On Triangle Inequality Based Approximation Error Estimation

A.K. Alekseev[1*], A.E.Bondarev[2], I. M. Navon[3]

[1]*Moscow Institute of Physics and Technology, Moscow, Russia*

[2]*Keldysh Institute of Applied Mathematics RAS, Moscow, Russia*

[3]*Department of Scientific Computing, Florida State University, Tallahassee, FL 32306-4120, USA*

Abstract

The distance between the true and numerical solutions in some metric is considered as the discretization error magnitude. If error magnitude ranging is known, the triangle inequality enables the estimation of the vicinity of the approximate solution that contains the exact one (*exact solution enclosure*). The analysis of distances between the numerical solutions enables discretization error ranging, if solutions' errors are significantly different. Numerical tests conducted using the steady supersonic flows, governed by the two dimensional Euler equations, demonstrate the properties of the exact solution enclosure. The set of solutions generated by solvers of different orders of approximation is used. The success of this approach depends on the choice of metric.

Keywords: *discretization error, ensemble of numerical solutions, metric, Euler equations.*

1. Introduction

The abundant set of numerical methods with the wide range of approximation orders, which is available at present, provides additional opportunities for the analysis of CFD results. From this viewpoint, we consider the distances between approximate solutions. These distances are caused by the approximation errors, so, some structure in the set of numerical solutions may be engendered by differences in the order of approximation. The order of approximation of a finite-difference/finite volume scheme is related to the truncation error order. The truncation error δu may be obtained via Taylor series decomposition of the discrete operator $A_h u_h = f_h$, which approximates the system of PDE, formally denoted herein as $Au = f$. The truncation error dependence on the spatial step h is usually written 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 may be described by the formal solution $\Delta u = A^{-1}\delta u$. For linear problems, the approximation error $\Delta u = O(h^n)$ has the same order n (Lax theorem, [1]) if the discrete operator is well-posed (the inverse operator is uniformly bounded $\left\| A_h^{-1} \right\| < C$). For the case of nonlinear equations [2, 3, 4, 5, 6, 7] the error order is essentially local and varies significantly depending on the type of flow structures. In this case, the observed order of convergence is not equal to the nominal order of the approximation error even in the asymptotic range. A similar situation may be caused by discontinuities in the coefficients [8]. There is no convergence, if the discrete operator is not well-posed, for example, this may occur for the Kelvin-Helmholtz instability [9].

The commonly used single-grid classes of the discretization error estimation are based on certain *norms* (*a priori* and *a posteriori* error estimation). The present discussion is addressing the estimation of the distances in *metrics*, which may either be engendered by a norm or may be norm independent. We hope to gain additional generality and flexibility by transition from norms to metrics.

*corresponding author
Email addresses: [1]alekseev.ak@phystech.edu, [2]a_bond2001@mail.ru, [3]navonim@gmail.com

A priori error norm estimation may be expressed in the form $\|\Delta u\| < C \cdot h^n$, which contains unknown constants independent of the numerical solution. It is a common approach to error analysis in the design of numerical algorithms. *A priori* error estimation justifies the practice to stop mesh refining when the dependence of numerical solution on the step size becomes unobservable.

A posteriori error estimation [10, 11, 12] may be presented in the form $\|\Delta u\| \le \eta_h(u_h)$, where the computable error indicator $\eta_h(u_h)$ depends only on the approximate solution u_h. Sometimes [13], the form $\|\Delta u\| \le C(u_h) \cdot \eta_h$ is used, where $C(u_h)$ is a computable constant, which depends on the numerical solution, and η_h is the computable bound on the residual $\|\delta u\| \le \eta_h$. At present, the best results in this direction are achieved for elliptic equations and finite element methods starting from the seminal work by Babushka [10]. In the finite element notations δu corresponds to residual r_h and the approximation error $\Delta u = u_h - u$ is usually noted as e_h. In most of practical applications the constant $C(u_h)$ is not estimated, while the error indicator is used for the mesh adaptation. However, *a posteriori* error estimation may provide additional information regarding both the error and the exact solution. For example, [13] demonstrated that the estimation of the stability constant and the residual may be used for the determination of a vicinity of the numerical solution, which contains the exact solution. The feasibility of single grid rigorous estimations is the significant merit of this approach when compared with the standard mesh refinement or the Richardson extrapolation [14, 15].

There are different approaches to the estimation of the truncation error δu. It may be computed by the action of the high order scheme stencil on a precomputed flowfield [16, 17], by the action of the differential operator on the interpolation of the numerical solution [18] or via a differential approximation [19, 20].

Surveys of the global (approximation) error $\Delta u = A^{-1}\delta u$ calculation methods may be found in [21, 22]. In the simplest option, the estimation of this error may be performed using the defect correction [16, 23, and 24]. In the defect correction frame, the truncation error δu is used as the source term intended for the correction of 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 engender their own error. The estimation of the error may be performed also via the linearized problem [24], complex differentiation [25] or by adjoint equations [17, 18, 20, 26]. Usually, adjoint equations are applied to the estimation of a valuable functional (drag, lift etc.) uncertainty. Nevertheless, the variant of adjoint method, described in [20], enables estimation of the norm of the solution error. Unfortunately, it implies the solution of a number of adjoint problems that is proportional to the number of grid nodes that implies an extremely high computational burden.

