HHS Public Access

Author manuscript

Eur Biophys J. Author manuscript; available in PMC 2019 October 01.

Published in final edited form as:

Eur Biophys J. 2018 October ; 47(7): 837–844. doi:10.1007/s00249-018-1309-z.

Two-dimensional Grid Optimization for Sedimentation Velocity Analysis in the Analytical Ultracentrifuge

Haram Kim[1], **Emre Brookes**[2], **Weiming Cao**[3], and **Borries Demeler**[2]

[1]Cornell University, Department of Computer Science, 402 Gates Hall, Ithaca, NY 14853

[2]The University of Texas Health Science Center at San Antonio, Department of Biochemistry and Structural Biology, 7703 Floyd Curl Drive, San Antonio, Texas 78229-3901

[3]The University of Texas at San Antonio, Department of Mathematics, One UTSA Circle San Antonio, TX 78249

Abstract

Sedimentation velocity experiments performed in the analytical ultracentrifuge are modeled using finite element solutions of the Lamm equation. During modeling, three fundamental parameters are optimized: the sedimentation coefficients, the diffusion coefficients, and the partial concentrations of all solutes present in a mixture. A general modeling approach consists of fitting the partial concentrations of solutes defined in a two-dimensional grid of sedimentation and diffusion coefficient combinations that cover the range of possible solutes for a given mixture. An increasing number of grid points increases the resolution of the model produced by the subsequent analysis, with denser grids giving rise to a very large system of equations. Here we evaluate the efficiency and resolution of several regular grids and show that traditionally defined grids tend to provide inadequate coverage in one region of the grid, while at the same time being computationally wasteful in other sections of the grid. We describe a rapid and systematic approach for generating efficient two-dimensional analysis grids that balance optimal information content and model resolution for a given signal-to-noise ratio with improved calculation efficiency. These findings are general and apply to one- and two-dimensional grids, although they no longer represent regular grids. We provide a recipe for an improved grid point spacing in both directions which eliminates unnecessary points, while at the same time providing a more uniform resolution that can be scaled based on the stochastic noise in the experimental data.

Introduction

Sedimentation velocity (SV) experiments performed in an analytical ultracentrifuge provide information about composition, size and anisotropy, and for some experimental designs information about density of colloidal molecules in solutions. They measure the sedimentation and diffusion transport of a colloidal particle in a centrifugal force field, and provide the partial concentration of each solute in a mixture. The observed signal is typically convoluted with systematic and stochastic noise. Where possible, systematic noise contributions can be removed mathematically [1], leaving only the stochastic noise to the residuals of a fit. We have developed a number of optimization routines to solve the problem of fitting experimental data in an unbiased approach, and to extract the sedimentation and diffusion coefficients and partial concentrations of mixtures of analytes [2, 3, 4, 5, 6, 7]. For

all of these methods, the ability to recover these parameters is limited by the magnitude of the stochastic noise present in the data. The magnitude of the noise determines the minimum amount of signal that the fitting method needs to be able to resolve. Any signal larger than the noise is not lost in the noise and the grid must therefore be able to resolve differences between grid points that are equal or a slightly smaller than the noise signal. In other words, the underlying model must be able to explain the sedimentation and diffusion transport present in the experimental data with slightly higher resolution than the resolution necessary to account for the magnitude of the stochastic noise. This transport, when performed in a sector-shaped centrifugation cell under ideal solution conditions (constant temperature, absence of pressure dependence, constant speed, and under dilute conditions), is described by the Lamm equation *L* (Equ. 1) [8]:

$$\frac{\partial C}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}[rD\frac{\partial C}{\partial r} - s\omega^2 r^2 C] \quad \text{Equ. 1}$$

where *r* is the radial distance from the rotor center, *s* and *D* are the sedimentation and diffusion coefficients, *C* is the partial concentration of a solute, and ω is the angular velocity of the rotor. Inspection of Equ. 1 reveals that fitting an experimental dataset consists of adjusting the sedimentation and diffusion coefficient, and finding the appropriate concentration *C*. In the general case, one must allow for the presence of multiple solutes C_i, where *i* indicates the i^{th} species in a mixture. For non-interacting mixtures of solutes, the general solution for a multi-component mixture with *n* unknown species is given by:

$$C_{total} = \sum_{i=1}^{n} c_i L_i(s, D) \quad \text{Equ. 2}$$

where c_i is the partial concentration of the i^{th} solute. In the general case, *n*, c_i, and L_i are not known and need to be determined with a degenerate fitting approach that does not impose any user bias or prior knowledge upon the solution. Furthermore, a rigorous solution to this problem requires that *s* and *D* for each solute are allowed to vary independently, requiring a two-dimensional fitting approach that can account for variable distributions in both sedimentation and diffusion coefficient. Previously we proposed a two-dimensional spectrum analysis (2DSA) approach to solve this problem [2, 3]. 2DSA begins by building a regular two-dimensional grid of sedimentation coefficients in one dimension and frictional ratios in the second dimension. This results in a two-dimensional grid of unique solutes, where each solute is defined by a unique combination of sedimentation and diffusion coefficients. Next, the finite element solution for the entire experiment is calculated and a full set of scans and radial absorbances is simulated for each individual solute, using the experimental and boundary conditions of the actual experiment (rotor speed, buffer conditions, meniscus position, and bottom of cell). The simulated datapoints for each unique solute represent a basis vector of a linear combination of all solutes represented by the two-dimensional grid. The optimization problem is solved by forming a linear combination (Equ. 2) of the basis vectors $\boldsymbol{l_i = L_i(s, D)}$, representing simulated solutions for each *s*, *D* for all *n*

