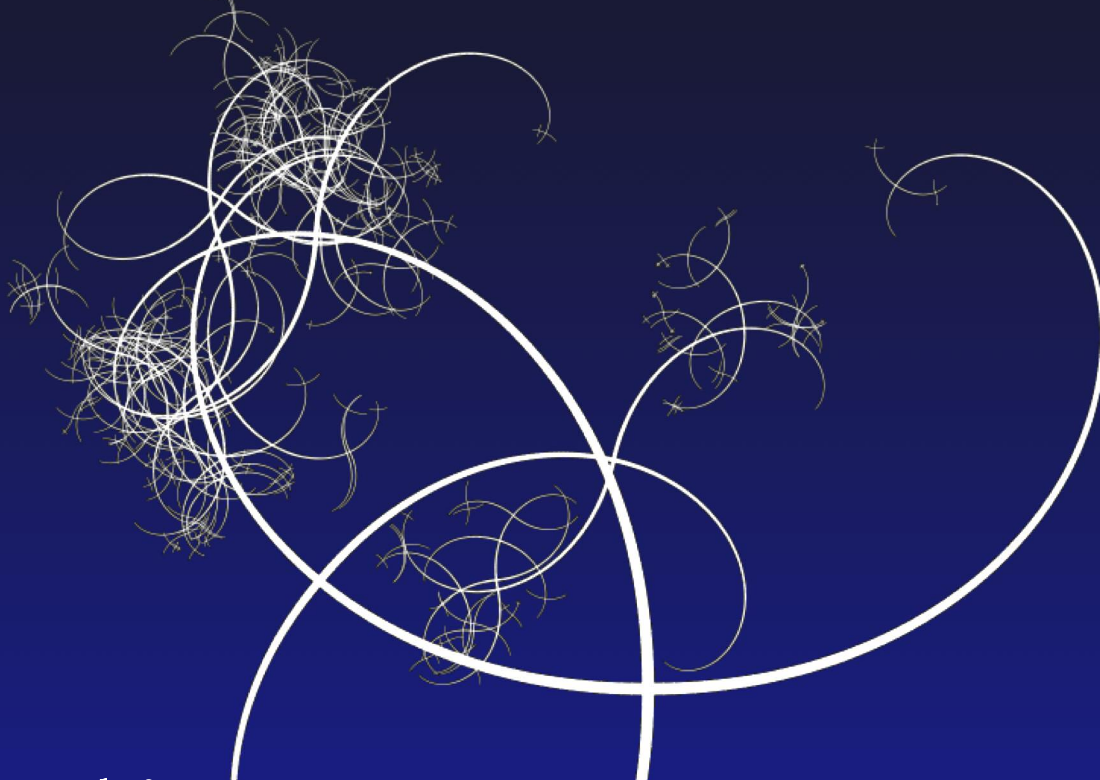
NISSUNA UMANA INVESTIGAZIONE SI PUO DIMANDARE
VERA SCIENZIA S'ESSA NON PASSA PER LE
MATEMATICHE DIMOSTRAZIONI
LEONARDO DA VINCI



vol. 8 no. 1 2020

Mathematics and Mechanics of Complex Systems

msp

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS

msp.org/memocs

MEMOCS (ISSN 2325-3444 electronic, 2326-7186 printed) is a journal of the International Research Center for the Mathematics and Mechanics of Complex Systems at the Università dell’Aquila, Italy.

Cover image: “Tangle” by © John Horigan; produced using the *Context Free* program (contextfreeart.org).

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
Vol. 8, No. 1, 2020
dx.doi.org/10.2140/memocs.2020.8.1

SECOND-ORDER WORK CRITERION AND DIVERGENCE CRITERION: A FULL EQUIVALENCE FOR KINEMATICALLY CONSTRAINED SYSTEMS

JEAN LERBET, NOËL CHALLAMEL, FRANÇOIS NICOT AND FELIX DARVE

This paper presents stability results for rate-independent mechanical systems, associated with general tangent stiffness matrices including symmetric and nonsymmetric ones. Conservative and nonconservative as well as associate and nonassociate elastoplastic systems are concerned by such a theoretical study. Hill's stability criterion, also called the second-order work criterion, is here revisited in terms of kinematically constrained systems. For piecewise rate-independent mechanical systems (which may cover inelastic and elastic evolution processes), such a criterion is also a divergence Lyapunov stability criterion for any kinematic autonomous constraints. This result is here extended for systems with nonsymmetric tangent matrices. By virtue of a new type of variational formulation on the possible kinematic constraints, and thanks to the concept of kinematical structural stability (KISS), both criteria, Hill's stability criterion and the divergence stability criterion under kinematic constraints, are shown to be equivalent.

1. Introduction

The aim of this paper is to contribute to close a sixty-year-old debate initiated by Hill [1958; 1959] concerning the stability of rate-independent mechanical systems. In these papers, Hill proposed a new criterion of stability today called the second-order work criterion of stability, which leads to critical values of loading which are not always in accordance with the usual ones calculated from the divergence criterion. Among the concrete examples where the second-order work criterion performs well, one of the most demonstrative is probably the liquefaction of water-saturated loose sands. In situ, liquefaction occurs most often in a saturated sand layer during seismic events. In lab experiments, the so-called "undrained" triaxial loading simulates these in situ conditions well, where the fast seismic loading

Communicated by Angelo Luongo.
MSC2010: 34D30, 37C20.
PACS2010: 46.32.+x.
Keywords: Hill's criterion, kinematic structural stability, nonconservative systems, nonassociate materials.

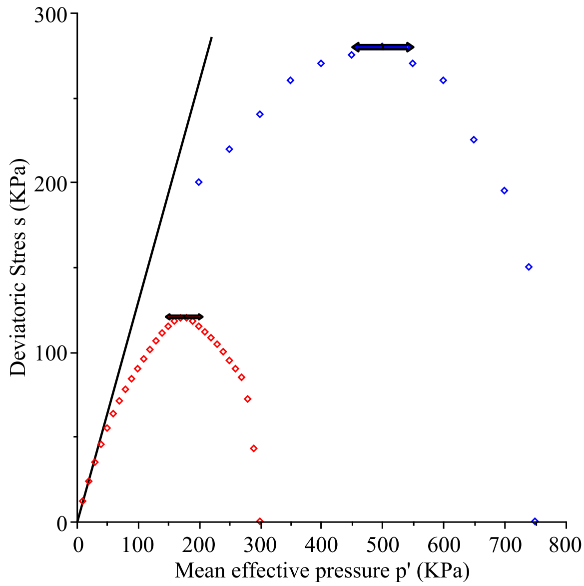


Figure 1. Undrained (isochoric) axisymmetric triaxial compressions of two loose sands plotted in blue and red: stress paths q versus the mean pressure p'. The experimental results correspond to the points, while the Mohr–Coulomb plastic limit condition is given by the straight black line. The q maxima reached along the undrained stress paths clearly occur before the Mohr–Coulomb line (the figure is reconstituted from [Daouadji et al. 2010]). In these experiments, the loading is axially strain-controlled, so q peaks can be passed. If the same loading would have been axially q-controlled, a sudden dynamic instability would have developed at the q peak.

prevents the water drainage of the layer to occur and thus enforces undrained conditions. So, in lab, if the sand is loose enough to be contractant in drained shear (for more details see [Darve 1994; 1996]), the deviatoric stress denoted by s passes through a maximum clearly before reaching the Mohr–Coulomb limit line (see Figure 1).

Thus, if the loading path is s stress-controlled (as in situ by the weight of the above constructions), at the s peak a sudden failure occurs largely before the plastic limit criterion. In this example, the second-order work criterion gives the limit load whereas the usual Mohr–Coulomb condition fails. Then, how can one explain this strong instability? First it is to be noted that the axisymmetric undrained triaxial compression is a mixed stress-strain loading, axially s stress-controlled ($ds =$ constant) with a kinematic constraint given by the isochoric condition related to the undrained conditions ($d\epsilon_1 + 2\,d\epsilon_3 = 0$). Now the second-order work for axisymmetric conditions in stress-strain principal axes can be rewritten as

$$d^2W = d\sigma_1\, d\epsilon_1 + 2\, d\sigma_3\, d\epsilon_3 = ds\, d\epsilon_1 + d\sigma_3\, d\epsilon_V$$

where $d\epsilon_V = d\epsilon_1 + 2\,d\epsilon_3$ is the relative incremental volume variation and $ds = d\sigma_1 - d\sigma_3$ characterizes the incremental deviatoric stress. Thus, for an undrained (isochoric) loading ($d\epsilon_V = 0$), the second-order work d^2W vanishes at the s peak ($ds = 0$) [Darve and Chau 1987]. So it is shown that the second-order work criterion can properly describe sand liquefaction and the explanation of the failure thanks to the second-order work criterion is linked to the kinematic constraint. Indeed, due to the kinematic constraint (constant volume), the failure is not described here by a plastic limit condition (the sand still behaves in a hardening regime) but by an instability condition given by the loss of positive-definiteness of the elastoplastic matrix. This sand liquefaction example contains all the ingredients (second-order work criterion, failure criterion, and kinematic constraint) that are involved in these investigations. However, to highlight the deep relationship between these three ingredients, we will use here a framework allowing us to perform calculations and analytical developments which are not limited to some examples but which allow us to analyze the most complex situations involving any rate-independent materials and systems.

The paper is organized as follows.

We will start in Section 2 with the simple case of a two degree of freedom system that contains all the key ingredients to well understand the foundation of both the problem and the solution as well. In this example, both stability criteria are surprisingly linked in a way which is the key of these developments. In Section 3 we focus on the different concepts of stability involved in this question. Whereas the divergence criterion is linked to the well known Lyapunov point of view on stability, the second-order work criterion can be viewed as the criterion involved in another type of stability we decided to call Hill stability. Hill [1958; 1959] did not define a concept of stability that he projected to use, but he only proposed a criterion of stability. In order to be clear from a logical point of view, we define the Hill stability by the corresponding criterion, this Hill stability definition not being in accordance with the Lyapunov one. Fortunately for the rationality of the approach, one may link this Hill stability to a type of perturbation of the equilibrium (called mixed perturbations) distinct from the perturbations (small purely kinematic perturbations) used to investigate the Lyapunov stability of the equilibrium. To summarize, we then will have on hand two types of stability, the Lyapunov stability and the Hill stability, both leading for quasistatic investigations to the divergence criterion for the first type and to the second-order work criterion for the second type. The main object of this paper is then to propose an explicit equivalence thanks to kinematic constraints.

This equivalence necessitates the investigation of Lyapunov stability of any subsystem obtained by imposing on the initial system additional kinematic constraints. Such investigations lead to the key concept of KISS, to which Section 4 is devoted and which was highlighted in the initial example in the next section. We define quite simply KISS, and we present a brief review of some results that may be related to this concept despite its relatively recent emergence in 2014. We then give its main properties involved in the solution of our problem. In Section 5, we deduce the claimed equivalence between the two criteria in a fully symmetric way which appears as a natural generalization of the well accepted equivalence of both the elastic conservative and the associated elastoplastic cases in the case of symmetric stiffness operators. Finally, in Section 6, we show how the geometric method is performing to provide the appropriate destabilizing constraint by investigating a discrete model of the Leipholz column. The nonconservative part is distributed on each link of the system through follower forces $\vec{P}_i$. This example is a model of the device realized by Bigoni and Noselli [2011]. Unlike the introductory example where the calculations could be performed by hand, the use of the geometric point of view is then essential.

2. An introductory example

Before any general consideration, we start by observing some facts on the well known example of the two degree of freedom Ziegler system Σ. Some of these facts are also well known whereas others are not or very little known.

The system Σ is made up of two bars $S_1 = OA$ and $S_2 = AB$ of the same length ℓ so that the two links of the system are torsion elastic springs at O and A with the same stiffness k (see Figure 2). The load is a follower force $\vec{P} = -P(\overrightarrow{AB}/\|\overrightarrow{AB}\|)$ with $P \geq 0$. The dimensionless expression of the load parameter P is $p = P\ell/k$, and the kinematics of Σ is described by θ_1 and θ_2 as in Figure 2.

We introduce $\Theta = (\theta_1, \theta_2)$, and we identify a couple (θ_1, θ_2) and the corresponding two-dimensional column vector $\binom{\theta_1}{\theta_2}$.

This system is interesting due to the double advantage of its simplicity and of the fact that it has all the characteristics of nonconservative systems. There has been a long debate about the physical meaning of follower forces [Koiter 1996; Elishakoff 2005], but Bigoni and Noselli [2011] showed the applicability of these forces by creating an experimental device illustrating the model of the two degree of freedom Ziegler system Σ used here.

The stiffness matrix at the (unique) equilibrium position $0 = (0, 0)$ reads

$$K(p) = \begin{pmatrix} 2-p & -1+p \\ -1 & 1 \end{pmatrix}$$

and we may note that $K(p)$ is nonsymmetric.

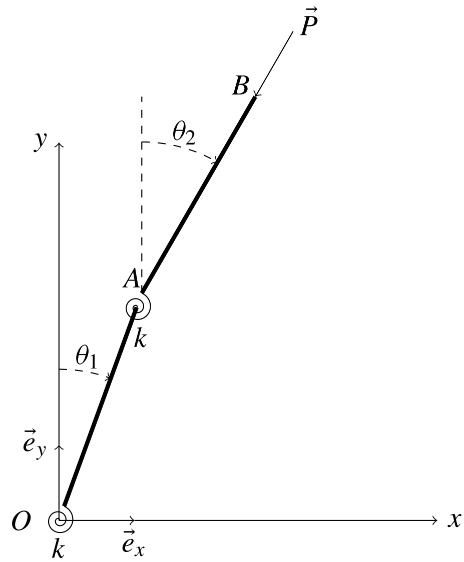


Figure 2. Two degree of freedom Ziegler system.

Referring to the usual framework of the linear Lyapunov stability approach, the divergence-type stability or static stability criterion is investigated thanks to the determinant of $K(p)$. Calculations give $\det(K(p)) = 1$ independently of p. It means that no divergence instability of the equilibrium $0 = (0, 0)$ may occur and that the critical value of divergence stability is $p^*_{\text{div}} = +\infty$. Note that it is usual to conclude that the only way to investigate the Lyapunov linear stability of $0 = (0, 0)$ is to involve inertial terms via the mass matrix. For this system, flutter-type instability occurs for the value $p^*_{\text{fl}} = 2.54$ for a uniform mass repartition.

Whereas Lyapunov stability questions the (dynamic) behavior of the system subjected to the loading of the equilibrium when it is subjected to a purely kinematic perturbation at $t = 0$ (here made up by a quadruplet $(\delta\theta_1, \delta\theta_2, \delta\dot{\theta}_1, \delta\dot{\theta}_2)$), another kind of stability may be defined thanks to the concept of mixed perturbation (see [Absi and Lerbet 2004; Challamel et al. 2009] for example). For the current example, it means that the system can be subjected to a set of perturbations involving a kinematic part related to $\delta\Theta = (\delta\theta_1, \delta\theta_2)$ and a "force" counterpart related to $\delta C = (\delta C_1, \delta C_2)$ where δC_i is any "small" torque acting on S_i for $i = 1, 2$. Both $(\delta\theta_1, \delta\theta_2)$ and $(\delta C_1, \delta C_2)$ must satisfy the fundamental principle of energy conjugation in mechanics: we cannot force together the motion and the mechanical action controlling this motion. This means that $\delta\theta_i$ and δC_i cannot be chosen arbitrarily together.

We do emphasize that, in mixed perturbations, the system is no longer subjected to only the mechanical external given actions involved in the equilibrium whose stability is investigated. For Σ, the system of external forces involved in the (unique) equilibrium position $(0, 0)$ is reduced to $\vec{P}$. For a mixed perturbation, defined by example by $\delta\theta_1$ for the kinematic part and by δC_2 for the force part, a new equilibrium position $(\delta\theta_1, \delta\theta_2) \neq (0, 0)$ is reached when Σ is subjected to $\vec{P}, \delta C_2$. This new equilibrium equation reads

$$\delta C = K(p)\,\delta\Theta \tag{1}$$

where $K(p)$ is the (tangent) stiffness matrix of Σ at the equilibrium position $(0, 0)$ (for the intrinsic geometrical meaning of $K(p)$, see [Lerbet et al. 2018]).

We then define a second type of stability of the equilibrium configuration (here $(0, 0)$), denoted Hill stability of the equilibrium (explanations of this name will appear hereafter), by requiring that the system reaches another equilibrium position close to the one whose stability is investigated when it is subjected to any mixed perturbation. The condition of Hill stability of the equilibrium position then reads (see [Absi and Lerbet 2004; Challamel et al. 2009] for example) $\delta\Theta^T\,\delta C > 0$ for any mixed perturbation. Using (1) it is equivalent to $\delta\Theta^T\,K(p)\,\delta\Theta > 0$ for any $\delta\Theta \neq 0$. Because this expression involves only the symmetric part $K_s(p)$ of $K(p)$, it is equivalent to require that $\delta\Theta^T\,K_s(p)\,\delta\Theta > 0$ for any $\delta\Theta \neq 0$. It is nothing else but the Hill second-order work criterion: the Hill stability of equilibrium is preserved as long as the symmetric part of the stiffness matrix remains positive definite. It is the reason why this type of stability is called the Hill stability. Obviously the framework of [Hill 1958; 1959] is not the same as the one considered in this example. It dealt with associate or nonassociate plasticity, and the object involved in the criterion was the tangent stiffness matrix along an incremental evolution. Nevertheless, the stiffness matrix of this elastic nonconservative system plays the role of the general tangent stiffness matrix of incremental elastoplastic evolutions and captures all the essential ingredients of the problem.

It is however worth noting that the way Hill [1958] introduced his stability criterion — which is now called the second-order work criterion — did not involve explicitly mixed perturbations of equilibrium but emerged from energetic considerations. In [Hill 1958], it is also difficult to identify a clear distinction between his definition of stability and his criterion of stability. Roughly speaking, Hill claimed the condition $\dot{f}\dot{u} \geq 0$ for all $\dot{u}$ where $f = Ku$, which explained the term of second order to characterize this criterion, the term $\dot{f}\dot{u}$ involving two terms of first order. This leads obviously to the same property of positive definiteness for the matrix K. Another approach to derive the second-order work criterion can be found for example in [Nicot et al. 2012a], where the investigation of the infinitesimal variation of the kinetic energy shows that in certain conditions, it is governed by a term $\dot{x}K\dot{x}$. However, in the two last expressions $\dot{f}\dot{u}$ and $\dot{x}^T K\dot{x}$, the dot sign over the vectors does not have the same meaning. In the first one, it means the derivative with respect to the process evolution whereas in the second one, it means the physical time.

Calculations give

$$K_s(p) = \begin{pmatrix} 2-p & -1+p/2 \\ -1+p/2 & 1 \end{pmatrix}.$$

Since the loading path is monotone, the second-order work criterion can also be investigated through the equation $\det(K_s(p)) > 0$, which is equivalent here.

We find $\det(K_s(p)) = 1 - p^2/4$ so that the critical value for the increasing loading path for the Hill stability is $p_H^* = 2$. We then have two thresholds ($p_{\text{div}}^* = +\infty$ and $p_H^* = 2$) for both types of stability, and an important issue since 1958 is to find a sound bridge between them. Note that the universal relation $p_H^* \le p_{\text{div}}^*$ was well known.

We now investigate another question which will actually be the key to these issues. We propose to investigate the divergence (Lyapunov) stability of all the systems obtained from Σ by adding linear kinematic constraints. Because the system has two degrees of freedom, it can be subjected to only one kinematic constraint, which is a linear relation between the small values $x_1 = \theta_1 - \theta_{e,1}$ and $x_2 = \theta_2 - \theta_{e,2}$ of the deviation of angles θ_1 and θ_2 with respect to the equilibrium configuration $\theta_e = (\theta_{e,1}, \theta_{e,2}) = (0,0)$. The general form reads $a_1x_1 + a_2x_2 = 0$ with $(a_1, a_2) \neq (0,0)$. It can be stressed that the isochoric condition related to the undrained conditions as mentioned in the introduction ($d\epsilon_1 + 2d\epsilon_3 = 0$) has exactly this form.

Introducing the Lagrange multiplier λ, the (static) equation system of the constrained system reads

$$\begin{cases} (2-p)x_1 + (p-1)x_2 - \lambda a_1 = 0, \\ -x_1 + x_2 - \lambda a_2 = 0, \\ a_1x_1 + a_2x_2 = 0, \end{cases} \tag{2}$$

whose determinant reads $D = -2a_2(a_2 + a_1)p + a_2^2 + (a_2 + a_1)^2$. The divergence critical value for the constrained system is given by the (minimal positive) value $p_{\text{div}}^*(a_1, a_2)$ that makes D vanish. Elementary calculations give

- for $a_2 = 0$ or $a_1 = -a_2$, $D > 0$ for all p and the corresponding critical value of divergence stability is again $p_{\text{div}}^* = +\infty$, and
- for $a_2 \neq 0$ and $a_1 \neq -a_2$ then the critical value of divergence stability reads

$$p_{\text{div}}^*(\alpha) = \frac{\alpha^2 + 2\alpha + 2}{\alpha + 1}, \tag{3}$$

corresponding to a one-parameter problem with parameter $\alpha = a_1/a_2$.

A straightforward calculation shows that the minimal positive value of $p^*_{\text{div}}(\alpha)$ is 2 for $\alpha = 0$, namely for the constraint $x_2 = 0$.

Surprisingly, $p = 2$ has then two distinct meanings for the two degree of freedom Ziegler system. On the one hand, it is the critical value for Hill stability of the structure Σ by applying the Hill second-order work criterion; on the other hand it is the minimal critical value regarding the Lyapunov divergence stability but for any constrained subsystem of Σ. Indeed, for $p < 2$, no constrained system can be divergence unstable. This astonishing result is in fact general and is the key result for the equivalence between both stability criteria.

It has to be stressed that this stability analysis does not involve the inertia of the system, namely, for a linear analysis, the mass matrix M. This stability analysis is full in the framework of quasistatic evolutions. This framework is the one of Hill's papers. A dynamic linear stability analysis for this two degree of freedom Ziegler system, like for any mechanical system, needs to investigate flutter-type instability. It occurs, for a homogeneous mass distribution of the two degree of freedom Ziegler system, for the critical value $p^*_{\text{fl}} \approx 2.54$. When this system is subjected to a kinematic constraint, it becomes a one degree of freedom system and the flutter-type instability can no longer occur. In this paper, except for some remarks, the flutter-type instability is neither investigated nor mentioned since we are concerned by the links between the second-order work criterion and the divergence criterion.

3. Two distinct approaches of stability, and respective strengths and weaknesses

As already mentioned, both criteria refer to two distinct points of view of the stability of an equilibrium state. The apparent conflict between the two corresponding criteria should not be a real issue. However, for 60 years, the question has always been tackled in a competitive way. But, as will be shown hereafter, according to an original variational formulation, both kinds of stability will be fully reconciled.

As a usual result of linear algebra gives that $\det A_s \leq \det A$ for any real matrix ($A_s = \frac{1}{2}(A + A^T)$ is the symmetric part of A), the Hill stability of any system Σ prevents the Lyapunov divergence instability of Σ. The Hill stability criterion, namely the second-order work criterion, then goes towards safety regarding the stability of the (equilibrium configuration of the) system. In fact originally, both approaches question the equilibrium from two different points of view which are complementary. Referring to the above result and reasoning with the concern of safety, the second-order work criterion should have been universally adopted for the quasistatic evolution of systems.

However, it was not. Several reasons can explain this irrational situation. A first reason is that the Lyapunov point of view of stability is an old, well established, and general framework, largely developed with a lot of deep results involving nonlinear dynamic analysis. Its theorems are applied daily with success in any field of human activity.

On the other hand, even if the Hill criterion of stability is fundamentally (since 1958) a nonlinear criterion used at each step of an incremental loading path, it has provided hitherto no natural extension to dynamics. Some recent papers such as [Nicot et al. 2011; 2012a; 2012b] suggest however that the quadratic part $\dot{X}^T K_s(p)\dot{X}$ whose sign is governed by the Hill second-order work criterion may be involved in a transition from purely incremental quasistatic evolution to a dynamic one. A complete and larger point of view, governed by the safety of structures, could then use

- the Hill criterion for quasistatic evolutions and transitions towards dynamics and
- the Lyapunov dynamic criterion for dynamic evolution.

We however must emphasize that there is no continuous transition between the two criteria. For example, for the two degree of freedom Ziegler system with a uniform mass distribution, the first threshold is $p_H^* = 2$ for the Hill stability whereas $p_{\text{fl}}^* = 2.54$ for the flutter-type stability. It can be proved for this system that no flutter-type instability may occur for $p < 2$ for any continuous mass distribution. However, there exist concentrated mass distributions such that the corresponding flutter-type instability critical value is < 2. The general conclusion is that the Hill approach is especially well founded for quasistatic evolution and Lyapunov stability for dynamic evolution, the transition between the two regimes needing to be more deeply investigated.

Note however that for conservative (and associate elastoplastic) systems, all these considerations are meaningless since the only mode of instability is the divergence-type instability; as the stiffness matrix is symmetric, the divergence stability criterion and second-order work criterion are the same and read

$$\det(K(p)) = \det(K_s(p)) > 0.$$

Finally, another mental habit also inherited from the study of conservative systems is the fact that the stability of a system ensures the stability of any system obtained from the initial system by adding kinematic constraints.

The above example shows that it is not generally right for any mechanical system and it is even a signature of the nonconservativity. This paradoxical fact may be balanced with the other paradox of the destabilizing effect of additional damping, which has been, for its part, very deeply investigated (see for example [Bolotin 1963; Kirillov and Verhulst 2010]). This paradox is called the Hermann or Ziegler paradox according to whether the damping is internal or external. Conversely, to the best of our knowledge very few references deal with the destabilizing effect of additional kinematic constraints. As will be systematically investigated in the next section, the Lyapunov criteria of stability (both divergence-type or flutter-type criteria as well) fail with respect to this property whereas the Hill stability meets this requirement. We will say that the Hill stability criterion is kinematically structurally stable.

To conclude this section, the two approaches have then their own strengths and weaknesses and none should be systematically rejected, especially since the Hill stability criterion emerges as the best criterion offering the divergence stability the required property of kinematic structural stability. It is the purpose of the next section. We first present a brief historical review about this concept. Secondly, we provide the formal definition of the so-called kinematic structural stability (KISS) and finally, we outline its main properties.

4. The KISS

4.1. ***A brief history.*** Since 2010 [Challamel et al. 2010], we have investigated the behavior of nonconservative elastic and nonassociate elastoplastic systems firstly under only one additional kinematic constraint and from 2012 [Lerbet et al. 2012] for any number of kinematic constraints. However, the key concept of kinematic structural stability (KISS) emerged only in 2014. As mentioned in the introduction of the paper, the framework for the current presentation is the linear discrete mechanics even though similar reasoning may be given for a material REV with the corresponding tangent stiffness matrix. From a historic point of view and to the best of our knowledge, only a few papers deal with some issues in relation with KISS. Thompson [1982] noted the paradoxical possibility of destabilizing a nonconservative column with an additional constraint whereas Ingerle [2018] computed approximate loadings that may destabilize nonconservative columns by investigating some special constrained systems. More accurately, for the continuous Beck column, the dimensionless divergence stability value for any kinematic constraint has been calculated in [Lerbet et al. 2017] and is equal to π^2. This value had been empirically obtained by Ingerle [2013] from a discrete approach using numerical arguments.

It is worth mentioning that the stability limit under kinematic constraints is the generalization of the one under some specific constraints especially applied to the boundaries of the system. For instance, Ingerle [1969] found a dimensionless divergence buckling load $p = 20.19$ in the presence of a specific constraint applied to the end of the column (the application point of the follower load), whereas the free Beck column admits a flutter instability value of 20.05, as calculated by Beck [1952] (see also [El Naschie 1976; 1977] for this result). However, consider that any kinematic constraint reduces this value to π^2 as mentioned above as was observed by Ingerle [2013] and as has been definitively proved by Lerbet et al. [2017].

4.2. ***General framework.*** In the chosen framework, the kinematics of the holonomic system Σ is described by a Lagrange coordinate system $q = (q_1, \dots, q_n)$, where $\tilde{q}$ is the current equilibrium configuration whose stability is investigated, and $p = (p_1, \dots, p_m)$ is a family of loading parameters. A loading path Λ_p is a one-dimensional curve in the loading parameter space $\mathcal{P} = \{p_k \geq 0 \mid k = 1, \dots, m\}$. This curve is given by $\sigma \in [0, \infty[\mapsto p(\sigma) = (p_1(\sigma), \dots, p_m(\sigma)) \in \mathcal{P}$. We suppose that $p(0) = 0 \in \mathcal{P}$ and that, for the unloaded system (namely for $p(0) = 0$), $K_s(0)$ is positive definite. The stiffness matrix at $\tilde{q}$ is then a function $K(p)$ of $p = (p_1, \dots, p_m)$. $X = q - \tilde{q}$ is the vector giving the infinitesimal or incremental (according to the point of view) response of the structure. For a complete nonlinear description with the use of differential geometry, see [Lerbet et al. 2018].

When any set $\mathscr{C} = \{C_1, \dots, C_r\}$ of r kinematic constraints is acting on Σ, we denote by $\Sigma_{\mathscr{C}}$ the corresponding kinematic constrained system and we say that $\Sigma_{\mathscr{C}}$ is r-constrained if $C_1, \dots, C_r$ viewed as vectors of $\mathbb{R}^n$ are linearly independent. That also means that the constraint corresponding to C_i reads $X^T C_i = 0$. We also suppose that the equilibrium $\tilde{q}$ is not perturbed by the additional kinematic constraints. The case where the equilibrium position is changed with the kinematic constraints is a full different mechanical problem (in this case, see for example [Tarnai 2004]) which is not investigated here. For the Hill stability, we denote by $D_H \subset \mathcal{P}$ the stability domain and $\Gamma_H = \partial D_H \subset \mathcal{P}$ is the corresponding critical domain for the system Σ. For the divergence stability, we denote by $D_{\text{div}} \subset \mathcal{P}$ the divergence stability domain and $\Gamma_{\text{div}} = \partial D_{\text{div}} \subset \mathcal{P}$ is the corresponding critical domain for the system Σ. For any constrained system $\Sigma_{\mathscr{C}}$, the corresponding domains are denoted $D_{H,\mathscr{C}}$ and $D_{\text{div},\mathscr{C}}$.

Remarks. (1) The (second-order work) criterion of Hill stability is that $K_s(p)$ is positive definite. For elastoplastic materials, this domain is not intrinsic, namely it is path-dependent and can be defined only for each loading path Λ_p. Then, D_H is investigated through a priori an infinite number of one-dimensional (along each loading path Λ_p) analyses. At the other extreme, like in the introductory example, for elastic nonconservative systems with only one loading parameter, D_H is a simple interval of $\mathbb{R}_+$ and $\Gamma_H = \partial D_H$ is reduced to a point. For the two degree of freedom Ziegler system, $D_H = [0, 2[$ and $\Gamma_H = \{2\}$.

(2) Because of the continuity of involved applications and since $\det(K_s(0)) > 0$, the boundary $\Gamma_H = \partial D_H \subset \mathcal{P}$ of critical values for Hill stability can also be found by solving the equation $\det(K_s(p)) = 0$ which defines a nonconnected hypersurface in $\mathcal{P}$, one of whose components is $\Gamma_H = \partial D_H$.

(3) When the D_H is not path-dependent, for example for any elastic nonconservative system like for the discrete Leipholz column investigated in Section 6, some general properties for D_H can be underlined. D_H is an open set of $\mathcal{P}$. To give more properties about D_H, we need to know the dependency $p \mapsto K_s(p)$. For example, if the dependency is linear like for the discrete Leipholz column, the parametrization $p \mapsto K_s(p)$ defines an affine set S of $\mathcal{M}_n(\mathbb{R})$. Then, since the set $\mathcal{S}_n^{++}(\mathbb{R})$ of symmetric definite matrices is a convex semicone of $\mathcal{M}_n(\mathbb{R})$, $S \cap \mathcal{S}_n^{++}(\mathbb{R})$ is a convex set of $\mathcal{M}_n(\mathbb{R})$. Because the inverse of $p \mapsto K_s(p)$ from S to $\mathcal{P}$ is again an affine map, D_H is then a convex set of the loading space $\mathcal{P}$.

(4) If $p \in D_H$, the isotropic cone $C_0(p) = \{X \in \mathcal{M}_{n1}(\mathbb{R}) = \mathbb{R}^n \mid X^T K_s(p) X = 0\}$ is reduced to the zero vector of $\mathbb{R}^n$. For $p^* \in \partial D_H$, the isotropic cone $C_0(p^*) = \ker K_s(p^*)$ is always a vector space and generally a one-dimensional space. Considering $p^* = p(\sigma^*)$ belongs to a loading path Λ_p, then for $\sigma > \sigma^*$ in the vicinity of σ^*, $C_0(p) = \{X \in \mathcal{M}_{n1}(\mathbb{R}) = \mathbb{R}^n \mid X^T K_s(p) X = 0\}$ is a cone no longer reduced to a vector space. We then put $C_-(p) = \{X \in \mathcal{M}_{n1}(\mathbb{R}) = \mathbb{R}^n \mid X^T K_s(p) X \leq 0\}$ and $C_+(p) = \{X \in \mathcal{M}_{n1}(\mathbb{R}) = \mathbb{R}^n \mid X^T K_s(p) X > 0\}$.

(5) From the general statement $\det A_s \leq \det A$ it follows that $D_H \subset D_{\text{div}}$ and, in the present investigations, $D_H \subsetneq D_{\text{div}}$.

(6) A characteristic property of elastic conservative or associate elastoplastic systems is that $D_H = D_{\text{div}}$.

(7) A characteristic property of elastic conservative or associate elastoplastic systems is that $D_{\text{div}} = D_{\text{div},\mathscr{C}}$ for any constraint system $\mathscr{C}$.

(8) In general for a system there is family of constraints $\mathscr{C}$ such that $D_{\text{div}} \subsetneq D_{\text{div},\mathscr{C}}$.

(9) A characteristic property of the second-order Hill criterion is that $D_H \subset D_{H,\mathscr{C}}$ for any constraint system $\mathscr{C}$.

The well known properties (7) and (9) may be deduced from the Rayleigh quotient.

4.3. *The kinematic structural stability (KISS).* KISS refers to the behavior of the stability of the equilibrium positions when the system is subjected to additional kinematic constraints. According to path-independent or path-dependent stability domains, the definitions can be defined globally on $\mathcal{P}$ or defined for each loading path Λ_p. In order to apply this definition to the case of the discrete Leipholz column investigated in Section 6, we present both definitions:

Definition (intrinsic aspects). Suppose the system is hypoelastic such that the stability issue is intrinsic.

- The KISS is said to be universal (for the corresponding criterion and equilibrium) if and only if $D \subset D_{\mathscr{C}}$ for all $\mathscr{C}$.

- The KISS is said to be conditional (for the corresponding criterion and equilibrium) if and only if there is $D_{\text{co}} \subsetneq D$ such that $D_{\text{co}} \subset D_{\mathscr{C}}$ for all $\mathscr{C}$. This notation also means that there is at least a value $p^*_{\text{co}} \in \partial D_{\text{co}} \cap D$ and a constraint set $\mathscr{C}^*$ with $p^*_{\text{co}} = p^*_{\mathscr{C}^*}$

Definition (path-dependent aspects). Let

$$\Lambda_p : \sigma \in [0, \infty[\mapsto p(\sigma) = (p_1(\sigma), \dots, p_m(\sigma)) \in \mathscr{P}$$

be a loading path drawn in $\mathscr{P}$. The critical load for the involved stability criterion (Hill, divergence-type, or flutter-type) is supposed reached on this loading path Λ_p for $p^* = p(\sigma^*)$ (which can be infinite). Then

- the KISS is said to be universal (for the corresponding criterion and equilibrium) if and only if $p^* \leq p^*_{\mathscr{C}}$ for all $\mathscr{C}$ and
- the KISS is said to be conditional (for the corresponding criterion and equilibrium) if and only if there is $p^*_{\text{co}} = p(\sigma^*_{\text{co}}) < p^*$ such that $p^*_{\text{co}} \leq p^*_{\mathscr{C}}$ for all $\mathscr{C}$. This notation also means that σ^*_{co} is optimal (it corresponds to the minimal value of the parameter σ with this property) and there is then a constraint set $\mathscr{C}^*$ with $p^*_{\text{co}} = p^*_{\mathscr{C}^*}$

The rest of the reasoning is given with the path-dependent formalism. It means that a loading path Λ_p is fixed.

As a consequence, when the KISS is conditional, there is an appropriate set of constraints $\mathscr{C}^*$ such that $p^*_{\mathscr{C}^*} < p^*$, namely making the constrained system $\Sigma^*_{\mathscr{C}}$ unstable whereas Σ is still stable: it is the paradoxical destabilizing effect of additional kinematic constraints.

In order to highlight the link between the KISS and the second-order work criterion, the way to tackle constrained mechanical systems should be commented on further. In the example investigated in Section 2, we used, as usual, Lagrange multipliers. It leads to a problem of larger size ($2+1=3$ for the example) whereas the constrained system has a lower degree of freedom ($2-1=1$ for the example). During our investigations, it appeared that a better way to systematically tackle constrained systems consists of using so-called compressions of operators. It provides objects not only appropriate to physical systems (for example a matrix of size r for an r degree of freedom mechanical system) but also applicable to various other situations such as constrained continuous media involving infinite-dimensional spaces (see [Lerbet et al. 2017] for its use leading to the complete solution of the constrained Beck column). The formal definition of the compression of a linear map reads:

Definition. Let $u \in \mathscr{L}(E)$ be a linear map of a euclidean space E and F a vector subspace of E. The compression u_F of u on F is the element of $u \in \mathscr{L}(F)$ defined by $u_F = p_F \circ u \circ i_F$ where $i_F : F \to E$ is the canonical injection map from F to E and $p_F : E \to F$ is the orthogonal projection on F. In other words, $u_F = p_F \circ u_{|F}$ where $u_{|F}$ is the usual restriction of u to F.

The applicability of this concept is justified by:

Proposition 1. *Let Σ be a mechanical system. Suppose a mechanical property of Σ is described in a linear framework by a linear map u of $\mathbb{R}^n$. The same property for a constrained system $\Sigma_{\mathscr{C}}$ is described by the compression $u_{F_{\mathscr{C}}^{\perp}}$ of u on $F_{\mathscr{C}}^{\perp}$ where $F_{\mathscr{C}}$ is the space spanned by the vectors $C_1, \dots, C_r$ of $\mathbb{R}^n$ defining the set of constraints $\mathscr{C}$.*

Thanks to this concept of compression, the KISS issue for the divergence instability criterion may be reformulated through the following purely geometric approach.

Let Σ be a mechanical system and $u(p)$ the linear map of $\mathbb{R}^n$ associated with the stiffness matrix $K(p)$. The divergence stability of Σ means the invertibility of $K(p)$, namely the one-to-one property of $u(p)$. The KISS issue means

(1) find a threshold p^*_{co} such that all the compressions u_F are still one-to-one for all $p < p^*_{\text{co}}$ and for all subspaces F of $\mathbb{R}^n$ and

(2) as $p = p^*_{\text{co}}$, find a subspace F^* of $\mathbb{R}^n$ such that u_{F^*} is no longer one-to-one. The corresponding set of critical constraints $\mathscr{C}^*$ will be a generator system of the orthogonal $(F^*)^{\perp}$ of F^*.

The solution of these issues is contained in:

Theorem 2. *Let $u \in \mathscr{L}(E)$ be an one-to-one linear map of a euclidean space E. All these compressions on (strict) subspaces are still one-to-one if and only if the symmetric part u_s of u is definite. As u_s loses its definiteness, one may build a compression on a hyperplane F^* of E for example such that u_{F^*} is not one-to-one. More specifically, if $x^* \in E$ is a nonzero vector on the isotropic cone of u_s, one may choose $F^* = \langle u(x^*)\rangle^{\perp}$ ($u(x^*) \neq 0$ because u is supposed to be one-to-one).*

This result extends to Hilbert spaces with compressions to closed spaces. It has been used in [Lerbet et al. 2017].

Geometric proof. Suppose first that u_s ceases to be definite for a value p^* of the parameter. Then, there is $x^* \neq 0$ such that $u_s(x^*) = 0$. Let $F^* = \langle u(x^*)\rangle^{\perp}$, which is a hyperplane because $u(x^*) \neq 0$. Then $F^* = \ker c^*$ with c^* a linear form (namely the appropriate constraint) $c^*(y) = (u(x^*) \mid y)$. We have to prove that the constrained system is divergence unstable or equivalently that the compression u_{F^*} of u on F^* is not one-to-one. But $u_s(x^*) = 0$ implies $(u(x^*) \mid x^*) = (u_s(x^*) \mid x^*) = (0 \mid x^*) = 0$, which proves that $x^* \in F^*$.

Thus, $(u_{F^*}(x^*) \mid y) = (p_{F^*} \circ u(x^*) \mid y) = (u(x^*) \mid p_{F^*}(y)) = (u(x^*) \mid y) = 0$ for all $y \in F^*$, which means that $u_{F^*}(x^*)$ is not one-to-one as an endomorphism of F^*.

