PROCEEDINGS

Industrial Computation Seminar

SEPTEMBER

1950

PROCEEDINGS

Industrial Computation Seminar

SEPTEMBER

1950

EDITED BY IBM APPLIED SCIENCE DEPARTMENT

CUTHBERT C. HURD, *Director*

INTERNATIONAL BUSINESS MACHINES CORPORATION

NEW YORK ✦ NEW YORK

International Business Machines Corporation
590 Madison Avenue, New York 22, N. Y.
Form 22-8415-0

PRINTED IN THE UNITED STATES OF AMERICA

FOREWORD

An Industrial Computation Seminar, sponsored by the International Business Machines Corporation, was held in the IBM Department of Education, Endicott, New York, from September 25 to September 29, 1950. The ninety research engineers and scientists who participated in this Seminar met to discuss the fundamental computational methods which are applicable in a wide variety of research problems. Particular attention was drawn to computational techniques recently developed in the fields of chemistry and petroleum. The International Business Machines Corporation wishes to express its appreciation to all who participated in this Seminar.

CONTENTS

PARTICIPANTS

Ackerman, Herman A., *Geophysicist*
Socony-Vacuum Oil Company
New York, New York

Aronofsky, Julius S., *Senior Research Engineer*
Magnolia Petroleum Company
Dallas, Texas

Bejarano, Gabriel G., *Research Engineer*
California Research Corporation
Richmond, California

Bell, Clarence J., *Research Engineer-Mathematician*
Battelle Memorial Institute
Columbus, Ohio

Bloom, Charles A., *Stress Group Leader*
Canadair, Limited
Montreal, Quebec

Brillouin, Leon, *Director*
Electronic Education, IBM Corporation
New York, New York

Brinkley, Stuart R., Jr., *Physical Chemist*
U.S. Bureau of Mines
Pittsburgh, Pennsylvania

Brown, William F., Jr., *Research Physicist*
Sun Oil Company
Philadelphia, Pennsylvania

Buchanan, Alva C., Jr., *Chief Accountant*
Tabulating, Magnolia Petroleum Company
Dallas, Texas

Buchholz, Werner
Engineering Laboratory, IBM Corporation
Poughkeepsie, New York

Carlson, Harrison C., *Research Project Engineer*
E. I. duPont deNemours and Company
Wilmington, Delaware

Chancellor, Justus
Applied Science Department, IBM Corporation
New York, New York

Clamons, Eric H., *Research Engineer*
Minneapolis-Honeywell Regulator Company
Minneapolis, Minnesota

Collins, Francis, *Associate Reservoir Engineer*
Atlantic Refining Company
Dallas, Texas

Danforth, Clarence E., *Technical Engineer*
General Electric Company
Lynn, Massachusetts

de Finetti, Bruno, *Professor*
Department of Mathematics, University of Trieste
Trieste, Italy

Dempsey, Carl W., *Assistant Research Mathematician*
Sun Oil Company
Philadelphia, Pennsylvania

Donnell, John W., *Professor*
Department of Chemical Engineering, Michigan State College
East Lansing, Michigan

DuFort, Edward C., *Associate Reservoir Engineer*
Continental Oil Company
Ponca City, Oklahoma

Eaton, Thomas T., *Engineering Group Supervisor*
Radio Corporation of America, RCA Victor Division
Camden, New Jersey

Eckert, Wallace J., *Director*
Department of Pure Science, IBM Corporation
New York, New York

Elkins, Thomas A., *Geophysicist*
Gulf Research and Development Company
Pittsburgh, Pennsylvania

Feigenbaum, David, *Associate Research Engineer*
Cornell Aeronautical Laboratory
Buffalo, New York

Freud, Oliver, *Senior Research Engineer*
The Budd Company
Philadelphia, Pennsylvania

Fullerton, Paul W., Jr.
Applied Science Department, IBM Corporation
New York, New York

Glauz, Roy L., Jr., *Process Engineer*
Standard Oil Company
Cleveland, Ohio

Greene, Charles H., *Manager*
Melting Development, Corning Glass Works
Corning, New York

Greenfield, Alexander, *Senior Electronic Engineer*
Research Laboratories, Bendix Aviation Corporation
Detroit, Michigan

Greenlaw, David S.
Color Control Department, Eastman Kodak Company
Rochester, New York

Grosch, H. R. J., *Senior Staff Member*
Watson Scientific Computing Laboratory, IBM Corporation
New York, New York

Grosh, L. E., Jr., *Research Associate*
Statistical Laboratory, Purdue University
West Lafayette, Indiana

Harrington, Robert A., *Physicist*
B. F. Goodrich Research Center
Brecksville, Ohio

Hastings, Brian T., *Project Engineer*
Office of Air Research, USAF, Wright Field
Dayton, Ohio

Hoagbin, Joseph E., *Physicist*
AC Spark Plug Division, General Motors Corporation
Flint, Michigan

Hoelzer, Helmut, *Chief*
Computing Laboratory, Ordnance Guided Missile Center
Redstone Arsenal
Huntsville, Alabama

Hunter, G. Truman
Applied Science Department, IBM Corporation
New York, New York

Hurd, Cuthbert C., *Director*
Applied Science Department, IBM Corporation
New York, New York

Hurley, Wesley V., *Air Design Specialist*
General Electric Company
Lynn, Massachusetts

Johnson, Walter H.
Applied Science Department, IBM Corporation
New York, New York

Keefer, Karl H., *Assistant Project Engineer*
Aeroproducts Division of General Motors Corporation
Dayton, Ohio

Kinckiner, Roy A., *Assistant Director*
Engineering Research Laboratory, E. I. duPont Experimental Station
Henry Clay, Delaware

King, Gilbert W., *Research Chemist*
Arthur D. Little, Inc., and Research Laboratory for Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts

Krawitz, Eleanor
Watson Scientific Computing Laboratory, IBM Corporation
New York, New York

Leslie, John D., *Research Engineer*
Standard Oil Development Company
Linden, New Jersey

Ley, Darwin M., *Senior Systems Analyst*
Ford Motor Company
Dearborn, Michigan

Liggett, Irving C.
Applied Science Department, IBM Corporation
New York, New York

Lindley, Charles A., *Development Engineer*
Thompson Aircraft Products
Cleveland, Ohio

Lucas, Robert R., *Comptroller*
Monmouth Products Company
Cleveland, Ohio

McAdams, H. T., *Research Chemist*
Aluminum Ore Company
East St. Louis, Illinois

McIntire, Robert L., *Chemical Engineer*
Phillips Petroleum Company
Bartlesville, Oklahoma

Merrick, Elsie V., *Engineer*
Technical Service Division, Standard Oil Company
Cleveland, Ohio

Moncreiff, Bruse, *Junior Methods Analyst*
Methods Division, Prudential Insurance Company of America
Newark, New Jersey

Nichols, Nathaniel B., *Professor*
Department of Electrical Engineering, University of Minnesota
Minneapolis, Minnesota

Nims, Paul T., *Staff Engineer—Research*
Chrysler Corporation
Detroit, Michigan

O'Brian, Wade B., *Supervisor*
Payroll, Tabulating, Timekeeping
The Cleveland Graphite Bronze Company
Cleveland, Ohio

Olsen, John L., *Development Engineer*
Sun Oil Company
Marcus Hook, Pennsylvania

Opler, Ascher, *Project Leader*
Physics Laboratory, Great Western Division
The Dow Chemical Company
Pittsburg, California

Orr, S. Robert, *Assistant Research Physicist*
Monsanto Chemical Company
Miamisburg, Ohio

Parker, Robert W.
Engine Performance Calculations, Allison Division
General Motors Corporation
Indianapolis, Indiana

Peters, Leo J., *Chief*
Geophysical Operations Division
Gulf Research and Development Company
Pittsburgh, Pennsylvania

Ramser, John H., *Physical Chemist*
The Atlantic Refining Company
Philadelphia, Pennsylvania

Randall, Lauros M., *Project Engineer*
Allison Division, General Motors Corporation
Indianapolis, Indiana

Randels, Robert, *Physicist*
Corning Glass Works
Corning, New York

Roberts, John B., *Group Supervisor*
E. I. duPont deNemours and Company
Wilmington, Delaware

Roggenbuck, Robert A., *Research Engineer*
Engine Section, Ford Motor Company
Dearborn, Michigan

ROSE, ARTHUR, *Associate Professor*
Department of Chemical Engineering
Pennsylvania State College
State College, Pennsylvania

RUBINOFF, MORRIS, *Research Assistant Professor*
University of Pennsylvania, Moore School of Electrical Engineering
Philadelphia, Pennsylvania

SCHUMACHER, LLOYD E., *Assistant Chief*
Flight Research Section, Headquarters Air Materiel Command
Dayton, Ohio

SCIFRES, EUGENE M., *Research Engineer*
Gates Rubber Company
Denver, Colorado

SELLS, BERT E., *Turbine Engineer*
Aircraft Gas Turbine Division
General Electric Company
West Lynn, Massachusetts

SHELDON, JOHN W.
Applied Science Department, IBM Corporation
New York, New York

SHERMAN, JACK, *Mathematician*
The Texas Company
Beacon, New York

SHIVELY, RICHARD D., *Assistant Manager*
Office Services Department, Gates Rubber Company
Denver, Colorado

SMITH, EDWARD A., *Chief Mechanical Engineer*
Raymond Concrete Pile Company
New York, New York

SMITH, EDGAR L., JR., *Acting Chief*
Data Analysis, Long Range Proving Ground Division, USAF
Cocoa, Florida

SMITH, ROBERT W., JR., *Mathematician*
U.S. Bureau of Mines
Pittsburgh, Pennsylvania

TANNICH, RICHARD E., *Research Specialist*
Humble Oil and Refining Company
Baytown, Texas

TAYLOR, CHARLES R., *Supervising Metallurgist*
Armco Steel Corporation
Middletown, Ohio

ULLOCK, DONALD S., *Staff Engineer*
Union Carbide and Carbon Corporation
South Charleston, West Virginia

WAKEHAM, HELMUT, *Section Head*
Textile Research Institute
Princeton, New Jersey

WALKER, JACK K., *Physicist*
Socony-Vacuum Research and Development
Paulsboro, New Jersey

WATSON, FREDERIC R., *Engineer*
Products Application Department, Shell Oil Company
San Francisco, California

WATSON, H. J. MICHAEL, *Special Cost Clerk*
Steel Company of Canada, Limited
Hamilton, Ontario

WEINKAMER, WILLIAM A., *Test Engineer*
Harris Products Company
Cleveland, Ohio

WELKER, E. L., *Associate in Mathematics*
American Medical Association
Chicago, Illinois

WHITNEY, ALICE M.
Applied Science Department, IBM Corporation
New York, New York

WILLIAMS, THEODORE J., *Research Fellow*
Department of Chemical Engineering
The Pennsylvania State College
State College, Pennsylvania

WILSON, L.
IBM Corporation
New York, New York

ZEIGLER, MARTIN L., *Assistant Supervisor*
Tabulating Division, The Pennsylvania State College
State College, Pennsylvania

ZIEGLER, GEORGE E., *Director of Research*
Midwest Research Institute
Kansas City, Missouri

The Role of the Punched Card in Scientific Computation

WALLACE J. ECKERT

International Business Machines Corporation

AS I LOOK over the list of occupations of the members of this Seminar, I am impressed by the wide range of fields represented—engineering, physics, chemistry, accounting, even astronomy. It is interesting to note how experience in one field can influence seemingly unrelated activities in other fields. I might illustrate by a trivial example of a procedure mentioned here in connection with an accounting problem which also occurred recently in an astronomical problem at the Watson Laboratory. It was mentioned that it is frequently more convenient to produce the calendar date on the accounting machine than to copy it from the calendar. In our problem we required calendar dates at forty-day intervals from 1653 to 2060, taking into account the complicated leap-year rules of our calendar; the list was prepared by a single run on the IBM Type 602-A Calculating Punch.

The close relationship between apparently unrelated things is not new in science; it is necessary to look at the picture in the proper perspective to see the relationship. Let us consider the ancient astronomer who was intrigued by the small spots of light in the sky, called planets. He spent many hours measuring and recording their positions, and I am sure that his contemporaries could not see how these activities would ever put food into the mouths of men. His contemporaries could not see enough of the picture. We now know how the study of such planetary observations has led not only to our understanding of the motions of the planets, but also to our knowledge of the fundamental principles of mechanics, the basis of all mechanical design. From this vantage point, when we see a farmer riding a properly designed tractor and pulling carefully designed implements, we realize that the early astronomer has done more to feed the multitudes than all of his contemporaries.

The computing profession has always incorporated mathematical and mechanical techniques; benefits from one field of science have been carried into other fields. At the time of the early astronomical observations computing was being done on a considerable scale, and from that time forward man has tried to develop computing aids. The mathematician and the scientist have tried to devise both mechanical aids and mathematical aids. The first astronomer needed trigonometric tables in order to make his computations. Later, the development of the logarithmic table greatly facilitated his arithmetic operations. The adding machine was invented by Pascal in 1642 and the desk calculator by Leibnitz in 1693. Thus, we have the invention by scientists of these two tools that were greatly needed by scientists; yet they were of little use to science for over two centuries. Although the desk calculator of Leibnitz was, in principle, our present-day machine, two centuries were required to develop it into a generally useful implement.

There are two reasons for this long delay: one is mechanical, and the other is mental. It is a long, hard pull from the gleam in the inventor's eye, or even from the first model, to the point where a scientist can use a device as a help and a tool, rather than as a problem in itself. It is not difficult for the inventor to make his model work under his own benevolent criticism, but to have it developed to the point where it is accurate, fool-proof, and efficient is another matter. Of course, things did not proceed as rapidly in those days as they do now, but one should not overlook the great number of technical developments necessary to fill in the details that make a complicated machine work.

On the mental side there was the fact that people had been trained in the use of logarithms, and computational work had been organized for the use of logarithms. If you look through many books in applied mathematics of that era, you will find that large portions of many of them were devoted to the conversion of basic formulae into a form suitable for the effective use of logarithms. Then, too, there was the matter of tables. To replace logarithmic tables with natural tables required some time. This seems like a modern age, yet I am not an octogenarian and I can remember the dying gasp of the logarithmic table as the standard method of computation. I have seen the desk calculator become a necessary instrument for every scientist who is doing quantitative work. They are now so efficient and so reliable that the scientist has merely to insert the proper number, push the proper control key to perform his desired operation, and read the results.

In medicine there is a well-known phenomenon. While smallpox and diphtheria were decimating mankind, little attention was paid to some of the lesser diseases, but now that smallpox and diphtheria have been brought under control the lesser diseases are considered very important. Similarly in computation, when multiplication and division were performed longhand, or with the aid of logarithms, the computer did not worry about the associated clerical operations. The advent of the desk calculator, however, enabled him to undertake larger problems, and he has become painfully aware of the details of reading and writing data and of initiating the proper control operations. The need for automatic handling of data and instructions becomes important for further progress.

Here again, we have a situation similar to that of the development of the desk calculator. In 1893, as you have heard, the punched card was introduced as a means of reading data and instructions into a machine, but many years elapsed before the method became accepted as a regular part of scientific research. This delay also has been due to the necessity of both mechanical and mental development. When Mr. Watson became president of the International Business Machines Corporation in 1914, he immediately organized a development program to make the machines more versatile. From 1914 to 1930 there was extensive development of the various functions of the machines such as reading cards, sorting, printing, adding, subtracting, multiplying and recording. During the following two decades further remarkable technical development in the machines occurred, but in my opinion the mental development among the scientists and engineers has been more striking. During this time there has come the general realization that the punched card had already provided the means of automatically handling scientific and engineering data. The recent introduction of the electronic circuit, which has greatly increased the speed of some operations, has dramatized the automatic process, but the hundreds of successful automatic computing installations now in operation have their roots in the mental revolution of the past two decades.

It is interesting to note that during this period Mr. Watson saw the importance of automatic computation to science more clearly than the scientist or the engineer, and the present widespread use of such facilities is due in large measure to his early efforts. These efforts included not only the rapid development of standard machines to make them more generally useful, but the development of many special devices for academic purposes. In 1928 he established the Columbia University Statistical Bureau for educational research, and soon after installed there a special statistical calculator. The operation of this machine was very striking even by today's standards. It would read data and limited operating instructions from the cards at the rate of nearly a million digits an hour; it would add 100 digits simultaneously from the cards or from other parts of the machine according to a complicated program, and print the results at the rate of nearly one-half a million digits an hour. In 1933 a second laboratory was established at Columbia capable of handling general scientific calculations such as the solution of differential equations, and the reduction of observational data. This laboratory was in full-time operation on basic research until the advent of the war when it was converted to military research. By 1940 a number of laboratories about the country were using standard punched card machines for technical computation. During the war new laboratories were quickly established in all phases of the defense effort, including atomic energy, aircraft design and construction, air and sea navigation, and many others that are still classified. In 1944 the IBM Automatic Sequence Controlled Calculator was completed in Endicott and presented to Harvard University; this machine, known as Mark I, has since been in continuous service. In the same year two relay calculators were installed at the Aberdeen Proving Ground; these machines were more limited in capacity and flexibility but were about twenty times as fast as the sequence calculator. They are still the fastest relay calculators in operation.

In 1945, a special table printing device was installed at the Naval Observatory which enabled the scientist to print mathematical tables from punched cards in a form suitable for direct reproduction by the printer. In the same year, the recording equipment for the great wind tunnel at California Institute of Technology was installed so that the observational data could be recorded directly in cards without hand transcription. Then came the 603, the first commercial electronic calculator, which has been replaced by the more versatile IBM Type 604 Electronic Calculating Punch. The Selective Sequence Electronic Calculator, which was dedicated in January, 1948, provided electronic speed of operation together with an internal storage capacity of half a million digits and completely automatic programming. The Card Programmed Electronic Calculator is the most recent addition.

From such a wide variety of available punched card machines—sorters, accounting machines, collators, reproducers, the 602, the 602-A, the 604, the CPC, and the SSEC—many of you are probably wondering which equipment you should use and why. Since much of the later part of the program deals with the more recent machines such as the 604 and the CPC and their detailed application, I shall confine my attention to some of the more basic uses of the punched card and of the simpler punched card machines. I shall take my examples from some of the earlier work; it is interesting to note that in many applications these early techniques are still the most efficient in spite of all the advances that have been made in design. Moreover, many of these early techniques clearly illustrate basic principles that can be readily applied in general.

A question of basic importance in the application of the punched card is whether the calculation should be done sequentially or in parallel. In a computation with a desk calculator the problem also arises, and we shall illustrate it with a simple calculation. Let us assume that we wish to evaluate the formula

$$f(x) = a + bx + cx^2 + d\sin(a + bx + cx^2)\ .$$

Ordinarily we should write the formula across the top of the sheet and assign successive columns for the intermediate results. Successive rows would be assigned for successive values of x. We could perform the computation by evaluating the formula completely for the first value of x, including the use of the sine table. In this case the work would be completed a row at a time, or sequentially. On the other hand, we could perform the first operation, say $b \cdot x$, for all values of x before proceeding to the next operation. This procedure by columns is termed parallel computation. The experienced hand computer knows that he can perform operations such as table lookup most efficiently by the parallel method. The user of the desk calculator, however, can frequently avoid the recording and reading of intermediate results by using the sequential method of operation, for example, the formation of $a + bx + cx^2$. Thus, he might combine the sequential and the parallel methods. In the use of the punched cards, parallel operations become more important because of the ease of reading and recording intermediate results in the cards.

The ability to store numbers temporarily in the machine largely determines the length of the sequence that can be handled before the recording of intermediate results. The 602-A, for example, can store a dozen numbers while the SSEC can handle 20,000. Perhaps we should all like to use a machine that can handle a million digits, but economic considerations must be taken into account. It must be remembered that a small box of cards costing about ten dollars has more storage capacity than a million-dollar calculator. The desirable amount of storage in the machine is thus an economic matter, and the amount of storage that is needed depends to a considerable extent on the nature of the problem. Where parallel computation can be employed, and a very large portion of technical work can be so handled, the small machine with card storage is indicated.

The length of sequence that a given machine will handle depends on the facilities for handling instructions as well as on the storage capacity. On the 602-A it is easy to wire, say, ten, twenty or thirty successive operations; many more than that are possible, but the wiring becomes somewhat troublesome. The CPC has greater storage capacity than the 602-A and also the ability to read operating instructions from punched cards. It can therefore handle sequences of any length, limited only by the storage capacity. The SSEC has a tremendous storage capacity and facilities for the most intricate programs. Not only can instructions be read from cards, but the entire storage capacity of the machine can be used for manipulating the operating instructions.

Efficient use of the punched card also requires that full advantage be taken of the facilities for rearranging data. The simple sorter can rearrange cards at the rate of 500 or 600 a minute. When you consider that each card carries eighty digits, it is easy to see that in a day you can rearrange and reassociate millions of digits. Previous to the punched card there was no facility for automatically rearranging data; the computer wished to avoid such manipulation rather than encourage it. In adopting a new medium for computation it is desirable to profit from past experience, but not to limit one's thinking by the limitations of the past.

To illustrate new possibilities introduced by the ability to rearrange data, I shall describe a large-scale operation carried out in 1928. The results of a survey contained 50,000 cases with a dozen variables. These data were punched in 50,000 cards, and in a week or two all correlations between all of the variables were computed. Those of you who are familiar with correlation analysis know that this work involved the formation of the sums of millions of products. There was no multiplying machine available, but by sorting the cards and adding from them in the proper order, we obtained the required results and printed them automatically. A small mental calculation shows the efficiency of this method. By feeding eighty-digit cards at the rate of 9,000 cards an hour it is possible to accumulate partial products at the rate of 720,000 digits an hour. It is necessary to make a run for each digit of the multiplier and to allow space in the counters for accumulation. Allowing for three-digit factors it is easy to obtain 40,000 products an hour, a considerable achievement even in modern terms. This method, known as progressive digiting, or similar techniques continue to recur as an effective method in many problems including harmonic analysis and the formation of normal equations. Since the method requires only a key punch, a sorter, and an accounting machine, it has been applied where only simple accounting installations are available.

An important change in emphasis brought about by the ability to rearrange data is in the use of mathematical tables. In hand computation the use of tables is laborious and is frequently avoided. The punched card, however, makes the consulting of tables one of the most efficient operations. The sorter will rearrange cards, and each card may contain eighty columns of tabular data. For example, a single card can carry an angle, the sine, the cosine, the tangent and any other data associated with the angle. The sorter will put the cards in the proper place, and the reproducer will transfer the data into the computing cards at the rate of 6,000 cards an hour. This means that with this simple equipment you can *look up* half a million digits of tabular data in an hour. This ability to find millions of digits of tabular data in a day with the simplest equipment and a few cards is so fundamental that the scientist must re-examine his whole concept of computation. He no longer says "How can I avoid table lookup?" but "How can I use more tables?"

Before leaving the subject of tables, I might mention one example of their use. In 1940 the U. S. Naval Observatory in Washington had the task of producing the American Air Almanac for the use of navigators at sea and in the air. It contains 730 pages a year, and each page contains several thousand digits. Since the lives of the crew depend upon it, the highest accuracy is demanded. The entire computation was performed automatically on standard punched card machines. Three months elapsed from the time the book was decided upon until the machines were delivered and in operation. The necessary planning and preparation of basic data had been done meanwhile. The first volume was prepared and checked, published by the Government Printing Office and distributed to the fleet before the first of the following year. For the benefit of those who instinctively think of the most recent and most powerful equipment for a given problem, the work here described was performed with only the key punch, the sorter, the accounting machine, and the reproducer-summary punch. The multiplier was available, but it was not used because the other method was more efficient. The whole volume was produced by continuous summation in parallel columns involving the sexagesimal system. The summations were recorded for a ten-minute interval, and the data to be summed were obtained from special punched card tables prepared for the purpose. The standard accounting machine performed the accumulation in the sexagesimal system by means of fractional counters that carry on 6 instead of 10 (these fractional counters are regularly available). After the initial volumes had been prepared, a special table printing device was installed to transcribe the resulting data from the cards in a form suitable for reproduction by means of line cuts and electroplates. In the twelve years that the publication has been produced, there have been computed and published about twenty-five million digits, which have been examined and used by thousands of navigators, and as yet not a single error has been reported.

Another example of the use of simple punched card equipment dates back to 1935; this is a large-scale reduction of observational data, which should interest many here who are associated with organizations that regularly handle large quantities of industrial, engineering, or scientific data. Here, again, we have an example of a procedure that applies equally well in any field; the punched card machines cannot tell the difference between data concerning a chemical compound or the wing of an airplane. The program handled in 1935 concerned the measurement of astronomical photographs. Each plate contained several hundred black spots or star images. Since the blackness and size of the image depend upon the brightness of the star, the brightness may be determined by the measurement of the amount of light obscured by the image in a photometer. The total program involved several hundred plates with several hundred thousand images to be measured. The problems involved were those common to all measurements: identifying the object to be measured, recording the instrumental reading, calibrating the instrument, applying instrumental corrections, performing mathematical transformations, combining results, discussing the errors statistically, and recording the results for publication. The punched card method permits all these operations to be carried out more thoroughly than before, without error, and with a minimum of drudgery. The identification of the image at the instrument was made with the aid of scale settings computed for each image on the card; these settings were printed on the card by the interpreter. The observer placed the card in a punch and recorded the instrumental settings directly. The operator applied the instrumental corrections to the cards by sorting the cards into groups and gang punching the appropriate correction.

The conversion from corrected instrumental reading to star brightness was determined by an empirical process; this conversion was made previously by means of a calibration curve for each plate. Each plate contained several hundred images whose brightness was to be determined and about forty comparison images of known brightness. Previously, the known images were plotted and the results for the unknown ones read from the curve. The plotting and reading of these curves had been very laborious. For the punched card method investigation showed that, if a standard calibration curve were used for all of the plates, the resulting errors for a given plate could be represented as a linear function of the abscissas. Therefore, it was decided to use a standard calibration curve in the form of a punched card file and to determine the two constants of the linear function for each plate. We determined these constants by the method of least squares, using an equation for each of the thirty or forty comparison stars.

The remaining operations consisted of arithmetic operations which were performed mostly in the parallel manner on simple equipment such as the 601. The statistical analysis of the errors was primarily a matter of sorting and counting.[a]

As a final example of simplifying the punched card procedure we shall describe a case where the longest way around is the shortest way home. Frequently a computing problem appears intricate and complicated, involving iterative processes and long sequential expansions. Yet, if the problem is turned around, the advantages of the punched card method can be applied directly, and the solution is obtained. The problem arises in connection with a method of radio navigation developed during the war and known as Loran. In this method each of two fixed stations broadcasts a signal which is received by the ship or plane. The receiving instrument indicates the difference between the dis-

[a] Details of this work are given in *Punched Card Methods in Scientific Computation,* W. J. Eckert (Thomas J. Watson Astronomical Computing Bureau, New York, 1940), Chapter X.

tances from the two stations; the problem for the navigator is to find the positions his ship could occupy in order to obtain the observed difference in distance. Since long distances are involved, it is necessary to treat the earth as a spheroid, and when the equations are examined it is found that troublesome expansions are involved. However, when you turn the problem around and compute the distance from an assumed point on the spheroid to one of the broadcasting stations, the computation becomes simple and straightforward. The solution was to adopt a uniform grid of points covering the desired area and to compute by simple parallel methods all the distances involved. The resulting distances, recorded in cards, were used as a table, and the required positions were determined by inverse interpolation. Although the grid consisted of a two-dimensional array of points, it was necessary to perform interpolation in only one coordinate, the other being held fixed. Since the data were in cards, it was possible to separate them into sections to facilitate the work. In some areas interpolation along a parallel of latitude gave greater determinateness, and in others interpolation along a meridian was preferable. The tabular interval was chosen to permit linear interpolation in most areas, and the critical areas could be removed for the use of second-order interpolation where necessary. Finally, the data were arranged with the sorter and listed in the most convenient order for the draftsman who was to plot them on a chart.

In conclusion I wish to emphasize the value of examining each problem as a job for the simplest equipment rather than for the most recent and most powerful equipment in operation. In the two and one-half years that the SSEC has been in operation we have been impressed by the number of problems proposed for solution on it that can be handled effectively on a standard accounting machine.

Machine Calculation of the Plate-by-Plate Composition of a Multicomponent Distillation Column

ASCHER OPLER ROBERT G. HEITZ

The Dow Chemical Company

ONE of the most frequently performed groups of chemical engineering calculations is that involving fractional distillation columns. In a continuous distillation tower, a mixture of two or more volatile liquids is fed into the column. Depending on their relative volatilities, the components redistribute themselves on a series of trays or plates (sometimes these "plates" are merely theoretical equilibrium points) above and below the feed. At each theoretical plate, equilibrium is reached between the descending liquid stream and the ascending vapor stream. At the top and the bottom of the column (and often at intermediate points) the partially separated portions are taken from the column. At the top, part of the vapor is condensed and returned as liquid reflux to the column; at the bottom, some of the liquid is boiled and goes up the column.

The properties of each compound in the feed mixtures are known. The total composition is also known. The column may operate at atmospheric pressure, or it may be working at high or low pressures. The chemical engineer desires to know the necessary amount of reflux, the height of the column (i.e., number of plates), the temperature and composition on each plate, and the composition of the top and bottom product. For a multicomponent mixture, these calculations can become quite complex and laborious.

There is considerable variation in the detail used in performing these calculations. For example, if one wants only an estimate of the number of plates required for a particular separation, an empirical formula will provide this answer. It is also possible to perform an extremely rigorous calculation, accepting no simplications or shortcuts. However, the most frequent choice lies somewhere in between.

In our work we use the Lewis and Matheson method, which is quite well-known to chemical engineers. In this method, the composition of each plate is determined from the composition of the previous plate by combining the equations stating the equilibrium between the liquid and vapor leaving a plate with the material balance equations relating the compositions and quantities of liquor and vapor at any level in the column. This may be expressed as in the following four equations:

1. For use above the feed in calculating up the column:

$$x_n = x_{n-1}\frac{p}{P}\frac{V}{L} - x_0\frac{D}{L}$$

2. For use below the feed in calculating up the column:

$$x_n = x_{n-1}\frac{p}{P}\frac{V'}{L'} + x_b\frac{W}{L'}$$

3. For use above the feed in calculating down the column:

$$x_n = x_{n+1}\frac{P}{p}\frac{L}{V} + x_0\frac{D}{V}\frac{P}{p}$$

4. For use below the feed in calculating down the column:

$$x_n = x_{n+1}\frac{P}{p}\frac{L'}{V'} - x_b\frac{W}{V'}\frac{P}{p}$$

NOMENCLATURE

L = liquid flow leaving plate—above feed (mols).

L' = liquid flow leaving plate—below feed (mols).

V = vapor leaving plate—above feed (mols).

V' = vapor leaving plate—below feed (mols).

D = product leaving top of column (mols).

W = product leaving bottom of column (mols).

p = vapor pressure of a component (millimeters of mercury).

P = total operating pressure (millimeters of mercury).

x_n = the concentration of a component in the liquid leaving the nth plate numbering from the bottom (mol per cent).

x_b, x_0 = concentration of bottom and overhead products, respectively (mol per cent).

Simplifying Assumptions

It is to be noted that the use of the vapor pressure of the pure component implies that the equilibria in the system follow Raoult's Law. This is a common assumption used in absence of data to the contrary; if more exact data were at hand, we believe it would be practical in many cases to calculate a suitable pseudo vapor pressure to use for a given system. It might be remarked at this point that the use of vapor pressure data, rather than *alpha* or *k* values, was chosen, because the same cards can then be used for a given compound in many different mixtures without further work. We also expect to use these same cards for other calculations such as flash vaporizations.

It will be found necessary to make one further simplifying assumption in common use. This assumption is that the reflux ratios L/V are constant throughout the column. As will be brought out later, we expect to be able to handle variable L/V by expressing it as a function of temperature.

Machine Computation Method

In determining the composition of one plate from the preceding plate, it is necessary to set up one of the above equations for each component and solve it for the new concentration. Only when the set of vapor pressures is selected at the correct temperature does the sum of the calculated concentrations equal 100 per cent. As a rule in calculating manually, an approximately correct set of vapor pressures is obtained by estimating the temperature of the plate. In that case, each calculated value is divided by the sum of the concentrations to yield a distribution totaling 100 per cent.

In the machine method, advantage is taken of the high speed of calculation (four substitutions into one of the above equations requires eight seconds). A group of temperatures with one degree centigrade intervals is selected and a calculation made for *each* set of vapor pressures. For every set, the composition is totaled and the sum is inspected for proximity to 100 per cent. The calculation yielding the closest set of values is taken as the correct one. This new composition is used to calculate the composition of the next plate and so on. Although many more calculations are made than are actually used, the rapidity of the machine more than compensates for the extra computation.

While this operation has been performed using an IBM Type 602 Calculating Punch, there is no reason to believe that it cannot be performed with the IBM Type 602-A Calculating Punch or the IBM Type 604 Electronic Calculating Punch. We have been able to calculate columns with up to four components on the 602 machine. Six component calculations could be calculated with the 602-A if all counters and storage units are provided.

At first glance, the solution and testing of four sets of equations as above seems difficult for the machine to perform on each card cycle. However, as several of the terms are constant for one portion of the column, they may be combined. Thus, the equations may be reduced to the forms:

$$x_n = x_{n-1} \cdot K \cdot p \pm C$$

for calculating up the column;

$$x_n = x_{n+1} \cdot K' \cdot \left(\frac{1}{p}\right) \pm C' \cdot \left(\frac{1}{p}\right)$$

for calculating down the column;

$$\text{where } K = V/PL$$

$$K' = PL/V$$

$$C = x_0 \frac{D}{L} \text{ or } x_b \frac{W}{L'}$$

$$C' = x_0 \frac{DP}{V} \text{ or } x_b \frac{WP}{V'} \ .$$

In a preliminary step, the constants are multiplied by the vapor pressures or their reciprocals as required. The equations then reduce to the form:

$$x_n = x_{n+1} \cdot K'' \pm C''$$

$$\text{where } K'' = K \cdot p \text{ or } K'\left(\frac{1}{p}\right)$$

$$C'' = C \text{ or } C'\left(\frac{1}{p}\right)$$

The 602 may be wired to calculate four sets of these equations and to accumulate the sum of the results. The over-all procedure used in the calculation is determined when the chemical engineer fills out a form which contains sufficient information to enable the tabulating department operator to proceed with the calculation. The steps are as follows:

Preparation of Vapor Pressure Cards

These need be prepared only once for each compound used (unless systems are encountered requiring different pseudo vapor pressures). After calculation, the cards are filed for future use. As the cards may be readily duplicated, it is possible that prepared sets of vapor pressure data for common compounds will become available. The data are calculated by means of the Antoine equation: $\log_{10}p = A - B/(T+C)$, although it is perhaps better to interpolate actual physical measurement of the vapor pressure. Briefly, the preparation of the cards begins with the gang punching of A and B on a reproduced set of $(T+C)$ masters. These masters contain the reciprocals of $T + 230$ (the value of C used here) with the corresponding T (temperature) from -100 to 270°C. $A-B\ [1/(T+230)]$ is calculated quite simply. The results are merged with a set of logarithm cards, and both p and $1/p$ are punched on each detail card. Thus, we have a set of 370 cards containing temperatures, vapor pressures, and reciprocal vapor pressures. If several sets are prepared at once, the time required per set reduces to approximately twenty minutes.

Preparation of Master System Cards

A portion of the vapor pressure cards is selected which will safely include any temperatures encountered in the column. Cards are selected for each compound and are successively reproduced into a single set of system masters. These cards form a set of operating vapor pressures (and reciprocals) for the problem. They may be used for any problems involving the same mixture. Figure 1 contains a listing of the vapor pressure portion of cards for mixtures of carbon tetrachloride, trichloroethylene, 1,1,2 trichloroethane, and perchloroethylene.

Preparation of Master Working Cards

In this operation, the vapor pressures (or reciprocals) are multiplied by the appropriate constants. These cards are used for one problem only.

Reproduce Detail Working Cards

The master cards are reproduced into the cards that will be actually used in the plate calculation. Several sets of details are prepared. If additional cards are required, they may be reproduced later from the masters.

C°	CCl_4	C_2HCl_3	$C_2H_3Cl_3$	C_2Cl_4
75	720	513	219	171
76	742	531	227	178
77	765	548	235	184
78	789	565	244	191
79	813	583	252	198
80	838	602	261	206
81	863	621	271	213
82	889	640	280	221
83	916	660	290	229
84	943	680	300	237
85	970	701	311	246
86	999	723	322	254
87	1020	746	333	263
88	1050	767	344	273
89	1088	789	356	282
90	1110	813	367	292
91	1150	837	380	302
92	1180	862	392	312
93	1210	887	405	323
94	1250	913	419	334
95	1280	939	432	345
96	1320	966	446	356
97	1350	993	460	368
98	1390	1030	475	380
99	1420	1050	490	393
100	1460	1080	506	405
101	1500	1110	522	418

FIGURE 1

Actual Plate-by-plate Calculation

The working cards contain K'', C'', and the $X78$ that determines the sign of the C. Starting at the bottom, for example, the initial composition is gang punched into a set of details. The set is run through the machine, which stops and signals when the total composition passes 100%. The *correct* (closest to 100% total) card is selected and marked x_1. This card is placed at the head of another group of detail cards, gang punched, and then set aside for the listing. This new group is calculated in the same way. This successive calculation continues until the terminating conditions are encountered. At this point, either a new section of the column is calculated or the calculation is discontinued.

Printing the Results of the Calculation

The correct cards are accumulated into a deck during the plate calculation with the bottom of the column on the bottom of the deck and the remaining cards in ascending order of the plate numbers. The cards are run through the 405 accounting machine and listed. A partial listing for a distillation involving the same compounds shown in Figure 1 is given in Figure 2. This listing is sent back to the chemical engineer as the results of the calculations.

To illustrate the method using a well-known example, the three component system, benzene-toluene-xylene, was calculated using the conditions described by Robinson and Gilliland in *Elements of Fractional Distillation*. Figure 3 shows the comparative results obtained by Robinson and Gilliland and by the use of this method. As these writers made exactly the same assumptions as are made here, the results should be comparable. While the feed plate, xylene disappearance plate, and total number of plates are the same, there are definite numerical discrepancies. The use of smaller temperature intervals (one degree steps instead of five) should slightly increase our accuracy. The three factors discussed below, however, make the machine calculations less accurate.

Use of Calculated Vapor Pressure Data

One cause of the discrepancy is the use of different sources of vapor pressure data. Robinson and Gilliland have read points from graphed physical data. Our vapor pressures were calculated by means of the Antoine equation.

In the range used for the calculation, the vapor pressure discrepancy had a mean value of 1.15% for benzene, 0.57% for toluene, and 2.25% for xylene. If the same data were used and correctly interpolated, this discrepancy would disappear.

Plate	T°	CCl_4	C_2HCl_3	$C_2H_3Cl_3$	C_2Cl_4	Sum
41	78	98.68	.08			98.74
40	77	97.22	.11			97.31
39	78	98.57	.15			98.70
32	78	99.03	.88			99.89
31	77	97.56	1.08			98.62
30	78	98.80	1.38	.01	.01	99.82
29	77	97.33	1.73	.01	.03	99.08
28	78	98.68	2.29	.01	.28	101.25
27	79	97.21	2.80	.05	1.15	101.22
26	82	92.72	3.37	.23	4.27	100.60
FEED						
25	81	81.48	3.63	.62	13.48	99.21
24	82	82.32	4.58	.63	13.62	101.16
23	82	80.79	5.62	.63	13.62	100.67
6	96	5.82	68.48	1.76	23.14	99.20
5	101	3.83	61.36	2.60	32.35	100.15
4	107	2.24	47.69	3.63	47.10	100.67
3	113	1.19	31.59	4.52	63.69	101.00
2	117	.52	17.68	4.85	77.59	100.65
1	120	.28	8.88	4.60	87.50	101.27
W		.10	4.00	4.00	92.00	100.10

Figure 2

	Temp°C	Benzene		Toluene		Xylene	
Plate No.	O&H	O&H	R&G	O&H	R&G	O&H	R&G
D	80	94.26		.01			
16	80	97.26	98.8	.56	1.1		.0007
15	81	99.30	97.4	1.56	2.4		.003
14	82	96.75	95.3	3.22	4.5		.015
13	83	93.28	92.2	5.91	7.65	.01	.06
12	84	87.72	87.3	9.88	12.3	.09	.18
11	86	82.29	80.2	15.65	18.9	.42	.6
10	89	74.57	71.2	23.45	26.7	1.60	2.0
9	92	63.73	60.5	31.23	33.6	5.58	5.8
8	95	54.86	52.1	38.63	42.0	5.77	5.9
7	100	43.47	40.2	50.36	53.5	6.12	6.3
6	105	29.92	27.6	63.66	65.7	6.55	6.7
5	108	17.97	16.9	73.86	75.9	6.90	7.1
4	110	9.99	9.5	81.65	83.0	7.35	7.5
3	112	5.24	4.98	86.39	86.5	8.25	8.5
2	114	2.59	2.48	86.87	86.8	10.28	10.65
1	116	1.19	1.16	83.41	83.5	15.09	15.3
W		.50	.50	74.40	74.4	25.10	25.1

Figure 3. Comparison of Composition of Benzene-Toluene-Xylene System
(Robinson and Gilliland and by this Method)

Failure to Balance to Exactly 100 Mol Per Cent

At each plate, the calculation is stopped at (usually) some value between 99 and 101 mol per cent. If each composition were divided by the total, the distribution would remain the same, but the values used for the next calculation would be more accurate. This division by the sum is perfectly feasible in machine operation but requires the use of a more elaborate calculating machine.

Digital Errors

The four multiplications are carried to six significant figures, but two of the multiplicands are limited by machine capacity to the first three significant figures. This results in a digital error of .01% for two of the factors and .1% for two others. The use of a 602-A or another larger capacity machine would eliminate this inaccuracy.

Because of the iterative nature of the calculation, the errors accumulate and are compounded. The nature and extent of these inaccuracies should be kept in mind.

ONCE the required vapor pressure cards are on hand for the desired calculation, the time of calculation is as follows for a four-component calculation over a fifty-degree boiling range that requires forty plates for separation:

Machine Operation	*Machine Time*
Prepare master system cards (1 set)	2.0 minutes
Prepare master working cards (2 sets)	12.5 minutes
Reproduce 5 sets of working cards (10 sets)	5.0 minutes
Calculate plate composition	26.6 minutes
Print column composition	0.5 minutes
TOTAL	46.6 minutes

The number of cards per test will vary from ten to two. An average of five was used in calculating machine time.

The additional operator manipulation time depends on the skill and experience of the operator. It should certainly not exceed the machine time.

Calculations with Vanishing Components

If one or more of the feed components in the upper or lower part of the column is reduced to less than .01 mol per cent, the calculation may be handled as follows:

1. No vanishing components—Calculate up column or down column.
2. One or more components vanish in upper section—Calculate up the column.
3. One or more components vanish in the lower section—Calculate down the column.
4. At least one component vanishes in each section of the column—Calculate both up and down the column.

The practicability of applying machine methods of calculation to chemical engineering calculation hinges on many factors. Rigorous calculations of this type should always be made by or under the direct observation of one who understands the problems involved. The method described here is offered as a rapid method of accumulating information about column design and performance. Some of the cases in which this rapid collection of information should be exceptionally useful are suggested below:

Calculation of columns for compounds and classes of compounds frequently handled. For organizations calculating large numbers of columns, this should be quite valuable.

Calculations in which several assumed conditions must be calculated in order to select the correct one. This is the case where the boiling point of one or more components lies between those of the components whose separation is desired.

Preparation of correlations between properties, operating conditions, and column performance. The effect of reflux ratio on column height, optimum feed condition, position for side streams, etc., may be found for a variety of systems.

Calculation of temperatures and compositions for a number of variations in the operating conditions (e.g., estimating optimum temperature control locations).

In rigorous calculations the chemical engineer often makes a heat balance on each tray as he calculates down the column. This determines the change in L and V due to heat losses from the column, the change in liquid and vapor sensible heats with temperature, and the differences between the heats of vaporization of the various components. Experience indicates that this change may be often closely approximated as a function of temperature alone for a given problem. This function may be found by first calculating the compositions for a constant L/V followed by thermal balancing at selected plates in the column. Interpolation at 1°C. intervals will then give a first approximation to the L/V throughout the column.

Since an individual detail card contains the vapor pressure and the reflux data for each temperature, it is clearly possible to use the interpolated L/V's on successive cards at different temperatures. By repeating the plate-by-plate calculations with the more exact L/V's, a more accurate calculation will be obtained.

DISCUSSION

Professor Donnell: The general advantage of doing the calculation by the punched card machine rather than by a slide rule is that very quick short-cut methods using abridged formulae yield insufficient information for most purposes. They may give one answer, the number of plates, but will not tell you temperature gradient or the composition of a given point.

Professor Nichols: How do you go about calculating the feed plate in your position?

Mr. Opler: We assume that we are given the composition that is entering the column at that point. Knowing that composition, we simply use it as our initial condition in a distillation. I actually insert a zero for the amount of the material which is removed at the top or bottom of the column.

Professor Donnell: I have a question in regard to the disappearance of compounds in the bottom and the top. How do you know the composition of the feed tray?

Mr. Opler: We are assuming that the feed composition is the composition of the feed plate. This may be a weakness.

