On the inference of viscoelastic time constants from stress relaxation experiments

Kumar Vemaganti[a,*], Sandeep Madireddy[b]

[a] *Department of Mechanical and Materials Engineering, University of Cincinnati, Cincinnati, Ohio 45221*
[b] *Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois 60439*

Abstract

Several constitutive theories have been proposed in the literature to model the viscoelastic response of soft tissue, including widely used rheological constitutive models. These models are characterized by certain parameters ("time constants") that define the time scales over which the tissue relaxes. These parameters are primarily obtained from stress relaxation experiments using curve-fitting techniques. However, the question of how best to estimate these time constants remains open.

As a step towards answering this question, we develop an optimal experimental design approach based on ideas from information geometry, namely Fisher information and Kullback-Leibler divergence. Tissue is modeled as a standard linear solid and described using a one- or two-term Prony series. Treating the time constants as unknowns, we develop expressions for the Fisher information and Kullback-Leibler divergence that allow us to maximize information gain from experimental data. Based on the results of this study, we propose that the largest time constant estimated from a stress relaxation experiment for a linear viscoelastic material should be at most one-fifth of the total time of the experiment in order to maximize information gain.

Keywords: Viscoelastic, Time constant, Optimal experimental design, Fisher information, Kullback-Leibler divergence

1. Introduction

Soft tissues are generally considered to be viscoelastic and several constitutive theories have been proposed and used in the literature to model their mechanical response. These include linear viscoelasticity, quasilinear viscoelasticity, and nonlinear viscoelasticity [1, 2, 3, 4, 5, 6, 7, 8, 9]. Linear and quasilinear viscoelastic theories based on rheological spring-dashpot models (Fig. 1) are extensively used in the literature [10]. These models consist of a spring of stiffness E_0 in parallel with several spring-dashpot branches. The ratio of the spring

*Corresponding author

Email addresses: `Kumar.Vemaganti@uc.edu` (Kumar Vemaganti), `smadireddy@mcs.anl.gov` (Sandeep Madireddy)

stiffness (E_k) to the dashpot damping coefficient (η_k) in a branch is the "time constant" τ_k associated with the branch, and represents one of the possibly many time-scales at which the material relaxes.

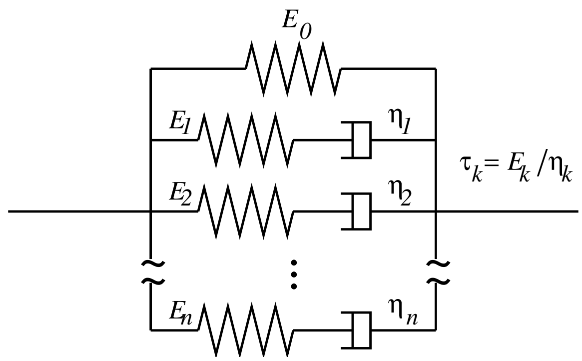


Figure 1: The standard linear solid model of linear viscoelastic materials.

The parameters of the model (E_0, E_k, τ_k) are generally estimated from stress relaxation experiments, where a tissue sample is subjected to a step strain and the stress response is recorded. Various approaches have been proposed in the literature as to how to choose the number of time constants and bounds on these constants. For instance, some authors use one time constant per decade of the time scale of interest, while others carry out unconstrained optimization [11]. Often, the number of time constants is selected based on the goodness of fit and the analyst's experience. A more detailed discussion of various parameter estimation methods for viscoelastic materials may be found in the work of [12] and the references cited therein.

Generally speaking, the characterization of viscoelastic materials in the literature has been uni-directional, i.e., experiments are first conducted and a suitable viscoelastic model is then used to obtain the best fit to the data. But the question of designing relaxation experiments in order to best inform the relaxation model remains an open one. Our overall goal in this work is to understand the relationship between the duration of the relaxation experiment and the time constants we seek to infer from the relaxation data. *In particular, we ask: how long should a relaxation experiment last so that we gain the most information about the time constants of a linear viscoelastic material*? This is posed as an optimal experimental design problem in which a utility function that depends on the goal of the experiment is maximized to obtain the optimal configuration of experimental parameters.

There are two broad classes of methods that are employed to perform optimal experimental design. The first one is the classical approach where the utility function is based on the Fisher information matrix (e.g., [13, 14]), which is well suited for linear models. For nonlinear models, additional simplifications and assumptions (such as linearization of the forward model or Gaussian approximation of the utility function) are needed. These assumptions result in a locally optimal design.

The other class of methods come under Bayesian optimal design [15, 16, 17] in which the utility function is based on the posterior distributions of the quantities of interest. The roots of the this class of methods can be traced back to Lindley and Smith [18] who present a decision-theoretic approach to experimental design, upon which Bayesian experimental design is based. Some of the commonly used utility functions include mutual information based on entropy, Kullback-Leibler (KL) divergence, squared loss error, etc.; see [17] for a detailed discussion. Some of the methods used to evaluate this utility function in the Bayesian context include Laplace approximation, numerical quadrature, and Monte Carlo approximation.

In this study we employ both Fisher information from classical experimental design and Kullback-Leibler divergence from Bayesian experimental design for linear viscoelastic materials. For the Fisher information case, we use analytical and numerical approaches to maximize the information gained about time constants of a linear viscoelastic material. Also, the Kullback-Leibler divergence between the prior and posterior distributions of the time-constant parameter is calculated numerically for such a material. Our results show that once the duration of the relaxation experiment exceeds five times the largest time constant of a linear viscoelastic material, there is no further information gain and additional stress relaxation data are of little use. Put another way, from the point of view of maximizing information gain, the largest time constant estimated from a stress relaxation experiment should not exceed one-fifth of the duration of the experiment. The approach adopted here can be extended to generalized linear viscoelastic, and nonlinear viscoelastic models.

The paper is structured as follows. Following this introduction, section 2 describes the linear viscoelastic material constitutive model used in this study. In section 3, we introduce the necessary ideas from classical and Bayesian optimal experimental design, and then the calculation of Fisher information and KL divergence in section 4. Results and discussion are presented in Section 5 and 6, respectively, followed by some concluding remarks.

2. Materials and Methods

For a linear viscoelastic material, the stress response for a step strain history corresponding to a relaxation experiment can be written as

$$\sigma(t) = \epsilon_0 \mathcal{G}\{(t,s)|_{s=-\infty}^{s=t}\}, \tag{1}$$

where ϵ_0 is the instantaneous strain, and $\mathcal{G}$ is the relaxation function. Equation (1) can be written in an integral form using the Riesz theorem [2] as

$$\sigma(t) = \int_{-\infty}^{t} \mathcal{G}(t-s) \ \mathrm{d}\epsilon(s). \tag{2}$$

In practice, a step strain history is difficult to achieve. A more realistic strain history for a relaxation experiment consists of an initial ramp up to ϵ_0 at a constant strain rate followed by a long hold, as shown in Fig. 2. This deformation leads to a rapid increase in the stress in the initial loading stage, following which the material relaxes to a constant equilibrium stress. The rate at which the material is deformed in the loading stage generally depends on the available experimental setup.

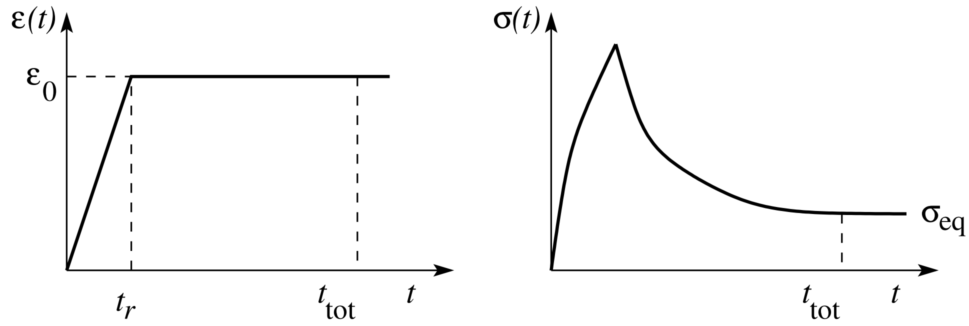


Figure 2: Stress relaxation experiment: a constant-rate strain load is applied over a time period t_r and then held constant (left), and the resulting stress response is recorded for the duration of the experiment, t_{tot} (right).

