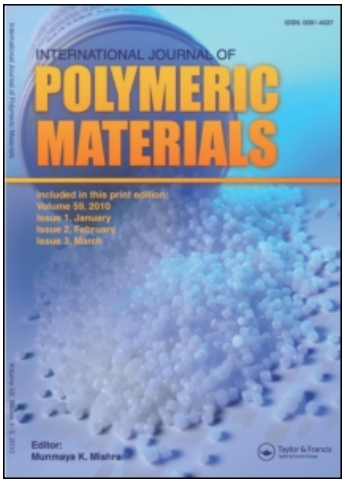

International Journal of Polymeric Materials
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713647664

An Iterative Computer Algorithm for Generating Line Spectra from Linear Viscoelastic Response Functions

Igor Emri[a]; N. W. Tschoegl[b]
[a] University of Ljubljana, Ljubljana, SI, Slovenia [b] California Institute of Technology, Pasadena, California, U.S.A.

To cite this Article Emri, Igor and Tschoegl, N. W.(1998) 'An Iterative Computer Algorithm for Generating Line Spectra from Linear Viscoelastic Response Functions', International Journal of Polymeric Materials, 40: 1, 55 — 79
To link to this Article: DOI: 10.1080/00914039808050143
URL: http://dx.doi.org/10.1080/00914039808050143

Intern. J. Polymeric Mater., 1998, Vol. 40, pp. 55–79
Reprints available directly from the publisher
Photocopying permitted by license only

An Iterative Computer Algorithm for Generating Line Spectra from Linear Viscoelastic Response Functions

IGOR EMRI[a] and N. W. TSCHOEGL[b]

[a]*University of Ljubljana, Ljubljana, SI 61000, Slovenia*
[b]*California Institute of Technology, Pasadena, California 91125, U.S.A.*

(Received 15 June 1997)

This paper examines the reasons for the success of a simple iterative computer algorithm proposed earlier, that is useful in generating line spectra from the experimental response functions commonly used in linear viscoelastic theory. Generating spectra from these response functions is mathematically an ill-posed problem because the inversion of the response function leads to instabilities in the calculations. The paper identifies those regions of the response functions in which experimental and computational rounding errors lead to the instabilities and which are thus responsible for the ill-posedness. It shows that these regions can be excluded from the calculations without impairing the line spectra obtained.

Keywords: Fredholm integral equations (first kind); ill-posed problems; line spectra; viscoelastic response functions

1. INTRODUCTION

In a series of earlier papers we had described an iterative computer algorithm for obtaining line spectra from the most frequently used experimental response functions of linear viscoelastic theory [1–4]. In those papers we showed *how* the algorithm works but did not discuss *why* – in view of the ill-posedness of the inversion of Fredholm integral equations of the first kind – it is as successful as it is. It is the purpose of this paper to address the latter question.

To examine the reasons for the success of the algorithm, we consider the (shear) relaxation modulus,

$$G(t) = \{G_e\} + \sum_{i=1}^{i=N} H_i \exp(-t/\tau_i), \tag{1}$$

as a typical example of an experimental response function. The braces around the equilibrium modulus, G_e, signify that the modulus is absent when the material is rheodictic (shows steady-state flow), but is present when it is arrheodictic (does not show steady-state flow) [5]. The H_i in Eq. (1), are modulus values representing the strengths associated with a particular relaxation time, τ_i. The number, N, of such pairs forms the set $\{H_i, \tau_i; i = 1, 2, \ldots, N\}$ that constitutes the *discrete* relaxation, or *line*, spectrum, $H(\tau_i)$. In Eq. (1) the type of excitation, a step function of time, is represented by the exponential kernel function, $K(t, \tau_i) = \exp(-t/\tau_i)$, where t is the current time.

If the relaxation times may be deemed to be sufficiently closely spaced, the sum in Eq. (1) may be replaced by an integral. We then have

$$G(t) = \{G_e\} + \int_0^\infty H(\tau) \exp(-t/\tau) \frac{d\tau}{\tau}, \tag{2}$$

where $H(\tau)$ is the *continuous* relaxation spectrum.

Neither $H(\tau_i)$, nor $H(\tau)$, can be determined by direct experiment [5] since there is no excitation that would not depend on time in one way or another. (Constancy with time is just another form of time-dependence). Equation (2) is a specific form of a Fredholm integral equation of the first kind. Attempts at inversion, i.e., attempts to retrieve $H(\tau)$ from under the integral sign, generally lead to instabilities in the calculations because in certain regions of $G(t)$ experimental errors and rounding errors in digital computations, coupled with the finite resolving power of the computer, may obscure the information contained in the concomitant regions of $H(\tau)$. Thus, a small change in the data can lead to a large change in the solution. For this reason the inversion is recognized as an *ill-posed* problem. The mathematical theory of ill-posed problems and ways to obtain usable solutions has been well treated in the literature [6] and it is not our aim

to review it here. Several specific methods have been proposed to cope with the problem as encountered in linear viscoelastic theory. These will be reviewed briefly in Section 5.1, where we will also consider older, classical methods of approximations to $H(\tau)$ in the context of our own work.

The algorithm we have described earlier [1–4] is applied to equations of the type represented by Eq. (1), the digital (discretized) form of Eq. (2). Discretization, of course, does not remove ill-posedness. It must have occurred to readers of our earlier publications – as it had indeed occurred to us – to ask why the algorithm is successful in view of the ill-posedness of the problem it deals with? Briefly, we show that in the calculation of a line, H_i, from the given response function, here $G(t)$, not all regions of the function contribute equally to the instabilities that may affect the calculation of the line. We found it possible to localize and exclude the trouble-causing regions without impairing the determination of the spectrum. The task – as will be discussed in detail below – is accomplished by scanning a given data set through, as it were, 'windows', and operating only on the information contained in them. These windows effectively constitute band-pass filters that move along the abcissa. In the following section we present this procedure which we call the *Windowing (or Filtering) Approach.*

A line spectrum calculated by this approach from any of the experimental response functions necessarily constitutes an *approximation* to the (unknown) 'true' spectrum. However, it can reproduce the source function from which it is derived sufficiently closely to be eminently useful in practical applications. If determined or known over a sufficiently wide range of the time, t, (or, equivalently, the frequencly, ω, of the Laplace transform variable, s) the line spectrum derived from a particular response function contains complete information on the time dependence of the material, unaffected by the type of excitation chosen to elicit the response represented by the kernel function. A line spectrum calculated from data generated in response to a strain excitation can be converted readily into those that will generate the responses to a stress excitation, and vice versa [4]. Once a line spectrum has been obtained, the time-dependent part of any response function within the same mode of deformation (shear, bulk, etc.) is readily obtained from it.

