A historical oversight: Vladimir P. Kolgan and his high-resolution scheme

Bram van Leer
Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109-2140, USA

ARTICLE INFO

Article history:
Received 12 October 2010
Received in revised form 1 December 2010
Accepted 16 December 2010
Available online 25 December 2010

Keywords:
History of Computational Fluid Dynamics
Euler equations
Godunov-type methods
Second-order methods
High-resolution schemes
Non-oscillatory schemes
Limiters

ABSTRACT

The Russian landmark paper “Application of the principle of minimizing the derivative to the construction of finite-difference schemes for computing discontinuous solutions of gas dynamics” by Vladimir P. Kolgan is discussed in a historical–technical perspective. The 1972 paper already featured a Godunov-type scheme for the Euler equations with second-order spatial accuracy, and a limiter to make it monotonicity-preserving. The work remained little known within and completely unknown outside the USSR, largely because of Kolgan’s untimely death in 1978. To honor Kolgan’s contribution to CFD, an English translation of his paper is included in this issue of the Journal of Computational Physics. Facts about Kolgan’s life are presented here, and an appraisal is given of the originality of his ideas.

© 2011 Published by Elsevier Inc.

1. Introduction

This article is meant as a historical and technical commentary to the article immediately following it: the 1972 research paper “Application of the principle of minimizing the derivative to the construction of finite-difference schemes for computing discontinuous solutions of gas dynamics” by Vladimir P. Kolgan. The article originally appeared in Russian in the TsAGI Research Notes 3 (1972) 68–76, and is a true milestone in the history of CFD. It was translated by Konstantin Kabin and Valeriy Tenishev for publication in the Journal of Computational Physics. The reasons for reprinting this article in the English language will be made clear in the present article.

2. The birth of high-resolution scheme

Computational Fluid Dynamics (CFD) owes its present success to a large degree to the advent of the so-called high-resolution (HR) schemes [1] 40 years ago. Such schemes are at least second-order accurate in regions where the flow solution is smooth, while capturing discontinuities as narrow, monotone structures.

For an appreciation of the problem of designing a high-resolution scheme for the Euler equations it is useful to first consider modeling the linear advection of a step function. Here we immediately run into a famous theorem included by Godunov in his 1959 paper [2]: if an advection scheme preserves the monotonicity of the solution it is at most first-order accurate. This result could discourage anyone attempting to improve advection schemes; fortunately, there is a way to circumvent it. In the proof of this theorem it is tacitly assumed that the linear advection equation is approximated by a linear discretization; once nonlinear discretizations are admitted the theorem no longer stands and high-resolution schemes become possible.

DOI of original article: 10.1016/j.jcp.2010.12.033
E-mail address: bram@umich.edu

0021-9991/$ - see front matter © 2011 Published by Elsevier Inc.
The realization that Godunov’s theorem could be circumvented came at the start of the 1970s, when, within the span of one year, three independent approaches were launched for the construction of oscillation-free higher-order advection schemes.


These approaches, while more or less distinct, are similar in their use of nonlinearity to prevent numerical oscillations in regions of strongly varying solution gradients, a theme that has persisted in hyperbolic method development ever since. It is instructive to follow the careers of their originators.

Boris and collaborators published their views on nonlinear limiters and transport schemes in a series of 4 papers in the Journal of Computational Physics (JCP) [6–9], coining the term “Flux-Corrected Transport” (FCT). Boris’s seminal contribution to CFD is evident from the reprinting of his 1973 journal paper on FCT in the 30th anniversary issue of JCP [10], dedicated to the journal’s most-cited papers. Among the honors Boris has received is the AIAA’s 2005 Fluid Dynamics Award for his contributions to and applications of CFD.

Van Leer published his findings in a series titled “Towards the Ultimate Conservative Difference Scheme:” with the conference paper as first installment and four more papers in JCP [11–14], unfolding non-oscillatory higher-order Godunov-type schemes. The last, keystone paper of the series, describing the “MUSCL” scheme (for: Monotonic Upstream Scheme for Conservation Laws), also found a place in JCP’s 30th-anniversary issue [10]. For his contributions to CFD and to aerospace-engineering education Van Leer received the AIAA Fluid Dynamics Award in 2010.

