ANALYTICAL BIOCHEMISTRY 57, 137–156 (1974)

Protein Molecular Weight Computation from Sedimentation Velocity Data

J. GRIEVINK AND R. T. B. HOUTERMAN

Twente University of Technology, Enschede, The Netherlands

AND

K. DE GROOT[1]

Free University, Department of Material Science, Schools of Dentistry and Medicine, de Boelelaan 1115, Amsterdam, The Netherlands

Received April 19, 1973; accepted July 30, 1973

In ultracentrifugation, the concentration gradient of mono-disperse samples obtained by sedimentation velocity experiments is described by Gehatia's equation which holds several parameters including the sedimentation and diffusion constants. Once these two constants are known, the molecular weight follows from the Svedberg equation. A least squares method has been developed to derive the transport constants from the refractive index gradient curves. The method employs a mathematical model based on Gehatia's theory. A main feature of the model is the application of two sets of intermediate parameters via which the transport coefficients are much easier calculated than along a direct way. Furthermore some difficult to observe quantities cancel out. The square residues are minimised numerically. The potential errors introduced by this numerical minimalisation are shown to be unimportant compared to the unavoidable experimental errors.

I. INTRODUCTION

Molecular weight determination of proteins by means of ultracentrifugation usually proceeds via equilibrium sedimentation or the approach to equilibrium method according to Archibald. Sedimentation velocity experiments are mostly analyzed using Svedberg's method, yielding sedimentation coefficients S.

Gehatia (1), Gehatia and Hubner (2), (3) however, has provided methods to derive both S and the diffusion constant D and thus the molecular weight M, from a single sedimentation velocity run on monodisperse samples. The advantages of obtaining molecular weights in

[1] To whom correspondence should be addressed.

Gehatia's way are obvious: the time needed for a velocity run is much smaller than the time necessary for equilibrium or approach to equilibrium runs. Furthermore, the usual methods yield only M, not S and D separately.

The disadvantages, however, which have limited the practical use, are twofold. Firstly, Gehatia's method requires a few experimentally difficult to observe parameters such as the initial time (t_0), the initial position of the phase boundary (r_0). Secondly, the arithmetical analysis necessary to obtain S and D is very cumbersome.

In this article a least squares method is described, employing a model that has been derived from Gehatia's theory (1). By means of two parameter transformations the determination of r_0 and t_0 can be bypassed. A scheme for automated data handling facilitates the execution of the tedious calculations considerably.

II. THEORY

The theory of sedimentation velocity experiments has been discussed extensively in the literature [Gehatia (1), Fujita (4)]. The results of Gehatia's analysis will be used as a starting point for the development of a model relating S, D and M to the measurable variables, i.e., refractive index gradient versus time and radial distance. This section describes the development of the model, while Sec. III deals with the application of the model to the data processing.

The sedimentation velocity experiment is assumed to satisfy the conditions listed below:

a. Initially the liquid mixture consists of two phases of the same solute and solvent with different concentrations of the solute and separated by an indefinitely sharp interface (step-distribution).

b. The solute and the solvent are both mono-disperse nonelectrolytes and completely miscible.

c. The mixture has a negligible compressibility and shows no viscous effects or convection during ultracentrifugation.

d. The temperature is constant. The angular velocity is increased rapidly to the desired level and remains constant till the end of the experiment.

e. During the experiment the concentration profile must not be affected by the presence of the tangential walls (bottom and meniscus).

f. A sector cell is used.

g. The refractive index depends linearly on the concentration of the solute.

h. The transport coefficients of the solute are virtually independent of the concentration in the range occurring in the system, thus S and D are independent of concentration.

i. The gradient of the refractive index is registered as a function of time and radial distance.

Having specified the experimental conditions, the equation of the concentration gradient of the solute as developed by Gehatia can be introduced:

$$\frac{\partial c(r,\tau)}{\partial r} = \frac{2 \cdot \Delta c \cdot e^{-\lambda} \cdot e^{-\rho^2}}{(4D\alpha\tau)^{1/2}} \cdot (e^{-\rho_0^2} \cdot I_1(2\rho\rho_0) \cdot \rho_0 - e^{-\rho_b^2} \cdot I_1(2\rho\rho_b) \cdot \rho_b) \quad (1)$$

where:

r	radius (the distance to the center of rotation)
τ	time elapsed since the beginning of the experiment $= t - t_0$; t is the time with an arbitrary origin; t_0 is the time at the start of the experiment
r_0	initial position of the interface
r_b	position of the bottom
c	concentration of the solute
Δc	initial concentration difference between both phases

$$\lambda = 2\beta\tau = 2 \cdot S \cdot \omega^2 \cdot \tau, \quad (1.1)$$

ω is the angular velocity.

$$\alpha = (e^{\lambda} - 1)/\lambda,$$

$$\rho = r/(4D\alpha\tau)^{1/2}, \quad (1.2)$$

$$\rho_0 = r_0 e^{\lambda/2}/(4D\alpha\tau)^{1/2}, \quad (1.3)$$

$$\rho_b = r_b e^{\lambda/2}/(4D\alpha\tau)^{1/2},$$

$$I_1(2\rho\rho_0) = \sum_{k=0}^{\infty} (\rho\rho_0)^{2k+1}/[k!(k+1)!].$$

(Bessel function of the first kind and the first order.)