solutes defined in the grid. This linear system can be written as **Ax=b** where **A** is the matrix of basis vectors $\boldsymbol{l_i}$, **x** is the vector of unknown concentrations c_i, and **b** is vector with experimental data. This problem is solved with the non-negatively constrained least squares algorithm (NNLS) [9], which results in a vector **x** containing positive concentrations c_i for solutes C_i contributing to the NNLS fit and zero for all other solutes not found in the experimental data. Clearly, the model resolution obtained from the fit will be proportional to the number of solutes included in matrix **A**, with the resolution increasing with the size of **A**. In any case, for a typical experiment **A** will be very large (on the order of several gigabytes). As the size of **A** increases, so does the computational effort and the required calculation time. The exact scaling of the computational effort with resolution is difficult to generalize, since it depends on the number of components present in the experimental data, the size of **A**, the number of parallel processors available, and the number of partitions employed in the 2DSA. Therefore, a compromise has to be made between the desired resolution and the available computational resources. An obvious question therefore is: What exactly is the best set of grid points to use in a two-dimensional grid to obtain a desired model resolution for a given problem? A good rule of thumb is to use a grid layout where elimination of any grid point in the two-dimensional grid would introduce an error slightly less in magnitude than the stochastic noise inherent in the experimental data. If the grid point density is high enough to where the removal of a grid point does not affect the root mean square deviation (RMSD) of the fit within the magnitude of the noise, then any solute present in the experimental data can be distinguished reliably, and the chance for missing a solute is minimized.

According to Equ. 1, each solute measured in an analytical ultracentrifugation (AUC) experiment gives rise to a sedimentation and diffusion coefficient, and NNLS optimization recovers the partial concentration of each solute in the grid of solutes (which may be zero). Once sedimentation and diffusion coefficients with non-zero concentrations are determined, additional properties of the found solutes are available. From the diffusion coefficient, we can derive the frictional coefficient:

$$f = \frac{RT}{ND} \quad \text{Equ. 3}$$

where R is the gas constant, T is the temperature in Kelvin, and N is Avogadro's number. If the partial specific volume($\bar{\nu}$) is available, we can derive the molar mass, M:

$$M = \frac{sNf}{1 - \bar{\nu}\rho} \quad \text{Equ. 4}$$

where ρ is the density of the solvent. Once molar mass and partial specific volume are available, we can assume a hypothetical spherical particle with the same volume as the actual solute and calculate the volume V and hydrodynamic radius r_0 of the spherical particle:

$$V = \frac{M\bar{\nu}}{N}, r_0 = \left(\frac{3V}{4\pi}\right)^{1/3} \quad \text{Equ. 5}$$

Using the Stokes-Einstein relationship, we can derive the frictional coefficient of this hypothetical sphere:

$$f_0 = 6\pi\eta r_0 \quad \text{Equ. 6}$$

Finally, the frictional ratio, or anisotropy k can be derived:

$$k = \frac{f}{f_0}$$

The latter property describes how non-globular a molecule is. For a perfectly spherical molecule, $f = f_0$, and $k = 1.0$, for all other molecules $k > 1.0$.

To aid in the interpretation of AUC results, it is frequently more convenient to express the results by using parameterizations of the sedimentation and diffusion coefficients, and to present the results in terms of more intuitive parameters, for example, as functions of molar mass and anisotropy or partial specific volume and molar mass instead of sedimentation and diffusion coefficients. As was shown in [6], it is straightforward to express a range of solute properties of interest in terms of any combination of another type of grid. In this work we evaluate the resolution, and contrast the computational requirements of several regular grid layouts, and show that all of these regular grids are either computationally wasteful or lack the ability to describe an experimental system with the desired resolution for each region of the grid equally. With the recent introduction of the Beckman Optima AUC instrument, a significant enhancement of the data quality and signal to noise ratio is realized, which suggests that commensurate enhancements in the data analysis resolution are desirable. This raises the question of the exact distribution of solutes in Equ. 2 that will provide the optimal compromise between resolution and computational requirements. In this manuscript we present a systematic evaluation of the performance of traditionally employed regular grids, and propose an adaptive grid layout providing improved solute point positions for s and D, which are easily computed, and which still can be converted to any custom grid application proposed in [6]. The new grid optimizes the retrieval of available information while at the same time minimizing the computational effort as a function of resolution, performing significantly better than any other regular grid layout tested by us.