The incompleteness of truncation error estimation is the general drawback of the above listed residual-based methods. The differential approximation methods based on Taylor series [20] do not account for high order terms of the expansion. The postprocessor based methods do not account for the higher scheme truncation errors [17] or the interpolation errors [18].

Herein, the truncation error is accounted for completely, although implicitly, since the analysis is conducted in the space of numerical solutions. We consider a single-grid analysis of non-intrusive post-processor type. The ensemble of calculations, performed by the solvers of different approximation order, is used for the search of the numerical solution vicinity that contains an exact solution. We denote this operation as the "*exact solution enclosure*". In contrast to above mentioned norm oriented approaches, the current analysis is based on the ensemble of distances (distance matrix) in different metrics, an approach that provides a more general and flexible analysis. The norm oriented variant of this approach is presented in [27, 28]. The Multidimensional Scaling (MDS) [29] concerns similar problems, however, we consider the cases when MDS cannot be applied, since the data vector (numerical solution) length greatly exceeds the number of data vectors. Numerical tests demonstrated that various metrics have significantly different properties

from exact solution enclosure perspective. The best characteristics are observed for the IMED metric [30].

The paper is organized as follows. In Section 2 we discuss the opportunities for the discretization error estimation that are provided by *a priori* information regarding error rating. Section 3 considers *a posteriori* analysis of error relations provided by the ensemble of numerical solutions performed by different solvers. The supersonic shocked flows, described by the two dimensional Euler equations, are considered as test problems in Section 4. Section 5 presents the set of metrics, which are used for comparing of solutions. In Section 6 we present the results of the ensemble based error measure estimation, performed using different metrics, in comparison with the true error. The solvers, used for the computations are listed in this section. Several issues, concerning the applications of the metric based error analysis, are surveyed in Section 7. Conclusions are presented in the final Section 8.

2. Exact solution enclosure for approximate solutions with ranged errors.

We analyze the ensemble of numerical solutions obtained on the same grid using finite volume schemes of different approximation order. We denote the numerical solution as the vector $u^{(i)} \in R^N$ (i is the scheme number, N is the number of grid points). The values of an unknown exact solution at nodes of this grid ("exact" solution) is denoted as $\widetilde{u} \in R^N$. The approximation error magnitude is considered as the distance between the exact and approximate solutions $d(u^{(k)}, \widetilde{u}) = \delta_{0,k}$ in some metric (for example, $d(u^{(k)}, \widetilde{u}) = \left\| u^{(k)} - \widetilde{u} \right\|_{L_2}$). Let the relation of these approximation error values be known *a priori*.

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* $\delta_{1,2} = d(u^{(1)}, u^{(2)})$ *between two numerical solutions* $u^{(1)} \in R^N$ *and* $u^{(2)} \in R^N$ *be known from computations and distances between numerical and exact solutions be related as*

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

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

$$\delta(u^{(2)}, \widetilde{u}) \leq \delta_{1,2} \tag{2}$$

Proof. The triangle inequality [31] for distances $\delta_{0,1}, \delta_{1,2}, \delta_{0,2}$ between points $u^{(1)}, u^{(2)}, \widetilde{u}$ may be presented as $\delta_{0,1} \leq \delta_{1,2} + \delta_{0,2}$ or $\delta_{0,1} - \delta_{0,2} \leq \delta_{12}$. By accounting (1) in the form $\delta_{0,1} - \delta_{0,2} \geq \delta_{0,2}$, one obtains $\delta_{0,2} \leq \delta_{0,1} - \delta_{0,2} \leq \delta_{12}$ and, finally, the desired expression $\delta_{0,2} \leq \delta_{12}$.

The *Theorem 1* may be stated in a slightly more general form: if two solutions are ranged by the error as

$$\delta_{0,1} > (1 + \alpha)\delta_{0,2}, \alpha > 0, \tag{3}$$

then $\delta_{0,1} - \delta_{0,2} > \alpha\delta_{0,2}$, $\alpha\delta_{0,2} < \left| \delta_{0,1} - \delta_{0,2} \right| \leq \delta_{1,2}$ and

$$\delta_{0,2} < \delta_{1,2} / \alpha. \tag{4}$$

This means that two numerical solutions, having the error relation $1 + \alpha, (\alpha > 0)$ in some metric, define the domain, which contains the exact solution as the hypersphere around more accurate solution $\delta_{0,2} < \delta_{1,2} / \alpha$. As a result, the distance between two numerical solutions enables finding the

vicinity of more accurate solution that contains an exact solution, if the relation of errors is known *a priori* in some metrics.

3. A posteriori analysis of the error relations

Despite the widespread opinion that the schemes of higher order are more accurate, the evident weakness of *Theorem 1* is the assumption of the existence of solutions with *a priori* ranged error. For this reason, we consider some options for *a posteriori* check of error rating.

