

MULTIPARAMETRIC CURVE FITTING—VIII

THE RELIABILITY OF DISSOCIATION CONSTANTS ESTIMATED BY ANALYSIS OF ABSORBANCE-pH CURVES

MILAN MELOUN

Department of Analytical Chemistry, College of Chemical Technology, CS-532 10 Pardubice,
 Czechoslovakia

and

MILAN JAVŮREK

Computer Centre, College of Chemical Technology, CS-532 10 Pardubice, Czechoslovakia

(Received 20 November 1984. Revised 29 March 1985. Accepted 19 April 1985)

Summary—The program SPOPT estimates stability constants β_{pq} and molar absorptivities ϵ_{pq} of all light-absorbing species $M_pL_qH_r$ by analysis of the absorbance-concentration (or absorbance-pH) curve. The program DCMINUIT estimates dissociation constants and molar absorptivities of protonated species. Both programs have been tested and compared with DCLET and LETAGROP-SPEFO for analysis of the overlapping equilibria of a triprotic acid. Computer plots of the residual-square-sum function are used to test the conditioning of parameters. Two approaches are made to formulation of the mathematical model, and several optimization algorithms are tested to find a reliable minimization procedure. The accuracy of ill-conditioned parameters is shown to be dependent on the precision of the absorbance measurements. General rules for investigation of A -pH curves are recommended.

The analysis of the absorbance-pH curve for a polyprotic acid to determine dissociation constants and molar absorptivities is not a straightforward procedure. In the application of regression analysis attention must be paid to formulation of a suitable mathematical model for the system and to the choice of an efficient minimization subroutine and minimization strategy. The precision of the spectrophotometric data should be as high as possible, especially for evaluation of ill-conditioned parameters. It must be discovered which parameters are well-conditioned and which ill-conditioned, and finally, criteria must be chosen to characterize the reliability of parameter estimates, and to terminate the minimization process.

This paper considers all these problems, warns against unthinking application of regression analysis, and suggests general rules for use of regression procedures to obtain reliable estimates of dissociation constants and molar absorptivities. Four regression programs are discussed, *viz.* DCLET,¹ LETAGROP-SPEFO² and two new programs, SPOPT and DCMINUIT in which several minimization subroutines and two formulations of the residual-square-sum function are available. The A -pH curve for a triprotic acid involving one overlapping protonation

equilibrium is analysed and the reliability of the computer treatment is assessed.

THEORY

Structure of the regression programs used

Our "old" regression program DCLET,¹ Sillén's LETAGROP-SPEFO,² and our new programs SPOPT and DCMINUIT, were modified to have nearly the same structure (Fig. 1), as previously described for the ABLET regression system.

SPOPT AND DCMINUIT

The RESIDUAL-SQUARE-SUM block

The regression analysis of spectrophotometric data to provide molar absorptivities and stability constants for the individual absorbing species requires the minimization of the residual-square-sum function

$$U = \sum_{i=1}^n w_i (A_{\text{exp},i} - A_{\text{calc},i})^2 \quad (1)$$

where $A_{\text{exp},i}$ is the measured absorbance at a given wavelength for given concentration of the three main components M, L and H, for the equilibria



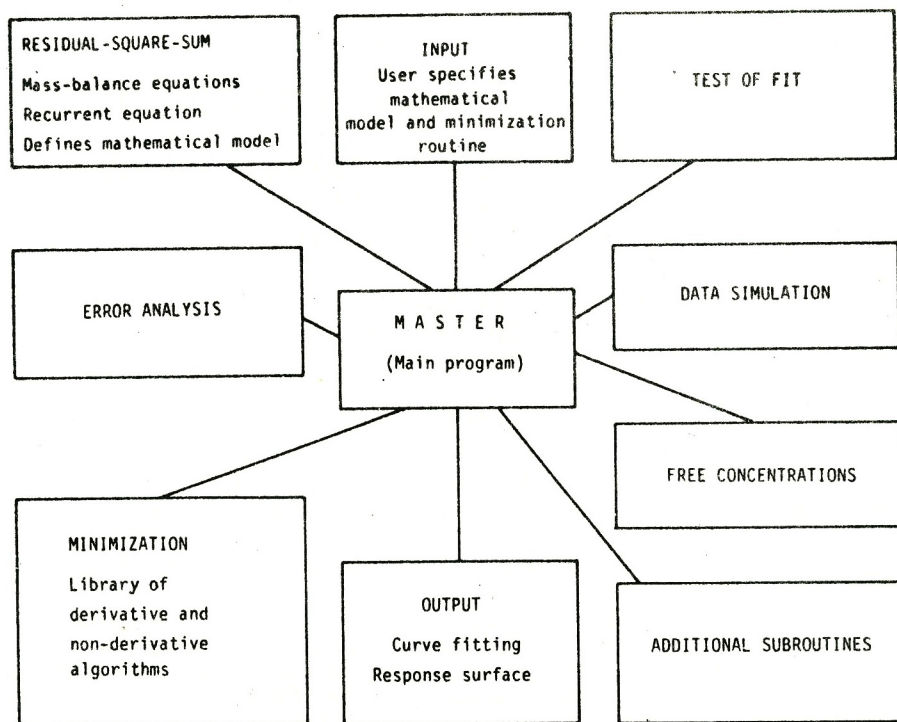


Fig. 1. Functional blocks of the program SPOPT.

with stability constants

$$\beta_{pq} = [M_p L_q H_r] / ([M]^p [L]^q [H]^r) \quad (3)$$

If J species absorb at the given wavelength, the absorbance A of the solution (in a cell of path-length d cm) is given by

$$\begin{aligned} A &= d \sum_{j=1}^J \epsilon_{pqr,j} [M_p L_q H_r]_j \\ &= d \sum_{j=1}^J \epsilon_{pqr,j} (\beta_{pqr} [M]^p [L]^q [H]^r)_j \end{aligned} \quad (4)$$

where $\epsilon_{pqr,j}$ is the molar absorptivity of the j th species and $[M_p L_q H_r]_j$ the free concentration of this species^{4,5} and the mass-balance equations for the three basic components are given by

$$S = \sum_{j=1}^J (\beta_{pqr} [M]^p [L]^q [H]^r)_j \quad (5)$$

$$c_M = [M] + pS \quad (6)$$

$$c_L = [L] + qS \quad (7)$$

$$c_H = [H] + rS \quad (8)$$

U is assumed to be a second-degree function of m unknown parameters in $(m+1)$ dimensional (and hence parametric) space, m being sufficient for calculation of a position for the minimum. Parameters are estimated by optimization in the multiparametric space in which values are given for absorbance and concentrations, and stability constants and molar absorptivities are adjusted by the computer.

There are two approaches to the formulation of U in the SPOPT program, corresponding to two different ways of determining A_{calc} . The first is a general method based on equation (4), in which an adjusted set of stability constants and molar absorptivities, and free concentrations of the components $[M]$ and $[L]$ are calculated ($[H^+]$ is known from pH measurement), and hence A_{calc} . This version of SPOPT is referred to as SPOPT(MB) (mass-balance equation approach). In the second, the A -pH curve for a mononuclear acid is written with the assumption that base L is protonated to form the various forms LH_1, LH_2, \dots, LH_r , etc. of the mononuclear acid LH_r . As discussed previously, the equation for the absorbance-pH curve may be written as

$$A = dc_L \left\{ \frac{\epsilon_L + \sum_{r=1}^R \epsilon_{LH_r} 10^{(r \log a_{H^+} + \log \beta_{01r})}}{1 + \sum_{r=1}^R 10^{(r \log a_{H^+} + \log \beta_{01r})}} \right\} \quad (9)$$

In the expression $(r \log a_{H^+} + \log \beta_{01r})$, the conventional activity pH scale may be used and the protonation constants β_{01r} may be expressed as a function of the mixed stepwise dissociation constant $K_{ai} = a_{H^+} [LH_{i-1}] / [LH_i]$ and so

$$r \log a_{H^+} + \log \beta_{01r} = \sum_{i=1}^r pK_{a,i} - r \text{pH} \quad (10)$$

This version of SPOPT is referred to as SPOPT(DC) (recurrent equation of the A -pH curve for dissociation constants determination).

DCMINUIT also contains a residual-square-sum function formulated by use of equations (9) and (10). Both SPOPT(DC) and DCMINUIT determine dissociation constants $pK_{a,i}$ and SPOPT(MB) determines stability constants or protonation constants $\log \beta_{0qr}$.

