An Inverse Problem in Agglomeration Kinetics

RAMACHANDRAN MURALIDAR AND DORAISWAMI RAMKRISHNA

School of Chemical Engineering, Purdue University, West Lafayette, Indiana 47907

Received June 24, 1985; accepted October 21, 1985

An inverse problem is formulated for agglomerating particle populations with a view toward determining the agglomeration frequencies from measurements of particle size distributions. The methodology is developed for the case where the agglomeration frequency is homogeneous and where self-preserving distributions exist. Since the inverse problem is ill-posed, regularization techniques have been used to facilitate inversion. The method has been successfully employed to recover known agglomeration frequencies from self-preserving distributions obtained from forward computation for the cases of approximate Brownian, continuum Brownian, and free-molecular Brownian agglomerations as well as diffusion-controlled growth of supported metal crystallites. Accurate measurements of size distributions are required for satisfactory inversion to obtain agglomeration frequencies. © 1986 Academic Press, Inc.

I. INTRODUCTION

The evolution of size distributions in a population of agglomerating particles uniformly dispersed in space is determined by a population balance equation (PBE) which features as its key ingredient an agglomeration frequency generally a function of particle sizes. This equation is given by

\[
\frac{\partial n(v, t)}{\partial t} = \frac{1}{2} \int_0^\infty q(v - v', v') n(v - v', t) n(v', t) dv' - n(v, t) \int_0^\infty q(v, v') n(v', t) dv' \tag{1}
\]

where \(n(v, t)\) is the number density of particles of volume \(v\) and \(q(v, v')\) is the size-specific agglomeration frequency. When this agglomeration frequency is known, the population balance equation \([1]\) can be solved numerically subject to some initial distribution. The frequency \(q(v, v')\) depends on the physical mechanism by which agglomeration occurs. Generally, agglomeration occurs as a result of relative motion between particles leading to their approach within sufficient proximity of each other. Indeed such relative motion must depend on the state of motion in the continuous phase in which the particles are suspended. The reciprocal effects of motion in the continuous phase and that among the local particle population prohibit a detailed and rigorous mechanical analysis of the actual situation.

Several agglomeration frequencies exist in the literature which have been obtained based on different simplifying assumptions. A classic is the work of Smoluchowski (1) on Brownian motion (for other situations see Levich (2)). The validity of such frequencies can be put to test only by actually analyzing the population for size distributions and comparing with experimental data. This procedure is somewhat indirect and at best only a substitute in the absence of more direct methods of identifying the frequency function from experiments. This is particularly so in situations where reasonable models are not available for analyzing the relative motion between particles. The most direct method of identifying the frequency functions calls for microscopic observation of the particles and examining individual agglomeration events. This procedure is impractica-
AGGLOMERATION KINETICS

Thus experimental observation is limited to dynamic measurements of particle size distributions. The identification problem is then one of extracting the agglomeration frequency from experimental data on particle size distributions. It is this problem which we refer to as the "inverse" problem of coagulation kinetics.

While it is a relatively simple task to predict the evolution of the size spectra of the dispersed phase particles once the breakage and coalescence frequencies are known, the identification of these functions from size spectra is more difficult. The objective of this work is to address the inverse problem of coagulation kinetics.

II. FORMULATION OF THE INVERSE PROBLEM

As mentioned earlier, the inverse problem of coagulation kinetics is concerned with the estimation of the agglomeration frequency \( q(v, v') \) from transient size spectra. The task of identification is considerably simplified by restricting to systems where the population evolves due to agglomeration alone. Such systems are indeed realizable in practice. For example, in stirred liquid-liquid dispersions one can neglect breakage effects in the initial period if the initial particles are very small. Ultimately however breakage will also become important. The direct estimation of a bivariate function \( q(v, v') \) from Eq. [1] using experimentally measured \( n(v, t) \) is a formidable task and it is essential to consider various means of simplification of this inverse problem. Such means of simplification must arise from exploiting any additional properties by ways of symmetry or other upon which we will dilate at a later stage. We note here that \( q(v, v') \) is clearly a symmetric function of its arguments. In the treatment that follows it will be convenient to work with the cumulative volume fraction of the dispersed phase particles which is defined by

\[
F(v, t) = \frac{1}{\phi} \int_0^v v' n(v', t) dv'
\]

where \( \phi \) is the volume fraction of the dispersed phase particles and \( F(v, t) \) satisfies the conditions

\[
F(0, t) = 0, \quad \lim_{v \to \infty} F(v, t) = 1.
\]

