High Order Weighted Essentially Non-Oscillatory WENO-Z schemes for Hyperbolic Conservation Laws

Marcos Castro ∗ Bruno Costa † Wai Sun Don ‡

July 20, 2010

Abstract

In ([10], JCP 227 No. 6, 2008, pp. 3101–3211), the authors have designed a new fifth order WENO finite-difference scheme by adding a higher order smoothness indicator which is obtained as a simple and inexpensive linear combination of the already existing low order smoothness indicators. Moreover, this new scheme, dubbed as WENO-Z, has a CPU cost which is equivalent to the one of the classical WENO-JS ([2], JCP 126, pp. 202–228 (1996)) and significantly lower than that of the mapped WENO-M, ([5], JCP 207, pp. 542–567 (2005)), since it involves no mapping of the nonlinear weights. In this article, we take a closer look at Taylor expansions of the Lagrangian polynomials of the WENO substencils and the related inherited symmetries of the classical lower order smoothness indicators to obtain a general formula for the higher order smoothness indicators that allows the extension of the WENO-Z scheme to all (odd) orders of accuracy. We further investigate the improved accuracy of the WENO-Z schemes at critical points of smooth solutions as well as their distinct numerical features as a result of the new sets of nonlinear weights and we show that regarding the numerical dissipation WENO-Z occupies an intermediary position between WENO-JS and WENO-M. Some standard numerical experiments such as the one dimensional Riemann initial values problems for the Euler equations and the Mach 3 shock density-wave interaction and the two dimensional double-Mach shock reflection problems are presented.

Keywords
Weighted Essentially Non-Oscillatory, WENO-Z, Smoothness Indicators, Nonlinear Weights

AMS
65P30, 77Axx

∗Department of Mathematics, Hong Kong Baptist University, Hong Kong, China. E-Mail: wdson@hkbu.edu.hk
†Departamento de Matemática Aplicada, IM-UFRJ, Caixa Postal 68530, Rio de Janeiro, RJ, C.E.P. 21945-970, Brazil. E-Mail: bcosta@ufrj.br
‡Departamento de Matemática Aplicada, IM-UFRJ, Caixa Postal 68530, Rio de Janeiro, RJ, C.E.P. 21945-970, Brazil. E-Mail: marcoscastro@ufrj.br
1 Introduction

The weighted essentially non-oscillatory conservative finite difference schemes (WENO) \cite{1, 2} are a popular choice of numerical methods for solving compressible flows modeled by means of hyperbolic conservation laws in the form

$$\frac{\partial u}{\partial t} + \nabla \cdot F(u) = 0,$$

and in the presence of shocks and small scales structures. WENO schemes owe their success to the use of a dynamic set of stencils where a nonlinear convex combination of lower order polynomials adapts either to a higher order polynomial approximation at smooth parts of the solution, or to an upwind spatial discretization that avoids interpolation across discontinuities and provides the necessary dissipation for shock capturing. It is an evolution of the Essentially Non-Oscillatory (ENO) schemes, introduced in \cite{7}, which choose only the smoothest stencil, instead of forming a combination of all the stencils available in order to optimize accuracy.

The local computational grids of ENO and WENO schemes are composed of $r$ overlapping substencils of $r$ points, forming a larger stencil with $(2r - 1)$ points and yielding a local rate of convergence that goes from order $r$, in the case of ENO when only one substencil is used, to order $(2r - 1)$ when the WENO weighted combination is applied at smooth parts of the solution. The nonlinear coefficients of WENO’s convex combination, hereafter referred to as nonlinear weights, are based on lower order local smoothness indicators that measure the sum of the normalized squares of the scaled $L^2$ norms of all derivatives of local interpolating polynomials \cite{2}. An essentially zero weight is assigned to those lower order polynomials whose underlying substencils contain high gradients and/or shocks, aiming at an essentially non-oscillatory solution close to discontinuities. At smooth parts of the solution, higher order is achieved through the mimicking of the central upwind scheme of maximum order, when all smoothness indicators are about the same size. Hence, an efficient and careful design of these smoothness indicators is a delicate and important issue for WENO schemes.

The first set of nonlinear weights of widespread use has been presented in \cite{2}, hereafter denoted as $\omega_k$, where $\beta_k$ are the associated lower order smoothness indicators. The scheme resulting from these will be hereafter referred as the classical WENO scheme (WENO-JS). In \cite{5} it was pointed out that these smoothness indicators were the cause of a reduction of the convergence rate at critical points (points of zero derivatives) of the function. In the same article, a fixing was also proposed in the form of a mapping on the classical WENO-JS weights, leading to corrected weights that recovered the formal order of accuracy at critical points. We call the scheme composed by this mapped set of weights as the mapped WENO scheme (WENO-M). Alternatively, in \cite{10}, it was shown that the incorporation of a global higher order smoothness indicator, which we call $\tau_{2r-1}$ (in order to emphasize the utilization of the whole set of $(2r - 1)$ points available) into the WENO-JS weights definition also improved the convergence at critical points with no need of mapping (in the case of $r = 3$). This last scheme has been named the WENO-Z scheme by its authors. However, a direct comparison of the schemes showed that the improvement on numerical solutions of the WENO-M over the WENO-JS was not due to the increase of the convergence rate at critical points, as it was claimed in \cite{5}, but to the decreased dissipation of WENO-M when correcting the weights to a disposition closer to the central upwind scheme. In the same way, the new set of nonlinear weights of WENO-Z provided even less dissipation than WENO-M, obtaining sharper results among all of the schemes. Moreover, the mapping procedure of WENO-M incurs in extra computational cost, while the WENO-Z modifications are obtained through a simple and inexpensive linear combination.
of the WENO-JS smoothness indicators $\beta_k$.

In this article, we extend the fifth order WENO-Z scheme introduced in [10] to higher orders of accuracy by providing a closed-form formula for the generation of the associated global smoothness indicators $\tau_{2r-1}$ as linear combinations of the classical lower order local smoothness indicators $\beta_k$. This was achieved through a thorough investigation of the Taylor expansions of the $\beta_k$ that revealed symmetries that were used to build the $\tau_{2r-1}$ for all values of $r$. As we shall see, these symmetries originate in the Lagrangian interpolation bases of the local substencils due to their symmetrical dispositions with respect to the global stencil. We also determine the maximum order the higher order smoothness indicator $\tau_{2r-1}$ can achieve as a linear combination of the $\beta_k$.

As mentioned above, WENO’s main idea to avoid oscillations is to generate a convex combination of low order polynomials giving smaller weights to substencils containing discontinuities or high gradients. In [10] it was pointed out that the differences of resolutions among the three schemes, WENO-JS, WENO-M and WENO-Z, were due to their distinct dissipative properties, which in turn were consequence of the sizes of the weights assigned to the discontinuous substencils. Thus, the WENO-Z sharper results that were obtained in [10] were due to the assignment of weights to discontinuous substencils that were larger than the ones of WENO-M. Analogously, in [5], WENO-M also assigned larger weights than the ones of WENO-JS. In this article, we show that the general situation is a little bit distinct from what was shown in [5] and [10], for if we take a closer look at all the parameters involved at the definition of the weights, we conclude that the weights power parameter $p$ (see formula (25)) plays a major role in the amount of dissipation a particular WENO scheme might have, since it affects the weights relative scaling in the presence of discontinuities. We will also see that the schemes can be classified in a dissipation scale with WENO-JS being the most dissipative and the mapped WENO being the less for the same value of $p$, with WENO-Z occupying an intermediary position. For higher values of $r$, this classification determines which scheme keeps the ENO behavior for the values of $p$ currently used. The power parameter $p$ also has an active role on recovering the formal order of accuracy at critical points of the solution and we shall see that this is a distinct property of the WENO-Z scheme.

This paper is organized as follows: In Section 2, brief descriptions of the several WENO schemes are given. The main results of the article will be presented in Section 3, where we derive a general formula for the higher order smoothness indicators $\tau_{2r-1}$ of all odd orders WENO-Z scheme. We also derive a closed-form formula, for arbitrary values of $r$, for the optimal order that $\tau_{2r-1}$ can achieve when expressed as a linear combination of the lower order local smoothness indicators $\beta_k, k = 0, \ldots, r - 1$. The issue of the degradation of the order of convergence at critical points of smooths solutions is discussed in Section 4. In Section 5, dissipation of the various schemes and the influence of the values of the power parameter $p$ are discussed through the numerical simulation of the linear advection of discontinuous functions, the one dimensional Euler equations with Riemann initial values problems and the Mach 3 shock density-wave interaction. The two dimensional double-Mach shock reflection problem was also simulated using the high order WENO-Z finite difference scheme. Concluding remarks are given in Section 6.
2 Weighted essentially non-oscillatory schemes

In this section we describe all three versions of the \((2r-1)\) order weighted essentially non-oscillatory conservative finite difference scheme with \(r \geq 3\) when applied to hyperbolic conservation laws as in (1). Namely, we first recall the essentials of the classical WENO scheme as designed in \([2]\), along with the mapped version introduced in \([5]\) and then the WENO-Z scheme introduced in \([10]\), which we denote as WENO-JS, WENO-M and WENO-Z, respectively. We will recall the fifth \((r = 3)\) as an example in the discussion below.

2.1 The Classical WENO-JS scheme

Consider an uniform grid defined by the points \(x_i = i\Delta x, i = 0, \ldots, N\), which are called cell centers, with cell boundaries given by \(x_{i+\frac{1}{2}} = x_i + \frac{\Delta x}{2}\), where \(\Delta x\) is the uniform grid spacing. The semi-discretized form of (1), by the method of lines, yields a system of ordinary differential equations

\[
\frac{du_i(t)}{dt} = -\frac{\partial f}{\partial x}\bigg|_{x=x_i}, \quad i = 0, \ldots, N,
\]

where \(u_i(t)\) is a numerical approximation to the point value \(u(x_i, t)\).

A conservative finite-difference formulation for hyperbolic conservation laws requires high-order consistent numerical fluxes at the cell boundaries in order to form the flux differences across the uniformly-spaced cells. The conservative property of the spatial discretization is obtained by implicitly defining the numerical flux function \(h(x)\) as

\[
f(x) = \frac{1}{\Delta x} \int_{x - \frac{\Delta x}{2}}^{x + \frac{\Delta x}{2}} h(\xi)d\xi,
\]
such that the spatial derivative in (2) is exactly approximated by a conservative finite difference formula at the cell boundaries,
\[
\frac{du_i(t)}{dt} = \frac{1}{\Delta x} \left( h_{i+\frac{1}{2}} - h_{i-\frac{1}{2}} \right),
\]
where \( h_{i+\frac{1}{2}} = h(x_{i+\frac{1}{2}}) \).

High order polynomial interpolations to \( h_{i+\frac{1}{2}} \) are computed using known grid values of \( f_i = f(x_i) \). The classical \((2r - 1)\)-order WENO scheme uses \((2r - 1)\)-points global stencil, which is subdivided into \( r \) sub-stencils \( \{S_0, S_1, \ldots, S_{r-1}\} \) with each sub-stencil containing \( r \) grid points. For instance, the classical fifth-order WENO scheme uses a 5-points stencil, hereafter named \( S_5 \), which is subdivided into three 3-points sub-stencils \( \{S_0, S_1, S_2\} \), as shown in Fig. 1. The \((2r - 1)\) degree polynomial approximation \( \hat{f}_{i+\frac{1}{2}} = h_{i+\frac{1}{2}} + O(\Delta x^{2r-1}) \) is built through the convex combination of the interpolated values \( \hat{f}^k(x_{i+\frac{1}{2}}) \), in which \( \hat{f}^k(x) \) is the \( r \)-th degree polynomial below, defined in each one of the sub-stencils \( S_k \):
\[
\hat{f}_{i+\frac{1}{2}} = \sum_{k=0}^{r-1} \omega_k \hat{f}^k(x_{i+\frac{1}{2}}),
\]
where
\[
\hat{f}^k(x_{i+\frac{1}{2}}) = \hat{f}^k_{i+\frac{1}{2}} = \sum_{j=0}^{r-1} c_{kj} \hat{f}_{i-k+j}, \quad i = 0, \ldots, N.
\]
The \( c_{kj} \) are Lagrangian interpolation coefficients (see [2]), which depend on the left-shift parameter \( k = 0, \ldots, r - 1 \), but not on the values \( f_i \).

