determination of confidence intervals for parameters in nonlinear models. We have applied the approach to the analysis of equilibrium sedimentation data and have demonstrated that, although models for analyzing such data are formally nonlinear they are functionally linear. As such, linear approximation confidence intervals for the parameters are adequate for these models and data sets.

Further, we have been able to examine the effect of implementing a multiple independent variable approach (in this case, using multiple rotor speeds) on the precision of the analysis. We found that the standard errors of the parameters were reduced and that this is accounted for by either the increase in the number of data points or the decreases in parameter correlation. In this case, profiling helped to visualize the effect on the sum of squares surface of reducing parameter correlation, making the effect of the small decreases in the correlation of some parameters more evident.

Using profiling, it should be easy to explore other methods for the improvement of the analysis of ultracentrifugation data and to be able to quantitate the improvement. With the above discussion as an example, it is likely that the profiling approach should be quite useful and broadly applicable in the analysis of data in terms of nonlinear models.

Acknowledgments

The authors gratefully acknowledge Michael L. Doyle and Paul McAllister for input to and careful reading of the manuscript, and Sergio Rose for help with some of the original experiments. We also thank Michael D. Minnich and Jeffrey Culp for supplying the purified p24gag and Christine Debouck for constructing the expression system for the protein.

[22] Boundary Analysis in Sedimentation Velocity Experiments

By WALTER F. STAFFORD III

Introduction

Measurement of the sedimentation velocity of macromolecular particles was the first type of analysis to which the analytical centrifuge was applied.[1-3] This chapter describes briefly the earlier methods of analysis

[1] T. Svedberg and J. B. Nichols, *J. Am. Chem. Soc.* **45,** 2910 (1923).

of sedimentation velocity data that were used before the advent of digital computers. Then, using this discussion as a point of reference, it proceeds to describe approaches that have become practical because of the availability of computers. It then goes on to describe newer techniques that have become possible through the introduction of modern on-line digital data acquisition systems. Special emphasis is placed on techniques employing the time derivative of the concentration distribution.[4-6]

Traditionally, sedimentation transport experiments have been observed in two main ways, as either the concentration or concentration gradient as a function of radius depending on the type of optical system employed. The schlieren optical system displays the boundary in terms of refractive index gradient as a function of radius; the Rayleigh optical system, in terms of refractive index as a function of radius; and the absorption optical system, in terms of optical density as a function of radius. This chapter concerns itself mainly with treatment of data obtained with the latter two types of systems since the acquisition and analysis of data from these systems can be nearly completely automated.

Subjects to be discussed are techniques for smoothing and differentiation as well as a new analysis technique that uses the time derivative of the concentration profile. Use of the time derivative results in an automatic baseline elimination with a consequent increase in accuracy. Combination of the time derivative with an averaging procedure has resulted in an increase of 2–3 orders of magnitude in precision. The time derivative technique produces what is referred to as an apparent sedimentation coefficient distribution function, $g(s^*)$, where the symbols are defined below. The function $g(s^*)$ versus s^* is very nearly geometrically similar to the corresponding schlieren pattern that represents dn/dr versus r (where n is the refractive index and r is the radius) and, therefore, can be used in any situation that one would have used a schlieren pattern. The advantage of using the averaged $g(s^*)$ is that it has about 2 orders of magnitude higher signal-to-noise ratio than dn/dr and, therefore, can be used to study interacting systems that previously were inaccessible to either the absorbance or Rayleigh optical systems. Extrapolation procedures for minimizing the effects of diffusion on the resolution of boundaries are also discussed briefly.

---

[2] T. Svedberg and H. Rinde, *J. Am. Chem. Soc.* **45,** 943 (1923).

[3] T. Svedberg and H. Rinde, *J. Am. Chem. Soc.* **46,** 2677 (1924).

[4] W. F. Stafford, *in* "Methods for Obtaining Sedimentation Coefficient Distributions" (S. E. Harding, A. J. Rowe, and J. C. Horton, eds.) p. 359. Royal Society of Chemistry, Cambridge.

[5] W. F. Stafford, *Anal. Biochem.* **203,** 295 (1992).

[6] W. F. Stafford, *Biophys. J.* **61,** A476 (1992).

For convenience, we refer to the data from either system in terms of concentration rather than optical density or refractive index. However, various averages computed from these data will be weighted according to either the extinction coefficients or the refractive index increments of the various components in a mixture depending on the optical system employed.

The concentration data acquired by either optical system can be thought of as composed of two main parts: a time-dependent part that is due to the transport of macromolecule and a time-independent background part due to inhomogeneities in the optical system, detector, and cell windows. If the concentration is sufficiently high, the background part often can be ignored. However, in those cases for which the background contribution is a significant fraction of the signal, a correction for the background must be carried out by some means. The usual ways have been to perform a separate run without protein in the cell or to allow the material to pellet completely before taking a background scan. After the sedimentation run has been completed, the background contribution is subtracted from each scan. The first method is usually satisfactory as long as there have been no changes in the background between the main scans and the background scan. The second method requires that there be no slowly sedimenting material. Accumulation of dirt from drive oil, fingerprints, or the effects of moisture condensation, for example, could be sources of variation in the background between the main run and the background run. Ideally, a background run should be performed both before and after the main run to determine whether there have been any changes during the run. The importance of a background correction becomes evident if the concentration data are to be treated by numerical differentiation to produce concentration gradient curves. Variations in the background can often be of the same magnitude as those of the concentration and can obscure the true gradient curves.

Methods of Data Acquisition

Traditionally, acquisition of data from the schlieren and Rayleigh optical systems has been done by photography on glass plates or estar thick based film. Subsequent analysis of the photographs was carried out visually using an optical microcomparator with manual logging of the data. Various schemes for automating this process have been devised. A particularly useful procedure was devised by DeRosier *et al.* for analysis of Rayleigh interferograms.[7] Essentially, this procedure, along with various

[7] D. J. DeRosier, P. Munk, and D. Cox, *Anal. Biochem.* **50,** 139 (1972).

derivatives of it, uses a Fourier transform analysis to extract the phase changes, and hence the concentration changes, associated with deflection of fringes in interferograms. A second advance was the introduction of on-line video photography to eliminate the wet photography steps. With the introduction of charge-coupled device (CCD) video cameras, Laue[8] has devised an on-line video acquisition system that has become the basis for many current video systems. A video acquisition system for the Beckman Instruments (Palo Alto, CA) Model E analytical ultracentrifuge that allows very rapid capture and analysis of the whole cell image of Rayleigh interferograms every 4 sec has been devised by Liu and Stafford[9] and allows the rapid data acquisition necessary for time derivative analysis with multicell rotors. For example, images from five cells in a Model-E AN-G rotor can be acquired, converted to fringe displacement as a function of position, and the results stored on disk every 20 sec.

The Spinco Division of Beckman Instruments has introduced the Optima XL-A analytical ultracentrifuge equipped with modern UV scanning optics. Profiles of absorbance as a function of radius are automatically acquired and stored as data files readable by other software.