The equation for \( F(v, t) \) from [1] may be identified as

\[
\frac{\partial F(v, t)}{\partial t} = \frac{\phi}{2} \int_0^v dF(v', t) \int_0^{v-v'} dF(v'', t) \times \left[ \frac{1}{v'} + \frac{1}{v''} \right] q(v', v'') - \phi \int_0^v dF(v', t) \int_0^\infty dF(v'', t) \frac{1}{v''} q(v', v''). \tag{3}
\]

The inverse problem is now one of determining the coalescence frequency \( q(v', v'') \) given experimental measurements of \( F(v, t) \). The motivation to work with \( F(v, t) \) arises out of the fact that when the so-called "self-preserving" distributions exist (see, for example, Swift and Friendlander (3) and Hidy (4)) they are most simply described by

\[
F(v, t) = f(z) \tag{4}
\]

where \( z \) is a "similarity" variable representing some combination of \( v \) and \( t \). The primary stress of this paper is the investigation of the inverse problem when such similarity solutions exist.

Similarity Concept and Self-preserving Spectra

The term "similarity" has many implications. In the broadest sense it implies a reduction in the number of independent variables in the problem (generally by one) as a result of some invariance relationships. Physically, the process harbors some behavioral symmetry which manifests in some quantitative manner. More specifically, consider some of the well-known agglomeration frequencies displayed in Table I. Clearly, all the frequencies listed satisfy the following two properties

\[
q(v, v') = q(v', v) \quad \text{(symmetry)} \tag{5}
\]

\[
q(\lambda v, \lambda v') = \lambda^n q(v, v') \quad \text{(homogeneity)} \tag{6}
\]

TABLE I

<table>
<thead>
<tr>
<th>Mechanism of agglomeration</th>
<th>Agglomeration frequency $q(v, v')$</th>
<th>$m$</th>
<th>$Q(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approximate Brownian coagulation</td>
<td>$k$</td>
<td>0</td>
<td>$k$</td>
</tr>
<tr>
<td>Limited case of gravitational coagulation</td>
<td>$k(v + v')$</td>
<td>1</td>
<td>$k(1 + x)$</td>
</tr>
<tr>
<td>Brownian coagulation (Continuum regime)</td>
<td>$(kv^{1/3} + v'^{1/3})(v^{-1/3} + v'^{-1/3})$</td>
<td>0</td>
<td>$k(2 + x^{1/3} + x^{-1/3})$</td>
</tr>
<tr>
<td>Brownian coagulation (Free molecular regime)</td>
<td>$k\left(\frac{1}{v} + \frac{1}{v'}\right)^{1/2}(v^{1/3} + v'^{1/3})^2$</td>
<td>1/6</td>
<td>$k\left(\frac{1}{x} + \frac{1}{x'}\right)^{1/2}(1 + x^{1/3})^2$</td>
</tr>
<tr>
<td>Laminar or turbulent shear</td>
<td>$k(v^{1/3} + v'^{1/3})$</td>
<td>1</td>
<td>$k(1 + x^{1/3})^3$</td>
</tr>
<tr>
<td>Diffusion-controlled growth of supported metal crystallites</td>
<td>$k(v^{-2/3} + v'^{-2/3})$</td>
<td>$-2/3$</td>
<td>$k(1 + x^{-2/3})$</td>
</tr>
</tbody>
</table>

While the first of the properties above is an obvious result of the manner in which the frequency is defined, the second implies a "similarity" in agglomeration between particle pairs whose sizes bear a constant ratio. More precisely, if we set $\lambda = 1/v'$ in [6] we obtain

$$q(v, v') = v^m Q\left(\frac{v}{v'}\right)$$  \[7\]

where

$$Q\left(\frac{v}{v'}\right) = q\left(\frac{v}{v'}, 1\right).$$

Thus the bivariate nature of $q(v, v')$ is reduced to a simpler characterization involving a single parameter $m$ and a univariate function $Q(v/v')$. The symmetry property [5] yields

$$Q(x) = x^m Q\left(\frac{1}{x}\right).$$  \[8\]

It is clear from [8] that the values of $Q(x)$ in the unit interval $(0, 1)$ completely determine its values in the interval $(1, \infty)$. It is also thus evident that the parameter $m$ and $Q(x)$, $x \in (0, 1)$ completely determine $q(v, v')$. Our concern will therefore be to obtain from experiments the value of $m$ and $Q(x)$ so that the agglomeration frequency can be identified via [7].

There are some deep reasons why the foregoing quest is very reasonable. The homogeneity property implies that by examining agglomeration in a limited size interval, it is possible to deduce the agglomeration characteristics of particles outside that interval. A consequence of such "similarity" in agglomeration properties is possibly the development, in due course of time, of size distributions in which some sort of similarity manifests. It is then a question of determining the coalescence frequency from an analysis of similarity behavior. In some systems, as for example stirred liquid-liquid dispersions, breakage occurs in conjunction with coalescence. If however, one begins with very small particles (negligible breakup rates) then the initial transient period must reflect the effects of coalescence alone. Eventually however breakage will also become important. If the initial particles are sufficiently small, it is likely that the period during which only coalescence occurs is sufficiently long to allow for similarity behavior to manifest. That only small particles were present during the agglomeration period.
need not concern us because of the implications of a homogeneous coalescence frequency.

