




# A third-order weighted essentially non-oscillatory scheme in optimal control problems governed by nonlinear hyperbolic conservation laws

David Frenzel<sup>1</sup> · Jens Lang<sup>1</sup> 

Received: 25 September 2020 / Accepted: 21 June 2021 / Published online: 2 July 2021  
© The Author(s) 2021

## Abstract

The weighted essentially non-oscillatory (WENO) methods are popular and effective spatial discretization methods for nonlinear hyperbolic partial differential equations. Although these methods are formally first-order accurate when a shock is present, they still have uniform high-order accuracy right up to the shock location. In this paper, we propose a novel third-order numerical method for solving optimal control problems subject to scalar nonlinear hyperbolic conservation laws. It is based on the first-discretize-then-optimize approach and combines a discrete adjoint WENO scheme of third order with the classical strong stability preserving three-stage third-order Runge–Kutta method SSPRK3. We analyze its approximation properties and apply it to optimal control problems of tracking-type with non-smooth target states. Comparisons to common first-order methods such as the Lax–Friedrichs and Engquist–Osher method show its great potential to achieve a higher accuracy along with good resolution around discontinuities.

**Keywords** Nonlinear optimal control · Discrete adjoints · Hyperbolic conservation laws · WENO schemes · Strong stability preserving Runge–Kutta methods

**Mathematics Subject Classification** 34H05 · 49M25 · 65L06 · 65M22

## 1 Introduction

We consider the optimal control problem

---

✉ Jens Lang  
lang@mathematik.tu-darmstadt.de  
David Frenzel  
frenzel@gsc.tu-darmstadt.de

<sup>1</sup> Department of Mathematics, Technical University of Darmstadt, Dolivostraße 15, 64293 Darmstadt, Germany

$$u_0^{min} = \arg \min_{u_0 \in \mathcal{U}_{ad}} J(y(T, \cdot; u_0), y_d) \quad (1)$$

with the tracking-type functional

$$J(y(T, \cdot; u_0), y_d) = \int_I G(y(T, x; u_0), y_d(x)) dx, \quad (2)$$

where

$$G(y(T, x; u_0), y_d(x)) = \frac{1}{2} |y(T, x; u_0) - y_d(x)|^2 \quad (3)$$

and  $y = y(T, x; u_0)$  is the scalar entropy solution at the final time  $T > 0$  of the nonlinear hyperbolic conservation law (later referred to as state equation)

$$\begin{aligned} \partial_t y + \partial_x f(y) &= 0, & (t, x) \in \Omega_T &:= (0, T] \times \mathbb{R}, \\ y(0, x) &= u_0(x), & x \in \mathbb{R}. \end{aligned} \quad (4)$$

Here,  $u_0 \in \mathcal{U}_{ad} \subseteq L^\infty(\mathbb{R})$  is the control and  $y_d \in L^2(\mathbb{R})$  denotes a given target towards which we strive to optimize. We assume that the flux function satisfies  $f \in C^m(\mathbb{R})$  with sufficiently large  $m \in \mathbb{N}$  and is convex, the admissible set  $\mathcal{U}_{ad}$  is non-empty, convex and closed, and the region of integration  $I$  in (2) is a bounded interval. Weak solutions to (4) are in general not unique, which implies that the physically relevant solution has to be chosen. As a fact we cite the well-known result from [21], which states that for  $u_0 \in L^\infty(\mathbb{R}) \cap BV(\mathbb{R})$  there exists a unique entropy solution in the sense of Krůžkov in the class  $C([0, T], L^1_{loc}(\mathbb{R})) \cap L^\infty(\mathbb{R} \times [0, T])$ . Using well-posedness and compactness properties of this solution, the existence of a minimizer  $u_0^{min}$  in (1) can be established under some natural additional assumptions on the class of admissible data  $\mathcal{U}_{ad}$ , see e.g. [5, Theorem 2.1] and [16, Proposition A.1]. In general, uniqueness is not guaranteed due to the occurrence of discontinuous solutions in (4), which can be equal for different initial values. An illustrative example can be found in [5]. These statements generalize to the case where a regularization term  $\mathcal{R}(u)$  is added to the objective function [28].

In this work, we focus on the numerical treatment of optimal control problems (1) governed by hyperbolic conservation laws, which has been studied amongst others in [1, 2, 5, 6, 10–12, 16–18, 22–24, 29–31]. We will follow the *first-discretize-then-optimize* approach, i.e., Eq. (4) is first discretized in space and time by applying a weighted essentially non-oscillatory (WENO) scheme and a strong stability preserving Runge-Kutta (SSPRK) method. This leads to a finite dimensional optimal control problem, for which the first-order discrete optimality system can be derived and solved by existing optimization solvers such as nonlinear Newton-type algorithms. In spite of the large size of the resulting problems, the flexibility of this approach naturally allows the incorporation of additional constraints and bounds. Further advantages are the direct use of automatic differentiation techniques and the computation of discrete adjoints, which are consistent with the discrete optimal control problem. Symmetric approximations of Hessian matrices can be easily derived and result in a computational speedup.

The application of common methods from nonlinear optimization requires the computation of directional derivatives of the target functional  $J$  with respect to the control. An efficient computation of the gradient can be effectuated by using the so-called adjoint approach, in which the derivative is represented via the adjoint state. The crucial issue of hyperbolic conservation laws is the possible formation of shocks even for smooth initial data, for which reason the classical adjoint calculus does not apply. To overcome these difficulties, nonstandard variational concepts have been developed in [4, 29, 31], which incorporate the shock sensitivity in order to derive rigorous optimality conditions. The resulting non-conservative equation has been studied in [3, 7, 29]. Their numerical resolution is intricate, since the interior boundary condition defined on a set of Lebesgue measure zero—existing for the continuous setting—is not present for the discrete counterpart. This inherent problem has been addressed in [1, 10–12]. The theory is, however, restricted to differentiable monotone schemes which have sufficiently large numerical diffusion and are of first order only.

To avoid unwanted smearing of the solution by large numerical diffusion and to overcome the lower order restriction of monotone schemes, we propose a novel approach based on WENO schemes introduced in [20, 25, 26]. These schemes have proven to approximate hyperbolic equations comprising both shocks and complex smooth solution structure with higher accuracy and adequate stability along with good resolution around discontinuities. Although these methods are formally first-order accurate when a shock is present, they still have uniform high-order accuracy right up to the shock location. WENO schemes are extensions of the ENO procedure, i.e., they perform essentially non-oscillatory, but overcome shortcomings of the ENO approximation, see [27] for a detailed discussion. By employing a global flux-splitting, the numerical flux function becomes classically differentiable and therefore allows to develop discrete adjoint WENO methods of higher order. Since the third-order WENO method is often applied in applications, we consider this method in the context of optimal control in more detail. We prove that the discrete adjoint WENO3 method is third-order consistent in space for smooth solutions. A fully discrete method is derived by applying a third-order SSPRK method. We present numerical results and study the approximation behaviour of the adjoint WENO3 scheme. Finally, we solve an optimal control problem with discontinuous target and compare the performance of our novel scheme to common first-order schemes such as the modified Lax–Friedrichs and the Engquist–Osher scheme. Further examples can be found in [9].

## 2 Adjoint equation and reversible solutions

In this section, we briefly recall some theoretical basics in order to set up appropriate adjoint equations for hyperbolic conservation laws. As pointed out in [4, Example 1], the solution operator  $S_t : u_0 \mapsto y(t, \cdot; u_0)$  is generically not differentiable in  $L^1_{loc}(\mathbb{R})$ , for which reason the classical adjoint calculus does not apply. However, in [29] it has been shown that entropy solutions to hyperbolic

conservation laws admit a generalized differentiable structure called *shift-differentiability*. Under suitable assumptions, a generalized Taylor expansion in  $L^1_{loc}$  of the form

$$y(t, \cdot; u_0 + \delta u_0) = y(t, \cdot; u_0) + S_y^{(x_i)}(T_{u_0}(\delta u_0))(\cdot) + o(\|\delta u_0\|_{L^\infty(\mathbb{R})}) \tag{5}$$

exists for all  $\delta u_0 \in L^\infty(\mathbb{R})$ , where  $T_u : \delta u \in L^\infty(\mathbb{R}) \mapsto (\delta y^T, \delta x_1, \dots, \delta x_N) \in L^r(I) \times \mathbb{R}^N$ ,  $r \in (1, \infty]$ , is a bounded linear operator and  $S_y^{(x_i)}$  is the shift variation defined by

$$S_y^{(x_i)}(\delta y^T, \delta x_1, \dots, \delta x_N)(x) = \delta y^T(x) + \sum_{i=1}^N (y(T, x_i-) - y(T, x_i+)) \text{sign}(\delta x_i) \mathbf{1}_{\Omega_i}, \tag{6}$$

where  $\Omega_i = [\min(x_i, x_i + \delta x_i), \max(x_i, x_i + \delta x_i)]$  and  $x_1, \dots, x_N$  denote the locations of the down-jumps of the entropy solution. The important advantage of shift-variations is that this framework allows to develop an adjoint calculus for hyperbolic conservation laws by using an *averaged sensitivity equation* which avoids the linearization of (4) in the usual way, see [31] for further details. The directional derivative of  $J$  in (2) in the direction of  $\delta u_0$  can then be represented by

$$\partial_{u_0} J(y(T, \cdot; u_0), y_d) \delta u_0(\cdot) = \int_I p(0, x) \delta u_0(x) dx, \tag{7}$$

where  $p$  is the solution of the adjoint equation

$$\begin{aligned} \partial_t p + f'(y) \partial_x p &= 0, & (t, x) \in \Omega_T, \\ p(T, x) &= p^T(x), & x \in \mathbb{R}. \end{aligned} \tag{8}$$

