{O}(h^{\zeta+\max(r,k+1)}) = \mathcal{O}(h^{\min(2r-1,\zeta+\max(r,k+1))}). \end{aligned}$$

Utilizing the definition of ζ in (3.16), one can easily verify that

$$\zeta + \max(r, k + 1) \geq 2r - 1 \quad \forall k \leq 2r - 3,\ s_1 \geq 1.$$

This completes the proof. □

Proof of Theorem 3.1. Lemmas 3.1 and 3.2 establish (3.12) in the alternative cases (1) and (2) identified above. □

Remark 3.1. *All these precautions on the possibility of having smoothness indicators that vanish asymptotically are not void, since the function*

$$f(x) = \begin{cases} \mathrm{e}^{-1/x^2} & \text{for } x > 0, \\ 0 & \text{for } x \leq 0 \end{cases}$$

satisfies $f \in \mathcal{C}^\infty(\mathbb{R})$ and $f^{(n)}(0) = 0$ for all $n \in \mathbb{N}$, therefore, for $z = 0$, it follows that $I_0(h) = 0$ for all $h > 0$.

Theorem 3.2. *If f has a discontinuity at z and is r times continuously differentiable in $(z - \delta_0, z) \cup (z, z + \delta_0)$, for some $\delta_0 > 0$, then*

$$e(\varepsilon, h) = f(x_{1/2}) - \bar{q}(x_{1/2}) = \mathcal{O}(h^r) + \mathcal{O}(\varepsilon^{s_2}). \tag{3.18}$$

Proof. Let $J_r := \{0 \leq j \leq r - 1 : S_{r,j}$ does not cross $z\}$. Note that $J_r \neq \varnothing$, since the stencils $S_{r,j}$ do not overlap for $0 \leq j \leq r - 1$.

Then, if $i \notin J_r$, (3.15) reads

$$\frac{\beta_j}{\beta_i} = \frac{1 + d/I_j^{s_1}}{1 + d/I_i^{s_1}} = \frac{1 + d/I_j^{s_1}}{1 + d/I_i^{s_1}} = \frac{1 + \bar{\mathcal{O}}(1)/I_j^{s_1}}{1 + \bar{\mathcal{O}}(1)/\bar{\mathcal{O}}(1)} = \bar{\mathcal{O}}(1)\left(1 + \frac{\bar{\mathcal{O}}(1)}{I_j^{s_1}}\right).$$

Therefore,

$$\frac{\beta_j}{\beta_i} = \begin{cases} \bar{\mathcal{O}}(1) & \text{if } j \notin J_r, \\ \bar{\mathcal{O}}(h^{-2(k+1)s_1}) & \text{if } j \in J_r. \end{cases}$$

Similarly, if $i \in J_r$,

$$\frac{\beta_j}{\beta_i} = \frac{1 + d/I_j^{s_1}}{1 + d/I_i^{s_1}} = \frac{1 + \bar{\mathcal{O}}(1)/I_j^{s_1}}{1 + \bar{\mathcal{O}}(1)/\bar{\mathcal{O}}(h^{2(k+1)s_1})} = \bar{\mathcal{O}}(h^{2(k+1)s_1})\left(1 + \frac{\bar{\mathcal{O}}(1)}{I_j^{s_1}}\right).$$

Hence,

$$\frac{\beta_j}{\beta_i} = \begin{cases} \bar{\mathcal{O}}(h^{2(k+1)s_1}) & \text{if } j \notin J_r, \\ \bar{\mathcal{O}}(1) & \text{if } j \in J_r. \end{cases}$$

Therefore, the denominator of (3.15) reads

$$\sum_{j=0}^{r-1} c_j \left(\frac{\beta_j}{\beta_i}\right)^{s_2} = \sum_{j\in J_r} c_j \left(\frac{\beta_j}{\beta_i}\right)^{s_2} + \sum_{j\notin J_r} c_j \left(\frac{\beta_j}{\beta_i}\right)^{s_2}.$$

Thus, if $i \notin J_r$

$$\sum_{j=0}^{r-1} c_j \left(\frac{\beta_j}{\beta_i}\right)^{s_2} = \sum_{j\notin J_r} \bar{\mathcal{O}}(h^{-2(k+1)s_1 s_2}) + \sum_{j\in J_r} \bar{\mathcal{O}}(1) = \bar{\mathcal{O}}(h^{-2(k+1)s_1 s_2}).$$

Therefore, if $i \notin J_r$ then

$$\omega_i = \mathcal{O}(h^{2(k+1)s_1 s_2}).$$

On the other hand, if $i \in J_r$

$$\sum_{j=0}^{r-1} c_j \left(\frac{\beta_j}{\beta_i}\right)^{s_2} = \sum_{j\notin J_r} \bar{\mathcal{O}}(1) + \sum_{j\in J_r} \bar{\mathcal{O}}(h^{2(k+1)s_1}) = \bar{\mathcal{O}}(1).$$

Hence, if $i \in J_r$, then $\omega_i = \mathcal{O}(1)$. Therefore, following a similar accuracy analysis as in the smooth case, the result holds by imposing $2s_1 s_2 \geq r$. □

Remark 3.2. *During the whole theoretical analysis we have considered the definition of $\bar{d}$ in* (3.5).

We point out that it is straightforward to check that the choice $\bar{d} = d_2 = d_{r,2r-1}$, with $d_{r,2r-1}$ defined in (3.4) *also attains the required properties both in case of smoothness and discontinuity in the data, except the case in which a discontinuity is located between x_{r-1} and x_r. In this case, one ideally would like to obtain optimal accuracy since the reconstructions are performed in the region $[x_{-r+1}, x_{r-1}]$, which contains smooth data. This is satisfied by $\bar{d}$ defined as in* (3.5)*, but not for $\bar{d} = d_{r,2r-1}$, which in this case is $\bar{\mathcal{O}}(1)$. However, we will also use the choice $\bar{d} = d_2 = d_{r,2r-1}$ in the numerical experiments for the sake of comparison.*

3.1.3. *Summary of the algorithm.* For the ease of reference we summarize here the steps of the new WENO reconstruction.

Input: $\bar{S} = \{f_{-r+1}, \ldots, f_r\}$, with $f_i = \mathcal{L}_m[f](x_i)$, and $\varepsilon > 0$.

(1) Compute $p_{r,i} = \mathcal{I}_m(x_{-r+1+i}, \ldots, x_i; f_{-r+1+i}, \ldots, f_i)$, $0 \leq i \leq r-1$, the corresponding interpolating polynomials, computed assuming the data as pointwise data for $m = 1$ and as cell averages of a certain unknown function if $m = 2$, with $\mathcal{I}_m$ the associated interpolating operator, at $x = x_{1/2}$.

(2) Compute the Jiang-Shu smoothness indicators:

$$I_{r,i} = \sum_{l=1}^{r-1} \int_{x_{-1/2}}^{x_{1/2}} h^{2l-1} (p_{r,i}^{(l)}(x))^2 \, \mathrm{d}x, \quad 0 \leq i \leq r-1.$$

(3) Obtain the squares of the corresponding undivided differences of order $2r-2$ and $2r-1$:

$$d_{r,2r-2} = \left(\sum_{j=-r+1}^{r-1} (-1)^{j+r-1} \binom{2r-1}{j+r-1} f_j \right)^2,$$

$$d_{r,2r-1} = \left(\sum_{j=-r+1}^{r} (-1)^{j+r} \binom{2r}{j+r-1} f_j \right)^2 .$$

(4) Compute the harmonic mean of the aforementioned terms:

$$\bar{d}(\varepsilon, h) = \frac{d_{r,2r-2} d_{r,2r-1}}{d_{r,2r-2} + d_{r,2r-1} + \varepsilon}. \tag{3.19}$$

(5) Compute the terms

$$\alpha_{r,i} = c_{r,i} \left(1 + \frac{\bar{d}(\varepsilon, h)^{s_1}}{I_{r,i}^{s_1} + \varepsilon} \right)^{s_2}, \quad 0 \le i \le r-1$$

with $c_{r,i}$ the ideal linear weights, for some s_1, s_2 chosen by the user such that $s_1 \ge 1$ and $s_2 \ge \frac{r}{2s_1}$.

(6) Generate the WENO weights:

$$\omega_{r,i} = \frac{\alpha_{r,i}}{\alpha_{r,0} + \cdots + \alpha_{r,r-1}}.$$

(7) Obtain the OWENO reconstruction at $x_{1/2}$:

$$q_r(x_{1/2}) = \sum_{i=0}^{r-1} \omega_{r,i} p_{r,i}(x_{1/2}). \tag{3.20}$$

Output: $q_r(x_{1/2})$.

3.1.4. *Application to conservation laws.* The optimal WENO interpolators presented along this section have been proven to attain unconditionally the optimal order, at the cost of adding an additional node to the stencil used for the reconstruction. We next describe the procedure to apply them to semi-discrete finite-difference conservative schemes for hyperbolic conservation laws. We here limit the discussion of the initial value problem (1.1), (1.2) to $d = 1$ space dimension, for which we write x instead of $\boldsymbol{x}$ and drop the index $i = 1$. The methods under consideration correspond to a spatial discretization of (1.1) [18] in the form

$$\boldsymbol{u}_t(x_j, t^n) + \frac{\hat{\boldsymbol{f}}_{j+1/2} - \hat{\boldsymbol{f}}_{j-1/2}}{h} = 0$$

for approximations $\hat{\boldsymbol{f}}_{j+1/2} \approx \boldsymbol{f}(\boldsymbol{u}(x_{j+1/2}, t^n))$, which is further advanced in time on a discretization of the time variable $t^n = nk$, $k = T/N$, $N \in \mathbb{N}$, by means of an ODE solver, so as to approximate the value $u(x, T)$.

