

## ON EXACT SOLUTION ENCLOSURE ON ENSEMBLE OF NUMERICAL SIMULATIONS

A.K. ALEXEEV<sup>1</sup>, A.E. BONDAREV<sup>2</sup>

<sup>1</sup> Moscow Institute of Physics and Technology  
Moscow, Russia

e-mail: [aleksey.k.alekseev@gmail.com](mailto:aleksey.k.alekseev@gmail.com)

<sup>2</sup> Keldysh Institute of Applied Mathematics Russian Academy of Sciences  
Moscow, Russia

e-mail: [bond@keldysh.ru](mailto:bond@keldysh.ru)

**Summary.** The estimation of a vicinity of the approximate solution that contains the exact one (*exact solution enclosure*) may be performed using an ensemble of numerical solutions if the information on their error ranging is available *a priori*. A posteriori analysis of distances between numerical solutions enables error ranging by magnitudes, if the ensemble of numerical solutions separates into clusters of “accurate” and “inaccurate” solutions. For nonlinear problems this enclosure may serve as the computational proof of the exact solution existence. The impact of metric selection on the solution enclosure is observed. The numerical tests for the supersonic flows, governed by two dimensional Euler equations, demonstrate the exact solution enclosure using the set of solvers that have different orders of accuracy.

### 1 INTRODUCTION

We consider some additional opportunities for analysis of CFD results that may be provided by the abundant set of numerical methods with wide range of orders of approximation, which is available at present.

Usually, the order of approximation of the finite-difference/finite volume scheme is related with the truncation error order. The truncation error  $\delta u$  is obtained via Taylor series decomposition of the discrete operator  $A_h u_h = f_h$ , which approximates the system of PDE, formally noted herein as  $Au = f$ . The truncation error dependence on the spatial step  $h$  is usually presented as  $\delta u = O(h^n)$ , where the order  $n$  is equal to the minor order of series terms.

The approximation error  $\Delta u = u_h - u$  is caused by the truncation error and is of the real practical interest. It may be described by the tangent linear equation  $A\Delta u = \delta u$  having the formal solution  $\Delta u = A^{-1}\delta u$ .

For linear problems, the approximation error  $\|\Delta u\| = O(h^n)$  tends to zero as  $h$  decreases with the same order  $n$  (Lax theorem, [1]) if the discrete operator is well-posed (the inverse operator is uniformly bounded,  $\|A_h^{-1}\| < C$ ).

For the case of nonlinear equations with discontinuities [2,3,4,5], the error order is essentially local and varies significantly depending on the type of flow structure elements. In

**2010 Mathematics Subject Classification:** 35Q93, 35R30, 94A15.

**Key words and Phrases:** truncation error, discretization error, ensemble of numerical solutions, distance matrix, Euler equations, exact solution enclosure

this event, the observed order of convergence may be not equal to the nominal order of the approximation error even in the asymptotic range.

Two single-grid based approaches to the discretization error estimation are of interest herein.

**A priori error estimation** is the common approach to the error analysis and may be expressed in the form  $\|\Delta u\| < C \cdot h^n$ , which contains unknown constants independent on current numerical solution. A priori error estimation justifies the common practice to stop the mesh refining when the dependence of numerical solution on the step size becomes unobservable. This technique may be used for the proof of the exact solution existence [6] for linear problems.

**A posteriori error estimation** [7,8,9] has the form  $\|\Delta u\| \leq C_h e_h$ , where  $C_h$  is the computable stability constant, which depends on the numerical solution, and  $e_h$  is the computable indicator of the truncation error. At present, the most successes in this direction are achieved for elliptic equations and finite element methods starting from the work by Babushka [7]. In most of practical applications the stability constant is not estimated, while the error indicator is used for the mesh adaptation.

The feasibility for rigorous estimations of the exact solution without mesh refinement is the significant merit of this approach. This is another way if compare with the standard mesh refinement approach, the Richardson extrapolation [12,13] and a multigrid approach, presented, for example, by [14].

However, *a posteriori* error estimation may provide more information regarding the exact solution, for example, [10,11]. *A posteriori* check that can be applied to a numerical solution of Navier-Stokes equations to guarantee the existence for the sufficiently smooth solution of the exact problem is considered in [10]. The paper [11] demonstrated for nonlinear elliptic equations that the estimation of the stability constant (inverse operator norm  $\|A_h^{-1}\|$ ) and the residual may be used for the determination of the vicinity of the numerical solution, which contains the exact solution. The results [11] are interpreted as the proof of the existence of the exact nonlinear solution nearby the approximate solution. This information may be of use due to problems with the existence for the compressible multidimensional Euler equations [15]. According [15] the standard weak solution may not exist for the compressible multidimensional Euler equations. The use of the measure-valued solutions (in Young measures), considered in [15], may cause unacceptable computational burden. So, the *a posteriori* proof of exact solution existence (even local) in the vicinity of the ensemble of numerical solutions may be of practical interest.

We consider a single-grid analysis of another type, if compare with [10,11] herein.

The truncation error  $\delta u$  may be computed by the action of the high order scheme stencil on the precomputed flowfield [16, 17], by the action of the differential operator on the interpolation of the numerical solution [18] or via the differential approximation [19,20]. The practical application of the truncation error  $\delta u$  implies the calculation of the discretization (global) error  $\Delta u = A^{-1} \delta u$ . The surveys of the global error calculation methods may be found in [21, 22]. In the simplest option, the estimation of this error may be performed using defect correction [16] or nearby equation methods [22,23]. In defect correction frame, the truncation error  $\delta u$  is used as the source term inserted in the discrete algorithm in order to correct the

solution. However, the total subtraction of the error implies the elimination of the scheme viscosity that may cause oscillations in the vicinity of discontinuities or activation of some addition dissipation sources, which engenders their own error. Also, the estimation of the error may be performed via the linearized problem [24], complex differentiation [25] or by adjoint equations [17,18,20,26,27]. Usually, adjoint equations are applied to estimation of the uncertainty of certain valuable functional (drag, lift etc.). Nevertheless, the approach described in [20] enables to estimate the norm of the solution error. Unfortunately, it implies the solution of the number of adjoint problems that is proportional to the number of grid nodes that causes the extremely high computational burden.

The unknown components of truncation error causes the general disadvantage of above discussed residual-based methods for the error estimation. The differential approximation based methods using Taylor series [20] do not account for senior terms of expansion. The postprocessor based methods do not account for the higher scheme truncation errors [17] or the interpolation errors [18].

