CHAPTER IV

NON-LINEAR METHODS OF PARAMETER OPTIMIZATION:
GENERAL PRINCIPLES
CHAPTER IV

NON-LINEAR METHODS OF PARAMETER OPTIMIZATION: GENERAL PRINCIPLES

1. Preliminaries.

As stated before, the integrated rate equations pertaining to concentration and system property changes during consecutive reactions are non-linear and transcendental, i.e., cannot be solved by using simple algebraic manipulations. Nor can they be linearized by taking logarithms. They require time-consuming, cumbersome calculations if direct methods, such as the method of successive approximations, are employed for the evaluation of parameters from experimental data which are assumed to obey the system equation. When the data are accurate, with no systemic or random errors being present, but only round-off errors, strictly speaking, only as many readings will be required as there are parameters in the equation. Thus, for example, in the case of the simple first-order reaction, if $A_0$ is accurately known, just a single accurate reading will suffice to evaluate $k$, in the case of data of product formation in a two-step, consecutive, first-order reaction, just two accurate readings will suffice to evaluate both $k_1$ and $k_2$, again if $A_0$ is accurately known. This will require a numerical method, which falls under the purview of numerical analysis, which is a branch of mathematics: the results that are obtained are accurate to a number of digits which depends on the number of digits to which the data were accurate and on the type of the system equation. It is possible generally to arrive at these results by assuming some initial values of the same and adopting a scheme of calculations, usually presented in the form of an algorithm, which is repeated a number of times: at the end of each cycle of calculations, known as an iteration, the initial value is corrected, the corrections
diminishing in magnitude with successive iterations and the estimates of the parameters steadily approach their final values. The procedure is terminated when the desired level of accuracy in the digits of the parameters, within a specified ‘tolerance value’, has been attained. The numerical procedure is said to have converged. Obviously, convergence could be expected to quicker, requiring fewer iterations if the initially assumed values are closer to the final values of the parameters. There are occasions, however, when, for a variety of reasons, convergence is not observed: the corrections effected at the end of each iteration, instead of becoming successively smaller, become larger and the results move rapidly away from the finally expected values, leading to the failure or breakdown of the computational process.

For solving the same equation for its parameters or roots, a number of different numerical methods are usually available and these generally differ in their efficiency and also in their convergence criteria, some routines performing better than others under a given set of conditions. Some of the numerical methods, such as the Bolzano or bisection method for determining the roots of an equation, are of ancient origin [71]. A number of numerical techniques were invented by mathematicians like Newton, Seidel and Gauss, and many of these have been modified and improved upon by subsequent workers Good surveys of these numerical methods are available [47, 66, 71 - 76] These routines can be readily programmed for running on computers

When the data contain random experimental/observational errors, it will not be possible to estimate the parameters accurately by a numerical method A statistical approach will be needed, normally, many more readings are taken than would be needed if the data were accurate, in the hope or belief that during the computations, the errors would cancel out one another to a large extent, leading to estimates which are quite close to their “true” values. The parameters are then said to have been optimized, i.e., their “best”, under the circumstances, though not necessarily accurate, values have been obtained. The various statistical techniques
of parameter evaluation are known broadly as regression or optimization methods. Excellent treatises and surveys of various optimization procedures are available [77-80]. These too can be programmed for running on computers, enabling the saving of much of the time and labour involved in manual calculations, especially when the number of readings is large and the computations are intricate [81].

Many of these techniques have found their way into the ever-expanding field of operations research [82 - 85]. Originally developed by British scientists concerned principally with the optimum allocation of limited resources of war materials during World War II, it now deals with a variety of problems where decision making depends on the determination of the optimum or best course of action, and it has used simulation techniques in a large way.