The MINIMIZATION block

Regression analysis finds estimates for the unknown parameters $\bar{\beta}_{pqr}$ and $\bar{\epsilon}_{pqr}$ by minimizing the difference between the experimental and calculated data. The general problem is to find the best values of parameters $\beta_{pqr,j}$ and $\epsilon_{pqr,j}$, $j = 1, J$, for which U is minimal. Here, $A_{exp,i}$ is a measured absorbance for the i th point and $A_{calc,i}$ is calculated from $A_{calc,i} = f(pH_i, \bar{\beta}_{pqr}, \bar{\epsilon}_{pqr})$, where pH is known as an experimental quantity. The non-linear estimation problem is really simply a problem of optimization in the parameter space in which the A and pH values are known and the β and ϵ values are the variables. The function U must have its minimum at a point where either (i) all derivatives $\partial U / \partial \beta_j$ are zero ($j = 1, J$), (a stationary point), or (ii) some derivatives $\partial U / \partial \beta_j$ do not exist (a cusp), or (iii) the point β_j is on the boundary of the allowed region (an edge point). When it is realized that there may be any number of stationary points, cusps and edge points, all of which may be arbitrarily hard to find by simple sampling of the function value, the whole problem begins to appear hopeless unless some simplifying assumptions are made. The usual simplification consists of abandoning the attempt to find the global minimum and being satisfied with a local minimum for which the parameters have a physical meaning.

There are 30 minimization algorithms in the SPONA routines library,⁶ and these may be divided into derivative and non-derivative routines. All may be called by SPOPT. If in the search for a minimum of U , the partial derivatives of U with respect to the parameters must be calculated, then the method is classified as derivative; otherwise, it is termed non-derivative.

Derivative methods used and selected for mention here are the Steepest Descent Method,⁷ Gradient Method⁸ and the Conjugate Gradients Method;⁹ the Grid Method,¹⁰ the Simplex Method¹¹ and Rosenbrock's Method¹² represent the non-derivative methods. However, the user is free to use any other minimization routine in the SPONA library.⁶

DCMINUIT incorporates three different minimization methods from the MINUIT regression system,¹³ each of which may be used alone or in combination with the others, depending on the behaviour of U and on the requirements of the user. First, a Monte Carlo searching non-derivative subroutine¹⁴ may be used at the beginning of a minimization when no reasonable initial value for the parameters can be guessed, or when it is suspected that there are several minima. Second, the Nelder and Mead non-derivative simplex method¹¹ is "safe" and

fast when far from a minimum, and may also be used to converge to the exact minimum. Third, a derivative method developed by Fletcher¹⁵ is extremely fast near a minimum or in any "nearly-quadratic" region, but slower if U is badly behaved. It uses the partial first derivatives of U , calculated either analytically or numerically. The program employs some "global" logic; if one method fails, the other is automatically caused to make another attempt. In addition, the minimization can be guided or separated into steps by the input data, which may cause a variable parameter to be fixed at a constant value or restored to variable status between minimization steps.

The ERROR ANALYSIS block

This block finds estimated confidence intervals for the parameters. The partial first derivatives of U for the estimates of the parameters may be calculated analytically or numerically for all the A - pH points. The square roots of the diagonal elements of the covariance matrix are the estimated errors, or standard deviations, of the parameters.

DCMINUIT also prints the correlation coefficients between the parameters (represented by the off-diagonal elements), and the global correlation coefficient for a given parameter, which is the correlation between it and that linear combination of the other parameters most highly correlated with it.¹³

The FITNESS TEST block

This test-of-fit block contains the STATS subroutine;¹⁶ its function has been already described.¹³

The DATA SIMULATION block

This block contains the random error generator in the RANDOM subroutine, and also the additional subroutines SIMUL and NORAND.^{3,18} It calculates a simulated A - pH curve by addition of generated random errors to the calculated precise values of absorbance for given pH values.

The user may select values for the errors in the parameters and for the instrumental error of the spectrophotometer. Then, for given values of the independent variable pH , the precise values of the independent variable A are calculated. Each precisely calculated point is then transformed into a simulated "experimental" one by addition of a random error having Gaussian distribution. The actual distribution of errors generated is then tested, and four statistical moments, Pearson's Chi-Square test¹⁶ and the Hamilton R -factor test¹⁹ are applied.

The FREE CONCENTRATIONS block

This block calculates for each experimental point the free concentrations $[M]$ and $[L]$ from the current set of stability constants β_{pqr} , the chemical composition of the solution and the stoichiometric coefficients pqr of each species, and given value of $[H^+]$. Subroutines COGSNR and CCSCC are used. CCSCC is a "book-keeping" routine of COGSNR,

which is used in SCOGS²⁰ to evaluate free concentrations of species by the Newton-Raphson method.

The ADDITIONAL SUBROUTINES block

This block contains subroutines for format-free reading of integers (READI) and reals (READR).³

The INPUT block

This contains subroutine DATA which reads the experimental data, the independent variable, pH, the dependent variable, A_{exp} , and does some preliminary calculations.²² The measured values of pH_{read} are corrected for any deviation of the glass electrode from Nernstian slope (WK), for any difference in temperature from 298.16 K (WT), and for the liquid-junction potential correction in pH units (WW). WZ is $\text{p}a_{\text{H}^+}$ for the standard buffer solution used for calibration.

$$\begin{aligned} \text{pH}_{\text{corr}} = &[(\text{pH}_{\text{read}} - \text{WZ}) \times 59.16 \\ &\times \text{WT}/(\text{WK} \times 298.16)] \\ &+ \text{WZ} + \text{WW} \end{aligned} \quad (11)$$

The OUTPUT block

This prints the parameter estimates and their standard deviations, and a printer-plotting subroutine PLOTT²¹ makes a graph of experimental and calculated A -pH curves. A graph of U as a function of parameter values in the region of the pit may be plotted by DIGIGRAPH equipment to allow investigation of the conditioning of the parameters in the U function.²³ Computer-drawn plots of the hyperparaboloid response surface ($1 - U$) as a function of any two chosen parameters are of great assistance in deciding whether parameters are ill-conditioned.

Computation

The computations were done with an EC 1033 (500 K) computer and the DCLET,¹ LETAGROP-SPEFO,² DCMINUIT, SPOPT, SQUAD(84),^{26,27} and FA608 + EY608²⁵ programs in the Computing Centre of the College of Chemical Technology, Pardubice, Czechoslovakia.

DISCUSSION

Identification of ill-conditioned parameters

The reliability of any estimates of ill-conditioned parameters depends on the choice of minimization method. The refinement of parameters should lead to a minimum, preferably local rather than global, for which the parameter values have physical meaning. In the analysis of the absorbance-pH curve of a triprotic acid, for example, the seven parameters ϵ_L , ϵ_{LH} , $\text{p}K_{\text{a1}}$, ϵ_{LH_2} , $\text{p}K_{\text{a2}}$, ϵ_{LH_3} , $\text{p}K_{\text{a3}}$ have to be estimated. If $|\text{p}K_{\text{ai}} - \text{p}K_{\text{ai}-1}| > 3$ the protonation equilibria do not overlap, but when $|\text{p}K_{\text{ai}} - \text{p}K_{\text{ai}-1}| < 3$, overlap does occur.

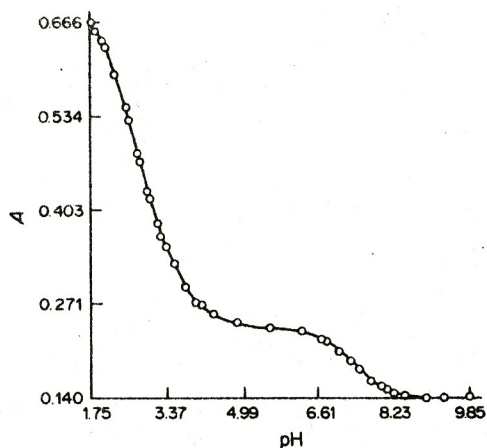


Fig. 2. Experimental A -pH curve for 7-(carboxyphenylazo)-8-hydroxy-quinoline-5-sulphonic acid measured by external titration, with spectrophotometer VSU2-G (Zeiss, Jena, GDR). Experimental conditions: $c_L = 4.18 \times 10^{-5} M$, $\lambda = 540$ nm, path-length 0.998 cm, $I = 0.1$ ($\text{NaClO}_4 + \text{Na}_3\text{PO}_4 + \text{HClO}_4 + \text{EDTA}$), 25° .