Methods

1. Testing grid performance and simulation

We define grid performance as the reciprocal product of the computational effort times the number of grid points required to obtain a constant grid resolution. In order to compare grid performance, a resolution metric needs to be established. A convenient resolution metric is

the signal difference between the experimental data from two simulated solutes with equal loading concentration [10]. Here, the simulations are for two adjacent grid points, and simulated to match the experimental run conditions. An optimal grid layout will feature a resolution and grid spacing such that the difference between Lamm equation solutions from adjacent grid points equals tolerance t, which should be slightly less than the RMSD originating from stochastic noise in the data. We suggest a constant value e which should be half of the expected RMSD. This difference needs to be satisfied in both dimensions of the grid:

$$L(s_{i+1}, D_i) - L(s_i, D_i) = t \quad \text{Equ. 7}$$
$$L(s_i, D_{i+1}) - L(s_i, D_i) = t$$
$$with\, t = RMSD - e$$

It is important to point out that contributions to the signal difference between two points in the grid depend on several experimental conditions, including rotor speed, the radial range of the fitted data, the interval between scans, the partial concentration of a solute, and the duration of the experiment, and should be derived from the UltraScan edit profile, which sets data range limits. We investigated regular grid types parameterized by k vs. s, k vs. M, D vs. s, and a new improved k vs. s grid with point spacings based on the first derivative of the Lamm equation with respect to s and D. In each case, we attempted to cover the same domain in s and D, regardless of parameterization. For all grids, the number of total grid points, N_{grid}, was kept constant at 210 points in order to approximate equal computational effort across all grids. The total number of grid points was chosen such that the grid coverage was visually comparable across all grids. Generating grids should be fast and efficient, so numerical routines that empirically identify grid spacings satisfying a given resolution, for example through a line search or root finding algorithm, are not desirable due to their large computational overhead (data not shown). In contrast, our proposed improved grid can be generated quickly and is suitable for parallel methods implemented on supercomputers [11]. To compare the efficiency of all grid types an empirical method using finite element simulations was needed. For this purpose, a new UltraScan module was developed, reusing already available data structures and processing methods in the UltraScan C++ class library. RMSD values were determined by subtracting two finite element solutions representing adjacent grid points from each other as follows: Scans for each component were simulated with equal time increments, and over the experimental duration. Only points having less than one-half of the plateau concentration were included in the calculation, and any scans where the midpoint of the boundary was to the right of the bottom b of the cell with meniscus m according to $b = me^{s\omega^2 t}$ were excluded from the RMSD calculations. Any points located to the left of a point 0.025 cm to the right of the meniscus were also excluded from the RMSD computation. This approach assured that steep gradients in the back-diffusion region are excluded from the fitting range due to refractive artifacts in this region, and because absorbance values at the bottom of the cell tend to exceed the dynamic range of the detector. The simulated time of the experiment was 6.8 hours, and was chosen such that the midpoint of the boundary from the average sedimentation coefficient of the grid's s-value range would cross the bottom of the cell according to Equ. 14. For each experiment,

100 equidistant scans in time were simulated. All finite element simulations were performed for a 40,000 rpm rotor speed, using 200 simulation points for ASTFEM grid. For sedimentation coefficients larger than the mean sedimentation coefficient, sedimentation time was shortened such that the RMSD calculation ignored scans after the faster of the two components pelleted. This prevented calculated RMSD values from being underestimated due to the inclusion of baseline values from pelleted solute states. For all grids, we used a sedimentation coefficient range from $s_1 = 1.1 \times 10^{-13}$ s to $s_2 = 9.9 \times 10^{-13}$ s, and a frictional ratio range from $k_1 = 1.2$ to $k_2 = 3.8$. These ranges were chosen to prevent the program from simulating unreasonable frictional coefficients below 1.0 during RMSD isobar calculations. To measure grid resolution, RMSD isobars were calculated around each grid point by measuring the RMSD difference along the polar coordinate lines from zero to 2π in two degree increments with each grid point at its center, producing 180 RMSD points around each grid point. The required length of the polar coordinate line was determined by testing the RMSD at points from the 4 corners of the grid. We found that constant multipliers proportional to the regular $s \cdot k$ grid spacing were more than sufficient to capture all RMSD isobars of interest. Furthermore, the chosen constant scaling also ensured that all RMSD values along the polar coordinate line allowed for a linear extrapolation (data not shown). This approach was repeated by iterating over all other solute points in a test grid projected on to the $s \cdot k$ plane. Linear interpolations between 0 and 0.5% RMSD were used to generate RMSD error ellipsoid isobars (see Figure 1). Equations for the linear approximations needed for the generation of the ellipsoid isobars were then stored in an output file to allow ellipsoids for different RMSD values to be generated without simulating every grid point and its associated sample points again.

