

Article

An Information-Theoretic Framework for Optimal Design: Analysis of Protocols for Estimating Soft Tissue Parameters in Biaxial Experiments

Ankush Aggarwal ^{1,†} , Damiano Lombardi ^{2,3,†}  and Sanjay Pant ^{4,*,†} 

¹ Glasgow Computational Engineering Centre, James Watt School of Engineering, University of Glasgow, Glasgow G12 8LT, UK; Ankush.Aggarwal@glasgow.ac.uk

² COMMEDIA, Inria Paris, 2 rue Simone Iff, 75012 Paris, France; Damiano.Lombardi@inria.fr

³ Laboratoire Jacques-Louis Lions, Sorbonne Université, 75006 Paris, France

⁴ Zienkiewicz Centre for Computational Engineering, College of Engineering, Swansea University, Swansea SA1 8EN, UK

* Correspondence: Sanjay.Pant@swansea.ac.uk

† These authors contributed equally to this work.

Abstract: A new framework for optimal design based on the information-theoretic measures of mutual information, conditional mutual information and their combination is proposed. The framework is tested on the analysis of protocols—a combination of angles along which strain measurements can be acquired—in a biaxial experiment of soft tissues for the estimation of hyperelastic constitutive model parameters. The proposed framework considers the information gain about the parameters from the experiment as the key criterion to be maximised, which can be directly used for optimal design. Information gain is computed through k -nearest neighbour algorithms applied to the joint samples of the parameters and measurements produced by the forward and observation models. For biaxial experiments, the results show that low angles have a relatively low information content compared to high angles. The results also show that a smaller number of angles with suitably chosen combinations can result in higher information gains when compared to a larger number of angles which are poorly combined. Finally, it is shown that the proposed framework is consistent with classical approaches, particularly D-optimal design.

Keywords: optimal design; soft tissue mechanics; mutual information; biaxial experiment; inverse problems; information theory

MSC: 62K05; 94A15; 92C10



Citation: Aggarwal, A.; Lombardi, D.; Pant, S. An Information-Theoretic Framework for Optimal Design: Analysis of Protocols for Estimating Soft Tissue Parameters in Biaxial Experiments. *Axioms* **2021**, *10*, 79. <https://doi.org/10.3390/axioms10020079>

Academic Editor: Gabriella Bretti

Received: 31 March 2021

Accepted: 22 April 2021

Published: 1 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Soft tissues exhibit complex biomechanical behaviour, including nonlinearity, anisotropy and heterogeneity [1]. Moreover, the tissues also demonstrate inelastic properties, such as rate-dependence, hysteresis and permanent set. The important link between biomechanics and their physiological function has motivated a large number of ex-vivo studies aimed at characterising their biomechanical properties. Given the complex interplay between the different aspects of their biomechanical properties, the experimental design of ex-vivo soft tissues is extremely challenging and has been a subject of investigation, and a variety of experiments have been proposed [2–6].

Since a variety of soft tissues are thin—e.g., blood vessels, heart valves and skin—biaxial testing is a widely used experimental technique that allows the independent stretching of the tissue in two orthogonal directions and for the corresponding forces to be measured [7,8]. Applying different stretches in two directions allows the characterization of the in-plane anisotropic behavior of a given tissue, while a range of stretches provides us with its nonlinear elastic response. However, even with this relatively simple set of options,

the choices of which stretches to apply are unclear. Moreover, it is not obvious upon what these choices will depend.

A variety of hyperelastic models have been developed to describe the anisotropic and nonlinear elastic properties of specific soft tissues [4,9–11]. Biaxial experimental data are commonly fit to these models in order to determine the model parameters. As the unknown parameters depend on specific models, the choice of experimental setup—the problem of optimal design—might depend on the choice of model. However, in practice, a predetermined set of experimental protocols is used.

In the present work, an optimal design problem is defined to find the most suitable protocol in view of estimating the parameters of the material model. A comprehensive overview of the optimal design problem can be found in [12,13], and several criteria for optimal design have been proposed in the literature, often based on the minimisation of the variance of the parameters and sensitivities [14,15]. In the present work, we investigate a criterion based on information theoretic quantities, in the spirit of what has been proposed in [16,17] (from a Bayesian point of view) and [18]. Several works have recently proposed information-based criteria to better define experimental protocols. In [19], the authors proposed the maximisation of the mutual information between the parameters and the observations under the assumption that the model error is a Gaussian process. In [20], the authors proposed a framework based on mutual information maximisation to deal with the design of chemistry experiments. The same criterion is proposed in [21]; the authors maximise the mutual information by using a stochastic gradient ascent method. An application to system biology is investigated in [22]. In [23], the maximisation of the information is exploited in order to choose high-fidelity model resolutions in a multi-fidelity modelling framework.

While mutual information has been used for optimal design in previous studies, the novelty of this work is in the proposal of a combination of information-theoretic quantities of both mutual and conditional mutual information. A further novelty is the application of this framework in the optimal design of soft tissue experiments. Estimating information-theoretic quantities is in general a challenging problem, and this is especially the case in high-dimensional settings. In the present work, a model reduction method is coupled with non-parametric sample-based mutual information estimation in order to provide a pertinent estimation of the information-theoretic quantities involved in the optimal design problem and then apply this to the biaxial testing of soft tissues.

The structure of the work is as follows: in Section 2, the model and information-theoretical aspects of the problem are introduced. In particular, in Section 2.1, we detail the mathematical model of the biaxial experiments for soft tissues: after having introduced the notation and the non-linear elasticity model, in Section 2.1.1, we apply it to the biaxial testing experimental setup. In Section 2.1.2, we introduce the experimental protocol definition; the second part of the section is devoted to the description of the information-theoretic framework used to solve the optimal design problem. In Section 2.2.1, we introduce the problem; in Sections 2.2.2 and 2.2.3, the information-theoretic quantities and their numerical estimation are detailed. We then present the reduce order modeling method used and how to validate the results obtained by the proposed approach. The section ends with an overview of the method. The results and the discussion are presented in Section 3, followed by the conclusion and perspectives on future work.

2. Methods

The methodological aspects are divided into two broad categories: the mathematical model of the biaxial experiments and the information-theoretic optimal design framework.

2.1. Mathematical Model of the Biaxial Experiments

We begin by defining the notation: a material point at its reference position $\mathbf{X} \in \mathbb{R}^3$ moves to $\mathbf{x} \in \mathbb{R}^3$ after deformation. The elastic behaviour of soft tissues is described using the hyperelastic strain energy density Ψ , which depends on the deformation gradient

tensor $\mathbf{F} = \nabla_{\mathbf{x}}\mathbf{x}$. The ratio of the volume after deformation to that before deformation is given by $J = \det(\mathbf{F})$. Soft tissues are commonly regarded as incompressible due to their high water content; i.e. J is constrained to be unity.

We consider the hyperelastic model proposed by Gasser et al. [24], which defines the strain energy density as

$$\Psi = \frac{k_1}{2k_2} \left[e^{k_2(\kappa I_1 + (1-3\kappa)I_4 - 1)} - 1 \right] + \mu(I_1 - 3), \tag{1}$$

where $I_1 = \text{tr}(\mathbf{F}^T \mathbf{F})$ is the first invariant of the right Cauchy–Green strain tensor $\mathbf{C} = \mathbf{F}^T \mathbf{F}$ and $I_4 = \mathbf{M} \cdot \mathbf{C} \mathbf{M}$ is the fourth invariant representing the stretch along fiber direction \mathbf{M} . The resulting Cauchy stress is given by

$$\boldsymbol{\sigma} = 2\mathbf{F} \cdot \frac{\partial \Psi}{\partial \mathbf{C}} \cdot \mathbf{F}^T - p\mathbf{I}, \tag{2}$$

where p acts as the Lagrange multiplier to enforce incompressibility and \mathbf{I} is the identity matrix.

