Computer Sciences

By B. Gustafsson, J. Holmkvist and S. Hägglund

The Department of Computer Sciences is headed by Professor Heinz-Otto Kreiss. It was founded in 1965, and came as a natural consequence of the rapid development of fast electronic computers. These made it possible to solve large mathematical problems, which in the beginning mainly came from physics. However, they also made it necessary to develop new theories, so that numerical methods could be designed and analyzed.

The research in Uppsala was in the beginning concentrated to this field (numerical analysis), but the activities have later been widened, and they now also include non-numerical computer science and image processing.

Numerical analysis

Numerical Analysis, as a science, is a branch of applied mathematics. It is concerned with the approximate solution of various mathematical problems, where an explicit exact solution on closed form is impossible to obtain. It might also be the case that such a solution can be obtained, but is very impractical and inefficient to use. The numerical analyst starts formally with a given mathematical model, which might have been derived from a wide variety of domains within other sciences.

The main emphasis at the department is on the numerical solution of differential equations, in particular those of partial type. Typical application domains where this kind of equations arises, are fluid- and gas-dynamics, meteorology, oceanography, seismology, solid mechanics, high energy physics, reactor physics, plasmaphysics, chemistry, biology and medicin.

One might think that with the availability of high speed large computers, there should be no difficulty to choose a numerical method for a given problem. Just pick a simple crude method, and the computer produces the results. This is by no means the case. On the contrary, for most problems it is highly important to make a careful analysis before the choice of method. For some problems, crude methods can be used, but take to long computing time for a given accuracy. For other problems, much research must be done before any reasonable solutions can be obtained at all.

The dominating technique for solving partial differential equations numerically, is to use finite difference methods, that is the derivatives in the equation are replaced by finite differences. The resulting methods are

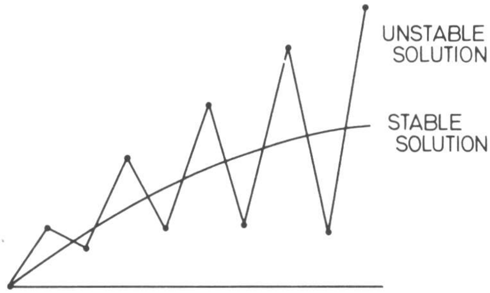

Fig. 1. Results obtained by two methods, where only one is stable.

often very simple to implement on a computer, and have therefore obtained a considerable interest. The finite element methods have also become popular during recent years. The main theoretical problem for difference methods is the question of stability and convergence. In particular, it is difficult to know what boundary conditions should be used with a given difference approximation to make sure that the solution stays bounded independent of the grid size. Fig. 1 shows a typical unstable solution.

Another natural question is what method within a class of stable methods should be chosen such that a given accuracy is obtained in shortest possible computing time and with smallest possible computer storage. This is very important for problems in several space dimensions, since the computing time and storage requirements grows very quickly with the number of dimensions.

Many problems in hydrodynamics give rise to singular perturbation problems. These are characterized by the existence of a small parameter which when it approaches zero changes the character of the differential equations. It represents for example the viscosity in a fluid, and gives rise to boundary layers, see Fig. 2. These layers are often much thinner than the smallest possible grid size in the remaining part of the domain. Even if the layer itself is of no interest, the numerical solution in the other part of the domain might be destroyed with an insuitable method.

All the problems mentioned there, becomes even more complicated when the differential equations are defined over a domain with an irregular boundary. The question is how should the grid and the difference approximation be defined near the boundaries, such that the calculations are stable and convenient to program.

The problem of getting a general software for numerical solution of partial differential equations is very important. Many of the difficulties associated with this are due to the great variety of the mathematical problems and the absence of good general purpose numerical algorithms. Most of the existing program packages for solving partial differential equations are very specialized to certain problems. The attempts to make more general packages has almost all been based on numerical general purpose algorithms but this leads in many cases to a drastic increase of computer time. It is important that the algorithms and programs used for a certain

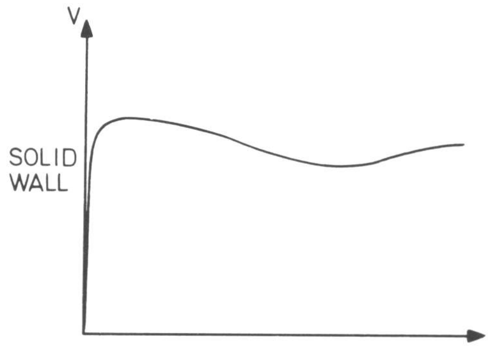

Fig. 2. Problem with a boundary layer: the velocity v of a fluid near a solid wall.

problem are efficient since numerical integration of PDE's often are very time and space consuming. Thus a general PDE-solver must include several difference schemes, boundary strategies, algebraic equation solvers etc. For efficiency it is desirable that the PDE-solver is able to make use of certain characteristics in a problem by recognizing absence of terms, constant coefficients, linear terms etc.