2. Improved grid generation

Our improved grid is based on the rate of change of the concentration as a function of s and D. This can be represented by the derivative of the Lamm equation with respect to s and D. Since an analytical solution to this problem is not readily available, and numerical solutions require computationally demanding algorithms, we chose to use the Faxén approximation to the Lamm equation [12]. Starting with the Lamm equation (Equ. 1), we first introduce dimensionless variables x, τ and ε:

$$x = (\frac{r}{m})^{2,} \tau = 2s\omega^2 t, \varepsilon = \frac{2D}{s\omega^2 m^2} \quad \text{Equ. 8}$$

where m and b are the meniscus and the bottom of the cell, respectively. Then the Lamm equation can be transformed to:

$$\frac{\partial C}{\partial \tau} = \frac{\partial}{\partial x}[x(\varepsilon \frac{\partial C}{\partial x} - C)] \quad \text{Equ. 9}$$

It is evident that the solution C depends on parameter ε only. When $\varepsilon \ll 1$, and x near 1, the solution to the Lamm equation can be approximated by the Faxén solution:

$$C(x,\tau) = \frac{1}{2}e^{-\tau}[1 - \Phi(\upsilon)] \quad \text{Equ. 10}$$

where

$$\Phi(\upsilon) = \frac{2}{\sqrt{\pi}}\int_0^{\upsilon} e^{-x^2}dx \quad \text{Equ. 11}$$

is the error function, and

$$v = \frac{1 - (xe^{-\tau})^{1/2}}{[\varepsilon(1 - e^{-\tau})]^{1/2}} \quad \text{Equ. 12}$$

Taking the partial derivative with respect to ε yields:

$$\frac{\partial C}{\partial \varepsilon}(x,\tau) = \frac{1}{2\sqrt{\pi}}e^{-\tau}\varepsilon^{-1}\upsilon e^{-\upsilon^2} \quad \text{Equ. 13}$$

To evaluate the magnitude of $\frac{\partial C}{\partial \varepsilon}$ we need to specify a meaningful time interval. We note that a typical experiment will be finished when the midpoint of the solute's boundary reaches the bottom of the cell, which occurs at:

$$t_* = ln(\frac{b}{m})\frac{1}{s\omega^2} \quad \text{Equ. 14}$$

Therefore, we use the time interval [0, t∗] to evaluate the magnitude of $\frac{\partial C}{\partial \varepsilon}$. More precisely, we introduce the norm of $\frac{\partial C}{\partial \varepsilon}$ in the domain $0 \le \tau \le \tau_* = 2s\omega^2 t_* = 2\ln(b/m)$ and $1 \le x \le x_* = (b/m)^2$ as follows:

$$\|\frac{\partial C}{\partial \varepsilon}\| = [\int_0^{\tau *}\int_1^{x *} |\frac{\partial C}{\partial \varepsilon}(x,\tau)|^2 dxd\tau]^{1/2} \quad \text{Equ. 15}$$

For fixed values of m, b and rotor speed ω this norm is dependent on ε only. Unfortunately, there is no explicit formula for the norm as a function of ε. A numerical evaluation of the

norm suggests that for typical ranges of s, D and ω, $\frac{\partial C}{\partial \varepsilon}$ is approximately proportional to $\varepsilon^{-3/4}$. See Figure 2 for a log-log plot of $\| \frac{\partial C}{\partial \varepsilon} \|$ as a function of ε in the case of $m = 6.5$, $b = 7.2$, and $\omega = 60{,}000$ rpm.

A careful study shows that ε is inversely proportional to the 3/2-th power of the product $s \cdot k$, more precisely,

$$\varepsilon = \frac{2}{9\sqrt{2}\pi\omega^2 b^2} \cdot \frac{RT}{N} \cdot \left(\frac{\eta^3 v}{1 - v\rho}\right)^{-1/2} \cdot (s \cdot k)^{-3/2} \quad \text{Equ. 16}$$

Let $\mu = (s \cdot k)^{-1}$, then we have $\varepsilon = O(\mu^{3/2})$, and thus:

$$\| \frac{\partial C}{\partial \varepsilon} \| \approx O(\varepsilon^{-3/4}) = O(\mu^{9/8}) \quad \text{Equ. 17}$$

Using the chain rule for differentiation

$$\frac{\partial C}{\partial \mu} = \frac{\partial C}{\partial \varepsilon} \cdot \frac{\partial \varepsilon}{\partial \mu} = \frac{\partial C}{\partial \varepsilon} \cdot O(\mu^{-1/2}) \quad \text{Equ. 18}$$

which implies that:

$$\| \frac{\partial C}{\partial \mu} \| \approx O(\mu^{-5/8}) \quad \text{Equ. 19}$$

Since $\| \frac{\partial C}{\partial \mu} \|$ is approximately constant along a curve $s \cdot k = const$, when designing an s-k grid system, the grid points can be picked along various curves $s \cdot k = \mu_j^{-1}$, j=1,2,…,N where the values of μ_j, j=1,2,…,N, are selected so that the RMSD error isobars are approximately uniformly distributed. We observed that when $\mu_j^{-1/4}$ is evenly spaced, the distribution of the RMSD error isobars is the closest to uniformity. Thus we select μ_j values accordingly for the grid generation. A detailed description of the creation of the s-k grid system follows:

Suppose in a 2DSA analysis the sedimentation coefficient s is between limits s_1, s_2 and the frictional ratio k is between k_1 and k_2, then the range for $\mu = (s \cdot k)^{-1}$ is between and $\mu_1 = (s_1 \cdot k_1)^{-1}$ and $\mu_2 = (s_2 \cdot k_2)^{-1}$. Let N be the number of partitions we would like to place in between μ_1 and μ_2. Then an equidistribution of $\mu^{-1/4}$ can be achieved approximately by using the dividing points:

$$\mu_j = [(1 - \frac{j}{N}) \cdot \mu_1^{-1/4} + \frac{j}{N} \cdot \mu_2^{-1/4}]^{-4}, 0 \le j \le N \quad \text{Equ. 20}$$

To generate the improved grid, we calculate all $y_j = 1/(\mu_j \cdot s_1)$ where $\mu_j^{-1} \ge s_1 \cdot k_2$ and all $x_{i,j} = 1/(\mu_i \cdot y_j)$, satisfying $s_1 \le x_{i,j} \le s_2$. Then the grids on the s-k plane is the collection of all points $(x_{i,j}, y_j)$, satisfying $s_1 \le x_{i,j} \le s_2$ and $\mu_j \le s_1 \cdot k_2$.

3. Adjusting the resolution of the improved grid

The resolution of the improved grid is proportional to the total number of grid points, N_{grid}. It can be controlled by adjusting N, the number of partitions between μ_1 and μ_2. An estimate of N_{grid} can be obtained as follows: First, ensuring $\mu_j \le 1/(s_1 \cdot k_2)$, we have $0 \le j \le J_a$ with:

$$J_a = N \cdot [1 - (\frac{\mu_1}{s_1 \cdot k_2})^{1/4}] \Big/ [1 - (\frac{\mu_1}{\mu_2})^{1/4}] \quad \text{Equ. 21}$$

For each $j \le J_a$, to ensure that $s_1 \le x_{i,j} \le s_2$, we have:

$$j \le i \le N \cdot [1 - (\frac{\mu_1}{s_2 \cdot y_j})^{1/4}] \Big/ [1 - (\frac{\mu_1}{\mu_2})^{1/4}] \quad \text{Equ. 22}$$

Consequently, the total number of gridpoints, N_{grid}, is given by:

$$N_{grid} = \sum_{j=1}^{J_a} ([1 - (\frac{\mu_1}{s_2 \cdot y_j})^{1/4}] \Big/ [1 - (\frac{\mu_1}{\mu_2})^{1/4}] - j) \quad \text{Equ. 23}$$

A plot of N_{grid} vs. N is displayed in Figure 3. Furthermore, a least squares fit shows that N_{grid} is approximately a quadratic function of the number of partitions as given by:

$$N_{grid} \approx N^2/e \quad \text{Equ. 24}$$

A comparison of the estimated total number of grid points using the above formula is also shown in Figure 3, which indicates a good match of Equ. 23 with the prediction by Equ. 24. Therefore, in practice, in order to generate an improved grid containing N_{grid} points we can select $N = \sqrt{e \cdot N_{grid}}$ as the number of partitions to produce the improved grid.

Results

One of the best metrics for grid performance is the RMSD distance between adjacent grid points. When the grid points are identical, the RMSD difference between these points is