- In the case of \( r = 3 \), it can be shown by Taylor series expansion of (5) that
\[
\hat{f}^k_{i+\frac{1}{2}} = h_{i+\frac{1}{2}} + A_k \Delta x^3 + O(\Delta x^4),
\]
where the values \( A_k \) are independent of \( \Delta x \).

The weights \( \omega_k \) are defined as
\[
\omega_k = \frac{\alpha_k}{\sum_{l=0}^{r-1} \alpha_l}, \quad \alpha_k = \frac{d_k}{(\beta_k + \epsilon)^p}.
\]
We refer to \( \alpha_k \) as the un-normalized weights. The parameter \( \epsilon \) is used to avoid the division by zero in the denominator and power parameter \( p = 2 \) is chosen to increase the difference of scales of distinct weights at non-smooth parts of the solution. The coefficients \( \{d_0, d_1, \ldots, d_{r-1}\} \) are called the ideal weights since they generate the \((2r - 1)\) order central upwind scheme using the \((2r - 1)\)-points stencil. For example, the coefficients \( \{d_0 = \frac{3}{10}, d_1 = \frac{3}{5}, d_2 = \frac{1}{10}\} \) generate the fifth-order central upwind scheme for the 5-points stencil \( S_5 \). Ideal weights for higher order WENO schemes can be found in [2].

The lower order local smoothness indicators \( \beta_k \) measure the regularity of the \((r - 1)\)-th degree polynomial approximation \( \hat{f}^k(x_i) \) at the sub-stencil \( S_k \) and are given by
\[
\beta_k = \sum_{l=1}^{r-1} \Delta x^{2l-1} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \left( \frac{d^l}{dx^l} \hat{f}^k(x) \right)^2 dx.
\]
Let us now check how the classical WENO-JS nonlinear weights \( \omega_\varepsilon \) value for the restrictions above. In [5], it was shown that if the smoothness indicators \( f \) in (7) are given by

\[
\begin{align*}
\beta_0 &= \frac{13}{12} (f_{i-2} - 2f_{i-1} + f_i)^2 + \frac{1}{4} (f_{i-2} - 4f_{i-1} + 3f_i)^2, \\
\beta_1 &= \frac{13}{12} (f_{i-1} - 2f_i + f_{i+1})^2 + \frac{1}{4} (f_{i-1} - f_{i+1})^2, \\
\beta_2 &= \frac{13}{12} (f_i - 2f_{i+1} + f_{i+2})^2 + \frac{1}{4} (3f_i - 4f_{i+1} + f_{i+2})^2,
\end{align*}
\]

and their Taylor series expansions at \( x_i \) are

\[
\begin{align*}
\beta_0 &= f_i^2 \Delta x^2 + \left( \frac{13}{12} f_i''^2 - \frac{2}{3} f_i'''^2 \right) \Delta x^4 - \left( \frac{13}{6} f_i'' f_i''' - \frac{1}{2} f_i''' \right) \Delta x^5 + O(\Delta x^6), \\
\beta_1 &= f_i^2 \Delta x^2 + \left( \frac{13}{12} f_i''^2 + \frac{1}{3} f_i'''^2 \right) \Delta x^4 + O(\Delta x^6), \\
\beta_2 &= f_i^2 \Delta x^2 + \left( \frac{13}{12} f_i''^2 - \frac{2}{3} f_i'''^2 \right) \Delta x^4 + \left( \frac{13}{6} f_i'' f_i''' - \frac{1}{2} f_i''' \right) \Delta x^5 + O(\Delta x^6).
\end{align*}
\]

The general idea of the weights definition (7) is that on smooth parts of the solution the smoothness indicators \( \beta_k \) are all small and about the same size, generating weights \( \omega_k \) that are good approximations to the ideal weights \( d_k \). On the other hand, if the substencil \( S_k \) contains a discontinuity, \( \beta_k \) is \( O(1) \) and the corresponding weight \( \omega_k \) is small relatively to the other weights. This implies that the influence of the polynomial approximation of \( h_{i+\frac{1}{2}} \) taken across the discontinuity is diminished up to the point where the convex combination (4) is essentially non-oscillatory. For instance, in the case \( r = 3 \), Fig. 1 shows the case where substencil \( S_2 \) is discontinuous, yielding \( \beta_0 \) and \( \beta_1 \) to be much smaller than \( \beta_2 \). By (7), this results on \( \omega_2 \) being a small number in (4).

The process synthesized by (4)-(5) is called the WENO reconstruction step, for it reconstructs the values of \( h(x) \) at the cell boundaries of the interval \( I_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \) from its cell averaged values \( f(x) \) in the substencils \( \{S_k, k = 0, \ldots, r - 1\} \). In [5], truncation error analysis of the finite difference equation (3) led to the following necessary and sufficient conditions on the weights \( \omega_k \) for the WENO scheme to achieve the formal \( (2r - 1) \) order of convergence at smooth parts of the solution:

\[
\begin{align*}
\sum_{k=0}^{r-1} A_k (\omega_k^+ - \omega_k^-) &= O(\Delta x^r), \\
\omega_k^\pm - d_k &= O(\Delta x^{r-1}).
\end{align*}
\]

In the case \( r = 3 \), we see that a sufficient condition for fifth-order convergence is simply given by:

\[
\omega_k^\pm - d_k = O(\Delta x^3).
\]

It was also found that at first order critical points \( x_c \), points where the first derivative of the function vanishes \( (f'(x_c) = 0) \), the rate of convergence degraded to only third order \( (O(\Delta x^3)) \), a fact that was hidden by the homogenization of the weights caused by the use of a relatively large value for \( \varepsilon \) in (7).

Let us now check how the classical WENO-JS nonlinear weights \( \omega_k \) in (7) behave with respect to the restrictions above. In [5], it was shown that if the smoothness indicators \( \beta_k \) satisfy \( \beta_k =
\(D(1 + O(\Delta x^2))\), then the weights \(\omega_k\) satisfy \(\omega_k = d_k + O(\Delta x^2)\), where \(D\) is a nonzero constant independent of \(k\). Looking at the Taylor series expansions of the smoothness indicators \(\beta_k\) in (12)-(14), we see that \(\beta_k = D(1 + O(\Delta x^2))\), implying that \(\omega_k = d_k + O(\Delta x^2)\). This requires that condition (15) must be satisfied as well for the classical WENO to have the expected fifth-order convergence, which indeed happens and can be easily confirmed with any symbolic calculation software.

Nevertheless, at critical points this situation becomes more complex depending on the number of vanishing derivatives of \(f(x_i)\). For instance, if only the first derivative vanishes, then \(\beta_k = D(1 + O(\Delta x^2))\) and \(\omega_k = d_k + O(\Delta x^2)\), degrading the order of convergence of the scheme to third order only. If the second derivative also vanishes, then the rate of convergence decreases even more to second order.

\[2.2\] The Mapped WENO-M scheme

A fix to this deficiency of the classical weights \(\omega_k\) was proposed in [5]. It consisted of the application of a mapping function that increased the approximation of \(\omega_k\) to the ideal weights \(d_k\) at critical points to the required third order \(O(\Delta x^3)\) as in (17). The mapping function \(g_k(\omega)\) used in [5] is defined as

\[
g_k(\omega) = \frac{\omega (d_k + d_k^2 - 3d_k\omega + \omega^2)}{d_k^2 + \omega (1 - 2d_k)}, \tag{18}
\]

and is a non-decreasing monotone function with the following properties:

1. \(0 \leq g_k(\omega) \leq 1\), \(g_k(0) = 0\) and \(g_k(1) = 1\).
2. \(g_k(\omega) \approx 0\) if \(\omega \approx 0\); \(g_k(\omega) \approx 1\) if \(\omega \approx 1\).
3. \(g_k(d_k) = d_k\), \(g_k'(d_k) = g_k''(d_k) = 0\).
4. \(g_k(\omega) = d_k + O(\Delta x^3)\), if \(\omega = d_k + O(\Delta x^2)\).

Numerical results in [5] confirmed the usefulness of the mapping, since with the modified weights the resulting WENO-M scheme recovered the formal fifth-order convergence at critical points of a smooth solution. Note, however, that if at a critical point the second derivative also vanishes, \(\beta_k = D(1+O(1))\), implying \(\omega_k = d_k+O(1)\) (see equations (12)-(14)) and the mapping is unable to improve the weights approximation, maintaining the same second order of convergence as the classical WENO-JS scheme. The downside of using the mapping function is an additional \(\approx 20\%-30\%\) cost of cpu time when compared to the classical WENO-JS scheme.

\[2.3\] The WENO-Z Scheme

The novel idea of the WENO-Z scheme introduced in [10] is the modification of the \(\beta_k\) with information obtained from a higher order smoothness indicator, which we denote here by \(\tau_{2r-1}\) for any given order \((2r-1), r \geq 3\). This new smoothness indicator is built using the values of the numerical solution at the whole \((2r-1)\) points stencil in the form of a simple linear combination
of the $\beta_k$. For instance, as shown in [10], for $r = 3$, $\tau_5$ is simply defined as the absolute difference between $\beta_0$ and $\beta_2$ at $x_i$, namely

$$\tau_5 = |\beta_0 - \beta_2|.$$  \hspace{1cm} (19)

It is straightforward to see from (12)-(14) that the truncation error of $\tau_5$ is

$$\frac{13}{3} |f_i'' f_i'''| \Delta x^5 + O(\Delta x^6),$$  \hspace{1cm} (20)

and that it is a measure of higher derivatives of $f$, when they exist, and is indeed computed using the whole 5-points stencil $S^5$. The relevant properties of $\tau_5$ to be used in the redefinition of the WENO weights are:

- If $S^5$ does not contain discontinuities, then $\tau_5 = O(\Delta x^5) \ll \beta_k$, for $k = 0, 1, 2$;
- if the solution is continuous at some of the $S_k$, but discontinuous in the whole $S^5$, then $\beta_k \ll \tau_5$, for those $k$ where the solution is continuous;
- $\tau_5 \leq \max_k \beta_k$.

The lower order local smoothness indicators $\beta^Z_k$ of the WENO-Z scheme are then defined with the help of $\tau_5$ as

$$\beta^Z_k = \left( \frac{\beta_k + \epsilon}{\beta_k + \tau_5 + \epsilon} \right), \quad k = 0, 1, 2,$$  \hspace{1cm} (21)

and the new normalized nonlinear weights $\omega^Z_k$ and the un-normalized nonlinear weights $\alpha^Z_k$ becomes

$$\omega^Z_k = \frac{\alpha^Z_k}{\sum_{l=0}^{2^r-1} \alpha^Z_l}, \quad \omega^Z_k = \frac{d_k}{\beta^Z_k} = d_k \left( 1 + \frac{\tau_5}{\beta_k + \epsilon} \right), \quad k = 0, 1, 2,$$  \hspace{1cm} (22)

where $\epsilon$ is, as usual, a small number used to avoid the division by zero in the denominators of (21) and (22).