Herein, the analysis is conducted in the space of numerical solutions, so, the truncation error is accounted implicitly and completely. The feasibility to find the vicinity of the numerical solution that contains an exact solution using the ensemble of calculations performed by the solvers of different approximation order is addressed. In contrast to above mentioned norm oriented approaches, the current analysis is addressed to the ensemble of distances (distance matrix) in different metrics. The Multidimensional Scaling (MDS) [28] concerns similar problems, however, we consider the events when the vector length is much greater the number of vectors, so MDS cannot be applied.

## 2 EXACT SOLUTION ENCLOSURE VIA THE SET OF NUMERICAL CALCULATIONS WITH RANGED ERRORS

The approximation error is considered herein as the distance between the exact and approximate solutions. Let us consider the ensemble of numerical solutions obtained using finite difference (finite volume) schemes of different order on the same grid. Let the relation of the approximation error of these schemes to be *a priori* known.

We note the numerical solution as the vector  $u^{(i)} \in R^N$  ( $i$  is the scheme number,  $N$  is the number of grid points), values of unknown exact solution at nodes of this grid (further denoted as exact solution) as  $\tilde{u} \in R^N$  and use some metrics  $d(u, v)$  in the space of solutions. The unknown deviation of exact solution values at grid points  $\tilde{u} \in R^N$  from computed solution is estimated using  $d(u^{(k)}, \tilde{u}) = \delta_{0,k}$  (for example,  $d(u^{(k)}, \tilde{u}) = \|u^{(k)} - \tilde{u}\|_{L_2}$ ). The numerical solutions  $u^{(k)}$  are located at surfaces of nested concentric hyperspheres with the centre at  $\tilde{u}$  and radii  $\delta_{0,k}$ .

The following theorem may be stated for two numerical solutions  $u^{(1)}$  and  $u^{(2)}$  having *a priori* known errors relation  $\delta_{0,1} \geq 2 \cdot \delta_{0,2}$ .

**Theorem 1.** *Let the distance  $\delta_{1,2}$  between two numerical solutions  $u^{(1)} \in R^N$  and  $u^{(2)} \in R^N$  be known from computations and there is available a priori information*

$$\delta_{0,1} \geq 2 \cdot \delta_{0,2} \tag{1}$$

then the exact solution is located within the hypersphere of radius  $\delta_{1,2}$  with the centre at  $u^{(2)}$ :

$$\delta(u^{(2)}, \tilde{u}) \leq \delta_{1,2} \quad (2)$$

**Proof.** The analysis is founded on the triangle inequality [29]:  $\delta_{ij} \leq \delta_{ik} + \delta_{kj}, i \neq j \neq k$ . For our problem, (points  $u^{(1)}, u^{(2)}, \tilde{u}$  and distances  $\delta_{0,1}, \delta_{1,2}, \delta_{0,2}$  between them) it has the form  $\delta_{0,1} \leq \delta_{1,2} + \delta_{0,2}$ , which may be transformed to  $\delta_{0,1} - \delta_{0,2} \leq \delta_{1,2}$ . By accounting (1) as  $\delta_{0,1} - \delta_{0,2} \geq \delta_{0,2}$  one obtains  $\delta_{0,2} \leq \delta_{0,1} - \delta_{0,2} \leq \delta_{1,2}$  and, finally, the desirable expression  $\delta_{0,2} \leq \delta_{1,2}$ .

This theorem may be easily stated in  $L_2$  norm, statements in other norms are not straightforward, however, so the general metric based approach is of great advantage. Herein, we usually consider the metrics determined by  $L_2$  and  $L_1$  norm, the metrics of Mahalonobis form [30] is used also in several tests.

### 3 THE ANALYSIS OF THE ERROR RELATIONS FOR AN ENSEMBLE OF CALCULATIONS

The evident weakness of the *Theorem 1* is the assumption of the existence of solutions with *a priori* ranged error. Despite the widespread opinion that the schemes of higher order are more accurate, it should be checked numerically. Herein, we consider some options for the check of error rating. The collection of distances between solutions  $\delta_{i,j}$  enables a detection of the close and distant solutions. For example, if  $\delta_{0,1} \gg \delta_{0,i}$ , the set  $\delta_{i,j}$  is split into a cluster of inaccurate solutions with great values  $\delta_{1,j}$  and the cluster of more accurate solutions  $\delta_{i,j} (i \neq 1)$ . It is caused by the asymptotics  $\delta_{1,j} / \delta_{0,1} \rightarrow 1$  and  $\delta_{i,j} (i \neq 1) / \delta_{0,1} \sim (\delta_{0,i} + \delta_{0,j}) / \delta_{0,1} \rightarrow 0$  at  $\delta_{0,i} / \delta_{0,1} \rightarrow 0$ .

The separation of the collection of distances between solutions into clusters is the evidence of the existence of solutions with significantly different errors that may be considered as a proof of error ranging. The quantitative criterion based on dimension of clusters and the distance between them is of interest. Let us compare the set of distances  $\delta_{1,j}$  and  $\delta_{k,j}$ , where  $u^{(1)}$  is maximally incorrect solution and  $u^{(k)}$  is the selected accurate solution (the localization of exact solution is performed in its vicinity),  $\delta_{i,\max}$  is the maximum error in the subset of accurate solutions.

We state the following heuristical criterion (**Conjecture 1**):

*The exact solution may be enclosed if the distance between clusters is greater the size of the cluster of accurate solutions. Then the condition (1) is valid and the exact solution is located within a hypersphere of radius  $\delta_{i,k}$  with the center at  $u^{(i)}$ :  $\delta_{0,i} \leq \delta_{i,k}$ , where  $u^{(i)}$  belongs to the cluster of more accurate solutions and  $u^{(k)}$  is maximally inaccurate solution.*

This conjecture is based on the assumptions that the dimension of the accurate cluster to be  $r_{i,\max} + r_k$ , and a cluster of inaccurate solutions belongs the interval  $(\delta_{0,1} - \delta_{i,\max}, \delta_{0,1} + \delta_{i,\max})$ ,

the relation of accurate cluster dimension and the distance between clusters has an appearance  $\delta_{0,1} - 2\delta_{i,\max} - \delta_{i,k} > \delta_{i,\max} + \delta_{i,k}$ . This leads to the relation  $\delta_{0,1} > 2\delta_{0,k}$ , that corresponds condition (1).