For this model, the set of unknown parameters can be written as $\{k_1, k_2, \kappa, \mu\}$, assuming that the fiber direction \mathbf{M} is known a priori (based on another experiment; e.g., light scattering [6]). κ represents the dispersion of collagen fibers, which is usually measured from optical experiments. Its value lies between 0 (perfectly anisotropic) and 1/3 (perfectly isotropic). The value of μ corresponds to the shear modulus of the neo-Hookean term in (1), which represents the amorphous and non-fibrous extracellular matrix. Its role in the mechanics of soft tissues is limited to small strains and is largely constant across different tissues. In this paper, in order to simplify the problem, we assume that $\kappa = 0.1$ and $\mu = 1$ kPa are known and fixed. Thus, the aim of an ex-vivo biomechanical experiment is to determine parameters $k_1 \in [5, 100]$ kPa and $k_2 \in [5, 80]$ robustly and with high confidence [25,26]. A commonly used experiment called biaxial testing is described bellow.

2.1.1. Biaxial Experiments for Soft-Tissues

Many of the soft tissue types are planar with a small thickness. In a biaxial experiment, a square-shaped tissue sample is mounted via clamps or rakes and stretched along two orthogonal directions aligned with the sample edges (Figure 1a). If these directions are used as the two coordinate axes and incompressibility is assumed, the stretching results in a diagonal deformation gradient tensor:

$$\mathbf{F} = \text{diag} \left[\lambda_1, \lambda_2, \frac{1}{\lambda_1 \lambda_2} \right], \tag{3}$$

where λ_1 is the stretch along the first in-plane direction and λ_2 is the stretch along the second in-plane direction. The fiber direction \mathbf{M} is generally aligned with the first coordinate axis, which results in only normal stress components. As no force is applied along the thickness of the tissue, $\sigma_{33} = 0$ is used to determine the Lagrange multiplier p . Thus, we obtain

$$\sigma_{11} = 2 \frac{\partial \Psi}{\partial I_1} \left[\lambda_1^2 - \frac{1}{\lambda_1^2 \lambda_2^2} \right] + 2 \frac{\partial \Psi}{\partial I_4} \lambda_1^2 \tag{4}$$

$$\sigma_{22} = 2 \frac{\partial \Psi}{\partial I_1} \left[\lambda_2^2 - \frac{1}{\lambda_1^2 \lambda_2^2} \right]. \tag{5}$$

The applied stresses σ_{11} , σ_{22} are controlled using load cells. The resulting strains, defined as $e_1 := \lambda_1 - 1$ and $e_2 := \lambda_2 - 1$, are measured from the marker positions (although e_1 and e_2 are not the usual strain measures, we use these as our observations). It is important to note that a homogeneous stress and strain state is assumed in the middle of the sample (Figure 1a). Therefore, an implicit assumption is that the material properties and sample thickness are homogeneous. Moreover, these measurement techniques carry an error

due to the limitations in measurement tools and/or the deviation from homogeneity, incompressibility and material direction.

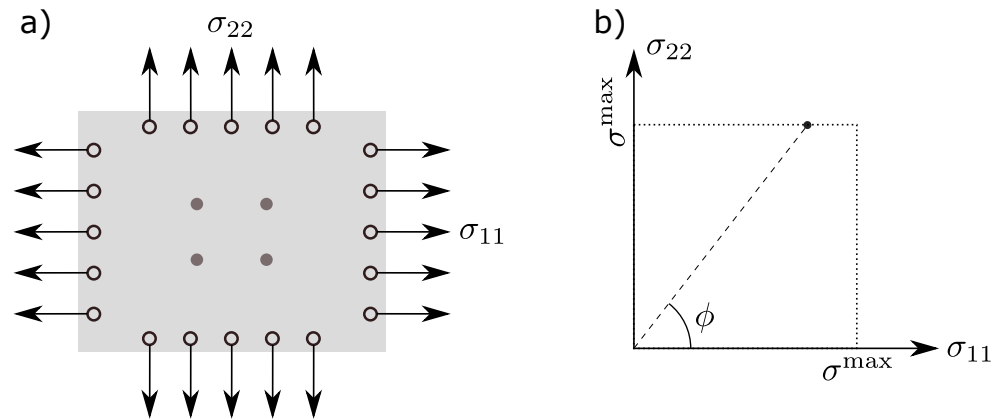


Figure 1. (a) A schematic of a biaxial experimental setup in which a thin planar tissue sample (in light gray) is mounted via rakes and two orthogonal forces are applied to induce stresses σ_{11} and σ_{22} , and the resulting strains are measured by tracking the locations of the markers (in dark gray). (b) The $\sigma_{11} - \sigma_{22}$ space, where the applied stresses lie on the dotted line with a finite number of protocol angles ϕ used.

2.1.2. Protocol Definition

In practice, there are two approaches to the biaxial experiment: (1) displacement-controlled, where known stretches are imposed and forces are measured; and (2) force-controlled, where known forces are applied and stretches are measured. Generally, the force-controlled approach is used as it is easier to implement. Therefore, in the force-controlled approach, different values of stresses σ_{11} and σ_{22} can be applied.

A single-angle biaxial protocol is defined as a straight line in the σ_{11} - σ_{22} space (Figure 1b). That is, the ratio between the two stresses is kept constant while the applied forces are increased until a maximum value $\sigma^{\max} = 200$ kPa. Thus, for a chosen angle ϕ , we apply

$$\sigma_{11} = \begin{cases} \sigma & \text{if } \phi \leq \frac{\pi}{4} \\ \tan(\phi)\sigma & \text{else} \end{cases} \quad (6)$$

$$\sigma_{22} = \begin{cases} \cot(\phi)\sigma & \text{if } \phi \leq \frac{\pi}{4} \\ \sigma & \text{else} \end{cases}, \quad (7)$$

where $\sigma \in [0, \sigma^{\max}]$. For σ , 100 linearly spaced observation points between zero and the maximum stress ($\sigma^{\max} = 200$ kPa) are used. The resulting strains are calculated by iteratively solving Equation (5) for $\lambda_{1,2}$ and thereby obtaining $e_{1,2}$. In practice, a combination of angles can be successively tested. We refer to this combination as the experimental protocol that needs to be optimally designed.

For each angle, it is easy to acquire large numbers of points as the sample is continuously stretched. However, to vary between angles, it is essential to restart the experiment at zero applied force, which further requires the “pre-conditioning” of the sample by cyclically applying small stretches. This makes it practically difficult to apply an arbitrarily large number of angles. Therefore, in practice, usually only five angles are tested.

2.2. Information-Theoretic Framework for Optimal Design

The problem of optimal design typically refers to the choice of a design of experiments such that the design is optimal with respect to a pre-determined statistical criterion. We propose that the information-theoretic measures naturally define such statistical criteria.

The central idea is that information gain [27,28] from an experiment or protocol—as quantified by the information-theoretic quantities of mutual information and conditional mutual information—can be directly used as a reasonable statistical criterion for optimal design. These quantities are described next after presenting the framework for optimal design.

2.2.1. Optimal Design Problem

Consider the following general model:

$$\mathbf{y} = \mathcal{M}(\boldsymbol{\theta}), \tag{8}$$

where \mathcal{M} denotes a forward model that takes $\boldsymbol{\theta} \in \mathbb{R}^m$ and outputs $\mathbf{y} \in \mathbb{R}^n$. Note that $\boldsymbol{\theta}$ may contain initial and boundary conditions of the model and that \mathbf{y} may subsume the output at many time-points in the case of a dynamic system. Subsequently, consider that the measurement model is as follows:

$$\mathbf{z} = \mathcal{H}_p(\mathbf{y}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}, \tag{9}$$

where \mathcal{H}_p represents the observation operator, $\mathbf{z} \in \mathbb{R}^d$ represents the measurement vector, and $\boldsymbol{\varepsilon}$ represents the vector of measurement error/noise. Note that the observation operator \mathcal{H}_p depends on the design of experiments, which specifies which quantities are measured. Given a set of possible $\mathcal{H}_p = \{\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_h\}$, and a statistical criterion $\mathcal{S}(\mathcal{H}_p)$ to be maximised, the optimal design is given by

$$\hat{\mathcal{H}}_p = \arg \max_{\mathcal{H}_p} \mathcal{S}(\mathcal{H}_p). \tag{10}$$

