

# TENSOR BASIS GAUSSIAN PROCESS MODELS OF HYPERELASTIC MATERIALS

Ari L. Frankel,\* Reese E. Jones, & Laura P. Swiler

Sandia National Laboratories, Livermore, California, 94551, USA

\*Address all correspondence to: Ari L. Frankel, Sandia National Laboratories, Quantitative Modeling and Analysis, Livermore, CA 94551, USA, E-mail: alfrank@sandia.gov

Original Manuscript Submitted: 12/18/2019; Final Draft Received: 5/5/2020

*In this work, we develop Gaussian process regression (GPR) models of isotropic hyperelastic material behavior. First, we consider the direct approach of modeling the components of the Cauchy stress tensor as a function of the components of the Finger stretch tensor in a Gaussian process. We then consider an improvement on this approach that embeds rotational invariance of the stress-stretch constitutive relation in the GPR representation. This approach requires fewer training examples and achieves higher accuracy while maintaining invariance to rotations exactly. Finally, we consider an approach that recovers the strain-energy density function and derives the stress tensor from this potential. Although the error of this model for predicting the stress tensor is higher, the strain-energy density is recovered with high accuracy from limited training data. The approaches presented here are examples of physics-informed machine learning. They go beyond purely data-driven approaches by embedding the physical system constraints directly into the Gaussian process representation of materials models.*

**KEY WORDS:** *Gaussian process regression, hyperelastic materials, physics-informed machine learning*

## 1. INTRODUCTION

Machine learning models have seen an explosion in development and application in recent years due to their flexibility and capacity for capturing the trends in complex systems (Hastie et al., 2016). Provided with sufficient data, the parameters of the model may be calibrated in such a way that the model gives high fidelity representations of the underlying data generating process (Raissi et al., 2017; Jones et al., 2018; Frankel et al., 2019a,b). Moreover, computational capabilities have grown such that constructing deep learning models over datasets of tens of thousands to millions of data points is now feasible (Dean et al., 2012). There remain, however, many applications in which the amount of data present is insufficient on its own to properly train the machine learning model. This may be due to a prohibitively large model that requires a correspondingly large amount of data to train and where training data is expensive to acquire. Furthermore, even with a wealth of data, it is possible that the machine learning model may yield behavior that is inconsistent with the expected trend of the model when the model is queried in an extrapolatory regime.

In such cases it is appealing to turn to a framework that allows the incorporation of physical principles and other *a priori* information to supplement the limited data and regularize the behavior of the model. This information can be as simple as a known set of constraints that the

regressor must satisfy, such as positivity or monotonicity with respect to a particular variable, or can be as complex as knowledge of the underlying data-generating process in the form of a partial differential equation. Consequently, the past few years have seen great interest in “physics-constrained” machine learning algorithms within the scientific computing community (Brunton et al., 2016; Jones et al., 2018; Lee and Carlberg, 2020; Ling et al., 2016; Lusch et al., 2018; Pan and Duraisamy, 2018; Raissi and Karniadakis, 2018). The overview paper by Karpatne et al. (2017) provides a taxonomy for theory-guided data science, with the goal of incorporating scientific consistency in the learning of generalizable models. Much research in physics-informed machine learning has focused on incorporating constraints in neural networks (Ling et al., 2016; Jones et al., 2018), often through the use of objective/loss functions that penalize constraint violation (Magiera et al., 2020).

In contrast, the focus in this paper is to incorporate rotational symmetries directly and exactly into Gaussian process (GP) representations of physical response functions. This approach has the advantages of avoiding the burden of a large training set that comes with a neural network model, and the inexact satisfaction of constraints that come with penalization of constraints in the loss function. There has been significant interest in the incorporation of constraints into Gaussian process regression (GPR) models recently (Bachoc et al., 2019; Da Veiga and Marrel, 2012; Jensen et al., 2013; López-Lopera et al., 2018; Raissi et al., 2017; Riihimäki and Vehtari, 2010; Solak et al., 2003; Yang et al., 2018). Many of these approaches leverage the analytic formulation of the GP to incorporate constraints through the likelihood function or i.e. the covariance function.

In this paper, the task of learning the six components of a symmetric stress tensor from the six components of a symmetric stretch tensor is formulated through a series of transformations so that it becomes a regression task of learning three coefficients that are a function of three invariants of the problem. The main contribution of this paper is the extension of GPR to enforce rotational invariance through a tensor basis expansion.

The paper is organized as follows: Section 2 presents an overview of constitutive models for hyperelastic materials. Sections 3 and 4 present GPR and the extension to a tensor basis GP, respectively. Section 5 presents a further extension of the tensor-basis GP to handle the strain energy potential. Section 6 provides results for a particular hyperelastic Mooney–Rivlin material, and Section 7 provides concluding discussion.

## 2. HYPERELASTIC MATERIALS

A hyperelastic material is a material that remains elastic (nondissipative) in the finite/large strain regime. In this context the fundamental deformation measure is the  $3 \times 3$  deformation gradient tensor  $\mathbf{F}$ :

$$\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}}, \quad (1)$$

which is the derivative of the current position  $\mathbf{x}$  with respect to position  $\mathbf{X}$  of the same material point in a chosen reference configuration. In an Eulerian frame (in an inertial frame from which the tensors are measured), the Finger tensor  $\mathbf{B}$ ,

$$\mathbf{B} = \mathbf{F}\mathbf{F}^T, \quad (2)$$

is the typical finite stretch measure, which is directly related to the Almansi strain, which measures the total deformation that a material has undergone relative to its initial configuration. Note

that this tensor is symmetric positive definite, as its eigenvalues are equal to the relative changes in length along principal axes, and negative lengths are not possible. [The choice of the Finger tensor is not limiting in terms of the generality of this formulation, given the equivalence of strain measures provided by the Seth–Hill (Hill, 1968) and Doyle–Ericksen (Doyle and Ericksen, 1956) formulae.] The deformation of a hyperelastic material requires an applied stress state, associated with a certain amount of energy, to arrive at that deformed state. For a hyperelastic material, the stress is solely a function of the current stretch (or strain) of the material. Hence, the major goal of material modeling of hyperelastic materials is to construct constitutive relations between the kinematic variable  $\mathbf{B}$  and the corresponding dynamic variable, the  $3 \times 3$  Cauchy stress tensor

$$\boldsymbol{\sigma} = \mathbf{f}(\mathbf{B}) = \frac{2}{|\mathbf{B}|^{1/2}} \mathbf{B} \frac{\partial \Phi}{\partial \mathbf{B}}, \quad (3)$$

which for a hyperelastic material is given by the derivative with respect to all the tensor components of a potential, namely the strain energy density  $\Phi$ , and  $|\mathbf{B}|$  is the determinant of the Finger tensor. Without additional arguments such as a structure tensor, this formulation is appropriate for isotropic hyperelastic materials. For further details please consult Malvern (1969), Ogden (1997), and Gurtin (1982).

Typical approaches to model these relations seek semiempirical formulations for the strain energy density with some parameters to be fit, which are then fit to experimental data. An example of this type of formulation will be discussed in a later section. In this work we consider nonparametric modeling of hyperelastic material responses.

### 3. GAUSSIAN PROCESS REGRESSION

GPR provides a nonparametric model for a response function given an input set of training data through a Bayesian update involving an assumed prior distribution and a likelihood tying the posterior distribution to observed data. We denote a GP prior for a function  $f$  by

$$f \sim \mathcal{GP}(0, K), \quad (4)$$

where we assume the GP has a nominal mean of 0, without loss of generality, and is described by a covariance function  $K$ . We adopt the commonly employed squared-exponential covariance function:

$$K(x, x') = \theta_1 \exp(-\theta_2 |x - x'|^2), \quad (5)$$

which has a scale parameter  $\theta_1$  and a (inverse-square) length parameter  $\theta_2$ .