The application of the WENO scheme to solve hyperbolic conservation laws amounts to approximate the values $\hat{\boldsymbol{f}}_{j+\frac{1}{2}}$ by means of the WENO reconstruction described in this Section, i.e., either the value $q_r(x_{1/2})$ in (3.20) or the value $q_r(x_{-1/2})$ which is computed analogously. The choice of one or another depends on the direction of movement of the *characteristics* of the system, which in turn depends on the sign of the eigenvalues of the Jacobian matrix of the conservation law, in a process called *upwinding* (see [15]). In general, to obtain a $(2r-1)$-th order accurate upwind reconstruction of the flux at a given cell interface $x_{j+1/2}$, both left-biased and right-biased reconstructions are required for different flux components, so that one uses the information of the $2r$-point stencil $S = \{\boldsymbol{f}_{j-r+1}, \ldots, \boldsymbol{f}_{j+r}\}$ with

$$\boldsymbol{f}_i \approx \frac{1}{h} \int_{x_{i-1/2}}^{x_{i+1/2}} \boldsymbol{f}\big(\boldsymbol{u}(x)\big)\, \mathrm{d}x$$

(for the sake of simplicity, we drop the time dependence). For instance, if all characteristics move to the right, upwinding imposes a left-biased reconstruction, and one would just take the substencil of S defined by $S' = \{\boldsymbol{f}_{j-r+1}, \dots, \boldsymbol{f}_{j+r-1}\}$ to obtain $\hat{\boldsymbol{f}}_{j+1/2} \approx \boldsymbol{f}(\boldsymbol{u}(x_{j+1/2}))$. The only difference in this case is that we instead pass to the interpolation routine the whole stencil S, that is, adding the rightmost node $\boldsymbol{f}_{j+r}$ to S'. On the other hand, if characteristics move in both directions then a *flux-splitting* algorithm is required, so that the reconstruction acts on *characteristic fluxes* $\boldsymbol{f}_i^{\pm}$ obtained from $\boldsymbol{f}_i$ by transformations that depend on the direction of movement. The *transformed stencils* are $S^+ = \{\boldsymbol{f}^+_{j-r+1}, \dots, \boldsymbol{f}^+_{j+r-1}\}$ and $S^- = \{\boldsymbol{f}^-_{j-r+2}, \dots, \boldsymbol{f}^-_{j+r}\}$ with $\boldsymbol{f}_i^{\pm} = \boldsymbol{G}_{\pm}(\boldsymbol{f}_i)$ for a pair of invertible functions $\boldsymbol{G}_{\pm}$, and that the corresponding left- and right-biased reconstructions are $\hat{\boldsymbol{f}}^+_{j+1/2}$ and $\hat{\boldsymbol{f}}^-_{j+1/2}$, respectively, so that the reconstruction is given by

$$\hat{\boldsymbol{f}}_{j+1/2} = \frac{1}{2}\Big(\boldsymbol{G}_+^{-1}(\hat{\boldsymbol{f}}^+_{j+1/2}) + \boldsymbol{G}_-^{-1}(\hat{\boldsymbol{f}}^-_{j+1/2})\Big).$$

In this case one must also compute the corresponding transformation of each missing node, that is, consider instead $\bar{S}^+ = \{\boldsymbol{f}^+_{j-r+1}, \dots, \boldsymbol{f}^+_{j+r}\}$ with $\boldsymbol{f}^+_{j+r} = \boldsymbol{G}_+(\boldsymbol{f}_{j+r})$, and $\bar{S}^- = \{\boldsymbol{f}^-_{j-r+1}, \dots, \boldsymbol{f}^-_{j+r}\}$ with $\boldsymbol{f}^-_{j-r+1} = \boldsymbol{G}_-(\boldsymbol{f}_{j-r+1})$, and pass these extended stencils to the corresponding left and right-biased reconstruction routines, respectively.

3.2. **Approach 2: Optimal WENO schemes of order using only the original nodes.** In this section we describe a way to obtain a reconstruction that keeps the optimal order under the same circumstances as in Section 3.1, through a different definition of $\bar{d}(\varepsilon, h)$ that does not require an additional node. For clarity we state here the construction of the WENO weights and move the complete theoretical justification of the accuracy of the method to Appendix A.

Let $p_h(x)$ be the interpolating polynomial associated to the stencil S. By Lemma A.3 (see Appendix A for further details) with $n = 2r-2$, $z = 0$ and $s = 2r-4$, the polynomial $P_h(w) := p_h(wh)$ satisfies

$$P_h^{(2r-4)}(w) = \sum_{j=0}^{2} L_j(f_{-r+1,h}, \dots, f_{r-1,h}) w^j,$$

where $L_j : \mathbb{R}^{2r-1} \to \mathbb{R}$, $j = 0, \dots 2$ are linear functions. Now, by Theorem A.1 with $n = 2r-2$, the expression

$$D_r := B_h^2 - 4A_hC_h \tag{3.21}$$

satisfies

$$D_r = \begin{cases} \bar{\mathcal{O}}(1) & \text{if a discontinuity crosses the stencil,} \\ \mathcal{O}(h^{4r-3}) & \text{if the stencil converges to a critical point of order } k = 2r-3, \end{cases}$$

with

$$\begin{aligned} A_h &:= L_2(f_{-r+1,h}, \dots, f_{r-1,h}), \\ B_h &:= L_1(f_{-r+1,h}, \dots, f_{r-1,h}), \\ C_h &:= L_0(f_{-r+1,h}, \dots, f_{r-1,h}). \end{aligned}$$

For instance, for a WENO5 scheme with reconstructions from point values these terms can be written as

$$A_h = \frac{1}{2}f_{-2} - 2f_{-1} + 3f_0 - 2f_1 + \frac{1}{2}f_2 = \frac{1}{2}d_{3,4},$$
$$B_h = -\frac{1}{2}f_{-2} + f_{-1} - f_1 + \frac{1}{2}f_2,$$
$$C_h = -\frac{1}{12}f_{-2} + \frac{4}{3}f_{-1} - \frac{5}{2}f_0 + \frac{4}{3}f_1 - \frac{1}{12}f_2,$$

and for reconstructions from cell averages

$$A_h = \frac{1}{2}f_{-2} - 2f_{-1} + 3f_0 - 2f_1 + \frac{1}{2}f_2 = \frac{1}{2}d_{3,4},$$
$$B_h = -\frac{1}{2}f_{-2} + f_{-1} - f_1 + \frac{1}{2}f_2,$$
$$C_h = -\frac{1}{8}f_{-2} + \frac{3}{2}f_{-1} - \frac{11}{4}f_0 + \frac{3}{2}f_1 - \frac{1}{8}f_2.$$

Therefore, the parameter $\bar{d}(\varepsilon, h)$ given by

$$\bar{d}(\varepsilon, h) := \frac{d_{r,2r-2}^{s_1}|D_r|^{s_1}}{d_{r,2r-2}^{s_1} + |D_r|^{s_1} + \varepsilon} \tag{3.22}$$

satisfies, given a critical point of maximal order k,

$$\bar{d}(\varepsilon, h) = \begin{cases} \mathcal{O}(h^{(4r-4)s_1}) + \mathcal{O}(\varepsilon) & \text{if } 0 \leq k \leq 2r-4, \\ \mathcal{O}(h^{(4r-3)s_1}) + \mathcal{O}(\varepsilon) & \text{if } k = 2r-3, \\ \bar{\mathcal{O}}(1) + \mathcal{O}(\varepsilon) & \text{if a discontinuity crosses the stencil.} \end{cases} \tag{3.23}$$

By the same argument as in Section 3.1 and reasoning as in Theorem 3.1 and Theorem 3.2 it can be readily checked that in this case the suitable bounds for s_1 and s_2 are the same as those obtained by using Approach 1, that is,

$$s_1 \geq 1, \quad s_2 \geq \frac{r}{2s_1}. \tag{3.24}$$

We refer to Appendix A for the justification of (3.23).

Remark 3.3. *Since D_r is not guaranteed to be positive, we included its absolute value $|D_r|$. If one wants to avoid using an absolute value (and thus a Boolean condition in a WENO scheme), one has simply to chose an even s_1 satisfying the bounds in Equation* (3.24).

Summary of the algorithm. The algorithm in this case has the same steps as the one expounded for the additional node modality in Section 3.1.3, with the only difference that now the input is the original stencil S instead of the extended stencil $\bar{S}$ and that (3.19) is replaced by (3.22), i.e., the occurrences of $d_{r,2r-1}$ in (3.19) are replaced by D_r from (3.21).

4. Numerical experiments

In this section, the chosen exponents are $s_1 = \lceil r/2 \rceil$ when considering the additional node modality, $s_1 = 2\lceil r/4 \rceil$ when considering the modality with original nodes (taking into account Remark 3.3) and $s_2 = 1$. The reason for this choice is due to the fact that the choice of ε in (3.2) is ill-conditioned by the exponent s_2, since one ought to take $\varepsilon \geq\approx \sqrt[s_2]{\varepsilon_0}$, with ε_0 the lowest positive number of the

working precision, in order to avoid arithmetic underflow/overflow. Moreover, although unnecessary according to the accuracy requirements in case of smoothness, the greater the parameter s_1 is, the closer are simultaneously the weights to the ideal weights in case of smoothness and to zero in case of discontinuity.

4.1. **Algebraic test cases.** We start our numerical tests with several numerical experiments devoted to stress out the accuracy properties analyzed theoretically beforehand. We will perform tests involving JS-WENO (with the Jiang-Shu weight design [13]), YC-WENO (with the improved version of the Yamaleev-Carpenter weight design [2,21]) and OWENO (with our design) schemes of order $2r-1$, with $2 \leq r \leq 5$. All the tests are performed both with the modality of reconstructions from cell average values to pointwise values and from pointwise values to pointwise values.

In order to perform these experiments, we will use the multiple-precision library MPFR [17] through its C++ wrapper [12], using a precision of 3322 bits (≈ 1000 digits) and taking in all the cases $\varepsilon = 10^{-10^6}$.

