CHROM. 4485

THE IDENTIFICATION OF A SMALL AMOUNT OF A SUBSTANCE OVERLAPPED BY A LARGE AMOUNT OF ANOTHER SUBSTANCE

S. TURINA and M. KAŠTELAN-MACAN

University Institute of Inorganic and Analytical Chemistry, Ul. soc. revolucije 8, Zagreb (Yugoslavia)

SUMMARY

The recognition of a small amount of component B overlapped by a large amount of a similar component A is possible by means of numerical analysis of the chromatographic peak. The authors describe a new method based on the comparison of the experimental curve with the theoretical Gaussian curve. The numerical analysis has been successfully applied to an asymmetrical and a symmetrical peak.

INTRODUCTION

The distribution of any pure substance on a chromatogram can be presented by a curve where the abscissae are the distance from the start, and the ordinates are the concentration of the observed points. In the cases of symmetrical spots the curves obtained are similar to the theoretical Gaussian curve.

If the major component (A) overlaps the other component, a change in the shape of the curve occurs. The position and ratio of the overlapped component to the major component are interesting in practical work. Fig. 1 illustrates the three possible zones where the minor component B can be overlapped by the major component.

The presence of component B is evident immediately in the diagram if it is in zone I. Its presence in zone II, however, is not evident from Fig. 1. Zone III is of no interest chromatographically because there is no significant difference between the R_F value of the components B and A in that zone.

The problem investigated was the recognition of component B in zone II. A mathematical method was used, the amount of component A being several times larger than the amount of component B.

Many mathematical methods have been developed for symmetrical peaks which should permit the detection of the overlapped component B[1-3]. Unfortunately these methods are not suitable in the case of TLC where asymmetrical distribution curves often occur.

In these cases it is necessary to find the regularity of the asymmetry because one may assume that pure substances would give the regular symmetric or asymmetric peaks. The asymmetry or symmetry of distribution can be arrived at by comparing the experimental curve to the ideal Gaussian curve, whose maximum is at a value of o on the abscissae (Fig. 2). The negative values of the abscissae are on the
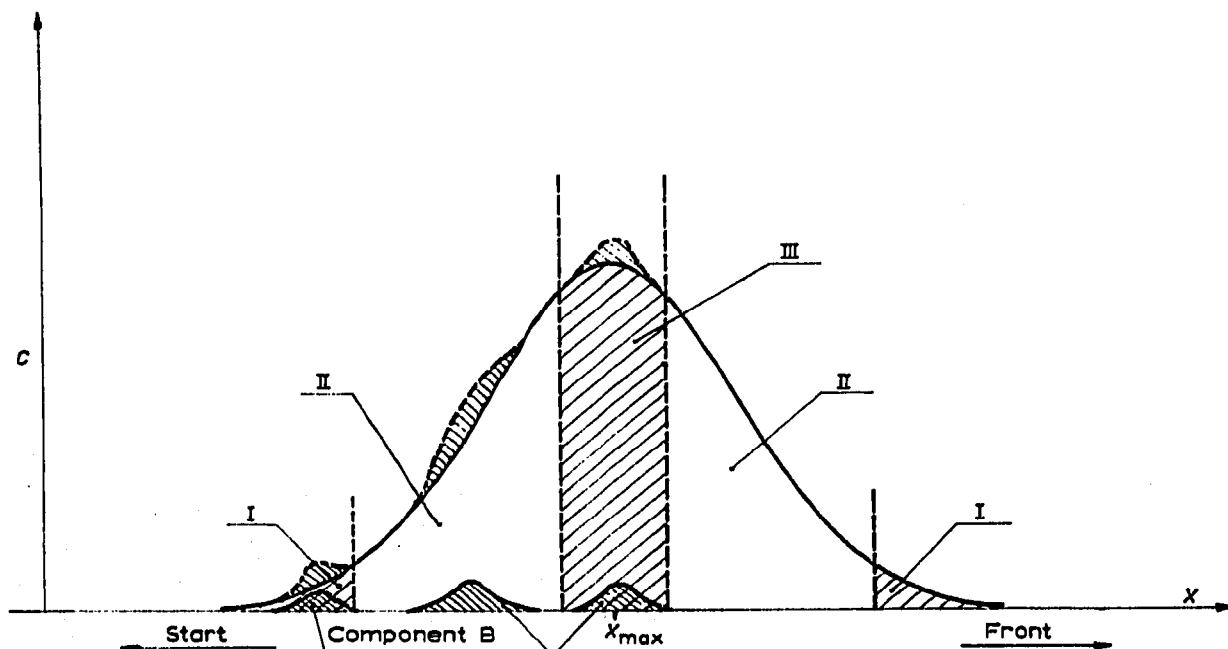
Fig. 1. Zones in the chromatographic peak of major component A where it is possible to detect the component B. The zones I and III (cross-hatched) were not of interest for this investigation. Only zone II is of interest for mathematical interpretation.

left side of the curve and positive values of abscissae on the right. The standard deviation (σ) is the abscissa unit. The value 0.4 is the ordinate maximum. Before making the comparison, the experimental curve must be normalised, *i.e.* transformed into a curve in which the maximum value is 0.4. This can be done by changing the ordinate values as follows:

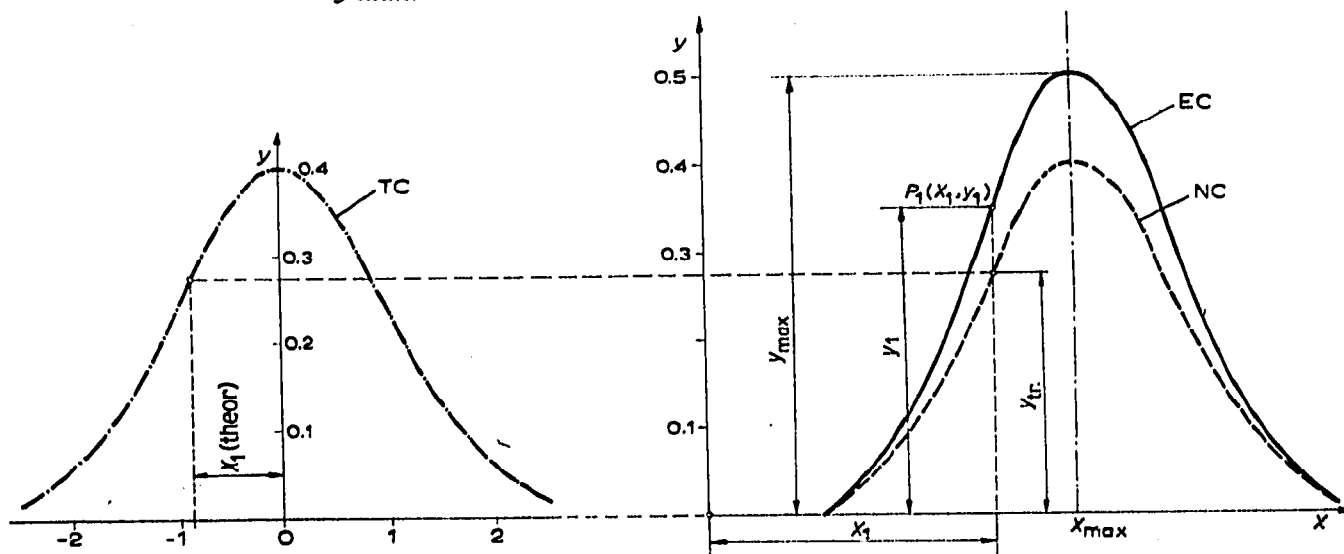$$y_{theor} = y_{exp} \frac{0.4}{y_{max.}} \tag{1}$$

Fig. 2. The curve obtained from smoothed data (EC), the normalised curve (NC) and theoretical Gaussian curve (TC). The hatched lines shows the course of the mathematical operation. The smoothed data are normalised and compared with the theoretical values.
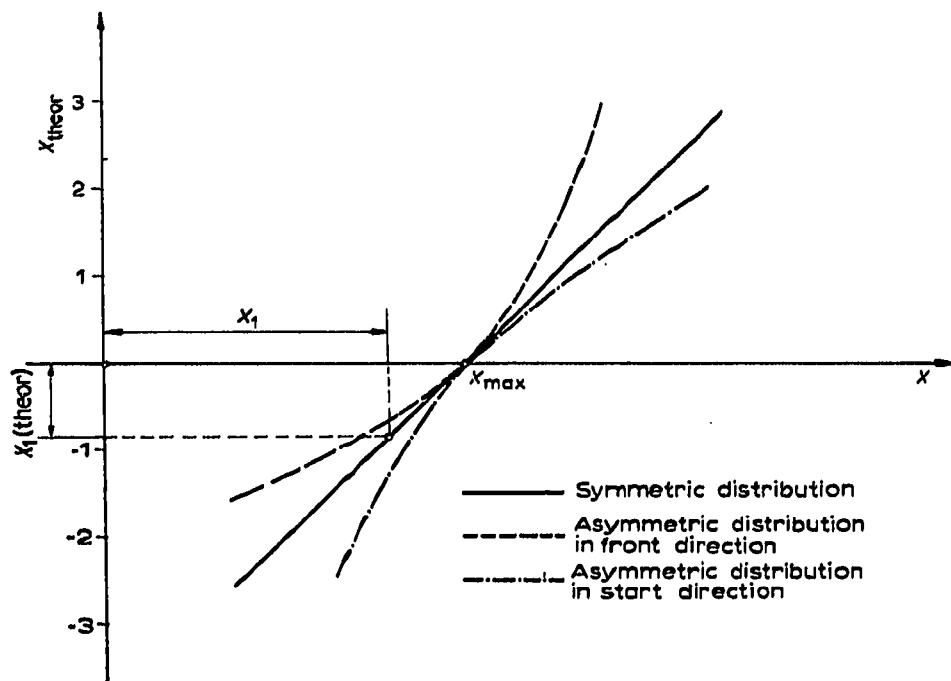
Fig. 3. Diagram obtained by comparison of the normalised curve and the theoretical curve (Fig. 2). The straight line shows the symmetrical distribution of the spot; the concave and convex lines represent asymmetric distribution.

For every abscissa value on the experimental curve there is a corresponding abscissa value on the theoretical curve, but the left side of the experimental curve must be compared with the left side of the theoretical curve, and *vice versa*.

The abscissae values of the experimental and corresponding theoretical curves give the coordinates which, when drawn graphically, give the line of asymmetry (Fig. 3). The values of the abscissae in Fig. 3 are the same as in the experimental curve, and the abscissae of the theoretical curve are ordinates.

A symmetrical distribution would give a straight line in the diagram, but if the concentration maximum of the substance is towards the front of the spot, a concave curve is obtained; if it is towards the rear of the spot, a convex curve is obtained.