Numerical tests demonstrate that the collection of distances between solutions $\delta_{i,j}$ enables a detection of the close and distant solutions in certain events. For example, if $\delta_{0,1} >> \delta_{0,i}$, the set of distances $\delta_{i,j}$ is split into a cluster related to inaccurate solution (with great values $\delta_{1,j}$) and the cluster of more accurate solutions $\delta_{i,j}(i \neq 1)$. This is caused by the asymptotics $\delta_{1,j}/\delta_{0,1} \to 1$ and $\delta_{i,j}(i \neq 1)/\delta_{0,1} \sim (\delta_{0,i} + \delta_{0,j})/\delta_{0,1} \to 0$ at $\delta_{0,i}/\delta_{0,1} \to 0$.

The separation of the distances between approximate solutions into clusters is the evidence of the existence of solutions with significantly different errors and 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 $\delta_{i,j}$, where $u^{(1)}$ is maximally incorrect solution and $u^{(i)}$ is some more accurate solution (the localization of exact solution is performed in its vicinity), $\delta_{0,i}^{\max}$ is the maximum error in the subset of accurate solutions. The maximum of $\delta_{i,j}(i \neq 1)$ (the distance from zero to maximum error in the cluster of accurate solutions) is noted as the upper bound of the accurate solutions' cluster d_1, the minimum of $\delta_{1,j}$ is noted as the low bound of the second cluster d_2.

The following heuristic criterion for the *Theorem 1* applicability may be stated as:

Conjecture 1: *If the set of distances between solutions is split into clusters and the distance between clusters is greater than the size of the cluster of accurate solutions: $d_2 - d_1 > d_1$, then the exact solution is located within a hypersphere of radius $\delta_{i,1}$ with its center at $u^{(i)} : \delta_{0,i} \leq \delta_{i,1}$, where $u^{(i)}$ belongs to the cluster of more accurate solutions and $u^{(1)}$ is the maximally inaccurate solution.*

This conjecture is based on the assumptions that the dimension of the accurate cluster is equal to $d_1 = 2\delta_{0,i}^{\max}$ $(i \neq 1)$, and the cluster of inaccurate solutions belongs to the interval $(\delta_{0,1} - \delta_{i,\max}, \delta_{0,1} + \delta_{i,\max})$, so $d_2 = \delta_{0,1} - \delta_{0,i}^{\max}$. Since both these evaluations correspond to collinear vectors of error, they are overestimated. If one assumes them to be valid, the relation of accurate cluster dimension and the distance between clusters has the form $\delta_{0,1} - \delta_{0,i}^{\max} > 4\delta_{0,i}^{\max}$. This leads to the relation $\delta_{0,1} > 5\delta_{0,i}^{\max}$, which ensures the condition (1) $\delta_{0,1} > 2\delta_{0,i}^{\max}$.

This criterion may be rigorous only in the limit of the infinite set of solutions obtained by independent methods.

Nevertheless, numerical tests for two dimensional supersonic inviscid flows confirm the applicability of this heuristic criterion, however, with a significant dependence on applied metrics.

4. Test problems

Several flow patterns, governed by two dimensional unsteady Euler equations,

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

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

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

are considered as the tests problems.

Here $U^1 = U, U^2 = V$ are the velocity components, $h_0 = (U^2 + V^2)/2 + h$, $h = \frac{\gamma}{\gamma - 1}\frac{P}{\rho} = \gamma e$,

$e = \frac{RT}{\gamma - 1}$, $E = \left(e + \frac{1}{2}(U^2 + V^2)\right)$ are enthalpies and energies (per unit volume), $P = \rho RT$ is the

state equation and $\gamma = C_p / C_v = 1.4$ is the specific heat ratio.

The flow structures engendered by the single oblique shock wave, the interaction of shock waves of I and VI kinds according Edney classification [32, 33] were used as the test problems due to the availability of analytic solutions. The flow patterns were determined by the selection of the inflow and lateral boundaries conditions. The computations were performed for a Mach number range of $M = 3 \div 5$, flow deflection angles range $\alpha = 10 - 30^o$. All tests concern the steady state solutions.

The values of analytical solution at grid points are considered herein as the 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 presents the computed isolines of density for Edney-I flow structure ($M = 3$ and flow deflection angles $\alpha_1 = 20^o$ and $\alpha_2 = 15^o$). The isolines are provided with a small step to illustrate the presence of a numerical error. The crossing shock waves and contact discontinuity line, engendered at the shocks crossing point, are the main elements of this flow structure.

Fig. 2 presents the density distribution for Edney-IV flow structure ($M = 4$, two consequent flow deflection angles $\alpha_1 = 10^o$, $\alpha_2 = 15^o$). The flow is determined by the merging shock waves, the contact line and the expansion fan.
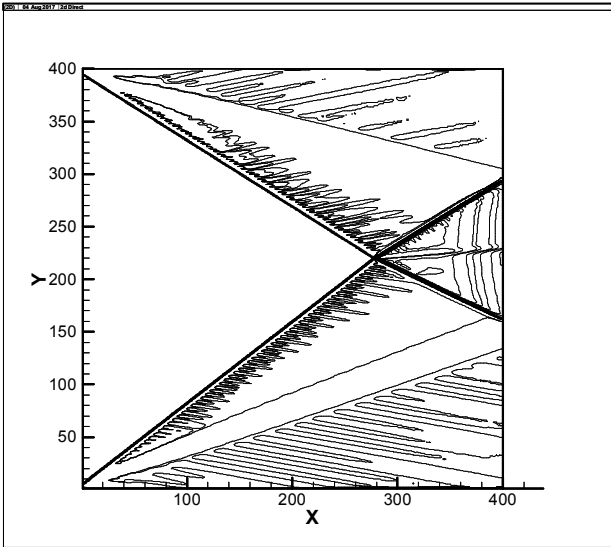