Example 1: Smooth problem. Let us consider the family of functions $f_k : \mathbb{R} \to \mathbb{R}$, $k \in \mathbb{N}$, given by

$$f_k(x) = x^{k+1}\mathrm{e}^x.$$

The function f_k has a smooth extremum at $x = 0$ of order k. Results involving the different values of r and k considered ($0 \leq k \leq 2r-3$) are shown for $3 \leq r \leq 5$ in Table 1 for the case of the traditional JS-WENO and YC-WENO schemes, and in Table 2 for the optimal WENO schemes, being OWENO+1H the modality with an additional node where $\bar{d}$ is defined by (3.5), OWENO+1 the modality with an additional node where $\bar{d} = d_{r,2r-1}$ and OWENO the modality with the original nodes, i.e., with $\bar{d}$ defined by (3.22). The error is given by $E_{k,n} = |P_n(0) - f_k(0)|$, with P the corresponding reconstruction at $x_{1/2} = 0$, with the grid $x_i = (i-1/2)h$, $-r+1 \leq i \leq r-1$, with $h = 1/n$ for $n \in \mathbb{N}$, when pointwise values are taken, namely, $f_{k,i} = f_k(x_i)$ and reconstructions from pointwise values to pointwise values are performed. On the other hand, it is also presented in the tables the same setup when cell average values are taken instead: $f_{k,i} = \int_{x_i-h/2}^{x_i+h/2} f(x)\mathrm{d}x$, by performing reconstructions from cell average values to pointwise values. In all the cases, the tables show the corresponding average reconstruction orders, $O_k = \frac{1}{80}\sum_{j=1}^{80} o_{k,j}$, where $o_{k,j} = \log_2(E_{k,n_{j-1}}/E_{k,n_j}))$, with $n_j = 5 \cdot 2^j$, $0 \leq j \leq 80$.

As we can see, the JS-WENO looses accuracy near critical points, presenting the order $r + |k - r + 1|$, with k the order of the critical point, whereas the YC-WENO scheme looses accuracy in the corner case $k = 2r-3$, as suggested in our theoretical considerations. In contrast, we can see that our proposed optimal schemes attain the optimal accuracy in all cases.

Example 2: Discontinuous problem. We next consider the function $g : \mathbb{R} \to \mathbb{R}$ given by

$$g(x) = \begin{cases} \mathrm{e}^x & \text{if } x \leq 0 \\ \mathrm{e}^{x+1} & \text{if } x > 0 \end{cases}$$

and test the accuracy of the methods with the same parameters as above, where, in order to highlight the behaviour of our optimal scheme at discontinuities, in this case we change the location of the discontinuity by considering a grid of the form $x_i = (i - \frac{1}{2} + \theta)h$, $-r+1 \leq i \leq r-1$, for $-r+2 \leq \theta \leq r-1$. Since $x_{1/2} = \theta h$, the

k	JS-WENO	YC-WENO	JS-WENO	YC-WENO
	Order 5 (from point values)		Order 5 (from cell averages)	
0	4.9915	4.9983	4.9909	4.9983
1	3.9742	4.9980	3.9802	4.9981
2	3.0198	5.0331	3.0348	5.0324
3	3.9946	3.9945	3.9928	3.9928
	Order 7 (from point values)		Order 7 (from cell averages)	
0	6.9902	6.9984	6.9899	6.9984
1	5.9743	6.9981	5.9699	6.9981
2	5.0494	7.0002	5.0432	7.0001
3	4.0005	7.0627	4.0001	7.0600
4	5.0747	7.0040	5.0655	7.0108
5	6.0008	6.0008	6.0011	6.0011
	Order 9 (from point values)		Order 9 (from cell averages)	
0	8.9831	8.9984	8.9829	8.9985
1	8.0225	8.9983	8.0226	8.9983
2	7.0368	8.9981	7.0229	8.9981
3	6.0712	8.9978	6.0625	8.9978
4	5.0133	9.0628	5.0072	9.0625
5	5.9855	9.0325	5.9815	9.0283
6	7.0409	9.0121	7.0746	9.0143
7	7.9898	7.9898	7.9880	7.9880

TABLE 1. Example 1 (smooth problem): Fifth-order, seventh-order, and ninth-order traditional schemes.

k	OWENO+1H	OWENO+1	OWENO	OWENO+1H	OWENO+1	OWENO
	Order 5 (from point values)			Order 5 (from cell averages)		
0	4.9983	4.9983	4.9983	4.9983	4.9983	4.9983
1	4.9979	4.9979	4.9980	4.9979	4.9979	4.9980
2	5.0136	5.0161	5.0324	5.0131	5.0157	5.0317
3	5.0070	5.0070	5.0056	5.0052	5.0053	5.0035
	Order 7 (from point values)			Order 7 (from cell averages)		
0	6.9984	6.9984	6.9984	6.9984	6.9984	6.9984
1	6.9981	6.9981	6.9981	6.9981	6.9981	6.9981
2	6.9979	6.9979	7.0000	6.9979	6.9979	6.9998
3	7.0535	7.0543	7.0548	7.0439	7.0452	7.0482
4	7.0039	7.0040	7.0040	7.0107	7.0108	7.0108
5	6.9907	6.9907	6.9907	6.9970	6.9970	6.9970
	Order 9 (from point values)			Order 9 (from cell averages)		
0	8.9984	8.9984	8.9984	8.9985	8.9985	8.9985
1	8.9983	8.9983	8.9983	8.9983	8.9983	8.9983
2	8.9981	8.9981	8.9981	8.9981	8.9981	8.9981
3	8.9978	8.9978	8.9978	8.9979	8.9979	8.9979
4	9.0175	9.0179	8.9976	9.0173	9.0177	8.9976
5	9.0325	9.0325	9.0185	9.0282	9.0283	9.0082
6	9.0121	9.0121	9.0121	9.0143	9.0143	9.0143
7	8.9856	8.9856	8.9541	8.9875	8.9875	8.9872

TABLE 2. Example 1 (smooth problem): Fifth-order, seventh-order, and ninth-order optimal schemes.

error in this case is thus given by $|P(\theta h) - g(\theta h)|$. The results are shown in Table 3 for the traditional schemes and in Table 4 for the optimal schemes.

θ	JS-WENO	YC-WENO	JS-WENO	YC-WENO
	Order 5 (from point values)		Order 5 (from cell averages)	
-2	2.9955	2.9951	2.9955	2.9951
-1	2.9927	2.9925	2.9935	2.9934
0	3.0029	3.0045	3.0033	3.0050
1	3.0271	3.0390	3.0294	3.0411
	Order 7 (from point values)		Order 7 (from cell averages)	
-3	3.9970	4.0035	3.9971	4.0041
-2	4.0088	4.0089	4.0071	4.0072
-1	3.9509	3.9493	4.0086	4.0087
0	4.0086	4.0086	4.0086	4.0086
1	4.0234	4.0234	4.0234	4.0234
2	4.0206	4.0368	4.0211	4.0370
	Order 9 (from point values)		Order 9 (from cell averages)	
-4	4.9937	4.9937	4.9938	4.9938
-3	4.9933	4.9933	4.9933	4.9933
-2	4.9928	4.9928	4.9927	4.9927
-1	4.9925	4.9925	4.9924	4.9924
0	4.9886	4.9886	4.9917	4.9917
1	5.0561	5.0561	5.0561	5.0561
2	5.0564	5.0564	5.0574	5.0574
3	5.0129	5.0992	5.0154	5.1006

TABLE 3. Example 2 (discontinuous problem): Fifth-order, seventh-order, and ninth-order traditional schemes.

θ	OWENO+1H	OWENO+1	OWENO	OWENO+1H	OWENO+1	OWENO
	Order 5 (from point values)			Order 5 (from cell averages)		
-2	2.9951	2.9955	2.9917	2.9951	2.9955	2.9929
-1	2.9925	2.9927	2.9923	2.9934	2.9935	2.9933
0	3.0051	3.0034	3.0070	3.0056	3.0039	3.0071
1	3.0454	3.0389	3.0517	3.0474	3.0410	3.0517
	Order 7 (from point values)			Order 7 (from cell averages)		
-3	4.0035	3.9970	4.0140	4.0041	3.9971	4.0297
-2	4.0089	4.0088	4.0090	4.0072	4.0071	4.0073
-1	3.9492	3.9508	3.9473	4.0087	4.0086	4.0088
0	4.0086	4.0086	4.0086	4.0086	4.0086	4.0086
1	4.0234	4.0234	4.0234	4.0234	4.0234	4.0235
2	4.0368	4.0368	4.0344	4.0369	4.0370	4.0353
	Order 9 (from point values)			Order 9 (from cell averages)		
-4	4.9937	4.9937	4.9937	4.9938	4.9938	4.9938
-3	4.9933	4.9933	4.9933	4.9933	4.9933	4.9933
-2	4.9928	4.9928	4.9928	4.9927	4.9927	4.9927
-1	4.9925	4.9925	4.9925	4.9924	4.9924	4.9924
0	4.9886	4.9886	4.9886	4.9917	4.9917	4.9917
1	5.0561	5.0561	5.0561	5.0561	5.0561	5.0561
2	5.0564	5.0564	5.0564	5.0574	5.0574	5.0574
3	5.1095	5.0992	5.1073	5.1109	5.1006	5.1042

TABLE 4. Example 2 (discontinuous problem): Fifth-order, seventh-order, and ninth-order optimal schemes.

The table shows clearly that the optimal accuracy is also attained in all the cases when a discontinuity is on the data.

	JS-WENO5				YC-WENO5			
	$\|\cdot\|_1$		$\|\cdot\|_\infty$		$\|\cdot\|_1$		$\|\cdot\|_\infty$	
n	Error	rate	Error	rate	Error	rate	Error	rate
10	8.44e-03	—	1.28e-02	—	9.52e-04	—	1.45e-03	—
20	3.59e-04	4.56	6.93e-04	4.20	2.95e-05	5.02	4.65e-05	4.96
40	1.09e-05	5.04	2.37e-05	4.87	9.03e-07	5.03	1.42e-06	5.03
80	3.29e-07	5.05	7.00e-07	5.08	2.78e-08	5.02	4.37e-08	5.02
160	1.02e-08	5.01	2.21e-08	4.98	8.63e-10	5.01	1.36e-09	5.01
320	3.19e-10	5.00	6.65e-10	5.06	2.68e-11	5.01	4.22e-11	5.01
640	9.96e-12	5.00	2.02e-11	5.04	8.37e-13	5.00	1.32e-12	5.00

TABLE 5. Example 3 (linear advection equation, solution at $T = 1$): traditional fifth-order schemes.