One possible approach is to let the user specify his problem i.e. equations, boundary conditions etc. in a mathematical notation. The system then analyzes the problem symbolically, and numerical algorithms are selected according to the result of the analysis. Finally a complete FORTRAN program for numerical integration of the equations is generated.

The type of problems described above and other problems closely related to them, are currently treated at the department. They will probably also achieve most of the interest for a few years.

The ideal situation, where the physicist defines the mathematical model, where the mathematician shows that there is a unique solution, and where the numerical analyst provides a suitable method of solution, hardly ever occurs. The experience at the department of Computer Sciences is that a close cooperation between the different categories is necessary. In the first place, it is often difficult to define the problem, since the mathematicians and the physicists/engineers do not speak "the same language". It is therefore very important that the numerical analyst/mathematician has a good knowledge of the topic from which the problem emanated. Furthermore, he must be able to simplify the very complicated equations that sometimes arise, but to do it in such a way that the basic properties of the original problem still are present. The education at the universities has often been limited to the treatment of very simple model equations, which do not present the difficulties, which usually are contained in practical applications. It is of course very important to understand the simple problems first, but more emphasis on the solution of more difficult problems is necessary to prepare the student for the hard reality. Already at the basic level, the student must be introduced to some problem from an applied area. At the post graduate level it is very important that the numerical analyst learns to understand the language used by the specialist in an applied field, and to translate it into mathematical-numerical terms.

Furthermore, the physicist/engineer very seldom can tell a priori the domains, in which his parameters can vary. Since the properties of most numerical methods are very different depending on the values of the parameters, there must be a continuous contact between the two categories, when a certain problem has to be solved.

More generally, there must be an effective flow of information between on one side the basic research and education at the universities, and on the other side the specialities from the applied fields. This information flow has improved considerably at our department during recent years, and one can foresee a further development during the coming years. This is of course important in both directions: those physicists who want to device the numerical methods themselves have to know about recent results in numerical analysis.

Image processing

In close collaboration with the Department of Clinical Cytology of the University Hospital in Uppsala, research is carried on aiming at quantitative and automatic methods in cytology. The general approach is to use image processing methods to develop algorithms for accurate extraction of various cell parameters from digitized microscopic images of the cells. These parameters are analyzed by using pattern recognition, mainly statistical techniques. In this way accurate measurements of cell parameters such as nuclear size, cytoplasm size, light absorbtion of nuclear can be obtained. Also more abstract parameters, usually not used by human observers, such as spectral information of nucleus, distribution of grey scale values of the cell etc. are possible to quantitate.

Presently work is carried out in four different areas, namely:

1. Automatic classification of cells, as malignant or normal, from the uterine cervix. This is an important problem which has to be solved before it is possible to build automatic systems for mass screening of cervical smears for early detection of cancer of the uterine cervix. Promising results have been obtained using the distribution of grey scale values of the cells in order to classify them.

2. Development of image processing algorithms which can automatically detect and isolate cells in a digitized image containing several cells. This is a general problem in the design of automatic screening systems, which in our case is directed to be applied on cervical smears.

3. Development of a software system for interaction between microscope–man–computer to enable measurements on cells, such as size, shape, structure etc. This system is general and also applicable to other kind of microscopic objects. In the system is also included a statistical part for evaluation of measured parameters.

4. Characterization of tumor cell populations from breast cancers using computer analysis of digitized images of cell nuclei. The object of this work is to obtain an accurate and reproducible grading of breast cancers based on the morphology of the cell nuclei, and to relate the grading to epidemiological findings of the patients.

Non-numerical computer science

"Datalogilaboratoriet", headed by Professor Erik Sandewall, was started in 1969 and consisted in January 1976 of about fifteen persons. The main theme of the activities in the group has been to apply methods from Artificial Intelligence research to computer problems that involve intense processing of moderately sized but complexly structured data bases. One characteristic of such a "small" data base is that there is only one memory level concerned, in contrast to large business oriented data bases with problems of physical allocation on tapes and discs. The need for intense processing of the data base will presumably arise from inherent complexity in the given task, rather than complexity in the efficient application of computer techniques. One important utilization of this type of data base is when the information to be stored is expressed in natural language, say English or Swedish. This application has been one justification for a concentration in this particular field and on programming systems supporting this work. Others are e.g. formula and program manipulation, automatic problem solving and interactive information systems. Still other research areas of importance at Datalogilaboratoriet are methods for query languages and other communication with data bases as well as methods for management of large and complex programs.

