Polish Academy of Sciences

Institute of Fundamental Technological Research

Archives of Mechanics

Archiwum Mechaniki Stosowanej

volume 48 issue 5

Polish Scientific Publishers PWN

Warszawa 1996

ARCHIVES OF MECHANICS IS DEVOTED TO
Theory of elasticity and plasticity • Theory of nonclassical continua • Physics of continuous media • Mechanics of discrete media • Nonlinear mechanics • Rheology • Fluid gas-mechanics • Rarefied gas • Thermodynamics

---

Printed in Poland, Editorial Office: Świętokrzyska 21,
00-049 Warszawa (Poland)

Arkuszy wydawniczych 11,5. Arkuszy drukarskich 10,5
Papier offset. kl. III 70 g. B1. Oddano do składania w lipcu 1996 r.
Druk ukończono we wrześniu 1996 r.
Skład i łamanie: "MAT-TEX"
Druk i oprawa: Drukarnia Braci Grodzickich, Żabieniec ul. Przelotowa 7

Arch. Mech., 48, 5, pp. 791–812, Warszawa 1996

Thermodynamic potentials and extremum principles for a Boltzmann gas

Z. BANACH and S. PIEKARSKI (WARSZAWA)

IN THIS PAPER, a thermodynamic interpretation of the kinetic theory and Boltzmann's equation is explicitly obtained; the nonequilibrium thermodynamic space consists of the nonnegative distribution functions. Beginning from a molecular expression for entropy in the form of Boltzmann's H functional, the systematic construction of thermodynamic potentials is demonstrated for gaseous systems beyond local equilibrium ("not infinitesimally near to equilibrium"). Since the nonequilibrium thermodynamic space provides natural comparison states for the principles of maximum entropy or minimum energy, a simple criterion for the choice of gas-state variables can be given which shows that the maximization of the Legendre transforms of entropy is logically equivalent to the minimization of the Legendre transforms of energy. This criterion is sought after in such a way that the concepts of temperature and pressure need no reformulation out of equilibrium. After these preparations, the technique of functional differentiation is used to derive the generalized Gibbs equation (relation) for Boltzmann's entropy. Finally, the paper presents an analysis of how the functional representation of this equation relates to the method of moments.

1. Introduction

IN GIBBSIAN THERMOSTATICS [1–3], one postulates that all macroscopic properties of a thermodynamic system are contained in a fundamental equation representing either the entropy or the energy as a function of additive conserved quantities. Thus in both the entropy and energy representations the extensive parameters play the roles of mathematically independent variables, whereas the intensive parameters are introduced in a systematic manner as derivatives of the fundamental equation. The energy-language fundamental equation can be taken as the basic relation on which the Legendre transformation is performed. Such Legendre transforms of the energy are usually called thermodynamic potentials. Another set of thermodynamic functions (Massieu's functions) can be obtained by performing the Legendre transformations on the entropy rather than on the energy. In the entropy representation the entropy is maximum for constant energy, and from this it follows that each Legendre transform of the entropy is maximum for constant values of the transformed (intensive) variables. Similarly, in the energy representation the energy is minimum for constant entropy, and from this it follows that each Legendre transform of the energy is minimum for constant values of the transformed (intensive) variables.

However, it has been known for a long time that the consistent development of these ideas is contingent on the solution of a methodological problem of Gibbsian thermostatics illustrated by the following paradox: how are we to give a precise meaning to the statement that entropy is maximum for constant energy, whereas

entropy is defined only for systems in equilibrium? Many authors have grappled with this dilemma until a partial solution was found in terms of the composite system. The basic purpose of this paper is to present a simple alternative solution. Based on the kinetic theory of rarefied gases, we approach the question by means of a molecular expression for entropy in the form of Boltzmann's H functional. As we shall soon see, the introduction of nonequilibrium distribution functions enables us to define comparison states for the extremum principles and thus to solve the aforementioned paradox.

Our method is, in fact, quite straightforward. We provide a thermodynamic interpretation of the kinetic theory in which the nonequilibrium states of the gas are described by means of the nonnegative distribution functions f. This description can be used in various ways to introduce also other variables. For example, an interesting alternative is to use the specific volume v, the internal energy density ε, and an appropriately defined phase-space function G, and then to express Boltzmann's entropy h (per unit mass) in terms of (v, ε, G) rather than f. These two descriptions appear on an equal footing, and we can choose either one to suit the problem at hand.

If we choose (v, ε, G), we will be able to exhibit the decomposition of Boltzmann's entropy h into two physically different parts:

$$h(v, \varepsilon, G) = h_E(v, \varepsilon) + \Delta(G). \tag{1.1}$$

The first part, denoted for brevity by h_E, represents the thermostatic entropy (which is a function of v and ε), while the second part, denoted by Δ, gives the functional contribution to h independent of (v, ε) and vanishing at equilibrium; Δ depends only on G. The resulting expression (1.1) for h is such that the derivatives of h with respect to v and ε are the same functions as in equilibrium, and Gibbsian thermostatics is not to be viewed simply as a first approximation to the full description of the system, but, instead, as an exact theory valid for a suitably chosen parametrization of the space of nonequilibrium states. The motivation for the introduction of Eq. (1.1) may be understood in terms of our desire to insure that the entropy maximum principle will go over into an energy minimum principle on inversion of $h = h(v, \varepsilon, G)$ with respect to ε: $\varepsilon = \varepsilon(v, h, G)$. The extremum principles in the Legendre transformed representations are then proved to be logically equivalent to the maximization of the entropy or to the minimization of the energy.

Mathematically, the nonequilibrium method presented in this paper is parallel to the equilibrium method of CALLEN [2], but has the following feature distinguishing it from that encountered in thermostatics: when the system is removed from equilibrium, the theory of Legendre transforms and thermodynamic potentials depends very much on the choice of variables in terms of which it has to characterize the state of a system. Clearly, if one is interested only in the discussion of the equilibrium case, one knows full well what the "right" variables are

and these sorts of complications may be avoided altogether. For the general case, however, the optimum definition of state variables is not a trivial problem. As a matter of fact, one will be unable to obtain any successful theory of thermodynamic potentials unless one formulates some adequate criterion for the choice of state variables. In our approach we introduce this criterion in such a way that the concepts of temperature and pressure need no reformulation out of equilibrium and the extremum principles are valid for the Legendre transforms of entropy and energy.

If the space of nonequilibrium states carries a structure of a finite-dimensional manifold, the entropy maximum principle allows one to draw upon results from the critical point theory, as formulated by MORSE [4]. Using this theory, it is possible to find a coordinate system for the manifold of nonequilibrium states such that the specific entropy can exactly be written as a sum of two physically different terms [5–7]: the first term represents the thermostatic entropy (which is a function of conserved variables), while the second term is given by a quadratic form depending only on nonequilibrium variables. Due to the existence of this particularly simple representation of the specific entropy, one easily arrives at the natural definitions of temperature, pressure, and thermodynamic potentials for systems "not infinitesimally near to equilibrium." However, there does not as yet exist a kinetic-theory framework in which these and similar problems may be addressed in a very satisfactory way. Thus, the underlying philosophy here is not to formulate a completely systematic extension of the aforementioned results to the general infinite-dimensional case, or even to propose some modification of Morse's lemma. Rather, the objective is to exploit the specific properties of a classical rarefied gas of massive particles and to obtain the required coordinatization of the space of nonequilibrium states by direct guessing.

Another remark is also in order. For fixed values of v and ε, $h(v, \varepsilon, \bullet)$ denotes a functional, that is, a function whose argument is G. Consequently, any derivation of the Gibbs relation for h is necessarily based on the technique of functional differentiation. It can be most simply conceived of as a straightforward generalization of the concept of partial derivative. The foregoing functional justification of the existence of thermodynamic potentials for gaseous systems beyond local equilibrium is different from the method of moments [8]. Formally, *this method gives the same Gibbs relation* for the Boltzmann entropy density h as in the technique of functional differentiation; but its precise definition and mathematical status are complex, and lose direct physical meaning because the Hermite expansions of G and $\ln(1 + G)$ are applied at the outset of the analysis. Furthermore, there are difficulties in proving the existence and convergence of various series involved. Nevertheless, to understand the conceptual problems associated with this method, we decide to present an explanation of how the moment representation of entropy relates to our formalism.

The layout of this paper is as follows. In Secs. 2 and 3, we describe the properties of nonequilibrium thermodynamic potentials. In Secs. 4 and 5, the status

of the method of moments is examined *vis-a-vis* the technique of functional differentiation. Section 6 is for discussion and conclusion. Some auxiliary material is included as Appendices A and B.

2. The relation between Boltzmann's entropy and mass density

2.1. Preliminaries

The kinetic theory describes the state of a gas by the distribution function $f(x, c, t)$ defined, according to Boltzmann, in such a way that $f(x, c, t)\, d^N c$ is the number density of molecules at the point x and at the time t that have velocities in the "volume" element $d^N c$ around c; N is the dimension of the vector space to which c belongs. The distribution function obeys the kinetic equation of the form

$$\partial_t f + c \cdot \partial_x f = J(f), \tag{2.1}$$

where $J(f)$ is the collision term.

In the kinetic theory of rarefied gases, local entropy S (per unit volume) is sometimes required, and it is locally defined by the functional expression

$$S(f) := -k_B \int f \ln(\mathbb{C} f)\, d^N c, \tag{2.2}$$

where

$$\mathbb{C} := (2\pi\hbar m^{-1})^N \tag{2.3}$$

and where m is the molecular mass and k_B and $2\pi\hbar$ are constants of Boltzmann and Planck, respectively. Differentiating S with respect to time and using Eq. (2.1) yield the entropy balance equation

$$\dot{h} = -\frac{1}{\varrho}(\partial \cdot \Phi) + \sigma \tag{2.4}$$

in which

$$\begin{aligned} h &:= S/\varrho, \\ \varrho &:= m \int f\, d^N c, \\ \Phi &:= -k_B \int \bar{c} f \ln(\mathbb{C} f)\, d^N c, \\ \sigma &:= -(k_B/\varrho) \int J(f) \ln(\mathbb{C} f)\, d^N c, \\ \bar{c} &:= c - u, \\ u &:= (m/\varrho) \int c f\, d^N c. \end{aligned} \tag{2.5}$$

Here and henceforth, $\partial := \partial_x$ and an overdot indicates the substantial time derivative defined by $\dot{A} := \partial_t A + u \cdot \partial_x A$. Interpreting Eqs. (2.5), h is the specific entropy, ϱ is the mass density, Φ is the entropy flux, σ is the nonnegative entropy production, $\bar{c}$ is the peculiar velocity, and u is the macroscopic velocity.

2.2. The direct Legendre transform of Boltzmann's entropy

In order to compare and contrast the predictions of various theories of thermodynamic potentials, we begin our discussion by introducing the standard Legendre transform of Boltzmann's entropy. Given a clear statement as to what this transform is, one should be in a much better position to understand what the alternative method of Sec. 3 really entails. Since Boltzmann's entropy is a functional of f, the variation of $S(f)$ can be written as

$$\delta S = \int \lambda \, \delta f \, d^N c, \tag{2.6}$$

where

$$\lambda(c) := \delta S / \delta f(c) = -k_B \left[1 + \ln(\mathbb{C} f)\right]. \tag{2.7}$$

In Eq. (2.7) the dependence of λ on x and t is not shown explicitly in order to make the resulting formulas shorter. The relation (2.7) can be solved uniquely for f in the form

$$f = \frac{1}{\mathbb{C}} \exp\left(-1 - \frac{\lambda}{k_B}\right). \tag{2.8}$$

We define the Legendre transform of $S(f)$ as

$$F(\lambda) := S(f) - \int \lambda f \, d^N c. \tag{2.9}$$

In view of the kinetic-theory definition (2.2) of S, we then find from Eqs. $(2.5)_2$, (2.7), and (2.9) that F is proportional to the mass density ϱ:

$$F = (k_B/m)\varrho. \tag{2.10}$$

The variation of $F(\lambda)$ as a functional of λ is given by

$$\delta F = -\int f \, \delta\lambda \, d^N c, \tag{2.11}$$

where

$$f(c) = -\delta F / \delta\lambda(c). \tag{2.12}$$

The equivalence of $F = F(\lambda)$ with $S = S(f)$ is evident from the fact that the latter can be regained from the former. The relation is dual in the sense that the inverse and direct relations have the same form, except for a sign in the equation of the Legendre transformation.

Following these lines, a *formal* theory of nonequilibrium potentials can easily be founded parallel to the theory of equilibrium thermodynamics, but in such a method of dealing with S there does not appear to be a way that the "nonequilibrium inverse temperature" $1/T$ and the "thermodynamic pressure" p could be associated with the derivatives of $h = S/\varrho$ with respect to the specific internal energy ε (per unit mass) and the specific volume $v = 1/\varrho$:

$$\frac{1}{T} = \frac{\partial h}{\partial \varepsilon}, \qquad p/T = \frac{\partial h}{\partial v}. \tag{2.13}$$

Thus another way of dealing with Boltzmann's entropy must be proposed, and in Sec. 3 a calculation is made to *prove* that the generalized Gibbs formulas (2.13) are valid *only* for a suitably chosen parametrization of the space of nonequilibrium states.

The problem has to do with the freedom of choosing independent variables in terms of which we could describe the nonequilibrium state of a Boltzmann gas. If, instead of considering the "fundamental equation" $S = S(f)$ with f as independent "variable," we had replaced f by (v, ε, G) and then considered v, ε, and G as new independent variables([1]), we could have defined another set of thermodynamic potentials by performing the Legendre transformations on $h = h(v, \varepsilon, G)$ rather than on $S = S(f)$. The corollary of this observation is as follows: the kinetic theory in itself does not provide a precise definition of what one means by the Legendre transform of Boltzmann's entropy, and some additional specifications are still necessary to make this definition precise. They are formulated in the text below.

3. Further Legendre transformations

3.1. A Maxwellian molecular density

To carry on the intended analysis of the aforementioned questions, it is useful to define a few mathematical quantities. First, we define the specific internal energy ε and the reduced peculiar velocity κ by

$$\varepsilon := (m/2\varrho) \int |\bar{c}|^2 f \, d^N c \tag{3.1}$$

and

$$\kappa := \alpha \bar{c}, \tag{3.2}$$

([1]) See Sec. 3.1 for the definition of v, ε and G.

where

$$\alpha := (N/2\varepsilon)^{1/2}. \tag{3.3}$$

Inspection shows that κ is dimensionless. With the specific internal energy and the reduced peculiar velocity so defined, we now construct f_M as follows:

$$f_M(x,\kappa,t) := [\varrho(x,t)/m][\alpha(x,t)]^N \Omega(\kappa), \tag{3.4}$$

where ϱ is the mass density of Eq. $(2.5)_2$ and $\Omega(\kappa)$ is given by

$$\Omega(\kappa) := \left(\frac{1}{2\pi}\right)^{N/2} \exp\left(-\frac{1}{2}|\kappa|^2\right). \tag{3.5}$$

Such an f_M is called a Maxwellian molecular density.

As a measure of the deviation of f from f_M we suggest

$$G := \frac{1}{f_M}(f - f_M). \tag{3.6}$$

The natural, independent variables of this function are x, κ, and t; thus $G = G(x,\kappa,t)$. The same remark concerns f_M. *However, for simplicity, the dependence of f_M and G on x and t will not be shown explicitly.* Hence we have for f

$$f(x,c,t) = f_M(\kappa)[1 + G(\kappa)]. \tag{3.7}$$

Also, in virtue of the definitions of $(\varrho, u, \varepsilon)$, we immediately see that

$$\int \Omega \mathcal{G} G \, d^N\kappa = 0, \tag{3.8}$$

where

$$\mathcal{G} := 1, \quad \kappa, \quad |\kappa|^2. \tag{3.8'}$$

To summarize, even though f_M does not satisfy the Boltzmann equation (2.1), we can always write f in the form (3.7) and thus uniquely represent f in terms of ϱ, u, ε, and G. This representation and the relations (3.8) are *exact and are automatically assured* if the distribution function f obeys the following conditions:

$$\int |c|^n f \, d^N c < \infty, \qquad n = 0, \ldots, 3. \tag{3.9}$$

The use of these conditions introduces a natural class of distribution functions which are considered to prove the existence of the equations of balance of ϱ, u and ε.

3.2. A canonical form of Boltzmann's entropy

If we substitute the decomposition (3.7) into Eq. (2.2), then by Eqs. (3.4) and (3.8) we obtain for $h = S/\varrho$

$$h(v, \varepsilon, G) = h_E(v, \varepsilon) + \Delta(G), \tag{3.10}$$

where $v = 1/\varrho$ and

$$\begin{aligned} h_E(v, \varepsilon) &:= \frac{N}{2}(k_B/m) - (k_B/m) \ln \left[\mathbb{C} \frac{1}{mv} \left(\frac{N}{4\pi\varepsilon} \right)^{N/2} \right], \\ \Delta(G) &:= -(k_B/m) \int \Omega(1 + G) \ln(1 + G)\, d^N\kappa. \end{aligned} \tag{3.11}$$

The separation of h in Eq. (3.10) into h_E and Δ has a clear physical significance: h_E represents the thermostatic entropy (which is a function of v and ε), while Δ gives the functional contribution to h independent of (v, ε) and vanishing at equilibrium. This result shows that the change of dynamical variables, namely, the transition from f to (v, ε, G), enables one to obtain a particularly useful parametrization of the space of nonequilibrium states. Indeed, on applying such a parametrization, one can easily prove that $\partial h/\partial v$ and $\partial h/\partial \varepsilon$ are the same functions of v and ε as in equilibrium. Precisely speaking, the infinitesimal variation δh and the substantial time derivative of h assume the form

$$\begin{aligned} \delta h &= (p/T)\delta v + \frac{1}{T}\delta\varepsilon + \int \Theta\, \delta G\, d^N\kappa, \\ \dot{h} &= (p/T)\, \dot{v} + \frac{1}{T}\, \dot{\varepsilon} + \int \Theta\, \dot{G}\, d^N\kappa, \end{aligned} \tag{3.12}$$

where $\dot{G}(x, \kappa, t) := (\partial_t + u \cdot \partial_x) G(x, \kappa, t)$, δG is the infinitesimal variation of G consistent with the constraints (3.8) and the obvious inequality $G > -1$([2]), and

$$\begin{aligned} p/T &:= \frac{\partial h}{\partial v} = \partial h_E/\partial v = k_B(\varrho/m), \\ \frac{1}{T} &:= \frac{\partial h}{\partial \varepsilon} = \partial h_E/\partial \varepsilon = N k_B/2m\varepsilon, \\ \Theta(\kappa) &:= \delta h/\delta G(\kappa) = \delta\Delta/\delta G(\kappa) = -(k_B/m)\Omega[1 + \ln(1 + G)]. \end{aligned} \tag{3.13}$$

We shall refer to Eqs. (3.12) as the generalized Gibbs relations (equations).

From Eqs. (3.10) and $(3.12)_1$ it follows that we can determine h and δh without knowing the particular kinetic process occurring, and without regard to the time

([2]) It is not difficult to prove the existence of such variations δG of G. In this context, we wish to note that δG is a function of x, κ, and t.

and the place. In other words, the values of the specific entropy and its variation are ascertained from the information which is static and universal. This information consists in the form of the dependence of h upon v, ε, and G. Clearly, the relation $(3.12)_2$ is a direct consequence of Eq. $(3.12)_1$. However, this relation is not "static" and "universal," because it holds only for those distribution functions which are solutions of Boltzmann's equation.

In the theory of Boltzmann, the temperature T is regarded as but another name for the expected kinetic energy of relative motion; thus

$$T = \frac{2m\varepsilon}{Nk_B}. \tag{3.14}$$

Moreover, if we let $p = (2/N)\varrho\varepsilon$ denote the mean pressure, then from Eq. (3.14) it follows that the "ideal gas law" holds for every condition of the gas:

$$p = (\varrho/m)k_B T. \tag{3.15}$$

Another way to define T and p is through Eqs. $(3.13)_1$ and $(3.13)_2$; *both methods lead to the same result.*

We are now in a position to discuss certain problems regarding the structure of an expression for the entropy flux Φ. Examination of Eqs. $(2.5)_3$ and (3.7) makes it readily apparent that Φ can exactly be written as

$$\Phi = (q/T) - (k_B\varrho/m\alpha)\int \kappa\Omega(1+G)\ln(1+G)\,d^N\kappa, \tag{3.16}$$

where α is defined by Eq. (3.3) and q is the heat flux:

$$q := \frac{m}{2}\int |\bar{c}|^2\bar{c}f\,d^N c. \tag{3.17}$$

Our analysis here shows how the quantity q/T enters the general expression for Φ naturally. By Eq. (3.16) we see, however, that not only net heating flux gives rise to Φ. The above calculations also show that if we are to define the entropy flux Φ on the basis of kinetic theory, then it is necessary to identify the "nonequilibrium temperature" T with $2m\varepsilon/Nk_B$. Of course, for gas flows sufficiently near to local equilibrium in the sense that f differs little from the corresponding f_M, we can linearize the integral part of Eq. (3.16) with respect to G and so conclude from the constraints (3.8) that Φ approximately equals q/T. No such approximations are possible, however, in the nonlinear case.

One final word concerning the results just obtained. Given the natural condition (3.9) of Sec. 3.1, we have shown that use of the decomposition $f = F_M(1+G)$ in Eqs. $(2.5)_1$ and $(2.5)_3$ yields the specific entropy h in Eq. (3.10) and the entropy flux Φ in Eq. (3.16) in terms of v, ε, and G. Consequently, within the framework set up here, the formula (3.16) emerges in confirmation to the thermodynamic

principles, and an extended Gibbs-relation-like one-form $(3.12)_2$ is consistent with the entropy law as characterized by Eq. (2.4). The obvious reason for this consistency is the fact that Eqs. (3.10), (3.12), and (3.16) are identities. Of course, we can also test our results directly by substituting these identities and an appropriate expression for the entropy production σ into Eq. (2.4). The details of this somewhat elaborate programme will not be presented, however, because a hint of what to expect may be obtained from considerations of Sec. 5.

3.3. The extremum principles

We easily conclude from Eqs. (3.8), (3.10), (3.11), and the inequality

$$(1 + G)\ln(1 + G) - G \geq 0 \tag{3.18}$$

that of all states (v, ε, G) with given values of v and ε, the equilibrium state $(v, \varepsilon, 0)$ has the greatest specific entropy h (*the entropy maximum principle*):

$$h(v, \varepsilon, G) \leq h_E(v, \varepsilon). \tag{3.19}$$

Here h equals h_E if and only if $G = 0$.

As a further systematic step, it is plausible to express ε in terms of v, h, and G. In fact, by solving Eq. (3.10) for the specific internal energy ε we find that

$$\varepsilon(v, h, G) = \frac{N}{4\pi}\left(\frac{\mathbb{C}}{mv}\right)^{2/N} \exp\left\{\frac{2m}{Nk_B}[h - \Delta(G)] - 1\right\}. \tag{3.20}$$

Then, beginning from $\Delta \leq 0$, we see that among all states (v, h, G) having the same values of v and h, the equilibrium state $(v, h, 0)$ gives ε its smallest value (*the energy minimum principle*). Moreover, we have

$$\begin{aligned} p &= -\frac{\partial \varepsilon}{\partial v}, \qquad T = \frac{\partial \varepsilon}{\partial h}, \\ \Theta(\kappa) &= -\frac{1}{T}[\delta\varepsilon/\delta G(\kappa)]. \end{aligned} \tag{3.21}$$

Hence

$$\dot{\varepsilon} = -p\,\dot{v} + T\,\dot{h} - T\int \Theta\,\dot{G}\,d^N\kappa. \tag{3.22}$$

Clearly, this result is consistent with Eq. $(3.12)_2$.

The specific free energy

$$\begin{aligned} \varphi(v, T, G) &:= \varepsilon(T) - Th(v, \varepsilon(T), G) \\ &= (k_B T/m)\ln\left[\mathbb{C}\frac{1}{mv}\left(\frac{m}{2\pi k_B T}\right)^{N/2}\right] - T\Delta(G) \end{aligned} \tag{3.23}$$

is that partial Legendre transform of ε which replaces the entropy h by the temperature T as independent variable. The substantial time derivative $\dot{\varphi}$ is

$$\dot{\varphi} = -p\dot{v} - h\dot{T} - T\int \Theta \dot{G}\, d^N\kappa. \tag{3.24}$$

In addition, from Eq. (3.23) it follows that the equilibrium state $(v, T, 0)$ minimizes the specific free energy φ, not absolutely, but over the space of states (v, T, G) with given values of v and T (*the free-energy minimum principle*).

Now, we define the specific enthalpy H and the Gibbs function Q as $H := \varepsilon + vp$ and $Q := \varepsilon - Th + vp$, respectively. The specific enthalpy H is that partial Legendre transform of ε which replaces the specific volume v by the pressure p as independent variable. The Gibbs function Q in turn is the Legendre transform of ε which simultaneously replaces the specific entropy h by the temperature T and the specific volume v by the pressure p as independent variables. The physical meaning of these nonequilibrium thermodynamic potentials is apparent from the differential expressions obtained on using Eq. (3.22):

$$\dot{H} = v\dot{p} + T\dot{h} - T\int \Theta \dot{G}\, d^N\kappa, \tag{3.25}$$

$$\dot{Q} = v\dot{p} - h\dot{T} - T\int \Theta \dot{G}\, d^N\kappa. \tag{3.26}$$

Moreover, from

$$H = H(p, h, G) = \frac{(N+2)}{4\pi}\left(\frac{2\pi}{m}\mathbb{C}p\right)^{2/(N+2)} \times \exp\left\{\frac{2m}{(N+2)k_B}[h - \Delta(G)] - \frac{N}{N+2}\right\} \tag{3.27}$$

one can prove that among all states (p, h, G) having the same values of p and h, the equilibrium state $(p, h, 0)$ gives H its smallest value (*the enthalpy minimum principle*). Similarly, using

$$Q = Q(p, T, G) = (k_BT/m) + (k_BT/m)\ln\left[\mathbb{C}(p/k_BT)\left(\frac{m}{2\pi k_BT}\right)^{N/2}\right] - T\Delta(G), \tag{3.28}$$

one concludes that the equilibrium state $(p, T, 0)$ minimizes the Gibbs function Q, not absolutely, but over the space of states (p, T, G) with given values of p and T (*the Gibbs-function minimum principle*).

The total Legendre transform of ε is defined by

$$\mathcal{E}(p, T, \Theta) := Th - vp - T\int \Theta G\, d^N\kappa - \varepsilon = -(k_BT/m)\ln\left[\mathbb{C}(p/k_BT)\left(\frac{m}{2\pi k_BT}\right)^{N/2}\right] + T\int \Theta\, d^N\kappa. \tag{3.29}$$

A glance at Eq. (3.13)$_3$ shows that the equilibrium value Θ_E of Θ is $\Theta_E = -(k_B/m)\Omega$. Also, with Eq. (3.8) for $\mathcal{G} = 1$ and the inequality $\ln(1 + G) - G \leq 0$, we find that

$$\int (\Theta - \Theta_E)\, d^N\kappa \geq 0. \tag{3.30}$$

Hence

$$\mathcal{E}(p, T, \Theta) \geq \mathcal{E}(p, T, \Theta_E), \tag{3.31}$$

the equality holding if and only if $\Theta = \Theta_E$. The substantial time derivative of $\mathcal{E}$ assumes the form

$$(\mathcal{E})^{\cdot} = -v\dot{p} + h\dot{T} - \int G(T\Theta)^{\cdot}\, d^N\kappa. \tag{3.32}$$

Another set of functions (Massieu's functions) can be defined by performing the Legendre transformations on $h(v, \varepsilon, G)$ rather than on $\varepsilon(v, h, G)$. As the theory of these functions is very much analogous to that already made familiar, we will not discuss this theory further here; specifically, we will not derive the maximum principles for the Massieu functions([3]).