In the original derivation of Eq. (1) the initial concentration of the solute in one of the phases was equal to zero. In App. A it is proved that Eq. (1) is also valid in the more general case of an arbitrary concentration difference between the phases.

Equation (1) can be simplified considerably if the molecules of the solute are not too small and the duration of the experiment is relatively short. Under these circumstances the influence of the bottom on the concentration profile can be neglected [see (1)]:

$$\rho_0 \cdot e^{-\rho_0^2} \cdot I_1(2\rho\rho_0) \gg \rho_b \cdot e^{-\rho_b^2} \cdot I_1(2\rho\rho_b). \tag{2}$$

In the second place, the argument $(2\rho\rho_0)$ is normally large enough $(2\rho\rho_0 > 4 \times 10^2)$ to approximate I_1 $(2\rho\rho_0)$ by the first term of an appropriate asymptotic power expansion series [see (7), Sec. 9.7]:

$$I_1(2\rho\rho_0) \simeq [e^{2\rho\rho_0}/(4\pi\rho\rho_0)^{1/2}] \cdot \left(1 - \frac{3}{16\rho\rho_0} - \cdots\right). \tag{3}$$

Substitution of the simplifications (2) and (3) into Eq. (1) gives:

$$\frac{\partial c(r,\tau)}{\partial r} = \frac{\Delta c \cdot e^{-\lambda}}{(4\pi D\alpha\tau)^{1/2}} e^{-(\rho_0-\rho)^2} \left(\frac{\rho_0}{\rho}\right)^{1/2}. \tag{4}$$

The next step is to replace the concentration gradient by the refractive index gradient. The height $h(r,\tau)$ of a recorded sedimentation curve depends linearly on the concentration gradient:

$$h(d,\tau) = h(r,\tau) = F \cdot \frac{\Delta n}{\Delta c} \cdot \frac{\partial c(r,\tau)}{\partial r}, \tag{5}$$

where

$$r = G \cdot d + r_l;$$

G, radial magnification factor; d, radial distance of a point of a recorded curve with respect to a reference line; r_l, position of reference line; Δn, initial difference of refractive indices of the phases; F, vertical magnification factor. Eliminating the concentration gradient and replacing λ, ρ and ρ_0 in Eq. (4) by the right sides of (1.1) to (1.3) gives the following equation:

$$h(r,\tau) = \frac{F \cdot \Delta n}{(2\pi)^{1/2}} \cdot \frac{e^{-2\beta\tau}}{2D\alpha\tau} \cdot \exp\left[-\frac{(r - r_0 e^{\beta\tau})^2}{4D\alpha\tau}\right] \cdot \left(\frac{r_0 e^{\beta\tau}}{r}\right)^{1/2}. \tag{6}$$

Transformations. Equation (6) can be used as a model in the computation of β and D assuming the refractive index gradient is measured as a function of radial distance and time. The quantities r_0, Δn and F must also be determined. However, in doing so two difficulties arise. Firstly, Eq. (6) has not a suitable form to calculate β and D easily. Secondly, an accurate experimental determination of r_0 and τ is hardly possible due to some unavoidable mixing near the interface as the start and to an uncertainty with respect to the exact moment of interface formation. These difficulties can be solved by means of two transformations. The first one dissolves the necessity of measuring r_0, F and Δn. The second transformation allows the use of an arbitrary origin of the time axis

by the introduction of a reference time τ_r to be selected by the investigator.

The first transformation replaces β, D, F, Δn, r_0 and τ by a new set of time-dependent parameters m_0, m_1 and m_2 (shortly the vector $\mathbf{m}$):

$$m_0 = F \cdot \Delta n \cdot r_0^2/(2\pi)^{1/2} \qquad m_0 > 0, \tag{7.1}$$

$$m_1 = r_0 \exp(\beta\tau) \qquad m_1 > 0, \tag{7.2}$$

$$m_2 = 2D\alpha\tau \qquad m_2 > 0. \tag{7.3}$$

The parameters m_1 and m_2 are related to the reduced moments of the concentration gradient with respect to the radius:

$$m_1 = (\bar{w}_i)^{1/2}$$

$$m_2 = (\bar{w}_4 - \bar{w}_2^2)/2\bar{w}_2,$$

where $\bar{w}_i$ is a reduced moment of order i [see (1) for the definition]. Application of the first transformation to Eq. (6) yields:

$$h(r; \mathbf{m}) = m_0 \cdot m_1^{-2} \cdot \exp[-(r - m_1)^2/2m_2] \cdot [m_1/(m_2 \cdot r)]^{1/2} \tag{8}$$

The variables $h(r; \mathbf{m})$ and r are measurable quantities from which the parameter vector $\mathbf{m}$ can be derived.

The second transformation relates the components of the $\mathbf{m}$ vector to the transport coefficients via a new set of parameters. As the components of $\mathbf{m}$ are time dependent a subscript i is attached to refer to discrete time τ_i. The subscript r indicates the reference time and its corresponding parameter values, $m_{1,r}$ and $m_{2,r}$.

$$Z_{0,i} = \tau_i - \tau_r, \tag{9.1}$$

$$Z_{1,i} = 2\beta Z_{0,i} = \ln(m_{1,i}/m_{1,r})^2 \tag{9.2}$$

$$Z_{2,i} = 2DZ_{0,i} = \left(\frac{Z_{1,i}}{e^{Z_{1,i}} - 1}\right) \cdot (m_{2,i} - m_{2,r} \cdot e^{Z_{1,i}}). \tag{9.3}$$

The time derivatives of Z_1 and Z_2 are independent of the time and are equal to the transport coefficients times a factor two.