Kolgan published his initial findings in the Technical Notes of TsAGI in 1972 [5], but produced no follow-up papers, though further refinement was clearly desirable. Notably, improving the temporal accuracy from the first to the second order would have been a logical step forward. Though the 1972 paper shows Kolgan was several years ahead of Van Leer, he was eventually overtaken by the latter. Kolgan’s work, now largely of historical interest, is essentially unknown outside Russia, and known even in Russia only in limited circles outside TsAGI.

Why did Kolgan fail to build on his initial research success and collect the credit he clearly deserved? This question will be addressed in the next section.

3. An interrupted life

The most obvious reason for Kolgan’s lack of influence is his untimely death: he succumbed to lung cancer in 1978, at the age of 37. At that time the final papers by Boris’ group and by Van Leer had yet to appear.

Other circumstances may have contributed, in particular, the relative isolation of Kolgan’s place of employment. The research environment at TsAGI was industrial rather than academic. Dr. Vladimir Yumashev of TsAGI, a former collaborator of Kolgan, sent me an all-too-familiar story about the lack of communication between those in academia, interested in proving theorems and building idealized theories, and those in industry interested in solving practical, realistic problems. “Academic people usually had more publications and more international contacts. Industry people usually had more technical reports and fewer publications, and few international contacts because of the paranoid security policy.”

Circumstances, though, do not completely define a person’s actions. Yumashev correctly points out: “Besides, one should take into account personal interests, ambitions, and similar human factors”. It is likely that Kolgan was more interested in solving problems than in becoming famous. The information below gathered about his life and work appears to support this conclusion.

Vladimir Pavlovich Kolgan (Fig. 1) was born on October 16, 1940, in the ancient little town of Ostrogozhsk, in the Voronezh region. He graduated from the Moscow Institute for Physics and Technology (MPhTI, also known as PhysTech) in 1964 and obtained a Ph. D. in 1972, while employed at TsAGI. His final position at TsAGI was Senior Research Scientist, the highest non-administrative job at that time. His colleague Yumashev describes him as “a wonderful man: kind and tolerant with respect to his partners. He looked joyful and optimistic till his last days.”

The research field of V. P. Kolgan included 2D unsteady flows with strong shock waves. Just as for Boris and Van Leer, the need for greater computational accuracy and resolution drove his effort in developing a new method. Kolgan, however, remained focused on solving flow problems; after the 1972 paper detailing the new scheme, Kolgan’s numerical results went into technical reports of limited distribution. There is no evidence that he ever bothered to contact the mathematical establishment, Godunov in particular, in order to increase the impact of his discovery.

Kolgan’s extension of Godunov’s method was met by CFD people in TsAGI quite positively. It was implemented by several CFD teams in a number of codes, some of which are still in service.

1 In my historical review of HR schemes in [1], which appeared in 1997, I still wrote “two approaches,” not being aware of Kolgan’s work.
In 1977–1978 Kolgan extended his method to a 3D Euler solver on the platform of a new French mainframe that greatly exceeded the computational power available at TsAGI before that time. Vladimir Yumashev participated in this project; after Kolgan’s death, Yumashev and others continued to use the solver.

Outside of TsAGI, Kolgan’s method was recognized within the aeronautical community. For example, it was used at the Central Institute for Aviation Motors (TsNIIMash) by A. N. Kraiko, M. Ya. Ivanov and their colleagues and students. Eventually, A. V. Rodionov [15,16] from TsNIIMash introduced a predictor–corrector strategy making the method second-order accurate in time, but by then it was 1987, years after the publication of Van Leer’s fully-second-order MUSCL scheme (1979, [14]), Hancock’s predictor–corrector implementation of it (1982, [17]), and the advent of multi-stage time-marching in CFD [18].

In the summer of 1978 Vladimir Kolgan said good-bye to his colleagues, joyful as usual, and left for a vacation in his native town, never to return. He did not suffer long; his struggle with death happened in the span of a week. He died on July 28, 1978.