4. The method of moments

To study the consequences of using the method of moments, we introduce the Hilbert space $\mathcal{H}$ in which the scalar product $\langle \omega_1, \omega_2 \rangle$ is defined by

$$\langle \omega_1, \omega_2 \rangle := \int \Omega(\kappa)\omega_1(\kappa)\omega_2(\kappa)\, d^N\kappa. \tag{4.1}$$

We can determine the exact moment representations of h and $\dot{h}$ if we assume that $1 + G$, $\ln(1 + G)$, and $\dot{G}$ are elements of $\mathcal{H}$. Then by use of the *complete set* of tensor Hermite polynomials $B^n(\kappa)$, $n = 0, 1, ..., \infty$, it is possible to represent $1 + G$, $\ln(1 + G)$, and $\dot{G}$ by the expansions [9]

$$\begin{aligned} 1 + G &= Y := \sum_{n=0}^{\infty} \frac{1}{n!} b^n \cdot B^n, \\ \ln(1 + G) &= \sum_{n=0}^{\infty} X^n \cdot B^n, \\ \dot{G} &= \sum_{n=2}^{\infty} \frac{1}{n!} \dot{b}^n \cdot B^n, \end{aligned} \tag{4.2}$$

([3]) Of course, we can also derive the minimum principles, this being purely a matter of convention in the choice of the sign of the function.

where b^n and X^n are the expansion coefficients and the symbol $\cdot$ denotes the inner product of the tensors involved. Clearly, because of the constraints (3.8), we have

$$b^0 = 1, \qquad b^1 = 0, \qquad \mathrm{Tr}\, b^2 = 0, \tag{4.3}$$

where Tr is the trace operator. The above series converge in the sense of the norm in $\mathcal{H}$. However, to express X^n in terms of b^n, we must first assume that the series Y converges *both* pointwise for each κ and in the sense of $\mathcal{H}$ and then substitute Eq. $(4.2)_1$ into

$$X^n = \frac{1}{n!} \int \Omega B^n \ln(1 + G)\, d^N\kappa. \tag{4.4}$$

Here we remark that if the above conditions are not satisfied, then the moment representations of h and $\dot{h}$ described below are not expected to exist. This gives us necessary information about what is and is not possible. In the recent analysis [10] presented by EU [see, e.g., his equations (2.30) and (2.33)], an explicit assumption was made that the expansion Y in Eq. $(4.2)_1$ converges to $1 + G$ in the sense of means and that this rather weak condition is sufficient to see the method of moments in action (i.e., to express X^n in terms of b^n). The pointwise convergence of Y just deduced clearly suggests it to the contrary.

From Eqs. (3.10)–(3.13) and (4.2) plus the orthogonality properties of Hermite polynomials [9], the moment representations of h and $\dot{h}$ are as follows:

$$h = h_E - (k_B/m) \sum_{n=0}^{\infty} X^n \cdot b^n, \tag{4.5}$$

$$\dot{h} = (p/T)\, \dot{v} + \frac{1}{T}\, \dot{\varepsilon} - (k_B/m) \sum_{n=2}^{\infty} X^n \cdot \dot{b}^n. \tag{4.6}$$

Consistency([4]) between Eqs. (4.5) and (4.6) follows directly from the considerations of Appendix B [cf. Eqs. $(B.2)_6$ and $(B.3)_6$]. We can similarly analyze the kinetic-theory expression for the entropy flux Φ. In fact, putting the expansions $(4.2)_1$ and $(4.2)_2$ into Eq. (3.16), we find that [cf. also Eq. (B.7) in Appendix B]

$$\Phi = (q/T) - (k_B \varrho/m\alpha) \sum_{n=1}^{\infty} \left(X^n \cdot b^{n+1} + n X^n \cdot b^{n-1} \right), \tag{4.7}$$

where the heat flux q is related to $\mathrm{Tr}\, b^3$ by

$$\mathrm{Tr}\, b^3 = \frac{2}{\varrho} \left(\frac{N}{2\varepsilon} \right)^{3/2} q. \tag{4.8}$$

([4]) The series in Eqs. (4.5)–(4.7) converge absolutely if $1 + G$, $\kappa(1 + G)$, $\dot{G}$, and $\ln(1 + G)$ are elements of $\mathcal{H}$.

However, we have to worry about the convergence of the series in Eq. (4.7); this may require consideration of the situation in which also $\kappa(1+G)$ is an element of $\mathcal{H}$. From the viewpoint of the present paper, the method of moments is a formal and sophisticated way of deriving the generalized Gibbs relation $(3.12)_2$ for the specific Boltzmann entropy h. In Secs. 2 and 3, we have seen that there are simpler and more natural ways of deriving this relation. One obvious reason for this is that, with the technique of functional differentiation, we can draw *definite* and *exact* conclusions about the existence of Eqs. (3.12) without making any explicit or implicit reference to $\mathcal{H}$. This is crucial because representative and physically important cases are known [11] in which solutions of the kinetic equation do not exist in the Hilbert space chosen: $G \notin \mathcal{H}$. This fact detracts much from the usefulness of Hermite expansions (and of various *ad hoc* truncation and projection procedures) at the level of the *nonlinear* Boltzmann or Boltzmann-like equations.

We recall that the nonnegative entropy production σ is given by Eq. $(2.5)_4$ and conclude from Eqs. (3.7) and $(4.2)_2$ that

$$\sigma = -(k_B/m)\sum_{n=0}^{\infty} X^n \cdot P^n, \tag{4.9}$$

where

$$P^n := (m/\varrho)\int B^n J(f)\, d^N c \tag{4.10}$$

and where

$$P^0 = 0, \qquad P^1 = 0, \qquad \operatorname{Tr} P^2 = 0. \tag{4.11}$$

The exact and/or tractable moment representations of P^n are not expected to exist, except in the case of Maxwellian molecules.

In Sec. 5 we shall verify that if we use the formulas (4.5), (4.7), and (4.9) in the balance equation (2.4) for h, then this balance equation will be automatically satisfied, at least formally. Before doing so, however, it is necessary to derive the evolution equations for ϱ, u, ε, and b^n.

The equations of balance of ϱ, u, and ε are easily obtained from the Boltzmann equation under the natural assumption that f falls off sufficiently rapidly for large values of c:

$$\begin{aligned} \dot{\varrho} &= -\varrho(\partial \cdot u), \\ \dot{u} &= -\frac{1}{\varrho}\big[\partial \cdot (pI + \omega)\big], \\ \dot{\varepsilon} &= -\frac{1}{\varrho}(\partial \cdot q) - \frac{1}{\varrho}(pI + \omega) \cdot L. \end{aligned} \tag{4.12}$$

The abbreviated symbol I stands for the unit tensor of a Euclidean vector space $\mathbb{E}$ ($\dim \mathbb{E} = N$), L is the spatial gradient of u ($L := \partial u$), and ω is the stress deviator defined by

$$\omega := m \int \left(\bar{c} \otimes \bar{c} - \frac{1}{N} |\bar{c}|^2 I \right) f \, d^N c. \tag{4.13}$$

This stress deviator is proportional to b^2:

$$b^2 = \frac{N}{2\varrho\varepsilon} \omega. \tag{4.14}$$

By use of the notation introduced in Appendix A we obtain from

$$b^n = \frac{m}{\varrho} \int B^n f \, d^N c \tag{4.15}$$

and the Boltzmann equation (under usual assumptions) the following result:

$$\dot{b}^n = Z^n + P^n, \tag{4.16}$$

where

$$\begin{aligned} Z^n := & -\frac{1}{\alpha} \left(\partial \cdot b^{n+1} + n \partial \vee b^{n-1} \right) - \frac{1}{\alpha \varrho} \left(\varrho' \cdot b^{n+1} + n \varrho' \vee b^{n-1} \right) \\ & - n\alpha (\dot{u} \vee b^{n-1}) - n \left[L \cup b^n + (n-1) L \vee b^{n-2} \right] \\ & - \frac{\alpha}{N} \left[(n+1) \varepsilon' \cdot \psi^{n+1} + n(n-1) \varepsilon' \vee \psi^{n-1} \right] - \frac{n \dot{\varepsilon}}{2\varepsilon} \psi^n, \end{aligned}$$

$$\begin{aligned} & \varrho' := \partial \varrho, \qquad \varepsilon' := \partial \varepsilon, \\ & \psi^n := b^n + (n-1) I \vee b^{n-2}, \\ & b^{-3} := 0, \qquad b^{-2} := 0, \qquad b^{-1} := 0, \\ & n = 0, 1, \ldots, \infty. \end{aligned} \tag{4.16'}$$

This is the desired system of equations for the coefficients b^n in the expansion $(4.2)_1$ of $1 + G$. Inspection shows that Eqs. (4.16) are automatically satisfied if n equals 0 or 1, because of Eqs. (4.12); moreover, $(\operatorname{Tr} b^2)^{\cdot} = \operatorname{Tr} \dot{b}^2 = \operatorname{tr} Z^2 = \operatorname{tr} P^2 = 0$. Another remark is also in order. To obtain a manageable system of "extended" differential equations, the infinite set of moments has to be truncated and some procedures for expressing P^n in terms of b^n must be proposed. However, to the best of our knowledge, it still remains an open question whether such a truncution procedure is consistent with kinetic theory.

The manner in which these calculations form the first step in the formal derivation of $\dot{h} = -\varrho^{-1}(\partial \cdot \Phi) + \sigma$ from Eqs. (4.5), (4.7), and (4.9), will become clear in the text below.

5. Consistency between the generalized Gibbs relation and the entropy law

First of all, there is no question that Eqs. (3.12) and (4.6) are consistent with the entropy law (2.4) because these equations are identities. Nevertheless, since this point has been a subject of debate [12–14] in the past, here the internal consistency of the formalism will be demonstrated from still another viewpoint. To achieve the objective in mind, we first substitute Eqs. (4.7) and (4.9) into the entropy law (2.4) and then establish the following identity by using Eqs. (3.3), (4.8), (4.12), (4.14), (4.16), and the definitions of various quantities involved:

$$\dot{h} = (p/T)\,\dot{v} + \frac{1}{T}\,\dot{\varepsilon} - (k_B/m)\sum_{n=2}^{\infty} X^n \cdot \dot{b}^n + (k_B/m\varrho T)\,C_h\,, \tag{5.1}$$

where

$$\begin{aligned} C_h &:= \sum_{n=1}^{\infty}\Big[T(\partial\cdot\vartheta_n) + \varrho T(X^n\cdot Z^n) + (b^n\cdot\chi^n)\Big],\\ \vartheta_n &:= \varrho\alpha^{-1}\Big[X^n\cdot b^{n+1} + nX^n\cdot b^{n-1}\Big],\\ \chi^2 &:= \varrho T L, \qquad \chi^3 := \frac{\alpha}{N}\varrho T(I\vee\varepsilon'),\\ \chi^n &:= 0 \qquad \text{for} \quad n\neq 2,3. \end{aligned} \tag{5.2}$$

The expansion in Eq. $(5.2)_1$ starts from $n = 1$, because $\vartheta_0 = 0$ and $\vartheta_n \neq 0$ when $n > 0$. Combining Eqs. (4.6) and (5.1), we obtain

$$C_h = 0. \tag{5.3}$$

We call this equation the *consistency condition* because its role in essence is that of a guarantor of the generalized Gibbs relation (4.6) for entropy change. In order to demonstrate the internal consistency of the formalism, it is thus stimulating to show that Eq. (5.3) holds for all conditions of the Boltzmann gas.

Now, we shall prove that C_h can indeed be set equal to zero without encountering any internal contradiction. This proof generalizes to N-dimensional systems the conclusion([5]) formulated directly before Eq. (4.18) in [12]. A straightforward application of Eqs. (3.3), (4.16), (5.2), $(B.2)_1$, and $(B.3)_1$ yields

$$\frac{1}{\varrho T}C_h = -\alpha(\dot{u}\cdot\mathcal{D}_1) - \frac{\dot{\varepsilon}}{2\varepsilon}\mathcal{D}_2 - \frac{\alpha}{N}(\varepsilon'\cdot\mathcal{D}_3) - L\cdot\mathcal{D}_4 + \frac{1}{\alpha}\mathcal{D}_5\,, \tag{5.4}$$

([5]) The notation in [12] slightly differs from ours as follows: ours $\Rightarrow$ his; $N \Rightarrow 1$; $k_B \Rightarrow 1$; $m \Rightarrow 1$; $c \Rightarrow \lambda$; $\bar{c} \Rightarrow \bar{\lambda}$; $\partial \Rightarrow \nabla$; $\varepsilon \Rightarrow \epsilon$; $\dot{A} \Rightarrow d_t A$; $\psi^n \Rightarrow \alpha\varrho^{-1}\psi^{n-1}$; $X^0 \Rightarrow T^{-1}(\mu - T\nu)$; $X^n \Rightarrow -T^{-1}X^n$ for $n \neq 0$; $\vartheta_n \Rightarrow -T^{-1}(X^n\cdot\psi^n)$; $\chi^2 \Rightarrow 0$; $\chi^3 \Rightarrow -\chi^3$. Similar comparisons of our consistency condition $C_h = 0$ (specialized to the case $N = 1$) with the corresponding equation (25) of [14] are not possible, however, because in this equation the meaning of the symbol $\sum_n$ is not clearly explained.

where the objects $\mathcal{D}_1, ..., \mathcal{D}_5$ are defined by Eqs. (B.2). However, from the considerations of Appendix B it follows that $\mathcal{D}_k = 0$ when $k = 1, ..., 5$. This completes the proof of Eq. (5.3). The validity of $C_h = 0$ and hence of Eq. (4.6) is also obvious on the intuitive ground.

Thus contrary to the suggestion made in the literature [14], there does appear to be a way that the terms in Eq. $(5.2)_1$ cancel each other so neatly that $C_h = 0$ when Eqs. (4.2) and (4.4) hold; in other words, when the complete set is taken for $B^n(\kappa)$. This implies the second conclusion: there is no kinetic-theory foundation for including the *compensation function* (or the *calortropy*) in the thermodynamic description of Boltzmann's gas, as it is done in the so-called revised version of the modified moment method [13] or in its possible further corrections (see especially the discussion on p. 7177 after Eq. (3.30) in [10]). Such is indeed the case because the generalized Gibbs relation (4.6) holds for the entropy density h itself, and not for the compensation function which appears to be extraneous and redundant. The same observation concerns the notion of calortropy. To be more precise, substitution of the formula $(4.2)_2$ into Eq. (3.1) in [10] yields the conclusion that the calortropy does not differ from Boltzmann's entropy.

The gist of the point made by the present analysis is that the information contained in the generalized Gibbs relation formally does not contract as the level of description is passed from the phase-space level (3.12) *to that at the moment level* [cf. Eq. (4.6)], *since the passage essentially involves a complete set of Hermite polynomials*. Moreover, after expressing X^n in terms of b^n, as is formally always possible [cf., e.g., our analysis directly after Eq. (4.3)], the Boltzmann entropy h becomes a state function in a space spanned by the "thermodynamic" variables $\mathfrak{S} := \{v, \varepsilon, b^n \,|n = 2, 3, ..., \infty\}$, and thus one can think of $\dot{h}$ as being an exact differential in $\mathfrak{S}$. On the basis of such results, it is possible to infer that, as was already found in earlier work [12], a thermodynamic interpretation of kinetic theory may be erected on Boltzmann's entropy alone, i.e., without the necessity [10, 13] of referring to the "concepts" of compensation function and calortropy. However, the method of moments is surely not very useful in practice. This method obscures the real situation: it suggests that there is something very special about the way the theory of thermodynamic potentials is related to Hermite expansions, whereas in reality this is not so. A deductive mathematical way for exhibiting the generalized Gibbs relation has been proposed in Sec. 3 and is clearly linked to the technique of functional differentiation.

6. Discussion and conclusion

We have found a set of thermodynamic potentials for the description of a Boltzmann gas. We have obtained the extremum principles for these potentials, and examined their physical meaning in the simplest case (a one-component gas).

The basis for the initial analysis was Boltzmann's entropy which is a functional of the single particle distribution function, not of the fields. This entropy was then divided into two parts, that associated with the local distribution (and hence yielding a standard function of conserved variables) plus the remainder. Using the technique of functional differentiation, we have proved that if the independent gas-state variables are (v, ε, G), the quantities $Nk_B/2m\varepsilon$ and $2\varrho\varepsilon/N$, which everybody would write down as the only natural concepts in the situation of a classical rarefied gas, exactly correspond to the "nonequilibrium inverse temperature" $1/T$ and the "thermodynamic pressure" p as defined by investigating the partial derivatives of Boltzmann's entropy $h(v, \varepsilon, G)$ with respect to v and ε.

Comparison with the usual approach shows that we can obtain the same results as usual but in a much more transparent way, because in the standard approach (the method of moments) the existence of thermodynamic potentials depends on the convergence of the following series:

$$\Im(x,t) := \sum_{n=0}^{\infty} X^n(x,t) \cdot b^n(x,t), \tag{6.1}$$

where (x,t) is an arbitrary space-time point. On the other hand, in order to use the method of moments to draw valuable conclusions about the "thermodynamic branch" [12] of solutions of Boltzmann's equation, it would be necessary to have not only the convergence of $\Im$ for arbitrary space-time points but also some information about uniformity (in space-time) of convergence; the existence theorems for Boltzmann's equation give no indication that there will be any such uniformity. Furthermore, the divergence of Y [cf. Eq. $(4.2)_1$] in certain important cases makes uniformity of convergence problematical. For a discussion of these divergences, see, for example, [11].

The situation is different, however, with the formalism of Sec. 3, for its equations give rise to the exact theory of thermodynamic potentials independent of any *ad hoc* assumptions and artificial constructions. To summarize, the technique of functional differentiation is an adequate tool to study the mathematical and physical status of the generalized Gibbs relation at the level of Boltzmann's equation. The point of this discussion is that instead of concentrating on the formal Hermite expansion (4.6) of $\dot{h}$, with the ambiguities that implies, we can deal directly with Eqs. (3.12).

At first sight, it seems that while the questions/problems posed here apply for general systems, their answers/resolutions must be limited to classical rarefied gases. *But this is not the case.* In fact, we have already verified that our ideas are quite universal and can be extended in a number of directions, one of them being the analysis of mixtures and quantum Bose-Einstein or Fermi-Dirac nonequilibrium ideal gases. However, since these extensions are not altogether trivial or immediate, they will be treated in a separate paper.

Appendix A. Some useful abbreviations

To make the resulting formulas shorter, in this paper we have introduced essentially the same notation as in [15–17]. Let $\mathbb{E}$ be a Euclidean vector space ($\dim \mathbb{E} = N$). Choose an orthonormal basis $\{e_1, ..., e_N\}$ in $\mathbb{E}$ and set $e_{r_1 ... r_\alpha} := e_{r_1} \otimes \cdots \otimes e_{r_\alpha}$.

1. The action of the symmetrizer Π on a tensor M^α of degree α is given by

$$\Pi M^\alpha := \sum_{r_1 ... r_\alpha = 1}^{N} M^\alpha_{(r_1 ... r_\alpha)} e_{r_1 ... r_\alpha}, \tag{A.1}$$

where the coefficients $M^\alpha_{r_1 ... r_\alpha}$ are components of M^α with respect to $\{e_{r_1 ... r_\alpha}\}$ and parentheses enclosing a set of α indices represent symmetrization of these indices, i.e., the sum over $\alpha!$ permutations of the indices, divided by $\alpha!$

2. Suppose that M^α and M^β are the tensors of degrees α and β, respectively. Then the equality

$$M^\alpha \vee M^\beta := \Pi \left(M^\alpha \otimes M^\beta \right) \tag{A.2}$$

defines the symmetric tensor product of M^α and M^β.

3. The action of $\cup$ on M^α and M^β is characterized by

$$M^\alpha \cup M^\beta := \Pi \operatorname*{Tr}_{(1,\alpha+1)} \left(M^\alpha \otimes M^\beta \right), \tag{A.3}$$

where $\operatorname{Tr}_{(1,\alpha+1)}$ is the trace operator with respect to the pair $(1, \alpha + 1)$.

4. Let us suppose that $\nu := \min(\alpha, \beta)$. Then in contracting M^α with M^β the ν-fold contraction is denoted by $\cdot$. The tensor $M^\alpha \cdot M^\beta$ of degree $\alpha + \beta - 2\nu$ is usually termed the inner tensor product of M^α and M^β. However, if M^α and M^β are not totally symmetric tensors, then some convention as to which of the 2ν indices are to be contracted, must be followed when doing the contraction. In this context, see the precise definition of $M^\alpha \cdot M^\beta$ in Appendix A of [17].

5. Suppose that M^α and $\overline{M}^\alpha$ are the tensors of degree α. Then the action of $\square$ on M^α and $\overline{M}^\alpha$ is described by

$$M^\alpha \square \overline{M}^\alpha := \sum_{r,s=1}^{N} \left[\left(e_r \cdot M^\alpha \right) \cdot \left(e_s \cdot \overline{M}^\alpha \right) \right] e_r \otimes e_s . \tag{A.4}$$

6. The effect of ∂ on a tensor field M^α is given by

$$\partial M^\alpha := \sum_{r=1}^{N} e_r \otimes \partial_r M^\alpha \tag{A.5}$$

with $\partial_r := \partial / \partial x^r$.

7. The action of $\partial\cdot$ on a tensor field M^α is defined by

$$\partial\cdot M^\alpha := \operatorname*{Tr}_{(1,\alpha+1)}(\partial M^\alpha). \tag{A.6}$$

8. The effect of $\partial\vee$ on a tensor field M^α is characterized by

$$\partial\vee M^\alpha := \Pi(\partial M^\alpha). \tag{A.7}$$

Appendix B. Auxiliary formal properties of b^n and X^n

The effect of $\overline{\partial}$ on a function $\Lambda(\kappa)$ or a tensor field $M^n(\kappa)$ of degree n is given by

$$\begin{aligned}\overline{\partial}\Lambda &:= \sum_{r=1}^{N}\left(\overline{\partial}_r\Lambda\right)e_r\,,\\ \overline{\partial}M^n &:= \sum_{r=1}^{N} e_r\otimes\overline{\partial}_r M^n,\end{aligned} \tag{B.1}$$

where $\overline{\partial}_r := \partial/\partial\kappa^r$.

The objects $\mathcal{D}_k$, $k = 1,\ldots,6$, are defined by

$$\begin{aligned}\mathcal{D}_1 &:= \sum_{n=1}^{\infty} nX^n\cdot b^{n-1},\\ \mathcal{D}_2 &:= \sum_{n=2}^{\infty} nX^n\cdot\psi^n,\\ \mathcal{D}_3 &:= -\operatorname{Tr} b^3 + \sum_{n=1}^{\infty}\left[(n+1)X^n\cdot\psi^{n+1} + n(n-1)X^n\cdot\psi^{n-1} - X^n\cdot b^{n+1}\right],\\ \mathcal{D}_4 &:= -b^2 + \sum_{n=1}^{\infty} n\left[b^n\square X^n + (n-1)X^n\cdot b^{n-2}\right],\\ \mathcal{D}_5 &:= \sum_{n=1}^{\infty}\left[(\partial X^n)\cdot b^{n+1} + n(\partial\cdot X^n)\cdot b^{n-1}\right],\\ \mathcal{D}_6 &:= \sum_{n=0}^{\infty}\dot{X}^n\cdot b^n,\end{aligned} \tag{B.2}$$

where, of course, $\partial = \partial_x$.

It is only a matter of labor to prove that

$$\begin{aligned}
\mathcal{D}_1 &= \frac{m}{\alpha^N \varrho} \int f \overline{\partial}[\ln(1+G)] d^N\kappa = 0,\\
\mathcal{D}_2 &= \frac{m}{\alpha^N \varrho} \int f \left\{\kappa \cdot \overline{\partial}[\ln(1+G)]\right\} d^N\kappa = 0,\\
\mathcal{D}_3 &= \frac{m}{\alpha^N \varrho} \int f\kappa \left\{N + 2 - |\kappa|^2 + \kappa \cdot \overline{\partial}[\ln(1+G)]\right\} d^N\kappa = 0,\\
\mathcal{D}_4 &= \frac{m}{\alpha^N \varrho} \int f \left\{I - \kappa \otimes \kappa + \kappa \otimes \overline{\partial}[\ln(1+G)]\right\} d^N\kappa = 0,\\
\mathcal{D}_5 &= \frac{m}{\alpha^N \varrho} \int f \{\kappa \cdot \partial[\ln(1+G)]\} d^N\kappa = 0,\\
\mathcal{D}_6 &= \frac{m}{\alpha^N \varrho} \int f \left[\ln(1+G)\right]^{\cdot} d^N\kappa = 0,
\end{aligned} \tag{B.3}$$

where $[\ln(1+G)]^{\cdot}$ is the substantial time derivative of $\ln(1+G)$. In these equations, f and G are functions of x, κ, and t.

The proof that the series $\mathcal{D}_k$, $k = 1, \ldots, 6$, can indeed be represented by the above *vanishing* integrals, is based on Eqs. $(2.5)_5$, (3.2), $(4.2)_2$, (4.15), and the following identities [15, 16] for Hermite polynomials $B^n(\kappa)$ [9]:

$$\begin{aligned}
\kappa \otimes B^n(\kappa) &= B^{n+1}(\kappa) + n \sum_{r=1}^{N} e_r \otimes \left[e_r \vee B^{n-1}(\kappa)\right],\\
\overline{\partial} B^n &= n \sum_{r=1}^{N} e_r \otimes \left(e_r \vee B^{n-1}\right),\\
n(n+1)\left[M^n \cdot (I \vee B^{n-1})\right] &= 2n\left(M^n \cdot B^{n-1}\right) + n(n-1)\left[(\mathrm{Tr}\, M^n) \cdot B^{n-1}\right],
\end{aligned} \tag{B.4}$$

where M^n is an arbitrary symmetric tensor of degree n.

The series in Eqs. (B.2) exist and are absolutely convergent if $1+G$, $[\ln(1+G)]^{\cdot}$, $\overline{\partial}\ln(1+G)$, $\kappa \otimes \overline{\partial}\ln(1+G)$, $\kappa \otimes \partial \ln(1+G)$, and $\kappa \otimes \kappa \otimes \overline{\partial}\ln(1+G)$ are elements of $\mathcal{H}$. Clearly, as usual, we must also assume that we can exchange the integral over c or κ with the derivative ∂_x; these assumptions are necessary in order to derive the evolution equations (4.16) for b^n. If these postulates are not satisfied, the method of moments fails to exist.