Reciprocally, suppose now that there is a compression u_{G^*} which is not one-to-one with G^* a subspace of E. There is $x^* \in G^*$, $x^* \neq 0$, such that $u_{G^*}(x^*) = 0$ or equivalently such that $(p_{G^*} \circ u(x^*) \mid y) = (u(x^*) \mid y) = 0$ for all $y \in G^*$. Applying this relation for $y = x^*$ shows that u_s is no longer definite. Moreover, we also deduce that $G^* \subset F^* = \langle u(x^*)\rangle^{\perp}$, which proves that only one constraint is sufficient to investigate the question. □

To be closer to the usual language of mechanics, we now propose a direct definition of the compression for the constrained system $\Sigma_{\mathscr{C}}$ of the stiffness matrix $K(p)$ of the system Σ. Moreover, since the theorem shows that the variational formulation with only one constraint is necessary and sufficient, we suppose that $\mathscr{C}$ is reduced to one constraint C which reads $C^T X = 0$. We put $F_C = \{X \in \mathbb{R}^n \mid C^T X = 0\}$. It is an $(n-1)$-dimensional subspace of $\mathbb{R}^n$ identified with the vector space of n-dimensional column matrices. F_C describes the kinematics of the constrained system Σ_C. We then provide:

Definition. Let $\mathscr{B}_C$ be an orthonormal basis of F_C. The compression $K_{\mathscr{B}_C}(p)$ of $K(p)$ on the kinematic space F_C of the constrained system Σ_C in $\mathscr{B}_C$ is the $n-1$ square matrix defined by

$$X^T_{\mathscr{B}_C} K_{\mathscr{B}_C}(p) Y_{\mathscr{B}_C} = X^T K(p) Y \quad \text{for all } X, Y \in F_C$$

where $X_{\mathscr{B}_C}$, $Y_{\mathscr{B}_C}$ are the coordinate column vectors of $X, Y \in F_C$ in the basis $\mathscr{B}_C$. So, obviously the (compressed) stiffness matrix $K_{\mathscr{B}_C}(p)$ of the constrained system $\Sigma_{\mathscr{C}}$ depends on $\mathscr{B}_C$ but not its invertibility nor its determinant, which depend only on F_C itself.

The main theorem then reads:

Theorem 3. *Suppose we have an increasing loading parameter p starting from* 0 *with* $\det K_s(0) > 0$. *As long as* $\det K_s(p) > 0$, *there is no kinematic constraint C destabilizing by divergence the corresponding constrained system*: $\det K_C(p) \neq 0$ *for all constraints C. As soon as $p = p^*$ such that* $\det K_s(p^*) = 0$ (*Hill's second-order work criterion failure*), *then there is a divergence destabilizing constraint C^** (*namely* $\det K_{C^*}(p^*) = 0$) *and the kinematic constraint C^* is explicit.*

Proof. We only prove the construction of the destabilizing constraint.

Thus, suppose that $\det K_s(p^*) = 0$ (failure of the second-order work criterion for Σ) and $\det K(p^*) \neq 0$ (divergence stability of Σ). Let $X^* \neq 0$ be in $\ker K_s(p^*)$ or equivalently on the isotropic cone of $K_s(p^*)$, and let us put $C^* = K(p^*)X^*$ so that $X^{*T} C^* = X^{*T} K(p^*) X^* = X^{*T} K_s(p^*) X^* = 0$. But, since $K(p^*)$ is invertible, then $C^* \neq 0$ and since $X^{*T} C^* = C^{*T} X^* = 0$, then $X^* \in F_{C^*} = \{Z \mid Z^T C^* = 0\}$. Moreover, $X^* \neq 0$ implies that $X^*_{\mathscr{B}_{C^*}} \neq 0$ in any orthonormal basis $\mathscr{B}_{C^*}$ of F_{C^*}.

But if $Y \in F_{C^*}$, then by a similar calculation, $0 = Y^T C^* = Y^T K(p^*) X^* = Y^T_{\mathcal{B}_{C^*}} K_{\mathcal{B}_{C^*}}(p^*) X^{*T}_{\mathcal{B}_{C^*}}$, which means that $K_{\mathcal{B}_{C^*}}(p^*) X^{*T}_{\mathcal{B}_{C^*}}$ is (the column vector of the coordinates in $\mathcal{B}_{C^*}$ of) a vector of F_{C^*} orthogonal to any vector of F_{C^*}. It is then the zero vector of F_{C^*}, which means that $K_{\mathcal{B}_{C^*}}(p^*) X^{*T}_{\mathcal{B}_{C^*}} = 0$. But, since $X^* \neq 0$, then $X^{*T}_{\mathcal{B}_{C^*}} \neq 0$ and then $K_{\mathcal{B}_{C^*}}(p^*)$ is not invertible, which implies the divergence instability of the constrained system Σ_{C^*}. □

Remark. We find nowhere in the literature neither the result nor its proof.

This theorem explicitly solves the two KISS issues. First it gives the equivalence between the divergence KISS and the Hill second-order work criterion: $p^*_{\text{co}} = p^*_H$. It also shows that the case of constrained systems with only one constraint (one-constrained system according to the above definition) is sufficient to investigate the KISS issues. This fact had been proved in [Lerbet et al. 2012] with the language of Lagrange multipliers thanks to the concept of r-definite matrices. A comparison between both approaches shows that the language of compressions greatly simplifies the problem. Finally, this above theorem also gives a constructive way to find the destabilizing constraint, because thanks to this theorem, the destabilizing constraint is given by the vector $K(p^*_H)X^*$ where X^* is any nonzero vector on the isotropic cone of $K_s(p^*_H)$. Before highlighting in the next section, thanks to this result, the announced full equivalence between divergence Lyapunov stability and Hill stability, let us conclude this current section by noting that the KISS issue has been investigated for various frameworks like flutter-type instability [Lerbet et al. 2016b], divergence-type instabilities for continuum mechanics [Lerbet et al. 2017], and instabilities of nonlinear incremental discrete mechanics [Lerbet et al. 2018].

5. Equivalence of the two criteria via an original variational approach

We now tackle the claimed equivalence regarding the Lyapunov divergence stability criterion and the Hill second-order work criterion. According to the previous section, we are led to investigate this question in terms of the variational formulation on all the possible kinematic constraints $\mathscr{C}$ that may be applied on the system Σ, keeping in mind that this large variational formulation may be reduced to families built by only one constraint, namely for one-constrained systems $\Sigma_{\mathscr{C}}$ as has been defined above. We first tackle the question of the kinematic structural stability of the Hill criterion itself.

A usual result of (bi)linear algebra — also called Sylvester's conditions for symmetric positive definite matrices — means in terms of compressions that all the compressions of a symmetric positive definite map are also positive definite. It is the exact characterization of the kinematic structural stability of the Hill second-order work criterion: if a mechanical system Σ is Hill stable, any constrained system is still Hill stable.

All above results of this paper may be summarized (for a monotone load path) as

(1) $\det K_s(p) \leq \det K(p)$, that is, the Lyapunov divergence instability of Σ leads to the Hill instability of Σ,

(2) the Hill instability of Σ by loss of definiteness of $K_s(p)$ leads to the existence of a set of constraints $\mathscr{C}$ such $\Sigma_{\mathscr{C}}$ is not Lyapunov divergence stable, and

(3) the Hill stability of Σ is equivalent to the Hill stability of $\Sigma_{\mathscr{C}}$ for any set of constraints $\mathscr{C}$.

These three results allow us to formulate the first statement for any rate-independent mechanical system Σ

Hill stability (s.o.w. criterion) of Σ $\iff$ Lyapunov divergence stability of $\Sigma_{\mathscr{C}}$ for all $\mathscr{C}$

and finally to conclude with the full and symmetric equivalence

Hill stability (s.o.w. criterion) of $\Sigma_{\mathscr{C}}$ for all $\mathscr{C}$ $\iff$ Lyapunov divergence stability of $\Sigma_{\mathscr{C}}$ for all $\mathscr{C}$

which, due to the remark on the compressions on hyperplanes, is also equivalent to

Hill stability (s.o.w. criterion) for all one-constrained $\Sigma_{\mathscr{C}}$ $\iff$ Lyapunov divergence stability for all one-constrained $\Sigma_{\mathscr{C}}$.

These last two equivalences may be interestingly compared with the usual statement valid for any conservative or associate elastoplastic system Σ:

Hill stability (s.o.w. criterion) of Σ $\iff$ Lyapunov divergence stability of Σ.

6. The discrete Leipholz column

The system Σ_n consists of n bars $OA_1, A_1A_2, \ldots, A_{n-1}A_n$ with $OA_1 = A_1A_2 = \cdots = A_{n-1}A_n = h$ linked with n elastic springs with the same stiffness k. Adopting the same device at the end of each bar of Σ leads to a family of follower forces $\vec{P}_1, \ldots, \vec{P}_n$. Figure 3 illustrates the case $n = 3$.

The pure follower forces $\vec{P}_1, \vec{P}_2, \ldots, \vec{P}_n$ are applied at the ends of OA_1, A_1A_2, $\ldots$, $A_{n-1}A_n$. The equilibrium position is $\theta = (\theta_1, \theta_2, \ldots, \theta_n) = (0, 0, \ldots, 0)$.

Adopting a dimensionless format, $p_i = \|\vec{P}_i\|h/k$ for $i = 1, \ldots, n$ are used as loading parameters. The corresponding physical system has been realized and described in [Bigoni and Noselli 2011]. For our concern, this system is interesting because of its nonreduced geometric degree of nonconservativity d (see [Lerbet et al. 2014; 2016a] for this concept). Roughly speaking, the geometric degree of nonconservativity d of a system is the minimal number of kinematic constraints necessary to make the system conservative. Indeed, unlike Ziegler systems, whose

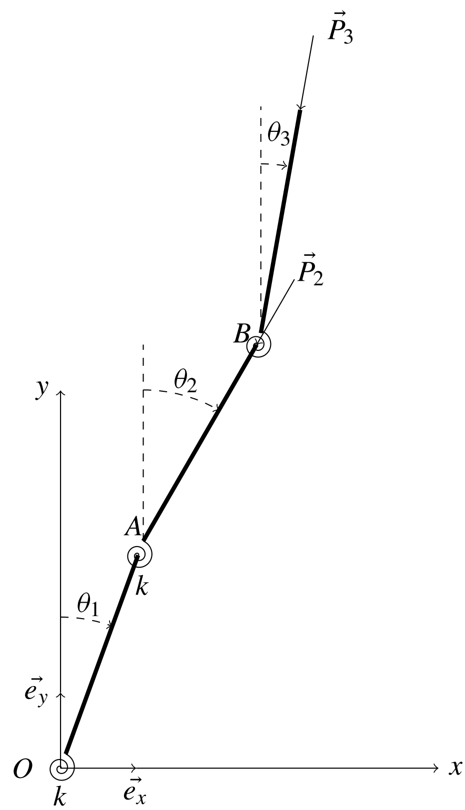


Figure 3. Three degree of freedom discrete Leipholz system.

geometric degree of nonconservativity is always equal to one whatever the degree of freedom is, the geometric degree of nonconservativity of the discrete Leipholz column is increasing as $\lfloor n/2 \rfloor$, the integer part of $n/2$.

The stiffness matrix K reads $K(p) = K(p_1, p_2, \ldots, p_n)$:

$$K(p) = \begin{pmatrix} Q_2 & -1+p_2 & p_3 & p_4 & p_5 & \cdots & p_{n-1} & p_n \\ -1 & Q_3 & -1+p_3 & p_4 & p_5 & \cdots & p_{n-1} & p_n \\ 0 & -1 & Q_4 & -1+p_4 & p_5 & \cdots & p_{n-1} & p_n \\ 0 & 0 & -1 & Q_5 & -1+p_5 & \cdots & p_{n-1} & p_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & -1+p_{n-1} & p_n \\ 0 & 0 & 0 & 0 & 0 & \cdots & 2-p_n & -1+p_n \\ 0 & 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}$$

where $Q_j = 2 - \sum_{i=j}^{n} p_i$.

We see that K does not depend on p_1, which is obvious from a mechanical point of view. We then may suppose that $p_1 = 0$.

A remarkable property of $K(p_1, p_2, \dots, p_n)$ is that its determinant does not depend on p:

$$\det K(p_1, p_2, \dots, p_n) = 1 \quad \text{for all } p_1, p_2, \dots, p_n.$$

That may be proved by applying $n-1$ times from $k = n$ to $k = 2$ (in this order) the same rule: the column C_{k-1} of the matrix at the step number k is replaced by $C_{k-1} + C_k$. At the end of the process, the determinant of the matrix is unchanged but the matrix is then upper-triangular with a diagonal of 1 and then its determinant is 1.

We may deduce that the condition "$K(p)$ invertible" of the main theorem (Theorem 2) holds without any condition on the loading parameters p_i, $i = 1, \dots, n$. It also means that $D_{\text{div}} = \mathcal{P}$.

The symmetric part of K then reads

$$K_s(p) = \begin{pmatrix} Q_2 & -1+R_2 & R_3 & R_4 & R_5 & \cdots & R_{n-1} & R_n \\ -1+R_2 & Q_3 & -1+R_3 & R_4 & R_5 & \cdots & R_{n-1} & R_n \\ R_3 & -1+R_3 & Q_4 & -1+R_4 & R_5 & \cdots & R_{n-1} & R_n \\ R_4 & R_3 & -1+R_4 & Q_5 & -1+R_5 & \cdots & R_{n-1} & R_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ R_{n-2} & R_{n-2} & R_{n-2} & R_{n-2} & R_{n-2} & \cdots & -1+R_{n_1} & R_n \\ R_{n-1} & R_{n-1} & R_{n-1} & R_{n-1} & R_{n-1} & \cdots & 2-p_n & -1+R_n \\ R_n & R_n & R_n & R_n & R_n & \cdots & -1+R_n & 1 \end{pmatrix}$$

where $R_j = p_j/2$. Then $p \mapsto K_s(p)$ is an affine map. So, D_H is a convex set of $\mathcal{P}$, but there is no chance to find a general formula for $\det(K_s(p))$. Thus, to make the computations analytically, we must fix a value of n. We investigate successively $n = 3$ and $n = 4$.

Case $n = 3$. The stiffness matrix then reads

$$K(p_2, p_3) = \begin{pmatrix} 2-(p_2+p_3) & -1+p_2 & p_3 \\ -1 & 2-p_3 & -1+p_3 \\ 0 & -1 & 1 \end{pmatrix}$$

and its symmetric part $K_s(p_2, p_3)$

$$K_s(p_2, p_3) = \begin{pmatrix} 2-(p_2+p_3) & -1+p_2/2 & p_3/2 \\ -1+p_2/2 & 2-p_3 & -1+p_3/2 \\ p_3/2 & -1+p_3/2 & 1 \end{pmatrix}.$$

The domain of investigation lies in the quadrant $\mathcal{P} = (p_2 \geq 0,\ p_3 \geq 0)$. As usual we suppose that any loading path starts from $(0, 0)$. The convex domain D_H of Hill

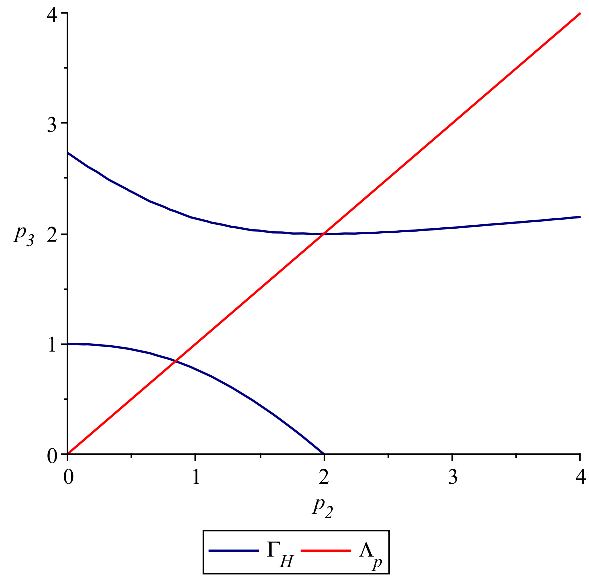


Figure 4. Hill stability domain D_H of the three degree of freedom discrete Leipholz column (D_{div} is the whole quadrant ($p_2 \geq 0$, $p_3 \geq 0$)). Inside D_H, $\det K_s(p) > 0$. On $\partial D_H = \Gamma_H$, for example for $p^* = ((1+\sqrt{33})/8, (1+\sqrt{33})/8) = \Gamma_H \cap \Lambda_p$, $\det K_s(p) = 0$. For $p_d = \left(\frac{3}{2}, \frac{3}{2}\right)$, $\det K_s(p) < 0$.

stability is delimited by the blue curve Γ_H in the vicinity of $(0, 0)$ (see Figure 4). D_H is given by the set of inequalities

$$\left\{0 \leq p_2 \leq 2,\ 0 \leq p_3 \leq 1,\ 1 - \tfrac{3}{2}p_3^2 + \tfrac{1}{2}p_2 p_3^2 + \tfrac{1}{2}p_3^3 - \tfrac{1}{4}p_2^2 - \tfrac{1}{2}p_2 p_3 \geq 0\right\},$$

and the explicit equation of the curve $\Gamma_H = \partial D_H$ delimiting the domain D_H is

$$p_2 = f(p_3) = p_3^2 - p_3 + \sqrt{p_3^4 - 5p_3^2 + 4}.$$

As proved in the general case (see above point (3) on page 12)), D_H is a convex set of the quadrant which can be directly checked by calculating the second derivative of f, which is always negative for $p_3 \in [0, 1]$. The main result (Theorem 2) means that, inside D_H defined by the second-order work criterion, there is no kinematic constraint that destabilizes the system Σ_3. On the contrary, on any $p^* \in \partial D_H = \Gamma_H$, there is a constraint $\mathscr{C}^* = \{c^*\}$ such that the constrained system $\Sigma_{3,\mathscr{C}^*}$ is Lyapunov divergence unstable.

The case $p_3 = 0$ corresponds to the introductory example whereas the case $p_2 = 0$ has been investigated in [Lerbet et al. 2012]. A loading path is determined by a curve Λ_p in the quadrant $\mathscr{P} = (p_2 \geq 0,\ p_3 \geq 0)$ and starting from $(0, 0)$.

Suppose for example that the loading path is given by the curve $\Lambda_p : p_2 = p_3$ or $\sigma \mapsto p(\sigma) = (\sigma, \sigma)$ plotted in red in the Figure 4. It is in accordance with Bigoni's device [Bigoni and Noselli 2011] where the friction forces are the same at each joint and because of the same velocity of the support which induces equal force friction at each joint. The line $\Lambda_p : p_2 = p_3$ intersects the curve Γ_H at the point $p^* = (p_2^*, p_3^*) = ((1+\sqrt{33})/8, (1+\sqrt{33})/8) = \Gamma_H \cap \Lambda_p$. For this critical value, the isotropic cone of $K_s(p^*)$ is no longer reduced to 0 but is reduced to one single direction. It is a vector space generated by the vector

$$X^* = \begin{pmatrix} 1986 - 350\sqrt{33} \\ 3678 - 642\sqrt{33} \\ 5370 - 934\sqrt{33} \end{pmatrix}.$$

Thus, applying the main theorem, the next step consists of calculating

$$K(X^*) = \begin{pmatrix} -2685 + 467\sqrt{33} \\ -993 + 175\sqrt{33} \\ 1692 - 292\sqrt{33} \end{pmatrix},$$

which means that the system Σ_3 subjected to the kinematic constraint

$$c^*(x_1, x_2, x_3) \\ = (-2685 + 467\sqrt{33})x_1 + (-993 + 175\sqrt{33})x_2 + (1692 - 292\sqrt{33})x_3 = 0$$

is divergence unstable when subjected to the load $p^* = ((1+\sqrt{33})/8, (1+\sqrt{33})/8)$.

Remarks. (1) If we would like to naively proceed as in the introductory example in order to investigate the divergence stability of the constrained systems, instead of searching for the maximum of one real function on one real variable as in (3), we should now solve a four-dimensional extremum formal (namely parametrized and nonnumerical) problem (in order to define any system of two linear constraints on three variables we need four variables) whose objective function is the determinant of a 5×5 formal matrix whose coefficients are functions of the parameter $t = p_2 = p_3$. Indeed, the involved matrix is no longer a 3×3 matrix as in (2) but is built with the five unknowns $(x_1, x_2, x_3, \lambda_1, \lambda_2)$ where λ_1 and λ_2 should be the corresponding Lagrange multipliers. Moreover, it should be calculated for all the possible loading paths $t \mapsto (p_2 = p_2(t),\ p_3 = p_3(t))$ or $h(p_2, p_3) = 0$, which becomes a quasi-impossible task. For the next case $n = 4$, the format of the involved matrix should be 7×7 whereas the loading path is described by any function $t \mapsto (p_2 = p_2(t),\ p_3 = p_3(t),\ p_4 = p_4(t))$. That shows that the second-order work criterion is an appropriate tool to tackle the constrained problem and that the main result (Theorem 2) is the appropriate geometrical approach that shortcuts these algebraic computations.

(2) Obviously in order to "see appear" the unstable behavior for the corresponding constrained system, the constrained system must be disturbed in one of the directions of the isotropic cone. At the boundary p^* it means in the X^* direction. If at $t = 0$ the perturbation is $U(0) = 0$ and $\dot{U}(0) = X^*$, then $t \mapsto U(t) = tX^*$ is a divergent solution of the (linear) dynamic equation of the constrained system and $\|U(t)\| \to +\infty$. On the contrary, in another direction of perturbation, no unstable behavior will be observed. It means that for concretely observing a divergence unstable evolution we have to first reach a threshold p^* in accordance with the loading path and given by the second-order work criterion, which also provides a direction X^*, second constrain the system by the appropriate constraint KX^*, and third perturb the system in the X^* direction. This direction is compatible with the kinematic constraint because, by construction, $(X^*)^T KX^* = 0$ since X^* is in the isotropic cone.

(3) From a geometric point of view, on the boundary ∂D_H the isotropic cone is degenerated into a vector space (a one-dimensional vector space "generally", which means that "generally" the rank of the matrix $K_s(p)$ drops from n to $n-1$ as p reaches p^* or as the determinant of $K_s(p)$ vanishes). Beyond the boundary, it is a real cone. For example, in Figure 4, consider the load $p = \left(\frac{3}{2}, \frac{3}{2}\right)$ which belongs to the domain $\det K_s(p) < 0$. According to Figure 4, for this loading we have $\det(K_s(p_d)) < 0$. For this loading, the isotropic cone is the blue surface plotted in Figure 5. The green line is in the direction

$$X_g = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

for which $X_g^T K(p_d) X_g > 0$, $X_g \in C_+\left(\left(\frac{3}{2}, \frac{3}{2}\right)\right)$, whereas the red line is in the direction

$$X_r = \begin{pmatrix} 1 \\ \frac{1}{2} \\ 0 \end{pmatrix}$$

for which $X_r^T K(p_d) X_r < 0$, $X_r \in C_-\left(\left(\frac{3}{2}, \frac{3}{2}\right)\right)$.

Case $n = 4$. The stiffness matrix reads

$$K(p_2, p_3, p_4) = \begin{pmatrix} 2-(p_2+p_3+p_4) & -1+p_2 & p_3 & p_4 \\ -1 & 2-(p_3+p_4) & -1+p_3 & p_4 \\ 0 & -1 & 2-p_4 & -1+p_4 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

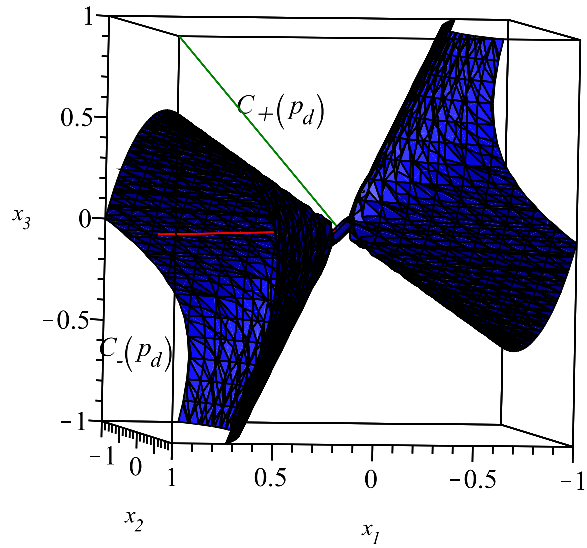


Figure 5. Isotropic cone (blue surface) for the load $p_d = \left(\frac{3}{2}, \frac{3}{2}\right)$ of the three degree of freedom discrete Leipholz column: $\det K_s(p_d) < 0$. Outside the isotropic cone (for example in the green line direction), then $X \in C_+(p_d)$: $X^T K_s(p_d) X > 0$. Inside the isotropic cone (for example in the red direction), then $X \in C_-(p_d)$: $X^T K_s(p_d) X < 0$. On the isotropic cone (blue surface), $X^T K_s(p_d) X = 0$.

and its symmetric part

$$K_s(p_2, p_3, p_4) = \begin{pmatrix} 2-(p_2+p_3) & -1+p_2/2 & p_3/2 & p_4/2 \\ -1+p_2/2 & 2-(p_3+p_4) & -1+p_3/2 & p_4/2 \\ p_3/2 & -1+p_3/2 & 2-p_4 & -1+p_4/2 \\ p_4/2 & p_4/2 & -1+p_4/2 & 1 \end{pmatrix}.$$

The domain of investigation lies now in the infinite cube ($p_2 \geq 0,\ p_3 \geq 0,\ p_4 \geq 0$). As above we suppose that any loading path starts from $(0, 0)$. The domain D_H of Hill stability is delimited by the blue surface Γ_H in the vicinity of $(0, 0)$ (see Figure 6).

D_H is given by the set of inequalities

$$D_H = \big\{0 \leq p_2 \leq 2,\ 0 \leq p_1 \leq 1,\ 0 \leq p_4 \leq 2-\sqrt{2},$$
$$1+\tfrac{1}{2}p_3^3-\tfrac{3}{4}p_4^4+4p_4^3+p_2p_3p_4-\tfrac{3}{4}p_2p_3p_4^2-\tfrac{1}{4}p_2^2+\tfrac{1}{16}p_2^2p_4^2-\tfrac{3}{4}p_2p_4^3+\tfrac{1}{2}p_2p_3^2-\tfrac{3}{4}p_3^2p_4^2$$
$$-\tfrac{3}{2}p_3p_4^3+\tfrac{3}{2}p_3^2p_4-\tfrac{1}{2}p_2p_3-\tfrac{1}{2}p_2p_4+2p_2p_4^2-\tfrac{3}{2}p_3^2-3p_3*p_4-5p_4^2+5p_3p_4^2 \geq 0\big\}.$$

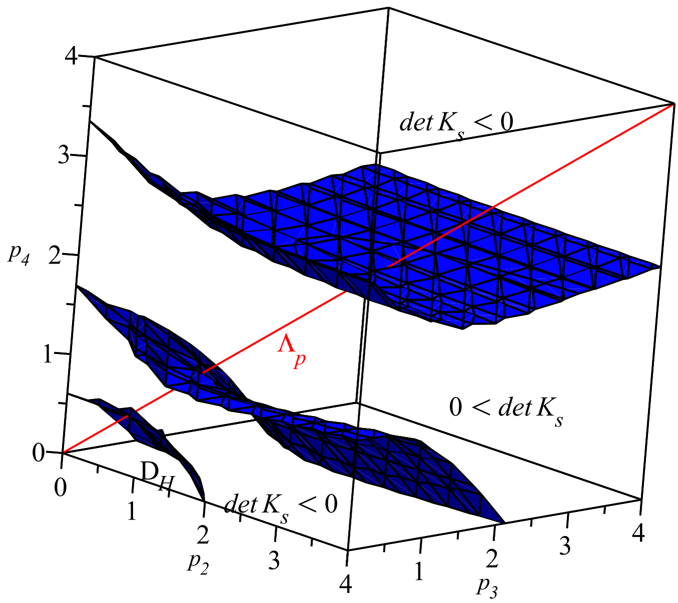


Figure 6. Hill stability domain D_H of the four degree of freedom discrete Leipholz column (D_{div} is the whole infinite cube ($p_2 \geq 0$, $p_3 \geq 0$, $p_4 \geq 0$)). Red line: Loading path $\Lambda_p : p_2 = p_3 = p_4$. Blue surface: three components of the hypersurface $\det(K_s(p)) = 0$.

Because the last inequality is a two-degree polynomial in the variable p_2, it can be solved explicitly. The explicit expression $p_2 = g(p_3, p_4)$ which is the explicit equation of the boundary $\partial D_H = \Gamma_H$ can be used to prove the convexity of D_H. The thresholds for the three intervals of variation for the variables p_2, p_3, p_4 are obtained by vanishing the two other variables and solving the remaining equalities.

As above, now choose the loading path $\Lambda_p = \{p = (p_2, p_3, p_4) \mid p_2 = p_3 = p_4\}$ plotted in red color in Figure 6. This line intersects the boundary $\partial D_H = \Gamma_H$ of the Hill stability domain at the point $p^* = (p_2^*, p_3^*, p_4^*)$ with $p_2^* = p_3^* = p_4^* \approx 0.4351852922$. Here p_2^*, p_3^*, p_4^* are solutions of the fourth-degree polynomial $\phi(t) = 1 - \frac{43}{4}t^2 + \frac{29}{2}t^3 - \frac{71}{16}t^4$ whose curve is plotted in Figure 7 and whose zeros give the values of the intersections of the red line Λ_p and the blue surface Γ_H in Figure 6. Then 0.4351852922 is a numerical approximation of the first positive root.

For this critical value, the isotropic cone of $K_s(p^*)$ is no longer reduced to $\{0\}$. It is generated by the vector

$$X^* = \begin{pmatrix} 0.742133032540456 \\ 0.625378368942899 \\ 0.0997261713362030 \\ -0.219533934555109 \end{pmatrix}.$$

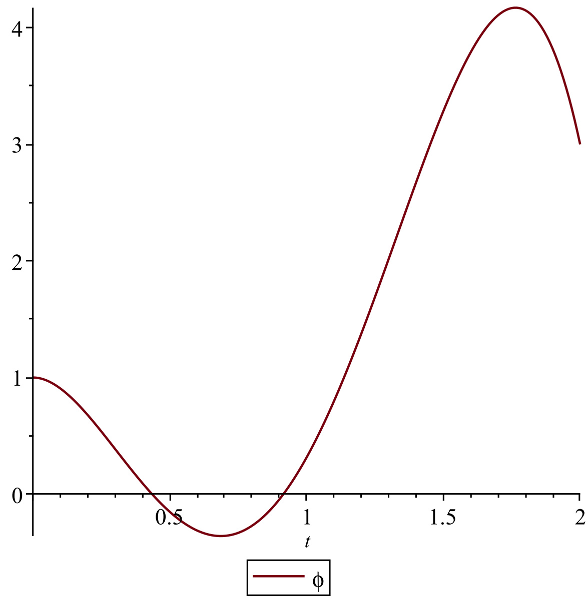


Figure 7. Function ϕ giving the critical load p^* on the loading path $\Lambda_p : p_2 = p_3 = p_4$.

Thus, applying the main theorem, the next step consists of calculating

$$K(X^*) = \begin{pmatrix} 0.110008445768715 \\ -0.187551978632908 \\ -0.345329394165552 \\ -0.319260105891312 \end{pmatrix},$$

which means that the system Σ_3 subjected to the kinematic constraint

$$c^*(\theta_1, \theta_2, \theta_3, \theta_4) = 0.110008445768715\theta_1 - 0.187551978632908\theta_2 \\ - 0.345329394165552\theta_3 - 0.345329394165552\theta_4 = 0$$

becomes, on the loading path $\Lambda_p = \{p = (p_2, p_3, p_4) \mid p_2 = p_3 = p_4\}$, divergence unstable when it is subjected to the critical load $p^* = (0.4351852922, 0.4351852922, 0.4351852922)$.

7. Conclusion

In this paper, the KISS concept is introduced and applied to shed new light on the sixty-year-old “competition” between the two criteria of stability investigated here: the second-order work criterion and the divergence criterion. We first start with an introductory example which possess all the necessary ingredients. Second we stress that the two criteria do not question exactly the same approach of stability. Third, by use of the KISS concept and thanks to the geometrical approach of constrained mechanics involving the compression of operators, a way is proposed to obtain a full equivalence.

This final equivalence between both criteria allowed us to highlight the significant original variational formulation on all the possible constrained systems keeping in mind that only one-constrained systems are finally involved in the solution. The multiparameter discrete Leipholz column example illustrates the involved concepts and shows the power of the geometric solution associated with the variational formulation.

The extension of wider frameworks (linear continuum mechanics and nonlinear discrete mechanics) have already been performed whereas the flutter-type instability does not lead to such beautiful results regarding the KISS issue. Indeed, according to the mass distribution, the full variety of situations may occur (universal or conditional KISS as well) and the KISS must then be investigated case by case. We believe that one of the last real challenges regarding both of these criteria is the investigation of the transition from the Hill stability criterion well adapted to a purely incremental quasistatic evolution to the Lyapunov dynamic stability approach.

References

[Absi and Lerbet 2004] E. Absi and J. Lerbet, "Instability of elastic bodies", *Mech. Res. Commun.* **31**:1 (2004), 39–44.

[Beck 1952] M. Beck, "Die Knicklast des einseitig eingespannten, tangential gedrückten Stabes", *Z. Angew. Math. Phys.* **3**:3 (1952), 225–228.

[Bigoni and Noselli 2011] D. Bigoni and G. Noselli, "Experimental evidence of flutter and divergence instabilities induced by dry friction", *J. Mech. Phys. Solids* **59**:10 (2011), 2208–2226.

[Bolotin 1963] V. V. Bolotin, *Nonconservative problems of the theory of elastic stability*, Pergamon, New York, 1963.

[Challamel et al. 2009] N. Challamel, F. Nicot, J. Lerbet, and F. Darve, "On the stability of non-conservative elastic systems under mixed perturbations", *Eur. J. Environ. Civ. En.* **13**:3 (2009), 347–367.

[Challamel et al. 2010] N. Challamel, F. Nicot, J. Lerbet, and F. Darve, "Stability of non-conservative elastic structures under additional kinematics constraints", *Eng. Struct.* **32**:10 (2010), 3086–3092.

[Daouadji et al. 2010] A. Daouadji, H. AlGali, F. Darve, and A. Zeghloul, "Instability in granular materials: experimental evidence of diffuse mode of failure for loose sands", *J. Eng. Mech.* **136**:5 (2010), 575–588.

[Darve 1994] F. Darve, "Liquefaction phenomenon: modelling, stability and uniqueness", pp. 1305–1319 in *Verification of numerical procedures for the analysis of soil liquefaction problems*, edited by K. Arulanandan and R. F. Scott, Balkema, Rotterdam, 1994.

[Darve 1996] F. Darve, "Liquefaction phenomenon of granular materials and constitutive stability", *Eng. Computation.* **13**:7 (1996), 5–28.

[Darve and Chau 1987] F. Darve and B. Chau, "Constitutive instabilities in incrementally non-linear modelling", pp. 301–310 in *Constitutive laws for engineering materials: theory and application* (Tucson, AZ, 1983), edited by C. S. Desai and R. H. Gallagher, Elsevier, Amsterdam, 1987.

[El Naschie 1976] M. S. El Naschie, "Post-critical behaviour of the beck problem", *J. Sound Vib.* **48**:3 (1976), 341–344.

[El Naschie 1977] M. S. El Naschie, "Some remarks on the Beck problem", *AIAA J.* **15** (1977), 1200–1201.

[Elishakoff 2005] I. Elishakoff, "Controversy associated with the so-called "follower forces": critical overview", *Appl. Mech. Rev.* **58**:2 (2005), 117–142.

[Hill 1958] R. Hill, "A general theory of uniqueness and stability in elastic-plastic solids", *J. Mech. Phys. Solids* **6**:3 (1958), 236–249.

[Hill 1959] R. Hill, "Some basic principles in the mechanics of solids without a natural time", *J. Mech. Phys. Solids* **7** (1959), 209–225.

[Ingerle 1969] K. Ingerle, "Die Verwendung des statischen Stabilitätskriteriums für Probleme der Elastostatik bei Fehlen nichttrivialer Gleichgewichtslagen", *Forsch. Ingenieurwes.* **35**:6 (1969), 196–197.

[Ingerle 2013] K. Ingerle, "Stability of massless non-conservative elastic systems", *J. Sound Vib.* **332**:19 (2013), 4529–4540.

[Ingerle 2018] K. Ingerle, *Non-conservative systems: new static and dynamic stability criteria*, CRC, Boca Raton, FL, 2018.

[Kirillov and Verhulst 2010] O. N. Kirillov and F. Verhulst, "Paradoxes of dissipation-induced destabilization or who opened Whitney's umbrella?", *Z. Angew. Math. Mech.* **90**:6 (2010), 462–488.

[Koiter 1996] W. T. Koiter, "Unrealistic follower forces", *J. Sound Vib.* **194**:4 (1996), 636–638.

[Lerbet et al. 2012] J. Lerbet, M. Aldowaji, N. Challamel, F. Nicot, F. Prunier, and F. Darve, "P-positive definite matrices and stability of nonconservative systems", *Z. Angew. Math. Mech.* **92**:5 (2012), 409–422.

[Lerbet et al. 2014] J. Lerbet, M. Aldowaji, N. Challamel, O. N. Kirillov, F. Nicot, and F. Darve, "Geometric degree of nonconservativity", *Math. Mech. Complex Syst.* **2**:2 (2014), 123–139.

[Lerbet et al. 2016a] J. Lerbet, N. Challamel, F. Nicot, and F. Darve, "Geometric degree of nonconservativity: set of solutions for the linear case and extension to the differentiable non-linear case", *Appl. Math. Model.* **40**:11–12 (2016), 5930–5941.

[Lerbet et al. 2016b] J. Lerbet, G. Hello, N. Challamel, F. Nicot, and F. Darve, "3-dimensional flutter kinematic structural stability", *Nonlinear Anal. Real World Appl.* **29** (2016), 19–37.

[Lerbet et al. 2017] J. Lerbet, N. Challamel, F. Nicot, and F. Darve, "On the stability of nonconservative continuous systems under kinematic constraints", *Z. Angew. Math. Mech.* **97**:9 (2017), 1100–1119.

[Lerbet et al. 2018] J. Lerbet, N. Challamel, F. Nicot, and F. Darve, "Coordinate free nonlinear incremental discrete mechanics", *Z. Angew. Math. Mech.* **98**:10 (2018), 1813–1833.

[Nicot et al. 2011] F. Nicot, A. Daouadji, F. Laouafa, and F. Darve, "Second-order work, kinetic energy and diffuse failure in granular materials", *Granul. Matter* **13**:1 (2011), 19–28.

[Nicot et al. 2012a] F. Nicot, N. Challamel, J. Lerbet, F. Prunier, and F. Darve, "Some insights into structure instability and the second-order work criterion", *Int. J. Solids Struct.* **49**:1 (2012), 132–142.

[Nicot et al. 2012b] F. Nicot, L. Sibille, and F. Darve, "Failure in rate-independent granular materials as a bifurcation toward a dynamic regime", *Int. J. Plasticity* **29** (2012), 136–154.

[Tarnai 2004] T. Tarnai, "Paradoxical behaviour of vibrating systems challenging Rayleigh's theorem,", pp. 11384 in *Mechanics of the 21st century: proceedings of the 21st International Congress of Theoretical and Applied* (Warsaw, 2004), edited by W. Gutkowski and T. A. Kowalewski, Springer, 2004.

[Thompson 1982] J. Thompson, "'Paradoxical' mechanics under fluid flow", *Nature* **296** (1982), 135–137.

Received 4 Jul 2019. Revised 6 Nov 2019. Accepted 7 Jan 2020.

JEAN LERBET: jlerbet@gmail.com
Université Paris-Saclay, CNRS, Université d'Evry, Laboratoire de Mathématiques et Modélisation d'Evry, Évry, France

NOËL CHALLAMEL: noel.challamel@univ-ubs.fr
Institut de Recherche Dupuy de Lôme, UMR CNRS 6027, Université de Bretagne Sud, Lorient, France

FRANÇOIS NICOT: francois.nicot@irstea.fr
INRAE, Université Grenoble Alpes, Unité de Recherche ETNA, Domaine Universitaire, Saint-Martin-d'Hères, France

FELIX DARVE: felix.darve@grenoble-inp.fr
Laboratoire 3SR, UMR CNRS 5521, Université Grenoble Alpes, Institut Polytechnique de Grenoble, Grenoble, France

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
Vol. 8, No. 1, 2020
dx.doi.org/10.2140/memocs.2020.8.29

ON THE GENERALIZATION OF THE BREWSTER LAW

FABRIZIO DAVÌ

In classical photoelasticity of transparent isotropic materials, the Brewster law states that the difference in principal refraction index (birefringence) is proportional to the difference in principal stress. Here we show that such a relation can be generalized to anisotropic crystals only for the high-symmetry classes of the cubic group and for a specific plane stress, also for the high-symmetry classes of the trigonal, tetragonal, and hexagonal groups. No further generalizations are possible.

1. Introduction

One of the most important relations in photoelasticity is the one which, for transparent isotropic materials, shows that the difference in the principal refraction index (birefringence) is proportional to the difference in the principal stress [Aben and Guillemet 1993]:

$$n_i - n_j = f_{\text{iso}}(\sigma_i - \sigma_j), \quad i, j = 1, 2, 3, \ i \neq j, \tag{1}$$

where n_i are the principal refraction indexes, σ_i are the principal values of the stress tensor $\boldsymbol{T}$, and f_{iso} is the photoelastic constant for isotropic materials which depends on the components of the piezooptic tensor and on the refraction index n_o of the unstressed material

Such a relation, obtained for the first time by David Brewster [1830] in 1818 and accordingly known as the "Brewster law", is at the basis of the experimental stress analysis based on photoelasticity. Besides the classical field of application in experimental structural mechanics (viz., [Kuske and Robertson 1974; Bain 2019]) such a relation allows, for instance, for the characterization and the quality control of high-energy physics crystals like the ones used in the CMS calorimeter at CERN or in the PANDA experiment at GSI in Darmstadt (viz., the recent review in [Montalto et al. 2019]).