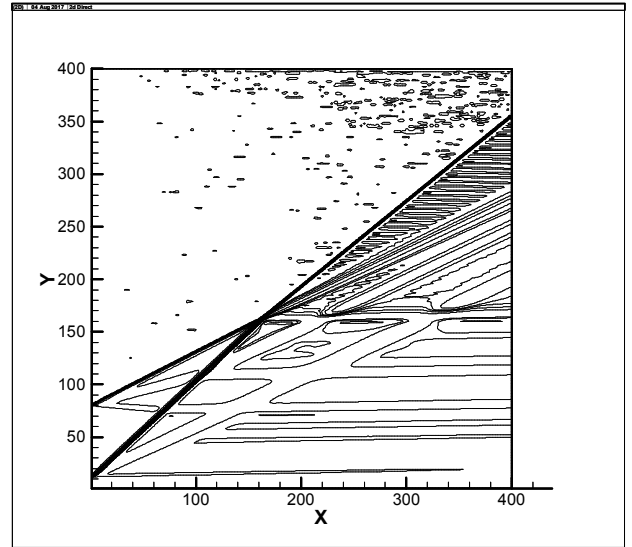Fig. 1. Edney I density isolines.

Fig. 2. Edney VI density isolines.

5. The set of metrics for flows comparison

Both the *Theorem 1 and Conjecture 1* are stated for the distances determined by some (unspecified) metrics. Metrics, used in CFD, significantly differ. L_1 norm based metric seems to be

most natural for problems related with shocks, since most results on approximation error are obtained in this norm. However, it is not informative, since most experience is related to the valuable functionals (lift, drag, etc). For norms engendered by some inner product (L_2, for example) the uncertainty of valuable functionals may be related to the error norm via the Cauchy–Bunyakovsky–Schwarz inequality

$$\left|\Delta\varepsilon\right| = \left|(\partial\varepsilon/\partial u \cdot (u_h - \tilde{u}))\right| < \left\|\partial\varepsilon/\partial u\right\|_{L_2} \cdot \left\|u_h - \tilde{u}\right\|_{L_2} \leq \left\|\partial\varepsilon/\partial u\right\|_{L_2} \cdot \left\|u^{(1)} - u^{(k)}\right\|_{L_2}.$$ From this viewpoint,

such norms may be more interesting in comparison with L_1. On other hand, the H^{-1} norm seems to be suitable for solutions with a low regularity.

Herein, we compare the metrics engendered by the L_1, L_2, H^{-1} norms, L_2 based metrics which imitates a relative error (REM-L_2), and IMED metrics [30]. The metrics, having some physical meaning, illustrative capabilities, and a potential for flow comparison are not limited by above considerations, so the search for optimal metric is of further interest.

We consider the four component solution $u^{(i)} = \{\rho^{(i)}, U^{(i)}, V^{(i)}, e^{(i)}\}$. For the metrics engendered by the L_1 and L_2 norms, the distance between solutions is expressed as

$$\left\|u^{(i)} - u^{(k)}\right\|_{L_1} = \left(\frac{1}{N}\sum_{j=1}^{N}\left(\left|\rho^{(i)} - \rho^{(k)}\right|_j + \left|U^{(i)} - U^{(k)}\right|_j + \left|V^{(i)} - V^{(k)}\right|_j + \left|e^{(i)} - e^{(k)}\right|_j\right)\right), \tag{8}$$

$$\left\|u^{(i)} - u^{(k)}\right\|_{L_2} = \left(\frac{1}{N}\sum_{j=1}^{N}\left((\rho^{(i)} - \rho^{(k)})_j^2 + (U^{(i)} - U^{(k)})_j^2 + (V^{(i)} - V^{(k)})_j^2 + (e^{(i)} - e^{(k)})_j^2\right)\right)^{1/2}. \tag{9}$$

For CFD problems, the vector of solution contains elements having different physical meanings, such as density, velocity components, and energy. So, in parallel to Expressions (8,9), the distance between solutions was calculated using the normalized expression

$$\left\|\{(\rho^{(i)} - \rho^{(k)})/\left\|\rho^{(i)}\right\|, (U^{(i)} - U^{(k)})/\left\|U^{(i)}\right\|, (V^{(i)} - V^{(k)})/\left\|V^{(i)}\right\|, (e^{(i)} - e^{(k)})/\left\|e^{(i)}\right\|\}\right\|_{L_2}, \tag{10}$$

which imitates a relative error (we note this expression as "relative error metric" (REM-L_2)). It should be noted that expression (10) corresponds to the distance

$$(\Delta u^{(i)}, M\Delta u^{(i)}) = (M_{j,k}\Delta u_j^{(i)}\Delta u_k^{(i)})^{1/2}. \tag{11}$$

This distance is determined by a metric tensor $M_{j,k}$ 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 [34]) one may 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 enclose the solution in the transformed space $Au^{(i)}$ (and L_2 norm) where the error may be described by a hypersphere.