It is straightforward to check from (12)-(14) and the properties of $\tau_5$ that, at smooth parts of the solution,

$$\frac{\tau_5}{\beta_k + \epsilon} = O(\Delta x^3), \quad k = 0, 1, 2,$$  \hspace{1cm} (23)

whenever $\epsilon <<< \beta_k$, and from (22),

$$\omega^Z_k = d_k + O(\Delta x^3), \quad k = 0, 1, 2.$$  \hspace{1cm} (24)

Thus, the new weights $\omega^Z_k$ satisfy the sufficient condition (17), providing the formal fifth order of accuracy to the WENO-Z scheme at noncritical points of a smooth solution.

The general definitions of the normalized and un-normalized nonlinear weights $\omega^Z_k$ and $\alpha^Z_k$, respectively, for $r \geq 3$ are respectively:

$$\omega^Z_k = \frac{\alpha^Z_k}{\sum_{l=0}^{2^r-1} \alpha^Z_l}, \quad \alpha^Z_k = \frac{d_k}{\beta^Z_k} = d_k \left( 1 + \frac{\tau_{2^r-1}}{\beta_k + \epsilon} \right)^p, \quad k = 0, \ldots, r - 1,$$  \hspace{1cm} (25)

where $p \geq 1$ is the power parameter, used to enhance the ratio between the smoothness indicators to guarantee convergence at a certain order. Analogously, we need

$$\omega^Z_k = d_k + O(\Delta x^r), \quad k = 0, 1, \ldots, r - 1,$$  \hspace{1cm} (26)
to obtain convergence of the WENO-Z at order \((2r - 1)\) and this is achieved if

\[
\frac{\tau_{2r-1}}{\beta_k + \epsilon} = O(\Delta x^r), \quad k = 0, \ldots, r - 1.
\]

(27)

As we shall see below, the \(\beta_k\) are always \(O(\Delta x^2)\), therefore, we need \(\tau_{2r-1} = O(\Delta x^{r+2})\). In the next section, we provide a closed-form formula for generating such a \(\tau_{2r-1}\) for all values of \(r\) as a linear combination of the \(\beta_k\).

**Remark 1** At critical points, the lower order local smoothness indicators \(\beta_k\) are no longer \(O(\Delta x^2)\) and the size of \(\tau_{2r-1}\) also varies in a way that the order of convergence of the ratio in (23) decreases monotonically from \(r\) to 1 as \(n_{cp}\), the order of the critical point, goes from 0 to \(r - 1\). The power parameter \(p\) can be used to recover this order. For instance, as shown in [10], for \(r = 3\), at a first order critical point the convergence of the WENO-Z scheme degrades to fourth order if \(p = 1\) in (25) and regains fifth order when \(p = 2\). This is unique to WENO-Z, changing the value of \(p\) in WENO-JS or in WENO-M does not alter their convergence rate at critical points. Numerical experiments at Section 4 further illustrate this property and more detailed computations that demonstrate this exclusive aspect of WENO-Z are shown in [10].

### 3 The global higher order smoothness indicators

In this section we formulate and prove the necessary theoretical results to obtain a general formula for the global higher order smoothness indicators \(\tau_{2r-1}\) for all values of \(r\). The proof of the main theorem explores symmetric structures of the underlying interpolating polynomials defining the \(\beta_k\). This will be done with the help of two lemmas and an auxiliary theorem. We also determine the existence for all \(r\), although no closed formula is provided, of an improved higher order smoothness indicator, \(\tau_{2r-1}^{opt}\), which has the optimal order among all the linear combinations of the \(\beta_k\). We shall see that the use of \(\tau_{2r-1}^{opt}\) in place of \(\tau_{2r-1}\) in the weights definition improves the ability of the WENO-Z scheme to capture higher order structures in the numerical solution.

#### 3.1 General Formula for \(\tau_{2r-1}\)

**Theorem 2** Given the order of the WENO reconstruction \((2r - 1)\) and the associated lower order local smoothness indicators \(\beta_0, \ldots, \beta_{r-1}\), the global higher order smoothness indicator \(\tau_{2r-1}\), defined as:

\[
\tau_{2r-1} = \begin{cases} 
|\beta_0 - \beta_{r-1}| & \text{mod } (r, 2) = 1 \\
|\beta_0 - \beta_1 - \beta_{r-2} + \beta_{r-1}| & \text{mod } (r, 2) = 0 
\end{cases},
\]

(28)

is of order \(O(\Delta x^{r+2})\).

Before we give the proof of the theorem at the end of this section, we will give some preliminary results and two lemmas that are necessary for the proof. Without loss of generality, we take \(i = 0\) \((x_0 = 0)\) and denote \(\frac{d^nf}{dx^n}f(x_0) = f_0^{(n)}\) unless stated otherwise. Furthermore, we shall denote the
lower order polynomial approximation \( \hat{f}_k(x) \) of degree \((r - 1)\) in the substencil \( S_k \) as \( p_k(x) \):

\[
p_k(x) = \sum_{j=0}^{r-1} a_{k,j} x^j,
\]

(29)

where \(a_{k,j}\) are the coefficients of the polynomial expansion of \( f(x) \) about the point \( x_0 \) in the substencil \( S_k \).

The \( n \)-th derivative of \( p_k(x) \), \( p_k^{(n)}(x) \), is given by

\[
p_k^{(n)}(x) = \sum_{j=0}^{r-1-n} b_{k,n,j} x^j, \quad 1 \leq n \leq r - 1,
\]

(30)

where \( b_{k,n,j} = \frac{(j + n)!}{j!} a_{k,n+j} \).

Thus, we may write \( \beta_k \) as

\[
\beta_k = \sum_{n=1}^{r-1} \Delta x^{2n-1} \int_{-\frac{1}{2}\Delta x}^{\frac{1}{2}\Delta x} \left( \sum_{j=0}^{r-1-n} b_{k,n,j} x^j \right)^2 dx
\]

\[
= \sum_{n=1}^{r-1} \Delta x^{2n-1} \int_{-\frac{1}{2}\Delta x}^{\frac{1}{2}\Delta x} \left( \sum_{j_1=0}^{r-1-n} b_{k,n,j_1} x^{j_1} \right) \left( \sum_{j_2=0}^{r-1-n} b_{k,n,j_2} x^{j_2} \right) dx
\]

\[
= \sum_{n=1}^{r-1} \Delta x^{2n-1} \int_{-\frac{1}{2}\Delta x}^{\frac{1}{2}\Delta x} \sum_{j_1=0}^{r-1-n} b_{k,n,j_1} \sum_{j_2=0}^{r-1-n} b_{k,n,j_2} x^{j_1+j_2} dx
\]

\[
= \sum_{n=1}^{r-1} \sum_{j_1=0}^{r-1-n} \sum_{j_2=0}^{r-1-n} C_{n,j_1,j_2} a_{k,n+j_1} a_{k,n+j_2} \Delta x^{j_1+j_2+2n},
\]

(31)

where

\[
C_{n,j_1,j_2} = \begin{cases} 
\frac{(j_1+n)!}{j_1!} & \text{mod } (j_1+j_2, 2) = 0, \\
\frac{(j_2+n)!}{j_2!} & \text{mod } (j_1+j_2, 2) = 1, \\
0 & \text{mod } (j_1+j_2, 2) = 0, \ldots, r - 1.
\end{cases}
\]

The idea of the proof of Theorem 2 is to rewrite \( \beta_k \) in (31) as another asymptotic expansion in \( \Delta x \) where the coefficients show an anti-symmetric behavior with respect to the substencils index \( k \). We start with the following two lemmas below, which express the numerical flux \( h(x) \) in terms of the derivatives of the physical flux function \( f(x) \) and establish the independence and anti-symmetry properties of the polynomial coefficients \( a_{k,j} \).

**Lemma 3** Consider the primitive function \( h(x) \), the numerical flux of \( f(x) \), as defined in (3), then

\[
h(x) = \sum_{\delta=0}^{\infty} \phi_{2\delta} f^{(2\delta)}(x) \Delta x^{2\delta},
\]

(32)

where \( \phi_0 = 1 \) and \( \phi_{2\delta}, \delta = 1, \ldots \) are constants in the expansion.
Proof. From (3), and expanding \( h_{\pm \frac{1}{2}} = h(x_{\pm \frac{1}{2}}) \) about \( x_0 = 0 \),

\[
   f' = h' + \Delta x^2 \frac{1}{4} h^{(3)} + \Delta x^3 \frac{1}{16} \frac{1}{5} h^{(5)} + \cdots, \tag{33}
\]

Integrating both sides,

\[
   f = h + \Delta x^2 \frac{1}{4} h^{(2)} + \Delta x^3 \frac{1}{16} \frac{1}{5} h^{(4)} + \cdots, \tag{34}
\]

and differentiating (33),

\[
   f^{(2)} = h^{(2)} + \Delta x^2 h^{(4)} + \cdots, \quad f^{(4)} = h^{(4)} + \Delta x^2 h^{(6)} + \cdots.
\]

Substituting \( h^{(2)} \) and \( h^{(4)} \) in (34) yields, after some algebra,

\[
   h = f - \frac{1}{24} f^{(2)} \Delta x^2 + \frac{7}{5760} f^{(4)} \Delta x^4 + \cdots.
\]

We have the constants \( \phi_{2j} \) as shown in the expansion with \( \phi_0 = 1 \).

The above process can be repeated to replace the higher order derivatives of \( h(x) \) by \( f(x) \) and to obtain (32). ■

Lemma 4 Let \( \rho_j = \left\lfloor \frac{r-j-1}{2} \right\rfloor \). If \( x \in [x_{-\frac{1}{2}}, x_{\frac{1}{2}}] \), then \( p_k(x) \) in (29) can be written as,

\[
   p_k(x) = \sum_{j=0}^{r-1} a_{k,j} x^j, \tag{35}
\]

where the coefficients \( a_{k,j} \) are expressed either as:

(a)

\[
   a_{k,j} = \frac{1}{j!} \sum_{\delta=0}^{\rho_j} \phi_{2\delta} f_0^{(j+2\delta)} \Delta x^{2\delta} + O(\Delta x^{r-j}), \tag{36}
\]

with the coefficients \( \phi_{2\delta} \) as given in lemma 3,

(b) Or alternatively as:

\[
   a_{k,j} = \sum_{l=0}^{\infty} \sigma_{k,j,l} f_0^{(j+l)} \Delta x^l, \tag{37}
\]

where

\[
   \sigma_{k,j,l} = (-1)^l \sigma_{r-1, k-j, l}, \tag{38}
\]

and, for \( l \leq \rho_j \), the above formula is independent of \( k \), that is

\[
   \sigma_{k,j,l} = \begin{cases} \frac{1}{j!} \phi_l & \text{mod } (l, 2) = 0 \\ 0 & \text{mod } (l, 2) = 1 \end{cases}. \tag{39}
\]
Proof. $p_k$ is a $(r - 1)$-th degree polynomial approximation of $h$:

$$p_k(x) = h(x) + O(\Delta x^r).$$

(40)

Combining equations (32) and (40) and expanding $f$ and its derivatives in Taylor series around $x_0 = 0$, we have

$$p_k(x) = \sum_{j=0}^{\infty} \frac{1}{j!} \sum_{\delta=0}^{\rho_j} \phi_{2\delta} f^{(j+2\delta)}(x_0) (x-x_0)^j \Delta x^{2\delta} + O(\Delta x^r).$$

One can see that the terms in the second summation are of $O(\Delta x^{j+2\delta})$ since $(x-x_0) = O(\Delta x)$.