The original problem to deduce β and D from the experimental data has been separated in two subproblems:

a. Determination of $\mathbf{m}_i$ from the experimental data.

b. Computation of Z_1 and Z_2 and their time derivatives, after selecting a reference time and a corresponding $\mathbf{m}$ vector. Both subproblems can be solved by the method of least squares as described in Sect. III.

Once the transport coefficients are known, the molecular weight of the solute follows immediately from the Svedberg equation (5):

$$M = \frac{RTS}{(1 - \rho v_1)D}\left(\frac{\partial \ln \alpha 1}{\partial \ln c1}\right)_T, \tag{10}$$

where

S	β/ω^2	
R	gas constant	
T	absolute temperature	
ρ	average density of the solution	
v_1	partial volume of the solute	
α_1	activity of the solute	both expressed in gmoles/cm³ or g/cm³
c_1	concentration of the solute	

The Eqs. (7), (8), (9) and (10) form the model to be applied to the data processing.

III. METHOD OF LEAST SQUARES

1. *Derivation of the* **m** *vector*

During an experiment a certain number of refractive index gradient curves are registered. Each curve must be represented numerically, which can be done by measuring the moment of recording t, the radial magnification factor G and the coordinates of at least three points on a curve: the height y with respect to the base line and the radial distance d relative to a reference line. The product of G and d added to the distance r_l between the reference line and the center of rotation, is the actual distance r of a point on a curve to the axis of rotation. It is convenient to select the points radially equidistant. The zero time may be chosen arbitrarily. Then the outcome of an experiment is given by the following set of data:

$$(y_k, r_k)_i, t_i \qquad \begin{matrix} k = 1, \ldots, K_i \quad (K_i \geqslant 3) \\ i = i, \ldots, I. \end{matrix}$$

The subscript k refers to a selected point; the subscript i refers to a curve. These data suffice to determine a **m** vector for each curve. As the derivation is identical for each curve, the suffix i will be deleted from now on.

The connection between the experimental data and the theoretical model is made by relating the experimental height y_k to the predicted height $h(r_k; \mathbf{m})$ is given by Eq. (8):

$$y_k = h(r_k; \mathbf{m}) + \epsilon_k \qquad k = 1, \ldots, K. \tag{11}$$

The residue ϵ_k accounts for the discrepancy between the model and reality. In order to find the closest fit of the model to the experimental

data the sum of square residues is minimised by adjusting the vector **m**. The sum of square residues (abbr. SSR) reads:

$$Q(\mathbf{m}) = \sum_{k=1}^{K} \epsilon_k^2 = \sum_{k=1}^{K} (y_k - h(r_k; \mathbf{m}))^2. \tag{12}$$

Q is a positive continuous, twice differentiable scalar function of **m**, defined on V, i.e., $m_0 > 0$, $m_1 > 0$, $m_2 > 0$. A local minimum exists at $\mathbf{m}^0 \epsilon V$ when the following conditions are satisfied at $\mathbf{m} = \mathbf{m}^0$:

$$(1)\ \frac{\partial Q(\mathbf{m})}{\partial m} = 0 \qquad j = 0, 1, 2. \tag{13.1}$$

(2) The matrix of second order derivatives of Q with respect to **m** is positive definite.

An analytical approach to the solution of (13.1) is not very promising due to the nonlinear occurrence of the parameters m_1 and m_2 in the expression of $h(r_k; \mathbf{m})$. Application of (13.1) yields symmetrical expressions after some substitutions:

$$\sum_{k=1}^{K} y_k \cdot h(r_k; \mathbf{m}) \cdot (r_k)^j = \sum_{k=1}^{K} (h(r_k; \mathbf{m}))^2 \cdot (r_k)^j \qquad j = 0, 1, 2. \tag{13.2}$$

During the operations the quantity $h(r_k; \mathbf{m})$ has been temporarily replaced by the right side of Eq. (8). This set of equations is not easily solvable either. This problem however, minimising a function of three variables, can be handled numerically by assuming the SSR to be an unimodal function on V and the matrix of second order derivatives is positive definite at the minimum.[2] Numerical optimalisation techniques [e.g., direct climbing routines, see (6) Chap. 7] deal with this class of problems successfully.

A simple but effective routine called SEARCH has been developed. The SSR is minimised by means of a series of explorations in a subset of V. The search to a minimum can be made more efficient by reducing the dimension of the SSR. The vector component m_0 being a linear parameter in Eq. (8) can be eliminated from Eq. (12) by means of condition (13.1), case $j = 0$. This yields:

[2] We are not able to prove these assumptions. No experimental evidence has been found to distrust the validity of the unimodality of the SSR. Nevertheless this point should be kept in mind when inspecting the numerical results. A theoretical analysis of the properties of the matrix of the second order derivatives does not produce very much as the calculations become very complicated.

$$Q(m_1, m_2) = \sum_{k=1}^{K} (y_k - f(r_k; m_1, m_2) \cdot m_0^*)^2, \tag{14.1}$$

where

$$f(r_k; m_1, m_2) = h(r_k; \mathbf{m})/m_0,$$

and

$$m_0^* = \sum_{p=1}^{K} y_p \cdot f(r_p; m_1, m_2) \Big/ \sum_{p=1}^{K} (f(r_p; m_1, m)_2)^2. \tag{14.2}$$

The variable Q (m_1, m_2) is a positive, continuous, twice differentiable scalar function of m_1 and m_2, defined for $m_1 > 0$ and $m_2 > 0$. The variable m_0^* is also a function of m_1 and m_2. Equation (14.1) is the objective function of the SEARCH routine. Appendix B describes the various activities of this module.