Mathematically, the implication of homogeneity is most directly felt in the cumulative distribution function \( F(v, t) \) which under some situations approaches a “similarity form” that is considered below.

**Establishment of the Identification Technique**

We focus on [3] which features \( F(v, t) \) rather than \( n(v, t) \) and propose the transformation

\[
 f(z) = F(v, t) \quad z = vg(t)
\]

where \( g(t) \) is a nonnegative function to be determined. This suggests that at large times, the cumulative volume fraction distribution may become invariant when plotted against the similarity variable “\( z \).” One surmises that \( g'(t) < 0 \) to account for the fact that particles must progressively increase in size in a purely agglomerating system. Substituting [9] in [3] one obtains

\[
 \frac{zf'(z)g(t)}{\phi} = \frac{1}{2} \int_0^z df(z') \int_0^{z-z'} df(z'')
\]

\[
 \times \left( \frac{1}{z} + \frac{1}{z''} \right) g(t) \left( \frac{z''}{z} \right)^m \frac{Q\left( \frac{z'}{z} \right)}{[g(t)]^m} - \int_0^z df(z')
\]

\[
 \int_0^\infty df(z') \frac{1}{z'} g(t) \left( \frac{z''}{z} \right)^m \frac{Q\left( \frac{z'}{z} \right)}{[g(t)]^m} \left( \frac{z''}{z} \right)^m \left( \frac{z'}{z} \right)^m Q\left( \frac{z'}{z} \right)
\]

or

\[
 \frac{zf'(z)g(t)[g(t)]^{m-2}}{\phi} = \frac{1}{2} \int_0^z df(z')
\]

\[
 \int_0^{z-z'} df(z') \left( \frac{1}{z} + \frac{1}{z''} \right) \left( \frac{z''}{z} \right)^m \frac{Q\left( \frac{z'}{z} \right)}{[g(t)]^m} \left( \frac{z''}{z} \right)^m \left( \frac{z'}{z} \right)^m Q\left( \frac{z'}{z} \right)
\]

\[
 - \int_0^z df(z') \int_0^\infty df(z'') \left( \frac{z''}{z} \right)^m Q\left( \frac{z'}{z} \right)
\]

A similarity transformation becomes feasible on letting

\[
 g'(t)[g(t)]^{m-2} = -a_m, \quad a_m > 0 \quad [12]
\]

to assure the negativity of \( g'(t) \). Equation [12] is readily integrated to get

\[
 g(t) = \left[ g_0^{m-1} - (m - 1)a_m(t - t_0) \right]^{1/(m-1)}, \quad m \neq 1
\]

\[
 g(t) = g_0 e^{-a_1(t-t_0)}, \quad m = 1. \quad [13]
\]

where \( t_0 \) is an arbitrary time fixed during the period of envisaged similarity behavior.

The number density function \( n(v, t) \) is obtained as

\[
 n(v, t) = \frac{\phi}{v} \frac{\partial F}{\partial v}(v, t) = \frac{\phi}{v} f'(z) \frac{\partial z}{\partial v},
\]

\[
 = \phi[g(t)]^2 \frac{f'(z)}{z}. \quad [14]
\]

The total population density is obtained as

\[
 N(t) = \frac{\phi g(t)}{\phi} \int_0^\infty \frac{f'(z)}{z} dz \quad [15]
\]

from which we obtain

\[
 g(t) = \frac{N(t)}{\phi} \left[ \int_0^\infty \frac{f'(z)}{z} dz \right]^{-1}. \quad [16]
\]

In [15] and [16] we assume the existence of the integrals occurring there in. Clearly the similarity variable can be expressed as

\[
 z = \frac{N(t)\phi}{\phi I} \quad [17]
\]

where

\[
 I = \int_0^\infty \frac{f'(z)}{z} dz. \quad [18]
\]

This is the same as that of Swift and Friedlander (3) and Hidy (4) except for a multiplicative constant \( 1/I \). Interestingly enough one has the freedom of choosing \( I \). On defining \( z = N(t)v/\phi, I \) becomes identically equal to 1. The similarity transformation is explicit, because the asymptotic decay law for \( N(t) \) is known. Combining [13] and [16] and choosing \( I = 1 \), yields

\[
 N(t) = \phi \left[ \frac{N_0}{\phi} \right]^{m-1} - a_m(m - 1)(t - t_0)^{1/(m-1)}, \quad m \neq 1
\]

\[
 N(t) = N_0 e^{-a_1(t-t_0)}, \quad m = 1. \quad [19]
\]