For $j + 2\delta > r - 1$, or $l > \rho_j = \left\lfloor \frac{r-j-1}{2} \right\rfloor$, one has $(x-x_0)^j \Delta x^{2\delta} \leq O(\Delta x^r)$, therefore

$$p_k(x) = \sum_{j=0}^{r-1} \frac{1}{j!} \sum_{\delta=0}^{\rho_j} \phi_{2\delta} f^{(j+2\delta)}(x_0) (x-x_0)^j \Delta x^{2\delta} + O(\Delta x^r),$$

(41)

with $\phi_0 = 1$.

Hence

$$p_k(x) = \sum_{j=0}^{r-1} a_{k,j} (x-x_0)^j,$$

(42)

with

$$a_{k,j} = \frac{1}{j!} \sum_{\delta=0}^{\rho_j} \phi_{2\delta} f^{(j+2\delta)}(0) \Delta x^{2\delta} + O(\Delta x^{r-j}).$$

Proof of (b): Our objective is to build $(r - 1)$-th degree polynomial approximations of $h$ in the substencils $S_k$ and $S_{r-1-k}$ in order to obtain the symmetry in (38). For that we use the functions $F_k$ and $G_k$, primitives of $h$ defined as

$$F_k(x) = \int_{x[k-(r-1)]}^{x} h(\xi) d\xi, \quad \text{and} \quad G_k(x) = -\int_{x}^{x[(r-1)-k]+\frac{1}{2}} h(\xi) d\xi.$$

We rewrite these as:

$$F_k(x_{i+\frac{1}{2}}) = \sum_{s=k-(r-1)}^{i} \int_{x_{s-\frac{1}{2}}}^{x_{s+\frac{1}{2}}} h(\xi) d\xi = \sum_{s=k-(r-1)}^{i} f_s \Delta x, \quad i \in \{k-(r-1), ..., k\},$$

(43)

$$G_k(x_{i+\frac{1}{2}}) = -\sum_{s=i+1}^{(r-1)-k} f_s \Delta x, \quad i \in \{-k, ..., (r-1)-k\},$$

where $f_s = \int_{x_{s-\frac{1}{2}}}^{x_{s+\frac{1}{2}}} h(\xi) d\xi$.

Let $P_k^F(x)$ be the only polynomial of degree less than or equal to $r$ that interpolates $F_k(x)$ in the $r + 1$ points $x_{i+\frac{1}{2}}, i \in \{k-r, ..., k\}$, and similarly, let $P_k^G(x)$ be the only polynomial of degree less
than or equal to $r$ that interpolates $G_k(x)$ in the $r + 1$ points $x_{i + \frac{1}{2}}, i \in \{-k - 1, \ldots, (r - 1) - k\}$.

One can show that (see [3] Shu)

$$h(x) = F_k'(x) + O(\Delta x^r), \quad x \in [x_{[k-(r-1)]-\frac{1}{2}}, x_{k+\frac{1}{2}}],$$
$$h(x) = G_k'(x) + O(\Delta x^r), \quad x \in [x_{-k-\frac{1}{2}}, x_{(r-1)-k+\frac{1}{2}}],$$

and thus

$$p_k(x) = (P_k^F)'(x), \quad x \in [x_{[k-(r-1)]-\frac{1}{2}}, x_{k+\frac{1}{2}}],$$
$$p_{(r-1)-k}(x) = (P_k^G)'(x), \quad x \in [x_{-k-\frac{1}{2}}, x_{(r-1)-k+\frac{1}{2}}],$$

are the approximations we are searching for. We can rewrite $P_k^F(x)$ and $P_k^G(x)$ using the following Lagrangian interpolation formulas:

$$P_k^F(x) = \sum_{i=k-r}^{k} F_k(x_{i+\frac{1}{2}}) L_{k,i}(x),$$
$$P_k^G(x) = \sum_{i=-k-1}^{(r-1)-k} G_k(x_{i+\frac{1}{2}}) L_{(r-1)-k,i}(x),$$

with

$$L_{k,i}(x) = \prod_{s=k-r, s \neq i}^{k} \frac{x - x_{s+\frac{1}{2}}}{x_{i+\frac{1}{2}} - x_{s+\frac{1}{2}}} = \prod_{s=k-r, s \neq i}^{k} \frac{x - (s + \frac{1}{2})\Delta x}{(i - s)\Delta x}$$

$$= \prod_{s=k-r, s \neq i}^{k} \frac{(x - (s + \frac{1}{2})\Delta x)}{(i - s)\Delta x} = \sum_{j=0}^{r} c_{k,i,j} \Delta x^{r-j} x^j$$

It is also easily seen that

$$L_{k,i}(x) = L_{(r-1)-k,-i-1}(-x).$$

(44)
Using (43) and (44), we obtain

\[
P^F_k(x) = \sum_{i=k-r}^{k} \sum_{s=k-(r-1)}^{i} f_s \Delta x L_{k,i}(x) = \sum_{i=k-r}^{k} \sum_{j=0}^{(r-1)-k} f_s \sum_{j=0}^{(r-1)-k} c_{k,i,j} \Delta x^{-(j-1)} x^j
\]

\[
P^G_k(x) = \sum_{i=-k-1}^{k} \sum_{s=i+1}^{k} -f_s \Delta x L_{(r-1)-k,i}(x) = \sum_{i=k-r}^{k} \sum_{s=-i}^{k} -f_s \Delta x L_{k,i-1}(-x)
\]

Moreover, differentiating the equalities above we obtain

\[
p_k(x) = \sum_{i=k-r}^{k} \sum_{s=k-(r-1)}^{i} f_s c_{k,i,j} \Delta x^{-(j-1)} jx^{j-1},
\]

\[
p_{(r-1)-k}(x) = \sum_{i=k-r}^{k} \sum_{s=k-(r-1)}^{i} -f_s c_{k,i,j} \Delta x^{-(j-1)} (-1)^j jx^{j-1}.
\]

Defining \( \Gamma_{k,i,j} = j c_{k,i,j} \) and reorganizing the indexes, we arrive at the following expressions for the interpolating polynomials of \( h \) in the subcells \( S_k \) and \( S_{(r-1)-k} \):

\[
p_k(x) = \sum_{j=0}^{r-1} \left( \frac{\sum_{s=k-(r-1)}^{s} \Gamma_{k,i,j}}{\Delta x^j} \right) f_s x^j
\]

\[
p_{(r-1)-k}(x) = \sum_{j=0}^{r-1} \left( \frac{\sum_{s=k-(r-1)}^{s} \Gamma_{k,i,j}}{\Delta x^j} (-1)^j \right) f_s x^j
\]
Thus, their respective coefficients are given by

\[
\begin{align*}
\sigma_{k,j,l} &= \sum_{s=k-(r-1)}^{k} \frac{\sum_{i=s}^{k} \Gamma_{k,i,j} f_s}{\Delta x^j} \\
&= \sum_{s=k-(r-1)}^{k} \frac{\sum_{i=s}^{k} \Gamma_{k,i,j} \sum_{\delta=0}^{\infty} \frac{1}{\delta!} f_0^{(\delta)} s^\delta \Delta x^\delta}{\Delta x^j}, \\
\sigma_{r-1-k,j,l} &= \sum_{s=k-(r-1)}^{k} \frac{\sum_{i=s}^{k} \Gamma_{k,i,j} \sum_{\delta=0}^{\infty} \frac{1}{\delta!} f_0^{(\delta)} (-1)^{\delta-j} s^\delta \Delta x^\delta}{\Delta x^j}.
\end{align*}
\]

where above we expanded \( f_s \) in Taylor series over the origin and reorganized the summations in terms of \( \Delta x \). We now define \( l = \delta - j \) and note from (36) that \( \delta - j \geq 0 \) to arrive at

\[
\begin{align*}
\sigma_{k,j,l} &= \sum_{l=0}^{\infty} \left( \sum_{s=k-(r-1)}^{k} \left( \sum_{i=s}^{k} \Gamma_{k,i,j} \frac{s^j+l}{(j+l)!} f_0^{(j+l)} \right) \right) \Delta x^l, \\
\sigma_{r-1-k,j,l} &= \sum_{l=0}^{\infty} \left( \sum_{s=k-(r-1)}^{k} \left( \sum_{i=s}^{k} \Gamma_{k,i,j} \frac{s^j+l}{(j+l)!} f_0^{(j+l)} (-1)^l \right) \right) \Delta x^l.
\end{align*}
\]

Thus, comparing both expressions we see that

\[
\begin{align*}
\sigma_{k,j,l} &= \sum_{s=k-(r-1)}^{k} \sum_{i=s}^{k} \Gamma_{k,i,j} \frac{s^j+l}{(j+l)!}, \\
\sigma_{r-1-k,j,l} &= \sum_{s=k-(r-1)}^{k} \sum_{i=s}^{k} \Gamma_{k,i,j} \frac{s^j+l}{(j+l)!} (-1)^l,
\end{align*}
\]

satisfy \( \sigma_{k,j,l} = (-1)^l \sigma_{r-1-k,j,l} \). 

**Definition 5** In order to simplify the notation we define:

\[
M^* = \left\lfloor \frac{M}{2} \right\rfloor,
\]

and

\[
E(k, n, j_1, j_2, l_1, l_2) = C_{n,j_1,j_2} \sigma_{k,n+j_1,l_1} \sigma_{k,n+j_2,l_2}.
\]
Using (38), we obtain the anti-symmetry/symmetry condition
\[ E(k, n, j_1, j_2, l_1, l_2) = (-1)^{l_1 + l_2} E(r - 1 - k, n, j_1, j_2, l_1, l_2). \] (47)

We now state and proof the new asymptotic expansion for the lower order local smoothness indicators $\beta_k$.

**Theorem 6** The lower order local smoothness indicators $\beta_k$ can be written as
\[ \beta_k = \sum_{M=2}^{\infty} \sum_{j=1}^{M^*} A_{k,M,j} f_0^{(j)} f_0^{(M-j)} \Delta x^M, \] (48)
with
\[ A_{k,M,j} = (-1)^M A_{(r-1)-k,M,j}. \] (49)
In addition, if $j < r$ and $M - j < r$, $A_{k,M,j}$ is independent of $k$, that is,
\[ A_{0,M,j} = A_{1,M,j} = \cdots = A_{r-1,M,j}. \]

**Proof.** Using the definition of $a_{k,j}$ in (37), one has
\[ a_{k,n+j_1} a_{k,n+j_2} = \sum_{l_1=0}^{\infty} \sum_{l_2=0}^{\infty} \sigma_{k,n+j_1,l_1} \sigma_{k,n+j_2,l_2} f_0^{(n+j_1+l_1)} f_0^{(n+j_2+l_2)} \Delta x^{l_1+l_2}. \]
Substituting (46) into (31), we obtain
\[ \beta_k = \sum_{n=1}^{r-1} \sum_{j_1,j_2=0}^{r-1-n} \sum_{l_1=0}^{\infty} \sum_{l_2=0}^{\infty} E(k, n, j_1, j_2, l_1, l_2) f_0^{(n+j_1+l_1)} f_0^{(n+j_2+l_2)} \Delta x^{j_1+j_2+2n+l_1+l_2}. \]
Fixing $M_1 = n + j_1 + l_1$ and $M_2 = n + j_2 + l_2$, we can reorganize the sum as
\[ \beta_k = \sum_{M_1=1}^{\infty} \sum_{M_2=1}^{\infty} \sum_{\Omega_{M_1,M_2}} E(k, n, j_1, j_2, l_1, l_2) f_0^{(M_1)} f_0^{(M_2)} \Delta x^{M_1+M_2}, \]
where
\[ \Omega_{M_1,M_2} = \{(n,j_1,j_2,l_1,l_2) \in \bigcup_{n=1}^{r-1} \Psi_n \mid n + j_1 + l_1 = M_1 \text{ and } n + j_2 + l_2 = M_2 \}, \]
with
\[ \Psi_n = \{1, \ldots, r-1\} \times \{1, \ldots, r-1-n\}^2 \times \mathbb{N}^2. \]
Defining $M = M_1 + M_2$, we have
\[ \beta_k = \sum_{M=2}^{\infty} \sum_{j=1}^{M-1} \sum_{\Omega_{j,M-j}} E(k, n, j_1, j_2, l_1, l_2) f_0^{(j)} f_0^{(M-j)} \Delta x^M. \]
Since \( f_0^{(j)} f_0^{(M-j)} = f_0^{(M-(M-j))} f_0^{(M-j)} \), we can reorganize the sum above as

\[
\beta_k = \sum_{M=2}^{\infty} \sum_{j=1}^{M^*} \sum_{\Omega_{j,M-j} \cup \Omega_{M-j,j}} E(k, n, j_1, j_2, l_1, l_2) f_0^{(j)} f_0^{(M-j)} \Delta x^M.
\]