Here,  $p^T(x)$  is given by

$$p^T(x) = \begin{cases} \frac{[G(y(T, x), y_d(x))]}{[y(T, x)]}, & x \in X_s, \\ \partial_y G(y(T, x), y_d(x)), & \text{otherwise,} \end{cases} \tag{9}$$

where  $X_s$  is the set of locations where  $y(T, \cdot)$  possesses a shock and  $[w(x)] := w(x-) - w(x+)$ , which naturally incorporates the shock sensitivity.

Equation (8) is a linear transport equation with, in general, discontinuous coefficients. It admits multiple solutions, which requires the selection of the correct adjoint state. This is achieved by so-called *reversible solutions* that are defined along generalized characteristics [8]. An illustrative demonstration is given in Example 2.1. Under suitable technical assumptions and for appropriate end data  $p^T$  it can be shown that there exists a unique reversible solution to (8) that is bounded,  $L^\infty$ -stable, and  $TV$ -stable [29, Theorem 4.2.10 and Corollary 4.2.11]. In what follows, we will work with formulation (8) to derive a discrete adjoint WENO3 method.

**Example 2.1** Let  $f(y) = \frac{1}{2}y^2$ ,  $u_0(x) = -\text{sign}(x)$ ,  $T = 0.5$ , and  $y_d(x) = 0$  with  $x \in \mathbb{R}$ . It is well-known that the unique entropy solution is given by  $y(t, x) = -\text{sign}(x)$ ,  $t \in [0, T]$ , and hence we have

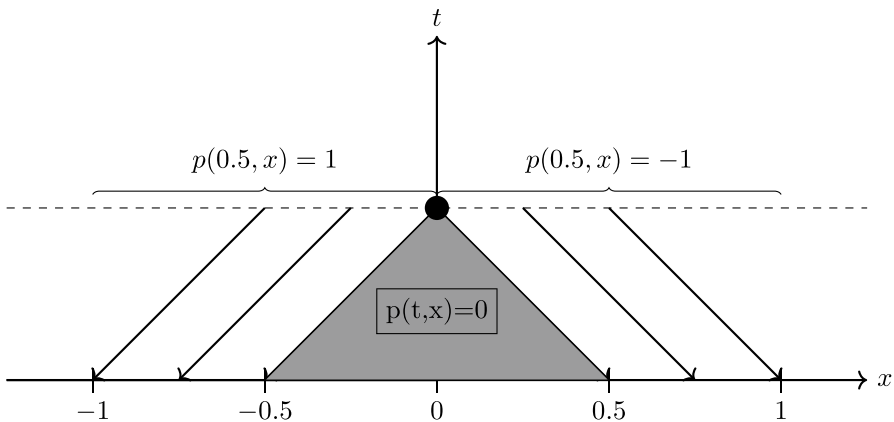
$$p^T(x) = \begin{cases} 0, & x = 0, \\ -\text{sign}(x), & x \neq 0. \end{cases} \tag{10}$$

The area that is not occupied by the classical characteristics is called shock funnel. It is represented by the grey-coloured triangle in Fig. 1. In this region,  $p$  takes the constant value zero. The adjoint remains constant along the classical backwards characteristics outside this region. Hence, the reversible solution  $p$  is given by

$$p(0, x) = \begin{cases} 1, & x < -\frac{1}{2}, \\ 0, & -\frac{1}{2} \leq x \leq \frac{1}{2}, \\ -1, & \frac{1}{2} < x. \end{cases} \tag{11}$$

### 3 Discrete adjoint WENO3 method

In order to discretize (4) in space, we now consider solutions with compact support  $[a, b]$  for the entire time interval  $[0, T]$  and take  $y(a) = y(b) = 0$  as boundary conditions. Thus, using as many stencil points outside as needed and the compactness of the solution, the implementation of the zero boundary condition does not bear any difficulty. The interval  $[a, b]$  is partitioned into subintervals  $[x_{j-1/2}, x_{j+1/2}]$  of the same size  $\Delta x$  and with midpoints  $x_j$  for  $j = 1, \dots, N$ . Setting  $\mathbf{u}_0 := (u_0(x_1), \dots, u_0(x_N))^T$  and defining spatial approximations  $\mathbf{y}(t) := (y_1(t), \dots, y_N(t))^T$  with  $y_j(t) \approx y(t, x_j)$ , a spatial semi-discretization of (4) reads



**Fig. 1** Construction of the reversible solution  $p(t, x)$  from the end data  $p^T(x)$  at  $T = 0.5$ . The shock funnel region is accentuated as grey-colored triangle

$$y'(t) = -F_{\Delta x}(y(t)), \quad y(0) = u_0 \in \mathbb{R}^N, \tag{12}$$

where the nonlinear operator  $F_{\Delta x} : \mathbb{R}^N \rightarrow \mathbb{R}^N$  represents the discretization of  $\partial_x f(y)$ . We choose a conservative finite difference

$$(F_{\Delta x}(y(t)))_j = \frac{1}{\Delta x} (\hat{f}_{j+\frac{1}{2}} - \hat{f}_{j-\frac{1}{2}}), \tag{13}$$

where  $\hat{f}_{j+1/2} : \mathbb{R}^m \rightarrow \mathbb{R}$  denotes the numerical flux at  $x_{j+1/2}$ , which is (at least) a Lipschitz continuous function of  $m$  neighboring values  $y_i(t)$ . In order to avoid the convergence of the scheme towards entropy violating solutions, we apply a global flux splitting

$$\hat{f}_{j+\frac{1}{2}} = \hat{f}_{j+\frac{1}{2}}^+ + \hat{f}_{j+\frac{1}{2}}^-. \tag{14}$$

Using the simple Lax–Friedrichs splitting  $f^\pm(y) = (f(y) \pm \alpha y)/2$  with  $\alpha := \max_u |f'(u)|$  yields the desired properties  $(f^+)'(y) \geq 0$  and  $(f^-)'(y) \leq 0$ . Then, the numerical flux functions of the WENO3 method [26] are defined by

$$\hat{f}_{j+\frac{1}{2}}^+(y_{j-1}, y_j, y_{j+1}) := \omega_1^+ \left( -\frac{1}{2}f^+(y_{j-1}) + \frac{3}{2}f^+(y_j) \right) + \omega_2^+ \left( \frac{1}{2}f^+(y_j) + \frac{1}{2}f^+(y_{j+1}) \right), \tag{15}$$

$$\hat{f}_{j-\frac{1}{2}}^-(y_{j-1}, y_j, y_{j+1}) := \omega_2^- \left( -\frac{1}{2}f^-(y_{j+1}) + \frac{3}{2}f^-(y_j) \right) + \omega_1^- \left( \frac{1}{2}f^-(y_j) + \frac{1}{2}f^-(y_{j-1}) \right), \tag{16}$$

where the weights are

$$\omega_m^\pm = \frac{\tilde{\omega}_m^\pm}{\sum_{i=1,2} \tilde{\omega}_i^\pm}, \quad \tilde{\omega}_m^\pm = \frac{\gamma_m^\pm}{(\varepsilon + \beta_m^\pm)^2}, \quad m = 1, 2. \tag{17}$$

The smoothness indicators are given by

$$\beta_1^\pm = (f^\pm(y_j) - f^\pm(y_{j-1}))^2, \quad \beta_2^\pm = (f^\pm(y_{j+1}) - f^\pm(y_j))^2 \tag{18}$$

and the linear weights are set to  $\gamma_1^+ = \gamma_2^- = 1/3$ ,  $\gamma_1^- = \gamma_2^+ = 2/3$ . Note that  $0 < \varepsilon \ll 1$  is chosen in order to avoid the denominator becoming zero. It is set to  $\varepsilon = 10^{-6}$  in our numerical calculations. We would like to emphasize the observation that by construction the numerical fluxes  $\hat{f}^\pm$  have the same smoothness dependency on its arguments as that of the physical flux function  $f(y)$ .