	OWENO5+1H				OWENO5+1				OWENO5			
	$\|\cdot\|_1$		$\|\cdot\|_\infty$		$\|\cdot\|_1$		$\|\cdot\|_\infty$		$\|\cdot\|_1$		$\|\cdot\|_\infty$	
n	Error	rate	Error	rate	Error	rate	Error	rate	Error	rate	Error	rate
10	9.48e-04	—	1.46e-03	—	9.51e-04	—	1.46e-03	—	9.52e-04	—	1.45e-03	—
20	2.95e-05	5.01	4.65e-05	4.98	2.95e-05	5.01	4.65e-05	4.98	2.95e-05	5.01	4.65e-05	4.96
40	9.03e-07	5.03	1.42e-06	5.03	9.03e-07	5.03	1.42e-06	5.03	9.03e-07	5.03	1.42e-06	5.03
80	2.78e-08	5.02	4.37e-08	5.02	2.78e-08	5.02	4.37e-08	5.02	2.78e-08	5.02	4.37e-08	5.02
160	8.63e-10	5.01	1.36e-09	5.01	8.63e-10	5.01	1.36e-09	5.01	8.63e-10	5.01	1.36e-09	5.01
320	2.68e-11	5.01	4.22e-11	5.01	2.68e-11	5.01	4.22e-11	5.01	2.68e-11	5.01	4.22e-11	5.01
640	8.37e-13	5.00	1.32e-12	5.00	8.37e-13	5.00	1.32e-12	5.00	8.37e-13	5.00	1.32e-12	5.00

TABLE 6. Example 3 (linear advection equation, solution at $T = 1$): optimal fifth-order schemes.

4.2. **Experiments for conservation laws.** In this section some numerical experiments involving hyperbolic conservation laws will be considered. For this purpose, we use a Local Lax-Friedrichs type flux splitting for smooth problems, and Donat-Marquina's flux formula [7] for problems with weak solutions. On the other hand, for the time discretization, the approximate Lax-Wendroff schemes proposed in [22] matching the spatial order will be considered. In this Section we work in all experiments with double precision representation and set $\varepsilon = 10^{-100}$. For all schemes we consider the case of fifth-order accuracy.

Example 3: Linear advection equation. We consider the linear advection equation with the following domain, boundary condition and initial condition:

$$u_t + f(u)_x = 0, \quad \Omega = (-1, 1), \quad u(-1, t) = u(1, t),$$
$$f(u) = u, \quad u_0(x) = 0.25 + 0.5\sin(\pi x),$$

whose exact solution is $u(x, t) = 0.25 + 0.5\sin(\pi(x - t))$. We run several simulations with final time $T = 1$, for resolutions $h = 2/n$, for some $n \in \mathbb{N}$ using the classical JS-WENO and YC-WENO schemes and our OWENO schemes, and compare them for the case of fifth-order accuracy, both with the $\|\cdot\|_1$ and $\|\cdot\|_\infty$ errors. Since the characteristics point to the right, we use left-biased reconstructions. The results are shown in Table 5 for the traditional fifth-order schemes and in Table 6 for the optimal fifth-order schemes.

The numerical results show that all schemes keep the fifth-order accuracy. The results of the OWENO schemes are almost identical to the ones of the YC-WENO scheme.

	JS-WENO5				YC-WENO5			
	$\|\cdot\|_1$		$\|\cdot\|_\infty$		$\|\cdot\|_1$		$\|\cdot\|_\infty$	
n	Err.	$\mathcal{O}$	Err.	$\mathcal{O}$	Err.	$\mathcal{O}$	Err.	$\mathcal{O}$
40	6.28e-05	—	2.73e-04	—	2.55e-05	—	2.62e-04	—
80	3.14e-06	4.32	4.26e-05	2.68	8.46e-07	4.91	1.04e-05	4.65
160	1.55e-07	4.35	2.87e-06	3.89	2.62e-08	5.01	3.27e-07	4.99
320	9.44e-09	4.03	2.75e-07	3.38	7.97e-10	5.04	1.02e-08	5.00
640	5.38e-10	4.13	3.29e-08	3.06	2.45e-11	5.02	3.14e-10	5.02
1280	3.46e-11	3.96	3.58e-09	3.20	7.59e-13	5.01	9.71e-12	5.02
2560	2.10e-12	4.04	4.80e-10	2.90	2.34e-14	5.02	3.03e-13	5.00

TABLE 7. Example 4 (Burgers equation, smooth solution at $T = 0.3$): traditional fifth-order schemes.

	OWENO5+1H				OWENO5+1				OWENO5			
	$\|\cdot\|_1$		$\|\cdot\|_\infty$		$\|\cdot\|_1$		$\|\cdot\|_\infty$		$\|\cdot\|_1$		$\|\cdot\|_\infty$	
n	Error	rate	Error	rate	Error	rate	Error	rate	Error	rate	Error	rate
40	2.45e-05	—	2.62e-04	—	2.44e-05	—	2.62e-04	—	2.49e-05	—	2.62e-04	—
80	8.46e-07	4.85	1.04e-05	4.66	8.46e-07	4.85	1.04e-05	4.66	8.46e-07	4.88	1.04e-05	4.65
160	2.62e-08	5.01	3.27e-07	4.99	2.62e-08	5.01	3.27e-07	4.99	2.62e-08	5.01	3.27e-07	4.99
320	7.97e-10	5.04	1.02e-08	5.00	7.97e-10	5.04	1.02e-08	5.00	7.97e-10	5.04	1.02e-08	5.00
640	2.45e-11	5.02	3.14e-10	5.02	2.45e-11	5.02	3.14e-10	5.02	2.45e-11	5.02	3.14e-10	5.02
1280	7.59e-13	5.01	9.71e-12	5.02	7.59e-13	5.01	9.71e-12	5.02	7.59e-13	5.01	9.71e-12	5.02
2560	2.34e-14	5.02	3.03e-13	5.00	2.34e-14	5.02	3.03e-13	5.00	2.34e-14	5.02	3.03e-13	5.00

TABLE 8. Example 4 (Burgers equation, smooth solution at $T = 0.3$): optimal fifth-order schemes.

Examples 4 and 5: Burgers equation. We now consider the Burgers equation with following setup involving the domain, boundary conditions and initial condition:

$$\begin{gathered} u_t + f(u)_x = 0, \quad \Omega = (-1,1), \quad u(-1,t) = u(1,t), \\ f(u) = \frac{u^2}{2}, \quad u_0(x) = 0.25 + 0.5\sin(\pi x). \end{gathered} \tag{4.1}$$

In this case, $f(u_0(x))$ has a first-order smooth extremum at $x = -1/2$ and $x = 1/2$. In Example 4, we consider the solution of (4.1) at $T = 0.3$, when it remains smooth, while in Example 5 we set $T = 12$, when the solution of (4.1) has become discontinuous.

In Example 4 we run simulations for different resolutions, with a Local Lax-Friedrichs flux splitting, showing the behaviour of the traditional fifth order schemes in Table 7 and the optimal fifth order schemes in Table 8. The exact solution is computed through a characteristic line method together with the Newton method, setting as tolerance the machine accuracy of the double precision. An accuracy loss is observed for the JS-WENO scheme. In contrast, the accuracy order of the YC-WENO and all the OWENO schemes is optimal.

In Example 5 we run the simulation instead until $T = 12$. At $t = 1$, the wave breaks and a shock is generated. Therefore, in this case we use the Donat-Marquina flux-splitting algorithm. The results are shown in Figure 1 with a resolution of $n = 80$ points, and are compared against a reference solution computed with a resolution of $n = 16000$ points. In this case, we can see that YC-WENO and OWENO schemes have a similar resolution, which in turn are slightly higher than

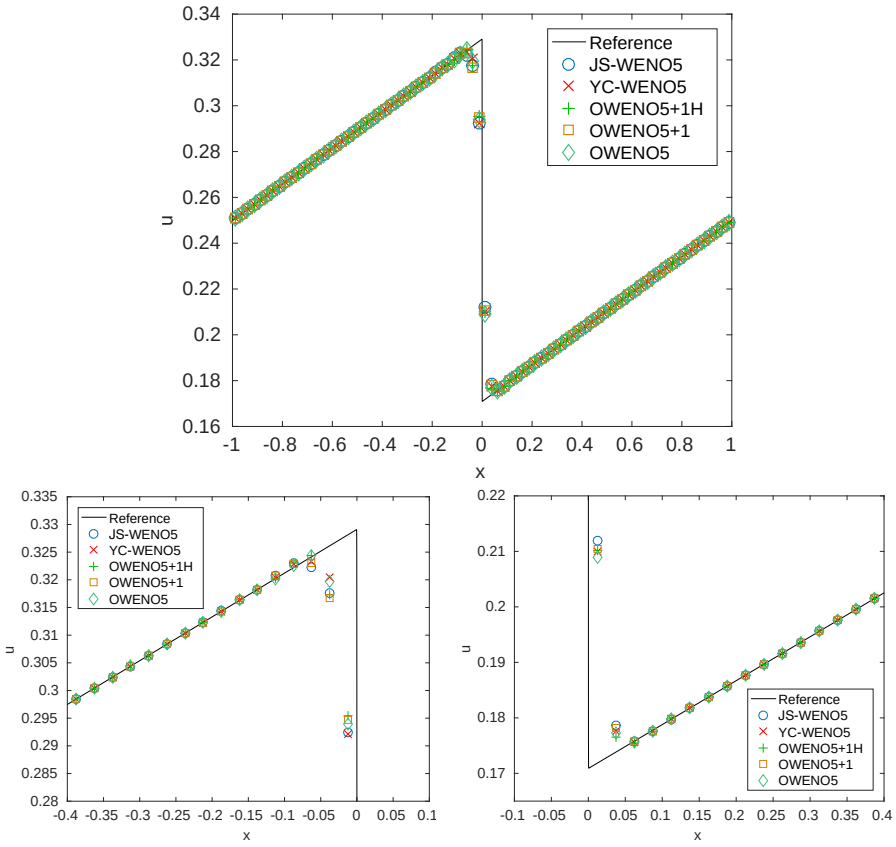

FIGURE 1. Example 5 (Burgers equation, discontinuous solution at $T = 12$): fifth-order schemes.

the resolution shown by the JS-WENO scheme. This ranking of resolution is also consistent with the behaviour that these schemes present in the smooth case.

Example 6: Customized equation with a third-order zero. We now consider the following initial-boundary value problem for a customized equation:

$$\begin{aligned} &u_t + f(u)_x = 0, \quad \Omega = (-1, 1), \quad u(-1, t) = u(1, t), \\ &f(u) = \frac{u^2}{2} + 0.25u, \quad u_0(x) = 0.25 + 0.5\sin(\pi x). \end{aligned} \tag{4.2}$$

In this case, $f(u_0(x))$ has a third-order smooth extremum at $x = -1/2$ and a first-order smooth extremum at $x = 1/2$.