In the case of the biaxial experiments, the model \mathcal{M} represents the model for the force controlled experiment (Sections 2.1.1 and 2.1.2) and \mathcal{H}_p essentially denotes the experimental protocol (see Section 2.1.2) representing the combination of angles—with each representing a straight line in the σ_{11} – σ_{22} plane—along which the strain measurements of e_1 and e_2 are acquired. With the possible variation of each angle between 0 and $\pi/2$, the set Φ of possible angles ϕ is constructed through a uniform discretisation of the space between 0 and $\pi/2$ into α levels; thus,

$$\Phi = \{\phi_0, \phi_1, \dots, \phi_\alpha\}. \tag{11}$$

The possible set of protocols is then given by any combination of elements in Φ with the restriction that the number of elements in a protocol must be limited to \mathcal{C} . Thus, if $\bar{\Phi} \subset \Phi$ is a subset of angles representing a protocol, our set of protocols is given by

$$\mathcal{H}_p = \{\bar{\Phi} \subset \Phi \mid 1 \leq |\bar{\Phi}| \leq \mathcal{C}\}, \tag{12}$$

where $|\cdot|$ represents the number of elements in the set. In other words, we choose at least 1 and up to \mathcal{C} elements from Φ , with the total elements in \mathcal{H}_p being

$$|\mathcal{H}_p| = \binom{\alpha}{1} + \binom{\alpha}{2} \dots + \binom{\alpha}{\mathcal{C}}. \tag{13}$$

2.2.2. Information-Theoretic Quantities for Optimal Design

In the framework of Section 2.2.1, we propose that information-theoretic quantities of mutual information and conditional mutual information are a natural choice for the statistical criterion \mathcal{S} . Denoting the random variables associated with $\boldsymbol{\theta}$ and \mathbf{z} as $\boldsymbol{\Theta}$ and \mathbf{Z} , respectively, the mutual information (MI) between the parameters $\boldsymbol{\Theta}$ and the measurements \mathbf{Z} is defined as [27]

$$\mathcal{I}(\boldsymbol{\Theta}; \mathbf{Z}) = \int_{x_{\boldsymbol{\Theta}} \times x_{\mathbf{Z}}} p_{\boldsymbol{\Theta}, \mathbf{Z}}(\boldsymbol{\theta}, \mathbf{z}) \frac{p_{\boldsymbol{\Theta}, \mathbf{Z}}(\boldsymbol{\theta}, \mathbf{z})}{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) p_{\mathbf{Z}}(\mathbf{z})} d\boldsymbol{\theta} d\mathbf{z}, \tag{14}$$

where $p_X(x)$ represents the probability density of a random variable X with a realisation $X = x$ and support \mathcal{X}_X . The mutual information $\mathcal{I}(\Theta; \mathbf{Z})$ quantifies the amount of information that can be gained on average by one random variable—e.g., \mathbf{Z} —knowing about the other—e.g., Θ . Indeed, with this interpretation, MI is a good candidate for the statistical criterion \mathcal{S} for optimal design. For an individual parameter, Θ_i , or indeed for any combination of parameters $\{\Theta_i, \Theta_j\}$, the corresponding information gains can be similarly computed through $\mathcal{I}(\Theta_i; \mathbf{Z})$ and $\mathcal{I}(\{\Theta_i, \Theta_j\}; \mathbf{Z})$, respectively. Thus, while $\mathcal{I}(\Theta_i; \mathbf{Z})$ quantifies the information gain individually for the parameter Θ_i , the quantity $\mathcal{I}(\{\Theta_i, \Theta_j\}; \mathbf{Z})$ quantifies information gain for the pair $\{\Theta_i, \Theta_j\}$ jointly. A measure of correlation between the parameters Θ_i and Θ_j is, however, missing and is provided by conditional mutual information (CMI), defined as

$$\underbrace{\mathcal{I}(\Theta_i; \Theta_j | \mathbf{Z})}_{\text{I}} = \underbrace{\mathcal{I}(\Theta_i; \{\Theta_j, \mathbf{Z}\})}_{\text{II}} - \underbrace{\mathcal{I}(\Theta_i; \mathbf{Z})}_{\text{III}} . \tag{15}$$

The CMI $\mathcal{I}(\Theta_i; \Theta_j | \mathbf{Z})$ represents the additional information gained about the parameter Θ_i when both Θ_j and \mathbf{Z} are known (term II) relative to when only the measurements Θ_i alone are known (term III). Note that CMI is symmetrical—i.e., $\mathcal{I}(\Theta_i; \Theta_j | \mathbf{Z}) = \mathcal{I}(\Theta_j; \Theta_i | \mathbf{Z})$ —and can be interpreted as a measure of dependence between the parameters given the measurements \mathbf{Z} . It should also be noted that both MI and CMI are non-negative.

With the above background, many statistical measures can be constructed. For example:

1. The mutual information for any single parameter may be maximised, giving $\mathcal{S} = \mathcal{I}(\Theta_i; \mathbf{Z})$. This approach only concerns the posterior of the parameter Θ_i and ignores all other parameters;
2. The joint mutual information may be maximised, giving $\mathcal{S} = \mathcal{I}(\Theta; \mathbf{Z})$. In the sense of classical optimal design, this can be interpreted as D-optimal design. This is because D-optimal designs minimise the determinant of the inverse Fisher Information Matrix, and $\mathcal{S} = \mathcal{I}(\Theta; \mathbf{Z})$ measures the information gain in the joint Θ space;
3. The sum of individual parameter mutual information may be maximised, giving $\mathcal{S} = \sum_{i=1}^m \mathcal{I}(\Theta_i; \mathbf{Z})$. In the sense of classical optimal design, this can be interpreted as A-optimal design. This is because A-optimal design minimises the trace of the inverse Fisher Information Matrix, and $\mathcal{S} = \sum_{i=1}^m \mathcal{I}(\Theta_i; \mathbf{Z})$ measures the sum of the information gains for all the parameters;
4. Alternatively, one may seek to maximise individual parameter information gain while minimising pairwise CMI, thus seeking both small posterior variances and minimising pairwise correlations between the parameters. In this case, the statistical criterion is

$$\mathcal{S} = \sum_{i=1}^m \mathcal{I}(\Theta_i; \mathbf{Z}) - \tau \sum_{i=1}^m \sum_{j=i}^m \mathcal{I}(\Theta_i; \Theta_j | \mathbf{Z}), \tag{16}$$

where $\tau > 0$ is a regularisation parameter. Note that high CMI implies that a large amount of information can be gained only about a combination of the two parameters (for instance, their sum or product), but not for each parameter individually. Thus, we seek to minimise the CMI.

Note that the above list is not exhaustive, and based on the interpretations of MI and CMI, other criteria may be constructed based on the desired sense of optimality.

2.2.3. Estimating Mutual Information

In general, the forward model in Equation (8) is non-linear, and thus even if the observation operator is linear (implying linear combinations of the state are measured), the analytical computation of mutual information is intractable. Thus, the information-theoretic quantities of MI and CMI must be estimated. A common method is to generate samples of Θ through the specification of an appropriate prior probability density $p_{\Theta}(\theta)$. Denoting these N_s samples as $\theta^{(i)}, i = \{1, 2, \dots, N_s\}$, each $\theta^{(i)}$ can be propagated through the forward and observation models of Equations (8) and (9) to produce corresponding

samples of \mathbf{Z} , denoted as $\mathbf{z}^{(i)}$. The samples of $\theta^{(i)}$ and $\mathbf{z}^{(i)}$ can subsequently be used on non-parametric estimators of MI and CMI. Such non-parametric estimators can broadly be classified into two categories: kernel density estimators (KDE) [29] and k -nearest neighbour (kNN) estimators [30,31]. For an overview of such methods, we refer to [32]. While the estimator proposed by Kraskov et. al. [30] is widely used and performs very well across a range of scenarios, one of its drawbacks is that it suffers from higher errors when extreme correlations are present between the variables and/or when the data are effectively in a lower-dimensional manifold. Since we are working with models that specify explicit relationships between the variables through the forward and observation model, this is likely to be true for the data set of $(\theta^{(i)}, \mathbf{z}^{(i)})$. Thus, in this study, we employ the local non-uniformity correction (LNC) proposed in [33], which includes a correction term to the original estimator by Kraskov et al. [30]. This term accounts for strong dependencies between the variables through local principle component analysis [33]. The method of [33] is used for the estimation of all MIs, and CMIs are estimated from the difference of two MIs; see Equation (15).