Professor Donnell: Don't you think that is quite far from the case?

Mr. Opler: The trick is to match the feed as closely as possible to the composition of the feed plate. I am using a little inverse logic in trying to match the feed plate as closely as possible to the feed.

Professor Donnell: The reason we are so particular is that we work in petroleum and have disappearing compounds in the bottom and the top. We have found that the composition of the feed tray is different; the reflux ratio is about four to one. You could go on and say, "We will just match the ratio of the key component on the feed tray with the feed. That does not give you the maximum feed tray location."

Mr. Opler: That is very probably one of the weaknesses.

The problem of choosing the optimum feed tray location is often solved when using the Lewis and Matheson method at the desk by trial and error, introducing the disappearing components as a chosen feed tray is approached from both top and bottom and carrying the calculation a few plates past the chosen feed tray. When the proper tray at which to introduce these disappearing components has been established by trial so that the compositions match sufficiently close at the feed tray, one completely balanced solution has been found. Repetition with choice of a different tray for the feed will disclose whether more or fewer total plates for the column are required. Further repetition will locate the optimum feed location for minimum total plates. Although we have not worked out examples of this in detail, we believe that the machine procedure would save considerable time in carrying through the tedious repetitions.

Continuous Distillation Design Calculations with the IBM Card-Programmed Electronic Calculator[*]

ARTHUR ROSE THEODORE J. WILLIAMS WILLIAM S. DYE, III

The Pennsylvania State College

READILY available large-scale computing machines such as the IBM Type 602-A Calculating Punch and Type 604 Electronic Calculating Punch have been shown to be very advantageous for any calculation problem where multiple repetition of a relatively simple calculation procedure was necessary. For instance, tables of values of plate compositions at finite reflux ratios for binary mixtures in 100-plate distillation columns, and for ternary mixtures in 25-plate columns, have been readily obtained using the 604.[4] For more complex problems these machines are restricted by their limited storage capacity and by the fact that a new set of wiring for the control panels is necessary for each new calculation procedure used.

The new IBM Card-Programmed Electronic Calculator, however, has approximately five times the storage capacity of the type 604, and a single set of control panels can be used for any arithmetical calculation procedure, provided fractional exponents are not involved. Thus, this machine overcomes many of the shortcomings of the previously mentioned machines. The card-programmed calculator owes its versatility to the fact that the control panels are wired to make possible each of the four basic operations of arithmetic. The particular calculation procedure desired is carried out by means of a deck of specially punched IBM cards, each card in its turn choosing the proper arithmetical operation and the proper factors to carry out one step in the calculation. A new equation can be evaluated merely by the use of a new set of program cards, properly punched to carry out the steps of the equation desired. Thus, by the proper combination of program cards any length problem can be handled, provided only that the storage capacity of the calculator is not exceeded during the calculation period.

This versatility of the card-programmed calculator led the authors to investigate its use for the trial-and-error type of problem which arises in many chemical engineering design problems as well as in other engineering fields. Within the field of chemical engineering the unit operation of distillation was chosen because of its wide usage and because of the large number and variety of problems encountered. Also, in the field of distillation, as well as all the other diffusional operations, the trial-and-error problem is especially important, because in a large proportion of the cases studied, insufficient data are available to establish the value of all the independent variables present. Thus, the values of some quantities must be assumed, and the accuracy of these assumptions must be checked—hence, the trial-and-error calculation. This situation arises almost constantly in the problems encountered in design studies on distillation columns. Solution of these problems by machine methods should result in large savings in time and money for all organizations involved in this and related fields. The methods of solution of several variations of problems of this type are discussed in this paper.

The General Problem

A problem frequently arising in distillation involves determining either the head composition (x_D), the number of plates (n), or the reflux ratio (R) corresponding to a particular feed condition, with other column variables known or specified. In nearly every case the problem becomes a trial-and-error type of calculation, because an explicit function of either x_D, n, or R, in terms of the other variables of distillation, is either very complicated or entirely inexpressible. This is especially true if the distillation system under study involves variation in relative volatility (α), heats of vaporization of the components, or plate efficiency.

In order to define a distillation system properly, the values of a certain minimum group of the variables must be established. This group need not always consist of the same variables, of course. The values of all other column variables can then be calculated from these known quantities.

For the case of an adiabatic continuous column (Figure 1) operating on an ideal binary system (the simplest possible case), the minimum group of variables might consist of the following:

[*] Solution of trial-and-error-type problems.

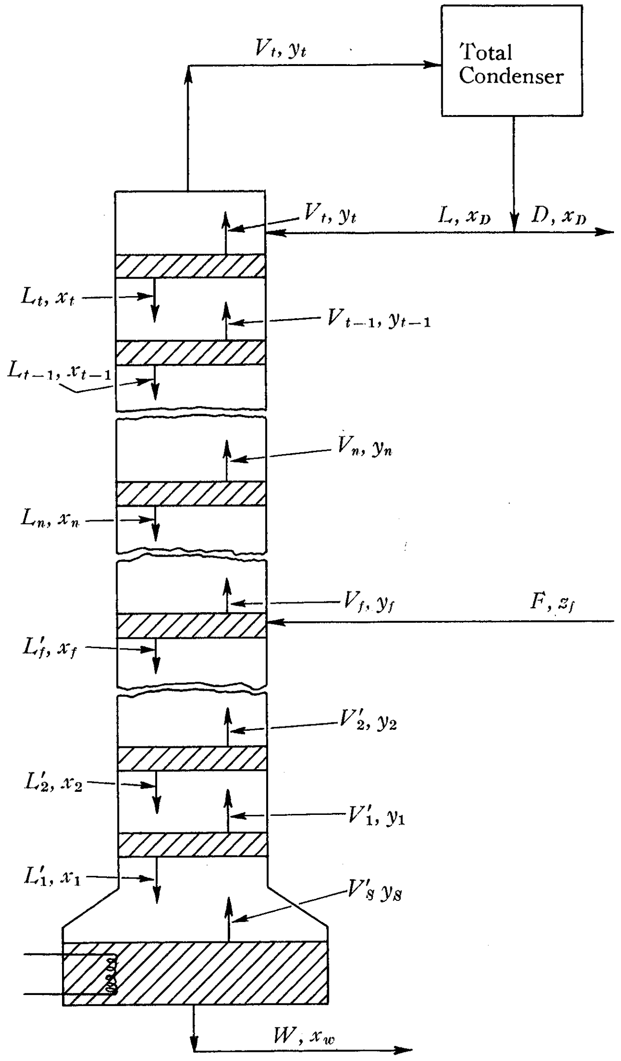


Figure 1. Idealized Diagram of a Typical Continuous Distillation Column

1. Feed rate (F) in mols per unit time.
2. Feed composition (z_f) as the mol fraction of the more volatile component.
3. Condition of feed (q) as a ratio of the heat required to vaporize one mol of the feed to that required to vaporize one mol of saturated liquid of the same composition.
4. Distillate composition (x_D) as the mol fraction of the more volatile component.
5. Distillate takeoff rate (D) as mols per unit of time.
6. Enriching section vapor rate (V) as mols per unit of time.
7. Relative volatility of the mixture (α).
8. Number of theoretical plates (n) in the column.

From the above factors the following additional functions, which are necessary to define the distillation system completely, can be calculated. The additional functions are:

1. Enriching section liquid return rate (L) as mols per unit time.
2. Stripping section vapor rate (V') as mols per unit time.
3. Stripping section liquid rate (L') as mols per unit time.
4. Bottoms takeoff rate (W) as mols per unit time.
5. Bottoms composition (x_W) as the mol fraction of the more volatile component.
6. Feed plate composition (x_i) as the mol fraction of the more volatile component.
7. Plate compositions (x_n) as the mol fraction of the more volatile component.

It must be kept in mind that variations of the problem involve other combinations of known and unknown factors.

The usual design problem arises because not all of the minimum group of factors are known; and, in addition, the relations between variables are too complex to permit direct calculation. Therefore, the required procedure must be: (1) to guess at some probable values of the unknowns among the minimum group of factors; (2) use these trial values to calculate trial values of at least one of the dependent functions in two or more different ways; (3) from a comparison of these calculated values, determine a probable error in the original trial values; and finally, (4) choose new trial values and repeat the above procedure until the chosen trial values show themselves to be correct. This trial-and-error procedure is important not only in distillation, but also in the design calculations of nearly every phase of chemical engineering.

Use of the Card-Programmed Calculator in Solving the Problem

As a basis for a description of the use of the card-programmed calculator in trial-and-error calculations, assume that the following problem is at hand: The variables whose values are known, or specified, are the same as those listed in the general case discussed above. Let the distillate composition (x_D) be the variable for which trial values are to be assumed until a satisfactory value is obtained. The complete calculation procedure is now carried out with the card-programmed calculator according to the steps of the following plan. The equations are numbered to correspond to the steps wherein they occur.

1. The values of the known and assumed quantities are read into the calculator and stored in designated storage units.
2. The bottoms rate is calculated as

$$W = F - D \ . \qquad (2)$$

3. The bottoms composition is calculated as

$$x_W = \frac{Fz_f - Dx_D}{W} \ . \qquad (3)$$

4. The enriching section liquid rate is calculated as

$$L = V - D \ . \qquad (4)$$

5. The stripping section liquid rate is calculated as

$$L' = L + qF \ . \qquad (5)$$

6. The stripping section vapor rate is calculated as

$$V' = V + (q-1)\,F \ . \qquad (6)$$

7. The feed plate composition is calculated as

$$x_i = \frac{(L+D)z_f + D(q-1)x_D}{(L+qD)} \ . \qquad (7)$$

8. The top plate liquid composition is calculated as

$$x_t = \frac{x_D}{\alpha - (\alpha-1)x_D} \ . \qquad (8)$$

The immediately subsequent steps involve plate-to-plate calculations down the column by the well-known McCabe-Thiele procedure.[1] That is, we calculate the vapor and liquid composition on each plate of the column in turn, until the bottom plate (reboiler) is reached. The liquid composition in the reboiler is the bottoms composition, here designated as x'_W, if perfect mixing occurs. If the trial value of x_D were correct, then x'_W would be identical with x_W from step 3. In general, the first trial value of x_D will not achieve this equality. A comparison of x'_W with x_W gives a measure of the error in the original assumed value of x_D. From knowledge of this error we are in a position to choose a much more accurate value of x_D for the next trial.

The plate-to-plate calculation of liquid and vapor compositions by the McCabe-Thiele method is carried out on the card-programmed calculator as follows:

9. The composition of vapor leaving the plate next to the top plate is calculated as

$$y_{t-1} = \frac{L}{V} x_t + \frac{D}{V} x_D \ . \qquad (9)$$

10. The liquid composition on the plate next to the top plate is calculated as

$$x_{t-1} = \frac{y_{t-1}}{\alpha - (\alpha-1)y_{t-1}} \ . \qquad (10)$$

Liquid and vapor compositions corresponding to the remaining plates above and including the feed plate, that is, in the enriching section, are calculated similarly by repeated alternate use of the equations:

11. $$y_n = \frac{L}{V} x_{n+1} + \frac{D}{V} x_D \ . \qquad (11)$$

12. $$x_n = \frac{y_n}{\alpha - (\alpha-1)\,y_n} \ . \qquad (12)$$

However, below the feed plate (that is, in the stripping section) a different set of equations must be used to calculate the values of y_n and x_n because of the different flow characteristics of the column, caused by the introduction of the feed. The required equations are:

$$y_m = \frac{L'}{V'} x_{m+1} - \frac{Wx_W}{V'} \qquad (11a)$$

and

$$x_m = \frac{y_m}{\alpha - (\alpha-1)y_m} \ . \qquad (12a)$$

The basis of determining whether a particular plate is in the enriching section or the stripping section of the column is to note whether or not x_n is greater or less than the value of x_i, the composition corresponding to the point of intersection of the operating lines with the q line.

This can be used by the calculator to give it a basis of choice between equations 11 and 11a in calculating the value of y for the next lower plate.

It can be seen that the equation

$$x_n - (x_i + 0.0001) = u \qquad (13)$$

will give a negative value of u for any x_n less than or equal to x_i and a positive value of u for x_n greater than x_i.

Use is now made of the negative balance feature of the machine to cause a selector to transfer when u becomes negative. This selector controls the region of the program card from which the program punches are read, and thus can cause the calculator to use equations 11 and 12 or 11a and 12a, if the proper control signals are punched in the two different regions on the card.

The test described above must of necessity follow the calculation of every value of plate liquid composition and, therefore, will take place between calculation steps 10 and 11 and after step 12 as they are described above.

To recapitulate regarding the plate-to-plate calculations, the calculator starts with steps 9 and 10 for the plate next to the top plate, then makes the comparison of equation 13. Unless the unusual condition of introduction of feed on the top plate is involved, the results of the comparison of equation 13 will direct the calculator to proceed through steps 11 and 12, following which the comparison of equation 13 is again made. This cycle of steps 11 and 12 followed by the comparison of equation 13 is repeated until the test indicates the need for use of equations 11a and 12a instead of equations 11 and 12. These are then used repeatedly until the plate-to-plate calculations have been made for the specified total number of plates in the column. The value of x_m from the last use of equation 12a is, as mentioned before, the desired trial x'_W.

When x'_W has been calculated, the difference,

$$x_W - x'_W = e_{x_W} \ , \tag{14}$$

is a measure of the error in the original assumed value of x_D. Because of the hyperbolic form of the equilibrium diagram for an ideal system, the function relating e_{x_W} and C_{x_D} takes on the following form:

$$C_{x_D} = \frac{(1.0 - x_D)^k \, e_{x_W}}{K \, (x_W)^k} \ , \tag{15}$$

where

$$k = f(\alpha) \ , \tag{16}$$

and K is a constant.

From the value of C_{x_D} calculated above, it is possible to determine a new value of x_D as:

$$x_{D \text{ new trial value}} = x_D + C_{x_D} \ . \tag{17}$$

Since the calculator must use integer powers of numbers in repetitive calculation, k must of necessity take the integer form, although this may require several more trial values of x_D because C_{x_D} cannot be calculated exactly.

Once the new trial value of x_D has been calculated, the whole procedure is repeated with the exception of step 1 to obtain a new trial value of x_D, etc. The calculation is finished whenever C_{x_D} equals zero, that is, the correct value of x_D has been assumed. Figure 2 shows two such trials as graphed on an equilibrium diagram.

Other Variations of the Distillation Problem and Their Solution

Determining Reflux Ratios by Trial and Error

As mentioned previously, the above-described problem is only one ramification of the general distillation problem. Another aspect which presents itself is as follows: A specific distillate composition (x_D) is desired, and the problem is to determine the proper reflux ratio (L/D) to attain this result from a known feed. This type of problem is at least as important as the one just discussed in detail and is solved in the same manner, except for two changes. In this case trial values of enriching section liquid rate (L) are

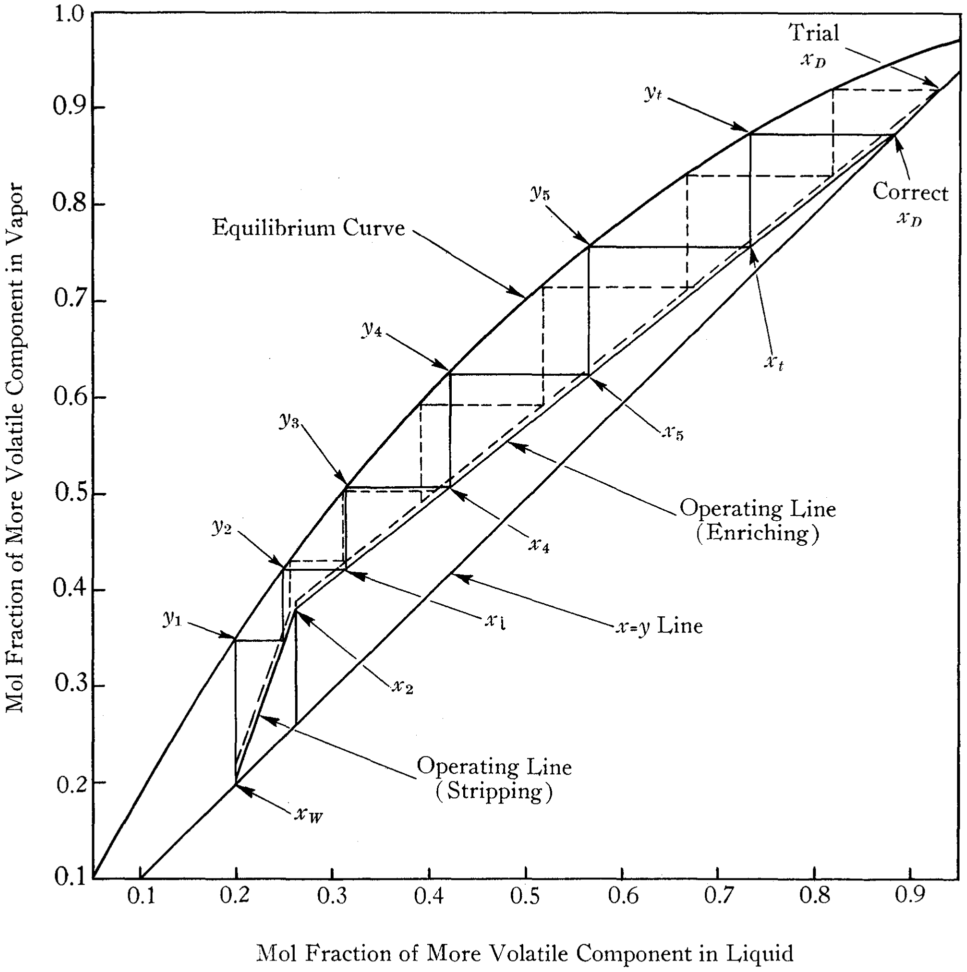


Figure 2. Graph to Show McCabe-Thiele, Trial-and-Error Method of Solution of Distillation Problems

chosen, and a system must be devised to provide a basis for choosing new trial values of L. The inclusion of the liquid rate (L) among the minimum group of required variables, of course, places the vapor rate (V) among the dependent variables.

The method used for the choice of the new trial value of L, and thus eventual determination of the proper value of R, takes the following form:

$$R_2 = L_2/D = \frac{R_1}{r} = L_1/Dr \ , \tag{18}$$

where r is > 1 when e_{x_W} is positive and r is < 1 when e_{x_W} is negative. This condition is satisfied if

$$r = 1 + K'(e_{x_W}) \ , \tag{19}$$

and thus

$$L_2 = \frac{L_1}{1 + K'(e_{x_W})} \ . \tag{20}$$

Effects of Non-ideality

Both of the problems considered previously have been for the case of a column in which all of the simplifying assumptions of distillation hold true. This was done in order to give maximum clarity to the trial-and-error aspects of the problem. A note is in order here, however, concerning some of the methods of handling the various types of non-ideality and to show that their only effect upon the problem is to lengthen it somewhat.

The four most common causes of non-ideality in column operation are:

1. Non-constant relative volatility (α) of the mixture being distilled.
2. Non-adiabatic operation of the column itself due to heat gains or losses through the column walls.
3. Unequal heats of vaporization of the components of the mixture being distilled.
4. A column plate efficiency which is not equal to 100%.

A discussion of the method of handling each of these will now be given.

In most cases the components of the mixture do not exhibit a constant relative volatility over the complete composition range; that is, the relative ease with which one component is boiled away from the other is not constant. This is due usually to physicochemical effects within the mixture itself.

The method of solution involves one of two choices. Either the quantity α can be expressed as a function of x, usually a series, or x can be expressed directly as a function of y, again usually a series. Of these, the second choice is undoubtedly the best, for the other would involve an iterative calculation, because α is a function of x, and x is, of course, the quantity we desire to establish. In the second case, the calculations above would proceed the same as before, except that equations 8, 10, 12 and 12a would be in the form

$$x = f(y) \ , \tag{21}$$

different from and probably more complicated than that previously used. Equation 21 is commonly in the form of the series

$$x = ay \pm by^2 \pm cy^3 \pm \ldots \ . \tag{22}$$

Non-adiabatic operation of the column has the effect of changing the values of L and V, and of course, L' and V' from plate to plate as one proceeds up the column. In the case of heat loss from the column, both L and V will be decreased by the amount

$$\Delta L = \Delta V = \frac{Q}{H_v} \ , \tag{23}$$

and

$$L_n = L_{n+1} + \frac{Q}{H_v} \ , \tag{24}$$

where Q is the amount of heat lost per unit time, and H_v is the heat of vaporization of one mol of the mixture being distilled. If the heat loss is constant for each plate, usually a good approximation, this merely means that steps 5 through 8 must be repeated for each plate rather than occur only at the start of the calculation. In other words, the calculation for each plate now involves the use of equation 24, followed by steps 5 through 10 rather than merely steps 9 and 10 or 11 and 12.

The case of unequal heats of vaporization which also has the effect of varying L and V is best handled by the well-known Peters method[3] which involves the use of a fictitious molecular weight for the components of the mixture so that H_v is again constant per mol of liquid. This has the effect of making the values of L and V, expressed in mols, constant although the values of L and V, expressed in pounds, may be decidedly different from plate to plate. The effect of the above change upon the calculations is usually only a change in the value of α.

Theoretical plates—that is, column plates which fulfill equations 10 and 21 exactly—are uncommon in actual practice. Therefore, an efficiency term must be inserted to compare the actual relation of x and y which exists in the column, to that of a column with theoretical plates. The efficiency term is usually expressed[2] as

$$E = \frac{x_{n+1} - x_n}{x_{n+1} - x_n{}^*} \ , \tag{25}$$

where $x_n{}^*$ is the value of x_n expressed by equations 11 and 12, and x_n is the value of x_n actually existing in the column. Therefore,

$$x_n = x_{n+1} - E(x_{n+1} - x_n{}^*) \ , \tag{26}$$

and the efficiency can be used in the calculations above by using equation 26 after step 12 for each plate.

Consideration of the above discussion shows that for the general case any or all of the above causes of non-ideality can be taken into account in either of the calculations described previously with no major effect except to lengthen the calculations in some cases.

General Solution of Trial-and-Error Problems

Consideration of the method of attack of the problems already discussed leads to some general statements regarding trial-and-error problems as a whole.

The first and most important point to be brought out is that trial-and-error calculations can be done independently by the machine only if expressions such as equations 15 and 20 can be derived. Other than this, the only restrictions are that arithmetical operations must be used, and the storage capacity cannot be exceeded.

The general method of approach in calculating trial-and-error problems is as follows:

1. Express the known independent variables of the system along with a trial value of the variable to be established.
2. Calculate the values of the dependent variables from known and assumed values of the independent variables, being sure to express at least one of the dependent variables in two different ways, thus getting two possible answers for this variable.
3. From the magnitude and sign of difference between the two possible answers of the dependent variable, determine a new trial value of the assumed independent variable.
4. Repeat the above steps until no difference is noted in the dependent variable as calculated by the two different methods. The last trial value of the unknown independent variable is then the correct value.

Conclusions

The work of this paper has shown that the method of calculation control employed on the card-programmed calculator makes its use possible for nearly all types of trial-and-error calculations. The machine is capable of carrying out the calculations at a great saving in time compared with hand calculation, and this saving increases greatly as the complexity and length of the problem increases.

The main requirements in applying the machine to common engineering problems where the trial-and-error method is involved are:

1. The relations involved must be arithmetical or must be capable of being expressed arithmetically.
2. A basis must be available to enable the machine to choose a new trial value of the desired quantity.
3. The storage capacity of the machine cannot be exceeded.

Nomenclature

$a, b, c,$ ... Constant coefficients used in equation 22.

C_{x_D} ..... Correction to be applied to x_D to obtain a new x_D for the next trial.

D .... Distillate takeoff rate in mols per unit time; as a subscript it refers to distillate.

E .... Plate efficiency as defined by equation 25.

e_{x_W} ..... Difference in the value of x_W as calculated by two different methods.

F .... Feed rate in mols per unit time.

H_v ...... Heat of vaporization of the liquid mixture at the point in question.

i .... As a subscript it refers to the intersection of operating lines, i.e., feed plate.

K .... Constant in equation 15.

K' ...... Constant used in equation 20.

k .... Exponent in equation 15.

L .... Enriching section liquid return rate as mols per unit time.

L' ...... Stripping section liquid rate as mols per unit time.

m .... As a subscript it refers to the general value of the plate number in the stripping section.

n .... As a subscript it refers to the general value of the plate number in the enriching section of the column.

Q .... Amount of heat loss from column per unit time.

q .... Condition of the feed as a ratio of the heat required to vaporize one mol of the feed to that required to vaporize one mol of saturated liquid of the same composition.

R .... Reflux ratio as L/D.

r .... Divisor used in derivation of equation 20.

t .... As a subscript it refers to the top plate.

u .... Value of the difference between the liquid composition on the plate in question and the liquid composition on the feed plate.

V .... Enriching section vapor rate as mols per unit time.

V' ...... Stripping section vapor rate as mols per unit time.

W ...... Bottoms takeoff rate as mols per unit time; as a subscript it refers to the bottoms.

x .... Liquid composition as the mol fraction of the more volatile component; as a subscript it refers to the location in the column.

x_n^* ..... Value of the liquid composition as predicted by theory under ideal conditions.

y .... Vapor composition as the mol fraction of the more volatile component; as a subscript it refers to the location in the column.

z_f .... Feed composition as the mol fraction of the more volatile component.

α .... Relative volatility of the mixture under study.

Δ .... Symbol signifying the rate of change of the quantity in question.

Appendix

Problem 1. Calculation of the Correct Value of x_D.

I. The initial conditions are specified and supplied to computer:

$\alpha = 2.23$ $F = 10.0000$ $z_f = 0.2000$
$L = 4.0000$ $D = 1.0000$ $q = 1.0000$

II. The machine first uses equations (2) to (8) as indicated to calculate remaining initial conditions: $W = 9.000$ (2); $x_W = 0.1333$ (3); $V = 5.0000$ (4); $L' = 14.000$ (5); $V' = 5.0000$ (6); $x_i = 0.2000$ (7); $x_t = 0.2000$ (8).

III. The value of x_T is checked for feed plate location: $u = 0.4420$ (13)

IV. Calculation of plates is then carried out in order as:

(a) Plate 5:
$y_5 = 0.6737$ (9); $x_5 = 0.4808$ (10);
$u = +0.2807$ (13).

(b) Plate 4:
$y_4 = 0.5446$ (11); $x_4 = 0.3491$ (12);
$u = +0.1490$ (13).

(c) Plate 3:
$y_3 = 0.4393$ (11); $x_3 = 0.2600$ (12);
$u = +0.0599$ (13).

(d) Plate 2:
$y_2 = 0.3680$ (11); $x_2 = 0.2070$ (12);
$u = +0.0069$ (13).

(e) Plate 1:
$y_1 = 0.3256$ (11); $x_1 = 0.1780$ (12);
$u = -0.0221$ (13).

The calculator would now begin to use stripping line equations (11a) and (12a), if there were more plates to be calculated. However, plate 1 is the last plate.

V. Calculation of the new value of x_D is carried out using the equations (14), (15), (16) and (17) with x_D new trial value = 0.7666.

VI. The above calculations are repeated for the new trial with the following results.

a.

Plate or Calculation Step	x_W	y_n	x_n	u	e_{xW}	x_D
Material Bal.	0.1370					
Top			0.5956	+0.3954		
5		0.6298	0.4328	+0.2327		
4		0.4996	0.3093	+0.1092		
3		0.4008	0.2307	+0.0307		
2		0.3379	0.1862	−0.0139		
Calculator transfers to stripping line equation here.						
1		0.2748	0.1452			
New x_D					−0.0082	0.7595

b.

Plate or Calculation Step	x_W	y_n	x_n	u	e_{xW}	x_D
Material Bal.	0.1378					
Top			0.5862	+0.3860		
5		0.6209	0.4234	+0.2233		
4		0.4906	0.3016	+0.1015		
3		0.3932	0.2251	+0.0250		
2		0.3320	0.1813	−0.0178		
Calculator transfers to stripping line equation here.						
1		0.2624	0.1376			
New x_D					−0.0002	0.7595

0.7595 is, therefore, the true value of x_D.

Problem 2. Calculation of the True Value of x_D for a Case of Non-Constant Relative Volatility.

Conditions are same as for problem 1, except that x is related to y by means of the equation below rather than equation (12);

$$x = ay + by^2 = 0.1429y + 0.8571y^2 \qquad (22a)$$

Use of this relation also necessitated a change in equations for the calculation of the new trial x_D so that

$$x_{D\text{ new trial value}} = \frac{x_D + (1.0 - x_D)^1}{5(x_W)^1}\, e_{xW} \;. \qquad (15b)$$

This was necessary in order to damp the calculations properly because of the small spread in the equilibrium curve at its upper end (Figure 3). The actual operations are as follows:

I. Initial conditions, both calculated and specified, are the same as for problem 1 except as specified above and below.

II. The trials are carried out as follows:

1.

Plate or Calculation Step	x_W	y_n	x_n	u	e_{xW}	x_D
Material Bal.	0.1333					
Top			0.6628	+0.4627		
5		0.6902	0.5069	+0.3068		
4		0.5655	0.3549	+0.1548		
3		0.4439	0.2322	+0.0321		
2		0.3458	0.1519	−0.0482		
Calculator transfers to stripping line equation here.						
1		0.1854	0.0560			
New x_D					+0.0773	0.8233

2.

Plate or Calculation Step	x_W	y_n	x_n	u	e_{xW}	x_D
Material Bal.	0.1307					
Top			0.6985	+0.4984		
5		0.7235	0.5521	+0.3520		
4		0.6063	0.4017	+0.2016		
3		0.4860	0.2718	+0.0717		
2		0.3821	0.1797	−0.0204		
Calculator transfers to stripping line equation here.						
1		0.2679	0.0998			
New x_D					+0.0309	0.8317

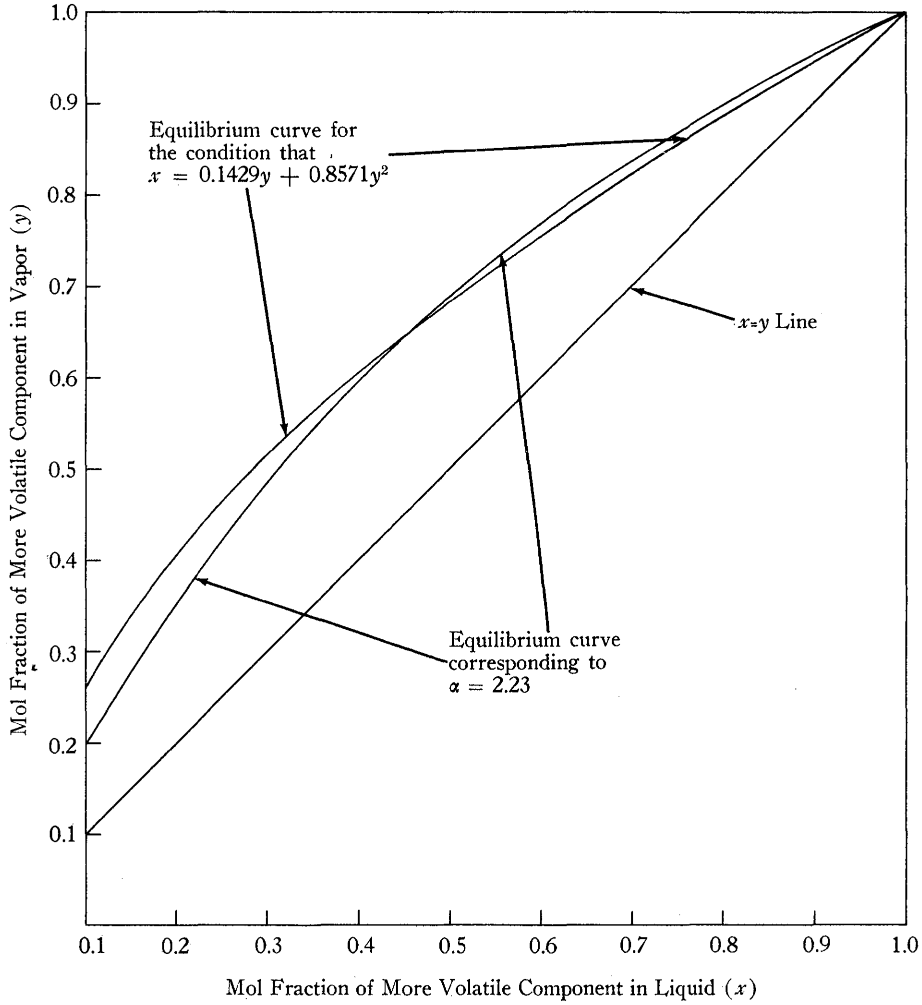


FIGURE 3

3.

Plate or Calculation Step	x_W	y_n	x_n	u	e_{x_W}	x_D
Material Bal.	0.1298					
Top			0.7117	+0.5116		
5		0.7357	0.5690	+0.3689		
4		0.6215	0.4199	+0.2198		
3		0.5023	0.2880	+0.0879		
2		0.3967	0.1916	−0.0085		
Calculator transfers to stripping line equation here.						
1		0.3028	0.1219			
New x_D					+0.0079	0.8337

4.

Plate or Calculation Step	x_W	y_n	x_n	u	e_{x_W}	x_D
Material Bal.	0.1296					
Top			0.7149	+0.5148		
5		0.7387	0.5733	+0.3732		
4		0.6254	0.4246	+0.2245		
3		0.5064	0.2922	+0.0921		
2		0.4005	0.1947	−0.0054		
Calculator transfers to stripping line equation here.						
1		0.3119	0.1280			
New x_D					+0.0016	0.8342

5.

Plate or Calculation Step	x_W	y_n	x_n	u	e_{x_W}	x_D
Material Bal.	0.1295					
Top			0.7157	+0.5156		
5		0.7394	0.5743	+0.3742		
4		0.6263	0.4257	+0.2256		
3		0.5074	0.2932	+0.0931		
2		0.4014	0.1955	+0.0046		
Calculator transfers to stripping line equation here.						
1		0.3143	0.1296			
New x_D					+0.0001	0.8342

0.8342 is, therefore the correct value of x_D.

REFERENCES

1. WARREN L. MCCABE and E. W. THIELE, *Industrial and Engineering Chemistry, 17,* 605 (1925).
2. E. V. MURPHREE, *Industrial and Engineering Chemistry, 17,* 747, 960 (1925).
3. W. A. PETERS, *Industrial and Engineering Chemistry, 14,* 476 (1922).
4. ARTHUR ROSE and T. J. WILLIAMS, *Industrial and Engineering Chemistry 42,* 2494 (1950).

DISCUSSION

[There was an informal discussion of this paper during the demonstration of the problem on the card-programmed calculator.]

Application of Automatic Computing Methods to Infrared Spectroscopy

GILBERT W. KING

Arthur D. Little, Incorporated

ALTHOUGH it cannot be said that computing machines have been indispensable to progress in infrared spectroscopy, they have allowed progress in certain fields, and as more people become active in the mixed field, advances in almost all directions will be made. To illustrate both these points, we shall catalogue the instances where our computing laboratory has supported the infrared.

Interpretation of Rotational Structure

The earliest application was in the interpretation of spectra. The foundations for this are firmly based on quantum mechanics, and the dimensions of many symmetrical molecules have been deduced from the rotational structure of their infrared spectra. In the case of the majority of molecules, asymmetry prevented such an analysis because of the numerical work required. The most troublesome step is the calculation of the energy levels of a rotor with three different moments of inertia. These levels are simply related to the roots of certain matrices, and give rise to the well-known problem of finding characteristic values of sets of linear equations. In the case of the asymmetric rotor, the matrices (and their determinants) have only three non-zero diagonals, and the problem of finding the energy levels is simply that of finding the roots, η, of a series of continued fractions,

$$a_0 - \eta - \cfrac{b_1}{a_1 - \eta + \cfrac{b_2}{a_3 - \eta \cdots}} = 0. \qquad (1)$$

This can only be done by successive substitutions of guesses at η; however, a characteristic of continued fractions is that such a process rapidly converges.

This problem is ideally suited to punched card equipment because the operation has to be repeated many times (spectroscopists would like a table consisting of a million η's) and because, further, the whole process depends on a simple arithmetic algorithm

$$a - b/c\,, \qquad (2)$$

which is iterated several times. These roots, incidentally, are the characteristic values of the Lamé functions.

A coarse table was constructed by these means. Five-point interpolation carried out on computing machines gave good approximations for η at finer intervals. These approximate η's were substituted in the continued fraction to give the correct answers. In this way, a fundamental table of the energy levels of the asymmetric rotor has been built up. It has found great use by infrared and microwave spectroscopists, and we are at present enlarging it.

Having this basic table on cards, we can proceed with the analysis of rotational band spectra. The approach has been a stochastic one in which a structure of the molecule is assumed, the spectrum calculated and compared with the observed. Since several successive guesses have to be made, the repetitive nature of the procedure is suited to punched card techniques.

The computation required is as follows: First, a parameter, κ, a characteristic of the molecule, is computed by hand, and the appropriate reduced energy levels, $\eta_i(\kappa)$, are looked up in the fundamental table. We then calculate

$$E_i(\kappa) = J(J+1)\ [(a-c)\eta_i(\kappa) + (a+c)]\,, \qquad (3)$$

where a and c are further parameters of the assumed dimensions of the molecule (actually the reciprocals of the least and greatest moments of inertia), and J is a second *serial number* labeling the levels.

In the meantime, another basic table, giving the probability Φ_{ij} of a line occurring in the spectrum because of the absorption of energy transferring the molecule from energy level $E_i(\kappa)$ to $E_j(\kappa)$, is taken for the κ in hand, and reproduced. This is a two-way table: First, the energies $E_i(\kappa)$ are collated with it, each $E_i(\kappa)$ card preceding all the Φ_{ij} cards with the same i. The value of $E_i(\kappa)$ is gang punched into the Φ_{ij} cards. Next, these cards are sorted on j, the $E_j(\kappa)$ cards are collated, and $E_j(\kappa)$ is gang punched.

We now have the basic cards representing the expected spectrum. On them we calculate the position of the line

$$\nu = E_j(\kappa) - E_i(\kappa) \ , \qquad (4)$$

and the intensities of the line

$$\alpha_{ij} = \Phi_{ij} e^{-E_i(\kappa)/kT} \ . \qquad (5)$$

The exponential function of the argument is obtained by collating the Φ cards with $E_i(\kappa)/kT = x$ on them with a basic table of the exponential function e^{-x}. The latter is gang punched from the basic table into the cards. Multiplication by Φ_{ij} is an elementary operation.

On sorting the cards by ν, we have the basic line spectrum, a card for each line, having its position ν (in wave numbers) and intensity α. Unfortunately, as a rule, infrared spectroscopy does not resolve the individual lines, and we come to a real contribution of punched cards. The recorded spectrum is actually a certain integral of intensity over the finite slits of the spectroscope. We know the slit function $p(\nu)$ well enough[a] and can calculate what our completely resolved spectrum (the cards with α and ν) would look like with finite slits by calculating the integral

$$I(\nu) = \int_{-\infty}^{+\infty} p(\nu - \nu') \ \alpha(\nu') d\nu' \ . \qquad (6)$$

The function of $p(\nu)$ is replaced by a set of discrete values, and the integration obtained by summing over a finite interval. With card cycle transfer, this sliding integration can be done in one pass through the accounting machine.

The above procedure is repeated many times for various structures until a satisfactory fit with the observed is obtained. The comparison is done graphically, and because many trials are made we have found our method[1] of plotting with the 405 accounting machine very useful. Several infrared bands have been analyzed in this way.[2,3,4] Each one, a few years ago, could have been a doctoral thesis.

Vibrational Spectra

The majority of molecules have so many atoms that the rotational structure discussed above is smeared out and the infrared spectrum consists of many bands, each one associated with an excitation of vibrations in the molecules. An analysis of these spectra could proceed along the lines outlined above, although the details are considerably different. Here, one would have to assume certain force constants in the molecule, calculate the vibrational frequencies by finding the roots of simultaneous equations, which in this case are more general than the ones discussed above, and reconstruct the spectrum from these results. Unfortunately, there are so many parameters to be introduced into the fundamental equations, this procedure cannot be followed without further considerations. However, there is no question that computing machines will be necessary to solve the equations once a procedure for setting them up has been defined.

At the present time, an empirical approach is being made in many laboratories to assign frequencies to certain structural features. This is an empirical and, in a sense, a statistical procedure, and now calls for large-scale computing methods since several thousand spectra have been recorded. However, one should approach these data with considerable caution because there is no sense in trying to make empirical rules out of incorrect data. In particular, the present-day infrared spectroscope is incapable of resolving all the frequencies which are characteristic of a molecule, and before any large-scale computing is done, the data should be better digested. One aspect of this that we are studying in our laboratory is the use of low temperatures to improve the resolution of the bands. Low temperatures decrease the natural widths of the bands and, therefore, separate them; but to see this decrease in width of the bands, one needs instruments of high resolving power. Computing methods to improve resolving power of an instrument are discussed below. To make the empirical assignments of band frequencies to structural features, the original data should be put on cards and comparisons made with the collator. The direct recording of spectra on cards is described below.

Reflection Data

Another example where present data are misleading and could give rise to a lot of fruitless calculation is shown by the reflexion spectrum of materials which are highly absorbing, such as glass, quartz and other minerals. It is conventional to interpret the maxima in the reflexion curves as vibrational frequencies of the crystal. A closer examination[b] of the theory shows that the reflexion curves are functions of both the absorption coefficient and the refractive index, and the characteristic vibrational frequencies of the crystal are determined from the maxima of the square of the refractive index times the absorption coefficient. It is, therefore, necessary to get the refractive index and the absorption coefficient from the reflectivity data. This can be done if the reflectivities are measured at two angles. In this case, the required parameters n and κ can be obtained by a solution of the transcendental equations:

$$R_{\phi} = \frac{(g - 2f\cos\phi + \cos^2\phi)(g + \sin^2\phi \tan^2\phi)}{(g + 2f\cos\phi + \cos^2\phi)(g + 2f\sin\phi\tan\phi + \sin^2\phi\tan^2\phi)} , \qquad (7)$$

[a] Recently we have carried out the necessary Fourier transforms giving the precise nature of $p(\nu)$, taking into account all the experimental conditions, using punched card methods.

[b] These experiments will be presented in detail elsewhere by H. O. McMahon and I. Simon.

with

$$g = \frac{1}{2}\left[n^2\,(1-\kappa^2) - \sin^2\phi\right] + \frac{1}{2}\sqrt{[n^2\,(1-\kappa^2) + \sin^2\phi]^2 + 4n^4\kappa^2} \qquad (8)$$

$$+\frac{2n^4\kappa^2}{[n^2\,(1-\kappa^2) - \sin^2\phi] + \sqrt{[n^2\,(1-\kappa^2) - \sin^2\phi]^2 + 4n^4\kappa^2}}$$

and $f =$ (9)

$$\sqrt{\frac{1}{2}[n^2(1-\kappa^2) - \sin^2\phi] + \frac{1}{2}\sqrt{[n^2(1-\kappa^2) - \sin^2\phi]^2 + 4n^4\kappa^2}}$$

Clearly, a solution for n and κ, given two reflectivities at two angles, is a considerably complicated problem. This has been done in our computing laboratory, and a fundamental table giving n and κ in terms of two reflectivities has been made. Given two reflectivity curves, at each wave length one can enter the table by use of a collator and obtain the refractive index and the absorption coefficient. When these are plotted, it is seen that there is considerable difference between the true vibrational characteristics of the crystal and the observed reflectivities. Again, we note that the conversion of the recorded spectra into the final significant curves requires the reading of the signal at every point along the curve. Clearly, the direct recording of the data as digital numbers on cards, rather than as a graph, saves a great deal of manual labor.

Card-Programmed Selective Sequence Calculations

The solution of the equations given above involves a great number of operations and was only made possible by certain developments in computing methods with IBM equipment. Specifically, we have converted our 602-A calculating punch into a card-programmed selective sequence calculator. A single control panel is permanently wired. One of twenty-four different operations can be called for on this control panel by a lead card in which the required operation is coded. Other instructions have to be given to the machine, such as clearing counters and calling for a new card. The code numbers are punched on a lead card. Several successive operations can be calculated on each card by having a sequence of codes punched on the lead card. At the present time, we have only extended the calculations to three successive operations. With this control panel, we have found the 602-A more flexible than a desk calculator. With the use of this card-programmed control panel, the equations for the reflectivity are not formidable, and the number of operations required to evaluate them can be carried out in a reasonable time.

Reading of the Data on Cards

The foregoing examples are typical of the processing of experimental data by various theoretical and calibration procedures in which the recorded data have to be read point-by-point over a long strip. One spectrum alone may require 2,000 points to be read off in a continuous curve. Obviously, this is the bottleneck when a great deal of experimental data are accumulated. To overcome this problem we have built an instrument, which instead of giving a continuous record of a signal taken experimentally, converts the signal into a number, which it punches in a card. Thus, the data are recorded as digital numbers on cards as well as on a continuous curve. Any processing of the data can be done with standard IBM equipment. For certain practical and theoretical reasons, the most efficient method is to record a single number in a single column of a card. To do this, the number has to be represented in a binary system. Fundamentally, punched cards allow a representation of binary, rather than decimal numbers, for either there is or is not a hole in a certain location, corresponding to the digits 1 or 0, which is all that is required on the binary scale. We have found that the most convenient method of using standard IBM equipment is to indicate the binary digits by punches down the rows so that each column represents a single reading of the instrument. A card then carries 80 successive readings.