Measurement of Transport in Analytical Ultracentrifuge

The basis for the analysis of transport in the ultracentrifuge is the continuity equation presented by Lamm[10]

$$\left[\frac{\partial c(r,t)}{\partial t}\right]_r = \frac{1}{r}\frac{\partial}{\partial r}\left\{Dr\left[\frac{\partial c(r,t)}{\partial r}\right]_t - \omega^2 r^2 s c(r,t)\right\}_t$$

where c is the concentration as a function of radius and time, r is the radius, t is time, s is the sedimentation coefficient, D is the diffusion coefficient, and ω is the angular velocity of the rotor. Flux, J, owing to sedimentation alone is given by the product of the velocity of the particles and their concentration:

$$J_{\text{sed}} = \frac{dr}{dt}c = \omega^2 src$$

Flux owing to diffusion alone is given by Stokes' first law and is proportional to the concentration gradient:

$$J_{\text{diff}} = -D\left(\frac{dc}{dr}\right)_t$$

[8] T. M. Laue, Ph.D. Dissertation, University of Connecticut, Storrs (1981).
[9] S. Lin and W. F. Stafford, *Biophys. J.* **61,** A476 (1992).
[10] O. Lamm, *Ark. Mat. Astron. Fysik.* **21B** (1929).

Traditional Methods of Analysis

A detailed account of the traditional manual methods of data analysis is not given since they are adequately represented in the literature[11-13] and have been supplanted mainly by methods made practical by the introduction of computers. For sedimentation velocity experiments using the schlieren optical system, the main piece of information obtained was the radial position of the peak in the refractive index gradient curve as a function of time. One obtained the sedimentation coefficient from the slope of a plot of the logarithm of the radial position of the peak versus time. Similarly, for the UV photoelectric scanner and the Rayleigh optical systems, one plotted the logarithm of the radial position of the midpoint of the sedimenting boundary versus time. However, neither of these methods will give highly accurate results unless the boundary is symmetrical; therefore, they can be applied rigorously only to a hydrodynamically ideal, homogeneous, monodisperse solution.

The correct method for obtaining accurate estimates of the sedimentation coefficient for cases that exhibit asymmetric or polymodal boundaries is the so-called second moment method in which an equivalent boundary position is computed. It gives the position of a hypothetical boundary that would have been observed in the absence of diffusion and polydispersity (Fig. 1). The rate of movement of the equivalent boundary position gives the weight average sedimentation coefficient for the system and corresponds to the weight average velocity of particles in the plateau region at the plateau concentration. The equivalent boundary position can be computed in several different ways depending on the form of the data.[11,13,14] The term second moment is generally applied to data obtained with the schlieren optical system from which the equivalent boundary position is obtained as the second moment of the refractive index gradient curve:

$$\langle r^2 \rangle = \frac{\int_{r_m}^{r_p} \frac{dc}{dr} r^2 \, dr}{\int_{r_m}^{r_p} \frac{dc}{dr} \, dr}$$

[11] C. Chervenka, "A Manual of Methods," Spinco Division of Beckman Instruments, Palo Alto, California, 1973.
[12] T. Svedberg and K. O. Pederson, "The Ultracentrifuge." Oxford Univ. Press, New York, 1940.
[13] H. K. Schachman, "Ultracentrifugation in Biochemistry." Academic Press, New York, 1959.
[14] H. Fujita, "Foundations of Ultracentrifugal Analysis." Wiley, New York, 1976.

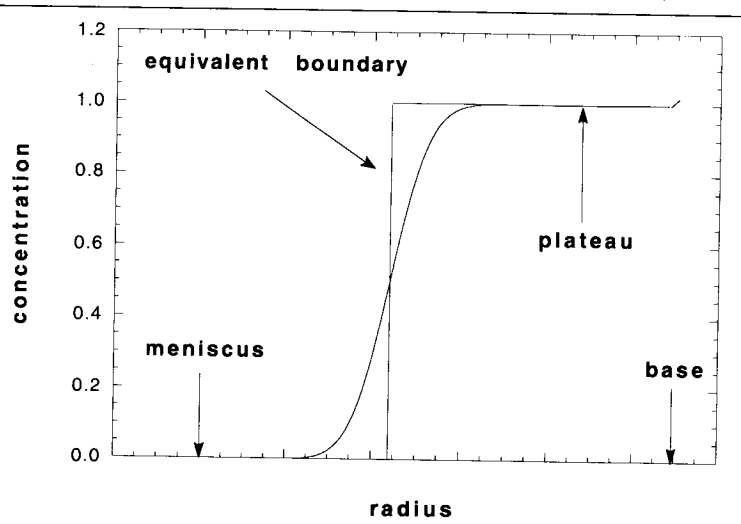

FIG. 1. Hypothetical sedimentation boundary showing the meniscus, equivalent boundary position, plateau, and base of the cell.

However, when the data are obtained as concentration versus radius instead of concentration gradient versus radius, it is convenient to compute the equivalent position from the following relation:

$$\langle r^2 \rangle = \frac{\int_{r_m}^{r_p} r^2 \, dc}{\int_{r_m}^{r_p} dc}$$

After obtaining the equivalent boundary position at various times, one then computes the weight average sedimentation coefficient from the slope of a plot of the logarithm of the equivalent boundary position versus time. The slope of a plot $\ln(\langle r^2 \rangle)$ versus t is $2\omega^2 s_w$, where s_w is the weight average sedimentation coefficient. It should be noted that this method is valid only if the concentration gradient is zero at r_p. An appropriate baseline correction is required for high accuracy.

Transport Method

The transport method is used in cases for which diffusion does not allow the boundary to move away completely from the meniscus. It is similar in computational complexity to the equivalent boundary method. Essentially, one measures the amount of material between the meniscus and some point in the plateau region that is removed by sedimentation

during a given time interval. This amount can be converted to J_{sed}, and, with a knowledge of the plateau concentration, one can compute the sedimentation coefficient. The validity of this method also depends on a negligible gradient in the plateau region. The relation may be derived in the following way (for other derivations, see refs. 13 and 14).

By integrating the Lamm equation,

$$\left[\frac{\partial c(r,t)}{\partial t}\right]_r = \frac{1}{r}\frac{\partial}{\partial r}\left\{Dr\left[\frac{\partial c(r,t)}{\partial r}\right]_t - \omega^2 r^2 s c(r,t)\right\}_t$$

we have

$$\int_{r_m}^{r_p}\left[\frac{\partial c(r,t)}{\partial t}\right]_r r\,dr = \left\{Dr\left[\frac{\partial c(r,t)}{\partial r}\right]_t - \omega^2 r^2 s c(r,t)\right\}_{r_m}^{r_p}$$

and now, after bringing the differential operator outside of the integral, we have

$$\frac{\partial}{\partial t}\left[\int_{r_m}^{r_p} c(r,t) r\,dr\right]_r = \left\{Dr\left[\frac{\partial c(r,t)}{\partial r}\right]_t - \omega^2 r^2 s c(r,t)\right\}_{r_m}^{r_p}$$

and since there is no net transport across the meniscus, we have the condition that $\left\{Dr_m\left[\frac{\partial c(r_m,t)}{\partial r}\right]_t - \omega^2 r_m^2 s c(r_m,t)\right\} = 0$ at r_m. We also have $\partial c/\partial r = 0$ at r_p, giving us an expression for mass transport from the region between r_m and r_p across the cylindrical surface at r_p:

$$\frac{d}{dt}\left[\int_{r_m}^{r_p} c(r,t) r\,dr\right] = -\omega^2 r_p^2 s c(r_p,t)$$