Most of the optimization methods are based on some numerical technique or other. In addition, they employ a criterion, such as the Least Squares norm, to specify when the parameters have been optimized. The number of datum points/observations must be greater than the number of parameters to be optimized in order to minimize the impact of the errors therein. An objective function is defined, based on the chosen norm, for the minimization of the errors in the data. A cycle of computations, which starts with initial inputs or rough guesses of the parameters which are sought to be optimized, is repeated a number of times. At the end of each iteration, corrections are effected in their values, the magnitudes of the corrections diminishing with each successive iteration. If the corrections decrease rapidly, convergence is achieved quickly; if the corrections diminish slowly, convergence requires a large number of iterations. In some cases, towards the end, the iterates (estimates at the end of the iteration) may not converge to a limit, but oscillate about a mean value. And, quite often, due to wrong choice of initial inputs, there may be no convergence, successive iterates becoming increasingly larger. Breakdown of the computational process will be also encountered if a step at some stage involves "division by zero", storing an intermediate numerical result that is beyond the capacity of the computer to hold ("overflow"), or taking the square-root or logarithm of a quantity which has

53
turned out to be negative ("illegal function call") Such failures can be met by restarting the computations with (hopefully) better initial inputs of the parameters.

It is absolutely essential that the computational program is written correctly, for even a small error, including a slightly wrong sequencing of the program lines, can lead to a failure of the program (in which case remedial measures can be taken) or, worse, despite the error, produces outputs which appear to be reasonable but are actually wrong and may not be recognized as such, and so, accepted in good faith. Hence the necessity to perform sufficient test runs of the program with various inputs and to check the correctness of the outputs by independent methods.


Assuming a situation where all systematic errors in an experiment have been eliminated or adjusted for in the computations, there still remain the unavoidable random errors in the data. In the rest of this thesis, it is assumed that such a situation has been achieved and the data contain only indeterminate or random errors. Given that, using such data, it is impossible to arrive at the exact or true values of the various parameters, such as rate coefficients from kinetic data, the question then arises as how best to minimize the impact of these errors and arrive at optimal estimates, that are close to the true values, of these parameters.

Some norm or criterion has to be adopted to specify the situation where the impact of the (random) errors on the estimate of a parameter becomes the
minimum. Several such norms have been prescribed, but the three principal ones have been

1. the minimization of the sum of (absolute) deviations;
2. the minimization of the maximum deviation, also known as the minimax, or Chebyshev norm
and,
3. the minimization of the sum of squares of deviations, also known as the Least Squares (LS) or the Gaussian norm

In each of these procedures, the optimal estimate of the parameter is the one that minimizes the objective function (sum of the deviations, the maximum deviation or the sum of the squares of the deviations) of the observed values for all the datum points from the values computed for each of these points for the assumed values of the parameters in the system equation.

Minimizing the sum of the absolute deviations is logically the simplest norm if the different readings in an experiment can contain errors of different magnitudes, what else could be more reasonable than to accept as optimal that estimate of the parameter which minimizes the sum of the (absolute) errors in the entire datum set? The disadvantage in employing this norm is that the sum of absolute deviations, though is a continuous function, is not differentiable throughout the essential condition for differentiability of a function at a point is that the right-hand derivative and the left-hand derivative at the point must be equal (both in magnitude and sign), absolute quantities do not possess this property and hence are not differentiable [86]. So the minimum cannot be reached by the fast iterative methods which employ derivatives but only by direct search.

In kinetic situations, the experimental data usually consist of a number (N) of observations of the system property Y, which may be a titre value, absorbance reading, etc., collected at different reaction times, t Y is expected to obey the model equation,
\[ Y = f(A, B, C, \ldots, t) \quad \ldots (42) \]

where \( Y \) is a function of several parameters \( A, B, C, \ldots \) (one or more of which can be the rate coefficients), and variable \( t \) (reaction time). If \( \hat{Y} \) is the value of the system property \( Y \) predicted by the model equation for assumed values of \( A, B, C, \ldots \), at \( t \), then this norm stipulates that the optimal values of \( A, B, C, \ldots \) are those which minimize the trial function, also known as the objective function,

\[ F_0 = \sum |(Y - \hat{Y})| = \min \quad \ldots (43) \]

\[ \sum |Y - f(A, B, C, \ldots, t)| = \min \quad \ldots (44) \]

The Chebyshev norm minimizes the deviation of the single reading that deviates most from the value predicted by the system model equation for the assumed estimates of the parameters.

\[ F_0 = (Y - \hat{Y})_{\max} = \min \quad \ldots (45) \]

\[ [Y - f(A, B, C, \ldots, t)]_{\max} = \min \quad \ldots (46) \]

The minimax norm has been very little used in chemistry as such but is widely employed in decision theory and in game theory in management ([82], Ch 10 and 11.) The maximum deviation, again though a continuous function, is not differentiable throughout, and therefore suffers from the same drawback as the sum of absolute deviations norm in that the fast, gradient search methods cannot be employed for finding the minimum.