When the loading in the relaxation experiment is done at a finite strain rate $\dot{\epsilon}$, Eqn. (2) is modified to be

$$\sigma(t) = \int_{-\infty}^{t} \mathcal{G}(t-s)\,\dot{\epsilon}(s)\;\mathrm{d}s. \tag{3}$$

The relaxation function in this study is chosen to be that of the well known standard linear solid (SLS) with several spring-dashpot elements in parallel with a spring, as shown in Fig. 1. For this configuration, we can write

$$\mathcal{G}(t) = E_0 + \sum_{i=1}^{N} E_i \exp\left(-\frac{t}{\tau_i}\right), \tag{4}$$

where E_0 and E_i represent the spring stiffness values, τ_i are the time constants associated with the dashpot elements, and N is the number of spring-dashpot elements. This relaxation function is also called the Prony series in the literature.

We assume that the duration of the loading stage in the relaxation experiment is t_r and the total time of the experiment is t_{tot}. This is shown in Fig. 2. Assuming a constant loading rate $\dot{\epsilon}$, the corresponding expressions for the stress are derived using Eqn. (3) as follows

$$\sigma(t) = \begin{cases} \dot{\epsilon}\displaystyle\int_0^t \mathcal{G}(t-s)\;\mathrm{d}s, & \text{for } 0 \le t \le t_r, \\ \dot{\epsilon}\displaystyle\int_0^{t_r} \mathcal{G}(t-s)\;\mathrm{d}s, & \text{for } t_r < t \le t_{\text{tot}}. \end{cases} \tag{5}$$

For a one-term Prony series ($N = 1$) with $\tau_1 = \tau$, we obtain

$$\sigma(t) = \begin{cases} f_1(t) = \dot{\epsilon}\left[E_0 t + E_1\tau\left(1 - \exp\left(-\dfrac{t}{\tau}\right)\right)\right], & \text{for } 0 \le t \le t_r, \\ f_2(t) = \dot{\epsilon}\left[E_0 t_r + E_1\tau\left(\exp\left(-\dfrac{(t-t_r)}{\tau}\right) - \exp\left(-\dfrac{t}{\tau}\right)\right)\right], & \text{for } t_r < t \le t_{\text{tot}}. \end{cases} \tag{6}$$

We now consider an optimal experimental design approach for the relaxation testing of a linear viscoelastic solid represented by the standard linear solid. Specifically, we seek to determine the conditions under which information gain about τ, as measured by Fisher information and KL divergence, is maximized.

3. Optimal Experimental Design Methodologies: Introduction

3.1. Fisher Information

Fisher information is a measure of the amount of information that a given experimental dataset carries about the unknown model parameters to be estimated from that experiment. The experimental data are represented by a random variable $\mathbf{d}$ and their relationship with the model parameter θ is established through a likelihood function $P(\mathbf{d}|\theta, I)$, which is the probability of the model predicting the observed experimental data $\mathbf{d}$ for a given value of the parameter θ and background information I. For instance, I might consist of information about physically meaningful values of the model parameter θ.

Roughly speaking, the less likely an event is to occur, the more information it provides. Conversely, the more likely an event is to occur, the less information it provides. This means that information is linked to the slope of the likelihood function, or, more commonly, the slope of the logarithm of the likelihood function, $\log(P)$.

Formally, for a one parameter model, the Fisher information is given by

$$\mathcal{I}(\theta) = \int_{\mathbf{d}} \left(\frac{\partial \log(P(x|\theta, I))}{\partial \theta} \right)^2 P(x|\theta, I) \, \mathrm{d}x, \tag{7}$$

which is the second moment of $\left(\frac{\partial \log(P(x|\theta,I))}{\partial \theta}\right)$ with respect to $P(x|\theta, I)$ for a given value of θ. This can be written as

$$\mathcal{I}(\theta) = \left\langle \left(\frac{\partial \log(P(x|\theta, I))}{\partial \theta} \right)^2 \right\rangle, \tag{8}$$

where $\langle \cdot \rangle$ represents the expectation of the argument with respect to $P(x|\theta, I)$ given θ. If certain regularity conditions of the likelihood are satisfied [19], an equivalent expression for the Fisher information is given by

$$\mathcal{I}(\theta) = -\left\langle \left(\frac{\partial^2 \log(P(x|\theta, I))}{\partial \theta^2} \right) \right\rangle, \tag{9}$$

which is the expectation of the curvature of the log-likelihood function. Intuitively, a log-likelihood function with a sharp peak has a higher Fisher information value than a function with a blunt peak. It is worth noting that the Fisher information depends on the value of the parameter θ, and if the true value of θ (assuming such a value exists) is unknown, then so is $\mathcal{I}(\theta)$. The higher the value of the Fisher information, the more information we gain about the model parameter θ from the data $\mathbf{d}$. Thus, the Fisher information serves as an objective criterion by which one can design experiments for specific models of interest in a physical phenomenon.

In many cases, the true value of the model parameter θ is unknown and we must resort to estimates of the Fisher information. One such approach to estimate the Fisher information is to compute the quantity

$$\mathcal{J}(\hat{\theta}) = - \left. \frac{\partial^2 \log\left[P(\mathbf{d}|\theta, I)\right]}{\partial\theta^2} \right|_{\theta=\hat{\theta}}, \tag{10}$$

where $\hat{\theta}$ is the maximum likelihood estimate (MLE) value of θ, i.e., the value of θ that maximizes the likelihood function $P(\mathbf{d}|\theta, I)$. The quantity $\mathcal{J}(\hat{\theta})$ is called the observed Fisher information at $\hat{\theta}$. Note that this quantity does not require the calculation of the expectation in Eqn. (9) and that $\mathcal{I}(\hat{\theta}) = \langle \mathcal{J}(\hat{\theta}) \rangle$.

For the case where the model is characterized by multiple parameters, i.e., θ is a vector, we can define a Fisher information matrix. If θ is an m-dimensional vector, then the observed Fisher information is an $m \times m$ matrix given by

$$(\mathcal{J}(\hat{\theta}))_{i,j} = - \left. \frac{\partial^2}{\partial\theta_i \partial\theta_j} \log\left[P(\mathbf{d}|\theta, I)\right] \right|_{\theta=\hat{\theta}}, \quad 1 \leq i, j \leq m. \tag{11}$$

The information contained in this matrix can be characterized by various estimators [20] that can help to optimize the design of an experiment and maximize information-gain about the parameters θ. These are termed optimality criteria. In this work, we use the D-optimality criterion, defined as the determinant of the Fisher information matrix given by Eqn. (11). This criterion results in maximizing the differential Shannon information content of the parameter estimates [21].

3.2. Relative Entropy or Kullback-Leibler Divergence

Another approach to characterize information is based on the concepts of relative entropy or Kullback-Leibler divergence. We now briefly describe these ideas.

If X is a discrete random variable and $\psi(x)$ is the value of its probability distribution at $X = x$, then the entropy of X is

$$H(X) = - \sum_{x \in X} \psi(x) \log_2 \psi(x). \tag{12}$$

Entropy measures the amount of information (or uncertainty) in a random variable. It is measured in bits and is always greater than or equal to zero. For a continuous random variable, the summation in Eqn. (12) is replaced by an integral. Given two continuous probability distributions, $\psi(x)$ and $\eta(x)$, the relative entropy of $\psi(x)$ with respect to $\eta(x)$, or the Kullback-Leibler divergence of $\psi(x)$ from $\eta(x)$ is given by

$$D_{KL}(\psi || \eta) = \int_{x \in X} \psi(x) \log_2 \left(\frac{\psi(x)}{\eta(x)} \right) \, \mathrm{d}x. \tag{13}$$

Note that $D_{KL}(\psi||\eta) \neq D_{KL}(\eta||\psi)$. The KL divergence essentially compares the entropy of the two distributions over the same random variable. When the logarithm is taken with respect to base 2, the obtained KL divergence is in the units of nats [22].