This criterion may be rigorous only in the limit of the infinite set of solutions obtained by independent methods. Nevertheless, most numerical tests for two dimensional supersonic inviscid flows confirm the applicability of this heuristic criterion. The violation of the enclosure condition  $\|\tilde{u} - u^{(i)}\|_{L_2} \leq \|du_{i,k}\|_{L_2}$  above 15% was not observed.

#### 4 THE SELECTION OF METRICS

The vector of solution for CFD problems contains elements having different physical meanings, such as density, velocity components, energy, etc. Herein, we consider two dimensional Euler equations with four component  $u^{(i)} = \{\rho^{(i)}, U^{(i)}, V^{(i)}, E^{(i)}\}$ . We consider the metrics engendered by the simplest  $L_1$  and  $L_2$  norms.

The norm

$$\|u^{(i)} - u^{(k)}\|_{L_2} = \|\{(\rho^{(i)} - \rho^{(k)}), (U^{(i)} - U^{(k)}), (V^{(i)} - V^{(k)}), (e^{(i)} - e^{(k)})\}\|_{L_2} \quad (3)$$

enables to calculate the distance between solutions. In parallel to Expression (3), the distance between solutions was calculated using the expression

$$\|\{(\rho^{(i)} - \rho^{(k)})/\|\rho^{(i)}\|, (U^{(i)} - U^{(k)})/\|U^{(i)}\|, (V^{(i)} - V^{(k)})/\|V^{(i)}\|, (e^{(i)} - e^{(k)})/\|e^{(i)}\|\}\|_{L_2}, \quad (4)$$

which imitates a relative error.

It should be noted that expression (4) corresponds not to the norm but to the distance  $(\Delta u^{(i)}, M \Delta u^{(i)}) = (M_{j,k} \Delta u_j^{(i)} \Delta u_k^{(i)})^{1/2}$ . This distance is determined by a metric tensor with the matrix  $M_{j,k}$  of the diagonal form that describes some ellipsoid. With account of the presentation  $M = A^* A$  (valid for a metric tensor as the symmetric positively defined matrix, a Mahalanobis distance metric [30]) one can state  $(\Delta u^{(i)}, M \Delta u^{(i)}) = (\Delta u^{(i)}, A^* A \Delta u^{(i)})^{1/2} = (A \Delta u^{(i)}, A \Delta u^{(i)})^{1/2} = (\Delta z^{(i)}, \Delta z^{(i)})^{1/2}$ . So, we can enclosure the solution in the transformed space  $Au^{(i)}$ , where the error may have the form of hypersphere.

The search for more complicated metrics, having some physical meaning and illustrative capabilities, is of interest also. The suitable metrics should not compare different variables, i.e. the matrix should have block-diagonal form with blocks related to different variables. It would be useful, if small flowfield deformations (shift by single step of grid, for example) corresponds small distances. So, the comparison of variables at adjacent grid nodes may be of use, i.e. blocks may have a band form.

#### 5 NUMERICAL TESTS

The tests of the exact solution enclosure are presented below for flows governed by two dimensional unsteady Euler equations.

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho U^k)}{\partial x^k} = 0; \quad (5)$$

$$\frac{\partial(\rho U^i)}{\partial t} + \frac{\partial(\rho U^k U^i + P \delta_{ik})}{\partial x^k} = 0; \quad (6)$$

$$\frac{\partial(\rho E)}{\partial t} + \frac{\partial(\rho U^k h_0)}{\partial x^k} = 0; \quad (7)$$

The elementary structures such as the single oblique shock wave, the interaction of shock waves of I and VI kinds according Edney classification [31] were used as the test problems. All tests concern the steady state solutions. The analytical solution may be easily constructed for these problems. The values of analytical solution at grid points is considered herein as “exact” solution. The flowfield contains undisturbed domains (nominal order of error is expected), shock waves (error order about  $n = 1$  [5]), contact discontinuity line (error order about  $n = 1/2$ , [4]). In result, one may hope to obtain the nontrivial error composed of components with different orders of accuracy. The estimation of this error and the capture of exact solution in certain hypersphere around a numerical solution are the main purposes of the paper.

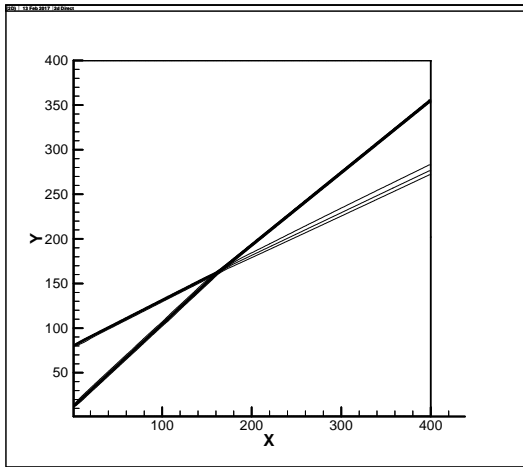


Fig. 1. Edney VI density isolines.

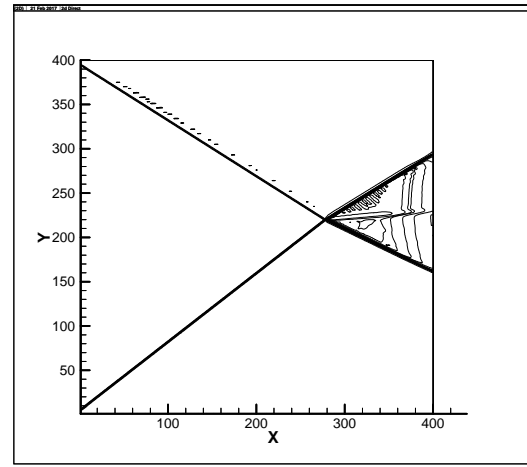


Fig. 2. Edney I density isolines.

The computations were performed for  $C_p/C_v = 1.4$ , Mach number range  $M = 3 \div 5$  and flow deflection angles range  $\alpha = 10 - 30^\circ$ . All tests contain discontinuities in the flow field. Fig. 1 presents the density distribution for Edney VI flow structure ( $M = 4$ , two consequent flow deflection angles  $\alpha_1 = 10^\circ$ ,  $\alpha_2 = 15^\circ$ ). The flow is determined by the merging shock waves, the contact line and the expansion fan. Fig. 2 presents the density isolines for Edney I flow structure ( $M = 3$  and flow deflection angles  $\alpha_1 = 20^\circ$  and  $\alpha_2 = 15^\circ$ ). The crossing shock waves and contact discontinuity line, engendered at the shocks crossing point, are the main elements of this flow structure.

