UNIVERSITY OF WEST BOHEMIA

Applied and Computational Mechanics 2 (2008) 315–322

Applied and Computational Mechanics

Computability limits within the scope of computational mechanics

M. Okrouhlík[a,*]

[a] *Institute of Thermomechanics, Academy of Sciences of the Czech Republic, Prague*

Received 2 October 2008; received in revised form 23 October 2008

Abstract

The contribution deals with limits of computational mechanics as they unreel from those of physics, instrumental limits of computers and experimental instruments. It is shown that, more than on the above limits, our ability to predict the behavior of investigated phenomena depends on validity of models being employed.

Author ponders about limits of applicability of continuum model as well as about limits of finite element approaches when applied to approximate solutions of fast transient tasks in solid continuum mechanics.

Keywords: continuum, validity of models, threshold, finite element method

1. Introduction

We are living in the word of Newtonian physics that admits the notion of absolute time, independent of space. The Newtonian description is unchanged under a Galileo transformation. According to Newton's concept, while the positions of two observers differ by virtue of their relative 'slow' motions; both have an identical perception of time, which does not depend on their frames of reference. Newtonian mechanics provides a vital tool, which still perfectly works in all manners of ways from engineering to astronomy. The Newtonian concept is deterministic. If positions, velocities and masses of various bodies are given at one time, and if the history of pertinent state variables is known, then their positions, velocities and accelerations are determined for all later times.

The concept of continuum ignores the molecular or atomic structure of matter and can be equally well applied to specimens of any size even if their volumes are vanishing to zero.

So the Newtonian continuum mechanics is a paradigm constraining all our activities far away from the world of fast relativistic approaches as well as from the world of non-deterministic quantum microcosms.

We are not inventing completely new concepts of the world, nature — we are not proposing any fundamental principles. Rather, we are dealing with more and more precise models designed within the scope the Newtonian mechanics, testing them on non-standard and/or large-scale problems.

Today, material non-linearities, large strains and deformations, high-velocity impact problems and many others are almost routinely treated by sophisticated discrete tools as for example boundary elements, finite elements expressed in Eulerian, Lagrangian and/or ALE formulations. [1]

*Corresponding author. Tel.: +420 266 053 158, e-mail: ok@it.cas.cz.

The progress in the field would be unthinkable without the fast development and wide availability of computer platforms, especially the vector and parallel ones.

The models we are designing and using either work or do not work. In this respect they cannot be proclaimed to be true or false — the question of their truth becomes irrelevant. We are mainly interested in the design of reliable models — reliable within the scope of their validity.

And ascertaining the validity limits of models we are working with is one of our ultimate goals.

2. Models and their solutions

Model is a purposefully simplified concept of a studied phenomenon invented with the intention to predict – what would happen if... Accepted assumptions (simplifications) specify the validity limits of the model and in this respect the model is neither true nor false. Regardless of being simple or complicated, it is good, if it is approved by an experiment. [2]

The methods for investigating stress wave propagation tasks will be discussed and used as a vehicle for assessing the limits of our prediction and observation.

3. Model versus experiment

When trying to reveal the 'true' behavior of a phenomenon we are using the experiment. When trying to predict the 'true' behavior of a mechanical phenomenon we are accepting a certain model of it and then solve it analytically and/or numerically. Eventually, we test it by an experiment.

Physical laws (models) as Newton's, energy conservation, theory of relativity, etc. cannot be proved (in mathematical sense).

Often we say that it is the experiment which ultimately confirms the model. But experiments, as well as the numerical treatment of models describing the nature, have their observational thresholds.

Sometimes, the thresholds of computational analysis are narrower than those of experiment. From this point of view the experiment is a model of nature as well.

4. Frequency limits of continuum and of FE analysis

For fast transient problems, as shock and impact, the high frequency components of solutions are of utmost importance. In continuum, there is no upper limit of the frequency range of the response. In this respect continuum is able to deal with infinitely high frequencies. This is a sort of singularity embedded in the continuum model.

As soon as we apply any of discrete methods for the approximate treatment of transient tasks in continuum mechanics, the value of upper cut-off frequency is to be known in order to 'safely' describe the frequencies of interest. Furthermore the temporal and spatial dispersion effects enter the stage. [6]

Continuum mechanics deals with response of solid and/or fluid medium to external influences. By response we mean description of motion, displacement, force, strain and stress expressed as functions of time and space. By external influences we mean loading, constraints, etc, expressed as functions of time and space.