A GP is defined such that any finite collection of realizations from the process are governed by a multivariate normal distribution. That is, for any set of observed realizations  $\mathbf{X}$  and prediction points  $\mathbf{X}^*$  with corresponding function values  $f(\mathbf{X})$  and  $f(\mathbf{X}^*)$ , the probability distribution  $p(f(\mathbf{X}^*), f(\mathbf{X}) | \mathbf{X}^*, \mathbf{X})$  is given by

$$\begin{bmatrix} f(\mathbf{X}) \\ f(\mathbf{X}^*) \end{bmatrix} = \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) & K(\mathbf{X}, \mathbf{X}^*) \\ K(\mathbf{X}^*, \mathbf{X}) & K(\mathbf{X}^*, \mathbf{X}^*) \end{bmatrix} \right), \quad (6)$$

where it is understood that the vectors and matrices presented are given in block form for multiple instances in  $\mathbf{X}$  and  $\mathbf{X}^*$ . GPR uses a GP as a prior over the function space for the data-generating process, and predictions proceed through the use of Bayes' rule. Upon observation of some initial set of noisy data points  $\mathbf{y} = y(\mathbf{X}) \sim \mathcal{N}(f(\mathbf{X}), \varepsilon^2)$ , where  $\varepsilon^2$  is the variance of

Gaussian noise in the function values, the probability distribution of the values of the GP at  $\mathbf{X}$  may be determined by forming the (posterior) conditional distribution of  $p(\mathbf{f}|\mathbf{y}, \mathbf{X})$ :

$$p(\mathbf{f}|\mathbf{y}, \mathbf{X}) = \frac{p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{X})}{p(\mathbf{y}|\mathbf{X})} = \frac{p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{X})}{\int p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{X})d\mathbf{f}}, \quad (7)$$

where  $\mathbf{f} = f(\mathbf{X})$  and the Gaussian likelihood is given by

$$p(\mathbf{y}|\mathbf{f}) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\varepsilon^2} \exp\left[-\frac{(y_i - f_i)^2}{2\varepsilon^2}\right]. \quad (8)$$

The push-forward posterior distribution for predictions at a new set of points  $\mathbf{X}^*$  has the following analytical solution:

$$\mathbf{f}^* = f(\mathbf{X}^*)|\mathbf{X}, \mathbf{y}, \mathbf{X}^* \sim \mathcal{N}[\mathbf{K}^*(\mathbf{K} + \varepsilon^2\mathbf{I})^{-1}\mathbf{y}, \mathbf{K}^{**} - \mathbf{K}^*(\mathbf{K} + \varepsilon^2\mathbf{I})^{-1}\mathbf{K}^{*T}], \quad (9)$$

where  $\mathbf{I}$  denotes the identity matrix,  $\mathbf{K} = K(\mathbf{X}, \mathbf{X})$ ,  $\mathbf{K}^* = K(\mathbf{X}^*, \mathbf{X})$ , and  $\mathbf{K}^{**} = K(\mathbf{X}^*, \mathbf{X}^*)$ . Then the predictive mean of the distribution at any new points  $\mathbf{X}^*$  is given by

$$\mathbb{E}[\mathbf{y}^*|\mathbf{X}, \mathbf{y}, \mathbf{X}^*] = \mathbf{K}^*(\mathbf{K} + \varepsilon^2\mathbf{I})^{-1}\mathbf{y}, \quad (10)$$

and the predictive variance, assuming the same noise level, is given by

$$\mathbb{V}[\mathbf{y}^*|\mathbf{X}, \mathbf{y}, \mathbf{X}^*] = \mathbf{K}^{**} - \mathbf{K}^*(\mathbf{K} + \varepsilon^2\mathbf{I})^{-1}\mathbf{K}^{*T} + \varepsilon^2\mathbf{I}, \quad (11)$$

where  $\mathbf{y}^* = y(\mathbf{X}^*)$ . This result shows the combination of uncertainty in the prediction due to epistemic uncertainty in the mean process (the first two terms on the right side) plus the aleatoric uncertainty of inherent variability in the measurements (the last term on the right side). In this work, although we will work with noiseless data, we assume a value of  $\varepsilon^2 = 10^{-10}$  in order to regularize the inversion of the covariance matrix.

The task that dominates the computational expense in constructing this model is the inversion of  $(\mathbf{K} + \varepsilon^2\mathbf{I})$ , or, equivalently, the solution of the linear system based on  $(\mathbf{K} + \varepsilon^2\mathbf{I})$  for either the mean or variance evaluations. Since  $\mathbf{K}$  is dense, the scaling is typically  $\mathcal{O}(N^3)$  for  $N$  training points. This limitation constrains GPR to at most  $N = 10,000$  since memory and computation time become prohibitive, although for this work we found that  $N = \mathcal{O}(100)$  was sufficient to reach reasonable accuracy, as will be shown later. Nominally the matrix is symmetric positive semidefinite, which enables efficient solution by Cholesky decompositions, although ill-conditioning is frequently an issue, especially for large dataset sizes. Ill-conditioning requires adding a nugget or large noise ( $\varepsilon \gg 1$ ) term to the covariance matrix to regularize the solution, using pseudoinverses via the singular value decomposition of the covariance matrix, or other greedy subset selection to reduce the matrix size.

It appears at first glance that the variance in Eq. (11) is nominally independent of the actual point values  $\mathbf{y}$ , and only depends on the locations of the selected data points  $\mathbf{X}$ . This is true for a fixed covariance function; however, we are typically interested in changing the GP hyperparameters to maximize the accuracy of the GP while balancing the model complexity. Traditionally, this is managed by tuning the hyperparameters to optimize  $p(\mathbf{y}|\mathbf{X})$ , which is the marginal-likelihood of the GP, and is frequently called the ‘‘model evidence.’’ Equivalently, we may optimize the logarithm of the model evidence  $L = \log p(\mathbf{y}|\mathbf{X})$  for numerical stability reasons:

$$L = -\frac{1}{2}\mathbf{y}^T (\mathbf{K} + \varepsilon^2\mathbf{I})^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K} + \varepsilon^2\mathbf{I}| - \frac{N}{2} \log 2\pi. \quad (12)$$

That is, we choose to tune the covariance hyperparameters  $\theta_1$  and  $\theta_2$  in Eq. (5) in order to maximize  $L$ . It is worth noting that solving this optimization problem requires multiple inversions of a dense covariance matrix, further increasing the cost of performing the inference. Further discussion of this approach can be found in Rasmussen and Williams (2006).

#### 4. TENSOR BASIS GAUSSIAN PROCESS

In this section we show how the standard GPR described in the previous section may be adapted to enforce rotational invariance through a tensor basis expansion. We call this formulation a tensor basis Gaussian process (TBGP).

