

# Direct-expansion forms of ADER schemes for conservation laws and their verification

Yoko Takakura

*Department of Mechanical Systems Engineering, Tokyo Noko University, 2-24-16 Nakacho, Koganei, Tokyo 184-8588, Japan*

Received 29 March 2005; received in revised form 18 March 2006; accepted 1 May 2006

Available online 3 July 2006

---

## Abstract

To seek general-purpose numerical schemes for hyperbolic problems, the ADER approach has been reviewed on the state-series expansion forms and the direct expansion forms in the viewpoint of the numerical procedure and the accuracy, and the advantages and disadvantages of the latter forms have been discussed. As ADER direct expansion schemes, ADER-D (standard ones with Godunov states/fluxes) and ADER-waf (ones with WAF states/fluxes) are adopted. Then, verification has been carried out on the scalar conservation laws with a linear flux, nonlinear convex fluxes, and various types of nonlinear non-convex fluxes. Convergence studies have shown that all the ADER schemes achieve the designed order of accuracy up to small cell sizes, yield small errors even in large cell sizes, and have computational efficiency with keeping the CFL number close to unity. Capturability of discontinuity and rarefaction has been investigated. As results, the ADER schemes have worked well for the problem of long-time propagation in the linear cases and for the problems of complicated wave formation and interaction in the nonlinear cases corresponding to various types of convex and non-convex fluxes. It is remarkable that ADER-waf schemes have shown sharper resolvability than the other ADER schemes, but have less robustness.

© 2006 Elsevier Inc. All rights reserved.

*Keywords:* ADER approach; Direct-expansion forms; Convex fluxes; Non-convex fluxes; Wave formation

---

## 1. Introduction

It is very challenging to seek and develop more general-purpose numerical schemes for hyperbolic problems, because conventionally numerical schemes are optimized either for a linear or for a non-linear flow models. Even in a few cases which are optimized for both models, the model equations are the linear advection equation and the Burgers' equation with convex flux, and problems on non-convex fluxes have not been taken into account. However, in some scalar model equations there appears the non-convex property such as the Buckley–Leverett equations for two phase fluid flow in a porous media. Furthermore even in the Euler equations, there are occasions where the state equation is different from that of the usual fluid, for example, in the high-temperature gas with dissociation and in the atmosphere at entry of vehicles to planets

---

*E-mail address:* [takakura@cc.tuat.ac.jp](mailto:takakura@cc.tuat.ac.jp).

other than Earth, etc. In such cases fluxes are not always convex. Therefore for these future problems it is very important to seek the numerical algorithms to treat linearity, convexity, and non-convexity accurately and generally.

A candidate for the general-purpose schemes is the arbitrary accuracy derivative riemann problem (ADER) approach recently developed as extension of Godunov-type schemes [2]. The ADER approach [18–20] is to construct explicit, one-step advection schemes with very high order of accuracy in both space and time on the basis of the solution of GRP (generalized Riemann problem) obtained by use of the solutions of the conventional RP (Riemann problem) and derivative RPs (DRPs). Two methods are possible for ADER schemes to solve general nonlinear conservation laws: methods based on state-series expansion [19] and direct expansion [11]. Recently it has been extended, with high accuracy, to nonlinear conservation laws with source terms [12,13] and diffusion terms [14] and to those in multi-space dimensions [22], and applied to the practical fluid dynamics problems such as the Euler equation system [15].

This paper is intended to investigate, as a basic research, if the ADER approach can be the general-purpose algorithm. First, the ADER schemes based on the state-series expansion and the direct expansion are reviewed in the viewpoint of the numerical procedure and the accuracy. Special emphasis is placed on verification of ADER schemes in the direct expansion forms, and another version with high resolvability is also included in the verification. The advantages and the disadvantages of the ADER schemes in direct-expansion forms are discussed in comparison with those in state-series expansion forms. Then, numerical verifications are carried out on the scalar conservation laws:  $\partial_t q + \partial_x f(q) = 0$ . On the ADER schemes, verification has been successfully shown for the linear advection equations with  $f(q) = q$  and the Burgers' equations with  $f(q) = q^2/2$  (a nonlinear case with a convex flux) till now. This paper shows the verification of ADER schemes more inclusively to investigate the applicability for wider range of problems. Convergence is studied for the linear problem with  $f(q) = q$ , the nonlinear problems with convex fluxes  $f(q) = q^2/2, q^4/4$ , and nonlinear problems with non-convex ones  $f(q) = q^3/3, q^5/5$ . Furthermore verification is carried out on various patterns of wave formation and interaction on shocks and expansions for the fluxes above and another type of non-convex fluxes,  $f(q) = (1/4)(q^2 - 1)(q^2 - 4)$  and  $f(q) = q^2/(q^2 + a(1 - q)^2)$  (Buckley–Leverett equations).

## 2. Governing equation and conservative scheme

Here the initial value problem of the scalar conservation law is considered,

$$\partial_t q + \partial_x f(q) = 0, \tag{1}$$

together with the IC

$$(q, 0) \equiv q_0(x), \tag{2}$$

where  $q(x, t)$  is the conserved variable,  $f(q)$  is the flux function, and  $q_0(x)$  is the initial distribution of  $q$ . Eq. (1) can be described as follows:

$$\partial_t q + \lambda(q) \partial_x q = 0, \tag{3}$$

where  $\lambda(q)$  is the characteristic speed defined by

$$\lambda(q) \equiv \frac{df}{dq}. \tag{4}$$

Integrating (1) in time and space  $[\mathbf{x}_n, \mathbf{x}_{n+1}] \times [t_{i-\frac{1}{2}}, t_{i+\frac{1}{2}}]$  gives the conservative form

$$q_i^{n+1} - q_i^n - \frac{\Delta t}{\Delta x} [f_{i+\frac{1}{2}} - f_{i-\frac{1}{2}}], \tag{5}$$

where  $q_i^n$  is the spatial average of  $q$  at time  $t = t_n$

$$q_i^n = \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} (q(x, t_n)) dx, \tag{6}$$

$f_{i+\frac{1}{2}}$  is the time average of  $f(q)$  at cell interface  $x = x_{i+\frac{1}{2}}$

$$f_{i+\frac{1}{2}} = \frac{1}{\Delta} \int_{\mathbf{x}_n}^{\mathbf{x}_{n+1}} f(\mathbf{x}_{i+\frac{1}{2}}) d\mathbf{x}, \tag{7}$$

and cells, cell sizes, time-step sizes are defined by

$$I_i \equiv [ \mathbf{x}_{i-\frac{1}{2}}, \mathbf{x}_{i+\frac{1}{2}} ], \tag{8}$$

$$\Delta = \mathbf{x}_{i+\frac{1}{2}} - \mathbf{x}_{i-\frac{1}{2}}, \tag{9}$$

$$\Delta = \mathbf{x}_{n+1} - \mathbf{x}_n. \tag{10}$$

The ADER approach is a type of Godunov schemes [2] based on the conservative form (5).

### 3. ADER approach

Here the ADER approach with  $m$ th order of accuracy in time and  $r$ th order of accuracy in space is presented in the viewpoint of the numerical procedure and the accuracy. As to basic derivation, see Ref. [11]. When  $m = r$  is taken, the resulting ADER schemes have the  $r$ th order of accuracy in both time and space. First, the ADER schemes based on state-series expansion are presented, and then those based on direct expansion are explained.

#### 3.1. Method based on state-series expansion

The ADER approach in the state-series expansion consists of the following steps 3.1.1–3.1.4.

##### 3.1.1. Reconstruction and GRP

At each time  $t_n$ , the data in the form of cell-averages  $\bar{u}_i^n$  are reconstructed by piece-wise smooth functions  $p_i(x)$  for cell  $I_i$ . To avoid spurious oscillations in the vicinity of discontinuities, the ENO or WENO [3,9] polynomial interpolations are adopted. With  $r$  stencils, polynomial  $p_i(x)$  of degree at most  $r - 1$  can be constructed, and the spatial accuracy is  $r$ th order in the case of the ENO interpolations, and  $(2r - 1)$ th order in the case of the WENO interpolations:

$$\text{ENO} : p_i(\mathbf{x}) = \bar{u}_i^n + O(\Delta^n), \tag{11}$$

$$\text{WENO} : p_i(\mathbf{x}) = \bar{u}_i^n + O(\Delta^{2r-1}), \tag{12}$$

but the spatial accuracy for the  $k$ th order derivative for  $q$  is  $(r - k)$ th order in the both ones:

$$\text{ENO/WENO} : \partial^{(k)} p_i(\mathbf{x}) = \partial^{(k)} \bar{u}_i^n + O(\Delta^{r-k}). \tag{13}$$

From this reason, the order of spatial accuracy in the ADER schemes is  $r$ th, even if either the ENO or the WENO interpolations may be used. However, as the values of errors are smaller with the WENO, the WENO interpolations are adopted here. About the discussion on the accuracy of ADER approach, see Section 4.

Near each cell interface  $\mathbf{x}_{i+\frac{1}{2}}$  at time  $t_n$ , introduce

$$\begin{cases} \xi = \mathbf{x} - \mathbf{x}_{i+\frac{1}{2}}, \\ \tau = \mathbf{x} - \mathbf{x}_n \\ Q(\xi, \tau) = \bar{u}_{i+\frac{1}{2}} + \xi \bar{u}_{\mathbf{x}_n} + \tau \end{cases} \tag{14}$$

and consider the GRP having the following PDE and the IC on  $\xi \in (-\infty, +\infty)$  and  $\tau \in [0, +\infty)$ :

$$\text{PDE} : \partial_\tau Q + \partial_\xi f(Q) = 0, \tag{15}$$

$$\text{IC} : Q(\xi, 0) = \begin{cases} p_i(\mathbf{x}_{i+\frac{1}{2}} + \xi) & \text{if } \xi < 0, \\ p_{i+1}(\mathbf{x}_{i+\frac{1}{2}} + \xi) & \text{if } \xi > 0, \end{cases} \tag{16}$$

where the initial data are the reconstructed polynomial functions translated by  $-\mathbf{x}_{i+\frac{1}{2}}$ .