It is from this line, which represents the regularity of distribution, that it is possible to detect in a simple manner the position of another component B. In the case of symmetrical distribution the presence of the component B is obvious because at that position the straight line diverges.

If the lines obtained are not straight lines, the detection of component B is not directly possible, and mathematical methods must be used; hence we used numerical analysis.

Every regular curve can be approximately represented by the equation[4]:

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \ldots \tag{2}$$

The equation for a straight line being: $\qquad\qquad\qquad\qquad\qquad$ (3)

$$f(x) = a_0 + a_1 x$$

If one takes the difference between the x_{theor} ordinates at equal distances along the abscissae, the first differences (Δ^1) give a constant value and the second differences (Δ^2), i.e. the differences between the Δ^1 values, are or have values very close to zero. The points at the position of component B give values significantly greater than zero for the second differences.

The curve representing a slight asymmetry of distribution can be very well approximated by the equation:

$$f(x) = a_0 + a_1x + a_2x^2 \qquad (4)$$

The first differences (Δ^1) are given by:

$$f^I(x) = a_1 + 2a_2x \text{ (straight line)} \qquad (5)$$

Δ^2 give a constant:

$$f^{II}(x) = 2a_2 \qquad (6)$$

and Δ^3 give values very near zero:

$$f^{III}(x) \approx 0 \qquad (7)$$

The curve presented by a very asymmetric spot can be approximated by the equation:

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 \qquad (8)$$

The fourth differences give values that are close to zero: $f^{IV}(x) \approx 0$ \qquad (9)

In those cases where there is a divergence from zero the overlapped component B will be present at that position.
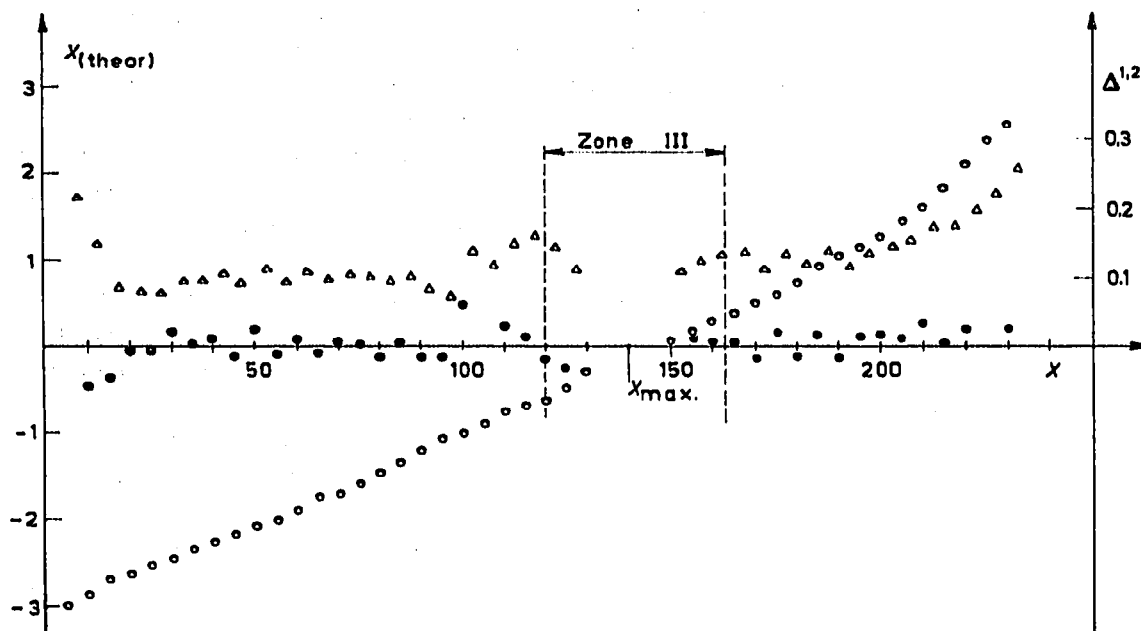

Fig. 4. Diagram showing the data (○) of arabinose spot asymmetry without smoothing, and the corresponding data after numerical analysis. (△), first differences; (●), second differences. It shows too many deviations.

All the mathematical operations may be carried out by means of a mini-computer of a type that is in common use in analytical laboratories.

EXPERIMENTAL

The substances investigated were the sugars arabinose and rhamnose.

Chromatography was carried out on plates 20 × 20 cm (E. Merck) precoated with a layer of Silica Gel H. The developing solvent was isopropanol–trichloroethylene–acetic acid (3:1:1). The substance was applied at the start in the form of a line.

Visualisation was by spraying with 2% aniline in acetic acid + 2% diphenylamine in acetic acid + 85% phosphoric acid (5:5:1) (ref. 5).

The corresponding peaks were measured by means of a Photovolt TLC densitometer.