*Journal of Colloid and Interface Science, Vol. 112, No. 2, August 1986*
In the above \( N_0 \) is the total number density at time \( t_0 \). Equation \([19]\) can be recast as
\[
N(t) = N_0 e^{-a(t-t_0)}, \quad m = 1
\]
where
\[
A_m = a_m \phi^{m-1}.
\]

The above can be written in terms of dimensionless variables as shown below
\[
y(r) = y_0 e^{-a(t-t_0)k}, \quad m = 1.
\]

In the above \( \tau \) is the dimensionless time defined by
\[
\tau = \frac{\phi^m k t}{[N(0)]^{m-1}}. \quad [22]
\]
The quantity \( k \) is the multiplicative constant in the coalescence frequency. Equation \([21]\) specifies the asymptotic decay law for the total number of particles in dimensionless form. Suppose now that experimental data are available for \( N(t) \). After a sufficiently long time, say, \( t_0 \) one assumes the validity of \([20]\) and glean the data for a direct fit of the parameters \( m \) and \( A_m \) or alternatively adopt the procedure outlined below. Partition the interval \( t > t_0 \) using the increasing sequence \( t_1, t_2, \ldots \) and let \( N_i = N(t_i) \). It is readily shown that for \( m \neq 1 \),
\[
N_i^{m-1} - N_{i+1}^{m-1} = A_m (1 - m) (t_i - t_{i+1}) \quad [23]
\]
through which the fit of constants \( m \) and \( A_m \) from data is facilitated. Consider, for example, the kernel \( q(v, v') = k vv' \) which has \( m = 2 \) and \( Q(x) = x \). Substitution of \( m = 2 \) in \([19]\) yields
\[
N(t) = N_0 - a_m \phi (t - t_0).
\]
The case \( m = 1 \) is even simpler and \([23]\) is replaced by
\[
N_{i+1} = N_0 e^{-ai(t_{i+1} - t_i)}. \quad [24]
\]
Such an exponential decay is identified experimentally without much difficulty. From the previous discussion it is clear that asymptotic data can provide the value of \( m \), the degree of homogeneity of the kernel. What remains therefore is the identification of the function \( Q(x) \). For this we must return to \([11]\) and rewrite it as
\[
-zf'(z) \frac{a_m}{\phi} = \frac{1}{2} \int_0^z df(z') \int_0^{z'} df(z'')
\]
\[
\times \left[ \frac{1}{z'} + \frac{1}{z''} \right] (z'')^{m} Q \left( \frac{z'}{z''} \right) - \int_0^z df(z')
\]
\[
\times \int_0^\infty df(z'')(z'')^{m-1} Q \left( \frac{z'}{z''} \right). \quad [25]
\]
It is possible to process the above equation further and rewrite it as
\[
g(z) = zf'(z) \frac{a_m}{\phi} = \int_0^\infty k(z, \xi) Q(\xi) d\xi \quad [26]
\]
where
\[
k(z, \xi) = \frac{1}{2^\xi} \int_0^{\xi^{1+\xi}} u^m f'(u) f'(u/\xi) du
\]
\[
+ \int_0^{\xi} u^m f'(u) f'(u/\xi) du. \quad [27]
\]
The above equations can be simplified to obtain
\[
g(z) = zf'(z) \frac{a_m}{\phi} = \int_0^1 k(z, \xi) Q(\xi) d\xi \quad [28]
\]
in which
\[
k(z, \xi) = \kappa(z, \xi) + \kappa(z, \frac{1}{\xi}) \frac{1}{2^\xi}. \quad [29]
\]
We have thus obtained an integral equation of the first kind for \( Q(x) \). The function \( g(z) \) appearing in Eq. \([26]\) is not the same as that used in deriving the similarity transformation. The left-hand side and the kernel \( k(z, \xi) \) are obtained through processing of experimental data. Thus the similarity analysis based on the physical principles of symmetry and homogeneity of \( q(v, v') \) have been used to obtain the identification technique by the application of a similarity analysis.

As is typical of most inverse problems, one has to deal with an integral equation of the

\[\text{Journal of Colloid and Interface Science, Vol. 112, No. 2, August 1986} \]
first kind which falls in the category of ill-posed problems. Such problems call for special solution strategies.

III. NUMERICAL EXAMPLES

To demonstrate the inversion technique presented in the previous section the following procedure has been adopted. The agglomeration equation was solved for coagulation frequencies which are well known to yield self-preserving spectra. The self-preserving distributions obtained were then used to recover back the coagulation frequencies which went into the PBE. The coalescence frequencies considered are well known to exhibit the similarity behavior outlined earlier and are listed below along with the references.