In the context of optimal Bayesian experimental design, the prior information on the parameters ($P(\theta|I)$) is updated in light of the experimental data $\mathbf{d}$ to obtain the posterior distribution of the parameters $P(\theta|\mathbf{d}, I)$. This is accomplished using the Bayes theorem

$$P(\theta|\mathbf{d}, I) = \frac{P(\mathbf{d}|\theta, I)P(\theta|I)}{P(\mathbf{d}|I)}, \tag{14}$$

where $P(\mathbf{d}|\theta, I)$ is the likelihood function introduced previously. The quantity in the denominator, $P(\mathbf{d}|I)$, is called the marginal likelihood or evidence.

The KL divergence between the prior and the posterior distributions of the parameters is obtained using the expression

$$D_{KL}(P(\theta|\mathbf{d}, I)||P(\theta|I)) = \int_{\theta\in\Theta} P(\theta|\mathbf{d}, I) \log_2 \left[\frac{P(\theta|\mathbf{d}, I)}{P(\theta|I)}\right] \, \mathrm{d}\theta. \tag{15}$$

In the context of experimental design for characterizing viscoelastic materials, we seek to maximize the information gained from the data. This information gain is characterized by the KL divergence between the prior and posterior distributions of the model parameter(s).

4. Optimal Experimental Design Methodologies: Calculation

4.1. Fisher Information

We now turn to the question of calculating the Fisher information and KL divergence for the specific case of viscoelastic material relaxation. We consider the rheological model introduced in Section 2, which is characterized by the parameters: E_0, E_i, $\tau_i, 1 \leq i \leq N$. For simplicity, we will consider the case where $N = 1$. Furthermore, we will assume that the parameters E_0 and E_1 are known, so that the uncertainty lies primarily in the lone time constant $\tau = \tau_1$. The question now becomes this: under what experimental conditions do we gain the most information about τ?

To answer this, we hypothesize that the mathematical model for the stress is related to the experimental data $\mathbf{d}$ through the equation

$$\sigma(t, \tau) = \mathbf{d} + \boldsymbol{\eta}, \tag{16}$$

where $\mathbf{d}$ is the hypothetical true data, and $\boldsymbol{\eta}$ is the noise that contributes uncertainty to the experiment. This noise generally arises from a number of sources: the sensors used to take the measurements, sample-to-sample variability, and sometimes uncertainty in the environmental conditions. Note that $\mathbf{d} + \boldsymbol{\eta}$ is the actual system response obtained from the experiment.

The measured response typically consists of the stress values at a number of times, say, $t_i, 1 \leq i \leq n$, and the measurement is carried out for multiple samples. For simplicity, we will assume that these data are available in the form of (a) a mean vector D_i representing the mean stress measured at time t_i, and (b) a vector of standard deviations Δ_i, representing the noise in the data at time t_i.

Based on maximum entropy considerations [23], we assume that the noise associated with experimental measurements follows a Gaussian process. Then the likelihood function for the i-th data point $d_i = \{D_i, \Delta_i\}$ can be written as

$$P(d_i|\tau) = \frac{1}{\sqrt{2\pi{\Delta_i}^2}} \exp\left[-\frac{(\sigma(t_i,\tau) - D_i)^2}{2{\Delta_i}^2}\right]. \tag{17}$$

If the n measured values are independent and identically distributed (i.i.d), the overall likelihood function is found using the product rule as

$$P(\mathbf{d}|\tau) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi{\Delta_i}^2}} \exp\left[-\frac{(\sigma(t_i,\tau) - D_i)^2}{2{\Delta_i}^2}\right]. \tag{18}$$

So, with the notation $L(\tau) = \log(P(\mathbf{d}|\tau))$, the logarithm of the likelihood is given by

$$L(\tau) = \log\left[P(\mathbf{d}|\tau)\right] = \sum_{i=1}^{n} -\frac{1}{2}\log\left(2\pi{\Delta_i}^2\right) - \sum_{i=1}^{n}\left[\frac{(\sigma(t_i,\tau) - D_i)^2}{2{\Delta_i}^2}\right]. \tag{19}$$

Making use of the separate expressions for the stress in the loading and hold phases of the experiment from Eqn. (6), we obtain

$$L(\tau) = \sum_{i=1}^{N} -\frac{1}{2}\log(2\pi{\Delta_i}^2) - \sum_{i=1}^{n_r}\left[\frac{(f_1(t_i,\tau) - D_i)^2}{2{\Delta_i}^2}\right] - \sum_{i=n_r+1}^{n}\left[\frac{(f_2(t_i,\tau) - D_i)^2}{2{\Delta_i}^2}\right], \tag{20}$$

where n_r is the number of data points in the loading phase of the experiment.

We are interested in the behavior of the log-likelihood function in the neighborhood of $\hat{\tau}$, the MLE of the time constant τ. We therefore consider a Taylor-series expansion of the log-likelihood about $\hat{\tau}$:

$$L(\tau) = L(\hat{\tau}) + \left.\frac{\partial L}{\partial \tau}\right|_{\hat{\tau}} (\tau - \hat{\tau}) + \frac{1}{2}\left.\frac{\partial^2 L}{\partial \tau^2}\right|_{\hat{\tau}} (\tau - \hat{\tau})^2 + \ldots, \tag{21}$$

and since $\left.\frac{\partial L}{\partial \tau}\right|_{\hat{\tau}} = 0$ by definition, we have

$$L(\tau) = L(\hat{\tau}) + \frac{1}{2}\left.\frac{\partial^2 L}{\partial \tau^2}\right|_{\hat{\tau}} (\tau - \hat{\tau})^2 + \ldots. \tag{22}$$

This essentially means that we are assuming the likelihood function $P(\mathbf{d}|\tau)$ to be Gaussian in the neighborhood of the MLE. Comparing the expression to the Gaussian distribution yields that the variance Δ of the distribution is related to the curvature of the log-likelihood as follows:

$$\Delta = \left(-\left.\frac{\partial^2 L}{\partial \tau^2}\right|_{\hat{\tau}}\right)^{-\frac{1}{2}}. \tag{23}$$

From the definition of the observed Fisher information in Eqn. (10), it follows that

$$J(\hat{\tau}) = -\left.\frac{\partial^2 \log\left[P(\mathbf{d}|\tau)\right]}{\partial \tau^2}\right|_{\tau=\hat{\tau}} = \Delta^{-2}. \tag{24}$$

We now consider two approaches to calculating the observed Fisher information and KL divergence.

4.1.1. Analytical approach

First, we develop a closed form solution for the Fisher information in Eqn. (24). Evaluating Eqn. (24) using Eqn. (20) gives

$$\begin{aligned} J(\hat{\tau}) = &\sum_{i=1}^{n_r} \frac{(f_1'(t_i,\hat{\tau}))^2}{\Delta_i^2} + \sum_{i=1}^{n_r} \frac{f_1''(t_i,\hat{\tau})(f_1(t_i,\hat{\tau}) - D_i)}{\Delta_i^2} \\ &+ \sum_{i=n_r+1}^{n} \frac{(f_2'(t_i,\hat{\tau}))^2}{\Delta_i^2} + \sum_{i=n_r+1}^{n} \frac{f_2''(t_i,\hat{\tau})(f_2(t_i,\hat{\tau}) - D_i)}{\Delta_i^2}. \end{aligned} \tag{25}$$

Assuming the model response matches the experimental average well, we have $(f(t_i,\hat{\tau}) - D_i) \approx 0$, so that Eqn. (25) can be written as,

$$J(\hat{\tau}) = \sum_{i=1}^{n_r} \frac{(f_1'(t_i,\hat{\tau}))^2}{\Delta_i^2} + \sum_{i=n_r+1}^{n} \frac{(f_2'(t_i,\hat{\tau}))^2}{\Delta_i^2}. \tag{26}$$

Further simplification of the expression results in a closed form for $J(\hat{\tau})$, as shown in Appendix A. This analytical result is developed only for the one parameter case in this work.