The sum of squares of deviations (SSD), on the other hand, is a continuous function of the parameter(s) and the variable(s), and is also differentiable throughout; it is therefore eminently suited for reaching the
minimum by fast converging, iterative methods employing derivatives (gradients), these methods can be programmed for quick execution on computers. The objective function to be minimized is

\[ F_0 = \Sigma (Y - \hat{Y})^2 = \min \] .. (47)

i.e.,

\[ \Sigma [Y - f(A, B, C, \ldots, t)]^2 = \min \] (48)

The LS method is attributed to Gauss (1777-1855), but credit for its first publication goes to the French mathematician, Adrien-Marie Legendre (1752-1833). It was Gauss who exploited the LS method to the fullest, put it on a firm footing and popularized it, hence, it is rightly known as the Gaussian norm. The LS norm has now become so universally accepted and dominant, that it is adopted in optimization procedures employing scientific data to the near-total exclusion of the other norms.

In the present work, the least squares (LS) norm is adopted for all the optimizations.

3. Error Surfaces and Maps.

The parameters in reaction kinetics are generally the rate coefficients appearing in the exponential terms of the rate equations, which are non-linear (except in the case of zero-order reactions). Since these have to be estimated from experimental data containing random errors, the problem of error-reduction then involves the optimization of these parameters. This reduces to a task of searching for the minimum of the sum of the squares of the deviations.
The simplest approach, practically, would be a direct search, among all possible combinations of the parameters, for that combination which corresponds to the minimum. In reaction kinetics, X is reaction time and Y may be the concentration of a species that depends on time, or a system property such as optical absorbance, electrical conductance, optical rotation, etc, which is also changing with time. In the simplest case, that of a straight-forward first-order reaction,

\[ k \\
A \rightarrow B \]

where the time-dependence of the concentration of the reactant is given by the non-linear equation

\[ A = A_0e^{-kt}, \]

if \( A_0 \) is known exactly, the task simplifies to one of determining that value of k which leads to the minimum in the function \( F_0 = \sum [A - A_0e^{kt}]^2 \) that has been summed over all the N readings in the \((t, A)\) data. This involves a search in a single dimension, k. If, for all the possible values of k (from 0 to \( \infty \)) are plotted along a \( k \)-axis (say, the X-axis in a graph sheet), and \( F_0 \), i.e., the sum of deviations squared, along the Y-axis, one obtains the error curve which is continuous and will have a minimum in \( F_0 \) that corresponds to the optimized value of the first-order rate coefficient, k. The minimum is readily located and corresponds to the "true" value of the parameter if the data are fairly accurate. If the experimental errors become larger, the minimum in the curve becomes shallower and broader, leading to an increased uncertainty in the estimate of the parameter, further, the minimum gets shifted away from the true value. The net result is that the parameter is estimated with less accuracy as well as with less precision. If more complex situations, additional, shallower minima, known as
local or pseudo-minima, can occur besides the global or true minimum which gives the optimized estimate of the parameter(s).

If \(A_0\), the initial concentration of the reactant in the first-order reaction, is not known accurately, it can be treated as a second parameter to be optimized besides \(k\). This becomes a two-dimensional search. The two parameters, \(A_0\) and \(k\), can be represented along two Cartesian axes, say \(X\) and \(Y\), and the objective function, \(F_0 = \sum [A - A_0 e^{-kt}]^2\), plotted along the third (Z-) axis, one obtains the error surface. It can be mapped on plane paper just as any topographic map, the two parameters are represented along the \(X\)- and \(Y\)-axes while contours are employed to represent points having the same value of the objective function, \(F_0\).

If the readings are accurate, then the error surface has a sharp minimum lying in a valley with steep slopes. The minimum is easy to locate on the error map and its co-ordinates give the optimized values of the two parameters, \(A_0\) and \(k\). As the level of error in the data increases, the valley floor is raised, the minimum becoming shallower and difficult to locate as many points may have the same value of \(F_0\). There is considerable uncertainty or spread in the optimized estimates of the two parameters. Further, the valley as a whole is shifted from its original location in the map with the data where the datum errors have been negligible. That is, besides increased imprecision, the optimized estimates of the parameters become increasingly inaccurate.