If we designate the integral as

$$Q(t) = \int_{r_m}^{r_p} c(r,t) r\,dr$$

and note that $c(r_p,t) = c_0 \exp(-2\omega^2 s_w t)$ where c_0 is the initial concentration, then we have

$$\frac{dQ}{dt} = -\omega^2 r_p^2 s c_0 \exp(-2\omega^2 s t)$$

By rearranging and integrating and since

$$Q(t=0) = c_0 \int_{r_m}^{r_p} r\,dr = \frac{c_0}{2}(r_p^2 - r_m^2)$$

we have

$$Q(t) = \frac{c_0 r_p^2}{2} \exp(-2\omega^2 st) - \frac{c_0 r_m^2}{2}$$

By rearranging and taking the logarithm of both sides, we arrive at

$$\ln\left[\frac{2Q(t)}{c_0 r_p^2} + \left(\frac{r_m}{r_p}\right)^2\right] = -2\omega^2 st$$

A plot of the left-hand side versus t will have a slope of $-2\omega^2 s$.

For a polydisperse system, the value of s obtained is the weight average sedimentation coefficient, s_w, for the mixture, and c_0 is the total initial concentration. Note again that this method requires that there be no gradient in the plateau region at r_p. It also requires a knowledge of c_0.

Transport Method Using Time Derivative

A simple variant of the transport method can be used if the time derivative of the concentration is available.[15] Starting again with the Lamm equation, by rearranging and integrating, but now keeping the differential operator inside the integral, we have

$$\int_{r_m}^{r_p} \left(\frac{\partial c}{\partial t}\right)_r r\, dr = \left[Dr\left(\frac{\partial c}{\partial r}\right)_t - \omega^2 r^2 sc\right]_{r_m}^{r_p}$$

Noting again that transport across the meniscus is zero and that $\partial c/\partial r = 0$ in the plateau region at r_p, we have after rearranging

$$s = \frac{1}{-\omega^2 r_p^2 c_p} \int_{r_m}^{r_p} \left(\frac{\partial c}{\partial t}\right)_r r\, dr$$

Again this approach requires that there be no gradient in the plateau region at r_p. The values of $\partial c/\partial t$ as a function of r can be estimated with sufficient accuracy by subtracting concentration profiles closely spaced in time so that $(c_2 - c_1)/(t_2 - t_1)$ at each radial position can be used in place of $\partial c/\partial t$, and $(c_{p,1} + c_{p,2})/2$ in place of c_p, where the numerical subscripts refer to the two scans. For numerical computation, this equation can be recast as:

$$s = \frac{2}{-\omega^2 r_p^2 (c_{p,1} + c_{p,2})(t_2 - t_1)} \sum_{r_m}^{r_p} (c_2 - c_1)_i r_i \Delta r$$

[15] W. F. Stafford III, unpublished (1994).

Again, for a polydisperse system, the procedure gives the weight average sedimentation coefficient, s_w. This method is insensitive to noise because the integration averages out the random noise and is unaffected by the baseline contributions except for the determination of c_p.

Smoothing and Differentiating

To obtain a usable derivative curve, dc/dr, from the concentration profile by numerical differentiation, some degree of smoothing usually will be required. This section discusses some techniques used successfully in our laboratory. There are many variations of these techniques that will also work well if applied properly. The particular methods discussed below are presented because they are simple to use and produce reasonably good results.

Although smoothing will always introduce some degree of systematic dispersion and, therefore, should be carried out with caution, it can be justified in many cases as long one is aware of the magnitude and type of errors introduced. Several methods of smoothing that produce minimal distortion of the data are discussed along with a Fourier analysis of their frequency response.

As an aside, if one is using least squares fitting for parameter estimation, one should not smooth before fitting since the least squares procedure is itself a smoothing process, and the analysis of the residual noise is necessary for evaluation of the "goodness of fit" and for computation of the confidence limits of the fitting parameters. In the procedure described below, the noise of the original, unsmoothed data is used to compute the error bars for the smoothed $g(s^*)$ plots.

Smoothing and Differentiating as Filtering

Smoothing and differentiation can be considered as filtering processes and have an associated frequency response (see Diagram 1).[16]

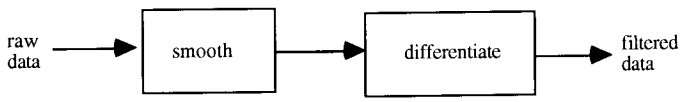

DIAGRAM 1

[16] C. S. Williams, "Designing Digital Filters." Prentice-Hall, Englewood Cliffs, New Jersey, 1986.

Analysis of Frequency Response of Smoothing Filters

The perfect smoothing filter is one that will remove the unwanted noise and reveal the pure signal without distorting it. In a real situation, this will not be possible if the noise and the signal have overlapping power spectra. If any of the spatial frequency components of the noise are in the same range as those of the true concentration distribution, then those components cannot be removed without also removing them from the data. Therefore, a compromise will have to be reached.

Procedure for Determining Frequency Response Given Filter Coefficients

Briefly, the procedure is as follows (see Ref. 16). First, determine the impulse response of the filtering operation to get the coefficients of the filter. Determining the coefficients in this manner is useful since it would allow one to create a filter that could be applied in one pass instead of several passes if desired. Second, compute the Fourier transform of the filter coefficients to obtain the power spectrum. The power spectrum is the frequency response of the filtering operation. This series of operations is represented in Diagram 2.

For example, if one wanted to know the frequency response for three passes of a simple moving average over the data, one could pass the moving average over the unit impulse three times to obtain the coefficients of the equivalent single-pass filter and then subject those coefficients to a Fourier transform. The power spectrum of that Fourier transform is the frequency response for the combined filtering operation. Figure 2a shows the results of passing a 5-point moving average over the unit impulse four times. Figure 2a shows a plot of the filter coefficients after each pass, and Fig. 2b shows the frequency response of each of the filters represented in Fig. 2a. A single pass of a moving average results in large side lobes in the frequency response curve allowing some of the high frequency components to pass through. The second and third passes reduce the side lobes to much smaller values, and subsequent passes do not improve the overall frequency response very much. Although each pass lowers the