##### 4.1 Tensor Basis Expansion

We consider the generic hyperelastic constitutive model of the form

$$\boldsymbol{\sigma} = \mathbf{f}(\mathbf{B}), \quad (13)$$

which, for any given analytic tensor valued function  $\mathbf{f}$ , may be expanded in an infinite series in terms of  $\mathbf{B}$  with fixed coefficients  $\bar{c}_n$ :

$$\boldsymbol{\sigma} = \sum_{n=0}^{\infty} \bar{c}_n \mathbf{B}^n. \quad (14)$$

Note that  $\mathbf{B}$  is symmetric, positive definite, and hence has a complete eigenbasis. Furthermore, it is clear that  $\boldsymbol{\sigma}$  and  $\mathbf{B}$  are coaxial, (i.e., have the same eigenbasis), due to the fact that we are considering an isotropic material. Since the tensors of interest are symmetric and of size  $3 \times 3$ , the Cayley-Hamilton theorem states that the tensor  $\mathbf{B}$  satisfies its corresponding characteristic polynomial

$$\mathbf{B}^3 - I_1 \mathbf{B}^2 + I_2 \mathbf{B} - I_3 \mathbf{I} = 0, \quad (15)$$

where we have defined the tensor invariants

$$\begin{aligned} I_1 &= \text{tr}(\mathbf{B}) &= \lambda_{B_1} + \lambda_{B_2} + \lambda_{B_3}, \\ I_2 &= \frac{1}{2} [\text{tr}(\mathbf{B})^2 - \text{tr}(\mathbf{B}^2)] &= \lambda_{B_1} \lambda_{B_2} + \lambda_{B_2} \lambda_{B_3} + \lambda_{B_3} \lambda_{B_1}, \\ I_3 &= \det(\mathbf{B}) &= \lambda_{B_1} \lambda_{B_2} \lambda_{B_3}, \end{aligned} \quad (16)$$

so-called because they are invariant under similarity transformations (i.e., rotations) of  $\mathbf{B}$ . Here,  $\lambda_{B_1}$ ,  $\lambda_{B_2}$ , and  $\lambda_{B_3}$  are the eigenvalues of  $\mathbf{B}$ , which are also a complete set of invariants. The theorem can be used as a recursion relation to write all powers of  $\mathbf{B}$  higher than 2 in terms of  $\mathbf{I}$ ,  $\mathbf{B}$ , and  $\mathbf{B}^2$  with coefficients that depend on the invariants and the unknown  $\bar{c}_i$ . Rather than seeking to identify the infinite number of fixed coefficients  $\bar{c}_i$  for a given constitutive relation, our task reduces to finding the three coefficients in the series expansion

$$\boldsymbol{\sigma} = c_0(I_1, I_2, I_3) \mathbf{I} + c_1(I_1, I_2, I_3) \mathbf{B} + c_2(I_1, I_2, I_3) \mathbf{B}^2, \quad (17)$$

where  $c_i$  is a function of the invariants; this notation will be suppressed for the remainder of this work for clarity.

One issue in many machine learning models of constitutive relations is that they do not always preserve rotational objectivity; that is, rotating the frame in which the deformations are measured should not change the physics predictions, but the machine learning models do not preserve this. For example, the GP formulation in Section 2 does not ensure that rotating the basis for  $\mathbf{B}$  (while keeping the components fixed) would give the corresponding fixed components for the stress predictions in the rotated basis. The advantage of this reduced expansion is that it embeds rotational objectivity in its structure. To see this, let  $\mathbf{R}$  be an orthogonal/rotation tensor with inverse given by  $\mathbf{R}^{-1} = \mathbf{R}^T$ . The rotation of  $\sigma$  in the original coordinate frame to the frame defined by  $\mathbf{R}$  is given by

$$\sigma' \equiv \mathbf{R}\sigma\mathbf{R}^T = \mathbf{R}f(\mathbf{B})\mathbf{R}^T. \quad (18)$$

Invoking the tensor basis expansion gives

$$\begin{aligned} \sigma' &= c_0\mathbf{R}\mathbf{R}\mathbf{R}^T + c_1\mathbf{R}\mathbf{B}\mathbf{R}^T + c_2\mathbf{R}\mathbf{B}^2\mathbf{R}^T \\ &= c_0\mathbf{R}\mathbf{R}\mathbf{R}^T + c_1\mathbf{R}\mathbf{B}\mathbf{R}^T + c_2\mathbf{R}\mathbf{B}\mathbf{R}^T\mathbf{R}\mathbf{B}\mathbf{R}^T = f(\mathbf{R}\mathbf{B}\mathbf{R}^T) \equiv f(\mathbf{B}'), \end{aligned} \quad (19)$$

which holds since the eigenvalues, and hence invariants and the coefficient functions, do not change upon application of  $\mathbf{R}$ . In general, an Eulerian tensor function of an Eulerian tensor argument must be objective in the sense it responds to a rotation of its argument with a corresponding rotation of the function value:

$$\mathbf{R}f(\mathbf{B})\mathbf{R}^T = f(\mathbf{R}\mathbf{B}\mathbf{R}^T). \quad (20)$$

This is precisely what we refer to as rotational invariance.

## 4.2 Application to Gaussian Process Modeling

The task of regression now falls to learning the coefficients  $c_0$ ,  $c_1$ , and  $c_2$  as a function of the invariants. This task is compressed from the original problem of having to learn six stress components from six strain components with the added benefit of enforcing the rotational invariance that provides this reduction. Since GPR guarantees theoretical asymptotic consistency of the predictions, using GPR to learn the coefficients as a function of the invariants indirectly guarantees asymptotic consistency of the stress tensor predictions since they are just linear combinations of the three coefficients.

Suppose we are given a dataset of pairs of tensors  $(\mathbf{B}, \sigma)$ . Under the assumption that the tensors may be diagonalized, the collinearity of the tensor basis expansion [Eq. (17)] implies that they are diagonalized by the same eigenvector matrix  $\mathbf{Q}$  but with different eigenvalue matrices,  $\Lambda_\sigma$  and  $\Lambda_B$ , respectively:

$$\sigma = \mathbf{Q}\Lambda_\sigma\mathbf{Q}^T, \quad (21)$$

$$\mathbf{B} = \mathbf{Q}\Lambda_B\mathbf{Q}^T. \quad (22)$$

Then for the given input-output pair, the values of the coefficients for the given set of eigenvalues are given by the solution to a  $3 \times 3$  linear system of equations as

$$\begin{bmatrix} \lambda_{\sigma_1} \\ \lambda_{\sigma_2} \\ \lambda_{\sigma_3} \end{bmatrix} = \begin{bmatrix} 1 & \lambda_{B_1} & \lambda_{B_1}^2 \\ 1 & \lambda_{B_2} & \lambda_{B_2}^2 \\ 1 & \lambda_{B_3} & \lambda_{B_3}^2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix}. \quad (23)$$

This linear system may be inverted easily to yield the coefficients  $c_i(\mathbf{B}, \boldsymbol{\sigma})$  for each training point in the dataset, and the invariants  $I_i$  of  $\mathbf{B}$  may also be computed straightforwardly from the eigenvalues given in Eq. (16). The three invariants  $I_i$  for each observation are accumulated into a matrix. They take the place of the feature matrix  $\mathbf{X}$  presented in the previous section, and a corresponding matrix of  $c_i$  replaces  $\mathbf{f}$ . We can thus readily extend the GPR approach to infer the coefficients as functions of the invariants of the input tensors and thus construct the representation of the function  $\boldsymbol{\sigma} = c_0\mathbf{I} + c_1\mathbf{B} + c_2\mathbf{B}^2$  given in Eq. (17).

