Round Off, Truncation Error, ODEs, MMS, and Ill-Posed IVPs
Part 2: The Lorenz System

Dan Hughes

September October 2010

Introduction

First for the nomenclature: ODEs means Ordinary Differential Equations, MMS means the Method of Manufactured Solutions, and IVPs means Initial Value Problems.

This previous post [http://models-methods-software.com/2010/01/03/ill-posed-ivps-and-the-mms/] provided some information on these subjects. So far as I know, that post presented the first results for application of MMS to ill-posed IVPs. That post suggested that for the numerical solution methods used therein, the original Lorenz system of 1963 has yet to be correctly solved.

I plan to present some results about each of the aspects of numerical solution methods;

(1) Effects of round off due to the finite approximation to numbers, and
(2) Effects of truncation errors due to discrete approximations to continuous equations.

These will be investigated using three number representations; Single Precision (SP), Double Precision (DP), and Multi-Precision (MP), the latter using 512 digits for these notes. The rounding options available through many Fortran compilers will not be used; that introduces way too many additional options.

For each of the following equation systems;

(1) A system of two ODEs for which there are analytical solutions; the Method of Exact Solutions (MES),
(2) The original Lorenz system of 1963 which shows chaotic response, and
(3) The Method of Manufactured Solution (MMS) for the original Lorenz system.

Using two numerical solution methods;

(1) The first-order explicit Euler method, and
(2) The fourth-order explicit Runge-Kutta method.

There are many possible combinations for all these options, so there'll be lots of figures. All the necessary background will be given in this part of the series.

Summary of the Results
There'll be several and after I finally get all the pieces documented it'll be much easier to write this section. The conclusions from the previous post continue to hold.

The Analytical Solution
The following two linear, autonomous, coupled ODEs

$$\frac{dX}{dt} = 998X + 1998Z$$
and (1.1)
$$\frac{dZ}{dt} = -999X - 1999Z$$

with initial conditions (ICs)

$$X(0) = 1 \text{ and } Z(0) = 0 \tag{1.2}$$

have analytical solutions which can be obtained by use of a Riccati transformation. The solutions are

$$X(t) = 2e^{-t} - e^{-1000t}$$
and (1.3)
$$Z(t) = -e^{-t} + e^{-1000t}$$

Equations (1.3) show that *X* and *Z* are basically curves of exponential decay with; (1) *X* starting at *X*=1.0 with a very rapid increase to the value *X*=2.0 and approaching *X*=0.0 from above, (2) *Z* starting at *Z*=0.0 with a rapid change to the value *Z*=-1.0 and then exponentially approaching *Z*=0.0 from below, and (3) *Z* approaches *Z*=0.0 faster than does *X*. These properties are dominated by the leading terms in the solutions.

Both solutions have a second, exceedingly small, contribution as the independent variable moves off the origin: e^{-1000t} has a small numerical value for almost all non-zero values for *t*. If, for example, we take the machine precision to be about 12 to 14 digits, the value for the exponential is less than this for *t* just a little bit larger than 0.030.

The equation system is usually employed as a model to illustrate several aspects about the stability of numerical solution methods. An explicit discrete approximation will give the results that for stability the step size must satisfy

$$h < \frac{1.0}{1000.0} \tag{1.4}$$

Other combinations of explicit and implicit approximations for the dependent variables will show that if one is handled explicitly and the other implicitly the step size must continue to be bounded by Eq. (1.4). Only when both are handled implicitly can the step size be greater than that. It is interesting to note that stability of the numerical solution is governed by the smallest numerical values in the calculations.

Accurate resolution of the rapid increase in the solutions will generally require that a step size much smaller that the stability limit be used in the numerical solution methods. The time at which the dependent variables hit the maximum value is

$$t_{X\max} = -\ln\left[\left(2.0/1000.0\right)^{1.0/1000.0}\right] \tag{1.5}$$

for X, and

$$t_{Z\max} = -\ln\left[\left(1.0/1000.0\right)^{1.0/1000.0}\right] \tag{1.6}$$

for Z.

Obviously, discrete step sizes can not hit these values exactly, so there will always be some smearing of the numerical solution near the maxima. The greatest differences between the analytical and numerical solutions will generally occur at the times that the dependent variables are changes most rapidly.