The paper presents the analysis of the ensemble of computations performed by methods listed below.

The first order scheme by Courant Isaacson Rees [32] designated as *S1* was used in the variant described by [33].

The second order scheme based on the MUSCL method [34] and using algorithm by [35] at cell boundaries is noted as *S2*.

Second order TVD scheme of relaxation type by [36], noted as *S2TVD*.

Third order modified Chakravarthy-Osher scheme [37, 38] marked as *S3*.

Fourth order scheme by [39] marked as *S4*.

Computations were performed on uniform grids containing  $100 \times 100$ ,  $200 \times 200$  or  $400 \times 400$  nodes. The vector of solution of Eq. (5-7) contains four components  $u^{(i)} = \{\rho^{(i)}, U^{(i)}, V^{(i)}, E^{(i)}\}$  having different physical meanings and different magnitudes. For example, for Edney VI flow (Fig.2) the norms of component are  $\|\rho^{(i)}\|_{L_2} \approx 2.5$ ,  $\|U^{(i)}\|_{L_2} \approx 0.87$ ,  $\|V^{(i)}\|_{L_2} \approx 0.24$ ,  $\|e^{(i)}\|_{L_2} \approx 0.17$ , herein

$$\|\Delta u\|_{L_2} = \left( \frac{1}{N} \left( \sum \Delta \rho_i^2 + \sum \Delta U_i^2 + \dots \right) \right)^{1/2}, \quad (8)$$

$$\|\Delta u\|_{L_1} = \frac{1}{N} \left( \sum |\Delta \rho_i| + \sum |\Delta U_i| + \dots \right). \quad (9)$$

The data under the consideration are extremely bulky, so for convenience of visualization, in Fig. 3-9 the norm of error is marked up along both axes, despite the data are one dimensional.

For the level of error less 0.1 the solutions visually can not be distinguished. The only visually distinguishable solution corresponds the scheme *S1* (level of error about 0.2) and is specified by high smearing of shock waves. The level of error about 0.1 ( $400 \times 400$ ) may be related with the shift of incident shock location by a single node.

It should be noted that methods *S1, S2, S3, S4* (1,2,3 and 4 nominal truncation orders) demonstrated the order of convergence a bit below  $n=1/2$  in norm  $L_2$ . In norm  $L_1$  the same computations demonstrated the order of convergence a bit higher  $n=1/2$ . The method *S2TVD* (nominal order 2) is the only exception with the order about  $n \sim 3/4$ .

In numerical tests, we first check *Conjecture 1* and, second, verify the enclosure. We consider the enclosure to be successful, if the error estimate  $\|u^{(i)} - u^{(k)}\|_{L_2}$  is greater the true error  $\|u^{(k)} - \tilde{u}\|_{L_2}$ , obtained in comparison with the analytical solution  $\tilde{u}$ .

The comparison with the analytical solution permits to conclude that using the scheme *S1* (as ‘‘inaccurate’’) and schemes *S2, S3, S4* (as ‘‘accurate’’) enables to find the vicinity of numerical solution that contains exact solution.

Second order *S2TVD* scheme [36] from standpoint of error norm is close to first order scheme *S1* for  $100 \times 100$  grid and to high order schemes for grid  $400 \times 400$ . When the clusters are detected, it also enables the enclosure of solutions generated by *S2, S3, S4*. If the clusters are not available, the exact solution is not enclosed. The calculations on the grid  $100 \times 100$  demonstrated the formation of clusters with “inaccurate” scheme *S2TVD* and successful enclosure of the exact solution. However, the scheme *S2TVD* on the grid  $400 \times 400$  does not form clusters. Paradoxically, the reason for this failure is the relatively rapid convergence of *S2TVD* in comparison with schemes *S2a, S2c, S3, S4*. In result, the scheme *S2TVD* on the grid  $400 \times 400$  transfers from “inaccurate” to “accurate” schemes approaching in error to *S2, S3, S4*.

The comparison of schemes *S2, S3, S4* ( $\|u^{(2)} - u^{(4)}\|_{L_2}$ ,  $\|u^{(3)} - u^{(4)}\|_{L_2}$ ,  $\|u^{(3)} - u^{(2)}\|_{L_2}$ ) does not enable to enclose the exact solution. Similarly, the enclosure of exact solution by pair *S2TVD, S1* fails. These schemes have the errors which are close in magnitude and splitting into clusters is not observed.

If the *Conjecture 1* is not satisfied (there are no clusters, or distance between them is less the dimension of the cluster of “accurate” solutions) the enclosure of true solution fails. However, the exact error is about two or three maximum distances between numerical solutions.

The numerical tests for the single oblique shock demonstrate the feasibility for the exact solution enclosure at the significant distance between clusters, if splitting occurs. However, the set of distances between solutions separates into clusters in about half of tests, usually for more fine meshes.

For Edney-VI shock interaction (Fig. 1), the set of distances between solutions also splits into clusters in about half of tests without dependence on the grid size. There is the enclosure of exact solution according expression (3), for the distance between clusters, which approximately equals the dimension of the cluster.

Fig. 3 (with “inaccurate” scheme *S1*,  $L_2$ , grid  $400 \times 400$ ) demonstrates the collection of distances between numerical solutions  $\|du_{i,k}\|_{L_2} = \|u^{(i)} - u^{(k)}\|_{L_2}$  to break into two clusters, one of them is related with the “inaccurate” scheme. It enables to enclose an exact solution. Fig. 4 presents results of  $L_2$  error enclosure based on clusters presented in Fig. 3. The norm  $L_1$  presents the same results as  $L_2$  for this test, Fig. 5.

For Edney-I shock interaction (Fig. 2), the set of distances between solutions splits by clusters in about half of tests, usually for more fine meshes. For the distance between clusters, which approximately equals the dimension of the cluster, the enclosure of exact solution for metrics engendered by  $L_2, L_1$  fails, see Figs 6,7.