Next we will derive the associated adjoint WENO3 scheme. Let  $f \in C^2(\mathbb{R})$ , i.e., there exists the Fréchet derivative of  $F_{\Delta x}$  defined in (13). The continuous optimal control problem is approximated by

$$u_0^{min} = \arg \min_{u_0 \in U_{ad}} \sum_{j=1}^N G(y_j(T), y_d(x_j)), \tag{19}$$

where  $U_{ad} = \{\mathbf{u} \in \mathbb{R}^N : \text{TV}(\mathbf{u}) \leq C\}$  is the discrete admissible set. Then, applying the common Lagrangian approach in  $\mathbb{R}^N$  with multipliers  $\mathbf{p}(t) = (p_1(t), \dots, p_N(t))^T$ , the adjoint equation to (12) reads

$$\mathbf{p}'(t) = \nabla_{\mathbf{y}} F_{\Delta x}(\mathbf{y}(t))^T \mathbf{p}(t), \quad \mathbf{p}(T) = (\partial_{\mathbf{y}} G(y_j(T), y_d(x_j)))_{j=1, \dots, N}, \tag{20}$$

where  $\nabla_{\mathbf{y}} F_{\Delta x}$  is the Fréchet derivative of  $F_{\Delta x}$  and gradients are treated as row vectors. The initial condition (the adjoint equation works backwards in time) is the discrete counterpart to (9). Observe that the interior boundary condition does not appear here. A short calculation yields the componentwise description

$$p'_j(t) = \frac{1}{\Delta x} \sum_{i=-2}^2 \partial_{y_j} L_{i,j}(\mathbf{y}(t)) p_{j+i}(t), \quad j = 1, \dots, N, \tag{21}$$

with the coefficients

$$\begin{aligned} L_{-2,j}(\mathbf{y}) &= \hat{f}_{j-3/2}^-, \\ L_{-1,j}(\mathbf{y}) &= \hat{f}_{j-1/2}^+ + \hat{f}_{j-1/2}^- - \hat{f}_{j-3/2}^-, \\ L_{0,j}(\mathbf{y}) &= \hat{f}_{j+1/2}^+ + \hat{f}_{j+1/2}^- - \hat{f}_{j-1/2}^+ - \hat{f}_{j-1/2}^-, \\ L_{1,j}(\mathbf{y}) &= \hat{f}_{j+3/2}^+ - \hat{f}_{j+1/2}^+ - \hat{f}_{j+1/2}^-, \\ L_{2,j}(\mathbf{y}) &= -\hat{f}_{j+3/2}^+. \end{aligned} \tag{22}$$

The indices of the numerical flux functions are directly related to their arguments, e.g.  $\hat{f}_{j+3/2}^+(y_j, y_{j+1}, y_{j+2})$  due to (15). For later use, we note that  $\sum_{i=-2, \dots, 2} L_{i,j}(\mathbf{y}) = 0$ .

We will now study the consistency order of the adjoint WENO3 scheme, i.e., how accurately does the semi-discretization (20) approximate the continuous adjoint equation (8) in the case of smooth solutions. Inserting exact solution values  $p(t, x_j)$  and  $\mathbf{y}(t, x_j)$  (still denoted by  $y_j$  to simplify notation) in the semi-discrete scheme (21) gives the residual-type local spatial errors

$$r_j(t) = \partial_t p(t, x_j) - \frac{1}{\Delta x} \sum_{i=-2}^2 \partial_{y_j} L_{i,j}(\mathbf{y}(t)) p(t, x_{j+i}). \tag{23}$$

Taylor expansion around  $x_j$  yields

$$r_j(t) = \partial_t p(t, x_j) - \sum_{k \geq 0} \Delta x^k \frac{1}{(k+1)!} \sum_{i=-2}^2 i^{k+1} \partial_{y_j} L_{i,j}(\mathbf{y}(t)) \partial_x^{k+1} p(t, x_j), \tag{24}$$

where we have already used that the sum of the  $L_{i,j}$  disappears. The method is said to have adjoint consistency order  $q$  if  $r_j(t) = \mathcal{O}(\Delta x^q)$ . In what follows, we will show that the adjoint WENO3 scheme satisfies all conditions for order  $q = 3$ .

First, we have to calculate  $\partial_{y_j} L_{i,j}$ , i.e., particularly the derivatives of the numerical flux functions defined in (15), (16). Since  $\omega_1^\pm + \omega_2^\pm = 1$  for all  $\mathbf{y}(t)$ , we deduce  $\partial_{y_k} \omega_1^\pm = -\partial_{y_k} \omega_2^\pm$ . Introducing the notation

$$\bar{f}_j^\pm(y_{j-1}, y_j, y_{j+1}) := \mp \frac{1}{2} f^+(y_{j-1}) \pm f^+(y_j) \mp \frac{1}{2} f^+(y_{j+1}), \tag{25}$$

we find

$$\begin{aligned} \partial_{y_{j-1}} \hat{f}_{j+1/2}^+ &= \partial_{y_{j-1}} \omega_1^+ \bar{f}_j^+ - \frac{1}{2} (f^+)'(y_{j-1}) w_1^+, \\ \partial_{y_j} \hat{f}_{j+1/2}^+ &= \partial_{y_j} \omega_1^+ \bar{f}_j^+ + (f^+)'(y_j) \left( \frac{3}{2} w_1^+ + \frac{1}{2} w_2^+ \right), \\ \partial_{y_{j+1}} \hat{f}_{j+1/2}^+ &= \partial_{y_{j+1}} \omega_1^+ \bar{f}_j^+ + \frac{1}{2} (f^+)'(y_{j+1}) w_2^+, \end{aligned} \tag{26}$$

and

$$\begin{aligned} \partial_{y_{j-1}} \hat{f}_{j-1/2}^- &= \partial_{y_{j-1}} \omega_1^- \bar{f}_j^- + \frac{1}{2} (f^-)'(y_{j-1}) w_1^-, \\ \partial_{y_j} \hat{f}_{j-1/2}^- &= \partial_{y_j} \omega_1^- \bar{f}_j^- + (f^-)'(y_j) \left( \frac{1}{2} w_1^- + \frac{3}{2} w_2^- \right), \\ \partial_{y_{j+1}} \hat{f}_{j-1/2}^- &= \partial_{y_{j+1}} \omega_1^- \bar{f}_j^- - \frac{1}{2} (f^-)'(y_{j+1}) w_2^-. \end{aligned} \tag{27}$$

We have the following three lemmata.

**Lemma 3.1** *Suppose  $f(y), y(t, \cdot) \in C^2(\mathbb{R})$ . Then*

$$\partial_{y_k} \omega_1^\pm \bar{f}_j^\pm(y_{j-1}, y_j, y_{j+1}) = \mathcal{O}(\Delta x^3), \quad k = j - 1, j, j + 1. \tag{28}$$

**Proof** Taylor expansion gives  $\bar{f}_j^\pm = \mathcal{O}(\Delta x^2)$ . It remains to show that  $\partial_{y_k} \omega_1^\pm = \mathcal{O}(\Delta x)$ . Indeed, we have

$$\partial_{y_k} \omega_1^\pm = \frac{\tilde{\omega}_2^\pm}{(\tilde{\omega}_1^\pm + \tilde{\omega}_2^\pm)^2} \partial_{y_k} \tilde{\omega}_1^\pm - \frac{\tilde{\omega}_1^\pm}{(\tilde{\omega}_1^\pm + \tilde{\omega}_2^\pm)^2} \partial_{y_k} \tilde{\omega}_2^\pm. \tag{29}$$

Since  $\beta_i^\pm = \mathcal{O}(\Delta x^2)$  for smooth flux functions  $f(y)$ , the two quotients are bounded by  $(\tilde{\omega}_i^\pm)^{-1} = \mathcal{O}(\varepsilon^2)$ ,  $i = 2, 1$ , respectively, for  $\Delta x \rightarrow 0$ . Taylor expansions of the derivatives  $\partial_{y_k} \tilde{\omega}_i^\pm = -2\gamma_i^\pm (\varepsilon + \beta_i^\pm)^{-3} \partial_{y_k} \beta_i^\pm$ ,  $i = 1, 2$ , show  $\mathcal{O}(\Delta x)$  for these terms and therefore also for  $\partial_{y_k} \omega_1^\pm$ .  $\square$

**Lemma 3.2** *Let  $\{x_{j-1}, x_j, x_{j+1}\}$  and  $\{x_j, x_{j+1}, x_{j+2}\}$  be two neighboring stencils and  $w_{i,j}^\pm, w_{i,j+1}^\pm, i = 1, 2$ , the corresponding weights. Suppose  $f(y), y(t, \cdot) \in C^3(\mathbb{R})$ . Then*

$$w_{i,j+1}^\pm - w_{i,j}^\pm = \mathcal{O}(\Delta x^4), \quad i = 1, 2. \tag{30}$$