zero, the further the two grid points move apart along either the *s* and *D* direction, the larger the RMSD difference will become between finite element solutions for these grid points. The comparison cannot just be made in one dimension, because both *s* and *D* contribute to this RMSD difference. As is shown in Figure 1, a constant level of RMSD difference around a grid point is best described by an ellipsoid, which varies in aspect ratio and orientation, depending on the position of the grid point in the two-dimensional grid space. Ideally, the RMSD difference between adjacent grid points should be slightly less than the RMSD level encountered in the stochastic noise in the data to assure all solute concentrations that exceed the noise level can be detected. To this end, we plotted the location and RMSD ellipsoids for five different RMSD levels (0.001 – 0.005), corresponding roughly to the noise level ranges observed in commercially available analytical ultracentrifuges, for a fixed number of grid points and several regular grid types, as well as for the improved grid based on the Faxén solution. Regular grid types offer the advantage of being intuitive in terms of the variable that they represent (for example, frictional ratio and molar mass) and can be quickly generated. They avoid the computational overhead of numerically optimized grids that will result in equi-distant RMSD grid points. Furthermore, such optimized grids are difficult to generate in more than one dimension. On the other hand, the computational overhead for the improved grid (Equ. 8 – Equ. 24) is trivial and well suited for methods like the 2DSA or genetic algorithms [4, 5], where hundreds of grids need to be computed. We investigated regular grids where *s* and *D* values were based on *s* vs. *k*, *M* vs. *k*, and *s* vs. *D* on a range consistent with a sedimentation coefficient range for *s* from 1–10 x 10^{-13} sec, and a frictional ratio range for *k* from 1–4. Our comparison of the performance of the improved grid with different regular grids revealed significant differences when error distances between adjacent grid points were evaluated. These differences are clearly seen when their RMSD error isobars are visually compared (Figure 4). We observe the following characteristics for each grid: the conventionally used *s* vs. *k* grid (Figure 4A) suffers from low resolution in *s* for the left half of the grid, but performs well for *s* in the right half of the grid. For *k*, the resolution increasingly suffers in the upper left quadrant of the grid, and is computationally very wasteful in the right half of the grid, where adjacent grid points overlap strongly in the *k* dimension. By far, the worst performing grid is a regular grid based on molar mass *M* and frictional ratio *k* (Figure 4B). Drastic loss of resolution in the lower left quadrant of the grid in both dimensions is accompanied by significant overlap in both dimensions in the entire right half of the grid, and especially strong in the center for *k*. A regular grid based on *s* vs. *D* (Figure 4C) performs reasonably well for s throughout the s-domain, with a slight loss in resolution in the upper left quadrant of the grid for *k*, similar to the error seen in the *s* vs. *k* grid (Figure 4A). In the right half of the grid significant overlaps are seen in the lower *k* regions of the grid, indicating significant computational inefficiencies. The most evenly distributed RMSD error over the entire grid is evident from the improved grid based on the Faxén solution (Figure 4D).

Remarkably, the improved grid provides excellent coverage for the lower left quadrant (see Figure 5 for a magnified view of the lower left quadrant for a 0.0005 (red) and 0.001 (blue) RMSD error level), demonstrating no overlap and nearly touching isobars. Similarly, overlaps in the right half of the grid are essentially absent, though spacing in the *s*-range suggests slight resolution loss in the upper right quadrant of the grid. It should be noted that

diffusion resolution is very low in the upper right quadrant since solutes in this portion of the grid have a small diffusion coefficient to begin with, and then they are sedimenting rapidly, leaving little time for diffusion, which decreases diffusion signal and explains lower resolution in *D*. Consequently, isobars are very elongated in the *k*-direction and white space at the upper right quadrant is caused by missing solute points that would be centered at frictional ratios > 4 which were not considered in this simulation. The same effect is very clear also in Figure 4C, where large regions were not simulated since they fell outside of the range of $1 \geq k \geq 4$, and solute regions outside of this range would be required to fill these white spaces. The current grid generation function for UltraScan's 2DSA produces a regularly spaced grid of solute points in terms of sedimentation coefficient and frictional ratio. Although this method can effectively analyze AUC experimental data, it does not necessarily do so in the most computationally efficient way. When using a regular $s \cdot k$ grid, there are often cases in which groups of two solute points on the grid are sufficiently similar in terms of their simulated behavior that when the stochastic noise of the experimental data is taken into account, the two are functionally identical. This is problematic because it requires the program to unnecessarily simulate a solute, a cost that can become significant for large grids with many redundant simulations.

Summary

We have presented a novel method for a computationally efficient $s \cdot k$ grid that substantially improves resolution for a given number of simulation points in a two-dimensional grid used for fitting sedimentation velocity experiments, while simultaneously minimizing computational effort and required memory. This innovation will reduce needed computer time on national supercomputing infrastructures like XSEDE and PRACE, and improve resolution when fitting sedimentation velocity data.

Acknowledgments

This work was supported by NIH grant GM120600 and NSF grant NSF-ACI-1339649 (to BD). Supercomputer calculations were performed on Comet at the San Diego Supercomputing Center (support through NSF/XSEDE grant TG-MCB070039N to BD) and on Lonestar-5 at the Texas Advanced Computing Center (supported through UT grant TG457201 to BD).

References

1. Demeler B. Methods for the Design and Analysis of Sedimentation Velocity and Sedimentation Equilibrium Experiments with Proteins. Cur Protoc Prot Sci. 2010; Chapter 7(Unit 7.13)
2. Brookes E, Boppana RV, Demeler B. Supercomputing '06 ACM. 2006. Computing Large Sparse Multivariate Optimization Problems with an Application in Biophysics. 0-7695-2700-0/06
3. Brookes E, Cao W, Demeler B. A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur Biophys J. 2010; 39(3):405–14. [PubMed: 19247646]
4. Brookes E, Demeler B. Genetic Algorithm Optimization for obtaining accurate Molecular Weight Distributions from Sedimentation Velocity Experiments. In: Wandrey C, Cölfen H, editorsAnalytical Ultracentrifugation VIII, Progr Colloid Polym Sci. Vol. 131. Springer; 2006. 78–82.
5. Brookes E, Demeler B. Parsimonious Regularization using Genetic Algorithms Applied to the Analysis of Analytical Ultracentrifugation Experiments. GECCO Proceedings ACM. 2007 978-1-59593-697-4/07/0007.