Using the identity

$$\frac{m}{\alpha^N \varrho} f = \Omega(1+G) \tag{B.5}$$

and integrating by parts, we easily conclude from the constraints (3.8) that the integrals in Eqs. (B.3) vanish; thus

$$\mathcal{D}_k = 0 \tag{B.6}$$

when $k = 1, \ldots, 6$. This observation completes the proof of Eq. (5.3).

A one-dimensional version of these calculations is represented by Eqs. (4.14) in [12]; see also the comments at the bottom of p. 359 in that paper. *All the essential ingredients for the proof of the consistency condition* (5.3) *were thus given seven years ago.*

From $\mathcal{D}_1 = 0$ it follows that Eq. (4.7) simplifies to

$$\Phi = (q/T) - (k_B \varrho / m\alpha) \sum_{n=1}^{\infty} X^n \cdot b^{n+1}. \tag{B.7}$$

References

1. J.W. GIBBS, *The Collected Works of J. Willard Gibbs*, Volume I, Thermodynamics, Yale University Press, 1948.
2. H.B. CALLEN, *Thermodynamics*, John Wiley, New York 1960.
3. L. TISZA, *Generalized thermodynamics*, MIT Press, Cambridge, MA, 1977.
4. M. MORSE, *Calculus of variations in the large*, Ann. Math. Soc. Coll., 4th ed., 1965.
5. V. CIANCIO, *Vectorial thermodynamic variables and heat propagation*, [in:] Proc. of the Vth International Meeting "Waves and Stability in Continuous Media", Sorrento, S. RIONERO [Ed.], World Sci. pp. 66–71, Singapore 1991.
6. Z. BANACH and S. PIEKARSKI, *A coordinate-free description of nonequilibrium thermodynamics*, Arch. Mech., **44**, 191–202, 1992.
7. Z. BANACH and S. PIEKARSKI, *Non-equilibrium thermodynamics and dissipative fluid theories. I. Consistency between the extended Gibbs relation and the entropy law*, Il Nuovo Cimento, **15D**, 1063–1086, 1993.
8. H. GRAD, *On the kinetic theory of rarefied gases*, Comm. Pure Appl. Math., **2**, 331–407, 1949.
9. H. GRAD, *Note on n-dimensional Hermite polynomials*, Comm. Pure. Appl. Math., **2**, 325–330, 1949.
10. B.C. EU, *Boltzmann entropy, relative entropy, and related quantities in thermodynamic space*, J. Chem. Phys., **102**, 7169–7179, 1995.
11. A. SANTOS, J.J. BREY, C.S. KIM and J.W. DUFTY, *Velocity distribution for a gas with steady heat flux*, Phys. Rev., **39A**, 320–327, 1989.
12. Z. BANACH, *On the mathematical structure of Eu's modified moment method*, Physica, **159A**, 343–368, 1989.
13. B.C. EU, *Revision of the modified moment method and a differential form for the compensated part of entropy*, Physica, **171A**, 285–312, 1991.
14. B.C. EU, *Response to Banach's analysis of the modified moment method*, Physica, **171A**, 313–324, 1991.
15. J. MEYER and J. SCHRÖTER, *Proper and normal solutions of the Fokker–Planck equation*, Arch. Rat. Mech. Anal., **76**, 193–246, 1981.
16. J. MEYER and J. SCHRÖTER, *Comments on the Grad procedure for the Fokker–Planck equation*, J. Stat. Phys., **32**, 53–69, 1983.
17. Z. BANACH and S. PIEKARSKI, *Irreducible tensor description. I. A classical gas*, J. Math. Phys., **30**, 1804–1815, 1989.

POLISH ACADEMY OF SCIENCES
INSTITUTE OF FUNDAMENTAL TECHNOLOGICAL RESEARCH

Received September 25, 1995.

Arch. Mech., 48, 5, pp. 813–823, Warszawa 1996

Magnetohydrodynamic boundary layer flow and heat transfer on a continuous moving wavy surface

M. A. HOSSAIN (DHAKA) and I. POP (CLUJ)

THE PROBLEM of the boundary layer flow and heat transfer on a continuous moving wavy surface in a quiescent electrically conducting fluid with a constant transverse magnetic field is formulated. The resulting parabolic differential equations are solved numerically using the Keller-box scheme. Detailed results for the velocity and temperature fields are presented, and also the results for the skin-friction coefficient and the local Nusselt number. These results are given for different values of the amplitude of the wavy surface and magnetic parameter when the Prandtl number equals 0.7. It is shown that the flow and heat transfer characteristics are substantially altered by both the magnetic parameter and the amplitude of the wavy surface.

1. Introduction

THE INTERACTION between an electrically conducting fluid and an applied magnetic field is an important practical problem which has been studied very often in relation to the magnetohydrodynamic (MHD) power generator and boundary layer flow control. Hydrodynamic behaviour of boundary layers along a flat plate in the presence of a constant transverse magnetic field was first analysed by ROSSOW [1], who assumed that magnetic Reynolds number was so small that the induced magnetic field could be ignored. This problem has been further investigated by many researchers, including LEWIS [2], KATAGIRI [3], LIRON and WILHELM [4], CHUANG [5], INGHAM [6], PATHAK and CHOUDHARY [7], SOUNDALGEKAR *et al.* [8], WATANABE [9], and WATANABE and POP [10], among others.

The purpose of this paper is to study the MHD boundary layer flow and heat transfer over a continuous moving wavy surface in an electrically conducting fluid at rest, in the presence of a constant transverse magnetic field. The transformed nonsimilar boundary layer equations were solved numerically using the Keller-box method [11] for some values of the amplitude of the wavy surface a, and magnetic parameter M with the Prandtl number Pr equal 0.7. We have studied the effect of the parameters a and M on the velocity and temperature fields, as well as on the skin-friction coefficient and the local Nusselt number. We expect that the physical insight gained in this paper will enable the understanding of the complex situations where boundary layer approximation is not made.

It is worth pointing out that the MHD flow and heat transfer over a wavy surface is of importance in several heat transfer collectors where the presence of roughness elements disturbs the flow past surfaces and alters the heat transfer

rate. On the other hand, a continuously moving surface in an electrically conducting fluid permeated by a uniform transverse magnetic field has many practical applications in manufacturing metallurgical processes involving the cooling of continuous strips or filaments by drawing them through a quiescent fluid. Mention may be made of drawing, annealing and tinning of coper wires. In all these cases the properties of the final product depend to a great extent on the rate of cooling. By drawing such strips in an electrically conducting fluid subject to a magnetic field, the rate of cooling can be controlled and final products of desired characteristics might be achieved. Another interesting application of hydromagnetics to metallurgy lies in the purification of molten metals from non-metallic inclusions by the application of a magnetic field.

2. Basic equations

Consider a wavy surface at wall temperature T_w moving tangentially from left to right with a constant velocity U through a stagnant electrically conducting fluid of temperature T_∞, where $T_w > T_\infty$. The wavy surface is electrically insulated and a constant magnetic field B_0 normal to the surface is imposed. The geometry and the coordinate system, which is fixed in space, are illustrated in Fig. 1. The wavy surface is described by

$$\bar{y} = \bar{S}(\bar{x}) = \bar{a}\sin(\pi\bar{x}/l), \tag{2.1}$$

where $\bar{a}$ is the amplitude of the wavy surface and l is the characteristic length scale associated with the waves. In the present analysis the magnetic Reynolds number is assumed to be small and therefore, the induced magnetic field will be very small and can be neglected compared to the applied field. Under this approximation,

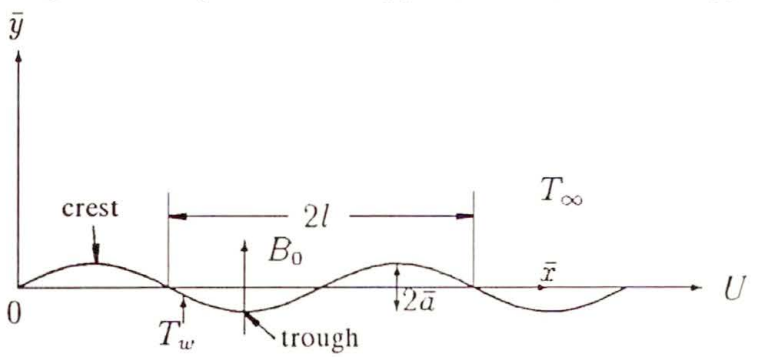

FIG. 1. Physical model and coordinate system.

the basic equations governing the steady flow of a viscous incompressible and electrically conducting fluid in presence of a uniform transverse magnetic field

are

$$
\begin{aligned}
&\frac{\partial \overline{u}}{\partial \overline{x}} + \frac{\partial \overline{v}}{\partial \overline{y}} = 0,\\
&\overline{u}\frac{\partial \overline{u}}{\partial \overline{x}} + \overline{v}\frac{\partial \overline{u}}{\partial \overline{y}} = -\frac{1}{\varrho}\frac{\partial \overline{p}}{\partial \overline{x}} + \nu\nabla^2\overline{u} - \frac{\sigma_0 B_0^2}{\varrho}\overline{u},\\
&\overline{u}\frac{\partial \overline{v}}{\partial \overline{x}} + \overline{v}\frac{\partial \overline{v}}{\partial \overline{y}} = -\frac{1}{\varrho}\frac{\partial \overline{p}}{\partial \overline{y}} + \nu\nabla^2\overline{v},\\
&\overline{u}\frac{\partial T}{\partial \overline{x}} + \overline{v}\frac{\partial T}{\partial \overline{y}} = \frac{\nu}{\mathrm{Pr}}\nabla^2 T,
\end{aligned}
\tag{2.2}
$$

where $\overline{u}$ and $\overline{v}$ are the components of velocity along the $\overline{x}$- and $\overline{y}$-directions, respectively, T is the temperature, $\overline{p}$ is the pressure, ϱ, ν and σ_0 are the density, kinematic viscosity and electric conductivity of the fluid, and ∇^2 is the Laplacian expressed in Cartesian coordinates.

The appropriate boundary conditions for the above equations are

$$
\begin{aligned}
&\overline{y} = \overline{S}(\overline{x}): \quad \overline{u}t_{\overline{y}} - \overline{v}t_{\overline{x}} = 0, \quad \overline{u}t_{\overline{x}} + \overline{v}t_{\overline{y}} = U, \quad T = T_w, \quad \text{all} \quad \overline{x} > 0,\\
&\overline{y} \to \infty: \quad \overline{u} = \overline{v} = 0, \quad \overline{p} = p_\infty, \quad T = T_\infty, \quad \text{all} \quad \overline{x} \geq 0,\\
&\overline{x} = 0: \quad \overline{p} = p_\infty, \quad T = T_\infty, \quad \text{all} \quad \overline{y} \neq 0,
\end{aligned}
\tag{2.3}
$$

where $t_{\overline{x}}$ and $t_{\overline{y}}$ are the components of the unit vector tangent to the wavy surface along $(\overline{x}, \overline{y})$-directions.

Equations (2.2) may now be nondimensionalized by using the following variables

$$
\begin{aligned}
&x = \overline{x}/l, \quad y = \overline{y}/l, \quad u = \overline{u}/U, \quad v = \overline{v}/U,\\
&p = (\overline{p} - p_\infty)/\varrho U^2, \quad \theta = (T - T_\infty)/\Delta T, \quad a = \overline{a}/l, \quad S(x) = \overline{S}(\overline{x})/l,
\end{aligned}
\tag{2.4}
$$

where $\Delta T = T_w - T_\infty$. Using these variables and introducing the nondimensional stream function ψ defined as

$$
u = \frac{\partial \psi}{\partial y}, \qquad v = -\frac{\partial \psi}{\partial x},
\tag{2.5}
$$

Eq. (2.5) can then be written as

$$
\begin{aligned}
&\frac{\partial \psi}{\partial y}\frac{\partial^2 \psi}{\partial x \partial y} - \frac{\partial \psi}{\partial x}\frac{\partial^2 \psi}{\partial y^2} = -\frac{\partial p}{\partial x} + \frac{1}{\mathrm{Re}}\nabla^2\left(\frac{\partial \psi}{\partial y}\right) - M\frac{\partial \psi}{\partial y},\\
&\frac{\partial \psi}{\partial y}\frac{\partial^2 \psi}{\partial x^2} - \frac{\partial \psi}{\partial x}\frac{\partial^2 \psi}{\partial x \partial y} = -\frac{\partial p}{\partial y} + \frac{1}{\mathrm{Re}}\nabla^2\left(\frac{\partial \psi}{\partial x y}\right),\\
&\frac{\partial \psi}{\partial y}\frac{\partial \theta}{\partial x} - \frac{\partial \psi}{\partial x}\frac{\partial \theta}{\partial y} = -\frac{1}{\mathrm{Pr}}\frac{1}{\mathrm{Re}}\nabla^2\theta.
\end{aligned}
\tag{2.6}
$$

Also, the boundary conditions (2.3) become

$$
\begin{aligned}
&y = S(x): \quad S_x\frac{\partial\psi}{\partial y} + \frac{\partial\psi}{\partial x} = 0, \quad \frac{\partial\psi}{\partial y} - S_x\frac{\partial\psi}{\partial x} = \sigma, \quad \theta = 1, \quad \text{all} \quad x > 0,\\
&y \to \infty: \quad \frac{\partial\psi}{\partial y} = \frac{\partial\psi}{\partial x} = 0, \quad p = 0, \quad \theta = 0, \quad \text{all} \quad x \geq 0,\\
&x = 0: \quad p = 0, \quad \theta = 0, \quad \text{all} \quad y \neq 0;
\end{aligned} \tag{2.7}
$$

here σ is defined according to

$$
\sigma = \left(1 + S_x^2\right)^{1/2} \tag{2.8}
$$

with $S_x = dS/dx$. Here $\mathrm{Re} = Ul/\nu$ is the Reynolds number and $M = \sigma_0 B_0^2 l/\varrho U$ is the magnetic field parameter. We notice that $(t_x, t_y) = (1/\sigma, S_x/\sigma)$ were used in (2.7). It should be noted that the value $\sigma = 1$, i.e. $a = 0$, corresponds to the case of a flat surface. In this case we take for l a characteristic length L along the flat surface.

The effect of the wavy undulations can be transferred from the boundary conditions (2.7) to the governing equations by means of the transformation given by (see REES and POP [12, 13]),

$$
\hat{x} = x, \qquad \hat{y} = y - S(x). \tag{2.9}
$$

Applying (2.9) to Eqs. (2.6) and dropping the hat we get the following equations:

$$
\begin{aligned}
&\frac{\partial\psi}{\partial y}\frac{\partial^2\psi}{\partial x\partial y} - \frac{\partial\psi}{\partial x}\frac{\partial^2\psi}{\partial y^2} = -\frac{\partial p}{\partial x} - S_x\frac{\partial p}{\partial y} + \frac{1}{\mathrm{Re}}L_1\psi - M\frac{\partial\psi}{\partial y},\\
&\frac{\partial\psi}{\partial y}\frac{\partial^2\psi}{\partial x^2} - \frac{\partial\psi}{\partial x}\frac{\partial^2\psi}{\partial x\partial y} + S_x\left(\frac{\partial\psi}{\partial x}\frac{\partial^2\psi}{\partial y^2} - \frac{\partial\psi}{\partial y}\frac{\partial^2\psi}{\partial x\partial y}\right) - S_{xx}\left(\frac{\partial\psi}{\partial y}\right)^2\\
&\qquad = \frac{\partial p}{\partial y} + \frac{1}{\mathrm{Re}}L_2\psi,\\
&\frac{\partial\psi}{\partial y}\frac{\partial\theta}{\partial x} - \frac{\partial\psi}{\partial x}\frac{\partial\theta}{\partial y} = \frac{1}{\mathrm{Pr}}\frac{1}{\mathrm{Re}}L_3\theta,
\end{aligned} \tag{2.10}
$$

and the boundary conditions (2.7) become

$$
\begin{aligned}
&y = 0: \quad \psi = 0, \quad \frac{\partial\psi}{\partial y} = 1/\sigma, \quad \theta = 1, \quad \text{ali} \quad x > 0,\\
&y \to \infty: \quad \frac{\partial\psi}{\partial y} = \frac{\partial\psi}{\partial x} = 0, \quad p = 0, \quad \theta = 0, \quad \text{all} \quad x \geq 0,\\
&x = 0: \quad p = 0, \quad \theta = 0, \quad \text{all} \quad y \neq 0,
\end{aligned} \tag{2.11}
$$

where L_1, L_2 and L_3 are the operators defined by

$$
\begin{aligned}
L_1 &= \sigma^2\frac{\partial^3}{\partial y^3} + \frac{\partial^3}{\partial y\partial x^2} - 2S_x\frac{\partial^3}{\partial x\partial y^2} - S_{xx}\frac{\partial^2}{\partial y^2}, \\
L_2 &= -S_x\sigma^2\frac{\partial^3}{\partial y^3} + (1+3S_x^2)\frac{\partial^3}{\partial x\partial y^2} - 3S_x\frac{\partial^3}{\partial y\partial x^2} \\
&\qquad - 3S_{xx}\frac{\partial^2}{\partial x\partial y} + 3S_xS_{xx}\frac{\partial^2}{\partial y^2} - S_{xxx}\frac{\partial}{\partial y} + \frac{\partial^3}{\partial x^3}, \\
L_3 &= \sigma^2\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial x^2} - 2S_x\frac{\partial^2}{\partial x\partial y} - S_{xx}\frac{\partial}{\partial y}.
\end{aligned}
\tag{2.12}
$$

Next, we introduce the boundary layer variables

$$
\tilde{x} = x, \quad \tilde{y} = \sqrt{\mathrm{Re}}\, y, \quad \tilde{\psi} = \sqrt{\mathrm{Re}}\,\psi, \quad \tilde{p} = p, \quad \tilde{\theta} = \theta. \tag{2.13}
$$

Substituting (2.13) into (2.10) and formally letting $\mathrm{Re} \to \infty$, we obtain, after dropping the tilde,

$$
\begin{aligned}
&\frac{\partial\psi}{\partial y}\frac{\partial^2\psi}{\partial x\partial y} - \frac{\partial\psi}{\partial x}\frac{\partial^2\psi}{\partial y^2} = -\frac{\partial p}{\partial x} + \mathrm{Re}^{1/2}S_x\frac{\partial p}{\partial y} + \sigma^2\frac{\partial^3\psi}{\partial y^3} - M\frac{\partial\psi}{\partial y}, \\
&S_x\left(\frac{\partial\psi}{\partial x}\frac{\partial^2\psi}{\partial y^2} - \frac{\partial\psi}{\partial y}\frac{\partial^2\psi}{\partial x\partial y}\right) - S_{xx}\left(\frac{\partial\psi}{\partial y}\right)^2 + S_x\sigma^2\frac{\partial^3\psi}{\partial y^3} = \mathrm{Re}^{1/2}\frac{\partial p}{\partial y}, \\
&\frac{\partial\psi}{\partial y}\frac{\partial\theta}{\partial x} - \frac{\partial\psi}{\partial x}\frac{\partial\theta}{\partial y} = \frac{1}{\mathrm{Pr}}\sigma^2\frac{\partial^2\theta}{\partial y^2}.
\end{aligned}
\tag{2.14}
$$

Equation $(2.14)_2$ indicates that the pressure gradient in the y-direction must be of $O(\mathrm{Re}^{-1/2})$. This implies that the lowest order pressure gradient in the x-direction can be determined from the inviscid flow solution. In the present problem, the inviscid flow field is at rest and hence $\partial p/\partial x = 0$. Now, elemination of $\partial p/\partial y$ between $(2.14)_1$ and $(2.14)_2$ results in the following boundary layer equations for the problem under consideration:

$$
\begin{aligned}
&\frac{\partial\psi}{\partial y}\frac{\partial^2\psi}{\partial x\partial y} - \frac{\partial\psi}{\partial x}\frac{\partial^2\psi}{\partial y^2} + \frac{\sigma_x}{\sigma}\left(\frac{\partial\psi}{\partial y}\right)^2 = \sigma^2\frac{\partial^3\psi}{\partial y^3} - \frac{M}{\sigma^2}\frac{\partial\psi}{\partial y}, \\
&\frac{\partial\psi}{\partial y}\frac{\partial\theta}{\partial x} - \frac{\partial\psi}{\partial x}\frac{\partial\theta}{\partial y} = \frac{1}{\mathrm{Pr}}\sigma^2\frac{\partial^2\theta}{\partial y^2},
\end{aligned}
\tag{2.15}
$$

subject to the corresponding boundary conditions

$$
\begin{aligned}
&y = 0 : \quad \psi = 0, \quad \frac{\partial\psi}{\partial y} = 1/\sigma, \quad \theta = 1, \quad \text{all } x > 0, \\
&y \to \infty : \quad \frac{\partial\psi}{\partial y} = \frac{\partial\psi}{\partial x} = 0, \quad \theta = 0, \quad \text{all } x \geq 0, \\
&x = 0 : \quad \theta = 0, \quad \text{all } y \neq 0.
\end{aligned}
\tag{2.16}
$$

To solve Eqs. (2.15) along with the boundary conditions (2.16), we introduce the following group of transformations:

$$\psi = \sigma \xi^{1/2} f(\xi, \eta), \qquad \theta = g(\xi, \eta), \tag{2.17}$$

where

$$\eta = \frac{y}{\sigma} \xi^{-1/2}, \qquad \xi = x. \tag{2.18}$$

Equations (2.15) then become

$$\begin{aligned} f''' + \frac{1}{2} f f'' + \frac{\sigma_\xi}{\sigma} \xi (f f'' - f'^2) - \frac{M}{\sigma^2} \xi f' &= \xi \left(f' \frac{\partial f'}{\partial \xi} - f'' \frac{\partial f}{\partial \xi} \right), \\ \frac{1}{\mathrm{Pr}} g'' + \frac{1}{2} f g' + \frac{\sigma_\xi}{\sigma} \xi f g' &= \xi \left(f' \frac{\partial g}{\partial \xi} - g' \frac{\partial f}{\partial \xi} \right) \end{aligned} \tag{2.19}$$

subject to the boundary conditions

$$\begin{aligned} f(\xi, 0) &= 0, \qquad f'(\xi, 0) = 1/\sigma, \qquad g(\xi, 0) = 1, \\ f'(\xi, \infty) &= h(\xi, \infty) = 0, \qquad g(\xi, \infty) = 0, \end{aligned} \tag{2.20}$$

where primes denote partial differentiation with respect to η. We notice that Eqs. (2.19) reduce to those derived by REES and POP [14] when there is no applied magnetic field ($M = 0$) in the flow field.

The physical quantities of interest are the skin-friction coefficient and the local Nusselt number defined as

$$C_f = \frac{\overline{\tau}_w}{\varrho U^2}, \qquad \mathrm{Nu}_x = \frac{\overline{x}\, \overline{q}_w}{k \Delta T}, \tag{2.21}$$

where the skin-friction $\overline{\tau}_w$ and the heat flux $\overline{q}_w$ at the wall are given by

$$\overline{\tau}_w = \mu \left(\frac{\partial \overline{u}}{\partial \overline{y}} + \frac{\partial \overline{v}}{\partial \overline{x}} \right)_{\overline{y}=0}, \qquad \overline{q}_w = -k \mathbf{n} \cdot \nabla T. \tag{2.22}$$

Here μ and k are the viscosity and thermal conductivity of the fluid, and

$$\mathbf{n} = \left(-\frac{S_x}{\sigma}, \frac{1}{\sigma} \right) \tag{2.23}$$

is the unit vector normal to the wavy surface. Using (2.4), (2.9), (2.13) and (2.17), we get the skin-friction coefficient and the local Nusselt number from the following expressions:

$$C_f \mathrm{Re}_x^{1/2} = \frac{1}{\sigma} f''(\xi, 0), \qquad \mathrm{Nu}_x / \mathrm{Re}_x^{1/2} = -g'(\xi, 0), \tag{2.24}$$

where $\mathrm{Re}_x = U\overline{x}/\nu$ is the local Reynolds number.

3. Results and discussion

An implicit finite-difference method together with the Keller-box elimination technique [11] have been used to solve the parabolic differential equations (2.19) along with the boundary condition (2.20). Since a good description of this method is available in [15–17], it will not be repeated here. The accuracy of the predicted

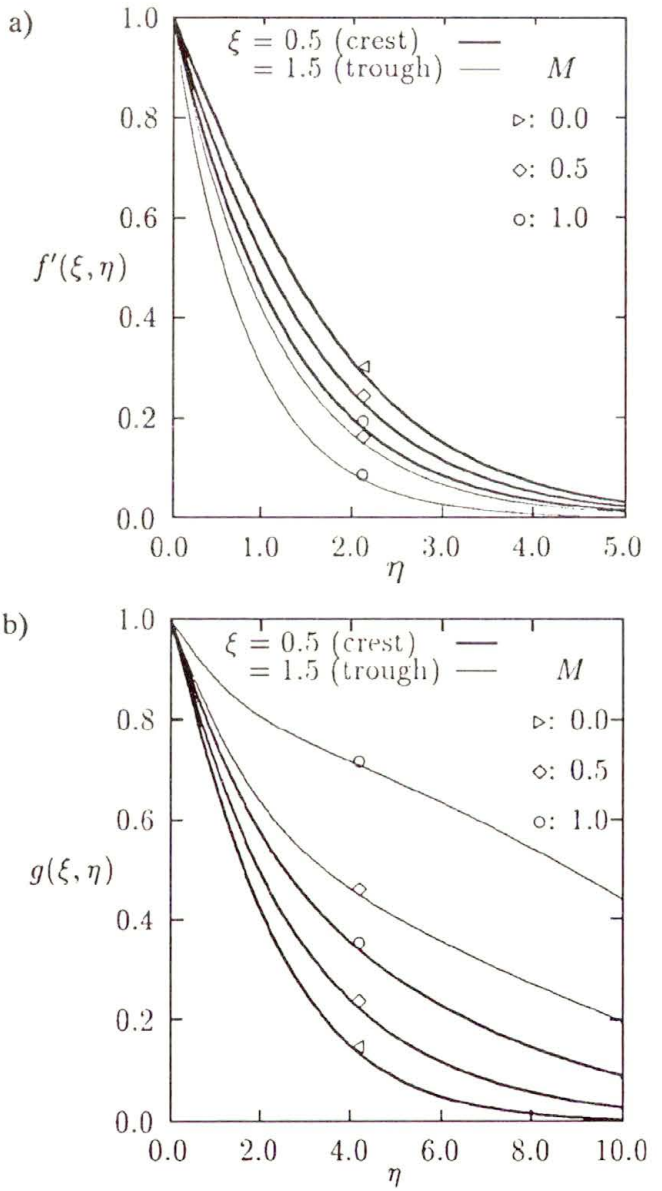

FIG. 2. a) Velocity profiles against η for different M with $a = 0.1$; b) temperature profiles against η for different M with $a = 0.1$ and Pr = 0.7.

results has been established by comparison with known results for the skin-friction coefficient and the local Nusselt number of a continuously moving flat plate ($a = 0$) in a viscous electrically non-conducting fluid with $M = 0$. Thus, REES and POP [14] found $C_f \mathrm{Re}_x^{1/2} = -0.4438$ and $\mathrm{Nu}_x/\mathrm{Re}_x^{1/2} = -0.3492$ for $\mathrm{Pr} = 0.7$, while the present calculations give $C_f \mathrm{Re}_x^{1/2} = -0.4439$ and $\mathrm{Nu}_x/\mathrm{Re}_x^{1/2} = -0.3509$. It is seen that these results are in excellent agreement and therefore we are confident that our present solution is very accurate.