Communicated by Francesco dell'Isola.
MSC2010: 78A10, 74F15, 74E10.
Keywords: photoelasticity, photoelastic constant, anisotropic crystals, Brewster law, scintillating crystals.

When we look at (1) from a mechanical point of view, by keeping in mind the much celebrated Mohr's circle [Sokolnikoff 1956], we see that it states the direct proportionality between the birefringence in the plane orthogonal to the k-th principal direction of stress and the shear stress in the same plane:

$$(\Delta n)_k = 2 f_{\text{iso}} \tau_k, \quad k = 1, 2, 3, \tag{2}$$

where

$$(\Delta n)_k = n_i - n_j, \quad \tau_k = \frac{\sigma_i - \sigma_j}{2}, \quad i, j = 1, 2, 3, \; i \neq j \neq k. \tag{3}$$

The experimental usefulness of such a relation thus becomes clear: the maximum birefringence is proportional to the maximum shear and accordingly can be related to failure criteria. On the other hand, provided we known the principal stress, a measurement of birefringence may allow for an experimental measurement of some components of the piezooptic tensor.

Here we look at the possibility to generalize the Brewster law for anisotropic materials to a more general relation in which the birefringences $(\Delta n)_k$, $k = 1, 2, 3$, are linear combinations of the three shear stresses τ_j, $j = 1, 2, 3$, with the coefficients f_{kj} of the linear combinations depending on the refraction index of the unstressed material, on the components of the piezooptic tensor, and on the angle between the principal optical and stress directions:[1]

$$(\Delta n)_k = (\Delta n)^o_k + f_{kj} \tau_j, \quad k, j = 1, 2, 3. \tag{4}$$

We look in detail at optically isotropic, uniaxial, and biaxial crystals and we show that, in order to arrive at a relation like (4), the birefringence must not depend on the spherical stress, an instance which depends on the crystal symmetry group (besides the trivial case of purely deviatoric (traceless) stresses). We show that this is possible for any stress in optically isotropic crystals, besides the isotropic case, only for the high-symmetry classes of the cubic group; moreover, when the stress has only diagonal component the obtained relation (4) simplifies into the isotropic one.

When we deal with optically uniaxial and biaxial crystals we show that in the general case it is not possible to arrive at (4). However, for uniaxial crystal we show that, for plane stress in the plane orthogonal to the material symmetry axis and provided the stressed crystal remains uniaxial, we can arrive at the isotropic relation (1) in the same plane for the high-symmetry trigonal, tetragonal, and hexagonal crystal classes.

[1] The same issue was addressed in [Rinaldi et al. 2018]; however, the results obtained there were related to specific cases of stress without the formal and more general treatment we give here; further, the treatment of cubic crystal was missing.

No other generalizations of the Brewster law are possible, besides for very specific material and particular cases of stress.

2. Crystal photoelasticity

2.1. *Propagation of light in crystals.* We consider a three-dimensional region composed of a polarized, nonconducting, and nonmagnetizable material: then for $\boldsymbol{e}$ and $\boldsymbol{h}$ the electric and magnetic fields, respectively, a monochromatic and linearly polarized light is the electromagnetic plane wave [Born and Wolf 1999]:

$$\boldsymbol{e} = \boldsymbol{e}_0 \cos \omega\Big(t - \frac{\boldsymbol{m}\cdot\boldsymbol{x}}{v}\Big), \qquad \boldsymbol{h} = \boldsymbol{h}_0 \cos \omega\Big(t - \frac{\boldsymbol{m}\cdot\boldsymbol{x}}{v}\Big), \qquad v = \frac{c}{n}, \tag{5}$$

where $\boldsymbol{e}_0$ and $\boldsymbol{h}_0$ are the constant amplitudes, ω is the *frequency*, $\boldsymbol{m}$ is the *direction of propagation*, $\|\boldsymbol{m}\| = 1$, v is the velocity of propagation of electromagnetic fields in a medium, c is the light velocity in the vacuum, and $|n| < 1$ is the *refraction index*.

The propagation condition for (5) is provided by the Maxwell equations with null total charge density,

$$\operatorname{curl} \boldsymbol{e} = -\mu_o \frac{\partial \boldsymbol{h}}{\partial t}, \qquad \operatorname{div} \boldsymbol{d} = 0, \qquad \operatorname{curl} \boldsymbol{h} = \frac{\partial \boldsymbol{d}}{\partial t}, \qquad \operatorname{div} \boldsymbol{h} = 0, \tag{6}$$

where $\boldsymbol{d}$ denotes the electric displacement and μ_0 is the vacuum permeability.

For an anisotropic dielectric material the constitutive relation between the electric field and the electric displacement is given by [Sirotin and Shaskolskaya 1982; Perelomova and Tagieva 1983; Nye 1985][2]

$$\boldsymbol{e} = \epsilon_0^{-1} \boldsymbol{B} \boldsymbol{d}, \tag{7}$$

where $\boldsymbol{B} \in \mathrm{Sym}^+$ is the *dielectric rigidity* (or dielectric impermeability or inverse permittivity) tensor, normalized with respect to the dielectric permittivity of the vacuum ϵ_0.

By using (5) and (7) in (6) and since $c^2 = \epsilon_o \mu_o$, then we arrive at the propagation condition

$$(\boldsymbol{M}(\boldsymbol{m}) - n^{-2} \boldsymbol{P}^{\perp}(\boldsymbol{m}))\boldsymbol{d} = \boldsymbol{0}, \tag{8}$$

where the *characteristic tensor* $\boldsymbol{M}(\boldsymbol{m}) \in \mathrm{Sym}$ is defined by

$$\boldsymbol{M}(\boldsymbol{m}) = \boldsymbol{P}^{\perp}(\boldsymbol{m}) \boldsymbol{B} \boldsymbol{P}^{\perp}(\boldsymbol{m}), \qquad \boldsymbol{P}^{\perp}(\boldsymbol{m}) = \boldsymbol{I} - \boldsymbol{m} \otimes \boldsymbol{m}; \tag{9}$$

here $\boldsymbol{P}^{\perp}(\boldsymbol{m})$ is the orthogonal projection on the plane normal to $\boldsymbol{m}$ and accordingly $\boldsymbol{M}(\boldsymbol{m})$ is the restriction of $\boldsymbol{B}$ to the plane orthogonal to $\boldsymbol{m}$.

[2] In the sequel we shall denote with Sym the subspace of symmetric tensors and Sym^+ the subspace of positive-definite symmetric tensors; Dev is the subspace of traceless symmetric tensors whereas Orth^+ denotes the proper orthogonal group.

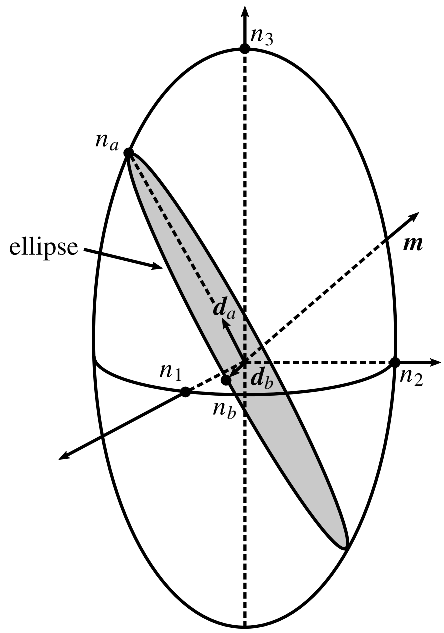


Figure 1. The optical indicatrix.

By the positive-definiteness of $\boldsymbol{B}$ we can define the *optical indicatrix* or *conjugate index ellipsoid* that is the locus of normalized constant dielectric energy [Born and Wolf 1999], which summarizes at a glance all the information about the crystal optical anisotropy:

$$\boldsymbol{B}z \cdot z = 1, \tag{10}$$

where z is a normalized electric displacement; hence $\boldsymbol{M}(\boldsymbol{m})$ is associated with the ellipse we get by intersecting the optical indicatrix with the plane orthogonal to $\boldsymbol{m}$ through the ellipsoid center. See Figure 1.

The eigenvalue problem associated with the propagation condition (8) admits, for each direction of propagation $\boldsymbol{m}$, at most two eigencouples

$$(n_a^{-2}, \boldsymbol{d}_a), \qquad (n_b^{-2}, \boldsymbol{d}_b) \tag{11}$$

whose eigenvectors are mutually orthogonal and lie in the plane orthogonal to $\boldsymbol{m}$: clearly the eigenvalues are the semiaxis of the ellipse described by $\boldsymbol{M}(\boldsymbol{m})$ and whose directions are spanned by the eigenvectors.

Whenever $n_a \neq n_b$ we have two different velocities $v_a \neq v_b$ associated with a given direction of propagation $\boldsymbol{m}$: such a phenomena is called *double refraction* and is measured in terms of the *birefringence* Δn:

$$\Delta n = n_a - n_b; \tag{12}$$

it makes sense to search for the directions of propagation $\boldsymbol{m}$ such that $\Delta n = 0$ and no double refraction holds. We define *optic axis* as these propagation directions and there are at most three possibilities, depending on the multiplicity of the eigenvalues of $\boldsymbol{B}$.

- Three equal eigenvalues:

$$\boldsymbol{B} = n_o^{-2}\boldsymbol{I}. \tag{13}$$

The material is *optically isotropic*, each direction is an optic axis, and there is no birefringence.

- Two equal eigenvalues:

$$\boldsymbol{B} = n_o^{-2}(\boldsymbol{I} - \boldsymbol{e}\otimes\boldsymbol{e}) + n_e^{-2}\boldsymbol{e}\otimes\boldsymbol{e}, \quad \|\boldsymbol{e}\| = 1. \tag{14}$$

The material is *optically uniaxial*, the direction $\boldsymbol{e}$ is the unique optic axis, and the maximum birefringence is attained for propagation directions orthogonal to $\boldsymbol{e}$:

$$(\Delta n)_{\max} = n_e - n_o. \tag{15}$$

We say n_e and n_o are the *extraordinary* and *ordinary* refraction indices and we define a crystal to be optically *positive* or *negative* when $n_e < n_o$ or $n_e > n_o$, respectively.

- Three different eigenvalues $B_1 > B_2 > B_3$:

$$\boldsymbol{B} = B_1\boldsymbol{e}_1\otimes\boldsymbol{e}_1 + B_2\boldsymbol{e}_2\otimes\boldsymbol{e}_2 + B_3\boldsymbol{e}_3\otimes\boldsymbol{e}_3, \quad B_i = n_i^{-2}. \tag{16}$$

The material is called *optically biaxial*; whenever the propagation direction coincides with an optic axis the ellipse associated with $\boldsymbol{M}(\boldsymbol{m})$ degenerates into a circle which lies in the plane containing $\boldsymbol{e}_2$; then there are two optic axes orthogonal to $\boldsymbol{e}_2$ and bisected by the direction spanned by $\boldsymbol{e}_1$.

We have three birefringences corresponding to directions of propagation $\boldsymbol{m} = \boldsymbol{u}_3$, $\boldsymbol{m} = \boldsymbol{u}_2$, and $\boldsymbol{m} = \boldsymbol{u}_1$, respectively:

$$(\Delta n)_3 = n_1 - n_2, \qquad (\Delta n)_2 = n_3 - n_1, \qquad (\Delta n)_1 = n_2 - n_3, \tag{17}$$

with

$$(\Delta n)_{\max} = \sup\{|(\Delta n)_1|, |(\Delta n)_2|, |(\Delta n)_3|\}. \tag{18}$$

In biaxial crystals, if the value of intermediate refractive index is closer to that of highest refractive index, the crystal is optically *negative*, and if it is closer to lowest refractive index, then the crystal is optically *positive*.

2.2. *Photoelasticity of crystals.* In *photoelastic crystals* the dielectric impermeability is a linear function of the Cauchy stress tensor $\boldsymbol{T} \in \mathrm{Sym}$:

$$\boldsymbol{B}(\boldsymbol{T}) = \boldsymbol{B}_o + \mathbb{M}[\boldsymbol{T}], \tag{19}$$

where $\mathbb{M}:\mathrm{Sym}\to\mathrm{Sym}$ is the fourth-order Maxwell piezooptic tensor and $\boldsymbol{B}_o$ is the dielectric permeability in the unstressed state. We define the *symmetry group* $\mathcal{G}$ for a photoelastic crystal as

$$\mathcal{G}\equiv\{\boldsymbol{Q}\in\mathrm{Orth}^+ \mid \boldsymbol{Q}\mathbb{M}[\boldsymbol{T}]\boldsymbol{Q}^T=\mathbb{M}[\boldsymbol{Q}\boldsymbol{T}\boldsymbol{Q}^T] \text{ and } \boldsymbol{Q}\boldsymbol{B}_o\boldsymbol{Q}^T=\boldsymbol{B}_o \text{ for all } \boldsymbol{T}\in\mathrm{Sym}\}.$$

Let $\Sigma\equiv\{\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_3\}$ be an orthonormal frame; then the components of the piezo-optic tensor obeys

$$\mathbb{M}_{ijhk}=\mathbb{M}_{jihk}=\mathbb{M}_{ijkh},\quad i,j,h,k=1,2,3, \tag{20}$$

and the tabular form of these components for the various crystallographic classes and groups is provided in the Appendix.

As far the tensor $\boldsymbol{B}_o$ is concerned, in the frame Σ its matrix has six independent components for crystals of the triclinic group, whereas for the crystals of the monoclinic group (all classes) we have

$$\boldsymbol{B}_o\equiv\begin{bmatrix} B_{11} & B_{12} & 0\\ \cdot & B_{22} & 0\\ \cdot & \cdot & B_{33}\end{bmatrix}, \tag{21}$$

where $\boldsymbol{e}_3$ is directed as the monoclinic b-axis. For orthorhombic crystals (all groups) we have instead

$$\boldsymbol{B}_o\equiv\begin{bmatrix} B_{11} & 0 & 0\\ \cdot & B_{22} & 0\\ \cdot & \cdot & B_{33}\end{bmatrix}; \tag{22}$$

in all the three cases $\boldsymbol{B}$ admits the representation (16) and hence triclinic, monoclinic, and orthorhombic crystals are optically biaxial.

Crystals of the tetragonal, trigonal, and hexagonal group (all classes) are optically uniaxial, since $\boldsymbol{B}_o$ admits the representation (14):

$$\boldsymbol{B}_o\equiv\begin{bmatrix} B_{11} & 0 & 0\\ \cdot & B_{11} & 0\\ \cdot & \cdot & B_{33}\end{bmatrix}, \tag{23}$$

where $\boldsymbol{e}_3$ is directed in this case as the symmetry c-axis.

The tensor $\boldsymbol{B}_o$ for cubic crystals and isotropic materials has the representation (13) and the material is optically isotropic.

The stress $\boldsymbol{T}$ changes the optical properties of materials: indeed an isotropic material can become uniaxial or biaxial when stressed, and a uniaxial one can behave biaxially upon the application of a stress: for a complete description of such changes see the analysis presented in [Davì 2015].

3. The generalization of the Brewster law for anisotropic crystals

For optically isotropic materials we have the following the stress-optic relation called the *Brewster law* (viz., [Aben and Guillemet 1993]):

$$n_i - n_k = f_{\text{iso}}(\sigma_i - \sigma_k), \quad i, k = 1, 2, 3, \ i \neq k, \ f_{\text{iso}} = n_o^3 \frac{\mathbb{M}_{1122} - \mathbb{M}_{1111}}{2}, \quad (24)$$

where n_o is the isotropic refraction index and $(\sigma_1, \sigma_2, \sigma_3)$ are the principal components of $\boldsymbol{T}$. From a mechanical point of view (24) relates the birefringences (17) with (twice) the shear stress in the same plane

Here we wish to obtain the restrictions on the piezooptic tensor and on the stress which allow us to generalize (24) for anisotropic crystals to a relation between the birefringences (17) and all (twice) the shear stresses:

$$(\Delta n)_k = (\Delta n)_k^o + f_{1k}(\sigma_2 - \sigma_3) + f_{2k}(\sigma_3 - \sigma_1) + f_{3k}(\sigma_1 - \sigma_2), \quad k = 1, 2, 3, \quad (25)$$

with $(\Delta n)_k^o$ depending on $\boldsymbol{B}_o$ and the constant f_{ik} depending on both $\boldsymbol{B}_o$ and $\mathbb{M}$.

Let $(\sigma_k, \boldsymbol{w}_k)$ be the eigencouples of a generic stress $\boldsymbol{T}$ with

$$\boldsymbol{w}_k = \boldsymbol{R}\boldsymbol{e}_k, \quad k = 1, 2, 3, \ \boldsymbol{R} \in \text{Orth}^+; \quad (26)$$

then by the decomposition of $\boldsymbol{T}$ into its spherical and deviatoric parts

$$\boldsymbol{T} = \sigma_m \boldsymbol{I} + \operatorname{dev} \boldsymbol{T}, \quad \sigma_m = \tfrac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3), \quad (27)$$

we get, in the frame $\Sigma_{\boldsymbol{T}} \equiv \{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3\}$,

$$\operatorname{dev} \boldsymbol{T} \equiv \frac{1}{3} \begin{bmatrix} 2\sigma_1 - \sigma_2 - \sigma_3 & 0 & 0 \\ \cdot & 2\sigma_2 - \sigma_1 - \sigma_3 & 0 \\ \cdot & \cdot & 2\sigma_3 - \sigma_1 - \sigma_2 \end{bmatrix}, \quad (28)$$

which can be rewritten in terms of the shear stresses

$$\tau_1 = \frac{\sigma_2 - \sigma_3}{2}, \quad \tau_2 = \frac{\sigma_3 - \sigma_1}{2}, \quad \tau_3 = \frac{\sigma_1 - \sigma_2}{2} \quad (29)$$

as

$$\operatorname{dev} \boldsymbol{T} \equiv \frac{2}{3} \begin{bmatrix} \tau_3 - \tau_2 & 0 & 0 \\ \cdot & \tau_1 - \tau_3 & 0 \\ \cdot & \cdot & \tau_2 - \tau_1 \end{bmatrix}. \quad (30)$$

By (19), (27), and (30) then

$$\begin{aligned} \boldsymbol{B}(\sigma_m, \tau_1, \tau_2, \tau_3) &= \boldsymbol{B}_o + \sigma_m \mathbb{M}[\boldsymbol{I}] + \tfrac{2}{3}(\tau_3 - \tau_2)\mathbb{M}[\boldsymbol{w}_1 \otimes \boldsymbol{w}_1] \\ &\quad + \tfrac{2}{3}(\tau_1 - \tau_3)\mathbb{M}[\boldsymbol{w}_2 \otimes \boldsymbol{w}_2] + \tfrac{2}{3}(\tau_2 - \tau_1)\mathbb{M}[\boldsymbol{w}_3 \otimes \boldsymbol{w}_3] \\ &= \boldsymbol{B}_o + \sigma_m \mathbb{M}[\boldsymbol{I}] + \tfrac{2}{3}\tau_3 \mathbb{M}[\boldsymbol{w}_1 \otimes \boldsymbol{w}_1 - \boldsymbol{w}_2 \otimes \boldsymbol{w}_2] \\ &\quad + \tfrac{2}{3}\tau_2 \mathbb{M}[\boldsymbol{w}_3 \otimes \boldsymbol{w}_3 - \boldsymbol{w}_1 \otimes \boldsymbol{w}_1] + \tfrac{2}{3}\tau_1 \mathbb{M}[\boldsymbol{w}_2 \otimes \boldsymbol{w}_2 - \boldsymbol{w}_3 \otimes \boldsymbol{w}_3]. \quad (31) \end{aligned}$$

Let $(B_k, \boldsymbol{u}_k)$ the eigencouples of $\boldsymbol{B}(\boldsymbol{T})$ with

$$\boldsymbol{u}_k = \widehat{\boldsymbol{R}}\boldsymbol{e}_k, \quad k = 1, 2, 3, \;\; \widehat{\boldsymbol{R}} \in \text{Orth}^+; \tag{32}$$

then we have

$$B_k(\sigma_m, \tau_1, \tau_2, \tau_3) = \boldsymbol{B}(\sigma_m, \tau_1, \tau_2, \tau_3) \cdot \boldsymbol{u}_k \otimes \boldsymbol{u}_k, \quad k = 1, 2, 3, \text{ not summed}, \tag{33}$$

and by $(16)_2$ the principal refraction index of the stressed material is given by

$$n_k(\sigma_m, \tau_1, \tau_2, \tau_3) = \frac{1}{\sqrt{B_k(\sigma_m, \tau_1, \tau_2, \tau_3)}}, \tag{34}$$

a relation that can be linearized into

$$\begin{aligned} &n_k(\sigma_m, \tau_1, \tau_2, \tau_3) \\ &\quad = \frac{1}{\sqrt{B_k(\mathbf{0})}} - \frac{1}{2B_k^{3/2}(\mathbf{0})}\left(\left.\frac{\partial B_k}{\partial \sigma_m}\right|_{\mathbf{0}} \sigma_m + \sum_{j=1}^{3} \left.\frac{\partial B_k}{\partial \tau_j}\right|_{\mathbf{0}} \tau_j\right) + o(\|\boldsymbol{T}\|^2). \end{aligned} \tag{35}$$

By (31) and (32) (in all of the following relation k is fixed and not summed)

$$\begin{aligned} B_k(\mathbf{0}) = \boldsymbol{B}_o \cdot \widehat{\boldsymbol{R}}\boldsymbol{e}_k \otimes \widehat{\boldsymbol{R}}\boldsymbol{e}_k \qquad &= B^o_{ij}\boldsymbol{e}_i \otimes \boldsymbol{e}_j \cdot \widehat{\boldsymbol{R}}\boldsymbol{e}_k \otimes \widehat{\boldsymbol{R}}\boldsymbol{e}_k \\ = B^o_{ij}(\widehat{\boldsymbol{R}}\boldsymbol{e}_k \cdot \boldsymbol{e}_i)(\widehat{\boldsymbol{R}}\boldsymbol{e}_k \cdot \boldsymbol{e}_j) &= B^o_{ij}\widehat{R}_{jk}\widehat{R}_{ik}; \end{aligned} \tag{36}$$

moreover,

$$\left.\frac{\partial B_k}{\partial \sigma_m}\right|_{\mathbf{0}} = \mathbb{M}[\boldsymbol{I}] \cdot \boldsymbol{u}_k \otimes \boldsymbol{u}_k = \mathbb{M}[\boldsymbol{I}] \cdot \widehat{\boldsymbol{R}}\boldsymbol{e}_k \otimes \widehat{\boldsymbol{R}}\boldsymbol{e}_k = \widehat{\boldsymbol{R}}^T\mathbb{M}[\boldsymbol{I}]\widehat{\boldsymbol{R}} \cdot \boldsymbol{e}_k \otimes \boldsymbol{e}_k = M_{kk}, \tag{37}$$

and by using also (26) finally we get

$$\begin{aligned} \left.\frac{\partial B_k}{\partial \tau_1}\right|_{\mathbf{0}} &= \tfrac{2}{3}\mathbb{M}[\boldsymbol{w}_2 \otimes \boldsymbol{w}_2 - \boldsymbol{w}_3 \otimes \boldsymbol{w}_3] \cdot \boldsymbol{u}_k \otimes \boldsymbol{u}_k \\ &= \tfrac{2}{3}\widehat{\boldsymbol{R}}^T\mathbb{M}[\boldsymbol{R}(\boldsymbol{e}_2 \otimes \boldsymbol{e}_2 - \boldsymbol{e}_3 \otimes \boldsymbol{e}_3)\boldsymbol{R}^T]\widehat{\boldsymbol{R}} \cdot \boldsymbol{e}_k \otimes \boldsymbol{e}_k \\ &= 2(\overline{\mathbb{M}}_{22kk} - \overline{\mathbb{M}}_{33kk}), \\ \left.\frac{\partial B_k}{\partial \tau_2}\right|_{\mathbf{0}} &= \tfrac{2}{3}\mathbb{M}[\boldsymbol{w}_3 \otimes \boldsymbol{w}_3 - \boldsymbol{w}_1 \otimes \boldsymbol{w}_1] \cdot \boldsymbol{u}_k \otimes \boldsymbol{u}_k \\ &= \tfrac{2}{3}\widehat{\boldsymbol{R}}^T\mathbb{M}[\boldsymbol{R}(\boldsymbol{e}_3 \otimes \boldsymbol{e}_3 - \boldsymbol{e}_1 \otimes \boldsymbol{e}_1)\boldsymbol{R}^T]\widehat{\boldsymbol{R}} \cdot \boldsymbol{e}_k \otimes \boldsymbol{e}_k \\ &= 2(\overline{\mathbb{M}}_{33kk} - \overline{\mathbb{M}}_{11kk}), \\ \left.\frac{\partial B_k}{\partial \tau_3}\right|_{\mathbf{0}} &= \tfrac{2}{3}\mathbb{M}[\boldsymbol{w}_1 \otimes \boldsymbol{w}_1 - \boldsymbol{w}_1 \otimes \boldsymbol{w}_1] \cdot \boldsymbol{u}_k \otimes \boldsymbol{u}_k \\ &= \tfrac{2}{3}\widehat{\boldsymbol{R}}^T\mathbb{M}[\boldsymbol{R}(\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 - \boldsymbol{e}_2 \otimes \boldsymbol{e}_2)\boldsymbol{R}^T]\widehat{\boldsymbol{R}} \cdot \boldsymbol{e}_k \otimes \boldsymbol{e}_k \\ &= 2(\overline{\mathbb{M}}_{11kk} - \overline{\mathbb{M}}_{22kk}), \end{aligned} \tag{38}$$

where

$$\overline{\mathbb{M}}_{hhkk} = \tfrac{1}{3}\widehat{\boldsymbol{R}}^T \mathbb{M}[\boldsymbol{R}(\boldsymbol{e}_h \otimes \boldsymbol{e}_h)\boldsymbol{R}^T]\widehat{\boldsymbol{R}} \cdot \boldsymbol{e}_k \otimes \boldsymbol{e}_k, \quad h, k \text{ fixed.} \tag{39}$$

By using (37), (38), and (29) in (35) we arrive at the linearized expression of the principal refraction index in terms of the differences between principal stress:

$$\begin{aligned} n_1(\sigma_1,\sigma_2,\sigma_3) &= \frac{1}{\sqrt{B_1(\mathbf{0})}} - \frac{1}{2B_1^{3/2}(\mathbf{0})}(\sigma_m M_{11} + (\sigma_2-\sigma_3)(\overline{\mathbb{M}}_{2211}-\overline{\mathbb{M}}_{3311}) \\ &\quad + (\sigma_3-\sigma_1)(\overline{\mathbb{M}}_{3311}-\overline{\mathbb{M}}_{1111}) + (\sigma_1-\sigma_2)(\overline{\mathbb{M}}_{1111}-\overline{\mathbb{M}}_{2211})), \\ n_2(\sigma_1,\sigma_2,\sigma_3) &= \frac{1}{\sqrt{B_2(\mathbf{0})}} - \frac{1}{2B_2^{3/2}(\mathbf{0})}(\sigma_m M_{22} + (\sigma_2-\sigma_3)(\overline{\mathbb{M}}_{2222}-\overline{\mathbb{M}}_{3322}) \\ &\quad + (\sigma_3-\sigma_1)(\overline{\mathbb{M}}_{3322}-\overline{\mathbb{M}}_{1122}) + (\sigma_1-\sigma_2)(\overline{\mathbb{M}}_{1122}-\overline{\mathbb{M}}_{2222})), \\ n_3(\sigma_1,\sigma_2,\sigma_3) &= \frac{1}{\sqrt{B_3(\mathbf{0})}} - \frac{1}{2B_3^{3/2}(\mathbf{0})}(\sigma_m M_{33} + (\sigma_2-\sigma_3)(\overline{\mathbb{M}}_{2233}-\overline{\mathbb{M}}_{3333}) \\ &\quad + (\sigma_3-\sigma_1)(\overline{\mathbb{M}}_{3333}-\overline{\mathbb{M}}_{1133}) + (\sigma_1-\sigma_2)(\overline{\mathbb{M}}_{1133}-\overline{\mathbb{M}}_{2233})). \end{aligned} \tag{40}$$

We are now in position to write the birefringences (17) as

$$\begin{aligned} (\Delta n)_1 &= (\Delta n)_1^o + f_1\sigma_m + f_{11}(\sigma_2-\sigma_3) + f_{21}(\sigma_3-\sigma_1) + f_{31}(\sigma_1-\sigma_2), \\ (\Delta n)_2 &= (\Delta n)_2^o + f_2\sigma_m + f_{12}(\sigma_2-\sigma_3) + f_{22}(\sigma_3-\sigma_1) + f_{32}(\sigma_1-\sigma_2), \\ (\Delta n)_3 &= (\Delta n)_3^o + f_3\sigma_m + f_{13}(\sigma_2-\sigma_3) + f_{23}(\sigma_3-\sigma_1) + f_{33}(\sigma_1-\sigma_2), \end{aligned} \tag{41}$$

where

$$\begin{aligned} (\Delta n)_1^o &= B_2^{-1/2}(\mathbf{0}) - B_3^{-1/2}(\mathbf{0}), \\ (\Delta n)_2^o &= B_3^{-1/2}(\mathbf{0}) - B_1^{-1/2}(\mathbf{0}), \\ (\Delta n)_3^o &= B_1^{-1/2}(\mathbf{0}) - B_2^{-1/2}(\mathbf{0}), \end{aligned} \tag{42}$$

$$\begin{aligned} f_1 &= \frac{1}{2B_3^{3/2}(\mathbf{0})}M_{33} - \frac{1}{2B_2^{3/2}(\mathbf{0})}M_{22}, \\ f_2 &= \frac{1}{2B_1^{3/2}(\mathbf{0})}M_{11} - \frac{1}{2B_3^{3/2}(\mathbf{0})}M_{33}, \\ f_3 &= \frac{1}{2B_2^{3/2}(\mathbf{0})}M_{22} - \frac{1}{2B_1^{3/2}(\mathbf{0})}M_{11}, \end{aligned} \tag{43}$$

and with the components of the nonsymmetric matrix $[f_{ik}]$ given by

$$
\begin{aligned}
f_{11} &= \frac{1}{2B_2^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{3322} - \overline{\mathbb{M}}_{2222}) + \frac{1}{2B_3^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{2233} - \overline{\mathbb{M}}_{3333}),\\
f_{12} &= \frac{1}{2B_2^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{1122} - \overline{\mathbb{M}}_{3322}) + \frac{1}{2B_3^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{3333} - \overline{\mathbb{M}}_{1133}),\\
f_{13} &= \frac{1}{2B_2^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{2222} - \overline{\mathbb{M}}_{1122}) + \frac{1}{2B_3^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{1133} - \overline{\mathbb{M}}_{2233}),\\
f_{21} &= \frac{1}{2B_3^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{3333} - \overline{\mathbb{M}}_{2233}) + \frac{1}{2B_1^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{2211} - \overline{\mathbb{M}}_{3311}),\\
f_{22} &= \frac{1}{2B_3^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{1133} - \overline{\mathbb{M}}_{3333}) + \frac{1}{2B_1^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{3311} - \overline{\mathbb{M}}_{1111}),\\
f_{23} &= \frac{1}{2B_3^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{2233} - \overline{\mathbb{M}}_{1133}) + \frac{1}{2B_1^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{1111} - \overline{\mathbb{M}}_{2211}),\\
f_{31} &= \frac{1}{2B_1^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{3311} - \overline{\mathbb{M}}_{2211}) + \frac{1}{2B_2^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{2222} - \overline{\mathbb{M}}_{3322}),\\
f_{32} &= \frac{1}{2B_1^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{1111} - \overline{\mathbb{M}}_{3311}) + \frac{1}{2B_2^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{3322} - \overline{\mathbb{M}}_{1122}),\\
f_{33} &= \frac{1}{2B_1^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{2211} - \overline{\mathbb{M}}_{1111}) + \frac{1}{2B_2^{3/2}(\mathbf{0})}(\overline{\mathbb{M}}_{1122} - \overline{\mathbb{M}}_{2222}).
\end{aligned}
\tag{44}
$$

Relations (41) differ from the desired expression (25) by the presence of the terms in σ_m: at a glance, we see that the necessary condition to have a generalized Brewster law (25) is that either $\boldsymbol{T}$ is traceless or, in the general case,

$$f_1 = f_2 = f_3 = 0, \tag{45}$$

a condition which, in terms of the piezooptic tensor $\mathbb{M}$ and the inverse permittivity $\boldsymbol{B}_o$, reads

$$
\begin{aligned}
B_2^{2/3}(\mathbf{0})M_{33} - B_3^{2/3}(\mathbf{0})M_{22} &= 0,\\
B_3^{2/3}(\mathbf{0})M_{11} - B_1^{2/3}(\mathbf{0})M_{33} &= 0,\\
B_1^{2/3}(\mathbf{0})M_{22} - B_2^{2/3}(\mathbf{0})M_{11} &= 0;
\end{aligned}
\tag{46}
$$

it is easy to show that (46) admits the solution

$$M_{kk} = \alpha B_k^{2/3}(\mathbf{0}), \quad k = 1, 2, 3, \ \alpha \in \mathbb{R}. \tag{47}$$

In the following subsections we shall deal in detail with condition (47) and with the expression of the components of (25) for the various symmetry groups and classes:

in particular we shall search for the cases in which (47) holds and those in which the matrix $[f_{ik}]$ is symmetric or diagonal.

3.1. *Optically isotropic.* For optically isotropic materials, the inverse permittivity has the representation (13): accordingly $B_k(\mathbf{0}) = n_o^{-2}$ and $(\Delta n)_k^o = 0$ for all k. As far as the piezooptic tensor and conditions (47) are concerned, we need to analyze in detail the various cases.

3.1.1. *Isotropic materials.* For isotropic materials the piezooptic tensor is symmetric and from (13), (19), and (78) of the Appendix we get

$$\boldsymbol{B}(\boldsymbol{T}) = n_o^{-2} + \sigma_m(\mathbb{M}_{1111} + 2\mathbb{M}_{1122})\boldsymbol{I} + (\mathbb{M}_{1111} - \mathbb{M}_{1122})\,\mathrm{dev}\,\boldsymbol{T}, \qquad (48)$$

and the principal directions of $\boldsymbol{B}(\boldsymbol{T})$ coincide with those of $\mathrm{dev}\,\boldsymbol{T}$ with $\boldsymbol{R} = \widehat{\boldsymbol{R}}$; by the definition of isotropic group

$$\widehat{\boldsymbol{R}}^T \mathbb{M}[\boldsymbol{R}\boldsymbol{A}\boldsymbol{R}^T]\widehat{\boldsymbol{R}} = \widehat{\boldsymbol{R}}^T \boldsymbol{R}\mathbb{M}[\boldsymbol{A}]\boldsymbol{R}^T \widehat{\boldsymbol{R}} = \mathbb{M}[\boldsymbol{A}] \quad \text{for all } \boldsymbol{A} \in \mathrm{Sym}, \qquad (49)$$

then we have

$$\begin{aligned} M_{kk} &= \mathbb{M}[\boldsymbol{I}] \cdot \boldsymbol{e}_k \otimes \boldsymbol{e}_k = \tfrac{1}{3}(\mathbb{M}_{1111} + 2\mathbb{M}_{1122})\boldsymbol{I} \cdot \boldsymbol{e}_k \otimes \boldsymbol{e}_k \\ &= \frac{\mathbb{M}_{1111} + 2\mathbb{M}_{1122}}{3} \quad \text{for all } k, \end{aligned} \qquad (50)$$

and condition (47) is verified for $3\alpha = (\mathbb{M}_{1111} + 2\mathbb{M}_{1122})n_o^3$.

Moreover, by (49), relation (39) yields

$$\overline{\mathbb{M}}_{hhkk} = \tfrac{1}{3}\mathbb{M}[\boldsymbol{e}_h \otimes \boldsymbol{e}_h)] \cdot \boldsymbol{e}_k \otimes \boldsymbol{e}_k = \tfrac{1}{3}\mathbb{M}_{kkhh}, \quad h, k \text{ fixed}, \qquad (51)$$

and (44) gives in turn

$$\begin{aligned} f_{11} = f_{22} = f_{33} &= \frac{n_o^3}{2}\frac{2}{3}(\mathbb{M}_{1122} - \mathbb{M}_{1111}), \\ f_{12} = f_{12} = f_{13} = f_{31} = f_{23} = f_{32} &= -\frac{n_o^3}{2}\frac{1}{3}(\mathbb{M}_{1122} - \mathbb{M}_{1111}), \end{aligned} \qquad (52)$$

and accordingly (41) reduces to (24).

3.1.2. *Cubic crystals, classes* 432*,* $\bar{4}3m$*, and* $m3m$*.* Also in this case the piezooptic tensor is symmetric and from (76) of the Appendix we have

$$\begin{aligned} M_{kk} &= \widehat{\boldsymbol{R}}^T \mathbb{M}[\boldsymbol{I}]\widehat{\boldsymbol{R}} \cdot \boldsymbol{e}_k \otimes \boldsymbol{e}_k = \tfrac{1}{3}(\mathbb{M}_{1111} + 2\mathbb{M}_{1122})\widehat{\boldsymbol{R}}^T \widehat{\boldsymbol{R}} \cdot \boldsymbol{e}_k \otimes \boldsymbol{e}_k \\ &= \frac{\mathbb{M}_{1111} + 2\mathbb{M}_{1122}}{3} \quad \text{for all } k; \end{aligned} \qquad (53)$$

then, even for these cubic classes condition (47) is verified for the same value of α as in the isotropic case and the Brewster law can be generalized to the form (25) with $(\Delta n)_k^o = 0$, $k = 1, 2, 3$.

By (39) and (44) the components of $[f_{ik}]$ depend not only on the components of $\mathbb{M}$ but also on both $\boldsymbol{R}$ and $\widehat{\boldsymbol{R}}$: however, if we assume that the frame Σ is also the principal frame for $\boldsymbol{T}$, that is $\boldsymbol{R}=\boldsymbol{I}$, then by (76) we obtain again relation (48); accordingly the results for isotropic material still apply with the components of $[f_{ik}]$ given by (52) and (24), which holds even in this case.

3.1.3. *Cubic crystals, classes* 23 *and* $3m$. The tensor $\mathbb{M}$ for these classes is not symmetric and hence we cannot apply the results of [Mehrabadi and Cowin 1990]. However, since

$$\mathbb{M}[\boldsymbol{I}]=(\mathbb{M}_{1111}+\mathbb{M}_{1122}+\mathbb{M}_{1133})\boldsymbol{e}_1\otimes\boldsymbol{e}_1+(\mathbb{M}_{1111}+\mathbb{M}_{1122}+\mathbb{M}_{2211})\boldsymbol{e}_2\otimes\boldsymbol{e}_2 + (\mathbb{M}_{1111}+\mathbb{M}_{2211}+\mathbb{M}_{3311})\boldsymbol{e}_3\otimes\boldsymbol{e}_3, \quad (54)$$

then condition (47) cannot be verified for any $\widehat{\boldsymbol{R}}$ for a unique value of α. Therefore for cubic crystal of these classes it is not possible, in general, to arrive at the Brewster-type relation (25).

3.2. *Optically uniaxial.* For optically uniaxial crystals, the inverse permittivity has the representation (14). We set $\boldsymbol{e}_3=\boldsymbol{e}$ and therefore,

$$(\Delta n)^o_1=-(\Delta n)^o_2=n_o-n_e, \qquad (\Delta n)^o_3=0, \quad (55)$$

whereas condition (47) requires

$$M_{11}=M_{22}=\alpha n_o^2, \qquad M_{33}=\alpha n_e^2. \quad (56)$$

Relation (56) may be verified for some specific crystals but, in the general case, cannot be satisfied and we may conclude that for uniaxial crystal the generalization (25) of the Brewster law is not possible, unless the stress is a purely deviatoric one.

Whereas it is still possible to arrive at (41), we may restrict our analysis to plane stress in the plane orthogonal to the optic axis, say $\boldsymbol{T}\boldsymbol{e}_3=\boldsymbol{0}$. In this case $\boldsymbol{e}_3=\boldsymbol{w}_3$ with $\sigma_3=0$ and $\boldsymbol{R}$ is a rotation about $\boldsymbol{e}_3$.

Further, if we choose α which satisfies (56), then we may write $(41)_3$ as

$$n_1-n_2=f_{13}\sigma_2-f_{23}\sigma_1+f_{33}(\sigma_1-\sigma_2), \quad (57)$$

which can be rewritten in the Brewster-like form

$$n_1-n_2=(f_{33}-f_{13})(\sigma_1-\sigma_2), \quad (58)$$

provided the following condition holds:

$$f_{13}=f_{23}. \quad (59)$$

From (44), the requirement that (59) be verified for any refraction indices n_o and n_e is equivalent to the two conditions

$$\overline{\mathbb{M}}_{1133}=\overline{\mathbb{M}}_{2233}, \qquad \overline{\mathbb{M}}_{1111}+\overline{\mathbb{M}}_{1122}=\overline{\mathbb{M}}_{2222}+\overline{\mathbb{M}}_{2211}. \quad (60)$$

Since $\boldsymbol{e}_3$ is an axis of symmetry for trigonal, tetragonal, and hexagonal crystals, the rotations about $\boldsymbol{e}_3$ belong to their symmetry group $\mathscr{G}$ and we can write (39) as

$$\overline{\mathbb{M}}_{hhkk} = \tfrac{1}{3}\boldsymbol{Q}\mathbb{M}[\boldsymbol{e}_h \otimes \boldsymbol{e}_h]\boldsymbol{Q}^T \cdot \boldsymbol{e}_k \otimes \boldsymbol{e}_k = \tfrac{1}{3}\mathbb{M}_{ijhh}(\boldsymbol{Q}\boldsymbol{e}_i \cdot \boldsymbol{e}_k)(\boldsymbol{Q}\boldsymbol{e}_j \cdot \boldsymbol{e}_k), \quad h, k \text{ fixed}, \tag{61}$$

where $\boldsymbol{Q} = \widehat{\boldsymbol{R}}^T \boldsymbol{R} \in \mathrm{Orth}^+$.

