The estimation of approximation error using inverse problem and a set of numerical solutions

A. K. Alekseev and A. E. Bondarev

Keldysh Institute of Applied Mathematics RAS, Moscow, Russia

ABSTRACT

In this paper, we consider the inverse problem for the estimation of a point-wise approximation error occurring at the discretization of the system of partial differential equations. We analyse the set of the solutions, obtained by the numerical algorithms of the dissimilar structures on the same grid. The differences between the numerical solutions are used as the input data for the inverse problem, which is posed in the variational statement with the zero-order Tikhonov regularization. The numerical tests, performed for the two-dimensional inviscid compressible flows corresponding to Edney-I and Edney-VI shock wave interference modes, are provided. The comparison of the estimated error and the exact error, obtained by subtraction of numerical and analytic solutions, is presented.

1. Introduction

The verification of software and numerical solutions for the systems of partial differential equations (PDE) implies the development of methods for the approximation error estimation. For instance, such estimation is required by modern standards [1,2] in the domain of computational fluid dynamics. A significant number of approaches for the evaluation of an approximation error has been developed [3–11]. However, some of them are highly computationally expensive or intrusive (require a modification of code); others are limited by the domain of application. For these reasons, the computationally inexpensive, nonintrusive universal methods for the discretization error estimation are of current interest. Some methods for the estimation of this error are in the focus of our paper. We write a system of partial differential equations (PDE) in the operator form $A(\tilde{u}) = f$. The discrete operator $A_h$ approximates this system on some grid: $A_h(u_h)u_h = f_h$. We denote the numerical solution (a grid function, vector $u_h \in R^M$) by $u_h$, and the projection of an exact solution onto the computational grid by $\tilde{u}_h \in R^M$. The exact approximation error can be expressed as $\Delta u_h = u_h - \tilde{u}_h$, however, unfortunately, it is usually unknown.

The widely used a priori error estimates have the following form $||\Delta u_h|| \leq Ch^p$, where $h$ is the step of discretization, $p$ is the order of approximation and $C$ is an unknown constant. These estimates are defined for the wide class of solutions and, usually, are not computable.
for a particular solution. For this reason the a priori error estimates have a limited impact on the practical applications and are used mainly at the design and analysis of the numerical algorithms.

The computable measure of the discretization error, which can be used in applications, is commonly denoted as a posteriori error estimate $||\Delta u_h|| \leq E(u_h)$ and is determined by an error indicator $E(u_h)$ that depends on the particular numerical solution $u_h$ and has no unknown constants. In the finite-element analysis, the efficient technique for the a posteriori error estimation has a long history that starts from papers [5,6]. At present, it has achieved maturity [7,8]. The finite element approaches can be successfully applied for the smooth enough solutions of the elliptic and parabolic equations. For the equations of hyperbolic or mixed type (CFD problems at supersonic flow modes, for example), the progress in the a posteriori error analysis is limited due to the discontinuities (shock waves, contact surfaces) that may occur and migrate in the flowfield. Some computationally cheap approaches for the approximation error norm estimation [12–15] have been proposed and tested for CFD problems. However, they do not provide point-wise information on the error.

The approximation error can be described by the norm of error $||\Delta u_h||$, by the error of some scalar valuable functional (integral (drag, lift, etc. or local (pressure, heat flux))) or by the point-wise error $\Delta u_m$ ($m$ is the index of the grid point). The norm of error conventionally relates to the theoretical analysis of algorithms. Errors of the valuable functionals are of the most practical interest. On contrary, the point-wise error has a universal nature. It can be easily converted to both the error norm and the error of a valuable functional. Also, the point-wise error allows detecting the local defects of the solution, for example, the spurious oscillations that may occur at flow discontinuities (shocks, contact lines). Once detected, such oscillations can be mitigated by the reduction of the local approximation order of the algorithm or by some tuning of the artificial viscosity.

The modern approaches to the point-wise error estimations in CFD are mainly based on the defect correction [9,10] or the Richardson extrapolation (RE). The latter is recommended by modern standards [1,2]. However, the implementation of these methods can be not straightforward. For example, in the case of the defect correction, the precision is limited by the uncertainties occurring at the truncation error estimation as the high order terms are neglected. The inconvenience of the defect correction is caused by the intrusive nature of this method that implies the changes in the solver’s code. On contrary, the RE method is nonintrusive. However, for the solutions with discontinuities, the practical applicability of RE is restricted by the spatially local unknown convergence order [11,16]. The situation can be improved via generalized Richardson extrapolation (GRE) [17,18], which provides an estimate of the error field by the computation of the spatially distributed order of the error convergence. However, GRE is implemented at the cost of the great computational burden (it requires four or a greater number of the consequent mesh refinements) and demonstrates a high level of oscillations (see [17,18]).