As an example, the absorbance-pH curve of 7-(3-carboxyphenylazo)-8-hydroxyquinoline-5-sulphonic acid²⁴ (I) illustrated in Fig. 2, was analysed by the regression program SPOPT(MB). A shortened output appears in Table 1. Since ϵ_L , $\text{p}K_{\text{a1}}$, ϵ_{LH} and ϵ_{LH_3} are well-conditioned in mathematical model (4), a graphical representation of the hyperparaboloid, simplified for two parametric co-ordinates ($m = 2$) in $(m + 1)$ -dimensional space shows a well-developed minimum U_{min} . In Fig. 3 this appears as a maximum of $(1 - U_{\text{min}})$. The shape of the hyperparaboloid for the ill-conditioned parameters is a rather flat-bottomed saucer; this pit cannot be improved and also cannot be reached by any minimization method (Figs. 4-6).

The search for true estimates of the parameters, then, cannot give a certain answer, and no method is able safely to find a pit in U . Careful choice of a minimization subroutine and strategy is necessary, because some algorithms will fail, some will lead quickly to the global minimum, and others will terminate at local minima in dependence on the initial guesses for the parameters.

Choice of minimization algorithm

From the 30 minimization algorithms included in SPOPT (265 K) three derivative and three non-derivative methods were selected. The non-derivative method LETAG from DCLET¹ (82 K), LETAGROP-SPEFO² (240 K) and three minimization methods of DCMINUIT (150 K), were compared with selected algorithms of SPOPT.

The derivative Gradient⁸ method and the non-derivative Grid,¹⁰ Rosenbrock¹² and Simplex¹¹ methods of SPOPT and the derivative Fletcher method¹⁵ of DCMINUIT, are algorithmic in nature and readily find the global minimum, $U_{\text{min}} = 3.60 \times 10^{-4}$ (Table 2). The derivative Steepest

Table 1. Non-linear regression of the experimental A -pH curve of (I) by algorithmic minimization by the Gradient method⁸ of the SPOPT(MB) program; the free concentrations of the protonated species are expressed as percentages of the total concentration c_L

Minimization terminated at $U_{\min} = 3.60 \times 10^{-4}$ with $s(A) = 0.00352$								
Estimated values of parameters:								
EPS(L) = 3290 ± 52								
EPS(LH) = 5731 ± 81								
PKA1 = 7.34 ± 0.06								
EPS(LH2) = 7028 ± 181								
PKA2 = 3.96 ± 0.08								
EPS(LH3) = 16706 ± 18								
PKA3 = 2.73 ± 0.08								
i	pH_i	A_{exp}	A_{calc}	Residual	[L], %	[LH], %	[LH ₂], %	[LH ₃], %
1	1.565	0.6600	0.6721	-0.0121	0.00	0.03	7.18	92.80
2	1.750	0.6660	0.6594	0.0066	0.00	0.07	10.58	89.36
3	1.817	0.6530	0.6536	-0.0006	0.00	0.09	12.13	87.79
4	2.000	0.6400	0.6336	0.0064	0.00	0.19	17.36	82.45
5	2.058	0.6310	0.6259	0.0051	0.00	0.24	19.35	80.41
6	2.244	0.5930	0.5963	-0.0033	0.00	0.52	26.83	72.66
7	2.500	0.5470	0.5436	0.0034	0.00	1.37	39.42	59.21
8	2.550	0.5300	0.5319	-0.0019	0.00	1.64	42.06	56.31
9	2.750	0.4830	0.4828	0.0002	0.00	3.24	52.45	44.31
10	2.788	0.4710	0.4732	-0.0022	0.00	3.66	54.31	42.03
11	2.956	0.4280	0.4318	-0.0038	0.00	6.10	61.54	32.35
12	3.000	0.4200	0.4214	-0.0014	0.00	6.92	63.10	29.98
13	3.185	0.3820	0.3807	0.0013	0.00	11.36	67.65	20.99
14	3.250	0.3660	0.3678	-0.0018	0.00	13.34	68.39	18.27
15	3.364	0.3500	0.3474	0.0026	0.00	17.38	68.53	14.08
16	3.518	0.3270	0.3239	0.0031	0.00	24.01	66.41	9.57
17	3.772	0.2950	0.2945	0.0005	0.01	37.52	57.82	4.64
18	4.000	0.2730	0.2761	-0.0031	0.02	51.14	46.62	2.21
19	4.082	0.2720	0.2709	0.0011	0.03	56.02	42.29	1.66
20	4.369	0.2570	0.2574	-0.0004	0.08	71.49	27.87	0.57
21	4.872	0.2450	0.2455	-0.0005	0.30	88.76	10.87	0.07
22	5.569	0.2380	0.2393	-0.0013	1.63	96.00	2.36	0.00
23	6.266	0.2320	0.2320	0.0000	7.77	91.78	0.45	0.00
24	6.691	0.2220	0.2210	0.0010	18.36	81.49	0.15	0.00
25	6.750	0.2200	0.2188	0.0012	20.49	79.38	0.13	0.00
26	7.056	0.2050	0.2047	0.0003	34.28	65.67	0.05	0.00
27	7.295	0.1920	0.1912	0.0008	47.50	52.48	0.02	0.00
28	7.500	0.1800	0.1792	0.0008	59.20	40.79	0.01	0.00
29	7.740	0.1640	0.1665	-0.0025	71.60	28.39	0.00	0.00
30	8.000	0.1550	0.1558	-0.0008	82.11	17.89	0.00	0.00
31	8.072	0.1530	0.1534	-0.0004	84.41	15.59	0.00	0.00
32	8.250	0.1470	0.1486	-0.0016	89.08	10.92	0.00	0.00
33	8.464	0.1440	0.1446	-0.0006	93.03	6.97	0.00	0.00
34	8.915	0.1400	0.1401	-0.0001	97.42	2.58	0.00	0.00
35	9.316	0.1400	0.1385	0.0015	98.96	1.04	0.00	0.00
36	9.855	0.1400	0.1378	0.0022	99.70	0.30	0.00	0.00
Statistical analysis of residuals:								
Residual mean = $7.65E-10$								
Mean residual = 0.00214								
Standard deviation = 0.00316								
Skewness = -1.023								
Curtosis = 7.336								
Pearson's $\chi^2 = 4.00$								
Hamilton R -factor = 0.008302								
CPU time (sec) = 1505.9								

Descent⁷ and Conjugate Gradients⁹ methods are algorithmic in nature, but did not reach the global minimum and terminated at the local minima $U_{\min} = 4.76 \times 10^{-4}$ and $U_{\min} = 4.013 \times 10^{-4}$. LETAG³ and LETAGROP,² and the algorithm MINUIT¹³ allow use of both heuristic (trial-and-error) and algorithmic minimization processes.

Generally, since algorithmic procedures lead mostly to the global minimum, heuristic minimization is used to allow the computer "to keep processing" near a local minimum which has a physical meaning and is supported by a preliminary graphical analysis. The local minimum may correspond to a higher value of U than the global value. The test for degree of fit

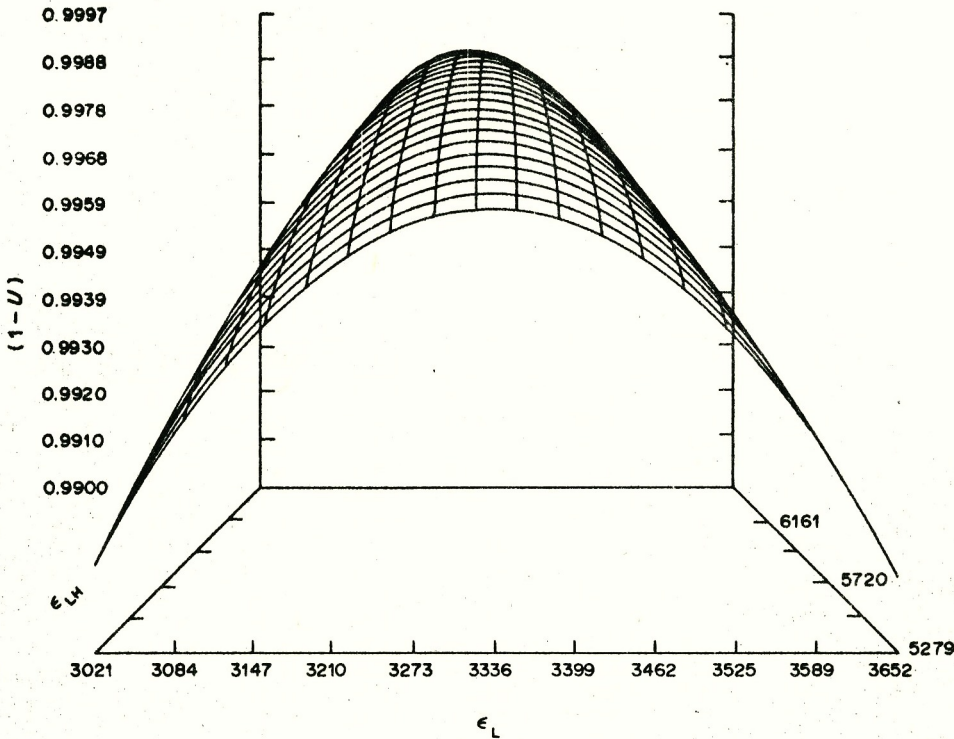


Fig. 3. The minimum of U for m parameters (for LH, $m = 7$) may be represented by a pit in the hyperparaboloid surface in $(m + 1)$ -dimensional space. A three-dimensional representation of the surface is often helpful. Here, the response surface $(1 - U)$ is drawn for two well-conditioned parameters, ϵ_{LH} and ϵ_L .