6. Demeler B, Nguyen TL, Gorbet GE, Schirf V, Brookes EH, Mulvaney P, El-Ballouli AO, Pan J, Bakr OM, Demeler AK, Hernandez Uribe BI, Bhattarai N, Whetten RL. Characterization of Size, Anisotropy, and Density Heterogeneity of Nanoparticles by Sedimentation Velocity. Anal Chem. 2014 Aug 5; 86(15):7688–95. [PubMed: 25010012]

7. Gorbet G, Devlin T, Hernandez Uribe B, Demeler AK, Lindsey Z, Ganji S, Breton S, Weise-Cross L, Lafer EM, Brookes EH, Demeler B. A parametrically constrained optimization method for fitting sedimentation velocity experiments. Biophys J. 2014; 106(8):1741–1750. DOI: 10.1016/j.bpj.2014.02.022 [PubMed: 24739173]

8. Lamm O. Die Differentialgleichung der Ultrazentrifugierung. Ark Mat Astr Fys. 1929; 21B:1–4.

9. Lawson CL, Hanson RJ. Solving Least Squares Problems. Prentice-Hall; Englewood Cliffs, New Jersey: 1974.

10. Brookes E, Demeler B. UltraScan: Improved 2DSA Resolution with Modified Parameter Space Grids. Analytical Ultracentrifugation Symposium; Nottingham. 2010.

11. Demeler B, Brookes E, Nagel-Steger L. Analysis of heterogeneity in molecular weight and shape by analytical ultracentrifugation using parallel distributed computing. Methods Enzymol. 2009; 454:87–113. [PubMed: 19216924]

12. Faxén H. Über eine Differentialgleichung aus der physikalischen Chemie. Ark Mat Astr Fys. 1929; 21B:1–6.

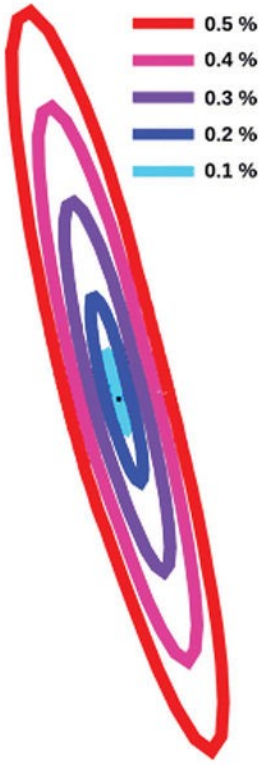

Figure 1.
RMSD error isobars for a solute point (black) in the 2DSA grid.

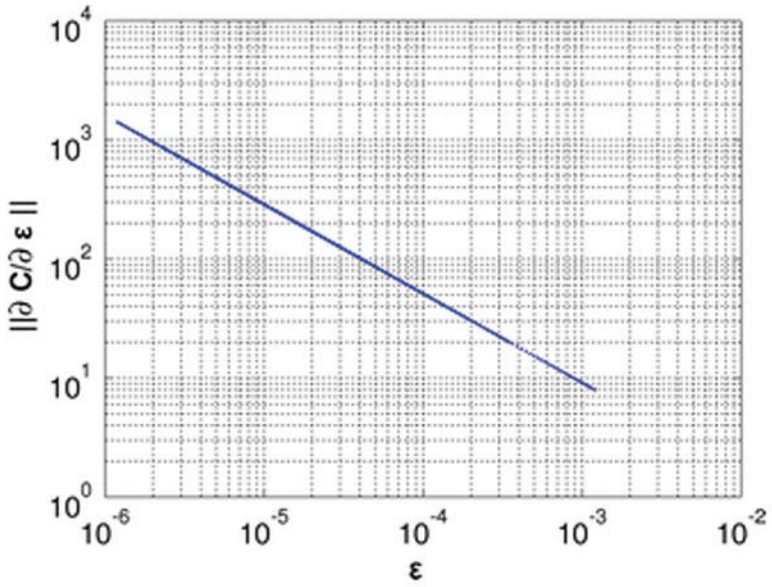

Figure 2.

A plot of the norm of $\frac{\partial C}{\partial \varepsilon}$ as a function of ε in the case of m = 6.5, b = 7.2, ω = 60000rpm.