To assemble any desired experimental response the time dependence must be supplemented with the so-called viscoelastic constants [5] such as the equilibrium or glassy moduli, G_e or G_g, or the equilibrium compliance, J_e, the steady-state compliance, J_e^o, and the steady-state fluidity, ϕ_f, (or its reciprocal, the steady-state viscosity, η_f) as required.

2. CONTRIBUTIONS OF THE SPECTRUM TO THE EXPERIMENTAL RESPONSE

To understand how certain regions of the response function influence the instabilities that may arise in the computations, we begin by considering Eq. (2) and examine how particular regions of the spectrum, $H(\tau)$, contribute to a given point of the relaxation modulus, $G(t)$. The contributions are govered by the kernel function, $\exp(-t/\tau)$, which, in effect, selects them from $H(\tau)$.

The upper portion of Figure 1 displays $\log G(t)$, while the lower portion shows $\log H(\tau)$. $H(\tau)$, shown as a continuous spectrum, should here be considered to be an envelope over the spectral strengths, H_i. We consider now a datum point located at $\log t_k$ along the logarithmic time scale and two spectrum lines located at $\log \tau_{k,l}$ and $\log \tau_{k,u}$ to the left and to the right of $\log t_k = \log \tau_k$. The interval $[\log \tau_{k,l}, \log \tau_{k,u}]$ is the 'Window' belonging to the spectrum line located at $\log \tau_k$. The long-dashed line marked $H(\tau_{k,l}) \exp(-t/\tau_{k,l})$, represents the contribution of the spectrum line located at $\log \tau_{k,l}$ to the datum point at $\log t_k$. Similarly, the short-dashed line marked $H(\tau_{k,u}) \exp(-t/\tau_{k,u})$, represents the contribution of the spectrum line located at $\log \tau_{k,u}$ to the same datum point at $\log t_k$. The distances, $\log \lambda_{k,l}$ and $\log \lambda_{k,u}$, constitute the width of the window. Their sizes depend on the type of the kernel function. They must be selected in such a manner that the contribution by the spectrum line located at $\log \tau_{k,l}$ to the datum point located at $\log t_k$ will be negligible, while the contribution of the spectrum line located at $\log \tau_{k,u}$ will make a constant (time-independent) contribution to the same datum point. Generalizing these observations, we can claim that the contributions to the line at $\log t_k$ made by all spectrum lines located to the left of $\log \tau_{k,l}$ will be negligible, while all spectrum lines located to the right of $\log \tau_{k,u}$ will make constant (i.e., time-independent) contributions.

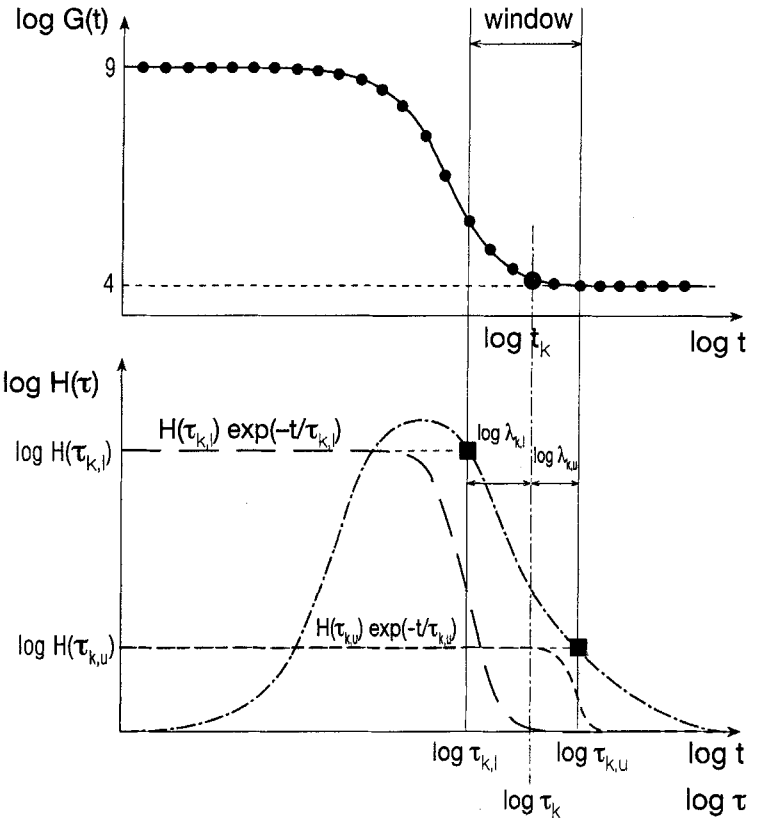

FIGURE 1 Schematic illustrating the contribution of $H(\tau)$ to the point $G(t_k)$.

Let now the experimental data be given by M discrete datum points, $\mathbf{G}(t_j)$, where $j=1,2,3,\ldots,M$. These are represented in the upper portion of Figure 1 by the small filled circles. The solid line representing $G(t)$ is obtained by putting a spline function through these points. We now focus our attention on a particular point, $\mathbf{G}(t_k)$, shown by the large filled circle located at t_k. Using Eq. (1), this point can be modeled in discretized form by

$$\mathbf{G}(t_k) = \{G_e\} + \sum_{i=1}^{i=N} H_i \exp(-t/\tau_i) + \Delta_k, \tag{3}$$

The term Δ_k has been added to account for any experimental errors and/or computational rounding errors. We write $\mathbf{G}(t_k)$ instead of $G(t_k)$ to indicate that such errors may be present.

We now split the sum in Eq. (3) to separate the contributions of the spectrum to the left of the window, within the window, and to the right of it.

$$\mathbf{G}(t_k) = \{G_e\} + \sum_{i=1}^{i=k-l} H_i \exp(-t/\tau_i) + \sum_{i=k-l+1}^{i=k+u} H_i \exp(-t/\tau_i) + \sum_{i=k+u+1}^{i=N} H_i \exp(-t/\tau_i) + \Delta_k, \tag{4}$$

where l and u denote the number of spectrum lines located within the *Window* $[\log \tau_{k,l}, \log \tau_{k,u}]$, while k indicates the spectrum line located at $\log \tau_k$.

