METHODS AND APPLICATIONS OF ANALYSIS.
Vol. 7, No. 3, pp. 473–478, September 2000
© 2000 International Press
004

ON THE ACCURACY OF GLIMM'S SCHEME*

PETER D. LAX†

Abstract. Glimm's random choice scheme constructs with probability close to 1 an approximate solution of an initial value problem for a hyperbolic system of conservation laws. This note describes an estimate for the probability of error that is much sharper than the one given by Glimm. A nontechnical discussion of this circle of ideas is given in [8] by Cathleen Morawetz, to whom this paper is affectionately dedicated.

In the 1950-s Sergei Godunov devised a finite difference scheme for solving approximately initial value problems for hyperbolic systems of conservation laws, in particular the equations of compressible gas flow, in one space dimension:

$$u_t + f_x = 0, \quad f = f(u), \tag{1}$$

u and f vectors with k components. The initial value $u(x,0)$ is prescribed.

Since solutions contain, in general, discontinuities such as shocks, they must be interpreted, as already observed by Riemann, in an integral sense. It is natural to make this the sense of distributions:

$$\int w(x,t)u(x,t)\,dx\Big|_0^T - \int_o\int [w_t u + w_x f(u)]dxdt = 0 \tag{1}'$$

for all differentiable test functions w.

In this difference scheme the initial data are approximated by one that is piecewise constant over x intervals I_j of length Δ. The first step in Godunov's method is to solve this piecewise constant initial value problem exactly. The exact solution is the union of centered waves–rarefaction, shock and contact–issuing from each point of discontinuity separating intervals of constancy, see figure 1.

This solution is valid only as long as the centered waves issuing from two adjacent points of discontinuity don't intersect. To make sure of that the time step is restricted to be less than $\Delta/2c$, where c is the maximum wave speed. Such restriction is similar to the classical Courant-Friedrichs- Lewy condition.

In the second step of Godunov's method the exact solution is replaced at the next time with a piecewise constant one, obtained by taking the average of the exact solution over each x interval I_j. Because of the conservation form of the equations these x averages can be obtained from the flux at the endpoints, without any integration.

The two steps are then repeated alternately until a desired time is reached.

Numerical evidence suggests strongly, almost overwhelmingly, that as Δ tends to zero the approximate solutions constructed by Godunov's method converge to an exact solution of the initial value problem. Yet, in spite of strenuous efforts, a rigorous mathematical proof of convergence is lacking.

In 1965 James Glimm showed, by a deep argument, how to estimate the *total variation* of the exact solution obtained in step 1 of Godunov's method in terms of the total variation of the initial value. He was unable to derive such an estimate for the averaged solutions; therefore he changed the second step of Godunov's scheme: the exact solution is replaced at the next time step by a piecewise constant one, obtained

*Received February 9, 2000.

†NYU-Courant, Rm. 912, 251 Mercer St., New York, NY10012-1110, USA (lax@CIMS.NYU.EDU).

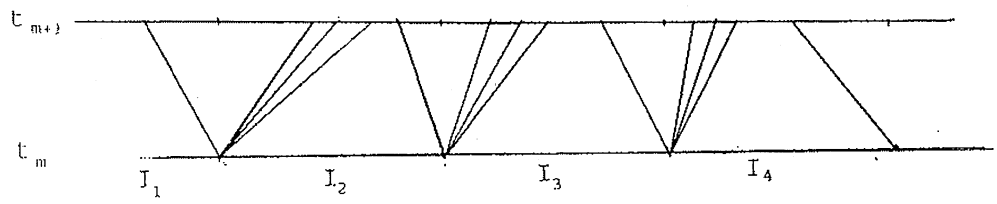

Fig. 1.

as the value of the exact solution in each interval I_j of length Δ at the point $\alpha\Delta, \alpha$ a random variable uniformly distributed over the unit interval.

The two steps are then repeated alternately; at each new time step a new random parameter α is chosen.

Let's denote the approximate solution obtained by Glimm's scheme as $v(x,t,\alpha)$ where $\alpha = (\alpha_1, \ldots, \alpha_N)$ are the values of the random parameters chosen in N steps. How close does v come to satisfy equation (1)'? Setting v into (1)' and using the fact that between two consecutive time steps v is an exact solution in the distribution sense we obtain, integrating by parts.

$$\sum_1^N \int w(x,t_m)[v(x,t_{m1}^+x) - v(x,t_m^-x)]dx \tag{2}$$

Here t_m are the intermediate times; since v is discontinuous at $t = t_n$, $v(x,t_m^+)$ and $v(x,t_m^-)$ denote the limits of $v(x,t)$ as t approaches t_m from above and from below respectively.