By defining

\[
A_{k,M,j} = \sum_{\Omega_{j,M-j} \cup \Omega_{M-j,j}} E(k, n, j_1, j_2, l_1, l_2),
\]

we have

\[
\beta_k = \sum_{M=2}^{\infty} \sum_{j=1}^{M^*} A_{k,M,j} f_0^{(j)} f_0^{(M-j)} \Delta x^M.
\]

Using the property (47) and the fact that \( j_1 + j_2 \) is always even (otherwise \( C_{n,j_1,j_2} = 0 \), we have

\[
\sum_{\Omega_{j,M-j} \cup \Omega_{M-j,j}} E(k, n, j_1, j_2, l_1, l_2) = \sum_{\Omega_{j,M-j} \cup \Omega_{M-j,j}} (-1)^{l_1 + l_2} E(r - 1 - k, n, j_1, j_2, l_1, l_2)
\]

\[
= \sum_{\Omega_{j,M-j} \cup \Omega_{M-j,j}} (-1)^{2n + j_1 + j_2 + l_1 + l_2} E(r - 1 - k, n, j_1, j_2, l_1, l_2)
\]

\[
= \sum_{\Omega_{j,M-j} \cup \Omega_{M-j,j}} (-1)^M E(r - 1 - k, n, j_1, j_2, l_1, l_2),
\]

yielding

\[
A_{k,M,j} = (-1)^M A_{(r-1)-k,M,j}.
\]

We can also see that if \( j < r \) and \( M - j < r \), \( A_{k,M,j} \) is independent of \( k \). Indeed, using (31) and (36),

\[
\beta_k = \sum_{n=1}^{r-1} \sum_{j_1, j_2=0}^{r-1-n} C_{n,j_1,j_2} \left( \sum_{\delta_1=0}^{\rho_{n+j_1}} \sum_{\delta_2=0}^{\rho_{n+j_2}} \phi_{2\delta_1} \phi_{2\delta_2} \Delta x^{2\delta_1 + 2\delta_2} \right) f_0^{(n+j_1 + 2\delta_1)} f_0^{(n+j_2 + 2\delta_2)}
\]

\[
= O(\Delta x^{r-\max(j_1,j_2)}) \Delta x^{j_1+j_2+2n}.
\]

The formula above shows that the coefficients associated to \( f_0^{(n+j_1 + 2\delta_1)} f_0^{(n+j_2 + 2\delta_2)} \) are independent of \( k \). Therefore the largest possible value for \( n + j_1 + 2\delta_1 \) (and for \( n + j_2 + 2\delta_2 \)) is

\[
n + j_1 + 2\rho_{n+j_1} = n + j_1 + 2 \left[ \frac{r - 1 - (n + j_1)}{2} \right] = \left\{ \begin{array}{ll} r - 1 & \text{mod } (r - 1 - (n + j_1), 2) = 0 \\ r - 2 & \text{mod } (r - 1 - (n + j_1), 2) = 1 \end{array} \right.
\]

Thus if \( j < r \) and \( M - j < r \), the coefficient \( A_{k,M,j} \) is independent of \( k \).

**Corollary 7** If \( M \leq r \), \( A_{k,M,j}, \ j = 1, \ldots, M^* \) is independent of the shifting parameter \( k \).
We are now ready to prove Theorem 2:

**Proof.** Using the definition of $\beta_k$ in (48) and the symmetry condition in (49), one has

$$\beta_0 = \sum_{M=2}^{M^*} \sum_{j=1}^{M^*} A_{0,M,j} f_0^{(j)} f_0^{(M-j)} \Delta x^M, \quad \beta_{r-1} = \sum_{M=2}^{M^*} \sum_{j=1}^{M^*} A_{0,M,j} f_0^{(j)} f_0^{(M-j)} \Delta x^M (-1)^M,$$

$$\beta_1 = \sum_{M=2}^{M^*} \sum_{j=1}^{M^*} A_{1,M,j} f_0^{(j)} f_0^{(M-j)} \Delta x^M, \quad \beta_{r-2} = \sum_{M=2}^{M^*} \sum_{j=1}^{M^*} A_{1,M,j} f_0^{(j)} f_0^{(M-j)} \Delta x^M (-1)^M.$$

We seek to find a linear combination of the above $\beta_k$ in such a way that the sum of all the terms of $O(\Delta x^k)$ is zero for all $k \leq r + 1$. In other words, we seek non-trivial constants $c_0, c_1, c_{r-2}, c_{r-1}$ such that the global higher order smoothness indicator

$$\tau_{2r-1} = c_0 \beta_0 + c_1 \beta_1 + c_{r-2} \beta_{r-2} + c_{r-1} \beta_{r-1} = O(\Delta x^{r+2}).$$

(50)

First of all, using Corollary 7, all the coefficients $\beta_0$ and $\beta_{r-1}$ associated to the terms $\Delta x^M, M \leq r$ are equal. Furthermore,

- assuming $r$ is odd, the coefficients of $\beta_0$ and $\beta_{r-1}$ associated to the terms $\Delta x^M, M = r + 1$ ($M$ is an even number) are equal. Hence

$$\tau_{2r-1} = |\beta_0 - \beta_{r-1}| = O(\Delta x^{r+2}).$$

- assuming $r$ is even, the coefficients associated with the $O(\Delta x^{r+1})$ (odd order) term of $(\beta_0 + \beta_{r-1})$ and $(\beta_1 + \beta_{r-2})$ are zero. Also, the coefficients of these two sums are equal to $2 \sum_{j=1}^{M^*} A_{k,M,j}$ and of order $O(\Delta x^M)$, for each $M \leq r$. This means that the coefficients of difference of these two sums $(\beta_0 + \beta_{r-1}) - (\beta_1 + \beta_{r-2})$ of order $\Delta x^M$ for all $M \leq r + 1$ are all zero, yielding

$$\tau_{2r-1} = |\beta_0 - \beta_1 - \beta_{r-2} + \beta_{r-1}| = O(\Delta x^{r+2}).$$

\[\Box\]

### 3.2 The optimal higher order smoothness indicator $\tau_{2r-1}^{opt}$

The linear combinations of the $\beta_k$ displayed in Theorem 2 are not unique, nor optimal, in the sense of generating the highest possible order for $\tau_{2r-1}$. Even though it is sufficient to have $\tau_{2r-1} = O(\Delta x^{r+2})$ to guarantee the formal $(2r - 1)$ order of accuracy of the WENO-Z scheme, the Taylor expansions of Theorem 6 may yield a higher order. In this section we perform a deeper investigation in the vector space generated by the linear combinations of the $\beta_k$ in order to find the global optimal order smoothness indicator, $\tau_{2r-1}^{opt}$, for each value of $r \geq 3$. We state our main result in the following proposition:

**Proposition 8** The greatest lower bound for the order of $\tau_{2r-1}^{opt}, M_{2r-1}, r \geq 3$, is given by the largest value of $m$ such that

$$\max \left( \rho^+, \left\lceil \frac{r}{2} \right\rceil \right) + \max \left( \rho^-, \left\lfloor \frac{r}{2} \right\rfloor \right) < r,$$

18
where

\[ \rho^+ = 1 + \phi_0, \quad \rho^- = \phi_1, \quad \phi_k = \begin{cases} \left\lfloor \frac{m-r}{2} \right\rfloor \left( \left\lfloor \frac{m-r}{2} \right\rfloor + 1 \right) \mod (m, 2) = k \\ \left\lfloor \frac{m-r}{2} \right\rfloor ^2 \mod (m, 2) = 1 - k \end{cases} \]

The proposition above yields the following table with the value of \( M_{2r-1} \) for \( r \) up to 20:

<table>
<thead>
<tr>
<th>( r )</th>
<th>( 2r - 1 )</th>
<th>( M_{2r-1} )</th>
<th>( r )</th>
<th>( 2r - 1 )</th>
<th>( M_{2r-1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5</td>
<td>5</td>
<td>12</td>
<td>23</td>
<td>17</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>7</td>
<td>13</td>
<td>25</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>8</td>
<td>14</td>
<td>27</td>
<td>19</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>9</td>
<td>15</td>
<td>29</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td>11</td>
<td>16</td>
<td>31</td>
<td>22</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>12</td>
<td>17</td>
<td>33</td>
<td>23</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>13</td>
<td>18</td>
<td>35</td>
<td>24</td>
</tr>
<tr>
<td>10</td>
<td>19</td>
<td>15</td>
<td>19</td>
<td>37</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>21</td>
<td>16</td>
<td>20</td>
<td>39</td>
<td>27</td>
</tr>
</tbody>
</table>

We see from the table that when derived as a linear combination of the lower order local smoothness indicators \( \beta_k \), the order of \( \tau_{2r-1}^{opt} \), \( M_{2r-1} \) is always smaller than the order of the scheme, \( 2r - 1 \), for \( r > 4 \), but it is also bigger than the order of the conventional \( \tau_{2r-1} \), \( r + 2 \), given by Theorem 2, for \( r > 3 \).

Let us now check by means of a numerical example the claim that the use of \( \tau_{2r-1}^{opt} \) improves over \( \tau_{2r-1} \) when capturing high order structures in the solution. For instance, consider the following test function,

\[ f(x) = x^k + \exp(lx), \quad x \in [-1, 1], \]

with parameters \( k = 8 \) and \( l = 5 \). The WENO-Z scheme built with the global high order smoothness indicator \( \tau_{11} = O(\Delta x^8) \) does not resolve well functions with information containing higher order than \( O(\Delta x^7) \). On the other hand, when equipped with \( \tau_{11}^{opt} = O(\Delta x^9) \), WENO-Z is able to capture high order information up to \( O(\Delta x^9) \). As illustrated in Table I, one can see that the WENO-Z11 scheme using \( \tau_{11} \) is significantly less accurate than the one using the \( \tau_{11}^{opt} \) when computing the derivative of the test function above.