Under the additional hypothesis that also $\widehat{\boldsymbol{R}}$ is a rotation about the material symmetry axis $\boldsymbol{e}_3$,[3] then we have that $\boldsymbol{Q}\boldsymbol{e}_3 = \boldsymbol{e}_3$ and $\boldsymbol{Q}\boldsymbol{e}_\alpha \cdot \boldsymbol{e}_3 = 0$, $\alpha = 1, 2$. Therefore from (61) we can write

$$\begin{aligned}
\overline{\mathbb{M}}_{1111} + \overline{\mathbb{M}}_{1122} &= \tfrac{1}{3}((\mathbb{M}_{1111} + \mathbb{M}_{1122})\cos^2\varphi + (\mathbb{M}_{2222} + \mathbb{M}_{2211})\sin^2\varphi \\
&\qquad - (\mathbb{M}_{1211} + \mathbb{M}_{1222} + \mathbb{M}_{2111} + \mathbb{M}_{2122})\sin\varphi\cos\varphi), \\
\overline{\mathbb{M}}_{2222} + \overline{\mathbb{M}}_{2211} &= \tfrac{1}{3}((\mathbb{M}_{1111} + \mathbb{M}_{1122})\sin^2\varphi + (\mathbb{M}_{2222} + \mathbb{M}_{2211})\cos^2\varphi \\
&\qquad + (\mathbb{M}_{1211} + \mathbb{M}_{1222} + \mathbb{M}_{2111} + \mathbb{M}_{2122})\sin\varphi\cos\varphi), \\
\overline{\mathbb{M}}_{1133} &= \tfrac{1}{3}\mathbb{M}_{3311}, \\
\overline{\mathbb{M}}_{2233} &= \tfrac{1}{3}\mathbb{M}_{3322},
\end{aligned} \tag{62}$$

where φ is the rotation angle about $\boldsymbol{e}_3$. Accordingly condition $(60)_1$ together with the requirement that $(60)_2$ holds for any value of φ leads to the following restrictions on the symmetries of $\mathbb{M}$:

$$\begin{aligned}
\mathbb{M}_{3311} &= \mathbb{M}_{3322}, \\
\mathbb{M}_{1111} + \mathbb{M}_{1122} &= \mathbb{M}_{2222} + \mathbb{M}_{2211}, \\
\mathbb{M}_{1211} + \mathbb{M}_{1222} + \mathbb{M}_{2111} + \mathbb{M}_{2122} &= 0.
\end{aligned} \tag{63}$$

From the tabular data in the Appendix it easy to show that these conditions hold only for the high-symmetry classes of trigonal ($\bar{3}m$, 32, $3m$), tetragonal ($4mmm$, 422, $4/mm$, and $\bar{4}2m$) and hexagonal ($\bar{6}m2$, 622, $6mm$, and $6/mmm$) crystals, in which cases (58) reduces once again to (24).

3.3. ***Optically biaxial.*** In the case of optically biaxial crystals, the inverse permittivity has either the representation (21) or (22). In any case it is $(\Delta n)_k^o \neq 0$, $k = 1, 2, 3$, and (47) leads to

$$M_{11} = \alpha n_1^2, \qquad M_{22} = \alpha n_2^2, \qquad M_{33} = \alpha n_2^2, \tag{64}$$

and it is not possible to obtain the relation (25), unless (64) is verified for some specific crystals. However, it is possible to arrive at the relation (41) which on the other hand didn't simplify the expression for birefringence to an appreciable degree. Accordingly we didn't go into further details.

[3]This assumption means that the stressed crystal remains uniaxial: a complete characterization of the stresses which leave uniaxial crystal still uniaxial is provided in [Davì 2015].

4. Conclusions

We looked at the possibility of generalizing the Brewster law for isotropic transparent materials to anisotropic crystalline materials. First we show that the birefringence can be written as a linear function of the spherical and shear stresses expressed as the difference in the principal stress, by taking into account the components of the unstressed inverse permittivity, those of the piezooptic tensor and the rotation of the stress, and the optical principal directions with respect to the symmetry axis of the material; then we show that this result can be formulated in a "Brewster-like" manner only when the birefringence is independent of the spherical part of the stress, a fact which is related to the crystal symmetry. We showed that the relation can be generalized for any stress only for the high-symmetry classes of the cubic group and that when the stress tensor is diagonal we obtain once more the isotropic relation.

For uniaxial crystals the relation can be generalized only for plane stress in the plane orthogonal to the optic axis and provided the stressed crystal remains uniaxial: also in this case we arrive at the isotropic-like relation in the plane of stress. We show that no general extension of the Brewster law is possible, in the general case, for biaxial crystals.

Appendix

In order to make the paper self-contained, in this appendix we list the tabular form of the piezooptic tensor $\mathbb{M}$ of the various crystallographic classes, from [Authier 2003].

Monoclinic. All classes:

$$[\mathbb{M}] \equiv \begin{bmatrix} \mathbb{M}_{1111} & \mathbb{M}_{1122} & \mathbb{M}_{1133} & 0 & \mathbb{M}_{1113} & 0 \\ \mathbb{M}_{2211} & \mathbb{M}_{2222} & \mathbb{M}_{2233} & 0 & \mathbb{M}_{2213} & 0 \\ \mathbb{M}_{3311} & \mathbb{M}_{3322} & \mathbb{M}_{3333} & 0 & \mathbb{M}_{3313} & 0 \\ 0 & 0 & 0 & \mathbb{M}_{2323} & 0 & \mathbb{M}_{2312} \\ \mathbb{M}_{1311} & \mathbb{M}_{1312} & \mathbb{M}_{1333} & 0 & \mathbb{M}_{1313} & 0 \\ 0 & 0 & 0 & \mathbb{M}_{1223} & 0 & \mathbb{M}_{1212} \end{bmatrix}. \tag{65}$$

Orthorhombic. All classes:

$$[\mathbb{M}] \equiv \begin{bmatrix} \mathbb{M}_{1111} & \mathbb{M}_{1122} & \mathbb{M}_{1133} & 0 & 0 & 0 \\ \mathbb{M}_{2211} & \mathbb{M}_{2222} & \mathbb{M}_{2233} & 0 & 0 & 0 \\ \mathbb{M}_{3311} & \mathbb{M}_{3322} & \mathbb{M}_{3333} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbb{M}_{2323} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbb{M}_{1313} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbb{M}_{1212} \end{bmatrix}. \tag{66}$$

Trigonal. Classes $\bar{3}m$, 32, and $3m$:

$$[\mathbb{M}] \equiv \begin{bmatrix} \mathbb{M}_{1111} & \mathbb{M}_{1122} & \mathbb{M}_{1133} & \mathbb{M}_{1123} & 0 & 0 \\ \mathbb{M}_{1122} & \mathbb{M}_{1111} & \mathbb{M}_{1133} & -\mathbb{M}_{1123} & 0 & 0 \\ \mathbb{M}_{3311} & \mathbb{M}_{3311} & \mathbb{M}_{3333} & 0 & 0 & 0 \\ \mathbb{M}_{2311} & -\mathbb{M}_{2311} & 0 & \mathbb{M}_{2323} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbb{M}_{2323} & 2\mathbb{M}_{2311} \\ 0 & 0 & 0 & 0 & \mathbb{M}_{1123} & \mathbb{M}_{1111} - \mathbb{M}_{1122} \end{bmatrix}. \tag{67}$$

Classes 3 and $\bar{3}$:

$$[\mathbb{M}] \equiv \begin{bmatrix} \mathbb{M}_{1111} & \mathbb{M}_{1122} & \mathbb{M}_{1133} & \mathbb{M}_{1123} & -\mathbb{M}_{2213} & 2\mathbb{M}_{1222} \\ \mathbb{M}_{1122} & \mathbb{M}_{1111} & \mathbb{M}_{1133} & -\mathbb{M}_{1123} & \mathbb{M}_{2213} & -2\mathbb{M}_{1222} \\ \mathbb{M}_{3311} & \mathbb{M}_{3311} & \mathbb{M}_{3333} & 0 & 0 & 0 \\ \mathbb{M}_{2311} & -\mathbb{M}_{2311} & 0 & \mathbb{M}_{2323} & \mathbb{M}_{2313} & 2\mathbb{M}_{1322} \\ -\mathbb{M}_{1322} & \mathbb{M}_{1322} & 0 & -\mathbb{M}_{2313} & \mathbb{M}_{2323} & 2\mathbb{M}_{2311} \\ -\mathbb{M}_{1222} & \mathbb{M}_{1222} & 0 & \mathbb{M}_{2213} & \mathbb{M}_{1123} & \mathbb{M}_{1111} - \mathbb{M}_{1122} \end{bmatrix}. \tag{68}$$

Tetragonal. Classes 4, $\bar{4}$, and $4/m$:

$$[\mathbb{M}] \equiv \begin{bmatrix} \mathbb{M}_{1111} & \mathbb{M}_{1122} & \mathbb{M}_{1133} & 0 & 0 & \mathbb{M}_{1112} \\ \mathbb{M}_{1122} & \mathbb{M}_{1111} & \mathbb{M}_{1133} & 0 & 0 & -\mathbb{M}_{1112} \\ \mathbb{M}_{3311} & \mathbb{M}_{3311} & \mathbb{M}_{3333} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbb{M}_{2323} & \mathbb{M}_{2313} & 0 \\ 0 & 0 & 0 & -\mathbb{M}_{2313} & \mathbb{M}_{2323} & 0 \\ \mathbb{M}_{1211} & -\mathbb{M}_{1211} & 0 & 0 & 0 & \mathbb{M}_{1212} \end{bmatrix}. \tag{69}$$

Classes $4mmm$, 422, $4/mm$, and $\bar{4}2m$:

$$[\mathbb{M}] \equiv \begin{bmatrix} \mathbb{M}_{1111} & \mathbb{M}_{1122} & \mathbb{M}_{1133} & 0 & 0 & 0 \\ \mathbb{M}_{1122} & \mathbb{M}_{1111} & \mathbb{M}_{1133} & 0 & 0 & 0 \\ \mathbb{M}_{3311} & \mathbb{M}_{3311} & \mathbb{M}_{3333} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbb{M}_{2323} & & 0 \\ 0 & 0 & 0 & 0 & \mathbb{M}_{2323} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbb{M}_{1212} \end{bmatrix}. \tag{70}$$

Hexagonal. Classes 6, $\bar{6}$, and $6/m$:

$$[\mathbb{M}] \equiv \begin{bmatrix} \mathbb{M}_{1111} & \mathbb{M}_{1122} & \mathbb{M}_{1133} & 0 & 0 & 2\mathbb{M}_{1222} \\ \mathbb{M}_{1122} & \mathbb{M}_{1111} & \mathbb{M}_{1133} & 0 & 0 & -2\mathbb{M}_{1222} \\ \mathbb{M}_{3311} & \mathbb{M}_{3311} & \mathbb{M}_{3333} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbb{M}_{2323} & \mathbb{M}_{2313} & 0 \\ 0 & 0 & 0 & -\mathbb{M}_{2313} & \mathbb{M}_{2323} & 0 \\ -2\mathbb{M}_{122} & 2\mathbb{M}_{1222} & 0 & 0 & 0 & \mathbb{M}_{1111} - \mathbb{M}_{1122} \end{bmatrix}. \tag{71}$$

Classes $\bar{6}m2$, 622, $6mm$, and $6/mmm$:

$$[\mathbb{M}] \equiv \begin{bmatrix} \mathbb{M}_{1111} & \mathbb{M}_{1122} & \mathbb{M}_{1133} & 0 & 0 & 0 \\ \mathbb{M}_{1122} & \mathbb{M}_{1111} & \mathbb{M}_{1133} & 0 & 0 & 0 \\ \mathbb{M}_{3311} & \mathbb{M}_{3311} & \mathbb{M}_{3333} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbb{M}_{2323} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbb{M}_{2323} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbb{M}_{1111}-\mathbb{M}_{1122} \end{bmatrix}. \tag{72}$$

Cubic. Classes 23 and $3m$:

$$[\mathbb{M}] \equiv \begin{bmatrix} \mathbb{M}_{1111} & \mathbb{M}_{1122} & \mathbb{M}_{1133} & 0 & 0 & 0 \\ \mathbb{M}_{2211} & \mathbb{M}_{1111} & \mathbb{M}_{1122} & 0 & 0 & 0 \\ \mathbb{M}_{3311} & \mathbb{M}_{2211} & \mathbb{M}_{1111} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbb{M}_{1212} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbb{M}_{1212} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbb{M}_{1212} \end{bmatrix}. \tag{73}$$

Classes 432, $\bar{4}3m$, and $m3m$:

$$[\mathbb{M}] \equiv \begin{bmatrix} \mathbb{M}_{1111} & \mathbb{M}_{1122} & \mathbb{M}_{1122} & 0 & 0 & 0 \\ \mathbb{M}_{1122} & \mathbb{M}_{1111} & \mathbb{M}_{1122} & 0 & 0 & 0 \\ \mathbb{M}_{1122} & \mathbb{M}_{1122} & \mathbb{M}_{111} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbb{M}_{1212} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbb{M}_{1212} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbb{M}_{1212} \end{bmatrix}. \tag{74}$$

Isotropic.

$$[\mathbb{M}] \equiv \begin{bmatrix} \mathbb{M}_{1111} & \mathbb{M}_{1122} & \mathbb{M}_{1122} & 0 & 0 & 0 \\ \mathbb{M}_{1122} & \mathbb{M}_{1111} & \mathbb{M}_{1122} & 0 & 0 & 0 \\ \mathbb{M}_{1122} & \mathbb{M}_{1122} & \mathbb{M}_{1111} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbb{M}_{1212} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbb{M}_{1212} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbb{M}_{1212} \end{bmatrix}, \tag{75}$$

with $2\mathbb{M}_{1212} = \mathbb{M}_{1111} - \mathbb{M}_{1122}$.

We remark that, for isotropic materials and cubic crystals of classes 432, $\bar{4}3m$, and $m3m$, the piezooptic tensor is symmetric, i.e., $\mathbb{M}_{ijhk} = \mathbb{M}_{hkij}$ and accordingly the results obtained in [Mehrabadi and Cowin 1990] for the eigenvalues and eigentensor of the symmetric elasticity tensor still apply.

In particular we shall make use of the two following results from [Mehrabadi and Cowin 1990].

- Cubic crystals, classes 432, $\bar{4}3m$, and $m3m$:

$$\mathbb{M} = (\mathbb{M}_{1111} + 2\mathbb{M}_{1122})\tfrac{1}{3}\boldsymbol{I} \otimes \boldsymbol{I} + (\mathbb{M}_{1111} - \mathbb{M}_{1122})\left(\sum_{k=1}^{3} \boldsymbol{E}_k \otimes \boldsymbol{E}_k - \tfrac{1}{3}\boldsymbol{I} \otimes \boldsymbol{I}\right) + 2\mathbb{M}_{1212}\sum_{k=4}^{6} \boldsymbol{E}_k \otimes \boldsymbol{E}_k, \quad (76)$$

where

$$\begin{aligned} \boldsymbol{E}_k &= \frac{1}{\sqrt{2}}\boldsymbol{e}_k \otimes \boldsymbol{e}_k, \quad k = 1, 2, 3, \\ \boldsymbol{E}_4 &= \tfrac{1}{2}(\boldsymbol{e}_2 \otimes \boldsymbol{e}_3 + \boldsymbol{e}_3 \otimes \boldsymbol{e}_2), \\ \boldsymbol{E}_5 &= \tfrac{1}{2}(\boldsymbol{e}_1 \otimes \boldsymbol{e}_3 + \boldsymbol{e}_3 \otimes \boldsymbol{e}_1), \\ \boldsymbol{E}_6 &= \tfrac{1}{2}(\boldsymbol{e}_2 \otimes \boldsymbol{e}_1 + \boldsymbol{e}_1 \otimes \boldsymbol{e}_2). \end{aligned} \quad (77)$$

- Isotropic materials:

$$\mathbb{M} = (\mathbb{M}_{1111} + 2\mathbb{M}_{1122})\tfrac{1}{3}\boldsymbol{I} \otimes \boldsymbol{I} + (\mathbb{M}_{1111} - \mathbb{M}_{1122})\mathbb{D}, \quad (78)$$

where $\mathbb{D} : \mathrm{Sym} \to \mathrm{Dev}$ is given by

$$\mathbb{D} = \sum_{k=1}^{6} \boldsymbol{E}_k \otimes \boldsymbol{E}_k - \tfrac{1}{3}\boldsymbol{I} \otimes \boldsymbol{I}, \quad (79)$$

i.e., $\mathbb{D}[\boldsymbol{T}] = \operatorname{dev} \boldsymbol{T}$.

Acknowledgments

The research leading to these results is within the scope of CERN R&D Experiment 18 "Crystal Clear Collaboration" and the PANDA Collaboration at GSI-Darmstadt. It has received funding from the Progetto Strategico di Ateneo 2017: "Scintillating crystals: an interdisciplinary, applications-oriented approach aimed to the scientific knowledge and process control for application concerning life-quality improvement".

References

[Aben and Guillemet 1993] H. Aben and C. Guillemet, *Photoelasticity of glass*, Springer, 1993.

[Authier 2003] A. Authier (editor), *International tables for crystallography*, vol. D: Physical properties of crystals, Springer, 2003.

[Bain 2019] A. K. Bain, *Crystal optics: properties and applications*, Wiley, Weinheim, Germany, 2019.

[Born and Wolf 1999] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*, 7th ed., Cambridge University, 1999.

[Brewster 1830] D. Brewster, *Papers on optics*, Royal Society of London, 1830.

[Davì 2015] F. Davì, "On the Bertin surfaces of photoelastic crystals", *J. Opt. Soc. Am. A* **32**:12 (2015), 2323–2337.

[Kuske and Robertson 1974] A. Kuske and C. Robertson, *Photoelastic stress analysis*, Wiley, London, 1974.

[Mehrabadi and Cowin 1990] M. M. Mehrabadi and S. C. Cowin, "Eigentensors of linear anisotropic elastic materials", *Quart. J. Mech. Appl. Math.* **43**:1 (1990), 15–41.

[Montalto et al. 2019] L. Montalto, P. P. Natali, L. Scalise, N. Paone, F. Davì, D. Rinaldi, G. Barucca, and P. Mengucci, "Quality control and structural assessment of anisotropic scintillating crystals", *Crystals* **9**:7 (2019), art. id. 376.

[Nye 1985] J. F. Nye, *Physical properties of crystals: their representation by tensors and matrices*, Oxford University, 1985.

[Perelomova and Tagieva 1983] N. V. Perelomova and M. M. Tagieva, *Problems in crystal physics: with solutions*, Mir, Moscow, 1983.

[Rinaldi et al. 2018] D. Rinaldi, F. Davì, and L. Montalto, "On the photoelastic constants and the Brewster law for stressed tetragonal crystals", *Math. Methods Appl. Sci.* **41**:8 (2018), 3103–3116.

[Sirotin and Shaskolskaya 1982] Y. I. Sirotin and M. P. Shaskolskaya, *Fundamentals of crystal physics*, Mir, Moscow, 1982.

[Sokolnikoff 1956] I. S. Sokolnikoff, *Mathematical theory of elasticity*, 2nd ed., McGraw-Hill, New York, 1956.

Received 30 Jul 2019. Revised 4 Aug 2019. Accepted 18 Jan 2020.

FABRIZIO DAVÌ: davi@univpm.it
Dipartimento di Ingegneria Civile, Edile e Architettura, Centro Interdipartimentale di Analisi Globale dei Cristalli, Università Politecnica delle Marche, Ancona, Italy

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
Vol. 8, No. 1, 2020
dx.doi.org/10.2140/memocs.2020.8.47

IBVP FOR ELECTROMAGNETO-ELASTIC MATERIALS: VARIATIONAL APPROACH

AMIRHOSSEIN AMIRI-HEZAVEH, POUYAN KARIMI
AND MARTIN OSTOJA-STARZEWSKI

This paper aims to establish a variational framework for materials having coupling interactions between electromagnetic and mechanical fields. Based on coupled constitutive equations and the alternative field equations, a general variational form, imposing no restriction on the fields involved, is given. Subsequently, the result is derived for the case when satisfaction of the strain-displacement equation is presumed as a restriction. Next, the variational forms for kinematically admissible processes and, in turn, for kinematically admissible displacement-potential processes are found. Finally, the principles characterizing the stress field instead of the displacement field are formulated. The results of the present work provide a framework in which the satisfaction of initial boundary conditions is inherently considered. The proposed framework furnishes an alternative path for the implementation of numerical approaches for PDEs governing the motion of electromagneto-elastic materials.

1. Introduction

Electromagneto-elastic materials, a category of materials that contains both piezoelectric and piezomagnetic phases, are being widely used in several devices including ultrasonic transducers and microactuators, thermal-imaging devices, health-monitoring devices, biomedical devices, biomimetics, and energy harvesting [Li 2003; Miehe et al. 2011]. Also, these materials have found applications in microwave electronic and optoelectronic instruments because of their flat frequency response as well as the capability of energy conversion [Li and Kardomateas 2006]. Consequently, to mathematically understand the physics of such materials, several studies have been carried out by employing a continuum approach in which classical laws are generalized to account for the coupling between mechanical, electric, and magnetic fields. Some of the most prominent contributions in this regard can be found in [Guggenheim 1936a; 1936b; Penfield et al. 1963; Brown 1966; Coleman

Communicated by David J. Steigmann.
MSC2010: 35A15, 74BXX, 74FXX.
PACS2010: 46.15.Cc, 75.80.+q, 77.65.-j, 83.10.Ff.
Keywords: variational principles, elastodynamics, electromagneto-elastic materials.

and Dill 1971; Tiersten and Tsai 1972; Mindlin 1972; Nelson 1979; Maugin 1988; Eringen and Maugin 1990; Landau et al. 1984].

In the linearized classical isothermal continuum mechanics, the governing equations to model a physical phenomenon are the balance of mass and balances of linear and angular momenta along with desired constitutive equations for the phenomenon at hand. These, accompanied by the initial and boundary conditions, often lead to a mixed initial-boundary value problem (IBVP) in terms of the displacement field. However, [Ignaczak 1959; 1963] proposed a robust alternative approach in which the governing equations are formulated in terms of stress. This formulation motivated several researchers to assess the potential of this strategy, which offers a much more straightforward framework when the boundary conditions are of Neumann type. For a comprehensive review of those works, the readers are referred to the review paper [Ostoja-Starzewski 2019]. Among them, the most pioneering works are convolutional variational principles, i.e., variational principles containing convolution products with respect to time, developed in [Gurtin 1963; 1964]. In these works, the framework has been rationally developed to be applicable for mixed initial-boundary value problems, leading to integro-partial-differential equations and the corresponding convolutional variational principles. From there on, [Nickell and Sackman 1968] generalized Gurtin's work to thermoelasticity and, subsequently, a specific form of such formulation has been obtained for piezoelectric materials in [Oden and Reddy 1983]. Recently, one can note the results given in [El-Karamany and Ezzat 2011] for two-temperature thermoelasticity.

Owing to the fact that the analytical methods are only sufficient tools for problems with simple geometry and rather strict assumptions, variational principles are of great importance in engineering sciences as they pave the way for developing numerical approaches to solve PDEs with either more relaxed assumptions or arbitrary/complicated boundary conditions. The finite element, mesh-free particle, and Ritz methods are examples stemming from variational principles. As an alternative application of such principles, we note the homogenization theory which supplies bounds for properties of materials (e.g., [Hashin and Shtrikman 1962]). In the case of solid mechanics, however, the classical variants of seminal work of [Washizu 1957] are not applicable to the case of a mixed problem of elastodynamics since the prescribed initial velocity is not realized and the knowledge of displacement field at a later time is only presupposed [Gurtin 1964]. Therefore, starting with [Gurtin 1964], as an alternative approach appropriate to elastodynamics, convolutional variational principles have been developed. The elegance of the approach consists of imposing the initial conditions implicitly in the form of a body force and thus assuring appropriate satisfaction of them.

Concerning the variational principles for electromagneto-elastic materials, several studies have been carried out dating back to [Toupin 1956].Variational principles for various problems including piezoelectric ceramics have been proposed in [He 2000; 2001a; 2001b; 2001c]. Convolutional regionwise variational principles for thermopiezoelectric media can be found in [Bo 2003]. Also, for the nonlinear case, some studies have been done, e.g., by making use of the first law of thermodynamics in [Kuang 2008]. Convolutional variational principles have then been proposed for the case of nonlinear electromagneto-elastic solids in [Wang et al. 2010]. Recently, on the basis of incremental variational principles, a general framework for functional dissipative materials has been obtained in [Miehe et al. 2011] and employed in [Miehe and Rosato 2011] to analyze piezoelectric ceramics.

To the best of the authors' knowledge, a comprehensive and systematic generalization of the results initially developed in [Gurtin 1964] for the case of linear electromagneto-elastic materials is not available in the literature and this challenge defines the focus of the present study. Indeed, it is of interest to enrich the numerical framework relevant to the analysis of electromagneto-elastic materials because of the progressive increase in the application of such materials in the realm of structural mechanics; see, for example, a recent review [Irschik et al. 2010]. As an example of the recent development in the use of smart materials, one can mention the recent paper [Schoeftner and Irschik 2016] in which the design of piezoelectric devices controlling the level of stress in thin bars has been discussed. The methodology to form and prove the results obtained in this study, similar to the presentation given in [Nickell and Sackman 1968], is based on [Gurtin 1964]. For the sake of completeness of the presentation, we collect in Appendix A the mathematical concepts and lemmas originally proved in [Gurtin 1964] along with a corollary obtained in [Nickell and Sackman 1968]. Alternative field equations, the main ground for establishing the corresponding convolutional variational principles, for electromagneto-elastic materials are described in Appendix B. Field equations, along with some continuity conditions, are given in detail in Section 2. Subsequently, analogously to [Gurtin 1964], the convolutional variational forms of the alternative integro-partial-differential field equations are obtained and proved comprehensively in Section 3. As mentioned earlier, the derivations in the main body of the present study can be useful in the sense of analysis and design of electromagneto-elastic materials in both practice and research.

2. Problem statement

In this section, the field equations for an electromagneto-elastic material are listed. Throughout the paper we indicate the position vector and time parameter, respectively, by $\boldsymbol{x}$ and t. Also, the standard index notation for Cartesian tensors is used. The mathematical notation used herein along with some lemmas and theorems that are the primary tools to obtain the results of this paper can be found in Appendix A.

Let $\overline{V}$ denote a closed and bounded subset of 3D Euclidean space, with interior V and boundary ∂V occupied by a deformable electromagneto-elastic material. Furthermore, assume that V is a regular region in the sense of [Gurtin 1964]. Let $u_i(\boldsymbol{x}, t)$, $\sigma_{ij}(\boldsymbol{x}, t)$, $\varepsilon_{ij}(\boldsymbol{x}, t)$, $f_i(\boldsymbol{x}, t)$, $E_i(\boldsymbol{x}, t)$, $D_i(\boldsymbol{x}, t)$, $B_i(\boldsymbol{x}, t)$, and $H_i(\boldsymbol{x}, t)$ with $(\boldsymbol{x}, t) \in \overline{V} \times [0, \infty)$, respectively, denote the components of the displacement vector, stress tensor, strain tensor, body force, electric field, electric displacement field, magnetic field, and magnetic field strength. In addition, scalar fields $\rho(\boldsymbol{x}, t)$, $q_e(\boldsymbol{x}, t)$, $\varphi(\boldsymbol{x}, t)$, and $\psi(\boldsymbol{x}, t)$ denote the mass density, charge density, electric potential, and magnetic potential, respectively.

Let ∂V_i denote a subset of ∂V over which $i = \boldsymbol{u}, \boldsymbol{\sigma}, \varphi, \boldsymbol{D}, \psi, \boldsymbol{B}$ is prescribed with the condition

$$\begin{aligned} \partial V_{\boldsymbol{u}} \cup \partial V_{\boldsymbol{\sigma}} &= \partial V, & \partial V_{\boldsymbol{u}} \cap \partial V_{\boldsymbol{\sigma}} &= \varnothing, \\ \partial V_{\varphi} \cup \partial V_{\boldsymbol{D}} &= \partial V, & \partial V_{\varphi} \cap \partial V_{\boldsymbol{D}} &= \varnothing, \\ \partial V_{\psi} \cup \partial V_{\boldsymbol{B}} &= \partial V, & \partial V_{\psi} \cap \partial V_{\boldsymbol{B}} &= \varnothing. \end{aligned} \tag{2-1}$$

Moreover, the symbol $\overline{\partial V_i}$ with $i = \boldsymbol{u}, \boldsymbol{\sigma}, \varphi, \boldsymbol{D}, \psi, \boldsymbol{B}$ stands for the closure of the aforementioned sets. Furthermore, the quasistatic electromagnetic condition is presumed; that is, it is assumed that the electric and magnetic fields are both curl free. This approximation leads to accurate results for instance, as a particular case, in analysis of nonmagnetizable elastic dielectrics when the wavelengths of mechanical waves are negligible if compared to wavelengths of electromagnetic waves of the same frequency [Eringen and Maugin 1990]. Accordingly, the governing equations read [Li 2003]

$$\begin{aligned} \rho \ddot{u}_i &= \sigma_{ij,j} + F_i && \text{on } V \times (0, \infty), \\ D_{i,i} &= q_e && \text{on } V \times (0, \infty), \\ B_{i,i} &= 0 && \text{on } V \times (0, \infty), \end{aligned} \tag{2-2}$$

in which $\sigma_{ij} = \sigma_{ji}$. Also, kinematic equations are

$$\begin{aligned} \epsilon_{ij} &= u_{(i,j)} = \tfrac{1}{2}(u_{i,j} + u_{j,i}) && \text{on } V \times (0, \infty), \\ E_i &= -\varphi_{,i} && \text{on } V \times (0, \infty), \\ H_i &= -\psi_{,i} && \text{on } V \times (0, \infty), \end{aligned} \tag{2-3}$$

in which $u_{(i,j)}$ denotes the symmetric part of the second-order tensor $u_{i,j}$.

Next, the constitutive equations need to be set. To that end, one needs to define which of the physical quantities are dependent variables and which are independent ones. Thus, one can define various forms of constitutive equations based on different independent variables. In general, in a nonlinear theory, it is a difficult task to obtain one form of the constitutive equations from the other. Nevertheless, in linear theory, the necessary Legendre transformations are easy to manipulate and readily obtain the desired variants of constitutive equations [Pérez-Fernández et al. 2009]. Assuming the isothermal condition and hyperelasticity, one can obtain the relations

$$\begin{aligned} \sigma_{ij} &= C_{ijkl}\varepsilon_{kl} - e^E_{kij}E_k - e^H_{kij}H_k && \text{on } V \times (0, \infty), \\ D_i &= e^E_{ikl}\varepsilon_{kl} + \kappa^E_{ij}E_j + \kappa^{EH}_{ij}H_j && \text{on } V \times (0, \infty), \\ B_i &= e^H_{ikl}\varepsilon_{kl} + \kappa^{EH}_{ji}E_j + \kappa^H_{ij}H_j && \text{on } V \times (0, \infty), \end{aligned} \tag{2-4}$$

with symmetry conditions

$$\begin{aligned} C_{ijkl} = C_{klij} = C_{jikl} &= C_{ijlk} && \text{on } V, \\ e^E_{kij} = e^E_{kji}, \qquad e^H_{kij} &= e^H_{kji} && \text{on } V, \\ \kappa^E_{ij} = \kappa^E_{ji}, \qquad \kappa^H_{ij} &= \kappa^H_{ji} && \text{on } V. \end{aligned} \tag{2-5}$$

Applying the Legendre transformation, one can find

$$\begin{aligned} \varepsilon_{ij} &= S_{ijkl}\sigma_{kl} + d^E_{kij}E_k + d^H_{kij}H_k && \text{on } V \times (0, \infty), \\ D_i &= d^E_{ikl}\sigma_{kl} + \chi^E_{ij}E_j + \chi^{EH}_{ij}H_j && \text{on } V \times (0, \infty), \\ B_i &= d^H_{ikl}\sigma_{kl} + \chi^{EH}_{ji}E_j + \chi^H_{ij}H_j && \text{on } V \times (0, \infty), \end{aligned} \tag{2-6}$$

with symmetry conditions

$$\begin{aligned} S_{ijkl} = S_{klij} = S_{jikl} &= S_{ijlk} && \text{on } V, \\ d^E_{kij} = d^E_{kji}, \qquad d^H_{kij} &= d^H_{kji} && \text{on } V, \\ \chi^E_{ij} = \chi^E_{ji}, \qquad \chi^H_{ij} &= \chi^H_{ji} && \text{on } V. \end{aligned} \tag{2-7}$$

In (2-4) and (2-6) the coefficients S_{ijkl}, C_{ijkl}, d^E_{kij}, d^H_{kij}, e^E_{kij}, e^H_{kij}, χ^E_{ij}, χ^H_{ij}, κ^E_{ij}, κ^H_{ij}, χ^{EH}_{ij}, and κ^{EH}_{ij}, all functions of position, represent, respectively, components of the compliance tensor, stiffness tensor, direct piezoelectric tensor, direct piezomagnetic tensor, reverse piezoelectric tensor, reverse piezomagnetic tensor, permittivity under constant stress, permeability under constant stress, permittivity under constant strain, permeability under constant strain, magnetoelectric tensor under constant stress, and magnetoelectric tensor under constant strain. The aforementioned variables are related as

$$\begin{aligned} S_{ijkl}C_{ijpq} &= \delta_{kp}\delta_{lq} && \text{on } V, \\ d^E_{kij} &= S_{pqij}e^E_{kpq} && \text{on } V, \\ d^H_{kij} &= S_{pqij}e^H_{kpq} && \text{on } V, \\ \chi^E_{ij} &= S_{pqrs}e^E_{ipq}e^E_{jrs} + \kappa^E_{ij} && \text{on } V, \\ \chi^H_{ij} &= S_{pqrs}e^H_{ipq}e^H_{jrs} + \kappa^H_{ij} && \text{on } V, \\ \chi^{EH}_{ij} &= S_{pqrs}e^H_{jpq}e^E_{irs} + \kappa^{EH}_{ij} && \text{on } V. \end{aligned} \tag{2-8}$$

To form the IBVP, define the boundary conditions

$$\begin{aligned} &\text{mechanical boundary conditions} \begin{cases} u_i = \hat{u}_i(\boldsymbol{x}, t) & \text{on } \partial V_{\boldsymbol{u}} \times [0, \infty), \\ t_i = \hat{t}_i(\boldsymbol{x}, t) & \text{on } \partial V_{\boldsymbol{\sigma}} \times [0, \infty), \end{cases} \\ &\text{electric boundary conditions} \begin{cases} \varphi = \hat{\varphi}(\boldsymbol{x}, t) & \text{on } \partial V_{\varphi} \times [0, \infty), \\ d = \hat{d}(\boldsymbol{x}, t) & \text{on } \partial V_{\boldsymbol{D}} \times [0, \infty), \end{cases} \\ &\text{magnetic boundary conditions} \begin{cases} \psi = \hat{\psi}(\boldsymbol{x}, t) & \text{on } \partial V_{\psi} \times [0, \infty), \\ b = \hat{b}(\boldsymbol{x}, t) & \text{on } \partial V_{\boldsymbol{B}} \times [0, \infty), \end{cases} \end{aligned} \tag{2-9}$$

where $t_i = \sigma_{ij} n_j$, $d = D_i n_i$, $b = B_i n_i$, and the initial conditions are

$$\begin{aligned} u_i(\boldsymbol{x}, 0) &= u_i^0(\boldsymbol{x}), \quad \boldsymbol{x} \in \overline{V}, \\ \dot{u}_i(\boldsymbol{x}, 0) &= v_i^0(\boldsymbol{x}), \quad \boldsymbol{x} \in \overline{V}. \end{aligned} \tag{2-10}$$

In (2-9), $\hat{u}_i(\boldsymbol{x}, t)$, $\hat{t}_i(\boldsymbol{x}, t)$, $\hat{\varphi}(\boldsymbol{x}, t)$, $\hat{d}(\boldsymbol{x}, t)$, $\hat{\psi}(\boldsymbol{x}, t)$, and $\hat{b}(\boldsymbol{x}, t)$ are, respectively, the prescribed displacement components, traction components, electric potential, electric displacement, magnetic potential, and magnetic field over the boundary. In addition, by the displacement-potential boundary conditions we mean

$$\begin{aligned} u_i &= \hat{u}_i(\boldsymbol{x}, t) \quad \text{on } \partial V_{\boldsymbol{u}} \times [0, \infty), \\ \varphi &= \hat{\varphi}(\boldsymbol{x}, t) \quad \text{on } \partial V_{\varphi} \times [0, \infty), \\ \psi &= \hat{\psi}(\boldsymbol{x}, t) \quad \text{on } \partial V_{\psi} \times [0, \infty). \end{aligned} \tag{2-11}$$

Similar to [Gurtin 1964], for reference in the remainder of the paper, the regularity assumptions are listed here:

(i) $\rho > 0$ is continuously differentiable on $\overline{V}$,

(ii) $\boldsymbol{C}$, $\boldsymbol{e}$, $\boldsymbol{\kappa}$ and $\boldsymbol{S}$, $\boldsymbol{d}$, $\boldsymbol{\chi}$ are continuously differentiable on $\overline{V}$ and meet (2-5), (2-7), and (2-8).

(iii) $\boldsymbol{u}^0(\boldsymbol{x})$ is continuously differentiable on $\overline{V}$,

(iv) $\boldsymbol{v}^0(\boldsymbol{x})$ is continuously differentiable on $\overline{V}$,

(v) $\boldsymbol{f}$ and q_e are continuously differentiable on $\overline{V}$,

(vi) $\hat{\boldsymbol{u}}$, $\hat{\varphi}$, and $\hat{\psi}$ are continuous on $\overline{\partial V_{\boldsymbol{u}}} \times [0, \infty)$, $\overline{\partial V_{\boldsymbol{D}}} \times [0, \infty)$, and $\overline{\partial V_{\boldsymbol{B}}} \times [0, \infty)$, respectively, and

(vii) $\hat{\boldsymbol{t}}$, $\hat{d}$, and $\hat{b}$ are piecewise continuous on $\overline{\partial V_{\boldsymbol{\sigma}}} \times [0, \infty)$, $\overline{\partial V_{\boldsymbol{D}}} \times [0, \infty)$, and $\overline{\partial V_{\boldsymbol{B}}} \times [0, \infty)$, respectively.

Since our goal is to obtain variational principles for electromagneto-elastic materials, analogous to [Gurtin 1964], we define what we mean by an admissible process:

Definition. An ordered array $S = [\boldsymbol{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}, \boldsymbol{E}, \boldsymbol{H}, \boldsymbol{D}, \boldsymbol{B}, \varphi, \psi]$ is called an admissible process on $\overline{V} \times [0, \infty)$ provided that $u_i \in C^{1,2}$, $\varepsilon_{ij} \in C^{0,0}$, $\sigma_{ij} \in C^{1,0}$, $D_i \in C^{1,0}$, $B_i \in C^{1,0}$, $\varphi \in C^{1,0}$, $\psi \in C^{1,0}$, $E_i \in C^{0,0}$, $H_i \in C^{0,0}$, $\varepsilon_{ij} = \varepsilon_{ji}$, and $\sigma_{ij} = \sigma_{ji}$.

In addition, a solution of the mixed initial-boundary value problem (i.e., IBVP) is an admissible process which meets (2-2), (2-3), (2-4), (2-9), and (2-10).

3. Variational principles

Now, for an electromagneto-elastic material, convolutional variational forms, originally developed for the case of pure elasticity in [Gurtin 1964], will be derived. The first part of the results, i.e., Theorems 1, 2, and 3, is devoted to the characterization of $[\boldsymbol{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}, \boldsymbol{E}, \boldsymbol{H}, \boldsymbol{D}, \boldsymbol{B}, \varphi, \psi]$. In the second part, as a corollary of Theorem 3, a variational form characterizing $[\boldsymbol{u}, \varphi, \psi]$ is set. Finally, the results of the third part are applied to the characterization of $[\boldsymbol{\sigma}, \varphi, \psi]$. In the results, t_i, $\tilde{t}_i$, d, $\tilde{d}$, b, and $\tilde{b}$ will be consistently used in place of $\sigma_{ij}n_j$, $\tilde{\sigma}_{ij}n_j$, $D_i n_i$, $\widetilde{D}_i n_i$, $B_i n_i$, and $\tilde{B}_i n_i$, respectively. Additionally, the definitions of $h(t)$ and $f_i^b(\boldsymbol{x}, t)$, which shall be used in the sequel, have been given in (B-1).

3.1. *Variational principles characterizing* $[\boldsymbol{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}, \boldsymbol{E}, \boldsymbol{H}, \boldsymbol{D}, \boldsymbol{B}, \varphi, \psi]$.