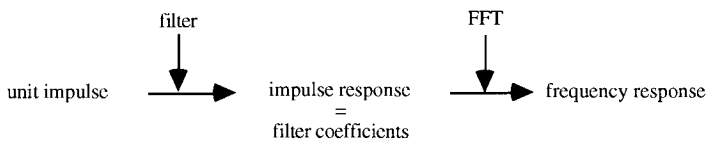

DIAGRAM 2

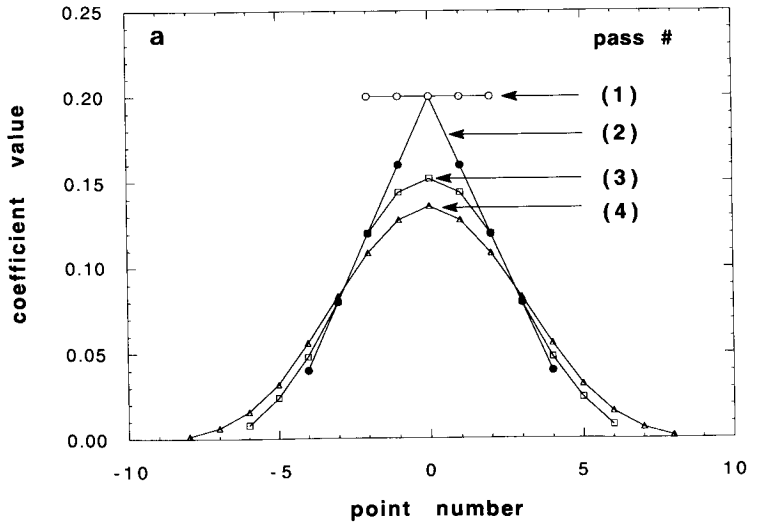

FIG. 2. (a) Plot of the filter coefficients obtained by passing a 5-point moving average over the unit impulse four times. For a 5-point centered moving average filter the coefficients are defined by the following relationship: $c_{-2}x_{i-2} + c_{-1}x_{i-1} + c_0x_i + c_{+1}x_{i+1} + c_{+2}x_{i+2}$, where c_k are the filter coefficients and x_i are a subset of the data to be filtered. For this moving average, the c_k values are all equal to 0.2. On the first pass, one of the $x_i = 1$ (i.e., the unit impulse) and the other x values are set equal to zero. After the first pass, the x_i are replaced by the filter coefficients which now become the data for the second pass of the moving average. Each new set of filter coefficients is generated by repeating this process. (b) Plot of the frequency response of each of the filters represented by the sets of filter coefficients obtained from the impulse response shown in (a). θ is defined by the following equations that were used to compute the Fourier transform of the filter coefficients. The sine transform is given by $S(\theta) = \Sigma\,\Sigma\, c_k \sin(k\theta)$; the cosine transform by $C(\theta) = \Sigma\,\Sigma\, c_k \cos(k\theta)$; and the power spectrum (i.e., frequency repsonse) by $P(\theta) = [S(\theta)^2 + C(\theta)^2]^{1/2}$, where $\theta = i\pi/N$ and $-N < i < N$, and the sums are over k and i, respectively. N was 300 in this case.

cutoff frequency somewhat, the cutoff frequency is best controlled by varying the number of points in the moving average. Therefore, smoothing and differentiation can be combined into one process, if the appropriate set of coefficients is chosen by the impulse response method.

The rest of this discussion is confined to the repeated application of moving averages (i.e., first-order polynomials) since they can be computed rapidly by recursion, and almost any desired low-pass frequency response can be obtained. For example, it is quicker to pass the 5-point moving average three times than it is to pass the equivalent 13-point filter once since the coefficients of the moving average are all equal and the first

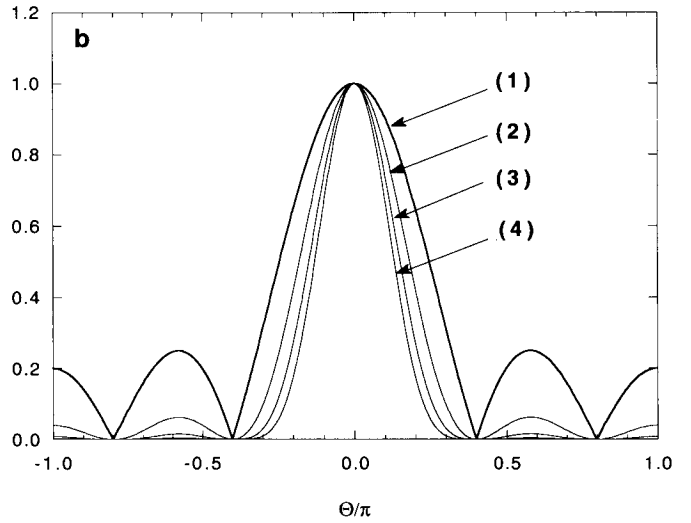

FIG. 2. (*continued*)

point can be removed and the next added without recomputing all the intermediate terms. However, the 13-term equation has unequal coefficients requiring that all 13 terms be recomputed before adding them to compute the value for the smoothed point. The recursion formula for the n-point moving average is

$$\langle y_{i+1}\rangle = \langle y_i\rangle - \frac{y_{i-(n-1)/2}}{n} + \frac{y_{i+1+(n-1)/2}}{n}$$

where the first smoothed value is given by $\langle y_{(n+1)/2}\rangle = \frac{1}{n}\sum_{j=1}^{j=n} y_j$.

The speed can be increased somewhat more, eliminating all the multiplications on each pass, by casting the process as

$$n\langle y_{i+1}\rangle = n\langle y_i\rangle - y_{i-(n-1)/2} + y_{i+1+(n-1)/2}$$

and then after the pth pass dividing each term by n^p (where p is the number of passes) for a total of N multiplications (where N is the number of data points) and $2pN$ additions. Incidently, smoothing with formulas for higher order polymonials also require at least three passes to reduce the side lobes to acceptable levels, and they also have unequal coefficients.

The justification for using repeated applications of a simple moving average, instead of either higher order polynomials or more complicated

filtering operations that necessitate convolution, is the speed of application. For example, the three passes of the recursive moving average (this is not a recursive filter, by the way) to a data set of N points requires approximately $6N$ additions with N multiplications to rescale the data at the end. Application of a higher order polynomial, for which all the terms would have to be recomputed after shifting to each new point, would require approximately N^2 multiplications and additions.

The filtering process is essentially the convolution of the smoothing polynomial with the data; therefore, application of a sliding polynomial to the data can be treated in terms of convolution. Convolution can be carried out rapidly using the fast Fourier transform (FFT).[16] The convolution theorem states that

$$\mathrm{FFT}(f * g) = \mathrm{FFT}(f)\mathrm{FFT}(g)$$

where g is the data and f is the filter and the asterisk denotes convolution. $\mathrm{FFT}(f)$ is just the frequency response of the filter and may be considered to be known since it need be computed only once ahead of time. Convolution by this method would require one FFT, followed by N complex multiplications and followed by an inverse FFT to arrive at the filtered data, $f * g$. Each FFT requires $N \log_2 N$ complex multiplications (4 multiplications and 2 additions) and complex additions (2 additions) for a total of $8N \log_2 N$ multiplications and additions. Because of the savings in computer time afforded by the repeated application of the recursive moving average, it is used almost exclusively in our laboratory. It is interesting to note that the frequency response of the 3-pass moving average filter is nearly identical to that of the popular von Hann window function often used in smoothing by convolution. Figure 3 shows a plot of the window coefficients for a 13-point von Hann window compared to those for the 13-point equivalent 3-pass moving average filter. It is gratifying that the faster recursive moving average procedure can give essentially the same results as the more time-consuming and complex convolution procedure.