4. The 1972 TsAGI paper

In this issue of JCP you will find an English translation of Kolgan’s 1972 TsAGI paper [5], describing his non-oscillatory Godunov-type method of second spatial and first temporal order of accuracy. Its title is: “Application of the principle of minimizing the derivative to the construction of finite-difference schemes for computing discontinuous solutions of gas dynamics.” The paper is straightforward; I will restrict myself to pointing out the significance of Kolgan’s findings.

I have identified four original contributions in Kolgan’s paper.

1. Sub-cell gradient. Introducing a sub-cell finite-difference-based gradient in order to raise the spatial accuracy of Godunov’s method was Kolgan’s first original contribution in this paper. Although it seems trivial now, no one had come up with this idea since the publication of Godunov’s [2] first-order scheme, in particular, neither Godunov himself or any of his collaborators or students.

2. Gradient limiting. Kolgan’s second original contribution was to limit this gradient so as to avoid numerical oscillations. This makes the scheme nonlinear even for a linear advection equation, thus circumventing Godunov’s barrier theorem (see Section 1).

3. Minmod limiter. Kolgan’s third original contribution was the particular limiter he adopted: a primitive form of what we now call the minmod limiter, after Osher and Chakravarthy [19]. Of the finite-difference gradients (forward and backward) that can be constructed with aid of a cell average and the two neighboring-cell averages, Kolgan selects the one with the minimum modulus:

\[
(\Delta u)_{\text{lim}} = \begin{cases} 
\Delta u & \text{if } |\Delta u| < |\Delta u|, \\
\Delta u & \text{otherwise.}
\end{cases}
\]

(1)
This limited gradient value has an undesirable property: it switches sign at the point where \( \Delta u \) and \( \Delta u \) are equal and opposite. Thus, it is a discontinuous function of its arguments. The problem is avoided in the standard minmod limiter by the introduction a zero-branch. If the two arguments have opposite signs, the limiter returns a zero value for the gradient:

\[
(\Delta u)_{\text{lim}} = \begin{cases} 
\min(|\Delta u|,|\Delta u|) \text{sgn}(\Delta u) & \text{if } \text{sgn}(\Delta u) = \text{sgn}(\Delta u), \\
0 & \text{otherwise.} 
\end{cases}
\] (2)

Kolgan found this technique so important that he included a description, “minimizing the derivative,” in the title of his paper, suggesting a fundamental principle underlying this choice. This emphasis may have had its draw-back: I suspect it kept Kolgan from realizing that the true principle is “limiting the derivative,” that there exists an infinity of choices of limiters, and that minmod is actually a rather crude one.

4. Nonlinear stability of “Forward Euler” time-marching. The least inspired part of Kolgan’s paper is the time-marching scheme he chose, viz., the simplest possible scheme known as “Forward Euler”; it is only first-order accurate in time. It is a bad match to the second-order spatial discretization, not only in view of accuracy but also in view of stability. Kolgan understood that achieving higher temporal accuracy would necessitate solving what is now called a Generalized Riemann Problem [20], one in which the initial values are not uniform but linear on either side of the discontinuity, but he was not prepared to do so. Ultimately, full and approximate formulas for the time derivative of Lagrangean Riemann fluxes were derived by Van Leer ([14]; soon enough, though, the appearance of Hancock’s [17] predictor–corrector version of MUSCL made it possible to avoid the use of those complex formulas.

Let us consider the matter of stability up close. For a linear ordinary differential equation (ODE) of the form

\[ u(t) = \lambda u, \quad \lambda \in \mathbb{C}, \] (3)

Forward Euler reads:

\[ u(t^{n+1}) = u(t^n) + \Delta t \lambda u(t^n). \] (4)

The stability domain of this method in the complex plane is given by

\[ |z + 1| \leq 1, \] (5)

\[ z = \lambda \Delta t. \] (6)

i.e., a circular disc in the negative-real plane, touching the imaginary axis in the origin. This condition may also be applied to a semi-discretization of a partial differential equation (PDE); in that case \( \lambda \) represents any eigenvalue of the spatial operator, obtainable by a Fourier analysis. The exact spatial eigenvalues of a hyperbolic PDE lie on the imaginary axis, hence, outside the stability domain (6). A spatial discretization of second- or higher-order accuracy has eigenvalues that lie close to the imaginary axis; the whole spectrum can be pulled into the stability domain only by multiplication with a vanishingly small \( \Delta t \). In other words, the combination of a second-order spatial discretization with Forward Euler time-marching is unstable, according to a linear stability analysis.