Our first application of this recording of an output of an infrared spectrometer as digital numbers on cards for processing was a study of the noise of the instrument. This is a very appropriate example, because the noise is essentially discontinuous, and it is very difficult for a galvanometer continuous recorder to give an accurate account of noise arising in the detector. The digital reader does not suffer from the discontinuity of the signal.

Contribution to Experimental Technique

Our study of noise was the autocorrelation function to see if there are any fundamental periods which would be due to pickup, rather than to Johnson noise from the detector,

$$\phi(\tau) = \frac{1}{2\mathrm{T}}\int_{-T}^{T} \mathrm{I}(t-\tau)\,I(t)\,dt\,. \qquad (10)$$

It is very easy to obtain autocorrelation functions on punched card equipment once the data are read off point-by-point as a digital number. One merely makes a new deck of cards, collates it into the primary deck, and sums all the cross products. The two decks are then separated, a card removed from the second deck, and the two collated again. This is repeated, removing as many cards as required, the sums of the cross products forming the autocorrelation function.

To carry out calculations of this sort, it is not necessary to convert the binary numbers in the column of the card into a decimal number. By means of digit selectors on the 602-A or the accounting machine, it is possible to read the binary number directly into the counters of the machine and carry out the multiplication on the decimal system. For example, if the value of the signal were 3 (in the binary notation 11), the cross product of 3 with itself would be read into the machine as 11 × 11, giving 121 which is a mixed binary in decimal code, indicating $4 + (2 \times 2) + 1 = 9$, the correct answer.

The digital reader has also been used to record the spectra directly into cards. Autocorrelations of spectra have been made in order to determine the characteristics of these spectra for the design of filters; and cross-correlations of spectra with noise have also been made for use with a more advanced theory of filtering. Now that the data are in digital form and can be handled by automatic computing equipment, a vast new field of processing data is available. For instance, it is now possible to apply any filter characteristics to the data without having to build special equipment. For example, it might be found desirable to use a square filter. The difficulty of building such a filter into the recording apparatus is that there are invariably phase shifts over such a filter which cause considerable unwanted oscillations of the recording equipment. When the values of the spectra are punched in cards, it is possible to process the data through the accounting machine and apply a filter of any characteristics without phase shift (or, strictly speaking, with a constant phase shift which is the same for all frequencies). In particular, a square filter would be obtained by applying an equation like (6) with $p(x) = \sin x/x$. We are at present developing the theory of time series to the recording of infrared spectra and anticipate employing various types of filters suitable for the spectrum in hand in order to improve the filtering and, therefore, improve the resolving power of the instrument.

Analysis of Infrared Spectra by Use of Punched Cards

It is quite clear when the experimental data are in punched cards in digital form, a great many other treatments of infrared spectra can be made. It is now possible to compare spectra not only at peaks, which has been the custom in the past, but in every wave length. One can also analyze spectra of mixed components by subtracting the desired amount of the spectrum of any one pure compound from the observed spectra. This would leave a residue which can then be studied for the presence of further components.

Conclusion

The application of computing machines to infrared spectroscopy has been treated here very superficially, but it should be clear that punched card machines are now part of the scientific equipment of a research laboratory.

REFERENCES

1. GILBERT W. KING, "A Method of Plotting on Standard IBM Equipment," *Mathematical Tables and Other Aids to Computation, III,* No. 25 (January, 1949).
2. A more detailed account of the above procedure can be found in GILBERT W. KING, PAUL C. CROSS and GEORGE B. THOMAS, "The Asymmetric Rotor. III Punched Card Methods of Constructing Band Spectra," *J. Chem. Phys.* 14, 35 (1946).
3. The results are given in GILBERT W. KING, "The Asymmetric Rotor IV. An Analysis of the 8.5-μ Band of D_2O by Punched Card Techniques," *J. Chem. Phys.,* 15, 85 (1947).
4. and in R. M. HAINER and GILBERT W. KING, "The Asymmetric Rotor. V. Analysis of the 3.7-μ Band of H_2S by Punched Card Techniques," *J. Chem. Phys.* 15, 89 (1947).

Correlation and Regression Analysis

E. L. WELKER

American Medical Association

IT IS MOST unusual for any variable to be of great interest when considered by itself. Usually interrelationships of different quantities are sought in an attempt to explain what influence one might have on the other. The simplest type of interrelationship which can be imagined is one in which one quantity completely determines the other. Perhaps one of the most common illustrations of this relationship is found in the arithmetic lesson of the grade school student who finds out the cost of three pencils if one pencil costs 5 cents. In a general formula form, if n is the number of pencils and c is the total cost, then the formula is

$$c = 5n\,.$$

Mathematicians call this a functional relationship, and it is said that c is a function of n because when n is given, c is completely determined. No doubt you can think of many others, especially in physics, because in this field many of the laws are expressed as precise functional relationships. It is for this reason that the use of statistics in physics is a rather new development. It has come into prominence in nuclear physics because of the possibility of electrons moving in a variety of directions. It is now popular for the physicist to be interested in statistics. He must study some of the phenomena which are not simple functional relationships and which properly are described as falling in the field of correlation theory.

Before attempting to define correlation, it might be well to consider the other extreme. The functional relationship is one of the boundaries of correlation, and this other extreme to which I refer is the other boundary. It can be illustrated by considering a somewhat senseless example. There is no relationship between the population of a state and the size of the shoe worn by the senator from that state. These two items are totally unrelated and they would be called uncorrelated. In the range from functional relationship to totally unrelated variables there is a broad field which is included under the term correlation, although it must be remembered that it is common to think of the extremes as being special cases still within the framework of correlation itself. If the age of a married man is known, a reasonable estimate of the age of his wife can be made—although in most cases, probably, it cannot be confirmed. This is due only to the reticence of most women to state their ages and not due to the lack of relationship between the ages of husbands and wives. Except for recently notable examples which were in the newspapers, it is common for husbands and wives to have approximately the same ages. It should be possible then to derive some formula for making a good guess, a guess which is based on the habits of the people and not purely on speculation.

This illustration is exceedingly simple because it involves only two variables. This restriction is not necessary, and indeed it must be removed if certain problems in industrial statistics are to be solved. An excellent illustration of this is found in the doctor's thesis of one of my colleagues, in which he made a study of the demand for copper and in which an attempt was made to evaluate two methods of finding a relationship. In his analysis he tried to estimate the amount of copper delivered by looking at the undeflated price, the private gross capital formation, the stocks of copper at the beginning of the year, and the undeflated price of the previous year. This is a more practical problem and gives us a better illustration of the types of relationships which might be developed in actual practice. The important point is that these variables, as listed, are not sufficient to determine completely the deliveries of the copper, but they are highly influential and will include most of the items which combine to determine the actual deliveries which are made.

In order to make these vague notions more precise, it is necessary to consider a few mathematical manipulations. It is best to consider first the simple case of only two variables, because the fundamental concept for handling more variables is really no different from that involved in handling only two. Naturally, a search for the relationship between two variables is started by looking at the values of these variables which have been observed in the past as occurring together. For example, find the ages of husbands and wives for 100 married couples and try to make an estimate of some formula from the data on the 100. Call the two variables X and Y, and assume that N is the number of pairs of values of these variables which are in the basic data. Consider X as being given and try to find a formula which would allow a prediction to be made of the most probable Y to associate with that X. Unfortunately, this problem is very difficult if the formula is complicated. If, however, a simple formula is assumed, as in the case of the cost of the pencils, the problem is quite simple. Suppose that the relationship is

$$Y = A + BX\,.$$

The best such formula will be found, and what is meant by the word best should be defined. Obviously, it should be concerned with the size of the errors involved. By errors it is meant that, in an individual case, there is no guarantee that the Y value, estimated from the known X, is exactly correct. It is hoped that if there is an error in one case in one direction it will probably be counterbalanced later with an error in the other direction. This is not quite enough, since it is not desirable to allow a large error to creep in under the assumption that later there will be a large error in the other direction. The desired condition can be imposed by saying that the squares of the errors are to be small. This will force all errors to be small and eliminate the chance of counterbalancing large errors, one against the other. This criterion is called the criterion of least squares. If A and B are chosen by the principle of least squares, then the resulting formula is the one which will be called best. It might be well to keep in mind that if the relationship is at all valid and that X does form a good basis for the prediction of Y by a formula of this type, this method or any other reliable method would give a close approximation to the correct answer. The pairs of values by X's and Y's can be denoted with subscripts in the following way:

$$(X_1, Y_1), (X_2, Y_2), \ldots., (X_N, Y_N)\,.$$

For each X there is a corresponding estimate of Y, and these X's and their estimates give N more pairs of values which will be written as:

$$(X_1, E_1), (X_2, E_2), \ldots, (X_N, E_N)\,.$$

The following formula is assumed:

$$Y = A + BX\,.$$

An application of the least squares condition requires that A and B be selected so that

$$\sum_{i=1}^{i=N} (Y_i - E_i)^2 \qquad .$$

is less than it would be for any other values of A and B which might be selected. Substituting the formula for the E in this expression, it becomes

$$\sum_{i=1}^{i=N} (Y_i - [A + BX_i])^2\,.$$

Take the derivative first with respect to the coefficient A, and secondly, with respect to the coefficient B, giving

$$-2\sum_{i=1}^{i=N} (Y_i - A - BX_i) \text{ and } -2\sum_{i=1}^{i=N} X_i(Y_i - A - BX_i)\,.$$

The minimum value is obtained by setting each of these expressions equal to 0. If this is done and the summations are separated to simplify the form, the following result is obtained:

$$NA + B\Sigma X_i = \Sigma Y_i\,, \text{ and}$$
$$A\Sigma X_i + B\Sigma X_i^2 = \Sigma X_i Y_i\,.$$

For convenience I have omitted the summation limits $i = 1$, $i = N$. Perhaps you remember your algebra sufficiently to see that it would be a simple matter to solve the two equations for A and B with the result:

$$A = \frac{\Sigma Y_i \Sigma X_i^2 - \Sigma X_i \Sigma X_i Y_i}{N\Sigma X_i^2 - (\Sigma X_i)^2}\,,$$

$$B = \frac{N\Sigma X_i Y_i - \Sigma X_i \Sigma Y_i}{N\Sigma X_i^2 - (\Sigma X_i)^2}\,.$$

At this point, as is often the case in mathematical derivations, it can be seen that it would be much easier if the sum of the X's and the sum of the Y's in these two expressions were each 0, for then A and B would have many fewer terms. Since the sum of the deviations of a variable about its mean is 0, this situation could have been obtained if the mean of X had been subtracted from each of the X values and the mean of Y from each of the Y values at the start. Denote these means by $\overline{X}$ and $\overline{Y}$, respectively, and let

$$x_i = X_i - \overline{X} \quad \text{and} \quad y_i = Y_i - \overline{Y} \text{ for all } i\text{'s}.$$

Then Σx_i and Σy_i would each equal 0. Furthermore, all of the steps above would be the same, but with small letters replacing capitals. The answers would have appeared in the simple form

$$a = 0, \quad b = \frac{\Sigma x_i y_i}{\Sigma x_i^2}\,;$$

this assumes that the original equation would have been written

$$y = a + bx\,.$$

The best estimating equation would be

$$y = \frac{\Sigma x_i y_i}{\Sigma x_i^2}\, x\,.$$

This is not a commonly used form. It shall be changed to the more usual one, and I will explain the advantages by the later discussion. This common form makes use of three constants: the two standard deviations σ_x and σ_y and the correlation coefficient r_{xy} which can be defined as:

$$\sigma_x = \sqrt{\frac{\Sigma x_i^2}{N}}\,, \quad \sigma_y = \sqrt{\frac{\Sigma y_i^2}{N}}\,, \quad r_{xy} = \frac{\Sigma x_i y_i}{N\sigma_x\sigma_y}\,.$$

These can be introduced by the following manipulation:

$$\Sigma x_i^2 = N\sigma_x^2 \quad \text{and} \quad \Sigma x_i y_i = N\sigma_x\sigma_y r_{xy}\,.$$

Therefore, $\dfrac{\Sigma x_i y_i}{\Sigma x_i^2} = \dfrac{N\sigma_x\sigma_y r_{xy}}{N\sigma_x^2} = \dfrac{\sigma_y r_{xy}}{\sigma_x}\,.$

Substituting in the equation, we have

$$y = \frac{\sigma_y r_{xy}}{\sigma_x}\,x\,.$$

Of course, this equation can be expressed in terms of the original X and Y values by writing

$$Y - \overline{Y} = \frac{\sigma_y r_{xy}}{\sigma_x}\,(X - \overline{X}), \text{ or}$$

$$Y = \frac{\sigma_y r_{xy}}{\sigma_x}\,(X - \overline{X}) + \overline{Y}\,.$$

It was noted above that the errors in the estimates should be small. The size of these errors can be determined by calculating the standard deviation of the difference $Y - E$. Denote this difference by d.

Then $d = Y - E = Y - \left[\dfrac{\sigma_y r_{xy}}{\sigma_x}\,(X - \overline{X}) + \overline{Y}\right]$

$$= Y - \overline{Y} - \frac{\sigma_y r_{xy}}{\sigma_x}\,(X - \overline{X}) = y - \frac{\sigma_y r_{xy}}{\sigma_x}\,x\,.$$

For convenience I shall omit summation subscripts, *i*, in the future.

$$\begin{aligned}
\sigma_d^2 &= \frac{1}{N}\,\Sigma d^2 - \left(\frac{1}{N}\,\Sigma d\right)^2 = \frac{1}{N}\,\Sigma\left(y - \frac{\sigma_y r_{xy}}{\sigma_x}\,x\right)^2 \\
&\qquad - \frac{1}{N^2}\left(\Sigma\left[y - \frac{\sigma_y r_{xy}}{\sigma_x}\,x\right]\right)^2 \\
&= \frac{1}{N}\,\Sigma\left(y^2 - 2\frac{\sigma_y r_{xy}}{\sigma_x}\,xy + \frac{\sigma_y^2 r_{xy}^2}{\sigma_x^2}\,x^2\right) \\
&\qquad - \frac{1}{N^2}\left(\Sigma y - \frac{\sigma_y r_{xy}}{\sigma_x}\,\Sigma x\right)^2 \\
&= \frac{1}{N}\,\Sigma y^2 - 2\,\frac{\sigma_y r_{xy}}{\sigma_x}\,\frac{\Sigma xy}{N} + \frac{\sigma_y^2 r_{xy}^2}{\sigma_x^2}\,\frac{\Sigma x^2}{N} - 0 \\
&= \sigma_y^2 - 2\,\frac{\sigma_y r_{xy}}{\sigma_x}\,\frac{N\sigma_x\sigma_y r_{xy}}{N} + \frac{\sigma_y^2 r_{xy}^2}{\sigma_x^2}\,\sigma_x^2 \\
&= \sigma_y^2 - 2\sigma_y^2 r_{xy}^2 + \sigma_y^2 r_{xy}^2 \\
&= \sigma_y^2 - \sigma_y^2 r_{xy}^2 \\
&= \sigma_y^2\,(1 - r_{xy}^2)\,.
\end{aligned}$$

The standard deviation of these differences is called the standard error of estimate and is usually denoted by S_y. The formula for S_y shows that the constant r_{xy} is never numerically larger than 1, since the quantity calculated is non-negative. The standard error of estimate is 0 only in case the r_{xy} is equal to plus or minus 1. Before leaving this subject to discuss the more complicated cases, it is well to find an answer to the question, "What is the variability present in the estimates, E, themselves?" This can be found by calculating the standard deviation of these E values. The mean of the E's, $\overline{E}$, is:

$$\begin{aligned}
\overline{E} &= \frac{1}{N}\,\Sigma\left[\frac{\sigma_y r_{xy}}{\sigma_x}\,(X - \overline{X}) + \overline{Y}\right] \\
&= \frac{1}{N}\,\frac{\sigma_y r_{xy}}{\sigma_x}\,\Sigma(X - \overline{X}) + \frac{1}{N}\,N\overline{Y} \\
&= \overline{Y}\,.
\end{aligned}$$

Then the variance of the E's, σ_E^2, is:

$$\begin{aligned}
\sigma_E^2 &= \frac{1}{N}\,\Sigma(E - \overline{E})^2 = \frac{1}{N}\,\Sigma\left[\frac{\sigma_y r_{xy}}{\sigma_x}\,(X - \overline{X}) + \overline{Y} - \overline{Y}\right]^2 \\
&= \frac{1}{N}\,\frac{\sigma_y^2 r_{xy}^2}{\sigma_x^2}\,\Sigma(X - \overline{X})^2 \\
&= \frac{\sigma_y^2 r_{xy}^2}{\sigma_x^2}\,\frac{1}{N}\,\Sigma x^2 \\
&= \frac{\sigma_y^2 r_{xy}^2}{\sigma_x^2}\,\sigma_x^2 \\
&= \sigma_y^2 r_{xy}^2\,.
\end{aligned}$$

This says that $\sigma_E = \sigma_y r_{xy}$.

The standard deviation of the E's, the standard error of estimate and the standard deviation of the original Y values are all related by an interesting formula which can most easily be expressed in terms of the squares of these various standard deviations. Reference to the formulas just derived immediately shows that

$$S_y^2 = \sigma_y^2 - \sigma_E^2\,,$$

which is more usually written

$$\sigma_y^2 = S_y^2 + \sigma_E^2\,.$$

This is the first time the statistics student encounters this formula, which is later termed *the analysis of variance;* the square of the standard deviation is frequently termed *variance.* In this case the variance of the original Y values is broken into two parts, one of which refers to the variability in the deviations between the estimates and the actual values, S_y^2, and the other which refers to the deviations in the estimates themselves, σ_E^2. It is common to speak of the latter as that amount of variability in the Y variable, which is explained by the variable X. The former is that portion of the variability of Y which is unexplained. It is clear that if S_y^2 is large and σ_E^2 is small, the variability in Y is not particularly associated with the variability in X. On the other hand, if S_y^2 is small and σ_E^2 is large, then a great deal of the variability in Y can be explained in terms of the independent variable X. The extreme case is that in which one of these two is 0. For example, if S_y^2 is 0, then *all* the deviations are 0 and every estimate is perfect. This situation occurs in examples such as the one involving the cost of the pencils. On the other hand, if σ_E^2 is 0, then none of the variability is explained in terms of X, and the situation is of the type illustrated by the population of the state and the size of the shoe of the senator. Usually the intermediate situation arises and the attempt to explain Y in terms of X is partly successful.

It might be well to consider a numerical example (Table I). For convenience, I have chosen one-digit values for X and Y, and I have assumed N to be 10 so that the division is easy. The arithmetic of the calculation is carried out in terms of the original variables, as well as in terms of the small letters which denote deviations from the mean. By the proper choice of values it was possible to obtain means which were whole numbers. This simplified the arithmetic considerably. In this instance r_{xy} was calculated to be .867, indicative of a relatively high degree of association between X and Y.

It is not the purpose of this paper to discuss significance tests. It is easily seen that the importance of a result of this type is dependent on the number of pairs of values on which the calculations have been based.

A very illuminating approach to the subject of correlation can be obtained by an examination of the relationship between the observed values of the dependent variable, Y, and the estimate of this variable, E. The correlation between these two variables is exactly the same as the correlation between X and Y. This can be easily established if the small letters are used instead of the big letters. I

TABLE I
EXAMPLE OF SIMPLE CORRELATION*

	X	Y	X^2	Y^2	XY	x	y	x^2	y^2	xy	E	YE	E^2
	9	8	81	64	72	4	4	16	16	16	7.125	57.	50.765625
	9	6	81	36	54	4	2	16	4	8	7.125	42.75	50.765625
	7	5	49	25	35	2	1	4	1	2	5.5625	27.8125	30.94140625
	6	7	36	49	42	1	3	1	9	3	4.78125	33.46875	22.8603515625
	5	3	25	9	15	0	−1	0	1	0	4.	12.	16.
	4	2	16	4	8	−1	−2	1	4	2	3.21875	6.4375	10.3603515625
	3	2	9	4	6	−2	−2	4	4	4	2.4375	4.875	5.94140625
	3	4	9	16	12	−2	0	4	0	0	2.4375	9.75	5.94140625
	2	2	4	4	4	−3	−2	9	4	6	1.65625	3.3125	2.7431640625
	2	1	4	1	2	−3	−3	9	9	9	1.65625	1.65625	2.7431640625
Σ	50	40	314	212	250	0	0	64	52	50	40.	199.0625	199.0625

*There is no rounding in the tabular values.

$$\overline{X} = 5, \quad \overline{Y} = 4, \quad N = 10$$

$$\sigma_x^2 = \frac{\Sigma X^2}{N} - \overline{X}^2 = 31.4 - 25 = 6.4 = \frac{\Sigma X^2}{N} \qquad \sigma_x = 2.5298$$

$$\sigma_y^2 = \frac{\Sigma Y^2}{N} - \overline{Y}^2 = 21.2 - 16 = 5.2 = \frac{\Sigma Y^2}{N} \qquad \sigma_y = 2.2804$$

$$r_{xy} = \frac{\Sigma XY - \overline{X}\Sigma Y}{N\sigma_x\sigma_y} = \frac{250 - 5(40)}{10(2.5298)(2.2804)} = \frac{50}{57.69} = 0.867 = \frac{\Sigma xy}{N\sigma_x\sigma_y}$$

Normal Equations: $NA + B\Sigma X = \Sigma Y \qquad 10A + 50B = 40$

$$A\Sigma X + B\Sigma X^2 = \Sigma XY \qquad 50A + 314B = 250$$

Solution $Y = 0.09375 + 0.78125X$

$$\sigma_E^2 = \frac{199.0625}{10} - 16 = 3.90625 \qquad \sigma_E = 1.9764$$

$$r_{EY} = \frac{\Sigma YE - \overline{Y}\Sigma E}{N\sigma_F\sigma_Y} = \frac{199.0625 - 4(40)}{10(1.9764)(2.2804)} = \frac{3.90625}{4.5070} = 0.867 = r_{xy} .$$

shall not bother to prove the following statement, but it is true that if each of the values of one of the variables in a correlation is multiplied by any constant other than 0, or if any constant is added to each of the values, or if both of these things are done, the correlation coefficient itself is unchanged. Since the difference between x and X is merely that X minus its mean is x, this will not change the correlation. A similar statement can be made with respect to Y. The truth of the above assertion about the correlation of Y with its estimate can now be established by calculating

$$r_{yE} = r_{ye} = \frac{\Sigma ye}{N\sigma_y\sigma_e},$$

remembering the relationship between the estimate and x. This formula can be expressed in terms of x by substituting

$$e = \frac{r_{xy}\sigma_y}{\sigma_x}x$$

in the formula. This gives

$$r_{ye} = \frac{1}{N\sigma_y\sigma_e}\Sigma y\frac{r_{xy}\sigma_y}{\sigma_x}x,$$

which can be reduced to

$$r_{ye} = r_{xy}\frac{\Sigma xy}{N\sigma_x\sigma_e}.$$

The only unknown term in this expression is the standard deviation of the estimated values which can be easily calculated as:

$$\sigma_e^2 = \frac{1}{N}\Sigma\frac{r_{xy}^2\sigma_y^2}{\sigma_x^2}x^2,$$

which reduces to

$$\sigma_e^2 = \frac{r_{xy}^2\sigma_y^2}{\sigma_x^2}\frac{\Sigma x^2}{N}$$

or

$$\sigma_e^2 = r_{xy}^2\sigma_y^2.$$

If the above relation is substituted in the original expression for the correlation, and is simplified by the following steps, the result is apparent:

$$r_{ye} = r_{xy}\frac{\Sigma xy}{N\sigma_x r_{xy}\sigma_y} = \frac{\Sigma xy}{N\sigma_x\sigma_y} = r_{xy}.$$

This formula states that the degree of agreement between x and y, as measured by correlation, is the same as the agreement between y and its estimate, when this estimate is based on x by the assumed formula. It is this approach which will be found most useful in the subject of multiple correlation to be discussed later.

There are other properties of simple correlation which might be of considerable interest. One of these important properties is the fact that if y is estimated from x and then an attempt is made to estimate x from y, the two formulas are not identical. By this I mean that the equation

$$Y = A + BX$$

could not be solved for X to obtain the best estimating equation for guessing at an X which is associated with a given Y. Therefore, there are actually two equations: one minimizes the squares of the errors in one direction, and the other minimizes the squares of the errors in the other direction. These two equations will be the same only in case r is equal to plus or minus 1. Geometrically, the graphs of these two equations are straight lines, called the regression lines. They intersect at a point whose X value is the mean of the X's and whose Y value is the mean of the Y's. In the case of perfect correlation, r is numerically equal to 1; the two lines coincide, and all points lie exactly on the lines. In case of no correlation, the lines are perpendicular—one horizontal and the other vertical.

I have spent a great deal of time considering the simple case involving two variables, because it is by analogy that this can be extended to multiple correlation in which any number of variables are being considered. It is convenient to think in geometric terms. The simple two-variable case, geometrically, is the problem of fitting a straight line to a set of points, which will all lie on the same straight line if r_{xy} is equal to either plus or minus 1, and which will scatter widely about the line if r_{xy} is close to 0. The d values are the vertical distances from the points on the line to the observed values of Y, and the E values, or the estimates of Y, are the distances from the horizontal base line to the point on the line.

If there are three variables, the geometric analogy is quite simple because it can still be thought of in terms of pictures which can be drawn. Unfortunately, it is difficult to visualize more than three dimensions, but the geometers who discuss these subjects carry over the language of solid geometry into higher dimensions by speaking of hyperplanes instead of planes. For convenience the three-dimensional terminology will be used. The formula requires the estimate of one variable in terms of two others. It shall be assumed to be a first-degree expression similar to the above for two dimensions, such as

$$Z = A + BX + CY.$$

Now Z is estimated from X and Y; before Y was estimated from X. Geometrically, this represents a plane in three dimensions. Values of d and E may be discussed just as before. The E value is the value of Z obtained by substituting a given pair of X and Y values in the right-hand side of the equation. It is a vertical distance from the base plane to the approximating plane. The d value is the vertical distance from this point on the estimating plane to the actual observed point. The only difference in the formulas between this and the previous case is that of the estimating equation,

$$Z = A + BX + CY\,.$$

The formula for d is identically the same,

$$d = Z - E\,,$$

with, of course, the substitution of Z in place of Y as the estimated variable. Of course, the formulas for A, B and C will be derived by a slightly more complicated process, since there are more letters involved; therefore, there will be more equations. The criterion of least squares still gives the methods necessary to obtain these answers through the media of calculus and algebra.

For the present those equations will not be discussed, and the correlation coefficient itself will be considered more directly. Assume that an analogous standard error of estimate could be calculated. If the formula were completely worked out for the coefficients A, B and C, the given values of X and Y could be substituted and the estimates, E, calculated. Then each E could be subtracted from its corresponding Z and the values of the d's found. Then the calculation of the standard deviation of these d values would give the standard error of estimate. The important point is that the standard error of estimate does not involve any complication due to the existence of the two variables X and Y on the right side of the equation. It is still merely the standard deviation of a set of d's. Now this standard error of estimate could be assumed to be of a similar form:

$$S_z^2 = \sigma_z^2\,(1-R_{z.xy}^2)\,.$$

This R^2 plays exactly the same role as r^2 did in the previous case. I assume it is clear that I was forced to use z in the subscript in place of y as I did before, because z is now the variable being estimated. However, in the other instance I put two subscripts on the r. There is an interesting point in reference to the r, as well as a new one about R. Reference to the formula for r will disclose the fact that it is symmetrical with respect to X and Y. If every X were replaced by a Y and every Y were replaced by X in the formula for r, the answer would be unchanged. It has not been established, but it is a fact that with the three variables this symmetry is not maintained. We must differentiate in our minds between the variable which is being estimated and the variables on which the estimate is based. For this reason it is common to use a notation as follows:

$$R_{Z.XY}\,.$$

You might naturally guess that the X and Y can be interchanged in this formula without changing the value of the answer. However, you could not interchange the X and Z or the Y and Z without changing the value.

The notion of multiple correlation, as incompletely described above, is the first and simplest of the extensions of the simple case. It is a measure of the association between a given variable Z and its estimate based on two other variables X and Y. The inherent relationship between Z and X, also, might be desired, for example, if the effects of Y on each of the variables X and Z were removed. Stated in a more direct form it is: calculate the amount of X that cannot be explained in terms of Y and calculate the amount of Z that cannot be explained in terms of Y. Then the relationship between these two amounts is the relationship measured by means of a partial correlation coefficient. In another form, it is asked, how much of the variability in Z, which cannot be explained in terms of Y, can be explained in terms of that part of variability in X, which also cannot be explained in terms of Y? When stated in this form, the formula and the arithmetic are almost obvious. What must be done is to find the differences between the values of Z and the estimated value of Z, if this estimate is based only on Y. The same must be done for X; namely, find the difference in the actual value of X and the value that would be obtained by estimating X from Y alone.

These two calculations will result in two variables which can be thought of in exactly the same way as X and Y in the case of simple correlation. The resulting correlation coefficient is called the partial correlation between X and Z with the effect of Y eliminated or with Y being held constant. The latter description is probably the least desirable of the two.

Now return to the problem of finding the estimating formula for a case with more than two variables. For convenience the formula

$$Z = A + BX + CY$$

referred to before, will be used. The extension to more variables will be very easy if this case is thoroughly understood. First express the deviations between the estimated values of Z and the observed values. Then square these deviations, add them all and then choose the numerical values of A, B and C so that this sum of squares of deviations is less than it would be for any other selection of the constants A, B and C. The expression for this sum of squares is

$$\Sigma[Z - (A+BX+CY)]^2.$$

At this point, again, calculus must be used to obtain the resulting equations. The steps involved are merely partial differentiation with respect to the three coefficients A, B and C. These partial derivatives are as follows:

$$-2\Sigma\;\;[Z - (A+BX+CY)]$$
$$-2\Sigma X[Z - (A+BX+CY)]$$
$$-2\Sigma Y[Z - (A+BX+CY)]\,.$$

The minimum sum of squares is obtained when these derivatives are set equal to 0 and the resulting equations solved for the unknown coefficients A, B and C.

$$NA + B\Sigma X + C\Sigma Y = \Sigma Z$$
$$A\Sigma X + B\Sigma X^2 + C\Sigma XY = \Sigma XZ$$
$$A\Sigma Y + B\Sigma XY + C\Sigma Y^2 = \Sigma YZ\,.$$

We shall consider the formal solution a little later. Now I would like to point out the general formation of these equations. Note that the first equation contains all the terms in the original predicting equation with a few summation signs and a coefficient N. The second equation is similar to the first one, except that each term involves an extra X, and the N does not appear. The third equation is similar to the second, except that it has an extra Y in each term instead of an extra X. If there were more than three variables, there would be more than three equations. The number of equations is always equal to the number of variables. The additional equations in the case of additional variables are formed like the last two mentioned above, in which an extra letter appears in each term. This means that there are, essentially, a set of equations with coefficients, which involve cross products of different variables and squares of variables all added together with some simple additions of the variables themselves. These simple additions always occur in the first equation and the first terms of the others. Again, notice that the whole set could have been materially simplified if x's, y's and z's had been selected instead of X's, Y's and Z's. This modification would have eliminated the first equation because all the terms except NA would have been 0. Thus, if the small letters are used, indicating deviations from the mean, for each variable, there would be only two equations when there are three variables. In general, the number of equations will be the same as the number of dependent variables on which the prediction is based. In the present example this is two, x and y. There would be also only two unknowns, b and c (a would be zero). Notice that in terms of the little letters the equations are

$$b\Sigma x^2 + c\Sigma xy = \Sigma xz$$

$$b\Sigma xy + c\Sigma y^2 = \Sigma yz\,,$$

and involve only sums of squares and sums of cross products. These sums contain deviations from the mean, and hence are numerators in the formulas for correlation coefficients and variances, the square of the standard deviation. For example, Σxy is the numerator in the formula for correlation coefficient

$$r_{xy} = \frac{\Sigma xy}{N\sigma_x\sigma_y}\,.$$

For completeness, I shall also write the formula for σ_x^2,

$$\sigma_x^2 = \frac{\Sigma x^2}{N}\,.$$

In this formula the numerator occurs as one of the coefficients in the two equations. It is, therefore, possible to express all of the coefficients in these equations directly in terms of correlation coefficients and variances. The result of this method of expressing it is to obtain an array of coefficients in which the variances occur in somewhat of a diagonal position, while the correlation coefficients themselves would enter in the positions off the diagonal. This can be done in the following way:

$$\frac{\Sigma x^2}{N}\,b + \frac{\Sigma xy}{N}\,c = \frac{\Sigma xz}{N}$$

$$\frac{\Sigma xy}{N}\,b + \frac{\Sigma y^2}{N}\,c = \frac{\Sigma yz}{N}\,,\ \text{or}$$

$$\sigma_x^2 b + r_{xy}\sigma_x\sigma_y c = r_{xz}\sigma_x\sigma_z$$

$$r_{xy}\sigma_x\sigma_y b + \sigma_y^2 c = r_{yz}\sigma_y\sigma_z\,.$$

This array of coefficients, when placed within the proper symbol bars, is called the variance-covariance matrix. The detail of the numerical solution of these equations will be discussed in another paper; so I shall not discuss that section of the theory, but say merely that the symmetry of the system of coefficients is the foundation for the development of a very simple solution routine, which is commonly known as the Doolittle method. The details of this method are unimportant for the discussion, because what is of interest here is the form of the relationship which does exist.

Indeed, further amplification of the theory of the normal equations needed to obtain the coefficients in the predicting equation is unnecessary, since the concept of multiple regression itself was attained without any necessity for finding the exact method by which the coefficients could be determined. This point is particularly important if you are interested in the subject of graphic multiple correlation methods or in the subject of curvilinear correlation. The concept of the correlation coefficient as a portion of the formula for the standard error of estimate easily suggests a method for doing correlations graphically. An approximating curve, not necessarily a line, can be drawn from an assumed equation or by a freehand fit to a given set of data. The deviations from this curve can be measured directly from the graph. The variance of these deviations will then be directly analogous to the standard error of estimate. The standard error of estimate can then be solved for the curvilinear correlation by means of the formula

$$r^2 = 1 - \frac{s^2}{\sigma^2}\,.$$

In this case r^2 and s^2 have meanings a little different from those in the straight line case. They are now measured from a curve.

Regression in Terms of Independent Variables

Consider a set of variables denoted by

$$X_1, X_2, X_3, \text{and } X_4\,.$$

For convenience the discussion will be restricted to the case of four variables; the extension to more is obvious. Let their deviations from their means be

$$x_1, x_2, x_3, \text{and } x_4\,.$$

Let x_1 be the dependent variable and consider the linear estimate of x_1 from x_2 alone by the least squares method. This can be written

$$x_1 = c_{12}\, x_2 \,.$$

The estimate of x_1 from x_2 is, in general, less reliable than the estimate from x_2 and x_3. Since x_2 and x_3 are usually correlated, use of only that portion of x_3 not included by previously using x_2 could be considered. This can be done by including, as an additional variable, the difference between x_3 and its estimate based on x_2. Denote this by $x_{3.2}$. The estimate is assumed to be the linear least square estimate. Then the estimating equation for x_1 would be

$$x_1 = c_{12}\, x_2 + c_{13.2}\, x_{3.2} \,.$$

This could be extended to include x_4 in a similar manner, namely, by adding that portion of x_4 not included in x_2 and x_3 as estimated by a linear least square estimate. Denote this by $x_{4.23}$. Then the estimating equation would be

$$x_1 = c_{12}\, x_2 + c_{13.2}\, x_{3.2} + c_{14.23}\, x_{4.23} \,.$$

This equation yields a set of normal equations with a single unknown in each equation. The variables on the right are orthogonal polynomials involving the original variables x_2, x_3 and x_4. The predictions resulting from this method are identical with those obtained from the usual regression method. Indeed, this procedure is the basis for an obvious method of graphic multiple correlation. It is of little value as a numerical tool because of the heavy arithmetic. It should be noted that the addition of other variables leaves previously calculated coefficients unchanged.

DISCUSSION

Mr. Bejareno: Could you give me a reference on correlation based on curve fitting which would be applicable to engineers' problems where they vary from a straight line, and it is apparent that there is a curve of some sort. The book known to me is *Graphical and Mechanical Computation* by Joseph Lipka (New York: John Wiley, 1918).

Dr. Sherman: In connection with curve fitting, I have a few remarks. One problem which comes up in engineering is that if you clock your data and try to fit it by means of a polynomial, there are cases in which the polynomial carried to rather a high degree is not a satisfactory fit, simply because the data may, for example, fit a hyperbola. There is a good section in William Edmund Milne, *Numerical Calculus* (Princeton University Press, 1950) on this point. Another book which I have found useful in curve fitting is Deming's book, *Adjustment of Experimental Data.* General problems of curve fitting in which the variables x and y may be subject to error are answered, and good examples are given.

Pile-Driving Impact

EDWARD A. SMITH

Raymond Concrete Pile Company

THIS CALCULATION, performed in the IBM Technical Computing Bureau, provided a complete numerical analysis of the behavior of a pile when struck by a pile-driving hammer. The results indicated what stresses occurred within the pile and capblock, as well as the penetration of the pile into the earth. This is an example of the replacement of costly experimentation by economical calculation.

Previously, using only key-driven calculators, it has been possible to study only partially a few isolated cases of the behavior of a pile under an impact. This was due to the high cost in time and money associated with each case studied. By employing the high speed and accuracy of IBM electronic calculating machines to perform this repeated formula evaluation, it is possible to study the behavior of numerous pile types. The cost of each case studied thereby will be substantially reduced.

When piles are driven as a foundation for a building or other structure, the load that they will carry safely usually is determined by measuring the penetration under the last blow of the hammer and substituting this figure in a formula. Many different formulas are in use, and they vary widely in the answers they give. This calculation provides a means of mathematically testing the accuracy of these formulas when applied to various types of piles and various ground conditions.

Furthermore, when the engineer has been asked what stresses the pile or the pile-driving equipment should be designed to withstand, he has often been at a loss for an answer. He has known only that certain practices and equipment have proved successful in the past. To do something new has meant proceeding by the often costly process of trial and error.

Because a pile is an object of considerable length, pile driving is a problem in longitudinal wave transmission and impact, the basic principles of which were first investigated 100 years ago by the French mathematicians, Saint Venant and Boussinesq. These principles have been ably set forth by L. H. Donnell in ASME Transactions APM-52-14. However, the problem is a very complicated one. The method of solution offered herein is based on approximate integration using a step-by-step calculation, and is of general interest because it can be applied to other impact problems. The calculation can be done by hand with a slide rule or desk calculating machine; however, modern electronic digital calculating machines are especially well adapted to make the calculation quickly and easily.

General Case—Basic Theory

For purposes of analysis, the hammer, pile, etc., will be represented by a series of concentrated weights separated by weightless springs. Subscripts $m-1$, m, $m+1$, etc., will be used to denote order of position, and superscripts $n-1$, n, $n+1$, etc., will be used to denote order of time.

Referring to Figure 1, let W_{m-1}, W_m and W_{m+1} represent successive weights, and S_{m-1}, S_m and S_{m+1} the springs underneath the respective weights. Let the initial positions

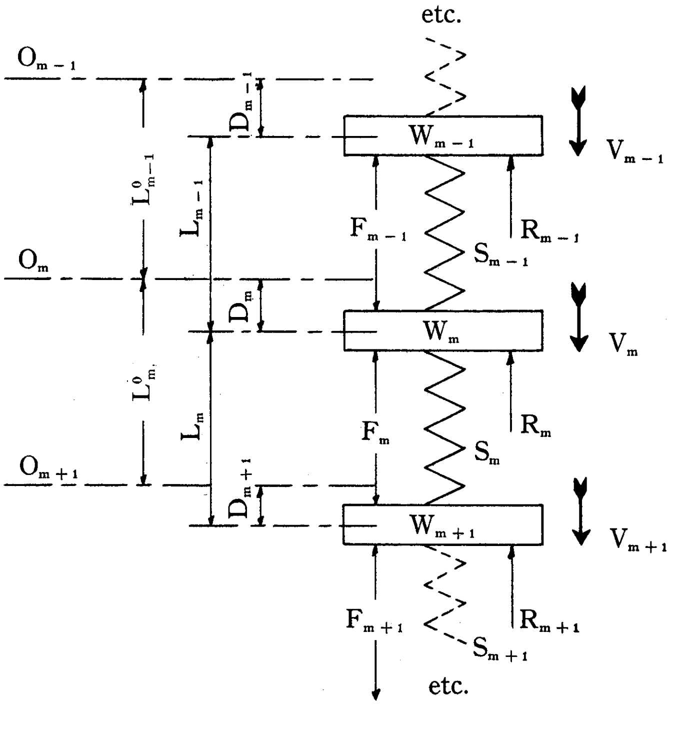


FIGURE 1

]

of weights W_m, etc., at the beginning of impact be indicated by O_m, etc., and the initial lengths of the springs, by L_m^0, etc. Also, let

D_m, etc. = instantaneous displacements (inches).

L_m, etc. = instantaneous lengths (feet).

C_m, etc. = $12[L_m^0 - L_m] = D_m - D_{m+1}$ = instantaneous amount of spring compression (inches).

K_m, etc. = elastic constants for springs S_m, etc. = force required to produce 1" of compression C_m, etc. (pounds per inch).

F_m, etc. = instantaneous forces (pounds) resulting from C_m, etc.

V_m, etc. = instantaneous velocities of W_m, etc. (fps).

R_m, etc. = external forces, such as ground resistance, affecting the motion of W_m, etc. (pounds).

Z_m, etc. = instantaneous net force acting on W_m, etc. = $F_{m-1} - F_m - R_m$, etc. (pounds).

t = time (seconds).

Also, let superscripts $n-1$, n, $n+1$, etc., denote successive time intervals Δt, so small that with negligible error it may be assumed that all forces and velocities remain constant during each time interval. Then, if by previous calculation or otherwise, the values of V and D are known for some particular time interval $n-1$, all values for C, F, Z, V and D for the next time interval can be calculated as follows:

Let V_m^{n-1} and D_m^{n-1}, etc., represent definite known values for time interval $n-1$, and let C_m^n, F_m^n, etc., represent definite values to be calculated for time interval n. Then the following general formulas apply:

$$C_m^n = D_m^{n-1} - D_{m+1}^{n-1} \qquad (1)$$

$$F_m^n = K_m C_m^n \qquad (2)$$

$$Z_m^n = F_{m-1}^n - F_m^n - R_m^n \qquad (3)$$

$$V_m^n = V_m^{n-1} + \Delta V_m^n = V_m^{n-1} + \left(\frac{32.17\Delta t}{W_m}\right) Z_m^n \qquad (4)$$

$$D_m^n = D_m^{n-1} + \Delta D_m^n = D_m^{n-1} + (12\Delta t)\, V_m^n \qquad (5)$$

ΔV above is evaluated by using the standard formula for change of velocity

$$(v_1 - v_2) = \frac{\text{Force} \times \text{Time}}{\text{Mass}} = \frac{g\,\text{Ft}}{W},$$

where W is the weight, and g is the acceleration of gravity. ΔD above is evaluated by the formula for distance traveled $s = vt$ with the coefficient 12 introduced to convert to inches.

Careful consideration of the above formulas will disclose that the force at the beginning of an interval is used to calculate the velocity at the end of the interval, and then this end velocity is used to calculate the distance traveled during the same interval; therefore, the forces, velocities, and displacements used are slightly out of step with one another, depending on the size of the time interval.

These five formulas are used repeatedly until the calculation has been carried as far as necessary. They apply whether or not the successive weights, elastic constants, and external forces are equal or unequal. Furthermore, the values of K, R, and Δt may be changed from interval to interval according to any definite formula, or suddenly as required. Sudden changes are required, for instance, when the stress in a material reaches the yield point, when a weight loses contact with a spring designed only for compression, when a coefficient of restitution is introduced, or when a ground resistance force must be made negative so as to resist temporary upward movements. Such conditions are called *boundary conditions.* Some of the more complicated digital calculators can take care of these boundary conditions automatically.

Choice of Lengths L and Time Intervals Δt

It should be borne in mind that formulas (1) to (5) involve the use of small but finite increments, not infinitesimals. The lengths L and the time intervals Δt must, therefore, be chosen small enough to suit each particular type of problem. For each new type of problem halving or quartering the size of the units and recalculating part of the problem must be tried until it is found that the use of smaller units makes a negligible difference in the peak stresses that occur soon after impact. The time interval can be changed while a calculation is in progress by inserting a new value for Δt in formulas (4) and (5), although this change is subject to the limitation pointed out in Step 2 below.

Illustrative Problem

A pile of non-uniform section as shown in Figure 2 is to be driven through water or very soft mud to a hard layer of ground which is capable of resisting a maximum force of 600,000 pounds under the pile point. If the point of the pile starts to move upward momentarily, a negative frictional force of 100,000 pounds must be assumed, acting to hold the point down. No other side frictional forces are to be considered. The calculation is made to determine the final penetration per blow at which the assumed point resistance of 600,000 pounds will be developed.

Side friction along the pile has been omitted from this problem so as to allow the stress wave to travel with the known speed of stress (or sound) in the pile material and thus provide one way of checking the calculation. The method would apply equally well no matter what values were assigned to side friction.