The metric engendered by the Sobolev norm of the negative order (H^{-1}) [35,36] also is of the great interest due to the low regularity of the Euler equation solutions. According to [35], the Sobolev norm in H^{-1} may be expressed as

$$\left\|f\right\|_{H^{-1}} = \sup_{\|u\|_{H^1}=1}\left|(f,u)\right|. \tag{12}$$

It was computed using the expression [36],

$$\|f\|_{H^{-1}} = (f, \widetilde{u})_{L_2},$$ (13)

where $\widetilde{u}$ is the solution of the screened Poisson equation

$$\frac{\partial \widetilde{u}}{\partial t} - \lambda \frac{\partial^2 \widetilde{u}}{\partial x^2} - \lambda \frac{\partial^2 \widetilde{u}}{\partial y^2} + \widetilde{u} = -f.$$ (14)

The coefficient λ determines the smoothing properties for the transformation $\widetilde{u} \to f$. The value of λ was varied in the range $10^{-4} \div 10^{-6}$. The calculations are performed by components for $f = \{(\rho^{(i)} - \rho^{(k)}), (U^{(i)} - U^{(k)}), (V^{(i)} - V^{(k)}), (e^{(i)} - e^{(k)})\}$ and corresponding $\widetilde{u}$. We used the divergent integro-interpolation method [37] and the time relaxation approach to solve this equation.

The above considered metrics are sensitive to small variations of the flowfield, such as shift of the shock wave location by single cell. So, two numerical flowfields, engendered by such shift, are considered as distant and describing different flow structures. However, such flows are identical from a practical viewpoint. Thus, these distances do not capture the structural proximity between solutions. The Euclidean Distance, modified for analysis of images (IMage Euclidean Distance (IMED)), is of interest from this standpoint [30]. It is described by the metric matrix

$$M_{ij} = \frac{1}{2\pi\sigma^2} \exp\{-\left|P_i - P_j\right|^2 / (2\sigma^2)\}.$$ (15)

The value $\left|P_i - P_j\right|$ is the distance between nodes P_i and P_j on the grid. For example, if P_i corresponds to the cell (k,l), and P_j corresponds to the cell (k_1, l_1), $\left|P_i - P_j\right|$ may be estimated as

$$\left|P_i - P_j\right| = ((k - k_1)^2 + (l - l_1)^2)^{1/2}.$$ (16)

For our two dimentional problem, the distance was estimated using the following form (presented here only for density)

$$(\Delta\rho^{(i)}, M\Delta\rho^{(i)}) = (\sum_{j,k,m,n} \frac{1}{2\pi\sigma^2} \exp\{-((j-m)^2 + (k-n)^2)/(2\sigma^2)\}\Delta\rho_{j,k}^{(i)}\Delta\rho_{m,n}^{(i)})^{1/2}.$$ (17)

This distance corresponds to the averaged error. At $\sigma \leq 0.25$ the probability distribution approximation is of poor quality. At $\sigma = 0.5 \div 1$ the values obtained by (17) are close to L_2 norm.

Asymptotically H^{-1} and IMED tends to L_2 as $\lambda \to 0$ or $\sigma \to 0$.

6. Results of numerical tests

The analysis used the ensemble of computations performed by the following methods.

The first order scheme by Courant Isaacson Rees [38] marked as $S1$ was used in the variant described by [39].

The second order scheme based on the MUSCL method [40] and using algorithm by [41] at cell boundaries is denoted as $S2$.

Second order TVD scheme of relaxation type by [42] is denoted as $S2TVD$.

Third order modified Chakravarthy-Osher scheme [43, 44] is marked as $S3$.

Fourth order scheme by [45] is marked as $S4$.

The FORTRAN codes by [42] are used for $S2TVD$. All other solvers were coded by the authors.

Computations were performed on uniform grids containing 100×100, 200×200 and 400×400 nodes.

Methods $S1, S2, S3, S4$ (1,2,3 and 4 nominal truncation orders) demonstrated the order of convergence (average) slightly below $n = 1/2$ in norm L_2. In norm L_1 the same computations demonstrated the order of convergence slightly higher than $n = 1/2$. The method S2TVD (nominal order 2) is the only exception with the order about $n \sim 3/4$. Second order $S2TVD$ scheme [42] from standpoint of error norm is close to first order scheme $S1$ for 100×100 grid and to high order schemes for grid 400×400. The calculations on the grid 100×100 demonstrated the formation of clusters in distances between $S2TVD$ and $S2, S3, S4$ engendered solutions and successful enclosure of the exact solution. However, the distances between solutions engendered by $S2TVD$ and $S2, S3, S4$ do not form clusters on the grid 400×400. Paradoxically, the reason for this failure is the relatively rapid convergence of $S2TVD$. So, the results, obtained by $S1, S2, S3, S4$ methods, are provided for illustration as more stable.

We first check *Conjecture 1* and, second, verify the exact solution enclosure. The enclosure is considered as successful, if the error estimate $d(u^{(i)} - u^{(k)})$ is greater than the true error $d(u^{(k)} - \tilde{u})$, obtained in comparison with the analytical solution $\tilde{u}$.