In cases where  $\mathbf{B}$  has repeated eigenvalues, Eq. (23), as it stands, is not invertible. Physically, this corresponds to the case where equal deformations are applied along different axes. In such a case, the Cayley-Hamilton theorem and continuity of the deformation to stress mapping require that the derivative of the stress eigenvalue with respect to the repeated Finger tensor eigenvalue must be zero. If  $\lambda_{B_1} = \lambda_{B_2} \neq \lambda_{B_3}$ ,

$$0 = c_1 + 2\lambda_{B_1}c_2, \quad (24)$$

should be substituted for one of the redundant equations. If  $\lambda_{B_1} = \lambda_{B_2} = \lambda_{B_3}$ , the second derivative must also be zero, giving the additional equation

$$c_2 = 0, \quad (25)$$

to replace another of the redundant equations in the system of equations, Eq. (23), in which case the result is trivially  $c_0 = \lambda_{\sigma_1}$  and  $c_1 = c_2 = 0$ . This can be interpreted physically as a volumetric deformation leads to a pressure for an isotropic, elastic material, as expected. Combining these auxiliary conditions with the fact that  $\mathbf{B}$  is symmetric positive definite, it is always possible to map the Finger deformation and Cauchy stress tensors to set of coefficients that satisfy the tensor basis expansion and forms a well-posed regression.

One issue with this approach is that the inclusion of measurement error of the stress tensor is not straightforward. Including Gaussian errors in the coefficients will lead to a more complex noise process in the stress predictions, where the stress tensor components will have unequal variances and potentially be correlated. While in this work we are considering noiseless data, extending GPR to consider noisy data is a problem worth considering in future work.

### 4.3 Example: Matrix Exponential

To illustrate the impact of embedding rotational invariance in the GP formulation, we consider the representation of the matrix exponential

$$\mathbf{S} = \exp(\mathbf{B}) = \mathbf{Q}\exp(\boldsymbol{\Lambda})\mathbf{Q}^T, \quad (26)$$

from a limited number of training samples. As in Section 4.2,  $\mathbf{B} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T$  is a symmetric diagonalizable matrix with eigenvector matrix  $\mathbf{Q}$  and diagonal eigenvalue matrix  $\boldsymbol{\Lambda}$ . Clearly the matrix exponential maintains rotational invariance under application of a rotation  $\mathbf{R}$ , and the series representation of  $\exp(\mathbf{B})$  demonstrates the coaxiality of  $\exp(\mathbf{B})$  and  $\mathbf{B}$ . Given Eq. (23), the expansion coefficients may be determined from

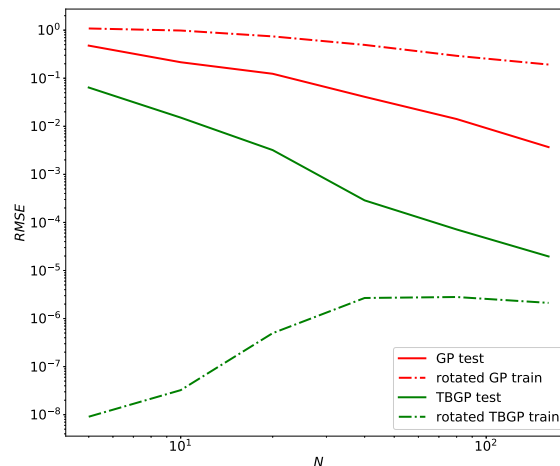
$$\begin{bmatrix} \exp \lambda_1 \\ \exp \lambda_2 \\ \exp \lambda_3 \end{bmatrix} = \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 \\ 1 & \lambda_2 & \lambda_2^2 \\ 1 & \lambda_3 & \lambda_3^2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix}. \quad (27)$$

To create the dataset, we draw random uniformly distributed  $3 \times 3$  matrices with entries between  $[0, 1]$  and compute their symmetric part. A single GP is formed over the six independent tensor components (the upper triangular part) of  $\mathbf{B}$  to predict the six independent output tensor components. This formulation is compared against the TBGP formulation, where the three invariants are used to predict the three expansion coefficients with a single GP. The *scikit-learn* library (Pedregosa et al., 2011) was used to train the GPs in both cases. The root-mean-square error is evaluated in each case at 10,000 testing points for validation. The input training points are then rotated randomly, and the GP prediction is evaluated at the inputs.

Figure 1 shows the results of the testing error as a function of increasing training points for the GP and TBGP formulations. Since the rotationally invariant formulation does not take the orientation of the eigenvectors into account, it is expected that the prediction error on the modified inputs in this case would be small, whereas the prediction error for the normal GP could be quite large. The TBGP error is indeed 1–2 orders of magnitude lower than the GP error on the testing sets, and the error on the randomly rotated training set is also over 5 orders of magnitude lower, demonstrating that the TBGP formulation learns the underlying function much more quickly than a standard GP. It also appears to take an order of magnitude or more data before the GP error catches up to the error that the TBGP had on the smallest dataset. The error on the rotated training set appears to increase, which we attribute to a combination of increasing condition number of the covariance matrix and accumulated interpolation error in the GP through the training points from the use of regularization to maintain invertibility. Even with these effects, the TBGP has very little error throughout the tests and is uniformly the better choice.

## 5. STRAIN ENERGY POTENTIAL GAUSSIAN PROCESS

The tensor basis GP formulation is quite powerful and could be made general to many different types of processes. For hyperelastic materials, where the stress is the derivative of a potential, it is possible to take this approach one step further with an alternate formulation. As in Eq. (3), we



**FIG. 1:** The root-mean-square error (RMSE) in predicting the matrix exponential for the GP and TBGP formulations as a function of training set size. The results also show the RMSE of the regressors evaluated at random rotations of the training set pairs.



define the strain-energy density function  $\Phi$  such that the stress tensor may be computed as

$$\boldsymbol{\sigma} = \frac{2}{I_3^{1/2}} \mathbf{B} \frac{\partial \Phi}{\partial \mathbf{B}}, \quad (28)$$

for some appropriate  $\Phi(\mathbf{B})$ . The strain-energy density is an invariant scalar quantity and cannot depend on the rotation of the frame; thus for an isotropic hyperelastic material it is preferable to express the strain-energy density as a function of the invariants of  $\mathbf{B}$ . Thus the expression for the stress tensor may be expanded as

$$\boldsymbol{\sigma} = \frac{2}{I_3^{1/2}} \mathbf{B} \left( \frac{\partial \Phi}{\partial I_1} \frac{\partial I_1}{\partial \mathbf{B}} + \frac{\partial \Phi}{\partial I_2} \frac{\partial I_2}{\partial \mathbf{B}} + \frac{\partial \Phi}{\partial I_3} \frac{\partial I_3}{\partial \mathbf{B}} \right), \quad (29)$$

where we have applied the chain rule for the strain-energy density partial derivatives. The derivatives of the invariants with respect to the original tensor may be computed directly using matrix calculus identities (Bonet and Wood, 1997). Specifically, we use

$$\begin{aligned} \frac{\partial I_1}{\partial \mathbf{B}} &= \mathbf{I}, \\ \frac{\partial I_2}{\partial \mathbf{B}} &= I_1 \mathbf{I} - \mathbf{B}, \\ \frac{\partial I_3}{\partial \mathbf{B}} &= I_3 \mathbf{B}^{-1}, \end{aligned} \quad (30)$$

and the expression for  $\boldsymbol{\sigma}$  simplifies to

$$\boldsymbol{\sigma} = \frac{2}{I_3^{1/2}} \left[ I_3 \frac{\partial \Phi}{\partial I_3} \mathbf{I} + \left( \frac{\partial \Phi}{\partial I_1} + I_1 \frac{\partial \Phi}{\partial I_2} \right) \mathbf{B} - \frac{\partial \Phi}{\partial I_2} \mathbf{B}^2 \right]. \quad (31)$$