4.1.2. Numerical approach

For cases in which more than one time constant is used to represent the response of the viscoelastic material, a numerical approach is more convenient to compute the Fisher information. In this case, the derivative terms, $\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}$, that populate the Fisher information matrix (11) are first evaluated symbolically. The resulting derivatives are evaluated numerically and the determinant of the matrix is computed using Python [24].

4.2. Relative Entropy or Kullback-Leibler Divergence

For the calculation of the KL divergence between the prior and posterior distributions of the time constant(s), we note that Eqn. (15) does not typically have a closed form solution and must be approximated numerically. In this study, a uniform distribution on the parameters is assumed as the prior distribution and the posterior distribution of the parameters is calculated using the nested Monte Carlo sampling-based algorithm MultiNest [25, 26]. This method is a fast and robust alternative to the traditional Markov Chain Monte Carlo methods that dominate the literature. This method also has the capability to efficiently handle multimodal posterior distributions.

The global nature of the posterior is investigated, and in the case when the posterior is unimodal and the number of parameters is small, the integral in Eqn. (15) can be approximated using numerical quadrature techniques. In this study, these conditions are met, so the global adaptive quadrature method in MATLAB [27] is used to evaluate the KL divergence integral.

5. Results

We now discuss the results of our study. First, for the one- and two-parameter cases, we evaluate the Fisher information and KL divergence as a function of the ratio of t_{tot}, the total time of the experiment to τ, the time constant. This quantity is denoted by K: $K = t_{\text{tot}}/\tau$. We then study the sensitivity of the Fisher information and KL divergence measures to the noise model.

5.1. One Parameter Case

5.1.1. Experimental Data

A one-term Prony series with one parameter of interest (τ) is considered in this section and the effect of the duration of the experiment on the information gained about the time constant τ is investigated. For this purpose, synthetic experimental data is generated by fixing the parameters in the model (Eqn. (6)) to the values shown in Table 1. The experimental noise is assumed to be independent and identically distributed Gaussian with a constant variance at each data point and equal to 5 Pa, which is small compared to the mean stress values. The resulting response is shown in Fig. 3. To study the effect of the duration of the experiment on the information gained about the parameter τ, the total experimental time is varied in increments of τ. The quantity $K = t_{\text{tot}}/\tau$ is varied from $K = 1$ to $K = 10$.

Table 1: One parameter case: Model parameters used to generate synthetic relaxation experimental data using a one-term Prony series.

t_r (sec)	$\dot{\epsilon}$ (1/sec)	t_{tot} (sec)	E_0 (Pa)	E_1 (Pa)	τ (sec)
1	1	20	600	500	2

5.1.2. Fisher Information & KL Divergence

The Fisher information is evaluated both analytically according to Eqn. (A.9) for different values of K and numerically as outlined in Section 4.1.2. The resulting values of the normalized Fisher information as a function of K are shown in Fig. 4(a), while Fig. 4(b) shows the computed KL divergence between the prior and the posterior for the same problem. There is no noticeable change in the normalized Fisher information or the KL divergence beyond a value of $K = 5$.

Note that, in the analytical approach, some terms are assumed to be negligible to simplify the analysis, whereas this assumption is not needed for the numerical evaluation of the Fisher information. Consequently, the two approaches do not lead to the exact same absolute values

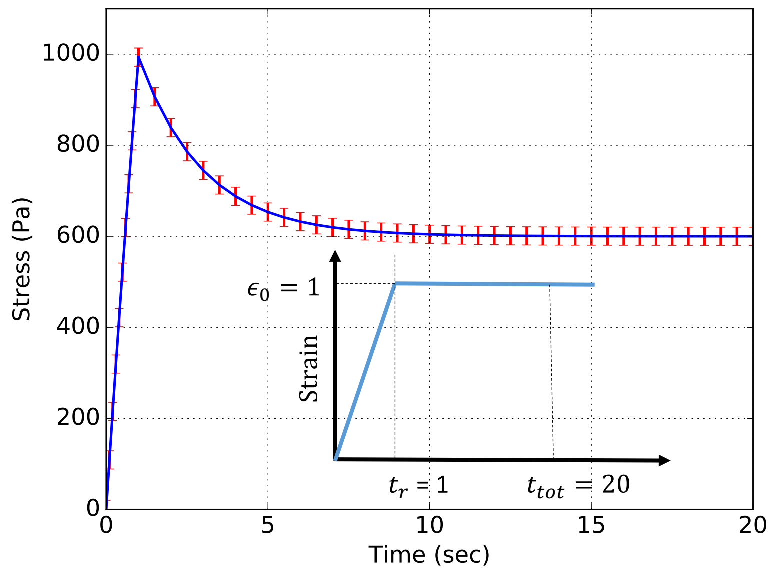


Figure 3: Stress-time synthetic experimental response of a linear viscoelastic material defined by a one-term Prony series.

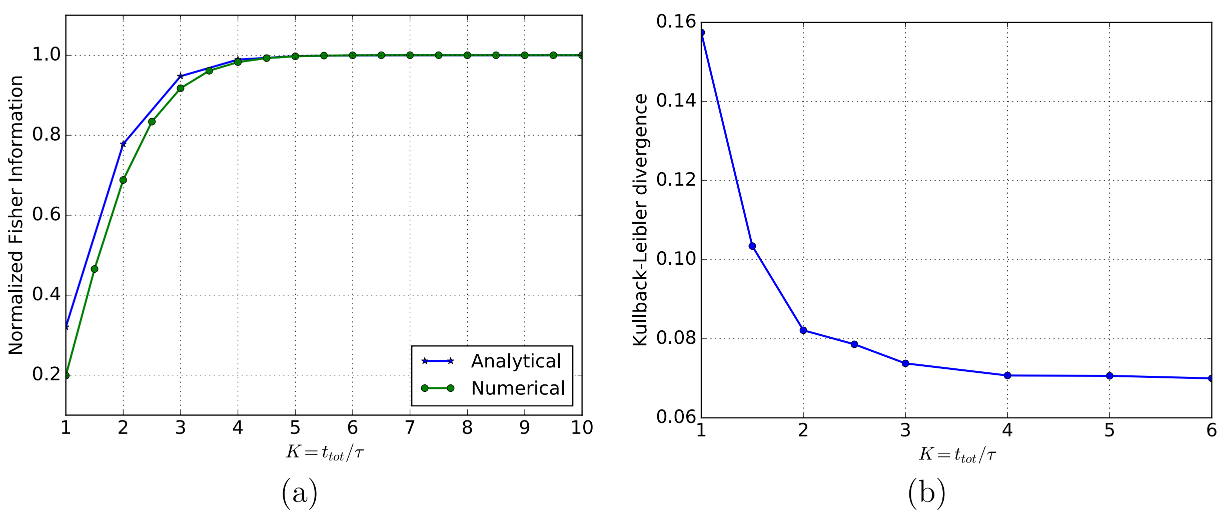


Figure 4: Information gain as a function of the duration of the experiment for the one-parameter case: (a) Fisher information calculated from analytical and numerical approaches, and (b) KL divergence calculated numerically.

of the Fisher information. Therefore, to facilitate comparison of the two methods, the Fisher information value from each approach is normalized with respect to the maximum from that approach.