3.1.2. Expansion with Cauchy–Kowalewski procedure

When the solution of GRP (15) and (16) is differentiable on time and space up to the  $(m - 1)$ th order near  $\xi = 0$  and  $\tau \rightarrow +0$ , it is expressed as a time Taylor-series expansion:

$$Q(0, \tau) = Q(0, +0) + \sum_{k=1}^{m-1} \frac{\tau^k}{k!} \partial_\tau^{(k)} Q(0, +0) + O(\tau^m), \tag{17}$$

When conservation law (3) is linear with constant  $\lambda$ , all time derivatives of  $q$  can be replaced with space derivatives of  $q$ , using the governing equation:

$$\partial_\tau^{(k)} = (-\lambda)^k \partial_\xi^{(k)}, \tag{18}$$

which we call Lax–Wendroff [6] procedure. When conservation law (3) is nonlinear, Cauchy–Kowalewski procedure [4] is adopted as follows:

$$\partial_\tau = -\lambda, \tag{19}$$

$$\partial_\tau^2 = -\lambda_\xi - \lambda_{\xi\xi}, \tag{20}$$

$$\partial_\tau^3 = -\lambda_{\xi\xi} - \lambda_{\xi\xi\xi}, \tag{21}$$

$$\partial_\tau^4 = -\lambda_{\xi\xi\xi} - 2\lambda_{\xi\xi\xi\xi} - \lambda_{\xi\xi\xi\xi} - \lambda_{\xi\xi\xi\xi}^2, \tag{22}$$

$$\begin{cases} \partial_\tau^5 = -\lambda_{\xi\xi\xi\xi} - \lambda_{\xi\xi\xi\xi\xi} - 2\lambda_{\xi\xi\xi\xi\xi} - \lambda_{\xi\xi\xi\xi\xi}^2, \\ \partial_\tau^6 = -\lambda_{\xi\xi\xi\xi\xi} - 3\lambda_{\xi\xi\xi\xi\xi\xi} - \lambda_{\xi\xi\xi\xi\xi\xi}, \end{cases} \tag{23}$$

⋮

and time derivatives of  $q$  can be expressed with space derivatives of  $q$ :

$$\partial_\tau^{(k)} = \alpha^{(k)}(\alpha^{(0)}, \dots, \alpha^{(k)}). \tag{24}$$

Thus, the time Taylor-series expansion (17) leads to the following equation:

$$Q(0, \tau) = \left( q_{i+\frac{1}{2}}^{(0)} n + 0 \right) + \sum_{k=1}^{m-1} \frac{\tau^k}{k!} \alpha^{(k)} \left( \alpha^{(0)} \left( q_{i+\frac{1}{2}}^{(0)} n + 0 \right), \dots, \alpha^{(k)} \left( q_{i+\frac{1}{2}}^{(0)} n + 0 \right) \right) + O(\tau^m). \tag{25}$$

3.1.3. Solution of GRP by use of RP and DRPs

The solution of the GRP at  $\xi = 0$  is approximated in the  $m$ th order of accuracy as

$$Q_{i+\frac{1}{2}}^{\text{GRP}m}(0, \tau) = q_{i+\frac{1}{2}}^{(0)} + \sum_{k=1}^{m-1} \frac{\tau^k}{k!} \alpha^{(k)} \left( q_{i+\frac{1}{2}}^{(0)}, \dots, q_{i+\frac{1}{2}}^{(k)} \right). \tag{26}$$

Here  $q_{i+\frac{1}{2}}^{(0)}$  is the solution at  $(\xi, \tau) = (0, +0)$  of the conventional RP:

$$\text{PDE} : \partial_\tau Q + \partial_\xi f(Q) = 0, \tag{27}$$

$$\text{IC} : Q(\xi, 0) = \begin{cases} q_{i+\frac{1}{2}}^{(0)} & \text{if } \xi < 0, \\ q_{i+\frac{1}{2}}^{(0)} & \text{if } \xi > 0, \end{cases} \tag{28}$$

which is given by the value at  $\xi/\tau = 0$  of the similarity solution for the above RP, and we call it Godunov state.  $q_{i+\frac{1}{2}}^{(k)} (k = 1, \dots, m - 1)$  are the solutions at  $(\xi, \tau) = (0, +0)$  of the  $k$ th order derivative RP (DRP) locally linearized:

$$\text{PDE} : \partial_\tau + \lambda \left( q_{i+\frac{1}{2}}^{(0)} \right) \partial_\xi = 0, \tag{29}$$

$$\text{IC} : (\xi, 0) = \begin{cases} q_{i+\frac{1}{2}}^{(k)} & \text{if } \xi < 0, \\ q_{i+\frac{1}{2}}^{(k)} & \text{if } \xi > 0, \end{cases} \tag{30}$$

where  $\partial_{\xi}^{(k)} Q$ , and they are given by the values at  $\xi/\tau = 0$  of the similarity solutions for the DRPs.  $\begin{matrix} (k) \\ L \\ R \end{matrix} i+\frac{1}{2}$  and  $\begin{matrix} (k) \\ L \\ R \end{matrix} i-\frac{1}{2}$  in ICs. Eqs. (28) and (30) are given by

$$\begin{cases} \begin{matrix} (k) \\ L \\ R \end{matrix} i+\frac{1}{2} \equiv \lim_{\rightarrow i+\frac{1}{2}-0} \partial^{(k)} p_i( ), \\ \begin{matrix} (k) \\ L \\ R \end{matrix} i-\frac{1}{2} \equiv \lim_{\rightarrow i+\frac{1}{2}+0} \partial^{(k)} p_{i+1}( ). \end{cases} \tag{31}$$

When  $Q$  is scalar, Godunov state  $\begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2}$  is obtained as  $\text{God} \left( \begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2}, \begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2} \right)$  in Osher’s formula [7], where the Godunov flux  $f^{\text{God}}$  is defined as follows [5]:

$$f^{\text{God}} = f \left( \text{God} \left( \begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2}, \begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2} \right) \right) = \begin{cases} \min_{\begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2} \leq \begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2}} f( ) & \text{if } \begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2} \leq \begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2}, \\ \max_{\begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2} \geq \begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2}} f( ) & \text{if } \begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2} \geq \begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2}, \end{cases} \tag{32}$$

and  $\begin{matrix} (k) \\ L \\ R \end{matrix} i+\frac{1}{2} (k = 1, \dots, m - 1)$  are given by

$$\begin{matrix} (k) \\ L \\ R \end{matrix} i+\frac{1}{2} = \begin{cases} \begin{matrix} (k) \\ L \\ R \end{matrix} i+\frac{1}{2} & \text{if } \lambda \left( \begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2} \right) > 0, \\ \begin{matrix} (k) \\ L \\ R \end{matrix} i-\frac{1}{2} & \text{if } \lambda \left( \begin{matrix} (0) \\ L \\ R \end{matrix} i+\frac{1}{2} \right) < 0. \end{cases} \tag{33}$$

### 3.1.4. ADER numerical flux function

The  $m$ th order ADER numerical flux in state-series expansion is obtained by time average

$$f_{i+\frac{1}{2}}^{\text{ADER}m\text{-S}} = \frac{1}{\Delta} \int_0^{\Delta} f \left( Q_{i+\frac{1}{2}}^{\text{GRP}m}(0, \tau) \right) d\tau, \tag{34}$$

which can be carried out numerically by a suitable Gaussian quadrature. When  $f_{i+\frac{1}{2}}^{\text{ADER}m\text{-S}}$  is used with conservative form (5), the  $m$ th order ADER state-series expansion scheme denoted by ADER $m$ -S is obtained.

### 3.2. Method based on direct expansion

Another approach relies on the Taylor-series expansion directly for flux function  $f$  and the conservation law for flux  $f$ , which is derived by multiplying conservation law (1) by  $\lambda(q) = df(q)/dq$ :

$$\partial_t f + \lambda( ) \partial_x f = 0. \tag{35}$$

#### 3.2.1. Reconstruction and GRP

At each time  $t_n$  fluxes should be also reconstructed to piece-wise smooth functions  $g_i(x)$  for cell  $I_i$ . Two ways are possible for reconstruction of  $f$ .

*Rf-1:* Compute cell-averaged values of  $f$  using the interpolation function for  $q, p_i(x)$ ,

$$\bar{f}_i = \frac{1}{\Delta} \int_{i-\frac{1}{2}}^{i+\frac{1}{2}} f(p_i( )) d, \tag{36}$$

which can be approximated by an appropriate Gaussian quadrature, and construct  $g_i(x)$  by ENO/WENO interpolations, and obtain  $\partial^{(k)} g_i( )$ .

*Rf-2:* Define

$$g_i( ) \equiv f(p_i( )). \tag{37}$$

Represent spatial derivatives  $\partial^{(k)} f$  by terms of spatial derivatives  $\partial^{(j)} g_i$  as

$$f = f( ) \tag{38}$$

$$f = \lambda( ) , \tag{39}$$

$$f = \lambda(\cdot) + \lambda(\cdot)(\cdot)^2, \tag{40}$$

⋮

$$f^{(k)} = \psi^{(k)}(\cdot^{(0)}, \cdot^{(1)}, \dots, \cdot^{(k)}), \tag{41}$$

and then, using interpolation functions  $p_i(x)$  instead of  $q(x)$ , obtain  $\partial^{(k)}g_i(\cdot)$ :

$$\partial^{(k)}g_i(\cdot) = \psi^{(k)}(p_i^{(0)}(\cdot), p_i^{(1)}(\cdot), \dots, p_i^{(k)}(\cdot)). \tag{42}$$

In both reconstruction ways, when  $r$  stencils are used, corresponding to the accuracy of  $p_i(x)$ , the following orders of accuracy are estimated for  $g_i(x)$  [11]

$$\text{ENO} : g_i(\cdot) = f(\cdot_{\mathbf{v}n}) + \mathcal{O}(\Delta^{-r}), \tag{43}$$

$$\text{WENO} : g_i(\cdot) = f(\cdot_{\mathbf{v}n}) + \mathcal{O}(\Delta^{-2r-1}), \tag{44}$$

but the spatial accuracy for the  $k$ th order derivative for  $f$  is  $(r - k)$ th order in the both cases:

$$\text{ENO/WENO} : \partial^{(k)}g_i(\cdot) = f^{(k)}(\cdot_{\mathbf{v}n}) + \mathcal{O}(\Delta^{-r-k}). \tag{45}$$

and then the spatial accuracy of the ADER approach is  $r$ th order (see Section 4). As way Rf-2 is effective because it only requires algebraic operations for reconstructed  $q$  and its spatial derivatives, here the second way is used.

Near each cell interface  $i+\frac{1}{2}$  at time  $t_n$ , introduce Eq. (14) and

$$F(\xi, \tau) = f\left(i+\frac{1}{2} + \xi_{\mathbf{v}n} + \tau\right), \tag{46}$$

and consider the GRP for  $F$

$$\text{PDE} : \partial_\tau F + \lambda(Q)\partial_\xi F = 0, \tag{47}$$

$$\text{IC} : F(\xi, 0) = \begin{cases} g_i\left(i+\frac{1}{2} + \xi\right) & \text{if } \xi < 0, \\ g_{i+1}\left(i+\frac{1}{2} + \xi\right) & \text{if } \xi > 0, \end{cases} \tag{48}$$

where the initial data are the reconstructed polynomial functions translated by  $-i+\frac{1}{2}$ .