For fixed test function w we call (2) the *weak residual* and denote it as $r(\alpha)$:

$$r(\alpha) = \sum_1^N r_m(\alpha), \quad r_m(\alpha) = \sum_j r_{m,j}(\alpha), \tag{3}$$

where

$$r_{m,j}(\alpha) = \int_{Ij} w(t_m)[v(t_m^+) - v(t_m^-)]dx\ . \tag{4}$$

Glimm has shown that the weak residuals are small, except for a set of α of small volume. Here is his argument:

LEMMA.

(i) $|r_m(\alpha)| \leq 0(\Delta)$.

(ii) $|\int r_m d\alpha_m| \leq 0(\Delta^2)$.

Proof. (i) For x in the interval I_j, $v(x,t_m^+) = v(x_j + \alpha_m\Delta, t_m^-)$, where x_j is the left endpoint of the interval I_j. We can estimate for each x in I_j:

$$|v(x,t_m^+) - v(x,t_m^-)| = |v(x_j + \alpha\Delta, t_m^+) - v(x,t_m^-)| \leq V_{m,j},$$

where $V_{m,j}$ is the total variation of $v(t_m^-)$ over I_j. Setting this into (4) gives

$$|r_{m,j}(\alpha)| \leq M\Delta V_{m,j}$$

where M is the maximum of $|w(x,t)|$. Setting this into (3) gives

$$|r_m(\alpha)| \leq M\Delta V \tag{5}$$

where V is the total variation of $v(x, t_m^-)$ which is bounded in terms of the initial value.

ii) We rewrite $r_{m,j}$ in (4) as

$$r_{m,j} = \int_{Ij} w(x,t_m) - w(x_j,t_m)[v(x,t_m^+) - v(x,t_m^-)]dx + w(x_j,t_m)\int_{Ij}[v(x,t_m^+) - v(x,t_m^-)dx, \tag{6}$$

where x_j is the left end point of the interval I_J. In performing the integration with respect to α we integrate first with respect to α_m. Since by definition

$$v(x,t_m^+) = v(x_j + \alpha_m\Delta, t_m^-),$$

integrating with respect to α_m the second term on the right in (6) gives zero. The first term is estimated as before, giving

$$\left|\int r_{m,j}(\alpha_m)d\alpha_m\right| \leq d\Delta^2 V_{m,j}$$

where d is the maximum of the x derivative of w. Summing we get

$$\left|\int r_m(\alpha_m)d\alpha_m\right| \leq d\Delta^2 V, \tag{7}$$

as asserted in the Lemma. □

Glimm then estimates the expected value of r^2:

$$\int r^2(\alpha)d\alpha = \int (\sum_1^N r_n)^2 d\alpha = \sum \int r_m^2 d\alpha + 2\sum_{n<m}\int r_m r_n d\alpha. \tag{8}$$

Using part (i) of the Lemma we see that the first sum on the right in (8) is $0(\Delta^2)N$. In the second sum integrate first with respect to α_m; since for $n < m, r_n$ is independent of α_m, we can estimate the integral using part (ii) of the lemma. Estimating r_n by part (i) we get

$$\left|\int r_m r_n d\alpha\right| \leq 0(\Delta^3).$$

Summing over all $n < m$ we see that the second sum is $\leq 0(\Delta^3)N^2$. Adding the two estimates we get

$$\int r^2 d\alpha \leq 0(\Delta^2)N + 0(\Delta^3)N^2,$$

Since the optimal time steps are proportional to Δ, the number N of time steps needed to reach a specified time is $0(\Delta^{-1})$. Setting this in the above estimate gives

$$\int r^2 d\alpha \leq 0(\Delta). \tag{9}$$

Denote by $m(\varepsilon)$ the measure of the set of x for which

$$|r(\alpha)| \geq \varepsilon. \tag{10}$$

It follows from (9) that

$$m(\varepsilon) \leq \text{const}\frac{\Delta}{\varepsilon^2}. \tag{11}$$

It follows from (11) that ε has to be large compared to $\Delta^{1/2}$, i.e. that the order of accuracy of the method is less than $1/2$. That is not the worst of it. If Δ is diminished by a factor of 2, estimate (11) indicates that $m(\varepsilon)$ diminishes merely by a factor of 2. Therefore, to make $m(\varepsilon)$ reasonably small one has to take Δ unreasonably small.