In the case of the numerical approach, the likelihood function can be computed for a given experimental configuration (i.e., a given K). The normalized distribution function for the likelihood is shown in Fig. 5 for various values of K. As K increases, the distributions

t_r (sec)	$\dot{\epsilon}$ (1/sec)	t_{tot} (sec)	E_0 (Pa)	E_1 (Pa)	E_2 (Pa)	τ_1 (sec)	τ_2 (sec)
0.5	1	100	600	500	500	1	2:2:12

Table 2: Model parameters used to generate synthetic experimental relaxation data using a two-term Prony series. Time constant τ_2 is varied from 2 sec to 12 sec in increments of 2 sec.

get narrower as the variance decreases and the curvature at the peak increases. Beyond $K = 5$, the curves overlap and are indistinguishable.

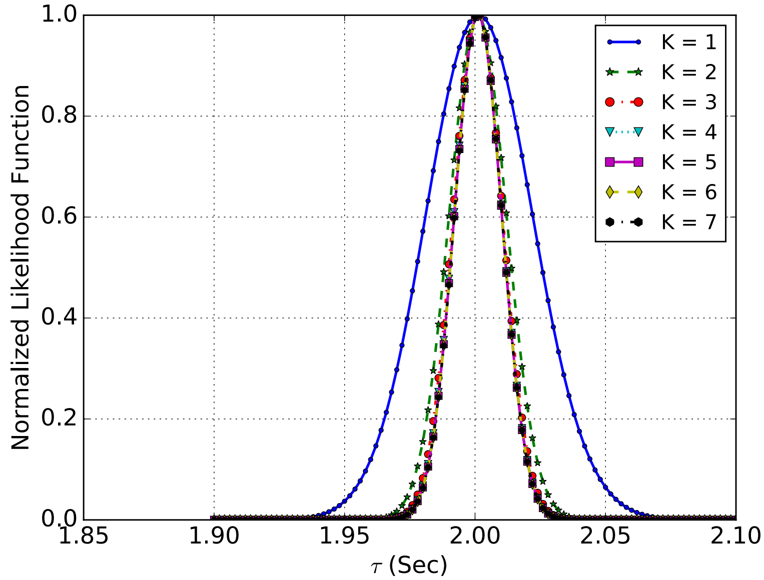


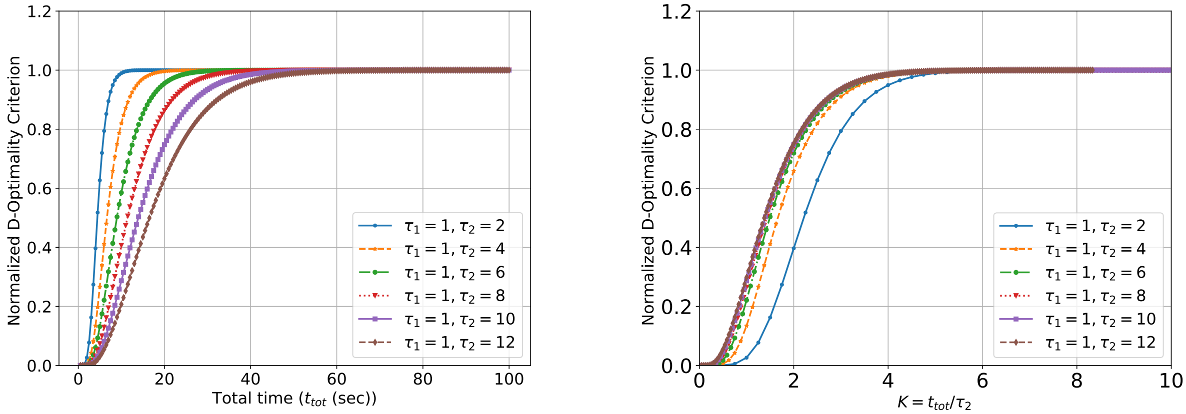
Figure 5: Normalized likelihood as a function of the duration of the experiment for the one-parameter case.

5.2. Two-parameter Case

A two-term Prony series with two parameters of interest (τ_1, τ_2) is considered in this section. This experiment is to study the relationship between the information gained about the time constants and total time of the experiment when there are more than one time constant. Synthetic experimental data are generated in a manner similar to the one-parameter case. The parameters used to generate the data are shown in Table 2. The value of τ_2 is varied to be $2, 4, 6, 8, 10, 12$ sec. Also, the experimental noise is assumed to be i.i.d Gaussian with a constant variance at each data point and equal to 5 Pa, as in the one-parameter case.

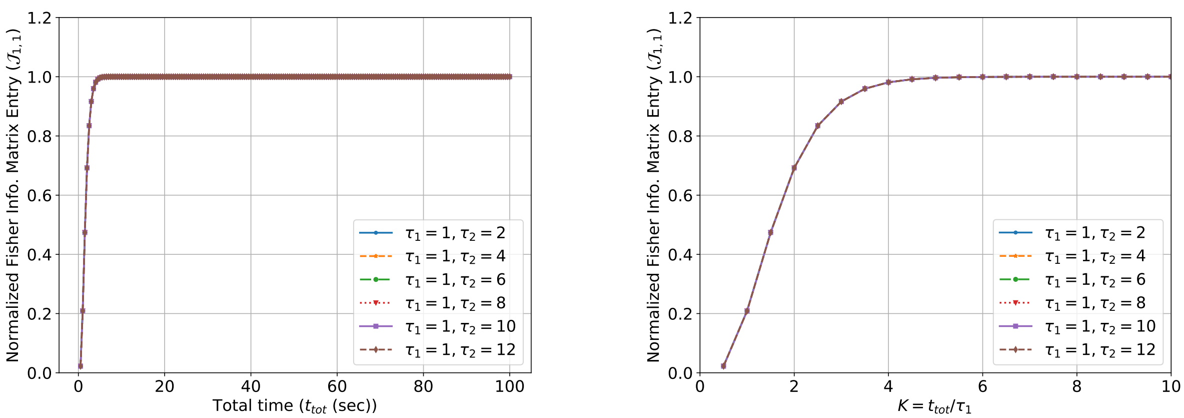
5.2.1. D-Optimality Criterion

The D-optimality criterion is calculated for each combination of τ_1 and τ_2, which leads to six different cases. For each case, the total time of the experiment is varied from 0.5 sec to 100 sec. The value of the D-optimality criterion is normalized by its highest value for each case and plotted as a function of total time as well as $K = t_{\text{tot}}/\tau_2$, as shown in



(a) (b)

Figure 6: Normalized D-Optimality criterion for various choices of τ_2 for the two-parameter case: (a) As a function of t_{tot}, and (b) as a function of $K = t_{\text{tot}}/\tau_2$.



(a) (b)

Figure 7: Normalized Fisher information matrix entry corresponding to τ_1 ($\mathcal{J}_{1,1}$) for various choices of τ_2: (a) As a function of t_{tot}, and (b) as a function of $K = t_{\text{tot}}/\tau_2$.

Fig. 6. In all cases, maximum information gain occurs around $K = 5$ and remains constant thereafter. In order to study the contribution from each of the time constants, the diagonal components corresponding to each time constant τ_1, τ_2 in the Fisher information matrix are shown in Figs. 7 and 8, respectively. In both cases, we see the diagonal entries converge to their maximum values at $K = 5$.