Continuum is a macroscopic model disregarding the corpuscular structure of matter based on the continuity hypothesis — it is assumed that matter within the observed specimen is distributed continuously. Quantities describing the continuum behaviour are expressed as piecewise continuous functions of time and space. And last but not least it is supposed that all *considered properties within the observed infinitesimal element are identical with those of a specimen of finite size*.

5. Using a model outside of its limit is a blunder

As a rule people designing the models are different from those who are using it. So the latter group might sin by overlooking the fundamental point, namely that using a model outside of its validity limits leads to erroneous results and conclusions. This is not, however, the fault of the model, but a pure consequence of a poor judgment of the model's user.

Blunders are easy to commit, since models give no warning of their misuse. Lot of checks might be satisfied and still...

In this paper there is presented a rather extended example, clarifying the intricacies of using two time step operators used in FE modeling of a case in transient mechanics. Other examples, devoted to *quantitative measures of agreements of 'close' solutions* and to *mesh refinement and frequency analysis* will be briefly mentioned in the oral contribution.

6. An example — strain distributions of the same task obtained by Newmark (NM) and central difference (CD) operators

Several time marching operators for solving the systems of ordinary differential equations, suitable for the FE modeling of transient tasks of solid continuum mechanics, are known today. The detailed description of their background and analyses of their properties can be found, e.g., in [10, 13, 5]. Commercial FE packages offer plethora of approaches. [11, 12]. The outlines and rules for their 'safe' usage are generally advocated; nevertheless it still might be of interest to analyze in detail the minute differences obtained by applying different integration methods to the same task.

Let's concentrate our attention to the comparison of results obtained by Newmark (NM) and central difference (CD) methods. [15]

The NM method is a classical representative of implicit methods. Used with consistent mass matrix and without algorithmic damping it conserves energy and is unconditionally stable. In order to minimize the temporal and spatial discretization errors the NM method is recommended [10], to be used with consistent mass matrix formulation.

The CD method, the representative of explicit methods, is only conditionally stable. When used within its stability limits with consistent mass matrix formulation it also fully conserves energy. To reduce the temporal and spatial discretization errors the CD method is recommended, [10], to be used with diagonal (lumped) mass matrix formulation. Using it with a consistent mass matrix is possible but practically prohibitive for two reasons. First, the problem becomes computationally coupled. Second, the data storage demands for the consistent mass matrix are substantially higher than those needed for a diagonal mass matrix. Today, the CD method is almost exclusively used with the diagonal mass matrix formulation, which is furthermore plausible from the point of view of minimization of dispersion effects. But using the CD method with diagonal mass matrix we are punished a little bit by the fact that the time dependence of total mechanical energy slightly fluctuates around its 'correct' value. [14]

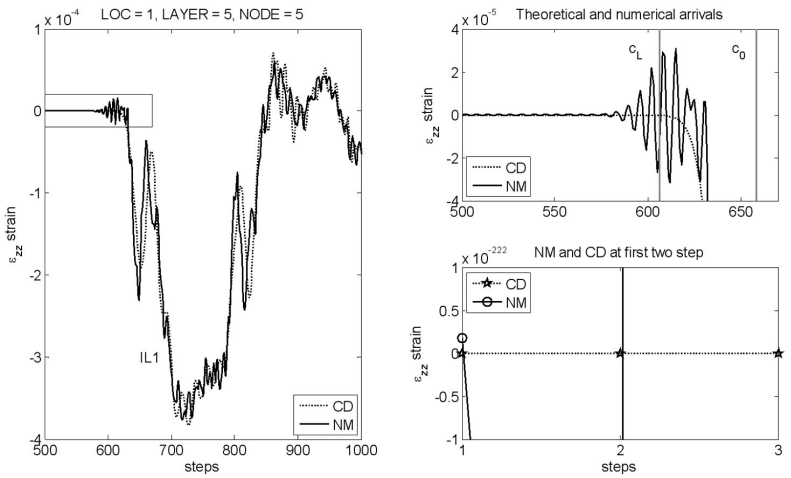

Fig. 1. Time distributions of surface axial strains obtained by NM and CD operators

Comparison of the time history of axial surface strains at a certain location of the studied case [9], whose distance from the impacted face of the tube is 340 mm, in layer 1 and node 5, obtained by NM and CD methods using 3D elements, is presented in Fig. 1. The same time integration step ($1e-7$ [s]) was used in both cases. The proper choice of the time step value is discussed in [18]. For the NM method the consistent mass matrix was employed, while the diagonal mass matrix was used for the CD method.