2.2.4. Dimensionality Reduction for the Biaxial Experiment

One of the main difficulties in estimating information-theoretic quantities is related to the data dimension. Non-parametric estimation is particularly challenging whenever the data are close to manifolds embedded in high-dimensional spaces. This is indeed the case when a physical model relates parameters and observable quantities. One of the possible ways to overcome this difficulty, or at least to mitigate it, is (dimension or) model reduction, which aims at discovering the underlying low-dimensional structure of a set of data (a comprehensive review of the topic can be found in [34–37]). A large spectrum of methods has been proposed in the literature. In the present contribution, we adopt a local reduced-basis method (similar in spirit to the methods proposed in [38,39]). Let the strains computed by the model be $e_{1,2}(\sigma; \phi; k_1, k_2)$, where k_1 and k_2 are the model parameters $(k_1, k_2) \in \Omega_k \subset \mathbb{R}^2$, and $\sigma \in \Omega_\sigma \subset \mathbb{R}$ is the variable defined in Section 2.1.2. Let $n \in \mathbb{N}^*$; thus, we introduce the following approximation:

$$e_{1,2} \approx \sum_{i=1}^n \eta_i r_i(\sigma, \phi) s_i(k_1, k_2, \phi), \tag{17}$$

which is well defined by virtue of the Eckart–Young theorem. First, let us observe that a given protocol consists of a set of known angles Φ . An efficient way to construct the local reduced basis is therefore to introduce a Proper Orthogonal Decomposition (POD) for each of the angles $\phi_j \in \Phi$. This corresponds to the search for an approximation of the form

$$e_{1,2}^{(j)}(\sigma; \phi_j; k_1, k_2) \approx \sum_{i=1}^n \eta_i^{(j)} r_i^{(j)}(\sigma) s_i^{(j)}(k_1, k_2), \tag{18}$$

where $\langle r_i^{(j)}, r_k^{(j)} \rangle_{\Omega_\sigma} = \delta_{ik}$ and $\langle s_i^{(j)}, s_k^{(j)} \rangle_{\Omega_k} = \delta_{ik}$ ($\langle \cdot, \cdot \rangle_{\Omega_\sigma}$ with Ω_k being the standard L^2 scalar product). The error in the approximation is related to the number n of modes retained:

$$\|e_{1,2}^{(j)} - \sum_{i=1}^n \eta_i^{(j)} r_i^{(j)}(\sigma) s_i^{(j)}(k_1, k_2)\|_{L^2(\Omega_\sigma \times \Omega_k)}^2 = \sum_{i=n+1}^\infty \eta_i^{(j)2} \tag{19}$$

In the present work, a number $n = 4$ of modes proved to be sufficient in order to obtain errors smaller than 10^{-3} in L^2 norm in the solution reconstruction. This means that the set of elements $e_{1,2}(\sigma; \phi_j; k_1, k_2)$ was close to the linear subspace spanned by the first $n = 4$ modes $r_i^{(j)}$. Henceforth, instead of considering the discretised $e_{1,2}$ we consider their coordinates in the subspace given by

$$z_{1,2}^{(j)}|_i = \langle e_{1,2}, r_i^{(j)} \rangle_{\Omega_\sigma} = \eta_i^{(j)} s_i^{(j)}(k_1, k_2). \tag{20}$$

2.2.5. Validation of Results against Existing Methods

Several methods and criteria to define and reach an optimal design of experiments have been proposed [12]. Among them, D-optimality criterion attempts at maximising the determinant of the information matrix. In the present case, this is equivalent to minimize the determinant of the inverse of the average Hessian of the loss function we would introduce in a classical parameter estimation method. In a noisy setting, and, in particular, when the noise is Gaussian, this cost function is equivalent to minus the logarithm of the likelihood function. Let the misfit function be $f(\theta)$ and \mathbb{E}_{Θ} denote the expectation operator. The average of the Hessian reads:

$$H = \mathbb{E}_{\Theta}[\partial_{\theta}^2 f|_{\theta_*}], \quad (21)$$

where θ_* is the value of the parameter minimising the loss function.

2.2.6. Overview of Approach for the Biaxial Experiments

In the context of the biaxial experiments, the parameters are k_1 and k_2 , represented as random variables K_1 and K_2 , respectively. The variability in these parameters is considered to be uniform (thus imposing a uniform prior distribution) in the following intervals: $k_1 \in [5, 100]$ kPa and $k_2 \in [5, 80]$. For a single value of angle ϕ , the measurements are the strain values e_1 and e_2 and are measured at 100 points along the line defined by the angle ϕ . Here, we consider $\alpha = 16$ discrete values of possible measurement angles ϕ uniformly distributed between, and including, 0° and 90° . For each angle ϕ , separate reduced bases of four modes for e_1 and e_2 are constructed through POD over 400 values of (K_1, K_2) sampled uniformly in the aforementioned parametric space. Thus, for any angle ϕ , the dimensionality reduction approach projects e_1 and e_2 measured at 100 points along the line defined by ϕ to a basis of $4 + 4$ modes. For a given protocol consisting of multiple angles, the measurement vector \mathbf{z} (with a corresponding random variable \mathbf{Z}) is the collection of all the reduced basis representations of e_1 and e_2 along the angles in the protocol. Lastly, the maximum number of angles in a protocol is restricted to $\mathcal{C} = 5$, giving a total of 6884 unique combinations of the $\alpha = 16$ angles.

For the estimation of MI and CMI, a total of $N = 10,000$ values of (K_1, K_2) are uniformly distributed in the parametric space. For each sample $(k_1^{(i)}, k_2^{(i)})$, the numerical model of the biaxial experiment is run to produce $e_1^{(i)}$ and $e_2^{(i)}$, which are then projected on to the reduced basis, giving $\mathbf{z}^{(i)}$. The N triplets of $(k_1^{(i)}, k_2^{(i)}, \mathbf{z}^{(i)})$ are subsequently used for the estimation of MI and CMI through the LNC estimator (see Section 2.2.3). In Equation (16), we use $\tau = 1$.

3. Results and Discussion

For all the 6884 combinations of angles, three statistical criteria are evaluated: (i) $\mathcal{I}(K_1; \mathbf{Z})$, (ii) $\mathcal{I}(K_2; \mathbf{Z})$ and (iii) $\mathcal{I}(K_1; \mathbf{Z}) + \mathcal{I}(K_2; \mathbf{Z}) - \mathcal{I}(K_1; K_2 | \mathbf{Z})$. While the first two criteria aim to maximise the information gain about K_1 and K_2 individually, the third criterion aims to maximise the information gain about K_1 and K_2 simultaneously while minimising the information dependence between them. Figures 2–4 show the variation in these three criteria when grouped by the number of angles in a protocol. In these figures, the values of information criterion when using two approaches to uniformly discretise the angular space within protocols are also presented. Observations from these plots are as follows:

1. Generally, all the three information criteria increase with the increasing number of angles in the protocol. Intuitively, this is expected, as a higher number of angles implies more measurement data and hence a higher potential for the improved estimation of the parameters. This observation is true for the maximum information gain, minimum information gain and the mean information gain;
2. Across all the three criteria, it is observed that the uniform discretisation is not necessarily reflective of the best protocol for estimating the parameters. In fact, in

- most cases, the performance of uniform discretisation is close to the mean information gain observed across all the angle combinations;
- From Figures 2 and 3, it is observed that the angular combinations that maximise information gain for K_1 are not identical—and vary significantly when more than two angles are simultaneously used—to those that maximise information gain for K_2 . This further motivates the use of a criterion that balances information gains in both the parameters while minimising their interdependence;
 - Figure 4 shows that the best combinations that maximise a balanced criterion, such as $\mathcal{I}(K_1; \mathbf{Z}) + \mathcal{I}(K_2; \mathbf{Z}) - \mathcal{I}(K_1; K_2 | \mathbf{Z})$, are a trade-off between the combinations of angles that maximise $\mathcal{I}(K_1; \mathbf{Z})$ and $\mathcal{I}(K_2; \mathbf{Z})$ individually. For example, when five angles are considered, the angles that maximise $\mathcal{I}(K_1; \mathbf{Z})$ are $\phi_a = [66, 72, 78, 84, 90]$ and those that maximise $\mathcal{I}(K_2; \mathbf{Z})$ are $\phi_b = [30, 36, 42, 48, 54]$, while the combination that maximises $\mathcal{I}(K_1; \mathbf{Z}) + \mathcal{I}(K_2; \mathbf{Z}) - \mathcal{I}(K_1; K_2 | \mathbf{Z})$ is $[30, 36, 48, 78, 90]$, which has two angles from ϕ_a and three angles from ϕ_b . It should be noted that such a trade-off between maximising individual parameter gains is still significantly different to a uniform discretisation;
 - Finally, it is observed that the worst combinations are all low angles: $[0, 6, 12, 18, 24]$. This can be related to the fact that, at low angles, the applied stress is largely aligned along the stiff fibers of the tissue, thus resulting in lower strain values. Thus, the lower angles provide a small range of the observations, while the larger angles provide a larger range (Figure 5a), thereby containing more information about the parameters.