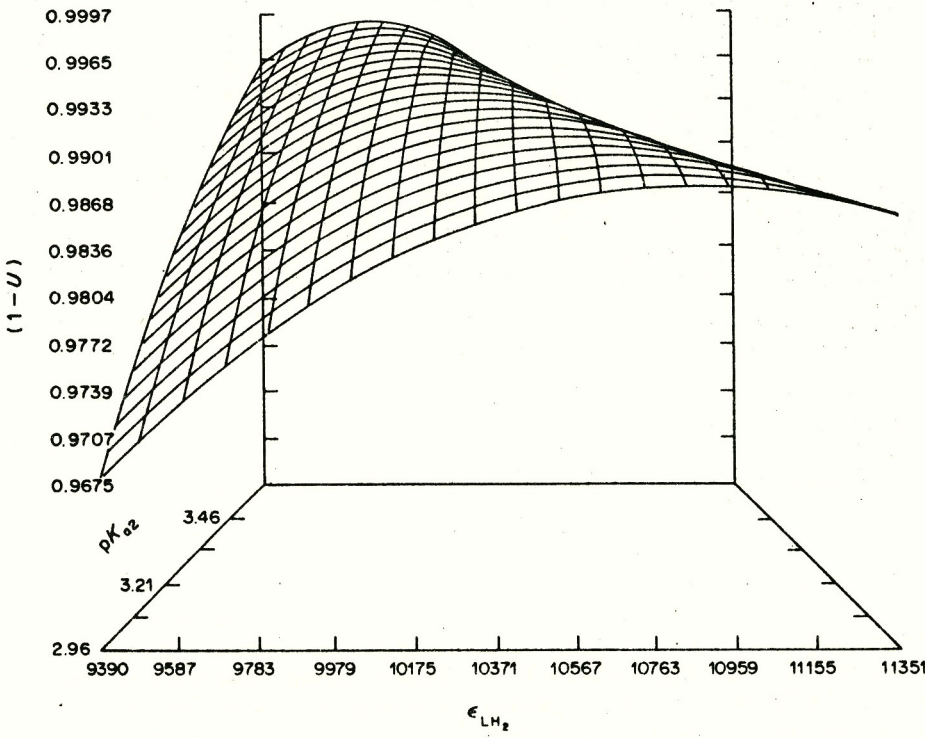


Fig. 4. Response surface $(1 - U)$ for the ill-conditioned parameters pK_{a2} and ϵ_{LH_2} .

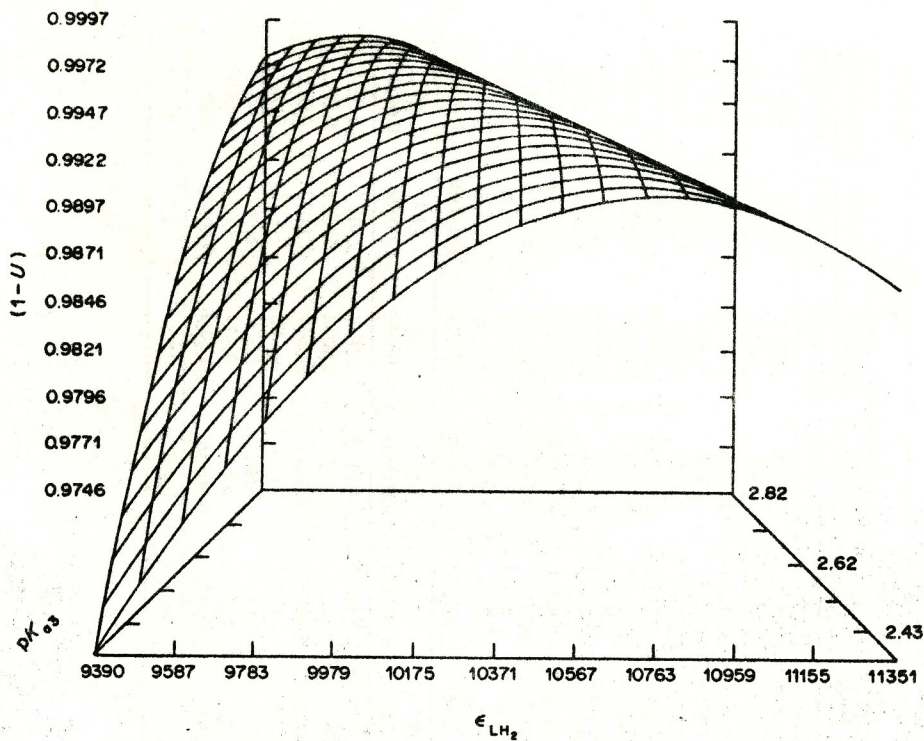


Fig. 5. Response surface $(1-U)$ for the ill-conditioned parameters pK_{a3} and ϵ_{LH_2} .

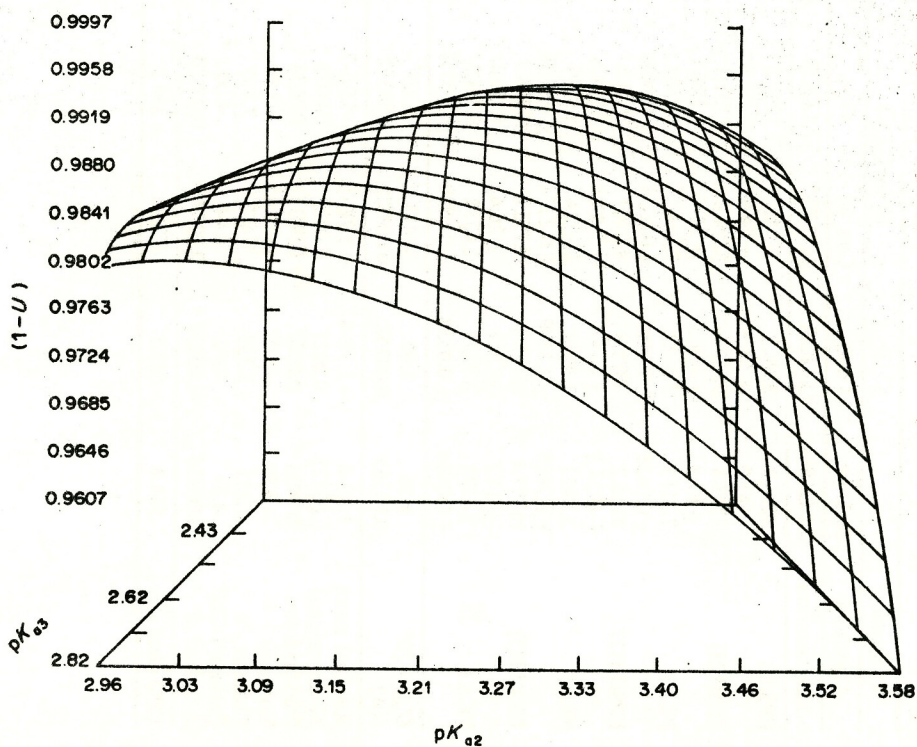


Fig. 6. Response surface $(1-U)$ for the ill-conditioned parameters pK_{a2} and pK_{a3} .