The left-hand subplot presents the axial strains as functions of timesteps in the above mentioned location. The negative peak, denoted IL1, corresponds to the immediate position of the loading pulse. There is a visible difference between NM and CD results, which — from the engineering point of view — seems to be small. Often, the differences are viewed by the prism of the plotting scale.

In the upper right-hand subplot of Fig. 1, which is the enlarged view of the small rectangle presented on the left-hand side of Fig. 1, the theoretical positions of arrivals of hypothetical 3D (c_{L}) and 1D (c_0) longitudinal waves are indicated by vertical lines. Of course in a bounded 3D body no pulse, being composed on infinitely many harmonics, can exclusively propagate by any of above mentioned velocities. But the theoretical wave speeds are useful bounds for our expectations. The detailed strain distributions, obtained by NM and CD methods, are shown as well. From the analysis of dispersion properties of finite elements and that of time integration methods, presented in detail in [10], it is known that the computed speed of wave propagation for the CD approach with diagonal mass matrix underestimates the actual speed, while using the NM approach with consistent mass matrix the actual speed is overestimated. The presented results nicely show this. When looking at the enlarged details of the wave arrivals, as modeled by NM and CD operators, a nagging question might intrude our minds. Where or actually when does the incoming pulse start? A similar subject was analyzed on experimentally obtained data in [9], where it was shown how the 'detected' moment of arrival depends on the observational threshold. Different frequency contents of both signals, as well as a more detailed analysis of CD and NM operators are treated in [9].

Less known is the fact that the speed of propagation, modeled by NM method with consistent mass matrix formulation, is actually 'infinitely' large. [14] A brief explanation of this curiosity is sketched out in the following paragraph.

7. Interlude — assessment of a 'variable' computational speed of wave propagation by analyzing two time marching algorithms for the numerical integration of the system of ordinary differential equations

Solving $\mathbf{M}\ddot{\mathbf{q}}+\mathbf{K}\mathbf{q}=\mathbf{P}(t)$ the central difference (CD) method and the Newmark (NM) method lead to the repeated solutions the system of algebraic equations

$$\frac{1}{\Delta t^2}\mathbf{M}\mathbf{q}_{t+\Delta t}=\tilde{\mathbf{P}}_t, \quad \text{(a)} \qquad \hat{\mathbf{K}}\mathbf{q}_{t+\Delta t}=\hat{\mathbf{P}}_{t+\Delta t}, \quad \text{(b)}$$

where the effective loading forces and the effective stiffness matrix are

$$\tilde{\mathbf{P}}_t=\mathbf{P}_t-\left(\mathbf{K}-\frac{2}{\Delta t^2}\mathbf{M}\right)\mathbf{q}_t-\frac{1}{\Delta t^2}\mathbf{M}\mathbf{q}_{t-\Delta t}, \quad \hat{\mathbf{P}}_{t+\Delta t}=\mathbf{P}_{t+\Delta t}+\mathbf{M}(c_1\mathbf{q}_t+c_2\dot{\mathbf{q}}_t+c_3\ddot{\mathbf{q}}_t),$$
$$\hat{\mathbf{K}}=\mathbf{K}+\frac{1}{\beta\Delta t^2}\mathbf{M}.$$

Definition of constants appearing above and more details are in [15].

Generally, the matrices $\mathbf{K},\mathbf{M},\hat{\mathbf{K}}$ are sparse and banded. Nevertheless their inversions $\mathbf{K}^{-1},\mathbf{M}^{-1}$; as well as $\hat{\mathbf{K}}^{-1}$ (needed for extracting the displacements $\mathbf{q}_{t+\Delta t}$ at the next time step from equations (a) and (b)) are full. From it follows that in both systems of equations the unknowns are coupled. This means that when calculating the i-th displacement, there are all other displacements, which — through the non-zero coefficients of a proper inverse matrix — are contributing to it.

It means that when (at the beginning of the integration) a nonzero loading is applied at a certain node, then (at the end of the first integration step) the displacements at all nodes of the mechanical system are non-zero, indicating that the whole system already 'knows' that it was loaded, regardless of the distance between the loading node and the node of interest.

The magic spell could only be broken if the matrix, appearing in the system of algebraic equations, is diagonal, because its inversion is then diagonal as well. This, however, could only be provided for the CD approach, operating with mass matrix, because it is the mass matrix only which can be meaningfully diagonalized. [15]

The above discussion is illustrated in the lower right-hand side subplot of Fig. 1 where one can see the strains computed by CD and NM operators (at a location whose distance from the loading area is 340 mm) during the first three steps of integration. The CD operator, with a diagonal mass matrix, gives the expected series of pure zeros, while the NM method gives values negligibly small (of the order of 10^{-222}) but still non-zero. It should be emphasized that this has nothing to do with round-off errors. The same phenomenon would have appear even if we had worked with symbolic (infinitely precise) arithmetics.