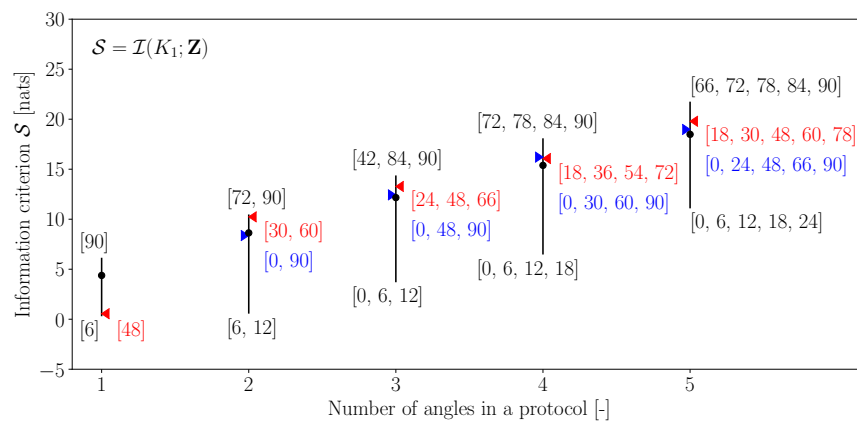


Figure 2. The variation of information criterion $\mathcal{S} = \mathcal{I}(K_1; \mathbf{Z})$ across the 6884 combinations grouped by the number of angles in a protocol. The vertical lines represent the variation around the mean value, which is shown in black circles. Black text shows the combinations that produce maximum and minimum values of \mathcal{S} . The red and blue pointers show \mathcal{S} for angle combinations that follow a uniform discretisation of the angular space between 0 and 90 degrees. Red and blue texts show the associated angle combinations.

From this point onward, we present results only for the balanced information criterion $\mathcal{S} = \mathcal{I}(K_1; \mathbf{Z}) + \mathcal{I}(K_2; \mathbf{Z}) - \mathcal{I}(K_1; K_2 | \mathbf{Z})$. Figure 6 shows the variation in \mathcal{S} across all the combinations (x-axis and in log-scale to capture the spread) grouped by the number of angles in a protocol and sorted according to the increasing order of \mathcal{S} within each such group. Within each group, observing the minimum and maximum values of \mathcal{S} shows that a better choice of angles can lead to more than a 100% increase in the information gain compared to a poor choice. Furthermore, this shows that good combinations of a lower number of angles can lead to higher information gain compared to a higher number of angles with poor combinations. For example, the maximum \mathcal{S} when only one angle is used is higher than many combinations with two to four angles. This emphasises the utility of optimal design and the proposed framework.

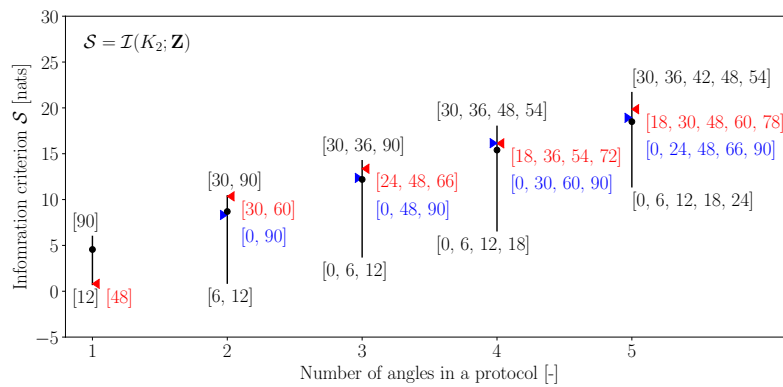


Figure 3. The variation of information criterion $\mathcal{S} = \mathcal{I}(K_2; \mathbf{Z})$ across the 6884 combinations grouped by the number of angles in a protocol. The vertical lines represent the variation around the mean value, which is shown in black circles. Black text shows the combinations that produce maximum and minimum values of \mathcal{S} . The red and blue pointers show \mathcal{S} for angle combinations that follow a uniform discretisation of the angular space between 0 and 90 degrees. Red and blue texts show the associated angle combinations.

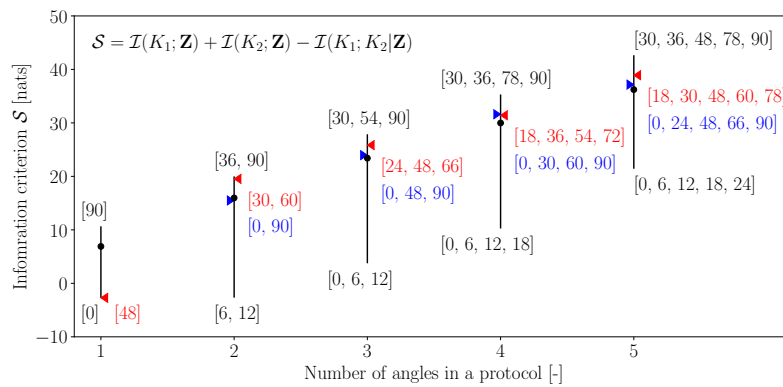


Figure 4. The variation of information criterion $\mathcal{S} = \mathcal{I}(K_1; \mathbf{Z}) + \mathcal{I}(K_2; \mathbf{Z}) - \mathcal{I}(K_1; K_2 | \mathbf{Z})$ across the 6884 combinations grouped by the number of angles in a protocol. The vertical lines represent the variation around the mean value, which is shown in black circles. Black text shows the combinations that produce maximum and minimum values of \mathcal{S} . The red and blue pointers show \mathcal{S} for angle combinations that follow a uniform discretisation of the angular space between 0 and 90 degrees. Red and blue texts show the associated angle combinations.

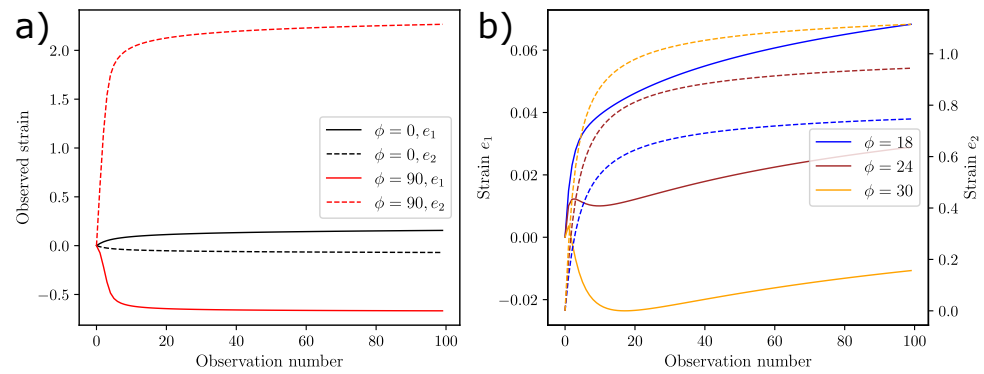


Figure 5. Representative observations from the model with $k_1 = 40$ kPa and $k_2 = 40$. (a) The observations using angles $\phi = 0$ and 90 degrees, with the latter covering a significantly larger range. (b) The change in observations the angle is changed from 18 , 24 to 30 degrees shows a transition in e_1 from positive to negative values, indicating a coupling between the two directions. Note that e_1 is shown here in solid lines (left y-axis) and e_2 is shown in dashed lines (right y-axis).