(i) \( q(v, v') = k \) 

(ii) \( q(v, v') = k(v^{1/3} + v'^{1/3})(v^{-1/3} + v'^{-1/3}) \)

(iii) \( q(v, v') = k(v^{1/3} + v'^{1/3})(v^{-1/3} + v'^{-1/3})^2 \)

(iv) \( q(v, v') = k(v^{-2/3} + v'^{-2/3}) \)

In each case a discretized set of equations corresponding to the continuous version of the PBE were solved using the IMSL subroutine DGEAR employing the stiff method of backward differentiation and a relative error tolerance of \( 1 \times 10^{-10} \). The equations are succinctly written as

\[
\frac{dy_i}{d\tau} = \frac{i-1}{2} \sum_{j=1}^{i-1} q^*(i-j, j) y_{i-j} y_j \\
- y_i \sum_{j=1}^{n_0} q^*(i, j) y_j \quad i = 1, 2, \ldots, n_0.
\]  

[29]

In the above \( y_i = n(v_i, t)/N(0) \), represents the dimensionless concentration of the \( i \)-mer while \( q^*(i, j) \) and \( \tau \) denote the dimensionless coalescence frequency and the dimensionless time, respectively. In each case a monodisperse initial distribution of \( n_0 \) particles was employed. The initial number of particles \( n_0 \) varied between 600 and 1200 and was chosen to be large enough to ensure negligible loss of mass during computation. This was checked periodically. At sufficiently large dimensionless times, \( \tau \), the distributions obtained were checked for self-preservation. The fractional number of particles were also recorded at different times \( \tau \). The maximum \( \tau \) was different in each case and depended upon the convergence rate at the lower end of the self-preserving spectrum. To avoid differentiation of discrete \( f(z) \) which is itself an ill-posed problem \( f'(z)/z \) was directly obtained. It turns out that \( f'(z)/z \) is the self-preserving distribution \( \Psi(z) \) as defined by Friedlander. Denoting the number density of particles consisting of \( i \) monomers by \( n_i \),

\[
\Psi_i = \frac{f'(z_i)}{z_i} = \frac{n_i \phi}{N^2 v_i}
\]  

[30]

and the above can be directly obtained from experiment.

The Inverse Problem

The results of the simulations were subjected to the similarity approach (outlined in Section II) to recover back the coalescence frequencies that determined the self-preserving distributions.

The total particle density \( N(t) \) at large times was obtained from its dimensionless counterpart \( y(\tau) \) and was used to identify the degree of homogeneity "\( m \)" and the parameter \( A_m \) by means of a direct least-squares fit of Eq. [20] using nonlinear regression SPSS package. The initial number of particles \( N(0) \), the multiplicative constant \( k \) and the dispersed phase fraction \( \phi \) were arbitrarily chosen for illustrative purposes. The regression model being nonlinear good initial guesses of "\( m \)" and \( A_m \) are required. Experimentally good initial estimates of these can be obtained as follows. For example, in a batch coagulation experiment with essentially monodisperse feed consisting of particles of volume \( v \) (i.e., minimum polydispersity), the coagulation frequency is approximately constant during the initial stages and the total particle density decays as

\[
- \frac{dN}{dt} = q(v, v) N^2
\]  

[31]

from which

$$q(v, v) = \frac{1}{t} \left[ \frac{1}{N(t)} - \frac{1}{N(0)} \right].$$  \[32\]

By repeating the experiment with a monodisperse feed, of, say, volume $\lambda v$, $q(\lambda v, \lambda v)$ is estimated. The degree of homogeneity is then estimated via \[6\] and subsequently $A_m$ via \[20\]. The least-squares procedure is computationally facilitated by dividing \[20\] by $N(0)^{m-1}$. The next stage involved the solution of the integral equation \[28\] for the function $Q(x)$. This problem as such is ill-posed in that the solution does not depend continuously on the data, i.e., two slightly differing sets of the data $g$ can lead to two entirely different solutions. This is critical because experimental data is subject to stochastic fluctuations and one seldom knows the exact data. To circumvent this difficulty, Tikhonov (8) suggested solving a well-posed approximation of the ill-posed problem. Instead of solving the operator equation $Kq = g$ one solves the variational problem of minimizing $\|Kq - g\|^2 + \lambda \Omega(q)$ where $\lambda$ is called the regularizing parameter and is chosen in accordance to the degree of uncertainty in $K$ and $g$. The quantity $\|Kq - g\|^2$ measures the fidelity to the data and $\Omega(q)$ is a smoothing functional representing certain undesirable traits to be minimized in the solution. Specifically, the undesirable traits are a large norm and oscillatory character of the solution. The regularizing parameter acts as a tradeoff parameter. As $\lambda$ increases, the residual error increases but the solution has less of the undesirable features. The above procedure is called regularization. For detailed descriptions of regularization we refer to Morozov (9) and Groetsch (10). A brief discussion with motivations is given in (11). In cases where $K$ is known accurately and $g_i = \bar{g}_i + \epsilon_i$ where $\epsilon_i$ are normally distributed random variables of mean $0$ and standard deviation $\sigma$, Wahba (12) and Crump and Seinfeld (13) have developed the cross-validation algorithm for determining the parameter $\lambda$. The validity of the cross-validated estimate in the presence of noise in the operator has not been established. The integral Eq. \[28\] was solved by the regularizing technique. The numerical procedures are given in (14–16) and are summarized in Appendix A. A hundred equally spaced data points $\{\Psi_i\}_{i=1, \ldots, 100}$ were interpolated using cubic splines (IMSL subroutine ICSSCV) to obtain the self-preserving distribution function and construct the kernel $k(z, \xi)$ in \[28\]. The data can be represented as