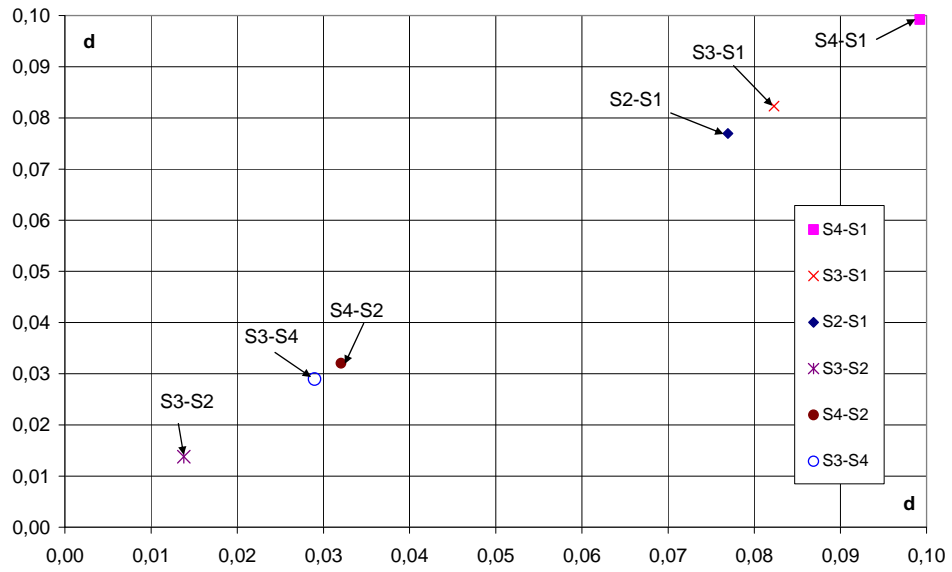


Fig. 3. Clusters related with “inaccurate” scheme S1 for Edney-VI in  $L_2$ , grid  $400 \times 400$ .

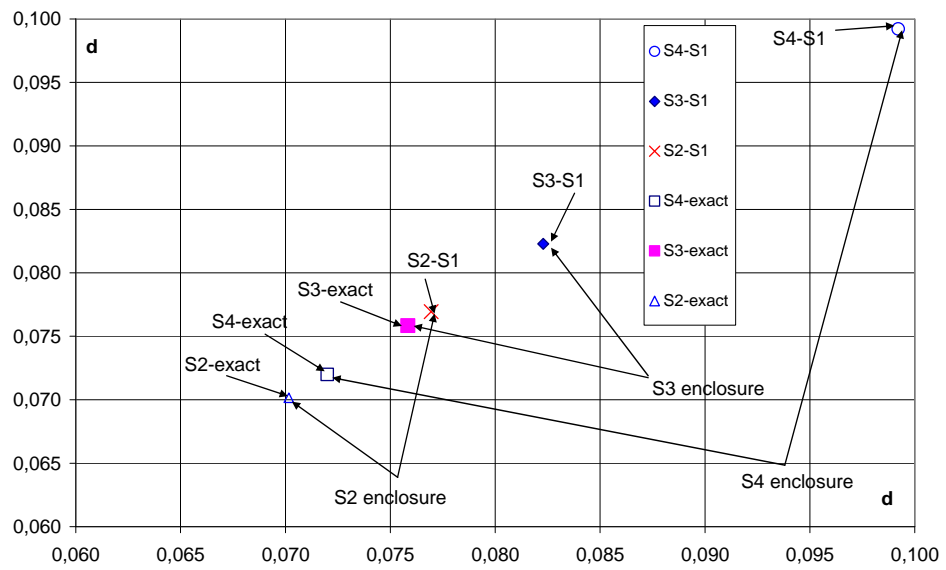


Fig. 4. Exact solution enclosure (Edney-VI) in vicinity of S2, S3, S4 using S1,  $L_2$ , grid  $400 \times 400$ .

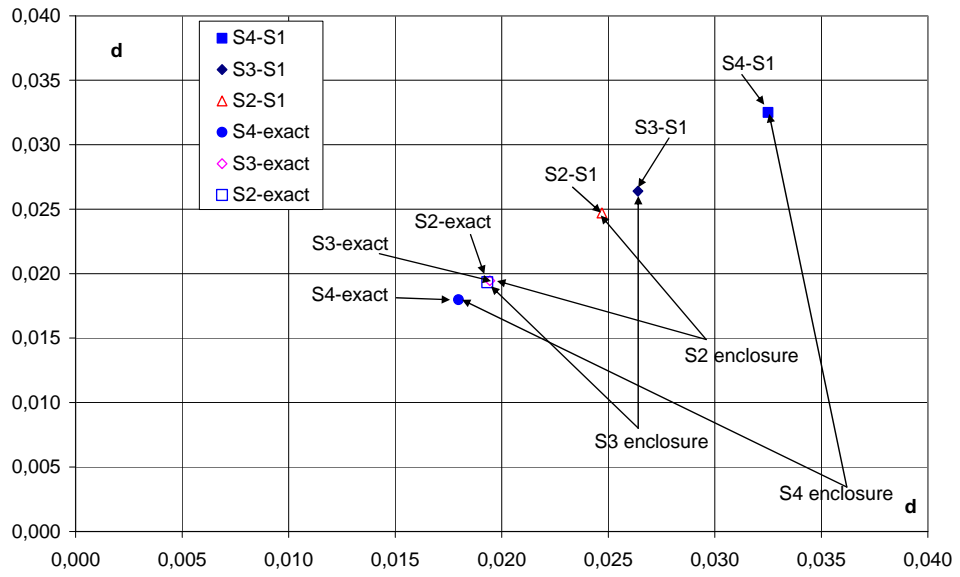


Fig. 5. Exact solution enclosure (Edney-VI) in vicinity of S2, S3, S4 using  $L_1$  norm, grid  $400 \times 400$ .