Thus, there is a high demand for the computationally inexpensive estimators of the point-wise approximation error in the CFD domain for flows with discontinuities. In this paper, we consider the robust and computationally cheap postprocessor-based nonintrusive method [19] for the a posteriori estimation of the point-wise approximation error. This method uses a set of numerical solutions that are obtained by the different methods on the same grid. The approximation error is computed using the point-wise difference
of numerical solutions that is treated by the inverse problem, formulated in the variational statement with the Tikhonov zero-order regularization [20,21]. The numerical tests for two-dimensional compressible flows, governed by the Euler equations, are performed to illustrate the features and potentialities of the considered approach. The Edney-I and Edney-VI flow patterns [22] are used due to the availability of analytical solutions. These flow structures are composed of shock waves and contact discontinuities that provides the most severe conditions for error estimation. The sets of numerical solutions (3, 5, 13) computed by distinct numerical algorithms are used for the approximation error estimation. The results are compared with the exact error estimated as the difference of analytic and numerical solutions.

The paper is structured as follows. In Section 2 we consider the system of linear equations that relate the approximation errors and the differences of the independent numerical solutions. Section 3 recasts the previous section statement in the form of the variational inverse problem with the zero-order Tikhonov regularization. Section 4 discusses the quality of considered postprocessors from the viewpoint of the error norm estimation. The effectivity index for the error estimator (in accordance with [7]) and the relative accuracy of the error estimation is considered. The dependence of the computed error on the regularization parameter magnitude is discussed in Section 5. The test problems (governing equations and flow patterns) are presented in Section 6. The results of the numerical experiments for the approximation error estimation are presented in Section 7. The discussion is presented in Section 8. The conclusions are provided in Section 9.

2. The relation of approximation errors and differences of numerical solutions

Let us consider a set of numerical solutions (grid functions) obtained at the discretization of some system of partial differential equations. We apply a vectorization of multidimensional flowfield data obtained in numerical tests \( u_m^{(i)} \) \( (m = 1 \ldots M) \). Herein, \( M \) is the number of the grid points multiplied by the number of flow variables. The considered set of solutions is computed on the same grid by \( n \) algorithms of the dissimilar inner structure (in several cases, by schemes of different approximation order) numbered by \( i = 1 \ldots n \). The projection of an exact (unknown) solution on the grid is denoted by \( \tilde{u}_{h,m} \), the approximation error (also unknown) for the \( i \)th solution is denoted by \( \Delta u_m^{(i)} \) \( (u_m^{(i)} = \tilde{u}_{h,m} + \Delta u_m^{(i)}) \). The differences between two numerical solutions (ith and jth) may be recast in the following point-wise form:

\[
d_{ij,m} = u_m^{(i)} - u_m^{(j)} = \tilde{u}_{h,m} + \Delta u_m^{(i)} - \tilde{u}_{h,m} - \Delta u_m^{(j)} = \Delta u_m^{(i)} - \Delta u_m^{(j)}
\]

(1)

As it can be seen from Expression (1) the computable differences of numerical solutions are equal to the differences of the approximation errors and, hence, contain some information regarding the unknown errors \( \Delta u_m^{(i)} \). This allows us to estimate the approximation error \( \Delta u_m^{(i)} \) as described by Alekseev et al. [19]. On the set of \( n \) numerical solutions we obtain \( N = n \cdot (n - 1)/2 \) relations that cast the following system of linear equations:

\[
D_{ij} \Delta u_m^{(j)} = f_{i,m}.
\]

(2)
Herein $f_{i,m}$ is a vectorized form of the differences $d_{ij,m}$, $D_{ij}$ is the rectangular $N \times n$ matrix, $j = 1 \ldots n; l = 1 \ldots N$. The summation over the repeating indexes is implied starting from this expression. At first glance, this system may be resolved for $n$ that is equal (or greater) three. For the simplest case ($n = 3, N = 3$) we use the following linear system:

$$
\begin{pmatrix}
1 & -1 & 0 \\
1 & 0 & -1 \\
0 & 1 & -1
\end{pmatrix}
\begin{pmatrix}
\Delta u^{(1)}_m \\
\Delta u^{(2)}_m \\
\Delta u^{(3)}_m
\end{pmatrix} =
\begin{pmatrix}
f_{1,m} \\
f_{2,m} \\
f_{3,m}
\end{pmatrix},
$$

(3)

where

$$
\begin{pmatrix}
f_{1,m} \\
f_{2,m} \\
f_{3,m}
\end{pmatrix} =
\begin{pmatrix}
d_{12,m} \\
d_{13,m} \\
d_{23,m}
\end{pmatrix} =
\begin{pmatrix}
\Delta u^{(1)}_m - \Delta u^{(2)}_m \\
\Delta u^{(1)}_m - \Delta u^{(3)}_m \\
\Delta u^{(2)}_m - \Delta u^{(3)}_m
\end{pmatrix} =
\begin{pmatrix}
u^{(1)} - u^{(2)}_m \\
u^{(1)} - u^{(3)}_m \\
u^{(2)} - u^{(3)}_m
\end{pmatrix}.
$$

(4)

In the numerical tests, we apply Equation (2) for 3 (minimum), 5 (optimal number, according to [19]) and 13 (maximum, the number of available solutions) variables. The matrix expressions for 5 and 13 variables are omitted, herein, for brevity.