We would like now to present more clearly the issues involved in the proof of Proposition 8. To obtain a \( \tau_{2r-1} \) of order \( m \), it is sufficient to find a nontrivial set of constants \( \{c_0, \ldots, c_{r-1}\} \) such that \( \sum_{k=0}^{r-1} c_k \beta_k = O(\Delta x^m) \). Rearranging the summation in (48), we want

\[ \sum_{M=2}^{m-1} \sum_{j=1}^{M} \sum_{k=0}^{r-1} c_k A_{k,M,j} f_0^{(j)} f_0^{(M-j)} \Delta x^M = 0, \]

or

\[ \sum_{k=0}^{r-1} c_k A_{k,M,j} = 0, \quad M = 2, \ldots, m - 1, \quad j = 1, \ldots, M^*. \]
We need to find a nontrivial solution of the system of linear equations $A_m c = 0$, where $c = (c_0, \ldots, c_{r-1})^T$ and

$$A_m = \begin{pmatrix}
A_{0,2,1} & A_{1,2,1} & \cdots & A_{r-1,2,1} \\
\vdots & \vdots & \ddots & \vdots \\
A_{0,r,1} & A_{1,r,1} & \cdots & A_{r-1,r,1} \\
\vdots & \vdots & \ddots & \vdots \\
A_{0,m-1,\lfloor \frac{m-1}{2} \rfloor - 1} & A_{1,m-1,\lfloor \frac{m-1}{2} \rfloor - 1} & \cdots & A_{r-1,m-1,\lfloor \frac{m-1}{2} \rfloor - 1} \\
A_{0,m-1,\lfloor \frac{m}{2} \rfloor - 1} & A_{1,m-1,\lfloor \frac{m}{2} \rfloor - 1} & \cdots & A_{r-1,m-1,\lfloor \frac{m}{2} \rfloor - 1}
\end{pmatrix}.$$  

Note that the matrix $A_m$ has $r$ column vectors and some of the row vectors of $A$ can be linearly dependent. Since the number of columns of $A_m$ is $r$, one needs $\text{rank}(A_m) < r$ to ensure that a non-trivial solution exists.
To clarify the notation above we consider \( r = 5 \) and look to the Taylor expansions of the lower order local smoothness indicators \( \beta_k \) for orders less than \( \Delta x^9 \) at \( x = x_0 \):

\[
\beta_0 = f_0'' x^2 + \frac{13 \Delta}{12} f_0'' x^4 + \left( -\frac{2}{5} f_0^{(5)} + \frac{1}{36} f_0^{(4)} + \frac{1}{720} f_0^{(2)} \right) \Delta x^6 + \left( \frac{2}{3} f_0^{(6)} - \frac{9}{5} f_0^{(5)} \right) \Delta x^7 +
\]

\[
\left( -\frac{13}{21} f_0^{(7)} + \frac{1235}{432} f_0^{(6)} - \frac{5467}{1440} f_0^{(5)} + \frac{32803}{30240} f_0^{(4)} \right) \Delta x^8 + O(\Delta x^9),
\]

\[
\beta_1 = f_0'' x^2 + \frac{13 \Delta}{12} f_0'' x^4 + \left( \frac{1}{10} f_0^{(5)} - \frac{1}{36} f_0^{(4)} - \frac{1}{720} f_0^{(2)} \right) \Delta x^6 + \left( \frac{1}{12} f_0^{(6)} + \frac{11}{60} f_0^{(5)} \right) \Delta x^7 +
\]

\[
\left( \frac{1}{21} f_0^{(7)} - \frac{251}{2160} f_0^{(6)} - \frac{781}{1440} f_0^{(5)} + \frac{32803}{30240} f_0^{(4)} \right) \Delta x^8 + O(\Delta x^9),
\]

\[
\beta_2 = f_0'' x^2 + \frac{13 \Delta}{12} f_0'' x^4 + \left( -\frac{1}{15} f_0^{(5)} - \frac{1}{36} f_0^{(4)} + \frac{1}{720} f_0^{(2)} \right) \Delta x^6 +
\]

\[
\left( -\frac{1}{126} f_0^{(7)} - \frac{53}{2160} f_0^{(6)} + \frac{781}{1440} f_0^{(5)} + \frac{32803}{30240} f_0^{(4)} \right) \Delta x^8 + O(\Delta x^9),
\]

\[
\beta_3 = f_0'' x^2 + \frac{13 \Delta}{12} f_0'' x^4 + \left( \frac{1}{10} f_0^{(5)} - \frac{1}{36} f_0^{(4)} - \frac{1}{720} f_0^{(2)} \right) \Delta x^6 + \left( \frac{1}{12} f_0^{(6)} + \frac{11}{60} f_0^{(5)} \right) \Delta x^7 +
\]

\[
\left( \frac{1}{21} f_0^{(7)} - \frac{251}{2160} f_0^{(6)} - \frac{781}{1440} f_0^{(5)} + \frac{32803}{30240} f_0^{(4)} \right) \Delta x^8 + O(\Delta x^9),
\]

\[
\beta_4 = f_0'' x^2 + \frac{13 \Delta}{12} f_0'' x^4 + \left( -\frac{2}{5} f_0^{(5)} - \frac{1}{36} f_0^{(4)} + \frac{1}{720} f_0^{(2)} \right) \Delta x^6 + \left( -\frac{2}{3} f_0^{(6)} + \frac{9}{5} f_0^{(5)} \right) \Delta x^7 +
\]

\[
\left( -\frac{13}{21} f_0^{(7)} + \frac{1235}{432} f_0^{(6)} - \frac{5467}{1440} f_0^{(5)} + \frac{32803}{30240} f_0^{(4)} \right) \Delta x^8 + O(\Delta x^9),
\]

The corresponding matrices \( A_m \) (omitting the zero lines) for \( m \) varying from 6 to 9 are

\[
A_6 = \left( \begin{array}{cccc} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \frac{13}{12} & \frac{13}{12} & \frac{13}{12} & \frac{13}{12} \\ -\frac{1}{5} & \frac{1}{5} & \frac{1}{5} & -\frac{1}{5} \end{array} \right), \quad A_7 = \left( \begin{array}{cccc} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \frac{13}{12} & \frac{13}{12} & \frac{13}{12} & \frac{13}{12} \\ \frac{13}{12} & \frac{13}{12} & \frac{13}{12} & \frac{13}{12} \end{array} \right)
\]

\[
A_8 = \left( \begin{array}{cccc} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \frac{13}{12} & \frac{13}{12} & \frac{13}{12} & \frac{13}{12} \\ \frac{13}{12} & \frac{13}{12} & \frac{13}{12} & \frac{13}{12} \end{array} \right), \quad A_9 = \left( \begin{array}{cccc} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \frac{13}{12} & \frac{13}{12} & \frac{13}{12} & \frac{13}{12} \\ \frac{13}{12} & \frac{13}{12} & \frac{13}{12} & \frac{13}{12} \end{array} \right)
\]

The rank of the matrices above are:

\[
\text{rank}(A_6) = 1, \quad \text{rank}(A_7) = 2, \quad \text{rank}(A_8) = 4 < r = 5, \quad \text{rank}(A_9) = 5.
\]
Therefore, the maximum order $\tau_0$ can achieve is given by $M_9 = 8$ since $\text{rank}(A_9) \geq r = 5$.

We now adopt the following strategy to prove proposition 8: For a given matrix $A_m$ with $r$ column vectors and with row vectors containing all the coefficients of $\beta_k$ of order up to and equal to $m - 1$ that does not have full rank, we increase the order by one to $m$ and append non-zero row vectors that correspond to the coefficients of $\beta_k$ of order $m$ to the row vectors of the old matrix $A_m$ to form a new matrix $A_{m+1}$. This process will be repeated until the newly formed matrix reaches full rank. This process will terminate in a finite number of iterations since only finite number of row vectors can be appended as the maximum possible order of $\tau_{2r-1} \leq 2r - 1$. These new non-zero row vectors are represented by the submatrix $B_m$ as below:

$$
B_m = \begin{pmatrix}
A_{0,m,1} & A_{1,m,1} & \cdots & A_{r-1,m,1} \\
A_{0,m,2} & A_{1,m,2} & \cdots & A_{r-1,m,2} \\
\vdots & \vdots & \ddots & \vdots \\
A_{0,m,\lfloor \frac{m}{2} \rfloor} & A_{1,m,\lfloor \frac{m}{2} \rfloor} & \cdots & A_{r-1,m,\lfloor \frac{m}{2} \rfloor}
\end{pmatrix}.
$$

**Remark 9** In the proof of Proposition 8 below, we search for a rule for the increase of the rank of $A_m$ when the submatrix $B_m$ is appended to $A_m$. Obviously, the constant vectors of $B_m$ do not count, on the other hand, we will assume that all the remaining ones do count, i.e., they are pairwise linearly independent. Although we do not have a proof for that, this is what we observed in all the cases we tested. We are computing the worst case scenario for the rank of $A_m$, or its lowest upper bound as stated in the proposition.

**Proof.** First of all, it is easy to see that all the row vectors of $A_{r+1}$ are linearly dependent since by Corollary 7, $A_{0,M,j} = \ldots = A_{r+1,M,j}$ for $M \leq r$ and thus $\text{rank}(A_{r+1}) = 1$. Let $m > r + 1$ be fixed. We want to find out how many row vectors of $B_m$ do not depend on $k$. These row vectors will be responsible for the increase of the rank of $A_m$. By Theorem 2, the elements that are not dependent on $k$ take the form of $A_{k,M,j}$, with $j < r$ and $m - j < r$. Therefore the number of new row vectors that are not dependent on $k$ is the cardinality of the set

$$
\left\{ j \in \left\{1, \ldots, \left\lfloor \frac{m}{2} \right\rfloor \right\} \mid j \geq r \quad \text{or} \quad m - j \geq r \right\},
$$

which is also the rank of $B_m$, $\text{rank}(B_m) = m - r$. It is not always true that $\text{rank}(A_{m+1}) = \text{rank}(A_m) + \text{rank}(B_m)$ and to understand the increase of the rank of $A_m$, we need to analyze the matrix $A_m$.

- If $r$ is odd, the row vectors take the form

$$
\begin{pmatrix}
A_{0,m,j} & A_{1,m,j} & \cdots & A_{r-1,m,j} & A_{r-2,m,j} & \cdots & A_{r-1,m,j}
\end{pmatrix}.
$$

Using the anti-symmetry/symmetry condition (49), it is equivalent to

$$
\begin{pmatrix}
A_{0,m,j} & A_{1,m,j} & \cdots & A_{r-1,m,j} & (-1)^m A_{1,m,j} & (-1)^m A_{0,m,j}
\end{pmatrix}.
$$

Considering the vector space formed by all the row vectors that takes the form above, one can see that each row vector has $\left\lceil \frac{m}{2} \right\rceil$ free variables $A_{0,m,j}, \ldots, A_{r-1,m,j}$, when $m$ is even, and $\left\lfloor \frac{m}{2} \right\rfloor$ free variables, when $m$ is odd, for in this case $A_{r-1,m,j} = -A_{r-1,m,j} = 0$.  

22
If $r$ is even, there is no central term $A_{\frac{r-1}{2}, m, j}$ and there are always $\frac{r}{2}$ free variables, regardless of the value of $m$.

Due to the factor $(-1)^m$, the $l$th element of each row vector can be anti-symmetric or symmetric to the $(r - 1 - l)^{th}$ element. This duality allows us to build two different vector spaces, $V^+$ and $V^-$, defined by

\[ V^+ = \{ (a_0, a_1, \ldots, a_{r-1}) \in \mathbb{R}^r | a_l = a_{r-1-l}, l = 0, \ldots, r-1 \}, \]
\[ V^- = \{ (a_0, a_1, \ldots, a_{r-1}) \in \mathbb{R}^r | a_l = -a_{r-1-l}, l = 0, \ldots, r-1 \}, \]

with dimensions equal to $\lceil \frac{r}{2} \rceil$ and $\lfloor \frac{r}{2} \rfloor$ respectively (note that when $r$ is even the dimensions are equal).

These vector spaces are in a direct sum; that is, if we intersect $A_m$ with each one of these vector spaces, we shall have two new matrices $A^+_{m}$ and $A^-_{m}$ and 

\[ \text{rank}(A_m) = \text{rank}(A^+_{m}) + \text{rank}(A^-_{m}). \]

We will take this moment to observe that $\text{rank}(A^+_{r+1}) = 1$ and $\text{rank}(A^-_{r+1}) = 0$.