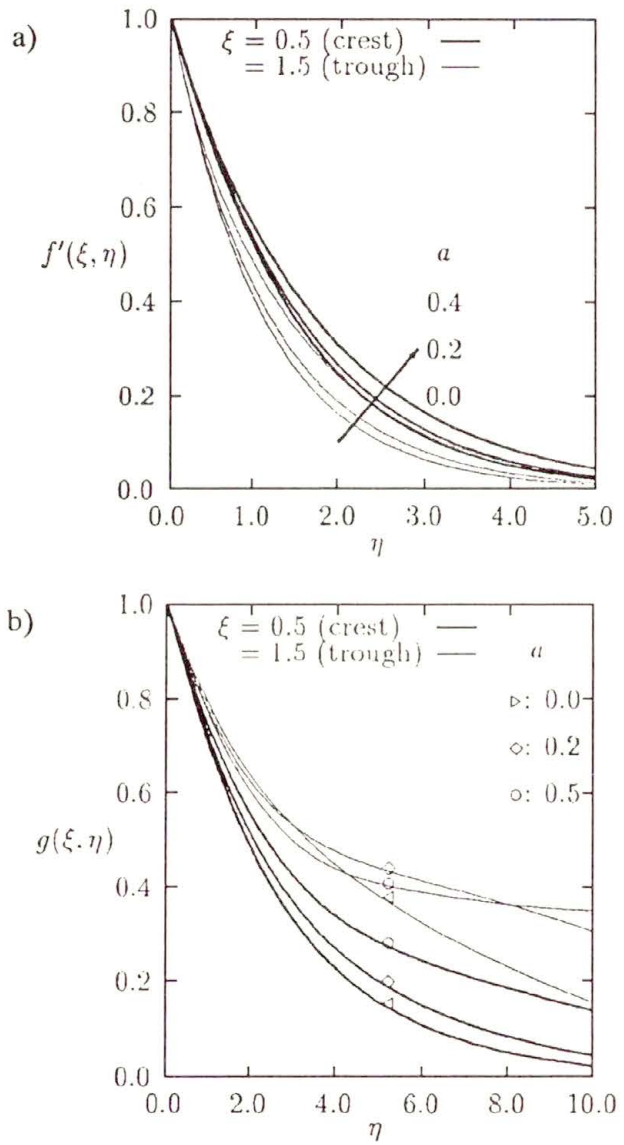

FIG. 3. a) Velocity profiles against η for different a with $M = 0.5$; b) temperature profiles against η for different a with $M = 0.5$ and $\mathrm{Pr} = 0.7$.

Representative velocity and temperature profiles are shown in Figs. 2 and 3 exhibiting the effects of the wave amplitude a and of the magnetic field parameter M. Results are given for Pr = 0.7 only. Then, since behaviour of these profiles at crest and trough positions is very similar, the case of $\xi = 0.5$ (crest) and $\xi = 1.5$ (trough) are only presented in this paper. Figures 2 and 3 show clearly that both the velocity and temperature profiles increase with the increase of M. However, Fig. 2 indicates that for $a = 0.1$ and $M = 0$ (non-magnetic field) at both the through and crest positions, the velocity and temperature profiles are almost identical due to which the differences between the thick and thin curves are not observable. But, at a larger value of a (0.5, say), there is a considerable difference at these two positions (trough and crest) in the velocity and temperature profiles for $M = 0$. On the other hand, the velocity profiles decrease, while the temperature profiles increase owing to the increase of the amplitude of the wavy surface.

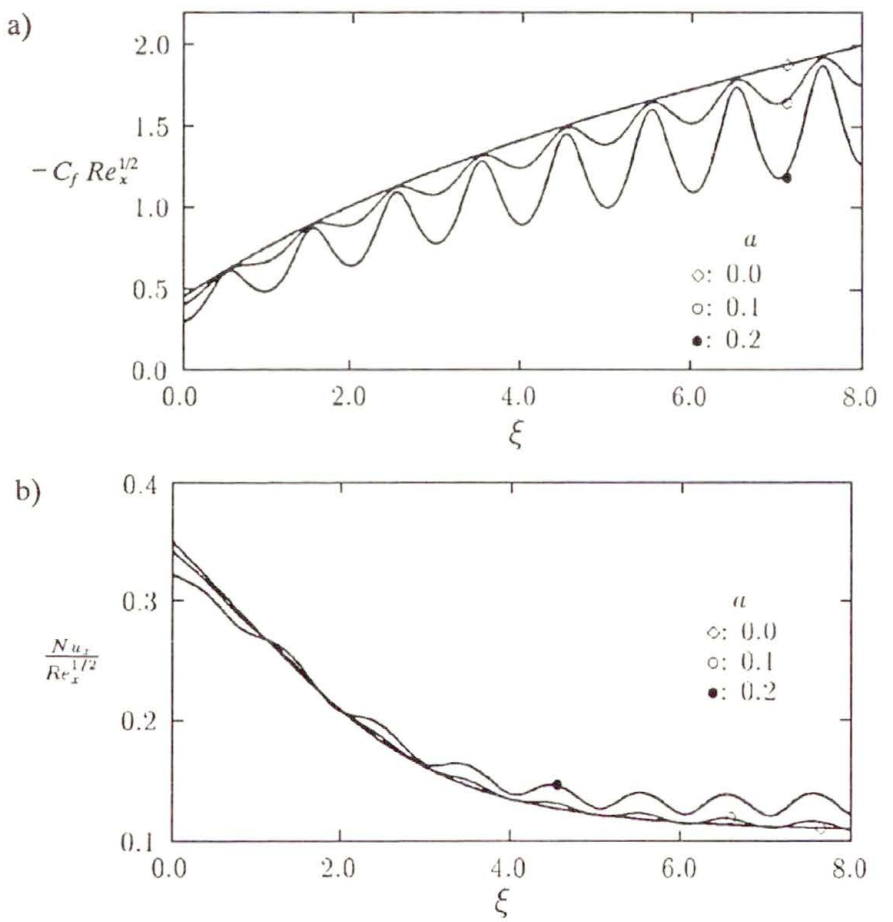

FIG. 4. a) Skin-friction coefficient for different a with $M = 0.5$; b) local Nusselt number for different a with $M = 0.5$ and Pr = 0.7.

In Figs. 4 and 5 the variation with a, M and ξ of the skin-friction coefficient and the local Nusselt number is illustrated. It is observed that these quantities vary periodically in the direction of ξ when $a \neq 0$ (wavy surface), while they vary smoothly for $a = 0$ (flat plate). Further, Fig. 4 a shows that the skin-friction coefficient is less than or equal to that corresponding to a flat surface ($a = 0$); this is due to the effect of centrifugal forces, the third term of Eq. $(2.19)_1$.

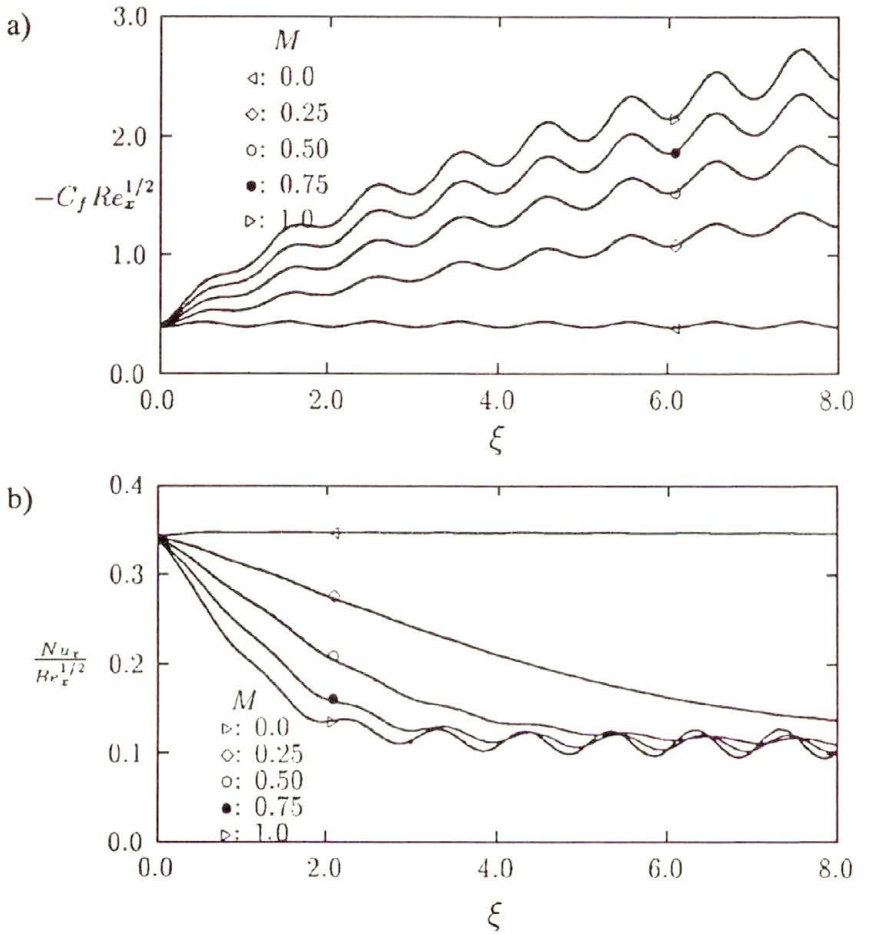

FIG. 5. a) Skin-friction coefficient for different M with $a = 0.1$; b) local Nusselt number for different M with $a = 0.1$ and $\mathrm{Pr} = 0.7$.

Acknowledgments

One of the authors (MAH) would like to thank Professor M. AHMED, Director, Network Project on Liquid and Metal Alloys, Department of Physics, University of Dhaka, for providing the necessary computational facilities. The authors are grateful to the referee for his constructive comments and suggestions.

References

1. V.J. ROSSOW, *On the flow of electrically conducting fluids over a flat plate in presence of a transverse magnetic field*, NACA Report No. 1358, 1958.
2. E. LEWIS, *The solution of MHD problem using a numerical method devised by Raetz*, Quart. J. Mech. Appl. Math., **21**, 401–411, 1968.
3. M. KATAGIRI, *The effect of Hall current on the magnetohydrodynamic boundary layer flow past a semi-infinite flat plate*, J. Phys. Soc. Japan., **27**, 1051–1059, 1969.
4. N. LIRON and H. E. WILHELM, *Integration of magnetohydrodynamic boundary-layer equations by Meksyn's method*, J. Appl. Math. Mech. (ZAMM), **54**, 27–37, 1974.
5. H. CHUANG, *Effect of transpiration on MHD boundary layer*, J. Appl. Phys., **47**, 5242–5244, 1976.
6. B.D. INGHAM, *MHD flows in the presence of a transverse magnetic field*, Nuclear Engng. Design, **52**, 325–329, 1979.
7. R.P. PATHAK and R. C. CHOUDHARY, *Magnetohydrodynamic boundary layer along a wall in a sink flow with magnetic field perpendicular to the wall*, Rev. Roumaine Phys., **25**, 139–146, 1980.
8. V. SOUNDALGEKAR, H.S. TAKHAR and M. SINGH, *Velocity and temperature field in MHD Falkner–Skan flow*, J. Phys. Soc. Japan., **50**, 3139–3143, 1981.
9. T. WATANABE, *Magnetohydrodynamic stability of boundary layers along a flat plate with pressure gradient*, Acta Mech., **65**, 41–50, 1986.
10. T. WATANABE and I. POP, *Hall effects on magnetohydrodynamic boundary layer flow over a continuous moving flat plate*, Acta Mech., **108**, 35–47, 1995.
11. H.B. KELLER, *Numerical methods in boundary layer theory*, Annual Rev. Fluid Mech., **10**, 417–433, 1978.
12. D.A.S. REES and I. POP, *A note on free convection along a vertical wavy surface in a porous medium*, J. Heat Transfer, **116**, 505–508, 1994.
13. D.A.S. REES and I. POP, *Free convection induced by a vertical wavy surface with uniform heat flux in a porous medium*, J. Heat Transfer, **117**, 547–550, 1995.
14. D.A.S. REES and I. POP, *Boundary layer flow and heat transfer on a continuous moving wavy surface*, Acta Mech., **112**, 149–158, 1995.
15. T. CEBECI and P. BRADSHAW, *Physical and computational aspects of convective heat transfer*, Springer, New York 1984.
16. M.A. HOSSAIN, *Viscous and Joule heating on MHD free convection flow with variable plate temperature*, Int. J. Heat Mass Transfer, **35**, 3485–3487, 1992.
17. M.A. HOSSAIN, S.K. DAS and I. POP, *Heat transfer response of MHD free convection flow along a vertical plate to surface temperature oscillations*, Int. J. Non-Linear Mech. [in press].

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF DHAKA, DHAKA, BANGLADESH
and
FACULTY OF MATHEMATICS
UNIVERSITY OF CLUJ, CLUJ, ROMANIA.

Received November 7, 1995.

Arch. Mech., 48, 5, pp. 825–833, Warszawa 1996

On the macroscopic parameters of brittle fracture

N.F. MOROZOV and Y.V. PETROV (ST. PETERSBURG)

THE PROBLEM of modeling of some of the specific effects of brittle fracture in the high loading rate conditions is discussed. An approach based on the system of fixed material constants describing macro-strength properties of the material is considered. New principles of material testing are analyzed. The corresponding incubation time criterion allows us to manage without the *a priori* given rate dependences of dynamic strength and fracture toughness. New applications of the criterion to the problems of disintegration and erosion are considered.

1. Basic structural characteristic of static fracture

ONE OF THE PRINCIPAL parameters of linear fracture mechanics is the material structure size d describing the elementary cell of failure. The classical approaches by GRIFFITH [3] and IRWIN [5] consider this characteristic as a latent quantity. It may be presented as dimensional combination of surface energy, critical stress intensity factor, static strength and elastic constants of the material:

$$d \sim \frac{\gamma \cdot E}{\sigma_c^2}, \qquad d \sim \frac{K_{Ic}^2}{\sigma_c^2}. \tag{1.1}$$

The elementary cell of fracture has no unique physical interpretation. It may be interpreted in various ways, depending on the class of problems. The corresponding Griffith–Irwin criterion is a universally recognized critical condition of brittle and quasi-brittle fracture. This criterion is based on the square root singularity of the stress field at the crack tip. Therefore its field of applicability is limited. For instance, in the case of an angular notch in a plate, the general energy balance equation cannot be satisfied for all methods of loading (MOROZOV [9]).

NEUBER [12] and NOVOZHILOV [14] suggested to consider the material structure directly. The corresponding criterion requires that the mean normal stress in the range of material structure size d must be equal to the static strength of the material. In the plane deformation state case we have:

$$\frac{1}{d}\int_0^d \sigma(r)\,dr \leq \sigma_c. \tag{1.2}$$

Assuming that in the simplest cases the criterion (1.2) gives the same results as the Irwin's critical stress intensity factor criterion, we obtain for the material structure size d the expression:

$$d = \frac{2}{\pi}\frac{K_{Ic}^2}{\sigma_c^2}. \tag{1.3}$$

Criterion (1.2) may be used in various cases in which the square root singularity and the appropriate energy balance do not work. Results obtained by means of the criterion (1.2) under the condition (1.3) are well confirmed by experiments in static cases (MOROZOV [9]).

The introduced material constant d is quite similar to the process zone size parameter for the short crack fracture assessment occurring in ceramics (ANDO *et al.* [1]).

2. Fast fracture processes

The constant progress observed in experimental mechanics during the last years enables us to understand the fact that dynamic fracture in brittle solids is remarkable for its specific nature. The corresponding experiments helped us to discover some principal effects that have no interpretation within the framework of the conventional models of brittle failure. Here are just some of them:

1) The dynamic branch of the time-dependence of strength and fracture delays in spalling (ZLATIN *et al.* [23]).

2) The extensive zones of failure (cavitation) in spalling and their unpredictable geometry (BROBERG [2], SEAMAN *et al.* [21], etc.).

3) The dependence of critical stress intensity factor of crack growth initiation on the loading history (RAVI-CHANDAR and KNAUSS [20]).

4) The behaviour of the short pulse load threshold amplitudes leading to failure at the crack tip (KALTHOFF and SHOCKEY [6], SHOCKEY *et al.* [22]).

Analysis of the experiments shows that the main contradictions of the traditional models appear when failure occurs during the short time intervals after the start of the loading process. MOROZOV and PETROV [10, 11] proposed an approach to the analysis of dynamic brittle failure based on the incubation time criterion:

$$\frac{1}{\tau}\int_{t-\tau}^{t}\frac{1}{d}\int_{0}^{d}\sigma(r,\theta,t')\,dr\,dt' \leq \sigma_c\,. \tag{2.1}$$

Here d and τ are material structure size and structure time of failure, respectively, σ_c is static strength of the material, (r,θ) are polar coordinates, $\sigma(r,\theta,t)$ is tensile stress at the crack tip ($r = 0$). The material structure size d is to be determined in accordance with the data of quasi-static tests of specimens containing cracks, and in the case of plane strain it may be expressed by the simple formula (1.3). The material structure time τ is responsible for the dynamic peculiarities of the macro-fracture process and for each material it should be found from experiments. In accordance with this approach, σ_c, K_{I_c} and τ constitute the system of fixed material constants describing macro-strength properties of the material. PETROV [15] has shown that the criterion (2.1) reflects the discrete nature of dynamic fracture of brittle solids.

In the case of virgin materials, the criterion (2.1) reduces to the form:

$$\frac{1}{\tau}\int_{t-\tau}^{t}\sigma(t')\,dt' \leq \sigma_c\,. \tag{2.2}$$

This form will now be used for the analysis of two particular problems.

The analysis of the particular problems of dynamic fracture mechanics is associated with the appropriate choice of the parameter τ. We shall mention two basic cases:

1. The incubation time is defined by the material structure size of fracture:

$$\tau = \frac{d}{c} = \frac{d\sqrt{\varrho}}{k}\,, \tag{2.3}$$

where c is the maximum wave velocity, ϱ is the density of continuum, k is the constant depending on the deformation material properties. According to this definition, the incubation time has a physical meaning of the minimum time period required for the interaction between two neighbouring material structure cells. The incubation-time criterion with the parameter τ selected according to the formula (2.3) allows us to describe effectively the time-dependence of strength and the fracture zone geometry in conditions of spalling (MOROZOV *et al.* [10, 11], PETROV and UTKIN [16]). Thus, the definition (2.3) provides a good analogy between the incubation time criterion and the well-known experiments in the case of "defectless" materials.

2. The incubation time does not directly depend on the material structure size of failure. This takes place when a problem of initiation of the macro-crack growth is considered. Nucleation, growth and coalescence of micro-defects in the special process zone region at the crack tip precedes the growth of the macro-crack. These processes are accompanied by a local stress relaxation and change the effective material properties. The incubation time τ is to be considered as the principal integral characteristic of the processes in the corresponding process zone region. PETROV and MOROZOV [18] proved that in the case of macro-cracks, the material structure time τ can be interpreted as an incubation time in the well-known minimum time criterion proposed and explored by KALTHOFF and SHOCKEY [6], HOMMA *et al.* [4], and SHOCKEY *et al.* [22]:

$$\tau = t_{\text{inc}}\,. \tag{2.4}$$

The aforementioned dependence of the fracture toughness on the loading history and the specific behaviour of the short loading pulse threshold amplitudes can be explained and effectively analyzed by means of the incubation time criterion under the condition (2.4) (PETROV and MOROZOV [18]).

3. Some basic principles of the material strength properties testing

In this section we outline some of the possible methods of description of the material strength properties. Table 1 represents the basic parameters and criteria to be used in testing of the materials. In Table 1 σ_c, K_{Ic} are the material constants, $\sigma_c^{\text{dyn}}(v)$, $K_{Id}(v)$ are the material functions that represent the dependences of critical characteristics on the loading rate v.

Table 1.

No.	Method	Material parameters	Criteria
1	Classical static	σ_c, K_{Ic}	$\sigma \leq \sigma_c$, $K_I \leq K_{Ic}$
2	Classical dynamic	$\sigma_c^{\text{dyn}}(v)$, $K_{Id}(v)$	$\sigma(t) \leq \sigma_c^{\text{dyn}}$, $K_I(t) \leq K_{Id}$
3	SRI International	$\sigma_c^{\text{dyn}}(v)$, $K_{Id}(v)$, t_{inc}	$\sigma(t) \leq \sigma_c^{\text{dyn}}$, minimum time criterion
4	Incubation time approach	σ_c, K_{Ic}, τ	incubation time criterion

The classical dynamic approach, resulting directly from the static strength theory and linear fracture mechanics, is based on two strength characteristics $\sigma_c^{\text{dyn}}(v)$, $K_{Id}(v)$, that are supposed to be material functions found from experiments.

The minimum time theory proposed by J.F. Kalthoff, D.A. Shockey and coworkers is based on the incubation time notion. It allows us to explain some of the principal dynamic fracture effects. On the other hand, the minimum time technique turns out to be too sophisticated for practical engineering.

It is seen from the Table 1 and the aforementioned results that the incubation time criterion combines the simplicity of the classical static method with the effectiveness of the SRI International approach. Basing on the system of fixed material constants, it enables us to predict the behaviour of dynamic strength and dynamic fracture toughness from a unified viewpoint. Thus, the rate strength dependences may be considered as calculated characteristics. The criterion may be applied for both the "defectless" and macro-cracked specimens.

4. Application to the problem of disintegration

The great interest in connection with the technology of disintegration of solids is presented by the issues concerning the speed of propagation of failure and the fracture zone geometry. Even elementary experiments on spalling show that the zone of destruction may have very diverse geometry. It can be either one spalling section (like a crack), or several ones. In some experiments, the zone of fracture has a form of a continuously damaged domain of finite extension. This domain was named a zone of continuous crushing (NIKIFOROFSKY and SHEMJAKIN [13]). There are also experiments, in which the zone of destruction represents the mixture

of damaged regions with intact parts of the material. Most of the experiments show, that the geometry of the failure zones strongly depends on the parameters of the applied loading, such as a speed of loading, its amplitude, duration, etc. Eventually, it can be said that the whole history of loading is very important.

In this section we consider the particular problem of disintegration of a solid ball caused by the instant discharge (unloading) of the external pressure (PETROV, SEMENOV and UTKIN [17]). We shall demonstrate that the incubation time criterion gives all the variety of the fracture zones just in the sonic approximation of the problem.

Let a ball be loaded on its surface by a uniform pressure. At a certain moment, the pressure is instantaneously taken away. The corresponding boundary value problem in the sonic approximation is described by the following system of equations:

$$
\begin{aligned}
& u = u(r,t), \qquad u = \frac{\partial \Phi}{\partial r}, \qquad \sigma = -\frac{\partial \Phi}{\partial t}, \\
& \frac{\partial^2 \Phi}{\partial t^2} = c^2 \frac{1}{r^2} \frac{\partial}{\partial r}\left(r^2 \frac{\partial \Phi}{\partial r}\right), \\
& -\frac{\partial \Phi}{\partial t}\bigg|_{r=R} = [1 - H(t)] \cdot \sigma_0 .
\end{aligned} \tag{4.1}
$$

Here $H(t)$ is the Heaviside step function; u and σ are the displacement and pressure inside the material, Φ is the potential function, σ_0 is the initial pressure, c is the wave velocity.

The discharge of the pressure produces a spherical wave of unloading moving to the center of the ball. This wave carries a tensile stress that produces fracture of the material. The aforementioned incubation time criterion allows us to find the extent and geometry of failure in this particular situation.

We assume that the material is homogeneous, isotropic and composed of the spherical layers of thickness b. All layers are assumed to have identical material properties. We assume that the layer is destroyed at the certain moment, when the critical condition in the center of the layer corresponding to the criterion (2.2) is fulfilled. The destroyed layer turns out to be, on the one hand, a shield for the moving waves, and on the other hand, a source of an additional unloading wave. The unloading waves running inside the ball are reflected from the destroyed layers and interact with other stress waves. Thus the process is characterized by a complicated picture of direct and reflected waves.

The scheme was realized in a complex of computer programs BALL. One of the main results of the calculations was the graphically submitted zone of fracture. It turned out that the region of failure strongly depends on the parameters of the problem and on the material constants. Some of the possible variants of the fracture zone for the hypothetical values $\sigma_c = 900\,\text{MPa}$, $\tau = 1\,\mu\text{s}$, $b = 3\,\text{mm}$, $c = 5000\,\text{m/s}$ are presented in the Fig. 1 (fracture zones are shown in black), where a large variety of geometry is seen.

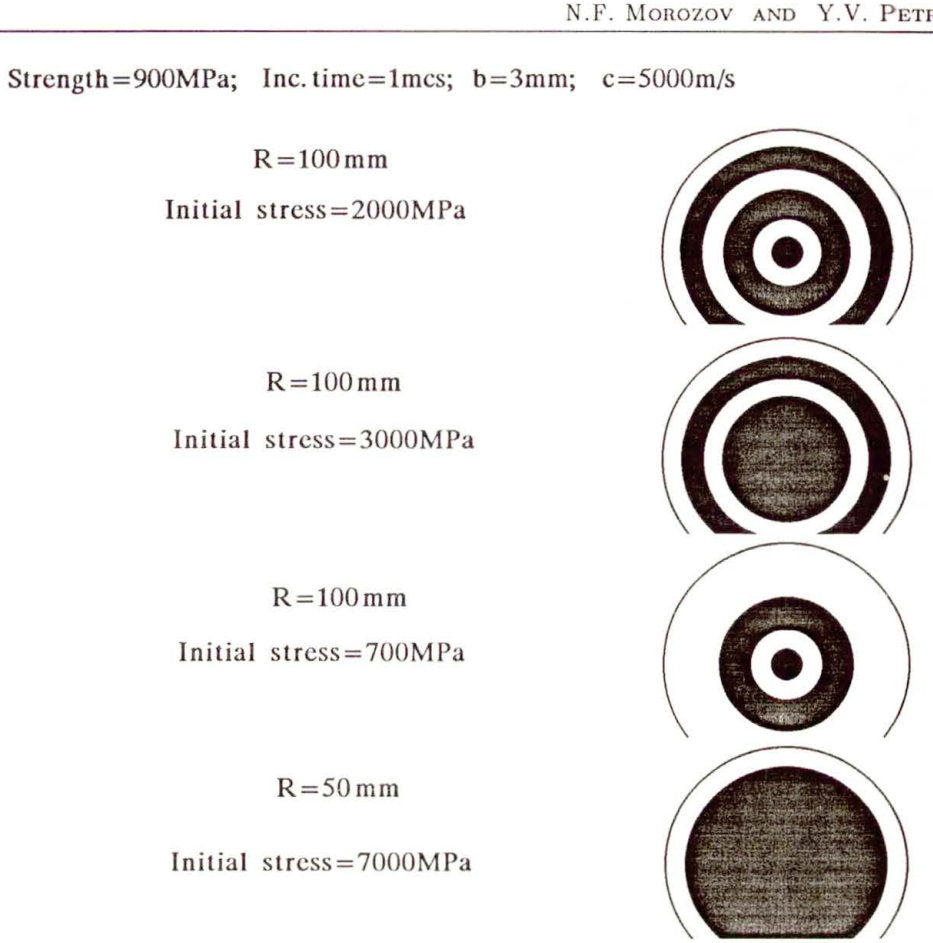

FIG. 1. Calculated fracture zones caused by instant unloading of the ball.