In the case of two-step, consecutive, irreversible, first-order reactions, if the kinetics have been monitored in terms of the changes, with respect to reaction time, in the concentration of the intermediate, product or co-product, the integrated rate equations for all these three species contain the two rate coefficients, \(k_1\) and \(k_2\), besides the initial concentration, \(A_0\), of the reactant. If \(A_0\) is known accurately, there remain only \(k_1\) and \(k_2\) as the two parameters to be optimized. In the case of product formation data, the objective function to be minimized is
The error map for product formation data in a two-step, consecutive, first-order reaction.
\[ F_0 = \sum (Y - \hat{Y})^2 = \sum [Y - f(k_1, k_2, t)]^2 = \text{min} \quad (49) \]

where \[ Y = 1 - \gamma = \frac{[k_2 \exp(-k_1t) - k_1 \exp(-k_2t)]}{(k_2 - k_1)} \quad (50) \]

While carrying out the computation of \( \hat{Y} \), and, therefore, of the deviations \( (Y - \hat{Y}) \) for different combinations of \( k_1 \) and \( k_2 \), the two parameters are treated as variables, while the various values of \( X \) (i.e., \( t \)) and the corresponding values of \( Y \) in the given set of data are treated as constants.

For a set of \( N \) readings containing random errors, the error surface can be mapped contour-wise with \( k_1 \) and \( k_2 \) represented along the \( X \)- and \( Y \)-axes of a graph. The optimized estimates of \( k_1 \) and \( k_2 \) are given by the minimum in the floor of the valley. Fig. 2 depicts part of the error map for product concentration data in a two-step, consecutive first order reaction in which \( k_1 = 0.1 \) and \( k_2 = 0.2 \) for 20 readings taken at uniform intervals of time, \( DT = 0.94 \) (units omitted). The data do not contain random errors, i.e., \( \sigma = 0 \), but contain errors due to round-off at the fourth decimal place. The narrow, curved valley has steep slopes in this case and there are two symmetrically located minima, depicted by ‘x’. The effect of the symmetry of eqn (6) for product formation with respect to the interchange of \( k_1 \) and \( k_2 \) is evident in the graph: the contours are symmetrical with respect to "reflection" about the \( k_1 = k_2 \) line. Due to this fact, the two rate coefficients are designated as \( k_{\text{slow}} \) (or \( J \)) and \( k_{\text{fast}} \) (or \( K \)) rather than as \( k_1 \) and \( k_2 \). But maps such are more ornamental in scope, for unless drawn to a large scale depicting many more contours, the full complexity of the error landscape cannot be properly appreciated, especially where the data contain errors.

When these principles are applied to the case of the kinetics of a three-step, consecutive, irreversible, first-order reaction followed in terms of the concentration of a species other than the reactant, the evaluation of the three rate coefficients, \( k_1, k_2 \) and \( k_3 \), involves optimization in three dimensions, representing...
the three rate coefficients, the object function, F0, extending into a fourth dimension; the error surface in the present case is called a hypersurface. It cannot be visualized but can be mapped contour-wise in sections. Nevertheless, the properties and behaviour adduced to the simpler error surface apply to the hypersurface as well. The optimal estimates of the three parameters are given by the minimum in it, which can be located with only decreasing accuracy and precision with increasing level of error in the concentration versus time data.

If the kinetics of a two-step, consecutive, irreversible, first-order reaction are followed by monitoring the change in a system property such as optical absorbance at a given wavelength, rather than the concentration of a species, since the equation contains four parameters, L and M, besides $k_1$ and $k_2$, their evaluation requires optimization in four dimensions. The error hypersurface extends into a fifth dimension along which F0 is represented. For a given set of data, the optimal estimates of the four parameters are given by the minimum in the hypersurface. While error hypersurfaces baffle imagination, they are eminently amenable to the mathematical and statistical treatments applicable to the simpler error curve and error surface, wherein lie the prospects of optimization of the parameters in higher dimensions.