Step 1. Decide on the time interval Δt. From previous experience a time interval of 1/4000 second has been chosen as being small enough to give accuracy within about 5%.

Step 2. Decide on lengths L. These must be at least as great as the distance stress will travel in the chosen time interval Δt; otherwise the stress wave will run ahead of the calculation, and the results will be meaningless. It is recommended that L be made equal to twice the distance that stress would travel in the chosen time interval. The upper part of this pile is entirely of steel, and the known speed of stress in steel is 16,800 fps; therefore, the recommended length for L is $(16{,}800 \times 2) \div 4000 = 8.4$ feet. The pile length, plus a little added for the follower, happens to be a multiple of this figure; therefore, 8.4 feet can be used throughout. If an odd length were required, it would be inserted at the point of the pile.

Step 3. Prepare a diagram as per Figure 3 showing how the ram, capblock, follower, and pile are to be represented for purposes of calculation. The individual weights W_1, etc., are calculated so as to give a weight distribution closely equivalent to that of Figure 2.

Step 4. Prepare a tabulation of all constants required for formulas (1) to (5) as per Figure 4. This is readily done by considering the weight, cross-section, and modulus of elasticity of each portion of Figure 2 equivalent to a single spring or weight in Figure 3. The elastic constant K_1 for the wooden capblock must be determined by experiment or must be assumed. For this problem, a value of 6,400,000 pounds per inch has been assumed, which represents a rather stiff capblock, perfectly elastic.

Step 5. If the work is to be done by hand, it is conveniently tabulated as shown in Figure 5, which covers only the first three time intervals. In order to start the calculation, it is necessary to have a value for V and a value for D in the first time interval. The value of V, representing the velocity of the ram at the beginning of

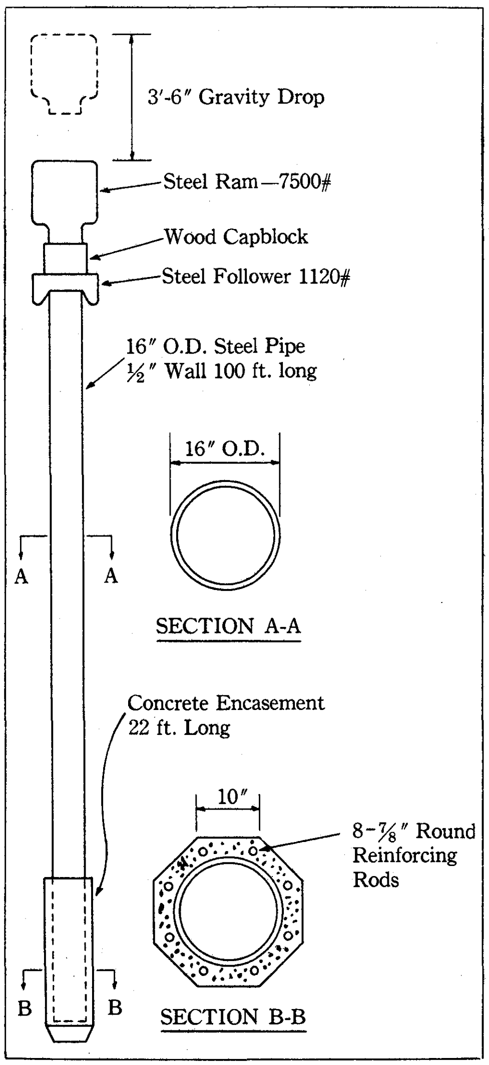


FIGURE 2

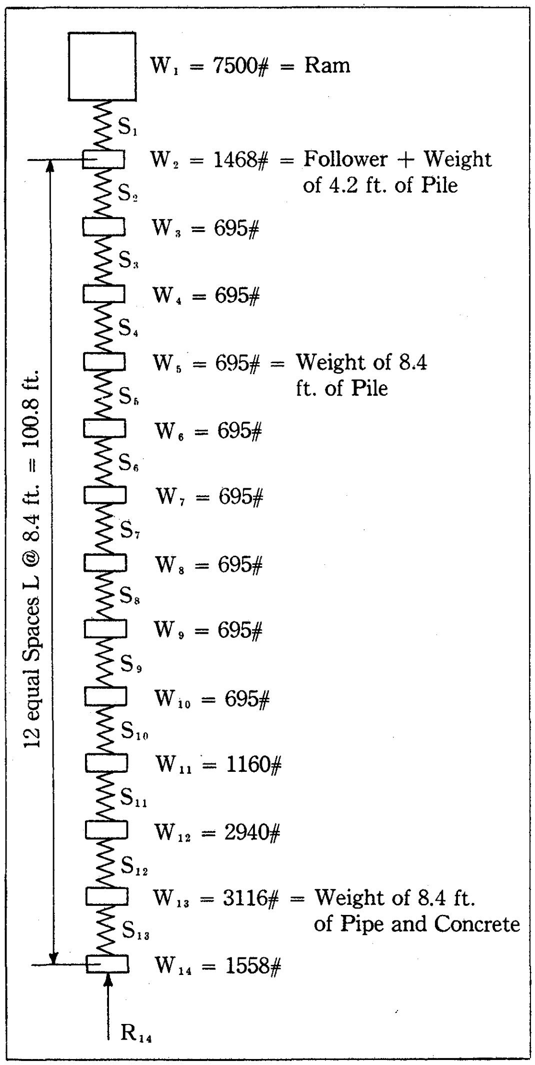

W₁ = 7500# = Ram
S₁
W₂ = 1468# = Follower + Weight of 4.2 ft. of Pile
S₂
W₃ = 695#
S₃
W₄ = 695#
S₄
W₅ = 695# = Weight of 8.4 ft. of Pile
S₅
W₆ = 695#
S₆
W₇ = 695#
S₇
W₈ = 695#
S₈
W₉ = 695#
S₉
W₁₀ = 695#
S₁₀
W₁₁ = 1160#
S₁₁
W₁₂ = 2940#
S₁₂
W₁₃ = 3116# = Weight of 8.4 ft. of Pipe and Concrete
S₁₃
W₁₄ = 1558#
R₁₄
12 equal Spaces L @ 8.4 ft. = 100.8 ft.

FIGURE 3

(Δt = 1/4000 Second)

Subscript "m"	W	K	R	$\frac{32.17\ \Delta t}{W}$	12 Δt
1	7500#	6,400,000#	0	1/932,500	0.003
2	1468	7,206,000	0	1/182,500	0.003
3	695	7,206,000	0	1/86,500	0.003
4	695	7,206,000	0	1/86,500	0.003
5	695	7,206,000	0	1/86,500	0.003
6	695	7,206,000	0	1/86,500	0.003
7	695	7,206,000	0	1/86,500	0.003
8	695	7,206,000	0	1/86,500	0.003
9	695	7,206,000	0	1/86,500	0.003
10	695	7,206,000	0	1/86,500	0.003
11	1160	14,900,000	0	1/145,000	0.003
12	2940	19,600,000	0	1/366,000	0.003
13	3116	19,600,000	0	1/388,000	0.003
14	1558		See Note	1/194,000	0.003

NOTE: *R_{14} assumed = F_{13} until F_{13} reaches 600,000#. Thereafter R_{14} = 600,000# if W_{14} is moving down and —100,000# if W_{14} is moving up.*

FIGURE 4

n	C_1	F_1	Z_1	V_1	D_1
	$D_1^{n-1} - D_2^{n-1}$	$6{,}400{,}000C_1^n$	$-F_1^n$	$V_1^{n-1} + \frac{Z_1^n}{932{,}500}$	$D_1^{n-1} + .003V_1^n$
	Inches	Pounds	Pounds	Ft. per Sec.	Inches
1	0	0	0	14.2500	.04275
2	.04275	273,600	−273,600	−.2934 13.9566	.04187 .08462
3	.08462 .00449 .08013	512,783	−512,783	−.5499 13.4067	.04022 .12484

n	C_2	F_2	Z_2	V_2	D_2
	$D_2^{n-1} - D_3^{n-1}$	$7{,}206{,}000C_2^n$	$F_1^n - F_2^n$	$V_2^{n-1} + \frac{Z_2^n}{182{,}500}$	$D_2^{n-1} + .003V_2^n$
	Inches	Pounds	Pounds	Ft. per Sec.	Inches
1	0	0	0	0	0
2	0	0	273,600	1.4992	.00449
3	.00449	32,409	512,783 32,409 480,374	2.6322 4.1314	.01240 .01689

NOTE: *Line 3 is not complete because the Force F_2^3 = 32,409 can be used to calculate values for Z_3^3, V_3^3 and D_3^3 which are not shown. Velocity of Ram at Impact = 14.250 fps.*

FIGURE 5

the impact, must be calculated by considering the distance it falls and allowing for hammer efficiency. For this problem, efficiency was assumed to be 90%, which gave a velocity of 14.25 fps to be used in starting the step-by-step calculation. The numerical value of the displacement D in the first time interval is obtained from the assumption that the ram continues to move with undiminished velocity through the first 1/4000 second after the impact. For an impact velocity of 14.25 fps this gives a displacement in the first time interval of 0.04275 inches. This displacement then is used to calculate force F_1 for the second time interval, and so on. As the stress wave travels down the pile, additional columns are needed in groups of five, all headed by the basic formulas (1) to (5) using constants taken from Figure 4.

If the work is done by an electronic digital calculator, the results will be tabulated by the machine as shown in Figure 6. This is a condensed tabulation which shows time intervals 1 to 5 and some of the later time intervals. For the later time intervals, only the data for the top

Time Interval "n"	Subscript "m"	Forces "F" pounds	Forces "R" pounds	Velocities "V" fps	Displacements "D" inches
1	1	0	0	14.2500	0.04275
2	1	273,600	0	13.9566	0.08462
	2	0	0	1.4992	0.00450
3	1	512,783	0	13.4067	0.12484
	2	32,409	0	4.1314	0.01689
	3	0	0	.3747	0.00112
4	1	690,869	0	12.6658	0.16284
	2	113,621	0	7.2944	0.03877
	3	8,100	0	1.5946	0.00591
	4	0	0	.0937	0.00028
5	1	794,001	0	11.8144	0.19828
	2	236,839	0	10.3473	0.06982
	3	40,547	0	3.8638	0.01750
	4	2,024	0	.5390	0.00190
	5	0	0	.0234	0.00007
20	1	238,945	0	5.7933	0.54314
	2	223,200	0	5.2403	0.50414
	13	522	0	0.0199	0.00009
	14	0	522	0	0
28	1	166,719	0	3.9070	0.65498
	2	221,163	0	4.6805	0.63125
	13	366,135	0	3.5019	0.02919
	14	0	366,135	0	0
30	1	139,864	0	3.5942	0.6770
	2	195,695	0	4.0635	0.6566
	13	834,305	0	5.1612	0.0581
	14	0	600,000	1.2734	0.0038
46	1	465,569	0	0.3514	0.78710
	2	539,162	0	3.7305	0.70211
	13	424,886	0	3.8138	0.26202
	14	0	600,000	0.0933	0.22918
56	1	182,869	0	−5.0749	0.69620
	2	365,779	0	−0.6060	0.68104
	13	334,104	0	0.4408	0.30442
	14	0	600,000	−0.9241	0.28328
57	1	97,066	0	−5.1790	0.68067
	2	280,059	0	−1.6087	0.67621
	13	414,362	0	−0.2967	0.30353
	14	0	−100,000	1.8713	0.28890
58	1	28,516	0	−5.2096	0.66504
	2	157,825	0	−2.3172	0.66926
	13	286,881	0	−0.5238	0.30196
	14	0	600,000	0.1696	0.28940
59	1	−27,017	0	−7.4977	0.64254
	2	15,191	0	−2.5485	0.66161
	13	246,113	0	−0.4272	0.30068
	14	0	600,000	−1.7537	0.28414

Figure 6

and bottom of the pile are included, because these are the controlling factors in the calculation.

The calculation begins in the first time interval with a ram velocity of 14.25 fps and a displacement of 0.04275 inches. The stress wave travels down the pile, bringing each successive section of the pile into action. For example, in the fourth time interval, the velocity V_3 amounts to only 1.5946 fps with a displacement of D_3 of only 0.00591 inches, and a total force F_3 of only 8100 pounds.

It will be observed that, to represent the specified ground resistance, R_{14} is given a value equal to F_{13} until F_{13} exceeds 600,000 pounds, as it does first at time interval 30. From this point on, R_{14} remains at 600,000 pounds, except that when V_{14} becomes negative, as it does in time interval 56, R_{14} is given a value of $-100{,}000$ pounds in the next time interval, which, in this case, is interval 57. In interval 57, the velocity V_{14} is again positive; therefore, R_{14} is again given a value of 600,000 pounds in interval 58, and so on. In the meantime, the values of the displacements D_{14} have increased gradually to a maximum of 0.28940 in interval 58, after which they remain practically unchanged. It will also be observed that the capblock force F_1 becomes negative in time interval 59. This means that the hammer ram has separated from the capblock; therefore, at this point, the hammer ram is dropped from the calculation.

Figure 7 is a graphic representation of the results of the calculation. A separate curve is plotted for each weight W,

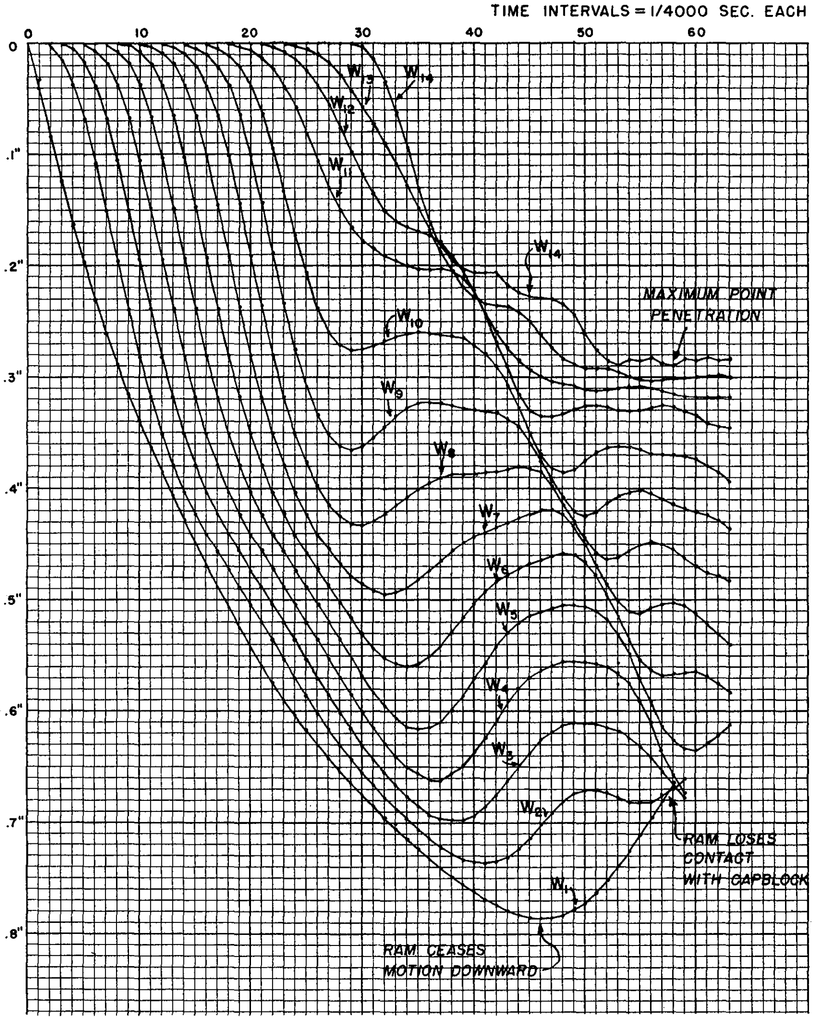


FIGURE 7. DISPLACEMENTS

and the curves represent the relationship between displacement and time. It can be seen from the curve for W_1 that the hammer ram curve crosses the pile head curve in the 58th time interval. On the curve for W_{14}, it can be seen that the penetration reaches a maximum in the 53rd time interval and then fluctuates slightly in the succeeding time intervals, reaching practically the same maximum again in interval 58. The calculation should always be carried beyond the first maximum of penetration in order to make sure that only slight fluctuations in penetration will occur thereafter.

Checking the Calculations

The total energy of the system for any particular time interval can be obtained by adding the kinetic energies of the individual weights, the potential energies of the individual springs, and the total work performed in overcoming the various external forces. The total should equal the energy of the ram just before impact.

The total momentum of the system for any particular time interval can be obtained by adding the products of each mass multiplied by its instantaneous velocity and the products of each external force multiplied by the total time it has acted. The total should equal the momentum of the ram just before impact.

Neither check is exact because of minor inaccuracies in this method, but a sudden variation between one time interval and the next indicates a numerical error. If the total varies by more than about 5%, consideration should be given to reducing the time interval and possibly the lengths L. If the work is done by hand, it is recommended that checks be made at every 10th interval. If the work is done by an automatic calculator, it may be possible to include a running check as part of the setup. The energy check is to be preferred to the momentum check as it is more complete. Plotting the calculated results for displacements D as per Figure 7 is also an excellent check on the reasonableness of the results. Curves for separate calculations may be readily compared.

Recalculation for Change in Ground Resistance

A change in the resistance near the point of the pile will change only a half or a third of the total calculation. Piles may, therefore, be recalculated for various point resistances with a considerable saving of effort as compared with an entirely new calculation.

DISCUSSION

Mr. Sheldon: I would like to make a few comments. I have been much impressed with Mr. Smith's handling and understanding of the phenomena which go into these piles. He has not employed calculus, but he has really derived for you the partial differential equation for the motion of the pile with boundary conditions.

The equation for the displacement $D(x,t)$ in a one-dimensional, inhomogeneous elastic medium is:

$$\rho(x)\frac{\partial^2 D}{\partial t^2} = \frac{\partial}{\partial x}\left\{Y(x)\frac{\partial D}{\partial x}\right\} + f(x,t)$$

where $\rho(x)$ is the mass density, $Y(x)$ is Young's modulus, and $f(x,t)$ is the applied stress. The simplest difference system by which we can replace the above differential equation is, in the notation of Mr. Smith,

$$\rho_m \frac{D_m^{n+1} - 2D_m^n + D_m^{n-1}}{(\Delta t)^2} = \frac{Y_{m+1}[D_{m+1}^n - D_m^n] - Y_m[D_m^n - D_{m-1}^n]}{(\Delta x)^2} + f_m^n .$$

The solution of this difference system is exactly equivalent to the solution of the difference system derived by Mr. Smith from first principals. To see this, note that

$$K_m = \frac{Y_m}{\Delta x}, F_m^n = K_m[D_m^n - D_{m-1}^n], R_m^n = f_m^n \Delta x ,$$
$$\frac{W}{32.17} = \rho_m \Delta x \quad \text{and} \quad V_m^n = \frac{D_m^n - D_m^{n-1}}{12\Delta t} .$$

Mr. Smith chose an interval $\Delta t = \Delta x/2c$ (c = sound speed), so that he had a safety factor of 2 in the Courant condition for the stability of numerical integration of hyperbolic type equations.

We solved this problem on the card-programmed calculator at the technical computing bureau. We chose the CPC for solution because Mr. Smith had quite complicated boundary conditions imposed. For one thing, the resistance of the ground is a non-linear function. It is 600,000 pounds upwards when the last weight is moving down and 100,000 pounds downward when the last weight is moving up. Another condition that has to be provided for is that the capblock and follower are not attached to the pile itself, so that after a certain period of time the capblock and follower fly off the top of the pile. Using the card-programmed calculator with its facility to list answers as we go along, we were able to observe the sign of the velocity at the bottom of the pile and also whether there was a tension or compression in the capblock. As soon as the condition which bounded this motion changed, we were able to insert a new instruction card which would take care of the new condition. This is a much simpler procedure than attempting to put these conditions into the control panel of a machine. We have solved, totally, eight cases of this pile-driving work, and for the last six cases there was an additional complication in the auxiliary conditions. Mr. Smith decided to take account of the fact that the capblock was made of wood and, therefore, was not perfectly elastic. We changed the elastic constant of the capblock according to whether the capblock was being compressed or was expanding.

The problem runs at about one step in the time every three minutes. This is an average figure, taking account of changing the program cards to take care of auxiliary conditions.

Dr. Aronofsky: I do not understand the boundary conditions at the top. The weight of the ram is 7,500 pounds. Is there any condition imposed?

Mr. Sheldon: The weight at the top is a freely falling weight; so the boundary condition at the upper end of the pile is that at $t = 0$ the ram has a certain definite velocity, and all the other weights are not moving.

Dr. Aronofsky: Is there any assumption about resistance along the lateral side of the pile all the way down?

Mr. Sheldon: In one case there was just the resistance at the bottom. In another case the resistance was applied in the middle.

Mr. Smith: In Figure 4 the only resistance that is inserted is the last one. All the other resistances are zero. However, if desired, you can put in as many resistances as there are weights.

Dr. Buchholz: I think such studies have been made on analog equipment. You replace this type of system by a network of little capacities and provide certain nonlinear elements to take care of boundary conditions and special conditions. You run into a bit of a problem in the case of the capblock leaving the rest of the system. I don't know whether this problem has been done, but I imagine it might be possible to do so.

Mr. Moncreiff: Was special wiring used for this problem?

Mr. Sheldon: No effort was made to change the standard setup at all. We made it very simple so that it took about one day to plan for the machine.

Mr. Moncreiff: The calculation is simple enough so that you could save time by wiring a special control panel.

Mr. Sheldon: That is true.

Punched Card Mathematical Tables on Standard IBM Equipment

ELEANOR KRAWITZ

International Business Machines Corporation

THE EFFICIENCY with which punched card tables can be constructed and used is so powerful that each problem should be examined for the extent to which tables can be used. Examples of tables commonly used in this manner are trigonometric and logarithmic tables. The cards are an ideal medium for mathematical tables; it is possible to look up millions of digits in a single day. In fact, the efficiency of the method increases as the number of cards increases.

An ordinary printed table has a tabular interval, arguments, function values, and it may have interpolating aids such as differences. The simplest table is a critical or turning point table in which the arguments are so chosen that the value of the tabular function changes by one unit. For example, consider a section of the sine table:

TABLE I

x	Sine x
2°.551	
	0.045
2°.608	
	0.046
2°.666	
	0.047
2°.723	
	0.048
2°.780	

For values of the argument between 2°.608 and 2°.666, the value of the sine is 0.046. If the required value of the argument may be one of the printed values, then the author must state which of the two adjacent function values corresponds to that argument. This is usually stated near the printed table. In this example, the lower of the two adjacent values is taken, that is, sin 2°.608 is 0.046. This critical table of the sine consists of 1001 cards for the first quadrant, and the error is 0.5 in the third decimal.

In most cases the number of necessary entries in a critical table is prohibitive; we may shorten the table by making use of the differences of the function. Consider the following small section of the sine table:

TABLE II

x_i	Sine x_i	δ'_i
2°.0	0.035	
		+17
3°.0	0.052	
		+18
4°.0	0.070	
		+17
5°.0	0.087	

To compute the value of sin 3°.38 using linear interpolation, we have

$$f_u = f_i + u\,\delta_i \text{ where } u = \frac{x - x_i}{x_{i+1} - x_i}\,,$$

$$\sin 3^\circ.38 = 0.052 + \frac{0.38}{1}\,(0.018) = 0.059\,.$$

The required argument serves two purposes in a table involving interpolation; it is used to enter the table, and it is used in the interpolating process. This sine table involves a linear interpolation, consists of 91 cards, and has the same accuracy as that of the critical table.

The number of entries in the table may be shortened still further by the introduction of higher order differences (Table III).

TABLE III

x_i	sin x_i	$\delta'_{i+1/2}$	δ''_i	$(\delta'_{i-1/2}+\delta'_{i+1/2})/2$	$\delta''_i/2$
0°	0.0000		0	1736	0
		+1736			
10°	0.1736		− 52	1710	− 26
		1684			
20°	0.3420		−104	1632	− 52
		1580			
30°	0.5000		−152	1504	− 76
		1428			
40°	0.6428		−196	1330	− 98
		1232			

Table III (*Continued*)

x_i	$\sin x_i$	$\delta'_{i+1/2}$	δ''_i	$(\delta'_{i-1/2}+\delta'_{i+1/2})/2$	$\delta''_i/2$
50°	0.7660		−232	1116	−116
		1000			
60°	0.8660		−263	868	−132
		737			
70°	0.9397		−286	594	−143
		451			
80°	0.9848		−299	302	−150
		152			
90°	1.0000		−304	0	−152

Using Stirling's interpolation formula, we have:

$$f_u = f_i + u\left[\frac{\delta'_{i-1/2}+\delta'_{i+1/2}}{2} + u\frac{\delta''_i}{2}\right].$$

To compute sin 37°.689,
sin 37°.689 = 0.5000

$$+ .7689\left[\frac{.1580+.1428}{2} + .7689\left(-\frac{.0152}{2}\right)\right] = .6111\,.$$

The necessary interpolation procedure involves three additions, two multiplications, and two divisions. In many problems, however, it is more efficient to tabulate quantities other than the differences themselves. If the quantities $(\delta'_{i-1/2}+\delta'_{i+1/2})/2$ and $\delta''_i/2$ are tabulated instead of the straight differences, the interpolation operation involves only two additions and two multiplications. This table contains ten entries, requires second-order interpolation, and the maximum error is 4 in the fourth decimal. A punched card table may be key punched directly from a printed table, it may be constructed with the use of machines, or it may be copied from someone else's punched card table. One value of the argument, its function value and interpolating aids are punched in a single card so that each card carries a line of the table. For example, a card for Table III would carry x_i, sin x_i, $(\delta'_{i-1/2}+\delta'_{i+1/2})/2$ and $\delta''_i/2$. A distinguishing X is punched in some column for identification purposes.

Assume that we are given a set of detail cards, each of which carries a value of the argument to be used in consulting a table. The arguments in the detail cards are presumably the result of a previous calculation. The first step in consulting a punched card table is to merge the detail cards with the table cards in such a manner that each table card is followed by all the detail cards whose arguments are greater than or equal to the argument in the table card but less than the argument in the following table card. This may be done on the sorter (if the arguments in the detail and table cards are in the same columns) or on the collator. If the sorter is used, the two sets of cards are arranged in ascending order of the argument, with the table cards entering the machine first. If the collator is available, and each set of cards is in ascending order of argument, then a simple run will accomplish the same purpose. The table cards are placed in the primary feed of the collator, and the detail cards are placed in the secondary feed. The argument of the first table card is compared with that of the first detail card for a high, low, or equal condition. If the arguments are equal, both cards are ejected simultaneously, with the card from the primary feed alighting first. If the arguments are not equal, the card with the lower argument is ejected. The control panel is wired to perform these operations.

In the case of the critical table it is only necessary to transfer the value of the function from a table card into the following detail cards with the use of the gang punch. The distinguishing X punch in the table card prevents punching from the last detail card of a group into the following table card. If the table is one in which interpolation is necessary, the cards may be passed through a calculating punch immediately after the merging operation. The values of the function and the differences for a table card are stored in the machine and used for the following detail cards. The next table card signals the machine to reset the previous data and replace it with new values.

Some features of the IBM Type 602-A Calculating Punch will be discussed, and a description of the control panel wiring for second-order interpolation will follow. There are 30 counter positions divided into six counter groups, and six storage registers of 12 positions each. All storage registers except the first one are divided into a right- and a left-hand group of six digits each. The first storage register is divided into an eight-digit right-hand group, which stores the multiplier and divisor, and a four-digit left-hand group. Two of these storage registers are capable of punching results. The counter groups are capable of adding and subtracting, and do not reset until they receive an impulse to do so. The storage units do not accumulate, and they automatically reset before a new number is accepted. There are 12 available program steps, which may be repeated or extended. In multiplication, the multiplier is read into the right-hand part of storage register 1 (IR), and the multiplicand is read out of any storage unit or counter into any other counter to accumulate the product.

The programming for the second-order interpolation, $f_u = f_i + u\,[(\delta'_{i-1/2}+\delta'_{i+1/2})/2 + u\,(\delta''_i/2)]$, is as follows (Figure 1):

Table cards: Read f_i (function) into storage unit 3R.
Read $(\delta'_{i-1/2}+\delta'_{i+1/2})/2$ (modified first difference) into storage unit 2L.
Read $\delta''_i/2$ (modified second difference) into storage unit 2R.

Detail cards: Read u into storage unit 1R.

CALCULATING PUNCH – TYPE 602A – CONTROL PANEL

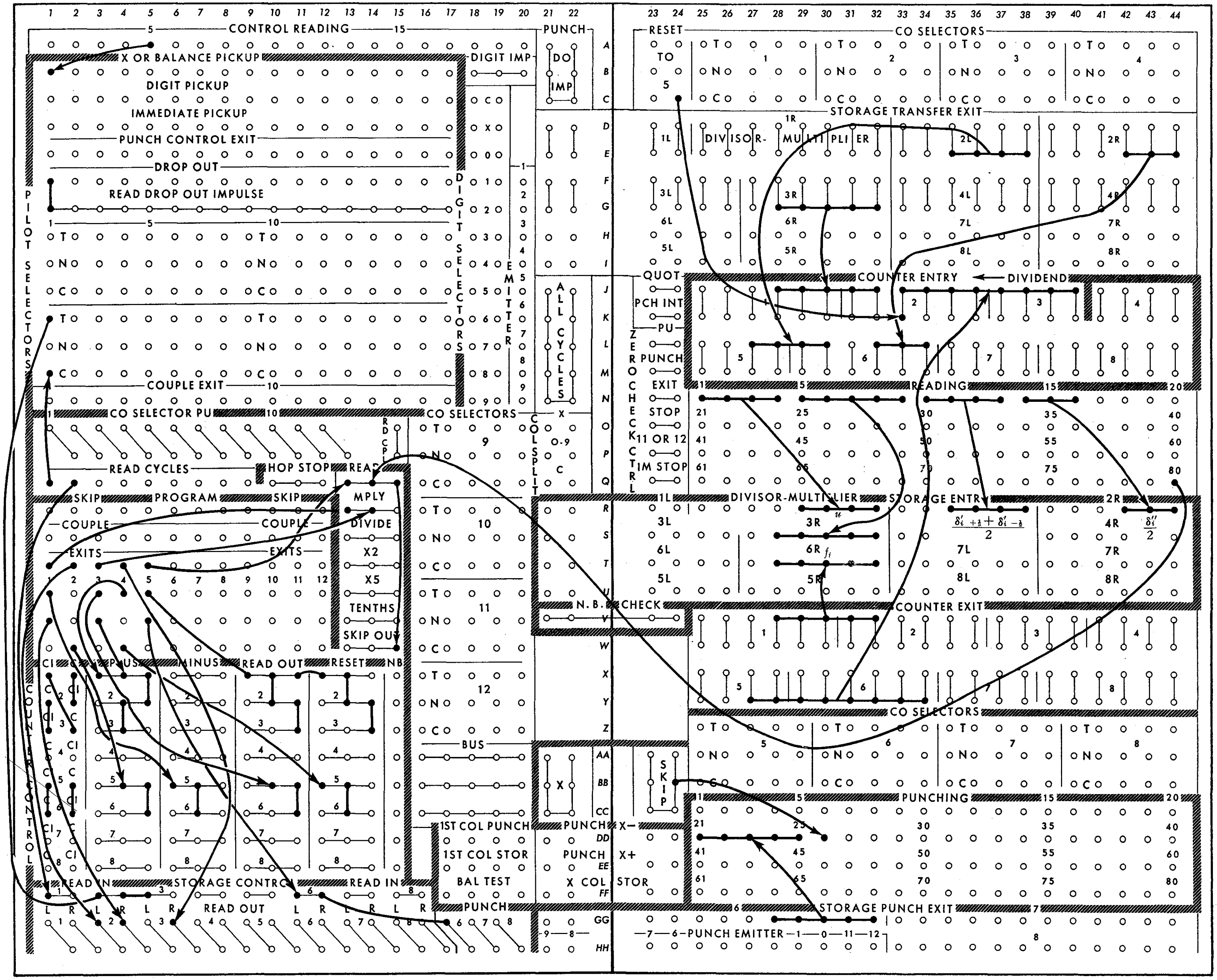


FIGURE 1

Program step 1: Multiply $u\ \delta''_i/2$: Read out $\delta''_i/2$ from storage unit 2R, and develop the negative product in counters 5, 6.

Program step 2: Add $(\delta'_{i-1/2}+\delta'_{i+1/2})/2$ to $u(\delta''_i)/2$: Read out $(\delta'_{i-1/2}-\delta'_{i+1/2})/2$ from storage unit 2L, and read positively into counters 5, 6.

Program step 3: Multiply $[(\delta'_{i-1/2}+\delta'_{i+1/2})/2+u\ (\delta''_i)/2]$ by u: Read out $[(\delta'_{i-1/2}+\delta'_{i+1/2})/2\ +\ u\ (\delta''_i)/2]$ from counters 5, 6 and develop the product in counters 1, 2, 3.

Program step 4: Add f_i to $u[(\delta'_{i-1/2}+\delta'_{i+1/2})/2+u\ (\delta''_i)/2]$ to obtain f_u: Read out f_i from storage unit 3R and read positively into counters 1, 2, 3. Reset counters 5, 6.

Program step 5: Punch the result f_u: Read f_u out of counters 1, 2, 3 into punching storage unit 6. Reset counters 1, 2, 3, reset to 5.

The design of a punched card table for a particular problem depends upon many factors. If the problem is of considerable magnitude, it is worthwhile to spend time and energy to prepare special tables. On the other hand, for a shorter problem a less refined table may be more efficient. In general, the selection and design of the punched card table depends upon the problem, the machine operators, and the machines available. Each table should be constructed so that it will be most economical.

Usually, it is necessary to trade table size for length of computation. Consider, for example, Newton's formula

$$f_u = f_i + u\,\Delta'_i + \frac{u(u-1)}{2}\,\Delta''_i + \ldots .$$

If a critical table is desired, the effect of the first differences must be negligible; therefore, all possible function values are needed. If only first differences are permitted, the effect of second differences must be negligible; however, a larger interval than that in the critical table is permitted, and fewer entries will be necessary in the table. The computation of the function now involves an addition and a multiplication. In the three tables we discussed previously, the critical table contained 1001 entries, and its accuracy was 0.5×10^{-3}; the table with linear interpolation contained 91 entries and had the same accuracy; the table with second-order interpolation contained 10 entries, and its accuracy was 0.4×10^{-3}. This process of the reduction of table size can be continued until one table entry remains, but the amount of necessary calculation is huge. In fact, in many cases this calculation may be equivalent to evaluating the function by its series expansion. In constructing a table, it should be kept in mind that the first differences vary directly with the interval, the second differences vary directly with the square of the interval, etc.

In most printed tables, the tabular interval is a unit in a given position of the argument, and it is usually constant throughout the table or varies by a power of 10. This choice of interval facilitates the interpolation process for the computer, and hence fewer errors result. In many problems it is desirable to use the largest tabular interval in which it is legitimate to interpolate with a given order of differences. A table constructed on this principle is called an *optimum interval table.* The punched card method conveniently handles varying intervals, and thus makes possible a considerable saving in the number of entries in the table. Moreover, the formula used for *n*th order interpolation in the optimum interval method is equivalent to that used for a constant interval table.

Consider a linear optimum interval table of the exponential function, $y = e^x$, in the range $x = 0$ to $x = 5$, where the intervals are so chosen that the error due to neglecting second differences does not exceed .05. The permissible intervals have been computed, and a portion of the table follows:

X_i	Interval	F	D
1.06			
	31	−.70	3.38
1.37			
	27		
1.64			
	.		
.			
	.		
.			
	.		
.			
	.		
4.85			
	5		
4.90			
	5		
4.95			

It is clear that a large interval may be chosen at the beginning of the table, but the interval must decrease sharply near the end of the table. To facilitate the interpolation operation, the formula for ordinary linear interpolation is transformed so that,

$$e^x = F + Dx \text{ where } D = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}\,,$$

$$\text{and } F = f(x_i) - x_i\,D\,.$$

For example:

$$e^{1.23} = -.70 + (1.23)(3.38) = 3.46\,.$$

The true value for $e^{1.23}$ is 3.42, which is within the limit of accuracy. This table consists of 40 cards; a table with uniform intervals of .01 contains 500 cards. It should be noted that the machine operations are the same as those necessary for ordinary linear interpolation.

Consider the problem of determining the sines of certain angles in the first quadrant with a given accuracy of 1×10^{-7}. The chart below gives the type of table, the order of interpolation, and the number of necessary cards.

Type of Table	*Order of Interpolation*	*Number of Cards*
critical table	first order	5,000,000
constant interval	first order	9,000
optimum interval	first order	1,500
optimum interval	second order	60
optimum interval	third order	15

It is apparent that for different problems different tables are more efficient. The third-order optimum interval table containing 15 cards looks attractive at first glance, but if it is to be used only a few times, it would not be economical to construct.

REFERENCES

1. W. J. ECKERT, *Punched Card Methods in Scientific Computation.* Revised edition in preparation.
2. H. R. J. GROSCH, "The Use of Optimum Interval Mathematical Tables," *Proceedings, Scientific Computation Forum, 1948* (IBM), pp. 23-27.

DISCUSSION

Dr. Sherman: In the problem of inverse interpolation, can optimum interval tables be used?

Miss Krawitz: It is not possible to use the optimum interval table directly for inverse interpolation. The values of *F* and *D* must be modified.

Mr. Opler: There is an article in the July *MTAC* on maximum interval tables. They appeared to be differentiating them from optimum interval tables.

Dr. Grosch: Stadler's article, which was the first sent to this country, is just an example of the confusion into which we are always getting. He prefers the term maximum interval, because if you apply the methods which he worked out independently of me for determining intervals, the calculation of the maximum allowable interval is obtained, but when Clemmons and Herget coined the phrase *optimum interval,* they implied a little more than choosing the largest possible interval. They implied the possibility of, having found the maximum interval, abandoning it in favor of a more suitable interval, if something is to be gained by it. In the list of tables of the Watson Laboratory there is a third-order uniform interval table of sines and cosines which is accurate to about 8½ decimal places. That is a uniform interval table of 25 lines per quadrant, because both sine and cosine were present in the same line of the table, and the allowable maximum interval governed, for half of the quadrant, the maximum function allowable for the cosine function. Actually, we would have only reduced the table to about 23 lines per quadrant if we had taken the maximum interval table.

We gained all sorts of advantages in implementing the table on the mechanical computing aids available. I think we ought to say that an optimum interval table is a table best adapted to the facilities at your disposal. That is, perhaps, the initial stage that you go through in arriving at an optimum table. We have many tables at the Watson Laboratory which started out to be extremely elaborate ones and finished in uniform interval form because we could use *f* instead of *F* and make the table more easily recognizable and usable.

As I see it, all phases of the problem, choosing the proper interpolation, forming the proper intervals of the table, and so forth, must be considered in order to make the table as economical as possible to use for the particular problem you are trying to solve.

Lt. Hastings: I would like to ask for some general comments on the extent to which it is no longer necessary to have tables of simple functions like sine and cosine, because of the program repeat on the CPC.

Dr. Grosch: There are certain mathematical functions that are very well adapted to the idea of direct mathematical handling. I think an example is the sine, cosine, where you have a very simple formula which converges very rapidly, for which only every other power of the power series involved is required, and in which all of the coefficients are simple rational numbers. Obviously, you can use the program repeat device and evaluate the sine function very nicely without reference to any of the tables discussed. The critical idea, in my mind, is whether or not you are able to step from value to value. If your problem is—such as it might very well be in a trajectory, for instance—that each table lookup is the value of the argument which is pretty near the preceding one, then you add the important advantage of being able to use, what I might think of as your past value of the X function, for instance, and obtain the $X + \Delta X$, using ΔX as a variable. In some cases you must be very careful about accumulation of errors. On the other hand if you have random lookups, evaluation of each function, without the aid of a table, is more advantageous.

Dr. Hurd: I think what prompted Lt. Hastings' question is the fact that he has made a survey of computing facilities around the country. He noticed in many places where a CPC is available that such functions as trigonometric functions and square root, etc., were being calculated directly on the calculator as these functions were needed.

The Solution of Simultaneous Linear Equations Using the IBM Card-Programmed Electronic Calculator

JUSTUS CHANCELLOR   JOHN W. SHELDON   G. LISTON TATUM

International Business Machines Corporation

SIMULTANEOUS linear algebraic equations up to and including order 21 may be readily solved on the IBM Card-Programmed Electronic Calculator by using a basic approach sometimes referred to as the Crout method.[1] A slight modification has been incorporated in the procedure in completing the back solution. Having obtained the auxiliary matrix at the end of the forward solution, the operator rearranges the elements in such a way that almost the same reduction technique may be used for the back solution.

The method involves the augmentation of the matrix of coefficients of the unknowns. To this original matrix is added one of the following:

a. One column of constants, if a single solution is desired; or

b. Several columns of constants, for several solutions; or

c. A unit matrix, if the inverse is desired.

Regardless of which end result is needed, the reduction calculations are the same.

By using the CPC, the solution is obtained much more rapidly and with considerably less card handling and processing than has been possible heretofore on IBM calculating punches.

Renumbering Elements of Original Matrix

To take advantage of the internal storage numbering and to facilitate coding instructions, the familiar notation for each element of the matrix has been altered. Instead of using 1, 2, 3, ... to designate row i or column j location, one numbers the elements after the code numbers of storage units and counters in the CPC.

It takes six hours to invert a 20 × 20 matrix on the 604, sorter, reproducer and accounting machine, and it takes two hours to perform the same inversion on the CPC.

The storage and counter designations are as follows:

Storage bank 1:	11, 12, 13, 14, 15, 16, 17, 18
Storage bank 2:	21, 22, 23, 24, 25, 26, 27, 28
Counter groups:	—, 2, 3, 4, 5, 6, 7

The code digit 7 before any counter group number is, here, an instruction code meaning *add into*. Thus, it is advisable to assign row and column numbers as follows:

Usual notation:	1	2	3	4	5	6	7	8	9	10	11
	12	13	14	15	16	17	18	19	20	21	
New notation:	11	12	13	14	15	16	17	18	21	22	23
	24	25	26	27	28	72	73	74	75	76	

Counter group 7 is reserved for the accumulation of the check sum. Counter group 1 is not used to store element values. Thus, it is possible to store 21 elements and provide for the accumulation of a check column. Each counter-group and storage location has the capacity for handling a ten-decimal digit number and its associated sign.

Augmenting the Original Matrix

Before the reduction calculations are made, the original matrix, consisting of the coefficients of the unknowns, must be augmented. The augmentation is always made by adding columns of elements to the right of the original matrix.

To facilitate identification, the *composite matrix,* including the original matrix, the augmented portion, and the check column, is divided into sections called subdivisions. The original matrix is subdivision 1, the augmented portion is subdivision 2, and the check column, each element of which consists of the sum of the elements in the corresponding row, is subdivision 3. The augmented portion, subdivision 2, will vary, depending on whether one or several solutions are desired, or whether the inverse matrix is to be obtained. In this paper the case of the inverse will be treated. The other cases would be handled as described below, but the augmented portion would consist of as many columns of constants as the number of solutions required—instead of the unit matrix.

Using the new notation and augmenting the original matrix so as to prepare for calculating the inverse, one has the twenty-first order matrix shown in Figure 1, page 58.

$$\begin{array}{lllll|llll|l}
\multicolumn{5}{c|}{\text{Original Matrix}} & \multicolumn{4}{c|}{\text{Unit Matrix}} & \text{Check Column}\\
\text{Subdivision:} & & 1 & & & & 2 & & & 3\\
a_{11,11} & a_{11,12} & a_{11,13}\ \ldots & a_{11,75} & a_{11,76} & 1_{11,11} & 0_{11,12}\ \ldots & 0_{11,75} & 0_{11,76} & c_{11,11}\\
a_{12,11} & a_{12,12} & a_{12,13}\ \ldots & a_{12,75} & a_{12,76} & 0_{12,11} & 1_{12,12}\ \ldots & 0_{12,75} & 0_{12,76} & c_{12,11}\\
\cdot & & & & & & & & & \\
\cdot & & & & & & & & & \\
\cdot & & & & & & & & & \\
a_{75,11} & a_{75,12} & a_{75,13}\ \ldots & a_{75,75} & a_{75,76} & 0_{75,11} & 0_{75,12}\ \ldots & 1_{75,75} & 0_{75,76} & c_{75,11}\\
a_{76,11} & a_{76,12} & a_{76,13}\ \ldots & a_{76,75} & a_{76,76} & 0_{76,11} & 0_{76,12}\ \ldots & 0_{76,75} & 1_{76,76} & c_{76,11}
\end{array}$$

FIGURE 1

1. Take the corresponding element $a_{i,j}$ in the composite matrix.

2. From it subtract the products resulting from pairing the elements of the *auxiliary* matrix in the row to the left and in the column above the desired element, taking the products in order, i.e., first in row by first in column, second in row by second in column, etc. (This rule will be illustrated later.)

3. If the desired element is on or below the principal diagonal, record the result obtained in 2. If the desired element is to the right of the diagonal, divide the result obtained in 2 by the diagonal element in the same row of the auxiliary matrix.