Since we are concerned here with the exponential kernel function, $\exp(-t/\tau_{k,l})$, the contribution of the spectrum line located at $\log \tau_{k,l}$ is already smaller than 3.72×10^{-44} when $\log \lambda_{k,l} \geq 2$. In line with the discussion above, contributions to the datum point, $\mathbf{G}(t_k)$, from the first sum are therefore already smaller than Δ_k, and thus effectively vanish. This is brought out in Figure 1 by the long-dashed line representing $H(\tau_{k,l}) \exp(-t/\tau_{k,l})$. The point $H(\tau_{k,l})$ is shown by the upper (left) filled square. On the other hand, $\exp(-t/\tau_{k,u})$ has already reached 0.99 when $\lambda_{k,u} = 2$, and is effectively unity when $\lambda_{k,u} > 2$. Thus, contributions to $\mathbf{G}(t_k)$ from the portion of the spectrum to the right of $\log \tau_{k,u}$ are constant, i.e, time-independent. This is shown in Figure 1 by the short-dashed line that represents $H(\tau_{k,u}) \exp(-t/\tau_{k,u})$. The point $H(\tau_{k,u})$ is shown by the lower (right) filled square. When the kernel is the exponential function, the distances, $\log \lambda_{k,l}$ and $\log \lambda_{k,u}$, are thus conveniently chosen as $\log 2$.

In view of these observations Eq. (4) simplifies to

$$\mathbf{G}(t_k) = \{G_e\} + \sum_{i=k-l+1}^{i=k+u} H_i \exp(-t/\tau_i) + \sum_{i=k+u+1}^{i=N} H_i + \Delta_k. \tag{5}$$

Equation (5) shows that contributions to $\mathbf{G}(t_k)$, apart from Δ_k, are made up of three portions:

(1) the constant contribution from the equilibrium modulus, $\{G_e\}$, that may be absent in a rheodictic material,
(2) time-dependent contributions from the sum within the window, i.e., bounded by log $\tau_{k,l}$ and log $\tau_{k,u}$, and
(3) the time-independent (constant) contributions from the region to the right of log $\tau_{k,u}$.

The situation is depicted schematically in Figure 2. The lower portion displays three regions of $H(\tau)$ marked I, II, and III. The portion of the spectrum in the first (right-hatched) region to the left of log $\tau_{k,l}$ does not contribute to $\mathbf{G}(t \geq t_k)$. The second (cross-hatched) region covers the window, i.e., that part of $H(\tau)$ whose contribution to $\mathbf{G}(t_k)$ and its neighbors to the left and right within the window are time-dependent. Finally, the third (left-hatched) region designates that portion of the spectrum whose contributions to $\mathbf{G}(t \leq t_k)$ are constant with respect to time.

These contributions are shown, again schematically, in the upper portion of Figure 2 which appropriately repeats the hatch patterns employed in the lower portion. Attention should be focused on the window. In Region III the contributions arise from the second sum in Eq. (5) and are time-independent. In Region II the contributions come from the first sum in Eq. (5) and are time-dependent, log t_k marking the point of steepest descent. There are no contributions from Region I since its contribution vanishes. It should be clear, of course, that these regions move along the abscissa for each experimental datum point.

In conclusion: the time-dependent contributions to $G(t)$ within the window arise from that portion of the spectrum that is covered by the window (Region II). The contributions from the portion of the spectrum to the left of the window (Region I) are negligible, and the contributions from the portion to the right (Region III) are constant.

3. INVERSION

We now use the insights just gained to design the algorithm in such a way that it successfully deals with the instabilities that would arise if the complete set of experimental data were included in the inversion. We first consider the case when the data are free of any experimental

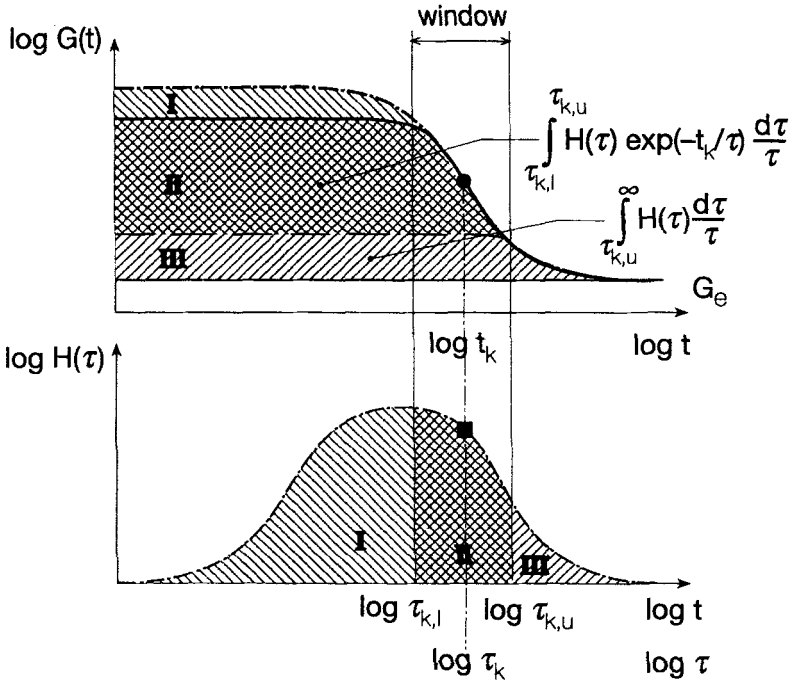

FIGURE 2 Three regions of $G(t)$ and of $H(\tau)$.

error so that the algorithm may need to contend solely with the computational errors resulting from the finite resolving power of the computer. We next derive a typical least squares inversion algorithm operating on the complete data set in the presence of experimental (and computational) error, and pinpoint the sources of instabilities that arise in that case.

3.1. Inversion in the Absence of Experimental Error

Even if the data are not encumbered by experimental errors, it follows from the above discussion that $G(t)$ data to the right of log t_k contain no information on spectrum lines located more than two decades to the left of log $\tau_{k,l}$ in Figure 1. Hence, spectrum lines to the left of log $\tau_{k,l}$ cannot be retrieved from $G(t)$ data to right of log t_k.

However, it further follows that $G(t)$ data to the left of log t_k contain time-independent (constant) contributions from the spectrum lines located to the right of log $\tau_{k,u}$. It would appear that these spectrum lines could therefore be retrieved from $G(t)$ data to the left of log t_k. Nevertheless, as we shall now show, the inclusion of these data in the calculations is a source of instability even in the absence of experimental errors.

To demonstrate this, let us consider that we deal with data that are far on the left side of $G(t)$ in Figure 1, so that even computational errors are of the order of magnitude of the first sum in Eq. (5). Note that proceeding towards the left, $H(\tau)$ becomes progressively smaller while $G(t)$ increases. Thus Eq. (5) then reduces to

$$\mathbf{G}(t_k) = \{G_e\} + \sum_{i=k+u+1}^{i=N} H_i + \chi_k. \tag{6}$$

This is tantamount to having subsumed any contributions below $i=k+u+1$ into χ_k, the unavoidable computational error. All datum points $\mathbf{G}(t_i < t_k)$ can be modeled in a similar manner. Since the spectrum is represented with N spectrum lines, we arbitrarily select N datum points for which $t_i \leq t_k$, yielding a set of N linear equations in N unknown H_i's.