So the detected value of speed of stress wave propagation (obtained by the registration of the first non-zero response at a certain time in a given distance from the loading point using the NM method) depends not only on the distance of the point of observation from the loading node but paradoxically on the timestep of integration as well. The CD method spares us of these troubles.

There is another computational threshold entering the consideration, namely the number of significant digits used for the mantissa representation of the floating number. [16]

The minimum floating point number that can be represented by the standard double precision format (that we have used for the computation) is of the order of 10^{-308}. This is our numerical observational threshold allowing distinguishing the value 10^{-222} in the first step of the lower right-hand side of Fig. 1.

If, for the same numerical integration in time, we had employed the single precision format (threshold of the order of 10^{-79}) we would have observed pure zero in the first step instead and the first non-zero value would appear later.

Of course, nobody would measure the wave speed this way. What would be a common sense approach? Sitting at a certain observational node, whose distance from the loaded node is known, one would estimate the speed by measuring the time needed for the arrival of the 'measurable' or 'detectable' signal. And the measurable signal is such that is in absolute value greater than a 'reasonable' observational threshold. And what is a proper value of it is a good question.

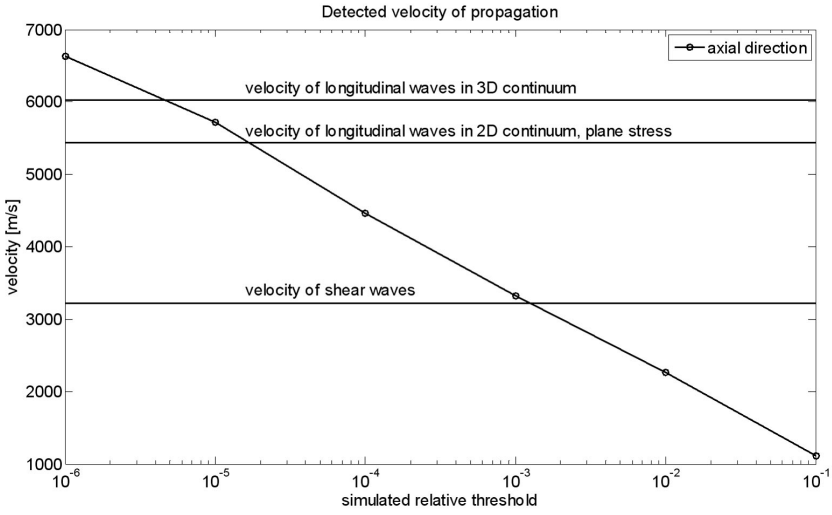

Fig. 2. Detected velocity of propagation vs. relative threshold

A thought experiment accompanied by FE computation might help. Imagine a standard finite element double-precision computation giving at a certain time the spatial distribution of displacements at a node on the surface of a body. Assume that the distance of our observational node from the loading node is known. Now, let's set a 'reasonable' value of the threshold and apply a sort of numerical filter on obtained displacements, which erases all the data whose absolute values are less than the mentioned value of the threshold. This way, for a given threshold value, we get a certain arrival time and from the known distance we obtain the propagation speed. Working with displacements normalized to their maximum values allows us to consider the threshold values as the relative ones. For more details see [14].

Varying the simulated threshold value in the range from 10^{-6} to 10^{-1} we will get a set of different velocities of propagation. As a function of threshold they are plotted in Fig. 2. Material constants for the standard steel were used. [9] The horizontal lines represent the theoretical speeds for longitudinal waves in 3D continuum, for longitudinal plane stress waves in 2D continuum as well as for the shear waves. Obviously, the shear wave speeds are identical both for 3D and 2D cases. [8, 17]

The previous discussion might appear rather academic. The threshold issue, however, is really important when the speed of propagation is being determined by experimental means. The procedure is the same as in the numerical simulation approach. Observing the first 'measurable' response at a certain time in a given distance from the loading point one can estimate the speed

of propagation. As before, the estimated velocity value depends on the observational threshold value. There is, however, a significant difference. While we could almost arbitrarily vary the simulated threshold value in the numerical treatment, the value of observational threshold is usually constant for the considered experimental setup being used for the measurement of a particular physical quantity.