5.3. Sensitivity to Experimental Variability

The synthetic experimental data used thus far is generated under the assumption that the experimental variability ($\Delta = 5$ Pa) is constant in the loading and relaxation phases of

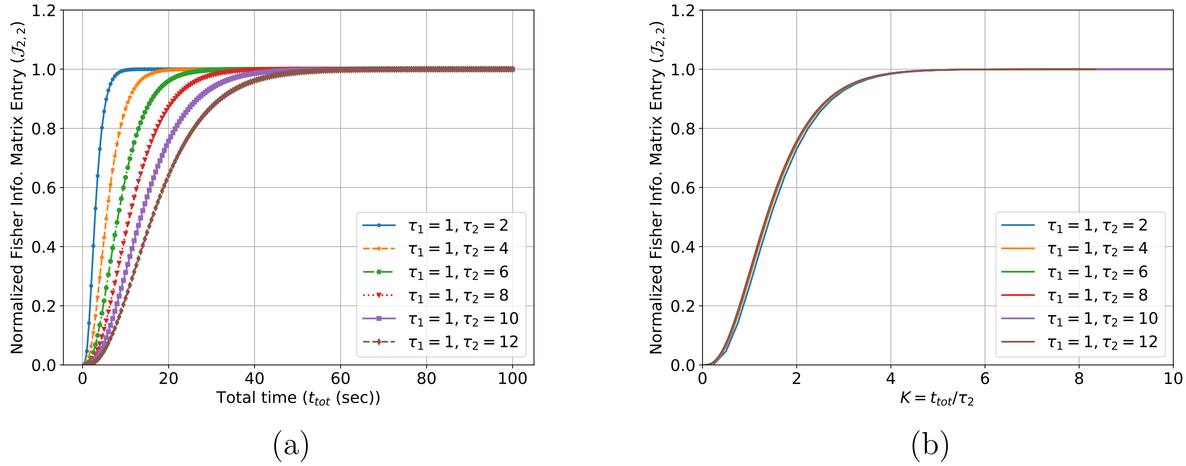


Figure 8: Normalized Fisher information matrix entry corresponding to τ_2 ($\mathcal{J}_{2,2}$) for various choices of τ_2: (a) As a function of t_{tot}, and (b) as a function of $K = t_{\text{tot}}/\tau_2$.

the experiment. In practice, however, we find that the experimental variability is higher in the loading stage and decreases as the experiment progresses into the relaxation phase. In order to study the effects of this variability on the Fisher information values, we consider experiments with different levels of experimental variability in the loading (Δ_l) and relaxation (Δ_r) stages.

Four cases are considered:

Case 1. The variabilities in the loading and relaxation phases are equal and constant: $\Delta_l = \Delta_r = 5$ Pa. This is the case considered thus far.

Case 2. The variabilities in the loading and relaxation phases are constant, with more variability in the loading phase: $\Delta_l = 25$ Pa, $\Delta_r = 5$ Pa.

Case 3. The variability is constant in the loading phase ($\Delta_l = 25$ Pa) and decreases linearly in the relaxation phase to a value of $\Delta_r = 5$ Pa as follows:

$$\Delta_r(t) = \left(\frac{\mu_2 - \mu_1}{t_{\text{tot}} - t_r}\right)(t - t_r),\ t \in (t_r, t_{\text{tot}}], \tag{27}$$

with $\mu_1 = 5$ Pa, $\mu_2 = 25$ Pa.

Case 4. The variability is constant in the loading phase ($\Delta_l = 25$ Pa) and decreases logarithmically in the relaxation phase to a value of $\Delta_r = 5$ Pa as follows:

$$\Delta_r(t) = \left(\frac{\mu_2 - \mu_1}{\log(t_{\text{tot}}) - \log(t_r)}\right)(\log(t) - \log(t_r)),\ t \in (t_r, t_{\text{tot}}], \tag{28}$$

with $\mu_1 = 5$ Pa, $\mu_2 = 25$ Pa.

The various models of experimental variability are shown in Fig. 9. First, we consider a viscoelastic material characterized by a one-term Prony series with the nominal value of

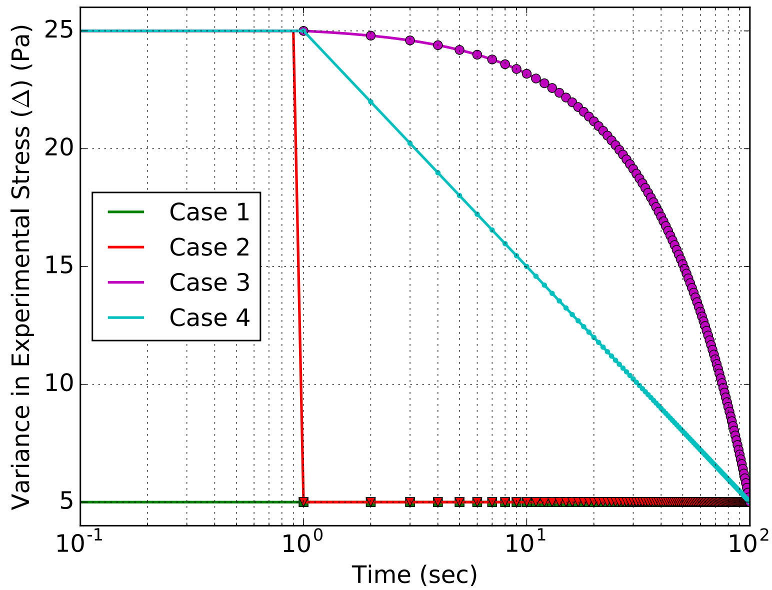


Figure 9: Four models of experimental variability in the stress as a function of time.

the time constant taken to be $\tau = 2$ sec. Setting the loading and total experimental times to $t_r = 1$ sec and $t_{\text{tot}} = 30$ sec, respectively, we calculate the Fisher information for each variability model. The calculations are done analytically for the first model ($\Delta_l = \Delta_r = 5$ Pa), and numerically for the rest. The results are shown in Figure 10(a) as a function of the quantity $K = t_{\text{tot}}/\tau$.

Similarly, the normalized D-optimality criterion is calculated for the two-term Prony series case with nominal time constant values $\tau_1 = 2$ sec, $\tau_2 = 10$ sec, and experimental conditions $t_r = 1$ sec, and $t_{\text{tot}} = 100$ sec. Figure 10(b) shows the plots for the normalized D-optimality criterion as a function of the quantity $K = t_{\text{tot}}/\tau_2$. For both the one- and two-term Prony series cases, we see that regardless of the variability model, there is no information gain beyond $K = 5$.

6. Discussion

For the one-term Prony series case, the predictions from the analytical and numerical predictions of the Fisher information are in close agreement: information gain from the data about the time constant does not change beyond $K = 5$. As such, performing the relaxation experiment for a longer duration is unnecessary. Conversely, given a relaxation experiment with some known duration, the time constant inferred from the data must not exceed one-fifth of the duration. This condition may be imposed as a constraint on the optimization algorithm used to infer the time constant. These conclusions are supported by considering the KL divergence or relative entropy between the prior and posterior distributions for τ.

For the case of the two-term Prony series, information gain (as defined by the D-optimality criterion) is maximized at a total time corresponding to five times the larger

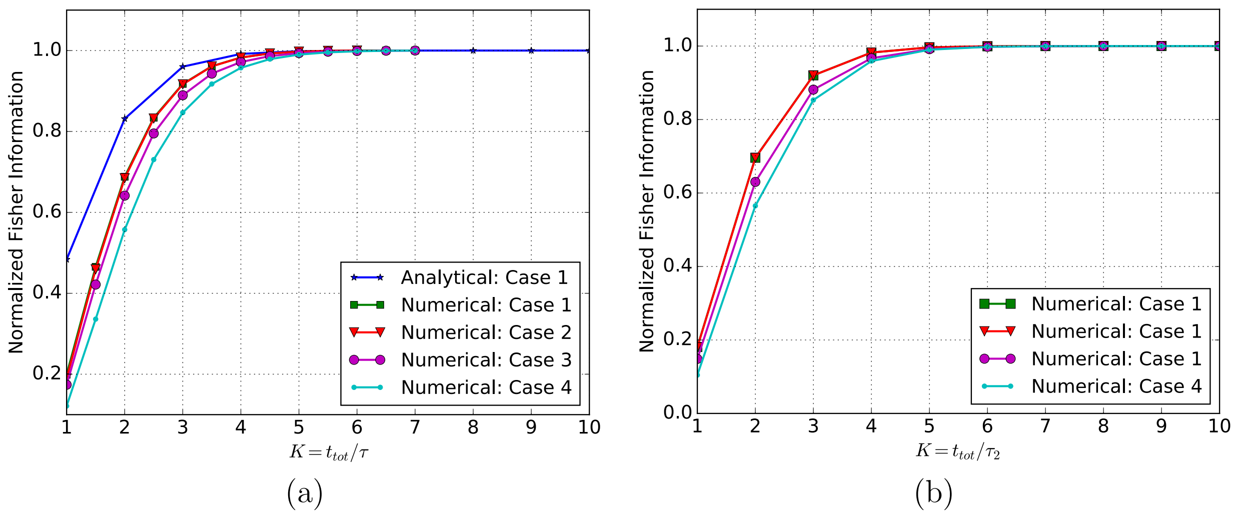


Figure 10: Sensitivity of the Fisher information criterion to the experimental variability model: (a) One-term case, (b) Two-term case.

time constant τ_2. This is seen in Fig. 6. The contribution of the diagonal term in the Fisher information matrix corresponding to τ_1 (Fig. 7) shows that there is no information gained about the smaller time constant beyond a total time of $5\tau_1$, and does not contribute significantly to the D-optimality criterion past this point. This conclusion is also supported by the response of the diagonal term in the Fisher information corresponding to τ_2 (Fig. 8).