First let us derive a general form which imposes no restriction on the fields:

Theorem 1. *Let Ω denote the set of all admissible processes. Let $S = [\boldsymbol{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}, \boldsymbol{E}, \boldsymbol{H}, \boldsymbol{D}, \boldsymbol{B}, \varphi, \psi]$ be an element of Ω and define the functional ϑ_t on Ω at each time, say $t \in [0, \infty)$, in the form of*

$$
\begin{aligned}
\vartheta_t(S) = {} & \frac{1}{2}\int_V C_{ijkl}(\boldsymbol{x})[h * \varepsilon_{ij} * \varepsilon_{kl}](\boldsymbol{x}, t)\, d\boldsymbol{x} - \int_V [h * \sigma_{ij} * \varepsilon_{ij}](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
& - \int_V e^E_{kij}(\boldsymbol{x})[h * \varepsilon_{ij} * E_k](\boldsymbol{x}, t)\, d\boldsymbol{x} - \int_V e^H_{kij}(\boldsymbol{x})[h * \varepsilon_{ij} * H_k](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
& - \frac{1}{2}\int_V \kappa^E_{ij}(\boldsymbol{x})[h * E_i * E_j](\boldsymbol{x}, t)\, d\boldsymbol{x} - \frac{1}{2}\int_V \kappa^H_{ij}(\boldsymbol{x})[h * H_i * H_j](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
& - \int_V \kappa^{EH}_{ij}(\boldsymbol{x})[h * E_i * H_j](\boldsymbol{x}, t)\, d\boldsymbol{x} + \int_V [h * D_i * E_i](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
& + \int_V [h * H_i * B_i](\boldsymbol{x}, t)\, d\boldsymbol{x} - \int_V [h * (D_{i,i} - q_e) * \varphi](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
& - \int_V [h * B_{i,i} * \psi](\boldsymbol{x}, t)\, d\boldsymbol{x} + \frac{1}{2}\int_V \rho(\boldsymbol{x})[u_i * u_i](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
& - \int_V [(h * \sigma_{ij,j} + f_i^b) * u_i](\boldsymbol{x}, t)\, d\boldsymbol{x} + \int_{\partial V_{\boldsymbol{\sigma}}} [h * (t_i - \hat{t}_i) * u_i](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
& + \int_{\partial V_{\boldsymbol{u}}} [h * t_i * \hat{u}_i](\boldsymbol{x}, t)\, d\boldsymbol{x} + \int_{\partial V_{\varphi}} [h * d * \hat{\varphi}](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
& + \int_{\partial V_{\psi}} [h * b * \hat{\psi}](\boldsymbol{x}, t)\, d\boldsymbol{x} + \int_{\partial V_{\boldsymbol{D}}} [h * (d - \hat{d}) * \varphi](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
& + \int_{\partial V_{\boldsymbol{B}}} [h * (b - \hat{b}) * \psi](\boldsymbol{x}, t)\, d\boldsymbol{x}. \qquad (3\text{-}1)
\end{aligned}
$$

Then, S is a solution of the mixed initial-boundary value problem if and only if $\delta\vartheta_t(S) = 0$ over Ω, within the time interval $t \in [0, \infty)$.

Proof. Let $\tilde{S} = [\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\varepsilon}}, \tilde{\boldsymbol{\sigma}}, \widetilde{\boldsymbol{E}}, \widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{D}}, \widetilde{\boldsymbol{B}}, \tilde{\varphi}, \tilde{\psi}] \in \Omega$ be an admissible process and suppose that $S + \lambda\tilde{S} \in \Omega$ for all real values of λ. Using (3-1), (A-5), the symmetry condition (2-5), the divergence theorem, and properties of convolution product given in Appendix A, we obtain

$$
\begin{aligned}
\delta_{\tilde{S}}\vartheta_t(S) = &\int_V [h * (C_{ijkl}\varepsilon_{kl} - e^E_{kij}E_k - e^H_{kij}H_k - \sigma_{ij}) * \tilde{\varepsilon}_{ij}](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
&+ \int_V [h*(-e^E_{kij}\varepsilon_{ij} - \kappa^E_{ik}E_i - \kappa^{EH}_{ki}H_i + D_k)*\tilde{E}_k](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
&+ \int_V [h*(-e^H_{kij}\varepsilon_{ij} - \kappa^{EH}_{ik}E_i - \kappa^H_{ik}H_i + B_k)*\tilde{H}_k](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
&+ \int_V [h*(E_i + \varphi_{,i})*\widetilde{D}_i](\boldsymbol{x}, t)\, d\boldsymbol{x} + \int_V [h*(H_i + \psi_{,i})*\tilde{B}_i](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
&- \int_V [h*(D_{i,i} - q_e)*\tilde{\varphi}](\boldsymbol{x}, t)\, d\boldsymbol{x} - \int_V [h*B_{i,i}*\tilde{\psi}](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
&+ \int_V [h*(u_{(i,j)} - \varepsilon_{ij})*\tilde{\sigma}_{ij}](\boldsymbol{x}, t) d\boldsymbol{x} - \int_V [(h*\sigma_{ij,j} + f^b_i - \rho u_i)*\tilde{u}_i](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
&- \int_{\partial V_{\boldsymbol{u}}} [h*(u_i - \hat{u}_i)*\tilde{t}_i](\boldsymbol{x}, t)\, d\boldsymbol{x} + \int_{\partial V_\sigma} [h*(t_i - \hat{t}_i)*\tilde{u}_i](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
&- \int_{\partial V_\varphi} [h*(\varphi - \hat{\varphi})*\tilde{d}](\boldsymbol{x}, t)\, d\boldsymbol{x} - \int_{\partial V_\psi} [h*(\psi - \hat{\psi})*\tilde{b}](\boldsymbol{x}, t)\, d\boldsymbol{x} \\
&+ \int_{\partial V_{\boldsymbol{D}}} [h*(d - \hat{d})*\tilde{\varphi}](\boldsymbol{x}, t)\, d\boldsymbol{x} + \int_{\partial V_{\boldsymbol{B}}} [h*(b - \hat{b})*\tilde{\psi}](\boldsymbol{x}, t)\, d\boldsymbol{x}. \qquad (3\text{-}2)
\end{aligned}
$$

First, based on Theorem B.2, for every $\tilde{S} \in \Omega$ $(0 \le t < \infty)$ we immediately find $\delta_{\tilde{S}}\vartheta_t(S) = 0$ when S is a solution of the IBVP, implying $\delta\vartheta_t(S) = 0$ over Ω. Conversely, suppose $\delta\vartheta_t(S) = 0$ over Ω. Let $\tilde{S} = [\tilde{\boldsymbol{u}}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}, 0, 0] \in \Omega$ where $\tilde{\boldsymbol{u}}$ and all its spatial derivatives are identical to zero on $\partial V \times [0, \infty)$. Then, based on $\delta\vartheta_t(S) = 0$, (3-2), and Lemma A.1, we obtain $h * \sigma_{ij,j} + f^b_i - u_i = 0$ on $V \times [0, \infty)$. Next, suppose $\tilde{\boldsymbol{u}}$ and all its spatial derivatives are identical to zero on $\partial V_{\boldsymbol{u}} \times [0, \infty)$. Based on Lemma A.2, $h * \sigma_{ij,j} + f^b_i - u_i = 0$ on $V \times [0, \infty)$, $\delta\vartheta_t(S) = 0$, and (3-2), we have $h * (t_i - \hat{t}_i) = 0$ on $\partial V_\sigma \times [0, \infty)$. Since $h \neq 0$, the property of convolution reads $(t_i - \hat{t}_i) = 0$ on $\partial V_\sigma \times [0, \infty)$. Considering (2-5), by the same token, $-e^E_{kij}\varepsilon_{ij} - \kappa^E_{ki}E_i - \kappa^{EH}_{ki}H_i + D_k = 0$ on $V \times (0, \infty)$ and $-e^H_{kij}\varepsilon_{ij} - \kappa^{EH}_{ik}E_i - \kappa^H_{ki}H_i + B_k = 0$ on $V \times (0, \infty)$ can be obtained. With the same logic mentioned so far, one can readily deduce $(2\text{-}2)_{2\text{–}3}$, $(2\text{-}3)_{2\text{–}3}$, and $(2\text{-}9)_{4,6}$. Next, let $\tilde{S} = [\boldsymbol{0}, \tilde{\boldsymbol{\varepsilon}}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}, 0, 0] \in \Omega$ in which $\tilde{\boldsymbol{\varepsilon}}$ is a symmetric second-order tensor and zero-valued on the whole boundary at all times. Thus, the symmetry of the constitutive equations and symmetry of $\boldsymbol{\sigma}$,

$\delta\vartheta_t(S)=0$, (3-2), and Lemma A.1 imply $h*(C_{ijkl}\varepsilon_{kl}-e^E_{kij}E_k-e^H_{kij}H_k-\sigma_{ij})=0$ on $V\times[0,\infty)$, leading to (2-4)$_1$. Similarly, by taking $\tilde{S}=[\mathbf{0},\mathbf{0},\tilde{\boldsymbol{\sigma}},\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0},0,0]$ in which $\tilde{\boldsymbol{\sigma}}$ is a symmetric second-order tensor and zero-valued on the whole boundary at all times, considering symmetry of ϵ_{ij}, we conclude (2-3)$_1$. Moreover, let us define $\tilde{S}=[\mathbf{0},\mathbf{0},\tilde{\boldsymbol{\sigma}},\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0},0,0]\in\Omega$ in which $\tilde{\boldsymbol{\sigma}}$ is a symmetric second-order tensor and zero-valued on the boundary $\partial V_{\boldsymbol{\sigma}}$ at all times. By taking into account (2-3)$_1$, $\delta\vartheta_t(S)=0$, (3-2), and Lemma A.3, we immediately find $u_i-\hat{u}_i=0$ on $\partial V_{\boldsymbol{u}}\times[0,\infty)$. Also, by having $\tilde{S}=[\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0},\widetilde{\boldsymbol{D}},\mathbf{0},0,0]$ in which $\widetilde{\boldsymbol{D}}$ is zero on $\partial V_{\boldsymbol{D}}$ at all times, using (2-3)$_2$, $\delta\vartheta_t(S)=0$, (3-2), and the Corollary A.4, we conclude $h*(\varphi-\hat{\varphi})=0$ on $\partial V_\varphi\times[0,\infty)$, which implies (2-9)$_3$. In a similar fashion, we have (2-9)$_5$. Hence, based on Theorem B.2, $\delta\vartheta_t(S)=0$ over Ω yields a solution of the mixed initial-boundary value problem, and the proof is complete. □

Next, as the first example in which there is a restriction on fields, analogous to [Gurtin 1964], we obtain a variational form of the mixed initial-boundary value problem of the admissible process for which the kinematic equation (2-3)$_1$ is identically satisfied.

Theorem 2. *Let Ω denote the set of all admissible processes which satisfy* (2-3)$_1$. *Let $S=[\boldsymbol{u},\boldsymbol{\varepsilon},\boldsymbol{\sigma},\boldsymbol{E},\boldsymbol{H},\boldsymbol{D},\boldsymbol{B},\varphi,\psi]$ be an element of Ω and define the functional Ξ_t on Ω at each time*, *say $t\in[0,\infty)$*, *in the form of*

$$
\begin{aligned}
\Xi_t(S)=&\int_V[h*\sigma_{ij}*\varepsilon_{ij}](\boldsymbol{x},t)\,d\boldsymbol{x}-\frac{1}{2}\int_V S_{ijkl}(\boldsymbol{x})[h*\sigma_{ij}*\sigma_{kl}](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&-\int_V d^E_{kij}(\boldsymbol{x})[h*\sigma_{ij}*E_k](\boldsymbol{x},t)\,d\boldsymbol{x}-\int_V d^H_{kij}(\boldsymbol{x})[h*\sigma_{ij}*H_k](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&-\frac{1}{2}\int_V \chi^E_{ij}(\boldsymbol{x})[h*E_i*E_j](\boldsymbol{x},t)\,d\boldsymbol{x}-\frac{1}{2}\int_V \chi^H_{ij}(\boldsymbol{x})[h*H_i*H_j](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&-\int_V \chi^{EH}_{ij}(\boldsymbol{x})[h*E_i*H_j](\boldsymbol{x},t)\,d\boldsymbol{x}+\int_V[h*D_i*E_i](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&+\int_V[h*B_i*H_i](\boldsymbol{x},t)\,d\boldsymbol{x}-\int_V[h*(D_{i,i}-q_e)*\varphi](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&-\int_V[h*B_{i,i}*\psi](\boldsymbol{x},t)\,d\boldsymbol{x}+\frac{1}{2}\int_V\rho(\boldsymbol{x})[u_i*u_i](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&-\int_V[f^b_i*u_i](\boldsymbol{x},t)\,d\boldsymbol{x}-\int_{\partial V_{\boldsymbol{u}}}[h*t_i*(u_i-\hat{u}_i)](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&-\int_{\partial V_{\boldsymbol{\sigma}}}[h*\hat{t}_i*u_i](\boldsymbol{x},t)\,d\boldsymbol{x}+\int_{\partial V_\varphi}[h*d*\hat{\varphi}](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&+\int_{\partial V_{\boldsymbol{D}}}[h*(d-\hat{d})*\varphi](\boldsymbol{x},t)\,d\boldsymbol{x}+\int_{\partial V_\psi}[h*b*\hat{\psi}](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&+\int_{\partial V_{\boldsymbol{B}}}[h*(b-\hat{b})*\psi](\boldsymbol{x},t)\,d\boldsymbol{x}.
\end{aligned}\tag{3-3}
$$

Then, S is a solution of the mixed initial-boundary value problem if and only if $\delta\,\Xi_t(S)=0$ over Ω, within the time interval $t\in[0,\infty)$.

Proof. Let $\tilde{S}=[\tilde{\boldsymbol{u}},\tilde{\boldsymbol{\varepsilon}},\tilde{\boldsymbol{\sigma}},\widetilde{\boldsymbol{E}},\widetilde{\boldsymbol{H}},\widetilde{\boldsymbol{D}},\widetilde{\boldsymbol{B}},\tilde{\varphi},\tilde{\psi}]\in\Omega$ be an admissible process and suppose that $S+\lambda\tilde{S}\in\Omega$ for all real values of λ. By employing (3-3) and using (A-5), the compatibility equation $(2\text{-}3)_1$, the symmetry condition (2-7), the divergence theorem, and properties of convolution product given in Appendix A, one can find

$$\begin{aligned}
\delta_{\tilde{S}}\,\Xi_t(S)=&-\int_V[h*(S_{ijkl}\sigma_{kl}+d^E_{kij}E_k+d^H_{kij}H_k-\varepsilon_{ij})*\tilde{\sigma}_{ij}](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&+\int_V[h*(D_k-d^E_{kij}\sigma_{ij}-\chi^E_{ik}E_i-\chi^{EH}_{ki}H_i)*\tilde{E}_k](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&+\int_V[h*(B_k-d^H_{kij}\sigma_{ij}-\chi^{EH}_{ik}E_i-\chi^H_{ik}H_i)*\tilde{H}_k](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&-\int_V[(h*\sigma_{ij,j}+f^b_i-\rho u_i)*\tilde{u}_i](\boldsymbol{x},t)\,d\boldsymbol{x}+\int_V[h*(E_i+\varphi_{,i})*\widetilde{D}_i](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&+\int_V[h*(H_i+\psi_{,i})*\tilde{B}_i](\boldsymbol{x},t)\,d\boldsymbol{x}-\int_V[h*(D_{i,i}-q_e)*\tilde{\varphi}](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&-\int_V[h*B_{i,i}*\tilde{\psi}](\boldsymbol{x},t)\,d\boldsymbol{x}-\int_{\partial V_{\boldsymbol{u}}}[h*(u_i-\hat{u}_i)*\tilde{t}_i](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&+\int_{\partial V_{\boldsymbol{\sigma}}}[h*(t_i-\hat{t}_i)*\tilde{u}_i](\boldsymbol{x},t)\,d\boldsymbol{x}-\int_{\partial V_{\varphi}}[h*(\varphi-\hat{\varphi})*\tilde{d}](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&-\int_{\partial V_{\psi}}[h*(\psi-\hat{\psi})*\tilde{b}](\boldsymbol{x},t)\,d\boldsymbol{x}+\int_{\partial V_{\boldsymbol{D}}}[h*(d-\hat{d})*\tilde{\varphi}](\boldsymbol{x},t)\,d\boldsymbol{x}\\
&+\int_{\partial V_{\boldsymbol{B}}}[h*(b-\hat{b})*\tilde{\psi}](\boldsymbol{x},t)\,d\boldsymbol{x}.
\end{aligned}\tag{3-4}$$

Due to Theorem B.2, if S is a solution to the IBVP, then we conclude $\delta_{\tilde{S}}\,\Xi_t(S)=0$ for every $\tilde{S}\in\Omega$ $(0\le t<\infty)$, leading us to $\delta\,\Xi_t(S)=0$ over Ω. Also, with the same path given in Theorem 1, by using Lemmas A.1, A.2, and A.3, Corollary A.4, (3-4), $\delta\,\Xi_t(S)=0$ over Ω, properties of convolution product, considering (2-6), (2-7), (2-8), and Theorem B.2, the implication in the other direction is proved. □

Theorem 1 is the most general variational form giving a solution of the elastodynamics IBVP for electromagneto-elastic materials. The displacement-strain kinematic equation in Theorem 2 is employed as the only restriction. Hence, one can further restrict the admissible process by which it automatically satisfies some of the field equations and boundary conditions. In doing so, define a kinematically admissible process and consequently obtain a relevant variational form.

Definition. An admissible process is called a kinematically admissible process if it satisfies the kinematic equations (2-3), the constitutive equations (2-4), and the displacement-potential boundary conditions.

Theorem 3. *Let Ω denote the set of all kinematically admissible processes. Let $S = [\boldsymbol{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}, \boldsymbol{E}, \boldsymbol{H}, \boldsymbol{D}, \boldsymbol{B}, \varphi, \psi]$ be an element of Ω and define the functional $\Sigma_t(S)$ on Ω at each time, say $t \in [0, \infty)$, in the form of*

$$\begin{aligned}\Sigma_t(S) = \frac{1}{2}\int_V [h * \sigma_{ij} * \varepsilon_{ij}](\boldsymbol{x}, t)\, d\boldsymbol{x} + \frac{1}{2}\int_V \rho(\boldsymbol{x})[u_i * u_i](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ - \int_V [f_i^b * u_i](\boldsymbol{x}, t)\, d\boldsymbol{x} - \frac{1}{2}\int_V [h * D_i * E_i](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ - \frac{1}{2}\int_V [h * B_i * H_i](\boldsymbol{x}, t)\, d\boldsymbol{x} + \int_V [h * q_e * \varphi](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ - \int_{\partial V_\sigma} [h * \hat{t}_i * u_i](\boldsymbol{x}, t)\, d\boldsymbol{x} - \int_{\partial V_{\boldsymbol{D}}} [h * \hat{d} * \varphi](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ - \int_{\partial V_{\boldsymbol{B}}} [h * \hat{b} * \psi](\boldsymbol{x}, t)\, d\boldsymbol{x}. \qquad (3\text{-}5)\end{aligned}$$

Then, S is a solution of the mixed initial-boundary value problem if and only if $\delta\Sigma_t(S) = 0$ over Ω, within the time interval $t \in [0, \infty)$.

Proof. Let $\tilde{S} = [\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\varepsilon}}, \tilde{\boldsymbol{\sigma}}, \widetilde{\boldsymbol{E}}, \widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{D}}, \widetilde{\boldsymbol{B}}, \tilde{\varphi}, \tilde{\psi}]$ be an admissible process and suppose that $S + \lambda\tilde{S} \in \Omega$ for all real values of λ. Obviously, it implies $\tilde{u}_i = 0$ on $\partial V_{\boldsymbol{u}} \times [0, \infty)$, $\tilde{\varphi} = 0$ on $\partial V_\varphi \times [0, \infty)$, and $\tilde{\psi} = 0$ on $\partial V_\psi \times [0, \infty)$. By making use of (3-5), (A-5), kinematic equation (2-3), the constitutive equation (2-4), the symmetry condition (2-5), and the divergence theorem, we obtain

$$\begin{aligned}\delta_{\tilde{S}}\Sigma_t(S) = -\int_V [(h * \sigma_{ij,j} + f_i^b - \rho u_i) * \tilde{u}_i](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ - \int_V [h * (D_{i,i} - q_e) * \tilde{\varphi}](\boldsymbol{x}, t)\, d\boldsymbol{x} - \int_V [h * B_{i,i} * \tilde{\psi}](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ + \int_{\partial V_\sigma} [h * (t_i - \hat{t}_i) * \tilde{u}_i](\boldsymbol{x}, t)\, d\boldsymbol{x} + \int_{\partial V_{\boldsymbol{D}}} [h * (d - \hat{d}) * \tilde{\varphi}](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ + \int_{\partial V_{\boldsymbol{B}}} [h * (b - \hat{b}) * \tilde{\psi}](\boldsymbol{x}, t)\, d\boldsymbol{x}. \qquad (3\text{-}6)\end{aligned}$$

As is clear when S is a solution of the mixed initial-boundary value problem, then $\delta_{\tilde{S}}\Sigma_t(S) = 0$ for every admissible $\tilde{S}$ $(0 \le t < \infty)$, leading to $\delta\Sigma_t(S) = 0$ over Ω. Conversely, similar to Theorem 1, since the array $[\tilde{\boldsymbol{u}}, \tilde{\varphi}, \tilde{\psi}]$ can be defined arbitrarily, for every $t \in [0, \infty)$, on the domain and the boundary ∂V, then by employing $\delta\Sigma_t(S) = 0$ over Ω, (3-6), Lemmas A.1 and A.2, properties of convolution product, and Theorem B.2, we obtain the desired result. □

3.2. ***Variational principles characterizing $[\boldsymbol{u}, \boldsymbol{\varphi}, \boldsymbol{\psi}]$.*** With the aid of Theorem 3, it is straightforward to obtain a variational form in terms of the displacement field, electric potential, and magnetic potential. In this regard, define an admissible and a kinematically admissible displacement-potential process as follows:

Definition. An array $S=[\boldsymbol{u},\varphi,\psi]$ is called an admissible displacement-potential process if $\boldsymbol{u}\in C^{1,2}$, $\varphi\in C^{1,0}$, and $\psi\in C^{1,0}$.

Definition. An array $S=[\boldsymbol{u},\varphi,\psi]$ is called a kinematically admissible displacement-potential process if it is an admissible displacement-potential process and meets the displacement-potential boundary conditions.

Now, to obtain the desired variational form as a corollary of Theorem 3, the constitutive equations (2-4) and kinematic relations (2-3) need to be employed in (3-5). Doing so, one can easily obtain the variational form corresponding to a kinematically admissible displacement-potential process:

Theorem 4. *Let Ω denote the set of all kinematically admissible displacement-potential processes. Let $S=[\boldsymbol{u},\varphi,\psi]$ be an element of Ω and define the functional Θ_t on Ω at each time, say $t\in[0,\infty)$, in the form of*

$$\begin{aligned}\Theta_t(S)=\frac{1}{2}\int_V&[h*(C_{ijkl}u_{k,l}+e^E_{kij}\varphi_{,k}+e^H_{kij}\psi_{,k})*u_{i,j}](\boldsymbol{x},t)\,d\boldsymbol{x}\\&+\frac{1}{2}\int_V[h*(e^E_{ikl}u_{k,l}-\kappa^E_{ij}\varphi_{,j}-\kappa^{EH}_{ij}\psi_{,j})*\varphi_{,i}](\boldsymbol{x},t)\,d\boldsymbol{x}\\&+\frac{1}{2}\int_V[h*(e^H_{ikl}u_{k,l}-\kappa^{EH}_{ji}\varphi_{,j}-\kappa^H_{ij}\psi_{,j})*\psi_{,i}](\boldsymbol{x},t)\,d\boldsymbol{x}\\&-\int_V[f^b_i*u_i](\boldsymbol{x},t)\,d\boldsymbol{x}+\int_V[h*q_e*\varphi](\boldsymbol{x},t)\,d\boldsymbol{x}\\&+\frac{1}{2}\int_V\rho(\boldsymbol{x})[u_i*u_i](\boldsymbol{x},t)\,d\boldsymbol{x}-\int_{\partial V_\sigma}[h*\hat{t}_i*u_i](\boldsymbol{x},t)\,d\boldsymbol{x}\\&-\int_{\partial V_D}[h*\hat{d}*\varphi](\boldsymbol{x},t)\,d\boldsymbol{x}-\int_{\partial V_B}[h*\hat{b}*\psi](\boldsymbol{x},t)\,d\boldsymbol{x}.\end{aligned}\tag{3-7}$$

Then, $S=[\boldsymbol{u},\varphi,\psi]$ is a solution of the mixed initial-boundary value problem if and only if $\delta\Theta_t(S)=0$ over Ω, within the time interval $t\in[0,\infty)$.

3.3. ***Variational principles characterizing $[\boldsymbol{\sigma},\boldsymbol{\varphi},\boldsymbol{\psi}]$.*** Theorem B.4 motivates us to develop variational forms in terms of the stress field rather than the displacement field, which is more desirable when the mechanical boundary conditions are traction-type. In other words, it is of interest to obtain conditions by which the array $S=[\boldsymbol{\sigma},\varphi,\psi]$ is a solution to the mixed initial-boundary value problems. To this end, let first define what we mean by a kinematically admissible electromagneto-stress process:

Definition. An array $[\boldsymbol{\sigma},\varphi,\psi]$ in which $\boldsymbol{\sigma}$ is a second-order symmetric tensor and $\boldsymbol{\sigma}\in C^{2,0}$, $\varphi\in C^{2,0}$, and $\psi\in C^{2,0}$ is called a kinematically admissible electromagneto-stress process if $\varphi=\hat{\varphi}(\boldsymbol{x},t)$ on $\partial V_\varphi\times[0,\infty)$ and $\psi=\hat{\psi}(\boldsymbol{x},t)$ on $\partial V_\psi\times[0,\infty)$.

Now, the following statement for the kinematically admissible electromagneto-stress processes holds true.

Theorem 5. *Let Ω denote the set of all kinematically admissible electromagneto-stress processes. Let $S = [\boldsymbol{\sigma}, \varphi, \psi]$ be an element of Ω and define the functional Υ_t on Ω at each time, say $t \in [0, \infty)$, in the form of*

$$\begin{aligned}\Upsilon_t(S) = \frac{1}{2}\int_V \Big[\frac{h}{\rho} * \sigma_{ij,j} * \sigma_{ik,k}\Big](\boldsymbol{x}, t)\, d\boldsymbol{x} - \int_V \Big[\Big(\frac{1}{\rho} f^b_{(i}\Big)_{,j)} * \sigma_{ij}\Big](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ + \frac{1}{2}\int_V [S_{ijkl}\sigma_{ij} * \sigma_{kl} + \chi^E_{ij}\varphi_{,i} * \varphi_{,j} + \chi^H_{ij}\psi_{,i} * \psi_{,j}](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ + \int_V [-d^E_{ikl}\sigma_{kl} * \varphi_{,i} - d^H_{ikl}\sigma_{kl} * \psi_{,i} + \chi^{EH}_{ij}\psi_{,j} * \varphi_{,i}](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ - \int_V [q_e * \varphi](\boldsymbol{x}, t)\, d\boldsymbol{x} + \int_{\partial V_{\boldsymbol{u}}} \Big[\Big(\frac{f^b_i}{\rho} - \hat{u}_i\Big) * t_i\Big](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ + \int_{\partial V_{\boldsymbol{\sigma}}} \Big[\frac{h}{\rho} * (\hat{t}_i - t_i) * \sigma_{ij,j}\Big](\boldsymbol{x}, t)\, d\boldsymbol{x} + \int_{\partial V_{\boldsymbol{D}}} [\hat{d} * \varphi](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ + \int_{\partial V_{\boldsymbol{B}}} [\hat{b} * \psi](\boldsymbol{x}, t)\, d\boldsymbol{x}. \qquad (3\text{-}8)\end{aligned}$$

Then, S is a solution of the mixed initial-boundary value problem if and only if $\delta\Upsilon_t(S) = 0$ over Ω, within the time interval $t \in [0, \infty)$.

Proof. Let $\tilde{S} = [\tilde{\boldsymbol{\sigma}}, \tilde{\varphi}, \tilde{\psi}]$ be an ordered array in which $\tilde{\sigma}_{ij} = \tilde{\sigma}_{ji}$, $\tilde{\boldsymbol{\sigma}} \in C^{2,0}$, $\tilde{\varphi} \in C^{2,0}$, and $\tilde{\psi} \in C^{2,0}$ such that $S + \lambda\tilde{S} \in \Omega$ for all real values of λ — which implies $\tilde{\varphi} = 0$ on $\partial V_\varphi \times [0, \infty)$ and $\tilde{\psi} = 0$ on $\partial V_\psi \times [0, \infty)$. By making use of (3-8), (A-5), and symmetry condition (2-7), applying the divergence theorem, properties of convolution, and the above-mentioned restriction, we find

$$\begin{aligned}\delta_{\tilde{S}}\Upsilon_t(S) = \int_V \Big[\Big(-\Big(\frac{h}{\rho} * \sigma_{(ik,k}\Big)_{,j)} + S_{ijkl}\sigma_{kl} - d^E_{kij}\varphi_{,k} - d^H_{kij}\psi_{,k} - \Big(\frac{1}{\rho} f^b_{(i}\Big)_{,j)}\Big) * \tilde{\sigma}_{ij}\Big](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ + \int_V [((d^E_{ikl}\sigma_{kl} - \chi^E_{ij}\varphi_{,j} - \chi^{EH}_{ij}\psi_{,j})_{,i} - q_e) * \tilde{\varphi}](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ + \int_V [(d^H_{ikl}\sigma_{kl} - \chi^{EH}_{ji}\varphi_{,j} - \chi^H_{ij}\psi_{,j})_{,i} * \tilde{\psi}](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ + \int_{\partial V_{\boldsymbol{u}}} \Big[\Big(\frac{h}{\rho} * \sigma_{ik,k} + \frac{f^b_i}{\rho} - \hat{u}_i\Big) * \tilde{t}_i\Big](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ + \int_{\partial V_{\boldsymbol{\sigma}}} \Big[\frac{h}{\rho} * (\hat{t}_i - t_i) * \tilde{\sigma}_{ij,j}\Big](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ - \int_{\partial V_{\boldsymbol{D}}} [((d^E_{ikl}\sigma_{kl} - \chi^E_{ij}\varphi_{,j} - \chi^{EH}_{ij}\psi_{,j})n_i - \hat{d}) * \tilde{\varphi}](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ - \int_{\partial V_{\boldsymbol{B}}} [((d^H_{ikl}\sigma_{kl} - \chi^{EH}_{ji}\varphi_{,j} - \chi^H_{ij}\psi_{,j})n_i - \hat{b}) * \tilde{\psi}](\boldsymbol{x}, t)\, d\boldsymbol{x}. \qquad (3\text{-}9)\end{aligned}$$

Obviously, if S is a solution of the mixed initial-boundary value problem, then $\delta_{\tilde{S}}\Upsilon_t(S) = 0$ for every above-defined $\tilde{S}$ ($0 \le t < \infty$) is implied from Theorem B.4, resulting in $\delta\Upsilon_t(S) = 0$ over Ω. Conversely, if $\delta\Upsilon_t(S) = 0$ over Ω, then, by utilizing Lemmas A.1, A.2, A.3, and A.5, and Theorem B.4, we obtain the desired result. □

When the mechanical boundary condition is entirely traction-type, say traction problems, one can establish a more convenient variational form in terms of the ordered array $S = [\boldsymbol{\sigma}, \varphi, \psi]$.

Definition. A kinematically admissible electromagneto-stress process is called a dynamically admissible electromagneto-stress process if $\sigma_{ij} n_j = \hat{t}_i(\boldsymbol{x}, t)$ on $\partial V_{\boldsymbol{\sigma}} \times [0, \infty)$.

Now, based on Theorem 5, it is straightforward to obtain a variational framework for the dynamically admissible electromagneto-stress processes:

Theorem 6. *Let Ω denote the set of all dynamically admissible electromagneto-stress processes. Let $S = [\boldsymbol{\sigma}, \varphi, \psi]$ be an element of Ω and define the functional $\Im_t$ on Ω at each time*, *say $t \in [0, \infty)$*, *in the form of*

$$\begin{aligned}\Im_t(S) = \frac{1}{2}\int_V \Big[\frac{h}{\rho} * \sigma_{ij,j} * \sigma_{ik,k}\Big](\boldsymbol{x}, t)\, d\boldsymbol{x} - \int_V \Big[\Big(\frac{1}{\rho} f^b_{(i}\Big)_{,j)} * \sigma_{ij}\Big](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ + \frac{1}{2}\int_V [S_{ijkl}\sigma_{ij} * \sigma_{kl} + \chi^E_{ij}\varphi_{,i} * \varphi_{,j} + \chi^H_{ij}\psi_{,i} * \psi_{,j}](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ + \int_V [-d^E_{ikl}\sigma_{kl} * \varphi_{,i} - d^H_{ikl}\sigma_{kl} * \psi_{,i} + \chi^{EH}_{ij}\psi_{,j} * \varphi_{,i}](\boldsymbol{x}, t)\, d\boldsymbol{x} \\ - \int_V [q_e * \varphi](\boldsymbol{x}, t)\, d\boldsymbol{x} + \int_{\partial V_{\boldsymbol{D}}} [\hat{d} * \varphi](\boldsymbol{x}, t)\, d\boldsymbol{x} + \int_{\partial V_{\boldsymbol{B}}} [\hat{b} * \psi](\boldsymbol{x}, t)\, d\boldsymbol{x}. \quad (3\text{-}10)\end{aligned}$$

Then, *S is a solution of the traction problem* (*i.e.*, $\partial V_{\boldsymbol{u}} = \varnothing$) *if and only if* $\delta\Im_t(S) = 0$ *over* Ω, *within the time interval* $t \in [0, \infty)$.

Proof. The proof is analogous to that of Theorem 5. □

4. Conclusion

In parallel to [Gurtin 1964], on the basis of alternative field equations for electromagneto-elastic materials, which are comprehensively given in Appendix B, the convolutional variational principles have been derived and proved rigorously. In Theorem 1, a general convolutional variational form, in which the admissible process is not required to meet any field equations and/or boundary/initial conditions, has been derived. The convolutional variational principle corresponding to an admissible process that meets only the strain-displacement relation has been formulated in Theorem 2. Next, the result for a more restricted process — namely, a kinematically admissible process — has been formulated in Theorem 3. As a corollary of Theorem 3, the convolutional variational principle corresponding to a kinematically admissible displacement-potential has been set in Theorem 4. Lastly, through Theorems 5 and 6, variational principles in terms of stress rather than displacement have been established, respectively, for general problems and traction problems. On the application side, the results of the present work provide a robust basis for numerical analysis of electromagneto-elastic materials with general material domain geometry and boundary/initial conditions.

Appendix A: Mathematical background

Here, for the sake of completeness, we summarize the basic concepts, originally developed in [Gurtin 1964; Nickell and Sackman 1968], that are employed in the main body of the paper. For a comprehensive discussion, the readers are referred to those references.

Smoothness of a vector (or scalar) function $\boldsymbol{f}$ (or f) is expressed mathematically by $C^{M,N}$, where M and N are nonnegative integers, with the following definition: $\boldsymbol{f}$ (or f) $\in C^{M,N}$, in which $\boldsymbol{f}$ (or f) is a function of position and time defined on $\overline{V} \times (0, \infty)$, if and only if the functions $f^{(n)}_{\underbrace{,ij\cdots k}_{m \text{ indices}}}$ ($m = 0, 1, \ldots, M$ and $n = 0, 1, \ldots, N$) exist and are continuous.

The pair $(\boldsymbol{x}, t) \in \overline{\partial V} \times [0, \infty)$ is called a regular point if the unit outward normal $\boldsymbol{n}$ at $\boldsymbol{x}$, and at any time, is continuous. Moreover, the function f is called a piecewise regular function on boundary $\overline{\partial V_i} \times [0, \infty)$ with $i = \boldsymbol{u}, \boldsymbol{\sigma}, \varphi, \boldsymbol{D}, \psi, \boldsymbol{B}$ if and only if it is piecewise continuous on $\overline{\partial V_i} \times [0, \infty)$ and continuous on every regular point of that region. Additionally, for piecewise regular functions f and $\hat{f}$ on $\partial \overline{V}_i \times [0, \infty)$, we say $f = \hat{f}$ if and only if the equality holds true for any regular point $(\boldsymbol{x}, t) \in \overline{\partial V_i} \times [0, \infty)$.

The symbol $f * g$, in which f and g are functions of the position and continuous functions of time defined on $\Re \times [0, \infty)$, with $\Re$ a subset of the Euclidean space, indicates the convolution of two functions in the sense of

$$[f * g](\boldsymbol{x}, t) = \int_0^t f(\boldsymbol{x}, t - \lambda) g(\boldsymbol{x}, \lambda)\, d\lambda, \quad (\boldsymbol{x}, t) \in \Re \times [0, \infty). \tag{A-1}$$

In this regard, one can show that the following properties hold true:

$$f * g = g * f, \tag{A-2}$$

$$f * g = 0 \iff f = 0 \vee g = 0, \tag{A-3}$$

$$f * (g * h) = (f * g) * h = f * g * h. \tag{A-4}$$

A functional is a real-valued function on a subset of a linear space. Denoting a linear space by L and a subset of L by K, and defining $\Phi(S)$ as a functional on K, we define

$$\delta_{\tilde{S}} \Phi(S) = \frac{d}{d\lambda} \Phi(S + \lambda \tilde{S}) \bigg|_{\lambda=0} \tag{A-5}$$

for all real numbers λ, where S, $\tilde{S} \in L$, and $S + \lambda \tilde{S} \in K$. And we say the variation of $\Phi(S)$ is zero and write $\delta\Phi(S) = 0$ over K if and only if $\delta_{\tilde{S}}\Phi(S)$ exists and equals zero for all $\tilde{S}$ such that S, $\tilde{S} \in L$, and $S + \lambda \tilde{S} \in K$.

Now, we list four lemmas and one corollary proved in [Gurtin 1964; Nickell and Sackman 1968]. However, we write them in such a way that they are applicable to the present study.

Lemma A.1 [Gurtin 1964]. *Let ϑ be a continuous function on $\overline{V} \times [0, \infty)$ and suppose*

$$\int_V \vartheta * \omega(\boldsymbol{x}, t)\, d\boldsymbol{x} = 0, \quad 0 \le t < \infty, \tag{A-6}$$

for every $\omega \in C^{\infty,\infty}$ which, together with its spatial derivatives, vanishes on $\partial V \times [0, \infty)$. Then

$$\vartheta = 0 \quad \text{on } \overline{V} \times [0, \infty). \tag{A-7}$$

Lemma A.2 [Gurtin 1964]. *Let ϑ be a piecewise regular function on $\overline{\partial V_i} \times [0, \infty)$ with $i = \boldsymbol{\sigma}, \boldsymbol{D}, \boldsymbol{B}$, and suppose*

$$\int_{\partial V_i} \vartheta * \omega(\boldsymbol{x}, t)\, d\boldsymbol{x} = 0, \quad 0 \le t < \infty, \tag{A-8}$$

for every $\omega \in C^{\infty,\infty}$ that vanishes on $\partial V_j \times [0, \infty)$ with, respectively, $j = \boldsymbol{u}, \varphi, \psi$. Then

$$\vartheta = 0 \quad \text{on } \overline{\partial V_i} \times [0, \infty). \tag{A-9}$$

Lemma A.3 [Gurtin 1964]. *Let ϑ_i be continuous on $\overline{\partial V_{\boldsymbol{u}}} \times [0, \infty)$, and suppose we have*

$$\int_{\partial V_{\boldsymbol{u}}} \vartheta_i * (\omega_{ij} n_j)(\boldsymbol{x}, t)\, d\boldsymbol{x} = 0, \quad 0 \le t < \infty, \tag{A-10}$$

for every $\omega_{ij} \in C^{\infty,\infty}$ which, together with all of its spatial derivatives, vanishes on $\partial V_{\boldsymbol{\sigma}} \times [0, \infty)$ and has the property $\omega_{ij} = \omega_{ji}$. Then

$$\vartheta_i = 0 \quad \text{on } \overline{\partial V_{\boldsymbol{u}}} \times [0, \infty). \tag{A-11}$$

The following statement is a corollary of Lemma A.3.

Corollary A.4 [Nickell and Sackman 1968]. *Let ϑ be continuous on $\overline{\partial V_\varphi}$ (or $\overline{\partial V_\psi}$)$\times [0, \infty)$ and suppose*

$$\int_{\partial V_\varphi \ (or\ \partial V_\psi)} [\vartheta * (\omega_i n_i)](\boldsymbol{x}, t)\, d\boldsymbol{x} = 0, \quad 0 \le t < \infty, \tag{A-12}$$

for every $\omega_i \in C^{\infty,\infty}$ which, together with its spatial derivatives, vanishes on $\partial V_{\boldsymbol{D}}$ (or $\partial V_{\boldsymbol{B}}$) $\times [0, \infty)$. Then

$$\vartheta = 0 \quad \text{on } \overline{\partial V_\varphi} \ (\text{or } \overline{\partial V_\psi}) \times [0, \infty). \tag{A-13}$$

Lemma A.5 [Gurtin 1964]. *Let ϑ_i be a piecewise regular function on $\overline{\partial V_{\boldsymbol{\sigma}}} \times [0, \infty)$, and suppose*

$$\int_{\partial V_{\sigma}} \vartheta_i * (\omega_{ij,j})(\boldsymbol{x}, t)\, d\boldsymbol{x} = 0, \quad 0 \le t < \infty, \tag{A-14}$$

for all $\omega_{ij} \in C^{\infty,\infty}$ with $\omega_{ij} = \omega_{ji}$. Then

$$\vartheta_i = 0 \quad \text{on } \overline{\partial V_{\boldsymbol{\sigma}}} \times [0, \infty). \tag{A-15}$$

Appendix B: Integro-partial-differential field equations

The alternative integro-partial-differential field equations of motion of an electromagneto-elastic body are derived in this part. To start with, define the functions [Gurtin 1964]

$$\begin{aligned} h(t) &= t, \quad 0 \le t < \infty, \\ f_i^b(\boldsymbol{x}, t) &= h * f(\boldsymbol{x}, t) + \rho(\boldsymbol{x})(t v_i^0(\boldsymbol{x}) + u_i^0(\boldsymbol{x})) \end{aligned} \tag{B-1}$$

in which f_i^b is a vector field obtained from the prescribed data (2-10) and the body force. We now have the following alternative formulation of $(2\text{-}2)_1$.