**Proof** We consider the weights  $w_{1,j}^+$  and  $w_{1,j+1}^+$ . Analogous calculations can be done for the other cases. We set  $h(x) := f^+(y(x))$  and define  $h_j := f^+(y(x_j))$ . Then



$$\begin{aligned}
 w_{1,j+1}^+ - w_{1,j}^+ &= \frac{\tilde{\omega}_{1,j+1}^+}{\tilde{\omega}_{1,j+1}^+ + \tilde{\omega}_{2,j+1}^+} - \frac{\tilde{\omega}_{1,j}^+}{\tilde{\omega}_{1,j}^+ + \tilde{\omega}_{2,j}^+} \\
 &= \frac{\frac{\tilde{\omega}_{2,j}^+}{\tilde{\omega}_{1,j}^+} - \frac{\tilde{\omega}_{2,j+1}^+}{\tilde{\omega}_{1,j+1}^+}}{\left(1 + \frac{\tilde{\omega}_{2,j}^+}{\tilde{\omega}_{1,j}^+}\right) \left(1 + \frac{\tilde{\omega}_{2,j+1}^+}{\tilde{\omega}_{1,j+1}^+}\right)}.
 \end{aligned}
 \tag{31}$$

Due to the strict positivity of the weights, it remains to study the asymptotic behaviour of the numerator. Using the definitions (17) and (18), we have

$$D_{\tilde{w}} := \frac{\tilde{\omega}_{2,j}^+}{\tilde{\omega}_{1,j}^+} - \frac{\tilde{\omega}_{2,j+1}^+}{\tilde{\omega}_{1,j+1}^+} = \frac{\gamma_2^+}{\gamma_1^+} \frac{(\epsilon + \beta_{1,j}^+)^2(\epsilon + \beta_{2,j+1}^+)^2 - (\epsilon + \beta_{1,j+1}^+)^2(\epsilon + \beta_{2,j}^+)^2}{(\epsilon + \beta_{2,j}^+)^2(\epsilon + \beta_{2,j+1}^+)^2}
 \tag{32}$$

with the smoothness indicators

$$\begin{aligned}
 \beta_{1,j}^+ &= (h_j - h_{j-1})^2, & \beta_{2,j}^+ &= (h_{j+1} - h_j)^2, \\
 \beta_{1,j+1}^+ &= (h_{j+1} - h_j)^2, & \beta_{2,j+1}^+ &= (h_{j+2} - h_{j+1})^2.
 \end{aligned}
 \tag{33}$$

Taylor expansion at  $x_j$  yields in (32)

$$D_{\tilde{w}} = \frac{\gamma_2^+}{\epsilon^4 \gamma_1^+} \left( 4\epsilon^3 \Delta x^4 \left( (h_j'')^2 + h_j' h_j''' \right) + \mathcal{O}(\epsilon^3 \Delta x^5) \right),
 \tag{34}$$

which shows the assertion. □

**Lemma 3.3** *Suppose  $f(y)$ ,  $y(t, \cdot) \in C^2(\mathbb{R})$ . Then*

$$\omega_1^\pm - \gamma_1^\pm = \mathcal{O}(\Delta x^3).
 \tag{35}$$

**Proof** We first consider  $\omega_1^+ - \gamma_1^+$ . The difference can be expressed by

$$\omega_1^+ - \gamma_1^+ = \frac{\tilde{\omega}_1^+ - \gamma_1^+(\tilde{\omega}_1^+ + \tilde{\omega}_2^+)}{\tilde{\omega}_1^+ + \tilde{\omega}_2^+}.
 \tag{36}$$

The denominator is bounded from below by  $(\gamma_1^+ + \gamma_2^+)\epsilon^{-2} = \epsilon^{-2} > 0$ . Further, we deduce for the numerator

$$N_{\tilde{\omega}} := \tilde{\omega}_1^+ - \gamma_1^+(\tilde{\omega}_1^+ + \tilde{\omega}_2^+) = \gamma_1^+ \gamma_2^+ \frac{(\epsilon + \beta_2^+)^2 - (\epsilon + \beta_1^+)^2}{(\epsilon + \beta_1^+)^2 (\epsilon + \beta_2^+)^2}.
 \tag{37}$$

Let  $h(x) := f^+(y(x))$  and define  $h_j := f^+(y(x_j))$ . Inserting the smoothness indicators  $\beta_1^+ = (h_j - h_{j-1})^2$  and  $\beta_2^+ = (h_{j+1} - h_j)^2$ , Taylor expansion at  $x_j$  yields

$$N_{\tilde{\omega}} = \frac{\gamma_1^+ \gamma_2^+}{\varepsilon^4} \left( 4\varepsilon \Delta x^3 h'_j h''_j + \mathcal{O}(\Delta x^5) \right). \tag{38}$$

Putting this together with the bound for the denominator stated above gives  $\omega_1^+ - \gamma_1^+ = \mathcal{O}(\Delta x^3/\varepsilon)$ , from which we can conclude the proof. The same arguments apply to the second difference  $\omega_1^- - \gamma_1^-$ .  $\square$

We are now ready to state the main result of this section.

**Theorem 3.1** *Let  $f(y), y(t, \cdot) \in C^3(\mathbb{R})$  and  $p(t, \cdot) \in C^4(\mathbb{R})$ . Then the adjoint WENO3 scheme (20) is adjoint consistent of order three, i.e.,  $r_j(t) = \mathcal{O}(\Delta x^3)$  in (24).*

**Proof** Let us define  $d_k := \sum_{i=-2, \dots, 2} i^{k+1} \partial_{y_j} L_{i,j}(\mathbf{y}(t))$ ,  $k = 0, 1, 2$ , and denote by  $w_{i,m}^\pm$  the weights that correspond to the stencils  $\{x_{m-1}, x_m, x_{m+1}\}$ ,  $m = j - 1, j, j + 1$ . From (22), we calculate

$$\begin{aligned} d_0 &= -\partial_{y_j} (2L_{-2,j}(\mathbf{y}(t)) + L_{-1,j}(\mathbf{y}(t)) - L_{1,j}(\mathbf{y}(t)) - 2L_{2,j}(\mathbf{y}(t))) \\ &= -\partial_{y_j} \left( \hat{f}_{j-3/2}^- + \hat{f}_{j-1/2}^+ + \hat{f}_{j-1/2}^- + \hat{f}_{j+1/2}^+ + \hat{f}_{j+1/2}^- + \hat{f}_{j+3/2}^+ \right), \end{aligned} \tag{39}$$

which gives by using (26), (27) for different stencils and Lemma 3.1 for all terms  $\partial_{y_j} \omega_1^\pm \tilde{f}_m^\pm$  with  $m = j - 1, j, j + 1$ ,

$$\begin{aligned} d_0 &= - \left( \frac{1}{2} w_{1,j}^- + \frac{1}{2} w_{1,j+1}^- - \frac{1}{2} w_{2,j-1}^- + \frac{3}{2} w_{2,j}^- \right) (f^-)'(y_j) \\ &\quad - \left( \frac{3}{2} w_{1,j}^+ - \frac{1}{2} w_{1,j+1}^+ + \frac{1}{2} w_{2,j-1}^+ + \frac{1}{2} w_{2,j}^+ \right) (f^+)'(y_j) + \mathcal{O}(\Delta x^3). \end{aligned} \tag{40}$$

Eventually, Lemma 3.2 and the property  $w_{1,j}^\pm + w_{2,j}^\pm = 1$  yields

$$d_0 = -((f^-)'(y_j) + (f^+)'(y_j)) + \mathcal{O}(\Delta x^3) = -f'(y_j) + \mathcal{O}(\Delta x^3). \tag{41}$$

Analogously, we derive

$$\begin{aligned} d_1 &= \left( \frac{1}{2} w_{1,j}^- - \frac{1}{2} w_{1,j+1}^- - \frac{3}{2} w_{2,j-1}^- + \frac{3}{2} w_{2,j}^- \right) (f^-)'(y_j) \\ &\quad + \left( -\frac{3}{2} w_{1,j}^+ + \frac{3}{2} w_{1,j+1}^+ + \frac{1}{2} w_{2,j-1}^+ - \frac{1}{2} w_{2,j}^+ \right) (f^+)'(y_j) + \mathcal{O}(\Delta x^3) \end{aligned} \tag{42}$$

and

$$\begin{aligned} d_2 &= \left( -\frac{1}{2} w_{1,j}^- - \frac{1}{2} w_{1,j+1}^- + \frac{7}{2} w_{2,j-1}^- - \frac{3}{2} w_{2,j}^- \right) (f^-)'(y_j) \\ &\quad + \left( -\frac{3}{2} w_{1,j}^+ + \frac{7}{2} w_{1,j+1}^+ - \frac{1}{2} w_{2,j-1}^+ - \frac{1}{2} w_{2,j}^+ \right) (f^+)'(y_j) + \mathcal{O}(\Delta x^3). \end{aligned} \tag{43}$$

Lemma 3.2 directly shows that  $d_1 = \mathcal{O}(\Delta x^3)$ . Using  $w_1^\pm + w_2^\pm = 1$  and again Lemma 3.2, the linear combinations of the weights in  $d_2$  can be simplified to

$2 - 3w_{1,j}^-$  and  $3w_{1,j}^+ - 1$  up to order  $\mathcal{O}(\Delta x^4)$ , respectively. Applying now Lemma 3.3 with  $\gamma_1^+ = 1/3$  and  $\gamma_1^- = 2/3$  to these expressions gives  $d_2 = \mathcal{O}(\Delta x^3)$ .