### 3.2.2. Expansion with Cauchy–Kowalewski procedure

As similar to the case of the state-series expansion, when the solution of the GRP (47) and (48) is differentiable on time and space up to the  $(m - 1)$ th order near  $\xi = 0$  and  $\tau \rightarrow +0$ , it is expressed as a time Taylor-series expansion:

$$F(0, \tau) = F(0, +0) + \sum_{k=1}^{m-1} \frac{\tau^k}{k!} \partial_\tau^{(k)} F(0, +0) + \mathcal{O}(\tau^m). \tag{49}$$

When conservation law (35) is linear with constant  $\lambda$ , all time derivatives of  $f$  can be replaced with space derivatives of  $f$  through Lax–Wendroff [6] procedure:

$$\partial_\tau^{(k)} f = (-\lambda)^k \partial_\xi^{(k)} f, \tag{50}$$

When conservation law (35) is nonlinear, Cauchy–Kowalewski procedure [4] is adopted as follows:

$$\partial_\tau f = -\lambda f, \tag{51}$$

$$\partial_{\tau\tau} f = -\lambda f_\tau - \lambda_\tau f, \tag{52}$$

$$\partial_{\tau\tau\tau} f = -\lambda f_{\tau\tau} - \lambda_{\tau\tau} f - \lambda_\tau f_\tau, \tag{53}$$

$$\partial_{\tau\tau\tau\tau} f = -\lambda f_{\tau\tau\tau} - 2\lambda_{\tau\tau} f_\tau - \lambda_{\tau\tau\tau} f - \lambda_\tau f_{\tau\tau}, \tag{54}$$

$$\begin{cases} f_{\tau}^{\prime} = -\lambda f_{\xi} - \lambda f_{\xi} - \lambda f_{\xi} - \lambda f_{\xi} - \lambda f_{\xi}, \\ f_{\xi}^{\prime} = -\lambda f_{\tau} - 2\lambda f_{\tau} - \lambda f_{\tau} - \lambda f_{\tau} - \lambda f_{\tau}^2, \\ \vdots \end{cases} \tag{55}$$

and time derivatives of  $f$  are expressed with the space derivatives of  $q$  and  $f$ :

$$\partial_{\tau}^{(k)} f = \beta^{(k)}(f^{(0)}, \dots, f^{(k-1)}, f^{(1)}, \dots, f^{(k)}). \tag{56}$$

Thus, the time Taylor-series expansion (49) leads to the following equation:

$$F(0, \tau) = f\left(i+\frac{1}{2}\Delta x, n+0\right) + \sum_{k=1}^{m-1} \frac{\tau^k}{k!} \beta^{(k)}\left(f^{(0)}\left(i+\frac{1}{2}\Delta x, n+0\right), \dots, f^{(k-1)}\left(i+\frac{1}{2}\Delta x, n+0\right), f^{(1)}\left(i+\frac{1}{2}\Delta x, n+0\right), \dots, f^{(k)}\left(i+\frac{1}{2}\Delta x, n+0\right)\right) + O(\tau^m). \tag{57}$$

### 3.2.3. Solution of GRP by use of RP and DRPs

The solution of the GRP for  $F$  at  $\xi = 0$  is approximated with  $m$ th order of accuracy:

$$F_{i+\frac{1}{2}}^{\text{GRP}m}(0, \tau) = f_{i+\frac{1}{2}}^{(0)} + \sum_{k=1}^{m-1} \frac{\tau^k}{k!} \beta^{(k)}\left(f_{i+\frac{1}{2}}^{(0)}, \dots, f_{i+\frac{1}{2}}^{(k-1)}, f_{i+\frac{1}{2}}^{(1)}, \dots, f_{i+\frac{1}{2}}^{(k)}\right). \tag{58}$$

To avoid entropy violation,  $f_{i+\frac{1}{2}}^{(0)}$  should be the flux of a monotone method, and here Godunov flux (32) is adopted.  $f_{i+\frac{1}{2}}^{(k)} (k = 1, \dots, m-1)$  are the solutions at  $(\xi, \tau) = (0, +0)$  of the  $k$ th order DRP on  $f$  locally linearized:

$$\text{PDE: } \partial_{\tau} + \lambda\left(i+\frac{1}{2}\right)\partial_{\xi} = 0, \tag{59}$$

$$\text{IC: } (\xi, 0) = \begin{cases} f_{\text{L } i+\frac{1}{2}}^{(k)} & \text{if } \xi < 0, \\ f_{\text{R } i+\frac{1}{2}}^{(k)} & \text{if } \xi > 0, \end{cases} \tag{60}$$

where  $f_{i+\frac{1}{2}}^{(k)} = \partial_{\xi}^{(k)} F$  and

$$\begin{cases} f_{\text{L } i+\frac{1}{2}}^{(k)} \equiv \lim_{\xi \rightarrow i+\frac{1}{2}-0} \partial_{\xi}^{(k)} g_i(\xi), \\ f_{\text{R } i+\frac{1}{2}}^{(k)} \equiv \lim_{\xi \rightarrow i+\frac{1}{2}+0} \partial_{\xi}^{(k)} g_{i+1}(\xi), \end{cases} \tag{61}$$

and they are given by the values at  $\xi/\tau = 0$  of the similarity solutions for the DRPs

$$f_{i+\frac{1}{2}}^{(k)} = \begin{cases} f_{\text{L } i+\frac{1}{2}}^{(k)} & \text{if } \lambda\left(i+\frac{1}{2}\right) > 0, \\ f_{\text{R } i+\frac{1}{2}}^{(k)} & \text{if } \lambda\left(i+\frac{1}{2}\right) < 0. \end{cases} \tag{62}$$

### 3.2.4. ADER numerical flux function

The  $m$ th order ADER numerical flux in direct expansion is obtained by time average

$$f_{i+\frac{1}{2}}^{\text{ADER}m\text{-D}} = \frac{1}{\Delta t} \int_0^{\Delta t} F_{i+\frac{1}{2}}^{\text{GRP}m}(0, \tau) d\tau = f_{i+\frac{1}{2}}^{(0)} + \sum_{k=1}^{m-1} \frac{(\Delta t)^k}{(k+1)!} \beta^{(k)}\left(f_{i+\frac{1}{2}}^{(0)}, \dots, f_{i+\frac{1}{2}}^{(k-1)}, f_{i+\frac{1}{2}}^{(1)}, \dots, f_{i+\frac{1}{2}}^{(k)}\right). \tag{63}$$

Notice that the numerical quadrature, which is included in the case of state-series expansion, is no more needed. When  $f_{i+\frac{1}{2}}^{\text{ADER}m\text{-D}}$  is used with conservative form (5), the  $m$ th order ADER direct expansion scheme denoted by ADER $m$ -D is obtained.

Regarding the form of  $\beta^{(k)}$  appearing in (63), different expressions than form (56) are possible by eliminating  $f^{(j)}$  or  $f_{\xi}^{(j)}$  by use of (39)–(41): hereafter form (56) is denoted by  $\beta_I^{(k)}$ ,

$$\partial_{\ddagger}^{(k)} f = \beta_I^{(k)}( \text{}^{(0)}, \dots, \text{}^{(k-1)}, f^{(1)}, \dots, f^{(k)} ), \tag{64}$$

a different form is

$$\partial_{\ddagger}^{(k)} f = \beta_{II}^{(k)}( \text{}^{(0)}, f^{(1)}, \dots, f^{(k)} ), \tag{65}$$

and another form is

$$\partial_{\ddagger}^{(k)} f = \beta_{III}^{(k)}( \text{}^{(0)}, \text{}^{(1)}, \dots, \text{}^{(k)} ). \tag{66}$$

The forms of  $\beta_I^{(k)}$ ,  $\beta_{II}^{(k)}$ ,  $\beta_{III}^{(k)}$  and  $\psi^{(k)}$  are summarized in [Appendix](#). As some of  $\beta_{II}^{(k)}$  include  $1/\lambda$ , manipulation is needed at  $\lambda = 0$ .

Thus five ways are possible by combination of function forms for  $\partial_{\ddagger}^{(k)} f$  with reconstructing methods for  $f$ :

A and B:  $\partial_{\ddagger}^{(k)} f = \beta_I^{(k)}( \text{}^{(0)}, \text{}^{(1)}, \dots, \text{}^{(k-1)}, f^{(1)}, \dots, f^{(k)} )$  with methods Rf-1 and Rf-2, respectively.

C and D:  $\partial_{\ddagger}^{(k)} f = \beta_{II}^{(k)}( \text{}^{(0)}, f^{(1)}, \dots, f^{(k)} )$  with methods Rf-1 and Rf-2, respectively.

E:  $\partial_{\ddagger}^{(k)} f = \beta_{III}^{(k)}( \text{}^{(0)}, \text{}^{(1)}, \dots, \text{}^{(k)} )$  without DRPs for  $f$ .

In Ref. [11] the verification for five ways A–E is shown, where the designed order of accuracy is obtained in convergence studies in all ways. Here way B or E is adopted for computational efficiency.

Another version of the ADER direct expansion schemes is suggested by Toro et al. in [21] by rearranging the schemes, (5) and (63), as:

$$f_i^{n+1} = f_i^n - \frac{(\Delta_{\ddagger})_0}{\Delta} (f_{i+\frac{1}{2}}^{(0)} - f_{i-\frac{1}{2}}^{(0)}) - \sum_{k=1}^{m-1} \frac{(\Delta_{\ddagger})_k}{\Delta} (\beta_{i+\frac{1}{2}}^{(k)} - \beta_{i-\frac{1}{2}}^{(k)}) \tag{67}$$

with

$$(\Delta_{\ddagger})_k = \frac{(\Delta_{\ddagger})^{k+1}}{(k+1)!} \tag{68}$$

If the flux is linear:  $f = \lambda q$ ,

$$\beta^{(k)} = (-\lambda)^k \partial^{(k)} f = \lambda \left( (-\lambda)^k \partial^{(k)} \right) \tag{69}$$

holds from the Lax–Wendroff procedure. Then, Eq. (67) can be interpreted as summation of solutions of evolutionary equations for state variable  $u = \partial^{(0)}$  (first line) and its gradients  $\partial^{(k)}$  (second line) by the Godunov first-order upwind method. This leads to the idea to replace the Godunov first-order flux with some high-order flux of total variation diminishing (TVD) schemes, not in the first line of Eq. (67) only, but in all terms in the expansion. Most of the modern TVD fluxes, however, achieve non-oscillatory behavior by imposing a certain monotonicity constraint on extrapolated values  $u_{L_{i+\frac{1}{2}}}$  and  $u_{R_{i+\frac{1}{2}}}$ , or on  $f_{L_{i+\frac{1}{2}}}$  and  $f_{R_{i+\frac{1}{2}}}$ , and use of the constraint prevents the ADER schemes from holding the designed order of accuracy. In [21] the TVD flux of the second-order weighted average flux (WAF) method [16] is recommended to use with the ADER schemes because it is the only high-order TVD flux which needs no constraints on the extrapolated values. ADER $m$ -waf schemes are obtained by applying the WAF flux as the leading flux in the first line of (67) and also the WAF states or WAF fluxes as the solutions of DRPs for  $q$  or  $f$  in the second line with a TVD limiter.

Thus, the advantages of the direct expansion in comparison with the state-series expansion are as follows: (1) in the ADER flux of the direct-expansion forms, the numerical quadrature, which is included in the state-series expansion forms, is not necessary; and (2) it is possible to combine with a suitable high-order TVD flux and obtain higher resolvability. The disadvantage is that the Cauchy–Kowalewski procedure becomes more complicated. However, the disadvantage is not essential, because this procedure can be carried out with the aid of software tools such as MAPLE or Mathematica. Because of the benefits above, solutions based on the direct-expansion are demonstrated in the verification.



#### 4. Accuracy

The accuracy of the method based on direct expansion is presented here. For the ADER approach (5) and (63), the local truncation error is defined by

$$\begin{aligned}
 T(i_{\mathbf{x}} n) &\equiv \frac{1}{\Delta} \{ (i_{\mathbf{x}} n + \Delta) - (i_{\mathbf{x}} n) \} + \frac{1}{\Delta} \{ f_{i+\frac{1}{2}}^{\text{ADER}m\text{-D}} - f_{i-\frac{1}{2}}^{\text{ADER}m\text{-D}} \} \\
 &= \frac{1}{\Delta} \{ (i_{\mathbf{x}} n + \Delta) - (i_{\mathbf{x}} n) \} + \sum_{k=0}^{m-1} \frac{(\Delta)^k}{(k+1)!} \left[ \frac{1}{\Delta} \{ \beta_i^{(k)}(i+\frac{1}{2}) - \beta_i^{(k)}(i-\frac{1}{2}) \} \right],
 \end{aligned}
 \tag{70}$$

where  $\beta_i^{(k)}(\cdot)$  are defined as follows:

$$\beta_i^{(0)}(\cdot) \equiv g_i(\cdot), \tag{71}$$

$$\beta_i^{(k)}(\cdot) \equiv \beta_1^{(k)}(p_i(\cdot), \partial^{(1)} p_i(\cdot), \dots, \partial^{(k-1)} p_i(\cdot), \partial^{(1)} g_i(\cdot), \dots, \partial^{(k)} g_i(\cdot)), \text{ for } k = 1, \dots, m-1. \tag{72}$$

The accuracy of  $p_i(x)$  and  $\partial^{(k)} p_i(\cdot)$  is shown in Eqs. (11)–(13), and the corresponding accuracy of  $g_i(x)$  and  $\partial^{(k)} g_i(\cdot)$  is in Eqs. (43)–(45). At each cell interface  $i+\frac{1}{2}$  the solution of the GRP is used, which is very effective to capture discontinuities, etc., clearly. However, the accuracy is discussed usually on the assumption of a smooth solution. When the reconstruction errors of the left and right states of  $q$  and  $f$  are those shown for  $p_i(x)$  and  $g_i(x)$ , then in the scalar conservation laws the solution of the GRP is within the error range.