Smoothing of the data obtained was necessary to eliminate the deviations caused by non-ideal conditions of the chromatographic process and non-ideal function
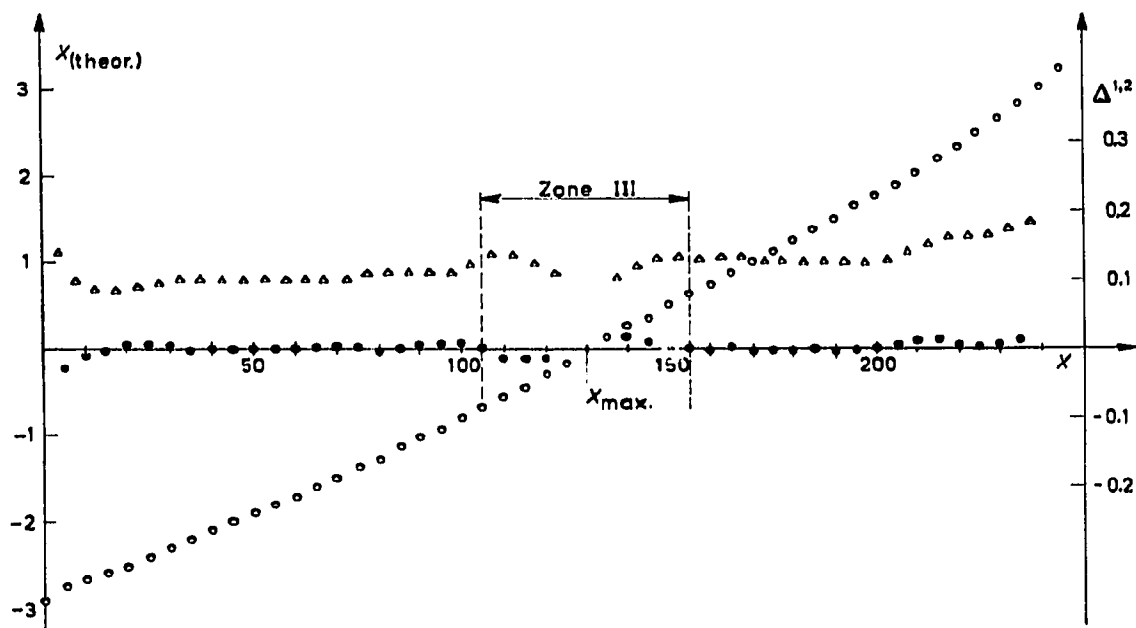
Fig. 5. Diagram of arabinose spot asymmetry obtained after smoothing data (◯), and the corresponding data after numerical analysis. (△), first differences; (●), second differences. It shows only small deviations.

of the TLC densitometer. In numerical analysis the error function cannot be established when the deviations are too great. For example, the numerical analysis of non-smoothed data for arabinose shown in Fig. 4 contrasts with the numerical analysis of smoothed data shown in Fig. 5. The difference can easily be seen. In Fig. 6, the scheme is shown for linear smoothing of the data by four degrees (Programme 1). It is carried out according to the following equations:

The first degree:

$$y_{1-2} = \frac{y_1 + y_2}{2} \; ; y_{2-3} = \frac{y_2 + y_3}{2} \tag{10}$$

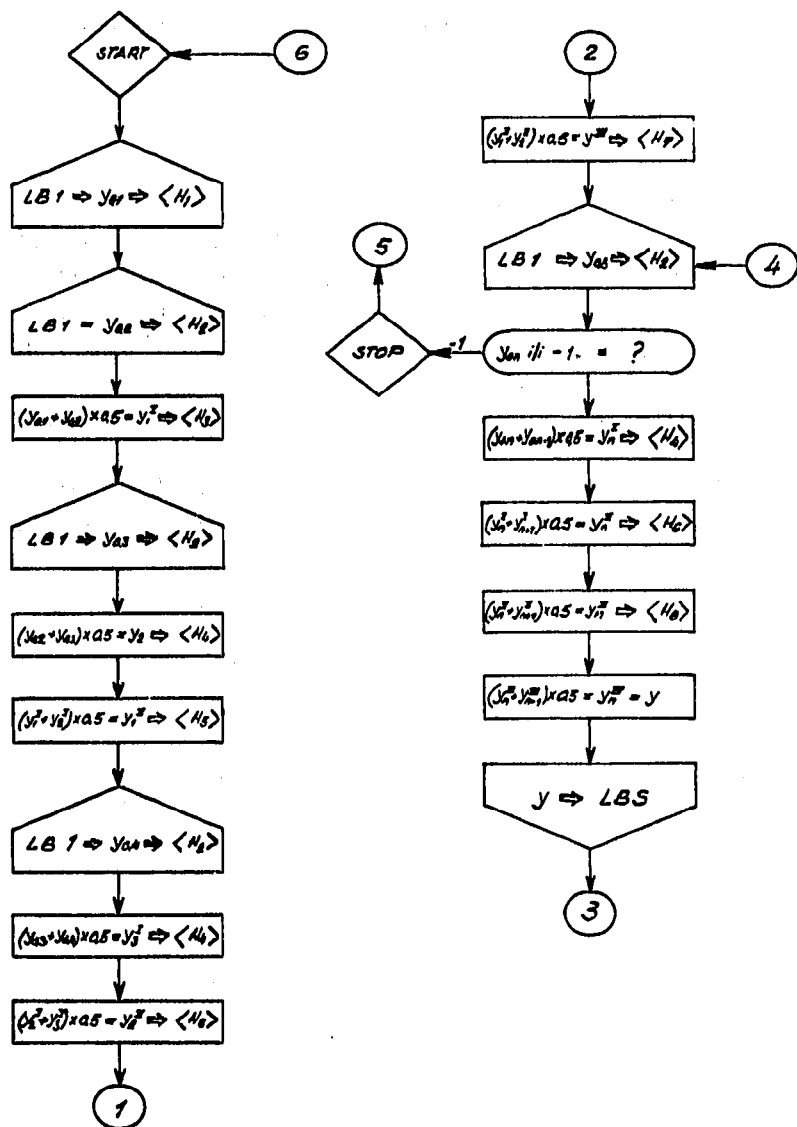Fig. 6. Scheme of the course of operation of the mini-computer (Celatron Ser-2) for quadruple data smoothing (Programme 1).