In a last step, we use the asymptotic expressions for  $d_i, i = 0, 1, 2$ , to calculate the residual-type local spatial error

$$\begin{aligned} r_j(t) &= \partial_t p(t, x_j) - \sum_{k=0}^2 \Delta x^k \frac{1}{(k+1)!} d_k \partial_x^{k+1} p(t, x_j) + \mathcal{O}(\Delta x^3) \\ &= \partial_t p(t, x_j) + f'(y(t, x_j)) \partial_x p(t, x_j) + \mathcal{O}(\Delta x^3) = \mathcal{O}(\Delta x^3). \end{aligned} \tag{44}$$

This concludes the proof. □

### 4 Numerical experiments

In this section, we will present some numerical examples for Burgers equation, i.e., we study problems with the nonlinear flux function  $f(y) = \frac{1}{2}y^2$  in (4). The first example with smooth initial data and solution is chosen in order to check to third-order convergence of the discrete adjoint WENO3 method as stated in Theorem 3.1. In the second example, the approximation property of the discrete adjoint in the case of a shock in the initial solution is investigated and compared to approximations computed by means of the first-order modified Lax–Friedrichs (LF) and Engquist–Osher (EO) schemes. These schemes read

$$\begin{aligned} y_j^0 &= u_0(x_j), \\ y_j^{n+1} &= y_j^n - \frac{\Delta x}{\Delta t} \left( \hat{f}(y_j^n, y_{j+1}^n) - \hat{f}(y_{j-1}^n, y_j^n) \right), \quad n = 0, \dots, n_T - 1, \end{aligned} \tag{45}$$

with  $y_j^n \approx y(n\Delta t, x_j)$ ,  $n_T \Delta t = T$ , and numerical fluxes given by

$$\begin{aligned} \hat{f}_{LF}(a, b) &= \frac{1}{2}(f(b) + f(a)) - \frac{\gamma}{2} \frac{\Delta x}{\Delta t} (b - a), \quad \gamma \in (0, 1), \\ \hat{f}_{EO}(a, b) &= f(0) + \int_0^a \max(0, f'(s)) ds + \int_0^b \min(0, f'(s)) ds. \end{aligned} \tag{46}$$

Applying a standard Lagrangian approach and discrete adjoint calculus, the discrete adjoint schemes can be derived from [16, Prep. 3.1] as

$$\begin{aligned} p_j^{n_T} &= \partial_y G(y_j^{n_T}, y_d(x_j)), \\ p_j^n &= c_{j-1} p_{j-1}^{n+1} + c_j p_j^{n+1} + c_{j+1} p_{j+1}^{n+1}, \quad n = n_T - 1, \dots, 0, \end{aligned} \tag{47}$$

with the coefficients

$$c_{j-1} = \frac{\gamma}{2} - \frac{\Delta t}{2\Delta x} f'(y_j^{n+1}), \quad c_j = 1 - \gamma, \quad c_{j+1} = \frac{\gamma}{2} + \frac{\Delta t}{2\Delta x} f'(y_j^{n+1}), \tag{48}$$

for the LF scheme and

$$\begin{aligned}
 c_{j-1} &= \frac{\Delta t}{2\Delta x} \left( |f'(y_j^{n+1})| - f'(y_j^{n+1}) \right), & c_j &= 1 - \frac{\Delta t}{\Delta x} |f'(y_j^{n+1})|, \\
 c_{j+1} &= \frac{\Delta t}{2\Delta x} \left( |f'(y_j^{n+1})| + f'(y_j^{n+1}) \right),
 \end{aligned}
 \tag{49}$$

for the EO scheme. Convergence of these schemes has been intensively studied in [1, 10–12, 29]. The choice  $\gamma = 1$  leads to the classical LF method. Stability requirements for the adjoint LF and EO schemes yield the optimal value  $\gamma^* = 0.5$  together with the CFL-condition  $\Delta t \leq \gamma^* \Delta x / \sup |f'(y)|$ , see e.g. [16]. Then, both schemes converge for Lipschitz continuous end data  $p^T(x)$  in (8). The stronger condition  $\Delta t \leq \gamma^* (\Delta x)^{2-q} / \sup |f'(y)|$ ,  $0 < q < 1$ , ensures the convergence of the modified LF scheme for discontinuous end data, too [11, 12]. Convergence for slightly modified end data and less numerical viscosity has been recently studied in [1].

In order to get a fully discrete scheme for WENO3, the differential equation (12) is numerically solved by the three-stage third-order strong-stability-preserving Runge-Kutta method SSPRK3, which offers good stability properties [13, 14, 16, 19]. In the Shu-Osher representation, it reads

$$\begin{aligned}
 y_0^n &= y^n, \\
 y_1^n &= y_0^n - \Delta t F_{\Delta x}(y_0^n), \\
 y_2^n &= \frac{3}{4}y_0^n + \frac{1}{4}y_1^n - \frac{1}{4}\Delta t F_{\Delta x}(y_1^n), \\
 y^{n+1} &= \frac{1}{4}y_0^n + \frac{2}{3}y_2^n - \frac{2}{3}\Delta t F_{\Delta x}(y_2^n), \quad n = 0, \dots, n_T - 1.
 \end{aligned}
 \tag{50}$$

The corresponding adjoint time discretization has the form (see e.g. [16])

$$\begin{aligned}
 p_0^n &= p^{n+1}, \\
 p_1^n &= \frac{2}{3}p_0^n - \frac{2}{3}\Delta t \nabla_y F_{\Delta x}(y_2^n)^T p_0^n, \\
 p_2^n &= \frac{1}{4}p_1^n - \frac{1}{4}\Delta t \nabla_y F_{\Delta x}(y_1^n)^T p_1^n, \\
 p^n &= \frac{1}{3}p_0^n + \frac{3}{4}p_1^n + p_2^n - \Delta t \nabla_y F_{\Delta x}(y_0^n)^T p_2^n, \quad n = n_T - 1, \dots, 0.
 \end{aligned}
 \tag{51}$$

We note that the adjoint scheme has only order two, which is the upper barrier for three-stage third-order SSPRK methods [16].

In the final experiment, we solve an optimal control problem with a discontinuous target, proposed in [16]. The discrete adjoint  $p^0$  provides gradient information, which can be directly used to set up the following algorithm:

0. Given a control  $u_0 := u^{(j)}$  at iteration  $j$ .
1. Compute the discrete adjoint  $p^0(u^{(j)})$  and update  $u^{(j+1)} = u^{(j)} - \alpha_j p^0(u^{(j)})$  with  $\alpha_j$  such that Armijo’s condition

$$J(\mathbf{y}^{n_T}(\mathbf{u}^{(j+1)}), \mathbf{y}_d) \leq J(\mathbf{y}^{n_T}(\mathbf{u}^{(j)}), \mathbf{y}_d) - \frac{1}{2} \alpha_j \| \mathbf{p}^0(\mathbf{u}^{(j)}) \|_{L^2(I)}^2$$

is fulfilled. If it is not satisfied, choose  $\alpha_j := 0.95 \alpha_j$  and check the condition again.

2. Stop if  $|J(\mathbf{y}^{n_T}(\mathbf{u}^{(j+1)}), \mathbf{y}_d) - J(\mathbf{y}^{n_T}(\mathbf{u}^{(j)}), \mathbf{y}_d)| \leq tol$ . Otherwise set  $j := j + 1$  and proceed with step 1.