The solution using the matrix inversion $\Delta u^{(j)}_m = (D_{ij})^{-1}f_{i,m}$ cannot be obtained, since the determinant of the matrix $D_{ij}$ should be equal to zero. This degeneracy is caused by the invariance of the system (2) solution under the shift transformation $\Delta u^{(j)}_m = \Delta \tilde{u}^{(j)}_m + b$ ($\Delta \tilde{u}^{(j)}_m$ is the exact error, obtained, herein, using the analytic solution) for any $b \in (-\infty, \infty)$ due to the use of the difference of solutions as the input data. Hence, the solution to the considered problem is nonunique. A problem is called well-posed in the Hadamard sense if its solution exists, is unique, and continuously depends on the input data [20]. Otherwise, as in the considered event, the problem is called ill-posed. Most inverse problems are ill-posed and for this reason suffer from the generic instability that causes the need for some regularization. The regularized method for the numerical solution of Equation (2) that was used in the paper is presented in the following section.

### 3. The approximation error estimation

In order to determine approximation error $\Delta u^{(j)}_m$ we pose the inverse problem in the variational statement [21] with the zero-order Tikhonov regularization term:

$$
\varepsilon_m = 1/2(D_{ij}\Delta u^{(j)}_m - f_{i,m}) \cdot (D_{jk}\Delta u^{(k)}_m - f_{i,m})
+ \alpha/2(\Delta u^{(j)}_m E_{jk}\Delta u^{(k)}_m), \quad i = 1 \ldots N, \; j, k = 1 \ldots n
$$

(5)

Herein, $\alpha$ is the regularization parameter and $E_{jk}$ is the unit matrix. The problem is stated in the local (point-wise) way, separately for each $m$th element of vectorized solution (no summation on the repeating $m$).

For the search of $\Delta u^{(j)}_m$, which minimizes the functional (5), we use the gradient-based method:

$$
\Delta u^{(j),k+1}_m = \Delta u^{(j),k}_m - \tau \nabla \varepsilon_m,
$$

(6)

where $k$ is the iteration number and $\tau$ is a step length parameter.
We apply the zero-order Tikhonov regularization for the following reason. It is natural to search for solutions to the minimum shift (unavoidable) error $|b|$ (ideally, $|b| \to 0$). For this purpose, we consider the search for the minimal $L_2$ norm of $\Delta u_m^{(j)}$ (normal solution, [20]) denoted as $\delta(b_m)$ in order to stress the dependence on the shift value:

$$\min_{b_m}(\delta(b_m)) = \min_{b_m} \sum_j (\Delta u_m^{(j)})^2/2 = \min_{b_m} \sum_j (\Delta \tilde{u}_m^{(j)} + b_m)^2/2. \quad (7)$$

This relation imposes some restrictions on the magnitude of $|b_m|$. One may see from the expression $\Delta \delta(b_m) = \sum_j (\Delta \tilde{u}_m^{(j)} + b_m)\Delta b_m$ that the minimum of (7) over $b_m$ (the norm depends only on $b_m$ at constant exact errors) occurs at

$$b_m^* = -\frac{1}{n} \sum_j \Delta \tilde{u}_m^{(j)} = -\Delta \tilde{u}_m. \quad (8)$$

Thus, the shift error $b_m^*$ is bounded at the minimum of (7) and the considered estimate of the approximation error has the appearance

$$\Delta u_m^{(j)} = \Delta \tilde{u}_m^{(j)} + b_m^* = \Delta \tilde{u}_m^{(j)} - \Delta \tilde{u}_m. \quad (9)$$

With an account of (9), expression (7) can be interpreted as the minimum of the exact error dispersion (the deviation of the exact error from the mean one) on the set of solutions.

Thus, for the considered problem, the minimum of (7) (which defines the normal solution) corresponds to the shifted exact solution. The shift error (8) is equal to the mean exact error (with the opposite sign), and, consequently, the accuracy of $\Delta u_m^{(j)}$ estimation is restricted by the unknown mean error value (so, $\Delta u_m^{(j)}$ contains the irremovable error).

However, the advantage is that the assumption of $\delta(b_m)$ minimality ensures the boundedness of the irremovable error $b_m^*$. Furthermore, the magnitude of the mean exact error $b_m^*$ should not be too great. It depends on the errors magnitude and the correlation between them, and thus, can decay at the enhancement of the ensemble of solutions.

The zero-order Tikhonov regularization is often considered to be primitive and a bit obsolete, however, it provides the unique possibility to obtain the estimate by Equation (8).