The second degree:

$$y_{2(II)} = \frac{y_{1-2} + y_{2-3}}{2}$$

$$y_{3(II)} = \frac{y_{2-3} + y_{3-4}}{2} \tag{11}$$

The third degree:

$$y_{2(II)-3(II)} = \frac{y_{2(II)} + y_{3(II)}}{2}$$

$$y_{3(II)-4(II)} = \frac{y_{3(II)} + y_{4(II)}}{2} \tag{12}$$

The fourth degree:

$$y_{3(IV)} = \frac{y_{2(II)-3(II)} + y_{3(II)-4(II)}}{2}$$

$$y_{4(IV)} = \frac{y_{3(II)-4(II)} + y_{4(II)-5(II)}}{2} \tag{13}$$

The principle of smoothing is illustrated in Fig. 7. The smoothed data were recorded on the paper-tape (LBS) and this was fed to the computer as the data for programme 2 (Fig. 8).
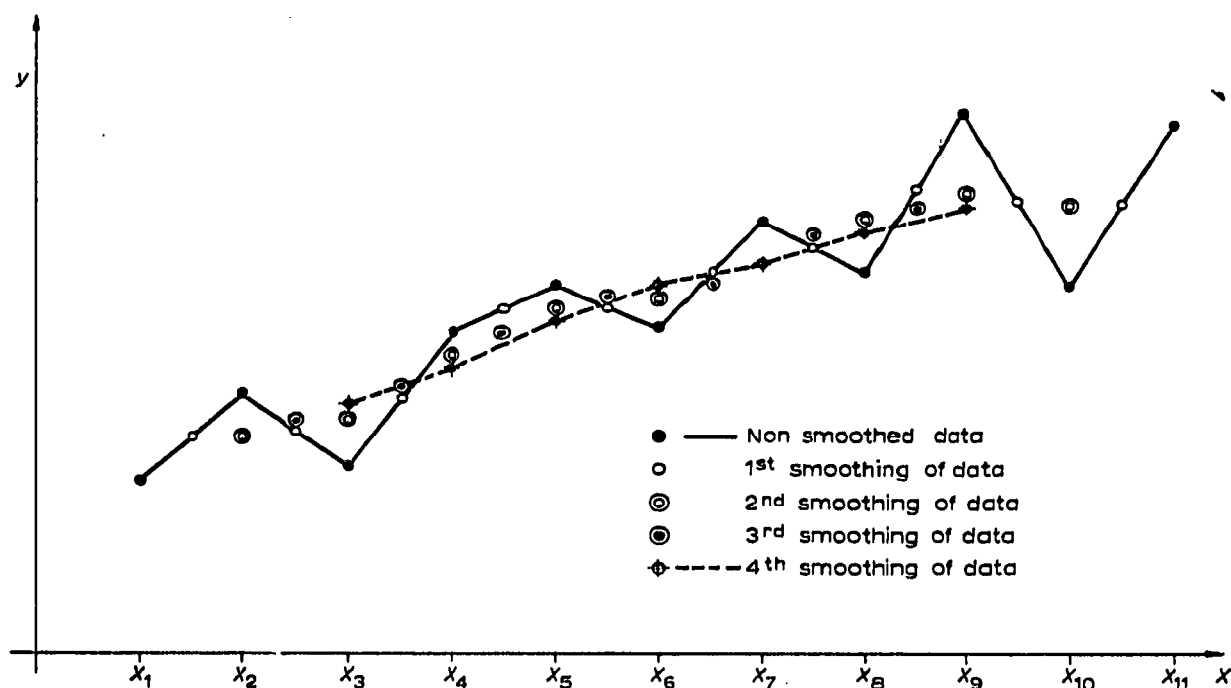
Fig. 7. The data smoothing process. The solid line shows the line obtained with non-smoothed data and the broken line represents the data after quadruple smoothing.

Programme 2 follows the mathematical operations described earlier. The computer gives the printed result on a sheet.

The numerical analysis of the smoothed data for arabinose is presented in Table I. The 1st column gives the position of the point observed on the spot and corresponds to the abscissae in the diagram (Figs. 2 and 3). The 2nd column gives the smoothed data corresponding to the concentration at the point observed. The data of the normalised curve ($y_{max.} = 0.4$) are in column 3. The 4th column gives the corresponding abscissae values (x_{theor}) of the theoretical Gaussian curve according to the equation:

$$x_{theor} = \pm \sqrt{2 \left(\ln \frac{1}{\sqrt{2\pi}} - \ln y_{theor} \right)} \tag{14}$$
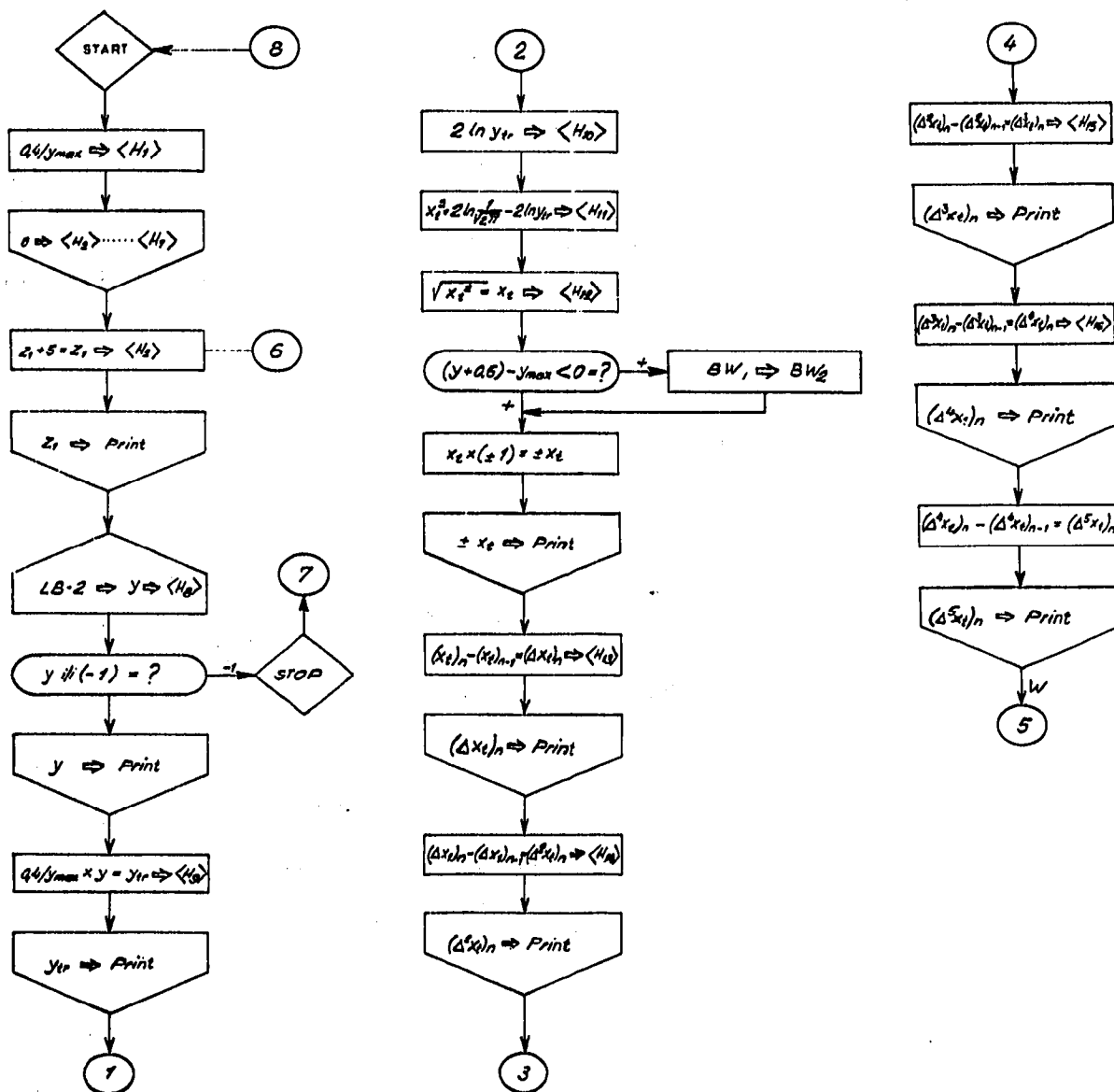
Fig. 8. The flow diagram for the numerical analysis (Program 2).

Before the maximum ($y \approx 0.4$) the computer gives negative values of x_{theor}, and then positive values after the maximum is passed.

Columns 5, 6, 7, 8 and 9 give the difference between the data in the respective preceding columns and represent the numerical analysis of the x_{theor} data.

Large deviations appear in zone III (Fig. 1) during the numerical analysis (Figs. 4 and 5) because small differences of y_{theor} near y_{max}. cause great differences in x_{theor}. Therefore data from the centre (framed part) must be eliminated.

In Table II, the numerical analysis of the smoothed data for the rhamnose spot is presented. The data in position D deviate from zero very significantly and it is possible to conclude that at this point in the spot another substance is present.

TABLE I

TREATED DATA FOR ARABINOSE[a]

The data in the centre (framed part) correspond to zone III, and they are of no interest for this investigation. The divergences from zero are not significant, and therefore there is no other component present in zone II.