5. Application to the problem of erosion

The solid particle impact velocity at the beginning of target material loss in the steady-state erosion process can be considered as a critical or threshold velocity. It is a principal characteristic that bears an information about dynamic strength properties of materials subjected to the impact loading. In this section, the relation between the threshold velocity W and the incubation time τ is investigated. The possibility of using the incubation time criterion in determining the threshold erosion characteristics is established.

One of the principal features of the erosion process is that the target material surface is subjected to extremely short impact actions. The evaluation of failure in these conditions may be done only on the basis of special criteria reflecting the specific nature of fast fracture phenomenon. The incubation time criterion (2.2) is an effective instrument for this analysis. Here we shall consider the simplest way to obtain some of the basic threshold erosion characteristics.

Let a spherical particle of radius R fall with velocity v on the surface of an elastic half-space. Using the classical Hertz impact theory approximation (KOLESNIKOV and MOROZOV [7]), we describe the motion of the particle by the following equation:

$$m\frac{d^2h}{dt^2} = -P, \tag{5.1}$$

where

$$P(t) = k(R)h^{3/2}(t), \qquad k(R) = \frac{4}{3}\sqrt{R}\,\frac{E}{(1-\nu^2)}. \tag{5.2}$$

At the beginning of the impact event we have $dh/dt = v$. The maximum penetration h_0 occurs when $dh/dt = 0$. Solving Eq. (5.1), we obtain

$$h_0(v,R) = \left(\frac{5mv^2}{4k}\right)^{2/5}, \qquad t_0(v,R) = \frac{2h_0}{v}\int_0^1 \frac{d\gamma}{\sqrt{1-\gamma^{5/2}}} = 2.94\frac{h_0}{v}, \tag{5.3}$$

where t_0 is the duration of the impact event. The penetration function $h(t)$ can be approximated by the simple formula (KOLESNIKOV and MOROZOV [7]):

$$h(t) = h_0 \sin(\pi t/t_0). \tag{5.4}$$

The maximum tensile stress occurring at the edge of the contact area is given by the expression (LAWN and WILSHAW [8]):

$$\begin{aligned} \sigma(t,v,R) &= \frac{1-2\nu}{2}\,\frac{P(t,v,R)}{\pi a^2(t,v,R)}, \\ a(t,v,R) &= \left[3P(t,v,R)(1-\nu^2)\frac{R}{4E}\right]^{1/3}, \end{aligned} \tag{5.5}$$

where the contact force $P(t,v,R)$ can be found by means of Eqs. (5.2)–(5.4).

Let $v = W$ denote the threshold velocity corresponding to the beginning of failure. We consider the function:

$$f(\tau,v,R) = \max_t \int_{t-\tau}^{t} \sigma(s,v,R)\,ds - \sigma_c\tau. \tag{5.6}$$

According to (2.2), we determine the threshold velocity $v = W$ as the minimum positive root of the equation:

$$f(\tau,v,R) = 0, \tag{5.7}$$

where τ is the incubation time for the target material.

The corresponding calculations were performed for the aluminum alloy B95 and the incubation time was determined according to formulae (1.3), (2.3): σ_c = 460 MPa, K_{Ic} = 37 MPa m$^{1/2}$, c = 6500 m/s, $\tau = 2K_{Ic}^2/(\pi\sigma_c^2 c) \approx 0.6\,\mu$s. The calculated dependence of the threshold velocity W on the value of radius R is presented in the Fig. 2 by the solid curve. The static branch shows a weak dependence of the threshold velocity on the length of the radius. On the contrary, the dynamic branch, corresponding to the small particles and very short loading pulses, represents a strong dependence of the critical velocity on the radius of particles. This behaviour of the threshold velocity is observed in numerous experiments (POLEZHAEV [19]), but it can not be explained on the basis of the traditional fracture mechanics. The dependence following from the conventional critical stress theory is also presented in the Fig. 2 by dashed line.

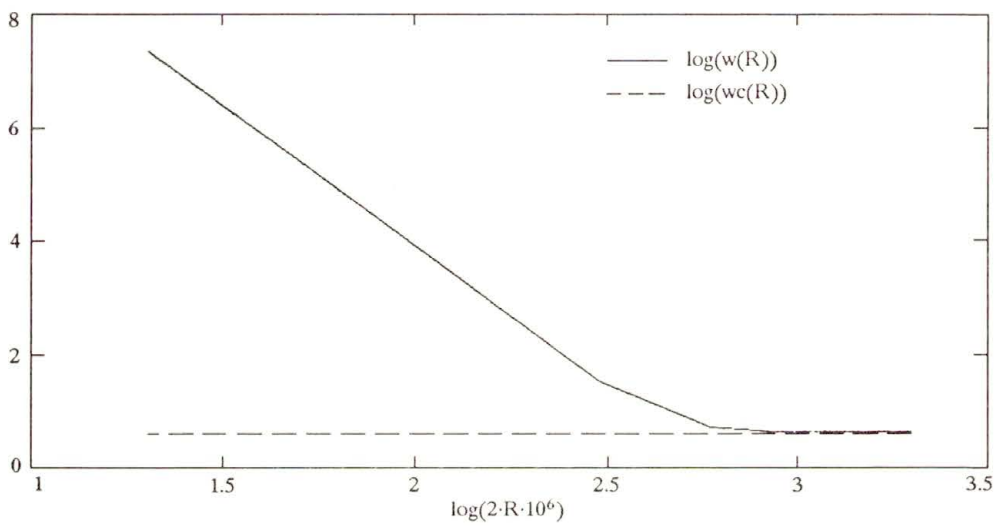

FIG. 2. Dependence of the threshold velocity W (m/s) on the radius R (m) of erodent particles calculated for aluminium alloy B95. The dependence corresponding to the classical fracture criterion: $\sigma \leq \sigma_c$ is plotted by dashed line.

Acknowledgments

The research described in this publication was made possible owing partly to the Grants 96-01-00391 and 95-01-00644 from the Russian Fundamental Research Foundation (RFFI).

References

1. K. ANDO, B.A. KIM, M. IWASA and N. OGURA, *Process zone size failure criterion and probabilistic fracture assessment curves for ceramics*, Fatigue Fract. Engng. Mater. Struct., **15**, 2, pp. 139–149, 1992.
2. K.B. BROBERG, *Some aspects of the mechanism of scabbing*, [in:] Stress Wave Propagate Material, N-Y.-L., Interscience, pp. 229–246, 1960.

3. A. GRIFFITH, *The phenomena of rupture and flow in solids*, Philos. Trans. Roy. Soc., London, Ser. A **221**, 1921.
4. H. HOMMA, D.A. SHOCKEY, and Y. MURAYAMA, *Response of cracks in structural materials to short pulse loads*, J. Mech. Phys. Solids, **31**, pp. 261–279, 1983.
5. G. IRWIN, *Analysis of stresses and strains near the end of a crack traversing a plate*, J. Appl. Mech., **3**, 1957.
6. J.F. KALTHOFF and D.A. SHOCKEY, *Instability of cracks under impulse loads*, J. Appl. Physics, **48**, pp. 986–993, 1977.
7. Y.V. KOLESNIKOV and E.M. MOROZOV, *Contact fracture mechanics* [in Russian], Nauka, Moscow 1989.
8. B.R. LAWN and T.R. WILSHAW, *Indentation fracture: principles and application*, J. Mater. Sci., **10**, 6, pp. 1049–1081, 1975.
9. N.F. MOROZOV, *Mathematical questions of crack theory* [in Russian], Nauka, Moscow 1990.
10. N.F. MOROZOV and Y.V. PETROV, *Dynamic fracture toughness in crack growth initiation problems* [in Russian], Izv. AN SSSR. MTT (Solid Mechanics), 6, pp. 108–111, 1990.
11. N.F. MOROZOV, Y.V. PETROV and A.A. UTKIN, *On the analysis of spalling in the framework of structural fracture mechanics* [in Russian], Dokl. AN SSSR, **313**, 2, pp. 276–279, 1990.
12. H. NEUBER, *Kerbspannungslehre (Notch Stresses)*, Julius Verlag, Berlin 1937.
13. V.S. NIKIFOROFSKY and E.I. SHEMJAKIN, *Dynamic fracture of solids* [in Russian], Nauka, Novosibirsk 1979.
14. V.V. NOVOZHILOV, *Necessary and sufficient criterion of brittle strength* [in Russian], Prikl. Matematika i Mechanika, **33**, 2, pp. 212–222, 1969.
15. Y.V. PETROV, *On "quantum" nature of dynamic fracture in brittle media* [in Russian], Dokl. AN SSSR, **321**, 1, pp. 66–68, 1991.
16. Y.V. PETROV and A.A. UTKIN, *Structure-time peculiarities of dynamic fracture of materials*, [in:] Macro- and Micro-Mechanical Aspects of Fracture, Abstracts, Proc. of the EUROMECH-291, St. Petersburg, Russia, June 22-27, 1992.
17. Y.V. PETROV, B.N. SEMENOV and A.A. UTKIN, *Dynamic fragmentation of solids caused by the instant discharge of intensive pressure* [in Russian], Proc. of the 1st Russian Symp. on Solid Mechanics, St. Petersburg, June 6-9, 1994.
18. Y.V. PETROV and N.F. MOROZOV, *On the modeling of fracture of brittle solids*, ASME J. Applied Mech., **61**, pp. 710–712, 1994.
19. Y.V. POLEZHAEV, *Thermo-gas-dynamic testing of aircrafts* [in Russian], Izd. MAI, Moscow 1986.
20. K. RAVI-CHANDAR and W.G. KNAUSS, *An experimental investigation into dynamic fracture. 1. Crack initiation and arrest*, Int. J. Fract., **25**, pp. 247–262, 1984.
21. L. SEAMAN, D.R. CURRAN and W.J. MURRI, *A continuum model for dynamic tensile microfracture and fragmentation*, J. Appl. Mech., **52**, pp. 593–600, 1985.
22. D.A. SHOCKEY, D.C. ERLICH, J.F. KALTHOFF and H. HOMMA, *Short pulse fracture mechanics*, Engng. Fract. Mech., **23**, pp. 311–319, 1986.
23. N.A. ZLATIN, G.S. PUGAVHEV, S.M. MOCHALOV and A.M. BRAGOV, *Time-dependence of strength of metals* [in Russian], FTT, **17**, 9, pp. 2599–2602, 1975.

ACADEMY OF SCIENCES OF RUSSIA
STRENGTH AND FRACTURE MECHANICS GROUP
ST. PETERSBURG, RUSSIA.

Received December 22, 1995.

Arch. Mech., 48, 5, pp. 835–841, Warszawa 1996

Stability of Couette flow in the wide gap of two circular concentric cylinders with rotating inner cylinder and finite growth rate

H.S. TAKHAR (MANCHESTER), M.A. ALI (BAHRAIN)
and V.M. SOUNDALGEKAR (THANE)

A FINITE-DIFFERENCE SOLUTION for the stability of flow of a viscous fluid in an annular wide-gap space is carried out by taking into account the effects of finite growth rate of the amplification factor σ. The numerical values of the minimum Taylor number Ta_m and the critical Taylor number Ta_c for different values of η and for $\sigma \gtrless 0$ and $\sigma = 0$, respectively, are derived and tabulated. The effects of η and σ on the radial component of disturbance and on the cell patterns are shown. It is observed that for increasing $\sigma(> 0)$, the cell patterns are reduced in size, while for decreasing $\sigma(< 0)$ they are enlarged.

1. Introduction

STABILITY OF AXISYMMETRIC FLOW of a viscous incompressible fluid between two concentric rotating cylinders has been investigated by TAYLOR [6], CHANDRASEKHAR [1], HARRIS and REID [2], WALOWIT, TSAO and DI PRIMA [7], SOUNDALGEKAR *et al.* [4], who used different methods and different boundary conditions. The usual mathematical procedure of the stability analysis is to assume that small disturbances are superimposed on the steady motion. These disturbances are assumed to be periodic in the z-direction and proportional to $e^{\sigma t}$, where σ is the amplification parameter or growth rate factor. Then the parameter which governs the stability of the motion is the exponential time factor σ, the motion being stable or unstable according to whether the real part of σ is less than or greater than zero, and when $\sigma = 0$, it is known as the marginal state. Almost all the papers in this field applying the linear stability analysis dealt with $\sigma = 0$, i.e. the marginal state of stability. ROBERTS [3] was the first to study the effects of non-zero values of the growth rate on the Taylor number in the wide-gap Couette flow, and computed the smallest characteristic values of the Taylor number for a given wave number a and the growth rate factor σ, by employing the numerical method given by Harris and Reid. However, the effects of σ on the radial velocity perturbation and on the cell pattern have not been studied or shown graphically. Hence it is now proposed to solve the eigen-value problem of ROBERTS [3] by using a finite-difference technique and to derive the minimum values of the Taylor number, Ta_m, for different values of $\pm\sigma$ of the radial velocity perturbations and the cell-patterns. In the next section, a short account of the finite-difference method is given and numerical values of Ta_m are tabulated for different values of

$\pm\sigma$ and wave number a. In order to verify our results, we have also computed the numerical values of Ta_c, the critical Taylor number for $\sigma = 0$ and critical wave number a_c.

2. Solution to the axisymmetric problem at finite growth rate

The axisymmetric linear stability of Couette flow at finite growth rate, when the outer cylinder is at rest and the two cylinders are separated by a wide gap, can be shown to be governed by the following system of sixth order (e.g. SOUNDALGEKAR *et al.* [5]).

$$(DD^* - a^2 - \sigma)(DD^* - a^2)u = -a^2\mathrm{Ta}\cdot g(x)v, \tag{2.1}$$

$$(DD^* - a^2 - \sigma)v = u, \tag{2.2}$$

$$u = v = Du = 0 \qquad \text{at} \quad x = 0, 1. \tag{2.3}$$

Here u, v are the radial and azimuthal components of the disturbances, a is the axial wave number, σ is the growth rate and Ta is the Taylor number. They are defined as follows:

$$d = R_2 - R_1, \qquad x = \frac{r - R_1}{d}, \qquad g(x) = 1 - x,$$

$$D = \frac{d}{dx}, \qquad D^* = \frac{d}{dx} + \frac{1-\eta}{\xi}, \qquad a = \lambda d, \tag{2.4}$$

$$\eta = R_1/R_2, \qquad \xi = \eta + (1-\eta)x, \qquad \mathrm{Ta} = \frac{4\eta^2 d^4}{1-\eta^2}\left(\frac{\Omega_1}{\nu}\right)^2,$$

where R_1 and R_2 are the radii of the inner and outer cylinders, respectively, and Ω_1 is the rate of rotation of the inner cylinder. Our Ta is equivalent to $2\mathrm{Ta}_R$ where Ta_R is the Taylor number defined by Roberts or Chandrasekhar, as

$$\mathrm{Ta}_R = \frac{2\eta^2 d^4}{1-\eta^2}\left(\frac{\Omega_1}{\nu}\right)^2.$$

3. Method of solution

By solving the above eigenvalue problem defined by Eqs. (2.1) – (2.3), we determine the smallest characteristic value of the Taylor number, denoted by Ta_m, for given wave number a and σ. To solve Eqs. (2.1) – (2.2) by the finite-difference technique, we first expand these as follows:

$$\left[D^4 + 2kD^3 - \left(3k^2 + 2a^2 + \sigma\right)D^2 + \left(3k^3 - 2ak - \sigma k\right)D\right.$$
$$\left. + (2a^2k^2 - 3k^4 + a^4 + \sigma(k^2 + a^2)\right] u = -a^2 \cdot \mathrm{Ta}\cdot g(x)v, \tag{3.1}$$

$$\left[D^2 + kD - (k^2 + a^2 + \sigma)\right] v = u, \qquad k = \frac{1-\eta}{\xi}. \tag{3.2}$$

We write the derivatives in terms of central differences and rearrange the terms. Then Eqs. (3.1) and (3.2) reduce to

$$m_i U_{i+2} + m_2 U_{i+1} + m_3 U_i + m_4 U_{i-1} + m_5 U_{i-2} = -h^4 a^2 \mathrm{Ta} \cdot g(x) \cdot V_i , \tag{3.3}$$

$$C_1 V_{i+1} + C_2 V_i + C_3 V_{i-1} = h^2 U_i , \tag{3.4}$$

where

$$\begin{aligned} m_1 &= 1 + hk, \\ m_2 &= -4 - 2hk - h^2(3k^2 + 2a^2 + \sigma) + \frac{1}{2}h^3(3k^3 - 2a^2 k - \sigma k), \\ m_3 &= 6 + 2h^2(3k^2 + 2a^2 + \sigma) + h^4(2a^2 k^2 - 3k^4 + a^4 + \sigma k^2 + \sigma a^2), \\ m_4 &= -4 + 2hk - h^2(3k^2 + 2a^2 + \sigma) - \frac{1}{2}h^3(3k^3 - 2a^2 k - \sigma k), \\ m_5 &= 1 - hk, \\ C_1 &= 1 + \frac{1}{2}hk, \\ C_2 &= -2 - h^2(k^2 + a^2 + \sigma), \\ C_3 &= 1 - \frac{1}{2}hk. \end{aligned} \tag{3.5}$$

The suffix i stands for the pivotal point under consideration. The step length $h = 1/N$, where N is the number of intervals into which the range $[0, 1]$ is divided. The boundary conditions (2.3) imply that

$$\begin{aligned} U_0 &= V_0 = U_N = V_N = 0, \\ U_{-1} &= U_1 , \qquad U_{N+1} = U_{N-1}. \end{aligned} \tag{3.6}$$

Equations (3.3) and (3.4) with conditions (3.6) can be written in matrix notation as

$$\begin{aligned} A_1 \overline{U} &= \mathrm{Ta}\, B_1 \overline{V}, \\ A_2 \overline{V} &= \overline{u}, \end{aligned} \tag{3.7}$$

where A_1, A_2 and B_1 are the coefficient matrices of order $n \times n$, and $n = N - 1$.

Equations (3.7) can be combined into an eigenvalue equation of the form

$$(C - \mathrm{Ta}\, I)\overline{V} = 0. \tag{3.8}$$

The eigenvalues are computed by using the QR algorithm for $\sigma = 0, \pm 0.5, \pm 1, \pm 1.5, \pm 2.0$ and are listed in Table 1 for $\eta = 0.85, 0.5, 0.1$ which corresponds to

the wide gap case. For $\sigma = 0$, the marginal state, the eigenvalues Ta_c are known as critical values of the Taylor number which corresponds to lowest values of the wave number a and are denoted by a_c, the critical wave number.

We observe from Table 1 that for the marginal state of stability ($\sigma = 0$), the critical values of the Taylor number and wave number increase by increasing the width of the gap between two concentric cylinders. However, when the amplification factor $\sigma(> 0)$ is increasing, there is also an increase in the minimum values of Taylor number, and opposite is the case when the amplification factor $\sigma(< 0)$ is decreasing; then the minimum values of Ta viz. Ta_m also decrease.

Table 1. Values of a_c, Ta_c ($\sigma = 0$) and a_m, Ta_m ($\sigma \neq 0$).

		η		a_c		Ta_c		$\mathrm{Ta}_{c(R)}$	
		0.85		3.130		3802		3805	
		0.5		3.162		6194		6199	
	η/σ	−2.0	−1.0	−0.5	−0.1	0.5	1.0	1.5	2.0
0.85	a_m	3.00	3.068	3.100	3.124	3.159	3.188	3.215	3.241
	Ta_m	3233	3515	3658	3773	3947	4097	4243	4393
0.5	a_m	3.033	3.100	3.132	3.156	3.191	3.219	3.247	3.273
	Ta_m	5289	5737	5964	6148	6425	6659	6895	7134
0.1	a_m	3.217	3.280	3.310	3.333	3.367	3.395	3.422	3.449
	Ta_m	56216	60318	62394	64065	66592	68716	70855	73011

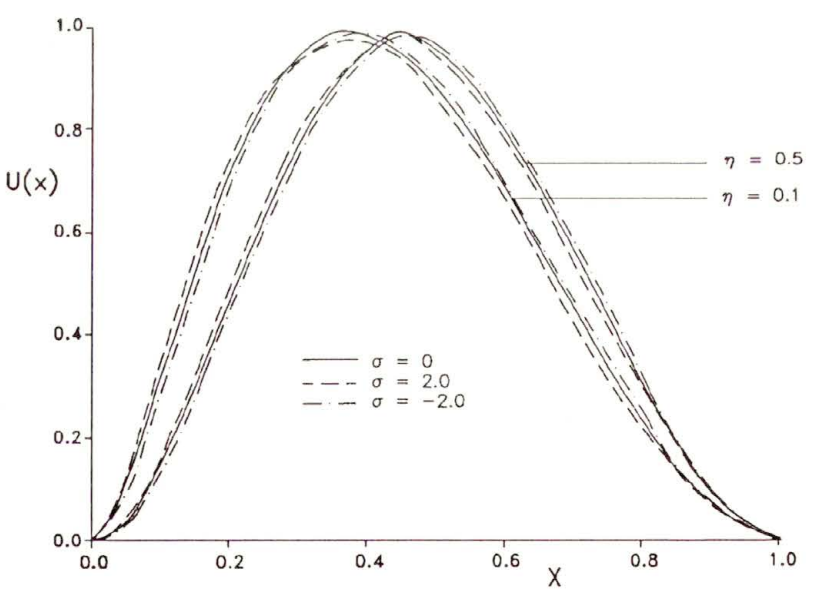

FIG. 1. Radial velocity $U(x)$.

The radial component of velocity disturbance $U(x)$ is shown in Fig. 1 for $\eta = 0.5, 0.1$ and for $\sigma = 0, \pm 2.0$. It is observed that $U(x)$ increases near the inner cylinder and decreases near the outer cylinder when $\sigma = 2.0$ as compared to that at the onset of instability ($\sigma = 0.0$), and opposite is the case when $\sigma = -2.0$. The cell patterns are shown for $\eta = 0.5$ and 0.1 for $\sigma = 0, \pm 2.0$ in Figs. 2–7. It is observed from these figures that the cells get reduced in size for $\sigma = 2.0$ and get enlarged in size for $\sigma = -2.0$ as compared to those at the onset of instability.

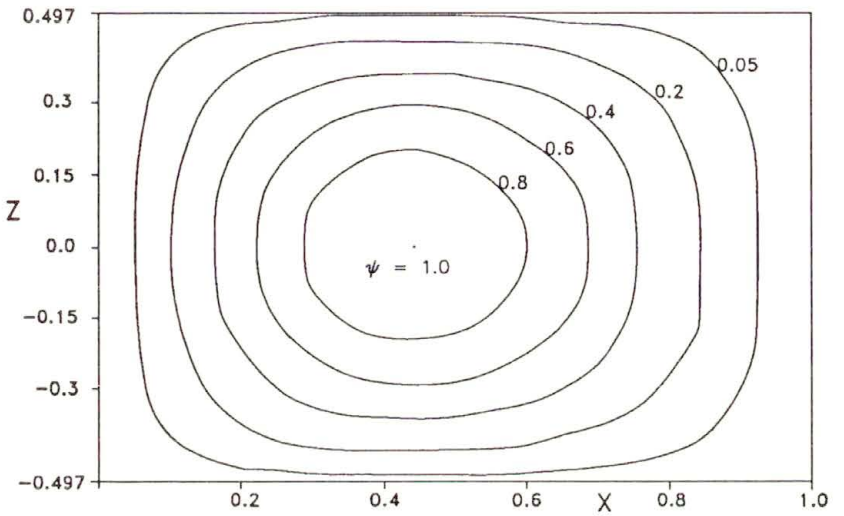

FIG. 2. Cell patterns at the onset of instability for $\eta = 0.5$ and $\sigma = 0.0$, $\psi = U(x)\cos aZ$.

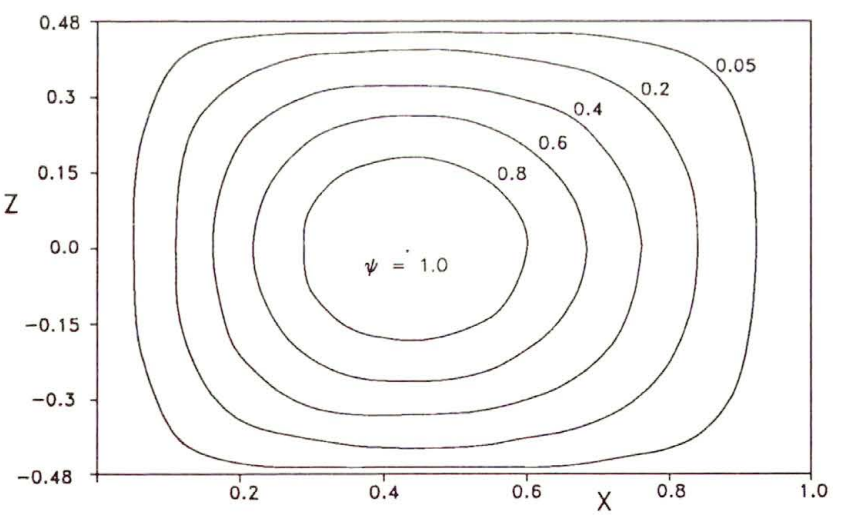

FIG. 3. Cell pattern instability for $\eta = 0.5$ and $\sigma = 2.0$.

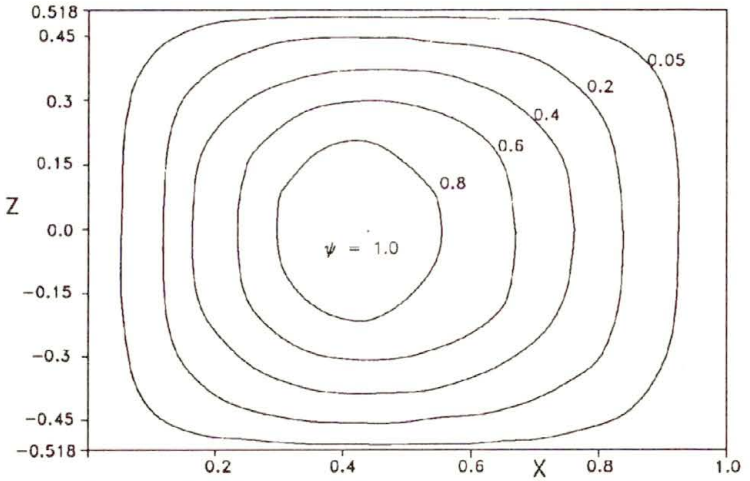

FIG. 4. Cell pattern instability for $\eta = 0.5$ and $\sigma = -2.0$.