4. Types of Iterative Search Methods.

The process of optimization of the parameters from the given set of data involves the search for the minimum of the objective function. Obviously, the simplest approach is to search directly for the minimum trying out every possible combination of the different parameters. The task in this exhaustive enumeration or brute force approach can be very time-consuming even carried out on a computer, as reaction rate coefficients can theoretically lie in the range 0 to $\infty$. Certainly, some à priori knowledge of the approximate extent of the ranges in which they lie would cut short the work as one can restrict the search to within
these ranges. Hence the importance of initial estimates or inputs. In the case of two-step, consecutive reactions, good initial estimates of both $k_{\text{slow}}$ and $k_{\text{fast}}$ can be obtained by the non-iterative methods, such as that of residuals.

A number of direct, iterative multi-dimensional search methods for locating a function minimum by selecting a search direction in which the objective function is improved (i.e., decreased) and then continuing the search in that direction, re-orienting the direction as and when required, have been described. These direct methods do not involve the use of gradients or derivatives and are most suited for functions for which the derivatives are difficult to compute or contain too many terms. Since a search direction is set up, the minimum is reached more quickly than in the comprehensive search. These include the simplex method [87] (q v), the cyclic coordinate method, Hookes and Jeeves’ method and the Rosenbrock method [88]; faster to converge are Powell’s method of conjugate search directions [89], and its various modifications, which employ the so-called conjugate search directions. Illustrations of these methods, except the simplex, are found in [85]. It is the simplex method, as formulated by Nelder and Mead, which has been employed as the main tool for optimization in the present work.

Indirect methods use derivatives of the function for reaching the minimum, thus optimizing the parameters. When the initial inputs are appropriate, they reach the minimum in just a few steps and so converge much more rapidly than the direct methods. Their main lacuna is that the initial estimates of the parameters must be quite close for convergence to be achieved, otherwise, the corrections to the parameters at each stage increase instead of decreasing, leading to convergence failure.

Indirect methods may be first-order or second-order. First-order methods, which use the first (partial) derivatives only of the objective function with respect to the independent variables (read, the parameters to be optimized), include the
classical gradient or steepest ascent/descent method and the conjugate gradient
method of Fletcher and Reeves [90]. Second-order methods use additional
information provided by the second (partial) derivatives as well. These include
Newton's method and its modifications and the Marquardt method [91].

The indirect method employed in this thesis, as an adjunct to the Nelder-
Mead simplex method, is the Deming [92] method, which is based on the well-
known Newton-Raphson iterative method in numerical analysis ([66] pp 575-
577, [79] pp.197 – 202.) While it is a simple method, it suffers from the
disadvantage that the initial inputs must be quite close, sometimes very close, to
their final values - not necessarily their "true" values, otherwise, as the truncation
of Taylor's approximation, which is employed to attain fast, quadratic
convergence, is valid only in close ranges, fails to hold good and convergence is
not achieved. This is particularly so with "noisy" data, i.e., data containing
significant levels of random error.

Another respect in which all iterative methods, both the direct and the
indirect, may fail arises with the presence of many local minima, and the iterative
procedure, if it is commenced close to a local minimum, may land in it, leading to
wrong estimates of the parameters, instead of converging to the global minimum.
There are "annealing" methods available to extricate the procedure of a local
minimum and make it to find ultimately the global minimum [81] (Sec. 10.9), but
these are not employed in the present thesis. A simpler strategy is adopted,
consisting in starting the iterations with more than one set of inputs and also
testing the minimum by approaching it with inputs lying in a direction opposite to
those first employed.
5. Optimization Techniques in Consecutive Reactions.

Despite so many developments in optimization techniques in the last century, their application to the kinetics of consecutive reactions has not kept pace. The main reason is the extreme paucity of experimental data in the kinetics of consecutive reactions. Apart from Kaufler’s three runs in the hydrolysis of 2,7-dicyanonaphthalene, dating back to 1906, there have been very few reports of kinetics of consecutive reactions followed in terms of the concentration of a species. Only in such cases, the rate equations, of a two-step, consecutive, irreversible, first-order reaction, of the three species (intermediate, product and co-product) contain just two parameters, viz, the two rate coefficients, $k_1$ and $k_2$, to be optimized. Since 1950, almost all the workers have followed the kinetics of consecutive reactions spectrophotometrically and, as the rate equation contains four parameters, many of these workers have preferred to use the simple method of residuals to obtain admittedly approximate estimates of $k_1$ and $k_2$, rather than resorting to optimization in four dimensions. Nevertheless, there have been occasional reports of use of optimization techniques in the kinetics of consecutive reactions:

Cull and Brenner [93] were among the first to demonstrate the power of computer-based, non-linear regression analysis in resolving the kinetics of complex reactions. They analyzed the reversible isomerization of hexane which involves five species; but as they determined the concentrations of all the five species, separated by column chromatography, the method has little relevance to the present work which is directed towards the resolution of the kinetics of consecutive reactions from the data on the concentration of a single species or a system property.

Wiberg [14] in 1964 was the first to use a computer program, written in FORTRAN, for evaluating the two consecutive first-order rate coefficients of the
oxidation of benzaldehyde by chromyl acetate from spectrosphotometric data. Starting with approximate initial estimates of the two rate coefficients as well as of the molar extinction coefficients of the reactant, intermediate and product, these were varied successively in small steps until the average deviation, rather than the sum of the squares of the deviations, was minimized. Convergence of the five parameters was observed after 122 iterations. Wiberg's program perhaps remains the one in reaction kinetics where a maximum number of parameters, viz., k₁, k₂, Eₐ, Eₖ and Eₖ, are sought to be optimized. The program included a modification by Lepse, which permitted the simultaneous variation of more than one parameter at a time. However, even with this modification, the number of iterations required for convergence may, in some cases, very large, especially with the computers of that time. Thus, Mathai et al. [94] found that, in the first trial, up to 500 iterations with 10% changes in the input parameters had to be employed to achieve the best fit of absorbance data pertaining to the kinetics of ethoxide ion-catalyzed isomerization of 1,4-diynes into 1,3-diynes via the dienyne intermediate.

Alcock et al. [13] computed, using a quasi-Newton least squares routine, the three parameters, k₁, k₂ and Eₖ from time-absorbance data of the reaction of nitrous acid with hydrogen peroxide to form nitric acid via the intermediate peroxynitrous acid.

In 1972, London and Hearon [95] presented a method for determining the minimum (and the maximum) number of precursors, which corresponds to the number of distinct steps in a sequence of first-order irreversible reactions by employing the rate of formation of the final product.

Bailey and Eadie [46] developed a non-linear regression method, based on mean transit times, for the evaluation of the rate coefficients of an n-step, consecutive, first-order, irreversible reaction sequence from the rate of formation of the final product. They applied the same to the data of Gale and Eadie, referred
to above, on the two-step hydrolysis of urea to form carbon dioxide, ammonia being the co-product. They incorporated a sub-routine to further minimize the sum of squares of deviations by adjusting the mean transit time where uncertainty existed about the exact moment of initiation of the reaction.

Moodie [15] investigated a routine based on the Kezdy-Mangelsdorf-Swinbourne algorithm [96 - 98] for the evaluation of the two rate coefficients of two-step, consecutive, first-order reactions, and the two apparent rate coefficients in cases where either of the steps is reversible, from spectrophotometric data, which obey eq. (14). In all the three cases, the method requires the collection of 600 absorbance readings at equally-spaced time intervals, stored on a microcomputer. The calculations then make alternate use of the later data (the last 400 points) and the earlier data (the first 60 points), to calculate and refine the estimates of the four parameters by the least-squares norm. They carried out some simulations with computer-generated data, incorporating normal random errors in the absorbance at the 0.1% error level (σ = 0.001). The \( k_{\text{fast}} / k_{\text{slow}} \) ratio was varied in the range 3 to 40 for the three situations (i) where only the intermediate absorbs, (ii) where the product alone absorbs, and (iii) where the molar absorbance of the product is twice that of the intermediate. The values of \( P \) and \( Q \) in eq. (14) were scaled in each case (as is done in this work) to give a total absorbance change of unity. In the limited range of parameters employed, they found that the mean estimates of \( k_{\text{slow}} \) and \( k_{\text{fast}} \) from twenty runs had errors of less than 1%, when the \( k_{\text{fast}} / k_{\text{slow}} \) ratio is above three, the results deteriorate rapidly when this ratio is less than 3 and the method becomes unreliable. He also found the errors in the estimate of \( k_{\text{fast}} \) to increase when \( k_{\text{fast}} \) is large, i.e., 40.