At the end of the optimalisation the routine produces the minimised sum of square residues (Q min) and the optimal values of m_1 and m_2. The corresponding value of m_0^* is calculated via (14.2). Finally a new parameter m_3 is introduced representing the average scatter of the experimental points along the optimal curve:

$$m_3 = [Q \text{ min}/K \cdot (m_0^*)^2]^{1/2}. \tag{15}$$

The division by the square of m_0^* is necessary to eliminate the influence of the vertical scaling factor.

The procedure as described in this part of the section has to be repeated for each curve. The obtained set of $\mathbf{m}$ vectors, $\mathbf{m}_i$, $i = 1, \ldots, I$ will be used to derive the transport coefficients.

2. *Computation of S and D*

The second transformation (9) establishes the relation between the transport coefficients and components of the $\mathbf{m}_i$ vectors. Firstly, the Z values are calculated. This calculation requires the reference parameters $m_{1,r}$ and $m_{2,r}$. The parameters of the curve having the smallest scatter (i.e., the smallest m_3 value of all curves) are taken as references. The factors $Z_{1,i}$ and $Z_{2,i}$ $(i = 1, \ldots, I)$ follow straight forward from Eq. (9). Then the time derivatives of the series $Z_{1,i}$ (equal to $2S\omega^2$) and $Z_{2,i}$ (equal to $2D$) can be calculated by the method of weighted least squares. Only one calculation will be given because the derivations are identical in both cases.

Owing to the second transformation a linear relationship between a Z factor and the experimental time t is introduced:

$$Z_i = A + Bt_i + \mu_i \qquad i = 1, \ldots, I, \tag{16}$$

where

Z_i	represents either $Z_{1,i}$ or $Z_{2,i}$
B	$= 2S\omega^2 = 2\beta$ in case of sedimentation
B	$= 2D$ in case of diffusion
A	is a quantity depending on the origin of the time axis
μ_i	is a deviation due to the approximate nature of the model and to possible errors in the computed Z values

The sum of the weighted square residues is given by:

$$R(A, B) = \sum_{i=1}^{I} w_i \cdot \mu_i^2 = \sum_{i=1}^{I} w_i \cdot (Z_i - A - B \cdot t_i)^2. \tag{17}$$

The symbol w_i denotes a weight factor. The weight factors will be used to suppress the contribution of those observations showing relatively much scatter. Therefore a (rather arbitrary) inverse relation between the weight factors and the scatter factors is applied:

$$w_i = \left(1 - m_{3,i} \Big/ \sum_{i=1}^{I} m_{3,i}\right) \Big/ (I - 1) \qquad I \geqslant 2. \tag{18}$$

The parameters A and B can be derived by minimising $R(A, B)$. Application of the necessary conditions for the existence of a minimum (A^0, B^0) yields:

$$A^0 = \bar{Z} - B^0 \cdot \bar{t}, \tag{19}$$

$$B^0 = (\overline{Zt} - \bar{Z} \cdot \bar{t})/N, \tag{20}$$

where

$$N = \bar{t}^2 - (\bar{t})^2,$$

$$\overline{Zt} = \sum_{i=1}^{I} w_i Z_i t_i,$$

$$\bar{t}^2 = \sum_{i=1}^{I} w_i t_i^2,$$

$$\bar{t} = \sum_{i=1}^{I} w_i t_i,$$

$$\bar{Z} = \sum_{i=1}^{I} w_i Z_i.$$

It is easy to prove that the matrix of second order derivatives is definite positive in (A^0, B^0). Furthermore, it should be noted that N, B^0 and $R(A^0, B^0)$ are invariant to uniform translations of the time axis.

The sample error of B^0 is given by the formulae [see (8)]:

$$\begin{aligned} \sigma(B^0) &= [R(A^0, B^0)/N(I-2)]^{1/2} \qquad I > 2 \\ &= 0 \qquad\qquad\qquad\qquad\qquad\quad I = 0. \end{aligned} \tag{21}$$

As soon as both transport coefficients are known the molecular weight M can be calculated via Eq. (10).

IV. APPLICATIONS AND DISCUSSION

In order to test the performance of the numerical solution method three sets of artificial "experimental data" have been generated and processed.[3] These data can be considered as being sampled from artificially constructed curves. The first test dealt with curves satisfying the model [the Eqs. (8) and (10)] exactly. In the second test the curves of the first case were shifted upward with respect to the base line (about 5% of the maximum height). This test accounts for those situations in which the height of a curve has a systematic deviation due to an uncertainty in the position of either the base line or the curve. In the third test the right sides of the curves have been omitted in order to check the sensitivity of the method to incomplete data. Table 1 lists the input values of the **m** vectors from which the curves have been constructed and the output values of the **m** vectors as derived by the SEARCH routine. Component m_1 is well reproduced and not very sensitive to inaccurate data. Component m_2 is also well reproduced in the first test case; in the other tests moderate systematic deviations occur. In the second test the SEARCH tries to minimise the deviation from the model by making m_2 systematically greater than the original input value. The absence of the right sides of the curves apparently causes a systematic underestimation of m_2. The output values of m_0, calculated according to (14.2), show a rather strong correlation with m_2. The consistent offset produced in the first curve of each case is caused by the very sharply peaked curves (ratio of lowest to highest value is 1.0 E −14) and to the small number of specified coordinates (five).