Table 2. Non-linear regression of the experimental A -pH curve, with algorithmic (ALGOR), heuristic (HEURIST) or combination minimization methods in SPOPT(DC), DCMINUIT, DCLET¹ and LETAGROP-SPEO² (data are from Table 1)

Program	SPOPT(DC)						DCMINUIT				DCLET		SPEO	
	Algorithm	St. Desc.	Grad.		Con. G.		Simplex		Rosenb.		Comb.	HEURIST	Comb.	HEURIST
			ALGOR	ALGOR	ALGOR	ALGOR	ALGOR	ALGOR	ALGOR	ALGOR				
Minimization														
EPS(L)	3294.9 ± 61.0 5763.6	3290.2 ± 53.3 5731.1	3295.5 ± 64.9 5781.2	3290.2 ± 53.3 5731.1	3290.2 ± 53.3 5731.1	3290.2 ± 53.3 5731.1	3290.2 ± 53.3 5731.1	3290.2 ± 53.3 5731.1	3290.2 ± 53.3 5731.1	3290.2 ± 53.3 5731.1	3292.9 ± 0.0 5765.8	3292.9 ± 0.0 5765.8	3292.9 ± 0.0 5765.8	3295.6 ± 31.6 5770.3
EPS(LH)	5763.6 ± 67.8 7321.5	5731.1 ± 106.5 73383	5781.2 ± 53.4 73134	5731.1 ± 106.5 73383	5731.1 ± 106.5 73383	5731.1 ± 106.5 73383	5731.1 ± 106.5 73383	5731.1 ± 106.5 73383	5731.1 ± 106.5 73383	5731.1 ± 106.5 73383	5765.8 ± 0.0 73235	5765.8 ± 0.0 73235	5765.8 ± 0.0 73235	5770.3 ± 33.8 73184
PKA1	7321.5 ± 0.0626 8166.2	73383 ± 0.0669 7028.4	73134 ± 0.0636 8779.7	73383 ± 0.0669 7028.4	73383 ± 0.0669 7028.4	73383 ± 0.0669 7028.4	73383 ± 0.0669 7028.4	73383 ± 0.0669 7028.4	73383 ± 0.0669 7028.4	73383 ± 0.0669 7028.4	73235 ± 0.0004 8167.1	73235 ± 0.0004 8167.1	73235 ± 0.0004 8167.1	73184 ± 0.1182 8149.0
EPS(LH2)	8166.2 ± 1124.2 3.5953	7028.4 ± 664.4 39600	8779.7 ± 1283.6 34714	7028.4 ± 664.4 39597	7028.4 ± 664.4 39597	7028.4 ± 664.4 39597	7028.4 ± 664.4 39597	7028.4 ± 664.4 39597	7028.4 ± 664.4 39597	7028.4 ± 664.4 39597	8167.1 ± 0.3 35932	8167.1 ± 0.3 35932	8167.1 ± 0.3 35932	8149.0 ± 59.0 35925
PKA2	3.5953 ± 0.2608 16737.4	39600 ± 0.3379 16706.4	34714 ± 0.2274 16752.0	39597 ± 0.3378 16706.5	39597 ± 0.3378 16706.5	39597 ± 0.3378 16706.5	39597 ± 0.3378 16706.5	39597 ± 0.3378 16706.5	39597 ± 0.3378 16706.5	39597 ± 0.3378 16706.5	35932 ± 0.0003 16737.8	35932 ± 0.0003 16737.8	35932 ± 0.0003 16737.8	35925 ± 0.1278 16731.0
EPS(LH3)	16737.4 ± 24.2 2.6578	16706.4 ± 20.1 2.7263	16752.0 ± 26.6 2.6218	16706.5 ± 20.1 2.7263	16706.5 ± 20.1 2.7263	16706.5 ± 20.1 2.7263	16706.5 ± 20.1 2.7263	16706.5 ± 20.1 2.7263	16706.5 ± 20.1 2.7263	16706.5 ± 20.1 2.7263	16737.8 ± 0.2 2.6596	16737.8 ± 0.2 2.6596	16737.8 ± 0.2 2.6596	16731.0 ± 42.9 2.6625
PKA3	2.6578 ± 0.0654	2.7263 ± 0.0358	2.6218 ± 0.0800	2.7263 ± 0.0358	2.7263 ± 0.0358	2.7263 ± 0.0358	2.7263 ± 0.0358	2.7263 ± 0.0358	2.7263 ± 0.0358	2.7263 ± 0.0358	2.6596 ± 0.0070	2.6596 ± 0.0070	2.6596 ± 0.0070	2.6625 ± 0.1466
$U_{\min} \times 10^4$	4.76 0.00365	3.60 0.00352	4.013 0.00372	3.60 0.00352	3.60 0.00352	3.60 0.00352	3.60 0.00352	3.60 0.00352	3.60 0.00352	3.60 0.00352	4.31 0.00385	4.31 0.00385	4.31 0.00385	3.82 0.00365
Residual mean	1.1E-6	-1.3E-8	-3.7E-5	-5.9E-7	-5.8E-10	-4.4E-10	-4.4E-10	-4.4E-10	-4.4E-10	-4.4E-10	-1.6E-5	-1.6E-5	-1.6E-5	2.5E-9
Mean residual	0.00221	0.0021	0.0023	0.0021	0.0021	0.0021	0.0021	0.0021	0.0021	0.0021	0.0022	0.0022	0.0022	0.0022
Standard dev.	0.0033	0.0032	0.0033	0.0032	0.0032	0.0032	0.0032	0.0032	0.0032	0.0032	0.0033	0.0033	0.0033	0.0033
Skewness	-0.96	-1.02	-1.14	-1.08	-1.02	-1.02	-1.02	-1.02	-1.02	-1.02	-1.05	-1.05	-1.05	1.05
Curtosis	7.481	7.335	7.437	7.337	7.336	7.336	7.336	7.336	7.336	7.336	7.602	7.602	7.602	7.335
Pearson's Chi ²	4.89	4.00	4.44	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.44	4.44	4.44	4.00
R-factor	0.00856	0.00830	0.00876	0.00830	0.00830	0.00830	0.00830	0.00830	0.00830	0.00831	0.00908	0.00908	0.00908	0.00856
CPU time (sec)	564.4	563.1	220.0	554.3	299.9	572.9	299.9	299.9	572.9	134.1	380.1	380.1	380.1	64.16

usually cannot help much, because that achieved for the local minimum seems to be nearly the same as that for the global one.

To test the algorithms selected, the initial guesses of parameters from graphical analysis were $\varepsilon_L = 3300$, $\varepsilon_{LH} = 5800$, $pK_{a1} = 7.3$, $\varepsilon_{LH_2} = 9800$, $pK_{a2} = 3.3$, $\varepsilon_{LH_3} = 16800$, $pK_{a3} = 2.55$. The routines of SPOPT(DC) were applied several times with different sets of initial guesses and most led to the same global minimum, $U_{\min} = 3.60 \times 10^{-4}$. This global minimum was also reached when mass-balance equations were used in the residual-square-sum function of SPOPT(MB) (Table 1).

In DCMINUIT, global logic was used (Table 2, Combination, ALGOR.); when one of the three minimization methods fails, another automatically makes another attempt. Minimization was guided by user's commands from data and three minimization routines of MINUIT¹³ were used (Table 2, Combination, HEURIST.). The algorithmic strategy leads to the global minimum obtained with SPOPT, but the heuristic strategy terminates at another local minimum, $U_{\min} = 4.31 \times 10^{-4}$.

DCLET¹ uses the heuristic strategy of LETAG,³ and gave a local minimum of $U_{\min} = 4.58 \times 10^{-4}$. A combined strategy (i.e., heuristic followed by algorithmic) gave better refinement of the parametric values and a slightly better fit was achieved, $U_{\min} = 3.67 \times 10^{-4}$.

LETAGROP-SPEFO² has a heuristic trial-and-error strategy totally under control of the user. Values for the minimization steps of all the parameters to be refined must be supplied, and good initial guesses are required to avoid divergence of the iterative process. Here, a local minimum of $U_{\min} = 3.82 \times 10^{-4}$ was achieved, with a good fit to the curve, and the parameter estimates had a physical meaning.

Because the four programs terminate with refined estimates that differ in value, the user must decide which program to choose as the best for getting reliable parameter estimates. It would be useful to classify the various regression algorithms in application. Two approaches were used: (1) comparison with results of analysis of *A*-pH curves, (2) study of simulated data.

For comparison, 23 absorbance-pH curves for 19 wavelengths were analysed by the FA608 + EY608 program;²⁵ the results were $pK_{a1} = 7.32 \pm 0.03$, $pK_{a2} = 3.28 \pm 0.03$, $pK_{a3} = 2.60 \pm 0.02$ with $s(A) = 0.0030$.²⁴ SQUAD(84)²⁷ gave $pK_{a1} = 7.332 \pm 0.066$, $pK_{a2} = 3.206 \pm 0.173$, $pK_{a3} = 2.503 \pm 0.294$ with $s(A) = 0.0022$. The absorbance matrix for 19 wavelengths gives more information about the protonation equilibria studied, so the estimates of the constants should be more accurate than those found by an analysis at one wavelength only. The results found from the absorbance matrix appear to be in agreement with the parameters corresponding to the local minimum.

To select reliable algorithms for analysis of *A*-pH curves, simulated data were investigated.