This expression explicitly takes the form of a tensor-basis expansion of the type in Eq. (17), where we have

$$\begin{aligned} c_0 &= 2I_3^{1/2} \frac{\partial \Phi}{\partial I_3}, \\ c_1 &= \frac{2}{I_3^{1/2}} \left( \frac{\partial \Phi}{\partial I_1} + I_1 \frac{\partial \Phi}{\partial I_2} \right), \\ c_2 &= -\frac{2}{I_3^{1/2}} \frac{\partial \Phi}{\partial I_2}. \end{aligned} \quad (32)$$

For a given set of coefficients and invariants, the corresponding partial derivatives can be evaluated using the following relations:

$$\begin{aligned} \frac{\partial \Phi}{\partial I_1} &= \frac{I_3^{1/2}}{2} (c_1 + c_2 I_1), \\ \frac{\partial \Phi}{\partial I_2} &= -\frac{c_2 I_3^{1/2}}{2}, \\ \frac{\partial \Phi}{\partial I_3} &= \frac{c_0}{2I_3^{1/2}}. \end{aligned} \quad (33)$$

Thus given a set of observed pairs of  $(\boldsymbol{\sigma}, \mathbf{B})$  and Eqs. (23) and (33), we can infer the corresponding values of the gradient of the strain-energy density function. Furthermore, we “ground” the function (i.e., remove the indeterminacy of calibration of  $\Phi$  from derivative data) by choosing a zero-point energy for a set of invariants where the material has not been deformed. Hence, we augment the gradient information with the datum corresponding to  $\boldsymbol{\sigma}(\mathbf{B} = \mathbf{I}) = 0$ :

$$\Phi(I_1 = 3, I_2 = 3, I_3 = 1) = 0. \quad (34)$$

The GP regression technique can be extended to take advantage of derivative information since the derivative of a GP is also a GP. Specifically, the covariance between the derivatives of the stress-energy density between two different points in invariant space,  $\mathbb{I} = (I_1, I_2, I_3)$  and  $\mathbb{I}' = (I'_1, I'_2, I'_3)$ , can be evaluated as

$$\text{Cov} \left[ \frac{\partial \Phi(\mathbb{I})}{\partial I_i}, \Phi(\mathbb{I}') \right] = \frac{\partial K(\mathbb{I}, \mathbb{I}')}{\partial I_i}, \quad (35)$$

$$\text{Cov} \left[ \frac{\partial \Phi(\mathbb{I})}{\partial I_i}, \frac{\partial \Phi(\mathbb{I}')}{\partial I'_j} \right] = \frac{\partial^2 K(\mathbb{I}, \mathbb{I}')}{\partial I_i \partial I'_j}. \quad (36)$$

Using these relations, a GP may be formed simultaneously over  $\Phi$  at the grounding point and its derivatives over the dataset by using the block covariance matrix

$$\mathbf{K}_\Phi(\mathbb{I}, \mathbb{I}') = \begin{bmatrix} \frac{\partial^2 K}{\partial I_1 \partial I'_1} & \frac{\partial^2 K}{\partial I_1 \partial I'_2} & \frac{\partial^2 K}{\partial I_1 \partial I'_3} \\ \frac{\partial^2 K}{\partial I_2 \partial I'_1} & \frac{\partial^2 K}{\partial I_2 \partial I'_2} & \frac{\partial^2 K}{\partial I_2 \partial I'_3} \\ \frac{\partial^2 K}{\partial I_3 \partial I'_1} & \frac{\partial^2 K}{\partial I_3 \partial I'_2} & \frac{\partial^2 K}{\partial I_3 \partial I'_3} \end{bmatrix}, \quad (37)$$

where each entry is a matrix corresponding to the covariance between the individual derivatives of  $\Phi$  evaluated between each of the training data points. The grounding point  $\mathbb{I}_g = (3, 3, 1)$  is included by augmenting this matrix with an additional row and column:

$$\mathbf{K}_g(\mathbb{I}, \mathbb{I}') = \begin{bmatrix} K(\mathbb{I}_g, \mathbb{I}_g) & \frac{\partial K(\mathbb{I}_g, \mathbb{I}')}{\partial I'_1} & \frac{\partial K(\mathbb{I}_g, \mathbb{I}')}{\partial I'_2} & \frac{\partial K(\mathbb{I}_g, \mathbb{I}')}{\partial I'_3} \\ \frac{\partial K(\mathbb{I}, \mathbb{I}_g)}{\partial I_1} & & & \\ \frac{\partial K(\mathbb{I}, \mathbb{I}_g)}{\partial I_2} & & \mathbf{K}_\Phi(\mathbb{I}, \mathbb{I}') & \\ \frac{\partial K(\mathbb{I}, \mathbb{I}_g)}{\partial I_3} & & & \end{bmatrix}. \quad (38)$$

In a slight abuse of notation, in the matrix  $\mathbf{K}_\Phi(\mathbb{I}, \mathbb{I}')$  in Eq. (37) and the matrix  $\mathbf{K}_g(\mathbb{I}, \mathbb{I}')$  in Eq. (38), the arguments  $\mathbb{I}$  and  $\mathbb{I}'$  should be interpreted as matrices of the invariants at the training points (as opposed to individual points).

The GP mean of the potential,  $\Phi$ , and its gradient with respect to the invariants,  $\nabla_{\mathbb{I}}\Phi = \{\frac{\partial\Phi}{\partial I_1}, \frac{\partial\Phi}{\partial I_2}, \frac{\partial\Phi}{\partial I_3}\}$ , at a new point  $\mathbb{I}^* = (I_1^*, I_2^*, I_3^*)$  may then be predicted with

$$\mathbb{E}[(\Phi(\mathbb{I}^*), \nabla_{\mathbb{I}}\Phi(\mathbb{I}^*))] = \mathbf{K}_g(\mathbb{I}^*, \mathbb{I}) \mathbf{w}, \quad (39)$$

where we have defined the weight vector

$$\mathbf{w} = \mathbf{K}_g^{-1}(\mathbb{I}, \mathbb{I}') \begin{bmatrix} 0 & \frac{\partial\Phi}{\partial I_1} & \frac{\partial\Phi}{\partial I_2} & \frac{\partial\Phi}{\partial I_3} \end{bmatrix}^T, \quad (40)$$

where the right-hand side partial derivatives are at the observed data points, and  $\mathbf{K}_g(\mathbb{I}^*, \mathbb{I})$  is the covariance between the test points and the training points augmented with the ground point. This expression is analogous to Eq. (10), although here we have omitted the noise term. The gradient of the potential  $\Phi$ , and hence the stress  $\boldsymbol{\sigma}$ , may be evaluated using the same weight vector.

Although this formulation sets a coregionalization model for multioutput GPR, the inclusion of derivative information in the covariance matrix has a tendency to exacerbate ill-conditioning and round-off errors in the inference process (Wu et al., 2017). Thus it is expected that prediction accuracy may deteriorate in certain circumstances.

## 6. RESULTS: MOONEY–RIVLIN MATERIAL

In this section we consider the application of GPR to predicting data drawn from the stress response of the deformation of a hyperelastic material. The underlying truth model will be assumed to be a compressible Mooney–Rivlin material with strain-energy density function

$$\Phi = k_1(I_3^{-1/2}I_1 - 3) + k_2(I_3^{-2/3}I_2 - 3) + k_3(I_3^{1/2} - 1)^2, \quad (41)$$

where we take  $k_1 = 0.162$  MPa,  $k_2 = 0.0059$  MPa, and  $k_3 = 10$  MPa (Marckmann and Verron, 2006) and we make  $k_3$  large ( $k_3 \gg k_1, k_2$ ) to effect a nearly incompressible material response.