In matrix form we have

$$\{\mathcal{G}\} = [\mathbf{A}] \cdot \{\mathbf{H}\}, \tag{7}$$

where

$$[\mathbf{A}] = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ A_{N1} & A_{N2} & \cdots & A_{NN} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \tag{8}$$

is the (square) kernel matrix,

$$\{\mathbf{H}\} = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_N \end{bmatrix} \tag{9}$$

and

$$\{\mathcal{G}\} = \begin{bmatrix} \mathbf{G}(t_1) & - & \{G_e\} & - & \chi_1 \\ \mathbf{G}(t_2) & - & \{G_e\} & - & \chi_2 \\ & & \vdots & & \\ \mathbf{G}(t_N) & - & \{G_e\} & - & \chi_N \end{bmatrix} \tag{10}$$

are the column vectors containing the spectrum lines, and the datum points, respectively. Since the kernel matrix is clearly singular, the set of equations has no solution. Hence, the use of data where the contribution of the spectrum lines is constant, necessarily leads to instabilities in the inversion even in the absence of experimental errors.

3.2. Least Squares Algorithm Operating on the Complete Set of Data in the Presence of Experimental Errors

To understand how the presence of experimental errors affects the inversion of equations of the type represented by Eq. (1), we now develop a typical least squares algorithm that operates on the complete set of experimental data. It may be noted that this algorithm is essentially the least-squares Multidata Method of Cost and Becker [7], when the latter is recast in matrix form.

Rewriting Eq. (3) for a general value of t_j gives

$$\mathbf{G}(t_j) = \{G_e\} + \sum_{i=1}^{i=N} H_i \exp(-t_j/\tau_i) + \Delta_j \tag{11}$$

for each datum point in the presence of experimental and/or computational error. The (absolute) error, Δ_j, is given by

$$\Delta_j = \mathbf{G}(t_j) - \left[\{G_e\} + \sum_{i=1}^{i=N} H_i \exp(-t_j/\tau_i)\right]. \tag{12}$$

Minimizing the error sum of squares according to

$$\delta = \sum_{j=1}^{M} \Delta_j^2 \tag{13}$$

leads to

$$\frac{\partial \delta}{\partial H_n} = -2\sum_{j=1}^{M} \Delta_j \exp(-t_j/\tau_n) \equiv 0 \tag{14}$$

where H_n is the spectrum line under consideration. Introducing Eq. (12) into (14) and using the abbreviations

$$E_{j,n} = \exp(-t_j/\tau_n) \tag{15}$$

$$E_{j,i} = \exp(-t_j/\tau_i) \tag{16}$$

we single out the nth spectrum line to obtain

$$\sum_{j=1}^{M} \left\{ \mathbf{G}(t_j) - \left[\{G_e\} + \sum_{i=1}^{i=n-1} H_i E_{j,i} + H_n E_{j,n} + \sum_{i=n+1}^{i=N} H_i E_{j,i} \right] \right\} E_{j,n} \equiv 0. \tag{17}$$

Introducing the further abbreviations

$$S_1 = \sum_{j=1}^{M} [\mathbf{G}(t_j) - \{G_e\}] E_{j,n}, \tag{18}$$

$$S_2 = \sum_{j=1}^{M} \left[\sum_{i=1}^{i=n-1} H_i E_{j,i} \right] E_{j,n} \tag{19}$$

and

$$S_3 = \sum_{j=1}^{M} \left[\sum_{i=n+1}^{i=N} H_i E_{j,i} \right] E_{j,n}, \tag{20}$$

the expression for the nth spectrum line becomes

$$H_n = \frac{S_1 - S_2 - S_3}{\text{Denom}} \tag{21}$$

where

$$\text{Denom} = \sum_{j=1}^{M} [E_{j,n}]^2. \tag{22}$$

We now proceed to pinpoint the source of the instability that renders the inversion of this algorithm an ill-posed problem. To do this, we split Eqs. (18), (19), and (20) into three parts. The first part will contain the data that lie to the left of the window, the second part those within the window, and the third part those to the right of the window. This leads to

$$S_1 = \sum_{j=1}^{j=k-l} [\mathbf{G}(t_j) - \{G_e\}] E_{j,n} + \sum_{j=k-l+1}^{j=k+u} [\mathbf{G}(t_j) - \{G_e\}] E_{j,n} + \sum_{j=k+u+1}^{j=M} [\mathbf{G}(t_j) - \{G_e\}] E_{j,n} \tag{23}$$

$$S_2 = \sum_{j=1}^{j=k-l} \left[\sum_{i=1}^{i=n-1} H_i E_{j,i} \right] E_{j,n} + \sum_{j=k-l+1}^{j=k+u} \left[\sum_{i=1}^{i=n-1} H_i E_{j,i} \right] E_{j,n} + \sum_{j=k+u+1}^{j=M} \left[\sum_{i=1}^{i=n-1} H_i E_{j,i} \right] E_{j,i} \tag{24}$$

and

$$S_3 = \sum_{j=1}^{j=k-l} \left[\sum_{i=n+1}^{i=N} H_i E_{j,i} \right] E_{j,n} + \sum_{j=k-l+1}^{j=k+u} \left[\sum_{i=n+1}^{i=N} H_i E_{j,i} \right] E_{j,n} + \sum_{j=k+u+1}^{j=M} \left[\sum_{i=n+1}^{i=N} H_i E_{j,i} \right] E_{j,n} \tag{25}$$

for the numerator and to

$$\text{Denom} = \sum_{j=1}^{j=k-l} [E_{j,n}]^2 + \sum_{j=k-l+1}^{j=k+u} [E_{j,n}]^2 + \sum_{j=k+u+1}^{j=M} [E_{j,n}]^2 \tag{26}$$

for the denominator in Eq. (22).

In all four expressions the three terms entail data along $G(t)$ that lie to the left, within, and to the right of the window. The symbols l and u indicate the number of datum points contained in the lower and upper portion of the window.