Additional Analytical Solutions
After completing some calculations with the Lorenz system, I decided to investigate a couple of very simple ODEs having analytical solutions. These are

$$\frac{dX}{dt} = -\sin(t) \quad \text{and} \quad \frac{dZ}{dt} = \cos(t) \tag{1.7}$$

These equations were selected because they are parts of the MMS functions used with the Lorenz system. The equations are not coupled and they are not autonomous. I use two because all I had to code these was change the coding from the original coupled ODEs having analytical solutions. You can do the solutions in your head.

The Lorenz Equation System
These equations are

$$\frac{dX}{dt} = -PrX + PrY$$
$$\frac{dY}{dt} = -Y + RaX - XZ \tag{1.8}$$
$$\frac{dZ}{dt} = -bZ + XY$$

We will calculate discrete approximations to these equations for (1) in the convergent region of parameter space, and (2) in the chaotic region of parameter space. The focus, however, will be on the former, for which

$$Pr = 1.0, \; Ra = 1.0, \text{ and } b = 4 \tag{1.8a}$$

And for the latter, the classical values will be used

$$Pr = 10.0, \; Ra = 28.0, \text{ and } b = 8/3 \tag{1.8b}$$

We will use the initial conditions generally associated with the classical studies of the equations

$$X(0) = 1.0, \; Y(0) = -1.0, \text{ and } Z(0) = 10.0$$

The MMS Functions
As in the previous post, the MMS functions are taken to be

$$\begin{aligned} \Phi_X &= a_X \cos t \\ \Phi_Y &= a_Y + \sin t \\ \Phi_Z &= a_Z \cos t \end{aligned} \tag{1.9}$$

For which the source functions to be appended to the right-hand sides of the original equations are

$$Q_X = -X(0)\sin(t) + X(0)\,Pr\cos(t) - Y(0)\,Pr - Pr\sin(t) \tag{1.10}$$

for the *X* variable,

$$Q_Y = \cos(t) + Y(0) + \sin(t) - Ra\;X(0)\;\cos(t) + \;X(0)\cos(t)\,Z(0)\cos(t) \tag{1.11}$$

for the *Y* variable, and

$$Q_Z = - Z(0)\sin(t) + b\, Z(0)\cos(t) - X(0)\cos(t)\, Y(0) - X(0)\cos(t)\sin(t) \qquad (1.12)$$

for the Z variable.

When these source terms are appended to the right-hand sides of the original equations, the solutions represented by Eqs. (1.9) should be returned.

It is important to note that the original autonomous system will no longer be autonomous with the addition of these source terms into the right-hand sides of the original equations.

Corrections for all incorrectos will be appreciated.

The Numerical Methods

There's not much to say here. The explicit Euler method is very simple and straightforward to code. Generally it's easier to roll-your-own for this method than to go looking for canned off-the-shelf versions. The explicit Runge-Kutta method is also straightforward and modification of off-the-self versions, of which there are dozens, can almost always be successfully completed.

All the calculations covered by these notes could just as easily be carried out by any of the many different versions of canned, off-the-shelf software for many different ODE numerical solution methods. I predict that none of the conclusions based on the calculations here will be changed by calculations by other numerical solution methods.

Multi-Precision Arithmetic

The multi-precision arithmetic routines for your favorite language can be obtained from this site: http://crd.lbl.gov/~dhbailey/mpdist/ among many others.

Results from the Lorenz System

I have way too many auxiliary results from investigations of round-off and truncation errors for the analytical solutions to include, because writing gets harder and harder and harder. I'll try to get back to these Real Soon Now.

The calculations of primary interest are those with the MMS version of the Lorenz System. A previous post on this subject is here: http://models-methods-software.com/2010/01/03/ill-posed-ivps-and-the-mms/ and the figures for that post are here: http://edaniel.files.wordpress.com/2010/01/ivp_mms_figs.pdf . The presents calculations will focus on the results shown in the last figure in the latter document, Figure 9. That figure shows the results of calculations with the MMS Lorenz system for a region of parameter space that does not exhibit chaotic response.; Eq. (1.8a) above. It is very important to note that the calculations have been carried out for values of the parameters for which the expected response is an approach to a stable equilibrium; see

Figure 7 in the previous post. I'm going to try to put that Figure 9 in this document right here