Now we can define $\text{rank}(A_{m+1})$ based on $\text{rank}(A_m)$. Indeed, by appending the row vectors of $B_m$, we can see that if $m$ is even (odd), only the rank of $A^+_{m+1}$ ($A^-_{m+1}$) will increase. On the other hand, this increase is limited by $\lceil \frac{r}{2} \rceil$ ($\lfloor \frac{r}{2} \rfloor$) for $\text{rank}(A^+_{m+1})$ ($\text{rank}(A^-_{m+1})$), because of the dimensions of the vector spaces. Thus,

\[ \text{rank}(A_{m+1}) = \begin{cases} \text{rank}(A^-_{m}) + \min \left( \text{rank}(A^+_{m}) + \text{rank}(B_m), \lceil \frac{r}{2} \rceil \right) & \text{mod} (m, 2) = 0 \\ \text{rank}(A^+_{m}) + \min \left( \text{rank}(A^-_{m}) + \text{rank}(B_m), \lfloor \frac{r}{2} \rfloor \right) & \text{mod} (m, 2) = 1 \end{cases}. \]

From this result we obtain a recurrence relation for $\text{rank}(A_m)$, with initial conditions $\text{rank}(A^+_{r+1}) = 1$ and $\text{rank}(A^-_{r+1}) = 0$, with the closed form stated in the Proposition.

For a large order $(2r - 1)$, it is difficult to find the kernel and rank of the matrix $A$ analytically. In this situation, symbolic computational system such as Maple can be employed. In Maple, the built-in function $\text{RowReduce}$ is used to find the kernel $c$, $\text{rank}(A)$ and the leading order term of order $O(\Delta x^{M_{2r-1}})$. Table II gives the parameters $r$, $(2r - 1)$, $\text{rank}(A)$, optimal order $M_{2r-1}$, coefficient vector $c$ and the leading order for the global optimal order smoothness indicator $\tau^{opt}_{2r-1}$, $r = 3, \ldots, 9$ for the $(2r - 1)$ order WENO-Z scheme.

**Remark 10** From this point on, we shall replace the definition of $\tau_{2r-1}$ in the definition of the nonlinear weights $\omega_k^p$ in the WENO-Z scheme with $\tau^{opt}_{2r-1}$ in the rest of the paper unless stated otherwise.

### 4 Critical points

In [5] it was shown that the classical fifth-order WENO-JS scheme loses convergence at critical points due to the inability of its nonlinear weights to distinguish between flat and rough parts of
Consider the following test function,

\[ f(x) = x^k \exp(lx), \quad x \in [-1, 1], \] 

(52)

in which its first \( k - 1 \) derivatives \( f^{(j)}(0) = 0, j = 0, \ldots, k - 1 \). That is, this function has a critical point of order \( n_{cp} = k - 1 \) at \( x = 0 \).

We show in Table III the convergence rates for the classical WENO-JS, the mapped WENO-M and for WENO-Z at a critical point of second order, i.e., \( k = 3 \). We used the fixed value of \( p = 2 \) for WENO-JS and WENO-Z, since their rates of convergence do not change when \( p \) varies. On the other hand, increasing the value of \( p \) from 2 to 3 allows WENO-Z to recover the formal order of accuracy of the scheme. Table IV shows that this is indeed a general behavior, that is, increasing \( p \), increases the rate of convergence of WENO-Z and, particularly, when \( p = r - 1 \) and \( n_{cp} < r - 1 \),
Table III: Rate of convergence at a second order critical point \((n_{cp} = 2)\) for the seventh order \((r = 4)\) WENO-JS, WENO-M and WENO-Z schemes.

Table IV: Rates of convergence at critical points of increasing order \(n_{cp}\) for \((2r-1)\) order WENO-JS, WENO-M and WENO-Z schemes. A fixed constant \(\epsilon = 1 \times 10^{-40}\) is used.

Critical points have become a point of discussion since when in [5] it was shown that the standard value of the parameter \(\epsilon = 10^{-6}\) was in fact hiding the loss of accuracy of WENO-JS. At critical points, the \(\beta_k\) have much smaller sizes and such an \(\epsilon\) dominates their relative variations towards a central upwind fifth order scheme. When a smaller value for \(\epsilon\) is used the order of accuracy of the scheme degrades to third order, since the variation of the \(\beta_k\) is understood as the indication of a high gradient.

WENO-M was presented as a fix to this situation, for it used a mapping that corrected the weights of WENO-JS and recovered the formal order of the scheme even when using very small values of
The numerical results with WENO-M were superior at shock problems and this was credited to the improvement of the weights at critical points. Nevertheless, it is well known that at problems with shocks one cannot expect better order of accuracy than O(1), and this dominates any eventual increase of convergence order at critical points. Thus, in [10] it was shown that the extra sharpness obtained at discontinuities by WENO-M was in fact due to its smaller dissipation. A smaller dissipation is achieved when close to discontinuities the scheme generates a set of nonlinear weights that provide a more centralized scheme and we will see in the next section that this is the case for the improved results of WENO-Z and WENO-M over WENO-JS.

Nevertheless, flat and high gradients regions of the solution can be distinguished by the absolute sizes of the $\beta_k$ and a conditional statement could be used to improve the WENO schemes convergence at critical points. However, a conveniently size of $\epsilon$ can be chosen in order to play the role of this conditional. Based on the structure disclosed by their Taylor series in the last section, such an $\epsilon$ could be of size $\Delta x^2$, since at critical points the sizes of the $\beta_k$ will be smaller than that. Table V below show numerical results for critical points of order $r - 1$, that is, $k = r - 1$ at (52), for the WENO-Z schemes. Note that the formal order of accuracy is recovered and the same occurs for WENO-JS and WENO-M (not shown).

<table>
<thead>
<tr>
<th>$\Delta x$</th>
<th>WENO-Z5</th>
<th>WENO-Z7</th>
<th>WENO-Z9</th>
<th>WENO-Z11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00000e-01</td>
<td>2.3e-05</td>
<td>2.7e-05</td>
<td>2.1e-05</td>
<td>6.3e-05</td>
</tr>
<tr>
<td>5.00000e-02</td>
<td>5.8e-07</td>
<td>2.2e-09</td>
<td>13.6</td>
<td>3.1e-09</td>
</tr>
<tr>
<td>2.50000e-02</td>
<td>1.9e-08</td>
<td>3.6e-11</td>
<td>5.9</td>
<td>8.8e-14</td>
</tr>
<tr>
<td>1.25000e-02</td>
<td>6.0e-10</td>
<td>2.8e-13</td>
<td>7.0</td>
<td>1.8e-16</td>
</tr>
<tr>
<td>6.25000e-03</td>
<td>1.9e-11</td>
<td>2.2e-15</td>
<td>7.0</td>
<td>3.5e-19</td>
</tr>
<tr>
<td>3.12500e-03</td>
<td>5.9e-13</td>
<td>1.7e-17</td>
<td>7.0</td>
<td>6.8e-22</td>
</tr>
<tr>
<td>1.56250e-03</td>
<td>1.9e-14</td>
<td>1.4e-19</td>
<td>7.0</td>
<td>1.3e-24</td>
</tr>
<tr>
<td>7.81250e-04</td>
<td>5.8e-16</td>
<td>1.1e-21</td>
<td>7.0</td>
<td>2.6e-27</td>
</tr>
</tbody>
</table>

Table V: Rates of convergence at a critical point of order $n_{cp} = r - 1$ for $(2r - 1)$ order WENO-Z schemes with power parameter $p = 1$ and $\epsilon = \Delta x^2$.

**Remark 12** The critical points issue has been the departure point of several investigations on improved sets of WENO nonlinear weights. However, the literature has yet to provide meaningful examples with shocks where the distinguished treatment of critical points provides any substantial improvement to the quality of the numerical solution. In the numerical experiments of next section we look closer to the issue of the numerical dissipation as a result of qualitative differences among the distinct sets of nonlinear weights of the WENO schemes studied in this article. Thus, we will pay closer attention to the power parameter $p$ and use the satisfactory variable value of $\epsilon = \Delta x^{r-1}$ for all the numerical experiments. This is a compromise value that avoids the original predominance of a fixed value of $\epsilon$ and also sets a lower bound for the ratio of the smoothness indicators.

## 5 Numerical Results

In this section We compare the sets of nonlinear weights generated by the three WENO schemes discussed in this article. We will see that the final amount of dissipation observed at the numerical solutions is a result of the sizes of the weights attributed to discontinuous substencils by each
scheme. We also show numerical experiments with the one dimensional scalar advection equation with initial condition consisted of a triangle, a Gaussian, a square and an ellipse functions, the one dimensional Euler Equations of compressible gas dynamics with Riemann initial condition such as the Lax problem and the one dimensional Mach 3 shock-density wave interaction. We end this section showing numerical simulations of the two dimensional Mach 10 double-Mach shock reflection problem using high orders WENO-Z schemes with increasing resolutions. The time dependent problems are all solved via the third order Runge-Kutta TVD scheme with CFL = 0.45.

5.1 Discontinuities and nonlinear weights

We now compare the different sets of normalized nonlinear weights $\omega_k$ generated by WENO-JS, WENO-Z and WENO-M. We fix $\epsilon = 1 \times 10^{-40}$ in the discussion of this section.

Consider the following discontinuous function

$$g(x) = \begin{cases}
-\sin(\pi x) - \frac{x^3}{2}, & -1 < x < 0 \\
1 - \sin(\pi x) - \frac{x^3}{2}, & 0 \leq x \leq 1
\end{cases}, \quad g(x) = g(x - 2),$$

with a single discontinuity located at $x = 0$. In figure 2, the nonlinear weights $\omega_k$ of the fifth order WENO-JS, WENO-Z and WENO-M schemes with power parameters $p = 1$ and $p = 2$ are displayed.

Table VI: Nonlinear weights $\omega_k$, $k = 0, 1, 2$ of the fifth order WENO-JS, WENO-Z and WENO-M schemes with power parameters $p = 1$ and $p = 2$ at locations $x = -0.0015$ and $x = -0.005$. The number of uniformly spaced grid points used is $N = 200$.

Analyzing the results shown at Table VI and at Figure 2, for the fifth order WENO schemes, we see that:

1. At $x = -0.015$, where the substencils $S_0$ and $S_1$ are smooth and only $S_2$ is nonsmooth:
   - WENO-Z assigns a larger weight $\omega_2$ for the nonsmooth substencil $S_2$ than WENO-JS, while WENO-M assigns an even larger one as evidenced in the Table.
   - Note also that $\omega_2^2 = 2\omega_2$; a straightforward computation justifying this fact is given in [10].
Figure 2: Nonlinear weights $\omega_k, k = 0, 1, 2$ of the fifth order ($r = 3$) WENO-JS, WENO-Z and WENO-M schemes with power parameters $p = 1$ and $p = 2$.

- Nevertheless, not always the weights of WENO-M are larger than the corresponding ones of WENO-Z and WENO-JS; for instance, $\omega_0 = \omega^Z_0 > \omega^M_0$. This occurs because the increase of the smallest weight is compensated due to the normalization of the weights.

2. At $x = -0.005$, only the substencils $S_0$ is smooth, while $S_1$ and $S_2$ both contain the discontinuity:

- Here, $\omega_2$ is the smallest weight since the local lower order polynomial approximation of $g(x)$ at $x = 0$ is computed through an extrapolation of the values in $S_2$. Note once again that this smallest weight, $\omega_2$, is significantly larger with WENO-Z than in WENO-JS, double the value, and also once again is even larger with WENO-M.