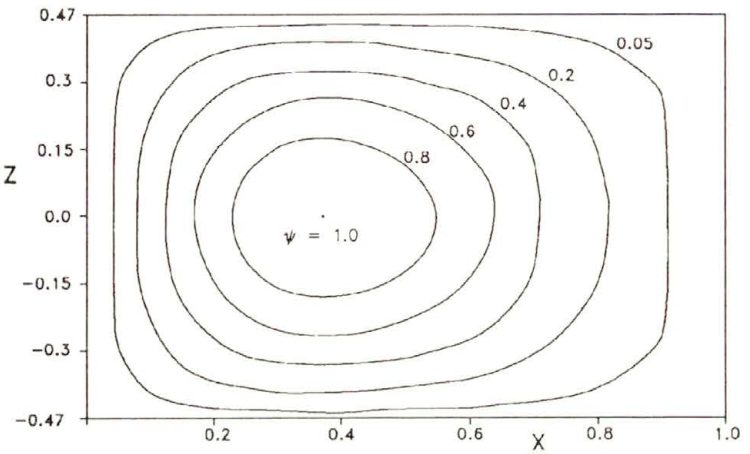

FIG. 5. Cell patterns at the onset of instability for $\eta = 0.1$ and $\sigma = 0.0$, $\psi = U(x)\cos aZ$.

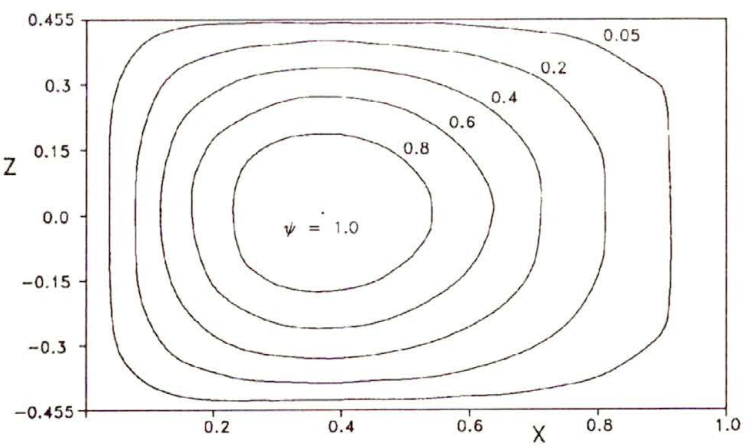

FIG. 6. Cell patterns for $\eta = 0.1$ and $\sigma = 2.0$.

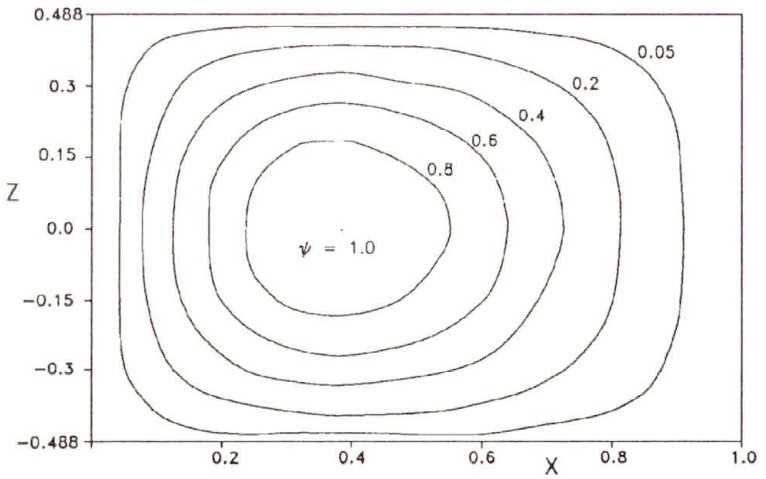

FIG. 7. Cell pattern for $\eta = 0.1$ and $\sigma = -2.0$

References

1. S. CHANDRASEKHAR, *The stability of viscous flow between rotating cylinders*, II, Proc. Roy. Soc. London, **268**, 145–152, 1962.
2. D.L. HARRIS and W.H. REID, *On the stability of viscous flow between rotating cylinders*, Pt 2. *Numerical analysis*, J. Fluid Mech., **20**, 95–101, 1964.
3. R.J. DONNELLY and K.W. SCHWARZ, *Experiments on the stability of viscous flow between rotating cylinders, VI. Finite amplitude experiments*, P.H. ROBERTS, (Appendix), Proc. Roy. Soc. London, **A 283**, 531–556, 1965.
4. V.M. SOUNDALGEKAR, H.S. TAKHAR and T.J. SMITH, *Effects of radial temperature gradient on the stability of viscous flow in an annulus with a rotating inner cylinder*, Wärme and Stoffübertragung, **15**, 233–238, 1981.
5. V.M. SOUNDALGEKAR, H.S. TAKHAR and M.A. ALI, *Stability of flow between rotating cylinders – Wide-gap problem*, ASME J. Fluids Engng., **111**, 97–77, 1989.
6. G.I. TAYLOR, *Stability of a viscous liquid contained between two rotating cylinders*, Phil. Trans. Roy. Soc. London, **A223**, 289–343, 1923.
7. J. WALOWIT, S. TSAO and R.C. DI PRIMA, *Stability of flow between arbitrarily spaced concentric cylindrical surfaces including the effect of a radial temperature gradient*, ASME J. Applied Mech., **31**, 585–593, 1964.

MANCHESTER SCHOOL OF ENGINEERING,
UNIVERSITY OF MANCHESTER, MANCHESTER, ENGLAND;
DEPARTMENT OF MATHEMATICS,
UNIVERISTY OF BAHRAIN, BAHRAIN (MIDDLE EAST)
and
BRINDAVAN SOCIETY, THANE, INDIA.

Received December 18, 1995.

Arch. Mech., 48, 5, pp. 843–868, Warszawa 1996

Singularities of aerodynamic transfer functions calculated on the basis of an unsteady lifting surface model in subsonic flow

M. NOWAK (WARSZAWA)

A DIRECT METHOD to calculate the Laplace transformed pressure distribution on subsonic lifting surfaces is considered. The kernel function is analytically continued in the entire p-plane (of the non-dimensional Laplace variable), and the discretizing procedure follows the lifting lines (or doublet-lattice) method developed for simple harmonic motion. The aerodynamic influence coefficient matrix is a function of Mach number M and the complex variable p. In the first part of the paper, some analytical properties of this matrix were investigated on the basis of numerical calculations performed for an aspect-ratio-3 rectangular wing. The main conclusion of this paper is that for $M \neq 0$, there exist a large (probably infinite) set of latent roots of the matrix in the left half of the p-plane which (usually) reflect in poles of the transfer functions. For $M \to 1$, all latent roots tend to the origin ($p = 0$). For $M \to 0$, all latent roots move to infinity and probably, for $M = 0$ there are no roots in the finite part of the p-plane. The distribution of latent roots in the p-plane does not depend on the number of aerodynamic elements introduced by the discretization (within the limits of accuracy of the calculation method). The algebraic equations are well-conditioned in the right half of the p-plane and in a strip parallel to the imaginary axis in left half of the p-plane. The width of this strip depends on the Mach number. In the second part of the paper, an approximation to the aerodynamic transfer functions based on the identified singularities and calculated left and right-hand latent vectors of the aerodynamic influence coefficients matrix is developed. It avoids the ill-posed analytical continuation from the imaginary axis in the whole p-plane. The results clarify also some unexpected phenomena observed in Laplace-domain calculations, and described in the literature.

1. Introduction

THE KNOWLEDGE of unsteady aerodynamic forces acting on a flexible aeroplane undergoing small perturbations from a steady equilibrium state of trimmed, rectilinear flight, is essential for stability analyses of the motion of the structure. The prediction of the unsteady aerodynamic loads is complicated by the fact that the unsteady flowfield surrounding the body is not determined solely by the instantaneous state variables of the structure, but it depends also on the past history of the motion of the body. The aerodynamic forces exhibit heredity due to the influence of vorticity shed into the wake at earlier instants of time.

The input data in a lifting surface aerodynamic model is the upwash distribution $w(x, y, t)$ on the wing surface S (Fig. 1). Assuming that all linear coordinates (x, y, z) are nondimensionalized by a reference length b (usual root semichord), and introducing nondimensional time t

$$t = \frac{U \cdot t_{\text{real}}}{b}, \tag{1.1}$$

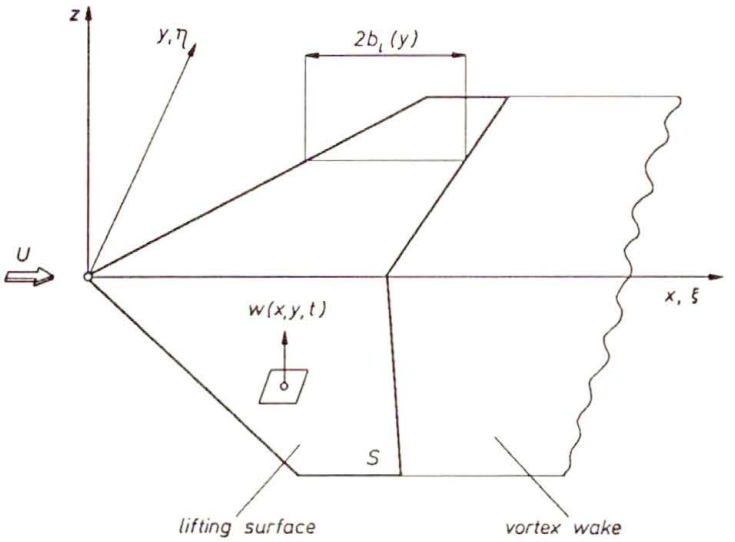

FIG. 1.

where U is the flight velocity, the expression for the upwash distribution can be put in the form

$$\frac{w(x,y,t)}{U} = \frac{\partial h}{\partial x} + \frac{\partial h}{\partial t}, \tag{1.2}$$

where $h(x,y,t)$ denotes the normal (nondimensionalized) displacements of the wing surface.

The lifting surface integral equation relates upwash and lifting pressure coefficient $c_p(x,y,t)$ (i.e. pressure difference $\Delta p(x,y,t)$ between the upper and lower surface, nondimensionalized by the dynamic pressure $\varrho U^2/2$) on the wing. The original form of the lifting surface equation, given in 1940 by KÜSSNER [1] applies to harmonic motion, when

$$w(x,y,t) = \hat{w}(x,y,ik)e^{ikt} \qquad \text{and} \qquad c_p(x,y,t) = \hat{c}_p(x,y,ik)e^{ikt}.$$

The lifting surface equation relates in this case the amplitudes of upwash and pressure coefficient

$$\frac{\hat{w}(x,y,ik)}{U} = \frac{1}{8\pi}\iint_S K(M,x_0,y_0,ik)\hat{c}_p(\xi,\eta,ik)\,d\xi\,d\eta, \tag{1.3}$$

where $x_0 = x - \xi$, $y_0 = y - \eta$, M stands for the Mach number and

$$k = \frac{\omega b}{U} \tag{1.4}$$

is the nondimensional frequency coefficient (called also reduced frequency).

The kernel of this equation is singular and the solution is sought in the class of functions vanishing on the trailing edge – this is a necessary condition for the uniqueness of the solution, and physically it expresses the Kutta condition.

There were developed many different methods for discretization of the lifting surface equation in the frequency domain. One of the most useful is the doublet-lattice technique of ALBANO and RODDEN [2]. The calculations in this paper were made mainly by the lifting-lines method [3, 4], with algorithms very similar to the doublet-lattice method, but usually with better convergence properties (with respect to the number of introduced aerodynamic elements).

For many years, the unsteady aerodynamic theories and its applications have focused primarily on the frequency domain, since the knowledge of aerodynamic forces at harmonic disturbances is sufficient for the determination of flutter boundaries. The advent of active control technology for flexible aircraft has renewed interest in unsteady aerodynamic forces given in the time and Laplace domains.

The displacements of the structure are usually described by means of a finite set of generalized coordinates $q_1(t)$, $q_2(t)$, ..., $q_n(t)$ defined on the basis of a set of assumed modes

$$h(x, y, t) = \sum_{k=1}^{n} h_k(x, y) \cdot q_k(t), \tag{1.5}$$

where the functions $h_k(x, y)$ $(k = 1, 2, \ldots, n)$ correspond in the most cases to natural vibration modes of the structure. The upwash distribution on the lifting surface may be expressed in terms of the generalized coordinates and generalized velocities

$$\frac{w(x, y, t)}{U} = \sum_{k=1}^{n} \frac{\partial h_k(x, y)}{\partial x} \cdot q_k(t) + \sum_{k=1}^{n} h_k(x, y) \cdot \dot{q}_k(t). \tag{1.6}$$

The generalized aerodynamic forces (related to the dynamic pressure and b^2) are defined by integrals taken over the surface

$$f_k(t) = \iint_S h_k(x, y) c_p(x, y, t)\, dS \qquad \text{for} \qquad k = 1, 2, \ldots, n. \tag{1.7}$$

The problem consist in determination of the generalized force vector $\{f(t)\}$ (with n elements (1.7)) for a given motion, described by the function $\{q(\tau)\}$ for $-\infty < \tau \leq t$. Independently of the details of the aerodynamic model, the aerodynamic operator which relates $\{q(t)\}$ to $\{f(t)\}$ possesses always some basic properties, such as single-valuedness, linearity, time-invariance and continuity. According to a theorem of SCHWARTZ [6], these four properties can be replaced

by the entirely equivalent condition that states that this operator has a (distributional) convolution representation.

$$\{f(t)\} = [A(M,t)] * \{q(t)\}, \tag{1.8}$$

where $[A(M,t)]$ is the unit impulse response matrix function (called also hereditary matrix [5]), the (j,k) element of which is the generalized indicial response in the j-th mode due to the pressure $c_p(x,y,t)$ generated by the motion in the k-th mode with $q_k(t) = \delta(t)$. The elements of this matrix depend also on the Mach number M. The aerodynamic forces can depend only on the history and not on the future of the motion. That means that the aerodynamic system is causal, and therefore

$$[A(M,t)] = 0 \qquad \text{for} \qquad t < 0.$$

Direct calculation of the elements of $[A(M,t)]$ for arbitrary time may be difficult and in practice, these functions are usually determined only by means of the inversion of Fourier or Laplace transform. Taking the Laplace transformation of the convolution (1.8) it follows that

$$\left\{\hat{f}(p)\right\} = \left[\hat{A}(M,p)\right] \{\hat{q}(p)\}, \tag{1.9}$$

where p is the Laplace variable, and the circumflex accents ($\hat{}$) denote transforms

$$\left\{\hat{f}(p)\right\} = \mathcal{L}\{f(t)\}, \quad \{\hat{q}(p)\} = \mathcal{L}\{q(t)\} \quad \text{and} \quad \left[\hat{A}(M,p)\right] = \mathcal{L}[A(M,t)]. \tag{1.10}$$

The aerodynamic transfer functions matrix $[\hat{A}(M,p)]$ is a Laplace transform of a real distribution and is real whenever p is real. Hence

$$\left[\hat{A}(M,p)\right]^* = \left[\hat{A}(M,p^*)\right], \tag{1.11}$$

where the star (*) denotes complex conjugate values.

The convolution (1.8) and the Laplace transformation should be interpreted on the basis of the theory of distributions [6]. The aerodynamic transfer functions grow with increasing $|p|$ like $O(|p|)$ in the case of compressible flow, and like $O(|p|^2)$ in the incompressible case. Additionally, the distributional Laplace transform does not contain explicitly the initial values and this simplifies the analysis.

If the Laplace variable is pure imaginary $p = ik$, then (1.9) determines the steady-state frequency response function, which relates the amplitudes of generalized coordinates to the amplitudes of generalized forces in harmonic motion.

$$\left\{\hat{f}(ik)\right\} = \left[\hat{A}(M,ik)\right] \{\hat{q}(ik)\}, \tag{1.12}$$

where k is the frequency coefficient defined in (1.4), and

$$\left[\widehat{A}(M, ik)\right] = \lim_{p \to ik} \left[\widehat{A}(M, p)\right]$$

is the matrix of harmonic transfer functions.

The elements of the matrix $[\widehat{A}(M, ik)]$ can be calculated numerically for given values M and k on the basis of the lifting surface equation (1.3), and the relations (1.6) and (1.7).

The aerodynamic transfer matrix $[\widehat{A}(M, p)]$ is the final product of aerodynamic calculations and it is usually determined by means of the analytic continuation of the elements of matrix $[\widehat{A}(M, ik)]$ from the imaginary axis into the whole complex plane. Two types of approximation have been used in practice for this purpose.

1. The first approach begins with calculating the values of harmonic transfer functions over a specified range of the frequency coefficients $k = k_1, k_2, \ldots, k_m$. Next, the harmonic transfer functions are approximated by rational function which fit best the calculated values. The last step is the analytic continuation of the resulting rational functions into the whole p-plane.

2. In the second (direct) approach, the kernel function of the lifting surface equation (1.3) is extended from the imaginary axis to the entire complex plane $K(M, x_0, y_0, ik) \to K(M, x_0, y_0, p)$ by means of an exact analytic continuation. The elements of the aerodynamic transfer functions are calculated directly for a given value of the Laplace variable p on the basis of this generalized lifting surface equation.

Both approaches have their own advantages and disadvantages. The elements of $[\widehat{A}(M, p)]$ are holomorphic functions with branch points $p = 0$ and $p = -\infty$ (for $M < 1$) which are neglected in the approximation by rational functions. When the transfer functions are approximated by polynomials or rational functions, it is possible to cast the aeroelastic (and aeroservoelastic) equations of motion in a linear time-invariant state-space form (instead of integro - differential form), although the size of the state vector increases due to the approximation. Currently there are three basic formulations used in approximating the aerodynamic transfer functions by means of rational functions: least-squares [7], modified matrix-Padé [8] and minimum-state [9]. The common disadvantage of these methods is the necessity of numerical realisation of an ill-posed analytic continuation.

The direct analytic continuation of functions which appear in the expression for harmonic aerodynamic forces gave rise in the past to arguments on the validity of the results in the left-hand half-plane of the Laplace variable and was rejected in a series of articles [10, 11, 12, 13]. This problem was later resolved by MILNE [14], EDWARDS [15] and others, but nowadays some doubts arose about the possibility of a practical utilisation of this approach [16]. It was stated, that the application of numerical solution techniques to the integral equation in the left half of the p-plane may result in a highly ill-conditioned set of linear equations

[17]. An unexpected phenomenon was observed in [18]. Some of the generalized forces for strongly decaying motions in high subsonic flow reveal a looping behaviour in the complex plane. Testing a new method for solving the lifting surface equation, UEDA [19] stated that "it looks probable that an aerodynamic singularity exists in the left half of the p plane when the flow becomes high subsonic".

Another interesting approach to the approximation of transfer functions was given by STARK [20]. He proposed an expression for the lift deficiency function in the time domain and assumed that this function is independent of the deflection mode of the wing. Laplace transform of his deficiency function possesses branch points $p = 0$ and $p = -\infty$, which are the only singularities of the transfer functions in the entire p plane. This approach leads to a good approximation in the incompressible case, but for non-zero Mach numbers the results were less satisfactory.

The knowledge of analytic properties of aerodynamic transfer functions in the p plane is until now only fragmentary. It is known that the matrix $[\widehat{A}(M, p)]$ cannot have any poles in the right half of the p plane, since the transient aerodynamic response is always stable. It is also known, that for a subsonic flow, the aerodynamic transfer functions have logarithmic branch points $p = 0$ and $p = -\infty$, as a result of the unlimited length of the wake. It is usually expected, that the aerodynamic transfer functions have no poles also in the left half of the p plane [5]. This is true for the exact solution of two-dimensional airfoil in incompressible flow but was newer proved for the compressible case. The problem is additionally complicated by the fact, that the solution of the singular integral equation is ill-posed, but the numerical methods used in the chordwise integration introduce a self-regularization and, after discretization, the resulting set of algebraic linear equations is usually well-conditioned. In the two-dimensional case (of an airfoil) the proof of this statement was given by LIFANOV [21].

The aim of this paper is to investigate the numerical problems which occur in solving the lifting surface equation in the Laplace domain and the analytical properties of the transfer functions in the left half of the p plane (for decaying motion). Particular attention will be paid to the conditioning of the linear algebraic equations obtained by the discretization of the lifting surface equation, and to the identification of singularities of the transfer functions.

2. Lifting surface equation in the Laplace domain

The lifting surface equation in the Laplace domain is the result of an analytic continuation of the kernel of (1.3). Formally, the variable ik should be replaced by the Laplace variable

$$\frac{\widehat{w}(x, y, p)}{U} = \frac{1}{8\pi} \iint_S K(M, x_0, y_0, p)\widehat{c}_p(\xi, \eta, p)\, d\xi\, d\eta. \tag{2.1}$$

The singular kernel (in the case of a flat surface) may be expressed in the form [4, 18]

$$K(M, x_0, y_0, p) = \frac{1}{r^2}\left[\left(1 + \frac{x_0}{R}\right) e^{-pru} - F(pr, u)\right] e^{-px_0}, \tag{2.2}$$

where

$$\begin{aligned} x_0 &= x - \xi, \\ y_0 &= y - \eta, \\ r &= |y_0|, \\ R &= \sqrt{x_0^2 + \beta^2 r^2}, \\ \beta &= \sqrt{1 - M^2}, \\ u &= \frac{MR - x_0}{\beta^2 r}. \end{aligned}$$

The function $F(z, u)$ is defined by means of following integrals

$$F(z,u) = \begin{cases} z\int\limits_u^{\infty}\left(1 - \dfrac{\eta}{\sqrt{1+\eta^2}}\right) e^{-z\eta}\, d\eta & \text{for} \quad \operatorname{Re}(z) \geq 0, \\ 2e^{zu} + z\int\limits_{-\infty}^{u}\left(1 + \dfrac{\eta}{\sqrt{1+\eta^2}}\right) e^{-z\eta}\, d\eta - i\pi z H_1^{(1)}(z) & \\ & \text{for} \quad \operatorname{Re}(z) < 0. \end{cases} \tag{2.3}$$

The second expression (for $\operatorname{Re}(z) < 0$) may be obtained from the first integral in (2.3), by an appropriate contour deformation.

Only a few papers (e.g. [18, 19, 22]) are known which are devoted to the numerical problems which occur in lifting surface calculations in the Laplace domain.

For small values of $|pru|$ it is convenient to split $F(pr, u)$ into two parts [18]

$$F(z, u) = F_1(z) - F_2(z, u), \tag{2.4}$$

where

$$F_1(z) = z\int\limits_0^{\infty}\left(1 - \frac{\eta}{\sqrt{1+\eta^2}}\right) e^{-z\eta}\, d\eta = 1 + z - \frac{\pi}{2} z(H_1(z) - Y_1(z)), \tag{2.5}$$

and the integral $F_2(z, u)$ defines an entire analytic function of the z variable, which may be expanded into a convergent series

$$F_2(z,u) = z\int\limits_0^{u}\left(1 - \frac{\eta}{\sqrt{1+\eta^2}}\right) e^{-z\eta}\, d\eta = \sum_{k=0}^{\infty} \frac{1}{(k+1)!} g_k(u) z^{k+1}, \tag{2.6}$$

with a recursive relationship for the coefficients

$$g_k(u) = \left(u\left(\sqrt{1+u^2}-u\right) - \frac{k}{k-1}\right)(-u)^{k-1} - \frac{k}{k-1}g_{k-2}(u) \qquad \text{for} \quad k \geq 2$$

and initial terms

$$g_0(u) = 1 - \left(\sqrt{1+u^2}-u\right),$$
$$g_1(u) = \ln\left(\sqrt{1+u^2}-u\right) + u\left(\sqrt{1+u^2}-u\right).$$

The Hankel function $H_1^{(1)}(z)$, Struve function $H_1(z)$ and Bessel function of the second kind $Y_1(z)$ may be calculated with high accuracy on the basis of the series given e.g. in [23]. The accuracy of the series (2.6) is limited by the numerical instability due to the round-off error in actual computation. For calculations performed with double precision, this limit depends on the values of parameters and sufficient accuracy can be achieved only if $|pru| < 6$. For larger values of parameters it is necessary to provide other approximations.

For very large values of $|pr|$ satisfactory results may be obtained from the asymptotic expansion derived by means of integrating by parts the integrals in (2.3)

$$F(z,u) \approx \left(1 - f(u) - \sum_{k=1} f^{(k)}(u)\frac{1}{z^k}\right) e^{-zu} - \begin{cases} 0 & \text{for} \quad \mathrm{Re}\,(z) \geq 0, \\ i\pi z H_1^{(1)}(z) & \text{for} \quad \mathrm{Re}\,(z) < 0, \end{cases} \tag{2.7}$$

where

$$f^{(0)}(u) = f(u) = \frac{u}{\sqrt{1+u^2}}, \qquad f^{(-1)}(u) = \sqrt{1+u^2},$$

and

$$f^{(k+1)}(u) = -\frac{1}{1+u^2}\left[(2k+1)uf^{(k)}(u) + (k+1)(k-1)f^{(k-1)}(u)\right].$$

The asymptotic series (2.7) is usually divergent and only a limited number of terms can be employed in the calculations.

Very useful in practice is an exponential approximation for the integrands of (2.3)

$$1 - \frac{\eta}{\sqrt{1+\eta^2}} \approx \sum_{k=1}^{12} a_k \exp(-2^k b_0 \eta),$$

proposed by Desmarais. The values of coefficients b_0 and a_k are given in [18]. The resulting rational approximation of $F(pr,u)$ may be used in the range $\pi/4 < |\arg(p)| < 3\pi/4$ and $-\infty < u < \infty$.

3. Discretization of the lifting surface equation

It is possible, independently of the used discretization procedure, to distinguish three steps of calculations of the values of transfer functions for a structure with n degrees of freedom (for a given Mach number M and value of the Laplace variable p):

a. Calculation of the substantial derivative to obtain a N-dimensional approximation of the upwash distribution

$$\underset{N\times 1}{\{\hat{w}(p)\}} = \Big(\underset{N\times n}{[D_1]} + p\,\underset{N\times n}{[D_2]}\Big)\underset{N\times 1}{\{\hat{q}(p)\}}; \tag{3.1}$$

b. Solution of a linear system of algebraic equations

$$\underset{N\times 1}{\{\hat{w}(p)\}} = \underset{N\times N}{[K(M,p)]}\,\underset{N\times 1}{\{\hat{c}_p(p)\}}; \tag{3.2}$$

c. Determination of the transforms of the generalized coordinates

$$\underset{N\times 1}{\big\{\hat{f}(p)\big\}} = \underset{n\times N}{[S]}\,\underset{N\times 1}{\{\hat{c}_p(p)\}}. \tag{3.3}$$

N is the size of the aerodynamic influence coefficients matrix which approximates the integral operator. The vectors $\{\hat{w}(p)\}$ and $\{\hat{c}_p(p)\}$ describe the upwash and pressure distributions on the wing surface. In practice, typical values are: $n = 20 \div 30$ and $N \sim$ some hundreds (but always $N \gg n$). The differentiation matrices $[D_1]$, $[D_2]$ are determined by the formula (1.6), and the integration matrix $[S]$ by the definition of the generalized forces (1.7). These constant matrices depend only on the used discretization method. Matrix $[K(M,p)]$ depends also on the Mach number and on the assumed value of p. The evaluation of this matrix is the most time-consuming part of the computation.