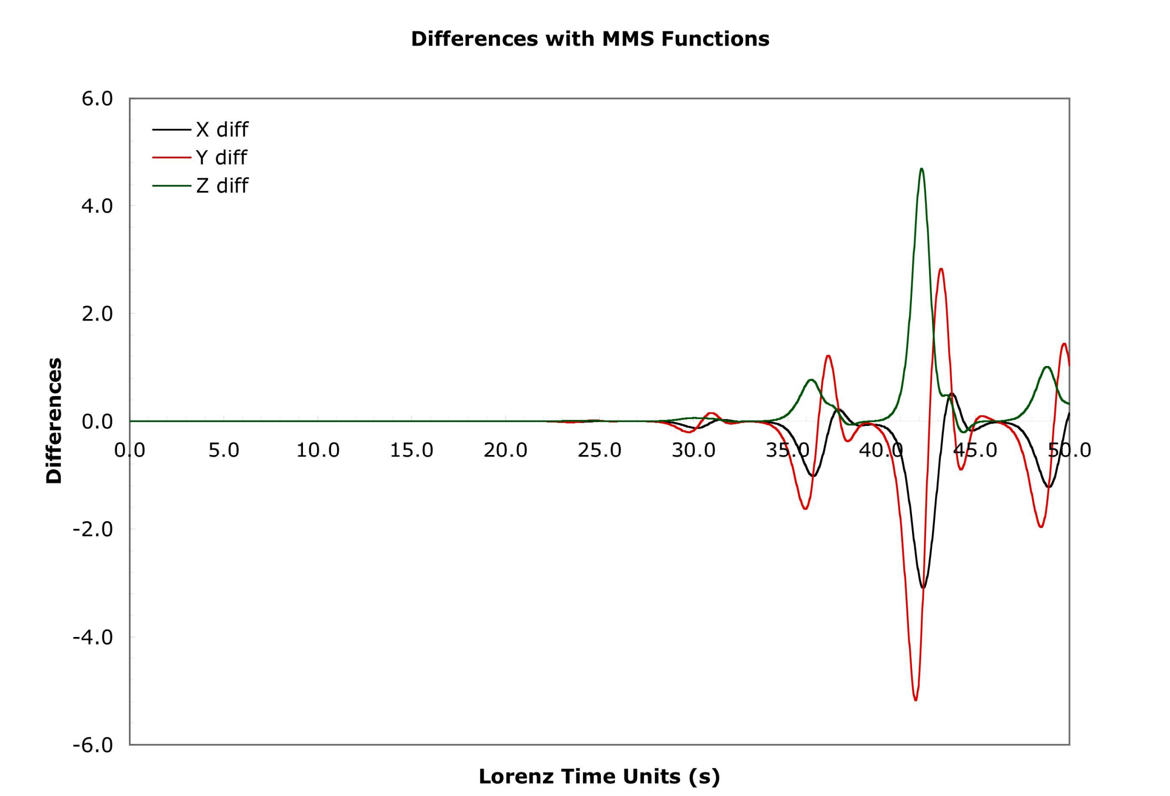

Figure 9. Differences between calculated solutions and expected solutions.

The figure shows the differences between the numerical integration for the dependent variables and the expected MMS solutions of Eqs. (1.9). The important aspect is that at about 25.0 Lorenz Time Units (LTUs) the differences are very significant. Actually, a closer look, what we're going to be doing in these notes, shows that the problem begins with the initial numerical integration steps; the differences are always increasing.

The objective of the MMS approach is to demonstrate that the numerical solution of the discrete approximations to the continuous equations are in fact correct. The results in Figure 9 show that this is not the case for the Lorenz system. The equations are not being solved. A very important result in my opinion.

What I've done now

With the above as the starting point, I started looking at the effects of the finite representation of numbers in computers, let's call this round-off, and the effects of the order of the discrete approximations in numerical solution methods.

My first investigations, as noted above, were conducted with ODEs for which analytical solutions are available. Note that this is in effect what the MMS approach does. The numerical solutions should correspond exactly to the MMS functions that are appended to the original equation system. The important parts of what I found out in that work is reflected in the present notes.