In the computer algorithm used in our laboratory, the end points are handled by starting with a 3-point moving average for the first three points and then by expanding the window 2 points at a time until the full window is reached. The full window is used until the end of the data is reached, where the process is reversed by decreasing the window size 2 points at a time. The moving average is passed over the data three times. Variation in the degree of smoothing is controlled by changing the window size. The typical default for smoothing our time derivative data is to use a window spanning about 2% of the distance from the meniscus to the base.

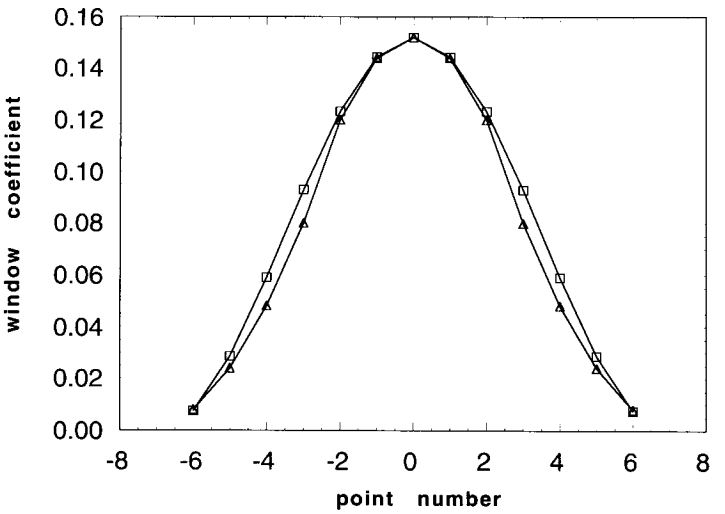

FIG. 3. Comparison of the filter coefficients of the 3-pass moving average (△) to those of the corresponding 13-point von Hann window function (□).

If more detail is desired, the smoothing can be turned off. It is sometimes necessary to turn the smoothing off to track down outliers.

Differentiation

The derivative of the concentration profile can be estimated by various polynomial fitting techniques. Differentiation and smoothing often have to be combined to get satisfactorily smooth curves. As mentioned above, a single polynomial fitting equation can be derived that will accomplish both smoothing and differentiation by using the coefficients obtained after application of the smoothing and the differentiating functions to the unit impulse function. The frequency response of the combined process can be computed by taking the Fourier transform of the coefficients. Application of the resulting polynomial can be accomplished by FFT convolution as mentioned above.

A simple and successful procedure used in our laboratory for differentiation has been to use a sliding 4-point cubic polynomial fit to equally spaced data spanning 13 points typically spaced at about 40 μm intervals in the centrifuge cell:

$$\left(\frac{dc}{dr}\right)_i = \frac{1}{12\Delta r}(c_{i-6} - 8c_{i-3} + 8c_{i+3} - c_{i+6})$$

where Δr is the radial spacing. This differentiating procedure combined with smoothing was used in a study of filament formation of *Acanthamoeba* myosin-II.[17]

Computation of Apparent Sedimentation Coefficient Distribution Functions

The methods to be described for computation of apparent sedimentation coefficient distribution functions are all derived from theoretical relationships predicated on the assumption that the diffusion coefficient of each species is zero. A rigorously correct distribution function will be obtained only for that case. The assumption of zero diffusion coefficient can be made with good approximation for many solutions of very high molecular weight polymers; however, when these relations are used in cases of nonnegligible diffusion, one obtains an *apparent* distribution function that has been broadened by the effects of diffusion. There are several extrapolation methods that can be used to correct for the effects of diffusion.[4,18]

The techniques described below concentrate mainly on the uses of the uncorrected apparent distribution function as a tool for sedimentation boundary analysis, especially as a technique for revealing details of boundary shape for the analysis of both heterogeneous and reversibly associating systems. As mentioned above, the apparent differential distribution function [$g(s^*)$ versus s^*] and the refractive index gradient (schlieren) curve (dn/dr versus r) are very nearly geometrically similar to one another, and, therefore, both show the same details of boundary shape. A derivative pattern is useful because it can reveal features that often are not otherwise obvious by direct inspection of a concentration boundary profile. In addition, the increase in signal-to-noise ratio achieved by using the time derivative combined with averaging to compute $g(s^*)$ has extended the ability to investigate interacting systems to much lower concentrations than previously possible with any given optical system.[4-6]

Differential Distribution Functions

Before the methods themselves are discussed, it is worth making a point concerning nomenclature. In the past, the apparent distribution function has been designated $g^*(s)$ versus s. However, when these patterns are used for boundary analysis it seems more appropriate to designate them as $g(s^*)$ versus s^* since s^* is a radial coordinate whose value can

[17] J. H. Sinard, W. F. Stafford, and T. D. Pollard, *J. Cell Biol.* **109,** 1537 (1989).
[18] K. E. Van Holde and W. O. Weischet, *Biopolymers* **17,** 1387 (1978).

be interpreted as a sedimentation coefficient only under special circumstances. One special circumstance already mentioned is in the case of negligible diffusion. Another is that the value of s^* at the peak of a symmetrical boundary for a monodisperse solute corresponds to within a very good approximation to the sedimentation coefficient, s, for that species.

Method Using Spatial Derivative

In 1942, Bridgman[19] presented an equation for computing the differential distribution function for systems exhibiting no diffusion:

$$g(s) = \frac{dc}{dr}\left(\frac{\omega^2 rt}{c_0}\right)\left(\frac{r}{r_m}\right)^2$$

where c_0 is the initial loading concentration and the other symbols have their usual meaning. The radial derivative can be obtained either directly from the schlieren optical system or by numerical differentiation of the concentration profile from either the absorbance or the Rayleigh optical system. Various useful differentiation and smoothing methods have been discussed above, and a recent example of the use of the radial derivative, computed by numerical differentiation of absorbance traces from the UV photoelectric scanner of the Beckman Instruments Model-E ultracentrifuge, to compute averaged $g^*(s)$ curves in a study of filament formation of *Acanthamoeba* myosin-II, has appeared[17] as already mentioned above.

Method Using Temporal Derivative

Sedimentation patterns in the form of concentration as a function of radius are acquired and stored in digital form. Pairs of curves appropriately spaced in time are subtracted from one another at corresponding radii to produce difference curves. If the time difference between the curves is made sufficiently small, the difference curve will be a good approximation to the time derivative of the original sedimenting boundary at each radial position.[5] The use of the time derivative results in an automatic baseline elimination. When used in conjunction with rapid data acquisition systems that allow signal averaging, the use of the time derivative can result in several (from 2 to 3) orders of magnitude increase in the signal-to-noise ratio.