We generate realizations of arbitrary mixed compression/tension/shear states using the following procedure. Let  $\mathbf{V}$  be a diagonal matrix of randomly sampled positive values between  $0 < l \leq 1 < u$ , and let  $\mathbf{R}$  be a random rotation matrix sampled uniformly on  $\text{SO}(3)$ . We then employ the polar decomposition of the deformation gradient tensor

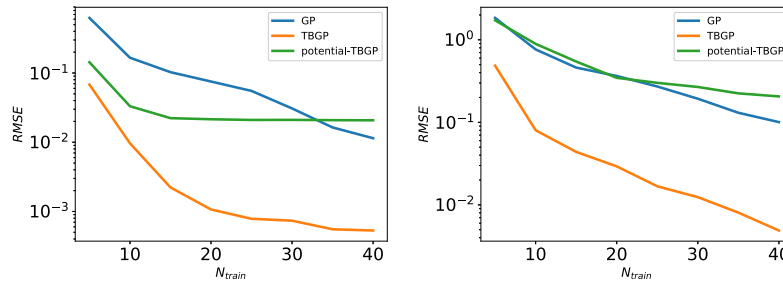
$$\mathbf{F} = \mathbf{R}\mathbf{V}, \quad (42)$$

with corresponding Finger tensor

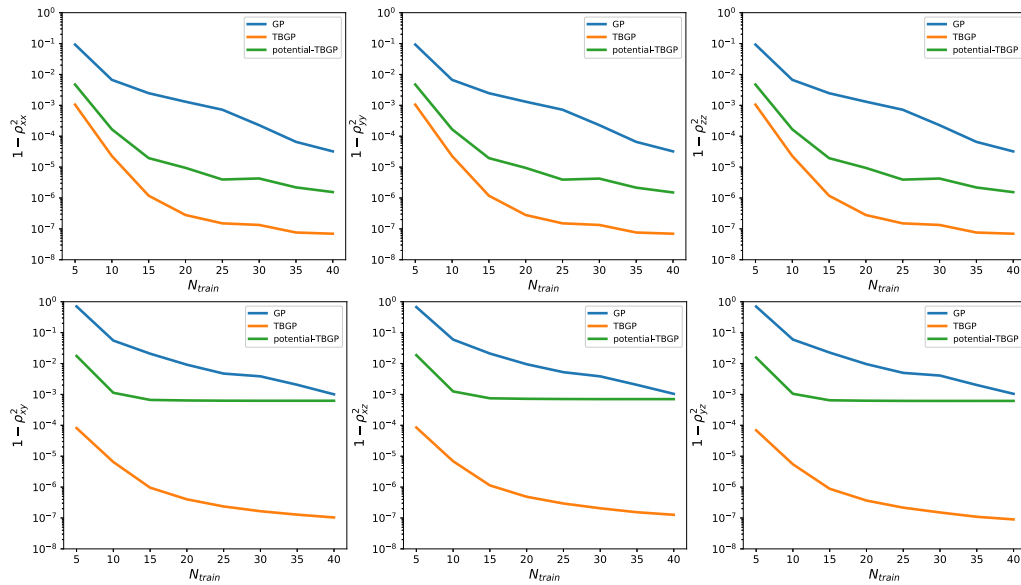
$$\mathbf{B} = \mathbf{R}\mathbf{V}^2\mathbf{R}^T, \quad (43)$$

thus guaranteeing that the determinant of  $\mathbf{F}$  is positive and that  $\mathbf{B}$  corresponds to a valid diagonalizable tensor with some superposition of tension/compression and shear in arbitrary directions. The corresponding eigenvalues of  $\mathbf{B}$  are randomly distributed in the interval  $[l^2, u^2]$ .

The results shown in Figs. 2, 3, and 4 are for two datasets, one sampling  $[l^2, u^2] = [1.0, 1.5]$  and another covering  $[0.9, 2.0]$ . This first choice includes strain values corresponding only to mild extension along the principal axes, while the second choice includes a more extreme range from mild compression to much more extension. The GP, TBGP, and potential-TBGP formulations were trained on 100 different random samples of datasets of varying size, and each trial's hyperparameters were selected with multistart L-BFGS optimization (Byrd et al., 1995) of the marginal likelihood with 20 random initializations.

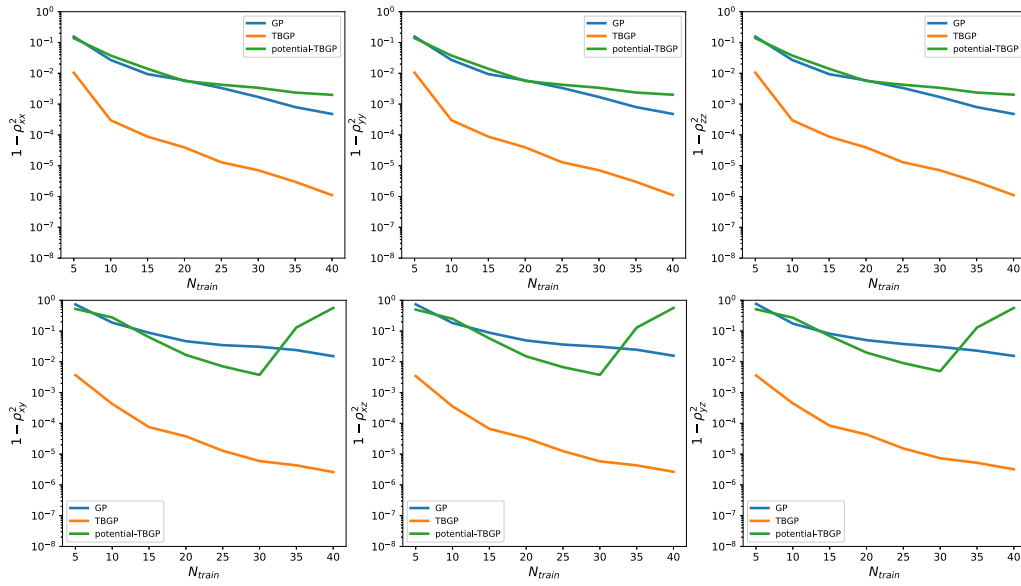


**FIG. 2:** The root-mean-square error of the stress tensor for the GP, TBGP, and potential-TBGP formulations as a function of training set size for  $[l^2, u^2] = [1.0, 1.5]$  (upper panel) and  $[0.9, 2.0]$  (lower panel)



**FIG. 3:** The component-wise fraction of variance unexplained ( $1 - \rho^2$ ) for the stress tensor predictions as a function of training set size for  $[l^2, u^2] = [1.0, 1.5]$ . Top row are the tension components, and the bottom are the shear components.