4. The quality of approximation error estimates

As stated in [7], the quality of a posteriori error estimate can be described with the effectivity index that is equal to the ratio of the estimated error norm to the exact error norm (usually, the global norm is used, which is computed over all grid nodes):

$$I_{\text{eff}}^{(j)} = \frac{||\Delta u^{(j)}||_{L_2}}{||\Delta \tilde{u}^{(j)}||_{L_2}}. \quad (10)$$

One may treat the norms of the exact error and estimate error as the radii of hyperspheres $r_{\text{exact}} = ||\Delta \tilde{u}^{(j)}||_{L_2}$ and $r_{\text{est}} = ||\Delta u^{(j)}||_{L_2}$ in the space of discrete solutions. Then the relation $I_{\text{eff}}^{(j)} \geq 1$ means that the hypersphere, containing the projection of exact error onto grid, belongs to the hypersphere defined by the estimator. Thus, in order to provide a
reliable estimation, the effectivity index should be greater than the unit. Also, the estimate should be not too pessimistic, so the value of the effectivity index should be not too great. For the finite element applications in the domain of elliptic equations (as usual, highly regular), the acceptable range of the effectivity index, according to [7], is $1 \leq I_{\text{eff}}^{(j)} \leq 3$. It should be noted that the upper bound is not strictly defined and may be problem dependent. The solutions, considered herein, contain discontinuities (shear lines, shock waves), so the acceptable range of the effectivity index may be greater and corresponding bounds should be established by an additional analysis (for example, from the acceptable errors of the valuable functionals, which are used in applications).

Taking into account expression (9), the effectivity index for the $j$th solution may be recast as

$$I_{\text{eff}}^{(j)} = \frac{||\Delta \tilde{u}^{(j)} - \Delta \tilde{u}||_{L_2}}{||\Delta \tilde{u}^{(j)}||_{L_2}}.$$ (11)

Considering the relation of $\Delta \tilde{u}^{(j)}$ and $\Delta \tilde{u}$, one can obtain the effectivity index value belonging to the range $1 - ||\Delta \tilde{u}||_{L_2}/||\Delta \tilde{u}^{(j)}||_{L_2} \leq I_{\text{eff}}^{(j)} \leq 1 + ||\Delta \tilde{u}||_{L_2}/||\Delta \tilde{u}^{(j)}||_{L_2}$. Thus, an underestimation of the error ($I_{\text{eff}}^{(j)} \leq 1$) is feasible if $\Delta \tilde{u}^{(j)}$ and $\Delta \tilde{u}$ have the same sign. For the case of $\Delta \tilde{u}^{(j)} \approx \Delta \tilde{u}$ the effectivity index is close to zero and the error estimation fails. Thus, some parts of the error estimates may have a low effectivity index. The expansion of the solutions set and the analysis of the distances between solutions (the selection of most remote solutions) may improve the value of the effectivity index for certain solutions. The best estimation $I_{\text{eff}}^{(j)} \to 1$ is possible, if $b = -\Delta \tilde{u} \to 0$. However, this condition can be violated due to the correlation of error in the vicinity of discontinuities that is caused by the artificial viscosity or limiters.

The effectivity index is not intuitively lucid, since the negative and positive deviations from the unit have quite different meanings. For this reason, the relative accuracy of the error estimation

$$I_{\text{rel}}^{(j)} = \frac{||\Delta u^{(j)} - \Delta \tilde{u}^{(j)}||_{L_2}}{||\Delta \tilde{u}^{(j)}||_{L_2}}$$ (12)

can be used as another quality indicator. This indicator is more intuitively transparent since the dependence of the estimated uncertainty on the error is monotonous ($I_{\text{rel}}^{(j)} \to 0$ for the precise solution). The values of indicators (10) and (12) are related by $I_{\text{rel}}^{(j)} \leq 1 + I_{\text{eff}}^{(j)}$ due to the triangle inequality.

5. The impact of regularization parameter on the accuracy of error estimation

The solution $\Delta u_m^{(j)}(\alpha)$, providing the minimum of the functional (5), depends on the regularization parameter $\alpha$. Due to the ill-posedness of this problem, $\Delta u_m^{(j)}(\alpha)$ is not bounded at $\alpha = 0$, so it is not acceptable. One may observe the trivial solution $\Delta u_m^{(j)}(\alpha) \to 0$ at $\alpha \to \infty$ that is not acceptable also. According to the practice of regularization [21], there exists a range of $\alpha$ with the weak dependence of the solution $\Delta u_m^{(j)}(\alpha)$ on $\alpha$. In this range
Figure 1. The dependence of the functional (5), mean error (8) and the effectivity index (10) on the regularizing coefficient (in logarithmic scale).

$\Delta u_m^{(j)}(\alpha)$ has the minimum deviation from the exact value $\Delta \tilde{u}_m^{(j)}$ and is adopted, herein, as the solution of the inverse problem.