The transport coefficients and their deviates are calculated from the

[3] A digital program has been developed named DISEMO. The size of the compiled program is about 9K; the execution time for the processing of the data of one experiment varies between 5 and 15 sec. Requests for a listing of the source module and a sample input form should be addressed to: Polymer Chemistry Laboratory, Department of Chemical Engineering, Twente University of Technology, Enschede, The Netherlands.

TABLE 1
Input and Output Values of Parameter Vector m

	Curve No. and time (sec)				
	1	2	3	4	5
Parameters	1.0E + 03[a]	2.0E + 03	3.0E + 03	4.0E + 03	5.0E + 03
m_0 Input	1.00E − 04	1.00E − 04	1.00E − 04	1.00E − 04	1.00E − 04
(−) 1. Test	0.92E − 04	1.00E − 04	1.00E − 04	1.00E − 04	1.00E − 04
2. Test	0.92E − 04	1.03E − 04	1.04E − 04	1.05E − 04	1.08E − 08
3. Test	0.92E − 04	0.88E − 04	0.98E − 04	1.00E − 04	0.98E − 04
m_1 Input	4.42E − 02	4.89E − 02	5.40E − 02	5.97E − 02	6.59E − 02
(m) 1. Test	4.42E − 02	4.89E − 02	5.40E − 02	5.97E − 02	6.59E − 02
2. Test	4.42E − 02	4.89E − 02	5.40E − 02	5.97E − 02	6.59E − 02
3. Test	4.42E − 02	4.88E − 02	5.40E − 02	5.97E − 02	6.59E − 02
m_2 Input	0.222E − 06	0.493E − 06	0.822E − 06	1.225E − 06	1.719E − 06
(m^2) 1. Test	0.222E − 06	0.493E − 06	0.822E − 06	1.222E − 06	1.716E − 06
2. Test	0.222E − 06	0.507E − 06	0.839E − 06	1.225E − 06	1.775E − 06
3. Test	0.222E − 06	0.368E − 06	0.790E − 06	1.221E − 06	1.659E − 06

[a] E + 03 means 10^3.

TABLE 2
Input and Output Values of Physical Variables[a]

Physical quantity		Input	1. Test	2. Test	3. Test
S	(sec^{-1})	2.46E − 11	2.46E − 11	2.46E − 11	2.46E − 11
$\sigma(S)$	(sec^{-1})	—	4.79E − 15	6.64E − 15	6.38E − 15
D	(m^2/sec)	1.00E − 10	1.00E − 10	1.03E − 10	0.98E 10
$\sigma(D)$	(m^2/sec)	—	0.18E − 12	2.16E − 12	5.43E − 12
M	(kg/kmole)	2.195E + 05	2.200E + 05	2.13E + 05	2.23E + 05

[a] Numerical values of the various constants used in the test computations: r_0 = 4.0E − 02 (m); ω = 6.09E + 04 (rpm); ρ = 1.0 (g/cm^3); $\bar{v}$ = 0.75 (cm^3/g); T = 298.15 (°K); $(\partial \ln \alpha/\partial \ln c)_T$ = 0.9.

output values of m_1 and m_2 as described in the second part of Sect. III. Table 2 lists the preassigned values and the reproduced values of the transport coefficients and the molecular weight. The values of the auxiliary physical quantities used in the computation of M are given in footnote *a* of Table 2. The sedimentation coefficient is not very sensitive to disturbances. Even in the worst case (third test) the relative error remains smaller than 0.1%. The diffusion coefficient and its sample error are much more affected by the upsets. The systematic deviations of m_2 in the second and third test case (see Table 1) propagate into the diffusion coefficient, but the second transformation and the weighted least squares method smooth the scatter to some extent.

These tests with artificial "experimental" data give an indication of the accuracy of the method under various circumstances. But the simultaneous effect of all sources of upsets, i.e., numerically, experimentally and inadequacies of the model can only be observed from real experimental data. Table 3 lists the experimental conditions and the calculated **m** vectors including the normalized weight factors of a test with α-chymotrypsinogen. Table 4 is a survey of the calculated transport coefficients, their sample errors and the molecular weight together with some corresponding data from the literature. The answers obtained by the method of least squares are in satisfactory agreement with the data found in the literature. The numerical errors do not prevail over the experimental inaccuracies. There are some indications that the overall accuracy has been improved by applying this method.

DISCUSSION

The discussion will be restricted to the presented method for data processing and the applicability of the model. The experimental procedure itself is not considered in the discussion because this procedure is common to each data processing method.