In general, taking the adjoint as a decent direction may increase the complexity of the optimization process due to the production of additional discontinuities [5, 23, 24]. A careful choice of the initial guess  $\mathbf{u}_0$  can remedy this serious problem. We follow the approach proposed in [17] and first solve the conservation law

$$\begin{aligned} \partial_t z + \partial_x f(z) &= 0, & (t, x) \in \Omega_T, \\ z(0, x) &= y_d(-x), \end{aligned} \tag{52}$$

where  $y_d$  is the target given in (1). The initial guess is then chosen as  $u_0 = z(T, -x)$ . Formally, as pointed out in [17], (52) is obtained by reverting  $t$  and  $x$  in (4) and taking  $y_d$  as initial condition. The advantage of this approach is that it delivers a control whose entropy solution is close to the target and the location of the discontinuities almost coincide. Hence, the production of additional discontinuities within each iteration step is avoided, which improves the performance of the algorithm drastically. We will exemplify the influence of the choice of the initial guess in our optimal control problem.

### 4.1 Order test for the discrete adjoint for smooth data

This section is devoted to numerically verify the third-order convergence of the adjoint WENO3 scheme. For this purpose, we choose the computational domain  $\Omega_T = (0, 0.5] \times [-1.5, 1.5]$  and the objective functional

$$J(y(0.5, \cdot; u_0), 0) = \frac{1}{2} \int_{-\frac{3}{2}}^{\frac{3}{2}} y(0.5, x; u_0)^2 dx \tag{53}$$

with the smooth initial data

$$u_0(x) = \begin{cases} e^{-\frac{1}{1-x^2}}, & |x| < 1, \\ 0 & , |x| \geq 1. \end{cases} \tag{54}$$

The exact solution  $y(t, x)$  can be directly computed from the method of characteristics, i.e.,  $y(t, x) = u_0(x_0(x(t), t))$  with  $x_0(x(t), t)$  being the solution of the nonlinear equation  $x(t) = x_0 + u_0(x_0)t$ . A reference solution  $y_T \approx y(0.5, x)$  at the final time is computed by Newton’s method with a high tolerance  $10^{-14}$ .

Since shocks are not present, we find  $p^T(x) = y(0.5, x)$  in (8). We also note that the characteristics curves of the adjoint problem coincide with the characteristic

curves of the forward problem. Thus, the corresponding reversible solution  $p(0, x)$  at time  $t=0$  is given by  $u_0(x)$ , which serves as reference solution for the adjoint.

We use a sequence of spatial meshes with a number of grid points  $N = 150 \cdot 2^i$ ,  $i = 0, \dots, 6$ , and set  $\Delta t = 0.5 \Delta x$ . In order to keep the temporal error below  $\mathcal{O}((\Delta x)^3)$ , we apply the classical fourth-order four-stage explicit Runge-Kutta method (ERK4). Its adjoint time discretization has also order four [15] for smooth solutions and therefore the overall scheme is suitable to check the order three of the adjoint WENO3 method. We also present results for the forward WENO3 method to document the error of the approximated starting value  $\mathbf{p}^{n_T} = \mathbf{y}^{n_T}$ . The  $L^\infty$ -errors collected in Table 1 clearly show asymptotic order three of the spatial WENO3 discretization for both forward and adjoint numerical solution.

### 4.2 Approximation of the discrete adjoint in the case of shocks

We now consider discontinuous solutions with shocks. Our test case is taken from Example 2.1 with computational domain  $\Omega_T = (0, 0.5) \times [-1, 1]$ . The reversible solution  $p(0, x)$  at  $t=0$  is given by (11). We apply the above described forward and adjoint LF, EO, and WENO3 schemes with  $\Delta x = 0.01, 0.002$ , and  $\Delta t = 0.25 \Delta x$ . The corresponding numerical approximations  $\mathbf{p}^0$  are shown in Fig. 2.

The first-order LF and EO schemes smear out the discontinuities, but deliver  $L^\infty$ -stable approximations and thus respect the analytical property of the adjoint. In the spirit of WENO schemes, the adjoint WENO3 delivers a quite sharp resolution of the shocks at the price of bounded over- and undershoots of around 5%. In Table 2, we plot the  $L^\infty$ -error in the shock funnel for  $x \in [-0.3, 0.3]$ . All schemes converge quite rapidly. Note that convergence in the shock funnel is not always achieved since the interior boundary condition at shock positions as given in (9) does not appear on the discrete level, see the discussions in [1, 11, 12].

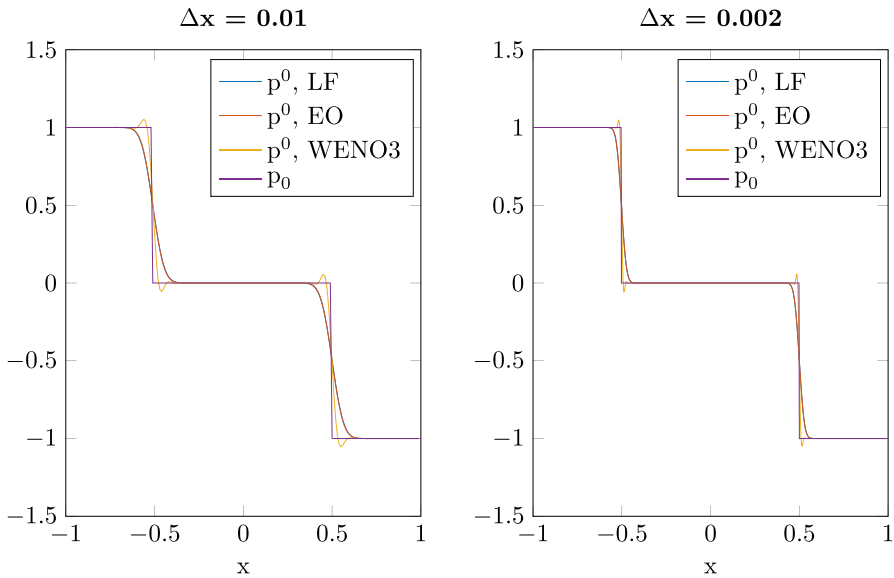
### 4.3 Optimal control problem with discontinuous target

We consider the optimal control problem (1) with the objective functional [16]

**Table 1** Burgers problem with smooth initial data and smooth solution:  $L^\infty$ -error of the forward solution  $\|y_T - y^{n_T}\|_\infty$  at the final time  $T = 0.5$  and adjoint solution  $\|u_0 - p^0\|_\infty$  at time  $t = 0$  for a sequence of spatial meshes with  $N = 150, 300, \dots, 9600$  grid points

N	$\ y_T - y^{n_T}\ _\infty$	Rate	$\ u_0 - p^0\ _\infty$	Rate
150	$2.00e-3$		$7.39e-3$	
300	$3.25e-4$	2.63	$9.37e-4$	2.98
600	$2.64e-5$	3.62	$7.14e-5$	3.71
1200	$2.16e-6$	3.62	$4.30e-6$	4.05
2400	$2.76e-7$	2.97	$5.49e-7$	2.97
4800	$3.46e-8$	2.99	$6.92e-8$	2.99
9600	$4.33e-9$	3.00	$8.66e-9$	3.00

The convergence rates are computed from  $\ln(E_N/E_{2N})/\ln(2)$ , where  $E_N$  stands for the corresponding error



**Fig. 2** Burgers problem with discontinuous initial and final solution taken from Example 2.1. Numerical approximations  $p^0$  to the reversible solution  $p_0 := p(0, x)$  given in (11) for the adjoint Lax–Friedrichs (LF), Engquist–Osher (EO) and WENO3 scheme applied with  $\Delta x = 0.01$  (left),  $\Delta x = 0.002$  (right), and  $\Delta t = 0.25 \Delta x$

**Table 2** Burgers problem with discontinuous initial and final solution taken from Example 2.1

$\Delta x$	LF	EO	WENO3
0.01	$4.91e-05$	$2.45e-05$	$3.92e-05$
0.002	$2.26e-17$	$5.79e-20$	$6.51e-16$

$L^\infty$ -error of the adjoint solution  $\|p_0 - p^0\|_\infty$  at time  $t = 0$  in the shock funnel  $x \in [-0.3, 0.3]$  for  $\Delta x = 0.01, 0.002$

$$J(y(0.5, \cdot; u_0), y_d(x)) = \frac{1}{2} \int_{-1}^1 |y(0.5, x; u_0) - y_d(x)|^2 dx \tag{55}$$

and the discontinuous target  $y_d$  defined by

$$y_d(x) = \begin{cases} 2x - \frac{1}{2}, & \frac{1}{4} \leq x \leq \frac{3}{4}, \\ 0, & \text{otherwise.} \end{cases} \tag{56}$$

The optimal control  $u_0^*$ , which serves as a reference solution, is

$$u_0^*(x) = \begin{cases} -2x + \frac{3}{2}, & \frac{1}{4} \leq x \leq \frac{3}{4}, \\ 0, & \text{otherwise.} \end{cases} \tag{57}$$

We will present results for two mesh sizes  $\Delta x = 0.005, 0.002$ , and time steps  $\Delta t = 0.25 \Delta x$ . The initial guess for the control is computed from (52) with the