Moodie illustrated the method with the absorbance data pertaining to the hydrolysis of methyl chloroformate, catalyzed by pyridine in dilute aqueous acid, monitored at 278 nm and interfaced to a microcomputer. The absorbance, due to the intermediate, the methoxycarbonylpyridinium ion, rises from zero to pass through a maximum before declining. Though the reaction is complicated by
simultaneous, uncatalyzed hydrolysis, the absorbance data obey eq (14). They estimated the two rate coefficients of the catalyzed reaction to be 3.64E-2 and 3.57E-3 sec\(^{-1}\). They found the rate coefficients to reproduce the data very closely, with a root mean square deviation of 6E-4. The closeness of the fit, they claimed, vouched for the accuracy of the estimates. However, since the maximum absorbance in the illustration is only 0.15 Abs, and the precision of the Unicam SP 3/60 spectrophotometer is ± 0.001 Abs. in this range, the error level in the data works out to 100 x 0.001/0.15, i.e., 0.67%, and it is doubtful whether the errors in their estimates of the rate coefficients would be in the same range (0 - 1%) as found in their simulations which employed data containing only 0.1% random error.

Bisby and Thomas [99] have written a program in BASIC for the evaluation of the two rate coefficients of the two-step, consecutive, first order, irreversible reaction from absorbance data, minimizing the sum of squares of deviations by the Marquardt algorithm [91].

Casado et al. [100] have proposed a method for optimizing \(k_1\), \(k_2\) and \(E_B\) in the same reaction scheme from spectrophotometric data, assuming \(E_A\) and \(E_C\) to be known exactly; using simulated, error-free absorbance data, they showed that their method, based on the Hartley [101] algorithm, converges more quickly than the same initial inputs of \(k_1\), \(k_2\) and \(E_B\) than does the method based on the quasi-Newton algorithm, which sometimes does not lead to convergence. They gave only a few illustrations of their method, and their simulated data did not incorporate any errors, except those of round-off.

Carlson and Provder [102] have used the Nelder-Mead downhill simplex routine [87] for the evaluation of the rate coefficients of a different type of reaction scheme, viz., a first-order reaction followed by one of second order, it illustrated the employment of contour plots and search trajectory for the test data.
during the minimization routine which determines the optimal estimates of the two rate coefficients.

Bacaloglu et al [104] have written a program based on relaxation theory and using a Powell minimization sub-routine [89] for the optimization of the rate coefficients of a multi-step, consecutive, first-order reaction.

Chrastil [104] has provided a program in BASIC for the determination of the two rate coefficients of two step reactions from product concentration vs time data, analyzed on the basis of a binomial, exponential equation,

$$\gamma = (1 - e^{-kt})^n$$  \hspace{1cm} (51)

The program can distinguish whether the reaction is simple first-order ($n = 1$), second-order, higher-order or diffusion-controlled ($n < 1$), or consecutive ($n > 1$), before proceeding to evaluate $k_{\text{slow}}$ and $k_{\text{fast}}$ in the last case.

Osborne [28] has formulated the solution of the general equation for the $n$-step, consecutive, first-order, irreversible reaction as a regression problem and given the computational details of a FORTRAN program, which is illustrated with the data pertaining to a two-step reaction. He cautions that the computations may lead to unsatisfactory estimates of the rate coefficients when applied to a reaction sequence containing more than two steps, because of the instability of the error matrix associated with the solution.

Kalonia and Simonelli [105] in 1989 published the results of a simulation study in which co-product data were generated for 20 readings, extending over 75-85% of co-product formation, without and with the incorporation of random errors at a single level, 2%. They employed three methods, viz., the method of residuals, linear regression and a non-linear, curve-fitting program, EUREKA, which used the steepest descent method, to evaluate the two rate coefficients.
the level of error introduced in the data, they found that the method of residuals could recover the rate coefficients with less than 15% error when $\kappa > 0.5$. When $0.05 < \kappa < 0.5$, there are significant errors in the estimates of the rate coefficients given by the three methods, even EUREKA returning errors of 20-30% in some simulations.