TABLE 3
Data of α-Chymotrypsinogen Experiment[a]

Curve	Reduced time (min)	No. of coordinates	m_0	m_1^2 (m²)	m_2 (m²)	w
1	0.0	9	0.1460E − 04	0.4354E − 02	0.1162E − 06	0.1865
2	4.0	11	0.1483E − 04	0.4366E − 02	0.1565E − 06	0.1702
3	8.0	11	0.1419E − 04	0.4386E − 02	0.1903E − 06	0.2104
4	12.0	12	0.1437E − 04	0.4395E − 02	0.2430E − 06	0.2086
5	16.0	12	0.1451E − 04	0.4408E − 02	0.2929E − 06	0.2243

[a] 50 mg/cm³ α-chymotrypsinogen solved in water with 0.9% NaCl and 0.01 M phosphate buffer (pH = 7.3). $\bar{v}$ = 0.736 (cm³/g) (9); ρ = 1.0 (g/cm³); ω = 48000 (rpm); T = 298.15°K; $(\partial \ln \alpha/\partial \ln c)_T$ = 1.0; G = 0.43428E − 03. Recording device: photographic plate.

TABLE 4
Molecular Weight and Transport Coefficients of α-Chymotrypsinogen

Physical quantity	Method of least squares[c]	2. Point method (3)	Svedberg method (3)	Literature
S (sec^{-1})	2.55E − 13	2.60E − 13	2.48E − 13	2.5E − 13 (3)
Sample variance (%)	6.23[a]	13.4[b]	—	—
D (m^2/sec)	0.906E − 10	1.214E − 10	—	1.04E − 10 (3)
Sample variance (%)	6.04[a]	10.4[b]	—	—
M (kg/kmole)	2.405E + 04	2.140E + 04	—	2.200E + 04 (10) 2.500E + 04 (11)

[a] Sample variance according Eq. (21).

[b] According to our deduction from the data in Table II of (3), the deviates given there, are the sample variances of the *sample means of S and D* which have been converted to the sample variances of S and D by multiplying with $\sqrt{n}$ (n = number of observations = 3).

[c] The temperature at which the values in this column have been determined is 25°C. All other values are valid at 20°C.

Problems with the proposed method may arise at three stages:

a. In sampling data from the recorded curves.
b. When applying the mathematical model outside its validity range.
c. During the numerical operations.

The main difficulties in sampling data concern mainly the accuracy of the ordinates of the selected points along a curve. The precision can become rather low, especially near the tails of a curve. It is recommended to sample only at the top and the sides of a curve. The difference in accuracy of the various points could be compensated by assigning weight factors. As no reliable method was known to convert the relative accuracy into a weight factor, in our method all points have the same weight.

The most important violations of the assumptions underlying the mathematical model are threefold. The purity and the mono-dispersity of the solute is sometimes questionable. In that case the basic equations of the model are strictly taken no longer applicable and maintaining the model as the base of the method yields less reliable results. An experimental investigation of these effects may clarify the quantitive aspects of this problem. The deviations from the assumed initial concentration distribution and the occurrence of boundary effects are the second and third source of errors. With increasing run time the consequences of the initial disturbances decrease while the boundary effects become more important. A compromise in run length must be found. This sort of model deviations however are inherent to each model based on Gehatia's method.

The crucial points in the numerical operations are:

a. The effectiveness of the optimalisation.
b. The use of reference parameters in the second transformation.
c. The rather arbitrary definition of the curve weight factors.
d. The assumption that S and D are constant over the concentration range employed.

The effectiveness of the optimalisation is an important factor as the optimalisation results are the starting point for the following numerical operations. The applied routine SEARCH works sufficiently accurately and rapidly, but it is not essential for the method. It can be replaced by any other routine according to the preference of the user.

The application of the reference parameters introduces the potential danger of loss of numerical accuracy because two numbers of the same order of magnitude are subtracted in the second transformation. Secondly, systematic deviations may occur due to incorrect values of the reference

parameters. This only applies to the $Z_{2,i}$ factors as can be seen from (9.2) and (9.3):

$$\frac{dZ_{1,i}}{dt} \text{ is independent of } Z_{1,r},$$

$$\frac{dZ_{2,i}}{dt} \text{ has a dependency on } Z_{1,r} \text{ and } Z_{2,r}.$$

The investigated problems gave no indications that these points had a serious bearing on the results, but they should be mentioned for sake of completeness.

As stated above the definition of the curve weight factors is an arbitrary choice, but it provides a means to account for the relative accuracy of each curve. A different definition would produce different numerical values for the weight factors. However, the selection of the reference curve is hardly affected by the definition. The influence of the value of the weight factors on the transport coefficients has been investigated for the α-chymotrypsinogen experiment. A repetition of the calculations with equal weights and the same reference curve yields: $S = 2.59\ S$ and $D = 0.903 \times 10^{-11}$ (m^2/sec). The differences with the values in Table 4 are rather small and can almost be neglected with respect to the sample errors.

The applicability of the method of least squares is not restricted to the model as presented in Sect. II. Under certain conditions (very small sedimentation coefficient) the factor $\alpha = (e^{2\beta t} - 1)/2\beta t$ becomes virtually equal to one and can be left out. In this case the definition of $Z_{2,i}$ (9.3) simplifies to $Z_{2,i} = m_{2,i} - m_{2,r}$. It is also possible to retain more terms of the power series expansion [see Eq. (3)]. The transformations do not change.

It should be noted that the method allows us to compare predicted values of r_0, t_0 and $F \cdot \Delta n$ with observed values. After the derivation of s and D the quantities r_0 and t_0 can be calculated from the definitions of m_1 and m_2. Applying the definition of m_0 yields the product of $F \cdot \Delta n$. This provides a means to check for a possible discrepancy between model and reality at the initial stage of an experiment.