The reduction calculations are applied to each element of one row of the original composite matrix to obtain the corresponding row of the auxiliary matrix, completing the solution for that row for the forward solution. Calculations begin on the first row and continue on succeeding rows until each row of the original matrix has been reduced to form the auxiliary matrix.

After one has developed a portion of the auxiliary matrix, it will appear as in Figure 2.
The zero elements to the right of the diagonal in the unit matrix need not be introduced until the back solution is started.

Reduction Calculations

To obtain the inverse matrix (or specific solutions) each element of the original composite matrix is reduced by a series of calculations which may be divided into two major portions. The first portion consists of the reduction calculations applied to each element of the original composite matrix. These reduce it to a matrix known as the *auxiliary matrix*. This first reduction is called the *forward solution*. The second portion of computations applied to each element of the auxiliary matrix (whose elements have been shifted as described later) reduce it to the final result which is the inverse matrix. This second reduction corresponds to the *back solution* of the Crout method.

Regardless of whether the reduction calculations are for the forward solution or the back solution they follow the same basic rules. These rules for calculating the value of a reduced element $b_{i,j}$ are:

As an example of how to apply the rules for calculating the value of a reduced element $b_{i,j}$ of the auxiliary matrix, consider the computation

$$b_{14,15} = \frac{a_{14,15} + (-b_{14,11}\,b_{11,15} - b_{14,12}\,b_{12,15} - b_{14,13}\,b_{13,15})}{b_{14,14}}\,.$$

During the calculation of the reduced elements $b_{i,j}$ in any one row of the auxiliary matrix, the elements to the left and on the principal diagonal are stored, and all of the elements to the right of the principal diagonal are summary punched as they are calculated. Each of these summary punched cards becomes the source of one of the pair of factors involved in negative multiplication. At the completion of the calculation for one row, the summary punched cards for that row are merged in with cards summary

$$\begin{array}{llllll|lll|l}
\text{Subdivision:} & & 1 & & & & & 2 & & 3\\
b_{11,11} & b_{11,12} & b_{11,13} & b_{11,14} & b_{11,15}\ \ldots & b_{11,76} & b_{11,11} & & & d_{11,11}\\
b_{12,11} & b_{12,12} & b_{12,13} & b_{12,14} & b_{12,15}\ \ldots & b_{12,76} & b_{12,11} & b_{12,12} & & d_{12,11}\\
b_{13,11} & b_{13,12} & b_{13,13} & b_{13,14} & b_{13,15}\ \ldots & b_{13,76} & b_{13,11} & b_{13,12} & b_{13,13} & d_{13,11}\\
b_{14,11} & b_{14,12} & b_{14,13} & b_{14,14} & & & & & & \\
 & & \multicolumn{3}{l}{\text{Principal diagonal}} & & & & &
\end{array}$$

FIGURE 2

$$
\begin{array}{lcccc|ccccc|c}
\text{Subdivision:} & & 1 & & & & & 2 & & & 3 \\
b_{11,12} & b_{11,13} & b_{11,14} \dots & b_{11,75} & b_{11,76} & b_{11,11} & & & & & d_{11,11} \\
 & b_{12,13} & b_{12,14} \dots & b_{12,75} & b_{12,76} & b_{12,11} & b_{12,12} & & & & d_{12,11} \\
 & & & \cdot & & & & & & & \cdot \\
 & & & \cdot & & & & & & & \cdot \\
 & & & \cdot & & & & & & & \cdot \\
 & & & b_{74,75} & b_{74,76} & b_{74,11} & b_{74,12} \dots & b_{74,74} & & & d_{74,11} \\
 & & & & b_{75,75} & b_{75,11} & b_{75,12} \dots & b_{75,74} & b_{75,75} & & d_{75,11} \\
 & & & & & b_{76,11} & b_{76,12} \dots & b_{76,74} & b_{76,75} & b_{76,76} & d_{76,11}
\end{array}
$$

FIGURE 3

punched from previous rows. These combined cards are arranged in groups to form the columns of elements above the particular row elements to be reduced by the next calculation. The original elements $a_{i,j}$ are also in this deck.

As the calculations proceed across a row, each result to the left of and on the principal diagonal is stored internally in the storage location corresponding to the identification of the column of the original element involved. In the example above, the $b_{14,15}$ element will be stored in the column designation of $a_{14,15}$ which is storage 15. Similarly, code 76 means *add into* counter group 6.

While the summary punched cards in the column above the particular element being calculated feed through the IBM Type 402 (or 417) Accounting Machine, each card is paired with its proper other factor for negative multiplication. This proper pairing is accomplished by having that storage location designated by the column number read-out to combine with the summary punched card containing the same row number. That is, the summary punched card containing one factor $b_{k,j}$ and row designation k calls out the second factor $b_{i,k}$ from the internal storage of the same column number k.

As the original element card feeds through the type 402 accounting machine, its value is added.

For any one row the divisor is always the same and is available once the principal diagonal element has been calculated. The *storage location* of the divisor is remembered for the entire row and is used to complete division at the proper time.

The same reduction operation applies throughout the entire composite matrix, including the unit matrix and check column. At the end of the forward solution the auxiliary matrix will be as shown in Figure 3.

To identify correctly, by row and column, results that will appear in the spaces occupied by the zero elements in subdivision 2, the following cards must be added:

$$
\begin{array}{cccc}
0_{11,12} & 0_{11,13} \dots & 0_{11,75} & 0_{11,76} \\
 & 0_{12,13} \dots & 0_{12,75} & 0_{12,76} \\
 & & \cdot & \cdot \\
 & & \cdot & \cdot \\
 & & \cdot & \cdot \\
 & & 0_{74,75} & 0_{74,76} \\
 & & & 0_{75,76}
\end{array}
$$

Back Solution

The Crout method back solution is modified so that the same calculating rules may be used as were employed in the forward solution, except that no division is involved at any time. To do this it is necessary to shift the arrangement of the elements in the auxiliary matrix before starting the reduction calculations.

This necessary rearrangement of elements is rapidly carried out on the sorting machine. The necessary shifts are:

1. Invert the order of the rows in each column of each subdivision, at the same time combining the zero elements with subdivision 2.
2. Invert the order of the columns in each row of subdivision 1 and 2.

Thus, the rearranged auxiliary matrix is as shown in Figure 4.

$$
\begin{array}{lcccc|cccccc|c}
\text{Subdivision:} & & 1 & & & & & & 2 & & & 3 \\
b_{75,76} & & & & & b_{76,76} & b_{76,75} & \dots & b_{76,13} & b_{76,12} & b_{76,11} & d_{76,11} \\
b_{74,76} & b_{74,75} & & & & 0_{75,73} & b_{75,75} & \dots & b_{75,13} & b_{75,12} & b_{75,11} & d_{75,11} \\
\cdot & & & & & \cdot & \cdot & & b_{74,13} & b_{74,12} & b_{74,11} & d_{74,11} \\
\cdot & & & & & \cdot & \cdot & & & \cdot & \cdot & \cdot \\
\cdot & & & & & \cdot & \cdot & & & \cdot & \cdot & \cdot \\
b_{12,76} & b_{12,75} \dots & \mathrm{b}_{12,14} & \mathrm{b}_{12,13} & & \cdot & \cdot & & & b_{12,12} & b_{12,11} & d_{12,11} \\
b_{11,76} & b_{11,75} \dots & b_{11,14} & b_{11,13} & b_{11,12} & 0_{11,76} & 0_{11,75} & 0_{11,74} & \dots & 0_{11,12} & b_{11,11} & d_{11,11}
\end{array}
$$

FIGURE 4

Repeating the calculations except for division, one obtains the rearranged auxiliary matrix reduced to its final form. Subdivisions 2 and 3 are of interest here:

$$\begin{array}{l|ccccc|c}
\text{Subdivision:} & & & 2 & & & 3 \\
 & e_{76,76} & e_{76,75} & \cdots & e_{76,12} & e_{76,11} & f_{76,11} \\
 & e_{75,76} & e_{75,75} & \cdots & e_{75,12} & e_{75,11} & f_{75,11} \\
 & \cdot & \cdot & & \cdot & \cdot & \cdot \\
 & \cdot & \cdot & & \cdot & \cdot & \cdot \\
 & \cdot & \cdot & & \cdot & \cdot & \cdot \\
 & e_{12,76} & e_{12,75} & \cdots & e_{12,12} & e_{12,11} & f_{12,11} \\
 & e_{11,76} & e_{11,75} & \cdots & e_{11,12} & e_{11,11} & f_{11,11}
\end{array}$$

Subdivision 2 is now the inverse matrix. The inverse matrix may, of course, be used in matrix by vector multiplication with any column of constants to obtain the specific solution corresponding to those particular constants. If one, or several, columns of constants had been used in the augmented portion of the matrix, the specific solutions would appear in as many columns of subdivision 2 above in place of the inverse matrix.

The idea of using the Crout method on the CPC was originated by Mr. William W. Woodbury, Northrop Aviation Company.

Details of planning charts and wiring diagrams may be obtained from the Applied Science Department, International Business Machines Corporation, 590 Madison Avenue, New York 22, New York.

REFERENCE

1. WILLIAM EDMUND MILNE, *Numerical Calculus* (Princeton University Press, 1949), pp. 17-25.

DISCUSSION

Mr. Elkins: Is this method the most efficient method now known for inverting a matrix?

Dr. Hurd: This is the most efficient method that we have so far discovered for the card-programmed calculator. As far as we know, this minimizes the amount of summary punching which is necessary. On standard machines, using the combination of any one of the calculating punches, the sorter and the reproducer particularly, if there are a number of matrices of the same order to invert, you will also have very efficient operation.

Dr. Sherman: There are several things I might say. If the original matrix is symmetrical, which is true for the case of normal equations, then the auxiliary matrix can be calculated slightly more easily because the elements below the diagonal are related to the elements to the left and right. You can either save almost half of the calculations, if you have a case of a symmetric matrix, or you can carry out this procedure and have an additional check. There is one obvious difficulty which actually arises; that is, if any one diagonal element is very small or close to zero, there are difficulties with the later calculations. How do you handle that case?

Dr. Grosch: That is a hard problem. I inverted a 19 by 19 where I knew in advance I was going to have a lot of indeterminates. I used the 602-A punch and carried 24 × 24 = 24 multiplications throughout. I lost 16 figures, which were so many that I could not iterate by the simple iteration procedure. I must do it over again with 32 figures, one of these days. There are other methods of using what I obtained. The trouble is that we have all sorts of interesting ideas about how to handle the problem of troublesome matrices but usually we don't know they are troublesome until we have started the problem. My own idea is that the Goldstein estimates of the laws of accuracy in inverting large matrices need not bother us unless the matrices arise from something like operational research where you have no idea at all of how the matrix is going to behave.

The kind of matrices that arise in flutter work are not likely to give trouble unless, in the few cases such as those with which I was experimenting, you have been warned in advance by the physical situation. I was using this as a matrix for fitting polynomials to curves which did not fit well, and I knew that determination of the coefficients was going to be hard; so I was not surprised at my troubles.

All of our discussion has been matrix arithmetic so far and has centered on the triangulation theme. There will be another paper on the Seidel or relaxation procedure. There is one other process worth considering on the card-programmed calculator. If you have a symmetrical matrix with which to deal and want an inverse in a simple manner, you may use the square root method which is now known under the name of Choleski. This is a process of finding a triangular matrix which, when multiplied by its transpose, equals the original given matrix. Having a triangular matrix you can find its inverse by operations similar to the back solution in the Gaussian elimination scheme.

This may be well worth considering, but the job of finding the triangular matrix is hard on any kind of standard equipment. With the card-programmed calculator, a pack of eight or ten cards is needed for a second-order matrix, then with a few more added for a third-order matrix and fourth-order, etc., up to several drawers full for 20th order. You would invert your matrix by loading the whole thing into the card-programmed calculator, and summary punch the results without any wasted time summary punching. Then, the cards just summary punched would be placed in the machine for the back solution, which is very simple.

Mr. Opler: I would like to make a general remark also about matrix inversion. Because the usual elimination procedures have so many different steps, I have spent considerable time investigating methods of matrix inversion other than the Gaussian elimination or row-by-row method. There is no doubt that this method which you have demonstrated is the only method that will work for virtually all cases of matrices. However, at times you have a borderline case which may be troublesome if you are not careful. There are a number of methods that appeal to me because I have done my calculating through the accounting department personnel, and I find that they have difficulty in a long procedure involving row-by-row elimination; so I prefer, in general, a method that will work by a simple procedure. There are several simple procedures for inverting matrices, but they do not apply to all matrices. One, which is without a doubt the simplest method and which I have carried out, is the Monte Carlo Method. The entire method consists in taking a pack of cards and putting them in an accounting machine. It's a very nice method, but unfortunately is only adapted to classes of matrices whose largest characteristic roots are less than 1.

Mr. Chancellor: I would like to second the motion of getting your operation as completely automatic as possible. That is the whole IBM goal in all of your applications—mathematical, accounting, or otherwise.

Lt. Hastings: I just wanted to remark that using the card-programmed calculator simultaneously for two systems, because of the alternation of the calculation and the collating steps, two systems could probably be done in very nearly the same time as one system.

Two Applications of the IBM Card-Programmed Electronic Calculator

IRVING C. LIGGETT

International Business Machines Corporation

The Gauss-Seidel Method of Solution of Simultaneous Linear Equations

THE GAUSS-SEIDEL method for the solution of simultaneous linear equations was first advanced by Gauss and Seidel about the turn of the 19th century in connection with the solution of normal equations. In addition to the solving of normal equations, this method may be used to solve various other systems if they meet certain requirements. We have used this method successfully, and the aircraft industry uses it very extensively, finding that it yields good results.

I would like to list the advantages of this method in an attempt to show you, as we go through the procedure, how these things are advantageous. We can obtain accurate answers to a large number of digits at a rapid rate. The problem is easy to set up for the machine. The machine runs continuously until an answer is obtained. In one 10 by 10 set of equations considered, results were obtained in thirty minutes, with the answers to seven decimal places. In about twenty minutes, the answers to five places were obtained. In less time than that, there were three or four places. The operator is in touch with the problem constantly and can tell how the problem is progressing immediately, since each new value of the variable and the correction factors are listed after they are calculated. The operations are completely repetitive, and almost any standard control panel can be used.

The mathematics is quite simple, and I would like to go through it briefly with you. Consider a set of three algebraic equations, bearing in mind that this method is general for any number of equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 - b_1 &= 0\,, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 - b_2 &= 0\,, \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 - b_3 &= 0\,. \end{aligned}$$

Each of the elements, $a_{i,j}$, is on a separate card. If the values of the x's are known, each of these equations will yield zero, to some accuracy. Assuming these values are unknown, let a first guess be made for each unknown x. Now the equations will not yield the answer zero, but something different from zero, say ϵ_i, and the equations will be as follows:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 - b_1 = \epsilon_1\,, \qquad (1)$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 - b_2 = \epsilon_2\,, \qquad (2)$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 - b_3 = \epsilon_3\,. \qquad (3)$$

The general procedure is this:

1. Using the first guess for x_1, x_2, and x_3, solve for ϵ_1.
2. Determine a new x_1 that will cause this equation to yield zero from this formula

   $$x_{1\,\mathrm{new}} = x_{1\,\mathrm{old}} - \frac{\epsilon}{a_{ii}}\,,$$

   where ϵ = residual from an equation and a_{ii} is a main diagonal coefficient of that equation. For equation (1) a is a_{11} and for (2) a is a_{22}, etc.
3. The new x_1 is used in place of the old x_1 from now on until a better value is found on the next cycle.
4. Using the new x_1, the old x_2 and x_3, solve for ϵ_2 from equation (2).
5. Solve for the new x_2 to make the equation go to zero.

   $$x_{\mathrm{new}} = x_{\mathrm{old}} - \frac{\epsilon_2}{a_{22}}$$

6. Substitute the new x_2 for the old x_2.
7. Solve for ϵ_3 from equation (3).
8. Solve for the new x_3.
9. Substitute the new x_3 for the old x_3. This is the end of a cycle for a set of three equations.

The process is repeated using one new value each time an equation is solved. As this process is continued, the values will converge to the correct answers for all equations, and the residual values will approach zero. This method has been found to converge for nearly all algebraic equations derived from engineering and physical problems and normal equations.

Machine Procedure

Assume that a standard 604 calculator control panel is available that will do at least the following operations necessary for this problem: multiply, divide, subtract, and transfer from channel C to channel B.

The 402 accounting machine control panel is the standard control panel that reads factors from channels A and B and coding instructions from the card. It is wired to LIST from from an X in the OP field channel A,B,C and instructions.

The additional wiring consists of the following: all of the a_{ij} coefficients are in cards; as the main diagonal, a_{ii}, is used to multiply by the proper x, it also must be stored to be used in the calculations for the new x.

$$x_{\text{new}} = x_{\text{old}} - \frac{\epsilon}{a_{ii}}$$

Channel A is wired to the entry hubs of counter group 7. An identifying X-40 is punched in all cards containing main diagonal, a_{ii}, coefficients. This X is used to pick up pilot selector 11 from the upper brushes. The sign of coefficient a will pick up pilot selector 12. When the main diagonal coefficient is read from the lower brushes, it will be added or subtracted into counter group 7 by the following wiring:

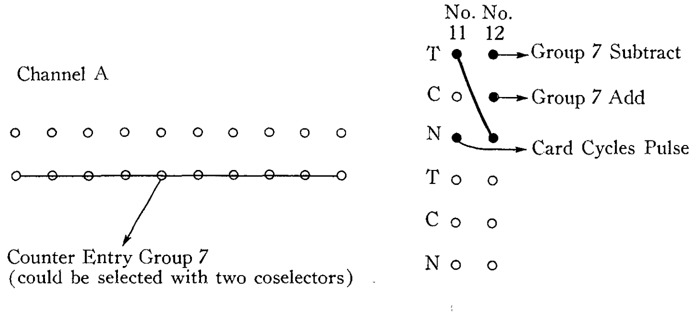


Figure 1

Channel A is also wired to the entry hubs of counter group 7.[a]

Coding of Cards

There are 22 ten-position storage registers in the CPC. Sixteen of these are registers, and six are counter groups. Let one counter group, say counter group 6, be used to accumulate the sum of products, $\Sigma ax - b = \epsilon$. Group 7 is used to store the main diagonal coefficient a_{ii}. That leaves 20 storage units for as many as 20 unknown x's. Each x is assigned to a storage unit and will be called for use from that unit at the proper time. When a new x is developed it will be sent to the storage unit to replace the proper old x value. x_1 and x_2 will be assigned to counter groups 2 and 3, respectively. All other values may be assigned in either counters or storage registers. This is done to avoid unnecessary blank cards. All of the coefficients, the values of the a_{ij}'s and the right-hand column, or b_j's, are key punched. There will be as many cards as there are a_{ij}'s and b_j's, one value on a card. Each card will also contain row and column identification of the coefficient as it appears in the matrix form, and instructions to the machine to perform the required operations.

Consider equation (1) of a 20 × 20 set of equations:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + \ldots + a_{120}x_{20} - b_1 = \epsilon_1 \, .$$

Card 1 contains a_{11} which is read on channel A. It contains instructions to read x_1 on channel B and multiply x_1 by a_{11}, adding the product in counter group 6.

Cards 2 through 20 contain similar instructions. Each card contains an a_{ij} coefficient and calls out the proper x, multiplies, and accumulates the answer in counter group 6.

Card 21 reads the right-hand side through the calculator to subtract in counter group 6.

Card 22 is a blank card necessary because the accumulated answer ϵ is wanted for the next operation, and it is not complete until the right-hand side b_j is subtracted.

Card 23 reads out ϵ which is in counter group 6 on channel A. It reads out the main diagonal coefficient, a_{ii}, which is in counter group 7, and divides ϵ/a_{ii}, and transfers the answer back to channel B. This answer will be listed from channel C.

Card 24 reads out the proper x_{old} on channel A and performs the operation

$$x_{\text{old}} - \frac{\epsilon}{a_{ii}} = x_{\text{new}} \, .$$

x_{new} is stored in place of x_{old} ready for use from then on. There are a total of 481 cards.

The cards for the next row initiate the same operations of multiplying and accumulating the answer in counter group 6. The main diagonal, a_{ii}, is stored automatically and used in card 23 of each row to calculate ϵ/a_{ii}. Card 24 calculates the new value of the x concerned and stores it in place of the old value to be used in the following operations.

This deck is fed through the machine as many times as necessary for desired convergence. For convenience, another deck may be 80-80 reproduced from the correct deck, thus giving 962 cards. Cards may be taken from the card stacker and replaced in the feed hopper, thereby keeping the machine in constant operation.

It would be very helpful if a good first guess could be entered in the machine. A good first guess would be the first digit of the actual answer. The machine would imme-

[a] With this added wiring, counter group 7 cannot be read out without resetting on channel A unless the wiring from channel A to the entry hubs is put through the transferred points of the coselectors coupled to pilot selector 11. This is unnecessary in this application as counter group 7 is always read out and reset. Counter group 7 cannot be read out on channel B at all; it will not read out properly.

diately show refinement of these answers. This in general is not necessary, fortunately, and the method usually works nicely with an initial guess of zero for all values. This is done by resetting all the storage registers and counters to zero in the beginning of the process.

One can tell that the process is functioning properly by watching the printed indications of the correction factors and the values of the x. As the process converges, the correction will become smaller and the x values will remain constant. In this process the machine may be run non-net balance to avoid conversion cycles. All answers are listed from channel C and are true figures with a sign.

If a master deck is available for the 20 × 20 set of equations, it will take approximately 2½ hours to key punch the a_{ij} and b_j values into the cards, prepare the deck by reproducing, and run the problem on the CPC. An example of the solution of a typical 20 × 20 after it was started on the CPC is as follows: A first guess of all values of $x = 0$ is made, with eight-decimal digits in original data and seven-decimal digits of accuracy required in the results. Twenty-seven iterations (times through the entire set of twenty equations) are completed in 90 minutes. Complete setup time, including punching of original data, should be less than one hour, using efficient methods.

Recommended Setup Method

A *master deck* for the solutions of a 20 × 20 set of equations is on file. It contains all necessary punching, except that the field in the cards set aside for the coefficients to be read on channel A will be blank. Assume that a set of 14 × 14 equations is to be solved:

The master deck is in order for operation on the CPC, that is, row, column. There are only 14 equations, so that the cards for row 15-20 will be taken out by hand and set aside. Sort the master deck in column order. This will put the card numbers for each row together. Recall that there are 24 cards for the solution of one equation in a set of 20 × 20 equations. Therefore, in column order, or card number in the row, there are 24 groups. Pull groups 15-20, as there is no use for these for a set of 14 × 14 equations. The remaining cards are the necessary cards for solution. Reproduce this deck (80-80). This deck is called the *work deck*. Put the master deck in order and file.

Key punch the a_{ij} coefficients and the right-hand sides, b_j, each in a card with column and row identification. There will be 210 cards punched in approximately 14 columns of the card. This punching must be correct; therefore, verify the punching! Sort the key punched cards in the same order as the work deck, column, and row. Take groups 1-14 and 21 out of the work deck. Reproduce the key punched coefficients into this part of the work deck, comparing the row and column for correct card punching. Place this deck back with the remainder of the work deck and sort into row and column order. This deck is ready for the solution on the CPC. It may be reproduced (80-80) several times to obtain a convenient number of cards if desired.

CPC Operation[b]

Reset all storage to zero. Commence running the cards with the setup switch on LIST, each card for one complete cycle. Check all 210 punched numbers against the original data. The machine will list from channel A. Switch off the accounting machine from this point, listing only card 23 and 24 of each row.

DISCUSSION

Mr. Liggett: If the matrix is positive definite, the Gauss-Seidel method may be used for the solution of the simultaneous linear equations. The process for discovering whether the matrix is positive definite is so laborious that it is not worth while. However, if the values on the main diagonal are large in comparison with the remainder of the elements of the matrix, the method is worth attempting.

Dr. Sherman: I have used the Gauss-Seidel method and found that it is very satisfactory for my purposes. In one case, in connection with infrared spectroscopy, I wished to solve a set of simultaneous equations with quadratic as well as linear terms. The Gauss-Seidel method may be applied to this problem. I neglected the quadratic terms and used only the linear terms for the first approximation. Having obtained the first approximation, I used it to solve the quadratic term. By this method the iteration may be continued and satisfactory results obtained.

Mr. Liggett: The problem I have discussed with you was a 10 × 10 matrix. It is part of a 20 × 20 matrix which was actually solved in our New York Technical Computing Bureau. The elements of the main diagonal consist of values between .6 and .9. The off-diagonal elements in some cases were as high as .7, with the major portion of the values smaller. Our problem of the 20 × 20 matrix was solved with 7 decimal-digit accuracy in 28 iterations. The 10 × 10 portion which I have used required 28 or 29 iterations to obtain the same accuracy. This indicates that the number of unknowns does not necessarily affect the rate of convergence of the iterative procedure.

[b] This method may be extended conveniently to 21 unknowns if the sums of products are accumulated electronically. Additional storage units allow the increase of the number of unknowns. If the sign of the answer is known, more than one element may be stored in the same unit, thus doubling or tripling the number of unknowns that can be accommodated.

One problem which has been solved on our SSEC contained 168 equations and required only nine iterations. Because this method does not take zero elements into consideration, these cards may be removed from the deck before calculation.

The Gauss-Seidel method may be adapted to the IBM Type 604 and 602-A Calculating Punches. Because there is not the added storage in these machines, it is necessary to punch intermediate results on a trailer card. These intermediate results are reproduced into another field of the card, the wires changed, and the iteration procedure repeated.

If a minor mistake is made during the calculation, there will not be any effect on the convergence, but if a major mistake is made, the effect would be the same as though the process had been begun again.

If, after a number of iterations, the values for the x_i's become increasingly large, convergence will not be accomplished. In general, after each iteration the values of the x's will improve, but not necessarily every value, every time.

Dr. Sherman: The Gauss-Seidel method has a more general application than the solution of simultaneous linear equations. It may be used for many types of simultaneous nonlinear equations.

Mr. Opler: The Fisher method can be used as an approximate test as to whether the matrix converges, that is, if $a_{ij}a_{ji} < a_{ii}a_{jj}$, the Gauss-Seidel method will converge. This is a sufficient condition, but not a necessary one.

REFERENCES

1. HAROLD HOTELLING, "Some New Methods in Matrix Calculation," *Annals of Mathematical Statistics,* Vol. 14, No. 1, pp. 1-34 (March 1943).
2. J. T. WYLIE, "Criteria of Convergence," *Annals of Mathematical Statistics* (1944).

Approximating the Roots of a Polynomial Equation

THE PROBLEM of approximating the roots of a polynomial equation arises frequently in industrial research. For example, in the aircraft industry it is desirable to calculate the roots of an equation which is obtained by equating the flutter determinant to zero.

A common method of solution of polynomial equations employs the Newton method of successive approximation. Thus, to approximate a root of $f(x) = 0$, one uses the iterative procedure described by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \,.$$

John Lowe has described a process for applying this procedure to polynomial equations with complex coefficients through the use of the IBM Type 602-A Calculating Punch.[1] Mr. Lowe uses synthetic division to calculate $f(x_n)$ and $f'(x_n)$ in one run through the 602-A, and calculates x_{n+1} in a second run.

The purpose of this paper is to indicate that the card-programmed electronic calculator has also been applied successfully to problems of this sort. We have wired a general-purpose control panel for complex arithmetic with ten-decimal digit accuracy. A total of sixty cards for data and programming will load the machine and carry out an iteration for a cubic polynomial equation with complex coefficients.

Extensions of the method of Newton to non-polynomial equations through the use of the CPC are clearly possible. For example, iterative solutions of equations containing trigonometric or exponential coefficients are possible if appropriate general purpose control panels are used.

REFERENCE

JOHN LOWE, "The Calculation of Roots of Complex Polynomials Using the IBM Type 602-A Calculating Punch," *Computation Seminar, December, 1949* (IBM), pp. 169-170.

Matrix by Vector Multiplication on the IBM Type 602-A Calculating Punch

ELEANOR KRAWITZ

International Business Machines Corporation

PUNCHED CARD equipment handles matrix arithmetic in a very efficient manner. Before giving a machine application, I should like to discuss briefly the basic principles of matrix arithmetic.

A matrix is a rectangular array of numbers; each number is referred to as an element of the matrix. Two indices are associated with each element: the first denotes the row, and the second denotes the column in which the element is found. For example, the element a_{ij} is located in the ith row and jth column of the matrix A. The following matrix A is referred to as an $n \cdot m$ matrix, where n (the number of rows) may or may not be equal to m (the number of columns).

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}$$

There is a matrix arithmetic that includes addition, subtraction, multiplication, and, in a sense, division. The sum of two matrices, $A + B = C$, is obtained by the addition of corresponding elements, i.e., $a_{ij} + b_{ij} = c_{ij}$. Subtraction is defined in a similar manner.

Consider a matrix A that has n rows and l columns, and a matrix B that has k rows and m columns. The product, $A \cdot B$, is defined only if $l = k$; that is, the number of columns in the matrix A must equal the numbers of rows in matrix B.

$$\begin{pmatrix} a_{11} & \dots & a_{1l} \\ a_{21} & \dots & a_{2l} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{n1} & \dots & a_{nl} \end{pmatrix} \cdot \begin{pmatrix} b_{11} & \dots & b_{1m} \\ b_{21} & \dots & b_{2m} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ b_{l1} & \dots & b_{lm} \end{pmatrix} = \begin{pmatrix} c_{11} & \dots & c_{1m} \\ c_{21} & \dots & c_{2m} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ c_{n1} & \dots & c_{nm} \end{pmatrix}$$

The product matrix, C, is an $n \cdot m$ matrix. The elements of the C matrix are defined in the following manner:

$$C_{ij} = \sum_{p=1}^{l} a_{ip} \cdot b_{pj} \, .$$

For example, $C_{12} = a_{11}b_{12} + a_{12}b_{22} + \dots + a_{1l}b_{l2}$.

If we have a set of twelve equations in twelve unknowns

$$\begin{array}{l} a_{1,1}\,x_1 + a_{1,2}\,x_2 + \dots + a_{1,12}\,x_{12} = b_1 \\ a_{2,1}\,x_1 + a_{2,2}\,x_2 + \dots + a_{2,12}\,x_{12} = b_2 \\ \cdot \qquad\qquad\qquad \cdot \\ \cdot \qquad\qquad\qquad \cdot \\ \cdot \qquad\qquad\qquad \cdot \\ a_{12,1}x_1 + a_{12,2}x_2 + \dots + a_{12,12}x_{12} = b_{12} \, , \end{array} \qquad (1)$$

it may be represented by a matrix and two single-column matrices. (A matrix that consists of a single row or column is referred to as a vector.)

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,12} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ a_{12,1} & a_{12,2} & \dots & a_{12,12} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_{12} \end{pmatrix} = \begin{pmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_{12} \end{pmatrix}, \text{ or } Ax = B. \qquad (2)$$

If the multiplication of the matrices is performed, the result will be the set of equations (1). To solve for the values of x, we must multiply both sides of the matrix equation (2) by $1/A$ or A^{-1}. The inverse of a matrix A, or A^{-1}, is defined as the matrix that, when multiplied by A, yields the unit matrix I.

$$AA^{-1} = A^{-1}A = I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

A discussion of matrix inversion will be given in a later paper.

Multiplying both sides of equation (2) we have

$$A^{-1}Ax = A^{-1}B \, .$$

Since any matrix multiplied by the unit matrix yields the matrix itself,

$$x = A^{-1}B\,,$$

or
$$\begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_{12} \end{pmatrix} = \begin{pmatrix} a'_{1,1} & \ldots & a'_{1,12} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a'_{12,1} & \ldots & a'_{12,12} \end{pmatrix} \begin{pmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_{12} \end{pmatrix}$$

or
$$x_i = a'_{i,1}b_1 + a'_{i,2}b_2 + \ldots + a'_{i,12}b_{12}\,, \qquad \text{where } i = 1, \ldots, 12.$$

We have now developed a problem that involves the multiplication of a vector by a matrix. There are many possible methods of performing this process on the calculating punches; the most efficient method depends upon the machines available, the order of the matrix, the number of digits in the elements, and the number of necessary multiplications. The following method was chosen for its simplicity. Assume that we have a vector with 12 elements, and a matrix of the 12th order;

$$(a_{1,1} \;\ldots\; a_{1,12}) \cdot \begin{pmatrix} b_{1,1} & \ldots & b_{1,12} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ b_{1,12} & \ldots & b_{12,12} \end{pmatrix}$$

The 12 columns in the B matrix are divided into four groups of three columns each, in the following manner:

$$\begin{array}{lll}
b_{1,1} & b_{1,2} & b_{1,3} \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
b_{12,1} & b_{12,2} & b_{12,3} \\
\\
b_{1,4} & b_{1,5} & b_{1,6} \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
b_{12,4} & b_{12,5} & b_{12,6} \\
\\
b_{1,7} & b_{1,8} & b_{1,9} \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
b_{12,7} & b_{12,8} & b_{12,9} \\
\\
b_{1,10} & b_{1,11} & b_{1,12} \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
b_{12,10} & b_{12,11} & b_{12,12}
\end{array}$$

Each row of three elements is punched into a card; there will be four groups of 12 cards each or a total of 48 cards. Three columns of each card carry the identification; one column denotes the group, and two columns denote the number of the card in the group. A group of 12 cards is punched for the A matrix, so that each card carries an element of the matrix. Two columns are used for the identification of the element. The elements of the A matrix are reproduced into each of the four groups of B-matrix cards. The cards are now in the following order:

1	1	a_1	$b_{1,1}$	$b_{1,2}$	$b_{1,3}$
1	2	a_2	$b_{2,1}$	$b_{2,2}$	$b_{2,3}$
.					
.					
.					
.					
1	12	a_{12}	$b_{12,1}$	$b_{12,2}$	$b_{12,3}$
.					
.					
.					
.					
4	1	a_1	$b_{1,10}$	$b_{1,11}$	$b_{1,12}$
4	2	a_2	$b_{2,10}$	$b_{2,11}$	$b_{2,12}$
.					
.					
.					
.					
4	12	a_{12}	$b_{12,10}$	$b_{12,11}$	$b_{12,12}$

A trailer card with a distinguishing X punch is inserted behind each group of (detail) cards. The sum of the products of the vector by three columns of the B matrix will be punched into a trailer card. The computation on the 602-A is outlined below (Figure 1, page 68).

Detail Cards:

Cards are skipped out without punching.

Read cycle: Read a_i into storage 1R.

Read $b_{i,j}$, $b_{i,j+1}$, $b_{i,j+2}$ into storage units 2R, 3R and 4R, respectively.

PI Multiply to obtain $a_ib_{i,j}$, $a_ib_{i,j+1}$, $a_ib_{i,j+2}$.

Read $\text{b}_{i,j}$ out of storage 2R, read in counters 1, 2.

Read $b_{i,j+1}$ out of storage 3R, read in counters 3, 4.

Read $b_{i,j+2}$ out of storage 4R, read in counters 5, 6.

P2 Read.

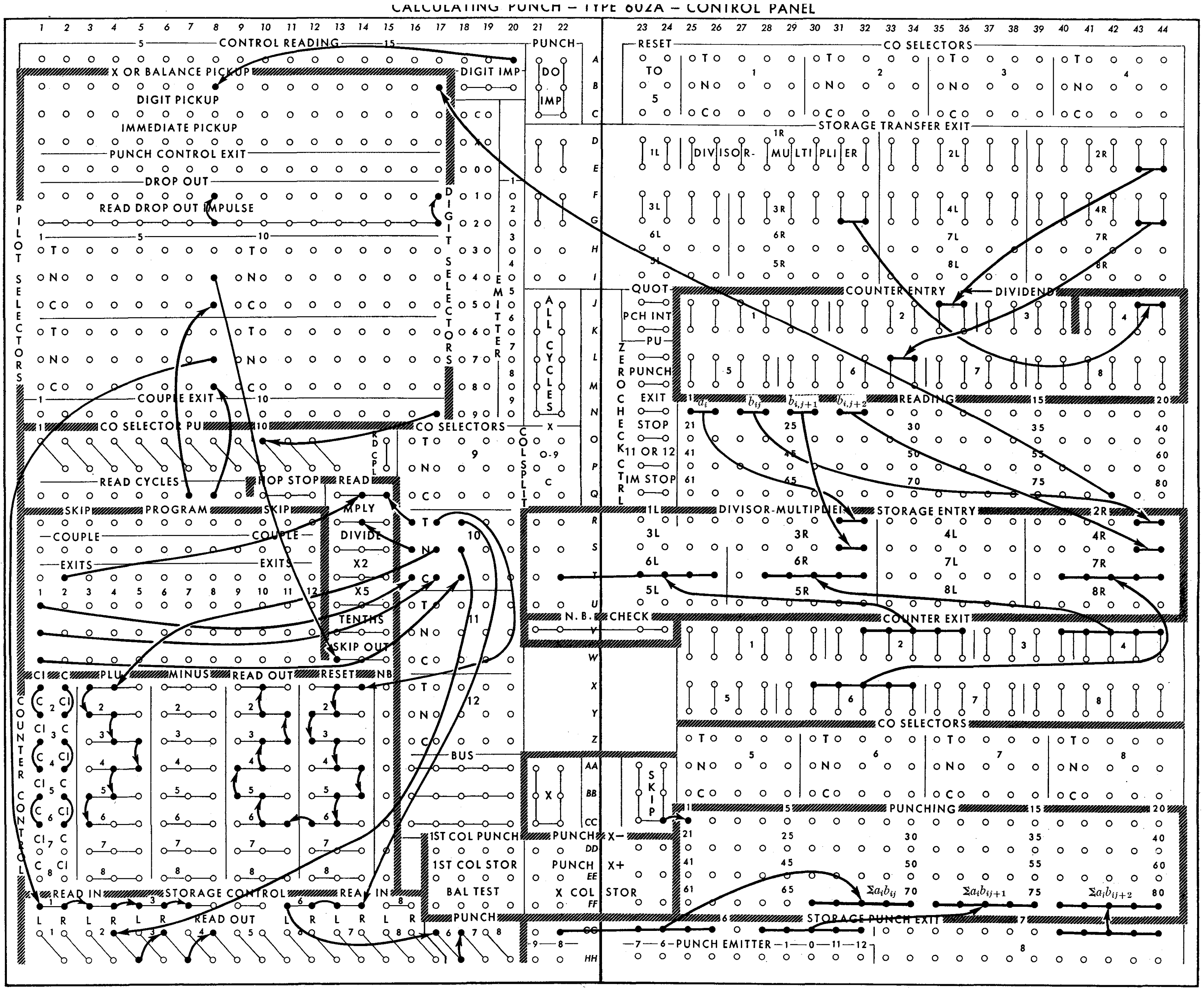


FIGURE 1. VECTOR BY MATRIX MULTIPLICATION (X card must be placed before group so that counters reset)

Trailer Cards:

PI Read out and reset counters 1, 2.

Read $\sum_{i=1}^{12} a_i b_{ij}$ into storage 6 and punch.

Read out and reset counters 3, 4.

Read $\sum_{i=1}^{12} a_i b_{i,j+1}$ into storage 6 and punch.

Read out and reset counters 5, 6.

Read $\sum_{i=1}^{12} a_i b_{i,j+2}$ into storage 7 and punch.

This procedure for multiplying a vector by a matrix is quite general. If the order of the matrices increases, the number of necessary cards increases, but the control panel wiring and the process remains the same. If the number of digits in each element is too large for the sum of the products to be accumulated in the designated counters, the matrix can be divided into six (or 12) groups, with two (or one) elements of the B matrix in a card. On the other hand, if the elements are two-digit numbers, four elements of the final matrix can easily be obtained simultaneously.

DISCUSSION

Dr. Sherman: Could the described method of multiplying a vector by a matrix be extended to include multiplication of nth order matrices?

Miss Krawitz: It could be extended very easily. It is only necessary to duplicate the number of cards in the B matrix as many times as there are rows in the A matrix.

Dr. Sherman: Is there any simpler method that anyone has developed?

Miss Krawitz: It is very difficult to answer your question directly. What may be the simplest method for one series of problems may be impossible for another. The simplest method, as I have stated before, depends upon the order of the matrix, the size of the elements, the machines available, and the number of multiplications in the problem. The method I have described is the most general, but is not the most efficient in all cases. For example, Dr. Petrie has used an entirely different approach. His problem was to multiply a 17th-order vector matrix by a 17th-order square matrix on the 602-A. All of the 17 elements of the vector are read into the machine on the first card cycle, and stored for the remainder of the calculation. Each element of the square matrix is put on a card (289 cards) and arranged in order of columns. After each element of a column of the square matrix is read by the machine it is multiplied by the corresponding element of the vector. The sum of the products of the vector by each column is accumulated and punched on a trailer card. This method is particularly efficient if a given matrix is to be used many times with different vector matrices. The fully equipped 602-A could handle a 24th-order matrix and vector, provided that the elements were four-digit numbers.

Dr. Sherman: There is one slight thing that can be done to speed this process. It is not necessary, of course, to punch the a's into the card that contains the b's. You can use one card as the master and the other as a detail card and, in general, eliminate the necessity for reproducing.

Miss Krawitz: It is true that the method you refer to eliminates the reproducing operation; it also keeps the original matrices intact, allowing them to be used over and over again. However, the machine time on the 602-A will be increased. The benefits of the two methods should be considered with each problem.

Dr. King: Have you ever modified this method to use complex numbers? The principle seems to be applicable.

Miss Krawitz: Complex numbers can be handled just as easily as real numbers. It is true that the number of multiplications and additions will be increased, but all of the methods described can be easily extended to include complex numbers.

Dr. Hurd: One application of matrix by vector multiplication known to many of you is the application in spectrometry. For a given instrument, you have a calibration which determines the coefficients in the left-hand side of an equation. With repeated routine analysis, the elements which change are the elements which are on the right-hand side of the equation. If the calibration is sufficient for three months and you have a good many routine analyses, it is profitable to invert the matrix and each time perform the matrix by vector multiplication. Similarly, many of you are interested in vibration analysis of mathematical models, which are in the form of linear ordinary differential equations with constant coefficients. It is possible to write such an equation in matrix form. In order to obtain a set of fundamental solutions of the original equation, one way to approach it is to start with a vector and find out successive powers of the matrix itself times the vector, which again brings in the matrix by vector multiplication. In the aircraft industry where one problem is flutter analysis, this iterative method is used so much that some of our calculators are busy 24 hours a day on this particular problem.

Mr. Walker: In the problem where there are a great number of zeros in a matrix, if it were possible to arrange these zeros so that it approaches a triangular matrix, would that not greatly simplify the problem? You could treat the original triangular matrix in a method of elimination to take care of the few coefficients that do occur below the diagonal term and then invert the top triangle of the matrix.

Dr. Grosch: Indeed, that is a fruitful idea. It is not always easy to arrange the matrix in a purely triangular form, but it is very easy to use a matrix with a large number of zeros with standard equipment if you make a very careful sorting code.

Dr. Sherman: In recalibrating from time to time there may be one critical calibration that has to be done often. You want to calculate the new inverse that results from just changing one column of the old one. I have worked with methods, and there are rather simple methods of calculating a new inverse from an old inverse as a result of making various changes, such as replacing one column by a new column or one row and column or one element, or deleting a row and column, or augmenting a row and column. The practical situation is much more encouraging because instead of having to go to the trouble of inverting the matrix, you can save a tremendous amount of work through a whole series of associated problems. We actually do that. In practice, you have a rather interesting situation of being able to take a large order inverse and make a rather large number of minor changes of rows and columns, or elements in rows, and maintain the original work of inverting that large matrix.

Numerical Solution of Two Simultaneous Second-Order Differential Equations

WALTER H. JOHNSON

International Business Machines Corporation

THIS PAPER describes the numerical solution, using the IBM Type 604 Calculating Punch, of the following set of two simultaneous second-order linear differential equations:

$$\ddot{\theta} + \omega_\theta^2\,\theta - K_\phi\phi = 0\,, \qquad (1)$$
$$\ddot{\phi} + \omega_\phi^2\,\phi - K_\theta\,\theta = F(t)\,.$$

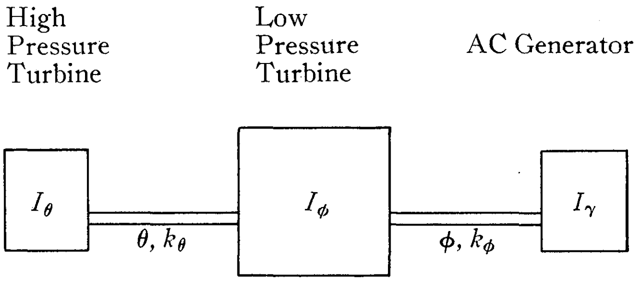


FIGURE 1

These equations express the relations that exist between the angular displacements, acceleration of angular displacements, reaction torque, moments of inertia, and tortional stiffness of the shafts for a turbine-generator set, where:

θ = angular displacement of the shaft between the high pressure and low pressure turbines.

ϕ = angular displacement of the shaft between the low pressure turbine and the AC generator.

$\omega_\theta^2 = k_\theta\left(\frac{1}{I_\theta}+\frac{1}{I_\phi}\right)$ = natural frequency of shaft θ.