Theorem B.1. *Let $u_i \in C^{0,2}$ and $\sigma_{ij} \in C^{1,0}$ be a vector field and a second-order symmetric tensor field*, *respectively. Then $\boldsymbol{u}$ and $\boldsymbol{\sigma}$ meet* $(2\text{-}2)_1$ *and the associated initial boundary conditions* (2-10) *if and only if*

$$\rho \boldsymbol{u} = h * \nabla . \boldsymbol{\sigma} + \boldsymbol{f}^b \quad \text{on } V \times [0, \infty). \tag{B-2}$$

Proof. See [Gurtin 1964]. □

Now, with the help of the following theorem, which is the direct result of Theorem B.1, one can define alternative field equations of the mixed initial-boundary value problem.

Theorem B.2. *The admissible process $S = [\boldsymbol{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}, \boldsymbol{E}, \boldsymbol{H}, \boldsymbol{D}, \boldsymbol{B}, \varphi, \psi]$ is a solution of the mixed initial-boundary value problem if and only if it satisfies* (B-2), $(2\text{-}2)_{2\text{–}3}$, (2-3), (2-4), *and* (2-9).

Now, through the next two theorems, we obtain two variants of field equations for electromagneto-elastic materials.

Theorem B.3. *Let $u_i \in C^{2,2}$, $\varphi \in C^{2,0}$, and $\psi \in C^{2,0}$. Then the ordered array $[\boldsymbol{u}, \varphi, \psi]$ corresponds to a solution of the mixed initial-boundary value problem if* *and only if the following equations hold true*:

$$
\begin{aligned}
&u_i = h * (C_{ijkl}u_{k,l} + e^E_{kij}\varphi_{,k} + e^H_{kij}\psi_{,k})_{,j} + f^b_i && \text{on } V \times [0,\infty),\\
&(e^E_{ikl}u_{k,l} - \kappa^E_{ij}\varphi_{,j} - \kappa^{EH}_{ij}\psi_{,j})_{,i} = q_e && \text{on } V \times [0,\infty),\\
&(e^H_{ikl}u_{k,l} - \kappa^{EH}_{ji}\varphi_{,j} - \kappa^H_{ij}\psi_{,j})_{,i} = 0 && \text{on } V \times [0,\infty),\\
&u_i = \hat{u}_i(\boldsymbol{x}, t) && \text{on } \partial V_{\boldsymbol{u}} \times [0,\infty),\\
&(C_{ijkl}u_{k,l} + e^E_{kij}\varphi_{,k} + e^H_{kij}\psi_{,k})n_j = \hat{t}_i(\boldsymbol{x}, t) && \text{on } \partial V_{\boldsymbol{\sigma}} \times [0,\infty),\\
&\varphi = \hat{\varphi}(\boldsymbol{x}, t) && \text{on } \partial V_{\varphi} \times [0,\infty),\\
&(e^E_{ikl}u_{k,l} - \kappa^E_{ij}\varphi_{,j} - \kappa^{EH}_{ij}\psi_{,j})n_i = \hat{d}(\boldsymbol{x}, t) && \text{on } \partial V_{\boldsymbol{D}} \times [0,\infty),\\
&\psi = \hat{\psi}(\boldsymbol{x}, t) && \text{on } \partial V_{\psi} \times [0,\infty),\\
&(e^H_{ikl}u_{k,l} - \kappa^{EH}_{ji}\varphi_{,j} - \kappa^H_{ij}\psi_{,j})n_i = \hat{b}(\boldsymbol{x}, t) && \text{on } \partial V_{\boldsymbol{B}} \times [0,\infty).
\end{aligned}
\tag{B-3}
$$

Proof. First, suppose that relations (B-3) hold true. Thus, $(2\text{-}9)_{1,3,5}$ are automatically satisfied. Define $\boldsymbol{\varepsilon}$, $\boldsymbol{E}$, and $\boldsymbol{H}$ through (2-3). Also, define $\boldsymbol{\sigma}$, $\boldsymbol{D}$, and $\boldsymbol{B}$ via (2-4). Then, $(2\text{-}9)_{2,4,6}$ are identically satisfied due to the symmetry (2-5); the above-defined $\boldsymbol{\varepsilon}$, $\boldsymbol{\sigma}$, and (2-5) together with $(\text{B-}3)_1$ give (B-2); $(\text{B-}3)_{2\text{–}3}$ together with the above-defined $\boldsymbol{D}$, $\boldsymbol{B}$, $\boldsymbol{\varepsilon}$, $\boldsymbol{E}$, $\boldsymbol{H}$, and symmetry (2-5) give $(2\text{-}2)_{2\text{–}3}$. Hence, by Theorem B.2, (B-3) is a solution to the mixed initial-boundary value problem. On the other hand, (B-2), $(2\text{-}2)_{2\text{–}3}$, (2-3), (2-4), (2-5), and (2-9) imply (B-3) and the proof is complete. □

Theorem B.4. *Let $\sigma_{ij} \in C^{2,0}$, $\varphi \in C^{2,0}$, and $\psi \in C^{2,0}$ with $\sigma_{ij} = \sigma_{ji}$. Then the ordered array $[\boldsymbol{\sigma}, \varphi, \psi]$ is a solution to the mixed initial-boundary value problem if and only if the following equations hold true*:

$$
\begin{aligned}
&S_{ijkl}\sigma_{kl} = \Big(\frac{h}{\rho} * \sigma_{(ik,k}\Big)_{,j)} + \Big(\frac{1}{\rho} f^b_{(i}\Big)_{,j)} + d^E_{kij}\varphi_{,k} + d^H_{kij}\psi_{,k} && \text{on } V \times [0,\infty),\\
&(d^E_{ikl}\sigma_{kl} - \chi^E_{ij}\varphi_{,j} - \chi^{EH}_{ij}\psi_{,j})_{,i} = q_e && \text{on } V \times [0,\infty),\\
&(d^H_{ikl}\sigma_{kl} - \chi^{EH}_{ji}\varphi_{,j} - \chi^H_{ij}\psi_{,j})_{,i} = 0 && \text{on } V \times [0,\infty),\\
&\frac{h}{\rho} * \sigma_{ik,k} + \frac{1}{\rho} f^b_i = \hat{u}_i(\boldsymbol{x}, t) && \text{on } \partial V_{\boldsymbol{u}} \times [0,\infty),\\
&\sigma_{ij}n_j = \hat{t}_i(\boldsymbol{x}, t) && \text{on } \partial V_{\boldsymbol{\sigma}} \times [0,\infty),\\
&\varphi = \hat{\varphi}(\boldsymbol{x}, t) && \text{on } \partial V_{\varphi} \times [0,\infty),\\
&(d^E_{ikl}\sigma_{kl} - \chi^E_{ij}\varphi_{,j} - \chi^{EH}_{ij}\psi_{,j})n_i = \hat{d}(\boldsymbol{x}, t) && \text{on } \partial V_{\boldsymbol{D}} \times [0,\infty),\\
&\psi = \hat{\psi}(\boldsymbol{x}, t) && \text{on } \partial V_{\psi} \times [0,\infty),\\
&(d^H_{ikl}\sigma_{kl} - \chi^{EH}_{ji}\varphi_{,j} - \chi^H_{ij}\psi_{,j})n_i = \hat{b}(\boldsymbol{x}, t) && \text{on } \partial V_{\boldsymbol{B}} \times [0,\infty).
\end{aligned}
\tag{B-4}
$$

Proof. First, suppose that (B-4) holds. Hence, $(2\text{-}9)_{2,3,5}$ are automatically satisfied. Define $\boldsymbol{u}$, $\boldsymbol{E}$, and $\boldsymbol{H}$ through (B-2), $(2\text{-}3)_2$, and $(2\text{-}3)_3$, respectively. Also, define $\boldsymbol{\varepsilon}$, $\boldsymbol{D}$, and $\boldsymbol{B}$ via (2-6). Then, $(2\text{-}9)_{1,4,6}$ are identically satisfied; (2-6) and (2-8) imply (2-4); $(2\text{-}3)_1$ holds because of $(\text{B-4})_1$, $(\text{B-4})_4$, and the above-defined $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$; $(\text{B-4})_{2–3}$ together with the above-defined $\boldsymbol{D}$, $\boldsymbol{B}$, $\boldsymbol{E}$, and $\boldsymbol{H}$ give $(2\text{-}2)_{2–3}$. Hence, by Theorem B.2, (B-4) is a solution to the mixed initial-boundary value problem. On the other hand, (B-2), $(2\text{-}2)_{2–3}$, (2-3), (2-4), (2-8), and (2-9) imply (B-4), and the proof is complete. □

Acknowledgments

The authors wish to thank an anonymous reviewer for his/her comments improving the manuscript.

This work has been partially supported by the NSF under grant CMMI-1462749.

References

[Bo 2003] H. Bo, "Gurtin-type region-wise variational principles for thermopiezoelectric elastodynamics", *Appl. Math. Mech.* **24**:5 (2003), 576–584.

[Brown 1966] W. F. Brown, Jr., *Magnetoelastic interactions*, Springer Tracts in Natural Philosophy **9**, Springer, 1966.

[Coleman and Dill 1971] B. D. Coleman and E. H. Dill, "Thermodynamic restrictions on the constitutive equations of electromagnetic theory", *Z. Angew. Math. Phys.* **22**:4 (1971), 691–702.

[El-Karamany and Ezzat 2011] A. S. El-Karamany and M. A. Ezzat, "Convolutional variational principle, reciprocal and uniqueness theorems in linear fractional two-temperature thermoelasticity", *J. Thermal Stresses* **34**:3 (2011), 264–284.

[Eringen and Maugin 1990] A. C. Eringen and G. A. Maugin, *Electrodynamics of continua, I: Foundations and solid media*, Springer, 1990.

[Guggenheim 1936a] E. A. Guggenheim, "On magnetic and electrostatic energy", *Proc. Roy. Soc. Lond. A* **155**:884 (1936), 49–70.

[Guggenheim 1936b] E. A. Guggenheim, "The thermodynamics of magnetization", *Proc. Roy. Soc. Lond. A* **155**:884 (1936), 70–101.

[Gurtin 1963] M. E. Gurtin, "Variational principles in the linear theory of viscoelasticity", *Arch. Rational Mech. Anal.* **13** (1963), 179–191.

[Gurtin 1964] M. E. Gurtin, "Variational principles for linear elastodynamics", *Arch. Rational Mech. Anal.* **16** (1964), 34–50.

[Hashin and Shtrikman 1962] Z. Hashin and S. Shtrikman, "On some variational principles in anisotropic and non-homogeneous elasticity", *J. Mech. Phys. Solids* **10** (1962), 335–342.

[He 2000] J.-H. He, "A variational approach to electroelastic analysis of piezoelectric ceramics with surface electrodes", *Mech. Res. Comm.* **27**:4 (2000), 445–450.

[He 2001a] J.-H. He, "Coupled variational principles of piezoelectricity", *Internat. J. Engrg. Sci.* **39**:3 (2001), 323–341.

[He 2001b] J.-H. He, "Hamilton principle and generalized variational principles of linear thermopiezoelectricity", *J. Appl. Mech.* **68**:4 (2001), 666–667.

[He 2001c] J.-H. He, “Variational theory for linear magneto-electro-elasticity”, *Int. J. Nonlinear Sci. Numer. Simul.* **2**:4 (2001), 309–316.

[Ignaczak 1959] J. Ignaczak, “Direct determination of stresses from the stress equations of motion in elasticity”, *Arch. Mech. Stos.* **11** (1959), 671–678.

[Ignaczak 1963] J. Ignaczak, “A completeness problem for stess equations of motion in the linear elasticity theory”, *Arch. Mech. Stos.* **15** (1963), 225–234.

[Irschik et al. 2010] H. Irschik, M. Krommer, and Y. Vetyukov, “On the use of piezoelectric sensors in structural mechanics: some novel strategies”, *Sensors* **10**:6 (2010), 5626–5641.

[Kuang 2008] Z.-B. Kuang, “Some variational principles in elastic dielectric and elastic magnetic materials”, *Eur. J. Mech. A Solids* **27**:3 (2008), 504–514.

[Landau et al. 1984] L. D. Landau, E. M. Lifshitz, and Pitaevskiĭ, *Electrodynamics of continuous media*, 2nd ed., Course of Theoretical Physics **8**, Pergamon, Oxford, 1984.

[Li 2003] J. Y. Li, “Uniqueness and reciprocity theorems for linear thermo-electro-magneto-elasticity”, *Quart. J. Mech. Appl. Math.* **56**:1 (2003), 35–43.

[Li and Kardomateas 2006] R. Li and G. A. Kardomateas, “The mode III interface crack in piezo-electro-magneto-elastic dissimilar bimaterials”, *J. Appl. Mech.* **73**:2 (2006), 220–227.

[Maugin 1988] G. A. Maugin, *Continuum mechanics of electromagnetic solids*, North-Holland Series in Applied Mathematics and Mechanics **33**, North-Holland, Amsterdam, 1988.

[Miehe and Rosato 2011] C. Miehe and D. Rosato, “A rate-dependent incremental variational formulation of ferroelectricity”, *Internat. J. Engrg. Sci.* **49**:6 (2011), 466–496.

[Miehe et al. 2011] C. Miehe, D. Rosato, and B. Kiefer, “Variational principles in dissipative electro-magneto-mechanics: a framework for the macro-modeling of functional materials”, *Internat. J. Numer. Methods Engrg.* **86**:10 (2011), 1225–1276.

[Mindlin 1972] R. D. Mindlin, “Elasticity, piezoelectricity and crystal lattice dynamics”, *J. Elast.* **2**:4 (1972), 217–282.

[Nelson 1979] D. F. Nelson, *Electric, optic, and acoustic interactions in dielectrics*, Wiley, New York, 1979.

[Nickell and Sackman 1968] R. E. Nickell and J. L. Sackman, “Variational principles for linear coupled thermoelasticity”, *Quart. Appl. Math.* **26** (1968), 11–26.

[Oden and Reddy 1983] J. T. Oden and J. N. Reddy, *Variational methods in theoretical mechanics*, 2nd ed., Springer, 1983.

[Ostoja-Starzewski 2019] M. Ostoja-Starzewski, “Ignaczak equation of elastodynamics”, *Math. Mech. Solids* **24**:11 (2019), 3674–3713.

[Penfield et al. 1963] P. L. Penfield, Jr., L. J. Chu, and H. A. Haus, “Electrodynamics of moving media”, quarterly progress report 70, Massachusetts Institute of Technology Research Laboratory of Electronics, 1963, Available at http://hdl.handle.net/1721.1/53884.

[Pérez-Fernández et al. 2009] L. D. Pérez-Fernández, J. Bravo-Castillero, R. Rodríguez-Ramos, and F. J. Sabina, “On the constitutive relations and energy potentials of linear thermo-magneto-electro-elasticity”, *Mech. Res. Comm.* **36**:3 (2009), 343–350.

[Schoeftner and Irschik 2016] J. Schoeftner and H. Irschik, “Stress tracking in thin bars by eigen-strain actuation”, *J. Sound Vibration* **383** (2016), 35–45.

[Tiersten and Tsai 1972] H. F. Tiersten and C. F. Tsai, “On the interaction of the electromagnetic field with heat conducting deformable insulators”, *J. Math. Phys.* **13**:3 (1972), 361–378.

[Toupin 1956] R. A. Toupin, “The elastic dielectric”, *J. Rational Mech. Anal.* **5** (1956), 849–915.

[Wang et al. 2010] Z.-J. Wang, D.-Z. Zheng, and C.-B. Zheng, "Simplified Gurtin-type generalized variational principles for fully dynamic magneto-electro-elasticity with geometrical nonlinearity", *Int. J. Solid. Struct.* **47**:22–23 (2010), 3115–3120.

[Washizu 1957] K. Washizu, "On the variational principles applied to dynamics problems of elastic bodies", tech. report 25-23, Massachusetts Institute of Technology Aeroelastic and Structures Research Laboratory, 1957.

Received 21 Aug 2019. Revised 17 Dec 2019. Accepted 21 Jan 2020.

AMIRHOSSEIN AMIRI-HEZAVEH: aa7@illinois.edu
Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States

POUYAN KARIMI: pkarimi2@illinois.edu
Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States

MARTIN OSTOJA-STARZEWSKI: martinos@illinois.edu
Department of Mechanical Science and Engineering, Institute for Condensed Matter Theory, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, United States

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
Vol. 8, No. 1, 2020
dx.doi.org/10.2140/memocs.2020.8.69

EQUILIBRIUM THEORY FOR A LIPID BILAYER WITH A CONFORMING CYTOSKELETAL MEMBRANE

BRETT HENDRICKSON, MILAD SHIRANI AND DAVID J. STEIGMANN

We discuss the mechanics of a lipid bilayer with a conforming cytoskeletal membrane in which the bilayer has the structure of a nematic liquid crystal and the cytoskeleton that of a simple elastic solid. Under certain conditions the cytoskeletal membrane mimics the effects of the so-called spontaneous curvature of the conventional theory of lipid membranes. The model is used to predict the classical biconcave discoid shape of red-blood cells in equilibrium.

1. Introduction

In this work we outline a model of the elastic response of a lipid bilayer with a conforming cytoskeletal membrane. This is intended for application to the mechanics of red-blood cells, which are known to consist of bilayers with subsurface cytoskeletal membranes formed by spectrin filaments arranged in networks that exhibit 6-fold hexagonal symmetry [Pan et al. 2018]. The basic framework of our model is similar to that underpinning Krishnaswamy's pioneering work [Krishnaswamy 1996] in which material points of the bilayer and cytoskeleton are assumed to be tethered by a so-called connector field while occupying distinct surfaces. The role of this connector is to maintain contact between the bilayer and cytoskeleton as they deform. In that work the bilayer is regarded as a fluid shell, as in Jenkins' model [Jenkins 1977], and the cytoskeleton is considered to be a perfectly flexible solid membrane. Current work on the mechanics of the cytoskeleton [Kamm and Mofrad 2006; Herant and Dembo 2006] suggests that the extent to which it convects with the bilayer is largely unknown. In the present work we therefore take the conservative view that the role of Krishnaswamy's connector is confined to maintaining congruency of the cytoskeletal and bilayer surfaces while playing no significant further role in the mechanical response.

In Section 2 we develop the model of the bilayer/cytoskeleton system via asymptotic expansion in which the bilayer is regarded as a thin nematic liquid crystal film and the cytoskeleton as a thin layer of a simple elastic solid. Certain vector

Communicated by Francesco dell'Isola.
MSC2010: 53B50, 74K25.
Keywords: lipid bilayers, membrane cytoskeleton, differential geometry.

fields arising in this procedure occur algebraically in the reduced model and are accordingly evaluated before proceeding further. This is explained in Section 3. In Section 4 we discuss material symmetry conditions for the cytoskeleton and bilayer. Some basic aspects of the differential geometry of surfaces [Naghdi 1972; Ciarlet 2005] are recalled in Section 5 and adapted there to the kinematics of congruent configurations of the bilayer and cytoskeleton. Equilibrium equations are deduced in Section 6 on the basis of a patchwise virtual-power postulate, and restrictions implied by the operative versions of the Legendre–Hadamard condition are discussed in Section 7. We conclude, in Section 8, with a derivation of a strain-energy function for the cytoskeleton which is such as to admit a surface having the shape of the characteristic biconcave discoid of a red-blood cell as an equilibrium state.

2. Leading-order asymptotic energy for small thickness

Consider a configuration of the bilayer-cytoskeletal combination in the shape of a prismatic cylinder generated by the parallel translation of a plane region Π forming the interface of the bilayer and cytoskeleton (Figure 1). The lipids of the bilayer are presumed to be straight, parallel and of uniform length in this configuration. The bilayer has thickness αh and the cytoskeleton $(1-\alpha)h$, where h is the thickness of the cylinder and $\alpha \in [0, 1]$.

The energy of the cylinder is

$$\mathcal{E} = \int_{\Pi} \mathcal{U}\, dA, \tag{1}$$

where

$$\mathcal{U} = \int_0^{\alpha h} \mathcal{U}_b\, d\varsigma + \int_{-(1-\alpha)h}^{0} \mathcal{U}_c\, d\varsigma, \tag{2}$$

in which $\mathcal{U}_b$ and $\mathcal{U}_c$ respectively are the volumetric energy densities of the bilayer and cytoskeleton and ς is a through-thickness coordinate.

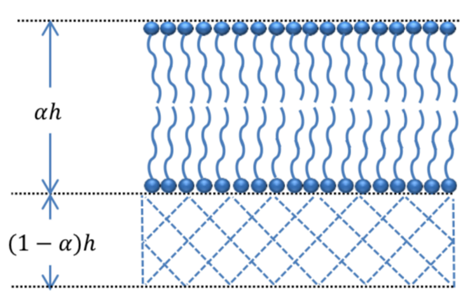


Figure 1. A patch of the bilayer and cytoskeleton.

A central aspect of the model to be developed is that Π is assumed to convect as a material surface with respect to both the bilayer and the cytoskeleton deformations so as to maintain congruency; that is, the (possibly distinct) images of Π under the bilayer and cytoskeletal deformations are subsets of a single surface ω. We elaborate on the kinematical implications of this restriction below. Here we assume that ω can be covered completely by the images of such patches, each of which is assumed, for the sake of notational convenience, to be parametrized by a single coordinate chart.

We suppose the thickness h to be much smaller than any other length scale, l say, in a given problem. If the latter is used as the unit of length ($l = 1$), then the dimensionless thickness $h \ll 1$. Regarding $\mathcal{U}$ as a function of h, we combine Leibniz's rule with a Taylor expansion to derive

$$\mathcal{U} = hU + o(h), \quad \text{with} \quad U = \alpha U_b + (1-\alpha)U_c, \tag{3}$$

in which U_b and U_c respectively are the values of $\mathcal{U}_b$ and $\mathcal{U}_c$ at $\varsigma = 0$, i.e., at their common interface Π. Accordingly,

$$\mathcal{E}/h = E + o(h)/h, \quad \text{where} \quad E = \int_\Pi U\,dA, \tag{4}$$

is the leading-order energy for small h.

Alternatively, in view of the fact that the thickness of the bilayer/cytoskeleton composite is on the order of molecular dimensions, it is appropriate to contemplate a direct theory based at the outset on the idea of a material surface without regard to thickness effects. However, the present asymptotic approach offers guidance as to the features that such a direct model should possess.

We assume the cytoskeleton to be a uniform elastic material with a strain energy given by

$$\mathcal{U}_c = \mathcal{W}_c(\widetilde{\boldsymbol{F}}), \tag{5}$$

where $\widetilde{\boldsymbol{F}}$ is the gradient of the cytoskeletal deformation $\tilde{\boldsymbol{\chi}}(\boldsymbol{x})$, with $\boldsymbol{x} \in \Pi \times [-(1-\alpha)h, 0]$, i.e., $\boldsymbol{x} = \boldsymbol{\xi} + \varsigma\boldsymbol{k}$, where $\boldsymbol{\xi}$ is the projection of $\boldsymbol{x}$ onto the plane region Π with unit normal $\boldsymbol{k}$ and $\varsigma \in [-(1-\alpha)h, 0]$. Thus, $\widetilde{\boldsymbol{F}} = \hat{\boldsymbol{F}}(\boldsymbol{\xi}, \varsigma)$, where

$$\hat{\boldsymbol{F}} = \nabla\hat{\boldsymbol{\chi}} + \hat{\boldsymbol{\chi}}' \otimes \boldsymbol{k}. \tag{6}$$

Here $(\cdot)' = \partial(\cdot)/\partial\varsigma$, $\nabla(\cdot)$ is the (two-dimensional) gradient with respect to $\boldsymbol{\xi}$, and $\hat{\boldsymbol{\chi}}(\boldsymbol{\xi}, \varsigma) = \tilde{\boldsymbol{\chi}}(\boldsymbol{\xi} + \varsigma\boldsymbol{k})$. Then,

$$U_c = \mathcal{W}_c(\boldsymbol{F}), \quad \text{where} \quad \boldsymbol{F} = \nabla\boldsymbol{r}_c + \boldsymbol{d} \otimes \boldsymbol{k}, \tag{7}$$

is the restriction to Π of the cytoskeletal deformation gradient, in which $\boldsymbol{r}_c(\boldsymbol{\xi}) = \hat{\boldsymbol{\chi}}_{|\Pi}$ is the interfacial cytoskeletal deformation and $\boldsymbol{d}(\boldsymbol{\xi}) = \hat{\boldsymbol{\chi}}'_{|\Pi}$ is the interfacial value of the normal derivative of the deformation.

Following Helfrich [1973], we model the lipid bilayer as a liquid crystal with an energy density

$$\mathcal{U}_b = \mathcal{W}_b(\tilde{\boldsymbol{n}}, \widetilde{\boldsymbol{D}}), \tag{8}$$

where $\tilde{\boldsymbol{n}}$ is a field of unit vectors specifying the local molecular orientation and $\widetilde{\boldsymbol{D}} = \operatorname{grad} \tilde{\boldsymbol{n}}$ is its (spatial) gradient. It is customary [Virga 1994] to specify a constitutive function for the energy per unit current volume and to regard the liquid crystal as an incompressible medium. Accordingly $\mathcal{U}_b$ is also the energy per unit reference volume, as assumed in the foregoing. Then,

$$U_b = \mathcal{W}_b(\boldsymbol{n}, \boldsymbol{D}), \tag{9}$$

where $\boldsymbol{n}$ and $\boldsymbol{D}$ are the interfacial values of $\tilde{\boldsymbol{n}}$ and $\widetilde{\boldsymbol{D}}$, respectively. Here, as in Helfrich's theory [Helfrich 1973], we suppress lipid *tilt* and thus take $\boldsymbol{n}$ to be the unit-normal field to the image π_b of the interface Π in the current configuration of the lipid/cytoskeleton system. In these circumstances, we have

$$\boldsymbol{D} = \nabla_s \boldsymbol{n} + \boldsymbol{\eta} \otimes \boldsymbol{n}, \tag{10}$$

where $\nabla_s(\cdot)$ is the surfacial gradient on π_b and $\boldsymbol{\eta}$ is the restriction to π_b of the derivative of $\tilde{\boldsymbol{n}}$ in the direction of $\tilde{\boldsymbol{n}}$. Because the latter is a field of unit vectors, we require $\boldsymbol{n} \cdot \boldsymbol{\eta} = 0$ and conclude that $\boldsymbol{\eta}$ is a tangential vector field on π_b.

The Gauss and Weingarten equations of differential geometry furnish

$$\nabla_s \boldsymbol{n} = -\boldsymbol{b}, \tag{11}$$

where $\boldsymbol{b}$ is the symmetric curvature 2-tensor on the local tangent planes of π_b. We elaborate further in Section 5 below.

The energy density of the composite is thus given, in an abuse of notation, by the function

$$U(\nabla \boldsymbol{r}_c, \boldsymbol{b}, \boldsymbol{d}, \boldsymbol{n}, \boldsymbol{\eta}) = \alpha\, U_b(\boldsymbol{b}, \boldsymbol{n}, \boldsymbol{\eta}) + (1-\alpha)\, U_c(\nabla \boldsymbol{r}_c, \boldsymbol{d}), \tag{12}$$

where

$$U_b(\boldsymbol{b}, \boldsymbol{n}, \boldsymbol{\eta}) = \mathcal{W}_b(\boldsymbol{n}, -\boldsymbol{b} + \boldsymbol{\eta} \otimes \boldsymbol{n}) \quad \text{and} \quad U_c(\nabla \boldsymbol{r}_c, \boldsymbol{d}) = \mathcal{W}_c(\nabla \boldsymbol{r}_c + \boldsymbol{d} \otimes \boldsymbol{k}). \tag{13}$$

We observe that the dependence of the energy on the fields $\boldsymbol{d}$ and $\boldsymbol{\eta}$ is purely algebraic. This suggests a strategy, pursued in the next section, whereby we attempt to render the energy stationary with respect to these fields *a priori*.

3. Determination of $\boldsymbol{d}$ and $\boldsymbol{\eta}$

3.1. ***Cytoskeletal deformation.*** We decompose $\boldsymbol{d}$ into normal and tangential parts as

$$\boldsymbol{d} = d_n \boldsymbol{n} + (\nabla \boldsymbol{r}_c)\, \boldsymbol{e}, \tag{14}$$

where $d_n = \boldsymbol{d} \cdot \boldsymbol{n}$, $\boldsymbol{e}$ is a 2-vector on Π and $J_c \boldsymbol{n} = \boldsymbol{F}^* \boldsymbol{k}$, in which $\boldsymbol{F}^*$ is the cofactor of $\boldsymbol{F}$, and we note that $\nabla \boldsymbol{r}_c$ maps Π to the tangent plane of the image π_c of Π under the deformation at the material point in question. Here J_c $(= |\boldsymbol{F}^* \boldsymbol{k}|)$ and $\boldsymbol{n}$ respectively are the areal stretch of the interface due to the deformation of the cytoskeleton and the unit normal to π_c; these are determined by $\nabla \boldsymbol{r}_c$. We then have $\det \boldsymbol{F} = \boldsymbol{F}\boldsymbol{k} \cdot \boldsymbol{F}^* \boldsymbol{k} = J_c d_n$ and thus require $d_n > 0$.

The cytoskeletal energy is frame-invariant if and only if it depends on $\boldsymbol{F}$ via the Cauchy–Green tensor $\boldsymbol{C} = \boldsymbol{F}^t \boldsymbol{F}$; we write $\mathcal{W}_c(\boldsymbol{F}) = F(\boldsymbol{C})$, where, from $(7)_2$ and (14),

$$\boldsymbol{C} = \boldsymbol{c} + \boldsymbol{\gamma} \otimes \boldsymbol{k} + \boldsymbol{k} \otimes \boldsymbol{\gamma} + \left(d_n^2 + \boldsymbol{e} \cdot \boldsymbol{c}\boldsymbol{e}\right) \boldsymbol{k} \otimes \boldsymbol{k}, \tag{15}$$

with

$$\boldsymbol{c} = (\nabla \boldsymbol{r}_c)^t \, \nabla \boldsymbol{r}_c \quad \text{and} \quad \boldsymbol{\gamma} = \boldsymbol{c}\boldsymbol{e}, \tag{16}$$

and we remark that

$$J_c^2 = \det \boldsymbol{c}. \tag{17}$$

Let $G(\boldsymbol{e}) = F(\boldsymbol{C}(\boldsymbol{e}))$, where $\boldsymbol{C}(\boldsymbol{e})$ is the function obtained by fixing d_n and $\nabla \boldsymbol{r}_c$ in (15). We seek 2-vectors $\boldsymbol{e}$ that render G stationary. Consider materials that exhibit reflection symmetry with respect to the plane Π, i.e., $F(\boldsymbol{C}) = F(\boldsymbol{R}^t \boldsymbol{C} \boldsymbol{R})$ with $\boldsymbol{R} = \boldsymbol{I} - 2\boldsymbol{k} \otimes \boldsymbol{k}$, in which $\boldsymbol{I}$ is the three-dimensional identity. Thus,

$$\boldsymbol{R}^t \boldsymbol{C} \boldsymbol{R} = \boldsymbol{c} - \boldsymbol{\gamma} \otimes \boldsymbol{k} - \boldsymbol{k} \otimes \boldsymbol{\gamma} + \left(d_n^2 + \boldsymbol{e} \cdot \boldsymbol{c}\boldsymbol{e}\right) \boldsymbol{k} \otimes \boldsymbol{k}, \tag{18}$$

and so reflection symmetry implies that G is an even function: $G(\boldsymbol{e}) = G(-\boldsymbol{e})$. It follows that there is a function S such that $G(\boldsymbol{e}) = S(\boldsymbol{E})$, where $\boldsymbol{E} = \boldsymbol{e} \otimes \boldsymbol{e}$ (see the Appendix). Accordingly, $G_{\boldsymbol{e}} = 2(S_{\boldsymbol{E}})\boldsymbol{e}$ and the stationarity condition is satisfied if $\boldsymbol{e} = \boldsymbol{0}$; equation (15) then reduces to

$$\boldsymbol{C} = \boldsymbol{c} + d_n^2 \boldsymbol{k} \otimes \boldsymbol{k}, \tag{19}$$

and the cytoskeletal energy is determined by $\boldsymbol{c}$ and d_n:

$$U_c = F\left(\boldsymbol{c} + d_n^2 \boldsymbol{k} \otimes \boldsymbol{k}\right). \tag{20}$$

This is stationary with respect to d_n (> 0) if and only if

$$\boldsymbol{k} \cdot (F_{\boldsymbol{C}}) \boldsymbol{k} = 0, \tag{21}$$

which fixes d_n in terms of $\boldsymbol{c}$.

As we are concerned with equilibria, it is appropriate to confine attention to deformations $\boldsymbol{F}$ that satisfy the strong-ellipticity condition; that is, to deformations satisfying

$$\boldsymbol{a} \otimes \boldsymbol{b} \cdot (\mathcal{W}_c)_{\boldsymbol{F}\boldsymbol{F}} [\boldsymbol{a} \otimes \boldsymbol{b}] > 0, \tag{22}$$

for all $\boldsymbol{a}\otimes\boldsymbol{b}\neq\boldsymbol{0}$. In these circumstances the stationarity conditions have unique solutions that minimize the energy absolutely [Steigmann 2010].

3.2. ***The lipid bilayer.*** We model the lipid bilayer as a nematic liquid crystal described by Frank's energy (see [Virga 1994, (3.63)])

$$\mathcal{W}_b(\boldsymbol{n},\boldsymbol{D})=k_1(\operatorname{tr}\boldsymbol{D})^2+k_2\big(\boldsymbol{W}(\boldsymbol{n})\cdot\boldsymbol{D}\big)^2+k_3|\boldsymbol{D}\boldsymbol{n}|^2+(k_2+k_4)[\operatorname{tr}(\boldsymbol{D}^2)-(\operatorname{tr}\boldsymbol{D})^2], \quad (23)$$

where k_1–k_4 are constants satisfying Ericksen's inequalities

$$2k_1\geq k_2+k_4,\quad k_2\geq|k_4|\quad\text{and}\quad k_3\geq 0, \quad (24)$$

in accordance with the assumed positive semidefiniteness of $\mathcal{W}_b(\boldsymbol{n},\cdot)$, and $\boldsymbol{W}(\boldsymbol{n})$ is the skew tensor with axial vector $\boldsymbol{n}$, i.e., $\boldsymbol{W}(\boldsymbol{n})\boldsymbol{v}=\boldsymbol{n}\times\boldsymbol{v}$ for all $\boldsymbol{v}$. Then, with (10) and (11), we have

$$\boldsymbol{W}(\boldsymbol{n})\cdot\boldsymbol{D}=\boldsymbol{\eta}\cdot\boldsymbol{W}(\boldsymbol{n})\boldsymbol{n}-\boldsymbol{W}(\boldsymbol{n})\cdot\boldsymbol{b}=0, \quad (25)$$

on account of the symmetry of $\boldsymbol{b}$.

Further,

$$\operatorname{tr}\boldsymbol{D}=-2H,\quad\text{where}\quad H=\tfrac{1}{2}\operatorname{tr}\boldsymbol{b}, \quad (26)$$

is the mean curvature of π_b. Combining

$$\boldsymbol{D}^2=\boldsymbol{b}^2-\boldsymbol{b}\boldsymbol{\eta}\otimes\boldsymbol{n}, \quad (27)$$

with the Cayley–Hamilton formula

$$\boldsymbol{b}^2=2H\boldsymbol{b}-K\boldsymbol{1},\quad\text{where}\quad K=\det\boldsymbol{b}, \quad (28)$$

is the Gaussian curvature of π_b and $\boldsymbol{1}=\boldsymbol{I}-\boldsymbol{n}\otimes\boldsymbol{n}$ is the (two-dimensional) identity on its local tangent plane, we arrive at

$$\operatorname{tr}(\boldsymbol{D}^2)=\operatorname{tr}(\boldsymbol{b}^2)=4H^2-2K. \quad (29)$$

Lastly, $\boldsymbol{D}\boldsymbol{n}=\boldsymbol{\eta}$ so that, altogether,

$$\mathcal{W}_b(\boldsymbol{n},\boldsymbol{D})=kH^2+\bar{k}K+k_3|\boldsymbol{\eta}|^2, \quad (30)$$

with

$$k=4k_1\quad\text{and}\quad\bar{k}=-2(k_2+k_4). \quad (31)$$

For k_3 nonzero this is stationary with respect to $\boldsymbol{\eta}$ at $\boldsymbol{\eta}=\boldsymbol{0}$, and so we recover the classical Canham–Helfrich energy [Helfrich 1973; Canham 1970]

$$U_b=kH^2+\bar{k}K, \quad (32)$$

for lipid bilayers, which of course covers the possibility that k_3 vanishes. For $k_3>0$, it is clear that (32) furnishes the minimum of (30).

It is well known that the term in square brackets in (23) is a null Lagrangian in three-dimensional liquid-crystal theory [Virga 1994]. This term is proportional to K, a null Lagrangian in the two-dimensional theory of lipid bilayers. Moreover, in this theory it is customary to model a possible asymmetry in bending response by introducing a variable C, the *spontaneous curvature*, via the modified energy [Ou-Yang et al. 1999]

$$U_b = k(H - C)^2 + \bar{k}K. \tag{33}$$

There are a number of physical effects that can give rise to a spontaneous curvature. Examples include diffusion of transmembrane proteins [Agrawal and Steigmann 2011] and flexoelectricity [Ou-Yang et al. 1999]. One of our objectives in this work is to demonstrate that a conforming cytoskeletal membrane effectively mimics a spontaneous curvature under certain conditions.

With reference to (3) and (4), the net leading-order composite energy is

$$E = \int_{\Pi} W\, dA, \tag{34}$$

where

$$W = W_b(H, K) + W_c(\boldsymbol{c}), \tag{35}$$

with

$$W_b(H, K) = \kappa H^2 + \bar{\kappa} K \quad \text{and} \quad W_c(\boldsymbol{c}) = (1-\alpha)\, F\big(\boldsymbol{c} + d_n^2(\boldsymbol{c})\, \boldsymbol{k} \otimes \boldsymbol{k}\big), \tag{36}$$

and with $\kappa = \alpha k$ and $\bar{\kappa} = \alpha \bar{k}$.

We adopt the conventional assumption [Evans and Skalak 1980] that deformations of the bilayer/cytoskeleton system conserve local surface area. This assumption is invoked for both the bilayer and cytoskeleton separately. For bilayers it is justified by bulk incompressibility in the parent theory of liquid crystals and by the suppression of lipid tilt. The presumed inextensibility of the lipids — expressed by the condition $|\boldsymbol{n}| = 1$ — then implies areal incompressibility. For the cytoskeleton it is justified by empirical evidence [Evans and Skalak 1980] indicating that areal compressibility of the bilayer/cytoskeleton system is typically negligible; areal incompressibility, in the case of a convecting cytoskeleton, then follows from that of the bilayer. Here we impose areal incompressibility of the cytoskeleton whether or not it convects with the bilayer (for a discussion of this issue, see [Krishnaswamy 1996]). Accordingly, the referential areal energy density W is also the areal density in the current configuration of the system in the sense that

$$E = \int_{\pi_b} W_b\, da + \int_{\pi_c} W_c\, da, \tag{37}$$

where $\pi_b \subset \omega$ and $\pi_c \subset \omega$ respectively are the images of Π under the bilayer and cytoskeletal deformations.

4. Material symmetry

4.1. ***The cytoskeleton.*** Little if anything is known about the symmetry group for the cytoskeleton, regarded as a three-dimensional continuum. However, on the basis of work reported in [Pan et al. 2018] we assume that the *two-dimensional* response of the cytoskeletal membrane exhibits hexatropic symmetry relative to the plane configuration Π, characterized by mechanically equivalent unit vectors $\boldsymbol{i}_1$, $\boldsymbol{i}_2$ and $\boldsymbol{i}_3$ aligned with the filaments of the cytoskeleton (Figure 2).

Thus the function $W_c(\boldsymbol{c})$ is assumed to be such that [Cohen and Wang 1984]

$$W_c(\boldsymbol{c}) = W_c(\boldsymbol{R}^t \boldsymbol{c} \boldsymbol{R}), \tag{38}$$

for all two-dimensional orthogonal $\boldsymbol{R}$ belonging to the hexatropic symmetry group. This group is characterized in [Zheng et al. 1992], where it is proved that the list $\{\operatorname{tr}\boldsymbol{c}, \operatorname{tr}(\boldsymbol{c}^2), \operatorname{tr}(\boldsymbol{h}_c\boldsymbol{c})\}$ is a function basis for hexatropic symmetry [Zheng et al. 1992, Table 1], with

$$\boldsymbol{h}_c = [(\boldsymbol{m}\cdot\boldsymbol{c})^2 - (\boldsymbol{m}'\cdot\boldsymbol{c})^2]\,\boldsymbol{m} - 2(\boldsymbol{m}\cdot\boldsymbol{c})(\boldsymbol{m}'\cdot\boldsymbol{c})\,\boldsymbol{m}', \tag{39}$$

in which the interposed dot is the inner product on the translation space Π' of Π, and

$$\boldsymbol{m} = \boldsymbol{e}_1\otimes\boldsymbol{e}_1 - \boldsymbol{e}_2\otimes\boldsymbol{e}_2, \quad \boldsymbol{m}' = \boldsymbol{e}_1\otimes\boldsymbol{e}_2 + \boldsymbol{e}_2\otimes\boldsymbol{e}_1, \tag{40}$$

with

$$\boldsymbol{e}_1 = \boldsymbol{i}_1 \quad\text{and}\quad \boldsymbol{e}_2 = (\boldsymbol{i}_2 - \boldsymbol{i}_3)/\sqrt{3}. \tag{41}$$

Alternatively, the Cayley–Hamilton formula yields the equivalent function basis $\{\operatorname{tr}\boldsymbol{c}, J_c, \operatorname{tr}(\boldsymbol{h}_c\boldsymbol{c})\}$ in which $J_c = 1$ by virtue of areal incompressibility. We suppress a possible explicit dependence of the strain energy on the material point $\boldsymbol{\xi}\in\Pi$ due to any nonuniformity of the material properties or of the orientation of the triad $\{\boldsymbol{i}_k\}$.