The tests permit to conclude that the solutions obtained by the scheme $S1$ (as "inaccurate") and by $S2$, $S3$, $S4$ (as "accurate") enable to find the vicinity of numerical solution that contains exact solution for all tested grids.

The comparison of results by schemes $S2, S3, S4$ does not enable to select clusters and to enclose the exact solution. These schemes produce solutions with errors which are close in magnitude and splitting into clusters is not observed.

If the *Conjecture 1* is not satisfied, the enclosure of true solution fails. However, the exact error value is about two or three maximum distances between numerical solutions that provides some additional way for the error estimation.

The numerical tests for the single oblique shock demonstrate the feasibility for the exact solution enclosure, for example, see Tables 1 and 2. Table 1 demonstrates the formation of clusters in different metrics (S2-S4, S3-S4 and S3-S2 are smaller than S2-S1, S3-S1, S4-S1). Table 2 demonstrates the successful exact solution enclosure with exclusion of the small violation in L_2. The tests correspond to $M = 4$, flow deflection angle $\alpha_1 = 10^o$, and the grid 100×100.

Table 1. Distances between solutions for single shock test.

$u^{(i)} - u^{(k)}$	S4-S2	S3-S2	S3-S4	S2-S1	S3-S1	S4-S1
L_1	0,00569	0,0052	0,0032	0.0186	0,023	0,024
L_2	0,0199	0,0287	0,011	0,0452	0,0566	0,060
H^{-1}	0,0086	0,0076	0,00278	0,029	0,035	0,036
REM-L_2	0,048	0,0145	0,022	0,0945	0,115	0,124
IMED	0,0145	0,0126	0,005	0,0519	0,067	0,0536

Table 2. Exact solution enclosure for single shock test.

$u^{(i)} - u^{(k)}$	S2-S1	S2-exact	S3-S1	S3-exact	S4-S1	S4-exact
L_1	0.0186	0,0116	0,023	0,0092	0,024	0,0066
L_2	0,0452	0,0459	0,0566	0,0407	0,060	0,0337
H^{-1}	0,029	0,0137	0,035	0,00934	0,036	0,00784
REM-L_2	0,0945	0,0923	0,115	0,079	0,124	0,0655
IMED	0,0519	0,0231	0,067	0,0155	0,0536	0,012

For Edney-VI shock interaction ($M = 4$, $\alpha_1 = 10^o$, $\alpha_2 = 15^o$, 100×100), the set of distances between solutions also splits into clusters. There is a successful enclosure of exact solution, Tables 3-4.

Table 3. Distances between solutions for Edney-VI test.

$u^{(i)} - u^{(k)}$	S4-S2	S3-S2	S4-S3	S2-S1	S3-S1	S4-S1
L_1	0,023	0,0098	0,021	0.0668	0,072	0,0874
L_2	0,059	0,025	0,051	0,149	0,16	0,191
H^{-1}	0,028	0,0107	0,0127	0,0976	0,0928	0,121
REM-L_2	0,051	0,0189	0,041	0,136	0,145	0,170
IMED	0,043	0,0195	0,035	0,179	0,192	0,171

Table 4. Exact solution enclosure for Edney-VI test.

$u^{(i)} - u^{(k)}$	S2-S1	S2-exact	S3-S1	S3-exact	S4-S1	S4-exact
L_1	0.0668	0,046	0,072	0,046	0,0874	0,0375
L_2	0,149	0,128	0,16	0,138	0,191	0,133
H^{-1}	0,0976	0,055	0,0928	0,0603	0,121	0,0607
REM-L_2	0,136	0,0898	0,145	0,093	0,170	0,0846
IMED	0,179	0,076	0,192	0,084	0,171	0,098

Tables 5 and 6 present the results for different metrics from the viewpoint of exact solution enclosure for Edney-I test ($M = 3$, flow deflection angles $\alpha_1 = 20^o$ and $\alpha_2 = 15^o$, 100×100).

Table 5. Distances between solutions for Edney-I test. Rough mesh.

$u^{(i)} - u^{(k)}$	S4-S2	S3-S2	S4-S3	S2-S1	S3-S1	S4-S1
L_1	0,017	0,018	0,019	0.0563	0,0673	0,0721
L_2	0,044	0,043	0,045	0,107	0,128	0,141
H^{-1}	0,0164	0,0154	0,0129	0,0609	0,0705	0,075
REM-L_2	0,05	0,039	0,043	0,122	0,14	0,159
IMED	0,028	0,028	0,022	0,126	0,148	0,159

Table 6. Exact solution enclosure for Edney-I test. Rough mesh.

$u^{(i)} - u^{(k)}$	S2-S1	S2-exact	S3-S1	S3-exact	S4-S1	S4-exact
L_1	0,0563	0,0436	0,0673	0,050	0,0721	0,039
L_2	0,107	0,124	0,128	0,146	0,141	0,139
H^{-1}	0,0609	0,0512	0,0705	0,587	0,075	0,0597
REM-L_2	0,122	0,163	0,14	0,178	0,159	0,176
IMED	0,126	0,114	0,148	0,0902	0,159	0,129

The tests by Tables 1-6 corresponds to the relatively rough mesh 100×100. A similar behaviour is observed for finer grids (200×200 and 400×400). Tables 7 and 8 present the results for different metrics from the viewpoint of exact solution enclosure for Edney-I test ($M = 3$, flow

deflection angles $\alpha_1 = 20^o$ and $\alpha_2 = 15^o$, 400×400). Edney-I test is selected since it demonstrates the worst results if compare with the single shock and Edney-VI tests.