$\omega_\phi^2 = k_\phi\left(\frac{1}{I_\phi}+\frac{1}{I_\gamma}\right)$ = natural frequency of shaft ϕ.

k_θ = torsional stiffness of shaft θ.

k_ϕ = torsional stiffness of shaft ϕ.

$$K_\theta = \frac{k_\theta}{I_\phi}$$

$$K_\phi = \frac{k_\phi}{I_\phi}$$

$$F(t) = \frac{\text{reaction torque}}{I_\gamma}\ \text{(Given for each 20°)}$$

Numerical Data

50,000 kw AC generator set, tandem compound, double flow, 5-bearing.

3,600 rpm

$$\begin{aligned} I_\theta &= 2{,}930 \text{ lb. in. sec.}^2 \\ I_\phi &= 9{,}500 \text{ lb. in. sec.}^2 \\ I_\gamma &= 15{,}650 \text{ lb. in. sec.}^2 \\ k_\theta &= 515 \text{ lb. in.} \\ k_\phi &= 595 \text{ lb. in.} \end{aligned} \qquad (2)$$

$f(t)$ is tabulated for each 20° of rotation from $t = 0$.

$$h = \frac{20°}{360°}\cdot\frac{1}{3600}\cdot 60 = 0.000926 \text{ seconds}$$

$$h = 20 \text{ degrees} = \frac{20°}{57.2958°/\text{radian}} \qquad (3)$$

$$\frac{1}{h} = 2.86479/\text{radian}$$

$$\frac{1}{h^2} = 8.20702$$

Approximation of Finite Differences

$F(t)$ is a known function of t over the range of integration. Some constant increment in time, h, is decided upon so that

$$t_{n+1} = t_n + h\,. \qquad (4)$$

In this case h is taken to be the time required for the generator to rotate 20°, since $F(t)$ was given for this time increment,

$$h = .000926 \text{ seconds.}$$

Suppose that in the region of an arbitrary point, t_n, within the interval of integration, that the solutions θ and ϕ and their second derivatives can be developed into a Taylor's series. Then,

$$\theta_{n+1} = \theta_n + h\overset{I}{\theta}_n + \frac{h^2}{2!}\overset{II}{\theta}_n + \frac{h^3}{3!}\overset{III}{\theta}_n + \frac{h^4}{4!}\overset{IV}{\theta}_n + \ldots.$$

and $$\phi_{n+1} = \phi_n + h\overset{I}{\phi}_n + \frac{h^2}{2!}\overset{II}{\phi}_n + \frac{h^3}{3!}\overset{III}{\phi}_n + \frac{h^4}{4!}\overset{IV}{\phi}_n + \ldots.$$

also $$\overset{II}{\theta}_{n+1} = \overset{II}{\theta}_n + h\overset{III}{\theta}_n + \frac{h^2}{2!}\overset{IV}{\theta}_n + \frac{h^3}{3!}\overset{V}{\theta}_n + \frac{h^4}{4!}\overset{VI}{\theta}_n + \ldots. \qquad (5)$$

and $$\overset{II}{\phi}_{n+1} = \overset{II}{\phi}_n + h\overset{III}{\phi}_n + \frac{h^2}{2!}\overset{IV}{\phi}_n + \frac{h^3}{3!}\overset{V}{\phi}_n + \frac{h^4}{4!}\overset{VI}{\phi}_n + \ldots.\,.$$

A modified function can be defined which will eliminate all odd derivatives and the fourth derivative. This modified function is

$$x = \theta - \frac{h^2}{12}\overset{II}{\theta} \qquad (6)$$

$$y = \phi - \frac{h^2}{12}\overset{II}{\phi}\,.$$

Insert the expansions of θ_{n+1} and $\overset{II}{\theta}_{n+1}$ from equations (5) to obtain x_{n+1}. Thus,

$$\begin{aligned} x_{n+1} = {} & \theta_n + h\overset{I}{\theta}_n + \frac{h^2}{2!}\overset{II}{\theta}_n + \frac{h^3}{3!}\overset{III}{\theta}_n + \frac{h^4}{4!}\overset{IV}{\theta}_n + \dots \\ & - \frac{h^2}{12}\overset{II}{\theta}_n - \frac{h^3}{12}\overset{III}{\theta}_n - \frac{h^4}{24}\overset{IV}{\theta}_n - \frac{h^5}{72}\overset{V}{\theta}_n - \dots \end{aligned}$$

or $x_{n+1} =$ (7)

$$\theta_n + h\overset{I}{\theta}_n + \frac{5h^2}{12}\overset{II}{\theta}_n + \frac{h^3}{12}\overset{III}{\theta}_n - \frac{h^5}{180}\overset{V}{\theta}_n - \frac{h^6}{480}\overset{VI}{\theta}_n - \dots\,.$$

By changing the sign of h, you can now write the expansion of the modified function at t_{n-1}. That is;

$$\begin{aligned} x_{n-1} = {} & \theta_n - h\overset{I}{\theta}_n + \frac{h^2}{2!}\overset{II}{\theta}_n - \frac{h^3}{3!}\overset{III}{\theta}_n + \frac{h^4}{4!}\overset{IV}{\theta}_n + \frac{h^5}{5!}\overset{V}{\theta}_n \\ & + \dots - \frac{h^2}{12}\overset{II}{\theta}_n + \frac{h^3}{12}\overset{III}{\theta}_n - \frac{h^4}{24}\overset{IV}{\theta}_n - \frac{h^5}{72}\overset{V}{\theta}_n \\ & - \dots \end{aligned}$$

or (8)

$$x_{n-1} = \theta_n - h\overset{I}{\theta}_n + \frac{5}{12}h^2\overset{II}{\theta}_n - \frac{1}{12}h^3\overset{III}{\theta}_n + \frac{1}{180}h^5\overset{V}{\theta}_n - \dots\,.$$

Similar series pertain for the modified function y.

The central difference of the modified function x about the point t_n is

$$\Delta^{II}x_n = x_{n+1} - 2x_n + x_{n-1}\,. \quad \left\{\begin{array}{llll} t_{n-2} & x_{n-2} & & \\ t_{n-1} & x_{n-1} & & \\ & & \Delta^I x_{n-1/2} & \\ t_n & x_n & & \Delta^{II}x_n \\ & & \Delta^I x_{n+1/2} & \\ t_{n+1} & x_{n+1} & & \\ t_{n+2} & x_{n+2} & & \end{array}\right.$$

$$\left\{\begin{aligned} \Delta^{II}x_n &= \Delta^I x_{n+1/2} - \Delta^I x_{n-1/2} \\ &= x_{n+1} - x_n - (x_n - x_{n-1}) \\ &= x_{n+1} - 2x_n + x_{n-1} \end{aligned}\right. \qquad (9)$$

Substituting from (6), (7) and (8) into (9) gives

$$\begin{aligned} \Delta^{II}x_n = {} & \theta_n + h\overset{I}{\theta}_n + \frac{5}{12}h^2\overset{II}{\theta}_n + \frac{1}{12}h^3\overset{III}{\theta}_n - \frac{1}{180}h^5\overset{V}{\theta}_n - \dots \\ & - 2\theta_n + h^2\overset{II}{\theta}_n \qquad (10) \\ & + \theta_n - h\overset{I}{\theta}_n + \frac{5}{12}h^2\overset{II}{\theta}_n - \frac{1}{12}h^3\overset{III}{\theta}_n + \frac{1}{180}h^5\overset{V}{\theta}_n - \dots \end{aligned}$$

or $\Delta^{II}x_n = h^2\overset{II}{\theta}_n - \frac{1}{240}h^6\overset{VI}{\theta}_n - \dots\,.$

By neglecting the terms of the 6th order and higher, and thus forming a truncation error, we have

$$\Delta^{II}x_n = h^2\overset{II}{\theta}_n \rightarrow \overset{II}{\theta}_n = \frac{\Delta^{II}x_n}{h^2}\,,$$

likewise

$$\Delta^{II}y_n = h^2\overset{II}{\phi}_n \rightarrow \overset{II}{\phi}_n = \frac{\Delta^{II}y_n}{h^2}\,. \qquad (11)$$

By substitution into equations (6) from (11) we can state the values of the functions θ and ϕ in terms of the modified functions x and y and their second differences:

$$x_n = \theta_n - \frac{h^2}{12}\left(\frac{\Delta^{II}x_n}{h^2}\right),$$

$$\theta_n = x_n + \frac{\Delta^{II}x_n}{12} = \frac{x_{n+1} + 10x_n + x_{n-1}}{12}\,,$$

and $\phi_n = y_n + \frac{\Delta^{II}y_n}{12} = \frac{y_{n+1} + 10y_n + y_{n-1}}{12}\,.$ (12)

Substituting in the original differential equations the expressions derived for the functions θ and ϕ and their second derivatives in terms of the modified functions x and y, we obtain

$$\frac{\Delta^{II}x_n}{h^2} + \omega_\theta^2\left(x_n + \frac{\Delta^{II}x_n}{12}\right) - K_\phi\left(y_n + \frac{\Delta^{II}y_n}{12}\right) = 0\,,$$

$$\frac{\Delta^{II}y_n}{h^2} + \omega_\phi^2\left(y_n + \frac{\Delta^{II}y_n}{12}\right) - K_\theta\left(x_n + \frac{\Delta^{II}x_n}{12}\right) = f(t_n)\,,$$

or (13)

$$\Delta^{II}x_n\left(\frac{1}{h^2} + \frac{\omega_\theta^2}{12}\right) - K_\phi\frac{\Delta^{II}y_n}{12} = K_\phi y_n - \omega_\theta^2 x_n\,,$$

$$\Delta^{II}y_n\left(\frac{1}{h^2} + \frac{\omega_\phi^2}{12}\right) - K_\theta\frac{\Delta^{II}x_n}{12} = K_\theta x_n - \omega_\phi^2 y_n + f(t_n)\,.$$

Solve for $\Delta^{II}x_n$ and $\Delta^{II}y_n$ to obtain the equations:

$$\Delta^{II}x_n = Ax_n + By_n + Cf(t_n)\,,$$

$$\Delta^{II}y_n = Ex_n + Fy_n + Gf(t_n)\,, \qquad (14)$$

where $A = \dfrac{K_\theta K_\phi - 12\omega_\theta^2\left(\frac{1}{h^2} + \frac{\omega_\phi^2}{12}\right)}{J}$ $\qquad E = \dfrac{12K_\theta}{h^2 J}$

$$B = \frac{12K_\phi}{h^2 J} \qquad F = \frac{K_\theta K_\phi - 12\omega_\phi^2\left(\frac{1}{h^2} + \frac{\omega_\theta^2}{12}\right)}{J}$$

$$C = \frac{K_\phi f(t)}{J} \qquad G = \frac{12f(t)\left(\frac{1}{h^2} + \frac{\omega_\theta^2}{12}\right)}{J}$$

$$J = 12\left(\frac{1}{h^2} + \frac{\omega_\phi^2}{12}\right)\left(\frac{1}{h^2} + \frac{\omega_\theta^2}{12}\right) - \frac{K_\phi K_\theta}{12}\,.$$

The coefficients A, B, C, E, F, and G are properties of the physical characteristics of the system only and remain constant for the entire integration—except for coefficients C and G, which contain the forcing function and must be computed for each integration step.

Consider again the central difference formula:

$$\begin{aligned} & \Delta^{II}x_n = x_{n+1} - 2x_n + x_{n-1} \\ \text{and} \quad & \Delta^{II}y_n = y_{n+1} - 2y_n + y_{n-1}\,, \\ \text{or} \quad & x_{n+1} = \Delta^{II}x_n + 2x_n - x_{n-1} \qquad (15) \\ \text{and} \quad & y_{n+1} = \Delta^{II}y_n + 2y_n - y_{n-1}\,. \end{aligned}$$

Substituting into (15) from (14) for the second differences we have:

$$x_{n+1} = Ax_n + By_n + Cf(t_n) + 2x_n - x_{n-1} \,,$$

or $\quad x_{n+1} = (2+A)\, x_n + By_n + Cf(t_n) - x_{n-1}$

and $\quad y_{n+1} = Ex_n + Fy_n + Gf(t_n) + 2y_n - y_{n-1} \,, \quad (16)$

or $\quad y_{n+1} = (2+F)\, y_n + Ex_n + Gf(t_n) - y_{n-1} \,.$

Starting with initial values for x_o, x, y_o, and y, the integration can be carried forward to obtain the modified functions, $x(t)$ and $y(t)$. By employing equations (12), these solutions for the modified function are immediately transformed to the functions $\theta(t)$ and $\phi(t)$;

$$\text{i.e., } \theta_n = \frac{x_{n+1} + 10x_n + x_{n-1}}{12} \,,$$

$$\phi_n = \frac{y_{n+1} + 10y_n + y_{n-1}}{12} \,. \qquad (12)$$

The constants for the problem are

$$(2+A) = +0.59410 \,,$$
$$B = +0.36467 \,,$$
$$(2+F) = +1.34706 \,,$$
$$E = +0.31564 \,.$$

The coefficients $Cf(t_n)$ and $Gf(t_n)$ are tabulated for integration points t_n.

Solution on the IBM Type 604 Calculating Punch

Cards are prepared in pairs; i.e., one card each for the θ and ϕ shaft for each point t_n (Figure 2).

Card Form

col. 1-2	col. 3-7 B or E	col. 8-13 $(2+A)$ or $(2+F)$	col. 14-18 $Cf(t)$ or $Gf(t)$	col. 19-23	col. 24-28
θ Card 1	.36467	.59410	.00015	θ	ϕ
ϕ Card 2	.31564	1.34706 etc.	.00481		

FIGURE 2

Assignment of Data in 604 Calculator

col. 3-7 = B or E assigned to MQ (Multiplier Quotient register)

col. 8-13 = $(2+A)$ or $(2+F)$ is assigned to FS 1 and 3 (Factor Storage units)

col. 14-18 = $Cf(t)$ or $Gf(t)$ is assigned to GS 1 and 3 (General Storage units)

Sequence card θ_n:

1.	RO GS 4, multiply +	By_n
2.	RO GS 2, RI MQ	x_n
3.	RO FS 1 and 3, multiply +	$(2+A)x_n + By_n$
4.	RO GS 1 and 3, RI EC + 6th	$Cf(t_n)$; $(2+A)x_n + By_n + Cf(t_n)$
5.	RO FS 2, RI EC—6th	$(2+A)x_n + By_n + Cf(t_n) - x_{n-1} = x_{n+1}$
6.	½ adjust 5th	
7.	RO and RS EC 6th, RI MQ	x_{n+1}
8.	Emit 1 + 2nd	
9.	Emit 2 + 1st	12
10.	RO and RS EC, RI GS 1 and 3	12
11.	RO MQ, RI EC + 1st	x_{n+1}
12.	RO GS 2, RI EC + 2nd	$x_{n+1} + 10x_n$
13.	RO FS 2, RI EC + 1st	$x_{n+1} + 10x_n + x_{n-1}$
14.	RO GS 2, RI FS 2	shift x_n to x_{n-1}
15.	RO MQ, RI GS 2	shift x_{n+1} to x_n
16.	RO GS 1 and 3, divide	$\theta_n = \dfrac{x_{n+1} - 10x_n - x_{n-1}}{12}$

DISCUSSION

Mr. Brown: I would like to ask about the possibility of getting these difference equations in a more direct manner. What is done here is to start with a differential equation and then go through an elaborate process of making approximations in terms of differences. Then, finally, one finishes with a set of difference equations. Now, in this case, the original differential equations themselves can be obtained by varying an integral. Instead of going through all the manipulation with the differential equation, one could conceivably start with an approximation in terms of differences for the kinetic and potential energy terms in Hamilton's integral. Then, carry out a variational process and obtain difference equations without ever handling the differential equations and without having to put all these approximations into the differential equations themselves. In many physical problems which can be expressed as the problem of varying an integral, the usual procedure is to do something like that; for example, to introduce a polynomial or other approximation with undetermined parameters in the integral and then to vary the parameters to minimize the integral. That would be the logical thing to do here. Has that been looked into in this type of problem, and if so what are the relative merits of each method?

Mr. Johnson: We did not look into that. I presume you could do it as you have outlined.

Mr. Collins: It is unnecessary to solve this problem by approximate methods as far as I can tell. If you have constant coefficients, it looks as though you could obtain two quadratic algebraic equations and reduce them to a fourth order polynomial and calculate the necessary roots. You would have to write out an explicit formula for the answer, and then evaluate all the points you wanted to from trigonometric exponential functions. I wondered why it is desirable to do the problem by this approximate method. There probably is a good reason.

Mr. Sheldon: I think it is really easier to solve this problem by the integration of the differential equation instead of evaluating the integral you would obtain if you solved it explicitly.

Numerical Evaluation of Integrals of the Form $\int_{\alpha}^{\beta} f(x)g(x)dx$

JOHN W. SHELDON

International Business Machines Corporation

SUPPOSE we wish to evaluate numerically

$$\int_{\alpha}^{\beta} f(x)g(x)dx\,,$$

and suppose $g(x)$ varies much more rapidly than $f(x)$ so that much finer intervals are required to represent $g(x)$ by an interpolating polynomial than would be required for $f(x)$. It is then worth while to consider calculating special Lagrangian integration coefficients, d_l, which take account of the variation of $g(x)$ so that

$$\int_{\alpha}^{\beta} f(x)g(x)dx = \sum d_l f_l$$

with the ordinates f_l chosen at intervals in x which need only be fine enough to represent $f(x)$ by an interpolating polynomial. Cases where this approach may be especially useful are:

1. $f(x)$ depends on one or more parameters, and integrations are desired over ranges of the parameters. Then the number of operations required to solve the problem may be very substantially reduced by this technique.

2. An integrand, $F(x)$, has a singularity $g(x)$ where $g(x)$ can be integrated analytically, and $\int g(x)dx$ is finite.

$$\text{Then } F(x) = \frac{F(x)}{g(x)}\,g(x) = f(x)g(x)\,.$$

3. $g(x)$ is periodic and varies so rapidly that we can represent $F(x)$ by an interpolating polynomial with intervals in x of one period of $g(x)$. In this case we obtain especially simple integration formulae.

Derivation of Formulae for the Special Integration Coefficients

For simplicity, let the range of integration be 0 to a. Let $(0,a)$ be divided into n equal intervals $(0,h)$, $(h,2h)$, . . . , $[(n-1)h,nh]$. Then

$$\int_0^a f(x)g(x)dx = \int_0^{h/2} f(x)g(x)dx + \sum_{j=1}^{n-1}\int_{(j-1/2)h}^{(j+1/2)h} f(x)g(x)dx + \int_{(n-1/2)h}^{nh} f(x)g(x)dx\,. \qquad (1)$$

Let $x = h(j+\xi)$ in the jth interval.

$$\text{Then} \int_0^a f(x)g(x)dx = h\left\{\int_0^{1/2} f(h\xi)g(h\xi)d\xi + \sum_{j=1}^{n-1}\int_{-1/2}^{1/2} f[h(j+\xi)]g[h(j+\xi)]d\xi + \int_{-1/2}^{0} f[h(n+\xi)]g[h(n+\xi)]d\xi\right\}. \qquad (2)$$

Let[a] $f[h(j+\xi)] = \sum_{k=0}^{m} b_{kj}\xi^k$ where m is a positive even number, (3)

$$\text{and where} \qquad b_{kj} = \sum_{l=-m/2}^{m/2} a_{lk}\, f_{j+l}\,. \qquad (4)$$

$$\text{Then,} \int_0^a f(x)g(x)dx = h\left\{\sum_{k=0}^{m} b_{k0}\int_0^{1/2} \xi^k g[h\xi]d\xi \right. \qquad (5)$$

$$\left. + \sum_{j=1}^{n-1}\sum_{k=0}^{m}\int_{-1/2}^{1/2} \xi^k g[h(j+\xi)]d\xi + \sum_{k=0}^{m} b_{kn}\int_{-1/2}^{0} \xi^k g[h(n+\xi)]d\xi\right\}.$$

$$\text{Let} \qquad c_{0k} = \int_0^{1/2} \xi^k g(h\xi)d\xi$$

[a] Here is a representation in Lagrangian form of an interpolating polynomial. I have found that in most cases the Lagrangian form is most convenient. An exception is case 3 above, where it is best to introduce interpolating polynomials with coefficients given in terms of differences. The particular interpolating polynomial introduced here involves ordinates outside the range of integration. Of course, if $f(x)$ does not exist outside the range of integration, different interpolating polynomials must be used near the ends of the range of integration.

$$c_{jk} = \int_{-1/2}^{1/2} \xi^k g[h(j+\xi)]d\xi \qquad j = 1,2,\ldots,n-1 \tag{6}$$

$$c_{nk} = \int_{-1/2}^{0} \xi^k g[h(n+\xi)]d\xi$$

$$\int_0^a f(x)g(x)dx = h\sum_{j=0}^{n}\sum_{k=0}^{m}\sum_{l=-m/2}^{m/2} a_{lk}c_{jk}f_{j+l} \tag{7}$$

Let $l = p-j$. Then,

$$\int_0^a f(x)g(x)dx = h\sum_{p=-m/2}^{m/2}\ \sum_{j=p-m/2\geqq 0}^{p+m/2\leqq n}\ \sum_{k\geqq|p-j|}^{m} a_{p-j,k}c_{jk}f_p \tag{8}$$

so that
$$d_p = h\sum_{j=p-m/2\geqq 0}^{p+m/2\leqq n}\ \sum_{k\geqq|p-j|}^{m} a_{p-j,k}c_{jk}\,, \tag{9}$$

and
$$\int_0^a f(x)g(x)dx = \sum_{p=-m/2}^{p=n+m/2} d_pf_p\,. \tag{10}$$

After the quantities d_p have been evaluated, the integral can be evaluated efficiently by machine for any choice of the function $f(x)$, provided only that $f(x)$ can be represented by interpolating polynomials in the interval chosen. The evaluation of the coefficients d_p is often the central problem. First of all one must obtain the coefficients c_{jk}. Sometimes the c_{jk} can be obtained by analytical integration, in which case they are represented by formulae which may be evaluated by hand or by machine. Otherwise, the c_{jk} must be obtained by numerical integration. Once the c_{jk} are found, the d_p's may be evaluated from equation (9). This summation is somewhat tedious to carry out if the card volume is small. Equation (9) is a formula for the d_p's which gives results with accuracy analagous to Bessel's integration formula in the ordinary theory of numerical integration, and it is the most accurate type of formula we can obtain using equal intervals and interpolating polynomials. A simpler procedure which we have used, but which is not so accurate, involves strip formulae and Vandermonde matrices as follows:

Divide the range of integration into subintervals $(0,h)$, $(h,2h)$, . . . , $[(n-1)h,nh]$ as before, but now let

$$\int_0^a f(x)g(x)dx = \sum_{p=0}^{n/m-1}\int_{mph}^{m(p+1)h} f(x)g(x)dx\,. \tag{11}$$

Put $x = h[m(p+1/2) + \xi]$ so that

$$\int_0^a f(x)g(x)dx = h\sum_{p=0}^{n/m-1}\int_{-m/2}^{m/2} f\{h[m(p+1/2) + \xi]\} \cdot g\{h[m(p+1/2) + \xi]\}d\xi\,, \tag{12}$$

and let
$$\int_{-m/2}^{m/2} f\{h[m(p+1/2) + \xi]\} \cdot g\{h[m(p+1/2) + \xi]\}d\xi = \sum_{j=-m/2}^{j=m/2} d_{mp+j}f_{mp+j}\,. \tag{13}$$

In order to evaluate the coefficients d_{mp+j}, let $f\{h[m(p+1/2) + \xi]\}$ equal successively 1, ξ, ξ^2, . . . , ξ^m,

and let
$$\int_{-m/2}^{m/2} \xi^n g\{h[m(p+1/2) + \xi]\}\mathrm{d}\xi = c^n_{mp}. \tag{14}$$

Then we obtain the system of $(m+1)$ simultaneous equations

$$c^n_{mp} = \sum_{j=-m/2}^{j=m/2} \xi^n_j d_{mp+j} \qquad n=0,1,2,\ldots,m \tag{15}$$

which can be solved for the d_{mp+j}, obtaining

$$d_{mp+j} = \sum_{n=0}^{n=m} (\xi^n_j)^{-1}\, c^n_{mp}\,. \tag{16}$$

After we have calculated d_{mp+j} from (16) we must add together $d_{mp+m/2}$ and $d_{m(p+1)-m/2}$ as they are both coefficients of the same ordinate. Here we have developed "strip" and "repeated strip" formulae which are analogous in accuracy to such formulae as Simpson's rule and Weddle's rule in the ordinary methods of numerical integration.

Applications

1. Evaluate $\int_0^\infty f(x)\cos kx\,dx$ when k is so large that we may take steps of one period in the cosine and still represent $f(x)$ accurately by an interpolating polynomial. In this case, we use the more accurate formula (9) for the d_l's and Stirling's interpolation formula in terms of differences in place of equation (3).

Because the average value of the cosine is zero, all the "zero-order" terms in the interpolating polynomial contribute nothing to the integral. Because the cosine is an even function, all the odd order terms cancel. We are left with a sum over the even order terms which are even order differences of $f(x)$ times constant coefficients. These are integrated to odd-order differences at the beginning and at the end of the range of integration. Taking account of all differences up to and including the 6th, we obtain

$$\int_0^\infty f(x)\cos kx\,dx = \frac{1}{k}\Big\{0.0016007662\left[f\left(\frac{-5\pi}{k}\right)-f\left(\frac{5\pi}{k}\right)\right] -0.0186667286\left[f\left(\frac{-3\pi}{k}\right)-f\left(\frac{3\pi}{k}\right)\right] +0.2071512979\left[f\left(\frac{-\pi}{k}\right)-f\left(\frac{\pi}{k}\right)\right]\Big\}\,. \tag{17}$$

For example, evaluate

$$\int_0^\infty e^{-x}\cos 50\pi x\,dx\,.$$

We obtain from (17)

$$\int_0^\infty e^{-x}\cos 50\pi x\,dx = 0.0000405268311.$$

By analytical evaluation we obtain

$$\int_0^\infty e^{-x}\cos 50\pi x\,dx = 0.0000405268310.$$

To obtain this result using, for instance, Simpson's rule would require us to add many thousand values of the integrand.

2. Evaluate $\int_0^\infty e^{-x/2}\,\mathrm{K}_0(x/2)\cos kx\,dx$ where $K_0(x)$ is a Bessel function of pure imaginary argument, defined in Watson's *Bessel Functions*.[1] Near the origin $K_0(x/2) \approx -\gamma - \log_e(x/4)$ where γ is the Euler-Maclaurin constant. Dividing by $\log_e 10$, we have

$$\frac{K_0(x/2)}{\log_e 10} = -\frac{\gamma}{\log_e 10} - \log_{10}(x/4)\,.$$

$$\left.\begin{aligned} \text{Let } f(x) &= \frac{e^{-x/2}K_0(x/2)\cos kx}{-\gamma/\log_e 10 - \log_{10}(x/4)} \\ g(x) &= -\gamma/\log_e 10 - \log_{10}(x/4) \end{aligned}\right\} \text{for } 0 \leqq x \leqq 0.48\,.$$

Then $f(x)$ is "slowly-varying" as compared to $g(x)$. To evaluate the integral, take steps of .04 in x and use $m=4$ (five-ordinate strips) in equations (14), (15) and (16). Then,

$$c^n_{4p} = \int_{-2}^{2} \xi^n\{-\gamma/\log_e 10 - \log_{10}[.16(p+1/2) + .04\xi]\}d\xi\,.$$

For $p=0$ we have

$c^0_0 = .2530485$
$c^1_0 = .0694871-$
$c^2_0 = .3682811$
$c^3_0 = .0926495-$
$c^4_0 = .9283464\,.$

The matrix (ξ^n_j) is

$$\begin{pmatrix} -1 & +1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \\ +4 & 1 & 0 & 1 & 4 \\ -8 & -1 & 0 & 1 & 8 \\ 16 & 1 & 0 & 1 & 16 \end{pmatrix},$$

and we find that $(\xi^n_j)^{-1}$ is

$$\begin{pmatrix} 0 & 1/12 & -1/24 & -1/12 & 1/24 \\ 0 & -2/3 & 2/3 & 1/6 & -1/6 \\ 1 & 0 & -5/4 & 0 & 1/4 \\ 0 & 2/3 & 2/3 & -1/6 & -1/6 \\ 0 & -1/12 & -1/24 & 1/12 & 1/24 \end{pmatrix},$$

so that from formula (16) we obtain

$d_{-2} = .0252662$
$d_{-1} = .1216794$
$d_0 = .0247837$
$d_1 = .0599131$
$d_2 = .0214058\,.$

Repeating this process for $p = 1,2,3$ we obtain

$d_{-2} = .0252662$
$d_{-1} = .1216794$
$d_0 = .0247837$
$d_1 = .0599131$
$-d_2 = .0355995$
$d_3 = .0600419$
$d_4 = .0203343$
$d_5 = .0516705$
$-d_6 = .0209827$
$d_7 = .0453289$
$d_8 = .0158488$
$d_9 = .0403588$
$-d_{10} = .0083192\,.$

($d_{mp+m/2}$ and $d_{m(p+1)-m/2}$ have been added together.)

For the range .48 to ∞, which we take to be .48 to 10.0, we use Gregory's interpolation formula in Lagrangian form, accurate to 6th differences, and continue integrating in steps in x of .04. We use the integrand $e^{-x/2}K_0(x/2)\cos kx$ without splitting it into product functions.

The integral can be evaluated analytically.[1]

$$\int_0^\infty e^{-x/2}K_0(x/2)\cos kx\,dx = 2\cdot\text{real part}\left\{\frac{\cos^{-1}(1-2iz)}{\sqrt{1-(1-2iz)^2}}\right\}.$$

We obtained the following results:[b]

[b] The evaluation of this integral was undertaken in connection with a problem at the IBM Technical Computing Bureau in New York, where we were really interested in the values of a somewhat similar integral which could not be evaluated analytically. The above integral was evaluated as a check on the theory and numerical work in obtaining the d's, which were then used for the actual integral which interested us.

k	*By Numerical Integration*	*Correct Value*
0.0	1.999975	2.00000
0.1	1.989402	1.98963
0.2	1.958604	1.95858
0.4	1.847696	1.84764
0.7	1.616215	1.61902
1.0	1.380436	1.38051
2.0	0.843857	0.84383
4.0	0.437586	0.43749
7.0	0.246750	0.24624

For larger values of k, we have developed special integration coefficients for the function

$$g(x) = [\gamma/\log_e 10 - \log_{10}(x/4)] \cos kx \, .$$

For this choice of $g(x)$ the c's themselves had to be evaluated by numerical integration.

In the application of this method to various problems, I acknowledge with pleasure the assistance of Mr. Liston Tatem and Mr. Daniel Ladd of the Applied Science Department, International Business Machines Corporation.

REFERENCE

1. G. N. WATSON, *Theory of Bessel Functions,* Second Edition (MacMillan Company, 1948), pp. 388.

DISCUSSION

Mr. Bejarano: Can this process be used for triple integrals?

Mr. Sheldon: Yes.

Dr. Buchholz: Is it still profitable to use this method if you must resort to numerical integration to obtain the c coefficients, instead of using analytical methods?

Mr. Sheldon: Yes, that is what we are doing in the case where $g(x) = \cos kx \log x \, dx$. The function can be integrated analytically, but it involves the cosine integral which is a tabulated function. Rather than evaluate the c's analytically, in this case it is simplier to take very fine intervals and evaluate the c_{jk}'s numerically.

Mr. Clamons: I have just one comment to make on this subject. Many times you run into a problem, in which there is a product function like this, in connection with the experimental data where one function is experimental and the other is analytical. It pays to establish the analytical function by means of that integral keeping the g function on the inside and then establishing a matrix deck.

Dr. Hurd: I think this whole discussion illustrates again the fact which we know very well, namely, that computing machines do not replace mathematicians. Here is an instance in which for a given integral, such as Mr. Sheldon has discussed, there are obvious methods of treating them directly by the definition of the integral. If you do this, you have a very slow computing process and the computing machine turns for days. If you apply some mathematical knowledge in advance, you cut the amount of computation time down. I am always glad to think that computing machines make the work of mathematicians more valuable.

*The Use of Orthogonal Polynomials in Curve Fitting and Regression Analysis**

JACK SHERMAN

The Texas Company

IN THE PROBLEM of two-variable curve fitting, a simple polynomial in powers of the independent variable is undoubtedly the most frequently used function in cases where the form is unknown from empirical or theoretical knowledge of the problem. Denoting the independent variable by X and the dependent variable by Y, the polynomial may be written as

$$Y = a_0 + a_1X + a_2X^2 + \dots . \tag{1}$$

The parameters a_0, a_1, a_2, ... are to be evaluated by the method of least squares; i.e., so that the sum of the squares of deviations between the calculated and observed values of Y is a minimum.

For the case that X values are without error and the Y values have equal weightings (precision), the so-called normal equations for the evaluation of the parameters are

$$Na_0 + (\Sigma X\ )a_1 + (\Sigma X^2)a_2 + \dots . = \Sigma Y \tag{2}$$

$$(\Sigma X\ )a_0 + (\Sigma X^2)a_1 + (\Sigma X^3)a_2 + \dots . = \Sigma XY \tag{3}$$

$$(\Sigma X^2)a_0 + (\Sigma X^3)a_1 + (\Sigma X^4)a_2 + \dots . = \Sigma X^2Y \tag{4}$$

$$\begin{matrix} \vdots & \vdots & \vdots & \vdots \end{matrix}$$

In the above equations, the index and limits of summation are omitted, for in each case the summation is carried out over all the values of X and Y—i.e., from 1 to N. A more general treatment of the least squares problem, in which both the dependent and the independent variables may be subject to error and in which the parameters may enter the equations nonlinearly, has been published.[1] The matrix of the coefficients of the parameters in a set of normal equations is always symmetrical about the main diagonal. These equations may be solved by any of the standard methods of solving linear simultaneous, algebraic equations.

The well-known method of curve fitting outlined in the foregoing paragraphs has two undesirable characteristics:

1. The computational labor of solving the normal equations becomes considerable when the number of parameters is large (e.g., greater than 5).
2. If the polynomial of the kth degree is fitted by the method of least squares and it is then decided to add the additional term $a_{k+1}X^{k+1}$, not only must a_{k+1} be evaluated but all the parameters as well, because their values will change by the inclusion of this additional term. Thus, in the usual method of curve fitting, the degree of the polynomial must be decided at the outset.

Orthogonal Polynomial Method[2,3]

Instead of expressing Y as a polynomial directly in powers of X, it may be written more generally as

$$Y = A_0 + A_1\xi_1 + A_2\xi_2 + \dots , \tag{5}$$

in which ξ_1 denotes a linear function of X, ξ_2 a quadratic function of X, etc.

If the ξ's are so chosen that

$$\sum_{k=1}^{N} \xi_i(X_k)\ \xi_j(X_k) = 0 \tag{6}$$

for all values of i and j, except when $i=j$, the polynomials are termed orthogonal. The usefulness of this property of orthogonality in curve fitting lies in the fact that all the coefficients of the parameters in the normal equations, except those on the main diagonal, are zero; thus, a complete separation of the parameters is achieved.

The normal equations then become:

$$NA_0 = \Sigma Y \tag{7}$$

$$(\Sigma\xi_1^2)A_1 = \Sigma\xi_1Y \tag{8}$$

$$(\Sigma\xi_2^2)A_2 = \Sigma\xi_2Y \tag{9}$$

$$\begin{matrix} \vdots & \vdots \end{matrix}$$

*This paper was presented informally.

For the case that the values of X are equally spaced, the orthogonal polynomials can be easily derived.[4] It is this case which will be discussed below.

In practice, it is more convenient to deal with the equation in the form

$$Y = A'_0 + A'_1\xi'_1 + A'_2\xi'_2 + \ldots\ldots , \qquad (10)$$

in which the ξ' polynomials are related to the ξ polynomials by means of equation

$$\xi'_i = \lambda_i\xi_i \; . \qquad (11)$$

The λ_i values are so chosen that the ξ' values are integers reduced to the lowest terms. The relationships between ξ' and X up to the fifth degree are:

$$\xi'_1 = \lambda_1\,(X-\bar{X})\,, \qquad (12)$$

$$\xi'_2 = \lambda_2\left[(X-\bar{X})^2 - \frac{N^2-1}{12}\right], \qquad (13)$$

$$\xi'_3 = \lambda_3\left[(X-\bar{X})^3 - (X-\bar{X})\left(\frac{3N^2-7}{20}\right)\right], \qquad (14)$$

$$\xi'_4 = \lambda_4\left[(X-\bar{X})^4 - (X-\bar{X})^2\left(\frac{3N^2-13}{14}\right) + \frac{3(N^2-1)(N^2-9)}{560}\right] \qquad (15)$$

$$\xi'_5 = \lambda_5\left[(X-\bar{X})^5 - (X-\bar{X})^3\left(\frac{5(N^2-7)}{18}\right) + (X-\bar{X})\,\frac{15N^4 - 230N^2 + 407}{1008}\right]. \qquad (16)$$

These equations enable one to transform the equation in the ξ's to an equation directly in terms of X, if so desired. However, the sum of the squares of the deviations from the regression equation may be obtained directly in terms of ξ's.

The equation is

$$\Sigma\Delta^2 Y = \Sigma Y^2 - A'_0\Sigma(Y) - A'_1\Sigma(Y\xi'_1) - A'_2\Sigma(Y\xi'_2) - \ldots\,.$$

The first term, ΣY^2, is the sum of the squares of the Y values about zero. The second term, $A'_0\Sigma(Y)$, is the amount by which the sum of squares about zero is decreased when taken about the mean. The third term, $A'_1\Sigma(Y\xi'_1)$, gives the amount by which the sum of the squares is further decreased when the deviations are taken about the best (least squares) straight line, etc.

Illustrative Numerical Example

As a simple numerical example illustrating the application of orthogonal polynomials to the determination of a regression equation, consider the X and Y values given in the first two columns of the table which follows:

X	Y	ξ'_1	ξ'_2
4	1.82	-3	5
5	6.13	-2	0
6	12.09	-1	-3
7	19.47	0	-4
8	29.80	1	-3
9	42.12	2	0
10	55.91	3	5
	$\Sigma\xi_i^2 =$	28	84

If the X and Y values are fitted to a polynomial

$$Y = A + BX + CX^2\,, \qquad (18)$$

by the standard method of least squares the normal equations are:

$$7A + 49B + 37C = 167.34 \qquad (19)$$

$$49A + 371B + 2989C = 1423.34 \qquad (20)$$

$$371A + 2989B + 25235C = 12481.56\,. \qquad (21)$$

The solution of these equations yields the polynomial

$$Y = 6.50494 - 5.18474X + 1.01309X^2\,. \qquad (22)$$

To obtain this same result by the application of orthogonal polynomials, the following quantities have been calculated:

$$\frac{\Sigma Y}{N} = 23.90571 \qquad (23)$$

$$\Sigma(\xi'_1 Y) = 251.96 \qquad (24)$$

$$\Sigma(\xi'_2 Y) = 85.1 \qquad (25)$$

Dividing the latter two steps by the number standing at the bottom of the corresponding ξ' columns, the following parameters are obtained:

$$A'_1 = \frac{\Sigma(\xi'_1 Y)}{\Sigma\xi'^2_1} = \frac{251.96}{28} = 8.99857\,, \qquad (26)$$

$$A'_2 = \frac{\Sigma(\xi'_2 Y)}{\Sigma\xi'^2_2} = \frac{85.1}{84} = 1.01309\,. \qquad (27)$$

The corresponding orthogonal polynomial expression for Y is

$$Y = \bar{Y} + A'_1\xi'_1 + A'_2\xi'_2 \qquad (28)$$

$$= 23.90571 + 8.899857\,\xi'_1 + 1.01309\,\xi'_2\,.$$

To transform the above equation in ξ' into an equation in X, the table of relationships previously given is used. (In this example λ_1 and λ_2 are equal to 1, so that $\xi'_1 = \xi_1$ and $\xi'_2 = \xi_2$. Also, 7 is the mean value of X.)

$$Y = 23.90571 + 8.99857(X-7) + 1.01309\,[(X-7)^2-4] \qquad (29)$$

$$= 19.85333 + 8.99857(X-7) + 1.01309\,(X-7)^2.$$

The above equation in powers of $X-7$ may be converted to an equation in powers of X by a method employing synthetic division. The result is

$$Y = 1.01309X^2 - 5.18476X + 6.50499\,. \qquad (30)$$

A comparison of equations (22) and (30) shows that they are identical to within a few units in the last figure.

In order to obtain the calculated values of Y or to carry out a regression or correlation analysis, it is not necessary to convert the equation in ξ' to an equation in X. The details need not be given here.[2]

Orthogonal polynomials may be used very efficiently to obtain an nth degree polynomial corresponding to a tabulation of $N-1$ equally-spaced values of Y.

REFERENCES

1. W. E. DEMING, *Statistical Adjustment of Data* (New York: Wiley, 1943).
2. "Tables of Orthogonal Polynomial Values Extended to $N=104$," Iowa State College of Agriculture, *Research Bulletin 297* (April, 1942).
3. R. A. FISHER and F. YATES, *Statistical Tables for Biological, Agricultural and Medical Research* (Oliver and Boyd, Ltd., 1943).
4. W. E. MILNE, *Numerical Calculus* (Princeton University Press, 1949).

DISCUSSION

Mr. Brown: As Dr. Sherman remarked, these tabular polynomials exist in convenient form only when the points are equally spaced. Experimentally, and on a small scale, I have tried calculating them and using them in a few cases when the points were not equally spaced. In that case, of course, the usual procedure is to form the normal equations of B^2 and go ahead and solve them. I have tried this alternative method which consists in calculating, for the particular set of given points, as many of the orthogonal polynomials as are needed for the particular problem. I am not sure about the relative amount of work involved in the conventional procedure, and in this procedure, but I am favorably impressed by what seemed at first glance a remarkably laborious task, namely, calculating a whole set of functions before you start your problem. I think the amount of work is about equal by the two methods.

The method has a great advantage. If you want to fit more than one function to the same set of points, you may perhaps have eight pages of work for the first function, but seven of those pages consist in calculating orthogonal polynomials, and the eighth consists of evaluating the coefficients. If I now want to fit a second function, there is one more page, not eight more. If it were calculated by the conventional method, that would be a great deal of calculation, but, on the whole, I am impressed with that way of doing it. I think it merits further investigation. I don't have any final evaluation of the two methods.

In some cases two variables may have almost the same effect on the variable, and they vary closely with each other, in which case there will be difficulty in solving because of an almost vanishing determinant. The two variables are so closely correlated that there may be some indeterminateness in determining these coefficients. On the other hand, certain variables may have so little effect on the variable Y that they can probably be left out. I think it would be interesting to investigate the usefulness of orthogonalizing such functions.

*General Purpose Ten-Digit Arithmetic on the IBM Card-Programmed Electronic Calculator**

STUART R. BRINKLEY, JR. G. L. WAGNER R. W. SMITH, JR.

U. S. Bureau of Mines

THE IBM Card-Programmed Electronic Calculator makes available to the computation laboratory of moderate size a computer of considerable flexibility. By means of appropriately designed control panels, it is possible to exploit the facilities for control provided by the accounting machine, and thus to employ the equipment for the efficient solution of a wide variety of problems.

We may conceive of two diametrically opposed methods for the utilization of the card-programmed electronic calculator. The control panels may be wired in such a manner as to permit the computer to perform the elementary arithmetic operations under control by suitable code punching in program cards which are read by the accounting machine. A given computational routine is then performed by reducing the problem to a series of arithmetical operations to be performed in sequence, and the program for the problem consists of the sequence of coded instructions punched into cards, by means of which the computer is given step-by-step instructions for carrying out the desired routine. When employed in this manner, the computer may be said to be a general-purpose computer. On the other hand, it is frequently possible to control simple sub-routines by means of control panel wiring, and it may be possible to generate some or all of the program within the machine instead of employing card control. In the extreme case, the computer may then be a single-purpose computer and may be said to be controlled by internal program.

Substantial savings of operating time and greatly increased efficiency of operation frequently result from the specialization of the control circuits to a particular problem. In this way the substantial facilities for control and storage provided by the accounting machine and the subsidiary storage unit can be employed in conjunction with the electronic calculator for the rapid solution of a particular problem. However, a considerable amount of time is normally required for the design and installation of the required control panels. This type of operation is, therefore, not usually justified unless the problem is of considerable magnitude.

Although the advantages to be gained by specialization of control panels should be examined for any particular problem, it will usually be the case that the average computation of moderate duration can be most expediently carried out on a general-purpose computer. In this way, the design and installation of control panels for the problem at hand is entirely eliminated, and the planning for the problem consists only of the formulation of the card program routine.

In view of the fact that the general-purpose computer is to be used in a wide variety of problems, it is desirable that it offer to the programmer the greatest possible flexibility consistent with the logical design of the equipment. It is desirable that it be possible to carry out the arithmetical operations with the minimum number of program cards. Furthermore, it is desired that facilities be provided for the automatic selection of alternative computational routines and for the selection of input data.

The electronic storage capacity of the IBM Type 604 Electronic Calculating Punch is assigned to provide four storage groups for ten-digit numbers. The units connected to channel A are designated electronic storage A. Similarly, electronic storage B and electronic storage C consist of the units connected to channel B and channel C, respectively. Auxiliary temporary storage is provided by additional units and is designated electronic storage D.