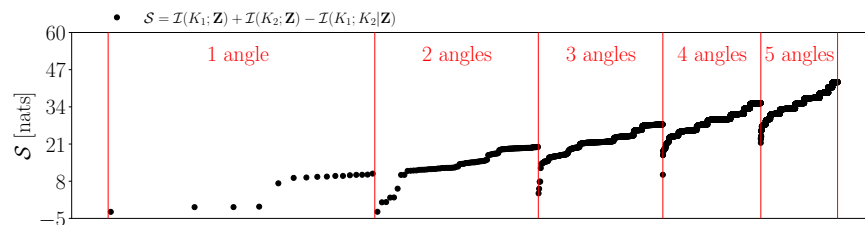


Figure 6. The variation of information criterion $\mathcal{S} = \mathcal{I}(K_2; \mathbf{Z}) + \mathcal{I}(K_2; \mathbf{Z}) - \mathcal{I}(K_1; K_2 | \mathbf{Z})$ across the 6884 combinations. The vertical red lines show the groupings with respect to the number of angles in a protocol and \mathcal{S} values are sorted in increasing order within each such grouping. The x-axis is represents the index associated with the protocol and is in logarithmic scale to capture the spread between one angle in a protocol (16 values) vs five angles in a protocol (4368 values).

Figure 7 shows \mathcal{S} for all the 6884 combinations in increasing order of magnitude, and Figure 8 shows a zoomed plot for the first 150 combinations along with the corresponding combinations of angles. Observing the index values of 26 (red) and 28 (blue) in Figure 8 shows that even though four combinations are used in the index 26 protocol, it produces a lower \mathcal{S} compared to when only a single angle is used in the index 28 protocol. Furthermore, since Figure 8 shows the first 150 out of 6884 combinations of Figure 7 (which is sorted in increasing order of \mathcal{S}), all combinations here are relatively low \mathcal{S} -producing protocols. Observing the high density of angles in the region $\phi < 24^\circ$ is indicative that lower values of angles—in particular, those less than 24° —are relatively less informative when compared to higher values of angles. This behaviour is also apparent in Figure 9, which shows \mathcal{S} values for protocols that use only one angle, and where a sharp jump can be observed when transitioning from 18° to 24° . This peculiar behaviour may be explained by the physics of the biaxial experiment. Looking at the resulting strains e_1 and e_2 of this transition (Figure 5b), we observe that the e_1 changes from positive to negative values. This behavior captures the important coupling between the two normal stresses and strains and is also related to the fiber dispersion in our constitutive model (Equation (1), [24]). It is remarkable and encouraging that the information-theoretic framework captures the physics of the problem without explicitly considering it in the framework. While for simpler low-dimensional models, the association between physics and optimal design may be relatively easy to see, inferring such behaviour is, in general, not trivial for more complex and higher-dimensional models.

Similarly to Figure 9, the results of the information-theoretic optimal design are further analysed for a higher number of angles. When two angles are considered, the \mathcal{S}

values in increasing order of magnitude and the corresponding angle combinations are shown in Figure 10. This figure re-iterates observations made previously: (i) the choice of combinations significantly affects the information gain, where the best combination gives approximately 20 nats more information compared to the worst combination; and (ii) generally speaking, higher angles are more informative compared to lower angles—in particular, angles below 24°. While a similar analysis for more than two angle combinations can be easily performed, the efficient visual representation of such results is cumbersome and avoided in this manuscript.

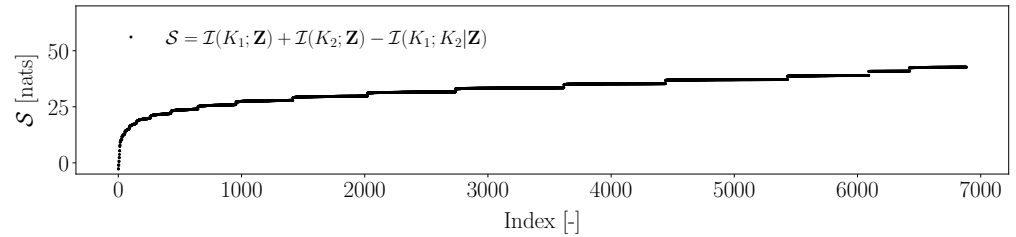


Figure 7. The variation of information criterion $\mathcal{S} = \mathcal{I}(K_1; \mathbf{Z}) + \mathcal{I}(K_2; \mathbf{Z}) - \mathcal{I}(K_1; K_2 | \mathbf{Z})$ across the 6884 combinations sorted in increasing order of \mathcal{S} .

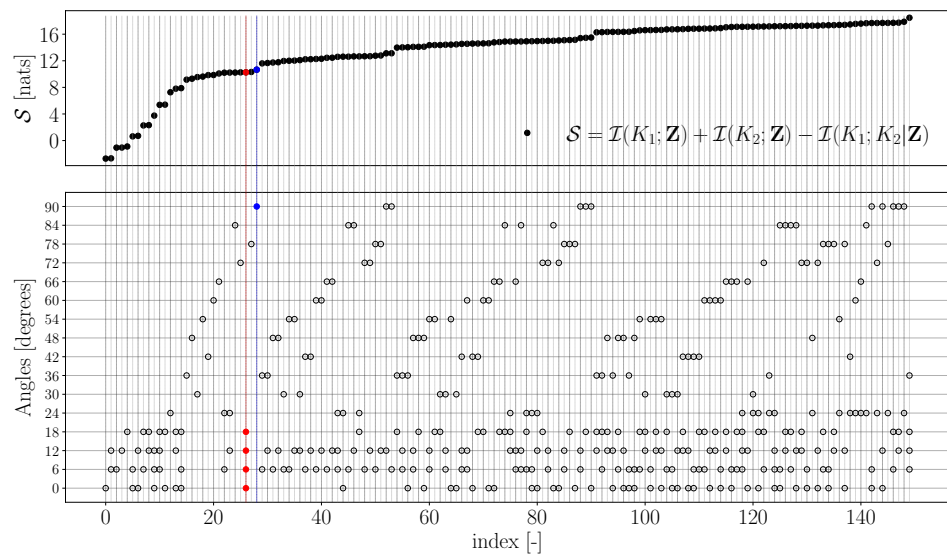


Figure 8. Zoomed view of the first 150 protocols from Figure 7. The upper panel shows \mathcal{S} and the lower panel shows the angles (by circles) in the corresponding protocol. The red points showcase a protocol with four measurement angles which yet produces a lower \mathcal{S} , implying a poorer protocol, with respect to the blue points which show a protocol with only one measurement angle.

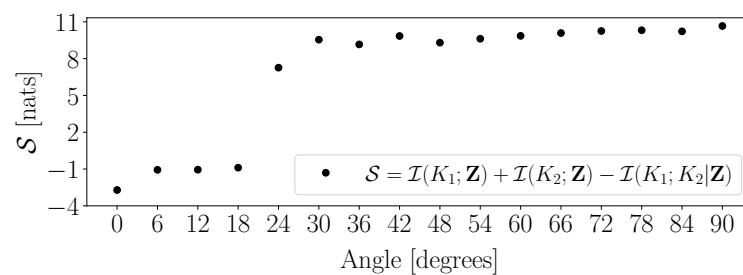


Figure 9. Information criterion against the angle when the protocols are restricted to a maximum of one angle.

To further illustrate the validity of the information-theoretic approach, a comparison with a classical method (see Section 2.2.5) is presented. For one and two angles in a protocol, Figure 11 shows a comparison between \mathcal{S} and the log of the determinant of the inverse Fisher Information Matrix, $\log |H^{-1}|$. It is encouraging that a high correspondence between the two metrics is observed. In particular, increases in \mathcal{S} , implying higher information gains, are accompanied by corresponding decreases in $\log |H^{-1}|$, implying a smaller volume of the parameter posteriors. A Pearson correlation coefficient of $r = -0.76$ is observed between \mathcal{S} and $\log |H^{-1}|$, implying a high similarity between the two metrics and validating the information-theoretic approach in part. We note that, when the number of the parameters become large, evaluating the Hessian would imply a non-negligible computational cost. On the contrary, the method used to evaluate the mutual information, as a primarily Monte Carlo-based estimation, is less severely dependent on the number of parameters. Furthermore, the computation of derivatives (either numerically or through adjoint based methods) may be cumbersome for certain types of models. Finally, we note that the effect of noise on information gain, and hence optimal design, can be easily assessed in the proposed framework by adding noise to the samples of \mathbf{Z} (see Equation (9)).