Modelling absorbance-pH curves by use of simulated data

The use of simulated *A*-pH data allows the analysis of real experimental data to be tested in a situation where the parameters are known. Such modelling serves (i) to test whether each parameter is ill-conditioned or well-conditioned in the mathematical model in question, (ii) to examine the influence of the instrumental error of the spectrophotometer on the precision and accuracy of the parameter estimates, and (iii) to find a reliable minimization routine which, for a similar shape of experimental *A*-pH curve, will give the best parameter estimates in the shortest time.

"True" values of seven parameters were chosen to correspond roughly to the experimental parameters for (I): $\varepsilon_L = 3300$, $\varepsilon_{LH} = 5800$, $\varepsilon_{LH_2} = 9800$, $\varepsilon_{LH_3} = 16800$, $pK_{a1} = 7.3$, $pK_{a2} = 3.3$ and $pK_{a3} = 2.6$. The instrumental error, $s_{\text{inst}}(A)$ was taken as 0.003. For set pH values, 35 absorbance values were calculated precisely, then loaded with random errors. The random errors had a normal distribution with the mean approximately zero, the standard deviation of the mean 0.003, the mean error 0.003, skewness 0, kurtosis 3 and a Pearson Chi-Square value of 12.60 for 6 degrees of freedom and 0.95 significance level. These errors appear to be Gaussian in nature (Table 3).

In the analysis of the simulated *A*-pH curve, the minimization was started either with a bad initial guess ($\varepsilon_L = 6000$, $\varepsilon_{LH} = 8000$, $pK_{a1} = 8.2$, $\varepsilon_{LH_2} = 12000$, $pK_{a2} = 3.7$, $\varepsilon_{LH_3} = 20000$, $pK_{a3} = 3.0$) or a good initial guess (the corresponding values being 3300, 5800, 7.3, 9800, 3.3, 16800, 2.6).

The Conjugate Gradients algorithm⁹ of SPOPT(MB) found the best estimates of all seven parameters either by the recurrent equation method or by formulation on the basis of the mass-balance equations (Tables 3 and 4).

The hyperparaboloid response surface shows that three parameters, ε_{LH_2} , pK_{a2} and pK_{a3} , are ill-conditioned, because the minima are broad and indefinite (Figs. 3–6). They cannot be determined accurately or precisely. The last *U*-contour (the "D boundary") may be expressed as the supercurve $U = U_{\min} + s^2(A)$, so the standard deviation in each parameter b_i is defined by the expression $s(b_i) = \max(b_D - b_{\min})_i$ which is the maximum difference between the value for b_i at any point on the "D boundary", and the value for b_i at the minimum. The standard deviations for the ill-conditioned parameters $s(\varepsilon_{LH_2})$, $s(pK_{a2})$ and $s(pK_{a3})$ have significantly greater values than those for the well-conditioned parameters. Because the response surface resembles a flat-bottomed saucer, there is a large amount of uncertainty in the location of the pit.

Table 3. Non-linear regression of a simulated A -pH curve by the Conjugate Gradients algorithm⁹ of SPOPT(DC), with $s_{\text{inst}}(A) = 0.003$

True values of parameters				Estimated values of parameters		
EPS(L) = 3300				3275 ± 30		
EPS(LH) = 5800				5811 ± 13		
PKA1 = 7.3				7.3 ± 0.03		
EPS(LH2) = 9800				9851 ± 2716		
PKA2 = 3.3				3.3 ± 0.31		
EPS(LH3) = 16800				16776 ± 52		
PKA3 = 2.6				2.60 ± 0.19		
<i>i</i>	pH	A_{accurate}	Error	A_{exp}	A_{calc}	Residual
1	1.450	0.68248	-0.00222	0.6803	0.6817	-0.0014
2	1.610	0.67426	-0.00100	0.6733	0.6735	-0.0002
3	1.770	0.66294	0.00350	0.6664	0.6622	0.0042
4	1.930	0.64766	-0.00286	0.6448	0.6471	-0.0023
5	2.090	0.62756	0.00030	0.6279	0.6272	0.0007
6	2.250	0.60200	-0.00244	0.5996	0.6018	-0.0022
7	2.410	0.57084	0.00188	0.5727	0.5708	0.0019
8	2.570	0.53472	0.00111	0.5358	0.5349	0.0009
9	2.730	0.49514	0.00138	0.4965	0.4955	0.0010
10	2.890	0.45434	-0.00242	0.4519	0.4549	-0.0030
11	3.050	0.41474	-0.00043	0.4143	0.4155	-0.0012
12	3.210	0.37850	0.00287	0.3814	0.3794	0.0020
13	3.370	0.34704	-0.00251	0.3445	0.3479	-0.0034
14	3.530	0.32099	0.00657	0.3276	0.3219	0.0057
15	3.690	0.30027	-0.00226	0.2980	0.3010	-0.0030
16	3.850	0.28433	0.00165	0.2860	0.2851	0.0009
17	4.010	0.27239	0.00288	0.2753	0.2731	0.0022
18	4.170	0.26363	0.00119	0.2648	0.2642	0.0006
19	4.330	0.25730	0.00057	0.2579	0.2579	0.0000
20	4.490	0.25276	0.00018	0.2529	0.2533	-0.0004
21	4.580	0.25081	-0.00352	0.2473	0.2514	-0.0041
22	4.670	0.24919	-0.00365	0.2455	0.2497	-0.0042
23	4.760	0.24784	0.00005	0.2479	0.2484	-0.0005
24	4.850	0.24672	-0.00416	0.2426	0.2473	-0.0047
25	5.235	0.24347	0.00169	0.2452	0.2440	0.0012
26	5.620	0.24109	0.00356	0.2447	0.2417	0.0030
27	6.005	0.23772	0.00380	0.2415	0.2384	0.0031
28	6.390	0.23112	0.00215	0.2333	0.2321	0.0012
29	6.775	0.21846	0.00297	0.2214	0.2198	0.0016
30	7.160	0.19856	0.00250	0.2011	0.2002	0.0009
31	7.545	0.17584	-0.00138	0.1745	0.1770	-0.0025
32	7.930	0.15779	-0.00138	0.1564	0.1582	-0.0018
33	8.315	0.14715	-0.00116	0.1460	0.1468	-0.0008
34	8.700	0.14194	-0.00258	0.1394	0.1412	-0.0018
35	9.085	0.13963	0.00161	0.1412	0.1387	0.0025
Statistical analysis of random errors:				Statistical analysis of residuals:		
Error mean = 2.41E-04				Residual mean = -1.14E-04		
Mean error = 0.00218				Mean residual = 0.00203		
Standard deviation = 0.00255				Standard deviation = 0.00246		
Skewness = 0.471				Skewness = -0.041		
Curtosis = 2.437				Curtosis = 2.482		
Chi ² (6, 0.95) = 5.46				Chi ² (6, 0.95) = 4.09		
Hamilton <i>R</i> -factor = 0.00660				Hamilton <i>R</i> -factor = 0.00633		

To test whether a model represents the data adequately, the residuals are analysed. These should be randomly distributed about the predicted regression curve, and systematic departures from randomness indicate that either the model or the parametric estimates are not satisfactory. To analyse the residuals, their statistics are compared with the statistics of the imposed random errors; it is checked whether both distributions are Gaussian in nature, and whether the errors agree in magnitude and/or sign. The degree of fit of the curves in Table 4 is good

enough, so the minimization process is assumed to have terminated successfully. Table 4 also shows how the instrumental error of absorbance affects the precision and accuracy of the parameter estimates, other things being equal.

Errors in absorbance cause systematic errors in the parametric estimates: the relative systematic error of the parametric estimates $e_{\text{rel}}(b_i)$ depends on the instrumental error $s_{\text{inst}}(A)$ approximately according to $e_{\text{rel}}(b_i) = q + k s_{\text{inst}}(A)$. The intercept q and slope k were calculated for seven parameters from the data of

Table 4, and found to be

Parameter b_i	ϵ_L	ϵ_{LH}	pK_{a1}	ϵ_{LH_2}	pK_{a2}	ϵ_{LH_3}	pK_{a3}
Intercept q	0.0091	-0.0164	0.0093	-4.9550	1.1236	-0.0998	2.1967
Slope k	195.79	29.45	92.18	6464.50	-1513.8	34.88	-2367.34

The higher values of the slope for the three ill-conditioned parameters ϵ_{LH_2} , pK_{a2} and pK_{a3} illustrate the greater importance of absorbance precision for accuracy of these estimates. Thus, for elucidation of overlapping equilibria, only high-precision data are suitable. Application of non-linear regression to imprecise data is very likely to lead to false values for the parameter estimates.