In the ENO case, these accuracy evaluation gives for  $k = 0, 1, \dots, r-1$

$$\beta_i^{(k)}(\cdot) = f_i^{(k)}(\cdot) + O(\Delta^{-k}), \tag{73}$$

whatever formulae for  $f_i^{(k)}$  among  $\beta_I^{(k)}$  in (64),  $\beta_{II}^{(k)}$  in (65) and  $\beta_{III}^{(k)}$  in (66) might be used. Therefore the following evaluation on space holds for  $k = 0, 1, \dots, m-1$  with  $r \geq m$ :

$$\frac{1}{\Delta} \{ \beta_i^{(k)}(i+\frac{1}{2}) - \beta_i^{(k)}(i-\frac{1}{2}) \} = (f_i^{(k)})(i_{\mathbf{x}}) + O(\Delta^{-k}). \tag{74}$$

On the other hand, the time Taylor-series expansion brings the following evaluation on time:

$$\frac{1}{\Delta} \{ (i_{\mathbf{x}} n + \Delta) - (i_{\mathbf{x}} n) \} = \sum_{k=0}^{m-1} \frac{(\Delta)^k}{(k+1)!} \partial^{(k+1)}(i_{\mathbf{x}} n) + O(\Delta^m). \tag{75}$$

Thus, with (74) and (75), local truncation error (70) becomes

$$\begin{aligned}
 T(i_{\mathbf{x}} n) &= \sum_{k=0}^{m-1} \frac{(\Delta)^k}{(k+1)!} \left[ \partial^{(k+1)}(i_{\mathbf{x}} n) + (f_i^{(k)})(i_{\mathbf{x}} n) + O(\Delta^{-k}) \right] + O(\Delta^m) \\
 &= \sum_{k=0}^{m-1} O(\Delta^k \Delta^{-k}) + O(\Delta^m),
 \end{aligned}
 \tag{76}$$

since the followings holds from governing Eq. (1),

$$\partial^{(k+1)}(i_{\mathbf{x}} n) + (f_i^{(k)})(i_{\mathbf{x}} n) = 0. \tag{77}$$

If  $\Delta x$  and  $\Delta t$  keep a constant ratio,  $\Delta x/\Delta t < +\infty$ , we have

$$|T(i_{\mathbf{x}} n)| \leq C \Delta^{-r} + C_i \Delta^m. \tag{78}$$

It has been shown here that, by using the ENO reconstruction of  $r = m$ , the ADER scheme (5) with (63), (65) or (66) is of order  $r$  in time and space.

However, summation of the spatial error in (76) might result in a large value of bound  $C_x$  in (78). Possibility to increase the convergence rate is to use the WENO interpolation [9,1] with  $r$  stencils which satisfies  $(2r-1)$ th order of accuracy for  $q$  and  $f$  (Eqs. (12) and (44)) instead of  $r$ th order (Eqs. (11) and (43)); then, as  $\lambda^{(j)}(\cdot)$  ( $j = 0, 1, \dots, k-1$ ) included in  $\beta_i^{(k)}(\cdot)$  are more accurately approximated in the case of WENO, the bound value  $C_x$  can be significantly reduced. In the WENO case, however, notice that the order of accuracy

for ADER schemes is same in the ENO case, because the order of accuracy for the  $k$ th derivatives of  $q$  and  $f$  is  $(r - k)$  in both cases (Eqs. (13) and (45)), and therefore the order of accuracy for  $\beta_i^{(k)}(\cdot)$  remains  $(r - k)$  as in Eq. (73).

## 5. Numerical verification and discussion

Numerical verification has been carried out for conservation laws with flux functions  $f(q) = (1/a)q^a$  ( $a = 1, 2, 3, 4, 5$ ) where  $a = 1$  corresponds to linear flux  $f(q) = q$ ,  $a = 2, 4$  to nonlinear convex fluxes  $f(q) = (1/2)q^2$  and  $f(q) = (1/4)q^4$ , and  $a = 3, 5$  to nonlinear non-convex fluxes  $f(q) = (1/3)q^3$  and  $f(q) = (1/5)q^5$ , and another types of non-convex fluxes,  $f(q) = (1/4)(q^2 - 1)(q^2 - 4)$  and  $f(q) = q^2/(q^2 + a(1 - q)^2)$ .

Here ADER schemes with  $r = m$ , i.e., with  $r$ th order of accuracy in time and space ( $r$  stencils) are verified. For the reconstruction of data, mainly the WENO interpolation is adopted, but there are some occasions where the ENO interpolation works better than the WENO one, for example, formation of complicated wave structure. In notation, ADER $r$ -D and ADER $r$ -S represent the  $r$ th order ADER schemes with the Godunov states/fluxes in direct expansion and state-series expansion, respectively. ADER $r$ -waf represents the schemes using the WAF states/fluxes averaged at half of the time stepping size [17] in the  $r$ th order direct expansion forms, and here the SUPERBEE limiter of Roe [8] is used together. ADER1 (ADER1-D and ADER1-S) and ADER1-waf correspond to the first-order Godunov upwind scheme and the second-order WAF scheme, respectively.

For the purpose of comparison, WENO [9,1] finite-volume schemes with the third-order TVD Runge–Kutta method [10] are adopted. In WENO schemes, use of  $r$  stencils yields  $(2r - 1)$ th order of accuracy in space, and the following time stepping size

$$\Delta t \sim \Delta x^{(2r-1)/3} \quad (79)$$

is used to hold the designed order of accuracy [1]. For  $r \geq 3$  the power of  $\Delta x$  is  $(2r - 1)/3 > 1$ , and therefore time stepping (79) causes very small value of  $\Delta t$  and consumes computing time. Here WENO schemes having  $r$  stencils with  $\Delta t_w$  by Eq. (79) and those with  $\Delta t$  by CFL condition are denoted by WENO $r$ :dtw and WENO $r$ :cfl, respectively.

In all the computation of ADER and WENO:cfl schemes, here the CFL number is taken as 0.8.

### 5.1. Convergence studies: verification on accuracy

For convergence studies, initial value problems defined on  $x \in [-1, 1]$

$$\text{PDE} : \partial_t \phi + \partial_x f(\phi) = 0, \quad f(\phi) = (1/a)\phi^a, \quad (a = 1, 2, 3, 4, 5) \quad (80)$$

$$\text{IC} : \phi(x=0) = 0.25 + 0.5 \sin(\pi x) \quad (81)$$

have been numerically solved on equally spaced grids with periodic boundary conditions; time is evolved until  $t = 1/\pi$  before shock waves are generated.

Figs. 1(a)–(e) show the convergence studies of ADER and WENO schemes for problems with linear flux  $f(q) = q$ , convex fluxes  $f(q) = (1/2)q^2$  and  $f(q) = (1/4)q^4$ , and non-convex fluxes  $f(q) = (1/3)q^3$  and  $f(q) = (1/5)q^5$ , respectively. Here  $L_1$  norm of global errors versus  $\Delta x$  is plotted in log scale, where the slopes indicate the order of accuracy. Figures show that all ADER schemes achieve the designed order of accuracy even if  $\Delta x$  is very small, and ADER5 schemes yield smallest errors among all schemes when  $\Delta x$  is large. Tendency of scheme errors can be summarized as follows: (1) in the linear problem (Fig. 1(a)), values of errors in ADER $r$ -D and ADER $r$ -S are same and very small ( $10^{-13}$  in the fifth order schemes), and somewhat smaller than those of ADER $r$ -waf except for  $r = 2$ ; (2) in the nonlinear problems (Fig. 1(b)–(e)), values of errors in ADER $r$ -D, ADER $r$ -S, and ADER $r$ -waf are almost same and small ( $10^{-10}$ – $10^{-11}$  in the fifth order schemes); (3) in all problems, as expected, errors of ADER1-waf (WAF scheme) are smaller than those of ADER1 (Godunov scheme) for  $r = 1$ .

For comparison, results of WENO $r$  schemes are included: with CFL number 0.8, the convergence rate remains the lower order of accuracy between those of time and space, i.e., third-order, while with time stepping (79) the convergence rate reaches the designed order of accuracy  $(2r - 1)$ , and errors of WENO schemes with

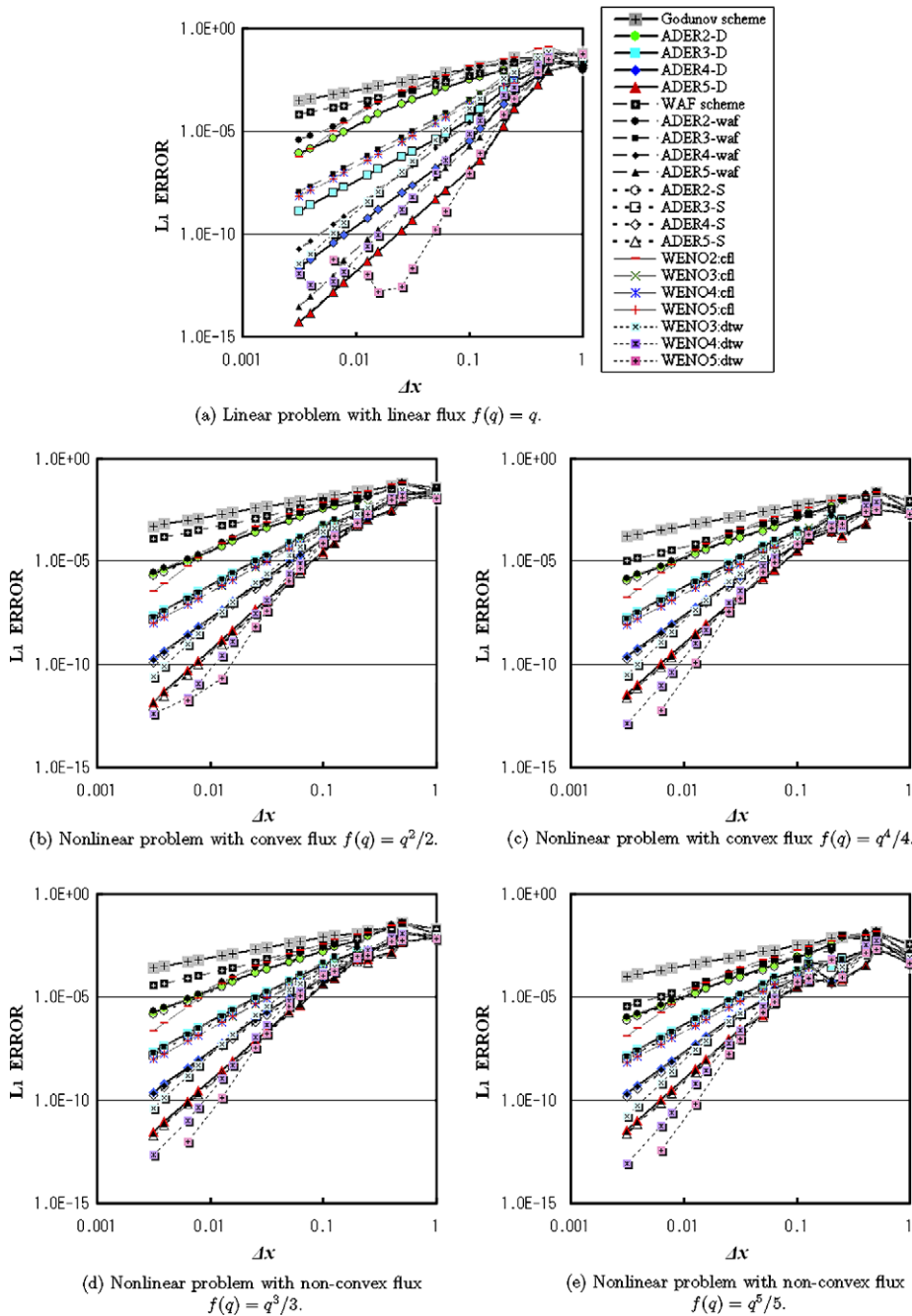
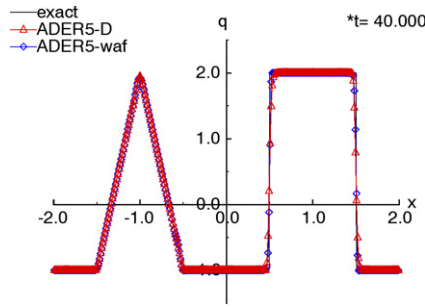