We see that WENO-M assigns larger weights than WENO-Z for discontinuous substencils for the same value of $p$. Also, in its turn, WENO-Z assigns larger weights to discontinuous substencils than WENO-JS for the same value of $p$. Moreover, a greater value of $p$ decreases the absolute sizes of the weights assigned to discontinuous substencils by all three schemes. Since a bigger weight for a discontinuous stencil contributes to a more central upwind approximation, we may conclude that there is a hierarchy of dissipation that puts WENO-M as the least dissipative scheme and WENO-JS as the most dissipative one, while WENO-Z assumes an intermediary position. Also, increasing the value of $p$, decreases the absolute sizes of the weights assigned to discontinuous substencils, making
for a more lateral linear combination of the substencils and, therefore, increases dissipation. We remark that this is a general behavior that does not depend on the order of the scheme; for instance, see Figure 3 where we show analogous graphs for the nonlinear weights of the seventh order case.

Figure 3: Nonlinear weights $\omega_k$, $k = 0, 1, 2, 3$ of the seventh order ($r = 4$) WENO-JS, WENO-Z and WENO-M schemes, with power parameters $p = 1$ and $p = 2$.

5.2 Linear Advection

Consider the one dimensional linear wave equation,

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = 0, \quad x \in [0, 1),$$

with an initial condition given by

$$u(x, t = 0) = \begin{cases} \frac{1}{6} [G(x, \beta, z - \delta) + 4G(x, \beta, z) + G(x, \beta, z + \delta)] & , x \in [-0.8, -0.6] \\ 1 & , x \in [-0.4, -0.2] \\ 1 - [10(x - 0.1)] & , x \in [0, 0.2] \\ \frac{1}{6} [F(x, \alpha, a - \delta) + 4F(x, \alpha, a) + F(x, \alpha, a + \delta \omega)] & , x \in [0.4, 0.6] \\ 0 & , \text{else} \end{cases}$$

$$G(x, \beta, z) = e^{-\beta(x-z)^2}, \quad F(x, \alpha, a) = \sqrt{\max(1 - \alpha^2(x-a)^2, 0)},$$
where \( z = -0.7, \, \delta = 0.005, \, \beta = \frac{\log 2}{\log 36}, \, a = 0.5 \) and \( \alpha = 10 \). This is the one dimensional scalar linear advection problem with an initial condition consisting of a smooth Gaussian, a discontinuous Heaviside function, a piecewise linear triangle function and a smooth elliptic function. Periodical boundary conditions are imposed on the two ends of the domain.

We compute the numerical solutions for all three schemes of order 11 (\( r = 6 \)) with \( N = 200 \) uniformly spaced grid points and final time \( T = 8 \). The third-order Runge-Kutta TVD scheme uses an adjusted time step \( \Delta t = \text{CFL} \times \Delta x \frac{2r-1}{3r} \) in order to maintain the convergence rate of the underlying spatial WENO scheme. The results in Figure 4 show the computed solutions and the absolute point-wise errors for \( p = 1, 2 \) and \( r - 1 \). Symbols represent the numerical solution while the lines indicate the absolute point-wise error at each grid point. The black solid line is the exact solution.

We interpret the behavior of the solution by looking at the error. A monotone error curve away from the discontinuity implies a non-oscillatory solution; otherwise, one may infer the presence of oscillations. Also, the oscillations are very small if variations in the error curve are close to the bottom of the log scale. In summary, one can observe that

- for \( p = 1 \), all three schemes do not have enough dissipation to simulate the advection of the square wave without oscillations.
- for \( p = 2 \), WENO-JS and WENO-Z achieve the ENO property for the square wave, however WENO-M is still oscillatory.
- for \( p = r - 1 \), all schemes show no oscillations, and their solutions are very similar in what regards the sharp approximation of corners.
- increasing the number of grid points will reduce the error away from the discontinuities in the solution and in its derivatives (not shown).

5.3 One dimensional Euler Equations: The Lax Problem

This same behavior is observed in Figure 5 where we used the ninth order, \( r = 5 \), WENO-JS, WENO-Z and WENO-M schemes, for the numerical simulation of the Lax problem. Once again we note that WENO-M requires the highest value of \( p \) among all three schemes in order to become non-oscillatory.

5.4 One dimensional Shock-density wave interaction

In the Lax problem, the solution is a piece-wise linear function which does not fully justify the cost of using the high order reconstruction process of the WENO schemes. The standard one dimensional shock-density wave interaction is generally preferred, since the solution of this problem consist of a main shock, a high gradient smooth post-shock region and multiple shocklets that develop in a later time, all of these requiring high order schemes in order to be efficiently and accurately represented. Details of the setup of this problem can be found in [2].
Figure 4: Numerical solutions of the linear advection equation at final time $t = 8$ computed by the eleventh order $(r = 6)$ WENO-JS, WENO-M and WENO-Z with power parameters $p = 1, 2, r - 1$. The number of grid points used is $N = 200$. The symbols indicate the numerical solution at the grid points. The lines show the absolute point-wise error at each grid point. The solid black line is the exact solution.
Figure 5: Numerical solution of the Lax problem as computed by the ninth order \((r = 5)\) WENO-JS, WENO-M and WENO-Z schemes with power parameters \(p = 1, 2, r - 1\).
Figure 6: Numerical solution of the shock-density wave interaction as computed by the WENO-JS, WENO-M and WENO-Z schemes of order $2r - 1 = 5, 7, 9, 11$ with power parameters $p = 1, 2, r - 1$. The number of grid points used is $N = 200$. 
Figure 7: Numerical solution of the shock-density wave interaction as computed by the WENO-JS, WENO-M and WENO-Z schemes of order $2r - 1 = 11$ with power parameters $p = 1, 2, r - 1$. The number of grid points used is $N = 250$.

Similarly, as in the examples before, one can observe from figures 6 and 7, using $N = 200$ and $N = 250$ number of uniformly spaced grid points, respectively, that the WENO-JS scheme is the most dissipative one, WENO-M scheme is the least one, and WENO-Z scheme occupies an intermediary position for any fixed order $(2r - 1)$ and power parameter $p = 1, 2, r - 1$. WENO-M scheme provides a better resolution of the small scales high frequency structures behind the main shock with respect to WENO-Z, and the same can be said about WENO-Z with regards to WENO-JS; and all of this is due to their distinct levels of dissipation, resulting from their distinct treatment of discontinuous substencils. Nevertheless, notice that all the WENO solutions are very similar once a sufficient number of grid points ($N = 250$) is used and this is a point against WENO-M, since its CPU costs, in this one dimensional test case, are approximately 20% to 30% more expensive than WENO-JS and WENO-Z schemes, which have very similar computational costs (see [10]).

5.5 Double-Mach shock reflection problem

Finally, we apply the high order conservative characteristic-wise WENO-Z finite difference scheme to the two dimensional double-Mach shock reflection problem [19] where a vertical shock wave moves horizontally into a wedge that is inclined by some angle. The domain of the problem is $[0, 4] \times [0, 1]$ and the shock moves diagonally at Mach 10, making an angle of 60 degrees with the horizontal axis. The equations are the two dimensional Euler equations ($\gamma = 1.4$) and initial conditions are given by

$$Q = (\rho, u, v, P) = \begin{cases} (8, 8.25 \cos \frac{\pi}{6}, -8.25 \sin \frac{\pi}{6}, 116.5), & x < x_0 + \frac{y}{\sqrt{3}} \\ (1.4, 0, 0, 1.0), & x \geq x_0 + \frac{y}{\sqrt{3}} \end{cases},$$

with $x_0 = \frac{1}{6}$. Boundary conditions at $x = 0$ are inflow, with post-shock values as above, and at $x = 4$ we have outflow boundary conditions with $\frac{\partial Q}{\partial x} = 0$. At $y = 0$, reflecting boundary conditions are applied to the interval $[x_0, 4]$, in the $x$-axis, simulating the wedge: $(\rho_y, u_y, v, P_y) = (0, 0, 0, 0)$ for $x_0 \leq x < 4$ and $y = 0$. At the upper boundary, $y = 1$, the flow has to be imposed such that there is no interaction with the moving shock. The exact location of the shockwave at $y = 1$, at
instant \( t \), is given by \( s(t) = x_0 + \frac{(1 + 20t)}{\sqrt{3}} \); we set post-shock and pre-shock conditions before and after this location, given respectively by:

\[
Q|_{y=1} = (\rho, u, v, P)|_{y=1} = \begin{cases} 
(8, 8.25 \cos \frac{\pi}{6}, -8.25 \sin \frac{\pi}{6}, 116.5), & 0 \leq x < s(t) \\
(1.4, 0, 0, 1.0), & s(t) \leq x \leq 4 
\end{cases}
\]

Numerical results of this problem are well documented in the literature (see also figure 8) and further details can be found in [19].

The global structure of the solution at time \( t = 0.2 \) is in general very similar across different schemes, resolutions and parameters. However, the resolving power of the higher order WENO schemes can be determined by the number of small vortices that can be captured along the slip line and the wall jet behind the lower half of the right moving shock by the underlying scheme with a given resolution.

Figure 8: Density contours of the double-Mach shock reflection as computed by the eleventh order WENO-Z scheme at time \( t = 0.2 \).

In Figure 9 we display the region around the double Mach stems in order to observe the numerical solutions of the WENO-Z scheme of orders \( 2r - 1 = 5, 9, 11 \) using three different uniform meshes with resolutions \( 400 \times 100 \), \( 800 \times 200 \) and \( 1600 \times 400 \) at time \( t = 0.2 \). Here, the power parameter \( p = r - 1 \) is used. For this problem, \( p \leq 2 \) is unstable for WENO-M \((r > 3)\) and WENO-Z \((r > 4)\) and this can be attributed to the insufficient dissipation of the schemes. We see from the figure that one may obtain similar results by increasing the order of approximation and decreasing the number of points in the spatial discretization. Additionally, as it was shown in [4], the CPU cost of increasing the order is smaller than the one of increasing the spatial resolution.

6 Conclusions

We extended the new WENO schemes introduced in [10] to rates of convergence higher than 5 by providing a formula in closed form for the higher order smoothness indicators \( \tau_{2r-1} \) as a simple linear combination of the lower order smoothness indicators \( \beta_k \). Another formula for the maximum order of convergence that such linear combinations can achieve was also provided. Both results were obtained from a thorough study of the properties of the Taylor expansions of the lower order smoothness indicators departing from the symmetric structure of their underlying Lagrangian interpolating polynomials with respect to the geometrical disposition of the global and the local WENO stencils. We also discussed the lack of convergence at critical points of smooth solutions and revisited the \( \epsilon \) issue of the classical WENO-JS weights showing that for practical matters it may work as an implicit conditional that distinguishes flat regions from discontinuities. The formation
Figure 9: Density contours of the double-Mach shock reflection problem as computed by the high order WENO-Z scheme of orders $2r - 1 = 5, 9, 11$, with resolutions at $400 \times 100$, $800 \times 200$ and $1600 \times 400$ at time $t = 0.2$. The power parameter $p = r - 1$ is used in these simulations.
of the distinct sets of nonlinear weights for the WENO-JS, WENO-Z and WENO-M were analyzed and numerical arguments were provided to explain their influence in the distinct numerical features of each scheme and their changes with the increasing of the power parameter \( p \). We concluded that regarding dissipation WENO-Z occupies an intermediary position between WENO-JS and WENO-M. Numerical results with the system of the Euler Equations in 1D and 2D were presented to show that the new high order WENO-Z schemes perform well at the treatment of numerical solutions containing both discontinuities and high order smooth structures.

7 Acknowledgments

The first and second authors have been supported by CNPq, grants 300315/98-8 and FAPERJ E-26/111.564/2008. The third author (Don) would like to thank the support provided by the FRG grant FRG08-09-II-12 from Hong Kong Baptist University and the RGC grant HKBU-092009 from Hong Kong Research Grants Council. The author would also like to thanks the Departamento de Matemática Aplicada, IM-UFRJ, for hosting his visit during the course of the research.

References