CONCLUSIONS

In the description of the least squares method the mathematical aspects have been more emphasised than the experimental sides. Summarising the features, the method requires neither the difficult to observe quantities r_0 and t_0 at the top or the area of a refractive index gradient curve, but uses sampled curve heights around the top. Apart from the transport constants and the molecular weight it produces also the same errors of the constants. The performance of the method with respect to the re-

producibility is good. As far as the method has been tested on experimental sedimentation data, the results are encouraging. Therefore, it may be an useful tool for the processing of sedimentation velocity data of proteins, where S and D are independent of the concentration.

APPENDIX A: EQUATIONS OF THE CONCENTRATION GRADIENT

In this appendix it is shown that Eq. (1) of Sect. II is valid for a two-component system, initially consisting of two solutions with a concentration difference. This includes also a two-component system initially with a pure solvent phase and a solution.

The sedimentation process in an infinite sector cell is described by the following set of equations:

$$\frac{\partial c(r, \tau)}{\partial \tau} = \frac{1}{r} \frac{\partial}{\partial r} \left[r \cdot D \cdot \frac{\partial c(r, \tau)}{\partial r} - \beta \cdot r^2 \cdot c(r, \tau) \right]. \tag{A.1}$$

Initial conditions:

$$c(r, o) = c_0 > 0 \qquad \tau = 0 \qquad 0 < r \leqslant r_0$$
$$c(r, o) = c_1 \geqslant c_0 \qquad \tau = 0 \qquad r_0 < r$$

Boundary conditions:

$c(r, \tau)$ and $\frac{\partial c(r, \tau)}{\partial r}$ both finite and continuous for all

$$r > 0 \text{ and all } \tau > 0.$$

The concentration and the concentration gradient can be calculated from this boundary value problem.

By means of a transformation of variables this boundary value problem is converted into a boundary problem that has already been solved [Fujita (4), Chap. 2.5]. The transformation is very similar to the one applied by Fujita when solving the problem (A.1) in the case $c_0 = 0$.

$$\begin{aligned} \lambda &= 2\beta\tau, \\ \epsilon &= 2D/\beta r_0^2 = 1/[\rho_0^2 \cdot (1 - e^{-\lambda})], \\ \chi &= (r/r_0)^2 = (\rho/\rho_0)^2 e^{\lambda}, \\ \theta(\chi, \lambda) &= \frac{c(r, \tau) - c_0 \cdot e^{-\lambda}}{c_1 - c_0}. \end{aligned} \tag{A.2}$$

The qualities ρ and ρ_0 have been defined in Sect. II.

The new boundary value problem reads:

$$\frac{\partial \theta(\chi, \lambda)}{\partial \lambda} = \frac{\partial}{\partial \chi} \left[\chi \left(\epsilon \frac{\partial \theta(\chi, \lambda)}{\partial \chi} - \theta(\chi, \lambda) \right) \right]. \tag{A.3}$$

Initial condition:

$$\theta(\chi, o) = 0 \qquad \lambda = 0 \qquad \chi < 1,$$
$$\theta(\chi, o) = 1 \qquad \lambda = 0 \qquad \chi > 1.$$

Boundary conditions: similar to those of (A.1).

Fujita (4) gives the solution of (A.3). Returning to the original variables yields an expression for $c(\rho,\lambda)$:

$$c(\rho, \lambda) = c_0 e^{-\lambda} + (c_1 - c_0) \cdot e^{-\lambda} \cdot \left[1 - 2 \cdot \int_0^{\rho_0} \gamma \cdot e^{-\rho^2-\gamma^2} \cdot I_0(2\gamma\rho) d\gamma \right]. \quad \text{(A.4)}$$

By putting c_0 equal to zero this equation becomes identical to the solution given by Fujita (4), Gehatia and Katchalski (12) in the case $c_0 = 0$.

Evaluation of the right side of (A.4) gives:

$$c(\rho, \lambda) = c_0 e^{-\lambda} + (c_1 - c_0) e^{-\lambda} \cdot \left[\int_{\rho_0}^{\infty} 2 \cdot \gamma \cdot e^{-\rho^2-\gamma^2} \cdot I_0(2\gamma\rho) d\gamma \right]. \quad \text{(A.5)}$$

Assuming the presence of the bottom does not yet affect the sedimentation process, the upper boundary of the integral can be replaced by ρ_b (defined in Sect. II):

$$c(\rho, \lambda) = c_0 e^{-\lambda} + (c_1 - c_0) e^{-\lambda} \cdot \left[\int_{\rho_0}^{\rho_b} 2 \cdot \gamma \cdot e^{-\rho^2-\gamma^2} \cdot I_0(2\gamma\rho) d\gamma \right]. \quad \text{(A.6)}$$

The derivative of this equation with respect to r gives Eq. (1) of Sect. II. Gehatia (1) proves this assertion in the case $c_0 = 0$. As the calculations are almost identical for $c_0 > 0$ the derivation is not repeated here.

APPENDIX B: SEARCH ROUTINE

The SEARCH routine minimises the objective function values by varying the independent variables, m_1 and m_2. The module has three main parts: (a) initiation; (b) explorations; (c) termination.