Figures 1 and 2 illustrate the influence of the regularization coefficient on the error for the set of three solutions $\Delta u^{(j)}$, $j = 1, 2, 3$. The regularization coefficient runs the values from the interval $\alpha = (10^{-10}, 1)$, exact errors $\Delta \tilde{u}^{(j)}$ are equal to $(1, -2, 3)$ ($\Delta \tilde{u} = 2/3$). Figure 1 presents the functional $\varepsilon(\alpha)$ (Expression (5)), mean sum of errors $\eta = \sum_{i=1}^{3} |\Delta u^{(i)}(\alpha)|/3$ and the averaged effectivity index

$$I_{\text{eff}} = \frac{\left(\sum_{i=1}^{3} (\Delta u^{(i)}(\alpha))^2\right)^{1/2}}{\left(\sum_{i=1}^{3} (\Delta \tilde{u}^{(i)})^2\right)^{1/2}}$$

in dependence on $\log(\alpha)$. One may see from the Figure 1 that the solution diverges at $\alpha \leq 10^{-7}$. The quality of results also deteriorates for $\alpha \geq 1$. The solution weakly depends on the regularization coefficient in the range $\alpha \in (10^{-6}, 10^{-1})$. Thus, a regularized solution from this range may be accepted as the solution of the considered inverse problem.

Figure 2 provides the error estimates in dependence on the magnitude of the regularizing coefficient $\alpha$ and the comparison with the exact errors. The systematic shift $b \approx -0.7$ is observed, which weakly depends on the regularization coefficient and the initial guess. The shift is close to the expected value $b = -\frac{1}{3} \sum_i \Delta \tilde{u}^{(i)} = -2/3$ that corresponds to the average of the exact errors $\Delta \tilde{u} = 2/3$ (with opposite sign) in this test.
6. Test problems

In order to verify the above analysis and to illustrate the essence of the proposed approach, we perform a set of numerical experiments. We consider the approximation error estimation for the tests problems governed by two-dimensional compressible Euler equations that follow

\[
\frac{\partial \rho}{\partial t} + \frac{\partial (\rho U^k)}{\partial x^k} = 0; \quad \frac{\partial (\rho U^i)}{\partial t} + \frac{\partial (\rho U^k U^i + P \delta_{ik})}{\partial x^k} = 0; \quad \frac{\partial (\rho E)}{\partial t} + \frac{\partial (\rho U^k h_0)}{\partial x^k} = 0. \tag{14-16}
\]

Here \( U^1 = U, U^2 = V \) are the velocity components, \( h_0 = (U^2 + V^2)/2 + h, \ h = (\gamma/(\gamma - 1))(P/\rho) = \gamma e, \ e = RT/(\gamma - 1), \ E = (e + \frac{1}{2}(U^2 + V^2)) \) 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 steady flow patterns, engendered by the interaction of shock waves of I and VI kinds according to Edney classification \[22\], were used as the test problems due to the availability of analytic solutions (defined by the Rankine–Hugoniot relations) and the high level (observability) of approximation errors. The exact error is obtained by the subtraction of the numerical solution from the projection of the analytic one on the corresponding grid.

Figure 3 presents the spatial density distribution for the Edney-I flow structure \((M = 4, \alpha_1 = 20^\circ \text{ and } \alpha_2 = 15^\circ)\). Figure 4 presents the density distribution for the Edney-VI flow structure \((M = 4, \text{ two consequent flow deflection angles } \alpha_1 = 10^\circ, \alpha_2 = \)
15°). Both fields provided in Figures 3 and 4 are computed using the method by Courant–Isaacson–Rees [23,24].

**7. The results of numerical tests**

The set of 13 numerical solutions obtained by 10 different algorithms [25–38] and (for methods, described in [30,31,32]) using several variants of the artificial viscosity is applied in tests in different combinations from the minimal one (3 solutions) to maximal (13 solutions).
The methods and their notations are listed below:

- First-order algorithm by Courant–Isaacson–Rees [23,24] marked as CIR.
- Second-order MUSCL [26] algorithm based on the approximate Riemann solver by Sun and Katayama [25] and marked as AUFS.
- Second-order MUSCL [26] algorithm based on the approximate Riemann solver by Toro [27].
- Second-order algorithm based on the relaxation approach [28,29].
- Second-order algorithm based on the MacCormack [30] scheme:
  1. without artificial viscosity,
  2. with second-order artificial viscosity with the viscosity coefficient $\mu = 0.01$ and $\mu = 0.002$,
  3. with fourth-order artificial viscosity ($\mu = 0.01$),
- Second-order algorithm based on the ‘two step Lax-Wendroff’ [31,32], with the artificial viscosity of the second-order ($\mu = 0.01$),
Table 1. The effectivity index (Equation (10)) for the Edney-I test.

|                | ICIR | IAUFS | ord3 | ord4 | W5  |
|----------------|------|-------|------|------|-----|
| Three solutions| 0.48 | 0.28  | –    | 0.56 | –   |
| Other solutions| 0.41 | 0.31  | 0.35 | –    | –   |
| Five solutions | 1.13 | 1.71  | 1.74 | 1.66 | 1.1 |
| 13 solutions   | 0.58 | 1.09  | 0.55 | 0.51 | 0.61|

Table 2. The relative accuracy of the error estimation (Equation (12)) for the Edney-I test.

|                | ICIR rel | IAUFS rel | ord3 rel | ord4 rel | W5  rel |
|----------------|----------|-----------|----------|----------|-------|
| Three solutions| 0.69     | 1.10      | –        | 0.99     | –     |
| Other solutions| 0.71     | 1.13      | 1.14     | –        | –     |
| Five solutions | 1.29     | 2.03      | 2.06     | 1.84     | 1.21  |
| 13 solutions   | 0.67     | 1.05      | 1.07     | 1.84     | 1.12  |

- Third-order algorithm based on the modification of the Chakravarthy–Osher scheme [33,34], marked as ord3,
- Third-order algorithm WENO [35–37],
- Fourth-order algorithm [38] marked as ord4,
- Fifth order algorithm WENO [35–37] marked as W5.