Equations (3.2), (3.3) and (3.4) may be put together in the form

$$\underset{n\times 1}{\big\{\hat{f}(p)\big\}} = \underset{n\times N}{[S]}\,\underset{N\times N}{[K(M,p)]^{-1}}\,\Big(\underset{N\times n}{[D_1]} + p\,\underset{N\times n}{[D_2]}\Big)\underset{N\times 1}{\{\hat{q}(p)\}}.$$

Hence, the aerodynamic transfer functions matrix is given by the formula

$$\underset{n\times n}{\big\{\hat{A}(M,p)\big\}} = \underset{n\times N}{[S]}\,\underset{N\times N}{[K(M,p)]^{-1}}\,\Big(\underset{N\times n}{[D_1]} + p\,\underset{N\times n}{[D_2]}\Big). \tag{3.4}$$

If the discretization procedure in the Laplace domain is the same as in the frequency domain (when $p = ik$), then the matrix (3.4) is the result of an exact

analytical continuation of the harmonic transfer matrix $[\widehat{A}(M, ik)]$. On the other hand, if the calculations for a harmonic motion are based on the discretized equation (1.3), then the values of analytic functions determined in the entire complex plane p are in fact calculated on the imaginary axis. Therefore, the knowledge of analytic properties of the transfer functions may be useful also in the case, when the calculations are restricted to the imaginary axis only.

In the case of subsonic flow ($M < 1$), the elements of $[K(M, p)]$ have a branch-point in the origin ($p = 0$) and from the expression for the kernel function, it follows that

$$[K(M, p)] = [K(M, 0)] + O(p^2 \ln(p)) \qquad \text{for} \quad p \to 0.$$

The transfer functions are holomorphic functions in the complex plane cut along the negative real axis.

Poles of the transfer matrix $[\widehat{A}(M, p)]$ may exist only in those points of the p plane, where the matrix $[K(M, p)]$ is singular

$$\det\left([K(M, p)]\right) = 0. \tag{3.5}$$

The number of latent roots of the equation (3.5) may be large or infinite, because the elements of the matrix $[K(M, p)]$ are transcendental functions of p.

4. Condition number and latent roots of the aerodynamic influence coefficients matrix

Most of the calculations in the following analysis were performed for a rectangular wing with an aspect ratio $A = 3$ in symmetric motion. This wing was also investigated in [20] and [22]. For the discretization, the lifting lines method [4] was used, but some of the calculations were repeated with the doublet-lattice method [2] (with the same or almost the same results).

The sensitivity of the solution of (3.2) to the perturbation of the data

$$\frac{\|\Delta \widehat{c}_p(p)\|}{\|\widehat{c}_p(p)\|} \leq \operatorname{cond}[K(M, p)] \frac{\|\Delta \widehat{w}(p)\|}{\|\widehat{w}(p)\|},$$

may be measured by the condition number of $[K(M, p)]$ defined as the product of two matrix norms

$$\operatorname{cond}[K(M, p)] = \|[K(M, p)]\| \cdot \left\|[K(M, p)]^{-1}\right\| \qquad (1 \leq \operatorname{cond} \leq \infty). \tag{4.1}$$

Logarithm to the base 10 of the condition number can be used to estimate the number of significant digits of the result which can be lost, independently of the accuracy of the method used to solve the linear equations. Hence, if the

calculations are performed with double precision, then the matrix is numerical singular when $\log_{10}\left(\text{cond}\,[K(M,p)]\right) > 16$.

The conditions numbers of the matrix $[K(M,p)]$ were calculated for different Mach numbers and for a large set of p-values by means of the SVD algorithms [24] for complex matrices.

Figure 2 shows the results of calculations made for Mach number $M = 0.8$ in a large region $-1.75 \leq \text{Re}\,(p) \leq 0$, $0 \leq \text{Im}\,(p) \leq 3.5$ of the complex p-plane. The size of the aerodynamic model was $N = 10 \times 20 = 200$ elements (10 lifting lines and 20 strips uniformly distributed on the half-span of the wing). It is seen, that the matrix in this region is well-conditioned, although for $\text{Re}\,(p) < -1.0$ the condition number grows very fast. There are also many local "spikes" which may indicate, that in its neighbourhood exist singular points of the matrix $[K(M,p)]$.

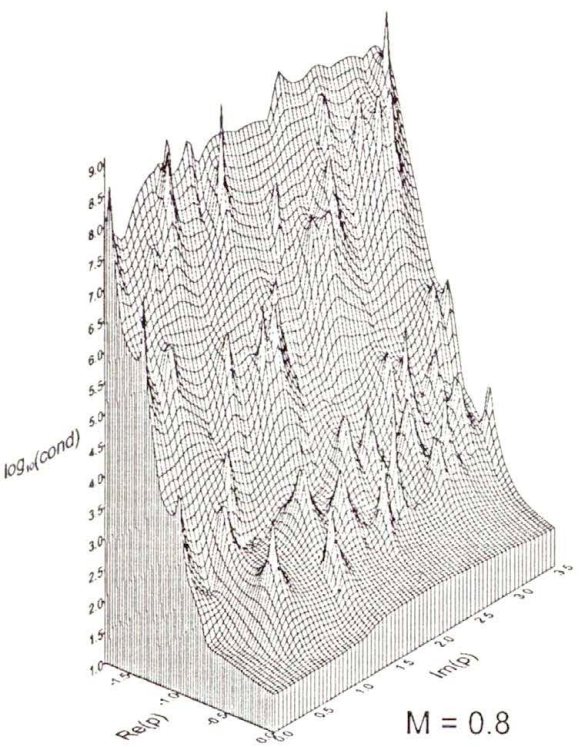

FIG. 2.

Figure 3 shows the same results in the form of a contour map. The latent roots were also calculated on the basis of Eq. (3.5) by means of the Muller method [25]. The results of these calculations are posted on the contour map in the form of black dots. In each of this calculated points $\log_{10}(\text{cond}\,[K(M,p)]) > 16$, hence the matrix is numerically singular. The initial values for the Muller iteration procedure were determined on the basis of the shape of contour lines. The condition

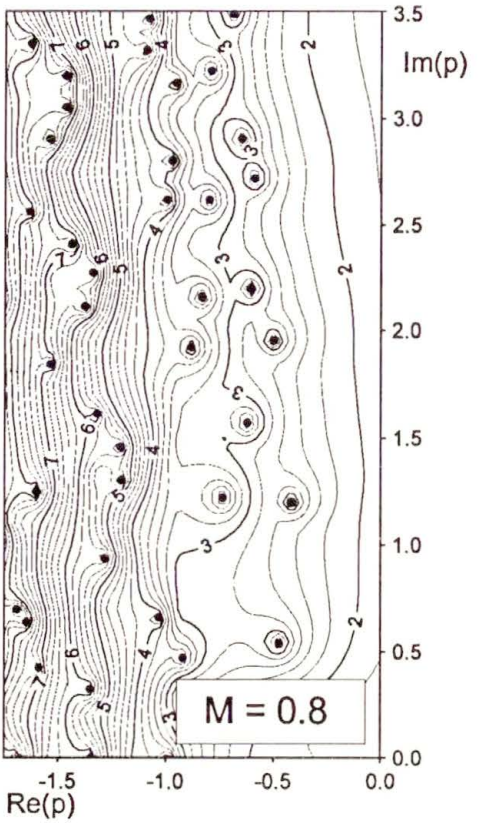

FIG. 3.

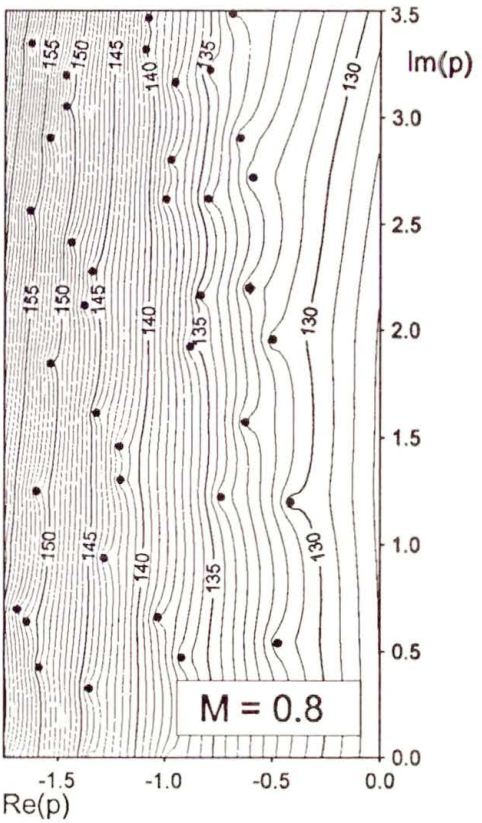

FIG. 4

number grows rapidly only in the vicinity of each root. In a very small region ($|\Delta p| < 10^{-8}$), the determinant of the matrix decreases usually by a factor about 10^{-10}, although its value may be still very large. This is shown in the Fig. 4, where the contour lines correspond to constant values of $\log_{10} |\det[K(M,p)]|$.

The singular points exist for each Mach number in the range $0 < M < 1$. The root distributions at Mach numbers $M = 0.5$, 0.7, 0.9 and 0.95 are shown in Fig. 5. It is seen that, as the Mach number increases, the width of the strip in the left half of the p plane where the matrix $[K(M,p)]$ is well-conditioned, decreases. At the same time, all latent roots move in the direction to the origin. This phenomenon is shown in Fig. 6 where the loci of about 20 selected roots are depicted. The outer ends of these curves correspond to the Mach number $M = 0.5$, and the inner ends to $M = 0.9$.

On the basis of Fig. 5 and Fig. 6 it is possible to formulate a hypothesis that for $M \to 0$, all roots move to infinity and in the incompressible case $M = 0$, there are no roots in the finite part of the plane $|p| < \infty$. On the other side, for $M \to 1$, all roots move to the origin and may significantly influence the behaviour of transfer functions at high subsonic Mach numbers.

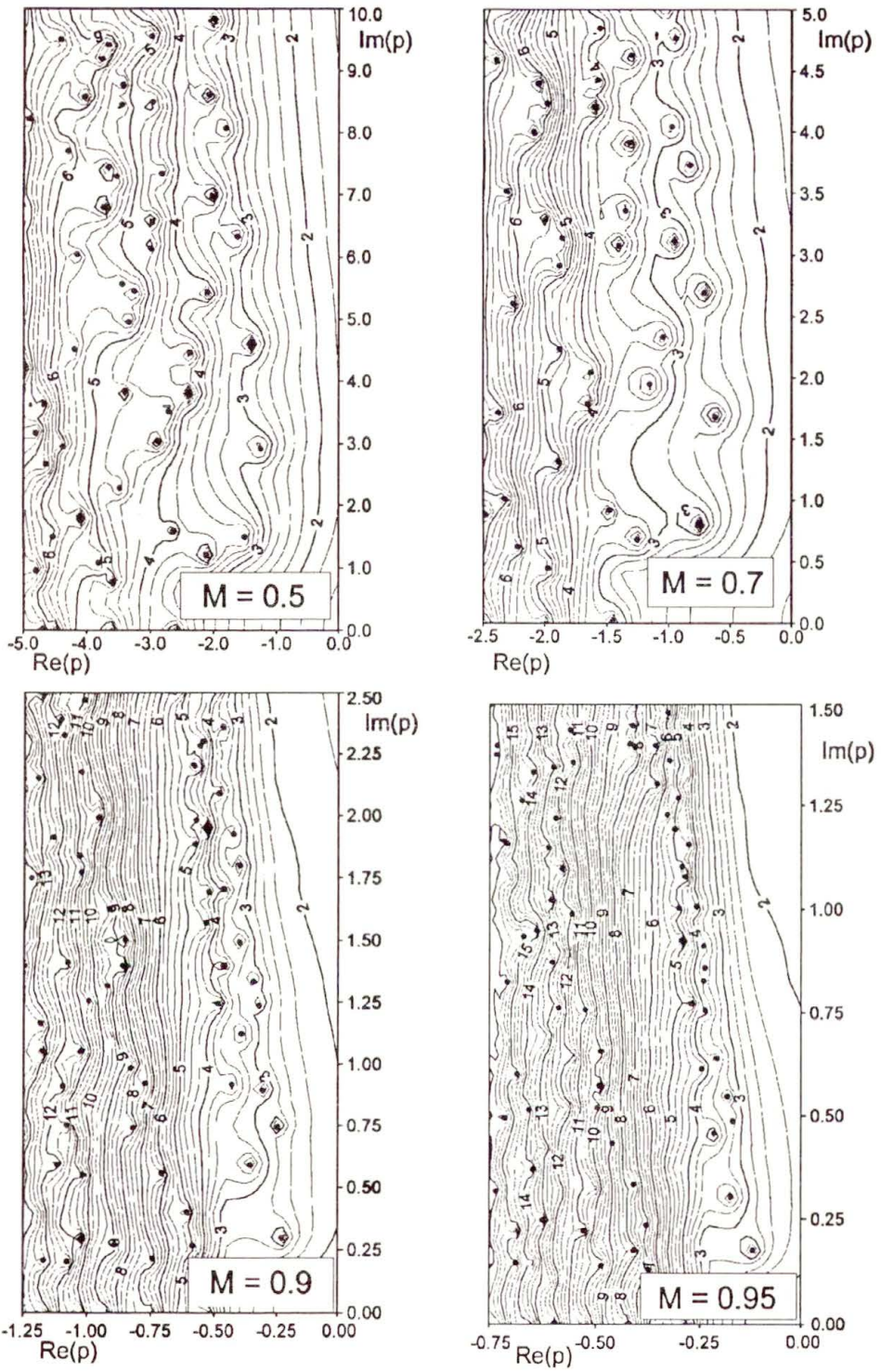

FIG. 5.

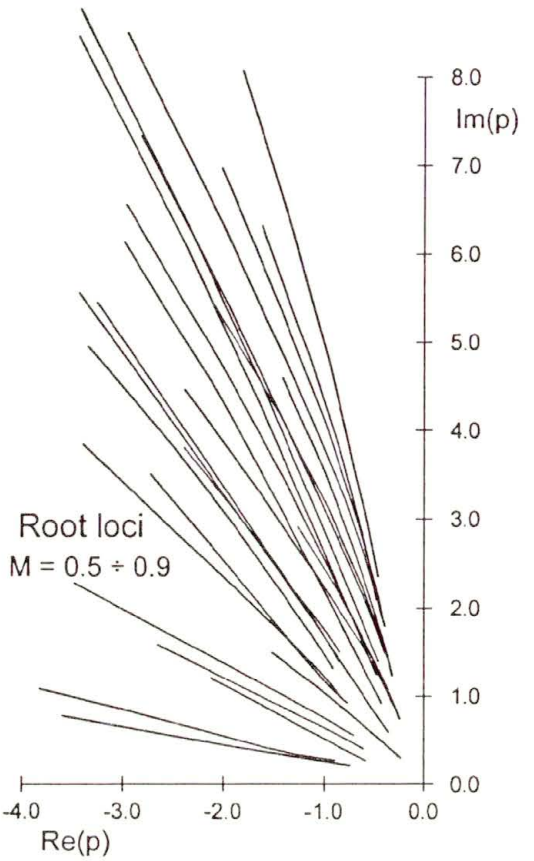

FIG. 6.

It has been found in the example of a rectangular wing, but also for other surface configurations, that the calculated roots of Eq. (3.5) were always simple roots only.

5. The influence of the discretization on the distribution of latent roots

It is not clear if the roots have a physical meaning and are related to the lifting surface equation or if they occur only in numerical calculations and are related to the discretized problem.

Figure 7 shows the influence of the size ($N = 48 \div 437$) of the matrix $[K(M, p)]$ on the distribution of latent roots in the p plane. The calculations were made by means of the lifting lines method, for a rectangular wing, at Mach number $M = 0.8$. It is seen that the differences may be related to the accuracy which may be achieved with the different models. For large values of the frequency coefficient, the pressure distribution is oscillating along the chord (Kutta waves)

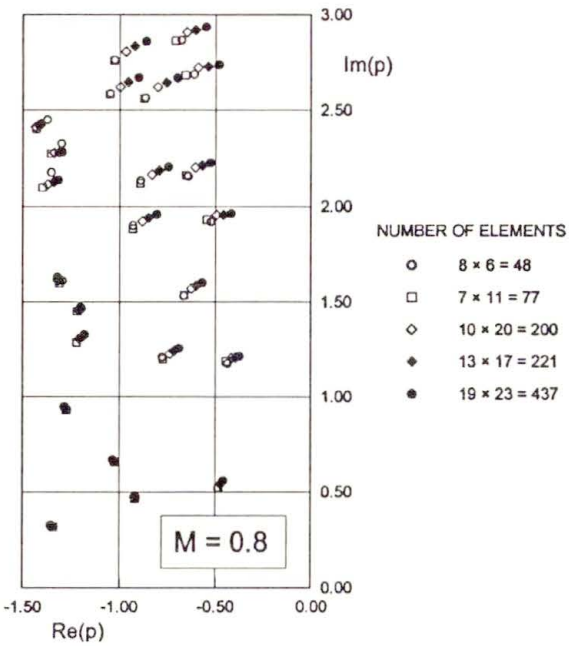

FIG. 7.

and a large number of aerodynamic elements is required at the discretization. It follows, that the number and distribution of latent roots do not depend on the size of matrix $[K(M,p)]$ within the limits of accuracy of the used method.

The most time-consuming part of the procedure to calculate latent roots is the search for a good initial approximation. The results presented in Fig. 7 suggest a practical approach, which may be applied for an arbitrary large N. The process should be divided into a sequence of steps, in which the number of aerodynamic elements increases $N_1 < N_2 < \ldots < N$. The results obtained in each step are used as the initial values for the next step. The choice of the initial approximations for the first step may be not strenuous if N_1 is small enough.

6. Approximation to the transfer matrix in the vicinity of its poles

The resolvent [26] of the matrix $[K(M,p)]$ for a given p has the form

$$\left([K(M,p)] - \lambda[I]\right)^{-1} = \sum_{j=1}^{N} \frac{\{u_j(p)\}\{v_j(p)\}^T}{\lambda_j(p) - \lambda}, \tag{6.1}$$

where the scalar parameter λ is distinct from the eigenvalues $\lambda_j(p)$, $j = 1, 2, \ldots, N$ of $[K(M,p)]$, while $\{u_j(p)\}$ and $\{v_j(p)\}$ are the right and left eigenvectors associated with $\lambda_j(p)$, and normalised in such a way, that $\{v_j(p)\}^T\{u_j(p)\} = 1$. The relation (6.1) is true also for $\lambda = 0$, because the matrix $[K(M,0)]$ is not singular.

The derivative of an eigenvalue of $[K(M,p)]$ is given by the expression

$$\frac{d}{dp}\lambda_j(p) = \{v_j(p)\}^T \left(\frac{\partial}{\partial p}[K(M,p)]\right)\{u_j(p)\}. \tag{6.2}$$

If the value $p = p_k$ is a latent root of (3.5), then at least one of the eigenvalues $\lambda_j(p_k)$, $j = 1,2,...,N$ is equal to zero and in the vicinity of p_k

$$\lambda_j(p) \approx \frac{d}{dp}\lambda_j(p)\Big|_{p=p_k} \cdot (p - p_k). \tag{6.3}$$

On the basis of (6.1), (6.2) and (6.3) it is possible to obtain an approximation to the matrix inverse $[K(M,p)]^{-1}$ in the proximity of the root $p = p_k$

$$[K(M,p)]^{-1} = \frac{\{u_k\}\{v_k\}^T}{p - p_k} + [\Phi(p)], \tag{6.4}$$

where $[\Phi(p)]$ is a regular function in the vicinity of p_k, while the latent vectors $\{u_k\}$ and $\{v_k\}$ are non-trivial solutions of the sets of homogeneous equations

$$\{v_k\}^T[K(M,p_k)] = 0 \qquad \text{and} \qquad [K(M,p_k)]\{u_k\} = 0, \tag{6.5}$$

normalised in such a way, that

$$\{v_k\}^T \frac{\partial}{\partial p}[K(M,p)]\Big|_{p=p_k} \{u_k\} = 1. \tag{6.6}$$

It follows from (6.4) that the latent roots of (3.5) usually reflect in poles of the transfer functions (3.4). However, there are two obvious exceptions to this rule.

If $\{s_i\}^T$ is the i-th row of the integration matrix $[S]$ which was defined in (3.3) and $\{s_i\}^T\{u_k\} = 0$, then the latent root p_k is not a pole of the functions (elements) in the i-th row of the transfer matrix $[\widehat{A}(M,p)]$. Similarly, if $\{d_{1j}\}$ and $\{d_{2j}\}$ are the j-th columns of the differentiation matrices $[D_1]$ and $[D_2]$ and at the same time $\{v_k\}^T\{d_{1j}\} = 0$ and $\{v_k\}^T\{d_{2j}\} = 0$, then the latent root p_k is not a pole of the functions in the j-th column of the transfer matrix.

The right latent vector which is a solution of the second homogeneous equation (6.5) determines a pressure distribution. In Fig. 8 and Fig. 9 two examples of such pressure distributions are shown which are associated with two latent roots. It has been numerically proved, that the shapes of these functions do not depend on the number of aerodynamic elements used to the discretization of the integral equation.

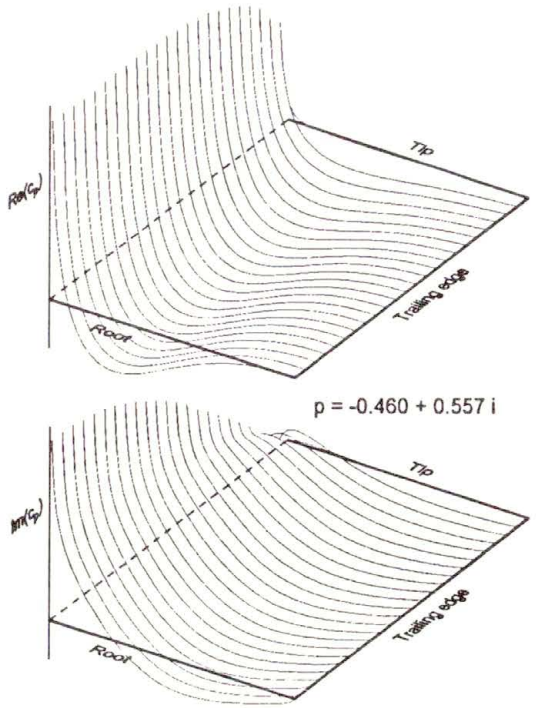

FIG. 8.

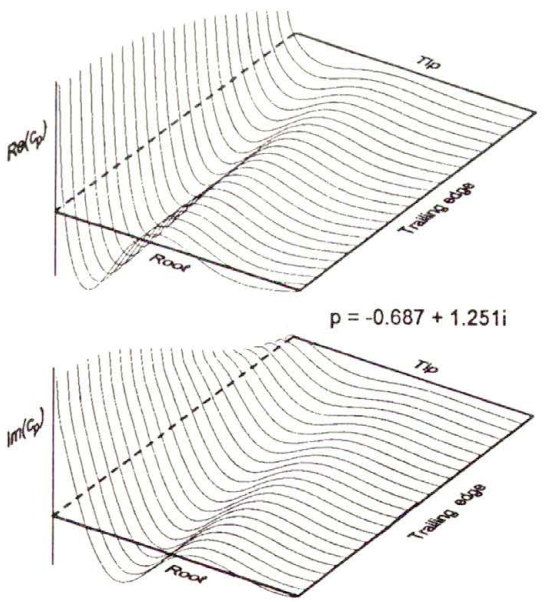

FIG. 9.

7. Approximation to the transfer functions based on their singularities

The knowledge of the singularities: branch points ($p = 0$ and $p = -\infty$) and poles (latent roots of (3.5)) makes possible an approximate reconstruction of the transfer functions in the entire complex p plane, without any use of analytic continuation (from the imaginary axis).

The solution of the equation may be put in the form

$$\{\hat{c}_p(p)\} = \left[\hat{A}e(p)\right]\{\hat{w}(p)\}, \tag{7.1}$$

where $\left[\hat{A}e(p)\right]$ denotes the matrix $[K(M,p)]^{-1}$.

To simplify the notation, in (7.1) and later in this section, the dependence of the matrices which define the aerodynamic system on the Mach number was not marked explicitly.

It should be emphasised that the relation (7.1) concerns the aerodynamic model only and does not depend on the definition of generalized coordinates used to describe the motion of the structure.

The inverse Laplace transform $\mathcal{L}^{-1}$ applied to (7.1) gives the relation between upwash and pressure distributions in the time domain in the form of a convolution

$$\{c_p(t)\} = [Ae(t)] * \{w(t)\}, \tag{7.2}$$

where the elements of $[Ae(t)]$ are the responses $\{c_p(t)\}$ which result from a unit impulse $\delta(t)$ in the elements of the discretized upwash distribution $\{w(t)\}$. In practice, it is usually more convenient to use inditial functions $[H(t)]$, which are responses to a unit step change in the (discretized) upwash distribution. From (7.2) it follows that

$$\{c_p(t)\} = [H(t)] * \{\dot{w}(t)\}, \tag{7.3}$$

where $\{\dot{w}(t)\}$ is the derivative with respect to time t of the upwash vector $\{w(t)\}$. The inditial functions $[H(t)]$ are related to the hereditary functions $[Ae(t)]$,

$$\left[\hat{H}(p)\right] = \frac{1}{p}\left[\hat{A}e(p)\right] \quad \text{and} \quad [H(t)] = \mathcal{L}^{-1}\left[\hat{H}(p)\right] = [Ae(t)] * 1_+(t), \tag{7.4}$$

where $1_+(t)$ is the unit step function (Heaviside function).

From the final value theorem [6] it follows

$$[H(\infty)] = \lim_{t\to\infty}[H(t)] = \lim_{p\to 0}\left[\hat{A}e(p)\right] = \left[\hat{A}e(0)\right] = [K(M,0)]^{-1}. \tag{7.5}$$

This limit corresponds to the steady solution (for constant boundary conditions on the surface). In compressible flow ($M \neq 0$), there exists also the limit given by the initial value theorem

$$[D] = \lim_{t\to 0+}[H(t)] = \lim_{p\to\infty}\left[\hat{A}e(p)\right] = \left[\hat{A}e(\infty)\right], \tag{7.6}$$

which can be calculated directly on the basis of the piston theory [27]

$$c_p(x, y, 0+) = \frac{4}{M} \frac{w(x, y, 0+)}{U} \tag{7.7}$$

(discretization of this relation in the method of the lifting lines is given in the Appendix).

In the incompressible flow ($M = 0$) the limiting values (7.6) do not exist, but

$$[M_A] = \lim_{t \to 0+} \left([H(t)] * 1_+(t) \right) = \lim_{p \to \infty} \left(\frac{1}{p} \left[\hat{A}e(p) \right] \right) \tag{7.8}$$

is the apparent mass matrix, which can be determined on the basis of a simplified model (without wake).