According to prevailing opinion [Evans and Skalak 1980; Tartibi et al. 2015], the cytoskeletal membrane exhibits response that is characteristic of an isotropic

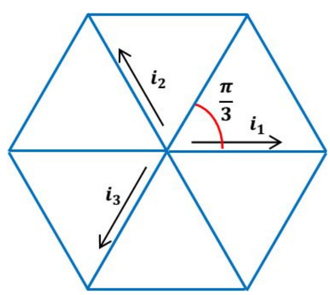


Figure 2. Hexagonal substructure of the cytoskeletal network.

material. This view must be qualified by the membrane-theoretic version of Noll's rule giving the symmetry group relative to any configuration when that relative to one of them is known, i.e., the membrane, if isotropic relative to one configuration, cannot be isotropic relative to all. Here, to avoid ambiguity, we interpret prevailing opinion as implying isotropy relative to Π and thus do not include $\operatorname{tr}(\boldsymbol{h}_c \boldsymbol{c})$ among the arguments of the strain-energy function. Thus we assume

$$W_c(\boldsymbol{c}) = \varpi(I), \quad \text{where} \quad I = \operatorname{tr}\boldsymbol{c}, \tag{42}$$

for some function $\varpi(\cdot)$. Naturally, the symmetry group is thereby enlarged to the orthogonal group. However, hexatropy may be reconciled with isotropy if the strain $\boldsymbol{\epsilon}$, defined by $2\boldsymbol{\epsilon} = \boldsymbol{c} - \mathbf{1}_\Pi$, where $\mathbf{1}_\Pi$ is the identity on Π', is sufficiently small.

Hexatropy implies that the strain energy, expressed as a function of the strain, has as arguments the elements of the function basis $\{\operatorname{tr}\boldsymbol{\epsilon}, \operatorname{tr}(\boldsymbol{\epsilon}^2), \operatorname{tr}(\boldsymbol{h}_\epsilon \boldsymbol{\epsilon})\}$, where $\boldsymbol{h}_\epsilon$ is defined by (39) with $\boldsymbol{c}$ replaced by $\boldsymbol{\epsilon}$. This function basis is approximated at quadratic order in $\boldsymbol{\epsilon}$ by the basis $\{\operatorname{tr}\boldsymbol{\epsilon}, \operatorname{tr}(\boldsymbol{\epsilon}^2)\}$ for isotropy. Thus, the view expressed in the literature is consistent with the substructure of the cytoskeletal network if terms through quadratic order in $\boldsymbol{\epsilon}$ are retained in the strain-energy function. Indeed, quadratic-order energies figure prominently in Evans' and Skalak's extensive treatment [Evans and Skalak 1980] of cytoskeletal membranes in which isotropy is assumed at the outset.

4.2. ***The bilayer.*** The bilayer energy may also be interpreted in the framework of material symmetry. It is known, in the case of areal incompressibility [Steigmann 2003; Zheng 2003], that any function of the mean and Gaussian curvatures H and K may be expressed as a function, B say, of $\boldsymbol{c} = (\nabla\boldsymbol{r})^t(\nabla\boldsymbol{r})$ and the bending strain $\boldsymbol{\kappa} = (\nabla\boldsymbol{r})^t \boldsymbol{b}(\nabla\boldsymbol{r})$, where $\boldsymbol{r}(\boldsymbol{\xi})$ is the bilayer deformation, provided that

$$B(\boldsymbol{c}, \boldsymbol{\kappa}) = B(\boldsymbol{R}^t\boldsymbol{c}\boldsymbol{R}, \pm\boldsymbol{R}^t\boldsymbol{\kappa}\boldsymbol{R}), \tag{43}$$

for all two-dimensional unimodular $\boldsymbol{R}$ ($|\det\boldsymbol{R}| = 1$), with the sign chosen in accordance with that of $\det\boldsymbol{R}$. Here the minus sign is associated with the reflection symmetry of bilayers. This restriction has its origins in Murdoch's and Cohen's extension [Murdoch and Cohen 1979] of Noll's concept [Noll 1958] of material symmetry to elastic surfaces, and comports with his use of the concept of material symmetry [Noll 2004] in the interpretation of the constitutive response of liquid crystals.

5. Surface differential geometry

A configuration of the bilayer/cytosleletal system occupies a surface ω, which we parametrize as $\boldsymbol{r}(\theta^\alpha)$ in which θ^α, $\alpha = 1, 2$, are surface coordinates. The surface parametrization induces the tangent basis $\{\boldsymbol{a}_\alpha\}$, where $\boldsymbol{a}_\alpha = \boldsymbol{r}_{,\alpha}$; the (invertible) surface metric $a_{\alpha\beta} = \boldsymbol{a}_\alpha \cdot \boldsymbol{a}_\beta$; the dual metric $a^{\alpha\beta}$, where $(a^{\alpha\beta}) = (a_{\alpha\beta})^{-1}$; and the dual tangent basis $\{\boldsymbol{a}^\alpha\}$, with $\boldsymbol{a}^\alpha = a^{\alpha\beta}\boldsymbol{a}_\beta$. The orientation of ω is specified by the unit-normal field $\boldsymbol{n}$ defined by $\varepsilon_{\alpha\beta}\,\boldsymbol{n} = \boldsymbol{a}_\alpha \times \boldsymbol{a}_\beta$, where $\varepsilon_{\alpha\beta} = \sqrt{a}\, e_{\alpha\beta}$, with $a = \det(a_{\alpha\beta})$, is the Levi–Civita alternating tensor and $e_{\alpha\beta}$ the permutation symbol ($e_{12} = -e_{21} = 1,\ e_{11} = e_{22} = 0$).

Central to our development are the Gauss and Weingarten equations [Ciarlet 2005; Naghdi 1972]

$$\boldsymbol{r}_{;\alpha\beta} = b_{\alpha\beta}\,\boldsymbol{n} \quad \text{and} \quad \boldsymbol{n}_{,\alpha} = -b_{\alpha\beta}\,\boldsymbol{a}^\beta, \tag{44}$$

respectively, where

$$\boldsymbol{r}_{;\alpha\beta} = \boldsymbol{r}_{,\alpha\beta} - \Gamma^\lambda_{\alpha\beta}\,\boldsymbol{r}_{,\lambda}, \tag{45}$$

is the (symmetric) second covariant derivative of the surface position field. Here $\Gamma^\lambda_{\alpha\beta}$ are the Levi–Civita connection coefficients and $b_{\alpha\beta}$ are the coefficients of the second fundamental form on ω; these are symmetric with respect to interchange of the subscripts, and the latter induce the curvature tensor

$$\boldsymbol{b} = b_{\alpha\beta}\,\boldsymbol{a}^\alpha \otimes \boldsymbol{a}^\beta. \tag{46}$$

The surfacial gradient of the field $\boldsymbol{n}$ is $\nabla_s \boldsymbol{n} = \boldsymbol{n}_{,\alpha} \otimes \boldsymbol{a}^\alpha$, in accordance with (11) and $(44)_2$. Here the connection coefficients are simply the Christoffel symbols and the connection is therefore metric compatible, i.e., the covariant derivatives of the metric components vanish.

The mean and Gaussian curvatures of ω are (see $(26)_2$ and $(28)_2$)

$$H = \tfrac{1}{2} a^{\alpha\beta}\, b_{\alpha\beta} \quad \text{and} \quad K = \tfrac{1}{2}\varepsilon^{\alpha\beta}\,\varepsilon^{\lambda\mu}\, b_{\alpha\lambda}\, b_{\beta\mu}, \tag{47}$$

respectively, where $\varepsilon^{\alpha\beta} = e^{\alpha\beta}/\sqrt{a}$, with $e^{\alpha\beta} = e_{\alpha\beta}$, is the contravariant alternator, and we note the relation

$$b^\beta_\mu\, \tilde{b}^{\mu\alpha} = K a^{\beta\alpha}, \tag{48}$$

where $b^\beta_\mu = a^{\beta\alpha}\, b_{\alpha\mu}$ and

$$\tilde{b}^{\alpha\beta} = \varepsilon^{\alpha\lambda}\,\varepsilon^{\beta\mu}\, b_{\lambda\mu}, \tag{49}$$

is the cofactor of the curvature, expressible as

$$\tilde{b}^{\alpha\beta} = 2H a^{\alpha\beta} - b^{\alpha\beta}, \tag{50}$$

this following on use of the identity

$$\varepsilon^{\alpha\lambda}\,\varepsilon^{\beta\mu} = a^{\alpha\beta}\, a^{\lambda\mu} - a^{\alpha\mu}\, a^{\beta\lambda}. \tag{51}$$

The Mainardi–Codazzi equations of surface theory are $b_{\lambda\mu;\beta} = b_{\lambda\beta;\mu}$ [Ciarlet 2005], or, more concisely, $\varepsilon^{\beta\mu} b_{\lambda\mu;\beta} = 0$. The metric compatibility of the connection implies that the covariant derivatives of $\varepsilon^{\alpha\lambda}$ vanish and the Mainardi–Codazzi equations are therefore equivalent to

$$\tilde{b}^{\alpha\beta}_{;\beta} = 0. \tag{52}$$

5.1. *Convected coordinates and surface-fixed coordinates.* The literature on lipid bilayers relies exclusively on the use of surface-fixed coordinates in the analysis of the so-called *shape equation* (see [Ou-Yang et al. 1999] for example). This formalism is entirely analogous to the spatial description of continuum mechanics in which problems are posed on a suitably parametrized fixed region of space. However, as in the latter setting, while this description often affords advantages in the solution of problems, it is a conceptual obstacle to the formulation of theories concerning material bodies. For the latter, convected coordinates that label material points furnish the appropriate alternative.

We encounter precisely the same issue in the mechanics of material surfaces, and thus pause to outline the distinction between parametrizations based on surface-fixed coordinates — analogous to the spatial coordinates of conventional continuum mechanics — and those based on convected coordinates. The relevant developments are due to Scriven [1960] and summarized in Chapter 10 of Aris' book [Aris 1989]. We present the main ideas in the present subsection for the sake of completeness.

Consider configurations of a surface regarded as a material manifold parametrized by a convected coordinate system ξ^α. This may be identified with the system θ^α of the previous subsection at the value $\epsilon = 0$, say, of a time-like parameter ϵ in a one-parameter family of configurations. The associated surface Ω, with parametric representation $\hat{\boldsymbol{r}}(\xi^\alpha)$, is fixed and may serve as a reference surface in a referential description of the motion. That is, we regard these coordinates as being convected in the sense that they identify, via a map $\boldsymbol{r} = \hat{\boldsymbol{r}}(\xi^\alpha, \epsilon)$, the position, associated with parameter value ϵ, of a material point occupying position $\hat{\boldsymbol{r}}(\xi^\alpha) \in \Omega$ at $\epsilon = 0$. This notion may be generalized by regarding Ω as a surface that is in one-to-one correspondence with that occupied at $\epsilon = 0$, so that it need not actually be occupied in the course of the deformation. The connection with the θ^α-parametrization of ω is provided by

$$\hat{\boldsymbol{r}}(\xi^\alpha, \epsilon) = \boldsymbol{r}\big(\theta^\alpha(\xi^\beta, \epsilon), \epsilon\big). \tag{53}$$

Thus we specify the fixed surface coordinates θ^α as functions of ξ^α and ϵ subject to $\theta^\alpha(\xi^\beta, 0) = \xi^\alpha$. We assume the relations giving θ^α in terms of ξ^α to be invertible, to reflect the notion that at fixed ϵ the coordinates θ^α can be associated with a unique material point (identified by fixed values of ξ^α). Any function, $f(\theta^\alpha, \epsilon)$, say, may then be expressed in terms of convected coordinates as $\hat{f}(\xi^\alpha, \epsilon)$, where

$$\hat{f}(\xi^\alpha, \epsilon) = f\big(\theta^\alpha(\xi^\beta, \epsilon), \epsilon\big). \tag{54}$$

The variational derivative of f is its partial derivative with respect to ϵ in the convected-coordinate representation, i.e., $\dot{f} = \partial \hat{f}(\xi^\alpha, \epsilon)/\partial\epsilon$, whereas its derivative in the fixed-coordinate parametrization is $f_\epsilon = \partial f(\theta^\alpha, \epsilon)/\partial\epsilon$; these are related by $\dot{f} = f_\epsilon + (\theta^\alpha)^\cdot f_{,\alpha}$.

The ϵ-velocity of a material point on Ω that has been convected by the deformation to ω is $\boldsymbol{u} = \dot{\boldsymbol{r}} = \partial\hat{\boldsymbol{r}}/\partial\epsilon$. We may write this in terms of components on the natural basis induced by the fixed-coordinate θ^α-parametrization:

$$\boldsymbol{u} = u^\alpha \boldsymbol{a}_\alpha + w\boldsymbol{n}. \tag{55}$$

This is related to the derivative $\boldsymbol{r}_\epsilon$ by

$$\boldsymbol{u} = (\theta^\alpha)^\cdot \boldsymbol{a}_\alpha + \boldsymbol{r}_\epsilon. \tag{56}$$

Following [Aris 1989; Scriven 1960] we adopt the fixed-coordinate parametrization defined by

$$\frac{d}{d\epsilon}\theta^\alpha = u^\alpha(\theta^\beta, \epsilon), \quad \theta^\alpha_{|\epsilon=0} = \xi^\alpha, \tag{57}$$

where the derivative is evaluated at fixed $\{\xi^\alpha\}$ and hence equal to $(\theta^\alpha)^\cdot$. The normal virtual velocity in (55) is then given by

$$w\boldsymbol{n} = \boldsymbol{r}_\epsilon, \tag{58}$$

and the convected and fixed-coordinate derivatives satisfy

$$\dot{f} = f_\epsilon + u^\alpha f_{,\alpha}. \tag{59}$$

We require the Lie derivative of the metric with respect to the velocity. This is simply the variational derivative $\dot{a}_{\alpha\beta}$ expressed in terms of the θ^α-parametrization. To this end we adopt convected coordinates ξ^α whose values coincide with θ^α at $\epsilon = 0$. The two sets of coordinate systems will of course differ at different values of ϵ due to the fact that material is moving with respect to the θ^α-system. Said differently, the material point located at the place with surface coordinates θ^α at $\epsilon = 0$ will have different locations at different values of ϵ and hence be associated with different values of θ^α, whereas the values of ξ^α remain invariant. Accordingly, while it is always permissible to identify ξ^α with θ^α at $\epsilon = 0$, say, it is not possible to do so over an interval of ϵ values. However, for our purposes this limitation is not restrictive. Using $\dot{a}_{\lambda\mu} = \dot{\boldsymbol{a}}_\lambda \cdot \boldsymbol{a}_\mu + \boldsymbol{a}_\lambda \cdot \dot{\boldsymbol{a}}_\mu$ and

$$\dot{\boldsymbol{a}}_\lambda = \left(\frac{\partial \boldsymbol{r}}{\partial\theta^\lambda}\right)^\cdot = \left[\frac{\partial \boldsymbol{r}}{\partial\xi^\mu}\left(\frac{\partial\xi^\mu}{\partial\theta^\lambda}\right)\right]^\cdot = \frac{\partial \boldsymbol{u}}{\partial\xi^\mu}\left(\frac{\partial\xi^\mu}{\partial\theta^\lambda}\right) + \frac{\partial \boldsymbol{r}}{\partial\xi^\mu}\left(\frac{\partial^2\xi^\mu}{\partial\theta^\lambda\,\partial\theta^\alpha}\right)u^\alpha, \tag{60}$$

together with $\partial\xi^\mu/\partial\theta^\lambda = \delta^\mu_\lambda$ (the Kronecker delta) and hence $\partial^2\xi^\mu/\partial\theta^\lambda\,\partial\theta^\alpha = 0$ *at* $\epsilon = 0$, we derive $\dot{\boldsymbol{a}}_\alpha = \partial\boldsymbol{u}/\partial\xi^\alpha$ and

$$\dot{a}_{\lambda\mu} = \boldsymbol{u}_{,\lambda}\cdot\boldsymbol{a}_\mu + \boldsymbol{a}_\lambda\cdot\boldsymbol{u}_{,\mu}, \tag{61}$$

where $\boldsymbol{u}_{,\lambda} = \partial\boldsymbol{u}/\partial\theta^\lambda$ *at* $\epsilon = 0$.

Combining (55) with the Gauss and Weingarten equations yields

$$\boldsymbol{u}_{,\lambda} = (u_{\alpha;\lambda} - wb_{\alpha\lambda})\,\boldsymbol{a}^\alpha + (u^\alpha\, b_{\alpha\lambda} + w_{,\lambda})\,\boldsymbol{n}, \tag{62}$$

where $\boldsymbol{a}^\alpha = a^{\alpha\beta}\boldsymbol{a}_\beta$ and $u_{\alpha;\lambda}$ is the covariant derivative defined by

$$u_{\alpha;\lambda} = u_{\alpha,\lambda} - u_\beta\,\Gamma^\beta_{\alpha\lambda}, \tag{63}$$

in which $\Gamma^\beta_{\alpha\lambda}$ are the connection symbols on ω pertaining to the induced metric in the θ^α-system. Hence the desired expression:

$$\dot{a}_{\lambda\mu} = u_{\mu;\lambda} + u_{\lambda;\mu} - 2w\,b_{\lambda\mu}. \tag{64}$$

For example, if $A_{\alpha\beta}$ is the (fixed) metric on the surface Ω induced by the parametrization $\hat{\boldsymbol{r}}(\xi^\alpha)$, then the areal stretch induced by the deformation is $J = \sqrt{a/A}$, where $A = \det(A_{\alpha\beta})$. The fact that the cofactor of $a_{\alpha\beta}$ is $(a)a^{\alpha\beta}$ then implies

$$\dot{J}/J = \tfrac{1}{2}a^{\alpha\beta}\,\dot{a}_{\alpha\beta}, \tag{65}$$

and with (61) this may be reduced to

$$\dot{J}/J = \boldsymbol{a}^\alpha\cdot\boldsymbol{u}_{,\alpha}. \tag{66}$$

5.2. ***Congruent configurations of the bilayer and cytoskeleton.*** This formalism may be adapted to the bilayer/cytoskeleton system by introducing one-parameter families, $\hat{\boldsymbol{r}}_c(\xi^\alpha;\epsilon_c)$ and $\hat{\boldsymbol{r}}_b(\eta^\alpha;\epsilon_b)$ of cytoskeleton and bilayer deformations respectively, in which ξ^α and η^α are convected coordinates. The surface-fixed coordinates on the cytoskeleton and bilayer are $\theta^\alpha_{(c)}(\xi^\alpha;\epsilon_c)$ and $\theta^\alpha_{(b)}(\eta^\beta;\epsilon_b)$, respectively. Congruency then implies that (see (53))

$$\hat{\boldsymbol{r}}_c(\xi^\alpha;\epsilon_c) = \boldsymbol{r}\big(\theta^\alpha_{(c)}(\xi^\beta;\epsilon_c),\epsilon_c\big) \quad\text{and}\quad \hat{\boldsymbol{r}}_b(\eta^\alpha;\epsilon_b) = \boldsymbol{r}\big(\theta^\alpha_{(b)}(\eta^\beta;\epsilon_b),\epsilon_b\big), \tag{67}$$

where $\boldsymbol{r}(\theta^\alpha,\epsilon)$ is the surface-fixed parametrization of ω.

We stipulate that $\xi^\alpha = \theta^\alpha_{(c)}(\xi^\beta;0)$ and $\eta^\alpha = \theta^\alpha_{(b)}(\eta^\beta;0)$; further, that $\theta^\alpha_{(c)}(\xi^\beta;0) = \theta^\alpha_{(b)}(\eta^\beta;0) = \theta^\alpha$, so that

$$\hat{\boldsymbol{r}}_b(\eta^\alpha;0) = \boldsymbol{r}(\theta^\alpha) = \hat{\boldsymbol{r}}_c(\xi^\alpha;0), \tag{68}$$

where, for the sake of brevity, we write $\boldsymbol{r}(\theta^\alpha)$ in place of $\boldsymbol{r}(\theta^\alpha, 0)$. In this way we construct convected coordinates ξ^α and η^α that coincide, *at* $\epsilon_c, \epsilon_b = 0$, with specified surface-fixed coordinates θ^α on ω. This is tantamount to adopting the place $\boldsymbol{r}(\theta^\alpha)$ occupied by material points of the bilayer (at $\epsilon_b = 0$) and cytoskeleton (at $\epsilon_c = 0$) as their common reference position.

With reference to $(57)_1$ we define the tangential *virtual velocities*

$$u^\alpha = \frac{d}{d\epsilon_c}\theta^\alpha_{(c)|\epsilon_c=0} \quad \text{and} \quad v^\alpha = \frac{d}{d\epsilon_b}\theta^\alpha_{(b)|\epsilon_b=0}, \tag{69}$$

of the cytoskeleton and bilayer, respectively, and assume, in keeping with congruency, that the normal virtual velocities have a common value, w say:

$$\left.\frac{\partial \boldsymbol{r}}{\partial \epsilon_b}\right|_{\epsilon_b=0} = \left.\frac{\partial \boldsymbol{r}}{\partial \epsilon_c}\right|_{\epsilon_c=0} = w\boldsymbol{n}, \tag{70}$$

(see (58)). Then the virtual velocities of the bilayer and cytoskeleton are

$$\boldsymbol{u}(\theta^\alpha) = \dot{\boldsymbol{r}}_b = u^\alpha \boldsymbol{a}_\alpha + w\boldsymbol{n}, \tag{71}$$

and

$$\boldsymbol{v}(\theta^\alpha) = \dot{\boldsymbol{r}}_c = v^\alpha \boldsymbol{a}_\alpha + w\boldsymbol{n}, \tag{72}$$

respectively, where

$$\dot{\boldsymbol{r}}_b = \left.\frac{\partial \hat{\boldsymbol{r}}_b}{\partial \epsilon_b}\right|_{\epsilon_b=0} \quad \text{and} \quad \dot{\boldsymbol{r}}_c = \left.\frac{\partial \hat{\boldsymbol{r}}_c}{\partial \epsilon_c}\right|_{\epsilon_c=0}. \tag{73}$$

The identification of $\boldsymbol{n}\cdot\boldsymbol{u}$ with $\boldsymbol{n}\cdot\boldsymbol{v}$ also features in a model proposed in [Herant and Dembo 2006].

The formula (64) for the variation of the surface metric applies as it stands to the cytoskeleton if the superposed dot is interpreted as a derivative with respect to ϵ_c (evaluated at $\epsilon_c = 0$). It also applies to the bilayer if the superposed dot is interpreted as a derivative with respect to ϵ_b (evaluated at $\epsilon_b = 0$), with v_μ substituted in place of u_μ.

To interpret the cytoskeletal deformation tensor $\nabla \boldsymbol{r}_c$ (see $(7)_2$) in this framework, let the patch Π be parametrized in the form $\boldsymbol{\xi}(\xi^\alpha)$. This parametrization induces the tangent basis $\boldsymbol{A}_\alpha = \boldsymbol{\xi}_{,\alpha}$, metric $A_{\alpha\beta} = \boldsymbol{A}_\alpha \cdot \boldsymbol{A}_\beta$, dual metric $A^{\alpha\beta}$, and dual basis $\boldsymbol{A}^\alpha$. Then,

$$\nabla \boldsymbol{r}_c = \boldsymbol{a}_\alpha \otimes \boldsymbol{A}^\alpha, \tag{74}$$

and the surfacial Cauchy–Green deformation tensor is

$$\boldsymbol{c} = a_{\alpha\beta}\, \boldsymbol{A}^\alpha \otimes \boldsymbol{A}^\beta. \tag{75}$$

The areal dilation induced by the deformation is

$$J_c = \sqrt{\det \boldsymbol{c}} = \sqrt{a/A}. \tag{76}$$

6. Energy, virtual power and equilibrium

6.1. ***Energy and power.*** To obtain equilibrium equations and edge conditions we invoke the virtual-power principle for the simply-connected patch Π. We account for areal incompressibility by extending the energy to unconstrained states and introducing appropriate Lagrange-multiplier fields. Reference may be made to Section 5.10 of [Berdichevsky 2009], for example, for an exposition of this idea together with some of its applications to continuum mechanics. From (34)–(37), the extended energy of the patch is

$$E = \int_{\Pi} [J_b W_b + J_c W_c + \lambda_b (J_b - 1) + \lambda_c (J_c - 1)]\, dA + \int_{\partial \Pi} \tilde{\mu} (J_b - 1)\, dS, \quad (77)$$

where $\lambda_{b,c}$ and $\tilde{\mu}$ are Lagrange multiplier fields. We have included a multiplier on the boundary because, as we show below, the tangential and normal derivatives of the virtual bilayer velocity $\boldsymbol{v}$, which figure in the expression for the variation of the energy, are constrained by areal incompressibility. To our knowledge this effect has not been discussed in the literature on bilayers. However, similar terms are known to play a role in the mechanics of continua of second grade [Guven et al. 2019; Steigmann 2018; Wang and Pipkin 1986] — as exemplified by lipid bilayers — in the presence of constraints on the first-order gradients.

Having proposed an expression for the extended energy, we identify equilibria with those states that satisfy

$$\dot{E} = P, \quad (78)$$

where P is the virtual power imparted to the patch. The form that this power takes is deduced in the course of the ensuing development. Here the superposed dot refers to a Gateaux derivative with respect to either ϵ_c or ϵ_b (evaluated at ϵ_c and ϵ_b equal to zero) or to both simultaneously.

6.2. ***Tangential equilibrium of the cytoskeletal membrane.*** For example, consider variations that preserve the bilayer configuration. These are $\boldsymbol{u}(\theta^\alpha) = u^\alpha \boldsymbol{a}_\alpha$ and $\boldsymbol{v} = \boldsymbol{0}$, and yield

$$\dot{E} = \int_{\pi_c} [\dot{W}_c + (W_c + \lambda_c) \dot{J}_c / J_c]\, da, \quad (79)$$

in which variation of λ_c has been suppressed as this merely returns the areal incompressibility constraint. In the extended (unconstrained) formalism, $J_c W_c$ is the cytoskeletal energy density on Π. Thus, in the case of isotropy, for example, we make the identification

$$J_c W_c = \varpi(I), \quad \text{with} \quad I = a_{\lambda\mu} A^{\lambda\mu}, \quad (80)$$

which reduces to (42) when the constraint is in effect. This depends via (75) and (76) on the surfacial Cauchy–Green tensor $\boldsymbol{c}$ and thus evolves in response to variations $\dot{a}_{\alpha\beta}$ of the surface metric. Accordingly, we write

$$(J_c W_c)^{\cdot} = \tfrac{1}{2} J_c \Sigma^{\alpha\beta}\, \dot{a}_{\alpha\beta}, \quad \text{with} \quad \tfrac{1}{2} J_c \Sigma^{\alpha\beta} = (J_c W_c)_{\boldsymbol{c}} \cdot \boldsymbol{A}^{\alpha} \otimes \boldsymbol{A}^{\beta}, \tag{81}$$

which we combine with (65) to obtain

$$\dot{W}_c + (W_c + \lambda_c)\, \dot{J}_c / J_c = \tfrac{1}{2} \sigma^{\alpha\beta}\, \dot{a}_{\alpha\beta}, \quad \text{with} \quad \sigma^{\alpha\beta} = \Sigma^{\alpha\beta} + \lambda_c\, a^{\alpha\beta}. \tag{82}$$

We note that $\Sigma^{\alpha\beta} = \Sigma^{\beta\alpha}$, and thus $\sigma^{\alpha\beta} = \sigma^{\beta\alpha}$, by virtue of the symmetry of $(J_c W_c)_{\boldsymbol{c}}$. For example, in the case of isotropy, we have from (75) and (80) that $(J_c W_c)_{\boldsymbol{c}} = \varpi'(I)\mathbf{1}_{\Pi}$, yielding

$$J_c \Sigma^{\alpha\beta} = 2\varpi'(I) A^{\alpha\beta}. \tag{83}$$

Combining this symmetry with (64) (with $w = 0$) we derive $\frac{1}{2}\sigma^{\alpha\beta}\, \dot{a}_{\alpha\beta} = \sigma^{\alpha\beta} u_{\alpha;\beta}$ and then convert (79) via Stokes' theorem to

$$\dot{E} = \int_{\partial\pi_c} \sigma^{\alpha\beta} \nu_\beta\, u_\alpha\, ds - \int_{\pi_c} \sigma^{\alpha\beta}{}_{;\beta}\, u_\alpha\, da, \tag{84}$$

where $\nu_\beta = \varepsilon_{\beta\alpha}\tau^\alpha$, in which $\tau^\alpha = d\theta^\alpha / ds$ are the components of the rightward unit normal to $\partial\pi_c$ with arclength parametrization $\theta^\alpha(s)$; i.e., $\boldsymbol{\nu} = \boldsymbol{\tau} \times \boldsymbol{n}$, where $\boldsymbol{\tau} = d\boldsymbol{r}(\theta^\alpha(s))/ds$ and $\boldsymbol{n}$ respectively are the unit tangent to $\partial\pi_c$ and the unit surface normal.

From (78) it follows that the virtual power is of the form

$$P = \int_{\partial\pi_c} t^{\alpha}_{(c)}\, u_\alpha\, ds + \int_{\pi_c} g^{\alpha}_{(c)}\, u_\alpha\, da, \tag{85}$$

and, with no further restrictions on u_α, that

$$\sigma^{\alpha\beta}{}_{;\beta} + g^{\alpha}_{(c)} = 0, \quad \text{in} \quad \pi_c \quad \text{and} \quad t^{\alpha}_{(c)} = \sigma^{\alpha\beta}\, \nu_\beta, \quad \text{on} \quad \partial\pi_c, \tag{86}$$

in which $g^{\alpha}_{(c)}$ and $t^{\alpha}_{(c)}$ respectively are the distributed tangential force (per unit area) and the tangential edge traction (force per unit length) acting on the cytoskeleton. From these relations it is clear that $\sigma^{\alpha\beta}$ plays the role of the cytoskeletal Cauchy stress. Equation $(82)_2$ then yields the interpretation of λ_c as a reactive surface tension. Here, to compensate for having suppressed variation with respect to the multiplier λ_c, it is necessary to impose $J_c = 1$ *a posteriori*. Thus, in the case of isotropy, we use $(82)_2$ in (86) with

$$\Sigma^{\alpha\beta} = 2\varpi'(I)\, A^{\alpha\beta}. \tag{87}$$

6.3. ***Variational derivative of the bilayer energy.*** We pause to discuss some formulae of a general nature valid for arbitrary bilayer virtual velocities $\boldsymbol{v}$ and subsequently specialize these to derive the tangential equilibrium equations.

First we note that because J_b and W_b depend on the surface position field through its first and second derivatives with respect to the coordinates, it follows that there are vector fields $\boldsymbol{N}^\alpha$ and $\boldsymbol{M}^{\alpha\beta}$ such that

$$\dot{W}_b + (W_b + \lambda_b)\dot{J}_b/J_b = \boldsymbol{N}^\alpha \cdot \boldsymbol{v}_{,\alpha} + \boldsymbol{M}^{\alpha\beta} \cdot \boldsymbol{v}_{;\alpha\beta}, \tag{88}$$

where $\boldsymbol{v} = \dot{\boldsymbol{r}}_b$ is the virtual velocity and $\boldsymbol{v}_{;\alpha\beta} = \boldsymbol{v}_{,\alpha\beta} - \Gamma^\lambda_{\alpha\beta}\boldsymbol{v}_{,\lambda}$ is the second covariant derivative of $\boldsymbol{v}$. This is symmetric in the subscripts; therefore, no generality is lost by imposing $\boldsymbol{M}^{\alpha\beta} = \boldsymbol{M}^{\beta\alpha}$.

For example [Agrawal and Steigmann 2009],

$$\dot{H} = \tfrac{1}{2}a^{\alpha\beta}\boldsymbol{n} \cdot \boldsymbol{v}_{;\alpha\beta} - b^{\alpha\beta}\boldsymbol{a}_\beta \cdot \boldsymbol{v}_{,\alpha} \quad \text{and} \quad \dot{K} = \tilde{b}^{\alpha\beta}\boldsymbol{n} \cdot \boldsymbol{v}_{;\alpha\beta} - 2K\boldsymbol{a}^\alpha \cdot \boldsymbol{v}_{,\alpha}, \tag{89}$$

whereas (see (66))

$$\dot{J}_b/J_b = \boldsymbol{a}^\alpha \cdot \boldsymbol{v}_{,\alpha}. \tag{90}$$

Using $\dot{W}_b = 2\kappa H\dot{H} + \bar{\kappa}\dot{K}$ (from $(36)_1$) we thus derive

$$\boldsymbol{N}^\mu = N^{\mu\beta}\boldsymbol{a}_\beta \quad \text{and} \quad \boldsymbol{M}^{\mu\beta} = M^{\mu\beta}\boldsymbol{n}, \tag{91}$$

with

$$N^{\mu\beta} = (\lambda_b + \kappa H^2 - \bar{\kappa}K)\,a^{\mu\beta} - 2\kappa H b^{\mu\beta} \quad \text{and} \quad M^{\mu\beta} = \kappa H a^{\mu\beta} + \bar{\kappa}\tilde{b}^{\mu\beta}. \tag{92}$$

Proceeding, we have

$$\boldsymbol{N}^\alpha \cdot \boldsymbol{v}_{,\alpha} + \boldsymbol{M}^{\alpha\beta} \cdot \boldsymbol{v}_{;\alpha\beta} = \varphi^\alpha_{;\alpha} - \boldsymbol{v} \cdot \boldsymbol{T}^\alpha_{;\alpha}, \tag{93}$$

where

$$\boldsymbol{T}^\alpha = \boldsymbol{N}^\alpha - \boldsymbol{M}^{\alpha\beta}_{;\beta}, \tag{94}$$

with

$$\boldsymbol{M}^{\beta\alpha}_{;\beta} = M^{\beta\alpha}_{;\beta}\,\boldsymbol{n} - M^{\beta\alpha}b^\mu_\beta\,\boldsymbol{a}_\mu, \tag{95}$$

and

$$\varphi^\alpha = \boldsymbol{T}^\alpha \cdot \boldsymbol{v} + \boldsymbol{M}^{\alpha\beta} \cdot \boldsymbol{v}_{,\beta}, \tag{96}$$

in which (91), (94) and (95) together give

$$\boldsymbol{T}^\alpha = \left(N^{\alpha\mu} + M^{\alpha\beta}b^\mu_\beta\right)\boldsymbol{a}_\mu - M^{\alpha\beta}_{;\beta}\,\boldsymbol{n}. \tag{97}$$

Combining (88) and (96) with Stokes' theorem furnishes

$$\int_{\pi_b} [\dot{W}_b + (W_b + \lambda_b)\dot{J}_b/J_b]\,da = \int_{\partial\pi_b} \varphi^\alpha \nu_\alpha\,ds - \int_{\pi_b} \boldsymbol{v} \cdot \boldsymbol{T}^\alpha_{;\alpha}\,da, \tag{98}$$

where $\boldsymbol{\nu} = \nu_\alpha \boldsymbol{a}^\alpha$ is the exterior unit normal $\partial\pi_b$ and

$$\boldsymbol{v} \cdot \boldsymbol{T}^\alpha_{;\alpha} = v_\mu \boldsymbol{a}^\mu \cdot \boldsymbol{T}^\alpha_{;\alpha} + w\boldsymbol{n} \cdot \boldsymbol{T}^\alpha_{;\alpha}, \tag{99}$$

with

$$\boldsymbol{a}^\mu \cdot \boldsymbol{T}^\alpha_{;\alpha} = \left(N^{\alpha\mu} + M^{\alpha\beta} b^\mu_\beta\right)_{;\alpha} + M^{\alpha\beta}_{;\beta} b^\mu_\alpha, \tag{100}$$

and

$$\boldsymbol{n} \cdot \boldsymbol{T}^\alpha_{;\alpha} = \left(N^{\alpha\mu} + M^{\alpha\beta} b^\mu_\beta\right) b_{\mu\alpha} - M^{\beta\alpha}_{;\beta\alpha}. \tag{101}$$

In the first term on the right-hand side of (98) we use the normal-tangential decomposition

$$\boldsymbol{v}_{,\beta} = \tau_\beta \boldsymbol{v}' + \nu_\beta \boldsymbol{v}_\nu, \tag{102}$$

where $\boldsymbol{\tau} = \tau_\alpha \boldsymbol{a}^\alpha = \boldsymbol{n} \times \boldsymbol{\nu}$ is the unit tangent to $\partial\pi_b$, $\boldsymbol{v}' = \tau^\alpha \boldsymbol{v}_{,\alpha} = d\boldsymbol{v}/ds$ is the tangential derivative of $\boldsymbol{v}$, and $\boldsymbol{v}_\nu = \nu^\alpha \boldsymbol{v}_{,\alpha}$ is the normal derivative. The term involving the tangential derivative is integrated by parts. If $\partial\pi_b$ is piecewise smooth in the sense that its tangent $\boldsymbol{\tau}$ is piecewise continuous, with discontinuities at a finite number of corners, then

$$\int_{\partial\pi_b} \varphi^\alpha \nu_\alpha \, ds = \int_{\partial\pi_b} \left(\{\boldsymbol{T}^\alpha \nu_\alpha - (\boldsymbol{M}^{\alpha\beta} \nu_\alpha \tau_\beta)'\} \cdot \boldsymbol{v} + \boldsymbol{M}^{\alpha\beta} \nu_\alpha \nu_\beta \cdot \boldsymbol{v}_\nu\right) ds \\ - \sum \boldsymbol{M}^{\alpha\beta} [\nu_\alpha \tau_\beta]_i \cdot \boldsymbol{v}_i, \tag{103}$$

in which the square bracket refers to the forward jump as a corner of the boundary is traversed, and the sum ranges over all corners. Thus, $[\cdot] = (\cdot)_+ - (\cdot)_-$, where the subscripts $_\pm$ respectively identify limits as a corner located at arclength station s is approached through larger and smaller values of arclength.

6.3.1. *Tangential bilayer equilibrium.* Consider variations with $\boldsymbol{v}$ and $\boldsymbol{v}_\nu$ vanishing on $\partial\pi_b$ (and at corners) that preserve the configuration of the cytoskeleton, i.e., $\boldsymbol{u} = \boldsymbol{0}$ and $\boldsymbol{v} = v_\mu \boldsymbol{a}^\mu$ in the interior of π_b. For these we have

$$\dot{E} = \int_{\pi_b} [\dot{W}_b + (W_b + \lambda_b)\dot{J}_b/J_b] \, da = -\int_{\pi_b} v_\mu \boldsymbol{a}^\mu \cdot \boldsymbol{T}^\alpha_{;\alpha} \, da, \tag{104}$$

in which variation of λ_b has been suppressed, and it follows, from (78), that the virtual power is of the form

$$P = \int_{\pi_b} g^\mu_{(b)} v_\mu \, da, \tag{105}$$

where $g^\mu_{(b)}$ is a tangential force (per unit area) acting on the bilayer. Because v_μ is unrestricted, we arrive at

$$\left(N^{\alpha\mu} + M^{\alpha\beta} b^\mu_\beta\right)_{;\alpha} + M^{\alpha\beta}_{;\beta} b^\mu_\alpha + g^\mu_{(b)} = 0, \quad \text{in} \quad \pi_b. \tag{106}$$

To reduce this we use (50), (52) and (92) to infer that

$$N^{\alpha\mu} + M^{\alpha\beta} b^{\mu}_{\beta} = (\lambda_b + \kappa H^2)\, a^{\alpha\mu} - \kappa H b^{\alpha\mu}, \tag{107}$$

with divergence

$$\big(N^{\alpha\mu} + M^{\alpha\beta} b^{\mu}_{\beta}\big)_{;\alpha} = a^{\alpha\mu}(\lambda_b)_{,\alpha} + 2\kappa H a^{\alpha\mu} H_{,\alpha} - \kappa b^{\alpha\mu} H_{,\alpha} - \kappa H b^{\alpha\mu}_{;\alpha}, \tag{108}$$

and combination with (see $(92)_2$)

$$M^{\alpha\beta}_{;\beta} = \kappa a^{\alpha\beta} H_{,\beta}, \tag{109}$$

furnishes

$$\big(N^{\alpha\mu} + M^{\alpha\beta} b^{\mu}_{\beta}\big)_{;\alpha} + M^{\alpha\beta}_{;\beta} b^{\mu}_{\alpha} = a^{\alpha\mu}(\lambda_b)_{,\alpha} + \kappa H \big(2a^{\alpha\mu} H_{,\alpha} - b^{\alpha\mu}_{;\alpha}\big), \tag{110}$$

in which the second parenthetical term on the right is $(2Ha^{\mu\alpha} - b^{\mu\alpha})_{;\alpha} = \tilde{b}^{\mu\alpha}{}_{;\alpha}$. Then, with (52), equation (110) reduces simply to

$$a^{\alpha\mu}(\lambda_b)_{,\alpha} + g^{\mu}_{(b)} = 0. \tag{111}$$

6.3.2. *Comoving bilayer and cytoskeleton.* If the cytoskeleton is anchored to the bilayer such as to convect with it, then $\boldsymbol{u} = \boldsymbol{v}$ in $\pi^* = \pi_b \cap \pi_c$. Choosing variations such that $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{v}_\nu$ vanish on $\partial\pi^*$ and $\boldsymbol{v} = v_\mu \boldsymbol{a}^\mu$ in π^*, with $\boldsymbol{u}$, $\boldsymbol{v}$ vanishing in $\omega \setminus \pi^*$, we obtain

$$\dot{E} = \int_{\pi^*} [\dot{W} + (W + \lambda)\dot{J}/J]\, da, \tag{112}$$

with $W = W_b + W_c$, $\lambda = \lambda_b + \lambda_c$ and $\dot{J}/J = v^{\mu}_{;\mu}$. We could proceed from this statement to derive the relevant balance equation directly, but it is more illuminating to combine (84) and (104) to arrive at

$$\dot{E} = -\int_{\pi^*} \big\{\sigma^{\mu\alpha}_{;\alpha} + a^{\mu\alpha}(\lambda_b)_{,\alpha}\big\}\, v_\mu\, da. \tag{113}$$

The associated virtual power therefore has the form

$$P = \int_{\pi^*} g^{\mu} v_\mu\, da, \tag{114}$$

and with v_μ unrestricted, $(86)_1$ and (111) then deliver

$$g^{\mu} = g^{\mu}_{(b)} + g^{\mu}_{(c)}. \tag{115}$$

Equivalently,

$$(\Sigma^{\mu\alpha} + \lambda a^{\mu\alpha})_{;\alpha} + g^{\mu} = 0, \tag{116}$$

in which the term in parentheses is the effective Cauchy stress for the bilayer/cytoskeleton composite subjected to a net tangential force g^{μ}.