But Kolgan’s spatial discretization is essentially nonlinear and thus circumvents linear theory. In his paper Kolgan proved that his space–time method, applied to a 1D advection equation, is monotonicity-preserving, hence stable, for CFL numbers not exceeding \( \frac{2}{3} \). In demonstrating that a nonlinear limiter can stabilize a linearly unstable scheme, Kolgan was twelve years ahead of Osher and Chakravarthy [19], who presented monotonicity/stability bounds for Forward Euler combined with a two-parameter family of minmod-type limiters.

The formulas of Osher and Chakravarthy indicate that Kolgan’s scheme should be monotonicity-preserving and stable for CFL numbers up to \( \frac{2}{3} \), provided that the standard minmod limiter is used; apparently the limiter’s zero branch yields extra stability. This is not surprising for odd–even oscillatory data it makes the scheme reduce to Godunov’s first-order method.

Kolgan did mention that the limit of \( \frac{2}{3} \) could be shown with a “more involved analysis,” but he did not provide the latter. I speculate that in this omitted analysis he did use the minmod limiter with the zero branch.

While numerically interesting, a purely nonlinear stabilizing mechanism does not make a great marching method. The limiter turns on and off at odd places and times for the sake of maintaining stability; its actions may show up in the form of little plateaus, steps or staircases in the numerical solution. In Figs. 1 and 2 of the Kolgan paper, showing pressure and density, respectively, of a shock-tube solution, one can clearly see a step at the head of the expansion wave.

It must be mentioned here that some of Boris’s limiters have this same side-effect; see, for instance, the square-wave profile rendered by Reversible FCT in Fig. 3 of [8], which shows similar steps.

In conclusion, Kolgan’s short article is rich in ideas, supported by rigorous analysis. Of the two shortcomings, viz., the crudeness of his limiter and the use of Forward Euler for time-marching, the former may already have been mended by the author at the time of publication, while the latter led him to discover a nonlinear stabilizing mechanism that was not to be rediscovered until the 1980s.
5. In search of Kolgan: a personal account

At the start of 1978 I was done with the writing of the series “Towards the Ultimate Conservative Difference Scheme.” I had submitted the last paper, “A Second-Order Sequel to Godunov’s Method,” to the Journal of Computational Physics in October 1977 and was waiting for the reviews to come in. I had explicitly asked the editor of JCP “to include a Soviet scientist among the referees.” I could well imagine that some little-known scientist in the USSR was ahead of me, and I wanted to do the best I could to find out whether or not this was the reality. Without knowing it I was searching for Vladimir Kolgan.

The Russian review was worth waiting for. The reviewer was generous; he wrote: “The method described in the paper is very interesting and new in principle.” These encouraging words seemed to directly address and answer my question; yet, for a reason I no longer remember, I continued to have doubts about the uniqueness of my work. To my luck an opportunity arose to actively take part in the search. I was able to obtain a travel grant in an exchange program existing between the Dutch Ministry of Education and Sciences and the Soviet Academy of Sciences. Travel between Amsterdam and Moscow would be paid for by the Dutch government, while the Academy would host me once in the USSR.

On my application I listed a number of institutes I wanted to visit: in Moscow the Institute of Applied Mathematics, where I knew Viktor V. Rusanov, and the Academy’s Computing Center, where I knew Yuriy Shmyglevsky; furthermore, on the Academy’s Novosibirsk campus, the Computing Center, where S. K. Godunov worked, and the Institute for Theoretical and Applied Mechanics, where N. N. Yanenko worked and where I knew Yuriy I. Shokin. The Academy wisely added a visit to St. Petersburg (then Leningrad), although I did not know anyone there in applied or numerical mathematics. I took the trip from June 25 to July 14, 1978.