x (1)	y (2)	ytr (3)	xtheor (4)	Δ^1 (5)	Δ^2 (6)	Δ^3 (7)	Δ^4 (8)	Δ^5 (9)
0	1.75	0.00595	2.9006	0.1286				
5	2.52	0.00857	−2.7720	0.0992	−0.0294	0.0171		
10	3.30	0.01122	−2.6728	0.0869	−0.0123	0.0110	−0.0061	0.0013
15	4.15	0.01411	−2.5859	0.0856	−0.0013	0.0062	−0.0048	0.0007
20	5.16	0.01754	−2.5003	0.0905	0.0049	0.0021	−0.0041	−0.0005
25	6.44	0.02190	−2.4098	0.0975	0.0070	−0.0025	−0.0046	0.0045
30	8.11	0.02757	−2.3123	0.1020	0.0045	−0.0026	−0.0001	0.0030
35	10.21	0.03471	−2.2103	0.1039	0.0019	0.0003	0.0029	−0.0058
40	12.78	0.04345	−2.1064	0.1061	0.0022	−0.0026	−0.0029	0.0050
45	15.89	0.05403	−2.0003	0.1057	−0.0004	−0.0005	0.0021	−0.0008
50	19.52	0.06637	−1.8946	0.1048	−0.0009	0.0008	0.0013	−0.0026
55	23.68	0.08051	−1.7898	0.1047	−0.0001	−0.0005	−0.0013	0.0028
60	28.40	0.09656	−1.6851	0.1041	−0.0006	0.0010	0.0015	−0.0005
65	33.66	0.11444	−1.5810	0.1045	0.0004	0.0020	0.0010	−0.0015
70	39.49	0.13427	−1.4765	0.1069	0.0024	0.0015	−0.0005	−0.0027
75	45.98	0.15633	−1.3696	0.1108	0.0039	−0.0017	−0.0032	0.0015
80	53.19	0.18085	−1.2588	0.1130	0.0022	−0.0034	−0.0017	0.0069
85	60.93	0.20716	−1.1458	0.1118	−0.0012	0.0018	0.0052	−0.0012
90	68.82	0.23399	−1.0340	0.1124	0.0006	0.0058	0.0040	−0.0074
95	76.82	0.26119	−0.9216	0.1188	0.0064	0.0024	−0.0034	−0.0005
100	85.10	0.28934	−0.8028	0.1276	0.0088	−0.0015	−0.0039	−0.0027
105	93.52	0.31797	−0.6752	0.1349	0.0073	−0.0081	−0.0066	0.0049
110	101.51	0.34513	−0.5403	0.1341	−0.0008	−0.0098	−0.0017	0.0084
115	108.16	0.36774	−0.4062	0.1235	−0.0106	−0.0031	0.0067	0.1712
120	112.86	0.38372	−0.2827		−0.0137		0.1779	
125	115.72	0.39345	−0.1729	0.1098	0.1611	0.1748	−0.5662	−0.7441
130	116.90	0.39746	−0.0980	0.0749	−0.2303	−0.3914	0.6850	1.2512
135	116.34	0.39556	0.1386	0.0406	0.0633	0.2936	−0.3411	−1.0261
140	114.06	0.38780	0.2425	0.1039	0.0158	−0.0475	0.0403	0.3814
145	110.00	0.37400	0.3622	0.1197	0.0086	−0.0072	0.0010	−0.0393
150	104.15	0.35411	0.4905	0.1283				
155	96.85	0.32929	0.6212	0.1307	0.0024	−0.0062		
160	88.50	0.30090	0.7525	0.1313	0.0006	−0.0018	0.0044	0.0034
165	70.55	0.27047	0.8829	0.1304	−0.0009	−0.0015	0.0003	−0.0041
170	70.44	0.23950	1.0112	0.1283	−0.0021	−0.0012	0.0003	0.0000
175	61.43	0.20886	1.1386	0.1274	−0.0009	0.0012	0.0024	0.0021
180	52.78	0.17945	1.2649	0.1263	−0.0011	−0.0002	−0.0014	−0.0038
185	44.66	0.15184	1.3908	0.1259	−0.0004	0.0007	0.0009	0.0023
190	37.25	0.12665	1.5156	0.1248	−0.0011	−0.0007	−0.0014	−0.0023
195	30.66	0.10424	1.6390	0.1234	−0.0014	−0.0003	0.0004	0.0018
200	24.84	0.08446	1.7628	0.1238	0.0004	0.0018	0.0021	0.0017
205	19.65	0.06681	1.8911	0.1283	0.0045	0.0041	0.0023	0.0002
210	14.95	0.05083	2.0305	0.1394	0.0111	0.0066	0.0025	0.0002
215	10.85	0.03689	2.1827	0.1522	0.0128	0.0017	−0.0049	−0.0074
220	7.63	0.02594	2.3385	0.1558	0.0036	−0.0092	−0.0109	−0.0060
225	5.25	0.01785	2.4932	0.1547	−0.0011	−0.0047	0.0045	0.0154
230	3.44	0.01170	2.6574	0.1642	0.0095	0.0106	0.0153	0.0100
235	2.10	0.00714	2.8370	0.1796	0.0154	0.0059	−0.0047	−0.0200
240	1.21	0.00411	3.0256	0.1886	0.0090	−0.0064	−0.0123	0.0076
245	0.62	0.00211	3.2383	0.2127	0.0241	0.0151	0.0215	0.0338

(The framed part corresponds to rows x = 120 to 145, columns xtheor through Δ^5: xtheor values −0.2827, −0.1729, −0.0980, 0.1386, 0.2425, 0.3622.)

[a] Obtained by the mini-computer Celatron Ser-2.

TABLE II

TREATED DATA FOR RHAMNOSE[a]

The data in the centre (framed part) represent zone III which is of no interest for this investigation. There are significant divergences from zero in columns 6, 7 and 8 which have their maximum for $x = 160$; this is the position of the minor component B.