We now compare the behaviour of the three schemes with the same setup as in Example 4, by running a simulation until $T = 0.3$, in which the solution is smooth. For the computation of the exact solution, we once again use the method of characteristic lines, with a Newton method matching the machine accuracy for the double precision. Since in this case the characteristics point always to the right, we use a left-biased upwind scheme. The results are shown in Table 9 for the traditional fifth-order schemes and in Table 10 for the optimal fifth-order schemes.

	JS-WENO5				YC-WENO5			
	$\|\cdot\|_1$		$\|\cdot\|_\infty$		$\|\cdot\|_1$		$\|\cdot\|_\infty$	
n	Error	rate	Error	rate	Error	rate	Error	rate
40	7.96e-05	—	5.17e-04	—	4.97e-05	—	3.15e-04	—
80	4.67e-06	4.09	7.31e-05	2.82	2.88e-06	4.11	5.58e-05	2.50
160	2.70e-07	4.11	9.73e-06	2.91	1.69e-07	4.09	7.98e-06	2.81
320	1.60e-08	4.08	1.25e-06	2.96	1.01e-08	4.06	1.06e-06	2.91
640	9.70e-10	4.04	1.59e-07	2.98	6.16e-10	4.03	1.36e-07	2.96
1280	5.95e-11	4.03	2.01e-08	2.99	3.81e-11	4.02	1.72e-08	2.98
2560	3.68e-12	4.02	2.52e-09	2.99	2.36e-12	4.01	2.17e-09	2.99

TABLE 9. Example 6 (customized equation, smooth solution at $T = 0.3$): traditional fifth-order schemes.

	OWENO5+1H				OWENO5+1				OWENO5			
	$\|\cdot\|_1$		$\|\cdot\|_\infty$		$\|\cdot\|_1$		$\|\cdot\|_\infty$		$\|\cdot\|_1$		$\|\cdot\|_\infty$	
n	Error	rate	Error	rate	Error	rate	Error	rate	Error	rate	Error	rate
40	3.07e-05	—	2.01e-04	—	3.09e-05	—	2.01e-04	—	2.93e-05	—	2.01e-04	—
80	9.74e-07	4.98	9.83e-06	4.35	9.75e-07	4.98	9.83e-06	4.35	1.01e-06	4.86	9.83e-06	4.35
160	2.69e-08	5.18	3.26e-07	4.91	2.69e-08	5.18	3.26e-07	4.91	3.05e-08	5.05	3.42e-07	4.85
320	7.94e-10	5.08	1.03e-08	4.99	7.94e-10	5.08	1.03e-08	4.99	8.82e-10	5.11	1.22e-08	4.81
640	2.46e-11	5.01	3.22e-10	5.00	2.46e-11	5.01	3.22e-10	5.00	2.61e-11	5.08	3.95e-10	4.95
1280	7.67e-13	5.00	1.01e-11	5.00	7.67e-13	5.00	1.01e-11	5.00	7.91e-13	5.04	1.25e-11	4.99
2560	2.47e-14	4.96	3.10e-13	5.02	2.47e-14	4.96	3.10e-13	5.02	2.51e-14	4.98	3.90e-13	5.00

TABLE 10. Example 6 (customized equation, smooth solution at $T = 0.3$): optimal fifth-order schemes.

In this example, one can see that the optimal accuracy order is lost for both the JS-WENO and YC-WENO schemes. In contrast, the fifth-order accuracy is solidly kept by the OWENO schemes.

Example 7: Shu-Osher problem. The 1D Euler equations for gas dynamics are given by $\boldsymbol{u} = (\rho, \rho v, E)^{\mathrm{T}}$ and $\boldsymbol{f}(\boldsymbol{u}) = \boldsymbol{f}^1(\boldsymbol{u}) = (\rho v, p + \rho v^2, v(E + p))^{\mathrm{T}}$, where ρ is the density, v is the velocity and E is the specific energy of the system. The variable p stands for the pressure and is given by the equation of state

$$p = (\gamma - 1)\left(E - \frac{1}{2}\rho v^2\right),$$

where γ is the adiabatic constant that will be taken as $\gamma = 1.4$. We now consider the interaction with a Mach 3 shock and a sine wave. The spatial domain is now given by $\Omega := (-5, 5)$, with the initial condition

$$(\rho, v, p)(x, 0) = \begin{cases} \left(\dfrac{27}{7}, \dfrac{4\sqrt{35}}{9}, \dfrac{31}{3}\right) & \text{if } x \leq -4, \\ \left(1 + \dfrac{1}{5}\sin(5x), 0, 1\right) & \text{if } x > -4, \end{cases}$$

with left inflow and right outflow boundary conditions. This problem was first considered in [19].

We run the simulation until $T = 1.8$ and compare the schemes against a reference solution computed with a resolution of $n = 16000$. Figures 2 and 3 correspond to resolutions of $n = 200$ and $n = 400$ points, respectively.

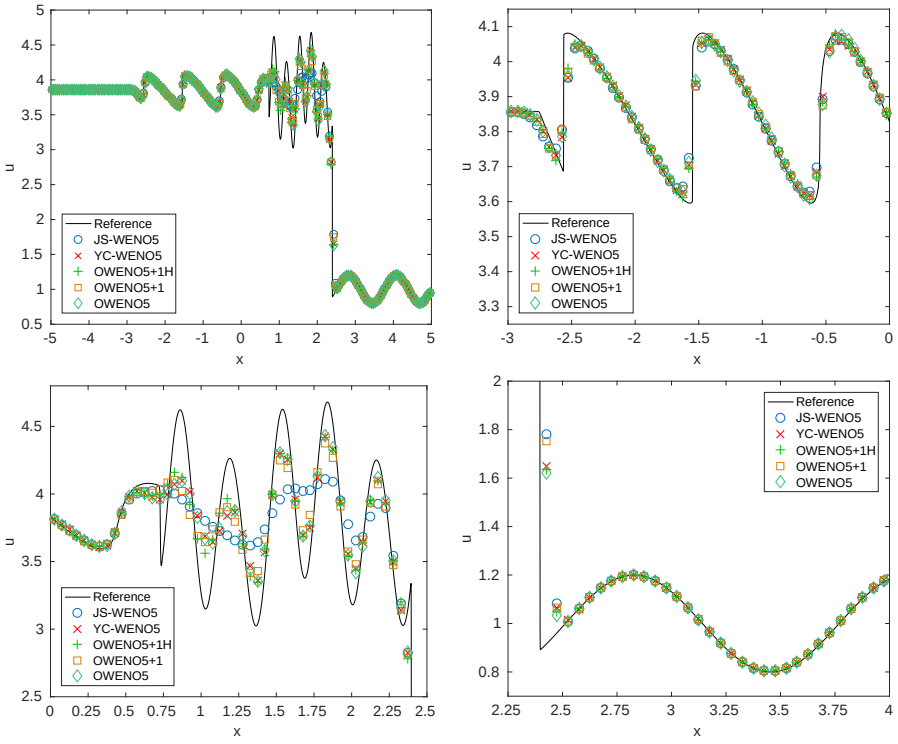

FIGURE 2. Example 7 (Euler equations, Shu-Osher problem): fifth-order schemes, $n = 200$.

We can see that both YC-WENO and OWENO schemes have similar resolutions, being the one presented by the OWENO schemes slightly higher in the case of the OWENO5+1H and OWENO5 schemes and slightly lower for the choice OWENO5+1. The lowest resolution clearly corresponds to the JS-WENO scheme, especially for the case $n = 200$. For $n = 400$ it can be seen that the OWENO5+1H and OWENO5 schemes capture notably better the shock than the other schemes. The slightly poorer performance of the OWENO5+1 scheme in this case might be due to the loss of accuracy that happens in the situation explained in Remark 3.2, or to the fact that it loses accuracy near a discotinuity when it is located at the extended stencil $\bar{S}$, but not at S.

Finally, we show in Figure 4 a comparison involving the error of each scheme with respect to the corresponding CPU time required to achieve it. We can see that the efficiency of all schemes is nearly the same in the case of fifth-order accuracy, although minor differences are found for lower resolution in benefit of both YC-WENO and OWENO schemes. Such asymptotic behaviour is probably due to the fact that there is no zero of order higher than 1 along the derivative of the composition of the flux with the solution. All the considered schemes can cope with the phenomena properly.

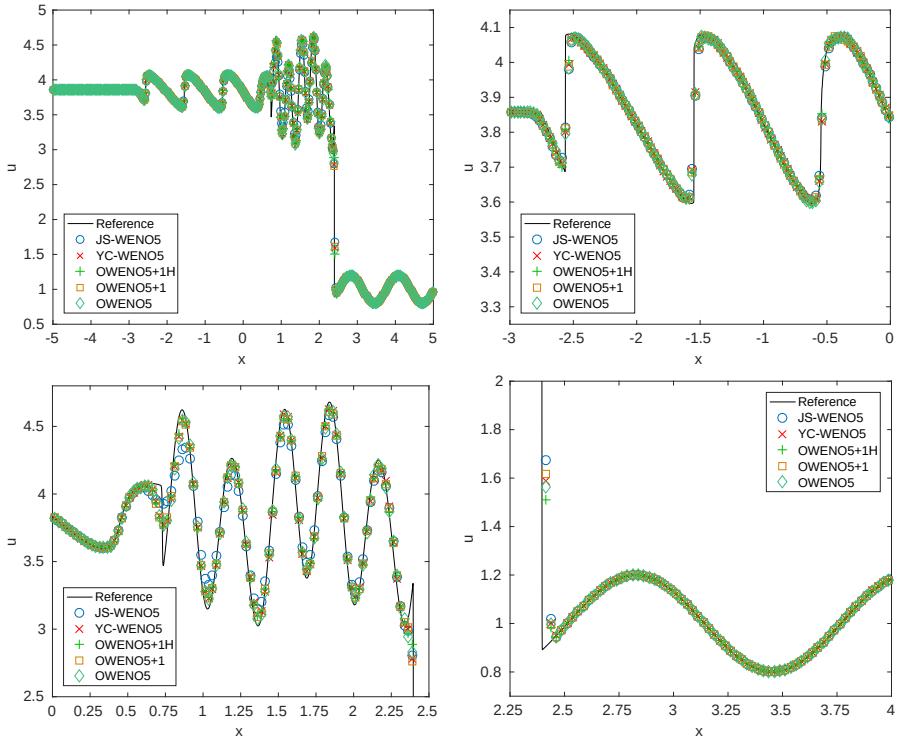

FIGURE 3. Example 7 (Euler equations, Shu-Osher problem): fifth-order schemes, $n = 400$.