Let us now consider a particular spectrum line, $H(\tau_n)$, located within the window so that $\log \tau_{n,l} < \log \tau_n < \log \tau_{n,u}$. Outside of the window the exponential function emulates the cut-off properties of the unit step function so that we may set

$$\exp(-t_j/\tau_n)|_{t_j > \tau_{n,u}} = E_{j,n}|_{t_j > \tau_{n,u}} \simeq 0 \tag{27}$$

and

$$\exp(-t_j/\tau_n)|_{t_j < \tau_{n,l}} = E_{j,n}|_{t_j > \tau_{n,l}} \simeq 1. \tag{28}$$

Thus, when the value of the kernel function becomes smaller than the experimental error, it is indistinguishable from 0 on the right hand side, and is equally indistinguishable from 1 on the left hand side. Utilizing these equations, we recast Eq. (23) to (26) as

$$S_1 = \sum_{j=1}^{j=k-l} [\mathbf{G}(t_j) - \{G_e\}] + \sum_{j=k-l+1}^{j=k+u} [\mathbf{G}(t_j) - \{G_e\}]\, E_{j,n} \tag{29}$$

$$S_2 = \sum_{j=1}^{j=k-l} \left[\sum_{i=1}^{i=n-1} H_i\, E_{j,i} \right] + \sum_{j=k-l+1}^{j=k+u} \left[\sum_{i=1}^{i=n-1} H_i\, E_{j,i} \right] E_{j,n} \tag{30}$$

$$S_3 = (k-l) + \sum_{i=n+1}^{i=N} H_i + \sum_{j=k-l+1}^{j=k+u} + \left[\sum_{i=n+1}^{i=N} H_i\, E_{j,i} \right] E_{j,n} \tag{31}$$

and

$$\text{Denom} = k - l + \sum_{j=k-l+1}^{j=k+u} [E_{j,n}]^2. \tag{32}$$

The sum in Eq. (32) contains the contribution within the window. This is always small in comparison to $k-l$. Since $E_{j,n}$ is always smaller

than 1, and $k-l$ is always larger, the value of the denominator is essentially equal to the number of datum points to the left of the window. Using Eqs. (29) to (31), Eq. (21) thus becomes

$$H_n = \frac{1}{\text{Denom}}\left[\sum_{j=1}^{j=k-l}[\mathbf{G}(t_j) - \{G_e\}] + \sum_{j=k-l+1}^{j=k+u}[\mathbf{G}(t_j) - \{G_e\}]\,E_{j,n}\right]$$
$$- \frac{1}{\text{Denom}}\left\{\sum_{j=1}^{j=k-l}\left[\sum_{i=1}^{i=n-l} H_i\,E_{j,i}\right] + \sum_{j=k-l+1}^{j=k+1}\left[\sum_{i=1}^{i=n-l} H_i\,E_{j,i}\right]E_{j,n}\right\} \quad (33)$$
$$- \frac{1}{\text{Denom}}\left\{(k-1)\sum_{i+n-1}^{j=N} H_i + \sum_{j=k-l+1}^{j=k+u}\left[\sum_{i=n+1}^{i=N} H_i\,E_{j,i}\right]E_{j,n}\right\}.$$

Now, each experimental datum point can be expressed as

$$\mathbf{G}(t_j) = G(t_j) + \Delta_j, \quad (34)$$

where $G(t_j)$ represents the value of the experimental datum point considered to be free of error, and Δ_j denotes the experimental (and computational) errors that encumber $\mathbf{G}(t_j)$. Equation (34) can therefore be recast in the form

$$H_n = \frac{1}{\text{Denom}}\left[\sum_{j=1}^{j=k-l}[G(t_j) - \{G_e\}] + \sum_{j=k-l+1}^{j=k+u}[G(t_j) - \{G_e\}]E_{j,n}\right]$$
$$- \frac{1}{\text{Denom}}\left\{\sum_{j=1}^{j=k-l}\left[\sum_{i=1}^{i=n-l} H_i\,E_{j,i}\right] + \sum_{j=k-l+1}^{j=k+u}\left[\sum_{i=1}^{i=n-1} H_i\,E_{j,i}\right]E_{j,n}\right\}$$
$$- \frac{1}{\text{Denom}}\left\{(k-1)\sum_{i=n+1}^{i=N} H_i + \sum_{j=k-l+1}^{j=k+u}\left[\sum_{i=n+1}^{i=N} H_i\,E_{j,i}\right]E_{j,n}\right\} \quad (35)$$
$$+ \frac{1}{\text{Denom}}\left[\sum_{j=1}^{j=k-l}\Delta_j + \sum_{j=k-l+1}^{j=k+u}\Delta_j\,E_{j,n}\right]$$

i.e.,

$$H_n = \bar{H}(\tau_n) + \frac{1}{\text{Denom}}\sum_{j=k-l+1}^{j=k+u}\Delta_j\,E_{j,n} + \frac{1}{\text{Denom}}\sum_{j=1}^{j=k-l}\Delta_j. \quad (36)$$

Here $\bar{H}(\tau_n)$ represents the true value of the spectrum line, i.e., the first three terms on the right hand side of Eq. (35). The next term represents the error within the window and this is generally small in comparison with $\bar{H}(\tau_n)$. The last term contains the average error in the data to the left of the window. This average error is rather large.

Equation (36) reveals the source of instabilities that arise upon inversion in the presence of experimental error. They arise from the last term in Eq. (36). Calculating a spectrum line H_n from the complete set of datum points will invariably contain the average experimental error from the data represented by that last term. Since the data on the left hand side can be up to 10^3 to 10^4 times larger than data on the right hand side, a 10% error will be vastly larger then the spectrum line that is being evaluated. Thus, the last term in Eq. (36) will dominate the first two. If the average error is negative, H_n may even become negative.

4. THE WINDOWING (FILTERING) APPROACH

We have shown that instabilities in the inversion of equations like Eqs. (1) or (2) can be avoided if only data within the window are used to calculate any given spectrum line. Thus, we need to exclude from the inversion algorithm those regions of the data in which any potentially useful information is overwhelmed by experimental and/or computational errors. Using the windowing approach to omit these datum points from the calculations provides the "filtered" nth spectrum line as

$$H_n = \frac{1}{\text{Denom}} \sum_{j=k-l+1}^{j=k+u} \left[\mathbf{G}(t_j) - \{G_e\} - \sum_{i=1}^{j=n-1} H_i E_{j,i} - \sum_{i=n+1}^{i=N} H_i E_{j,i} \right] E_{j,n} \tag{37}$$

where the denominator is now given by

$$\text{Denom} = \sum_{j=k-l+1}^{j=k+u} [E_{j,n}]^2. \tag{38}$$

This is essentially the algorithm that we had developed in the first of our papers [1]. The present paper deepens our understanding of why it is so simple and yet so effective. Although it is a least squares method (we call this the least squares *absolute error* approach), the calculations prove to be stable if the experimental error is not too severe. To contend with more serious experimental errors we found it expedient [3] to minimize with respect to the relative error (we call this the least squares *relative error* approach).

To demonstrate on hand of data that the procedure becomes unstable upon using data from the regions to the left and to the right of the window, i.e., the regions where Eqs. (27) and (28) apply, we use the synthetic 24-line spectrum displayed in Figure 3 that we had used earlier [3].