The numerical tests are performed using the uniform grids of 100 × 100 and 400 × 400 nodes. The results are qualitatively similar and their behaviour does not depend on the grid size, so only data for 100 × 100 grid are provided for the illustrations.

The inverse problem is solved numerically and results are compared with the above discussed exact error for Edney-I and Edney-VI flow structures. We minimize functional (5) at every grid point. The gradient of the functional is obtained by direct numerical differentiation. The regularization coefficient value $\alpha = 10^{-3}$ is used in most calculations. The neighbouring values ($10^{-2}$ and $10^{-4}$) of the regularization coefficient were used in several tests for the check with practically the same results.

We present the data for error of density $\Delta \rho^{(i)}$ only since the results for other flow variables are similar.

Figure 5 presents the exact error for Edney-I flow computed by the first-order scheme [23], while Figure 6 presents the results of the IP-based estimation of this error.

Table 1 provides the effectivity index of error estimation $I^{(j)}_{\text{eff}} = ||\Delta \rho^{(j)}||_{L_2}/||\Delta \tilde{\rho}||_{L_2}$ (Equation (10)) for the Edney-I test computed for the different sets of solutions (three solutions are used in two different combinations). Index $j$, which marks the numerical schemes, is replaced by the above described notations.

The relative accuracy $I^{(j)}_{\text{rel}} = ||\Delta \rho^{(j)} - \Delta \tilde{\rho}||_{L_2}/||\Delta \tilde{\rho}||_{L_2}$ (Equation (12)) is provided in Table 2. One may see that the error estimations for the first-order method [23] are (formally) superior due to the relatively great magnitude of the error.

Figure 7 presents the exact error and may be compared with Figure 8, which presents the IP-based estimation of this error for the Edney-VI flow pattern computed by the first-order scheme [23].
Figure 6. IP-based estimation of density error for Edney-I flow.

Table 3. The effectivity index (Equation (10)) for the Edney-VI test.

|          | $\eta_{\text{eff}}$ | $\eta_{\text{d3,eff}}$ | $\eta_{\text{d4,eff}}$ | $\eta_{\text{WS,eff}}$ |
|----------|----------------------|------------------------|------------------------|------------------------|
| Three solutions | 2.29                 | 3.46                  | 3.54                  | --                     |
| Other three solutions | 2.32             | 3.45                  | --                     | 3.29                   |
| Five solutions | 0.43                 | 0.36                  | 0.40                  | 0.62                   | 0.46             |
| 13 solutions | 0.64                 | 0.97                  | 0.98                  | 0.89                   | 0.61             |

Table 3 provides the effectivity index (Equation (10)) for several sets of numerical solutions for the Edney-VI test. The relative accuracy (Equation (12)) is provided by Table 4 for these solutions.

For uncorrelated exact errors, one may expect the improvement of the error estimation quality as the number of solutions is enhanced since $b_m = - (1/n) \sum_j^n \Delta u_m^{(j)}$. The corresponding values of the effectivity index (10) and the relative accuracy (12) are provided in Tables 1–4 for the sets of 3, 5 and 13 solutions. For the Edney-I test, the reliability (Table 1) and the accuracy (Table 2) of estimates demonstrate no correlated behaviour.
when the number of solutions increases. The accuracy of the estimate increases while reliability deteriorates. For the Edney-VI test, both indexes reveal the improvement of the error estimation quality with the increase in the number of used solutions. Both the reliability (Table 3) and the accuracy (Table 4) of estimates increase as the number of solutions is enhanced. The uncorrelated and non-monotonic behaviour of the reliability (10) and the accuracy (12) indexes at the solutions set expansion can be caused by the non-monotonic
Table 4. The relative accuracy of the error estimation (Equation (12)) for the Edney-VI test.

|                | \( \rho_{CR} \) rel | \( \rho_{IF} \) rel | \( \rho_{CR} \) rel | \( \rho_{IF} \) rel | \( \rho_{IV} \) rel |
|----------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Three solutions| 2.38                 | 3.62                 | 3.70                 | –                    | –                    |
| Other three solutions| 2.38 | 3.61 | – | 3.39 | – |
| Five solutions | 0.73                 | 1.12                 | 1.13                 | 1.03                 | 0.71                 |
| 13 solutions   | 0.61                 | 0.17                 | 0.19                 | 0.43                 | 0.63                 |

Figure 9. The comparison of the density error, estimated by the inverse problem, with the exact error in zone before shocks crossing (Edney-I).

convergence to zero of the sum of exact errors \( b_m \) due to the absence of errors ordering over the norm. The addition of an inaccurate solution can spoil the results obtained for the set of relatively precise solutions. Thus, in some cases, the optimal balance of computational efforts and the accuracy of results occur for the minimal set of solutions (three).