5. Conclusions

In this paper we have proposed a novel scheme, based on the approach of Yamaleev and Carpenter [21], in which the accuracy is not lost regardless of the order of the critical point to which the stencil converges. This approach requires the usage of an additional node or the computation of some additional quantities in order to build the weights so that the accuracy of the approximation is unconditionally optimal. However, the requirement of using an additional node in the former case is not a major problem in the context in which we have worked along the paper, focused on finite-difference schemes for hyperbolic conservation laws, in the sense that both the additional implementation and computational cost is not higher than for other WENO schemes, as is shown in the numerical experiments.

This new family of schemes has been proven to outperform the previous schemes under some circumstances, both in smooth and discontinuous solutions, and behave similarly under other situations. However, we expect a much more significative improvement for third-order schemes, whose original version proposed by Jiang-Shu [13] loses order near first order critical points, which in this case, unlike higher order smooth extrema, is a very common phenomena appearing on solutions of any type of ODE or PDE. Therefore, fixing this issue would entail a substantial improvement in the case of third-order WENO schemes. Since the procedure that we have described here is not valid for the case of third-order schemes, we are

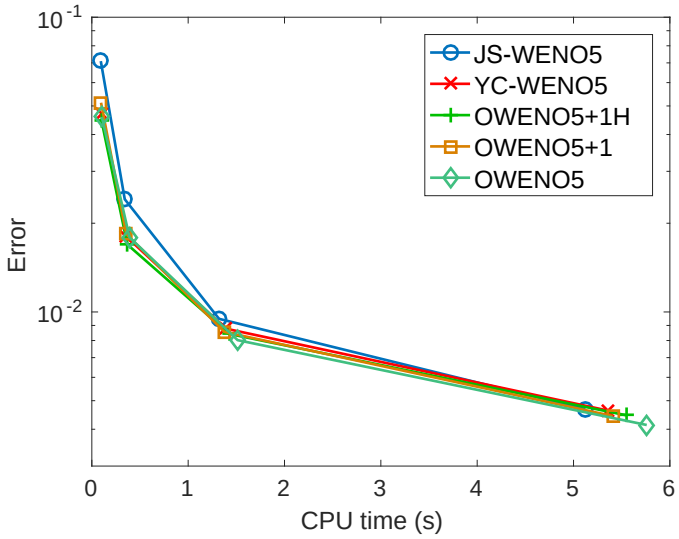

FIGURE 4. Example 7 (Euler equations, Shu-Osher problem): fifth-order schemes, efficiency plot.

currently working on the development of a third order scheme with unconditionally optimal accuracy for smooth data.

APPENDIX A. THEORETICAL RESULTS FOR THE ACCURACY OF OWENO SCHEMES WITHOUT ADDITIONAL NODES

This additional section stands for the theoretical grounds involving the motivation for the definition of the parameter D_r in (3.21) according to its accuracy properties. For the sake of simplicity, we will assume for now reconstructions from point values, albeit a similar analysis can be done for reconstructions from cell averages.

Lemma A.1. *Let $n \in \mathbb{N}$ and $p \in \bar{\Pi}_n$. If p has exactly n different roots, then $p^{(s)}$ has exactly $n-s$ roots, for $0 \leq s \leq n$.*

Proof. The result is direct consequence of applying inductively Rolle's theorem to deduce that the number of roots of $p^{(s)}$ is at least $n-s$ and the Fundamental Theorem of Algebra, together with the fact that $p^{(s)} \in \bar{\Pi}_{n-s}$, to conclude that $p^{(s)}$ has exactly $n-s$ roots. □

Lemma A.2. *Let $\{x_i\}_{i=0}^n$ be a stencil such that $x_i < x_j$ if $i < j$. Let $0 \leq i_0 \leq n-1$ and $p \in \Pi_n$ be an interpolating polynomial such that $p(x_i) = f_{\mathrm{L}}$ if $i \leq i_0$ and $p(x_i) = f_{\mathrm{R}}$ if $i > i_0$, with $f_{\mathrm{L}} \neq f_{\mathrm{R}}$. Then, $p^{(s)}$ has exactly $n-s$ roots, for $1 \leq s \leq n$, and $p^{(s)} \in \bar{\Pi}_{n-s}$ for $0 \leq s \leq n$. In particular, the parabola $p^{(n-2)}$ has two simple roots.*

Proof. Let $0 \leq i \leq n-1$ such that $i \neq i_0$. Then, by construction, we have $p(x_i) = p(x_{i+1})$, and therefore by Rolle's therem exists $\xi_i \in (x_i, x_{i+1})$ such that $p'(\xi_i) = 0$, $0 \leq i \leq n-1$. Therefore, $p' \in \Pi_{n-1}$ has at least $n-1$ roots. However, by hypothesis p is not a constant polynomial, and thus $p' \not\equiv 0$. Hence, $p' \in \bar{\Pi}_{n-1}$, and p' must have exactly $n-1$ roots. Thus, by Lemma A.1 we have that $(p')^{(s)} = p^{(s+1)}$

has exactly $(n-1)-s = n-(s+1)$ roots for $0 \le s \le n-1$. Replacing $s+1$ by s, we thus have that $p^{(s)}$ has exactly $n-s$ roots for $1 \le s \le n$. Since in particular $p^{(s)} \in \Pi_{n-s}$ with $n-s$ roots, $1 \le s \le n$, we deduce that the degree of $p^{(s)} \in \bar{\Pi}_{n-s}$, and thus it also deduced that the degree of $p \in \bar{\Pi}_n$, which proves the result. □

Lemma A.3. *Let $x_{i,h} = z + a_i h$, $0 \le i \le n$, be a grid such that $a_i < a_j$ if $i < j$ and $p_h \in \Pi_n$ the interpolating polynomial such that $p_h(x_{i,h}) = f_{i,h}$, for $f_{i,h} \in \mathbb{R}$, $0 \le i \le n$. Then, given $0 \le s \le n$, the s-th derivative of $P_h(w) := p_h(z+wh)$ can be written as*

$$P_h^{(s)}(w) = \sum_{j=0}^{n-s} L_{\boldsymbol{a}}^{s,j}(f_{0,h}, \ldots, f_{n,h}) w^j, \quad \boldsymbol{a} := (a_0, \ldots, a_n),$$

with $L_{\boldsymbol{a}}^{s,j} : \mathbb{R}^{n+1} \to \mathbb{R}$ a linear function, which does not depend on h. Furthermore,

$$L_{\boldsymbol{a}}^{s,j}(f_{0,h}, \ldots, f_{n,h}) = \frac{(s+j)!}{j!} L_{\boldsymbol{a}}^{0,s+j}(f_{0,h}, \ldots, f_{n,h}). \tag{A.1}$$

Moreover, if $f_{i,h} = f(x_{i,h})$, for some $f \in \mathcal{C}^{n+1}$, then

$$L_{\boldsymbol{a}}^{s,j}(f_{0,h}, \ldots, f_{n,h}) = h^{s+j} \frac{1}{j!} f^{(s+j)}(z) + \mathcal{O}(h^{n+1}). \tag{A.2}$$

Proof. The polynomial p_h can be written in Lagrange form as

$$p_h(x) = \sum_{i=0}^{n} p_{i,h}(x) f_{i,h}, \quad p_{i,h}(x) = \prod_{j=0, j\neq i}^{n} \frac{x - x_{j,h}}{x_{i,h} - x_{j,h}}. \tag{A.3}$$

Setting $P_{i,h}(w) := p_{i,h}(z+wh)$ and using $x_{i,h} = z + a_i h$, we easily verify that

$$P_{i,h}(w) = p_{i,h}(z+wh) = \prod_{j=0, j\neq i}^{n} \frac{w - a_j}{a_i - a_j}.$$

Therefore, $P_{i,h}(w) = P_i(w)$ is a polynomial in Π_n which does not depend on h, which in turn can be written as

$$P_i(z+wh) = \sum_{j=0}^{n} F_{i,j}(\boldsymbol{a}) w^j.$$

Now, replacing $x = z + wh$ in (A.3) and differentiating s times yields

$$P_h^{(s)}(w) = h^s \sum_{i=0}^{n} P_i^{(s)}(w) f_{i,h},$$

and thus

$$P_h^{(s)}(w) = \sum_{i=0}^{n} P_i^{(s)}(w) f_{i,h} = \sum_{i=0}^{n} \left(\sum_{j=0}^{n-s} F_{i,j}^s(\boldsymbol{a}) w^j \right) f_{i,h} = \sum_{j=0}^{n-s} \left(\sum_{i=0}^{n} F_{i,j}^s(\boldsymbol{a}) f_{i,h} \right) w^j,$$

with

$$F_{i,j}^s(a_0, \ldots, a_n) = \frac{(j+s)!}{j!} F_{i,j}(a_0, \ldots, a_n). \tag{A.4}$$

Therefore, the first statement of Lemma A.3 is proven by taking

$$L_{\boldsymbol{a}}^{s,j}(f_{0,h}, \ldots, f_{n,h}) := \sum_{i=0}^{n} F_{i,j}^s(\boldsymbol{a}) f_{i,h},$$

which is a linear map.