Provision is made for the eight arithmetic operations summarized in Table I. These consist of the elementary operations, square root, and certain combinations.

TABLE I. ARITHMETIC ORDERS

		Storage Result				
Code	*Order*	*A*	*B*	*C*	*D*	*Decimal Point*
0	$A = C$	C	—	C	D	$c = a$
1	$A + B = C$	C	B	C	D	$c = a = b$
2	$A - B = C$	C	B	C	D	$c = a = b$
3	$A \times B = C$	C	B	C	D	$c = a + b$
4	$A \div B = C$	C	B	C	A	$c = a - b + 1$
5	$\sqrt{A} = C$	C	$C/10$	C	A	$c = a/2$
6	$A + B + D = C$	C	B	C	D	$d = c = a = b$
7	$A - B + D = C$	C	B	C	D	$d = c = a = b$
8	$A \times B + D = C$	C	B	C	D	$d = c = a + b$

*This paper was presented informally. The supplementary notes were added subsequent to the presentation.

The table gives the condition of each of the electronic storage groups at the conclusion of the operation. The factors are identified by the label of the electronic storage group in which they are stored or by the channel over which they are transmitted. The decimal point rule for each operation is also given, where *a* denotes the number of decades between the leftmost decade of the number *A* and the decimal point, and *b*, *c*, and *d* are similarly defined for the numbers *B*, *C*, and *D*, respectively. The order code zero results in a transfer of the factor *A* to electronic storage C.

The result *C* of every operation is transferred to electronic storage A, where it may be employed as a factor in the succeeding operation. The factor *B* is preserved by every operation and may be employed in succeeding operations unless the square root operation intervenes. The number in electronic storage *D* is unaffected by all operations except division and square root, in which cases it is replaced by the factor *A*.

The provision for the retention through a sequence of operations of significant factors within the electronic storage of the computer contributes substantially to the efficiency and flexibility of the computer. It is frequently possible to carry out routines of considerable length before sending results to external storage. These provisions effectively increase the over-all storage capacity of the computer, a consideration that may be significant in problems for which storage capacity is at a premium.

The square root operation utilizes an iterative sub-routine, programmed on the 604 control panel. A special method of obtaining the first trial value of the root has been employed so that the complete operation requires a maximum time of two machine cycles.

In addition to the transfer of the result *C* to electronic storage A for every operation, provision has been made for three optional transfers between the different electronic storage groups. These transfer orders follow the arithmetic operations. They control access to electronic storage D and supplement the provisions of the arithmetic operations for the destination of the operation result. The optional transfer orders are summarized in Table II. Except as noted, they can accompany any of the arithmetic orders.

TABLE II. TRANSFER ORDERS

		Storage Result				
Code	*Order*	*A*	*B*	*C*	*D*	*Notes*
6	*D*→*B*	*C*	*D*	*C*	*D*	not with arith. orders 4, 5, 6, 7, 8
7	*C*→*B*	*C*	*C*	*C*	*D*	not with arith. order 5
8	*C*→*D*	*C*	*B*	*C*	*C*	not with arith. order 5

The optional transfer orders contribute significantly to the flexibility and efficiency of the computer.

The card-programmed calculator is regularly equipped with a shift unit that is associated with channel C. This unit makes it possible to shift a result *C* up to five decades to the left. The ease with which various problems can be programmed is markedly increased if provisions are made for shifting on channel A of the factor *A* before it enters the computing unit. The field selector of the accounting machine can be employed to provide a shift to the left of up to four decades, or a shift to the right of up to five decades. The shift orders are summarized in Table III.

TABLE III. SHIFT ORDERS

Code Channel A	*Result*	*Code Channel C*	*Result*
0	A	0	C
1	$10A$	1	$10C$
2	10^2A	2	10^2C
3	10^3A	3	10^3C
4	10^4A	4	10^4C
5	$10^{-1}A$	5	10^5C
6	$10^{-2}A$		
7	$10^{-3}A$		
8	$10^{-4}A$		
9	$10^{-5}A$		

The storage addresses of factors *A* and *B* and the address of the storage unit to which the result *C* is to be sent, together with operation and transfer codes and shifting instructions, constitute a ten-digit order code or word that normally is read by the accounting machine from columns 4 to 13 of an order card. The location of the various parts of the complete order is shown in Figure 1. Input to the computer is achieved either by entry from the order card into channels A and B under control of a 00 address for the channel involved, or by an order termed spread load (SL) under X control, permitting direct entry from the order card into six of the seven accounting machine counter groups. The card fields associated with each counter group during a spread load operation are indicated in Figure 1. The punch unit is wired to punch from certain accounting machine counter and electronic storage groups as shown in Figure 1. In both input and output, negative numbers are punched as true figures with an X in the units position of the card field, and sign control within the computer is automatic.

The provisions that have been described for the arithmetic, shift, and transfer operations are sufficient to permit the construction of efficient programs for computational routines involving purely algebraic operations. However, unless further implemented by provisions for program control, it is necessary that the machine operator perform all of the discriminatory operations involved in the solution of the problem at hand. In many problems—as, for example, those of iterative nature—it is necessary to provide alternative

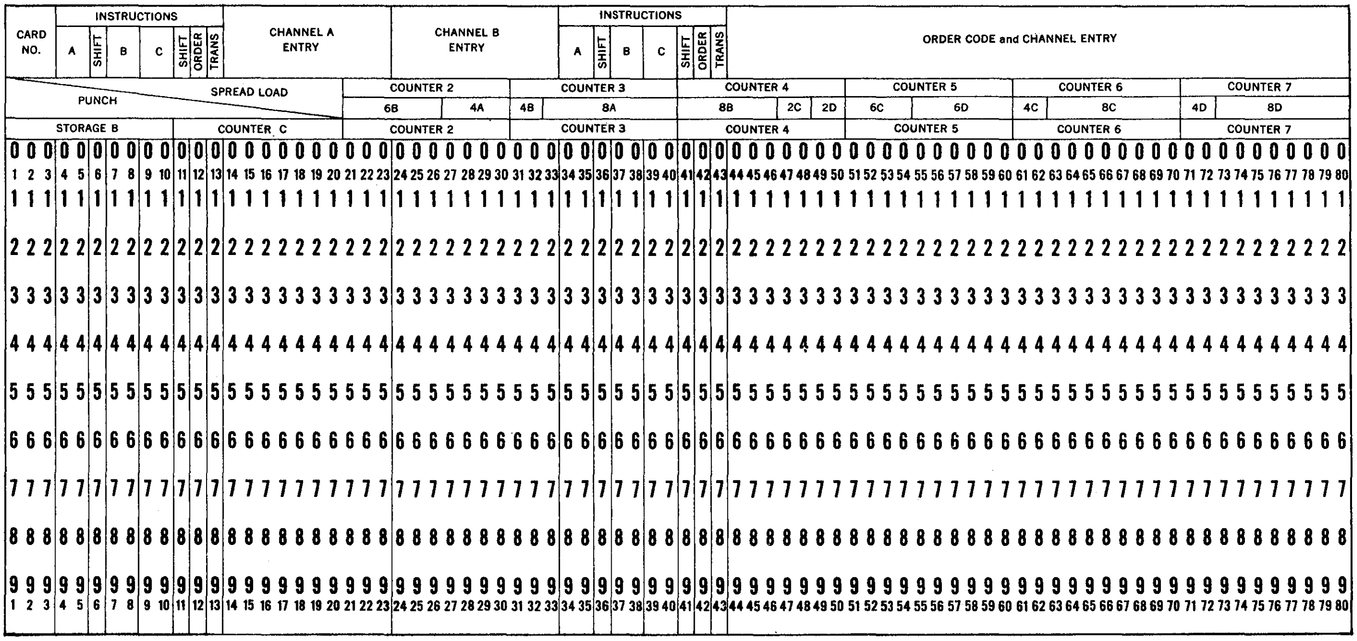


FIGURE 1. INSTRUCTION CARD

computational routines, and at some point in the calculation to select the appropriate routine on the basis of criteria developed by the calculation itself.

A considerable amount of automatic program control can be achieved by the addition of two program control orders. The ten-digit instruction word is normally contained in columns 4 to 13 of the order card, and this field is designated the normal instruction field. An alternative field, called the transfer instruction field, is located (Figure 1) in columns 34 to 43 of the order card. If, at some point in a calculation, the number occupying electronic storage C is negative, and if "conditional transfer" (CT) is ordered at this point by an X punch in the order card, the computer will transfer from the normal instruction field to the transfer instruction field, and it will obtain its instructions from the transfer field as long as "hold transfer" (HT) is consecutively ordered by another X punch in the order card. If the number occupying electronic storage C is positive at the time of the conditional transfer order, the computer continues to obtain its instructions from the normal instruction field. Alternative routines can be programmed in the normal and transfer fields, and the discrimination between the two routines can always be phrased in such a way as to involve the sign of an appropriate quantity.

If a given computation is to be performed for a series of different values of initially given parameters, it is necessary at each stage of the calculation to select the input parameters to be employed. It is frequently convenient to file with the order cards the input data necessary for a considerable number of individual problems and to provide a means for the selection of that portion of the data required for a particular calculation. Provision is made for a selective loading operation, termed "selective spread load" (SSL), controlled by an X punch in the order card. A load address, stored in the right-hand half of electronic storage D is compared with a load argument, punched in columns 14 to 18 of the order card. If the two numbers are equal, the X punch controls direct entry into the accounting machine counter groups 2 to 7 in the same manner as in the ordinary spread load operation. If the two numbers are not equal, the card passes through the machine without activity. This order may also be employed in table lookup operations for the introduction into otherwise algebraic routines of transcendental functions, the direct calculation of which cannot be expeditiously programmed. The program controls are summarized in Table IV.

TABLE IV. PROGRAM CONTROLS

Punch	*Column*	*Order*	*Name*
X	2	CT	conditional transfer
X	3	HT	hold transfer
X	5 or 35	SL	spread load
X	6 or 36	SSL	selective spread load
9	12 or 42	P	punch
X	12 or 42	L	list

The CPC has been employed in this laboratory in the manner described on a number of different computational problems. The facilities for control have been found to be adequate for the efficient operation on computation routines of considerable complexity. We have also developed a set of control panels that are completely analogous, except that operations are limited to eight-digit arithmetic, and the

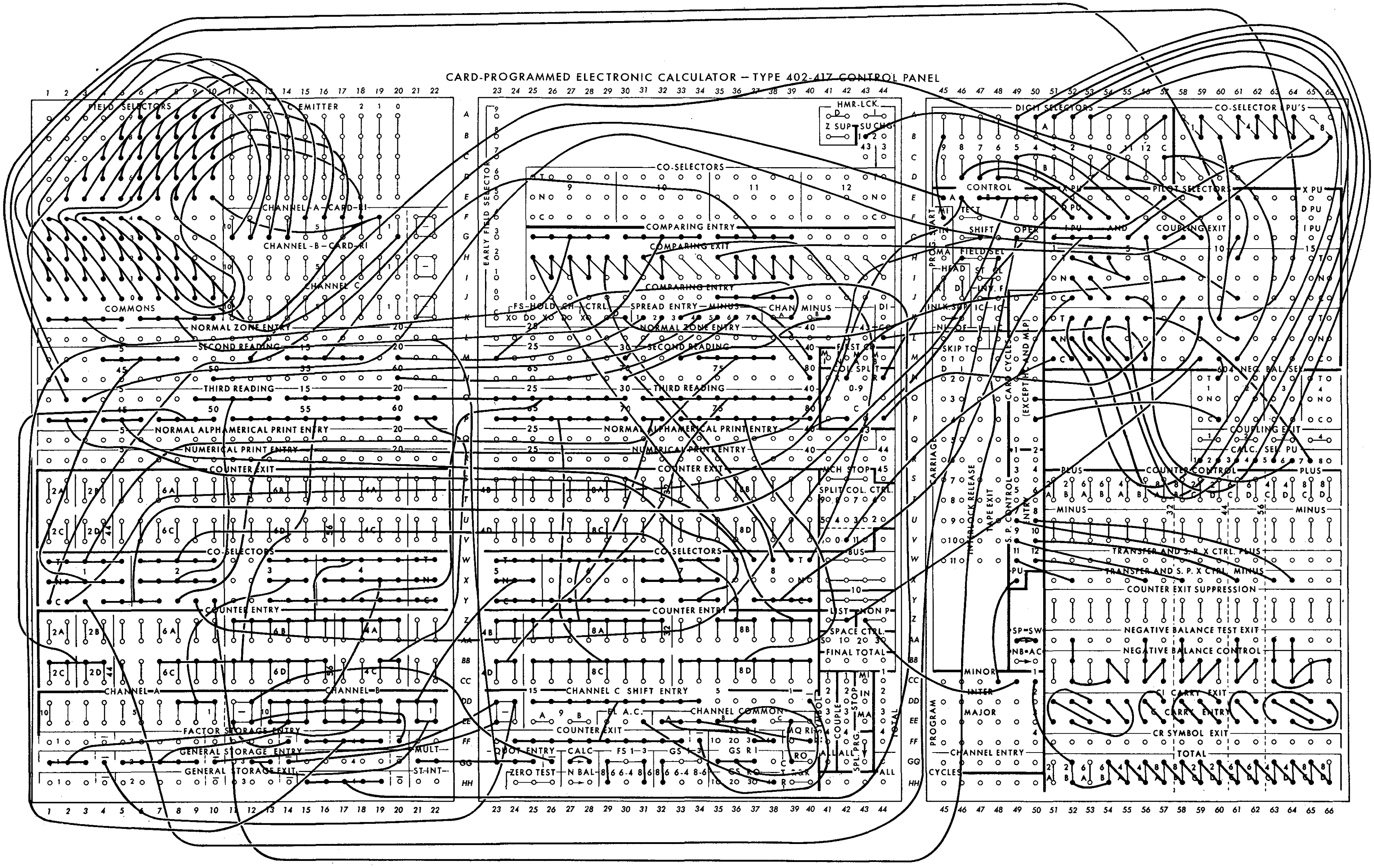


Figure 2. Control Panel for 417 Accounting Machine

square-root order is eliminated. In this way it is possible to preserve enough program capacity in the electronic calculator for internal programming on the control panel of the 604 unit for an iterative sub-routine for the calculating of transcendental functions such as the logarithm, exponential, trigonometric or hyperbolic trigonometric functions.

We append notes on the various operations and list the sixty program steps for the general purpose ten-digit arithmetic control panel. A diagram (Figure 2) is given for the wiring of the accounting machine control panel, and it is supplemented with diagrams presenting, with greater clarity, calculate selector (Figure 3), field selector (Figure 4), co-selector (Figure 5), and pilot selector (Figure 6) circuits. These notes are intended to supplement the material presented in the standard instruction manuals for the various machines.

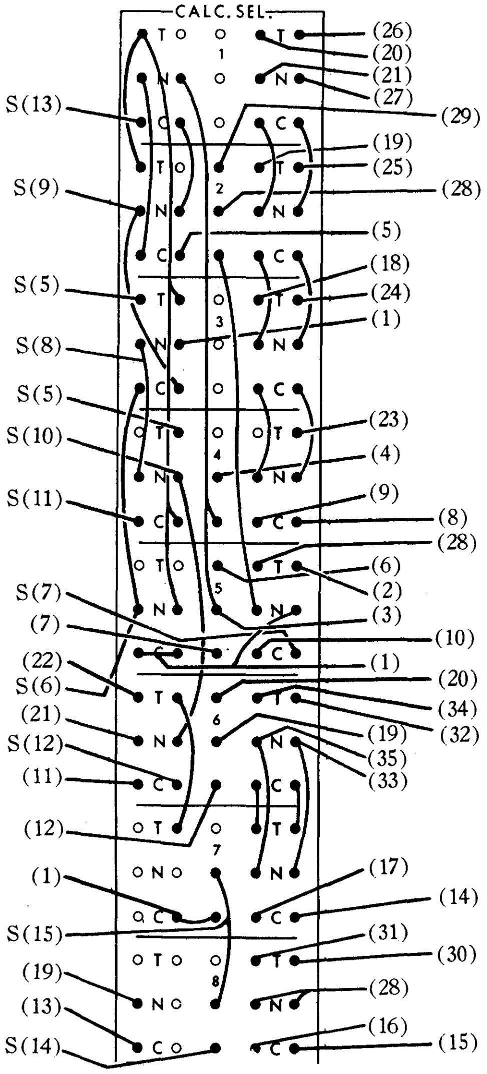
Notes

Addition: On order 1, the operation $A + B = C$ is performed. On order 6, the operation $A + B + D = C$ is performed.

Subtraction: On order 2, the operation $A - B = C$ is performed. On order 7, the operation $A - B + D = C$ is performed.

Multiplication: On order 3, the operation $A \times B = C$ is performed. On order 8, the operation $A \times B + D = C$ is performed. The product is rounded off with 1/2 adjustment to ten significant figures.

Division: On order 4, the operation $A \div B = C$ is performed. The formula $C = (A - A_1B_2/B_1)/B_1$ is employed,



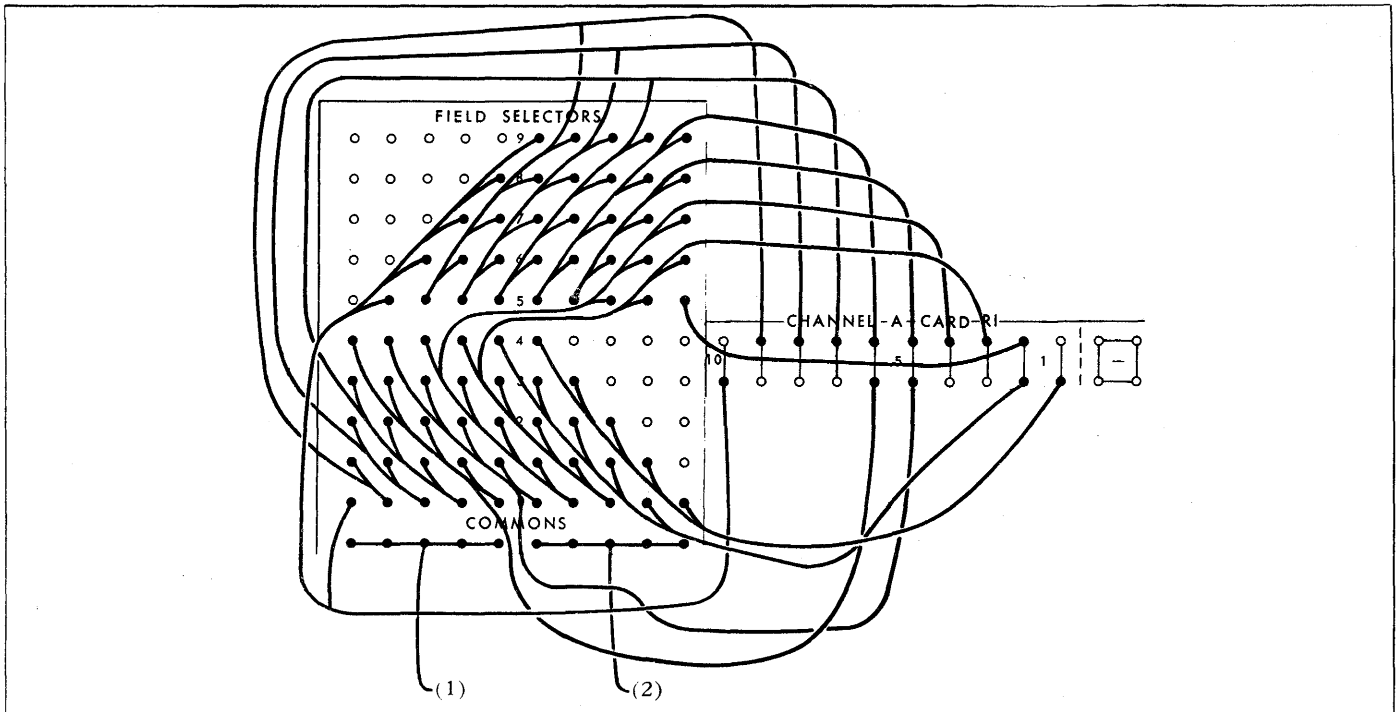
LEGEND

(1) sup. without test
(2) sup. on zero
(3) PU group sup. 1
(4) PU group sup. 2
(5) PU group sup. 3
(6) drop-out group sup. 4
(7) program 3
(8) program 10
(9) program 13
(10) program 15, 16
(11) program 41
(12) program 42
(13) program 52
(14) program 52
(15) program 52
(16) program 53
(17) program 53, 55
(18) read units into 2nd
(19) read units into 3rd
(20) read units into 4th
(21) read units into 5th
(22) read units into 6th
(23) read units out of 2nd
(24) read units out of 3rd
(25) read units out of 4th
(26) read units out of 5th
(27) read units out of 6th
(28) counter read in +
(29) counter read in −
(30) RI factor stor. 1 and 2
(31) RI general stor. 4
(32) RO factor stor. 1 and 2
(33) RO factor stor. 3 and 4
(34) RO general stor. 4
(35) RO mult. quot.

SUPPRESSION CIRCUITS

Group Suppression Exits		*Other Suppression Circuits*	
Suppression	*Program Lines*	*Suppression*	*Program Lines*
GS(1)	10-14	S (5)	6, 7,15,16
GS(2)	4- 5	S (6)	17,19,20,22,24,28,31,32
GS(3)	8- 9	S (7)	21
GS(4)	18,23	S (8)	25,29,30
		S (9)	26,27,37-39
		S(10)	33-35,40,46-50
		S(11)	36
		S(12)	41,42
		S(13)	43-45
		S(14)	52,53
		S(15)	54
		S(16)	55

FIGURE 3. CALCULATE SELECTOR CIRCUITS OF 604 ELECTRONIC CALCULATING UNIT

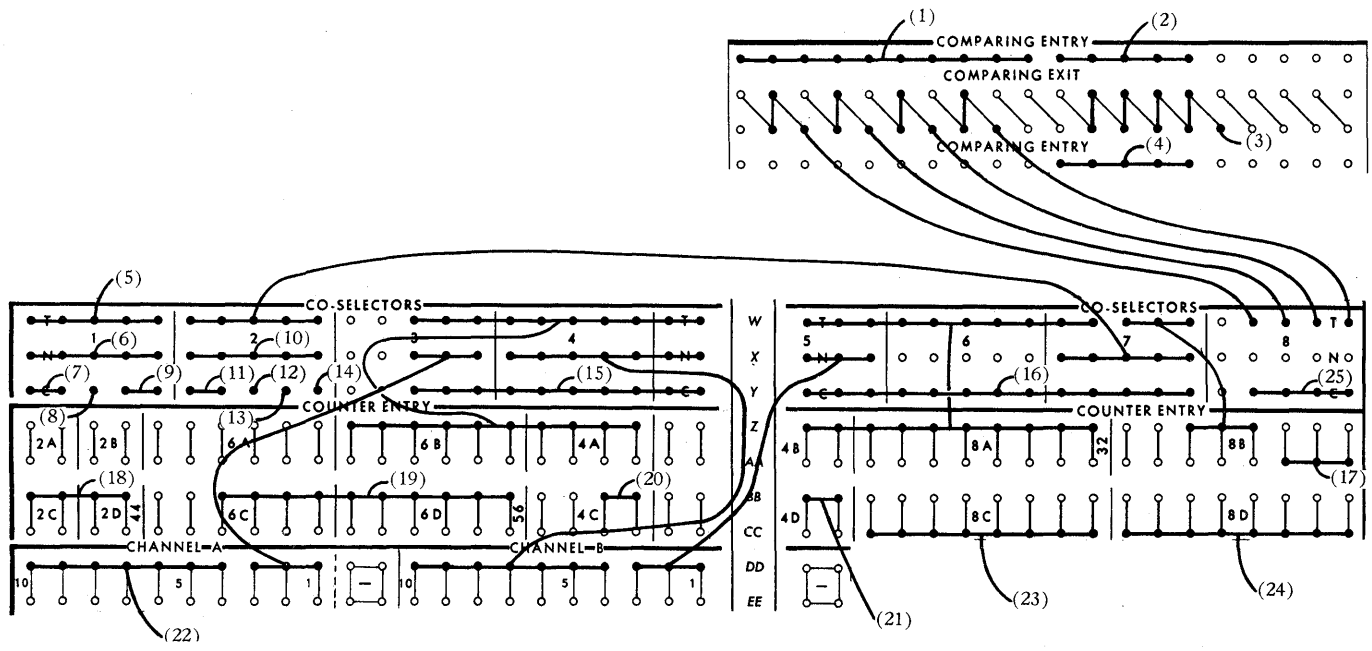


LEGEND

(1) Comparing Entry 1-5, Factor Storage 4 Entry (Split)

(2) Comparing Entry 6-10, Multiplier-Quotient Entry (Split)

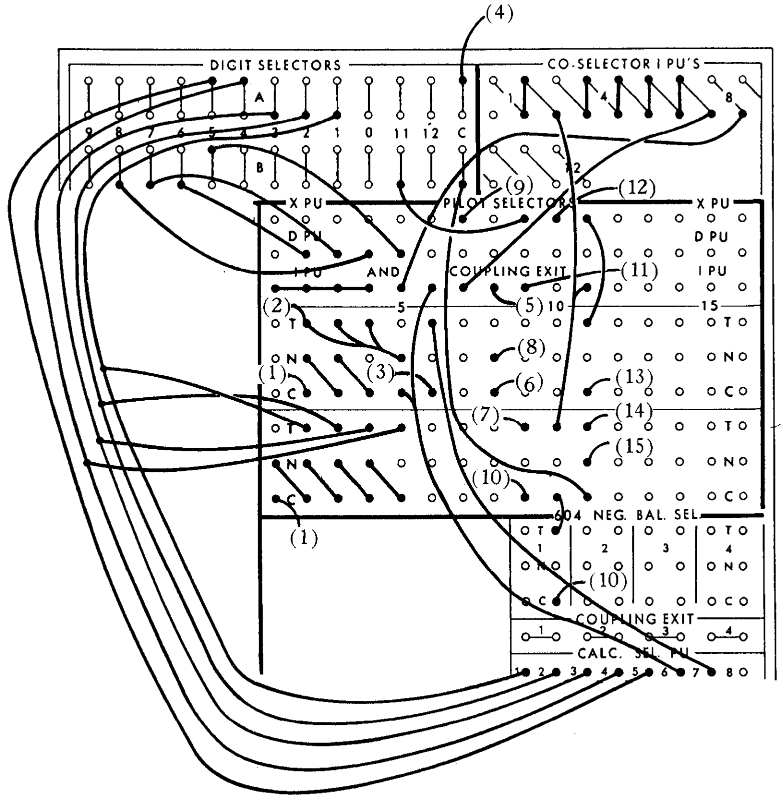
FIGURE 4. FIELD SELECTOR CIRCUITS OF 417 ACCOUNTING MACHINE



LEGEND

(1) field selector common
(2) 2nd reading, cols. 14-18
(3) pilot selector 8, IPU
(4) general storage 4, exit
(5) 2nd reading, cols. 34-38
(6) 2nd reading, cols. 4-8
(7) channel A control
(8) field selector, PU
(9) channel 13 control
(10) 3rd reading, cols. 9-13
(11) channel C control
(12) shift control
(13) digit selector 1, com.
(14) operation control
(15) 3rd reading, cols. 21-30
(16) 3rd reading, cols. 31-43
(17) 3rd reading, cols. 44-46
(18) 3rd reading, cols. 47-50
(19) 3rd reading, cols. 51-60
(20) 3rd reading, cols. 61-62
(21) 3rd reading, cols. 71-72
(22) 3rd reading, cols. 14-20
(23) 3rd reading, cols. 63-70
(24) 3rd reading, cols. 73-80
(25) pilot selectors 1-4, IPU

FIGURE 5. COSELECTOR CIRCUITS OF 417 ACCOUNTING MACHINE



LEGEND

(1) split col. control 0
(2) Co-selector 8, com.
(3) card count
(4) Co-selector 2, com.
(5) comparing exit
(6) field selector PU
(7) list
(8) spread RI through col. split
(9) spread RI and units position channel A control through col. split (split)
(10) card cycles
(11) set up change 1
(12) 3rd reading, col. 2
(13) 3rd reading, col. 3
(14) 3rd reading, col. 42
(15) 2nd reading, col. 12

FIGURE 6. PILOT SELECTOR CIRCUITS OF 417 ACCOUNTING MACHINE

where $A_1 = A\ (10-6)$, $B_1 = B\ (10-3)$, and $B_2 = B\ (2-1)$. The error is less than $(A_1/B_1)\ (B_2/B_1)^2$. The leftmost non-zero digit of B may not be more than one decade to the right of the leftmost non-zero digit of A.

Square Root: On order 5, the operation $\sqrt{A} = C$ is performed. The formula

$$C^{(n)} = [C^{(n-1)} + A/C^{(n-1)}]/2$$

is employed, where $C^{(n)}$ is the nth approximation to the square root. The successive approximations are continued until for some n, say n', $C^{(n')} = C^{(n'+1)}$. The zeroth approximation is obtained from the empirical rule,

$$C^{(0)} = 2(10^{k-10+a/2}) + A(10^{k-1+a/2}) \ ,$$

where the number of non-zero digit pairs of the 10-digit number A is $k+4$, and where a denotes the location of the decimal point of A, counting from the leftmost decade of A. The quantity A must be transmitted over channel A at the time square root is ordered. It is also essential that a be an even number.

Transfers: The $C \rightarrow A$ transfer accompanies every arithmetic order. The $C \rightarrow B$ and $C \rightarrow D$ transfers can be ordered with every arithmetic order except square root. The $D \rightarrow B$ transfer can accompany simple addition, subtraction, multiplication, and division (orders 1, 2, 3, 4) only.

Spread Load: Counters 2 to 7 read in from card fields. Counter 1 can be loaded simultaneously by card entry into channel A if channel C is given a 71 address.

Selective Spread Load: Counters 2 to 7 read in from card fields of the card for which the argument (cols. 14-18) is equal to a card address stored in electronic storage $D_2 = D\ (5-1)$. All cards for which the argument and card address are non-equal are passed through the machine without reading to the counters.

Conditional Transfer: The calculator transfers to the transfer instruction field instead of the normal instruction field if $C < 0$ when conditional transfer is ordered. Hold transfer must also be ordered.

Hold Transfer: The calculator reads instructions from the transfer instruction field instead of the normal instruction field if transfer has occurred as a result of conditional transfer, as long as hold transfer is continuously ordered. The first transfer instruction word must be blank, and the first normal instruction word after transfer must be blank. If either normal or transfer instructions call for entry into channel A or B from card fields, the opposing instruction word must be blank.

TYPE 604 PROGRAM STEPS FOR TEN-DIGIT ARITHMETIC

			Factor Storage				*General Storage*			*Activity*				
No.	*Sup*	*Sel*	*1(8-6)* *2*	*3(8-6)* *4*	*Mult Quot*	*Ctr*	*1(8-6)* *2*	*3*	*4*	+	−	×	÷	√
1	0		(PUGS 4)			RO,RC				×	×	×	×	×
2	0					RI(+)2		RO		×	×	×	×	×
3	0					RO,RC		RI		×	×	×	×	×
4	GS2		R14	RO									×	×
5	GS2		(PUGS 4)		RO				RI				×	×
6	5			RO		RI(+)4				×	×			×
7	5			RI		RO,RC				×	×			×
8	GS3			RO		RI(+)6				×	×			×
9	GS3				RO	RI(+)4				×	×			×
10	GS1	1N		RI		RO6,RC								×
		1T		RI		RO5,RC								×
		2T		RI		RO4,RC								×
		3T		RI		RO3,RC								×
		4T		RI		RO2,RC								×
11	GS1			RO		RI(+)2								×
12	GS1		(EMITTER 4)		RI5									×
13	GS1	1N			RO	RI(+)5								×
		1T			RO	RI(+)4								×
		2T			RO	RI(+)3								×
		3T			RO	RI(+)2								×
		4T			RO	RI(+)								×

(CONTINUED ON NEXT PAGE)

Type 604 Program Steps for Ten-Digit Arithmetic—*Continued*

			Factor Storage				*General Storage*			*Activity*				
No.	*Sup*	*Sel*	*3(8-6)* *2*	*3(8-6)* *2*	*Mult* *Quot*	*Ctr*	*1(8-6)* *2*	*3*	*4*	+	−	×	÷	√
14	GS1					RO3,RC	RI							×
15	5	2N 5N				RI(+)6	RO			×				
		2T 5N				RI(−)6	RO				×			
		5T				RI(+)6	RO							×
16	5	2N 5N				RI(+)3		RO		×				
		2T 5N				RI(−)3		RO			×			
		5T				RI(+)3		RO						×
17	6			RO		RI(−)5								×
18	GS4				RO	RI(−)3								×
19	6					RO3		RI						×
20	6					RO6	RI							×
21	7		(PROGRAM REPEAT) (ZERO TEST)			R16								×
22	6					RO,RC								×
23	GS4				RO	RI(+)2								×
24	6		(EMITTER 5)		RI4									×
25	8					MULT(+)	RO					×		×
26	9			RO	RI							×		
27	9			RI		RO6,RC						×		
28	6		(EMITTER 5)		RI									×
29	8					MULT(+)		RO				×		×
30	8			RO		RI(+)4						×		×
31	6					RO2		RI						×
32	6					RO5,RC	RI							×
33	10		(PUGS 1)		RI			RO					×	×
34	10		RO			MULT(+)							×	×
35	10		(PUGS 3)		DIV		RO						×	×
36	11			RI		RO4,RC						×	×	×
37	9					MULT(+)	RO					×		

(CONTINUED ON NEXT PAGE)

TYPE 604 PROGRAM STEPS FOR TEN-DIGIT ARITHMETIC—*Continued*

			Factor Storage				*General Storage*			*Activity*				
No.	*Sup*	*Sel*	*3(8-6) 2*	*3(8-6) 2*	*Mult Quot*	*Ctr*	*1(8-6) 2*	*3*	*4*	+	−	×	÷	√
38	9			RO		RI(+)						×		
39	9		(½ ADJUST)			RI3						×		
40	10				RO	RI(−)2							×	×
41	12	6N	RO			RI(+)5							×	×
		6T	RO			RI(+)6				C+D				
42	12	6N				RI(+)3			RO				×	×
		6T				RI(+)4			RO	C+D				
43	13				RI	RO4				×	×	×		
44	13				RO	RI(−)4				×	×	×		
45	13			RI		RO6,RC				×	×	×		
46	10				DIV		RO						×	×
47	10			RI		RO,RC	(PUGS 2)						×	×
48	10			RO		RI(+)6							×	×
49	10			RI4	RO								×	×
50	10				DIV		RO						×	×
51	0					RO,RC				×	×	×	×	×
52	14	6N 7T 8N		RO		RI(+)3				C→B				
		6N 7N 8T	RI	RO						C→D				
		6T 7T 8N	RO			RI(+)3				D→B				
53	14	6N 7T 8N			RO	RI(+)				C→B				
		6N 7N 8T			RO				RI	C→D				
		6T 7T 8N				RI(+)			RO	D→B				
54	15					RO3,RC	RI			C→B,D→B				
55	16	6N 7T			RO			RI2		C→B				
		6T 7T						RI2	RO	D→B				
56	0					RI(+)		RO						
57	0					RO2,RC		RI						
58	0			RO		RI(+)3								
59	0			RI		RO6								
60	0		(PU NEG BAL SELECTOR)		RO	RI(+)								

*Remarks on Distillation Calculations**

JOHN W. DONNELL

Michigan State College

THE WORK we have been doing with IBM calculating machines has been in distillation and absorption tower design.

First, I should like to say a few words regarding Mr. Opler's paper[1] on distillation and the previous paper by Dr. Rose[2] and his associates, to avoid repetition in describing our work which covers similar ground to that of these two excellent papers. It should be pointed out that where Mr. Opler assumed Raoult's law to hold, he is using the equation $y = PX/\pi$ to describe the relationship between the concentration of a given component in a boiling liquid and the concentration of that component in the vapor in equilibrium with the boiling liquid. In this equation, π is the total pressure on the system, P is the vapor pressure of the component in question in the pure state at the temperature in question, and x and y are the mol fractions, respectively, of the component in the liquid and vapor states.

In our work, especially in petroleum engineering, we have discovered that such an equation is not accurate enough for our design work, and the petroleum industry has quite universally adapted the equation $y = KX$ where, as you can see, P/π has been replaced by K, an experimentally determined factor, which is a function of temperature, pressure, and the component itself. These constants have been determined at great expense to the industry, and I think the industry should be called upon to pool its K factors. I suggest that IBM make a master deck of cards containing these values. In our work we are using K factors in both distillation and absorption calculations. In Dr. Rose's paper a symbol designated as α is, in reality, a ratio of K values; e.g., $\alpha_{1-2} = K_1/K_2$. While this varies much less with temperature than the individual K values, we in the petroleum industry cannot use this approximation in most design work, and must use the individual K factors and make no attempt to eliminate variations with temperature in either distillation or absorption columns.

In addition I should like to point out the similarity, in the approach to the distillation problem, of the method we are using to that of Dr. Rose. You recall his calculations concern themselves only with binary mixtures. He calculates the analysis of the distillate and bottoms products for a given feed, reflux ratio, and number of plates, by first estimating the total amount of distillate per unit feed and then calculating a new distillate quantity closer to the correct answer. With this new distillate quantity as a second estimate, he calculates a still closer answer and continues the iteration until the desired accuracy is obtained. We do a similar calculation except that we do not assume a constant α and we work with multicomponent mixtures. After obtaining our value of the amount of distillate by iteration, we must go back and correct for variations in K values caused by our inaccurate estimation of the correct temperature at the start of the first iteration. This, then, imposes a second iteration upon the first. Although this may sound complicated, it is really not so difficult, because we have prepared a detailed calculation chart showing just how each calculation comes from another (Figure 1). This chart is simple enough that one not skilled in the art—say a high-school student—could carry the calculations out on a desk calculator. Also, we have prepared a chart for absorption which we have adapted to the IBM Type 602 Calculating Punch, and we are in the process of doing the same for the distillation chart. In using the type 602 machine we have to use some seven control panels (wired separately) to carry out the various calculations. We are also considering, with the help of the IBM Applied Science Department, a somewhat simple and direct attack on the distillation calculation, using matrix algebra to solve a large number of equations used to describe the conditions in a distillation tower.

REFERENCES

1. ASCHER OPLER, "Machine Calculation of the Plate-by-Plate Composition of a Multicomponent Distillation Column," pp. 18-23.
2. ARTHUR ROSE, THEODORE J. WILLIAMS and WILLIAM S. DYE III, "Continuous Distillation Design Calculations with the Card-Programmed Electronic Calculator," pp. 24-31.

*This paper was presented informally.