Fig. 1. Convergence study.

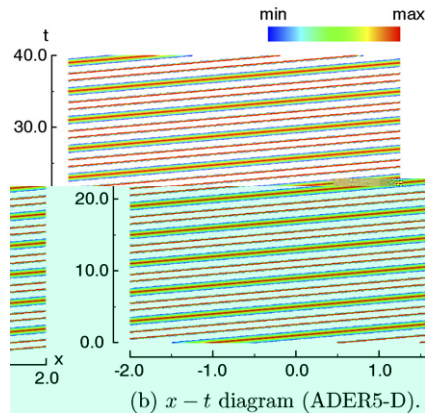
$r = 4, 5$  are smaller than those of ADER5 schemes in some range of  $\Delta x$ . However, in the latter time-stepping case,  $\Delta t$  in highly-accurate WENO schemes should be very small and therefore time-consuming. For example in the case of  $\Delta x = 10^{-2}$ , CFL condition gives  $\Delta t \approx 10^{-2}$ , while WENO time stepping condition (79) gives  $\Delta t \approx 2 \times 10^{-5}$  and  $\Delta t \approx 10^{-6}$  for  $r = 4$  and  $r = 5$ , respectively.

Therefore it is concluded that the advantage of ADER schemes is to achieve the designed order of accuracy up to small  $\Delta x$ , to yield smaller errors in large  $\Delta x$  compared with WENO schemes using the same stencils, and to have computational efficiency with the CFL number close to unity.





(a) Numerical and exact solutions at  $t = 40$  (ADER5-D and ADER5-waf).



(b)  $x - t$  diagram (ADER5-D).

Fig. 3. Wave propagation on linear flux (320 cells).

ADER2-waf–ADER5-waf on 160 cells. ADER1-waf captures discontinuities sharply, but bluntness appears in the solution of ADER2-waf. However, the higher the order of accuracy becomes through ADER3-waf–ADER5-waf, the clearer the discontinuities and apex are. Fig. 2(c) and (d) show the comparison of ADER-D, ADER-waf and WENO schemes with  $r = 4$  on 160 cells and 40 cells, respectively, and (e) and (f) show the comparison with  $r = 5$  on 160 cells and 40 cells, respectively. It is observed that in Figs. 2(c) and (e) on the finer grid, ADER-D, ADER-waf and WENO:dtw generate sharp and clear solutions, but WENO:cfl causes the overshoot and undershoot at the discontinuity when the time accuracy is largely different from the spatial one. In Figs. 2(d) and (f), it may be said that ADER schemes capture the discontinuity and apex better even on the coarse grid. Between ADER-D and ADER-waf, ADER-waf shows clearer capturability but has less robustness in numerical experiments.

Figs. 3(a) and (b) show the linear propagation of waves by ADER5 until time  $t = 40$  in a finer grid of 320 cells. It is observed that the apex of the triangle and the discontinuities of the rectangle are clearly resolved by ADER5-D and ADER5-waf without numerical oscillations, and the  $x - t$  diagram by ADER5-D shows the linearity in wave propagation.

5.2.2. Nonlinear problems with convex fluxes  $f(q) = (1/a)q^a$  ( $a = 2, 4$ )

The nonlinear conservation laws with convex fluxes  $f(q) = (1/a)q^a$  ( $a = 2, 4, \dots$ ) with IC:

$$q_0(x) = \begin{cases} -1 & \text{if } |x| \geq \frac{1}{2}, \\ 2 & \text{if } |x| < \frac{1}{2}, \end{cases} \tag{82}$$

are considered on region  $x \in [-2.5, 1.5]$ . As indicated in Fig. 4, the breakdown of the initial distribution results in a expansion fan including a sonic point from the discontinuity at  $x = -1/2$  with left-state value  $q_L = -1$  and right-state value  $q_R = 2$ , and does in a shock wave from the discontinuity at  $x = 1/2$  with  $q_L = 2$  and  $q_R = -1$  with speed  $S = (2^a - 1)/3a$ .

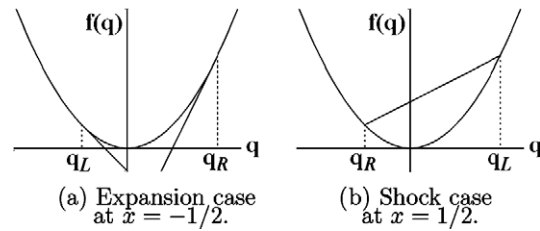


Fig. 4. Case of convex flux  $f(q) = (1/a)q^a$  ( $a$ : even number).

Let  $(x_O, t_O)$  represent the location and time where the head expansion wave overtakes the shock wave,

$$\begin{cases} O = \frac{(3a+2)2^a - 2}{2\{(3a-2)2^a + 2\}}, \\ \tau O = \frac{6a}{(3a-2)2^a + 2}. \end{cases} \tag{83}$$

Then for  $t < t_O$  the exact solution is given by:

$$\left( \begin{array}{c} \tau \\ \zeta \end{array} \right) = \begin{cases} -1 & \text{if } < -\frac{1}{2} - \tau, \\ \left( \frac{+1/2}{\tau} \right)^{\frac{1}{a-1}} & \text{if } -\frac{1}{2} - \tau \leq < -\frac{1}{2} + 2^{a-1} \tau, \\ 2 & \text{if } -\frac{1}{2} + 2^{a-1} \tau \leq \leq \frac{1}{2} + \frac{2^a - 1}{3a} \tau, \\ -1 & \text{if } > \frac{1}{2} + \frac{2^a - 1}{3a} \tau, \end{cases} \tag{84}$$

at  $t = t_O$  the expansion hits the shock at  $x = x_O$ , and for  $t \geq t_O$  when the expansion interacts with the shock, it is expressed by:

$$\left( \begin{array}{c} \tau \\ \zeta \end{array} \right) = \begin{cases} -1 & \text{if } < -\frac{1}{2} - \tau, \\ \left( \frac{+1/2}{\tau} \right)^{\frac{1}{a-1}} & \text{if } -\frac{1}{2} - \tau \leq \leq X_s(\tau), \\ -1 & \text{if } > X_s(\tau), \end{cases} \tag{85}$$

where  $X_s(t)$  is the location of the shock which satisfies the ordinary differential equation:

$$\frac{dX}{d\tau} = \frac{\left( \frac{X+1/2}{\tau} \right)^a - 1}{a \left( \frac{X+1/2}{\tau} + 1 \right)} \tag{86}$$

with the IC of  $X_s(t_O) = x_O$ .

Here as numerical verification, the problems with convex fluxes  $f(q) = (1/2)q^2$  and  $f(q) = (1/4)q^4$  are adopted. In the case of  $f(q) = (1/2)q^2$ , numerical solutions at  $t = 0.0, 0.4, 0.8, 1.6$  on 40 cells are shown in Fig. 5 and those on 320 cells are in Fig. 6, with the exact solution. Figs. 5(a)–(d) show the numerical solutions by ADER1 (Godunov scheme) and ADER1-waf (WAF scheme), those by ADER2-D–ADER5-D, those by ADER2-waf–ADER5-waf, and those by WENO3:dtw–WENO5:dtw, respectively. It is well-known that near the sonic point the change of the upwind direction sometimes results in overestimation of slopes, the so-called glitch phenomena. In the solution by ADER1, the glitch phenomena appear in the expansion fan, the wave front of the expansion fan is not clear for  $t < t_O$ , and the shock wave is not sharp, while in the solution by ADER1-waf the shock and expansion fan appear much clearer. The tendency of ADER1 is observed also in the solution by ADER2-D. However, the higher the order of accuracy becomes through ADER3-D–ADER5-D, the more the solution is improved in both expansion and shock waves. Also so as to ADER2-waf–ADER5-waf, the higher the order is, the better the solution is. Although ADER5-D, ADER5-waf and WENO5:dtw capture the numerical solutions clearly, the ADER5 schemes capture the shock wave slightly sharper than the WENO5 scheme. Figs. 6(a) and (b) show the wave formation by ADER5-D on a finer grid of 320 cells for  $t \leq 1.6$ . It is understood that the numerical solution is in excellent agreement with the exact solution without spurious oscillations, and in the  $x - t$  diagram the interaction between the expansion fan and shock wave is clearly observed.

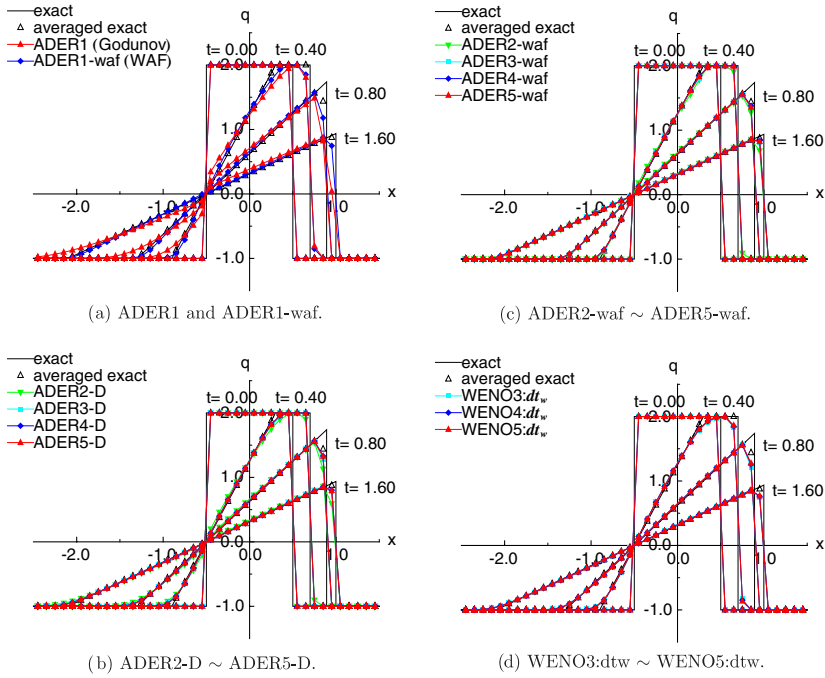


Fig. 5. Numerical and exact solutions on convex flux  $f(q) = (1/2)q^2$  (ADER-D, ADER-waf and WENO; 40 cells).