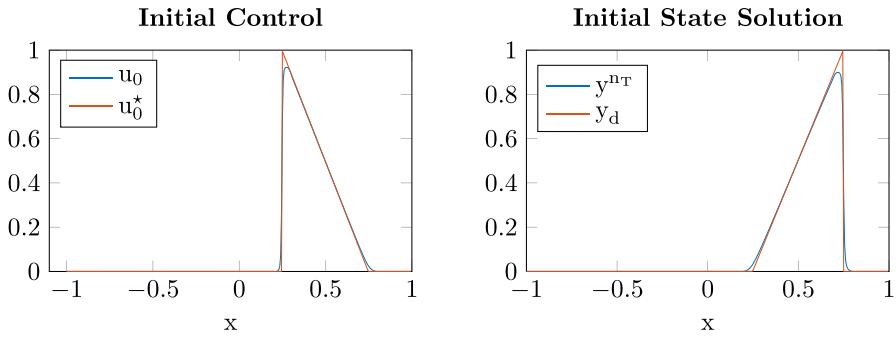


Fig. 3 Optimal control problem. Initial control  $u_0$  and optimal control  $u_0^*$  (left), initial state solution  $y^{nr}$  at  $T = 0.5$  and target  $y_d$  (right), computed with the WENO3 scheme and mesh size  $\Delta x = 0.005$

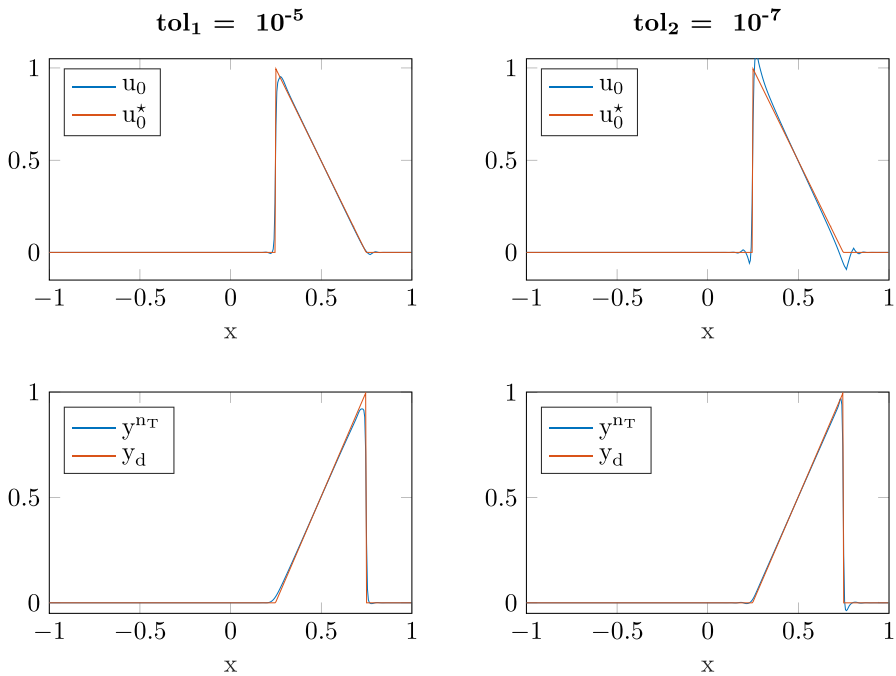


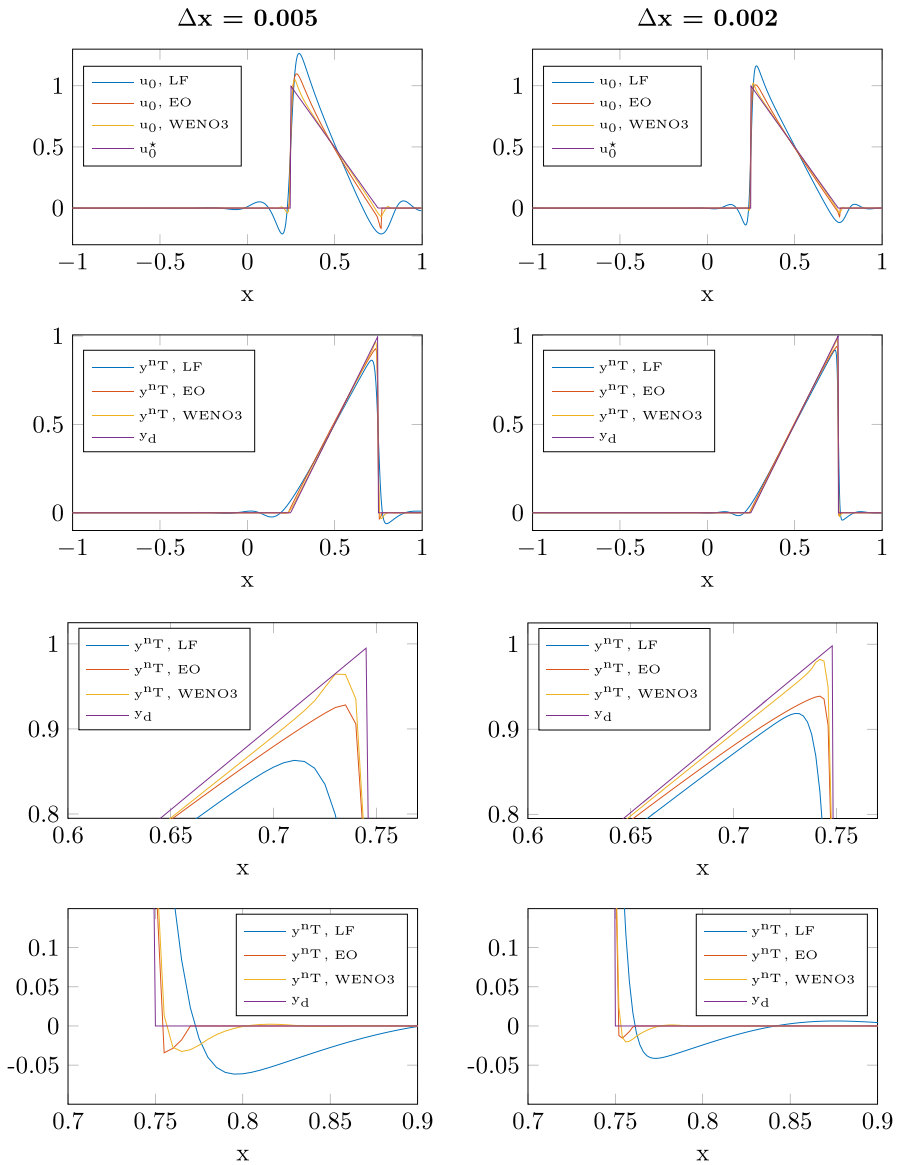
Fig. 4 Optimal control problem. Optimal control  $u_0^*$  and target  $y_d$ , numerically computed optimal control  $u_0$  and corresponding state solution  $y^{nr}$  for tolerances  $tol_1 = 10^{-5}$  (left) and  $tol_2 = 10^{-7}$  (right) using WENO3 with mesh size  $\Delta x = 0.005$

individual method under consideration. For WENO3 and the coarser mesh size, it is shown in Fig. 3 together with the corresponding state solution.

In Fig. 4, the results of the gradient based optimization procedure described above for tolerances  $tol_1 = 10^{-5}$ ,  $tol_2 = 10^{-7}$ , and mesh size  $\Delta x = 0.005$  for the adjoint WENO3 method are plotted. The value of the objective functional decreases



from  $4.75 \times 10^{-4}$  for  $tol_1$  to  $3.18 \times 10^{-4}$  for  $tol_2$ , resulting in a better approximation of the target  $y_d$ . We can conclude that the adjoint WENO3 method allows to recover the initial data together with the final state solution adequately. The shock of the target is sharply resolved and the rarefaction of the initial data is also recovered. In



**Fig. 5** Optimal control problem. Computed optimal control functions  $u_0$  (top) and corresponding state solution  $y^{nT}$  (above the middle) with a zoom into the shock region (below the middle, bottom) for 50 iterations of the gradient based optimization algorithm, using LF, EO, and WENO3 scheme with mesh size  $\Delta x = 0.005$  (left) and  $\Delta x = 0.002$  (right)

**Table 3** Optimal control problem

	LF	EO	WENO3	WENO3, $\mathbf{u}_0 = 0$
$\Delta x = 0.005$				
$\log(J_0)$	-4.68	-5.76	-7.30	-2.48
$\log(J_{50})$	-6.14	-7.80	-8.01	-6.14
$\Delta x = 0.002$				
$\log(J_0)$	-5.46	-6.47	-8.34	-2.48
$\log(J_{50})$	-7.02	-8.55	-8.96	-4.89

Logarithmic values of the objective functional (55) at the beginning and after 50 iterations of the optimization algorithm,  $J_0$  and  $J_{50}$ , respectively. For comparison, values for an initial control  $\mathbf{u}_0 = 0$  for WENO3 are shown, too

order to compare these results with those obtained from the LF and EO schemes, we perform 50 iterations of the optimization algorithm for both mesh sizes. The calculated optimal controls and their corresponding final state solutions are collected in Fig. 5. The adjoint WENO3 method resolves the shock sharply. In contrast, the LF method is too diffusive and only provides an unsatisfactory shock resolution. The numerical artifacts around the shocks are huge. The optimized final state solution obtained by the EO scheme possesses very small numerical artifacts, but the shock is less sharply resolved and the spike of the target is slightly smeared out. In Table 3, we depict the iteration history for all runs of the optimization. In every case, the LF method performs poorer than the others. In terms of a low cost functional, the adjoint WENO3 method performs best. We also see the influence of the initial guess on the performance of the algorithm. This is due to the fact that the use of  $\mathbf{u}_0 = 0$  as starting control value produces artificial discontinuities within each iteration step.