After the pairs of concentration curves are subtracted to give a set curves of dc/dt versus r, the x axis of each derivative curve is then

[19] W. B. Bridgman, *J. Am. Chem. Soc.* **64,** 2349 (1942).

converted to s^* by the following relationship, which transforms the dc/dt curves from the stationary reference frame of r to the moving reference frame of s^*:

$$s^* = \frac{1}{\omega^2 t} \ln\left(\frac{r}{r_m}\right) \tag{1}$$

In this moving reference frame, the dc/dt curves obtained over a small time interval will superimpose and can be averaged to reduce the random noise contributions. The value of t to be used in Eq. (1), and in others used for the numerical computation of $g(^*s)$, is the harmonic mean of the sedimentation times of the concentration profiles. For data obtained from the XL-A ultracentrifuge, one must get the sedimentation time from the value of $\omega^2 t$ included in each output file.

Detailed Procedure

The specific example described here applies to data obtained with the Beckman Instruments Optima XL-A analytical ultracentrifuge. The data are collected in digitized form and stored as text files from which two arrays of data can be extracted; one is the absorbance, and the other is the value of the radius at which the absorbance was measured at a particular sedimentation time, say, t_1:

$$c(1, t_1), c(2, t_1), c(3, t_1), \ldots, c(n, t_1)$$
$$r(1, t_1), r(2, t_1), r(3, t_1), \ldots, r(n, t_1)$$

The value of $(\partial c/\partial t)_r$ at each radial position is estimated by subtracting pairs of sedimentation patterns point by point at corresponding values of r. However, because the data from any two successive scans usually are not acquired at the same radial positions on the XL-A ultracentrifuge, the curves have to be interpolated onto the same radial grid before the subtraction can be performed. Typically, a grid spacing of 0.002 cm is chosen for a grid between 5.8 and 7.2 cm. After a pair of scans has been interpolated onto this grid, they can be subtracted to obtain $\Delta c/\Delta t$ at each value of r. The result is an estimate of $(\partial c/\partial t)_r$ with the automatic elimination of the time-independent optical background. After subtracting and dividing by Δt, we have

$$\begin{aligned} c(r, t_2)_{obs} &= c(r, t_2)_{true} + c(r)_{background} + \text{noise} \\ -[c(r, t_1)_{obs} &= c(r, t_1)_{true} + c(r)_{background} + \text{noise}] \\ \hline \frac{\Delta c(r)_{obs}}{\Delta t} &= \frac{\Delta c(r)_{true}}{\Delta t} + 0 + \sqrt{2} \cdot \text{noise} \end{aligned}$$

If the time interval is sufficiently small that $\Delta c/\Delta t \approx \partial c/\partial t$, then $\Delta c/\Delta t$ can be converted to an apparent sedimentation coefficient distribution function according to the procedure previously presented[5]:

$$g(s^*) = \left(\frac{\partial c}{\partial t}\right)_{\text{corr}} \left(\frac{1}{c_0}\right) \left[\frac{\omega^2 t^2}{\ln(r_m/r)}\right] \left(\frac{r}{r_m}\right)^2 \tag{2}$$

where $(\partial c/\partial t)_{\text{corr}}$ is the value of $(\partial c/\partial t)_r$ corrected for the plateau contribution as described[5] and repeated in detail below. The product of c_0 and $(r_m/r)^2$ is the plateau concentration at each point, and if these two factors are eliminated from Eq. (2) we have the unnormalized distribution function referred to below as $\hat{g}(s)$. This function can be used in much the same way as a schlieren pattern is used.

Calculation of $(\partial c/\partial t)_r$. For the purposes of this discussion, we assume that 10 scans have been acquired and will be converted to a single apparent sedimentation coefficient distribution, $g(s^*)$. If we designate the scans by C_1, C_2, C_3, etc., then the difference curves are computed as follows:

$$\left(\frac{\Delta C}{\Delta t}\right)_1 = \frac{C_1 - C_6}{t_6 - t_1}$$

$$\left(\frac{\Delta C}{\Delta t}\right)_2 = \frac{C_2 - C_7}{t_7 - t_2}$$

$$\left(\frac{\Delta C}{\Delta t}\right)_3 = \frac{C_3 - C_8}{t_8 - t_3}$$

$$\left(\frac{\Delta C}{\Delta t}\right)_4 = \frac{C_4 - C_9}{t_9 - t_4}$$

$$\left(\frac{\Delta C}{\Delta t}\right)_5 = \frac{C_5 - C_{10}}{t_{10} - t_5}$$

The x axis of each curve is converted to s^* using Eq. (1). The later curves are subtracted from the earlier ones to change the sign of $\Delta c/\Delta t$ so that the curves are positive for graphical purposes. Except near the base of the cell, beyond the so-called hinge point, $(\partial c/\partial t)_r$ is negative throughout the cell.

Averaging $\partial c/\partial t$. The five curves for $\Delta c/\Delta t$ as a function of s^* are then averaged at constant values of s^* to produce the averaged value, designated as $\langle \Delta c/\Delta t \rangle$. The averaged value is inserted into Eq. (2) for computation of the apparent distribution function.

Computation of Error Estimates. The standard deviation, σ_g, for each point on the averaged $g(s^*)$ curves is computed from the standard deviation of the average of $\Delta c/\Delta t$ using the following standard relation for propagation of errors:

$$\sigma_g^2 = \left[\frac{\partial g(s^*)}{\partial(\partial c/\partial t)}\right]^2 [\sigma^2_{(\partial c/\partial t)}]$$

where $\sigma_{(\partial c/\partial t)}$ is the standard deviation of $\Delta c/\Delta t$ and $\partial g/\partial c$ is obtained by differentiating Eq. (2) and is just

$$\frac{\partial g(s^*)}{\partial(\partial c/\partial t)} = \left(\frac{1}{c_0}\right)\left[\frac{\omega^2 t^2}{\ln(r_m/r)}\right]\left(\frac{r}{r_m}\right)^2 = \frac{t}{c_0 s^*}\exp(2\omega^2 s^* t)$$

The standard error of estimate for each averaged point on the curve is calculated by dividing the standard deviation by $n^{1/2}$, where n is the number of curves used in the average. It is worth pointing out that although the errors in $\delta c/\partial t$ are usually uniformly distributed across the cell, those in $g^*(s)$ are not because of their dependence on $1/s^*$, so that both the noise and the error bars tend to become larger at smaller values of s^*.

Correcting for Contribution to $\partial c/\partial t$ in Plateau Region

Equation (2) gives the normalized differential sedimentation coefficient distribution corrected for radial dilution. This equation was predicated on the assumption that the diffusion coefficient of each species was zero and, therefore, that each boundary contributing to the curve could be represented by a step function. Under this condition, the contribution from radial dilution to the total value of $\partial c/\partial t$ at the boundary position of any given species is affected only by the more centripetal components (i.e., those having smaller values of s). The value of $(\partial c/\partial t)_p$ at any given value of s is proportional to the product of s_w and c_p at s. Both s_w and c_p can be computed from $\hat{g}(s)$, where $\hat{g}(s)$ is the unnormalized distribution function as defined above. So, now, the plateau value of $\partial c/\partial t$ is given by

$$\left(\frac{\partial c}{\partial t}\right)_p = -2\omega^2 s_w c_p = -2\omega^2 \int s\hat{g}(s)\, ds$$

and the unnormalized distribution function is

$$\hat{g}(s) = \left(\frac{\partial c}{\partial t}\right)_{corr}\left[\frac{\omega^2 t^2}{\ln(r_m/r)}\right]$$

where

$$\left(\frac{\partial c}{\partial t}\right)_{\text{corr}} = \left(\frac{\partial c}{\partial t}\right)_{\text{obs}} - \left(\frac{\partial c}{\partial t}\right)_{\text{p}}$$

and "corr" indicates the corrected value and "obs" the observed value of $\partial c/\partial t$.

