Inorganica Chimica Acta, 18 (1976) 237–239

An Improved Computer Program for the Computation of Formation Constants from Potentiometric Data

P. GANS,* A. SABATINI and A. VACCA
Istituto di Chimica Generale dell'Università, Via J. Nardi 39, 50132 Firenze, Italy
Received October 22, 1975

An improved version of the computer program MINIQUAD is described. The new program, MINIQUAD 75, is faster by a factor of 2–3 and more reliable in extreme cases. The programs are compared on a benchmark consisting of eight chemical problems. Improvements in speed and reliability open the way to the development of a sound model selection procedure.

Introduction

The computation of formation constants relating to complicated solution equilibria is a process that requires, in practice, multiple computations, each one of which postulates a different chemical model for the equilibria obtaining. Selecting the most appropriate model is perhaps the most difficult part of the excercise, involving statistics and chemical judgement.[1] For this reason it is important that each computation should be efficient, fast and give a mathematically reliable result.

Our computer program MINIQUAD,[2] which is capable of dealing with most types of potentiometric titration data,[3–5] goes some way towards meeting these requirements. We describe briefly in this note improvements to the program which result in both increased speed and reliability.

Calculations and Results

MINIQUAD uses an iterative least-squares refinement process in which the crucial step is the setting-up and solution of the normal equations (1).

$$\begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \tilde{\mathbf{B}}_2 & \mathbf{B}_3 \end{pmatrix} \begin{pmatrix} \mathbf{s}_c \\ \mathbf{s}_\beta \end{pmatrix} = - \begin{pmatrix} \mathbf{g}_c \\ \mathbf{g}_\beta \end{pmatrix} \qquad (1)$$

$\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{B}_3$ are matrix partitions of the symmetric normal equations matrix $\mathbf{B}$ whose form and sparsity pattern have been illustrated before.[6] $\mathbf{s}_c$ and $\mathbf{g}_c$ are the 'shift' and 'gradient' vectors relative to the unknown free concentrations (*e.g.* [M] and [L] at each titration point) and $\mathbf{s}_\beta$ and $\mathbf{g}_\beta$ are the 'shift' and 'gradient' vectors relative to the formation constants β (notation as in ref. 6). The equations (1) can be rearranged to give the equations (2).

$$(\mathbf{B}_3 - \mathbf{B}_2\mathbf{B}_1^{-1}\mathbf{B}_2)\,\mathbf{s}_\beta = -\mathbf{g}_\beta + \tilde{\mathbf{B}}_2\mathbf{B}_1^{-1}\mathbf{g}_c \qquad (2a)$$

$$\mathbf{s}_c = -\mathbf{B}_1^{-1}\mathbf{g}_c - \mathbf{B}_1^{-1}\mathbf{B}_2\mathbf{s}_\beta \qquad (2b)$$

Equation (2b) is solved after equation (2a) and the 'shifts' are added to the current values of those free concentrations and formation constants which are parameters of the refinement. The refinement is made reliable by using that fraction of the shift vector $\{\mathbf{s}_c\ \mathbf{s}_\beta\}$ which optimises the reduction in the sum of squares. This 'linear optimisation' is not, however, an efficient process. In extreme cases, moreover, it does not guarantee reliability since the shift vector may be almost tangential to the iso-value contours of the sum of squares, so that a significant reduction may not be possible. Initially attempts to improve MINIQUAD were made independently in Florence and Leeds, and resulted in programs which we denote "A" and "B".

In the Florentine approach, $\mathbf{s}_\beta$ was calculated from equation (2a) as before, but no shifts were calculated for the free concentrations. Instead the latter were calculated in the subroutine MQ to optimise the fit with all mass-balance equations at each titration point, using the current values of the formation constants β. Since this implies $\mathbf{g}_c = \mathbf{0}$, the end result is clearly identical with that obtained previously. The term $\tilde{\mathbf{B}}_2\,\mathbf{B}_1^{-1}\,\mathbf{g}_c$ can be omitted from equation (2a) with a marginal saving in computer time. $\mathbf{s}_\beta$ may be (repeatedly) halved in order to avoid an increase in the sum of squares.