The calculations in the previous post were done using the first-order accurate explicit Euler method. For the calculations in these notes I have used the fourth-order accurate explicit Runge-Kutta method. First-order and fourth-order refer to the difference between the continuous equations and the discrete approximations; the truncation error. The truncation errors are one order higher than the order of the method. Part of the investigations with the analytical solutions were to determine that I have correctly coded the solution methods and that the expected theoretical performance is in fact reflected in the numerical solutions. That is the case.

Convergent Region and Equilibrium States without MMS Functions

I first integrated the original Lorenz system without the MMS modifications and for the parameter values corresponding to the convergent region using the Runge-Kutta method out to 200 Lorenz Time Units. The numerical solution did evolve to a stable equilibrium state and maintained rock-solid stable states for all three dependent variables; see Figure 7 in the previous post. I could put the new figure here, but it's very boring. These calculations were done with standard double precision IEEE arithmetic.

I did not see any adverse effects from DP arithmetic or step size. I will try to get the supporting information posted Real Soon Now.

Convergent Region and Equilibrium States with MMS Functions

These are the interesting results. I have used standard double precision arithmetic and multi precision arithmetic with 512 digits. Applying these to the MMS-modified Lorenz system continues to indicate that the system is not correctly integrated.

The basis of comparisons is the numerical solution compared with the expected MMS solutions. Showing results for all three dependent variables makes the plot crowded so instead I calculate the Eulerian distance between the expected and numerical solution by

$$D_{Eur} = \sqrt{X_{dif}^2 + Y_{dif}^2 + Z_{dif}^2} \qquad (1.13)$$

This number is larger than the individual contributions, but we are mostly interested in the trend of the values.

The first results are shown in the nearby Figure 1. The Figure shows the Eulerian distance for the results of using double precision arithmetic (DP) and multi-precision arithmetic (MP).

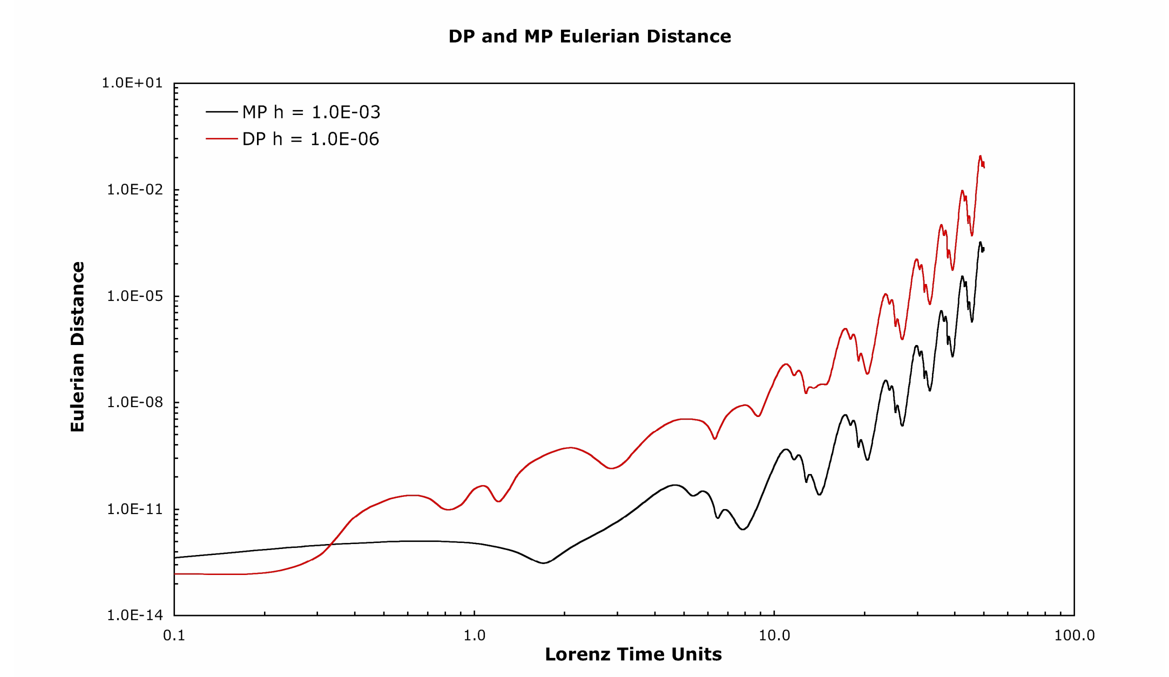

Figure 1. Eulerian distance for double and multi precision arithmetic for 50.0 LTUs.