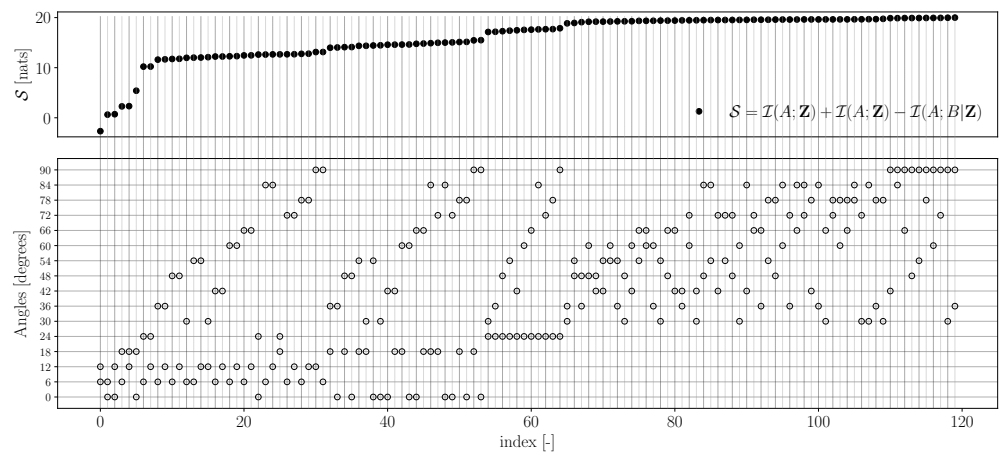


Figure 10. Information criterion against the angles when the protocols are restricted to only two angles. The upper panel shows \mathcal{S} and the lower panel shows the angles (with circles) in the corresponding protocol.

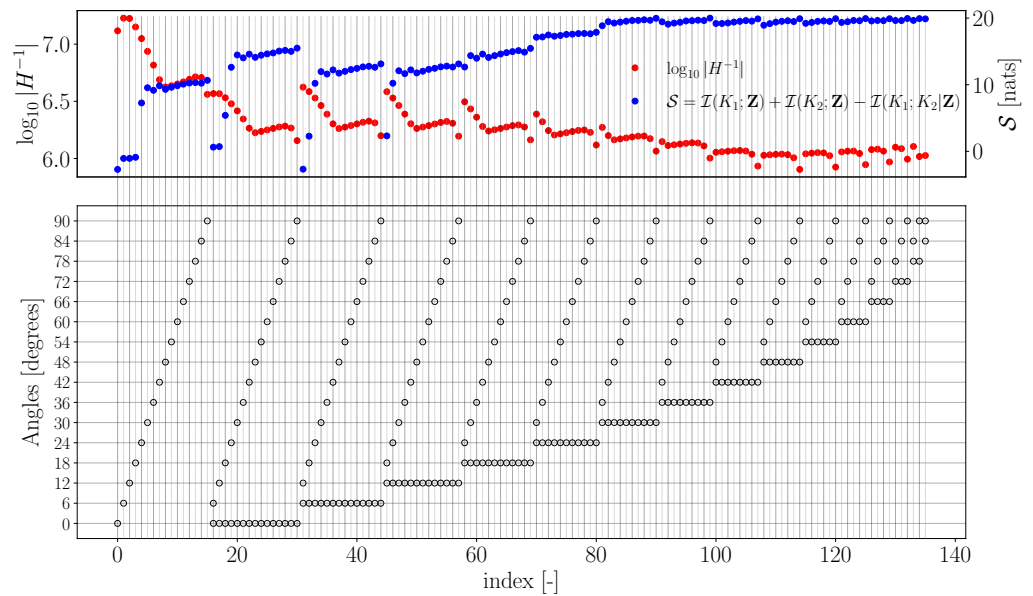


Figure 11. Information criterion \mathcal{S} (in blue) and the log of the determinant of the inverse Fisher Information Matrix $\log_{10} |H^{-1}|$ (in red) against the angles when the protocols are restricted to a maximum of two angles. The upper panel shows \mathcal{S} and $\log_{10} |H^{-1}|$, while the lower panel shows the angles (with circles) in the corresponding protocol.

4. Conclusions

A framework for optimal design based on information-theoretic quantities of mutual information and conditional mutual information is proposed. The framework treats information gain as the central criterion for inverse problems and proposes several information-theoretic frameworks for a desired sense of optimality. The capabilities of this framework are tested on the optimal design problem for biaxial experiments, where the effect of the angle combinations along which the strains are measured is assessed in terms of parameter estimation through information gain. Without including any physics-based reasoning, and purely through the information-theoretic measures, it is found that low angles $\leq 24^\circ$ are not very informative regarding the parameters relative to high angles. These observations are then found to be consistent based on physics-based reasoning, thereby showing the efficacy of the proposed framework. Furthermore, it is demonstrated that measurements for a low total number of angles which are carefully chosen can be more informative compared to the case when measurements along a high number of poorly chosen angles are acquired, thus highlighting both the importance of optimal design for biaxial experiments and the utility of the proposed framework in determining good angle combinations. The application of the proposed framework to classical optimal design is performed, and it is shown that the results produced by the new framework are consistent with classical frameworks.

5. Limitations and Future Work

While the proposed framework is shown to perform well on a two-parameter problem, its performance in higher parameter problems is not assessed. This assessment represents the primary limitation of this work and an area of future assessment. In particular, the problems envisaged are largely related to the performance of the MI and CMI estimators in higher dimensions of both parameters and the measurements. While a dimensionality reduction approach was adopted in this study to minimise the adverse effects of the latter, this may not be possible in many forward and inverse problems. Thus, a large area of future work is related to the development of efficient and robust MI and CMI estimators. Note that several approaches are being proposed by researchers to solve this problem; see for example [40–46]. Lastly, a thorough comparison against classical optimal design methods

(C, E, T and V-optimal designs, etc.) needs to be performed, along with the construction and analysis of corresponding information-theoretic metrics.