These three equations imply the following iterative procedure to correct $\partial c/\partial t$. First, compute an approximate value of $\hat{g}(s)$ using the observed value of $\partial c/\partial t$:

$$\hat{g}(s)_{\text{approx}} = \left(\frac{\partial c}{\partial t}\right)_{\text{obs}} \left[\frac{\omega^2 t^2}{\ln(r_{\text{m}}/r)}\right]$$

Then use this value to compute an approximation to the plateau value of $\partial c/\partial t$ at each point:

$$(\partial c/\partial t)_{\text{p,approx}} = -2\omega^2 s_w c_p = -2\omega^2 \int s\hat{g}(s)_{\text{approx}}\, ds$$

Subtract this from $(\partial c/\partial t)_{\text{obs}}$ to get an approximate value of $(\partial c/\partial t)_{\text{corr,approx}}$. Substitute $(\partial c/\partial t)_{\text{corr,approx}}$ into the above equation to compute a new value for $\hat{g}(s)_{\text{approx}}$. Repeat the cycle until the desired degree of convergence is attained. In practice, three iterations will give satisfactory convergence. The integration is carried out simply using the trapezoidal rule as follows:

$$\text{Area} = \left(\frac{y_1 + y_n}{2} + \sum_{i=2}^{i=n-1} y_i\right) \Delta s^*$$

where n is the number of points, Δs^* is the spacing in s^*, and, in this case, $y_i = s_i \hat{g}(s_i)$. In the algorithm used in this laboratory, the $g(s^*)$ curves are interpolated onto an equally spaced grid of s^*, mainly for subsequent extrapolation to correct for diffusion, so that Δs^* is a constant and, therefore, can be taken outside of the summation.

Weight Average Sedimentation Coefficient from $g(s^*)$

The weight average sedimentation coefficient can be estimated from $g(s^*)$, even in the case of significant diffusion, according to the following relationship with quite good accuracy:

$$s_w = \frac{\int_{s^*=0}^{s^*=s^*_{\text{p}}} \hat{g}(s^*) s^*\, ds^*}{\int_{s^*=0}^{s^*=s^*_{\text{p}}} \hat{g}(s^*)\, ds^*}$$

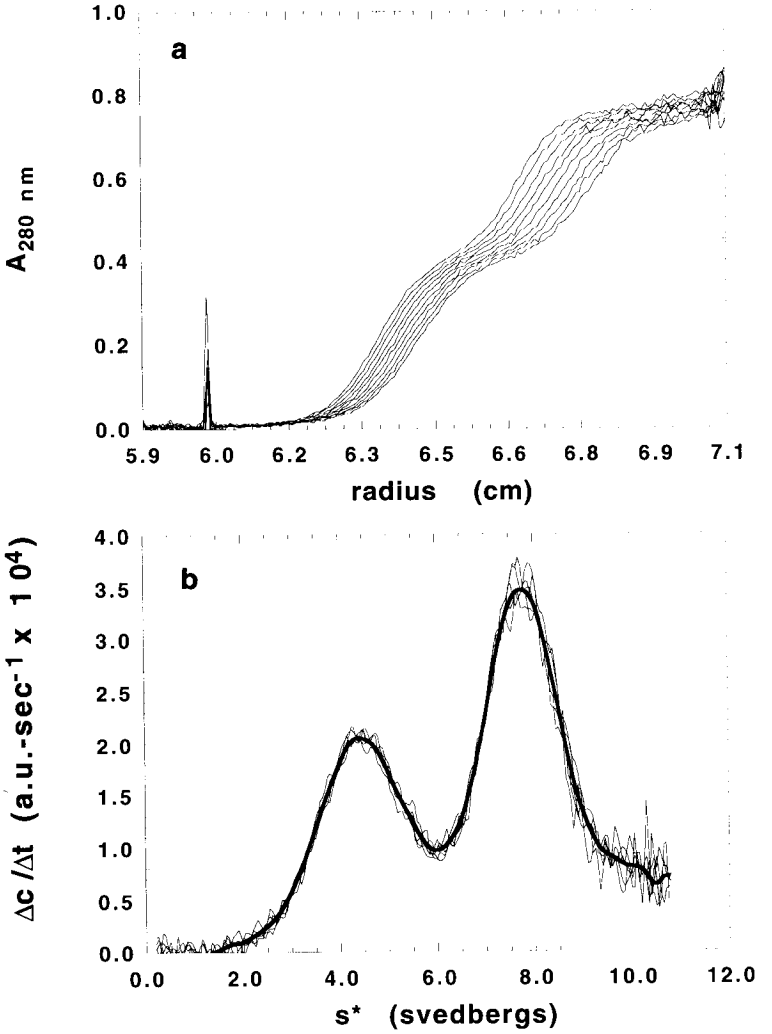

FIG. 4. (a) Data obtained from a sedimentation velocity experiment on the Beckman Instruments Optima XL-A analytical ultracentrifuge. Bovine serum albumin and aldolase (from Sigma, used without further purification) were dialyzed against 0.1 *M* NaCl, 10 m*M* phosphate, 1 m*M* dithiothreitol, pH 7.0, at 2° overnight. Ten plots of absorbance at 280 nm as a function of radius obtained at approximately 1-min intervals are given. The raw data were interpolated onto an equally spaced radial grid of 0.002 cm. Temperature of the run was 20°. (b) Plots of $\Delta c/\Delta t$ versus s^* obtained after subtracting pairs of concentration profiles as described in the text. The heavy line shows the average $\Delta c/\Delta t$ plot using a 2% window as outlined in the text. (c) Smoothed and unsmoothed $g(s^*)$ versus plots. Smoothing was performed on the averaged $\Delta c/\Delta t$ data with three passes of a simple moving average using a window spanning 2% of the distance from meniscus to base. (d) Averaged $g(s^*)$ plot for the system. The error bars are the standard error of the mean at each value of $g(s^*)$ propagated from the averaging of $\Delta c/\Delta t$.