Equation (A.1) is a direct consequence of (A.4). As for Equation (A.2), taking into consideration that $p_h^{(s)}(z) = f^{(s)}(z) + \mathcal{O}(h^{n-s+1})$, then, differentiating s times both sides of the equality $P_h(w) = p_h^{(s)}(z + wh)$ and evaluating at $w = 0$, we have

$$P_h^{(s)}(0) = h^s p_h^{(s)}(z) = h^s(f^{(s)}(z) + \mathcal{O}(h^{n+1-s})) = h^s f^{(s)}(z) + \mathcal{O}(h^{s+1}). \quad \text{(A.5)}$$

On the other hand, by definition, $P_h^{(s)}(0) = L_{\boldsymbol{a}}^{s,0}(f_{0,h}, \dots, f_{n,h})$. Combining this with (A.5), we have

$$L_{\boldsymbol{a}}^{s,0}(f_{0,h}, \dots, f_{n,h}) = h^s f^{(s)}(z) + \mathcal{O}(h^{n+1}),$$

and thus (A.2) holds using conveniently (A.1). □

Proposition A.1. *Let $n \geq 3$, $f \in \mathcal{C}(A)$, with $A = (-\infty, z) \cup (z, +\infty)$ with a discontinuity at z, that is, such that $\lim_{x \to z^-} f(x) =: f_{\mathrm{L}} \neq f_{\mathrm{R}} := \lim_{x \to z^+} f(x)$ and $f(z) = f_{\mathrm{L}}$ or $f(z) = f_{\mathrm{R}}$. Let $x_{i,h} = z + a_i h$, $0 \leq i \leq n$, a stencil such that $a_i < a_j$ if $i < j$, with $a_0 < 0$ and $a_n > 0$, and p_h the interpolating polynomial of degree at most n such that $p_h(x_{i,h}) = f(x_{i,h})$. Then exists $h_0 > 0$ such that for all h, $0 < h \leq h_0$, the parabola $P_h^{(n-2)}(w)$, with $P_h(w) := p_h(z + wh)$ has two simple roots for $0 < h < h_0$, $\lambda_{1,h}$, $\lambda_{2,h}$, which moreover satisfy $\lambda_1 \neq \lambda_2$, with $\lambda_i := \lim_{h \to 0^+} \lambda_{i,h}$, $1 \leq i \leq 2$.*

In particular, given the expression of the parabola $P_h^{(n-2)}$ as $P_h^{(n-2)}(w) = A_h w^2 + B_h w + C_h$, there holds

$$B_h^2 - 4A_h C_h = \bar{\mathcal{O}}(1).$$

Proof. Let

$$i_0 := \begin{cases} \min\{0 \leq i \leq n \mid a_i \leq 0 \wedge a_{i+1} > 0\} & \text{if } f(z) = f_{\mathrm{L}} \\ \min\{0 \leq i \leq n \mid a_i < 0 \wedge a_{i+1} \geq 0\} & \text{if } f(z) = f_{\mathrm{R}} \end{cases}$$

We now define p the interpolating polynomial in $\bar{\Pi}_n$ satisfying $p(a_i) = f_{\mathrm{L}} =: f_i$ if $i \leq i_0$ and $p(a_i) = f_{\mathrm{R}} =: f_i$ if $i > i_0$, $0 \leq i \leq n$. Then, by Lemma A.2, $P^{(n-2)}$, with $P(w) := p(z + wh)$ has two simple roots.

Since by Lemma A.3 we can write

$$P^{(n-2)}(w) = \sum_{j=0}^{2} L_{\boldsymbol{a}}^j(f_0, \dots, f_n) w^j$$

for some linear maps $L_{\boldsymbol{a}}^j(f_0, \dots, f_n)$, $0 \leq j \leq 2$ we deduce that

$$L_{\boldsymbol{a}}^1(f_0, \dots, f_n)^2 - 4L_{\boldsymbol{a}}^0(f_0, \dots, f_n) L_{\boldsymbol{a}}^2(f_0, \dots, f_n) > 0, \quad \text{(A.6)}$$

since the left-hand side of (A.6) is the discriminant of the equation $P^{(n-2)}(w) = 0$, which has two simple roots, λ_1, λ_2.

Now, using again Lemma A.3 on $P_h^{(n-2)}$, one has

$$P_h^{(n-2)}(w) = \sum_{j=0}^{2} L_{\boldsymbol{a}}^j(f(x_{0,h}), \dots, f(x_{n,h})) w^j. \quad \text{(A.7)}$$

By the countinuity of f outside z, $\exists h_1 > 0 : \forall 0 < h < h_1$ there holds $f(x_{i,h}) = f_i + \mathcal{O}(h)$. Therefore, using the linearity of the maps $L_{\boldsymbol{a}}^j$, (A.7) reads

$$P_h^{(n-2)}(w) = \sum_{j=0}^{2} L_{\boldsymbol{a}}^j(f_0 + \mathcal{O}(h), \dots, f_n + \mathcal{O}(h)) w^j$$

$$= \sum_{j=0}^{2} \left[L^j_{\boldsymbol{a}}(f_0, \ldots, f_n) + C_j \mathcal{O}(h) \right] w^j,$$

for some $C_j \in \mathbb{R}$, $0 \leq j \leq 2$.

Finally, defining

$$L^j_{\boldsymbol{a}}[h](f_0, \ldots, f_n) := L^j_{\boldsymbol{a}}(f(x_{0,h}), \ldots, f(x_{n,h})) = L^j_{\boldsymbol{a}}(f_0, \ldots, f_n) + C_j \mathcal{O}(h)$$

and taking into account that

$$\lim_{h \to 0^+} L^j_{\boldsymbol{a}}[h](f_0, \ldots, f_n) = L^j_{\boldsymbol{a}}(f_0, \ldots, f_n),$$

by continuity arguments and (A.6) one can deduce that exists h_0, $0 < h_0 < h_1$ such that for $0 < h < h_0$

$$L^1_{\boldsymbol{a}}[h](f_0, \ldots, f_n)^2 - 4 L^0_{\boldsymbol{a}}[h](f_0, \ldots, f_n) L^2_{\boldsymbol{a}}[h](f_0, \ldots, f_n) > 0,$$

and thus the equation $P_h^{(n-2)}(w) = 0$ has two simple roots, $\lambda_{1,h}$, $\lambda_{2,h}$, which moreover clearly satisfy $\lim_{h \to 0^+} \lambda_{i,h} = \lambda_i$, $1 \leq i \leq 2$, due to the continuity of the formula for the quadratic equations. □

Proposition A.2. *Let $n \geq 3$ $f \in \mathcal{C}^{(n+1)}$ and $z \in \mathbb{R}$ such that $f^{(s)}(z) = 0$ for $n-2 \leq s \leq n-1$ and $f^{(n)}(z) \neq 0$. Let $z_h \in \mathbb{R}$, such that $z_h - z = \mathcal{O}(h)$, the stencil $x_{i,h} = z_h + a_i h$, $0 \leq i \leq n$, $a_i < a_j$ if $i < j$, and consider the interpolating polynomial in Π_n satisfying $p_h(x_{i,h}) = f(x_{i,h}) =: f_{i,h}$. Then, exists $h_0 > 0$ such that for all h, $0 < h < h_0$, the roots $\lambda_{1,h}, \lambda_{2,h} \in \mathbb{C}$ of the parabola $P_h^{(n-2)}$, with P_h given by $P_h(w) := p_h(y + hw)$, satisfy $(\lambda_{2,h} - \lambda_{1,h})^2 = \mathcal{O}(h)$.*

In particular, given the expression of the parabola $P_h^{(n-2)}$ as $P_h^{(n-2)}(w) = A_h w^2 + B_h w + C_h$, there holds

$$B_h^2 - 4 A_h C_h = \mathcal{O}(h^{2n+1}).$$

Proof. By Lemma A.3, there holds

$$P_h^{(n-2)}(w) = \sum_{j=0}^{2} L^j_{\boldsymbol{a}}(f(x_{0,h}), \ldots, f(x_{n,h})) w^j, \tag{A.8}$$

with

$$L^j_{\boldsymbol{a}}(f_{0,h}, \ldots, f_{n,h}) = \frac{1}{j!} h^{n-2+j} f^{(n-2+j)}(z_h) + \mathcal{O}(h^{n+1}), \quad 0 \leq j \leq 2.$$

Now, denoting $A_h = L^2_{\boldsymbol{a}}(f_{0,h}, \ldots, f_{n,h})$, $B_h = L^1_{\boldsymbol{a}}(f_{0,h}, \ldots, f_{n,h})$ and $C_h = L^0_{\boldsymbol{a}}(f_{0,h}, \ldots, f_{n,h})$, using Taylor expansion centered at z and taking into account that $f^{(n-2)}(z) = 0$ and $f^{(n-1)}(z) = 0$, we have, defining $\delta_h := z_h - z = \mathcal{O}(h)$:

$$\begin{aligned}
A_h &= \frac{1}{2} h^n f^{(n)}(z_h) + \mathcal{O}(h^{n+1}) = \frac{1}{2} h^n (f^{(n)}(z) + \mathcal{O}(h)) + \mathcal{O}(h^{n+1}) \\
&= \frac{1}{2} h^n f^{(n)}(z) + \mathcal{O}(h^{n+1}), \\
B_h &= h^{n-1} f^{(n-1)}(z_h) = h^{n-1}(f^{(n-1)}(z) + \delta_h f^{(n)}(z) + \mathcal{O}(h^2)) + \mathcal{O}(h^{n+1}) \\
&= h^{n-1}(\delta_h f^{(n)}(z) + \mathcal{O}(h^2)) + \mathcal{O}(h^{n+1}) = \delta_h h^{n-1} f^{(n)}(z) + \mathcal{O}(h^{n+1}), \\
C_h &= h^{n-2} f^{(n-2)}(z_h) + \mathcal{O}(h^{n+1}) \\
&= h^{n-2} \left(f^{(n-2)}(z) + \delta_h f^{(n-1)}(z) + \frac{\delta_h^2}{2} f^{(n)}(z) + \mathcal{O}(h^3) \right) + \mathcal{O}(h^{n+1})
\end{aligned}$$

$$= h^{n-2}\left(\frac{\delta_h^2}{2}f^{(n)}(z)+\mathcal{O}(h^3)\right)+\mathcal{O}(h^{n+1}) = \frac{1}{2}\delta_h^2 h^{n-2}f^{(n)}(z)+\mathcal{O}(h^{n+1}).$$

Therefore, the discriminant of the quadratic equation reads

$$B_h^2-4A_hC_h = \left(\delta_h h^{n-1}f^{(n)}(z)+\mathcal{O}(h^{n+1})\right)^2 - 4\left(\frac{1}{2}\delta_h^2 h^{n-2}f^{(n)}(z)+\mathcal{O}(h^{n+1})\right)\left(\frac{1}{2}h^n f^{(n)}(z)+\mathcal{O}(h^{n+1})\right).$$