Interestingly, the experimental variability only seems to affect the rate at which the normalized Fisher information and D-optimality criterion increase, but not the K-value at which information gain is maximized (Fig. 10). This essentially says that, even in the case of multiple time constants, the largest time constant that can reliably be inferred from the experiment is about one-fifth of the total time of the experiment.

Traditionally, in the literature on viscoelasticity, it is common practice to use one relaxation time constant per decade of the experimental data time scale [11]. This essentially means that the time constants are equally spaced on the logarithmic time axis. This is done by fixing the time constant to be the decade values (e.g., $10^{-1}, 10^0, 10^1$), or fixing the intervals to span a decade and allowing the optimizer pick the time constant from this interval (e.g., $[10^{-1}, 10^0]$, $[10^0, 10^1]$, $[10^1, 10^2]$). Our results for the one- and two-term Prony series cases show that the largest time constant that can be obtained from the experiment is one-fifth of the total time of the experiment. This can be applied as a constraint to the optimizer in either of the aforementioned situations.

Also, it should be noted that the smallest time constant inferred from the experiment has to be greater than the acquisition frequency of the experiment, i.e., time interval between successive data points. Clearly, choosing a time constant smaller than this value is not physically meaningful.

7. Conclusions

In this study, we consider a linear viscoelastic material model and address the question of determining the duration of a relaxation experiment in order to maximize the information gained about the time constants. This can also be interpreted as a way to determine the largest time constant that can be inferred from the experiment. In this work, an analytical model is developed to estimate the Fisher information for the case of a linear viscoelastic material with one time constant parameter. The predictions from this analytical approach closely match those from the numerical approach as well as those from the calculating the KL divergence between the prior and posterior distributions of the time constant. The results clearly show that information gain about the time constant is maximized when $t_{\mathrm{tot}} \geq 5\tau$. The same conclusion is reached for the two-parameter case as well.

We plan to extend this method of optimal experimental design to generalized linear viscoelastic, and nonlinear viscoelastic models and also to study the relationship between time constant and the total time of the experiment for the case when the time constants are modeled as random variables due to the uncertainty in the experimental data.

8. Acknowledgements

We are grateful to the University of Cincinnati Simulation Center for providing financial support for this work.

9. Conflict of Interest Statement

The authors do not have any conflicts of interest to report.

References

[1] Y.-C. Fung, Biomechanics, Springer, 1990.

[2] A. Wineman, Nonlinear viscoelastic solids - a review, Mathematics and Mechanics of Solids 14 (3) (2009) 300–366.

[3] D. P. Pioletti, L. R. Rakotomanana, Non-linear viscoelastic laws for soft biological tissues, European Journal of Mechanics A-Solids 19 (LBO-ARTICLE-2000-002) (2000) 749–759.

[4] K. Miller, K. Chinzei, Mechanical properties of brain tissue in tension, Journal of Biomechanics 35 (4) (2002) 483–490.

[5] G. Limbert, J. Middleton, A transversely isotropic viscohyperelastic material: application to the modeling of biological soft connective tissues, International Journal of Solids and Structures 41 (15) (2004) 4237–4260.

[6] A. Nava, E. Mazza, M. Furrer, P. Villiger, W. Reinhart, In vivo mechanical characterization of human liver, Medical Image Analysis 12 (2) (2008) 203–216.

[7] B. Rashid, M. Destrade, M. D. Gilchrist, Mechanical characterization of brain tissue in simple shear at dynamic strain rates, Journal of the Mechanical Behavior of Biomedical Materials 28 (2013) 71–85.

[8] D. Sahoo, C. Deck, R. Willinger, Development and validation of an advanced anisotropic visco-hyperelastic human brain FE model, Journal of the Mechanical Behavior of Biomedical Materials 33 (2014) 24–42.

[9] A. Tamura, S. Hayashi, I. Watanabe, K. Nagayama, T. Matsumoto, Mechanical characterization of brain tissue in high-rate compression, Journal of Biomechanical Science and Engineering 2 (3) (2007) 115–126.

[10] R. De Pascalis, I. D. Abrahams, W. J. Parnell, On nonlinear viscoelastic deformations: a reappraisal of Fung's quasi-linear viscoelastic model, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 470 (2166).

[11] W. G. Knauss, J. Zhao, Improved relaxation time coverage in ramp-strain histories, Mechanics of Time-Dependent Materials 11 (3-4) (2007) 199–216.

[12] I. Saprunov, M. Gergesova, I. Emri, Prediction of viscoelastic material functions from constant stress- or strain-rate experiments, Mechanics of Time-Dependent Materials 18 (2014) 349–372.

[13] V. V. Fedorov, P. Hackl, Model-oriented design of experiments, Vol. 125, Springer, 1997.

[14] A. Atkinson, DT-optimum designs for model discrimination and parameter estimation, Journal of Statistical Planning and Inference 138 (1) (2008) 56–64.

[15] K. Chaloner, I. Verdinelli, Bayesian experimental design: A review, Statistical Science (1995) 273–304.

[16] X. Huan, Y. M. Marzouk, Simulation-based optimal Bayesian experimental design for nonlinear systems, Journal of Computational Physics 232 (1) (2013) 288–317.

[17] E. G. Ryan, C. C. Drovandi, J. M. McGree, A. N. Pettitt, A review of modern computational algorithms for Bayesian optimal design, International Statistical Review 84 (1) (2016) 128–154.

[18] D. V. Lindley, A. F. Smith, Bayes estimates for the linear model, Journal of the Royal Statistical Society. Series B (Methodological) (1972) 1–41.

[19] K. P. Burnham, D. R. Anderson, Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Springer, 2002.

[20] F. Pukelsheim, Optimal Design of Experiments, Vol. 50, Society for Industrial and Applied Mathematics, 1993.

[21] J. Andere-Rendon, D. C. Montgomery, D. A. Rollier, Design of mixture experiments using Bayesian D-optimality, Journal of Quality Technology 29 (4) (1997) 451.

[22] D. J. C. MacKay, Information Theory, Inference, and Learning Algorithms, Cambridge University Press, New York, 2003.

[23] D. S. Sivia, Data Analysis: A Bayesian Tutorial, Oxford University Press, 1996.

[24] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: Open source scientific tools for Python (2001–).
URL `http://www.scipy.org/`

[25] F. Feroz, M. Hobson, M. Bridges, MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics, Monthly Notices of the Royal Astronomical Society 398 (4) (2009) 1601–1614.

[26] F. Feroz, M. Hobson, E. Cameron, A. Pettitt, Importance nested sampling and the MultiNest algorithm, arXiv preprint arXiv:1306.2144.

[27] MATLAB, version 8.1.0.604 (R2013a), The MathWorks Inc., Natick, Massachusetts, 2013.