The root-mean-square error is shown in Fig. 2 and fraction of variance unexplained  $1 - \rho^2$  for correlation coefficient  $\rho$  for the stress tensor components are shown in Figs. 3 and 4 as a function of the training set size for both datasets. It is clear that the TBGP has a substantially lower error than the GP with approximately the same rate of convergence, with 1–2 orders of magnitude lower error and 5–6 orders of magnitude lower unexplained variance. As in the matrix exponential example, the TBGP formulation is uniformly the best choice at all tested numbers of datapoints. However, the potential-based TBGP has about the same error as the regular GP, and both show slightly degraded performance on predicting the shear components of the stress tensor compared to the tension components. In addition, for larger sets of data the accuracy of potential-TBGP stops improving and, in the larger deformation case, the error diverges. We attribute this trend to attempting to learn the full function behavior from gradient information alone, as well as the much larger and more ill-conditioned covariance matrix formed in the



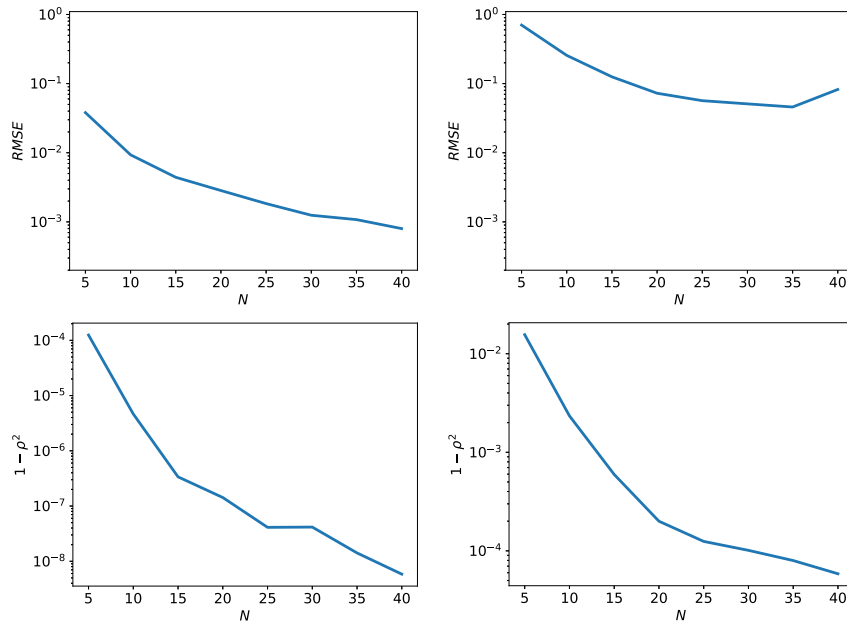
**FIG. 4:** The component-wise fraction of variance unexplained ( $1 - \rho^2$ ) for the stress tensor predictions as a function of training set size for  $[l^2, u^2] = [0.9, 2.0]$ . Top row are the tension components, and the bottom are the shear components.

inference process. Individual trials of the training process become more likely to yield higher-error models, increasing the average error. It is likely that using advanced low-rank factorizations of the covariance matrix or better regularization would reduce the magnitude of this diverging error.

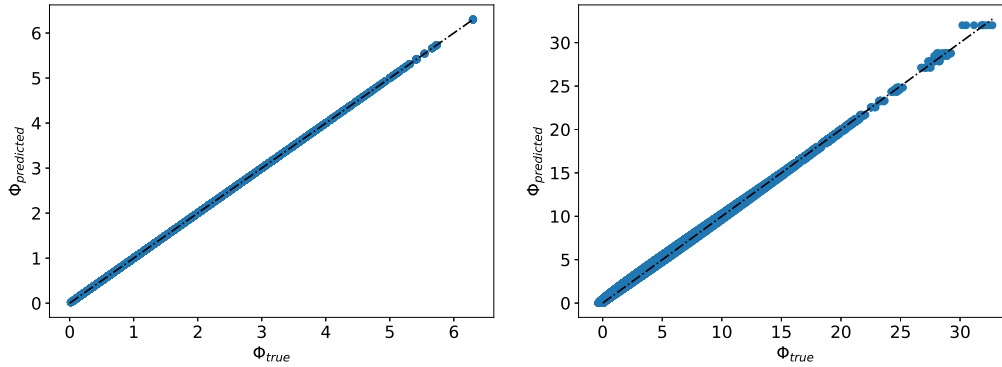
Although the potential-TBGP performance for predicting the stress components is no better than the GP, it is capable of predicting the strain-energy density function with fairly high accuracy, which is not a task that either of the other two formulations are capable of doing directly. Figure 5 shows the root-mean-square error and correlation coefficient as a function of training set size. The prediction accuracy does show substantial improvement with increasing training data for the lower-stretch case  $[l^2, u^2] = [1.0, 1.5]$ , but the ill-conditioning issues prevalent in the stress predictions for the higher-stretch case  $[l^2, u^2] = [0.9, 2.0]$  pervade the potential prediction as well. Figure 6 shows that the source of disagreement between the potential-TBGP and the Mooney–Rivlin potential originates from values at higher energy, where there is less training data available and a more complex trend to predict. One possible solution to improve the performance of the potential-TBGP is to use a greedy point selection approach that would use points that span a wider range of invariant space and energy density to train the GP.

## 7. CONCLUSION

This paper has developed and demonstrated an approach to embedding rotational invariance in the GPR framework for constitutive modeling of isotropic hyperelasticity. Embedding this physics knowledge led to a dramatic improvement in the accuracy and learning curves for the TBGP formulation compared to the traditional component-based approach. Also the potential-TBGP formulation demonstrated recovery of the potential function from stress-strain data with



**FIG. 5:** The root-mean-square error (upper) and unexplained variance (lower) of the strain energy density function  $\Phi$  as a function of training set size from the potential-TBGP for  $[l^2, u^2] = [1.0, 1.5]$  (left) and  $[l^2, u^2] = [0.9, 2.0]$  (right)



**FIG. 6:** Scatter-plot of the predicted and actual strain energy density values at the highest training set size for the cases  $[l^2, u^2] = [1.0, 1.5]$  (upper panel) and  $[0.9, 2.0]$  (lower panel)

comparable accuracy to the plain GP formulation for stress prediction. While the examples considered here are relatively simple, the application of the methodology to more complex hyperelastic materials and functions of kinematic variables is straightforward.

One important consideration for future work is the representation of anisotropic material response and functions of multiple tensors. In these cases, the tensor basis and corresponding invariants are more complex, but the underlying process remains the same. For example, the second Piola–Kirchhoff tensor  $\mathbf{S}$  may be expressed as a function of the Cauchy–Green deformation tensor  $\mathbf{C} = \mathbf{F}^T \mathbf{F}$  and a structure tensor  $\mathbf{G}$  that characterizes the anisotropy in the response:

$$\mathbf{S} = \mathbf{f}(\mathbf{C}, \mathbf{G}), \quad (44)$$

[see Zheng (1994) for more details]. For the case of transverse isotropy where the material response along a direction  $\mathbf{g}$  is different than in the plane perpendicular to the unit vector  $\mathbf{g}$ ,  $\mathbf{G} = \mathbf{g} \otimes \mathbf{g}$  can be employed as the structure tensor. The corresponding expansion for  $\mathbf{S}$  is

$$\mathbf{S} = \sum_{i=1}^6 c_i \mathbf{A}_i, \quad (45)$$

in terms of the tensors  $\mathbf{A}_i \in \{\mathbf{I}, \mathbf{C}, \mathbf{C}^2, \mathbf{G}, \mathbf{CG} + \mathbf{GC}, \mathbf{C}^2\mathbf{G} + \mathbf{GC}^2\}$  and coefficients  $c_i$ , which are functions of the extended invariant set  $\{\text{tr } \mathbf{C}, \text{tr } \mathbf{C}^2, \text{tr } \mathbf{C}^3, \text{tr } \mathbf{CG}, \text{tr } \mathbf{C}^2\mathbf{G}\}$  [refer to Boehler (1987) and use  $\mathbf{G}^2 = \mathbf{G}$ ]. Since  $\mathbf{S}$  is a symmetric tensor and the tensor basis  $\{\mathbf{A}_i\}$  is a linearly independent set, the expansion gives six equations for the six unknowns  $c_i$ , which may be solved readily. (If representation theory suggests more tensor basis elements than unknowns, the basis can be reduced to an independent set that spans the output, at which point the suggested procedure applies.) In this way, the corresponding coefficients as a function of the invariants may be inferred and a TBGP may be trained to make predictions for an anisotropic material response.