Tables 5 and 6 list the methods that did not fail, and which terminated reasonably quickly, selected from 15 derivative and 15 non-derivative algorithms of SPOPT, two of DCMINUIT, one of DCLET¹ and one of LETAGROP-SPEFO.² Derivative algorithms were tested with the use of good and bad initial guesses, but the non-derivative ones with just bad guesses. Good initial guesses were used in order to find a true minimum for comparison with any minimum found from a bad initial guess.

Starting from a good initial guess, the lowest value of U_{\min} (2.086×10^{-4} , perhaps representing a global minimum) was found by the Gradient and Rosenbrock methods. Close to this was $U_{\min} = 2.093 \times 10^{-4}$, found by the Steepest Descent, Conjugate gradients and Grid methods.

With good guesses, the relative systematic errors of the estimates found for the well-conditioned

and ill-conditioned parameters had nearly the same magnitude: $e_{\text{rel}}(\epsilon_{LH_2}) = 0.4\%$, $e_{\text{rel}}(pK_{a2}) = 0.2\%$, $e_{\text{rel}}(pK_{a3}) = 0.1\%$ for the ill-conditioned and $e_{\text{rel}}(\epsilon_L) = 0.8\%$, $e_{\text{rel}}(\epsilon_{LH}) = 0.1\%$, and $e_{\text{rel}}(\epsilon_{LH_3}) = 0.4\%$ for the well-conditioned parameters. Starting from a bad initial guess the algorithms found the same estimates as before only for the well-conditioned parameters. The relative systematic errors of the ill-conditioned parameters were about 10 times greater than those for the well-conditioned ones: $e_{\text{rel}}(\epsilon_{LH_2}) = 10\%$, $e_{\text{rel}}(pK_{a2}) = 3.4\%$ and $e_{\text{rel}}(pK_{a3}) = 2.7\%$.

Figure 7 demonstrates that well-conditioned parameters are not loaded by significant systematic errors, but the broader confidence interval for estimation of the ill-conditioned parameters indicates some uncertainty in the estimates.

The statistical analysis of the residues in Tables 5 and 6 shows that a good fit was achieved by most of the algorithms. The mean residual and its standard deviation are less than $s_{\text{inst}}(A) = 0.003$. The Hamilton R -factor is 0.6–0.7%. From a mathematical point of view, a satisfactory fit indicates that the parameter estimates have been refined sufficiently.

Table 4. Non-linear regression of the simulated A -pH curve by the derivative Conjugate Gradients algorithm⁹ of SPOPT(DC); data are simulated for various instrumental errors of the absorbance, $s_{\text{inst}}(A)$; bad initial guesses of the parameters were used

$s_{\text{inst}}(A)$	0.000001	0.0005	0.001	0.002	0.004	0.006
Error mean	1.40E-7	6.98E-5	1.40E-4	2.79E-4	5.58E-4	8.37E-4
Mean error	8.69E-7	0.00043	0.0009	0.0017	0.0035	0.0052
Standard deviation	9.71E-7	0.00049	0.0010	0.0019	0.0039	0.0058
Skewness	0.561	0.561	0.561	0.561	0.561	0.561
Curtosis	1.689	1.689	1.689	1.689	1.689	1.689
Pearson's Chi ²	9.57	9.57	9.57	9.57	9.57	9.57
R -factor	3.0E-6	0.00125	0.00251	0.00501	0.01002	0.01502
EPS(L) = 3300	3300	3303	3307	3314	3325	3339
EPS(LH) = 5800	5800	5801	5801	5802	5806	5809
PKA1 = 7.3	7.30	7.30	7.31	7.31	7.32	7.34
EPS(LH2) = 9800	9800.5	10191	9859	10453	12326	12854
PKA2 = 3.3	3.30	3.26	3.20	3.24	3.09	3.06
EPS(LH3) = 16800	16800	16800	16791	16790	16812	16816
PKA3 = 2.6	2.60	2.57	2.60	2.55	2.38	2.31
U_{\min}	2.93E-11	7.30E-6	3.00E-5	1.81E-4	4.67E-4	1.05E-3
$s(A)$	1.00E-6	0.00051	0.00103	0.00205	0.00408	0.00613
Residual mean	-8.29E-11	-3.06E-7	1.28E-6	-5.91E-6	6.85E-6	-6.27E-6
Mean residual	8.44E-7	0.00042	0.0008	0.0017	0.0033	0.0050
Standard deviation	9.15E-7	0.00046	0.0009	0.0018	0.0037	0.0055
Skewness	0.158	0.141	0.185	0.154	0.154	0.147
Curtosis	1.574	1.578	1.598	1.588	1.599	1.599
Pearson's Chi ²	11.86	8.66	14.14	11.86	10.03	8.66
R -factor	0.00090	0.00118	0.00237	0.00473	0.00942	0.01413
CPU time (sec)	150.02	411.44	173.98	305.42	159.46	279.52

Table 5. Non-linear regression of the simulated A -pH curve by various derivatives algorithms of SPOPT(DC) and DCMINUIT

Program	SPOPT(DC)						DCMINUIT			
	Steepest Descent		Gradient		Conjugate Gradient		Fletcher		Combination	
Algorithm	Initial	Guess	Bad	Good	Bad	Good	Bad	Good	Bad	Good
EPS(L) = 3300	3260.4	3259.6	3258.9	3258.9	3260.0	3275.4	3316.8	3318.5	3323.9	3313.9
	± 28.2	± 27.8	± 27.6	± 27.6	± 28.0	± 30.0	± 2.2	± 2.6	± 6.9	± 1.9
EPS(LH) = 5800	5810.9	5806.9	5803.5	5803.5	5809.2	5811.3	5794.8	5800.0	5805.4	5797.8
	± 12.6	± 13.0	± 13.8	± 13.8	± 12.7	± 13.0	± 1.5	± 1.3	± 2.6	± 0.7
PKA1 = 7.30	7.3323	7.3345	7.3364	7.3364	7.3332	7.3297	7.3270	7.3234	7.3150	7.3300
	± 0.0265	± 0.0260	± 0.0258	± 0.0258	± 0.0262	± 0.0263	± 0.0014	± 0.0000	± 0.0022	± 0.0000
EPS(LH2) = 9800	10842.8	9774.3	8949.5	8949.4	10333.8	9851.5	8707.6	9826.4	11242.6	9854.6
	± 2688.8	± 2557.6	± 2335.3	± 2335.3	± 2643.3	± 2716.3	± 4.2	± 4.0	± 8.3	± 7.5
PKA2 = 3.3000	3.1978	3.3087	3.4174	3.4174	3.2473	3.2988	3.4618	3.3096	3.1693	3.3070
	± 0.2428	± 0.2957	± 0.3446	± 0.3446	± 0.2665	± 0.3075	± 0.0001	± 0.0000	± 0.0009	± 0.0002
EPS(LH3) = 16800	16792.1	16781.9	16773.3	16773.3	16787.4	16776.3	16755.4	16769.3	16791.3	16771.7
	± 56.5	± 49.5	± 45.0	± 45.0	± 52.9	± 51.7	± 2.5	± 2.3	± 6.5	± 5.3
PKA3 = 2.60	2.5245	2.6025	2.6554	2.6554	2.5632	2.5985	2.6698	2.5974	2.4869	2.5949
	± 0.2197	± 0.1769	± 0.1447	± 0.1447	± 0.1988	± 0.1898	± 0.0003	± 0.0001	± 0.0016	± 0.0000
$U^{(0)}$	7.7E-1	7.67E-4	7.74E-1	7.67E-4	7.74E-1	7.67E-4	7.72E-1	2.97E-4	7.72E-1	2.97E-4
U_{\min}	2.11E-4	2.09E-4	2.08E-4	2.08E-4	2.10E-4	2.10E-4	2.71E-4	2.67E-4	2.64E-4	2.67E-4
$s(A)$	0.00275	0.00273	0.00273	0.00273	0.00274	0.00274	0.00310	0.00308	0.00307	0.00308
Residual mean	7.64E-7	-1.21E-6	-1.91E-9	4.23E-10	-4.45E-6	-1.14E-4	4.49E-6	1.58E-6	-2.52E-7	1.64E-5
Mean residual	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020	0.0025	0.0025	0.0025	0.0025
Standard deviation	0.0024	0.0024	0.0024	0.0024	0.0025	0.0025	0.0028	0.0027	0.0027	0.0027
Skewness	0.099	0.008	0.008	0.082	0.082	-0.041	0.215	0.182	0.141	0.213
Curtosis	2.531	2.446	2.416	2.416	2.486	2.482	1.629	1.595	1.584	1.607
Pearson's χ^2	8.66	2.26	1.34	1.34	3.63	4.09	14.60	13.23	11.86	14.14
R -factor	0.00633	0.00630	0.00629	0.00629	0.00632	0.00633	0.00717	0.00712	0.00707	0.00712
CPU time (sec)	560.9	550.0	545.2	510.3	80.6	75.5	70.1	53.9	110.9	57.5