$$\int_0^1 k(z_i, \xi)Q(\xi)d\xi = g(z_i), \quad i = 1, \ldots, M.$$  \[33\]

The collocation points $\{z_i\}_{i=1, \ldots, M}$ were logarithmically equally spaced so as to cover the distribution in some interval $[z_{\text{min}}, z_{\text{max}}]$. The cross-validated estimate $\bar{\lambda}$ was used as a lower bound in all the cases as uncertainty in the operator increases the estimate of $\lambda$. Although the choice of $\lambda$ is to some extent arbitrary, this is not a serious deficiency because one always has the liberty to use the regularized solution in the PBE, predict the transients and compare them with observed data. In fact in many cases, one solves for several values of $\lambda$. The choice of $\lambda$ is facilitated by the fact that an estimate of the solution can be obtained from the rate of decay of particles in the initial period starting with an essentially monodisperse feed. In this manner a satisfactory $\lambda$ can be obtained.

Influence of Noise

To simulate experiment, external noise was imposed on the self-preserving data. This was accomplished by replacing the data $\{\Psi_i\}_{i=1, \ldots, 100}$ by

$$\bar{\Psi}_i = \Psi_i + \epsilon_i, \quad i = 1, \ldots, 100$$  \[34\]

where $\epsilon_i$ are random numbers with a mean of zero and variance $\sigma^2 \Psi_i^2$. This illustrative model is based on the assumption that the noise is proportional to the size of the measurement. The data were smoothed using cubic splines (IMSL routine ICSSCV) to obtain an approximation to the self-preserving curve and the inverse problem was solved again.
is interesting to note that the same noise goes into the construction of the integral operator and the left-hand side of Eq. [28]. As such we have an approximation of Eq. [28] given by

$$Kq = g.$$  \[35\]

In cases where satisfactory solutions are not obtained, it may be necessary to filter noise from the operator $K$. The technique is given in Appendix B.

The results of inversion for the coalescence mechanisms considered previously are studied for different values of the signal to noise ratio [SNR] defined by

$$SNR = \frac{1}{M} \sum_{i=1}^{M} (\Psi_i/e)^2.$$  \[35\]

For the above noise model

$$SNR = \frac{1}{\sigma}.$$  \[35\]

### IV. DISCUSSION OF RESULTS

(i) **Approximate Brownian Coagulation**

If it is known that the coagulation frequency is a constant, then its estimation can be made by scanning only the total number density of particles at different times. However, if the frequency is not known a priori to be a constant, the methodology of this paper can be used effectively to diagnose the constancy of the frequency function. Figure 1a depicts the self-preserving distribution. The decay law for total particle density is shown in Fig. 1b. For this case, the analytical form of the self-preserving solution is available (5) and is given by

$$\Psi(z) = e^{-z}.$$  \[35\]

This is used to estimate $Q(x)$ via [28]. The result is shown in Fig. 1c [soln. A] along with the solution obtained from the simulation data [soln. B].

---

Fig. 1. (a) The self-preserving distribution. (b) Asymptotic decay of total particle density. $m$ estimate $= -0.0001$. (c) $Q(x)$ estimates.

*Journal of Colloid and Interface Science, Vol. 112, No. 2, August 1986*
(ii) Brownian Coagulation (Continuum)

The self-preserving distribution for this case is shown in Fig. 2a. The asymptotic decay law is displayed in Fig. 2b. The estimate of $m$ from regression was 0.002.

It is of course possible to accept this estimate of 0.002 for $m$ and proceed with the estimation of $Q(x)$ to complete the identification process. Figure 2c shows the result of this estimation.