6.4. ***Normal equilibrium of the bilayer and cytoskeleton.*** Having exhausted the consequences of the virtual-power statement for tangential variations, we proceed next to normal variations. In view of (71) and (72), these involve the bilayer and cytoskeleton together. Taking variations as in the previous subsection, now with $\boldsymbol{u} = \boldsymbol{v} = w\boldsymbol{n}$, with reference to (64), $(82)_1$ and (98) we obtain

$$\dot{E} = -\int_{\pi^*} w\big(\sigma^{\alpha\beta} b_{\alpha\beta} + \boldsymbol{n}\cdot\boldsymbol{T}^{\alpha}_{;\alpha}\big)\,da, \tag{117}$$

and conclude that the associated power has the form

$$P = \int_{\pi^*} pw\,da, \tag{118}$$

where p is the net lateral pressure acting on the surface in the direction of $\boldsymbol{n}$. Thus, with (101) and with w unrestricted, we arrive at

$$\sigma^{\alpha\beta} b_{\alpha\beta} + \big(N^{\alpha\mu} + M^{\alpha\beta} b^{\mu}_{\beta}\big)\, b_{\mu\alpha} - M^{\beta\alpha}_{;\beta\alpha} + p = 0, \quad \text{in} \quad \pi^*. \tag{119}$$

To reduce this we use (92), finding that

$$\big(N^{\alpha\mu} + M^{\alpha\beta} b^{\mu}_{\beta}\big)\, b_{\mu\alpha} = 2\lambda_b H + 2\kappa H^3 - \kappa H b^{\alpha\mu} b_{\mu\alpha}. \tag{120}$$

The final term on the right is $b^{\alpha\mu} b_{\mu\alpha} = \mathrm{tr}(\boldsymbol{b}^2)$, and with (29) and (109), equation (119) becomes

$$\kappa[\Delta H + 2H(H^2 - K)] - 2\lambda_b H = \sigma^{\alpha\beta} b_{\alpha\beta} + p, \tag{121}$$

where

$$\Delta H = a^{\alpha\beta} H_{;\alpha\beta} = \tfrac{1}{\sqrt{a}}(\sqrt{a}\, a^{\alpha\beta} H_{,\beta})_{,\alpha}, \tag{122}$$

is the surfacial Laplacian of H.

Equation (121) is the classical *shape equation* for lipid bilayers in which the right-hand side is the pressure transmitted to the bilayer [Dharmavaram and Healey 2015; Jenkins 1977; Nitsche 1993]. Thus the cytoskeleton, if curved, transmits an effective pressure to the bilayer that persists when the net pressure p acting on the system vanishes. Vice versa, the bilayer transmits an equal but opposite pressure to the cytoskeleton.

We may rewrite (121) in the form

$$\kappa[\Delta H + 2H(H^2 - K)] = (\Sigma^{\alpha\beta} + \lambda a^{\alpha\beta})\, b_{\alpha\beta} + p. \tag{123}$$

This is the appropriate equation to use if the cytoskeleton convects with the bilayer because the parenthetical term on the right is then subject to (116), and in this setting extends the system obtained in [Guven et al. 2019] for strain-free deformations in which the entire metric, and not just the local areal stretch, is constrained, with $T^{\alpha\beta} = \Sigma^{\alpha\beta} + \lambda a^{\alpha\beta}$, in which the $\Sigma^{\alpha\beta}$ are constitutively indeterminate, then serving as the operative Lagrange multipliers.

6.5. ***Edge conditions.*** Boundary conditions are of limited relevance in this subject because bilayers typically form closed surfaces. Nevertheless, in the present approach based on the notion of patchwise equilibrium, they deliver expressions for the various actions at the edge of a patch which are of independent interest. Further, a number of models that entail boundary interactions are available in the literature [Agrawal and Steigmann 2009; Guven et al. 2019; Rosso and Virga 1999].

With the foregoing Euler equations satisfied on ω, the variation of the energy reduces, with the aid of (77), (84) and (98), to

$$\dot{E} = \int_{\partial\pi^*} \big(\sigma^{\alpha\beta}\nu_\beta u_\alpha + \varphi^\alpha \nu_\alpha + \mu\, \dot{J}_b/J_b\big)\, ds, \tag{124}$$

where $\mu\, ds = (\tilde{\mu} J_b)\, dS$ and φ^α is defined by (96).

We note, from (90) and (102), that the constraint $J_b = 1$ yields $\boldsymbol{\tau}\cdot\boldsymbol{v}' + \boldsymbol{\nu}\cdot\boldsymbol{v}_\nu = 0$, implying that the normal and tangential derivatives of $\boldsymbol{v}$ on $\partial\pi^*$ are not independent. Because $\boldsymbol{v}'$ is determined by $\boldsymbol{v}_{|\partial\pi^*}$, it follows that $\boldsymbol{v}$ and $\boldsymbol{v}_\nu$ cannot be specified independently. In the extended formulation, this restriction is relaxed and an associated Lagrange multiplier μ is introduced. Then, with (103) we obtain

$$\dot{E} = \int_{\partial\pi^*} \big\{\sigma^{\alpha\beta}\nu_\beta u_\alpha + [\boldsymbol{T}^\alpha\nu_\alpha - (\boldsymbol{M}^{\alpha\beta}\nu_\alpha\tau_\beta + \mu\boldsymbol{\tau})']\cdot\boldsymbol{v} + (\boldsymbol{M}^{\alpha\beta}\nu_\alpha\nu_\beta + \mu\boldsymbol{\nu})\cdot\boldsymbol{v}_\nu\big\}\, ds \\ - \sum [\boldsymbol{M}^{\alpha\beta}\nu_\alpha\tau_\beta + \mu\boldsymbol{\tau}]_i \cdot \boldsymbol{v}_i. \tag{125}$$

The virtual power is thus expressible in the form

$$P = \int_{\partial\pi^*} (\boldsymbol{t}_c\cdot\boldsymbol{u} + \boldsymbol{t}_b\cdot\boldsymbol{v} + \boldsymbol{\mu}\cdot\boldsymbol{v}_\nu)\, ds + \sum \boldsymbol{f}_i\cdot\boldsymbol{v}_i, \tag{126}$$

where $\boldsymbol{t}_c$, $\boldsymbol{t}_b$, $\boldsymbol{\mu}$ and $\boldsymbol{f}_i$ respectively are the cytoskeletal and bilayer tractions and the double force and corner forces acting on the bilayer patch. Accordingly,

$$\begin{gathered} \boldsymbol{t}_c = \sigma^{\alpha\beta}\nu_\beta \boldsymbol{a}_\beta, \quad \boldsymbol{t}_b = \boldsymbol{T}^\alpha\nu_\alpha - (M^{\alpha\beta}\nu_\alpha\tau_\beta\,\boldsymbol{n} + \mu\boldsymbol{\tau})', \\ \boldsymbol{\mu} = M\boldsymbol{n} + \mu\boldsymbol{\nu} \;\text{ and }\; \boldsymbol{f}_i = -[M^{\alpha\beta}\nu_\alpha\tau_\beta\,\boldsymbol{n} + \mu\boldsymbol{\tau}]_i, \;\text{ with }\; M = M^{\alpha\beta}\nu_\alpha\nu_\beta. \end{gathered} \tag{127}$$

The first of these is just the condition $(86)_2$ on $\partial\pi^*$.

The couple acting on the interior of $\partial\pi^*$ is

$$\boldsymbol{c} = \boldsymbol{r}\times\boldsymbol{t} + \boldsymbol{r}_\nu\times\boldsymbol{\mu}, \tag{128}$$

where $\boldsymbol{t} = \boldsymbol{t}_b + \boldsymbol{t}_c$ is the net traction and $\boldsymbol{r}_\nu = \nu^\alpha \boldsymbol{r}_{,\alpha} = \boldsymbol{\nu}$. Thus,

$$\boldsymbol{c} - \boldsymbol{r}\times\boldsymbol{t} = -M\boldsymbol{\tau}, \tag{129}$$

a pure bending couple acting at the edge that does not involve the multiplier μ. However, it is not appropriate to assign the couple in a boundary-value problem. Rather, information about μ is furnished by the specification of the double force [Toupin 1962].

If the bilayer and cytoskeleton are comoving, then $(127)_{3,4}$ remain in effect but $(127)_{1,2}$ are replaced by the single equation

$$\boldsymbol{t} = (\boldsymbol{T}^{\alpha} + \sigma^{\alpha\beta}\boldsymbol{a}_{\beta})\,\nu_{\alpha} - (M^{\alpha\beta}\nu_{\alpha}\tau_{\beta}\,\boldsymbol{n} + \mu\boldsymbol{\tau})'. \tag{130}$$

7. Legendre–Hadamard conditions

If the cytoskeleton convects with the bilayer, then because the effective energy involves the spatial derivatives of a single deformation field through the second order, the operative Legendre–Hadamard necessary condition for energy minimizers entails perturbation of the latter only, at fixed values of the first derivatives [Hilgers and Pipkin 1993]. Because the cytoskeletal energy involves only first derivatives, the operative Legendre–Hadamard condition then involves the bilayer energy alone. For the energy $(36)_1$, this yields the nonnegativity of the bending modulus k [Agrawal and Steigmann 2008], as implied by $(24)_1$ and $(31)_1$.

If the cytoskeleton and bilayer are not comoving, then the membrane-theoretic version of the Legendre–Hadamard condition is applicable, and implies that, at an arbitrary material point p, say, the cytoskeletal energy, regarded as a function of $\nabla\boldsymbol{r}_c$, is locally convex with respect to perturbations of the form

$$\boldsymbol{u}_{,\alpha} = \boldsymbol{a}k_{\alpha}, \tag{131}$$

i.e.,

$$a^{\mu}k_{\alpha} = u^{\mu}_{;\alpha} - w\,b^{\mu}_{\alpha} \quad \text{and} \quad ak_{\alpha} = u^{\mu}b_{\mu\alpha} + w_{,\alpha}, \tag{132}$$

with $a^{\mu} = \boldsymbol{a}\cdot\boldsymbol{a}^{\mu}$ and $a = \boldsymbol{a}\cdot\boldsymbol{n}$, subject tof $\boldsymbol{a}^{\alpha}\cdot\boldsymbol{a}k_{\alpha} = 0$ on account of areal incompressibility (see (66)). Thus, areal incompressibility imposes the restriction

$$a^{\alpha\beta}a_{\beta}k_{\alpha} = 0, \tag{133}$$

where $a_{\beta} = a_{\beta\mu}a^{\mu}$.

The operative Legendre–Hadamard condition is [Steigmann 1990]

$$\boldsymbol{a}\cdot(\boldsymbol{E}^{\alpha\beta}k_{\alpha}k_{\beta})\,\boldsymbol{a} \geq 0, \tag{134}$$

for arbitrary $\boldsymbol{a}k_{\alpha}$ subject to (133), where

$$\boldsymbol{E}^{\alpha\beta} = 2\frac{\partial W}{\partial a_{\alpha\beta}}\boldsymbol{I} + 4\frac{\partial^2 W}{\partial a_{\alpha\mu}\,\partial a_{\beta\lambda}}\boldsymbol{a}_{\mu}\otimes\boldsymbol{a}_{\lambda}, \tag{135}$$

in which $W(a_{\alpha\beta}) = W_c(a_{\mu\lambda}\,\boldsymbol{A}^\mu \otimes \boldsymbol{A}^\lambda)$. Then, with $(81)_1$, specialized to $J_c = 1$, we require

$$\Sigma^{\alpha\beta} k_\alpha k_\beta |\boldsymbol{a}|^2 + 4\frac{\partial^2 W}{\partial a_{\alpha\mu}\,\partial a_{\beta\lambda}} a_\mu k_\alpha a_\lambda k_\beta \geq 0, \tag{136}$$

where $\Sigma^{\alpha\beta}$ is the constitutively determined part of the cytoskeletal Cauchy stress.

This condition yields a nontrivial restriction on W even if the bilayer remains undisturbed; i.e., if $w = 0$.

The choice $\boldsymbol{a} = a\boldsymbol{n}$ $(a_\beta = 0)$ conforms to (133) and reduces (136) to

$$\Sigma^{\alpha\beta} k_\alpha k_\beta \geq 0, \tag{137}$$

implying that the energetic part of the stress is positive semidefinite in energy minimizing states. In the absence of constraints, this implies, in accordance with a restriction proposed in [Stamenović 2006], that the Cauchy stress is positive semidefinite.

For example, in the case of isotropy (see (87)), (137) reduces to $\varpi'(I)|\boldsymbol{k}|^2 \geq 0$, where $|\boldsymbol{k}|^2 = A^{\alpha\beta} k_\alpha k_\beta$, and is thus satisfied if and only if

$$\varpi'(I) \geq 0, \tag{138}$$

whereas the full Legendre–Hadamard inequality (136), in the case of isotropy, is

$$\varpi'(I)|\boldsymbol{a}|^2|\boldsymbol{k}|^2 + 2\varpi''(I)(k^\alpha a_\alpha)^2 \geq 0, \tag{139}$$

with $k^\alpha = A^{\alpha\beta} k_\beta$.

8. Equivalent monolayers with spontaneous curvature

8.1. ***Equilibrium of monolayers.*** We expect the conforming cytoskeleton to confer asymmetry in the bending response of the bilayer/cytoskeleton composite, whereas that of an isolated bilayer is symmetric in the sense that the energy $(36)_1$ is the invariant under $\boldsymbol{b} \to -\boldsymbol{b}$. Asymmetric bending is also a feature of conventional monolayers, consisting of one sheet of oriented lipids instead of two of opposing orientation (Figure 1). Conventionally, this asymmetry is modelled by introducing a *spontaneous curvature* $C(\theta^\alpha)$ [Ou-Yang et al. 1999] via the energy

$$W(H, K; \theta^\alpha) = \kappa(H - C)^2 + \bar{\kappa} K. \tag{140}$$

The existence of these distinct models of asymmetric bending leads us to search for conditions under which they might be equivalent.

Proceeding as in Section 6.3, we derive (97) but with (92) replaced by

$$N^{\alpha\mu} = \{\lambda_m + \kappa(H - C)^2 - \bar{\kappa} K\} a^{\alpha\mu} - 2\kappa(H - C)\, b^{\alpha\mu}, \qquad \text{and} \quad M^{\alpha\mu} = \kappa(H - C)\, a^{\alpha\mu} + \bar{\kappa} \tilde{b}^{\alpha\mu}, \tag{141}$$

where λ_m is a Lagrange multiplier associated with the areal incompressibility of the monolayer. Then with some labor we find that (111) is replaced by

$$a^{\alpha\mu}[(\lambda_m)_{,\alpha} - 2\kappa(H - C)C_{,\alpha}] + g^{\mu}_{(m)} = 0, \tag{142}$$

where $g^{\mu}_{(m)}$ is a tangential distribution of force on the monolayer; and, in the absence of the cytoskeleton, that (121) is replaced by

$$\kappa[\Delta(H - C) + 2(H - C)(2H^2 - K) - 2H(H - C)^2] - 2\lambda_m H = p, \tag{143}$$

where p is the pressure exerted on the monolayer.

Evidently, (142) corresponds to (111) if $C_{,\alpha}$ vanishes, i.e., if the spontaneous curvature is uniform. In this case we have

$$\kappa[\Delta H + 2H(H^2 - K)] - 2\lambda_m H = p + 2\kappa C(CH - K), \tag{144}$$

which corresponds to (121), provided that $\lambda_m = \lambda_b$ and the cytoskeletal stress $\sigma^{\alpha\beta}$ satisfies

$$\sigma^{\alpha\beta} b_{\alpha\beta} = 2\kappa C(CH - K). \tag{145}$$

Equations (47) and (49) furnish $2H = a^{\alpha\beta} b_{\alpha\beta}$ and $2K = \tilde{b}^{\alpha\beta} b_{\alpha\beta}$, and so a *sufficient* condition for such correspondence is

$$\sigma^{\alpha\beta} = \kappa C(Ca^{\alpha\beta} - \tilde{b}^{\alpha\beta}), \tag{146}$$

provided that no tangential force is acting on the cytoskeleton. For, this expression for the stress is automatically divergence-free and $(86)_1$ requires that the tangential force vanish.

We observe, noting (123), that this same correspondence may be established between the monolayer and the comoving cytoskeleton if $\lambda_m = 0$ and if $\lambda_c = \lambda$ in $(82)_2$.

These correspondences must be qualified by the fact that the constitutive response of the cytoskeleton cannot be expected to yield (146) in general. Nevertheless, in the absence of tangential forces, the latter allows us to dispense with $(86)_1$ or (116), which would otherwise pose significant obstacles to analysis. Thus, we view (146) simply as a device for generating potential solutions by selecting from among a number of explicit solutions that are available for monolayers with constant spontaneous curvature [Ou-Yang et al. 1999]. Remarkably, these include the characteristic biconcave discoid shape of red-blood cells in equilibrium.

8.2. ***Biconcave discoid.*** Consider a surface of revolution described by

$$\boldsymbol{r}(\theta^\alpha) = r\boldsymbol{e}_r(\theta) + z(r)\boldsymbol{k}, \tag{147}$$

where $r(= \theta^1)$ is the radius from the symmetry axis directed along the fixed unit vector $\boldsymbol{k}$, $\theta(= \theta^2)$ is the azimuthal angle, and $\boldsymbol{e}_r(\theta)$ is a radial unit vector orthogonal to the axis of symmetry at azimuth θ. Let $\psi(r)$ be the angle defining the slope of a meridian: $\tan\psi = z'(r)$. Then with reference to Section 5, we compute

$$\boldsymbol{a}_1 = \boldsymbol{e}_r(\theta) + \tan\psi\,\boldsymbol{k}, \quad \boldsymbol{a}_2 = r\boldsymbol{e}_\theta(\theta), \tag{148}$$

where $\boldsymbol{e}_\theta = \boldsymbol{e}_r'(\theta)$; the metric and dual metric

$$(a_{\alpha\beta}) = \text{diag}(\sec^2\psi, r^2), \quad (a^{\alpha\beta}) = \text{diag}(\cos^2\psi, r^{-2}); \tag{149}$$

the curvature

$$(b_{\alpha\beta}) = \text{diag}(\psi'\sec\psi, r\sin\psi); \tag{150}$$

the mean and Gaussian curvatures

$$2H = r^{-1}(r\sin\psi)' \quad \text{and} \quad K = r^{-1}\psi'\sin\psi\cos\psi; \tag{151}$$

and the curvature cofactor

$$(\tilde{b}^{\alpha\beta}) = \text{diag}(r^{-1}\sin\psi\cos^2\psi, r^{-2}\psi'\cos\psi). \tag{152}$$

The Laplacian of the mean curvature, needed in (144), is (see (122))

$$\Delta H = r^{-1}\cos\psi[(r\cos\psi)H']'. \tag{153}$$

Consider the particular surface of revolution described by

$$\sin\psi = r(d\ln r + b), \tag{154}$$

where b, d are constants. Following the procedure outlined in Section 4.3 of [Ou-Yang et al. 1999] and adjusting for differences in notation, with some effort it may be verified that (154) solves the shape equation (144) for a monolayer with a constant spontaneous curvature, provided that

$$\lambda_m = 0, \quad p = 0 \quad \text{and} \quad d = 2C, \tag{155}$$

and no tangential distributed force is acting.

In [Ou-Yang et al. 1999] this surface is described in terms of the dimensionless radius

$$x = r/\bar{r}, \quad \text{where} \quad \bar{r} = \exp(-b/d), \tag{156}$$

is such that $\sin\psi(\bar{r}) = 0$, which we use to recast (154) as

$$\sin\psi = \beta x\ln x, \quad \text{with} \quad \beta = 2C\bar{r}. \tag{157}$$

Following [Ou-Yang et al. 1999], we fix $\beta < 0$ with $|\beta| < e$, corresponding to a negative spontaneous curvature. Evidently, $\sin\psi$ vanishes at $x = 0$ and $x = 1$ and is maximized at $x = e^{-1}$. Because $\sin\psi \le 1$ the domain of the variable x is $[0, x_e]$, where

$$x_e\ln x_e = |\beta|^{-1}, \tag{158}$$

which yields a unique $x_e > 1$ [Ou-Yang et al. 1999]. This is the dimensionless equatorial radius, where $\sin\psi = -1$.

To obtain the shape of the surface we integrate $\tan\psi = \varsigma'(x)$, where $\varsigma(x) = z(r)/\bar{r}$. Thus,

$$\varsigma(x) = \int_{x_e}^{x} \frac{\beta t \ln t}{\sqrt{1-\beta^2 t^2 (\ln t)^2}}\, dt, \tag{159}$$

in which we have chosen the positive root for the cosine and normalized to $\varsigma(x_e) = 0$. A numerical quadrature furnishes the upper half of a biconcave discoid, depicted in Figure 3. This is extended by rotational and reflection symmetry to the entire discoid.

Some insight into the mechanics of the system may be gained by computing the transverse shear traction S acting on a parallel of latitude. Assuming the component μ of the double force to vanish on a parallel, we find, from $(92)_2$ and $(127)_2$, that $S = \boldsymbol{n}\cdot\boldsymbol{T}^{\alpha}\nu_{\alpha}$, where

$$\boldsymbol{n} = \cos\psi\,\boldsymbol{k} - \sin\psi\,\boldsymbol{e}_r \quad \text{and} \quad \boldsymbol{\nu} = \cos\psi\,\boldsymbol{e}_r + \sin\psi\,\boldsymbol{k}, \tag{160}$$

are the surface normal and the normal to a parallel, respectively. Then $(92)_2$ and (97) furnish $S = -M^{\alpha\beta}_{;\beta}\nu_{\alpha} = -\kappa\nu^{\alpha}H_{,\alpha}$, i.e.,

$$S = -\kappa\cos\psi\, H'(r), \tag{161}$$

which may be reduced, using (151), (154) and $(155)_3$, to

$$S = -2\kappa C r^{-1}\cos\psi. \tag{162}$$

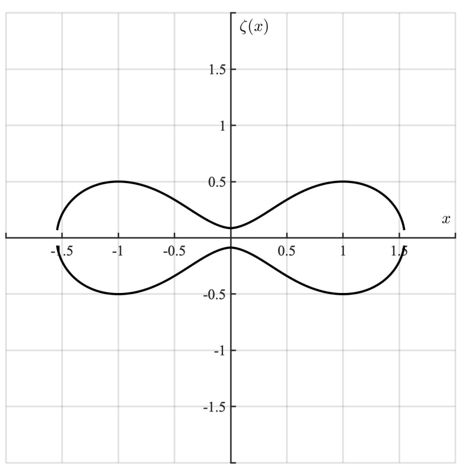


Figure 3. Biconcave discoid ($\beta = -1.4721$).

This vanishes at the equator, where $\psi = -\pi/2$, and therefore meets a necessary condition for reflection symmetry of the surface with respect to the equatorial plane. For, if there were a nonzero shear traction transmitted by the material below the equator to that above, then equilibrium would require that it be balanced by an equal and opposite traction exerted by the part of the membrane above the equator on that below, and this would destroy reflection symmetry. However, the biconcave discoid is not free standing. There is a point force $F\boldsymbol{k}$ acting at the pole, where $\psi = 0$, given by

$$F = -2\pi \lim_{r\to 0}(rS) = 4\pi\kappa C, \tag{163}$$

which was overlooked in [Ou-Yang et al. 1999].

8.3. ***Mapping a plane cytoskeletal disc to a biconcave discoid.*** To adapt (154) to the bilayer/cytoskeleton composite, we must select a suitable configuration relative to which the constitutive framework (87) for an isotropic cytoskeleton, say, may be implemented. Because the literature is ambiguous concerning this issue, we consider a plane disc for the sake of illustration, and seek a strain-energy function which is such as to admit (154) as an equilibrium configuration in the absence of any distributed tangential forces acting on the bilayer or cytoskeleton.

We parametrize the disc by the position function $\boldsymbol{\xi}(\theta^\alpha) = \rho(r)\,\boldsymbol{e}_r(\theta)$ (see (6)). The induced tangent basis elements, $\boldsymbol{A}_\alpha = \boldsymbol{\xi}_{,\alpha}$, are

$$\boldsymbol{A}_1 = \rho'(r)\,\boldsymbol{e}_r(\theta) \quad \text{and} \quad \boldsymbol{A}_2 = \rho(r)\,\boldsymbol{e}_\theta(\theta), \tag{164}$$

and the metric and dual metric are

$$(A_{\alpha\beta}) = \text{diag}[(\rho')^2, \rho^2] \quad \text{and} \quad (A^{\alpha\beta}) = \text{diag}[(\rho')^{-2}, \rho^{-2}]. \tag{165}$$

With $J_c = \sqrt{a/A}$, where $a = \det(a_{\alpha\beta})$ and $A = \det(A_{\alpha\beta})$, we obtain

$$J_c = r \sec\psi/(\rho\rho'), \tag{166}$$

and

$$I = a_{\alpha\beta}\,A^{\alpha\beta} = J_c^2(\rho/r)^2 + (r/\rho)^2. \tag{167}$$

Areal incompressibility then yields

$$I = (\rho/r)^2 + (r/\rho)^2, \tag{168}$$

and furnishes a differential equation for $\rho(r)$:

$$\rho\rho' = r \sec\psi. \tag{169}$$

This integrates to

$$\left(\frac{X}{x}\right)^2 = \frac{2}{x^2}\int_0^x t \sec\psi(t)\,dt, \quad \text{where} \quad X = \rho/\bar{r}, \tag{170}$$

and we have imposed $X = 0$ at $x = 0$ (Figure 4).

The constitutive part of the stress is given by (87). We combine this with $(82)_2$, (87) and (146) to derive the system

$$\begin{aligned} \lambda_c \cos^2\psi + 2\varpi'(I)(\rho')^{-2} &= -\kappa C r^{-1}\sin\psi\cos^2\psi + \kappa C^2\cos^2\psi, \\ \lambda_c r^{-2} + 2\varpi'(I)\,\rho^{-2} &= -\kappa C r^{-2}\psi'\cos\psi + \kappa C^2 r^{-2}, \end{aligned} \tag{171}$$

which also applies in the case of a comoving cytoskeleton if the multiplier λ_c is replaced by λ. Eliminating this multiplier, we obtain

$$2\varpi'(I)[(\rho')^{-2} - (r/\rho)^2\cos^2\psi] = -\kappa C\cos^2\psi(r^{-1}\sin\psi - \psi'\cos\psi), \tag{172}$$

which may be simplified by using (166) to reduce the left-hand side. On the right-hand side we use (154), finding that

$$r^{-1}\sin\psi - \psi'\cos\psi = -d. \tag{173}$$

Then, with (155) we have

$$\varpi'(I)[(\rho/r)^2 - (r/\rho)^2] = \kappa C^2, \tag{174}$$

where, from (168),

$$[(\rho/r)^2 - (r/\rho)^2]^2 = I^2 - 4. \tag{175}$$

With $\cos\psi \in (0, 1]$ almost everywhere on the biconcave discoid (Figure 3), equation (169) implies that $\rho/r (= X/x) > 1$ almost everywhere (Figure 4). Then

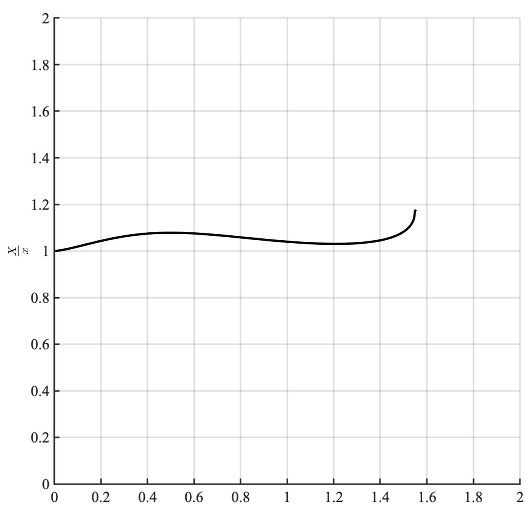


Figure 4. Map from the biconcave discoid to the plane disc ($\beta = -1.4721$).

$(\rho/r)^2 - (r/\rho)^2 > 0$ and (174), (175) deliver

$$\varpi'(I) = \kappa C^2/\sqrt{I^2-4}, \tag{176}$$

which is meaningful if $I > 2$ (as required by (175)) and satisfies (138). Thus,

$$\varpi(I) = \kappa C^2 \ln\bigl[\tfrac{1}{2}(I+\sqrt{I^2-4})\bigr], \tag{177}$$

normalized to $\varpi(2) = 0$.

We are not able to show that (176) satisfies the full Legendre–Hadamard inequality (139). However, as previously noted, the latter is not relevant if the cytoskeleton and bilayer are comoving.

Appendix

We show that $G(-\boldsymbol{e}) = G(\boldsymbol{e})$ if and only if there is a function S such that $G(\boldsymbol{e}) = S(\boldsymbol{e}\otimes\boldsymbol{e})$. Sufficiency is immediate. To establish necessity, we show that if $G(-\boldsymbol{e}) = G(\boldsymbol{e})$, then G is determined by $\boldsymbol{e}\otimes\boldsymbol{e}$, i.e., that $G(\boldsymbol{a}) = G(\boldsymbol{b})$ whenever $\boldsymbol{a}\otimes\boldsymbol{a} = \boldsymbol{b}\otimes\boldsymbol{b}$. The latter yields

$$a^2\boldsymbol{a} = (\boldsymbol{a}\cdot\boldsymbol{b})\,\boldsymbol{b} \quad \text{and} \quad b^2\boldsymbol{b} = (\boldsymbol{a}\cdot\boldsymbol{b})\,\boldsymbol{a},$$

where $a = |\boldsymbol{a}|$, etc. The combination of these gives $a = b$ and $a^2b^2 = (\boldsymbol{a}\cdot\boldsymbol{b})^2$. But there is $\theta \in \mathbb{R}$ such that $\boldsymbol{a}\cdot\boldsymbol{b} = ab\cos\theta$. Thus $\cos\theta = \pm 1$ and either of the two equations yields $\boldsymbol{b} = \pm\boldsymbol{a}$. The first alternative gives $G(\boldsymbol{a}) = G(\boldsymbol{b})$; the second yields $G(\boldsymbol{a}) = G(-\boldsymbol{b})$, so that if G is insensitive to the choice of sign, as assumed, then $G(\boldsymbol{a}) = G(\boldsymbol{b})$ whenever $\boldsymbol{a}\otimes\boldsymbol{a} = \boldsymbol{b}\otimes\boldsymbol{b}$.

Acknowledgments

Hendrickson gratefully acknowledges support provided by a National Science Foundation Fellowship. Shirani and Steigmann gratefully acknowledge the partial support of the NSF through grant number CMMI-1931064.

References

[Agrawal and Steigmann 2008] A. Agrawal and D. J. Steigmann, "Coexistent fluid-phase equilibria in biomembranes with bending elasticity", *J. Elasticity* **93** (2008), 63–80.

[Agrawal and Steigmann 2009] A. Agrawal and D. J. Steigmann, "Boundary-value problems in the theory of lipid membranes", *Contin. Mech. Therm.* **21** (2009), 57–82.

[Agrawal and Steigmann 2011] A. Agrawal and D. J. Steigmann, "A model for surface diffusion of trans-membrane proteins on lipid bilayers", *Z. Angew. Math. Phys.* **62**:3 (2011), 549–563.

[Aris 1989] R. Aris, *Vectors, tensors and the basic equations of fluid mechanics*, Dover Publications, 1989.

[Berdichevsky 2009] V. Berdichevsky, *Variational principles of continuum mechanics: part I—fundamentals*, Springer, Berlin, 2009.

[Canham 1970] P. B. Canham, "The minimum energy of bending as a possible explanation of the biconcave shape of the human red blood cell", *J. Theor. Biol.* **26**:1 (1970), 61–81.

[Ciarlet 2005] P. G. Ciarlet, *An introduction to differential geometry with applications to elasticity*, Springer, Dordrecht, 2005.

[Cohen and Wang 1984] H. Cohen and C.-C. Wang, "On the response and symmetry of elastic and hyperelastic membrane points", *Arch. Ration. Mech. Anal.* **85** (1984), 355–391.

[Dharmavaram and Healey 2015] S. Dharmavaram and T. J. Healey, "On the equivalence of local and global area-constraint formulations for lipid bilayer vesicles", *Z. Angew. Math. Phys.* **66** (2015), 2843–2854. arXiv

[Evans and Skalak 1980] E. Evans and R. Skalak, *Mechanics and thermodynamics of biomembranes*, CRC Press, Boca Raton, 1980.

[Guven et al. 2019] J. Guven, M. M. Müller, and P. Vázquez-Montejo, "Isometric bending requires local constraints on free edges", *Math. Mech. Solids* **24**:12 (2019), 4051–4077. arXiv

[Helfrich 1973] W. Helfrich, "Elastic properties of lipid bilayers: theory and possible experiments", *Z. Naturforsch.* **28** (1973), 693–703.

[Herant and Dembo 2006] M. Herant and M. Dembo, "Active cellular protrusion: continuum theories and models", pp. 204–224 in *Cytoskeletal mechanics: models and measurements in cell mechanics*, edited by M. R. K. Mofrad and R. D. Kamm, Cambridge University Press, Cambridge, 2006.

[Hilgers and Pipkin 1993] M. G. Hilgers and A. C. Pipkin, "Energy-minimizing deformations of elastic sheets with bending stiffness", *J. Elasticity* **31** (1993), 125–139.

[Jenkins 1977] J. T. Jenkins, "The equations of mechanical equilibrium of a model membrane", *SIAM J. Appl. Math.* **32**:4 (1977), 755–764.

[Kamm and Mofrad 2006] R. D. Kamm and M. R. K. Mofrad, "Introduction, with the biological basis for cell mechanics", pp. 1–17 in *Cytoskeletal mechanics: models and measurements in cell mechanics*, edited by M. R. K. Mofrad and R. D. Kamm, Cambridge University Press, Cambridge, 2006.

[Krishnaswamy 1996] S. Krishnaswamy, "A cosserat-type model for the red blood cell wall", *Int. J. Eng. Sci.* **34**:8 (1996), 873–899.

[Murdoch and Cohen 1979] A. I. Murdoch and H. Cohen, "Symmetry considerations for material surfaces", *Arch. Ration. Mech. Anal.* **72** (1979), 61–98.

[Naghdi 1972] P. M. Naghdi, "The theory of shells and plates", in *S. Flügge's Handbuch der Physik*, vol. VIa/2, edited by C. Truesdell, Springer, Berlin, 1972.

[Nitsche 1993] J. C. C. Nitsche, "Boundary value problems for variational integrals involving surface curvatures", *Quart. Appl. Math.* **51**:2 (1993), 363–387.

[Noll 1958] W. Noll, "A mathematical theory of the mechanical behavior of continuous media", *Arch. Ration. Mech. Anal.* **2** (1958), 197–226.

[Noll 2004] W. Noll, "On material frame-indifference", Five contributions to natural philosophy, 2004, Available at http://www.math.cmu.edu/∼wn0g/noll/FC.pdf.

[Ou-Yang et al. 1999] Z.-C. Ou-Yang, J.-X. Liu, and Y.-Z. Xia, *Geometric methods in the elastic theory of membranes in liquid crystal phases*, World Scientific, Singapore, 1999.

[Pan et al. 2018] L. Pan, R. Yan, W. Li, and K. Xu, "Super-resolution microscopy reveals the native ultrastructure of the erythrocyte cytoskeleton", *Cell Rep.* **22**:5 (2018), 1151–1158.

[Rosso and Virga 1999] R. Rosso and E. G. Virga, "Adhesive borders of lipid membranes", *Proc. R. Soc. Lond. A* **455**:1992 (1999), 4145–4168.

[Scriven 1960] L. E. Scriven, "Dynamics of a fluid interface: equation of motion for Newtonian surface fluids", *Chem. Eng. Sci.* **12**:2 (1960), 98–108.

[Stamenović 2006] D. Stamenović, "Models of cytoskeletal mechanics based on tensegrity", pp. 103–128 in *Cytoskeletal mechanics: models and measurements in cell mechanics*, edited by M. R. K. Mofrad and R. D. Kamm, Cambridge University Press, Cambridge, 2006.

[Steigmann 1990] D. J. Steigmann, "Tension-field theory", *Proc. R. Soc. Lond. A* **429**:1876 (1990), 141–173.

[Steigmann 2003] D. J. Steigmann, "Irreducible function bases for simple fluids and liquid crystal films", *Z. angew. Math. Phys.* **54** (2003), 462–477.

[Steigmann 2010] D. J. Steigmann, "Applications of polyconvexity and strong ellipticity to nonlinear elasticity and elastic plate theory", pp. 265–299 in *Poly-, quasi- and rank-one convexity in applied mechanics*, vol. 516, edited by J. Schröder and P. Neff, Springer, Vienna, 2010.

[Steigmann 2018] D. J. Steigmann, "Continuum theory for elastic sheets formed by inextensible crossed elasticae", *Int. J. Non-Linear Mech.* **106** (2018), 324–329.

[Tartibi et al. 2015] M. Tartibi, Y. X. Liub, G.-Y. Liub, and K. Komvopoulos, "Single-cell mechanics — an experimental-computational method for quantifying the membrane-cytoskeleton elasticity of cells", *Acta Biomater.* **27** (2015), 224–235.

[Toupin 1962] R. A. Toupin, "Elastic materials with couple-stresses", *Arch. Ration. Mech. Anal.* **11** (1962), 385–414.

[Virga 1994] E. G. Virga, *Variational theories for liquid crystals*, Chapman & Hall, London, 1994.

[Wang and Pipkin 1986] W.-B. Wang and A. C. Pipkin, "Inextensible networks with bending stiffness", *Quart. J. Mech. Appl. Math.* **39**:3 (1986), 343–359.

[Zheng 2003] Q.-S. Zheng, "Irreducible function bases for simple fluids and liquid crystal films: a new derivation", *Z. angew. Math. Phys.* **54** (2003), 478–483.

[Zheng et al. 1992] Q.-S. Zheng, J. Betten, and A. J. M. Spencer, "The formulation of constitutive equations for fibre-reinforced composites in plane problems: part I", *Arch. Appl. Mech.* **62** (1992), 530–543.

Received 24 Oct 2019. Revised 27 Feb 2020. Accepted 4 Apr 2020.

BRETT HENDRICKSON: bretth@berkeley.edu
Department of Mechanical Engineering, University of California, Berkeley, 6112 Etcheverry Hall, Berkeley, CA 94720-1740, United States

MILAD SHIRANI: milad_shirani@berkeley.edu
Department of Mechanical Engineering, University of California, Berkeley, 6112 Etcheverry Hall, Berkeley, CA 94720-1740, United States

DAVID J. STEIGMANN: dsteigmann@berkeley.edu
Department of Mechanical Engineering, University of California, Berkeley, 6133 Etcheverry Hall, Berkeley, CA 94720-1740, United States

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the submission page.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in MEMOCS are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and a Mathematics Subject Classification or a Physics and Astronomy Classification Scheme code for the article, and, for each author, postal address, affiliation (if appropriate), and email address if available. A home-page URL is optional.

Format. Authors are encouraged to use LaTeX and the standard amsart class, but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should normally be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages — Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc. — allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with as many details as you can about how your graphics were generated.

Bundle your figure files into a single archive (using zip, tar, rar or other format of your choice) and upload on the link you been provided at acceptance time. Each figure should be captioned and numbered so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript with all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables.

White Space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Mathematics and Mechanics of Complex Systems
vol. 8 no. 1 2020

MEMOCS is a journal of the International Research Center for the Mathematics and Mechanics of Complex Systems at the Università dell'Aquila, Italy.

2326-7186(2020)8:1;1-B