"A" is a constrained least-squares method in which the 'unknown' free concentration (of metal, ligand, *etc.*) are optimal for every set of formation constants. Since the residuals are linear functions of the formation constants[2] this device has transformed a non-linear minimisation problem into a linear one, which can be

* Visiting Research Fellow. Permanent address: Department of Inorganic and Structural Chemistry, The University, Leeds LS2 9JT, U.K.

expected to converge in one cycle without further iterations. However, non-linearity is still present because the constrained free concentrations are not constant and in practice a few iterations are needed. It is rare that the number of iterations exceeds five. Thus, the constrained optimisation is a very powerful method, particularly when the initial values of the formation constants are poorly estimated. The equations (2a) can be set up titration-point by titration-point so that $\mathbf{B}_1$ and $\mathbf{B}_2$ do not need to be stored in their entirety. The computer store requirements are thus less than for MINIQUAD in which $-\mathbf{B}_1^{-1}\,\mathbf{B}_2$ is stored for use in equation (2b). Against this must be placed the observation that "A" did not function well on the IBM 360 computer in single precision (approximately 7.2 decimal digits). This arises from the need to have compatible convergence criteria in MQ and the main minimisation. There is no difficulty with the ICL 1906A computer working in normal, 12 digit, precision, or with the IBM 360 working in double precision.

In the Leodensian approach, the linear optimisation is eliminated and instead equation (3) is solved directly.

$$(\mathbf{B}+\lambda\,\mathbf{D})\mathbf{s} = -\mathbf{g},\ \mathbf{s} = \{\mathbf{s}_c\mathbf{s}_\beta\},\ \mathbf{g} = \{\mathbf{g}_c\mathbf{g}_\beta\} \qquad (3)$$

$\mathbf{D}$ is a diagonal matrix consisting of the diagonal elements of $\mathbf{B}$. λ is the Marquart parameter whose effect is to rotate the shift vector towards the direction of steepest descent. The direct solution of equation (3) is achieved by using a method that utilises and retains the sparsity pattern inherent in $\mathbf{B}$. Typically $\mathbf{B}$ is a matrix of order 200×200 with less than 5000 non-zero elements. For the choice of λ we follow Fletcher's very efficient strategy.[7] Near the solution λ tends to zero so that "B" becomes a standard un-damped unconstrained least-squares method, with excellent convergence properties.

A direct comparison showed that neither "A" nor "B" was clearly the superior on all types of computation. The two programs have therefore been hybridized to form MINIQUAD 75. In the hybrid program each six iteration cycles consist of an "A" cycle followed by five "B" cycles. Thus the initial refinement power of "A" is complemented by the final convergence speed of "B".

TABLE I. Comparison of Programs' Execution Time (sec)[a].

Problem[b]	n_k[c]	n	n_{mbe}	n_c	n_p	MINIQUAD	"A"	"B"	MINIQUAD 75
1	9	4	3	1	25	6	4	1	2
2	6	5	2	1	145	201	19	53	20
3	6	5	2	1	145	19	9	5	6
4	7	2	3	2	89	31	19	38[g]	18
5	7	2	3	2	89	13	7	4	6
6	11	8	3	2	198	72	71	33	52
7	5	5	3	2	185	fails[d]	56	fails[e]	94[f]
8	15	2	4	3	26	7	10	4	5

[a] Indicated execution time on the ICL 1906A computer at Leeds. Since this machine lacks a "hardware clock" indicated times are accurate to only ± 10%.