On the other hand, estimation of physical quantities is frequently aided by notions about the physics of the process. For example, it is difficult to imagine that any direct physical argument would produce a fraction of 0.002 for the homogeneity parameter $m$. One may in that case surmise that the value of $m$ may in fact be zero and allow the veracity of the estimated frequency to be established by its ability to predict the observed size distribu-
(iii) Brownian Coagulation (Free Molecular)

The self-preserving distribution, asymptotic decay law, and regularized solutions are shown in Figs. 3a, b, and c, respectively.

(iv) Diffusion-Controlled Growth of Supported Metal Crystallites

The self-preserving spectrum, asymptotic decay law, and regularized solutions are displayed in Figs. 4a, b, and c, respectively.

It is observed that the regularized solutions approximate the actual curve very well except near the origin where the quality of the estimate decreases as the value of SNR decreases. Even in the absence of external noise, the difficulty near the origin is twofold. The self-preserving distribution itself does not carry adequate information on sizes of particles disappearing immediately. Moreover, the approximating scheme employed (expanding \( Q(x) \) in terms of splines) may not be very appropriate. It appears that the order of singularity of \( Q(x) \) near the origin must be known.

As in any identification process, the effect of noise obscures the high-frequency components of the unknown function. This is evident from Figs. 2d, 3c, and 4c. The self-preserving distributions for cases (i), (ii), and (iii) are very close as depicted in Fig. 5. Hence to accurately resolve between the corresponding \( Q(x) \)'s near the origin, exact data are required. The redeeming feature is the fact that the direct problem of predicting size distributions does not bear the same degree of sensitivity to the coalescence frequency. Indeed the constant coalescence frequency is a good working approximation of the Brownian case. It is however significant to observe that the identifica-
tion procedure reveals the rising characteristic of $Q(x)$ near the origin in a quantitative way although failing at the origin.

V. CONCLUSIONS

As expected, the solution of the inverse problem in agglomeration kinetics to obtain the agglomeration frequency from data on transient particle size distribution is very sensitive to the available data. The self-preserving curve must be determined as accurately as possible in order that reasonable estimates are obtained of the agglomeration frequency. Since the self-preserving distribution tends to wash out particle size-ratios with very high agglomeration rate, it is not reasonable to expect the identification to provide very good estimates of the agglomeration frequency in this range. Thus the inversions obtained here yield good estimates of the agglomeration frequency for all cases considered for all particle size-ratios except near the origin where the singularity of the frequency implies very high agglomeration rates. Detailed transients prior to
the attainment of self-preservation are needed to make the inversion in this range complete.

A limited amount of noise in the available data can be accommodated by the inversion procedure by suitable manipulation of the regularization parameter. Although the inversion procedure is basically unstable, the fact that reasonable predictions of particle size distributions can be made with relatively erroneous agglomeration frequencies mitigates the seriousness of this error. Indeed this paper establishes that accurate data can yield good estimates of the agglomeration frequency.

It must be recognized however that the inversion procedure provides a numerical solution for the unknown frequency within a homogeneous form. Thus it has no capacity to resolve differences between mechanisms which are numerically compatible with the same self-preserving distribution.

VI. SUMMARY

A technique for direct estimation of the coalescence frequencies of particles has been developed from direct observation of the transient size distributions of purely agglomerating systems. While the PBE has been employed in the past for predicting transient size distributions, we have adapted the population balance framework to recover homogeneous agglomeration frequencies from transient data in cases where self-preservation is observed. It is however true that not all agglomeration frequencies are homogeneous and that not all homogeneous agglomeration frequencies yield self-preserving spectra as described in Section II. Inversion techniques require development for nonhomogeneous agglomeration frequencies. Possibly, this lies in the identification of more general invariants in the distribution for different clans of kernels admitting suitable properties. Another possible drawback is the requirement of large quantity of accurate data. On the other hand, if data are garbled and scanty, several different agglomeration mechanisms may result and an accurate estimate of the coalescence frequency is probably not justifiably obtained.

APPENDIX A

Computational Procedure for Regularization

It is desired to solve the Eq. [28] approximately. The unknown function is expressed as a linear combination of B-splines (17), i.e.,

\[ q = \sum_{j=1}^{p} c_j B_j. \]  

[A.1]

The smoothing functional is given by

\[ \Omega = \sigma_0 \langle q, q \rangle + \sigma_1 \langle q', q' \rangle + \sigma_2 \langle q'', q'' \rangle \]  