## 5 Summary

We have developed a novel adjoint WENO3 scheme to provide approximations of the gradient for optimal control problems governed by hyperbolic conservation laws and proved third-order consistency in space for sufficiently smooth solutions. The adjoint WENO3 method is able to sharply resolve discontinuities of reversible solutions. For an exemplary optimal control problem with discontinuous target, the method works very well and outperforms common first-order methods as the Lax–Friedrichs and Engquist–Osher schemes.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This work was supported by the Graduate School CE within the Centre for Computational Engineering at Technische Universität Darmstadt and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the collaborative research centre TRR154 “Mathematical modeling, simulation and optimisation using the example of gas networks” (Project-ID 239904186, TRR154/2-2018, TP B01).

**Availability of data and material** The data are available on request.

**Code availability** All calculations have been done in MATLAB.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Aguilar, S.P., Schmitt, J.M., Ulbrich, S., Moos, M.: On the numerical discretization of optimal control problems for conservation laws. *Control Cybern.* **2/2019** (2019)
2. Banda, M.K., Herty, M.: Adjoint IMEX-based schemes for control problems governed by hyperbolic conservation laws. *Comput. Optim. Appl.* **51**(2), 909–930 (2012)
3. Bouchut, F., James, F.: One-dimensional transport equations with discontinuous coefficients. *Nonlinear Anal.* **32**(7), 891–933 (1998)
4. Bressan, A., Marson, A.: A variational calculus for discontinuous solutions of systems of conservation laws. *Commun. Partial Differ. Equ.* **20**(9–10), 1491–1552 (1995)
5. Castro, C.M., Palacios, F., Zuazua, E.: An alternating descent method for the optimal control of the inviscid Burgers equation in the presence of shocks. *Math. Models Methods Appl. Sci.* **18**(3), 369–416 (2008)
6. Chertock, A., Herty, M., Kurganov, A.: An Eulerian–Lagrangian method for optimization problems governed by multidimensional nonlinear hyperbolic PDEs. *Comput. Optim. Appl.* **59**(3), 689–724 (2014)
7. Conway, E.D.: Generalized solutions of linear differential equations with discontinuous coefficients and the uniqueness question for multidimensional quasilinear conservation laws. *J. Math. Anal. Appl.* **18**(2), 238–251 (1967)
8. Dafermos, C.M.: Generalized characteristics and the structure of solutions of hyperbolic conservation laws. *Indiana Univ. Math. J.* **26**, 1097–1119 (1977)
9. Frenzel, D.: Weighted Essentially Non-Oscillatory Schemes in Optimal Control Problems Governed by Nonlinear Hyperbolic Conservation Laws. Ph.D. thesis, Technical University of Darmstadt, Department of Mathematics, published in Reihe Mathematik, Verlag Dr. Hut, ISBN 978-3-8439-4454-0 (2020)
10. Giles, M.: Discrete adjoint approximations with shocks. In: Hou, T.Y., Tadmor, E. (eds.) *Hyperbolic Problems: Theory, Numerics, Applications: Proceedings of the Ninth International Conference on Hyperbolic Problems*, pp. 185–194. Springer, Berlin, Heidelberg (2003)
11. Giles, M., Ulbrich, S.: Convergence of linearized and adjoint approximations for discontinuous solutions of conservation laws. Part 1: linearized approximations and linearized output functionals. *SIAM J. Numer. Anal.* **48**(3), 882–904 (2010)
12. Giles, M., Ulbrich, S.: Convergence of linearized and adjoint approximations for discontinuous solutions of conservation laws. Part 2: adjoint approximations and extensions. *SIAM J. Numer. Anal.* **48**(3), 905–921 (2010)
13. Gottlieb, S., Ketcheson, D., Shu, C.-W.: *Strong Stability Preserving Runge–Kutta and Multistep Time Discretizations*. World Scientific Publishing Company (2011)

14. Gottlieb, S., Shu, C.-W., Tadmor, E.: Strong stability-preserving high-order time discretization methods. *SIAM Rev.* **43**, 89–112 (2001)
15. Hager, W.W.: Runge–Kutta methods in optimal control and the transformed adjoint system. *Numer. Math.* **87**, 247–282 (2000)
16. Hajian, S., Hintermüller, M., Ulbrich, S.: Total variation diminishing schemes in optimal control of scalar conservation laws. *IMA J. Numer. Anal.* **39**(1), 105–140 (2017)
17. Herty, M., Kurganov, A., Kurochkin, D.: Numerical method for optimal control problems governed by nonlinear hyperbolic systems of PDEs. *Commun. Math. Sci.* **13**(1), 15–48 (2015)
18. Herty, M., Kurganov, A., Kurochkin, D.: On convergence of numerical methods for optimization problems governed by scalar hyperbolic conservation laws. In: Klingenberg, C., Westdickenberg, M. (eds.) *Theory. Numerics and Applications of Hyperbolic Problems I*, pp. 691–706. Springer, Cham (2018)
19. Hintermüller, M., Strogies, N.: On the consistency of Runge Kutta methods up to order three applied to the optimal control of scalar conservation laws. *Springer Proc. Math. Stat.* **235**, 119–154 (2018)
20. Jiang, G.-S., Shu, C.-W.: Efficient implementation of weighted ENO schemes. *J. Comput. Phys.* **126**, 202–228 (1996)
21. Kruzkov, S.N.: First order quasilinear equations with several independent variables. *Math. USSR-Sbornik* **10**(2), 217–243 (1970)
22. Kurochkin, D.: Numerical Method For Constrained Optimization Problems Governed By Nonlinear Hyperbolic Systems Of PDEs. PhD thesis, University of New Orleans (2015)
23. Lecaros, R., Zuazua, E.: Tracking control of 1D scalar conservation laws in the presence of shocks. In: Ancona, V., Strickland, E. (eds.) *Trends in Contemporary Mathematics*, pp. 195–219. Springer, Cham (2014)
24. Lecaros, R., Zuazua, E.: Control of 2D scalar conservation laws in the presence of shocks. *Math. Comput.* **85**(299), 1183–1224 (2016)
25. Liu, X.-D., Osher, S., Chan, T.: Weighted essentially non-oscillatory schemes. *J. Comput. Phys.* **115**, 200–212 (1994)
26. Shu, C.-W.: Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws. In: Quarteroni, A. (ed.) *Advanced Numerical Approximation of Nonlinear Hyperbolic Equations. Lecture Notes in Mathematics*, vol. 1697, pp. 325–432. Springer, Berlin (1998)
27. Shu, C.-W.: Essentially non-oscillatory and weighted essentially non-oscillatory schemes. In: *Acta Numerica*, vol. 29, pp. 701–762. Cambridge University Press (2020)
28. Ulbrich, S.: Optimal Control of Partial Differential Equations (Chemnitz, 1998), Volume 133 of *Internat. Ser. Numer. Math.*, chapter On the existence and approximation of solutions for the optimal control of nonlinear hyperbolic conservation laws, pp. 287–299. Birkhäuser, Basel (1998)
29. Ulbrich, S.: Optimal Control of Nonlinear Hyperbolic Conservation Laws with Source terms. Technische Universität München, Germany, Habilitation, Zentrum Mathematik (2001)
30. Ulbrich, S.: A sensitivity and adjoint calculus for discontinuous solutions of hyperbolic conservation laws with source terms. *SIAM J. Control Optim.* **41**, 740–797 (2002)
31. Ulbrich, S.: Adjoint-based derivative computation for the optimal control of discontinuous solutions of hyperbolic conservation laws. *Syst. Control Lett.* **48**(3–4), 313–328 (2003)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.