The first section establishes the starting point for the first exploration (m_1^s, m_2^s) and the initial step sizes (Δm_1, Δm_2). The quantities m_1^s and m_2^s are derived from Eq. (8) for which purpose the left side has been put equal to the experimental height y_k. After some rearrangements the following equations are obtained:

$$m_{2,k}^s = \Delta r^2 / [\ln(y_k^2/(y_{k1} \cdot y_{k+1})) + 0.5 \ln(r_k^2/(r_{k+1} \cdot r_{k-1}))], \quad \text{(B.1)}$$

$$m_{1,k}^s = r_k + m_{2,k}^s \cdot [\ln(y_{k+1}/y_{k-1}) + 0.5 \ln(r_{k+i}/r_{k-i})]/(2 \cdot \Delta r), \quad \text{(B.2)}$$

where

$$\Delta r = r_{k+1} - r_k = r_k - r_{k-1} \qquad (k < K \text{ and } k \text{ even}).$$

Any three equidistant points suffice to calculate $m_{1,k}^s$ and $m_{2,k}^s$. But the initial guess becomes more accurate when this computation is repeated for several sets of three points and the linear average is taken and applied to form the first base point (m_1^s, m_2^s). The initial step size of m_1 has been set equal to Δr, while $\Delta m_2 = 0.1\ m_2^s$.

The second part controls the explorations. An exploration is defined by the actions listed below:

a. Computation of the value of the objective function in a base point.

b. Computation of the coordinates and the value of the objective function in four points surrounding the base point. If any of these points coincides with a previously explored point then this point is skipped.

c. The point with the lowest value of the objective function becomes the next base point.

d. Evaluation:

Case 1: The new selected base point coincides with the previous one. The step sizes are reduced proportional to the steepest slope found in the last exploration. However, when the number of explorations will exceed 50 or when all step sizes are already smaller than 0.1% of the actual magnitude of the corresponding independent variables, then the control is transferred to the termination section. Otherwise a new exploration starts.

Case 2: The new selected base point does not coincide with the previous one. The search will be continued unless the number of explorations exceeds 50 (preventing excessive calculations). When two consecutive shifts of the base point in the same direction occur, the step size in that direction is doubled.

Figure 1 gives a graphical representation of a few explorations. The numbers assigned to the points indicate in which order the space

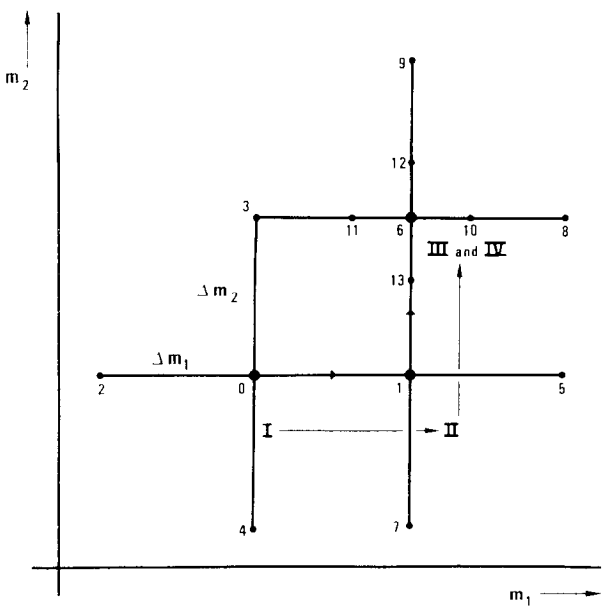

FIG. 1. Example of consecutive explorations.

is explored. The sum of square residues is calculated in base point 0. Afterwards the points 1, 2, 3 and 4 follow. Suppose point 1 has the smallest sum, then point 1 becomes the second base point (II). Point 6 is both third and fourth base point, assuming the step sizes have been reduced between the third and fourth exploration.

The termination section declares the last explored base point to be the minimum. A message is given when the number of explorations exceeds 50.

ACKNOWLEDGMENT

The authors thank Professor Schuyer (Dept. of Chemistry) for helpful discussions.

REFERENCES

1. GEHATIA, M. (1959) *Kolloid Z.* u. *Z. Polymere* **167,** 1.
2. GEHATIA, M., AND HUBNER, L. (1960) *Z. Naturforsch. B* **15,** 351.
3. JAENICKE, R., AND GEHATIA, M. (1960) *Biochim. Biophys. Acta* **45,** 217.
4. FUJITA, H. (1962) Mathematical Theory of Sedimentation Analysis, Academic Press, New York.
5. DE GROOT, S. R., AND MAZUR, P. (1963) Non-equilibrium Thermodynamics, North-Holland, Amsterdam.
6. WILDE, D. J., AND BEIGHTER, C. S. (1967) Foundations of Optimization, Prentice Hall, New York.
7. RECTORYS, E. K. (1970) Survey of Applicable Mathematics, M.I.T. Press, Cambridge, MA.
8. DRAPER, N. R., AND SMITH, H. (1966) Applied Regression Analysis, Wiley, New York.
9. BRANDRUP, J., AND IMMERGUT, E. H., (eds.) (1966) Polymer Handbook, Wiley (Interscience), New York.
10. H. M. RAUEN (ed.), (1964), *Biochemisches Taschenbuch,* 2nd ed., Springer–Verlag, Berlin.
11. Mann Research Laboratory, New York.
12. GEHATIA, M., AND KATCHALSKI, E. (1959) *J. Chem. Phys.* **30,** 1334.