The research strategy employed in the group strongly emphasizes the importance of combining theoretical work with practical programming. Thus the actual implementation of a computer program built upon a theory or an hypothesis corroborates the viability of the proposed solution to a problem. It is also important to realize the fertility of mutual influence of basic research and applied real-life projects in the subject of computer science. One must not isolate research from the problems of computer processing outside the university, provided that independence is not violated. At Datalogilaboratoriet there has been a declared policy to perform applied projects in co-operation with external participants, among which The National Swedish Environment Protection Board and The National Central Bureau of Statistics may be mentioned. Substantial attention is also devoted to questions of methodology for programming and system development.

Most important for the possibility of concentration upon essential problems in a research project, is the support by appropiate programming systems as well as program tools and modules. A considerable part of the work in the group has been concentrated in these areas. On top of this, specialized projects are performed. Thus, we may distinguish three levels of research, the language level, the support program level and the applied problem level. The border lines are however indistinct and there is a productive interactivity between levels.

On the base level a decision was made at an early stage to choose LISP as the common programming language in the group. This was motivated not the least from the qualities of LISP as an interactive programming system, since interactivity was considered a crucial point for the chosen line of research. The main effort on this level at Datalogilaboratoriet has been the development of INTERLISP for IBM 360/370 made in co-operation with Uppsala Computer Center (UDAC). This is an interactive, virtual memory LISP system including compiler, break and editing facilities. The system has presently been sold to about 10 places and is powerful tool for non-numeric data processing.

Several other LISP implementations have been performed in the group, the most important of which is the FORTRAN coded LISP F1 interpreter, now distributed to more than 75 places throughout the world. Since LISP has a somewhat unorthodox syntax, a LISP based language PLAST has been developed. PLAST is aimed to make advanced facilities available through what looks like an ordinary programming language to the user. Another project on the language level is the design and study of an experimental recursive language REC. SIMULA is still another language that has gained some interest in the group and a formal definition of the semantics of SIMULA is given by using LISP.

On the intermediate level the preferred approach in the group is to design program generators rather than general program modules. Starting with programs that are driven by parameters and are relevant to a large class of applications, a specializing program can make automatic reductions of the general program by using substitution of known parameters and partial evaluation. Thus a tailor-made and efficient program is generated for the specific task at hand. In many cases the general program was in the first place automatically produced from some high level description of a problem.

A major program package in this class is the PCDB program, which is a program generator for management of and deduction in a data base built from predicate calculus formulas. This means that the user describes his subject area in terms of logical expressions, i.e. predicates and relations applying to a set of objects. The system then helps him to retreive explicit and implicit information by converting declared axioms to program code that searches the data base.

As an attempt to define a formalized internal representation of natural language information, a predicate calculus formulation of semantic structures has been developed. This representation then acts as a canonic form to which a variety of surface language constructions can be transformed. In order to lay this as a foundation for natural language communication between men and computers, parsing and translating programs must be designed. Work in this area has favoured state transition grammars and semantically oriented parsing.

The formidable task of making computers understand human language does not yet appear to be near a general solution. However, for limited areas of applications, significant steps have been taken towards communication with computers in a way adapted to the needs of the human user. In application projects performed at Datalogilaboratoriet, it has been shown that much of the programming expertize needed to access computer stored information could be eliminated through the use of meta data bases and special user interfaces. A meta data base contains a description of the production data base and is used e.g. in order to generate data base processing programs from queries stated in some high level language.

Another line of research represented at Datalogilaboratoriet concerns methods for the development of large and complex program systems with high reliability. Within this project work has been made on semiformal descriptions of program structure and program development structure. Models of program development have been designed and will be tested using a data storage and retrieval program for programmer support. In addition to this, special studies have been made in the fields of error classification, testing and reliability estimation methods. There has also been a socio-cybernetic research project for development of a simulation system for social processes. This system is characterized by being modular, not completely numerical, self organising and using several levels of descriptions.

To summarize, the research activities at Datalogilaboratoriet are mainly directed towards interactive processing of data structures. Herein the most essential problems are

Constructing of programming systems for this type of computer processing.

Methods for query languages and other types of communication with data bases.

Methods for retrieval/deduction in data bases.

Methods for management of large and complex programs.

3–764370