x (1)	y (2)	ytr (3)	xtheor (4)	Δ^1 (5)	Δ^2 (6)	Δ^3 (7)	Δ^4 (8)	Δ^5 (9)
0	1.50	0.00495	−2.9634					
5	2.20	0.00726	−2.8310	0.1324	−0.0088	−0.0001	−0.0069	0.0065
10	3.31	0.01023	−2.7074	0.1236	−0.0089	−0.0070	−0.0004	0.0322
15	4.20	0.01386	−2.5927	0.1147	−0.0159	−0.0074	−0.0004	−0.0343
20	5.40	0.01782	−2.4939	0.0988	−0.0233	0.0244	0.0318	−0.0269
25	6.50	0.02145	−2.4184	0.0755	0.0011	0.0219	−0.0025	0.0279
30	7.80	0.02574	−2.3418	0.0766	0.0230	−0.0075	−0.0294	0.0059
35	9.80	0.03234	−2.2422	0.0996	0.0155	−0.0090	−0.0015	−0.0081
40	12.60	0.04158	−2.1271	0.1151	0.0065	−0.0046	0.0044	0.0235
45	16.20	0.05346	−2.0055	0.1216	0.0019	−0.0083	−0.0037	−0.0345
50	20.60	0.06798	−1.8820	0.1235	−0.0064	0.0115	0.0198	0.0149
55	25.50	0.08415	−1.7649	0.1171	0.0051	−0.0032	−0.0147	0.0031
60	31.40	0.10362	−1.6427	0.1222	0.0019	−0.0030	0.0002	0.0002
65	38.20	0.12606	−1.5186	0.1241	−0.0011	0.0003	0.0033	−0.0078
70	45.70	0.15081	−1.3956	0.1230	−0.0008	0.0038	0.0035	0.0045
75	53.80	0.17754	−1.2734	0.1222	0.0030	−0.0005	−0.0043	0.0005
80	62.60	0.20658	−1.1482	0.1252	0.0025	−0.0003	0.0002	−0.0055
85	71.90	0.23727	−1.0205	0.1277	0.0022	0.0004	0.0007	0.0114
90	81.40	0.26862	−0.8906	0.1299	0.0026	−0.0044	−0.0048	−0.0071
95	90.80	0.29964	−0.7581	0.1325	0.0018	0.0022	0.0066	0.0019
100	99.40	0.32802	−0.6274	0.1307	0.0004	0.0017	−0.0005	0.0157
105	107.00	0.35310	−0.4963	0.1311	0.0021	0.0031	0.0014	
110	113.30	0.37389	−0.3631	0.1332	0.0052		0.0171	
115	118.00	0.38940	−0.2247	0.1384	0.0254	0.0202	−0.0470	−0.0641
120	120.80	0.39864	−0.0609	0.1638	−0.0014	−0.0268	−0.0093	0.0377
125	122.00	0.40260	0.1015	0.1624	−0.0375	−0.0361	−0.1186	−0.1093
130	121.80	0.40194	0.2264	0.1249	−0.1922	−0.1547	0.5663	0.6849
135	119.50	0.39435	0.1591	0.0673	0.2194	0.4116	−0.6511	−1.2174
140	115.30	0.38049	0.3112	0.1521	−0.0201	−0.2395	0.2642	0.9153
145	109.70	0.36201	0.4432	0.1320	0.0046	0.0247	−0.0185	−0.2827
150	102.30	0.33759	0.5798	0.1366	0.0108	0.0062	−0.0482	−0.0297
155	92.90	0.30657	0.7272	0.1474	−0.0312	−0.0420	0.1532	0.2014
160	84.80	0.27984	0.8434	0.1162	0.0800	0.1112	−0.2339	−0.3871
165	70.50	0.23265	1.0396	0.1962	−0.0427	−0.1227	0.1681	−0.4020
170	59.40	0.19602	1.1931	0.1535	0.0027	0.0454	−0.0521	−0.2202
175	48.70	0.16071	1.3493	0.1562	−0.0040	−0.0067	0.0075	0.0596
180	39.20	0.12936	1.5015	0.1522	−0.0032	0.0008	0.0018	−0.0057
185	31.00	0.10230	1.6505	0.1490	−0.0006	0.0026	−0.0027	−0.0045
190	24.00	0.07920	1.7989	0.1484	−0.0007	−0.0001	0.0010	0.0037
195	18.20	0.06006	1.9466	0.1477	0.0002	0.0009	0.0033	0.0023
200	13.50	0.04455	2.0945	0.1479	0.0044	0.0042	0.0044	0.0011
205	9.70	0.03201	2.2468	0.1523	0.0130	0.0086	−0.0332	−0.0376
210	6.60	0.02178	2.4121	0.1653	−0.0116	−0.0246	0.0122	0.0454
215	4.50	0.01485	2.5658	0.1537	−0.0240	−0.0124	−0.0031	−0.0153
220	3.20	0.01056	2.6955	0.1297	−0.0395	−0.0155	0.0437	0.0468
225	2.50	0.00825	2.7857	0.0902	−0.0113	0.0282	−0.0395	−0.0832
230	2.00	0.00660	2.8646	0.0789	−0.0226	−0.0113	0.0679	0.1074
235	1.70	0.00561	2.9209	0.0563	0.0340	0.0566	−0.0948	−0.1627
240	1.30	0.00429	3.0112	0.0903	−0.0042	−0.0382		
245	1.00	0.00330	3.0973	0.0861				

[a] Obtained by the mini-computer Celatron Ser-2.

REFERENCES

1 V. CEJKA, M. H. DIPERT, S. A. TYLER AND P. D. KLEIN, *Anal. Chem.*, 40 (1968) 1614.
2 P. D. KLEIN AND B. KUNZE-FAULKNER, *Anal. Chem.*, 37 (1965) 1245.
3 P. D. KLEIN AND P. A. SZCREPANIK, *Anal. Chem.*, 39 (1967) 1276.
4 J. W. GRAHAM, in R. G. STANTON AND K. D. FRYER (Editors), *Topics in Modern Mathematics*, Prentice Hall, Englewood Cliffs, N.Y., 1964, p. 85.
5 P. D. BERGER AND A. S. AGATE, *J. Chromatog.*, 39 (1969) 232.