It is known that the longitudinal waves carry substantially less amount of energy than these of the shear and Rayleigh waves and that the surface response, measured in displacements or strains, is of substantially less magnitude for the former case.

From the experimental point of view one can conclude that for a correct capturing of the longitudinal velocity value, the relative precision of at least of the order of 10^{-6} is required. This is a tough request. The relative threshold of the order of 10^{-3} is more common in experimental practice. However, in an experiment with the relative precision of the order of 10^{-3}, one would not detect the arrival of longitudinal waves and might wrongly conclude that the first arriving waves are of the shear nature and would estimate the velocity of propagation of the order of 3 000 m/s.

All this fuzz is about the margins of our ability to distinguish something against nothing. This is, however, crucial for any meaningful human activity.

8. Conclusions

When trying to ascertain the reliability of modeling approaches and the extent of their validity one has to realize that the models as a rule do not have self-correction features. That's why we have to let the models to be checked by themselves, be checked by independent models and let the systematic doubt be our everyday companion.

When refining our approach to solid mechanics treatment by employing smaller elements and timesteps of integration we are revealing more and more details, but at the same time we witnessing certain new phenomena, some of which might, or might not be, artifacts of the employed model, for which, sometimes, we have no immediate explanations.

In this respect the statement initially applied to quantum physics might be consoling, namely *Science cannot solve the ultimate mystery of nature. And that is because, in the last analysis, we are part of the mystery that we are trying to solve*. The author of this assertion is Max Planck. It is cited from [7].

Our pronounced doubts about the behaviour of models, being tested on the brink of their capability, should not be understood as a sort of mistrust. On the contrary, they make us stronger in our qualified faith and confidence in computational mechanics.

Acknowledgements

The work has been supported by the grant project 1ET 400760509 of the Czech Academy of Sciences.

This paper covers the contents of invited lecture presented at the conference Computational Mechanics 2008.

References

[1] E. Stein, R. de Borst, J. R. D. Hughes, Encyclopedia of Computational Mechanics, Volumes 1 to 3, John Wiley & Sons, Ltd., Chichester, 2004.

[2] S. Flüge, Encyclopedia of Physics. Vol. III. Principles of Classical Mechanics and Field Theory. Springer Verlag, Berlin, 1960.

[3] A. Ralston, A First Course in Numerical Analysis, New York, McGraw-Hill, Inc., 1965.

[4] B. N. Parlett, The Symmetric Eigenvalue Problem, Englewood Cliffs, N. J. Prentice-Hall, Inc., 1978.

[5] M. Papadrakakis, Solving Large-Scale Problems in Mechanics, John Wiley and Sons Ltd., Chichester, Baffins Lane, 1993.

[6] M. Okrouhlík, C. Höschl, A Contribution to the Study of Dispersive Properties of One-dimensional Lagrangian and Hermitian Elements. Computers and Structures, Vol. 49, No. 5, 1993, pp. 779–795.

[7] K. Wilber, Quantum questions. Mystical Writings of the World's Great Physicists, New Science Library, Boston, 1985.

[8] A. E. H. Love, A Treatise on the Mathematical Theory of Elasticity. (First edition 1892). New York, Dover Publications 1944.

[9] M. Okrouhlík, S. Pták, FE and experimental study of axially induced shear stress waves in a tube with spiral slots. Part 1. Research Report of the Institute of Thermomechanics, Z 1417/2008, Prague, 2008.

[10] M. Okrouhlík, Mechanics of Contact Impact. Applied Mechanics Review, Vol. 47, No. 2, February 1994.

[11] Marc manuals, www.mscsoftware.com.

[12] Ansys manuals, www.ansys.com.

[13] T. Belytschko, W. K. Liu, B. Moran, Nonlinear Finite Elements for Continua and Structures, John Wiley, Chichester, 2000.

[14] M. Okrouhlík, S. Pták, Assessment of experiment by finite element analysis: Comparison, self-check and remedy, Strojnický časopis, Journal of Mechanical Engineering, Vol. 56, No. 1, 2005.

[15] K. J. Bathe, Finite Element Procedures, Prentice-Hall, Englewood Cliffs, New Jersey, 1996.

[16] W. Kahan, IEEE Standard 754 for Binary Floating-Point Arithmetic, Lecture Notes on the Status of IEEE 754, Elect. Eng. & Computer Science, UCLA, Berkeley, 1997.

[17] H. Kolsky, Stress Waves in Solids, Clarendon Press, Oxford, 1953.

[18] Dyna manuals, www.dynasupport.com/support/tutorial/users.guide.