Table 7. Distances between solutions for Edney-I test. Fine mesh.

$u^{(i)} - u^{(k)}$	S4-S2	S3-S2	S4-S3	S2-S1	S3-S1	S4-S1
L_1	0,0061	0,0052	0,0068	0.0169	0,0202	0,0223
L_2	0,0217	0,0226	0,0227	0,0545	0,0655	0,0709
REM-L_2	0,026	0,020	0,022	0,0644	0,0739	0,0830
H^{-1}	0,0148	0,0157	0,0135	0,0649	0,0764	0,0820
IMED	0,014	0,0145	0,011	0,0615	0,0764	0,0815

Table 8. Exact solution enclosure for Edney-I test. Fine mesh.

$u^{(i)} - u^{(k)}$	S2-S1	S2-exact	S3-S1	S3-exact	S4-S1	S4-exact
L_1	0.0169	0,0122	0,0202	0,0147	0,0223	0,0123
L_2	0,0545	0,0680	0,0655	0,0802	0,0709	0,0760
REM-L_2	0,0644	0,0662	0,0739	0,0767	0,0830	0,0754
H^{-1}	*0,0458*	0,0456	0,0546	0,0548	0,0577	0,0521
IMED	0,0615	0,0527	0,0764	0,0638	0,0815	0,0625

Tables 1-8 demonstrate that L_1 successfully performs the exact solution enclosure for all tests. L_2 and REM-L_2 fail for significant part of tests. This demonstrates the heuristic, approximate nature of *Conjecture 1*. However, the violation of enclosure condition in tests is moderate. H^{-1} engendered metric provides an intermediate quality. The IMED metric [30] enables a successful exact solution enclosure for most of tests. So, the choice of metric may be crucial for exact solution enclosure.

7. Discussion

The relation of errors, obtained in above analysis, is not necessarily attributed to properties of considered schemes. It may be caused by the imperfections of numerical realization by the authors. The authors do not pretend to assess the considered numerical schemes. We are mainly concerned with the verification of the non-intrusive single-grid error estimator based on the numerical solutions obtained by the solvers of different accuracy.

The standard grid convergence analysis is based on the heuristic rule by C. Runge [12]. From this viewpoint, if the difference of two approximate solutions on the coarse grid and on the fine grid is small, then numerical solutions are close to exact solution. This rule is not applicable, if there is no grid convergence, the examples of such problems are provided by [9]. Also, this rule may be wrong if the convergence rate is slow. For example, [6,7] considers orders of convergence $n = 1/4 \div 1/6$ for multidimensional finite volume methods, while [8] considers elliptic boundary value problems, whose finite element approximations converge arbitrarily slow.

In a more rigorous approach, one should desire the error estimate of form $\delta(u_h, \tilde{u}) \le \delta$ (or $\|u_h - \tilde{u}\| \le \delta$) with computable δ. Formally, the Richardson method [14,15] is close to this ideal. It enables to determine the refined solution and the error estimate, if the single error order exists in the total flowfield. The set of solutions, computed for different meshes, is used. Unfortunately, in most CFD problems the error order on different flow structures varies [2-8], a fact that hampers the application of the Richardson method.

We considered a single grid alternative to the Richardson method and Runge rule, based on the ensemble of solutions obtained by different solvers. The above considered method may be used

as a postprocessor similar to the Richardson extrapolation. However, it does not require mesh refinement and may be used away from the asymptotic range.

The feasibility of estimating the distance from the exact solution to numerical one $\delta(u_h, \tilde{u}) \leq \delta$ seems attractive. However, in practical applications, the numerical value of δ threshold, when two approximate solutions can be considered as describing the same flow, may not be evident. The magnitude of error ($\delta(u_h, \tilde{u})$ or norm $\|u_h - \tilde{u}\|$) is not very informative in CFD, since most experience is related with the valuable functionals (lift, drag, etc). Nevertheless, the error may be related with the uncertainty of some valuable functionals via the Cauchy–Bunyakovsky–Schwarz inequality, if the error magnitude estimate is engendered by some inner product.

The existence of "accurate" and "inaccurate" schemes is one of the main postulates of computational mathematics, unfortunately, it is valid only in the asymptotic sense.

The above numerical results demonstrate *a posteriori* feasibility to distinguish between "accurate" and "inaccurate" schemes in the sense of error ranging in certain metric. For example, the distributions of distances between solutions provided in Tables 1,3,5,7 show the presence of two clusters corresponding "accurate" and "inaccurate" schemes. This engenders the hope to enclose the exact solution only from observable values of distances between solutions (without *a priori* information on errors ranging), that is confirmed by Tables 2,4,6,8.

If there is no breaking into clusters, the maximum distance between solutions provides an opportunity for a rough estimation of numerical error, since it is relatively close to the distance between numerical and analytical solutions.