Table 6. Non-linear regression of the simulated A -pH curve by various non-derivative minimization algorithms of SPOPT(DC), DCMINUIT, DCLET and LETAGROP-SPEFO; the process was started with bad initial guesses of the parameters

Program	SPOPT(DC)			DCMINUIT		DCLET		SPEFO
Algorithm	Grid	Simplex	Rosenbr.	Simplex	LETAG-alg.	LETAG-comb.	LETAGROP	
EPS(L) = 3300	3259 ± 28	3413 ± 141	3259 ± 28	3331 ± 5	2573 ± 34	3322 ± 15	3324 ± 41	
EPS(LH) = 5800	5807 ± 13	3276 ± 832	5803 ± 14	5805 ± 12	5128 ± 43	5787 ± 3	5755 ± 59	
PKA1 = 7.30	7.33 ± 0.026	7.93 ± 0.97	7.336 ± 0.026	7.315 ± 0.0146	8.35 ± 0.02	7.325 ± 0.003	7.32 ± 0.22	
EPS(LH2) = 9800	9840 ± 2543	5826 ± 13	8945 ± 2334	11243 ± 20	8733 ± 31	7823 ± 1	5842 ± 23	
PKA2 = 3.30	3.30 ± 0.29	7.41 ± 0.14	3.41 ± 0.34	3.109 ± 0.008	3.56 ± 0.00	3.63 ± 0.00	5.84 ± 0.22	
PKA3 = 2.60	16784 ± 50	16694 ± 25	16773 ± 45	16791 ± 29	15818 ± 0	16741 ± 0	16668 ± 31	
	2.60 ± 0.18	2.83 ± 0.004	2.66 ± 0.14	2.486 ± 0.008	2.98 ± 0.00	2.72 ± 0.0000	2.83 ± 0.22	
$U^{(0)}$	0.774	0.774	0.773	0.771	0.773	0.773	0.773	
U_{\min} (terminated) $\times 10^4$	2.093	2.608	2.086	2.638	195.7	2.647	3.191	
Residual mean	-1.63E-6	-2.58E-5	4.064E-9	-2.01E-5	-6.92E-5	2.04E-5	-3.99E-11	
Mean residual	0.0020	0.0020	0.0020	0.0025	0.0021	0.0025	0.0026	
Standard deviation	0.0024	0.0027	0.0024	0.0027	0.0023	0.0027	0.0030	
Skewness	0.076	0.566	0.789	0.121	-0.498	0.316	0.188	
Curtosis	2.438	3.836	2.416	1.595	1.773	1.855	1.630	
Pearson's χ^2	2.26	9.11	1.34	8.20	25.11	10.49	9.11	
R-factor	0.00630	0.00704	0.00629	0.00708	0.06099	0.00709	0.00779	
CPU time (sec)	545.2	330.5	530.2	97.0	83.3	90.9	446.8	

CONCLUSIONS

To determine ill-conditioned parameters in an overlapping equilibrium system, an efficient min-

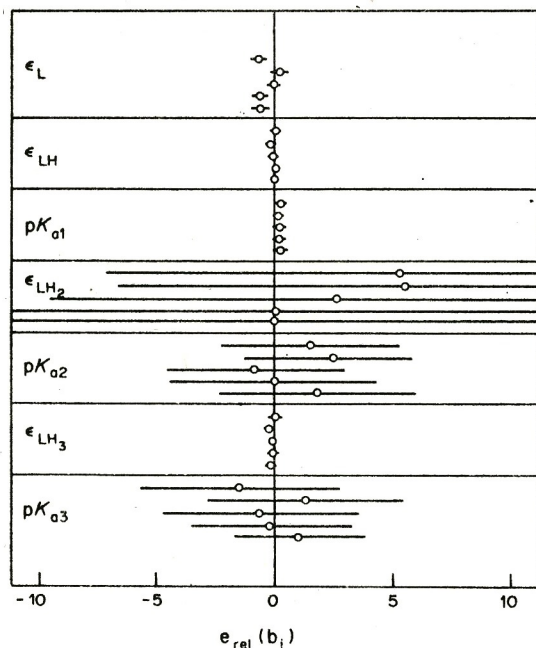


Fig. 7. Relative systematic errors of seven parameter estimates refined by the five minimization algorithms (Tables 5 and 6, from top to bottom) Steepest Descent, Fletcher, Conjugate Gradients, Grid and Rosenbrock methods. The positions of the parameter estimates and corresponding standard deviations are plotted on a relative percentage scale.

imization routine is needed, which is safe, fast, and able to find true estimates of the parameters. Study of the effects of instrumental error shows that data of the highest precision should be used; otherwise, estimates of the ill-conditioned parameters are not accurate enough. A criterion for determination of the minimization process should be decided with the use of a simulated data set.

Diagnostic tools such as (i) graphical interpretation of the hyperparaboloid response surface, in parametric co-ordinates, (ii) examination of the degree of fit by statistical analysis of the residuals, (iii) comparison of the calculated parametric estimates with the preselected true values, (iv) comparison of the statistics of the imposed errors with those of the residuals, (v) the size of the standard deviations calculated from the last U contour (D boundary), help in the examination of the conditioning of parameters in a particular model, and to determine the best minimization strategy, the termination criterion, etc.

The following procedure is recommended for regression analysis.

- (1) Formulate a suitable mathematical model.
- (2) Choose and test (by use of simulated data) a minimization algorithm suitable for safe determination of ill-conditioned parameters.
- (3) Set up computational conditions such as parameter limits, minimization steps, residuals statistics for best parametric estimates, an efficient minimization strategy and a termination criterion for the algorithm selected.
- (4) Analyse the experimental A -pH curve.

REFERENCES

1. M. Meloun and J. Čermák, *Talanta*, 1979, **26**, 569.
2. L. G. Sillén and B. Warnqvist, *Ark. Kemi*, 1969, **31**, 377.
3. M. Meloun and J. Čermák, *Talanta*, 1984, **31**, 947.
4. M. Meloun and J. Havel, *Computation of Solution Equilibria, Part I, Spectrophotometry, Folia Fac. Sci. Nat. Univ. Purk. Brunensis*, Brno, 1985.
5. J. Havel and M. Meloun, in *Computational Methods for the Determination of Stability Constants*, D. J. Leggett (ed.), Plenum Press, New York, 1985.
6. L. Lukšan, *Research Report N. V-4*, Czechoslovak Academy of Sciences, Prague, 1976.
7. W. Murray, *Numerical Methods for Unconstrained Optimization*, Academic Press, London, 1972.
8. B. V. Shah, R. J. Buehler and O. Kempthorne, *SIAM J. Appl. Math.*, 1964, **12**, 74.
9. M. R. Hestenes and E. Stiefel, *J. Res. Natl. Bur. Stds.*, 1952, **49**, 409.
10. R. Hooke and T. A. Jeeves, *J. Assoc. Comp. Mach.*, 1961, **8**, 212.
11. J. A. Nelder and R. Mead, *Comput. J.*, 1965, **7**, 308.
12. H. H. Rosenbrock, *ibid.*, 1960, **3**, 175.
13. F. James, *Comp. Phys. Commun.*, 1975, **10**, 343.
14. *Idem*, *Proc. 1972 CERN Computing and Data Processing School*, CERN 72-21.
15. R. Fletcher, *Comput. J.*, 1970, **13**, 317.
16. A. Sabatini, A. Vacca and P. Gans, *Talanta*, 1974, **21**, 53.
17. M. Meloun and M. Javůrek, *ibid.*, 1984, **31**, 1083.
18. J. R. Bell, *Commun. Assoc. Comp. Mach.*, 1968, **11**, 498, Algorithm 334.
19. W. C. Hamilton, *Statistics in Physical Science*, Ronald Press, New York, 1964.
20. I. G. Sayce, *Talanta*, 1968, **15**, 1397.
21. C. F. Moore, *Comp. Phys. Commun.*, 1971, **2**, 53.
22. M. Meloun and J. Pancl, *Collection Czech. Chem. Commun.*, 1976, **41**, 2365.
23. M. Javůrek, *Thesis*, College of Chemical Technology, Pardubice, 1985.
24. M. Meloun and J. Chýlková, *Collection Czech. Chem. Commun.*, 1979, **44**, 2815.
25. J. J. Kankare, *Anal. Chem.*, 1970, **42**, 1322.
26. D. J. Leggett and W. A. E. McBryde, *ibid.*, 1975, **47**, 1065.
27. M. Meloun, J. Havel and M. Javůrek, *Talanta*, 1985, in the press.