[A.2]

where the primes denote differentiation and \( \sigma_0 > 0, \sigma_1 \) and \( \sigma_2 \geq 0 \). In the above \( \langle u, v \rangle = \int_0^1 u(x)v(x)dx \). On substituting [A.1] in [A.2], the smoothing functional is given by \( \Omega = c^TWc \) where \( W \) is a \((p \times p)\) positive definite symmetric matrix with

\[ w_{ij} = \sigma_0 \int_0^1 B_i(s)B_j(s)ds + \sigma_1 \int_0^1 B'_i(s)B'_j(s)ds + \sigma_2 \int_0^1 B''_i(s)B''_j(s)ds. \]  

[A.3]

The regularized solution is given by

\[ \text{Min} \frac{1}{n} \| Xc - y \|^2 + \lambda c^TWc \]  

[A.4]

where \( X \) is a \((n \times p)\) design matrix whose components are given by

\[ X_{ij} = \int_0^1 k(x_i,s)B_j(s)ds \]  

[A.5]

and \( y \) is the \(n\)-dimensional data vector for the integral equation. We shall assume that \( n > p \) as is often the case. A Cholesky decomposition of the positive definite matrix \( W \) is performed to obtain

\[ W = R^TR \]  

[A.6]

where \( R \) is a \((p \times p)\) upper triangular matrix.
Letting $z = Re$, the regularizing problem is given by

$$\min \frac{1}{n} \|XR^{-1}z - y\|_2^2 + \lambda z^Tz. \quad [A.7]$$

A singular value factorization of $XR^{-1}$ is performed to obtain

$$XR^{-1} = UDV^T \quad [A.8]$$

where $U$ is $n \times p$ with orthogonal columns, $D$ is $(p \times p)$ diagonal consisting of singular values of $XR^{-1}$ and $U$ is $(p \times p)$ orthogonal matrix. Letting

$$h = UTy \quad [A.9]$$

and

$$\gamma = VTz \quad [A.10]$$

the regularized solution for a specific value of $\lambda$ is given by

$$\gamma_i = \frac{h_i d_i}{(d_i^2 + n\lambda)} \quad i = 1, 2, \ldots, p \quad [A.11]$$

whence the coefficient vector is obtained from $e = R^{-1}V\gamma$. Numerical algorithms to pick $\lambda$ in certain cases are given in (12-16).

### APPENDIX B

**Filtering Noise in the Integral Operators**

When the data is noisy, the design matrix $X$ is inaccurate and this could be an important source of trouble while solving first-kind problems. Let the matrix norm be defined by

$$\|A\| = \sqrt{\sum_i \sum_j a_{ij}^2}. \quad [B.1]$$

We replace the matrix $X$ by an approximate matrix $\chi$ such that $\|X - \chi\|$ is minimum. A singular value decomposition of $X$ yields

$$X = U_{np}D_{pp}V_{pp}^T. \quad [B.2]$$

Noise corrupts the smallest singular values of $X$ (these are approximately zero). To filter noise the last $p-k$ singular values are set to zero. The corresponding columns of $U, V$ are also made zero. The number of singular values set to zero depends on the noise to signal ratio and specific design matrix $X$. The difference in norms between the two matrices is given by

$$\|X - \chi\| = \sqrt{\sum_{j=k+1}^p d_j^2} \quad [B.3]$$

and $k$ can be estimated by observing the decay rate of the singular values.

### APPENDIX C: NOMENCLATURE

- $a_m$: parameter in decay law for total number of particles
- $f(z)$: cumulative volume fraction distribution as a function of the similarity variable
- $g(z)$: left-hand side of Eq. [28]
- $k$: multiplicative constant in coalescence frequency
- $m$: degree of homogeneity of the coalescence frequency
- $n(v, t)$: number density of the dispersed phase particles
- $n_0$: initial number of particles used in simulations
- $q(v, v')$: coalescence frequency of particles of volumes $v$ and $v'$
- $q^*(i, j)$: dimensionless coalescence frequency of an $i$-mer and $j$-mer
- $t$: time
- $t_0$: some arbitrary time during period of similarity behavior
- $v$: volume of particle
- $v_1$: volume of monomer
- $y(t)$: fractional number of original particles existing at time "$t$"
- $y_i(t)$: dimensionless number of $i$-mers at "$t$" time
- $z$: similarity variable

**Uppercase Letters**

- $A_m$: parameter in decay law for total number of particles
- $F(v, t)$: cumulative volume fraction distribution at time "$t$"
- $N(t)$: total number density of particles at time "$t$"
\( Q(x) \) univariate function required to describe a homogeneous coalescence frequency

Greek Symbols

\( \varepsilon_i \) random noise corresponding to \( \Psi_i \)
\( \Psi_i \) self-preserving data
\( \zeta \) particle size ratio

ACKNOWLEDGMENTS

The authors are grateful to the National Science Foundation for financial support under Grant CPE8414204. The authors are indebted to Professor D. Thoe of the Purdue University Mathematics Department for several illuminating discussions.

REFERENCES