[b] Problem 1. Ag^+, $Se(CH_2CO_2H)_2$ using glass and Ag/AgCl electrodes. Data kindly supplied by D.K. Laing.
Problem 2. $(Me_2NCH_2CH_2N(Me)CH_2CH_2)_2NMe$ (tetren).
Initial estimates of β_1, β_2 and β_3 incorrect by more than 2 log units.
Problem 3. As problem 2 but with good initial parameter estimates.
Problem 4. Ni^{2+}, $N(CH_2CH_2CH_2NH_2)_3$ (tpt). Input estimates $\beta_1 = 10^{14}, \beta_2 = 10^6$. *c.f.* ref. 6. Ligand protonation constants determined separately.
Problem 5. As problem 4. "Good" estimates for the β's.
Problem 6. Cu^{2+}, $H_2NC(CH_2OH)_3$. Ligand protonation constants determined separately.
Problem 7. Cd^{2+}, ascorbic acid. Data from P. Ulmgren and O. Wahlberg, *Acta Chem. Scand., 25,* 1079 (1971).
Problem 8. Cu^{2+}, L-phenylalanine, D-histidine, determination of constants for ternary complexes only. Constants for binary complexes and ligand protonation constants determined separately. Data kindly supplied by G. Brookes (ref. 4).

[c] n_k = total number of formation constants, n = number of refinable formation constants, n_{mbe} = number of mass-balance equations at each titration point, n_c = number of "unknown" free concentrations (*i.e.* n_{mbe}–number of electrodes), n_p = number of titration points.

[d] Failure in linear optimisation.

[e] Did not attain the specified convergence criteria.

[f] "A" eliminated one formation constant as negative. With MINIQUAD 75 this constant was retained but had a very high standard deviation.

[g] Converged to false minimum.

A benchmark of eight chemical problems reflecting a wide variety in type and difficulty was established and each program was run on each problem. Resulting computer times are given in Table I.

It is clear that on most problems MINIQUAD 75 is 2–3 times faster than the original MINIQUAD. It is markedly more reliable as shown by the result for the cadmium–ascorbate system, problem 7. This system is computationally difficult because the ionic strength is maintained by the cadmium ions and the degree of complex formation is small and almost constant. It is also clear that MINIQUAD 75 is only slightly less efficient than the best of "A" or "B", though storage requirements are greater than for "A".

The greatest advantage accrues in the process of model selection, where the same data are refined successively with slightly different models. Information, in the form of the 'unknown' free concentrations, is carried over from one model to the next with considerable saving in computing effort. In this way we are able to examine a batch of many models in very much less computer time than would be required if every model were refined individually with the old MINIQUAD. Thus the road is open, for the first time, for a serious consideration of the chemical and statistical factors involved in a sound model selection process. Many models, perhaps hundreds, can be examined, and further experiments can be performed so as to optimise experimental conditions with the approximate knowledge, provided by computation, of the extent of species' formation.

Further details concerning implementation and a FORTRAN listing of MINIQUAD 75 can be obtained from P. G.

Acknowledgments

We thank the C.N.R. (Italy) for a fellowship (to P. G.), L. D. Pettit, D. Laing, G. Brookes, J. Swash and R. Hefford for running their experimental data on development versions of the program. We also thank Prof. L. Sacconi for encouragement and for making continued collaboration possible.

References

1 L.G. Sillén and B. Warnqvist, *Arkiv Kemi, 31,* 341 (1969), *Idem, Ibid, 31,* 377 (1969).
2 A. Sabatini, A. Vacca and P. Gans, *Talanta, 21,* 53 (1974).
3 R.D. Graham and D.R. Williams, *J.C.S. Dalton,* 1123 (1974).
4 G. Brookes and L.D. Pettit, *J.C.S. Chem. Comm.,* 813 (1974).
5 M. Israeli, D.K. Laing and L.D. Pettit, *J.C.S. Dalton,* 2194 (1974).
6 A. Sabatini and A. Vacca, *J.C.S. Dalton,* 1693 (1972).
7 R. Fletcher, *A.E.R.E. Report,* R-6799 (1971).