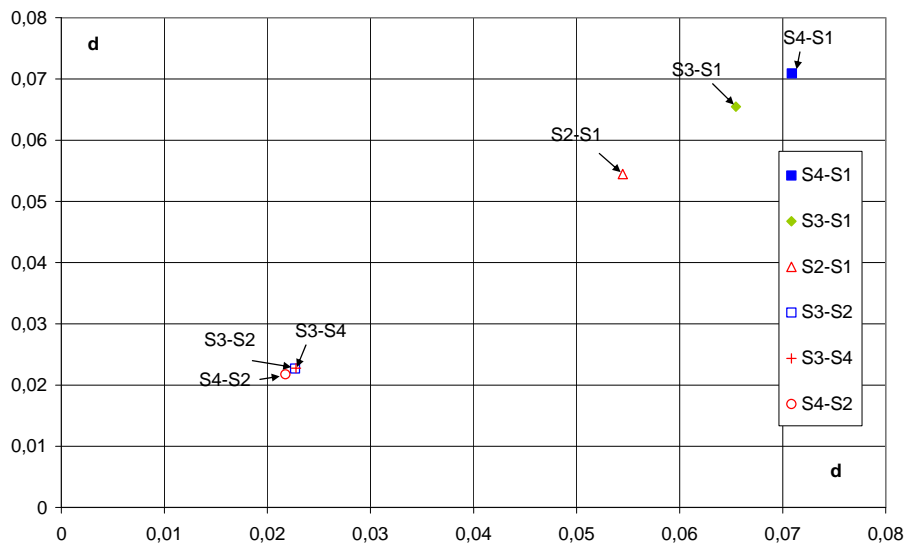


Fig. 6. Clusters related with S1 for Edney-I,  $L_2$ , grid  $400 \times 400$ .

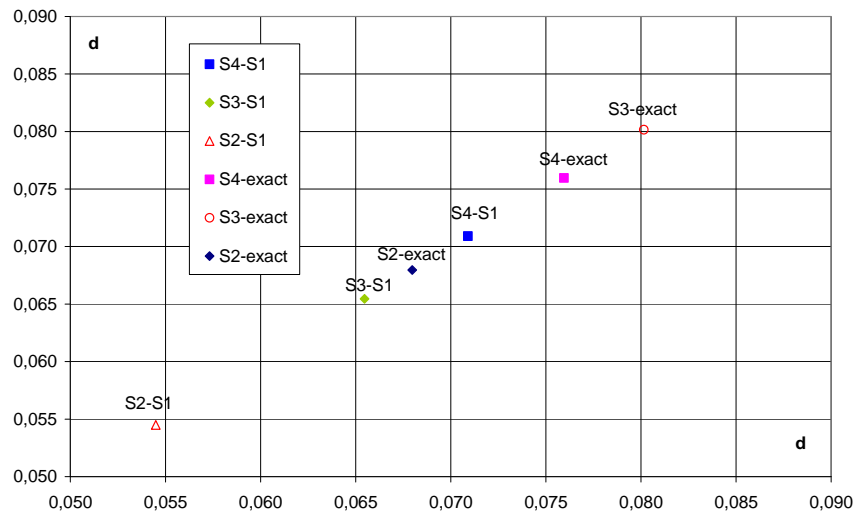


Fig. 7. Exact solution enclosure (Edney-I) in vicinity of S2, S3, S4 using S1,  $L_2$ , grid  $400 \times 400$ .

The results are obtained for  $M = 3$  and flow deflection angles  $\alpha_1 = 20^\circ$  and  $\alpha_2 = 15^\circ$ . The magnitude of capture condition  $\|\tilde{u} - u^{(2)}\|_{L_2} \leq \|du_{1,2}\|_{L_2}$  violation is about 10%. This demonstrates the heuristical, approximate nature of *Conjecture 1*. However, the violation of the enclosure condition is moderate.

The metrics (4) related with relative error enables an more reliable enclosure, Figs. 8,9.

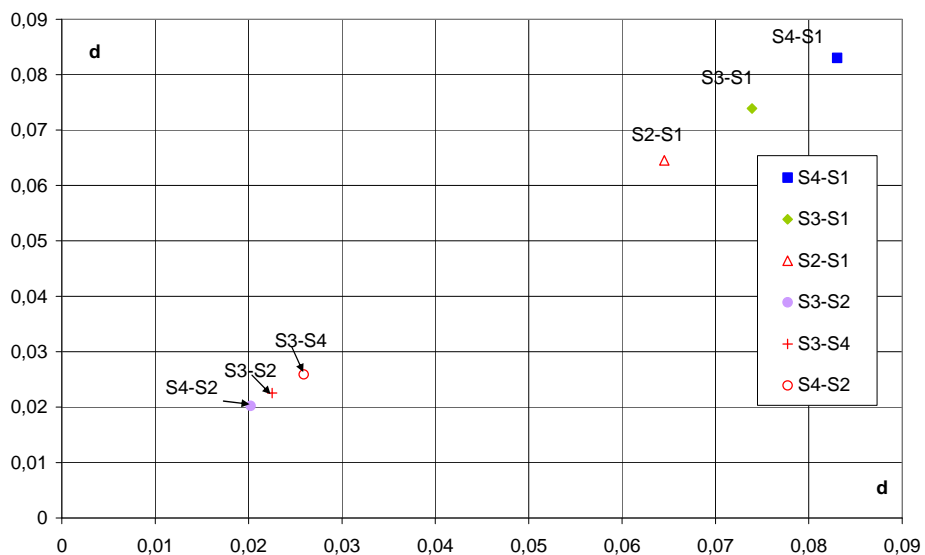


Fig. 8. Clusters related with S1 for Edney-I, grid  $400 \times 400$ .

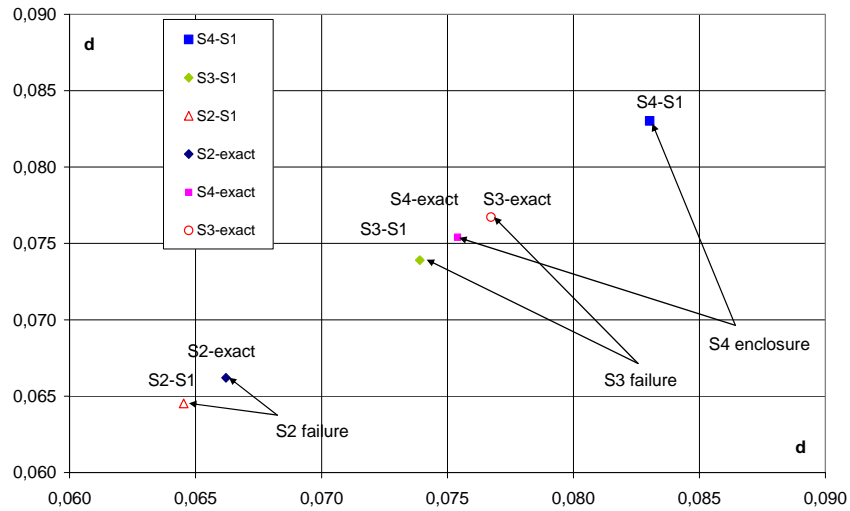


Fig. 9. Exact solution enclosure (Edney-I) in vicinity of S2, S3, S4 using S1, grid 400×400.