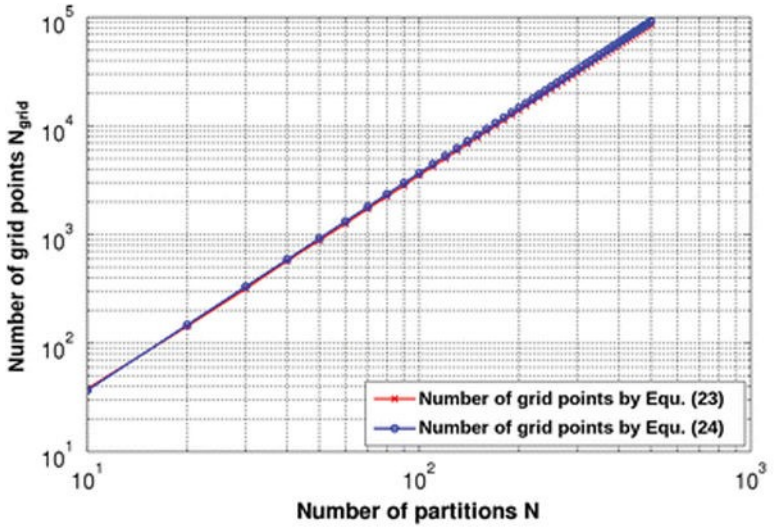

Figure 3.

Total number of grid points N_{grid} vs. the number of partitions *N* using Equ. 23 and approximate Equ. 24

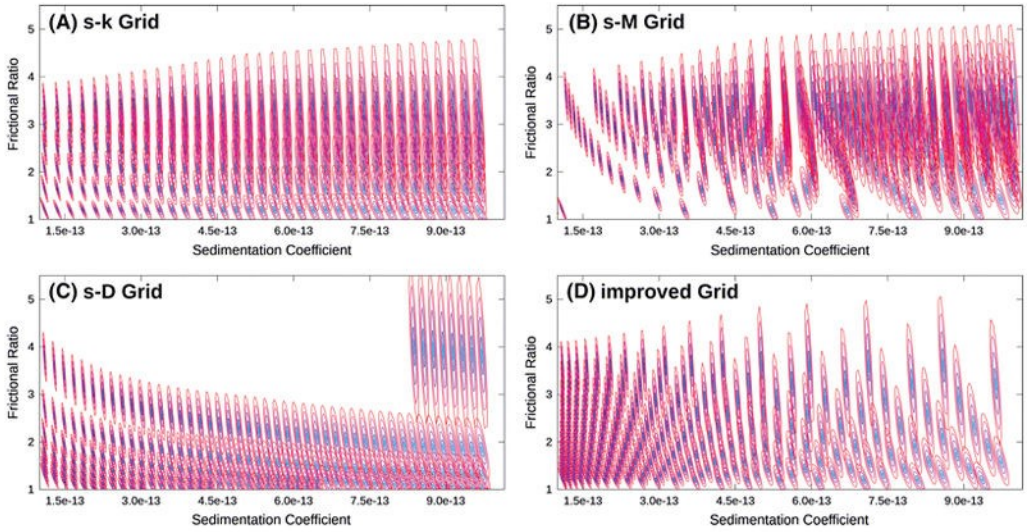

Figure 4.

RMSD error isobars for an equal number of grid points from three regular grids (A: s vs. k, B: M vs. k, C: s vs. D, and D: Improved grid based on the Faxen solution. Here, increasing white space between the outermost error isobar indicates reduced resolution, while overlaps between adjacent red isobars indicate wasteful inefficiencies. Ideally, red isobars should touch, but not overlap.

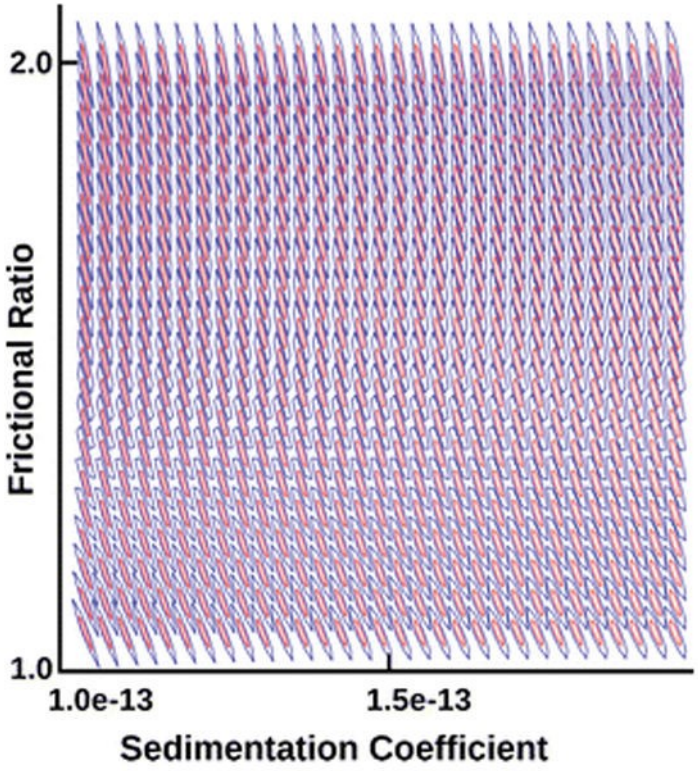

Figure 5.

Detail of lower left corner of improved grid based on the Faxen solution, demonstrating excellent coverage without overlaps and without resolution gaps (simulated resolution in blue: RMSD=0.001).