There is also room to improve the predictions of the potential-based TBGP. The squared-exponential kernel was selected for its simplicity and smoothness, but it is possible that a more complex or nonstationary kernel would be able to capture the behavior of the potential function in invariant-space more accurately and overcome the ill-conditioning issues seen in this formulation. Other work suggests that approaches involving low-rank or spectral representations of the covariance matrix with derivative information would likely improve the accuracy of the inference process as well.

## ACKNOWLEDGMENTS

This work was supported by the LDRD program at Sandia National Laboratories, and its support is gratefully acknowledged. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the U.S. government. This paper has been deemed UUR with SAND number SAND2020-5235J.

## REFERENCES

- Bachoc, F., Lagnoux, A., and López-Lopera, A., Maximum Likelihood Estimation for Gaussian Processes under Inequality Constraints, *Elect. J. Stat.*, vol. **13**, no. 2, pp. 2921–2969, 2019.
- Boehler, J., Representations for Isotropic and Anisotropic Non-Polynomial Tensor Functions, in *Applications of Tensor Functions in Solid Mechanics*, pp. 31–53, Berlin: Springer, 1987.
- Bonet, J. and Wood, R., *Nonlinear Continuum Mechanics for Finite Element Analysis*, Cambridge, UK: Cambridge University Press, 1997.
- Brunton, S.L., Proctor, J.L., and Kutz, J.N., Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems, *Proc. Nat. Acad. Sci.*, vol. **113**, no. 15, pp. 3932–3937, 2016.
- Byrd, R., Lu, P., and Nocedal, J., A Limited Memory Algorithm for Bound Constrained Optimization, *SIAM J. Sci. Stat. Comput.*, vol. **16**, pp. 1190–1208, 1995.

- Da Veiga, S. and Marrel, A., Gaussian Process Modeling with Inequality Constraints, *Annales de la Faculté des Sciences de Toulouse*, vol. **21**, pp. 529–555, 2012.
- Dean, J., Corrado, G., Monga, R., Chen, K., Devin, M., Mao, M., Ranzato, M., Senior, A., Tucker, P., and Yang, K., Large Scale Distributed Deep Networks, *Adv. Neural Inf. Process. Syst.*, pp. 1223–1231, 2012.
- Doyle, T. and Ericksen, J., Nonlinear Elasticity, *Adv. Appl. Mech.*, vol. **4**, pp. 53–115, 1956.
- Frankel, A., Jones, R., Alleman, C., and Templeton, J., Predicting the Mechanical Response of Oligocrystals with Deep Learning, *Comput. Mater. Sci.*, vol. **169**, p. 109099, 2019a.
- Frankel, A., Tachida, K., and Jones, R., Prediction of the Evolution of the Stress Field of Polycrystals Undergoing Elastic-Plastic Deformation with a Hybrid Neural Network Nodel, 2019b. arXiv: 1910.03172
- Gurtin, M., *An Introduction to Continuum Mechanics*, vol. **158**, Cambridge, MA: Academic Press, 1982.
- Hastie, T., Tibshirani, R., and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd Edition, Berlin: Springer, 2016.
- Hill, R., On Constitutive Inequalities for Simple Materials—I, *J. Mech. Phys. Solids*, vol. **16**, no. 4, pp. 229–242, 1968.
- Jensen, B., Nielsen, J., and Larsen, J., Bounded Gaussian Process Regression, in *2013 IEEE Int. Workshop on Machine Learning for Signal Processing (MLSP)*, Southampton, UK, 2013.
- Jones, R., Templeton, J., Sanders, C., and Ostien, J., Machine Learning Models of Plastic Flow based on Representation Theory, *Comput. Model. Eng. Sci.*, pp. 309–342, 2018.
- Karpatne, A., Atluri, G., Faghmous, J., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., and Kumar, V., Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data, *IEEE Transact. Knowledge Data Eng.*, vol. **29**, no. 10, pp. 2318–2331, 2017.
- Lee, K. and Carlberg, K., Model Reduction of Dynamical Systems on Nonlinear Manifolds Using Deep Convolutional Autoencoders, *J. Comput. Phys.*, vol. **404**, p. 108973, 2020.
- Ling, J., Jones, R., and Templeton, J., Machine Learning Strategies for Systems with Invariance Properties, *J. Comput. Phys.*, vol. **318**, pp. 22–35, 2016.
- López-Lopera, A., Bachoc, F., Durrande, N., and Roustant, O., Finite-Dimensional Gaussian Approximation with Linear Inequality Constraints, *SIAM/ASA J. Uncertain. Quant.*, vol. **6**, no. 3, pp. 1224–1255, 2018.
- Lusch, B., Kutz, J.N., and Brunton, S.L., Deep Learning for Universal Linear Embeddings of Nonlinear Dynamics, *Nat. Commun.*, vol. **9**, no. 1, p. 4950, 2018.
- Magiera, J., Ray, D., Hesthaven, J., and Rohde, C., Constraint-Aware Neural Networks for Riemann Problems, *J. Comput. Phys.*, vol. **409**, p. 109345, 2020.
- Malvern, L., *Introduction to the Mechanics of a Continuous Medium*, Engle Cliffs, NJ: Prentice Hall Inc., 1969.
- Marckmann, G. and Verron, E., Comparison of Hyperelastic Models for Rubber-Like Materials, *Rubber Chem. Technol.*, vol. **79**, no. 5, pp. 835–858, 2006.
- Ogden, R., *Non-Linear Elastic Deformations*, North Chelmsford, MA: Courier Corporation, 1997.
- Pan, S. and Duraisamy, K., Data-Driven Discovery of Closure Models, *SIAM J. Appl. Dyn. Syst.*, vol. **17**, no. 4, pp. 2381–2413, 2018.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., Scikit-Learn: Machine Learning in Python, *J. Mach. Learn. Res.*, vol. **12**, pp. 2825–2830, 2011.
- Raissi, M. and Karniadakis, G., Hidden Physics Models: Machine Learning of Nonlinear Partial Differential Equations, *J. Comput. Phys.*, vol. **357**, pp. 125–141, 2018.



- Raissi, M., Perdikaris, P., and Karniadakis, G., Machine Learning of Linear Differential Equations Using Gaussian Processes, *J. Comput. Phys.*, vol. **348**, pp. 683–693, 2017.
- Rasmussen, C. and Williams, C., *Gaussian Processes for Machine Learning*, Cambridge, MA: MIT Press, 2006.
- Riihimäki, J. and Vehtari, A., Gaussian Processes with Monotonicity Information, in *Proc. of the Thirteenth Int. Conf. on Artificial Intelligence and Statistics*, Sardinia, Italy, 2010.
- Solak, E., Murray-Smith, R., Leithead, W., Leith, D., and Rasmussen, C., Derivative Observations in Gaussian Process Models of Dynamic Systems, in *Advances in Neural Information Processing Systems*, pp. 1057–1064, 2003.
- Wu, A., Aoi, M.C., and Pillow, J.W., Exploiting Gradients and Hessians in Bayesian Optimization and Bayesian Quadrature, 2017. arXiv: 1704.00060
- Yang, X., Tartakovsky, G., and Tartakovsky, A., Physics-Informed Kriging: A Physics-Informed Gaussian Process Regression Method for Data-Model Convergence, 2018. arXiv: 1809.03461
- Zheng, Q., Theory of Representations for Tensor Functions—A Unified Invariant Approach to Constitutive Equations, *Appl. Mech. Rev.*, vol. **47**, no. 11, pp. 545–587, 1994.