Figures 5 and 6 provide the comparison of the exact error with the results of the IP-based error estimation for Edney-I flow. The corresponding analysis for the Edney-VI flow pattern may be performed using Figures 7 and 8. The errors in the vicinity of shocks have a wave-like shape with the positive and negative half-waves, which is quite natural for monotonous smoothing of the stepwise discontinuities. From the qualitative viewpoint, Figures 5–8 show that the estimates of the error demonstrate a similar spatial structure if compared with the exact error. The quantitative comparison of the estimates and exact errors can be obtained using Figures 9 and 10, which present the pieces of the vectorized grid function of density error obtained by the inverse problem in comparison with the exact error for the Edney-I test. The index along the abscissa axis is related to the indexes along \( X (k_X) \) and \( Y (k_Y) \) as \( m = N_X(k_X - 1) + k_Y \).

The periodical jump in the flow parameters corresponds to the transition through the shock waves, herein two transitions are presented. Figure 9 corresponds to the domain before shock crossing and Figure 10 illustrates the flow structure past shocks crossing.
8. Discussion

Formally, the IP-based method is less accurate compared to the Richardson extrapolation. This occurs due to the presence of the irremovable error, which is proportional to the approximation error averaged over the set of used solutions (Equation (9)). However, the generalized Richardson extrapolation, which may be correctly used for the shocked flows, suffers from the oscillations due to the variable convergence order that results in severe point-wise errors [17,18]. Contrary to the generalized Richardson extrapolation, the IP-based method is more robust. Details and numerical results of the direct comparison between the IP-based method and the Richardson extrapolation are presented in [39]. Additionally, the IP-based postprocessor is more computationally inexpensive, since it applies the computations without a mesh refinement (on the same grid).

The results in [19], obtained for the supersonic flows over the cone, demonstrate the successful error estimation with three or five solutions, computed by the different algorithms of the same (second) order and a certain increase of the quality of results at the expansion of the set. This paper provides the successful error estimation for the flows engendered by the interaction of the shock waves, if at least three numerical solutions, obtained by methods of different approximation orders, are available. The extension of the solutions number increases the reliability and accuracy of the results, however, in a nonmonotonous manner.

Despite the above-considered tests being selected from the CFD domain, the IP-based method can be applied for the numerical solution of any PDE system without the loss of generality, since no assumptions, which are specific for CFD, were made in this work.

9. Conclusions

The set of numerical solutions obtained using dissimilar algorithms contains information regarding the point-wise approximation error. This error can be estimated from the
differences of the solutions using the inverse problem, posed in the variational statement with the zero-order Tikhonov regularization. Since the considered problem is underdetermined, the results contain the irremovable error, which is equal to the exact error, averaged over the ensemble of solutions. The numerical tests demonstrate the feasibility for the estimation of the pointwise approximation error with the acceptable accuracy for the compressible inviscid flows.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Russian Science Foundation [grant number 18-11-00215].

ORCID

A. K. Alekseev http://orcid.org/0000-0001-8317-8688

References

[1] Guide for the Verification and Validation of Computational Fluid Dynamics Simulations, American Institute of Aeronautics and Astronautics, AIAA-G-077-1998, Reston, VA; 1998.
[2] Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer, ASME V&V 20-2009.
[3] Skeel RD. Thirteen ways to estimate global error. Numer Math. 1986;48:1–20.
[4] Roy Ch J, Oberkampf WL. Comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. CMAME; 2011;200: (25–28): 2131–2144.
[5] Prager W, Synge JL. Approximation in elasticity based on the concept of function spaces, quart. Appl Math. 1947;5:241–269.
[6] Babuska I, Rheinboldt W. A posteriori error estimates for the finite element method. Int J Numer Methods Eng. 1978;12:1597–1615.
[7] Repin SI. A posteriori estimates for partial differential equations. Vol. 4. Berlin: Walter de Gruyter; 2008.
[8] Ainsworth M, Oden JT. A posteriori error estimation in finite element analysis. New York (NY): Wiley-Interscience; 2000.
[9] Roy CJ, Raju A. Estimation of discretization errors using the method of nearby problems. AIAA J. 2007;45(6):1232–1243.
[10] Linss T, Kopteva N. A posteriori error estimation for a defect-correction method Applied to convection-diffusion problems. Int J Numer Anal Modeling. 2009; 1(N. 1):1–16.
[11] Banks JW, Hittinger JAF, Woodward CS. Numerical error estimation for nonlinear hyperbolic PDEs via nonlinear error transport. CMAME. 2012;213:1–15.
[12] Alekseev AK, Bondarev AE, Navon IM. (2017). On triangle inequality based approximation error estimation. arXiv:1708.04604.
[13] Alekseev AK, Bondarev AE. Estimation of the distance between true and numerical solutions. Comput Math Math Phys. 2019;59(6):857–863.
[14] Alekseev AK, Bondarev AE, Kuvshinnikov AE. Verification on the ensemble of independent numerical solutions. In: Rodrigues JMF, Cardoso PJS, Monteiro J, et al. editors. Computational science – ICCS 2019. ICCS 2019. Lecture Notes in Computer Science. Cham: Springer, 2019; Vol. 11540, p. 315–324.
[15] Alekseev AK, Bondarev AE, Kuvshinnikov AE. On uncertainty quantification via the ensemble of independent numerical solutions. J Comput Sci. 2020;42:10114.
[16] Carpenter MH, Casper JH. Accuracy of shock capturing in two spatial dimensions. AIAA J. 1999;37(9):1072–1079.