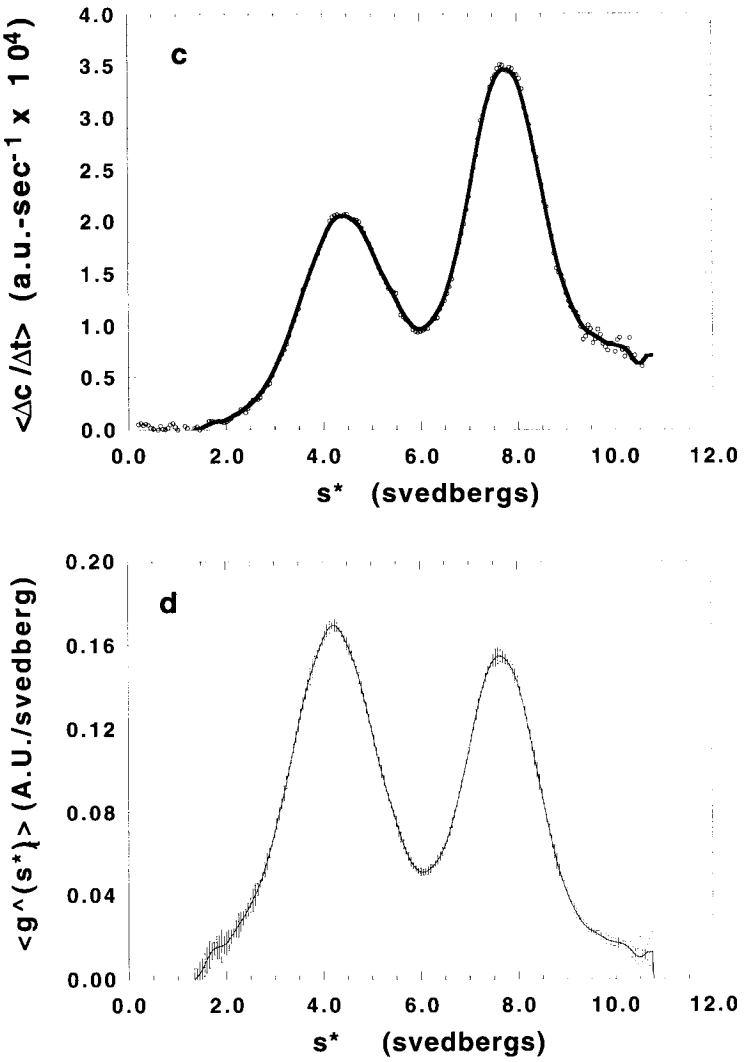

FIG. 4. (*continued*)

This method will give s_w to a very good approximation because the $\hat{g}(s^*)$ curves for each component are symmetrical and very nearly Gaussian on the s^* scale. A detailed description of the use of $\hat{g}(s^*)$ for the analysis of self-associating and heteroassociating systems will be presented elsewhere.

Experimental Example

As an experimental example of the method, a mixture of bovine serum albumin and aldolase (Sigma, St. Louis, MO) was run at 60,000 rpm in 0.1 *M* NaCl, 10 m*M* phosphate, 1 m*M* dithiothreitol at pH 7.0 and 20° in the Beckman Instruments Optimal XL-A analytical ultracentrifuge. Figure 4a shows 10 profiles of optical density ($A_{280\ nm}$) as a function of radius acquired at approximately 1-min intervals. The raw data were interpolated onto an equally spaced radial grid of 0.002 cm, and pairs of profiles were subtracted at equal radial positions as described above to compute $\Delta c/\Delta t$. Figure 4b shows the resulting $\Delta c/\Delta t$ versus s^* plots. The heavy line shows the average $\Delta c/\Delta t$ plot. Figure 4c shows smoothed and unsmoothed $g(s^*)$ versus s^* plots. Smoothing was performed on the averaged $\Delta c/\Delta t$ data with three passes of a simple moving average using a window spanning 2% of the whole data array from meniscus to base as outlined above. This degree of smoothing has produced a visually more pleasing result without introducing significant dispersion. Figure 4d shows the final plot of $g(s^*)$. The error bars are the standard error of the mean propagated from the averaged values of $\Delta c/\Delta t$. It is important to note that the error bars were computed from the original unsmoothed data and reflect the noise in the original curves.

Methods of Correcting Distribution Functions for Effects of Diffusion

The rate of boundary spreading owing to heterogeneity is roughly proportional to the first power of time, whereas that owing to diffusion is roughly proportional to the square root of time. Therefore, if the distribution functions are extrapolated to infinite time, the spreading from diffusion will become negligible compared to that from heterogeneity, and the true distribution function will be recovered. Various forms for extrapolation have been proposed in the literature and are discussed briefly below.

Differential Distribution Functions

Extrapolation of differential distribution functions to infinite time to remove the effects of diffusion was first described by Baldwin and Williams,[20,21] who extrapolated the functions at constant values of s^*. The functions of $g^*(s)$ were extrapolated versus $1/t$ to $1/t = 0$. Recently, several new forms that improve the extrapolation have been presented.[4] It

[20] J. W. Williams, R. L. Baldwin, W. Saunders, and P. G. Squire, *J. Am. Chem. Soc.* **74,** 1542 (1952).

[21] R. L. Baldwin and J. W. Williams, *J. Am. Chem. Soc.* **72,** 4325 (1950).

was sometimes observed that extrapolations versus $1/t$ resulted in negative values of the distribution function in regions where one expected them to go to zero. If the curves were extrapolated as $\ln[g^*(s)]$ versus a quadratic in $1/t^{1/2}$, better results could be obtained since the extrapolation is constrained to the positive domain of $g(s)$ by taking the logarithm.[4]

Integral Distribution Functions

A method for computation of integral sedimentation coefficient distribution functions was first introduced by Gralén and Lagermalm in 1952.[14,22] In that procedure, an apparent sedimentation coefficient was computed at each of several levels of the boundary. The apparent sedimentation coefficient is given by $s^* = \ln(r/r_m)/\omega^2 t$. The values of s^* computed at corresponding levels of successive boundary profiles are then extrapolated to infinite time to remove the contribution from diffusion. In the original method the values of s^* were extrapolated versus $1/t$ to $1/t = 0$. The method has been improved by Van Holde and Weischet,[18] who noted that a better approximation to the theoretically correct extrapolation could be attained if the values of s^* were extrapolated versus $1/t^{1/2}$.

Discussion

With the introduction of the Beckman Instruments Optima XL-A analytical ultracentrifuge, the processes of data acquisition in sedimentation experiments have now become largely automated. Analysis of these data, from both sedimentation equilibrium and velocity experiments, can be done entirely by digital computer. The speed and efficiency of analysis possible with a computer have allowed not only easier analysis by the older, more common methods but also the implementation of methods, both new and old, that would have been previously impractical.

In this chapter, I have briefly reviewed the older methods and introduced two new methods based on the time derivative of the concentration distribution. The ability to compute the time derivative, especially from data obtained with real-time Rayleigh optical systems,[8,9] has allowed a significant increase in the sensitivity of sedimentation velocity experiments. The process of computing the time derivative results in an automatic baseline correction. This baseline correction combined with averaging of the data can result in an increase in precision of 1–2 orders of magnitude with the UV photoelectric scanning system and 2–3 orders of magnitude with Rayleigh optics. The new methods are well suited to real-time sedimentation analysis.

[22] N. Gralén and G. Lagermalm, *J. Phys. Chem.* **56,** 514 (1952).