Figs. 6-9 demonstrate the visible dependence of the enclosure success on the metrics that is used for the distances calculation.

Thus, in order to enclosure the exact solutions, one should have a priori information (*Theorem 1*) or an ensemble of minimum three solutions with distances split into two clusters. The distance between clusters should be greater the dimension of cluster of more accurate solutions (*Conjecture 1*).

## 6 DISCUSSION

The relation of errors obtained in the paper is not necessarily attributed to properties of the considered schemes. In the strict sense, it may be caused by the imperfections of numerical realization by authors. So, authors do not pretend on the definitive assessment of considered methods. We only can compare solvers (algorithms realizations) from the viewpoint of numerical results.

The standard check of the grid convergence is based on heuristic rule by C. Runge [9]. From this standpoint, if the difference of two approximate solutions on coarse grid  $T_h$  with step  $h$  and on the fine grid  $T_{h,ref}$  with step  $h_{ref}$  is small, then  $u_{h,ref}$  and  $u_h$  are **probably** close to exact solution. From practical needs, one should desire the estimate of the form  $\|u_h - \tilde{u}\| \leq \delta$  with computable  $\delta$ . The Richardson method [12,13,22] is close to this ideal. It enables to determine the refined solution and the error estimate if the single error order exists in total flowfield. The set of solutions  $u_k^{(1)} = \tilde{u}_k + C_k h_1^n$ ,  $u_k^{(2)} = \tilde{u}_k + C_k h_2^n \dots$  computed on different meshes is used. Unfortunately, in most CFD problems the error order on different flow structures varies that hampers or prohibits the application of the Richardson method.

The present paper concerns a single-grid alternative to the Richardson method and Runge rule. The set of solutions is collected at the same mesh using different methods. Calculations may be terminated if a preassigned error level  $\delta(u_h, \tilde{u}) \leq \delta$  ( $\|u_h - \tilde{u}\| \leq \delta$ ) is achieved.

The existence of “accurate” and “inaccurate” schemes is one of the main postulates of computational mathematics, although having an asymptotic sense. The above results demonstrate the feasibility to distinguish “accurate” and “inaccurate” schemes in the sense of error norm ranging. For the events presented on Figs. 3, 5 the distribution of distances between solutions  $\|du_{i,j}\|_{L_2}$  shows the presence of two clusters corresponding “accurate” and “inaccurate” schemes. This engenders the hope to enclose the exact solution only from observable  $\|du_{i,j}\|_{L_2}$  (without a priori information on errors ranging), that is confirmed by Figs. 4,6,7. If clusters cannot be detected, the exact solution enclosure fails.

The feasibility to estimate the distance from the exact solution to numerical one  $\delta(u_h, \tilde{u}) \leq \delta$  seems attractive, however, it is difficult to define at what magnitude of  $\delta$  two solutions can be considered as coinciding (the calculations may be stopped) or describing different flows.

The estimation of uncertainty of certain valuable functionals (drag, lift etc.) is of interest in most of applications. This estimation may be performed using adjoint equations [17,18,20,26,27], nevertheless, it does not describe the flowfield in unique way.

The ensemble based method operates the total error including flowfield error, initial and boundary condition error and round-off errors. It may be used as a postprocessor similar to Richardson extrapolation [12,13,22]. However, it does not need the mesh refinement and may be used away the asymptotic range.

The dependence on the set of numerical methods and analyzed solution is the drawback of ensemble based method. The same set of methods may provide segregation by clusters on one flow pattern and may not provide on another. So, this approach cannot replace the standard accuracy control method (mesh refining) and is assigned to supplement it by non-expensive algorithm.

If there is no breaking into clusters, the distance between numerical solutions and analytical ones is 2-3 times greater the maximum distance between solutions that provide some opportunity for the rough estimation of numerical error.

So, it is feasible to obtain the information on the numerical error and exact solution location using the collection of solution obtained on the same grid by different solvers without mesh refinement.

## 7 CONCLUSIONS

The information on distances between numerical solutions in some metrics enables the enclosure of the exact solution and the estimation of the discretization error in this metrics.

If two numerical solutions with the discretization error relating twice or more in some metrics are available, the exact solution is located in the hypersphere with the centre at the more accurate solution and with the radius, which equals the distance between solutions.

If there is no *a priori* information on error ranging, the enclosure of the exact solution is feasible if the collection of solutions is available and it is split into separated clusters corresponding “accurate” and “inaccurate” schemes. The distance between clusters should be greater the dimension of "accurate" cluster.

The numerical tests confirmed the efficiency of this heuristic rule in metrics corresponding  $L_2$  and  $L_1$  norms for two dimensional supersonic problems governed by Euler equations. The

success of the enclosure is sensitive to the choice of the metric. The metrics, engendered by the norm, which imitate the relative error, provide an opportunity to enclosure the exact solution in certain events when  $L_2$  and  $L_1$  based enclosure fails. For nonlinear problems, the estimation of the vicinity of the numerical solution, which contains the exact solution, may be interpreted as the proof of the existence of the exact solution nearby the approximate solution.

### Acknowledgements

Authors acknowledge the partial support by grants of RFBR № 16-01-00553A and № 17-01-00444A.