$1 = x_f$	2	$3 = D$	$4 = B$	$5 = x_s$	$6 = Kx$	$7 = x_d$	$8 = \frac{y}{K}$	9* = Tray Temp.	10a	10b
	Data $P = 100$	$\frac{①_a + ①_b + ①_c - k_L}{1 - k_L - k_H}$	1 − ③	Calculate as Indicated Starting with Light Key	K⑤ K Selected at P and Proper Temp.	Calculate as Indicated Ending with Heavy Key	⑦$/K$ K Selected at P and Proper Temp.	$T_2 = T_1 + \frac{T_s - T_1}{M+N}$ ⋮ $T_I = T_{II} + \frac{T_s - T_1}{M+N}$	K at Temp. T_1 for Tray 1 133°	K at Temp. T_2 for Tray 2 155°
C_a .15	$R = 3$	.5435	.4565		242°	①$_a$/③ = .276	.1108	T_1 = 133	2.49	3.08
C_b .15	$V_m = CV_n$					①$_b$/③ = .276	.2706	T_2 = 155	1.02	1.31
C_c .25 = K	$C = 1$			k_L = .050	.1225	(①$_c - k_L$④)/③ = .418	.5291	T_x = 177	.79	1.04
C_d .10 = K_4	$N = 2$			(①$_d - k_H$③)/④ = .183	.2379	k_H = .030	.0857	T_{II} = 198	.35	.49
C_e .15	$M = 3$			①$_e$/④ = .329	.3718		Σ = 9962	T_I = 220	.26	.36
C_f .20	k_H = .03			①$_f$/④ = .438	.2672			T_s = 242	.13	.19
	k_L = .05				Σ = 9994 ΣKx must = 1		Σ y/K must = 1			
10f	10x	10y	10z	11 and 11a	$12a = A_1$	$12b = A_2$	$12f = S_f$	$12x = S_{II}$	$12y = S_I$	$13 = f_1(A)$
K at Temp. T_f for Feed Tray 177°	K at Temp. T_{II} for Tray II 198°	K at Temp. T_I for Tray I 220°	K at Temp. T_s for Still 242°	⑪ $= \frac{L_n}{V_n} = \frac{R}{R+1}$	For Tray 1 $\frac{⑪}{\textcircled{10a}}$	For Tray 2 $\frac{⑪}{\textcircled{10b}}$	For Tray f $\frac{\textcircled{10f}}{\textcircled{11a}}$	For Tray II $\frac{\textcircled{10x}}{\textcircled{11a}}$	For Tray I $\frac{\textcircled{10y}}{\textcircled{11a}}$	($\textcircled{12a}$ +1) $\textcircled{12b}$ +1
C_a 3.71	4.35	5.1	5.9	$\textcircled{11a} = \frac{L_m}{V_m} =$	.301	.244	3.064	3.593	4.213	1.317
C_b 1.64	1.97	2.37	2.83		.735	.573	1.355	1.627	1.958	1.994
C_c 1.34	1.66	2.02	2.45	$1+(RC+C-1)D$	.949	.721	1.107	1.371	1.669	2.405
C_d .65	.84	1.06	1.30	$CD(R+1)$	2.143	1.531	.537	.694	.876	5.812
C_e .50	.67	.89	1.13		2.885	2.083	.413	.553	.735	9.092
C_f .26	.35	.47	.61	1.21	5.769	3.947	.215	.289	.388	27.717
$14 = f_2(A)$	$15 = f_1(S)$	$16 = f_2(S)$	$17 = b$	$18 = C$	$19 = a + b$	$20 = d$	$21 = d - c$	$22 = x_f c$	$23 = x_f(d-c)$	$24 = a + b + c$
$\textcircled{12a}$ $\textcircled{12b}$	($\textcircled{12y}$ +1) $\textcircled{12x}$ +1	$\textcircled{12y}$ $\textcircled{12x}$	R⑭	$\textcircled{12f}$ ⑮	⑬ + ⑰	$(R+1)$ $\textcircled{10z}$ × $\textcircled{12f}$ ⑯	⑳ − ⑱	① ⑱	① ㉑	⑲ + ⑱
C_a .0734	19.730	15.137	.220	60.453	1.537	1094.56	1034.11	9.0680	155.12	61.990
C_b .421	5.813	3.186	1.263	7.877	3.257	48.870	40.993	1.1816	6.149	11.134
C_c .684	4.659	2.288	2.052	5.158	4457	24.823	19.665	1.2895	4.916	9.615
C_d 3.281	2.302	.608	9.843	1.236	15.655	1.698	.462	.1236	.0462	16.891
C_e 6.009	1.959	.406	18.027	.809	27.119	.7580	.051	.1214	.00765	27.928
C_f 22.770	1.401	.112	68.310	.301	96.027	.05880	.242	.06020	.04840	96.328
25 =	26 and 26a	27	28	29	$30 = x_d$	31	32	33	34	$35 = x_s$
$d - (a+b+c)$ ⑳ − ㉔	New D and D^2 Selected So Σ㉚ = 1	㉓ ㉖ + ㉒	㉔ ㉖	㉕ $\textcircled{26a}$ + ㉘	$\frac{㉗}{㉙}$	㉖ ㉚	㉛ ⑲	㉑ ㉖	⑱ + ㉝	$\frac{㉜}{㉞}$
C_a 1032.57	.541	92.987	33.537	335.749	.2770	.1499	.2303	559.45	619.90	.00037
C_b 37.736		4.508	6.024	17.068	.2641	.1429	.4654	22.18	30.057	.01548
C_c 15.208	.29268	3.949	5.202	9.653	.4091	.2213	.9864	10.639	15.797	.06244
C_d −15.193		.1486	9.138	4.691	.0317	.01715	.2685	.2499	1.4859	.1807
C_e −27.170		.1172	15.109	7.157	.0164	.00887	2406	.02759	.78141	.3078
C_f −96.269		.03402	52.113	23.937	.0014	.000757	.07269	.13092	.17008	.4274
					Σ = .9997					
36	37	$38a = (Lx)_1$	38a'	$38b = (Lx)_2$	38b'	$38f = (Lx)_f$	38f'	$38x = (Lx)_{II}$	38x'	$38y = (Lx)_I$
(1 − ㉖) ㉟	(R+1) ㉛	㊲ $\textcircled{12a}$	$\textcircled{38a}$ + ㉛	$\textcircled{38a'}$ 12b	$\textcircled{38b}$ + ㉛	$\frac{\textcircled{38b'}}{\textcircled{12f}}$	$\textcircled{38f}$ − ㊱	$\frac{\textcircled{38f'}}{\textcircled{12x}}$	$\textcircled{38x}$ − ㊱	$\frac{\textcircled{38x'}}{\textcircled{12y}}$
C_a .000170	.5994	.1804	.3303	.08059	.2305	.07521	.07504	.02089	.02072	.00492
C_b .007105	.5713	.4199	.5627	.3224	.4653	.3434	.3363	.2067	.1996	.1019
C_c .028660	.8851	.8399	1.0612	.7651	.9864	.8911	.8624	6290	.6004	.3597
C_d .082941	.06860	.1470	.1642	.2513	.2685	.5000	.4170	.6009	.5180	.5913
C_e .141280	.03548	.1023	.1112	.2317	.2406	.5825	.4412	.7978	.6565	.8932
C_f .196177	.00303	.01748	.01824	.07199	.07275	.3384	.1422	.4920	.2958	.7625
		1.7070		1.7231		2.7306		2.7473		2.7135
$39a = x_1$	$39b = x_2$	$39f = x_f$	$39x = x_{II}$	$39y = x_I$	40a	40b	40f	40x	40y	40z
$\frac{\textcircled{38a}}{\Sigma \textcircled{38a}}$	$\frac{\textcircled{38b}}{\Sigma \textcircled{38b}}$	$\frac{\textcircled{38f}}{\Sigma \textcircled{38f}}$	$\frac{\textcircled{38x}}{\Sigma \textcircled{38x}}$	$\frac{\textcircled{38y}}{\Sigma \textcircled{38y}}$	K $\textcircled{39a}$ K at T_1 = 137°	K $\textcircled{39b}$ 157° K at T_2 =	K· $\textcircled{39f}$ 177° K at T_f =	K $\textcircled{39x}$ 198° K at T_{II}	K $\textcircled{39y}$ 218° K at T_I =	K ㉟ K at T_s =
C_a .106	.047	.028	.0076	.0018	.2756	.1476	.1039	.0331	.0090	.0021
C_b .246	.181	.126	.075	.038	.2632	.2506	.2066	.1478	.0885	.0426
C_c .492	.444	.326	.229	.133	.4084	.4751	.4368	.3801	.2647	.1480
C_d .086	.146	.183	.218	.218	.0318	.0730	.1190	.1831	.2267	.2277
C_e .060	.134	.213	.290	.329	.0168	.0496	.1065	.1943	.2829	.3324
C_f .010	.042	.124	.179	.281	.0014	.0080	.0322	.0627	.1265	2479
					Σ = 9972	1.0039	1.0050	1.0011	.9983	1.0007

FIGURE 1. DISTILLATION CALCULATIONS

	Col. − $A = Y_0$		Col. − B = Data		Col. − $C1 = (L/G)_1$	Col. − $C2 = (L/G)_2$	Col. − $C3 = (L/G)_3$	Col. − $CN = (L/G)_N$
	$C_a =$ 78.50	$C_f =$ 1.35	$N =$ 20	$H_v =$ 9000	$O1 \div M1$	$O2 \div M2$	$O3 \div M3$	$ON \div MN$
	$C_b =$ 7.50	$C_g =$ 1.45	$L_{N+1} =$ 20	$Cp_g =$ 9.3				
	$C_c =$ 5.50	$C_h =$ .60	$G_o =$ 100	$P =$ 100#/IN2a	.200	.200	.200	.200
	$C_d =$ 1.70		$t_o =$ 100	$C =$ 10	.259	.252	.249	.232
	$C_e =$ 3.40		$Cp_1 =$ 94					
	Col. − $D1 = t_1 - t_0$	Col. − $E2 = t_2 - t_0$	Col. − $E3 = t_3 - t_0$	Col. − $EC = t_C - t_0$	Col. − $DC = t_C - t_0$	Col. − $EC + 1 = t_{C+1} - t_0$	Col. − $EC + 2 = t_{C+2} - t_0$	Col. − $EN = t_N - t_0$
	Assumed	$[D1\ (R1 + Q1) -$	$[D2\ (R2 + Q2) -$	$[DC-1\ (RC-1 +$	Trial	$[DC\ (RC + QC) -$	$[DC+1\ (RC+1 +$	$DN-1\ (RN-1 +$
	Maximum	$P1] \div R2$	$P2 - D1\ Q1] \div$	$QC-1) - NC-1 -$		$PC - DC-1 \times$	$QC+1) - PC+1 -$	$[QN-1) - PN-1 -$
	Rise		$R3$	$DC-2\ QC-2] \div RC$		$QC-1] \div RC+1$	$DC\ QC] \div RC+2$	$DN-2\ QN-2] \div RN$
I	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
II 1	10.0	11.30	10.19	− 0.14	10.0	13.65	14.81	2.76
2					9.0	12.26	13.29	.83
3					8.0	10.81	11.63	− 1.18
	Col. − $F1 = K_1$	Col. − $F2 = K_2$	Col. − $F3 = K_3$	Col. − $FN = K_N$	Col. − $G1 = A_1$	Col. − $G2 = A_2$	Col. − $G3 = A_3$	Col. − $GN = A_N$
	From Table	From Table	From Table	From Table	$C1 \div F1$	$C2 \div F2$	$C3 \div F3$	$CN \div FN$
C_a	37.8	37.8	37.8	37.8	.00529	.00529	.00529	.00529
C_b	5.72	5.72	5.72	5.72	.0350	.0350	.0350	.0350
C_c	1.80	1.80	1.80	1.80	.1111	.1111	.1111	.1111
C_d	.697	.697	.697	.697	.287	.287	.287	.287
C_e	.505	.505	.505	.505	.396	.396	.396	.396
C_f	.205	.205	.205	.205	.976	.976	.976	.976
C_g	.151	.151	.151	.151	1.324	1.324	1.324	1.324
C_h	.0782	.0782	.0782	.0782	2.56	2.56	2.56	2.56
	Col. − $H1 = b_1$	Col. − $H2 = b_2$	Col. − $H3 = b_3$	Col. − $HN = b_N$	Col. − $1 = \pi_1 Nb$	Col. − $J = Y_N$		
	$1 + G1$	$1 + G2 - G2 \div H1$	$1 + G3 - G3 \div H2$	$1 + GN - GN \div HN-1$	$H1\ H2\ H3 \ldots HN$	$A \div 1$		
C_a	1.00529	1.00004	1.00000	1.00000	1.00533	78.084		
C_b	1.0350	1.0012	1.0001	1.0000	1.0363	7.237		
C_c	1.1111	1.0111	1.0012	1.0000	1.1249	4.889		
C_d	1.287	1.064	1.017	1.0000	1.3982	1.216		
C_e	1.396	1.112	1.040	1.000	1.6560	2.050		
C_f	1.976	1.482	1.318	1.038	16.668	.08099		
C_g	2.324	1.754	1.669	1.325	1307.0	.00111		
C_h	3.56	2.841	2.659	2.560		.00000		
	Col. − $K1 = Y_1$	Col. − $K2 = Y_2$	Col. − $(KN-1) = Y_{N-1}$	Col. − $KN = Y_N$	Col. − $L1 = Y_1$	Col. − $L2 = Y_2$	Col. − $L3 = Y_3$	Col. − $LN = Y_N$
	$J + G2\ K2$	$J + G3\ K3$	$J + GN\ KN$	J	$[A - J] \div G1$	$[L1 - J] \div G2$	$[L2 - J] \div G3$	$[LN-1 - J] \div GN$
C_a	78.500	78.500	78.497	78.084				
C_b	7.500	7.500	7.490	7.237				
C_c	5.500	5.500	5.432	4.889				
C_d	1.700	1.700	1.565	1.216				
C_e	3.400	3.400	2.864	2.050				
C_f	1.2986	1.2476	.16003	.08099				
C_g					1.094	.826	.623	.001
C_h					.234	.0915	.0357	.0000
	Col. − $M1 = G_1$	Col. − $M2 = G_2$	Col. − $M3 = G_3$	Col. − $MN = G_N$	Col. − $N1 = \Delta G_1$	Col. − $N2 = \Delta G_2$	Col. − $N3 = \Delta G_3$	Col. − $NN = \Delta G_N$
	$\Sigma K1 + \Sigma L1$	$\Sigma K2 + \Sigma L2$	$\Sigma K3 + \Sigma L3$	$\Sigma KN + \Sigma LN$	$G_0 - M1$	$M1 - M2$	$M2 - M3$	$MN-1 - MN$
	99.227	98.766	98.454	93.558	.773	.461	.312	1.721
	Col. − $O1 = L_1$	Col. − $O2 = L_2$	Col. − $(ON-1) = L_{N-1}$	Col. − $ON = L_N$	Col. − $P1 = H_v \Delta G_1$	Col. − $P2 = H_v \Delta G_2$	Col. − $P3 = H_v \Delta G_3$	Col. − $PN = H_v \Delta G_N$
	$O2 + N1$	$O3 + N2$	$ON + (NN - 1)$	$L_{N-1} + NN$	$H_v\ N1$	$H_v\ N2$	$H_v\ N3$	$H_v\ NN$
	25.709	24.936	22.293	21.721	6960.	4150.	2810.	15,490.
	Col. − $Q1 = Cp_g G_1$	Col. − $Q2 = Cp_g G_2$	Col. − $Q3 = Cp_g G_3$	Col. − $QN = Cp_g G_N$	Col. − $R1 = Cp_1 L_1$	Col. $R2 = Cp_1 L_2$	Col. − $R3 = Cp_1 L_3$	Col. − $RN = Cp_1 L_{N+1}$
	$Cp_g\ M1$	$Cp_g\ M2$	$Cp_g\ M3$	$Cp_g\ MN$	$Cp_1\ O1$	$Cp_1\ O2$	$Cp_1\ O3$	$Cp_1\ L_{N+1}$
	923.	918.	916.	742.	2420.	2340.	2300.	1880

FIGURE 2. ABSORPTION CALCULATIONS

Some Applications of the Monte Carlo Method

Matrix Inversion on the IBM Accounting Machine[*]

ASCHER OPLER
The Dow Chemical Company

IN THE JULY, 1950, issue of *Mathematical Tables and Other Aids to Computation,* Forsythe and Liebler described a Monte Carlo method of matrix inversion. This has been adapted for use with an IBM Type 405 Accounting Machine. In the interest of brevity, I state a version of the original method without proof.

If B is a square matrix such that $A = | I - B |$, then $B^{-1} = (| I - A |)^{-1} = \sum_{k=0}^{\infty} A^k$. When the last expression is a convergent series, B^{-1} may be evaluated by playing an infinite number of games as described below. In practice it may be approximated by playing a large number.

Choose stop probabilities, p_j, such that $\sum_{j=1}^{n} a_{ij} + p_j = 1$. (This is one restriction of the class of matrices that may be inverted.) Prepare a deck of punched cards as follows: one half of the card is to contain random numbers; a playing field of n columns is prepared in the other half of the card. In each of the n columns, punch digits so that the probability of a card containing a_{ij} or p_j is equal to the corresponding element of A or the corresponding stop probability. "Shuffle" by repeated sorting in the randomizing field.

The game is played by passing this prepared deck through the accounting machine n times. To play for inverse element $(B_{ij})^{-1}$, the machine is instructed to start with the first card and read the ith column on that card. If the number read is m, the machine reads the mth column of the next card. Each card directs the reading of the following card. When a stop card is read, the game is ended. If the stop card was read in column j, the score is 1; otherwise the score is zero. Actually, n games are played at once with one of the n elements winning and others losing. The inverse elements are found by dividing the final scores by total games played and then finally dividing by the stop probability. In certain cases, the inverse matrix may be printed by the accounting machine, line-by-line.

This method appears to be the simplest and most economical way of inverting matrices. The operations are proportional to n^2 instead of n^3. The disadvantage is that the results obtained in reasonable times are approximations. (A fair approximation to a seventh-order inverse may be found in less than an hour; a good approximation might take four to six hours.) The class of matrices that may be inverted is limited, but, with ingenuity, many of these limitations may be overcome.

A paper describing the method fully has been submitted for publication in *Mathematical Tables and Other Aids to Computation.*

[*]This paper was presented informally.

Remarks on Finding Roots of, and Inverting, a Matrix[*]

GILBERT W. KING
Arthur D. Little, Incorporated

THE ESSENTIAL FEATURE of the Monte-Carlo Method of inverting a matrix, or of finding its roots, is the well-known iterative procedure of raising the matrix to a high power. The reason the method does this can be easily seen as follows: Choose a row and column at random and write down the matrix element, $a_{i\lambda}$. Then choose another element from the λth row lying in column μ, say, $a_{\lambda\mu}$. Again choose an element from the μth row, say $a_{\mu\nu}$. This is done N times. The choices are multiplied together,

$$a_{i\lambda}\, a_{\lambda\mu}\, a_{\mu\nu} \ldots a_{\kappa j}\,.$$

We recognize this as a term in the ijth element of the Nth power of the matrix. If we took all possible choices of paths from row to column, starting at the ith row and ending a jth column, and added the products, we would have precisely the value of ijth element of the Nth power. The random procedure described above merely picks some terms from some elements at random. By having the probability of picking any element, say, $a_{\mu\nu}$, proportional to its magnitude, the Monte-Carlo Method picks out the terms in proportion to their magnitude, and hence gets the principal term of the principal elements of the Nth power of the matrix.[1]

REFERENCE

1. See also GILBERT W. KING, "Stochastic Methods in Quantum Mechanics," *Seminar on Scientific Computation, November, 1949* (IBM).

[*]This paper was presented informally.

Remarks on the Monte Carlo Method

CUTHBERT C. HURD

International Business Machines Corporation

THE MONTE CARLO METHOD has aroused great interest and has found many applications. A general description of the method as applied to a class of problems in mathematical physics is given by Metropolis and Ulam.[1] The Proceedings of a Symposium on Monte Carlo Methods, which was held in Los Angeles from June 29 to July 1, 1949, is to be published by the National Bureau of Standards. Finally, papers on the subject have been presented at the IBM Seminar on Scientific Computation, November, 1949,[2] and the IBM Computation Seminar, December, 1949.[3]

In applying the method it is convenient to have available a set of random digits. These digits can be computed by several schemes which have been proposed. For example, 60,000 such digits can be computed in an hour on the IBM Type 604 Electronic Calculating Punch. Once the set of random digits is available, sets with a prescribed probability distribution can be obtained by performing a table lookup operation using the sorter and collator, or in some cases by direct calculation.

The Monte Carlo Method is useful in many instances in giving an estimate of an answer. The results of mathematical statistics then allow one to attach a measure of reliability to the answer.

REFERENCES

1. N. METROPOLIS and S. ULAM, "The Monte Carlo Method," *Jour. Am. Stat. Assoc.,* Vol. 44 (1949), pp. 335-341.
2. *Seminar on Scientific Computation, November, 1949* (IBM).
   WILLIAM W. WOODBURY, "Monte Carlo Calculations," pp. 17-19.
   HERMAN KAHN, "Modification of the Monte Carlo Method," pp. 20-27.
   Z. W. BIRNBAUM, "On the Distribution of Kolmogorov's Statistic for Finite Sample Size," pp. 33-36.
   GILBERT W. KING, "Stochastic Methods in Quantum Mechanics," pp. 42-48.
   JOHN H. CURTISS, "Sampling Methods Applied to Differential and Difference Equations," pp. 87-109.
3. *Computation Seminar, December, 1949* (IBM).
   MARK KAC and M. D. DONSKER, "The Monte Carlo Method and Its Applications," pp. 74-81.
   P. C. JOHNSON and F. C. UFFELMAN, "A Punched Card Application of the Monte Carlo Method," pp. 82-88.
   EVERETT C. YOWELL, "A Monte Carlo Method of Solving Laplace's Equation," pp. 89-91.
   GILBERT W. KING, "Further Remarks on Stochastic Methods in Quantum Mechanics," pp. 92-94.

Plotting Punched Card Data Using the IBM Type 405 Accounting Machine[*]

PAUL T. NIMS

Chrysler Corporation

A METHOD for plotting the points on a curve has been described in the writings[1] of Dr. Gilbert W. King. This method gives a plot having abscissa values from 1 to 80 and an unlimited range of ordinate. However, for the purpose of checking some automotive design calculations, it is convenient to have a larger range of ordinates and a relatively small range of abscissas, together with adaptability to multiple-valued functions.

A method similar to that of Dr. King's but differing in some particulars was worked out by Mrs. Virginia Johnson and Mr. F. F. Timpner of the Chrysler Corporation Engineering Division. First, a master deck was prepared, covering the range of ordinate values desired (positive and negative) with the ordinates punched in columns 79-80 as well as in the same field as the ordinates on the problem cards. The master deck was distinguished with an X punch in some suitable column.

The master deck and the problem deck were sorted together (in that order) on ordinate values to give a combined deck ready for the accounting machine.

As mentioned above, the range of abscissas is relatively small; in fact, determined by the available selector and type bar capacity, the range is 0 to 69. The two columns are read by the 405, and the tens and units digits enter separate digit selectors. The tens digits (one through six) control the six different 10-position field selectors on the type 405 machine to pick up one of them, depending on the tens digit read.

The ten units digits are wired, position for position, to the 10 common hubs on the field selector 6 and from its transferred hubs to counter input hubs 60-69. The normal hubs of field selector 6 are wired to the common hubs of the field selector 5, the transferred hubs of the field selector 5 are wired to counter inputs 50-59, and the remaining selectors are laced similarly. The normal hubs of the field selector 1 are wired to the first ten counter positions to take care of abscissas 0 through 9. The only exception is abscissa 43, which is not wired to any counter, as the 405 has a type bar support at this point. The counter exits are wired, position for position, to the total type bars.

Counters are all wired to add from *plug to C*, but carry wiring is unnecessary. The accounting machine is wired to print a minor total on the X in the master cards. Alternatively, if the ordinate values are punched in the same columns on both detail and X cards, a minor total pickup can be used. This minor total, also, advances the paper one line. With this arrangement, the abscissas of all cards having the same ordinate are stored in the counters, and the whole line is printed at once. All of the zero suppressors are raised.

There are a few special wires. Columns 79-80 of the master cards are listed to give an ordinate scale. As no zeros will print, the abscissa values ending in zero are wired to an X selector, and a *card count* pulse is entered in the counter in order to print a 1 for the zero values. A set of 70 list wires is also used to print headings on the curve and to print the abscissa scale.

This plotting method is very convenient for checking the results of engineering computations, and the illustration (Figure 1) shows a function which had a few deliberate errors introduced in the last place to check the sensitivity of error detection.

REFERENCE

1. GILBERT W. KING, *A Method of Plotting on Standard IBM Equipment, MTAC,* Vol. 3, No. 25 (January, 1949), pp. 352-355.

DISCUSSION

Mr. Bejarano: What is the possibility of using the reverse process and picking up the number from the graphs?

Dr. Hurd: Various systems have been tried and proposed. I will tell you what I know about it.

Mr. Bejarano: I have in mind distillation, using the graph of the case where the charts agree without any artificial function.

[*] This paper was presented informally.

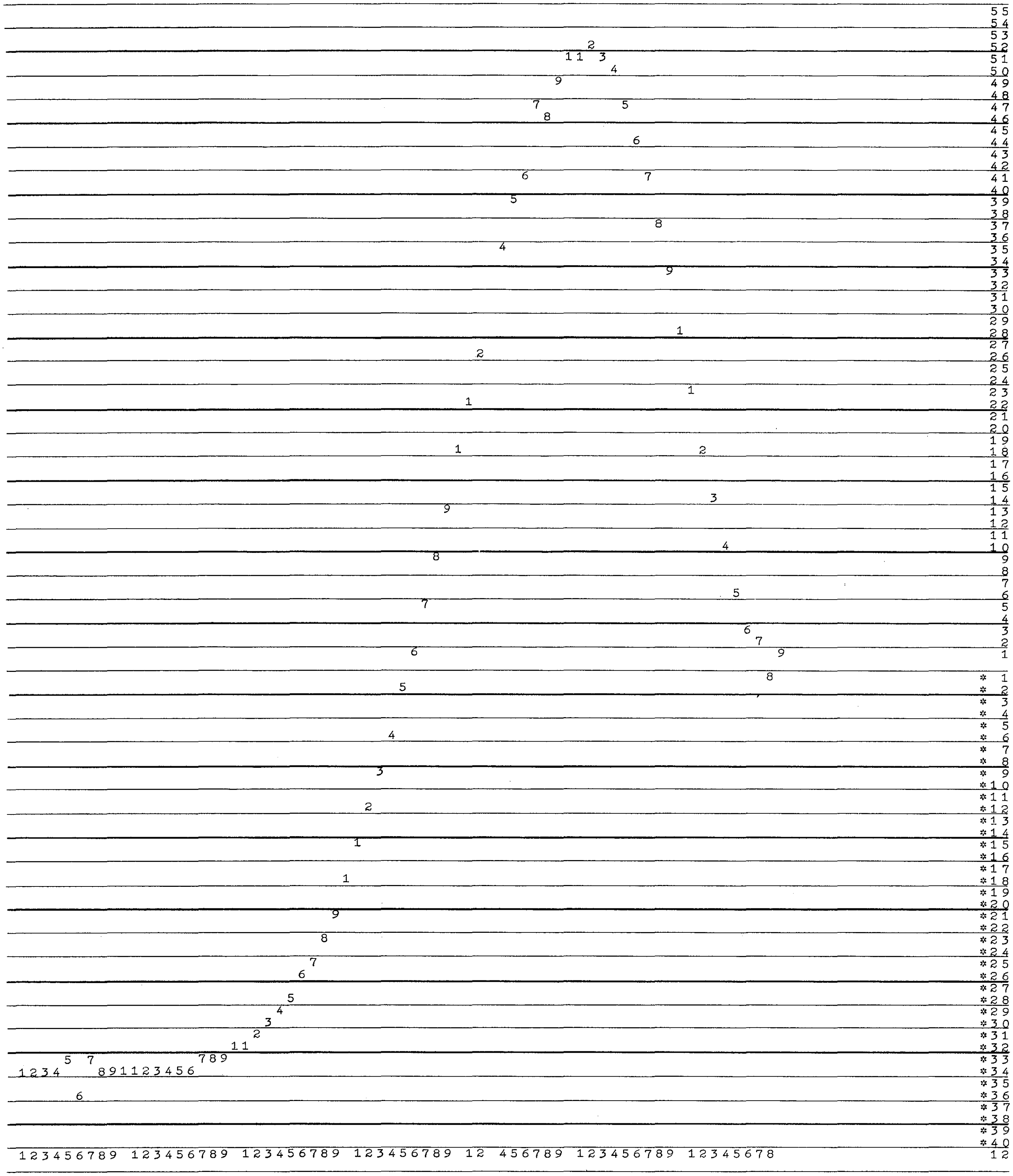

5 5
5 4
5 3
5 2
5 1
5 0
4 9
4 8
4 7
4 6
4 5
4 4
4 3
4 2
4 1
4 0
3 9
3 8
3 7
3 6
3 5
3 4
3 3
3 2
3 1
3 0
2 9
2 8
2 7
2 6
2 5
2 4
2 3
2 2
2 1
2 0
1 9
1 8
1 7
1 6
1 5
1 4
1 3
1 2
1 1
1 0
9
8
7
6
5
4
3
2
1
* 1
* 2
* 3
* 4
* 5
* 6
* 7
* 8
* 9
*1 0
*1 1
*1 2
*1 3
*1 4
*1 5
*1 6
*1 7
*1 8
*1 9
*2 0
*2 1
*2 2
*2 3
*2 4
*2 5
*2 6
*2 7
*2 8
*2 9
*3 0
*3 1
*3 2
*3 3
*3 4
*3 5
*3 6
*3 7
*3 8
*3 9
*4 0
1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9 1 2 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8
1 2

FIGURE 1

Dr. Hurd: One device, which is available at the present time, which goes from a plot to IBM cards is that which is done by manual operation to locate the ordinate and the abscissa. In fact, there are at least two devices of that kind. There has been experimentation on using photoelectric cells.

Dr. King: Some of the charts on the slides I have were made on the IBM accounting machine. We did not consider it a trivial proposition. We had hundreds of these plots to do. In our scheme we don't have any sorting and collating to do. We take the cards as they come out of the computer; the ordinate this way can be in any field of the card. Just change a few of the wires, and, as the cards feed through the machine, points are plotted just as fast as listing speed. There are no total cycles.

Dr. Hurd: Many variations of these schemes have been worked out. One scheme used at the Boeing Aircraft plant was to make a mark which is an 8. They also put in calibration marks and the expected average mark. Then they photograph it and reduce it, by which time these 8's look like dots and the graph looks like a continuous graph. You have essentially two-digit accuracy of plots here. As the cards are run through, the first differences are calculated and printed, so that if you want to interpolate, you have at least first order interpolating coefficients already plotted on the sheet.

Mr. Mongreiff: All of these schemes depend on selecting correct type bars. This is especially useful where the length of the graph is much greater than the height, but if you could turn the system around, that is, if your needs are that the height of the graph is greater than the length, it seems to me you could depend on successive reduction on the total in a counter, put the data in a counter to begin with, and then reduce it by one or two or three successively. You could even make that amount manually changeable just by shifting a wire or two and then emitting an X or another impulse when the counter turned negative. It would be possible to start off with a complement, say a 100 complement or a 500 complement of the original. In that way you would print a graph which was right side up.

Dr. Hurd: The 407 is a new accounting machine which we have not demonstrated here. Perhaps that would be capable of doing this.

Mr. Opler: In mentioning the 407, it reminds me that the 407 would be twice as powerful a tool for graphing as the 402, because it has 120 type bars across and lists 150 lines per minute. You can interpret in a hurry.

A Method for Evaluating Determinants and Inverting Matrices with Arbitrary Polynomial Elements by IBM Punched Card Methods

L. E. GROSH, JR. E. USDIN

Purdue University

THE METHOD of determinant evaluation proposed here is essentially a scheme for the collection of terms with like coefficients and like powers of x, where these terms are polynomials with non-numerical coefficients. The method is based upon the termwise expansion of a determinant; thus, it is concerned with the evaluation of $n!$ polynomials for an nth order determinant. If any zero elements appear in the determinant, a significant reduction in the number of cards handled may be made. The method works best when there are few distinct coefficients in the original determinant.

The inversion of a matrix with polynomial elements is based upon the expression for the general element of the inverse which involves the ratio of two determinants. Thus, for an nth order matrix, the inversion problem is reduced to the evaluation of n^2 determinants of order $(n-1)$ and one nth order determinant. Most of the cards necessary for the evaluation of the n^2 determinants are generated in the evaluation of the nth order determinant.

The equipment needed for the evaluation procedure is: a card punch, sorter, reproducer, 602-A calculating punch, and 405 accounting machine—direct subtracting.

The use of the calculating punch is not absolutely necessary; its role in the procedure could be performed by a sorting and gang punching operation. However, for a problem of any size, the sorting operation would become extremely complicated and impractical unless an IBM Type 101 Electronic Statistical Machine were available.

Evaluation of Determinants

The Coding Problem

The heart of the evaluation scheme is a method of coding. The coding is best understood when applied to a simple example. Consider the third order determinant shown below.

$$D = \begin{vmatrix} P_{11}(x) & P_{12}(x) & P_{13}(x) \\ P_{21}(x) & P_{22}(x) & P_{23}(x) \\ P_{31}(x) & P_{32}(x) & P_{33}(x) \end{vmatrix} = \begin{vmatrix} a_3^{11}x^3 + a_2^{11}x^2 + a_1^{11}x + a_0^{11} & a_1^{12}x + a_0^{12} & a_0^{13} \\ a_1^{21}x + a_0^{21} & a_1^{22}x + a_0^{22} & a_2^{23}x^2 + a_1^{23}x + a_0^{23} \\ a_0^{31} & a_1^{32}x + a_0^{32} & a_1^{33}x + a_0^{33} \end{vmatrix}$$

The termwise expansion of this determinant is:

$$\begin{aligned} D = \ & P_{11}(x)P_{22}(x)P_{33}(x) + P_{21}(x)P_{32}(x)P_{13}(x) \\ & + P_{31}(x)P_{12}(x)P_{23}(x) - P_{31}(x)P_{22}(x)P_{13}(x) \\ & - P_{21}(x)P_{12}(x)P_{33}(x) - P_{11}(x)P_{32}(x)P_{23}(x) \end{aligned}$$

If the factors of each term are ordered on the column subscript as above, these six terms may be represented by the permutations of the numbers 1, 2, 3, and an X punch to indicate a negative sign. Thus,

$$\begin{aligned} 123 &= P_{11}(x)P_{22}(x)P_{33}(x) \\ 231 &= P_{21}(x)P_{32}(x)P_{13}(x) \\ 312 &= P_{31}(x)P_{12}(x)P_{23}(x) \\ 321X &= -P_{31}(x)P_{22}(x)P_{13}(x) \\ 213X &= -P_{21}(x)P_{12}(x)P_{33}(x) \\ 132X &= -P_{11}(x)P_{32}(x)P_{23}(x)\ . \end{aligned}$$

Let each one of the six symbols 123, 321X, . . . be called a permutation number.

Now consider one of the products, say

$$\begin{aligned} P_{21}(x)P_{12}(x)P_{33}(x) = \ & \\ & (a_1^{21}x + a_0^{21})(a_1^{12}x + a_0^{12})(a_1^{33}x + a_0^{33}) \\ = \ & a_1^{21}\ a_1^{12}\ a_1^{33}x^3 \\ & + (a_1^{21}\ a_1^{12}\ a_0^{33} + a_1^{21}\ a_0^{12}\ a_1^{33} + a_0^{21}\ a_1^{12}\ a_1^{33})x^2 \\ & + (a_1^{21}\ a_0^{12}\ a_0^{33} + a_0^{21}\ a_1^{12}\ a_0^{33} + a_0^{21}\ a_0^{12}\ a_1^{33})x \\ & + a_0^{21}\ a_0^{12}\ a_0^{33}\ . \end{aligned}$$

Note that there are $2(2)(2) = 8$ different terms in the expanded polynomial. In general, there will be at most, $\prod_{j=1}^{n} (k_{ij} + 1)$ terms in this expansion, and there will be exactly this number if all $a_m^{ij} \neq 0$ and are distinct, where k_{ij} is the degree of $P_{ij}(x)$, and n is the order of the determinant. Each of these three-factor coefficients may be represented by a three-digit number—namely, the subscripts of the a^{ij}_m's involved. Thus,

$$\begin{array}{llll} 111 = a^{21}_1 \; a^{12}_1 \; a^{33}_1 & & 100 = a^{21}_1 \; a^{12}_0 \; a^{33}_0 \\ 110 = a^{21}_1 \; a^{12}_1 \; a^{33}_0 & & 010 = a^{21}_0 \; a^{12}_1 \; a^{33}_0 \\ 101 = a^{21}_1 \; a^{12}_0 \; a^{33}_1 & & 001 = a^{21}_0 \; a^{12}_0 \; a^{33}_1 \\ 011 = a^{21}_0 \; a^{12}_1 \; a^{33}_1 & & 000 = a^{21}_0 \; a^{12}_0 \; a^{33}_0 \end{array}$$

Let each of these three-digit numbers be called a selection number. The first digit i of the selection number ijk refers to the coefficient of x^i from a polynomial in the first column; the second digit j refers to the coefficient of x^j from a polynomial in the second column; and the third digit k refers to the coefficient of x^k from a polynomial in the third column. Also, the sum of the digits $i+j+k$ is the power of x with which this three-factor product is associated.

The combination of a permutation number and a selection number allows us to code each term in the expansion of D; also to indicate its sign and the power of x in the expansion with which the three factor coefficient is associated. Thus,

$$132\ X\ 302 = -a^{11}_3 \quad a^{32}_0 \quad a^{23}_2$$

and is a coefficient of x^5 in the final expansion.

If all of the a^{ij}_m's are distinct, the coding problem is finished, but in some important cases not all of the a^{ij}_m's will be distinct; thus there is one more coding step to be considered. For example, take the following special case of the above third order determinant D.

$$D_1 = \begin{vmatrix} b_3x^3 + b_2x^2 + b_1x + b_0 & b_3x + b_2 & b_0 \\ b_3x \;\; + b_0 & b_2x + b_1 & b_1x^2 + b_2x + b_3 \\ b_2 & b_1x + b_0 & b_0x \;\; + b_3 \end{vmatrix}$$

There are many ways in which the coefficient $\pm b_0b_2b_3$ may arise, e.g.,

$$\begin{aligned} 132\ X\ 200 &= -b_2b_0b_3 \\ 132\ X\ 301 &= -b_3b_0b_2 \\ 213\ X\ 101 &= -b_3b_2b_0 \\ 213\ X\ 000 &= -b_0b_2b_3 \\ &\text{etc.} \end{aligned}$$

There are only four distinct coefficients in our problem; therefore a four-digit number will be sufficient to identify any possible combination of these b_i's. These will be called term numbers. Let the first digit of this number be the power of b_0 in the term, the second digit be the power of b_1, etc. Thus,

$$\begin{aligned} 1011 &= b_2b_0b_3 \\ &= b_3b_0b_2 \\ &= b_3b_2b_0 \\ &= b_0b_2b_3\,. \end{aligned}$$

The main points of the coding scheme can be summarized as follows:

1. A three-digit permutation number indicates which three polynomials are multiplied together and indicates the sign of the resultant polynomial.
2. A three-digit selection number indicates which coefficient of each polynomial shall be multiplied together to obtain a resultant coefficient. The power of x, with which this resultant coefficient is associated, is the sum of the digits.
3. A four-digit number indicates the power of the original distinct coefficients as they appear in the resultant coefficient.

Card Generation

Although this coding scheme is relatively simple, it would be almost impossible to maintain 100% accuracy on even a medium-size problem if the coding were done by hand. Fortunately, all of the coding may be done by machines with very little key punching. To accomplish this, four different decks of cards are used. These are:

1. Permutation deck
2. Selection deck
3. Master gang punching deck
4. Term deck

The term deck is the final product, and the other three decks are used in its generation. A typical card layout for the problem considered above is as follows:

Card Layout

	col.	*Term Deck*
Permutation Deck	1	permutation number
	2	
	3	
	4	X if permutation number is positive
	5	X if permutation number is negative
Selection Deck	6	selection number
	7	
	8	
	9	the actual coefficients indicated by the permutation and term numbers $b_0=0$ $b_1=1$ $b_2=2$ $b_3=3$
	10	
	11	
	12	term number representing the power of the various b_m's as they appear in the final answer.
	13	
	14	
	15	
	16	the power of x with which the term number is associated (the sum of columns 6, 7, and 8).
	17	a check of the machine operation, the sum of columns 12, 13, 14, and 15. For this problem it should be a 3 for every term card.

The master gang punching deck will have a distinguishing X punch, say in column 80, and is divided into three parts:

1. punching in columns 1, 6, and 9
2. punching in columns 2, 7, and 10
3. punching in columns 3, 8, and 11.

This deck is used to facilitate the punching of columns 9, 10, and 11 of the term deck after the first eight columns have been generated from the permutation and selection decks. Consider part 2 of this deck; the required cards are:

Column			
2	7	10	
1	0	3	$= b_3$
1	1	2	$= b_2$
2	0	2	$= b_2$
2	1	1	$= b_1$
3	0	1	$= b_1$
3	1	0	$= b_0$.

The generation of the permutation and selection deck is quite simple, but is a very important part of the coding operation. The permutation deck for an nth order determinant is generated from the permutation deck for a determinant of the $(n-1)$st order. If the generation scheme is carried out correctly, it is only necessary to handle $n!$ cards for this operation. While it is possible to generate the permutation deck for any order determinant from a single key punched card, it is more practical to select some moderate-size order, say 4th, key punch the required 24 permutation cards, and start the generation procedure from this point. Since the example considered above is of 3rd order, we shall start with a 2nd order determinant for illustration. For ease of machine operation we shall use two columns to designate the sign of the permutation. These columns shall be the $(n+1)$st and the $(n+2)$nd columns, if we are generating cards for an nth order determinant. Thus, for our example, the sign will be in the 4th and 5th columns. If the generation procedure is started with permutation cards for an mth order determinant and cards for an nth order determinant are desired, the following steps must be repeated $(n-m)$ times:

1. Letting − denote a blank column, the permutations for a 2nd order determinant are
   12−X− and 21−−X,
   the X in column 4 denoting a positive permutation and the X in column 5 denoting a negative permutation.
2. Gang punch a 3 in column 3, giving
   123X− and 213−X
3. Reproduce the cards from step 2, interchanging columns 2 and 3 and columns 4 and 5. Thus,
   123X− gives 132−X,
   213−X gives 231X−.
4. Reproduce the cards from step 3, interchanging columns 1 and 2 and columns 4 and 5. Thus,
   132−X gives 312X−,
   231X− gives 321−X.
5. For larger n, this procedure of interchanging columns is repeated until the n gang punched in step 2 in the nth columns has been moved over to the first column. Thus, there are $(n-1)$ interchanges.

The six cards we have generated are:

123X− 132−X 312X−
213−X 231X− 321−X,

and are all the permutation cards necessary for a 3rd order determinant.

The steps in the generation of the selection cards are:

1. Determine the highest degree appearing in each column of the determinant. The maximum number of cards needed will be $\prod_{j=1}^{n}(k_{ij}+1)$ as stated above, where k_{ij} is the highest degree appearing in column j.
2. Using blank cards, key punch the numbers 0, 1, . . . , k_{i1} into column $n+3$, one number to each card. For our example this would be 0, 1, 2, and 3 in column 6.
3. Gang punch a 0 in column $n+4$ of the cards made in step 2, giving 00, 10, 20, and 30.
4. Reproduce column $n+3$ of step 3 and gang punch 1 into column $n+4$, giving 01, 11, 21, and 31. Repeat this operation until k_{2j} has been gang punched into column $n+4$.
5. Gang punch a 0 in column $n+5$ of the cards generated in steps 3 and 4.
6. Reproduce columns $(n+3)$ and $(n+4)$ and gang punch a 1 in column $(n+5)$.
7. Repeat step 6 until k_{3j} has been gang punched into column $(n+5)$.
8. The operations of gang punching and reproducing are repeated until k_{nj} has been gang punched into column $(2n+2)$.

The 24 cards generated for the determinant D by the above procedure are:

000	010	100	110	200	210	300	310
001	011	101	111	201	211	301	311
002	012	102	112	202	212	302	312

The next step in the evaluation procedure is the generation of the term cards. The general procedure to be followed here is to place the selection cards in the reproducing hopper and a blank deck of cards preceded by a single permutation card in the gang punch hopper. Then, gang punch the permutation number and reproduce the selection number into the blank deck, column-for-column. This procedure would yield $n!\prod_{j=1}^{n}(k_{ij}+1)$ term cards, many of which would have no meaning unless all elements of the same column had the same degree. Referring to the example above, $3!(24) = 144$ term cards would be generated by the unmodified reproducing and gang punching procedure. By a study of the degree of the polynomial represented by each permutation card, it is easily seen that only 60 term cards are needed for the above example.

Permutation Number	$k_{1j}+1$	$k_{2j}+1$	$k_{3j}+1$	*Number of Term Cards*
123X−	4	2	2	16
132−X	4	2	3	24
213−X	2	2	2	8
231X−	2	2	1	4
312X−	1	2	3	6
321−X	1	2	1	2
Total number of term cards				60

The extraneous term cards can be eliminated and much time and many cards saved if a sorting procedure is used before the term cards are punched for each permutation card. As an aid to the sorting operation, rewrite the determinant in the following symbolic form:

$$\begin{vmatrix} 0,\ 1,\ 2,\ 3 & 0,\ 1 & 0 \\ 0,\ 1 & 0,\ 1 & 0,\ 1,\ 2 \\ 0 & 0,\ 1 & 0,\ 1 \end{vmatrix}$$

where each element has been replaced by the subscripts of the a^{ij}_m s appearing in that element. From this symbolic form it is easily seen which selection numbers are needed for any permutation number. The sorting procedure to be followed for each permutation number is:

	1	2
Permutation Number	*Sort on col. 6 Reject*	*Sort cards from 1 on column 8 Reject*
123X−	none	2
132−X	none	none
213−X	2, 3	2
231X−	2, 3	1, 2
312X−	1, 2, 3	none
321−X	1, 2, 3	1, 2.

The selection cards left after this sorting procedure are the ones to be used in punching the term cards.

Machine Operations

After the generation of the term cards, the following machine operations must be performed:

1. Using the master gang punch deck, gang punch in the appropriate columns (9 to 11) the coded values of the b_i's.
2. Using the 602-A calculating punch, count the number of times the various b_i's appear in columns 9 to 11 and punch this information in columns 12 to 15.
3. On the 602-A, crossfoot columns 6, 7, and 8 to determine the power of x with which each term card is associated; punch this sum in column 16. At the same time, crossfoot columns 12, 13, 14, and 15 and punch in column 17 to obtain a check on operation 2.
4. By sorting, arrange the cards in order by term number, columns 12 to 15, within each power of x, column 16.
5. Using a 405 or any other direct subtraction accounting machine, tabulate the term cards, controlling on the power of x and the term number, columns 12 to 16, adding a 1 for each positive term card and subtracting 1 for each negative term card. Either column 4 or 5 may be used for this purpose. Naturally, the term number and power of x should be group indicated.

Inversion of Matrices

Using the expression

$$\bar{a}_{ij} = \frac{A_{ij}}{\det(a_{ij})}$$

for an element of the inverse of the matrix (a_{ij}), where A_{ij} is the co-factor of a_{ij}, it is easy to adapt the above determinant methods to evaluate this expression. The evaluation of $\det(a_{ij})$ needs no explanation; only the cofactor A_{ij} must be considered here. We can write

$$\det(a_{ij}) = \sum_{i \text{ or } j} a_{ij}A_{ij}\,.$$

To evaluate A_{ij}, it is only necessary to select from the term cards used in the evaluation of $\det(a_{ij})$ those cards which contain the element a_{ij}. If these cards are tabulated in the same manner as for $\det(a_{ij})$, the result will be $a_{ij}A_{ij}$, and to obtain A_{ij} we must factor out a_{ij} from this tabulation. Because we are considering polynomial elements,

$$a_{ij} = P_{ij}(x)\ ,$$

the factoring procedure would be cumbersome. To avoid this, instead of selecting all of the term cards for $P_{ij}(x)$, select only those for a particular a^{ij}_m of $P_{ij}(x)$. Then a^{ij}_m can be factored out of the tabulation record by subtracting 1 from each entry in the column of the record corresponding to a^{ij}_m which can be done automatically by the accounting machine itself.

This technique will not give us all the cofactors we need if some of the a_{ij}'s in the original matrix are zero, because in the evaluation of $\det(a_{ij})$ we have not punched the term cards corresponding to the $a_{ij} = 0$. In the case where there are some $a_{ij} = 0$, the following procedure may be used in the generation of term cards:

1. Select all $a_{ij} \neq 0$ permutation cards as before.
2. Of the permutation cards with at least one $a_{ij} = 0$, select those in which exactly one $a_{ij} = 0$ appears.
3. Generate the term cards as before. except that in the master gang punching step use a 12 punch to identify those term cards corresponding to an $a_{ij} = 0$.
4. Use these new cards to obtain A_{ij} corresponding to $a_{ij} = 0$. It is no longer necessary to factor out the a_{ij} because the 12 punch will be the extra punch, and it does not correspond to any coefficient.

Example

The following determinant arose in the evaluation of the integral:

$$I = \frac{1}{2\pi i}\int_{-\infty}^{\infty} dx \frac{\sum_{i=0}^{7} b_i x^{14-2i}}{\left\{\sum_{i=0}^{8} a_i x^{8-i}\right\}\left\{\sum_{i=0}^{8} (-1)^i a_i x^{8-i}\right\}}$$

$$D = \begin{vmatrix} A_0 & A_2 & A_4 & A_6 & 0 & 0 \\ 0 & A_0 & A_2 & A_4 & A_6 & 0 \\ 0 & 0 & A_0 & A_2 & A_4 & A_6 \\ A_1 & A_3 & A_5 & A_7 & 0 & 0 \\ 0 & A_1 & A_3 & A_5 & A_7 & 0 \\ 0 & 0 & A_1 & A_3 & A_5 & A_7 \end{vmatrix}$$

where

$$A_r = \sum_{v=0}^{r} a_v x^{r-v} .$$

If the above evaluation method were applied to the determinant in this form, approximately 130,000 term cards and 24,576 selection cards would be required. However, by judicious manipulation of rows and columns, D can be written in the form:

$$D = \begin{vmatrix} a_0 & a_2 + a_1 x & a_4 + a_3 x & a_6 + a_5 x & 0 & 0 \\ 0 & a_0 & a_2 + a_1 x & a_4 + a_3 x & A_6 & 0 \\ 0 & 0 & a_0 & a_2 + a_1 x & A_4 & A_6 \\ a_1 & a_3 - a_1 x^2 & a_5 - a_3 x^2 & a_7 - a_5 x^2 & 0 & 0 \\ 0 & a_1 & a_3 - a_1 x^2 & a_5 - a_3 x^2 & a_7 & 0 \\ 0 & 0 & a_1 & a_3 - a_1 x^2 & a_5 & a_7 \end{vmatrix}$$

This form of D was evaluated by the above procedure. Notice that the zero elements are the same in both forms but that the degree of the elements appearing in the first four columns has been reduced. In this form, approximately 3,000 term cards and 576 selection cards were required. The negative signs in columns 2, 3, and 4 were handled quite easily in the master gang punching step by double punching an X in the corresponding position of the term number. An extra crossfooting operation was used to determine the correct sign of the term card. The final answer involved about 340 terms associated with 19 different powers of x. The entire problem required about 13 hours of machine time.