From this spectrum, we obtained source data, $G(t)$, using

$$G(t) = G(e) + \sum_{i=1}^{24} H_i \exp(-t/\tau_i) \tag{39}$$

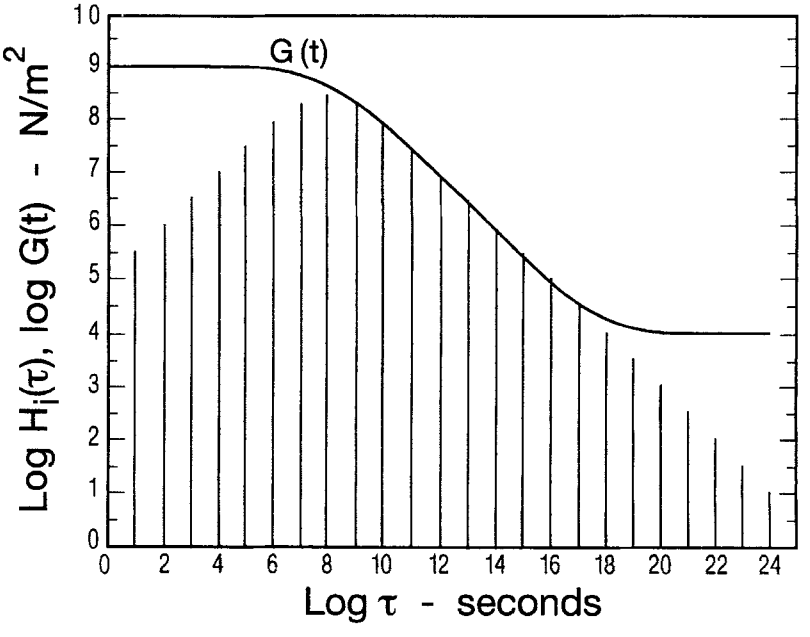

FIGURE 3 24-Line synthetic spectrum and $G(t)$ calculated from it.

where $G_g = 10^9$, $G_e = 10^4 \text{N/m}^2$, and the H_i are the spectral strengths tabulated in the second column of Table I under the heading of H_i.

Applying the Windowing Approach to these data produced the spectrum whose lines are shown in Table I under the heading H_i (ret) where (ret) marks these as *retrieved* values. Comparison of the two data columns shows that the algorithm satisfactorily reproduces the original spectrum. Now, if our theory is correct, then progressive widening of the window (or, in other words, widening of the pass band of the filter) should result in the emergence of some form of instability. Since our 24-line spectrum is synthetic, it is essentially free of experimental error. Any appearance of instabilities therefore demonstrates that these will arise even in the absence of experimental error because of the finite resolving power of the computer.

The width of our window extended from $\log \lambda_{k,l} = -0.6$ to $\log \lambda_{k,u} = 0.6$. Because of the "goodness" of our synthetic data combined with the use of double precision in the calculations, instabilities showed up only when the width ranged from $\log \lambda_{k,l} = -4.9$ to $\log \lambda_{k,u} = 4.9$. No instability appeared with a window that extended from $\log \lambda_{k,l} = -4.8$ to $\log \lambda_{k,u} = 4.8$. The onset was thus rather sudden. This is as it should be because, as already mentioned, with synthetic data instabilities may be expected to show up only when the resolving power of the computer is exceeded.

The "unstable" spectrum is represented in Figure 4 by the solid lines. Windows extending from $\log \lambda_{k,l} = -5.0$ to $\log \lambda_{k,u} = 5.0$, and

TABLE I Input and retrieved values of the 24-line synthetic spectrum (Relaxation times, τ, are in seconds, spectral strengths, H_i, in N/m^2)

$\log \tau_i$	$\log H_i$	$\log H_i(ret)$	$\log \tau_i$	$\log H_i$	$\log H_i(ret)$
1	5.56098	5.56206	13	6.55851	6.55851
2	6.06039	6.05983	14	6.06039	6.06039
3	6.55851	6.55885	15	5.56098	5.56100
4	7.05261	7.05252	16	5.06117	5.06112
5	7.53421	7.53422	17	4.56123	4.56137
6	7.97847	7.97847	18	4.06124	4.06080
7	8.32259	8.32259	19	3.56125	3.56256
8	8.45919	8.45919	20	3.06125	3.05734
9	8.32259	8.32259	21	2.56125	2.57257
10	7.97847	7.97847	22	2.06125	2.02682
11	7.53421	7.53421	23	1.56125	1.64320
12	7.05261	7.05261	24	1.06125	0.94864

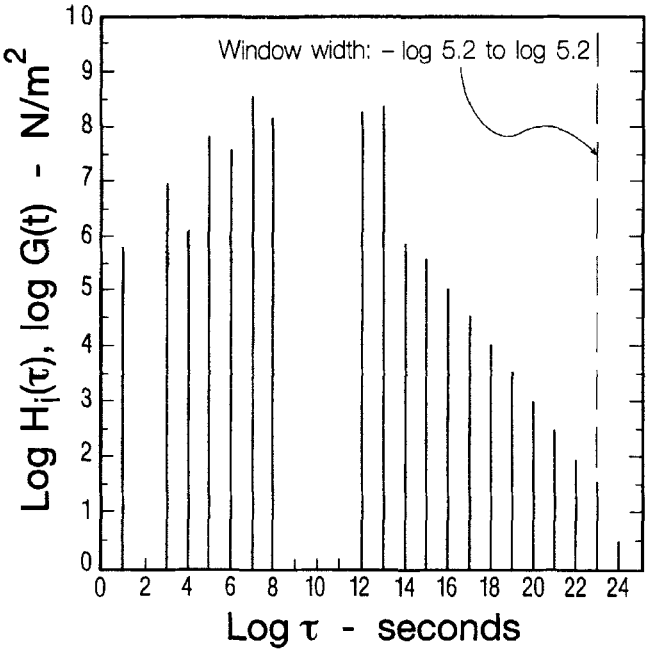

FIGURE 4 Line spectrum calculated with different widths of Window I.

from log $\lambda_{k,l} = -5.1$ to log $\lambda_{k,u} = 5.1$ produced spectra that were identical with that shown. However, letting the window range from log $\lambda_{k,l} = -5.2$ to log $\lambda_{k,u} = 5.2$ produced the singlet spectrum represented by the broken line in Figure 4. Further widening of the window from log $\lambda_{k,l} = -5.5$ to log $\lambda_{k,u} = 5.5$, and from log $\lambda_{k,l} = -6.0$ to log $\lambda_{k,u} = 6.0$ reproduced the same singlet spectrum.