Appendix A. Fisher Information: Analytical Approach

In this section, we detail the derivation of an analytical expression for the Fisher information when the viscoelastic material is characterized by a one-term Prony series, i.e., a single time constant. Using Eqn. (6), the first derivatives of f_1 and f_2 are evaluated as follows

$$
\begin{aligned}
f_1'(t,\hat{\tau}) &= \dot{\epsilon}\, E_1 \left[1 - \exp\left(-\frac{t}{\hat{\tau}}\right) - \frac{t}{\hat{\tau}} \exp\left(-\frac{t}{\hat{\tau}}\right)\right], \\
f_2'(t,\hat{\tau}) &= \dot{\epsilon}\, E_1 \left[\exp\left(-\frac{t-t_r}{\hat{\tau}}\right) - \exp\left(-\frac{t}{\hat{\tau}}\right) - \frac{t}{\hat{\tau}} \exp\left(-\frac{t}{\hat{\tau}}\right) + \left(\frac{t-t_r}{\hat{\tau}}\right) \exp\left(-\frac{t-t_r}{\hat{\tau}}\right)\right].
\end{aligned}
\tag{A.1}
$$

The expressions for $(f_1')^2$ and $(f_2')^2$ are

$$
(f_1'(t,\hat{\tau}))^2 = (\dot{\epsilon}\, E_1)^2 \left[1 + \exp\left(-\frac{2t}{\hat{\tau}}\right)\left(1 + \frac{2t}{\hat{\tau}} + \left(\frac{t}{\hat{\tau}}\right)^2\right) - 2\exp\left(-\frac{t}{\hat{\tau}}\right)\left(1 + \frac{t}{\hat{\tau}}\right)\right], \tag{A.2}
$$

and

$$
\begin{aligned}
(f_2'(t,\hat{\tau}))^2 = (\dot{\epsilon}\, E_1)^2 \Bigg[& 1 + \exp\left(-\frac{2t}{\hat{\tau}}\right)\left(1 + \frac{2t}{\hat{\tau}} + \left(\frac{t}{\hat{\tau}}\right)^2\right) \\
& + \exp\left(-\frac{2(t-t_r)}{\hat{\tau}}\right)\left(1 + \frac{2(t-t_r)}{\hat{\tau}} + \left(\frac{t-t_r}{\hat{\tau}}\right)^2\right) \\
& - 2\exp\left(-\frac{2(t-t_r)}{\hat{\tau}}\right)\left(1 + \frac{t}{\hat{\tau}} + \frac{(t-t_r)}{\hat{\tau}} + \frac{t}{\hat{\tau}}\frac{(t-t_r)}{\hat{\tau}}\right)\Bigg].
\end{aligned}
\tag{A.3}
$$

With $K = t_{\text{tot}}/\hat{\tau}$ and $t_i = \Delta t\, i = (t_{\text{tot}}/N)i$, we have $(t_i/\hat{\tau}) = (K/N)i$. Define $P_1 = \exp(-K/N)$ and $P_2 = \exp(-2K/N)$. Then, Eqn. (A.2) can be written as

$$
(f_1'(t,\hat{\tau}))^2 = (\dot{\epsilon}\, E_1)^2 \left[1 + P_2^i\left(1 + \frac{2K}{N}i + \left(\frac{K}{N}\right)^2 i^2\right) - 2P_1^i\left(1 + \frac{K}{N}i\right)\right]. \tag{A.4}
$$

Similarly Eqn. (A.3) yields

$$
\begin{aligned}
(f_2'(t,\hat{\tau}))^2 = (\dot{\epsilon}\, E_1)^2 \Bigg[& 1 + P_2^i\left(1 + \frac{2K}{N}i + \left(\frac{K}{N}\right)^2 i^2\right) + P_2^j\left(1 + \frac{2K}{N}j + \left(\frac{K}{N}\right)^2 j^2\right) \\
& - 2P_2^m\left(1 + \frac{K}{N}m + \frac{1}{4}\left(\frac{K}{N}\right)^2 (m^2 - n_r^2)\right)\Bigg],
\end{aligned}
\tag{A.5}
$$

where $j = i - n_r$ and $m = 2i - n_r$. Substituting Eqns. (A.4), (A.5) into Eqn. (26) gives

$$\begin{aligned} J(\hat{\tau}) = \left(\frac{\dot{\epsilon}\, E_1}{\Delta}\right)^2 \sum_{i=1}^{n_r} \left[1 + P_2^i \left(1 + \frac{2K}{N} i + \left(\frac{K}{N}\right)^2 i^2 \right) - 2P_1^i \left(1 + \frac{K}{N} i \right) \right] \\ + \sum_{i=n_r+1}^{N} 1 + P_2^i \left(1 + \frac{2K}{N} i + \left(\frac{K}{N}\right)^2 i^2 \right) \\ + \sum_{j=1}^{N-n_r} P_2^j \left(1 + \frac{2K}{N} j + \left(\frac{K}{N}\right)^2 j^2 \right) \\ + \sum_{m=n_r+2}^{2N-n_r} -2P_2^m \left(1 + \frac{K}{N} m + \frac{1}{4} \left(\frac{K}{N}\right)^2 (m^2 - n_r^2) \right). \end{aligned} \tag{A.6}$$

For $P = P_1, P_2$, define a summation term $S_i(P, N)$ as follows:

$$\begin{aligned} S_i(P, N) &= 1^i P^1 + 2^i P^2 + 3^i P^3 + ... + k^i P^k + ... + N^i P^N \\ &= \sum_{k=1}^{N} k^i P^k. \end{aligned} \tag{A.7}$$

Then, it is straightforward to show that the first few terms of this series are:

$$\begin{aligned} S_0(P, N) &= \frac{P - P^{n+1}}{1 - P}, \\ S_1(P, N) &= \frac{S_0(P, N) - N\, P^{n+1}}{1 - P}, \\ S_2(P, N) &= \frac{2S_1(P, N) - S_0(P, N) - N^2 P^{n+1}}{1 - P}, \\ S_3(P, N) &= \frac{3S_2(P, N) - 3S_1(P, N) + S_0(P, N) - N^3 P^{n+1}}{1 - P}, \\ S_4(P, N) &= \frac{4S_3(P, N) - 6S_2(P, N) + 4S_1(P, N) - S_0(P, N) - N^4 P^{n+1}}{1 - P}, \\ S_5(P, N) &= \frac{5S_4(P, N) - 10S_3(P, N) + 10S_2(P, N) - 5S_1(P, N) + S_0(P, N) - N^5 P^{n+1}}{1 - P}. \end{aligned} \tag{A.8}$$

Simplifying Eqn. (A.6) and using Eqn. (A.8), we find

$$
\begin{aligned}
J(\hat{\tau}) = \left(\frac{\dot{\epsilon}\, E_1}{\Delta}\right)^2 \Bigg[& n_r + S_0(P_2, N) + \left(\frac{2K}{N}\right) S_1(P_2, N) + \left(\frac{K}{N}\right)^2 S_2(P_2, N) \\
& - 2\left(S_0(P_1, n_r) + \left(\frac{K}{N}\right) S_1(P_1, n_r) \right) + S_0(P_2, N - n_r) + \left(\frac{2K}{N}\right) S_1(P_2, N - n_r) \\
& + \left(\frac{K}{N}\right)^2 S_2(P_2, N - n_r) - 2\Bigg(S_0(P_2, 2N - n_r) + \left(\frac{K}{N}\right) S_1(P_2, 2N - n_r) \\
& + \frac{1}{4}\left(\frac{K}{N}\right)^2 \left(S_2(P_2, 2N - n_r) - n_r^2 S_0(P_2, 2N - n_r)\right) \Bigg) + 2\Bigg(S_0(P_2, n_r + 2) \\
& + \left(\frac{K}{N}\right) S_1(P_2, n_r + 2) + \frac{1}{4}\left(\frac{K}{N}\right)^2 \left(S_2(P_2, n_r + 2) - n_r^2 S_0(P_2, n_r + 2)\right) \Bigg) \Bigg] .
\end{aligned}
\tag{A.9}
$$

This is the final expression used to analytically evaluate the Fisher information for the one-term Prony series case.