On the eve of my arrival in Moscow, Academician M. V. Keldysh, director of the Institute of Applied Mathematics and former Academy president, died. He was immured five days later in the wall of the Kremlin; the Institute was renamed Keldysh Institute of Applied Mathematics. Because of the commotion created by Keldysh’s death I did not get to meet Rusanov or any of the Institute’s other mathematicians. While this was a disappointment to me at the time, in retrospect it did not matter to my search for Kolgan, as explained toward the end of this story.

At the Computing Center in Moscow I found to my surprise a graduate student solving a compressible-flow problem with one of Boris’s FCT methods. Could it really be there was no home-grown non-oscillatory second-order scheme in the Soviet Union?

I gave formal and informal presentations about my work, with or without translator, in Moscow, Novosibirsk and St. Petersburg, and spoke at length with the applied and computational mathematicians on my list, including Godunov himself, and with others not on the list. Godunov answered emphatically negative when asked if ideas similar to mine had been developed in Russia. At that moment I must have shed any remaining doubts about the uniqueness of my work.

My stay in St. Petersburg was unexpectedly rewarding. An astrophysicist I knew there had managed to convince the Academy to let him host me. Vsevolod (Seva) V. Ivanov was not only the perfect host, he also had good friends in circles of applied mathematics, including Nina N. Ural'tseva of St. Petersburg State University. And so, on July 12, at the Steklov Institute of Mathematics of the Russian Academy of Sciences, I presented monotonicity-preserving second-order methods for hyperbolic equations in the spacious office of Ol’ga A. Ladyzhenskaya (+ 2004), who unfortunately was on travel, before a group of specialists in partial differential equations, most of them former students of hers, with Seva translating admirably.

Two days later I stepped on the plane back to Amsterdam in the certainty that there was no duplication of my work in the Soviet Union. And in a small town 600 km south of Moscow, the man I had been looking for was still alive – but not well. He was to die two weeks later.

6. Epilogue

My paper on the sequel to Godunov’s method appeared in 1979. A year later I met Rusanov at an international conference and learned he had been my Russian reviewer. He then shared with me a curious memory from the 1960s when he and Godunov still worked at the same institute. At that time Godunov actually discouraged Rusanov from trying to find a higher-order extension of his method, because he believed it was a unique, isolated method, one without a sequel. This could explain in part why in Russia the innovation did not spring from the academic applied-mathematics community.

It was not until about 1992 that I heard for the first time – from Ami Harten (+ 1994), who had picked this up when lecturing in Moscow – that some Russian scientist had developed a scheme like MUSCL way back in the 70s. This was affirmed by Dr. Vladimir Sabel’nikov, whom I met when he visited the University of Michigan in the fall of 1995 and who sent me proof of it: a copy of Kolgan’s 1972 TsAGI paper. A rough translation was provided by Valeriy Tenishev, then a student in Atmospheric, Oceanic and Space Sciences at the University of Michigan, a final translation by Konstantin Kabin, who had graduated from the same department. When the plan to publish an English translation of Kolgan’s article became solid, Dr. Sabel’nikov put me in contact with Dr. Yumashev at TsAGI, who sent me professional and personal information about Kolgan, including a photograph and a list of errata to the 1972 article.

The “Kolgan Project” progressed only slowly, being a task of category C: “important, not urgent.” Now that it finally has been completed, may it give food for thought to all of us who work hard to put computational science on an even higher plane. There are several lessons to be learned from the story of Vladimir Kolgan and his work: lessons relating to modesty,
secrecy, lack of communication, and fate. This introduction, and the translated article that follows, are meant to give Vladimir P. Kolgan the place in CFD history he deserves.

Acknowledgements

The author is indebted to Dr. Vladimir Sabel'nikov for providing the 1972 Kolgan paper and continued interest in the “Kolgan Project;” furthermore, to Dr. Vladimir Yumashev for providing detailed information about Kolgan’s life and work. He also thank Dr. Valeriy Tenishev for unlocking Kolgan's article with a rough translation, and Dr. Konstantin Kabin for providing the fine final translation. Without the interest of these people this project would not have been.

References