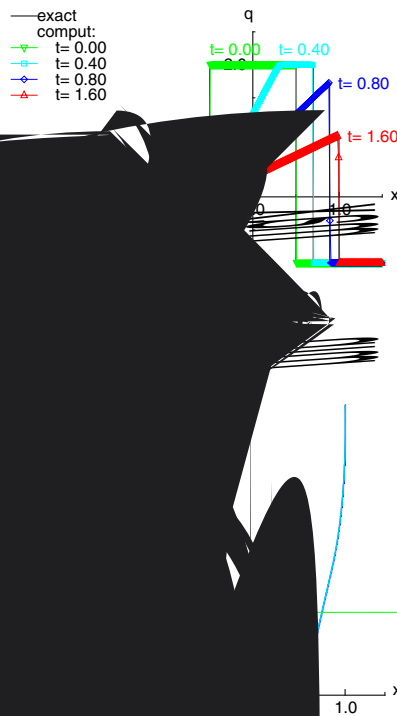


Fig. 6. Wave formation on convex flux  $f(q) = (1/2)q$  (ADER5-D; 320 cells).

In the case of  $f(q) = (1/4)q^4$ , Figs. 7 and 8 corresponds to Figs. 5 and 6, respectively. ADER1 overestimates slopes in the expansion fan, while ADER1-waf underestimates slopes. Otherwise the same tendency has been observed as mentioned in the case with  $f(q) = (1/2)q^2$ : the higher the order of accuracy is in ADER schemes, the more the numerical solution is improved in both expansion and shock waves, and ADER5-D and ADER5-waf capture the shock and expansion waves slightly shaper than WENO5-dtw. The interaction between the expansion fan and shock wave is clearly observed without numerical oscillations.

5.2.3. Nonlinear problems with non-convex fluxes  $f(q) = (1/a)q^a$  ( $a = 3, 5$ )

To investigate the capturability of shock and expansion waves the nonlinear conservation laws with non-convex fluxes  $f(q) = (1/a)q^a$  ( $a = 3, 5, \dots$ ) are investigated furthermore. The Riemann problem with IC:

$$q_0(x) = \begin{cases} q_L = -1.5 & \text{if } x < 0, \\ q_R = 1.1 & \text{if } x > 0, \end{cases} \tag{87}$$

is solved numerically on region  $x \in [-1, 2]$ . In problems with non-convex fluxes, the solutions can exist where from one discontinuity both shock wave and expansion fan are formed. Let  $q^*$  be the root of the equation obtained from  $S = \lambda(q^*)$ , where  $S$  is the speed of the shock with jump from  $q_L$  to  $q^*$ , and  $\lambda(q^*)$  is the characteristic speed at  $q^*$ ,

$$(a - 1)(q^*)^{a-1} - (q^*)^{a-2} q_L - (q^*)^{a-3} \frac{q_L^2}{2} \dots - \frac{q_L^{a-1}}{a-1} = 0. \tag{88}$$

If  $q_R > q^* > 0 > q_L$  or  $q_R < q^* < 0 < q_L$ , both waves are generated. Fig. 9(a) shows this case with IC (87) satisfying the former relation with non-convex flux, and Fig. 9(b) represents the solution with the shock and subsequent expansion waves, respectively. As  $S = (q^*)^{a-1}$ , the exact solution is given by

$$q(x, t) = \begin{cases} -1.5 & \text{if } x < (q^*)^{a-1} t, \\ (x - (q^*)^{a-1} t)^{\frac{1}{a-1}} & \text{if } (q^*)^{a-1} t < x < 1.1^{a-1} t, \\ 1.1 & \text{if } x \geq 1.1^{a-1} t, \end{cases} \tag{89}$$

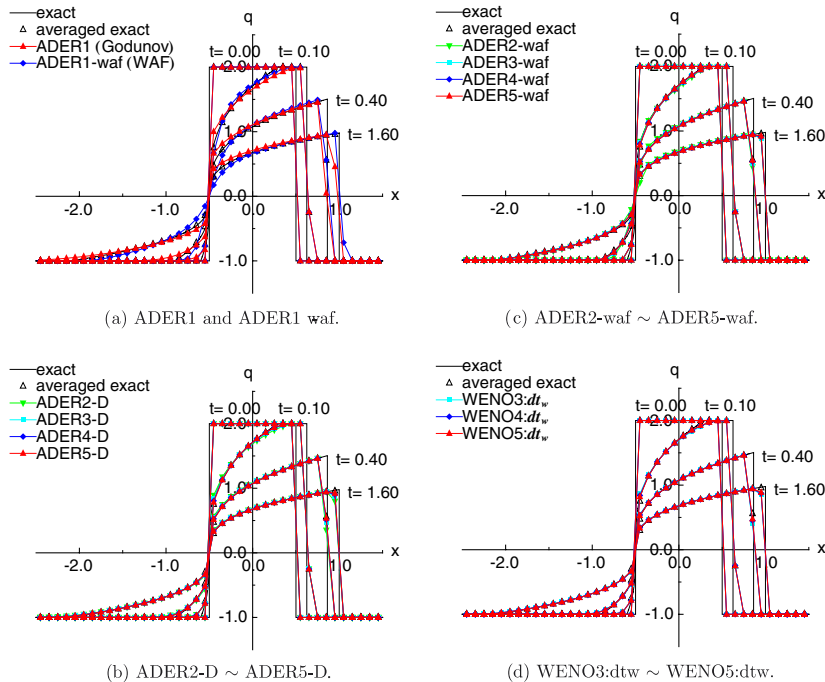
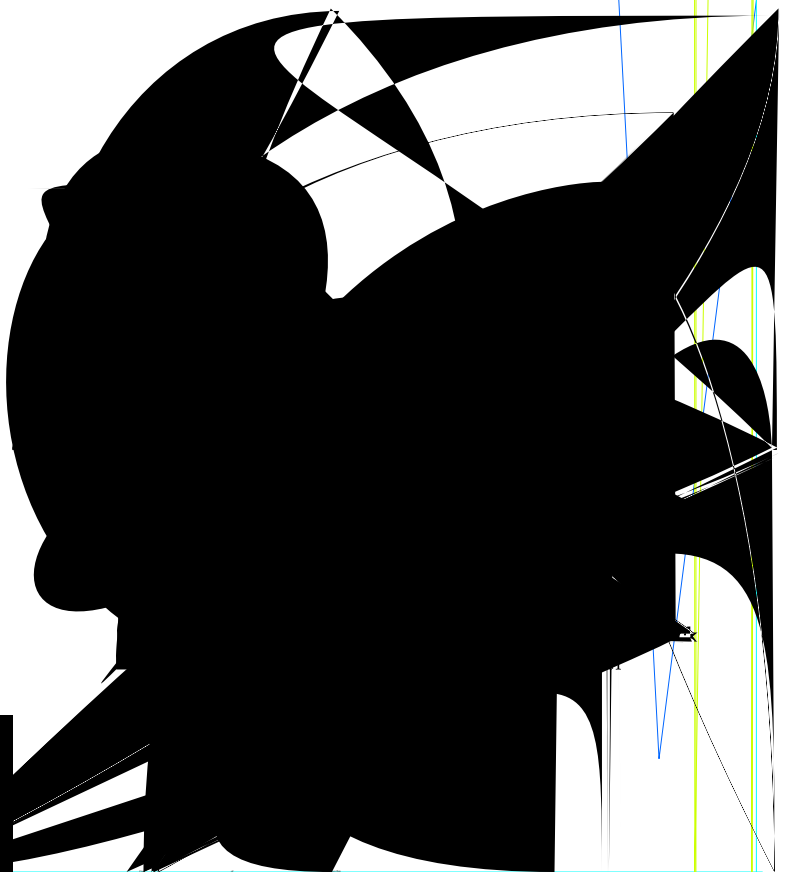
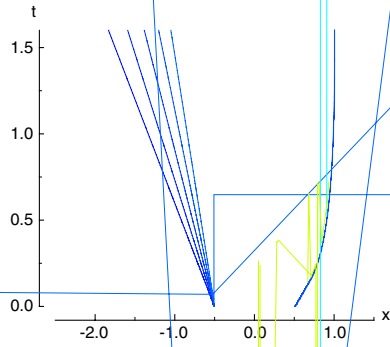
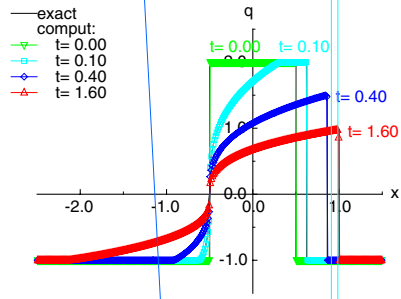
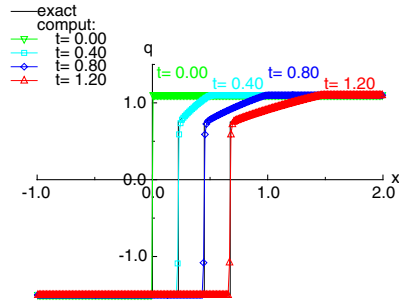


Fig. 7. Numerical and exact solutions on convex flux  $f(q) = (1/4)q^4$  (ADER-D, ADER-waf and WENO; 40 cells).

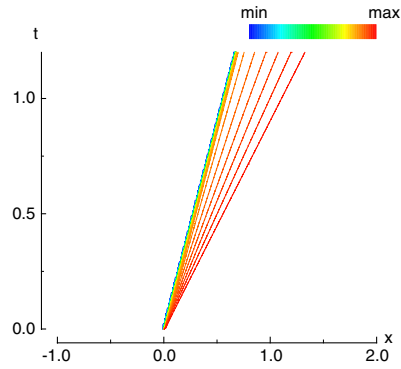








(a) Numerical and exact solutions.



(b)  $x - t$  diagram.

Fig. 11. Wave formation on non-convex flux  $f(q) = (1/3)q^3$  (ADER5-D; 240 cells).