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

8. Conclusions

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 located at the more accurate solution and with a radius, which equals the distance between numerical solutions.

If there is no *a priori* information regarding error ranging, the enclosure of the exact solution is feasible, if the set of distances between solutions is split into separated clusters corresponding "accurate" and "inaccurate" schemes. The distance between clusters should be greater than the dimension of "accurate" cluster.

The numerical tests confirmed the efficiency of this heuristic rule for two dimensional supersonic steady problems, governed by Euler equations, with the dependence on the choice of the metric. The L_1 based metric operates successfully in all tests. The L_2 based metric and metric, which imitate the relative error (REM-L_2), fail rather often, although with rather moderate violations. The metric, engendered by H^{-1} norm, provides an intermediate reliability. The IMED [30] metric demonstrated the quality of the exact solution enclosure comparable with the L_1 based metric.

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] T. Barth and M. Ohlberger. Finite Volume Methods: Foundation and Analysis. *Encyclopedia of Computational Mechanics.* 1(15): 1-57, 2004.

[7] M. Ohlberger and J. Vovelle. Error estimate for the approximation of non-linear conservation laws on bounded domains by the finite volume method. *Math. Comp.*, 75: 113-150, 2006.

[8] I. Babuska, J. Osborn. Can a finite element method perform arbitrarily badly? *Math. Comp.,* 69(230): 443-462, 2000.

[9] U.S. Fjordholm, S. Mishra, E. Tadmor. On the computation of measure-valued solutions, *Acta Numerica,* 567–679, 2016.

[10] I. Babuska and W. Rheinboldt. A posteriori error estimates for the finite element method. *Int. J. Numer. Methods Eng.* 12: 1597–1615, 1978.

[11] 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.

[12] S. I. Repin. *A posteriori estimates for partial differential equations.* Vol. 4. Walter de Gruyter, 2008.

[13] C. Ortner. A Posteriori Existence in Numerical Computations. *SIAM Journal on Numerical Analysis*, 47(4):2550-2577, 2009.

[14] G.I. Marchuk and V.V. Shaidurov. *Difference methods and their extrapolations*. Springer, NY, 1983.

[15] 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.

[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 post processing 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] Ch. J. Roy. Review of Discretization Error Estimators in Scientific Computing. *AIAA 2010-126:* 1-29.

[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 2011-3870:* 1-18.

[25] E. J. Nielsen and W. L. Kleb. Efficient construction of discrete adjoint operators on unstructured grids using complex variables. *AIAA Journal.* 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. Alekseev, A.E.Bondarev, I. M. Navon. On Estimation of Discretization Error Norm via Ensemble of Approximate Solutions, *arXiv:1704.04994* [physics.comp-ph] April 18, 2017

[28] A.K. Alexeev, A.E. Bondarev On Exact Solution Enclosure on Ensemble of Numerical Simulations. *Mathematica Montisnigri*. 38: 63-77, 2017.

[29] W.S. Torgerson. Multidimensional Scaling: Theory and Method. Psychometrika. 17: 401-419, 1952.

[30] L. Wang, Y. Zhang, J. Feng. On the Euclidean Distance of Images, IIEE Transactions on Pattern Analysis and Machine Intelligence. 27(8): 1334 – 1339, 2005.

[31] D. Burago, Yu.D. Burago, S. Ivanov. *A Course in Metric Geometry*. AMS, 2001

[32] B. Edney. Effects of Shock Impingement on the Heat Transfer around Blunt Bodies. *AIAA J.,* 6(1): 15-21, 1968.

[33] V.Ya. Borovoy, A. Yu. Chinilov, V.N. Gusev, I.V. Struminskaya, J. Delery, and B. Chanetz, Interference Between a Cylindrical Bow Shock and a Plane Oblique Shock. *AIAA J.* 35(11): 1721-1728, 1997.

[34] P. Ch. Mahalanobis. On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*. 2 (1): 49–55, 1936.

[35] J. H. Bramble, R.D. Lazarov and J. E. Pasciak. Least-squares Methods for Linear Elasticity Based on a Discrete Minus One Inner Product. *Comput. Methods in Appl. Mech. Engrg.* 152: 520-543, 2001.

[36] J. H. Bramble, R.D. Lazarov and J. E. Pasciak. A Least Squares Approach Based on a Discrete Minus One Inner Product For First Order Systems. *Mathematics of Computation*. 66(219): 935-955, 1997.

[37] A. A. Samarskii. *Theory of difference schemes.* Marcel Dekker. NY. 2001.

[38] 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.

[39] 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.

[40] 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.

[41] M. Sun, K. Katayama. An artificially upstream flux vector splitting for the Euler equations, *Journal of Computational Physics.* 189: 305-329, 2003.

[42] Hy Trac, Ue-Li Pen. A Primer on Eulerian Computational Fluid Dynamics for Astrophysics, arXiv:astro-ph/0210611v2, (2002) pp. 1-23.

[43] S. Osher, S. Chakravarthy. Very high order accurate TVD schemes, *ICASE Report*, 84-144: 229–274, 1984.

[44] C.-T. Lin et al. High resolution finite volume scheme for the quantum hydrodynamic equations. *Journal of Computational Physics*. 228(5): 1713-1732, 2009.

[45] 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.