These observations are specific for the (double) precision that was available in our computer. With another precision the outcome might be somewhat different. We have no doubt that the presence of even moderate experimenal error in non-synthetic data would have resulted in more severe instabilities and these would, in addition, have shown up with narrower windows than those where we found them in the results reported above. We do not deem it necessary to repeat the exercise on non-synthetic experimental data since we had already shown that our algorithm produced entirely satisfactory results even in

the presence of the errors normally encountered [3] in experimental data.

5. DISCUSSION

An understanding of this paper requires a brief summary of the mechanics of the algorithm whose specific forms have been described earlier [1–4]. These earlier publications should be consulted for details. We mention, however, that the number, N, of spectrum lines depends on M, the number of the available datum points. We found $N=2$ to be a convenient number. Increasing this number decreases the number of datum points available per logarithmic decade of the relaxation times, τ. The choice of τ_1 is somewhat arbitrary since it simply shifts the spectrum along the abscissa. A convenient choice is $\log \tau_1 = \log t_1$ where $\log t_1$ is the location of the first datum point.

Initially, the data are normalized and all spectrum lines are set to zero. We remark that the sum of the normalized spectrum lines must always add up to 1. In the first sweep through the data we start at the right-hand-side of the spectrum and calculate a provisional value for each line. We use linear regression to calculate the value, using only data within the window pertaining to that line. Instead of linear regression it would be possible to use Tikhonov regularization [6] (see Section 5.1) to refine the procedure but this does not seem to us to be necessary with the experimental errors commonly encountered.

Next, we begin iterating to refine the provisional values. We once more start on the right-hand-side and sweep over four consecutive lines to the left. Since we are using 2 lines per decade, the fifth line to the left will be two decades away from the line under consideration. Its contribution will therefore be negligible ($\sim 3.72\times10^{-44}$). In calculating each line we again use only data within its window. If any line becomes negative during the iteration, it is set to zero and the iteration is repeated. Negative spectrum lines must be disallowed since such lines have no physical meaning. The iteration is abandoned when the changes in the line under consideration are smaller than the preset exit criterion. After a run of iterations is completed, we move left to the next line and repeat the procedure, finishing eventually with the first line on the left-hand-side of the spectrum.

To obtain the results reported here, we used the relaxation modulus, $G(t)$, featuring the exponential (transient) kernel function.

$$K(t, \tau_i) = \exp(-t/\tau_i). \tag{40}$$

However, our discussion of the mechanics of the algorithm is equally valid for any other experimental response function whose kernel function has a shape identical with, or similar to, $K(t, \tau_i)$. Such response functions are the creep compliance,

$$J(t) = J_e^{\{o\}} - \sum_{i=1}^{i=N} L_i \exp(-t/\tau_i), \tag{41}$$

the storage modulus,

$$G'(\omega) = \{G_e\} + \sum_{i=1}^{i=N} H_i \frac{\omega^2 \tau_i^2}{1 + \omega^2 \tau_i^2}, \tag{42}$$

and the loss modulus

$$G''(\omega) = \sum_{i=1}^{i=N} H_i \frac{\omega^2 \tau_i}{1 + \omega^2 \tau_i^2}, \tag{43}$$

as well as their compliance counterparts, the storage compliance,

$$J'(\omega) = J_e^{\{o\}} - \sum_{i=1}^{i=N} L_i \frac{\omega^2 \tau_i^2}{1 + \omega^2 \tau_i^2}, \tag{44}$$

and the loss compliance

$$J''(\omega) = \sum_{i=1}^{i=N} L_i \frac{\omega \tau_{\mathrm{I}}}{1 + \omega^2 \tau_{\mathrm{I}}^2}, \tag{45}$$

The (operational) impulse responses, i.e., the relaxance,

$$\bar{Q}(s) = \{G_e\} + \sum_{\mathrm{I},=1}^{\mathrm{I},=N} H_{I,} \frac{\tau_{\mathrm{I}}, s}{1 + \tau_{\mathrm{I}}, s}, \tag{46}$$

and the retardance,

$$\overline{U}(s) = J_e^{\{o\}} - \sum_{i=1}^{i=N} L_i \frac{\tau_i s}{1 + \tau_i s}. \tag{47}$$

qualify also. In Eqs. (41), (44) and (45) the L_i are compliance values representing the strengths associated with a particular retardation time, τ_i.

The response to steady-state sinusoidal excitations contain the storage and loss kernels

$$K'(\omega, \tau_i) = \frac{\omega^2 \tau_i^2}{1 + \omega^2 \tau_i^2} \qquad \text{and} \qquad K''(\omega, \tau_i) = \frac{\omega \tau_i}{1 + \omega^2 \tau_i^2}, \tag{48}$$

where ω is the radian frequency, while responses to impulse excitation are characterized by the Laplace kernel

$$K(s, \tau_i) = \frac{\tau_i s}{1 + \tau_i s}. \tag{49}$$

where s is the Laplace transform variable. The kernels of the exponential, storage and impulse response functions share the property that they decay from a value of unity to a value of zero as $t \to 0$, $\omega, s \to \infty$, respectively, as $t \to \infty$, and $\omega, s \to 0$. $K''(\omega, \tau_i)$ possesses a maximum but can, nevertheless, be handled in a similar way to the others by operating separately on the two sides of the maximum [2]. The modifications in the computational methods necessary when using the non-exponential kernels have been described in earlier publications [2–4]. We remark that the speed with which convergence in the iteration is attained in our approach depends significantly on the nature of the kernel. The exponential kernel assures the most rapid convergence because its descent to zero is the steepest. The storage and the loss kernels descend progressively more slowly. Thus they require broader windows and need more time for convergence. Finally, the Laplace kernel is the least steep and the slowest to converge.

We emphasize that the algorithm is not restricted to the linear viscoelastic response functions. It operates equally well on other experimental responses as long as their kernel functions behave as just

described. Response functions that immediately spring to mind are the dielectric response functions.

We remark that a continuous spectral distributions can be obtained from a line spectrum by a rather simple calculation proposed by Baumgaertel and Winter [8].

Two additional aspects require attention. The first is concerned with other approaches within the realm of linear viscoelasticity designed to effectively circumvent the instabilities inherent in the inversion. The second takes a look in the light of the Windowing Approach at the traditional methods of deriving approximations to the spectra through graphical differentiation or finite difference calculus.

5.1. Other Approaches

In contrast to ours, all other methods designed to overcome the ill-posedness of the inversion problem make use of the complete set of data. In order to quash the data in trouble-causing regions, they thus require methods more sophisticated than the least squares (linear regression) approach we use.