Duggleby and Leonard have, in 1991, developed a BASIC program, DNRPEASY, based on an earlier, published program, DNRP53 [106]. DNRPEASY, which can be downloaded from the Internet, carries out the non-linear regression fit of data in selected or user-supplied equations, it can be used to optimize all the five parameters of equations such as (12) - (14), in cases where $Z_\infty$ is not known and is treated as the fifth parameter to be optimized. It employs the Marquardt [91] algorithm for optimization of the parameters and prints the standard errors in each as well.

Much before all these, in 1931, Lowry and Traill [107] had studied the mutarotation of 0.01 molar aluminium benzoylecamphor in carbon tetrachloride at 20°C using light of wavelength 564 1 nm.

<table>
<thead>
<tr>
<th>t, min</th>
<th>0 9</th>
<th>1 1</th>
<th>1 7</th>
<th>2 1</th>
<th>2 3</th>
<th>2 8</th>
<th>3 1</th>
<th>3 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_i$ (obsd)</td>
<td>962</td>
<td>1031</td>
<td>1100</td>
<td>1159</td>
<td>1200</td>
<td>1234</td>
<td>1259.5</td>
<td>1287</td>
</tr>
<tr>
<td>$a_i$ (L &amp; T)</td>
<td>933</td>
<td>984</td>
<td>1106</td>
<td>1164</td>
<td>1190</td>
<td>1236</td>
<td>1260.5</td>
<td>1285</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t, min</th>
<th>4 0</th>
<th>4 5</th>
<th>4 9</th>
<th>5 3</th>
<th>5 8</th>
<th>6 2</th>
<th>6 7</th>
<th>7 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_i$ (obsd)</td>
<td>1301</td>
<td>1319</td>
<td>1327</td>
<td>1333.5</td>
<td>1338</td>
<td>1340</td>
<td>1340</td>
<td>1338</td>
</tr>
<tr>
<td>$a_i$ (L &amp; T)</td>
<td>1306</td>
<td>320.5</td>
<td>1328</td>
<td>1334</td>
<td>1338</td>
<td>1339.5</td>
<td>1340</td>
<td>1338</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t, min</th>
<th>8 0</th>
<th>8 5</th>
<th>9 2</th>
<th>10 1</th>
<th>11 4</th>
<th>13 2</th>
<th>14 8</th>
<th>16 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_i$ (obsd)</td>
<td>1335.5</td>
<td>1332</td>
<td>1326</td>
<td>1322.5</td>
<td>1312</td>
<td>1300</td>
<td>1291</td>
<td>1281.5</td>
</tr>
<tr>
<td>$a_i$ (L &amp; T)</td>
<td>1335</td>
<td>1332</td>
<td>1327.5</td>
<td>1321</td>
<td>1312</td>
<td>1300</td>
<td>1291</td>
<td>1282</td>
</tr>
</tbody>
</table>

69
Mutarotations are not simple two-step, consecutive reactions, but ones in which both the steps are reversible; yet their system properties obey eq (14). By a direct method, Lowry and Traill fitted the optical rotation data into the equation

$$\alpha_t = \alpha_\infty + 258 \exp(-0.115t) - 867 \exp(-0.541t)$$

(52)

where 0.115 and 0.541 min\(^{-1}\) are the apparent, rather than the individual, rate coefficients of the two steps. Their estimates of the rotation calculated at the various reaction times are listed below the observed readings. There is close agreement between them except for the first two readings.

<table>
<thead>
<tr>
<th>t, min</th>
<th>18.4</th>
<th>20.0</th>
<th>23.7</th>
<th>29</th>
<th>37</th>
<th>51</th>
<th>(\infty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_t) (obsd.)</td>
<td>1275.5</td>
<td>1271</td>
<td>1261</td>
<td>1253.5</td>
<td>1247.5</td>
<td>1245</td>
<td>1244.5</td>
</tr>
<tr>
<td>(\alpha_t) (L&amp;T)</td>
<td>1275.5</td>
<td>1270</td>
<td>1261</td>
<td>1253.5</td>
<td>1248</td>
<td>1245</td>
<td>-</td>
</tr>
</tbody>
</table>