Multiplying all factors and gathering equal powers of h we can verify in a straigtforward manner, taking into consideration that $\delta_h = \mathcal{O}(h)$, that

$$B_h^2-4A_hC_h = \mathcal{O}(h^{2n+1}).$$

Now, since $f^{(n)}(z)\neq 0$, by continuity arguments $\exists h_0>0$ such that $\forall h$, $0<h<h_0$ there holds $A_h\neq 0$. In this case, assuming $0<h<h_0$ and substracting the roots of $P_h^{(n-2)}$ by applying the quadratic formula yields

$$\lambda_{2,h}-\lambda_{1,h} = \frac{B_h+\sqrt{B_h^2-4A_hC_h}}{2A_h}-\frac{B_h-\sqrt{B_h^2-4A_hC_h}}{2A_h} = \frac{\sqrt{B_h^2-4A_hC_h}}{A_h}.$$

Therefore,

$$(\lambda_{2,h}-\lambda_{1,h})^2 = \frac{B_h^2-4A_hC_h}{A_h^2} = \frac{\mathcal{O}(h^{2n+1})}{\bar{\mathcal{O}}(h^n)^2} = \frac{\mathcal{O}(h^{2n+1})}{\bar{\mathcal{O}}(h^{2n})} = \mathcal{O}(h).$$

□

Theorem A.1. *Let $n\geq 3$, $h>0$, $z\in\mathbb{R}$, z_h such that $z_h - z = \mathcal{O}(h)$, $x_{i,h} = z_h + a_i h$, $0\leq i\leq n$, $a_i<a_j$, $i<j$, a stencil, $f_{i,h} := f(x_{i,h})$ for some function f, p_h the interpolating polynomial of the stencil $\{x_{i,h}\}_{i=0}^n$ and $P_h(w) := p_h(z_h+hw)$. Let $P_h^{(n-2)}(w) = A_h w^2 + B_h w + C_h$ be the $(n-2)$-th derivative of P_h. Then there holds:*

$$B_h^2-4A_hC_h = \begin{cases} \bar{\mathcal{O}}(1) & \text{if } \exists h_0>0 : \forall 0<h<h_0, x_{0,h}<z<x_{n,h} \\ & \text{and } f \text{ has a discontinuity at } z, \\ \mathcal{O}(h^{2n+1}) & \text{if } f\in\mathcal{C}^{n+1}, \text{ with } f^{(l)}(z)=0,\ 0\leq l\leq n-1, \\ & \text{and } f^{(n)}(z)\neq 0. \end{cases}$$

Proof. The result is a direct consequence of applying Propositions A.1 and A.2, respectively. □

Remark A.1. *The consideration about analyzing the accuracy of $B_h^2-4A_hC_h$ at a point z_h close to but not necessarily equal to z is very important. Indeed, in the context of ODEs, PDEs, and in particular for hyperbolic conservation laws, it is interesting to analyze whether a scheme looses accuracy or not when a single stencil converges exactly at a critical point, but also within the neighbourhood of that point.*

Remark A.2. *Although we have presented only the results for reconstructions from point values, analogous results can be proven for the case of reconstructions from cell averages. The main difference between them dwells in Lemma A.3, in which the expression is more complicated for a general grid. In [1, Proposition 2], the authors provide an expression for uniform grids, whose procedure can be readily generalized to arbitrary grids, with a more complicated expression.*

Acknowledgements

AB, PM and DZ are supported by Spanish MINECO projects MTM2014-54388-P and MTM2017-83942-P. RB is supported by Fondecyt project 1170473; BASAL project PFB03 CMM, Universidad de Chile and Centro de Investigación en Ingeniería Matemática (CI^{2}MA), Universidad de Concepción; and CRHIAM, project CONICYT/FONDAP/15130015. PM is also supported by Conicyt (Chile), project PAI-MEC, folio 80150006. DZ is also supported by Conicyt (Chile) through Fondecyt project 3170077.

References

[1] F. Aràndiga, A. Baeza, A.M. Belda, and P. Mulet, *Analysis of WENO schemes for full and global accuracy*, SIAM J. Numer. Anal. **49** (2011), 893–915.
[2] F. Aràndiga, M.C. Martí, and P. Mulet, *Weights design for maximal order WENO schemes*, J. Sci. Comput. **60** (2004), 641–659.
[3] R. Borges, M. Carmona, B. Costa, and W. S. Don, *An improved weighted essentially non-oscillatory scheme for hyperbolic conservation laws*. J. Comput. Phys. **227** (2008), 3191–3211.
[4] M. Castro, B. Costa, and W. S. Don, *High order weighted essentially non-oscillatory WENO-Z schemes for hyperbolic conservation laws*. J. Comput. Phys. **230** (2011), 1766–1792.
[5] S. D. Conte, Carl de Boor, *Elementary Numerical Analysis: An Algorithmic Approach*. Third Edition, McGraw-Hill (1980), 65–67.
[6] W.-S. Don and R. Borges, *Accuracy of the weighted essentially non-oscillatory conservative finite difference schemes* J. Comput. Phys. **250** (2013), 347–372.
[7] R. Donat and A. Marquina, *Capturing shock reflections: An improved flux formula*, J. Comput. Phys. **125** (1996), 42–58.
[8] H. Feng, F. Hu, and R. Wang, *A new mapped weighted essentially non-oscillatory scheme*, J. Sci. Comput. **51** (2012), 449–473.
[9] G.A. Gerolymos, D. Sénéchal, and I. Vallet, *Very-high-order WENO schemes*. J. Comput. Phys. **228** (2009), 8481–8524.
[10] Y. Ha, C. H. Kim, Y. J. Lee, J. Yoon, *An improved weighted essentially non-oscillatory scheme with a new smoothness indicator* J. Comput. Phys. **232** (2013), 68–86.
[11] A.K. Henrick, T.D. Aslam, and J.M. Powers, *Mapped weighted essentially non-oscillatory schemes: Achieving optimal order near critical points*, J. Comput. Phys. **207** (2005), 542–567.
[12] P. Holoborodko, *MPFR C++*. Available at http://www.holoborodko.com/pavel/mpfr/
[13] G.S. Jiang and C.-W. Shu, *Efficient implementation of Weighted ENO schemes*, J. Comput. Phys. **126** (1996), 202–228.
[14] O. Kolb, *On the full and global accuracy of a compact third order WENO scheme*, SIAM J. Numer. Anal. **52**(5) (2014), 2335–2355.
[15] R. J. LeVeque, *Numerical methods for conservation laws*, 2^{nd} edition, Birkhäuser, 1992.
[16] X.-D. Liu, S. Osher, and T. Chan, *Weighted essentially non-oscillatory schemes*, J. Comput. Phys. **115** (1994), 200–212.
[17] The GNU MPFR library. http://www.mpfr.org/
[18] C.-W. Shu and S. Osher, *Efficient implementation of essentially non-oscillatory shock-capturing schemes*, J. Comput. Phys. **77** (1988), 439–471.
[19] C.-W. Shu and S. Osher, *Efficient implementation of essentially non-oscillatory shock-capturing schemes, II*, J. Comput. Phys. **83** (1989), 32–78.
[20] N.K. Yamaleev and M.H. Carpenter, *Third-order Energy Stable WENO scheme*, J. Comput. Phys. **228** (2009), 3025–3047.
[21] N.K. Yamaleev and M.H. Carpenter, *A systematic methodology to for constructing high-order energy stable WENO schemes*, J. Comput. Phys. **228** (2009), 4248–4272.
[22] D. Zorío, A. Baeza, and P. Mulet, *An approximate Lax-Wendroff-type procedure for high-order accurate schemes for hyperbolic conservation laws*, J. Sci. Comput. **71** (2017), 246–273.

Centro de Investigación en Ingeniería Matemática (CI^{2}MA)

PRE-PUBLICACIONES 2018

2018-03 Marcelo Cavalcanti, Wellington Correa, Mauricio Sepúlveda, Rodrigo Véjar: *Study of stability and conservative numerical methods for a high order nonlinear Schrödinger equation*

2018-04 Julio Aracena, Maximilien Gadouleau, Adrien Richard, Lilian Salinas: *Fixing monotone Boolean networks asynchronously*

2018-05 Ernesto Cáceres, Gabriel N. Gatica, Filander A. Sequeira: *A mixed virtual element method for a pseudostress-based formulation of linear elasticity*

2018-06 Jessika Camaño, Rodolfo Rodríguez, Pablo Venegas: *Convergence of a lowest-order finite element method for the transmission eigenvalue problem*

2018-07 Javier A. Almonacid, Gabriel N. Gatica, Ricardo Oyarzúa: *A posteriori error analysis of a mixed-primal finite element method for the Boussinesq problem with temperature-dependant viscosity*

2018-08 Gabriel N. Gatica, Bryan Gomez-Vargas, Ricardo Ruiz-Baier: *Formulation and analysis of fully-mixed methods for stress-assisted diffusion problems*

2018-09 Jay Gopalakrishnan, Manuel Solano, Felipe Vargas: *Dispersion analysis of HDG methods*

2018-10 Franco Fagnola, Carlos M. Mora: *Bifurcation analysis of a mean field laser equation*

2018-11 David Mora, Iván Velásquez: *A virtual element method for the transmission eigenvalue problem*

2018-12 Alfredo Bermúdez, Bibiana López-Rodríguez, Rodolfo Rodríguez, Pilar Salgado: *Numerical solution of a transient three-dimensional eddy current model with moving conductors*

2018-13 Raimund Bürger, Enrique D. Fernández Nieto, Victor Osores: *A dynamic multilayer shallow water model for polydisperse sedimentation*

2018-14 Antonio Baeza, Raimund Bürger, Pep Mulet, David Zorío: *Weno reconstructions of unconditionally optimal high order*

Para obtener copias de las Pre-Publicaciones, escribir o llamar a: Director, Centro de Investigación en Ingeniería Matemática, Universidad de Concepción, Casilla 160-C, Concepción, Chile, Tel.: 41-2661324, o bien, visitar la página web del centro: `http://www.ci2ma.udec.cl`







Centro de Investigación en
Ingeniería Matemática (CI²MA)
Universidad de Concepción

Casilla 160-C, Concepción, Chile
Tel.: 56-41-2661324/2661554/2661316
http://www.ci2ma.udec.cl