[17] Alexeev AK, Bondarev AE. On some features of Richardson extrapolation for compressible inviscid flows. Math Montisnigri. XL. 2017: 42–54.

[18] Roy CJ. Grid convergence error analysis for mixed–order numerical schemes. AIAA J. 2003; 41(4):595–604.

[19] Alekseev AK, Bondarev AE and Kuvshinnikov AE. A posteriori error estimation via differences of numerical solutions. In: Krzhizhanovskaya VV, Závodszky G, Lees MH, et al., editors. ICCS 2020. Lecture Notes in Computer science, Amsterdam: Springer Nature, 2020; Vol. 12143, p. 508–519.

[20] Tikhonov AN, Arsenin VY. Solutions of ill-posed problems. Washington (DC): Winston and Sons; 1977.

[21] Alifanov OM, Artyukhin EA, Rumyantsev SV. Extreme methods for solving ill-posed problems with applications to inverse heat transfer problems. New York (NY): Begell House; 1995.

[22] Edney B. Effects of shock impingement on the heat transfer around blunt bodies. AIAA J 1968;6(1):15–21.

[23] Courant R, Isaacson E, Rees M. On the solution of nonlinear hyperbolic differential equations by finite differences. Comm Pure Appl Math. V. 1952;5:243–255.

[24] Kulikovskii AG, Pogorelov NV, Semenov AY. Mathematical aspects of numerical solution of hyperbolic systems. Monographs and Surveys in Pure and Applied Mathematics. 2001; Vol. 188, Boca Raton (FL): Chapman&Hall/CRC.

[25] Sun M, Katayama K. An artificially upstream flux vector splitting for the Euler equations. JCP. 2003;189:305–329.

[26] van Leer B. Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov’s method. J Comput Phys. 1979;32:101–136.

[27] Toro E., Riemann solvers and numerical methods for fluid dynamics. Berlin: Springer; 2009.

[28] Trac H, Pen U-L. (2002). A primer on Eulerian computational fluid dynamics for astrophysics, arXiv:0210611v2.

[29] Jin S, Xin Z. The relaxation schemes for systems of conservation laws in arbitrary space dimensions. Commun Pure Appl Math. 1995;48:235–276.

[30] MacCormack RW. The effect of viscosity in hypervelocity impact cratering. AIAA Paper. 1969; 69–354.

[31] Lax PD, Wendroff B. Difference schemes for hyperbolic equations with high order of accuracy. Comm Pure Appl Math., 1964;17:381–379.

[32] Richtmyer RD. A survey of difference methods for non-steady Fluid dynamics. NCAR Tech Notes. 1963: 63–62.

[33] Osher S, Chakravarthy S. Very high order accurate TVD schemes. ICASE Report N. 1984;84-114:229–274.

[34] Lin CT, Yeh JY, and Chen JY, High resolution finite volume scheme for the quantum hydrodynamic equations, JCP. 228. 5. 2009 1713–1732.

[35] Johnsen E, Colomius T. Implementation of WENO schemes in compressible multicomponent flow problems. J Comput Phys. 2006;219(2):715–732.

[36] Antoniou AS, Karantasis KI, Polychronopoulos ED, et al. Acceleration of a finite-difference WENO scheme for large-scale simulations on many-core architectures. AIAA. 2010: 0525.

[37] Ha Y, Gardner CL, Gelb A, et al. Numerical simulation of high Mach number astrophysical jets with radiative cooling, J. Scientific Computing. 2005;24(1):29–44.

[38] Yamamoto S, Daiguji H. Higher-order-accurate upwind schemes for solving the compressible Euler and Navier–Stokes equations. Comput Fluids. 1993;22:259–270.

[39] Alekseev AK, Bondarev AE, Kuvshinnikov AE. 2021. A comparison of the Richardson extrapolation and the approximation error estimation on the ensemble of numerical solutions. In: Paszynski M, Kranzlmueller D, Krzhizhanovskaya VV , et al. editors. Computational Science – ICCS 2021. Lecture Notes in Computer science. Krakow: Springer, 2021; Vol. 12747, p. 554–566.