### REFERENCES

- [1] R.D. Richtmyer and K.W. Morton, *Difference Methods for Initial Value Problems*, John Wiley and Sons, NY (1967).
- [2] N.N. Kuznetsov, “Accuracy of some approximate methods for computing the weak solutions of a first-order quasi-linear equation”, *USSR Comp. Math. and Math. Phys.*, **16**, 105-119 (1976).
- [3] T. Tang and Z.H. Teng, “The sharpness of Kuznetsov's  $O(\sqrt{\Delta x})$   $L^1$  error estimate for monotone difference schemes”, *Math. Comput.*, **64**, 581-589 (1995).
- [4] J.W. Banks, T. Aslam, and W.J. Rider, “On Sub-linear Convergence for Linearly Degenerate Waves in Capturing Schemes”, *JCP*, **227**, 6985-7002 (2008).
- [5] M.H. Carpenter and J.H. Casper, “Accuracy of Shock Capturing in Two spatial Dimensions”, *AIAA J*, **37** (9), 1072-1079 (1999).
- [6] O.A. Ladyzhenskaya, “The Boundary Value Problems of Mathematical Physics”, *Applied Mathematical Sciences*, Springer, **49**, 1-43 (1985).
- [7] I. Babushka and W. Rheinboldt, “A Posteriori error estimates for the finite element method”, *Int. J. Numer. Methods Eng.*, **12**, 1597–1615 (1978).
- [8] I. Babushka and A.D. Miller, “The post-processing approach in the finite element method, III”, “A Posteriori error estimation and adaptive mesh selection”, *Int. J. Numer. Meth. Eng.*, **20**, 2311–2324 (1984).
- [9] S.I. Repin, *A Posteriori estimates for partial differential equations*, 4, Walter de Gruyter, Berlin-New York, (2008).
- [10] S.I. Chernyshenko, P. Constantin, J.C. Robinson, E.S. Titi, “A Posteriori regularity of the three-dimensional Navier-Stokes equations from numerical computations”, *J. of Mathematical Physics*, **48**, 065204 (2007).
- [11] C. Ortner, “A Posteriori Existence in Numerical Computations”, *SIAM Journal on Numerical Analysis*, **47**(4), 2550-2577 (2009).
- [12] G.I. Marchuk and V.V. Shaidurov, *Difference methods and their extrapolations*, Springer, N.Y. (1983).
- [13] Z. Zlatev, I. Dimov, I. Farago, K. Georgiev, A. Havasic, T. Ostromsky, “Application of Richardson extrapolation for multi-dimensional advection equations”, *Computers and Mathematics with Applications*, **67**, 2279–2293 (2014).
- [14] O.B. Feodoritova, N.D. Novikova, V.T. Zhukov, “Multigrid method for diffusion equations based on adaptive smoothing”, *Mathematica Montisnigri*, **XXXVI**, 14-26 (2016).
- [15] U.S. Fjordholm, S. Mishra, E. Tadmor, “On the computation of measure-valued solutions”, *Acta Numerica*, 567–679 (2016).
- [16] T. Linss and N. Kopteva, “A Posteriori Error Estimation for a Defect-Correction Method Applied to Convection-Diffusion Problems”, *Int. J. of Numerical Analysis and Modeling*, **1** (1), 1–16, (2009).
- [17] A.K. Alekseev, I.M. Navon, “A Posteriori Error Estimation by Postprocessor Independent of

- Flowfield Calculation Method”, *Computers & Mathematics with Applications*, **51**, 397-404 (2006).
- [18] M.B. Giles and E. Suli, “Adjoint methods for PDEs: a posteriori error analysis and postprocessing by duality”, *Acta numerica*, 145-206 (2002).
- [19] Yu.I. Shokin, *Method of differential approximation*, Springer-Verlag, (1983).
- [20] A.K. Alekseev, I.M. Navon, “Adjoint Correction and Bounding of Error Using Lagrange Form of Truncation Term”, *Computers & Mathematics with Applications*, **50** (8-9), 1311-1332 (2005).
- [21] R.D. Skeel, “Thirteen ways to estimate global error”, *Numer. Math.*, **48**, 1-20 (1986).
- [22] C.J. Roy, “Review of Discretization Error Estimators in Scientific Computing”, *48-th AIAA*, 126, 1-29 (2010).
- [23] Ch.J. Roy, and A. Raju, “Estimation of Discretization Errors Using the Method of Nearby Problems”, *AIAA J.*, **45** (6), 1232-1243 (2007).
- [24] T.S. Phillips and Ch.J. Roy, “Residual Methods for Discretization Error Estimation”, *AIAA*, Paper 2011-3870, 1-18 (2011).
- [25] E.J. Nielsen and W. L. Kleb, “Efficient construction of discrete adjoint operators on unstructured grids using complex variables”, *AIAA J.*, **44** (4), 827–836 (2006).
- [26] J.T. Oden and S. Prudhomme, “Estimation of modeling error in computational mechanics”, *Journal of Computational Physics*, **182**, 496-515 (2002).
- [27] A.K. Alexeev, A.E. Bondarev, “On Important Information Dynamics”, *Mathematica Montisnigri*, **XXXV**, 68-89 (2016).
- [28] W.S. Torgerson, “Multidimensional Scaling: Theory and Method”, *Psychometrika*, **17**, 401-419, (1952).
- [29] D. Burago, Yu. D. Burago, S. Ivanov, *A Course in Metric Geometry*, AMS, (2001).
- [30] P.Ch. Mahalanobis, “On the generalised distance in statistics”, *Proceedings of the National Institute of Sciences of India*, **2** (1), 49–55 (1936).
- [31] B. Edney, “Anomalous Heat Transfer and Pressure Distributions on Blunt Bodies at Hypersonic Speeds in the Presence of an Impinging Shock”, *FAA Report 115, Aeronautical Research Institute of Sweden*, (1968).
- [32] R. Courant, E. Isaacson, M. Rees, “On the Solution of Nonlinear Hyperbolic Differential Equations by Finite Differences”, *Comm. Pure Appl. Math.*, **5**, 243-255 (1952).
- [33] A.G. Kulikovskii, N.V. Pogorelov, and A.Yu. Semenov, *Mathematical Aspects of Numerical Solution of Hyperbolic Systems*, Monographs and Surveys in Pure and Applied Mathematics, **188**, Chapman&Hall/CRC, Boca Raton, Fl, (2001).
- [34] B. Van Leer, “Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov’s method”, *J. Comput. Phys.*, **32**, 101–136 (1979).
- [35] M. Sun, K. Katayama, “An artificially upstream flux vector splitting for the Euler equations”, *JCP*, **189**, 305-329 (2003).
- [36] Hy Trac, Ue-Li Pen, “A Primer on Eulerian Computational Fluid Dynamics for Astrophysics”, arXiv:astro-ph/0210611, **2**, 1-23 (2002).
- [37] S. Osher, S. Chakravarthy, “Very high order accurate TVD schemes”, *ICASE Report*, **84** (144), 229–274 (1984).
- [38] C.-T. Lin et al, “High resolution finite volume scheme for the quantum hydrodynamic equations”, *JCP*, **228** (5), (2009).
- [39] S. Yamamoto, H. Daiguji, “Higher-order-accurate upwind schemes for solving the compressible Euler and Navier-Stokes equations”, *Computers and Fluids*, **22**, 259-270 (1993).

Received January 15, 2017.