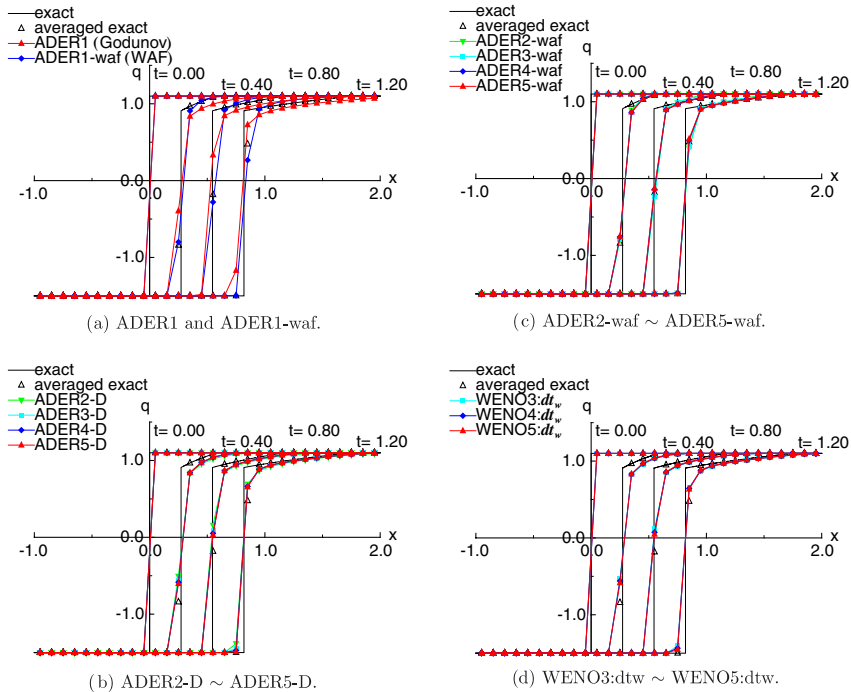
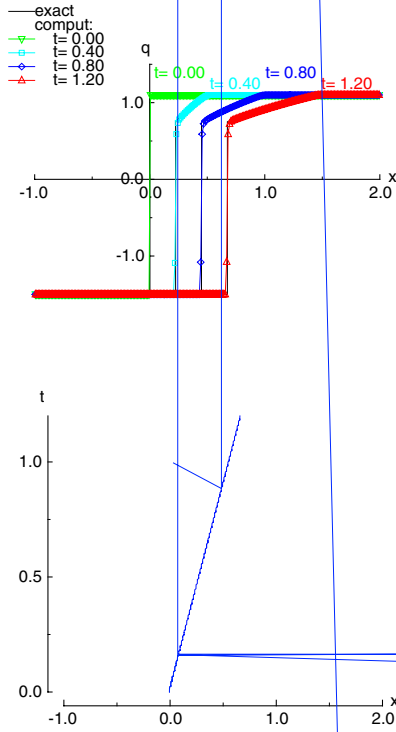


Fig. 12. Numerical and exact solutions on non-convex flux  $f(q) = (1/5)q^5$  (ADER-D, ADER-waf and WENO; 30 cells).



(ADER5-D; 240 cells).

$$q_0(x) = \begin{cases} q_L & \text{if } x < 0, \\ q_R & \text{if } x > 0, \end{cases} \quad (90)$$

is solved [4] numerically on region  $x \in [-1, 1]$ .

(1) Case of  $q_L < -\sqrt{5}$  and  $q_R > \sqrt{5}$

As indicated in Fig. 14(a),  $f(q)$  takes minimal values at  $q_R^* = \sqrt{2/5}$  and  $q_L^* = -\sqrt{2/5}$ , and then  $S = \lambda(q_L^*) = \lambda(q_R^*) = 0$  holds, where the shock speed and the characteristic speed agree with each other at  $q = q_R^*$  and  $q = q_L^*$ . Therefore the stationary shock can exist at  $x = 0$  together with expansion fans in both sides. The exact solution is given by:

$$q(x) = \begin{cases} q_L & \text{if } x \leq \lambda(q_L^*), \\ -h(-q_L^*) & \text{if } \lambda(q_L^*) \leq x < 0, \\ h(q_L^*) & \text{if } 0 < x \leq \lambda(q_R^*), \\ q_R & \text{if } \lambda(q_R^*) \leq x, \end{cases} \quad (91)$$

where  $h(\xi)$  is the solution of  $\xi = f(h(\xi))$  in the convex part  $f'(h(\xi)) > 0$ ;  $q_L^* = \sqrt{5/6}$  is an inflection point). Fig. 14(b) shows the solution of ADER1 and ADER5-D in the case of  $q_L = -3$  and  $q_R = 3$ . It is observed that wave formation of expansion–shock–expansion is clearly captured by ADER5-D.

(2) Case of  $\sqrt{5/6} < q_L < \sqrt{5}$  and  $-\sqrt{5/6} > q_R > -\sqrt{5}$

As indicated in Fig. 14(c), let  $q_L^*$  be the root of equation  $S_L = \lambda(q_L^*)$ , where  $S_L$  is the speed of the shock with jump from  $q_L$  to  $q_L^*$ , and  $q_R^*$  is the corresponding root of  $S_R = \lambda(q_R^*)$ . Then expansion waves are formed near  $x = 0$  and two shock waves propagate to both sides. The exact solution is given by:

$$q(x) = \begin{cases} q_L & \text{if } x < \lambda(q_L^*), \\ h(q_L^*) & \text{if } \lambda(q_L^*) < x < \lambda(q_R^*), \\ q_R & \text{if } \lambda(q_R^*) < x, \end{cases} \quad (92)$$

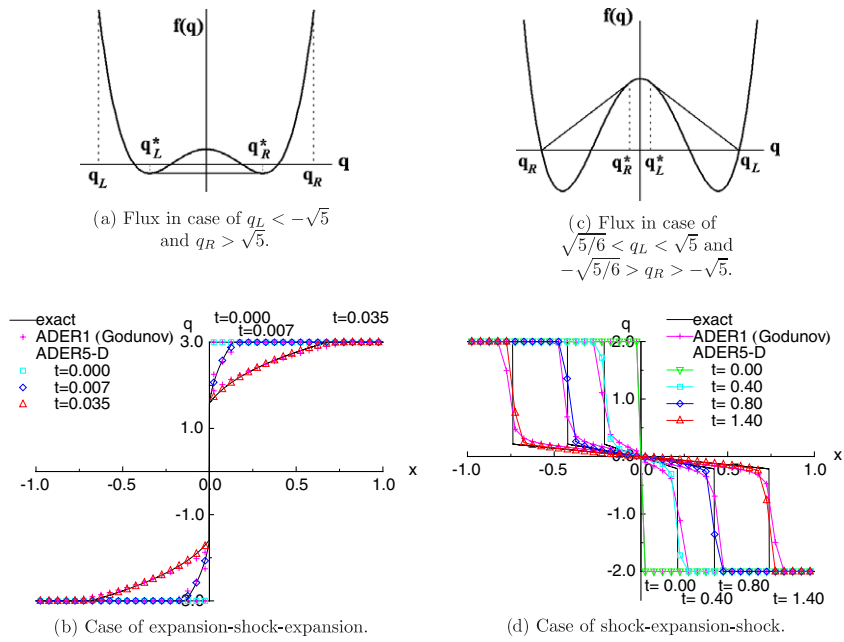


Fig. 14. Wave formation on non-convex flux  $f(q) = (1/4)(q^2 - 1)(q^2 - 4)$  (ADER1 and ADER5-D; 40 cells).

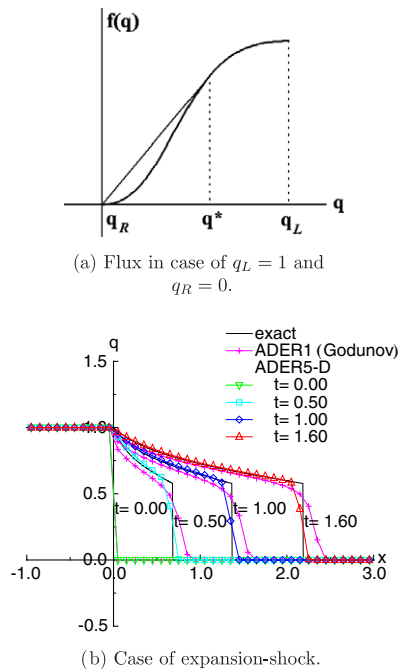


Fig. 15. Wave formation on non-convex flux  $f(q) = q^2 / (q^2 + a(1 - q^2))$  (ADER1 and ADER5-D; 40 cells).

where  $h(\xi)$  is the solution of  $\xi = f'(h(\xi))$  in the concave part  $f(q)$  ( $-q_1 < q < q_l$ ). In this case where numerical solution has complicated distribution with convexity and concavity, the WENO interpolation does not work well, and here the ENO interpolation is adopted in the ADER schemes. Fig. 14(d) shows the solution of ADER1 and ADER5-D in the case of  $q_L = 2$  and  $q_R = -2$ . It is observed that wave formation of shock–expansion–shock is clearly captured by ADER5-D.

5.2.5. Nonlinear problems with non-convex fluxes  $f(q) = q^2/(q^2 + a(1 - q)^2)$

Consider the Buckley–Leverett equations having the flux above, a simple model for two phase fluid flow in a porous media[5]. The Riemann problem with IC:

$$u_0(x) = \begin{cases} u_L = 1 & \text{if } x < 0, \\ u_R = 0 & \text{if } x > 0, \end{cases} \tag{93}$$

is numerically solved on region  $x \in [-1, 3]$ . Let  $q^*$  be the root of equation  $S = \lambda(q^*)$ , then the expansion and subsequent shock waves are formed, as indicated in Fig. 15 (a). The exact solution is given by

$$u(x, t) = \begin{cases} 1 & \text{if } x \leq 0, \\ h(\xi/q) & \text{if } 0 \leq x < \lambda(q^*)t, \\ 0 & \text{if } \lambda(q^*)t < x, \end{cases} \tag{94}$$

where  $h(\xi)$  is the solution of  $\xi = f'(h(\xi))$  in the concave part  $f(q)(q_I < q < 1; q_I$  is the inflection point). Also in this case the ENO interpolation is adopted in the ADER schemes because of the same reason stated in Section 5.2.4. Fig. 15(b) shows the solution of ADER1 and ADER5-D in the case of  $a = 0.5$ . It is observed that formation of expansion and subsequent shock waves is clearly captured by ADER5-D.

In this section solutions by ADER-D and ADER-waf have been displayed, and in numerical experiments the solutions by ADER-S have been almost same as those of ADER-D.

6. Conclusions

On the ADER approach, the state-series expansion forms and the direct expansion forms have been presented in the viewpoint of the numerical procedure and the accuracy, and the advantages and disadvantages of the latter forms are discussed in comparison with the former forms. As ADER direct expansion schemes, ADER-D (standard ones with Godunov states/fluxes) and ADER-waf (ones with WAF states/fluxes) are adopted for verification in comparison with ADER-S (state-series expansion schemes) and WENO. The verification has been carried out mainly for the ADER direct expansion schemes up to the fifth order of accuracy on the scalar conservation laws with a linear flux, nonlinear convex fluxes, and several types of nonlinear non-convex fluxes. Convergence studies with continuous initial distribution of states have shown that all the ADER schemes achieve the designed order of accuracy up to small cell sizes, yield small errors even in large cell sizes, and have computational efficiency with keeping the CFL number close to unity. In verification by discontinuous initial conditions, as the order of ADER schemes is made higher, the long-time propagation of apices and discontinuities is clearly captured in the linear problem, and the waves of shocks and expansions are correctly formed and interacted in the nonlinear problems, corresponding to each flux. It is remarkable that ADER-waf schemes have shown sharper resolvability than ADER-D and ADER-S schemes, but have less robustness. Therefore it is concluded that the ADER direct expansion schemes work well for both the linear problems and the nonlinear problems with convex and non-convex fluxes.

Acknowledgements

This is to develop the joint research with Prof. E.F. Toro. The author would like to thank to Prof. E.F. Toro and Dr. V.A. Titarev (Trento University, Italy) for their fruitful discussion, to Dr. B. Mueller and staffs of TDB (Uppsala University, Sweden) for their hospitality, and also Dr. H. Aiso (JAXA, Japan) for his advice on non-convex problems.