Of such methods the work of Honercamp and coworkers [9–14] is probably the most significant. In essence, these workers use Tikhonov regularization [6] in which weighting functions effectively suppress those data that our approach avoids entirely. Their method is efficacious but appears to us to require an unnecessary amount of effort resulting in slow conversion. In addition, and perhaps more significantly, their methods require *a priori* estimates of experimental errors. Recently, Yanovsky, Basistov, and Siginer [15] have pointed out some of the shortcomings of all methods based on Tikhonov regularization, such as the need for having *a priori* estimates of experimental errors. Their new method is free from this requirement (as is ours).

The approach of Baumgaertal and Winter [16] appears to avoid instabilities by judicially positioning the spectrum lines. This is not only time consuming but necessarily produces unequally spaced spectra that introduce "steps" in the reconstructed response functions where no physical reasons for these steps exist.

More recently, Mead [17] has developed an algorithm for use with rheodictic materials. His algorithm does not use error bounds as the

a priori information required to solve ill-posed problems. Instead, it makes use of the relations

$$\eta_f = \sum_{i=1}^{n} G_i \tau_i \tag{50}$$

and

$$J_e^o \eta_f^2 = \sum_{i=1}^{n} G_i \tau_i^2. \tag{51}$$

This algorithm will clearly only succeed if η_f and J_e^o have been determined independently, or if the data span a sufficiently large portion of the logarithmic time scale to extend effectively from minus to plus infinity. Synthetic data can satisfy the latter condition. Experimental data will almost never cover a wide enough range and the summations will give incorrect estimates [18] of η_f and J_e^o.

We remark finally that all methods based on matrix inversion such as the collocation method [19] and improvements thereof such as the multidata method [7] will work more or less successfully only with data that are effectively free of experimental error.

5.2. Traditional Methods

Turning attention now to the traditional "approximation" methods [5], the first thing to recognize is that these methods are not concerned with line spectra but furnish discrete data sets that represent continuous approximations to the spectra. The most common method of obtaining approximations to a continuous linear viscoelastic spectrum from experimental data is through logarithmic differentiation. Thus, the first approximation to the relaxation spectrum, $H_1(\tau)$ is obtained from

$$H_1(\tau) = -\left.\frac{dG(t)}{d\log t}\right|_{t=\tau}. \tag{52}$$

The reason that this approach works at all is due to the fact that each point of the approximation is obtained from a single experimental

datum point. This calculation is *local*, representing, as it were, a window of zero width. Consecutively higher approximations can – in principle – be obtained by calculating logarithmic derivatives of successively higher order. Each order produces a bell-shaped spectrum of successively narrower width. This is easily verified on hand of an analytic expression that is infinitely differentiable in principle, such as the equation for the so-called standard linear model or 3-parameter Maxwell model given by [5]

$$G(t) = G_e + G \exp(-t/\tau). \tag{53}$$

Only a derivative of infinite order produced in a computer of infinite resolving power – if such were possible – would produce a "true" continuous spectrum. From the windowing point of view the important aspect is that these calculations would always remain *local*, i.e, within a window of zero width.

The situation is different with the methods based on finite difference calculus [5]. Here, the approximation of first order to the relaxation spectrum is obtained from

$$H_{1h}(\tau/\beta_{1h}) = \left.\frac{G(t) - G(ht)}{\ln h}\right|_{t=\tau} \tag{54}$$

where $\log h$ is a suitably chosen spacing on the $\log t-$ axis, and β_{1h} is called a spectral shift factor. Each approximation of successively higher order requires successively more data to the left and right of the point at which the spectrum is calculated. Thus, the approximation of the second order becomes

$$H_{2h}(\tau/\beta_{2h}) = \left.\frac{hG(t) - (h+1)G(ht) + G(h^2t)}{(h-1)\ln h}\right|_{t=\tau}. \tag{55}$$

With each higher order of approximation there is thus a loss of information at the ends of the data set. In addition, an increase in the order of differentiation is tantamount to a widening of the effective window and would eventually necessarily lead to the appearance of instabilities, quite apart from reducing the width of the spectrum.

6. CONCLUDING REMARKS

The difference between the windowing (filtering) and the various regularization approaches used by others lies in the way in which the troublesome regions are weighted. A window in the filtering approach constitutes effectively a box-shaped weighting function. Since this information is part of the algorithm itself, no additional information needs to be supplied by the user. Thus, the windowing approach may be seen to be equivalent to a fast, simplified Tikhonoff regularization.

References

[1] Emri, I. and Tschoegl, N. W. (1993). *Rheol. Acta*, **32**, 31–32.
[2] Tschoegl, N. W. and Emri, I. (1993). *Rheol. Acta*, **32**, 322–327.
[3] Emri, I. and Tschoegl, N. W. (1994). *Rheol. Acta*, **33**, 60–70.
[4] Tschoegl, N. W. and Emri, I. (1992). *Intern. J. Polym. Mater.*, **18**, 117–127.
[5] Tschoegl, N. W. (1989). The Phenomenological Theory of Linear Viscoelastic Behavior (Springer-Verlag, Heidelberg).
[6] (a) Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solution of Ill-Posed Problems*, (V. H. Winston, Washington, D. C.). (b) Groetsch, C. W. (1993). *Inverse Problems in Mathematical Science*, Vieweg, Wiebaden.
[7] Cost, T. L. and Becker, E. B. (1970). *Intern. J. Num. Methods in Engg.*, **2**, 207.
[8] Baumgaertel, M. and Winter, H. H. (1992). *J. Non-Newtonian Fluid Mech.*, **44**, 15–30.
[9] Elster, C. and Honercamp, J. (1991). *Macromolecules*, **24**, 310–314.
[10] Elster, C. and Honercamp, J. (1992). *J. Rheology*, **36**, 911–927.
[11] Elster, C., Honercamp, J. and Weese, J. (1991). *J. Rheol. Acta*, **31**, 161–174.
[12] Honercamp, J. (1989). *Rheol. Acta*, **28**, 363–371.
[13] Honercamp, J. and Weese, J. (1989). *Macromolecules*, **22**, 4372–4377.
[14] Honercamp, J. and Weese, J. (1993). *Rheol. Acta*, **32**, 65–73.
[15] Yanovsky, Yu. G., Basistov, Yu. A. and Siginer, D. A. (1996). *Int. J. Eng. Sci.*, **34**, 1221–1245.
[16] Baumgaertel, M. and Winter, H. H. (1989). *Rheol. Acta*, **28**, 511–519.
[17] Mead, D. W. (1994). *J. Rheology*, **38**, 1769–1795.
[18] Emri, I. and Tschoegl, N. W. (1993). *J. Rheology*, **37**, 1103–1116.
[19] Schapery, R. A. (1962). *Proc. Fourth U.S. Nat. Congr. Appl. Mech.*, **2**, 107.