The calculations have been carried out to 50.0 LTUs with the step size $h = 1.0\text{x}10^{-6}$ for the DP case and $h = 1.0\text{x}10^{-3}$ for the MP case. I don't have the computer power to run the MP case out much farther or with smaller step size.

The results show that initially the differences would indicate that the equations have been solved. However, the differences increase throughout the calculation. I say the equations are not solved.

The effect of multi-precision finite-number representation is to delay the growth of the differences and to make them smaller. The DP results are generally greater than the MP results even with the three orders of magnitude less step size. An indication that the precision of the finite number representation does make a difference with the same order of truncation error.

So, how does this look for longer ranges of LTUs. That is shown in the nearby Figure 2. The calculation using DP arithmetic and step size $h = 1.0\text{x}10^{-6}$ shows that the growth eventually stops and the oscillations are bounded at the same order of magnitude as the expected analytical solutions.

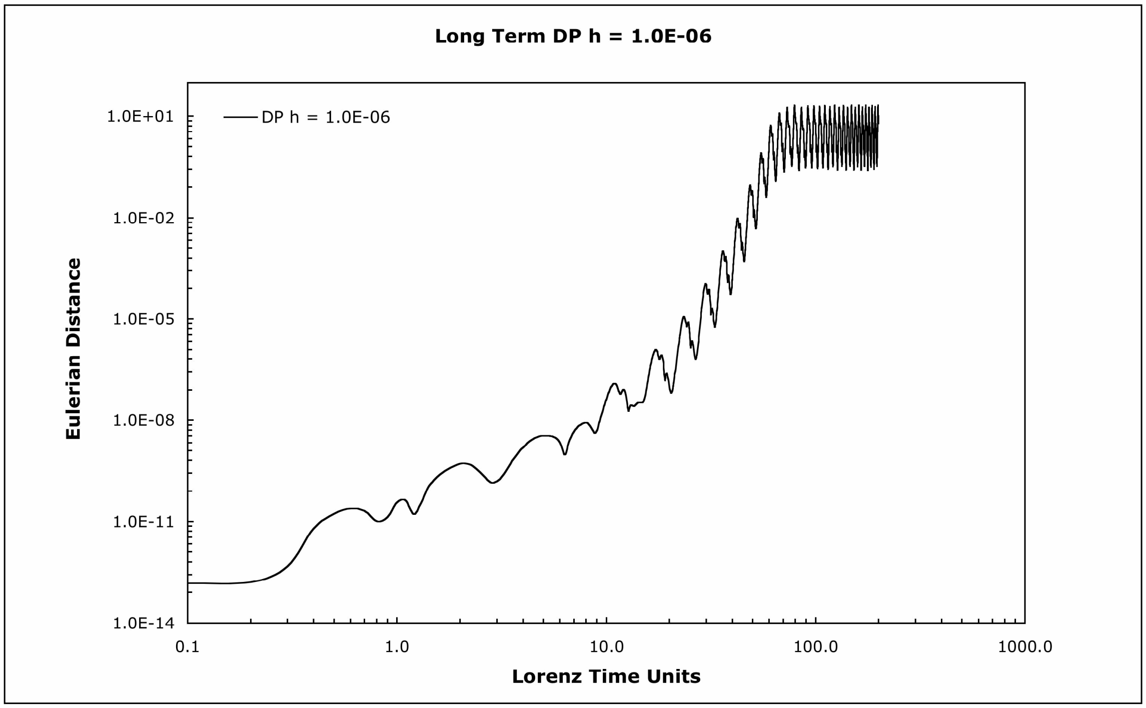

Figure 2. Eulerian distance for double and multi precision arithmetic for 200.0 LTUs.

This is the same behavior seen when the differences between two calculations of the un-modified Lorenz system with two different step sizes are plotted. The differences always increase to become of the same order as the numbers from a single calculation.

Conclusion

I think the Lorenz system has not yet been accurately integrated by any numerical solution methods. Higher-order methods plus at the same time higher precision representation of numbers will give results that might appear to be solutions. But, calculations for sufficiently long time spans will show that errors always increase.