Appendix. Forms of  $\alpha^{(k)}, \beta_I^{(k)}, \beta_{II}^{(k)}, \beta_{III}^{(k)}$  and  $\psi^{(k)}$

(1)  $\alpha^{(k)}(u^{(0)}, u^{(1)}, \dots, u^{(k)})$  for  $k = 1, 2, 3, 4$

$$\alpha^{(1)} = -\lambda(u^{(1)}),$$

$$\alpha^{(2)} = \lambda^2(u^{(2)}) + 2\lambda\lambda'(u^{(1)})^2,$$

$$\alpha^{(3)} = -\lambda^3 \lambda^{(3)} - 9\lambda^2 \lambda^{(1)} \lambda^{(2)} - 6\lambda(\lambda^{(1)})^2 (\lambda^{(1)})^3 - 3\lambda^2 \lambda^{(1)} (\lambda^{(1)})^3,$$

$$\alpha^{(4)} = \lambda^4 \lambda^{(4)} + 16\lambda^3 \lambda^{(1)} \lambda^{(3)} + 12\lambda^3 \lambda^{(2)} (\lambda^{(2)})^2 + 72\lambda^2 (\lambda^{(1)})^2 (\lambda^{(1)})^2 (\lambda^{(2)}) + 24\lambda(\lambda^{(1)})^3 (\lambda^{(1)})^4$$

$$+ 24\lambda^3 \lambda^{(1)} (\lambda^{(1)})^2 (\lambda^{(2)}) + 36\lambda^2 \lambda^{(1)} \lambda^{(1)} (\lambda^{(1)})^4 + 4\lambda^3 \lambda^{(1)} (\lambda^{(1)})^4.$$

(2)  $\beta_I^{(k)}(\lambda^{(0)}, \lambda^{(1)}, \dots, \lambda^{(k-1)}, f^{(1)}, \dots, f^{(k)})$  for  $k = 1, 2, 3, 4$

$$\beta_I^{(1)} = -\lambda f^{(1)},$$

$$\beta_I^{(2)} = \lambda^2 f^{(2)} + 2\lambda \lambda^{(1)} f^{(1)},$$

$$\beta_I^{(3)} = -\lambda^3 f^{(3)} - 6\lambda^2 \lambda^{(1)} f^{(2)} - 3\lambda^2 \lambda^{(2)} f^{(1)} - 6\lambda(\lambda^{(1)})^2 (\lambda^{(1)})^2 f^{(1)} - 3\lambda^2 \lambda^{(1)} (\lambda^{(1)})^2 f^{(1)},$$

$$\beta_I^{(4)} = \lambda^4 f^{(4)} + 12\lambda^3 \lambda^{(1)} f^{(3)} + 12\lambda^3 \lambda^{(2)} f^{(2)} + 4\lambda^3 \lambda^{(3)} f^{(1)} + 36\lambda^2 (\lambda^{(1)})^2 (\lambda^{(1)})^2 f^{(2)}$$

$$+ 36\lambda^2 (\lambda^{(1)})^2 (\lambda^{(1)})^2 f^{(1)} + 24\lambda(\lambda^{(1)})^3 (\lambda^{(1)})^3 f^{(1)} + 12\lambda^3 \lambda^{(1)} (\lambda^{(1)})^2 f^{(2)}$$

$$+ 12\lambda^3 \lambda^{(1)} (\lambda^{(1)})^2 f^{(1)} + 36\lambda^2 \lambda^{(1)} \lambda^{(1)} (\lambda^{(1)})^3 f^{(1)} + 4\lambda^3 \lambda^{(1)} (\lambda^{(1)})^3 f^{(1)}.$$

(3)  $\beta_{II}^{(k)}(\lambda^{(0)}, f^{(1)}, \dots, f^{(k)})$  for  $k = 1, 2, 3, 4$

$$\beta_{II}^{(1)} = -\lambda f^{(1)},$$

$$\beta_{II}^{(2)} = \lambda^2 f^{(2)} + 2\lambda (f^{(1)})^2,$$

$$\beta_{II}^{(3)} = -\lambda^3 f^{(3)} - 9\lambda \lambda^{(1)} f^{(1)} f^{(2)} - 3\frac{1}{\lambda} (\lambda^{(1)})^2 (f^{(1)})^3 - 3\lambda (f^{(1)})^3,$$

$$\beta_{II}^{(4)} = \lambda^4 f^{(4)} + 16\lambda^2 \lambda^{(1)} f^{(1)} f^{(3)} + 12\lambda^2 \lambda^{(2)} (f^{(2)})^2 + 48(\lambda^{(1)})^2 (f^{(1)})^2 f^{(2)}$$

$$+ 24\lambda \lambda^{(1)} (f^{(1)})^2 f^{(2)} + 20\frac{1}{\lambda} \lambda^{(1)} \lambda^{(1)} (f^{(1)})^4 + 4\lambda (f^{(1)})^4.$$

(4)  $\beta_{III}^{(k)}(\lambda^{(0)}, \lambda^{(1)}, \dots, \lambda^{(k)})$  for  $k = 1, 2, 3, 4$

$$\beta_{III}^{(1)} = -\lambda^2 \lambda^{(1)},$$

$$\beta_{III}^{(2)} = \lambda^3 \lambda^{(2)} + 3\lambda^2 \lambda^{(1)} (\lambda^{(1)})^2,$$

$$\beta_{III}^{(3)} = -\lambda^4 \lambda^{(3)} - 12\lambda^3 \lambda^{(1)} \lambda^{(2)} - 12\lambda^2 (\lambda^{(1)})^2 (\lambda^{(1)})^3 - 4\lambda^3 \lambda^{(1)} (\lambda^{(1)})^3,$$

$$\beta_{III}^{(4)} = \lambda^5 \lambda^{(4)} + 20\lambda^4 \lambda^{(1)} \lambda^{(3)} + 15\lambda^4 \lambda^{(2)} (\lambda^{(2)})^2 + 120\lambda^3 (\lambda^{(1)})^2 (\lambda^{(1)})^2 (\lambda^{(2)})$$

$$+ 60\lambda^2 (\lambda^{(1)})^3 (\lambda^{(1)})^4 + 30\lambda^4 \lambda^{(1)} (\lambda^{(1)})^2 (\lambda^{(2)}) + 60\lambda^3 \lambda^{(1)} \lambda^{(1)} (\lambda^{(1)})^4 + 5\lambda^4 \lambda^{(1)} (\lambda^{(1)})^4.$$

(5)  $\psi^{(k)}(\lambda^{(0)}, \lambda^{(1)}, \dots, \lambda^{(k)})$  for  $k = 1, 2, 3, 4$

$$\psi^{(1)} = \lambda \lambda^{(1)},$$

$$\psi^{(2)} = \lambda \lambda^{(2)} + \lambda (\lambda^{(1)})^2,$$

$$\psi^{(3)} = \lambda \lambda^{(3)} + 3\lambda \lambda^{(1)} \lambda^{(2)} + \lambda (\lambda^{(1)})^3,$$

$$\psi^{(4)} = \lambda \lambda^{(4)} + 4\lambda \lambda^{(1)} \lambda^{(3)} + 3\lambda (\lambda^{(1)})^2 (\lambda^{(2)}) + 6\lambda (\lambda^{(1)})^2 (\lambda^{(2)}) + \lambda (\lambda^{(1)})^4.$$

**References**

[1] D.S. Balsara, C.W. Shu, Monotonicity preserving weighted essentially non-oscillatory schemes with increasingly high order of accuracy, *J. Comput. Phys.* 160 (2000) 405–452.  
 [2] S.K. Godunov, Finite difference methods for the computation of discontinuous solutions of the equations of fluid dynamics, *Mat. Sb.* 47 (1959) 271–306.  
 [3] A. Harten, S. Osher, Uniformly high-order accurate non-oscillatory schemes I, *SIAM J. Numer. Anal.* 24 (2) (1987) 279–309.  
 [4] A. Harten, B. Engquist, S. Osher, S.R. Chakravarty, Uniformly high order accurate essentially non-oscillatory schemes III, *J. Comput. Phys.* 71 (1987) 231–303.  
 [5] R.J. LeVeque, *Numerical Methods for Conservation Laws*, second ed., Birkhaeuser Verlag, 1992.

- [6] P. Lax, B. Wendroff, Systems of conservation laws, *Comm. Pure Appl. Math.* 13 (1960) 217–237.
- [7] S. Osher, Riemann Solvers, The entropy condition and difference approximations, *SIAM J. Numer. Anal.* 21 (2) (1984) 217–235.
- [8] P.L. Roe, Some contributions to the modelling of discontinuous flows, *Lect. Appl. Math.* 22 (1985).
- [9] C.W. Shu, Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws, Technical Report, NASA/CR-97-206253, ICASE Report No. 97-65, 1997.
- [10] C.W. Shu, Total-variation-diminishing time discretizations, *SIAM J. Scienti. Statisti. Comput.* 9 (1988) 1073–1084.
- [11] Y. Takakura, E.F. Toro, Arbitrarily accurate non-oscillatory schemes for a nonlinear scalar conservation law, *CFD J.* 11 (1) (2002) 6–17.
- [12] Y. Takakura, E.F. Toro, Arbitrarily accurate non-oscillatory schemes for nonlinear scalar conservation laws with source terms, AIAA paper 2002-2736, 2002.
- [13] Y. Takakura, E.F. Toro, Arbitrarily accurate non-oscillatory schemes for nonlinear scalar conservation laws with source terms II, in: S. Armfield, P. Morgan, K. Srinivas (Eds.), *Computational Fluid Dynamics 2002, Proceedings of 2nd International Conference on CFD*, Springer-Verlag Pub., 2003, pp. 247–252.
- [14] V.A. Titarev, E.F. Toro, High order ADER schemes for scalar advection-reaction-diffusion equations, *CFD J.* 12 (1) (2003) 1–6.
- [15] V.A. Titarev, E.F. Toro, ADER schemes for three-dimensional non-linear hyperbolic systems, *J. Comput. Phys.* 204 (2005) 715–736.
- [16] E.F. Toro, A weighted average flux method for hyperbolic conservation laws, *Proc. Roy. Soc. London A423* (1989) 401–418.
- [17] E.F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics*, second ed., Springer-Verlag, 1999.
- [18] E.F. Toro, R.C. Millington, L.A.M. Nejad, Towards very high order Godunov schemes, in: E.F. Toro (Ed.), *Godunov Methods. Theory and Applications*, Kluwer/Plenum Academic Publishers, 2001, pp. 907–940.
- [19] E.F. Toro, V.A. Titarev, (2001) Very high order Godunov-type schemes for nonlinear scalar conservation laws, in: *ECCOMAS Computational Fluid Dynamics Conference 2001*, Swansea.
- [20] E.F. Toro, V.A. Titarev, Solution of the generalised riemann problem for advection-reaction equations, *Proc. Roy. Soc. London* 458 (2018) (2002) 271–281.
- [21] E.F. Toro, V.A. Titarev, TVD Fluxes for the high-order ADER schemes. Preprint NI03011-NPA, Isaac Newton Institute for Mathematical Sciences, University of Cambridge, 2003.
- [22] E.F. Toro, V.A. Titarev, ADER schemes for scalar non-linear hyperbolic conservation laws with source terms in three-space dimensions, *J. Comput. Phys.* 202 (2005) 196–215.