Author Contributions: All authors contributed equally to writing, editing, and reviewing the manuscript. S.P. and D.L. conceptualised the information-theoretic framework. A.A. and S.P. conceptualised the application to the biaxial experiments. AA wrote the numerical code for the biaxial experiment. S.P. and D.L. wrote the code for dimensionality reduction and the estimation of information-theoretic measures. All authors contributed equally to the analysis of the results. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Engineering and Physical Sciences Research Council of the UK (Grant reference EP/R010811/1 to SP and grant reference EP/P018912/1 and EP/P018912/2 to AA).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Holzappel, G.A. *Nonlinear Solid Mechanics*; Wiley: Chichester, UK, 2000; Volume 24.
- Zhang, W.; Feng, Y.; Lee, C.H.; Billiar, K.L.; Sacks, M.S. A generalized method for the analysis of planar biaxial mechanical data using tethered testing configurations. *J. Biomech. Eng.* **2015**, *137*, 064501.
- Labrosse, M.R.; Jafar, R.; Ngu, J.; Boodhwani, M. Planar biaxial testing of heart valve cusp replacement biomaterials: Experiments, theory and material constants. *Acta Biomater.* **2016**, *45*, 303–320.
- Humphrey, J.; Yin, F. On constitutive relations and finite deformations of passive cardiac tissue: I. A pseudostrain-energy function. *J. Biomech. Eng.* **1987**, *109*, 298–304.
- Laurence, D.; Ross, C.; Jett, S.; Johns, C.; Echols, A.; Baumwart, R.; Towner, R.; Liao, J.; Bajona, P.; Wu, Y.; Lee, C.H. An investigation of regional variations in the biaxial mechanical properties and stress relaxation behaviors of porcine atrioventricular heart valve leaflets. *J. Biomech.* **2019**, *83*, 16–27, doi:10.1016/j.jbiomech.2018.11.015.
- Jett, S.V.; Hudson, L.T.; Baumwart, R.; Bohnstedt, B.N.; Mir, A.; Burkhart, H.M.; Holzappel, G.A.; Wu, Y.; Lee, C.H. Integration of polarized spatial frequency domain imaging (pSFDI) with a biaxial mechanical testing system for quantification of load-dependent collagen architecture in soft collagenous tissues. *Acta Biomater.* **2020**, *102*, 149–168, doi:10.1016/j.actbio.2019.11.028.
- Billiar, K.L.; Sacks, M.S. Biaxial mechanical properties of the native and glutaraldehyde-treated aortic valve cusp: part II—a structural constitutive model. *J. Biomech. Eng.* **2000**, *122*, 327–335.
- Ross, C.; Laurence, D.; Wu, Y.; Lee, C.H. Biaxial Mechanical Characterizations of Atrioventricular Heart Valves. *J. Vis. Exp. JoVE* **2019**, doi:10.3791/59170.
- Maurel, W.; Thalmann, D.; Wu, Y.; Thalmann, N.M., Constitutive Modeling. In *Biomechanical Models for Soft Tissue Simulation*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 79–120.
- Holzappel, G.A.; Gasser, T.C.; Ogden, R.W. A new constitutive framework for arterial wall mechanics and a comparative study of material models. *J. Elast. Phys. Sci. Solids* **2000**, *61*, 1–48.
- May-Newman, K.; Yin, F.C.P. A constitutive law for mitral valve tissue. *J. Biomech. Eng.* **1998**, *120*, 38–47, doi:10.1115/1.2834305.
- Pukelsheim, F. *Optimal Design of Experiments*; Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, USA, 2006.
- Banks, H.T.; Holm, K.; Kappel, F. Comparison of optimal design methods in inverse problems. *Inverse Probl.* **2011**, *27*, 075002.
- Banks, H.T.; Dediu, S.; Ernstberger, S.L.; Kappel, F. Generalized sensitivities and optimal experimental design. *J. Inv. Ill-Posed Problems.* **2010**, *18*, 25–83, doi:10.1515/JIIP.2010.002.
- Banks, H.T.; Rubio, D.; Saintier, N.; Troparevsky, M.I. Optimal design techniques for distributed parameter systems. In Proceedings of the 2013 Conference on Control and Its Applications, San Diego, CA, USA, 8–10 July 2013; pp. 83–90.
- Lindley, D.V. On a measure of the information provided by an experiment. *Ann. Math. Stat.* **1956**, *27*, 986–1005.
- Sebastiani, P.; Wynn, H.P. Maximum entropy sampling and optimal Bayesian experimental design. *J. R. Stat. Soc. Ser. (Stat. Methodol.)* **2000**, *62*, 145–157.
- Capellari, G.; Chatzi, E.; Mariani, S.; others. Parameter identifiability through information theory. In Proceedings of the 2nd ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP), Rhodes Island, Greece, 15–17 June 2017; pp. 15–17.
- Bryant, C.; Terejanu, G. An information-theoretic approach to optimally calibrate approximate models. In Proceedings of the 50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition, Nashville, TN, 9–12 January USA, 2012; p. 153.
- Terejanu, G.; Upadhyay, R.R.; Miki, K. Bayesian experimental design for the active nitridation of graphite by atomic nitrogen. *Exp. Therm. Fluid Sci.* **2012**, *36*, 178–193.

21. Huan, X.; Marzouk, Y.M. Simulation-based optimal Bayesian experimental design for nonlinear systems. *J. Comput. Phys.* **2013**, *232*, 288–317.
22. Liepe, J.; Filippi, S.; Komorowski, M.; Stumpf, M.P. Maximizing the information content of experiments in systems biology. *PLoS Comput. Biol.* **2013**, *9*, e1002888.
23. Lewis, A.; Smith, R.; Williams, B.; Figueroa, V. An information theoretic approach to use high-fidelity codes to calibrate low-fidelity codes. *J. Comput. Phys.* **2016**, *324*, 24–43.
24. Gasser, T.C.; Ogden, R.W.; Holzapfel, G.A. Hyperelastic modelling of arterial layers with distributed collagen fibre orientations. *J. R. Soc. Interface* **2006**, *3*, 15–35, doi:10.1098/rsif.2005.0073.
25. Aggarwal, A. An improved parameter estimation and comparison for soft tissue constitutive models containing an exponential function. *Biomech. Model. Mechanobiol.* **2017**, *16*, 1309–1327, doi:10.1007/s10237-017-0889-3.
26. Aggarwal, A. Effect of Residual and Transformation Choice on Computational Aspects of Biomechanical Parameter Estimation of Soft Tissues. *Bioengineering* **2019**, *6*, 100, doi:10.3390/bioengineering6040100.
27. Pant, S.; Lombardi, D. An information-theoretic approach to assess practical identifiability of parametric dynamical systems. *Math. Biosci.* **2015**, *268*, 66–79.
28. Pant, S. Information sensitivity functions to assess parameter information gain and identifiability of dynamical systems. *J. R. Soc. Interface* **2018**, *15*, 20170871.
29. Moon, Y.I.; Rajagopalan, B.; Lall, U. Estimation of mutual information using kernel density estimators. *Physical Rev. E* **1995**, *52*, 2318.
30. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 066138.
31. Lombardi, D.; Pant, S. Nonparametric k-nearest-neighbor entropy estimator. *Phys. Rev. E* **2016**, *93*, 013310.
32. Beirlant, J.; Dudewicz, E.J.; Györfi, L.; Van der Meulen, E.C. Nonparametric entropy estimation: An overview. *Int. J. Math. Stat. Sci.* **1997**, *6*, 17–39.
33. Gao, S.; Ver Steeg, G.; Galstyan, A. Efficient estimation of mutual information for strongly dependent variables. In *Artificial Intelligence and Statistics; Proceedings of Machine Learning Research*, San Diego, CA, USA, 9–12 May 2015; pp. 277–286.
34. Benner, P.; Gugercin, S.; Willcox, K. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Rev.* **2015**, *57*, 483–531.
35. Benner, P.; Ohlberger, M.; Cohen, A.; Willcox, K. *Model Reduction and Approximation: Theory and Algorithms*; Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, USA, 2017.
36. Quarteroni, A.; Rozza, G.; others. *Reduced Order Methods for Modeling and Computational Reduction*; Springer: Berlin/Heidelberg, Germany, 2014; Volume 9.
37. Ma, Y.; Fu, Y. *Manifold Learning Theory and Applications*; CRC Press: Boca Raton, FL, USA, 2011.
38. Amsallem, D.; Haasdonk, B. PEBL-ROM: Projection-error based local reduced-order models. *Adv. Model. Simul. Eng. Sci.* **2016**, *3*, 1–25.
39. Maday, Y.; Stamm, B. Locally adaptive greedy approximations for anisotropic parameter reduced basis spaces. *SIAM J. Sci. Comput.* **2013**, *35*, A2417–A2441.
40. Belghazi, M.I.; Baratin, A.; Rajeshwar, S.; Ozair, S.; Bengio, Y.; Courville, A.; Hjelm, D. Mutual information neural estimation. In *Proceedings of the Machine Learning Research*, Stockholm, Sweden, 10–15 July 2018; pp. 531–540.
41. Singh, S.; Póczos, B. Generalized exponential concentration inequality for Rényi divergence estimation. In *International Conference on Machine Learning*; Proceedings of Machine Learning Research, International Conference on Machine Learning, Beijing, China, 22–24 June 2014; pp. 333–341.
42. Kleingesse, S.; Drovandi, C.; Gutmann, M.U. Sequential Bayesian experimental design for implicit models via mutual information. *Bayesian Anal.* **2021**, *1*, 1–30.
43. Fukumizu, K. Nonparametric Bayesian inference with kernel mean embedding. In *Modern Methodology and Applications in Spatial-Temporal Modeling*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 1–24.
44. Moon, K.R.; Hero, A.O. Ensemble estimation of multivariate f-divergence. In *Proceedings of the 2014 IEEE International Symposium on Information Theory*, Honolulu, HI, USA, 29 June–4 July 2014; pp. 356–360.
45. Brodu, N.; Crutchfield, J.P. Discovering Causal Structure with Reproducing-Kernel Hilbert Space ϵ -Machines. *arXiv* **2020** arXiv:2011.14821.
46. Gökmen, D.E.; Ringel, Z.; Huber, S.D.; Koch-Janusz, M. Phase diagrams with real-space mutual information neural estimation. *arXiv* **2021** arXiv:2103.16887.