Fortunately there is a better way to estimate $m(\varepsilon)$; we consider the expected value of $e^{sr(\alpha)}$, s a parameter to be chosen later:

$$\int e^{sr(\alpha)}d\alpha = \int e^{s\sum_1^N r_m}d\alpha = \int \prod_1^N e^{sr_m}d\alpha_N \dots d\alpha_1, \tag{12}$$

Since for $m < N$, r_m does not depend on α_N, we can rewrite (12) as

$$\int \prod_1^{N-1} e^{sr_m}d\alpha_{N-1}\dots d\alpha_1 \int e^{sr_N}d\alpha_N. \tag{12}'$$

Next we use the following simple inequality: for $|d| < 1$

$$e^d \leq 1 + d + d^2.$$

We will choose s so that $s\Delta << 1$; according to the lemma this makes $|sr_N| < 1$. Applying the above inequality to $d = sr_N$ we can estimate the last integral on the right of (12)′ as

$$\int e^{sr_N}d\alpha_N \leq \int (1 + sr_N + s^2r_N^2)d\alpha \leq 1 + s0(\Delta^2) + s^20(\Delta^2),$$

where in the last step we have made use of the estimates ii) and i) in the lemma. As we shall see, s is greater than 1, so

$$\int e^{sr_N}dx_N \leq 1 + s^20(\Delta^2) \leq e^{s^20(\Delta^2)}. \tag{13}$$

We estimate the remaining integrals in (12)′ similarly one by one, obtaining the estimate

$$\int e^{sr}dx \leq e^{s^20(\Delta^2)N}.$$

As we have seen, $N = 0(\Delta^{-1})$, so we deduce that

$$\int e^{sr}dx \leq e^{as^2\Delta}, \tag{14}$$

a some constant. Denote by $m_+(\varepsilon)$ the measure of the set of x for which

$$r(x) \geq \varepsilon,$$

It follows from (14) that

$$m_+(\varepsilon)e^{s\varepsilon} \leq e^{as^2\Delta},$$

the same as

$$m_+(\varepsilon) \leq e^{as^2\Delta - s\varepsilon}.$$

The optimum value of s is $\varepsilon/2a\Delta$, so

$$m_+(\varepsilon) \leq \exp(-\varepsilon^2/4a\Delta). \tag{15}$$

By estimating the expected value of e^{-sr} we obtain a similar estimate for the measure of the set where $r(x) < -\varepsilon$. So altogether we get that

$$m(\varepsilon) \leq 2\exp(-\varepsilon^2/4a\Delta), \tag{15}'$$

an estimate to be compared to the previous estimate (11). In both cases we must take ε large compared to $\Delta^{1/2}$. Note that our choice $s = \varepsilon/2a\Delta$ satisfies the conditions imposed previously on s:

$$\Delta s << 1, \qquad 1 < s.$$

Clearly (15)′ is a much sharper estimate of the measure $m(\varepsilon)$ of the set to be avoided. If we diminish Δ by a factor of 2, the right side of (15)′ decreases exponentially.

Of course the weak residual $r(\alpha)$ has to be small not only for a single test function w but for as many as are needed to resolve the approximate solution on a Δ grid. For a function of bounded total variation this may be as many as $|\log \Delta|$.

In nailing down the estimate for the total variation Glimm assumed that the initial data differ from a constant by a small amount. These conditions were substantially relaxed by Robin Young.

Although Glimm's method is of low accuracy, it has high resolution. In particular it presents shocks sharply, is free of overshoots and artificial oscillations. Chorin and Colella have shown it to be a practical method for computing one-dimensional flows.

Tai-ping Liu has shown how to use Glimm's estimate to prove the existence of solutions without resorting to random variables.

REFERENCES

[1] A. J. Chorin, *Random choice solutions of hyperbolic systems*, J. Comp. Phys., 22 (1976), pp. 517–536.

[2] A. J. Chorin, *Random choice methods with application to reacting gas flow*, J. Comp. Phys., 25 (1977), pp. 253–273.

[3] P. Colella, *Glimm's method for gas dynamics*, SIAM J. Sci Stat. Comp., 3 (1982), pp. 76–110.

[4] J. Glimm, *Solutions in the large for nonlinear hyperbolic systems of equations*, CPAM, 18 (1965), pp. 697–715.

[5] S. K. Godunov, *Bounds on the discrepancy of approximate solutions constructed for the equations of gas dynamics*, J. Comp. Math. and Math. Phys., 1 (1961), pp. 623–637. (in Russian)

[6] A. Harten and P. D. Lax, *A random choice finite difference scheme for hyperbolic conservation laws*, SIAM J. Num. Anal., 18 (1981), pp. 289–315.

[7] T. P. Liu, *The deterministic version of Glimm's scheme*, Comm. Math. Phys., 57 (1977), pp. 135–148.

[8] C. S. Morawetz, *Nonlinear conservation equations*, Amer. Math. Monthly, 86 (1979), pp. 284–287.

[9] R. Young, *Sup-norm stability of Glimm's scheme*, CPAM, 46 (1993), pp. 903–948.