Taking into account the properties of the elements of the matrix $[K(M, p)]$, it is possible to obtain (e.g. [5, 8, 14]) an asymptotic representation

$$\left[\hat{A}e(p) \right] - \left[\hat{A}e(0) \right] = O(p^2 \ln p) \qquad \text{for} \quad p \to 0. \tag{7.9}$$

It follows, that in the time domain

$$[H(t)] - [H(\infty)] = O(t^{-2}) \qquad \text{for} \quad t \to \infty. \tag{7.10}$$

The general form of the inditial matrix may be put in the form

$$[H(t)] = [H(\infty)] - [C(t)] + [M_A]\delta(t), \tag{7.11}$$

where the function $[C(t)]$ is usually called the deficiency function, and its asymptotic behaviour is determined by (7.10). The constant matrix $[H(\infty)]$ determines the steady-state limit and may be calculated on the basis of (7.5). The apparent mass matrix $[M_A]$ is involved only for incompressible flow.

This paper is focused on the poles of the transfer functions and their influence on the aerodynamic forces. It was shown that the latent roots of (3.5) appear only when $M > 0$, and therefore, the following analysis will be restricted to the compressible flow when $[M_A] = 0$ and the relations (7.6) and (7.7) may be used.

It is convenient to make a decomposition

$$\left[\hat{A}e(p) \right] = \left[\hat{A}e_1(p) \right] + \left[\hat{A}e_2(p) \right], \tag{7.12}$$

where the first term represents the influence of poles, and on the basis of (6.4) it may be put in the form

$$\left[\hat{A}e(p) \right] = \sum_k \left(\frac{\{u_k\}\{v_k\}^T}{p - p_k} + \frac{\{u_k^*\}\{v_k^*\}^T}{p - p_k^*} \right), \tag{7.13}$$

where the summation concerns all the (calculated) roots, which exist always as conjugate pairs. It may be assumed, that the second term $[\hat{A}e_2(p)]$ does not possess any poles and represents the influence of the branch-points $p = 0$ and $p = \infty$.

Similar decomposition of the indicial matrix has the form

$$[H(t)] = [H_1(t)] + [H_2(t)], \tag{7.14}$$

where on the basis of (7.13) it follows that

$$[H_1(0)] = \lim_{t\to 0+}[H_1(t)] = \lim_{p\to\infty}\left[\hat{A}e_1(p)\right] = 0, \tag{7.15}$$

$$[H_1(\infty)] = \lim_{t\to\infty}[H_1(t)] = \lim_{p\to 0}\left[\hat{A}e_1(p)\right] = \left[\hat{A}e_1(0)\right], \tag{7.16}$$

and for $t \to \infty$

$$[H_1(t)] - [H_1(\infty)] = O(e^{\alpha t}) \qquad \text{where} \quad \alpha = \max_k(\mathrm{Re}\,(p_k)) < 0. \tag{7.17}$$

Hence, the asymptotic behaviour of $[H(t)]$ is determined by $[H_2(t)] - [H_2(\infty)] = O(t^{-2})$.

From (7.6) and (7.15) follows also the limiting value

$$\lim_{t\to 0+}[H_2(t)] = \lim_{t\to 0+}[H(t)] = [D]. \tag{7.18}$$

The deficiency function matrix may be also represented in the form of a sum of two components

$$[C_1(t)] = [H_1(\infty] - [H_1(t)] \tag{7.19}$$

and

$$[C_2(t)] = [H_2(\infty] - [H_2(t)]. \tag{7.20}$$

The first component is determined by (7.13), but for the second component, only the limiting value is known

$$[C_2(0)] = [H_2(\infty)] - [D] = [K(M,0)]^{-1} - \left[\hat{A}e_1(0)\right] - [D] \tag{7.21}$$

and the asymptotic behaviour

$$[C_2(t)] = O(t^{-2}) \qquad \text{for} \quad t \to \infty. \tag{7.22}$$

Finally, the problem of approximating the response matrix with the use of identified singularities is reduced to the determination of deficiency functions which fulfils the conditions (7.21) and (7.22).

For this purpose it is possible to use a method proposed by STARK [20] in a similar problem. If the deficiency matrix function can be approximated by a scalar function

$$[C_2(t)] = [C_2(0)] \cdot g(t), \qquad \text{where} \quad g(0) = 1, \quad \text{and} \quad g(t) = O(t^{-2}) \quad \text{for} \quad t \to \infty, \tag{7.23}$$

then

$$[H(t)] = [H_1(t)] + [H_2(\infty)] - [C_2(0)] \cdot g(t). \tag{7.24}$$

Taking the Laplace transform of (7.24) and multiplying the result by p we obtain

$$[K(M,p)]^{-1} = \left[\hat{A}e(p)\right] \approx \left[\hat{A}e_1(p)\right] - \left[\hat{A}e_1(0)\right] + [K(M,0)]^{-1} - [C_2(0)] \cdot p\hat{g}(p), \tag{7.25}$$

where the matrix $[C_2(0)]$ is given by (7.21), and $\hat{g}(p) = \mathcal{L}g(t)$.

STARK proposed [20, 28] some forms of the function $g(t)$. The best results were obtained with the set

$$g_m(t) = \left(\frac{a}{a+t}\right)^m \qquad (m = 1,2,3,\ldots), \tag{7.26}$$

where a is a positive real number which can be chosen in numerical experiments. Laplace transforms of functions (7.26) may be expressed by the exponential integral functions. The conditions (7.23) fulfil the function $g_2(t)$.

8. Conclusions

The numerically calculated aerodynamic forces in the frequency domain are always the values of analytic functions determined in the entire complex plane of the Laplace variable. These functions have poles in the left half of the complex plane, which determine the limits for the approximation by means of rational functions (with analytic continuation from the imaginary axis into the complex plane) and which may significantly influence the aerodynamic forces in the time domain.

1. In the case when the discretizing procedure of the lifting surface equation follows the lifting lines or doublet-lattice methods, the resulting algebraic equations are well-conditioned in the right half of the p-plane and in a strip parallel to the imaginary axis in the left half of the p-plane. The width of this strip decreases with increasing Mach number, but is wide enough for almost all applications. Only for high subsonic flow the problem of conditioning may be severe.

2. In the compressible case ($M \neq 0$), there exist a large (probably infinite) set of latent roots of the aerodynamic coefficients matrix in the left half of the p-plane which reflects (usually) in poles of the transfer functions. The distribution of these latent roots in the p-plane does not depend on the number of aerodynamic elements introduced in the discretization procedure (only small differences were observed which may be related to the accuracy of the results).

3. Also the pressure distributions which correspond to latent vectors of the aerodynamic influence coefficients matrix do not depend on the number of aerodynamic elements introduced in the discretization procedure.

4. For decreasing Mach number $M \to 0$, all latent roots move away from the origin to infinity and probably, for $M = 0$ there are no roots in the finite part of the p-plane. It seems to agree with the results of the STARK method [20] which takes into account only one singularity – the branch-point in the origin. The remarkable accuracy of this method in the incompressible case and less satisfactory results for $M > 0$ may be caused by the influence of the poles of transfer functions.

5. In subsonic flow for $M \to 1$, all latent roots tend to the origin ($p = 0$). The proximity of many poles may cause significant difficulties in the calculation of transfer functions in the range of high subsonic flow.

6. The decomposition of the deficiency function into a part which expresses the influence of latent roots (7.19) and a part influenced by the branch point (7.20) enables the extraction of the part which is responsible for the starting pulse. This agrees with the results of EDWARDS [15] who stated, that the step response function obtained by integrating along the branch cut does not contain the starting pulse.

7. The looping behaviour of some of the generalized forces for strongly decaying motion observed in [18] may be explained as the result of influence of poles of the transfer functions. It may be regarded as an indirect confirmation of the existence of latent roots in the kernel-function results.

The calculations and all considerations presented in this paper concern the aerodynamic model and the results are independent of the choice of generalized coordinates used to describe the motion of the structure.

The approximation to the aerodynamic transfer functions based on the identified singularities and the calculated left and right-hand latent vectors of the influence coefficients matrix avoids the ill-posed analytical continuation from the imaginary axis into the whole p-plane. It may be applied also in regions which contain poles of the transfer functions.

Appendix. Discretization of the piston theory in the lifting lines method

In the lifting lines method of discretization (similarly to many other methods), the pressure distribution on a profile (cross-section of the wing) is approximated

by means of a truncated series of functions with appropriate singularities on the leading and trailing edges. The pressure distribution in the piston theory follows the upwash distribution and is a regular, continuos function. Therefore it is not possible to cast the piston theory in the lifting lines discretization scheme exactly. Nevertheless, the approximation can assure the exact values of moments of aerodynamic forces in the case if the upwash distribution is a polynomial of degree less than the number of lifting lines on the cross-section.

The procedure of calculating the (approximate) pressure distribution on the profile $c_p(x)$ on the basis of a known upwash distribution $w(x)/U$ in the lifting lines method [4] consists of the following steps

$$\frac{w(x)}{U} \Rightarrow \{f\} \Rightarrow \{w\} \Rightarrow \{c_p\} \Rightarrow \{a\} \Rightarrow c_p(x). \tag{A.1}$$

The vectors $\{f\}$ and $\{w\}$, as well as $\{c_p\}$ and $\{a\}$ describe in the cross-section the approximate distributions of the upwash and pressure coefficient, respectively. The sizes of these vectors are equal to the number of lifting lines (denoted later by m). The vectors $\{w\}$ and $\{c_p\}$ for all cross-sections of the wing create the vectors in (3.2) and $N = \sum m$.

The pressure distribution on a cross-section is, in the lifting lines method, approximated by a truncated series of Jacobi polynomials

$$c_p(x) = \frac{1}{b_l}\sqrt{\frac{1-x}{1+x}} \sum_{k=0}^{m-1} a_k P_k(x), \tag{A.2}$$

where $2b_l$ is the local chord, the coordinate x is normalised to the interval $-1 < x < 1$ and $P_k(x)$ are polynomials which fulfil the orthogonality condition

$$\int_{-1}^{1} \sqrt{\frac{1-x}{1+x}}\, P_j(x) P_k(x)\, dx = \delta_{jk}\pi. \tag{A.3}$$

The vector $\{a\}$ of the coefficients a_k is determined for a given pressure distribution by the expression

$$\{a\} = \frac{b_l}{\pi} \int_{-1}^{1} \{P(x)\}\, c_p(x)\, dx. \tag{A.4}$$

The elements of the vector $\{P(x)\}$ are the polynomials $P_k(x)$. The quantities calculated in the lifting lines method from the set of algebraic equations are the strengths of lifting lines (pressure doublets). They are related to the a_k coefficients directly

$$\{c_p\} = [W][P]^T\{a\}, \tag{A.5}$$

where $[W]$ is a diagonal matrix with weight coefficients of the Gauss-Jacobi quadrature, and the elements of the matrix $[P]$ are the values of the polynomials $P_k(x)$ in the nodes of this quadrature.

The upwash distribution is approximated by means of the polynomials $Q_k(x) = P_k(-x)$

$$\frac{w(x)}{U} \approx \frac{1}{\pi} \sum_{k=0}^{n-1} f_k Q_k(x) = \frac{1}{\pi} \{Q(x)\}^T \{f\}, \tag{A.6}$$

where the coefficients f_k are determined by the expression

$$\{f\} = \int_{-1}^{1} \sqrt{\frac{1+x}{1-x}} \{Q(x)\} \frac{w(x)}{U} dx. \tag{A.7}$$

They are next transformed to the form

$$\{w\} = \frac{1}{\pi}[P]^T \{f\} = ([P][W])^{-1} \{f\}. \tag{A.8}$$

The discretized model of the piston theory may be constructed on the basis of the following scheme

$$\{w\} \Rightarrow \{f\} \Rightarrow \frac{w(x)}{U} \Rightarrow c_p(x) \Rightarrow \{a\} \Rightarrow \{c_p\}, \tag{A.9}$$

with the use of the relations (A.8), (A.6), (7.7), (A.4) and (A.5).

$$\{c_p\} = \frac{4b_1}{\pi^2 M}[W][P]^T \left(\int_{-1}^{1} \{P(x)\}\{Q(x)\}^T dx \right) [P][W]\{w\}, \tag{A.10}$$

where

$$\int_{-1}^{1} P_k(x) Q_n(x)\, dx = \begin{cases} \dfrac{1+(-1)^{k+n}}{k+n+1} - \dfrac{1-(-1)^{n-k}}{n-k} & \text{for} \quad n \neq k, \\ \dfrac{2}{2n+1} & \text{for} \quad n = k \end{cases} \tag{A.11}$$

The matrix $[D]$ defined in (7.6) has, in the case of lifting lines method, a quasi-diagonal form and each diagonal block corresponds to a cross-section of the wing and has the form determined by (A.10).

Acknowledgment

This investigation was supported by the State Committee for Scientific Research Poland, partially under Grant 3 P404 001 04, which is gratefully acknowledged.

References

1. H.G. KÜSSNER, *Allgemeine Tragflächentheorie*, Luftfahrtforschung, **17**, pp. 370–378, 1940.
2. E. ALBANO and W.P. RODDEN, *A doublet-lattice method for calculating lift distributions on oscillating surfaces in subsonic flows*, AIAA J., **7**, 2, pp. 279–285, 1969.
3. J. GRZĘDZIŃSKI, *Application of the unsteady lifting-lines method to arbitrary configurations of lifting surfaces*, Arch. Mech., **41**, pp. 165–180, 1989.
4. M. NOWAK, J. GRZĘDZIŃSKI, *The lifting-lines method – An efficient tool for calculating unsteady aerodynamic forces on lifting surfaces*, Proc. of the Internat. Forum on Aeroelasticity and Structural Dynamics, June 3-5, 1991, Aachen, Paper 91-092, DGLR-Bericht 91-96, pp. 295–304, Bonn 1991.
5. D.L. WOODCOCK, *Aerodynamic modelling for studies of aircraft dynamics*, RAE TR 81016, Feb. 1981.
6. A.H. ZEMANIAN, *Distribution theory and transform analysis*, McGraw Hill, New York 1965.
7. K.L. ROGER, *Aeroplane math modelling methods for active control design*, Structural Aspects of Active Control Design, AGARD-CP-228, pp. 4.1–4.11, Aug. 1977.
8. R. VEPA, *Finite state modelling of aeroelastic systems*, NASA CR-2779, 1977.
9. M. KARPEL, *Design for active and passive flutter suppression and gust alleviation*, NASA CR-3482, 1981.
10. A.I. VAN DE VOOREN, *Generalisation of the Theodorsen function to stable oscillations*, J. Aero. Sci., pp. 209–211, March 1952.
11. E.V. LAITONE, *Theodorsen's circulation function for generalized motion*, J. Aero. Sci, pp. 211–213, March 1952.
12. W.P. JONES, *The generalized Theodorsen function*, J. Aero. Sci., p. 213, March 1952.
13. C. CHANG, *On Theodorsen function in incompressible flow and C-function in supersonic flow*, J. Aero. Sci., pp. 717–718, October 1952.
14. R.D. MILNE, *Asymptotic solutions of linear stationary integro-differential equations*, ARC R& M, no. 3548, London 1968.
15. J.W. EDWARDS, *Applications of Laplace transform methods to airfoil motion and stability calculations*, AIAA Paper 79-0772, April 1979.
16. R. VEPA, *Comment on "Active flutter control using generalized unsteady aerodynamic theory"*, J. Guidance and Control, **2**, 5, pp. 446–447, Sept.-Oct. 1979.
17. R. VEPA and A. SAEED, *Practical techniques of modelling aeroelastic systems for active control applications*, Proc. of the European Forum on Aeroelasticity and Structural Dynamics, Paper no. 89-058, pp. 529–537, 1989.
18. H.J. CUNNINGHAM and R.N. DESMARAIS, *Generalisation of the subsonic kernel function in the s-plane, with applications to flutter analysis*, NASA TP 2292, March, 1984.
19. T. UEDA, *Lifting surface calculations in the Laplace domain with application to root loci*, AIAA J., **25**, 5, pp. 698–704, May 1987.
20. V.J.E. STARK, *General equations of motion for an elastic wing and method of solution*, AIAA J., **22**, 8, pp. 1146–1152, August 1984.
21. S.M. BIELOTSERKOVSKII and I.K. LIFANOV, *Numerical methods in singular integral equations* [in Russian], Nauka, Moskva 1985.
22. T. UEDA, *Integral equation of lifting surfaces in Laplace domain and analytic continuation of its pressure kernel*, National Aerospace Laboratory, Chofu, Japan, Rept. TR-795T, Jan. 1984.
23. Y.L. LUKE, *The special functions and their approximations*, vol. II, Academic Press, New York and London 1969.
24. C.L. LAWSON and R.J. HANSON, *Solving least squares problems*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey 1974.
25. J. STOER and R. BURLISCH, *Introduction to numerical analysis*, Springer-Verlag, New York, Heidelberg, Berlin 1980.

26. P. LANCASTER, *Lambda - matrices and vibrating systems*, Pergamon Press, Oxford, London, Edinburgh, New York, Toronto, Paris, Braunschweig 1966.

27. H. ASHLEY and G. ZARTARIAN, *Piston theory – A new aerodynamic tool for the aeroelastician*, J.A.S., **23**, 12, pp. 1109–1118, Dec. 1956.

28. V.J.E. STARK, *Flutter calculation by a new program*, Collected papers of the Second International Symposium on Aeroelasticity and Structural Dynamics held in Aachen, DGLR-Bericht 85-02, pp. 276–285, April 1-3, 1985.

POLISH ACADEMY OF SCIENCES
INSTITUTE OF FUNDAMENTAL TECHNOLOGICAL RESEARCH

Received February 16, 1996.

Arch. Mech., 48, 5, pp. 869–892, Warszawa 1996

Travelling waves in laser sustained plasma Constant coefficient case

B. KAŹMIERCZAK (WARSZAWA)

WE USE the Conley index theory to prove existence of travelling waves to a system of partial differential equations describing a two-temperature model of plasma sustained by a laser beam. These waves connect two asymptotic state of gas: a cold one and a hot partially ionized one.

1. Introduction

THE AIM OF THIS PAPER is to prove the existence of travelling wave solution to the equations of a two-temperature model describing the laser-sustained plasma (see system (0)). The problem was positively solved by means of the implicit function theorem in [5] under the condition of sufficiently large values of the coupling parameter. This time we use the technique of Conley connection index theory (see [1, 2, 3, 4]). It seems interesting to compare these two methods. For simplicity, we consider the case of constant transport coefficients. The case of variable transport coefficients will be considered in the subsequent paper.

The evolution of temperatures T_1 and T_2 of electrons and heavy particles (i.e.atoms and ions) in plasma sustained by a laser beam under a constant pressure p are described by the following equations (see [5] and references therein):

$$
\begin{aligned}
\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\mathbf{grad}\right) c_1 &= \operatorname{div}(k_1 \,\mathbf{grad}\, T_1) + F_1 - (T_1 - T_2)\mathcal{W},\\
\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\mathbf{grad}\right) c_2 &= \operatorname{div}(k_2 \,\mathbf{grad}\, T_2) + F_2 - (T_1 - T_2)\mathcal{W}.
\end{aligned} \tag{0}
$$

Here k_i are effective heat conductivity coefficients, c_i their effective heat capacities per unit volume. F_1, F_2 are nonlinear source functions. The term $(T_1 - T_2)\mathcal{W}$ describes collisional energy exchange betweeen electrons and heavy particles. $\mathcal{W}$ is proportional to the frequency of electron-heavy particle collisions. This frequency tends to infinity as $p \to \infty$. So, we can write $\mathcal{W}(p; T_1, T_2) = \lambda(p) W(T_1, T_2)$, where λ is a real parameter, $\lambda(p) \to \infty$ as $p \to \infty$. The functions $k_i > 0$, $c_i > 0$, F_i, $i \in \{1, 2\}$, and $W > 0$ depend in general on T_1 and T_2. However, for simplicity of presentation we will assume that k_i, c_i, and W are constant. The dependence on T_1 and T_2 will be retained only in nonlinear source terms F_1 and F_2.

By looking for solutions in the form of a travelling wave, that is by making a substitution:

$$
T_1(\mathbf{x}, t) = u_1(\mathbf{x}\cdot\mathbf{n} + \chi t), \qquad T_2(\mathbf{x}, t) = u_2(\mathbf{x}\cdot\mathbf{n} + \chi t),
$$

where $n \in \mathbb{R}^3$ can be interpreted as the direction of propagation and χ as a speed of the wave, we are led to a system of ordinary differential equations of the form:

$$k_1 u_1'' - c_1\theta u_1' + F_1(u) + \lambda W(u_1 - u_2) = 0,$$
$$k_2 u_2'' - c_2\theta u_2' + F_2(u) + \lambda W(u_1 - u_2) = 0,$$

where $' := d/d\xi$, $\xi := \mathbf{x}\cdot\mathbf{n} + \chi t$, $\theta := (\chi + \mathbf{v}\cdot\mathbf{n})$ and $u := (u_1, u_2)$. It is obvious that by changing the scale of the independent variable and redefining the constants c_1, c_2 and λ we may obtain a simpler (but less symmetric) form of this system:

$$\begin{aligned} u_1'' - c_1\theta u_1' + F_1(u) - \lambda(u_1 - u_2) &= 0, \\ k u_2'' - c_2\theta u_2' + F_2(u) + \lambda(u_1 - u_2) &= 0. \end{aligned} \tag{1.1}$$

The roots of the corresponding algebraic system

$$F_1(u) - \lambda(u_1 - u_2) = 0,$$
$$F_2(u) + \lambda(u_1 - u_2) = 0,$$

are called constant states for (1.1). So, we are interested in *solutions defined for all* $\xi \in \mathbb{R}^1$ *whose derivatives vanish at* $\pm\infty$ *and such that* $(u_1(\xi), u_2(\xi))$ *tends to different constant states as* $\xi \pm \infty$. Such solutions are called *heteroclinics*. For a given λ such solutions can exist only for certain values of the parameter θ. (The problem considered is a sort of a nonlinear eigenvalue problem). Thus it makes sense to speak of heteroclinic triples (θ, u_1, u_2) satisfying Eqs. (1.1). Our aim is to prove existence of a heteroclinic connecting appropriate constant states of Eqs. (1.1). These constant states can be interpreted as the two states of gas: the cold incoming one (at $-\infty$) and the partially ionized hot one (at ∞). *The existence theorem is stated in Theorem at the end of Sec. 7.*

To analyze heteroclinic connections for Eqs. (1.1) we will consider the following family of systems:

$$\begin{aligned} u_1'' - c_1\theta u_1' + \mathcal{F}_{1\eta} - \lambda(u_1 - u_2) &= 0, \\ k u_2'' - c_{2\eta}\theta u_2' + \mathcal{F}_{2\eta} + \lambda(u_1 - u_2) &= 0, \end{aligned} \tag{1.η}$$

where $\eta \in [0, 1]$ and

$$c_{2\eta} = c_1 k(1 - \eta) + \eta c_2,$$
$$\mathcal{F}_{1\eta} := F_s + \eta(F_1 - F_s), \qquad \mathcal{F}_{2\eta} := F_s + \eta(F_2 - F_s),$$
$$F_s(u_1, u_2) := (F_1 + F_2)\left((1 + k)^{-1}(u_1 + k u_2),\ (1 + k)^{-1}(u_1 + k u_2)\right).$$

When we denote $w := (1 + k)^{-1}(u_1 + k u_2)$, $d := u_1 - u_2$, add and subtract the both sides of Eqs. (1.η), we obtain the system:

$$\begin{aligned} &w'' - c_1\theta w' + 2(1 + k)^{-1} F_s(w, w) \\ &\qquad + \eta(1 + k)^{-1}\left(F_1 + F_2 - 2F_s - (c_2 - c_1 k)\theta u_2'\right) = 0, \\ &d'' - c_1\theta d' - (1 + k^{-1})\lambda d + \eta k^{-1}\left(kF_1 - F_2 + (c_2 - c_1 k)\theta u_2'\right) = 0. \end{aligned} \tag{2}$$

Sections 2–5 have a preparatory character. Section 2 contains the assumptions imposed on the coefficients of system (1.1). In Sec. 3 we examine properties of constant states of system (1.η), and especially their behaviour for large λ. In Sec. 4 we prove *a priori* estimates for first derivatives of the solutions to (1.η) provided they are contained in a certain bounded region of (u_1, u_2)-space and prove that the set of θ, for which a heteroclinic orbit can exist is comprised in some bounded open interval (θ_0, θ_1), where $0 < \theta_0 < \theta_1$. Such estimations are necessary, because we want an isolating neighbourhood to be a compact subset of the phase space. In Sec. 5 we examine the eigenvalues and eigenvectors of the system linearized at its singular points. In Sec. 6 we construct an η-family of compact subsets of the phase space such that:

1) they are continuously varying with η,

2) each of them is an isolating neighbourhood with respect to the flow generated by Eqs. (1.η).

For $\eta = 0$ the system (1.η) has almost a "classical" structure and is relatively easy to analyze. Then, using the invariance of the connection index under continuation relation we can analyze existence of heteroclinics for the system (1.1). We did it in Sec. 7. For reader's convenience we have collected the necessary statements of the connection index theory taken from [1] in the Appendix A.

2. Assumptions

ASSUMPTION 1. All the considered functions are of C^2 class. □

ASSUMPTION 2. The constants k, c_1 and c_2 are positive. □

ASSUMPTION 3. In the interval $[-2\tau, 1 + 2\tau]$, $\tau > 0$, the equation

$$F(y, y) := F_1(y, y) + F_2(y, y) = 0 \tag{3}$$

has exactly three solutions 0, 1 and $y_0 \in (0, 1)$ such that $F_{,y}(0, 0) < 0$, $F_{,y}(1, 1) < 0$ and $F_{,y}(y_0, y_0) > 0$. □

ASSUMPTION 4. $\int_0^1 F(y, y)\, dy := I > 0$. □

3. Constant states during continuation

First of all, we will examine the behaviour of constant states for (1.η) i.e. solutions to the algebraic system:

$$\begin{aligned} \mathcal{F}_{1\eta}(u_1, u_2) - \lambda(u_1 - u_2) &= 0, \\ \mathcal{F}_{2\eta}(u_1, u_2) + \lambda(u_1 - u_2) &= 0, \end{aligned} \tag{4}$$

while the parameter η changes in the interval $[0, 1]$.

REMARK. Let us note that the functions $\mathcal{F}_{1\eta}$, $\mathcal{F}_{2\eta}$ satisfy the relation

$$(\mathcal{F}_{1\eta} + \mathcal{F}_{2\eta})(y, y) = F(y, y)\gamma(\eta),$$

where $\gamma(\eta) = 2 - \eta$. □

To begin with, we will characterize the properties of the solutions to the system (4) with respect to the solutions of Eq. (3). First, it is easy to note, by means of the implicit function theorem, that for sufficiently large λ and $u_1 \in [-2\tau, 1 + 2\tau]$ the equations $\mathcal{F}_{1\eta}(u_1, u_2) - \lambda(u_1 - u_2) = 0$ and $\mathcal{F}_{2\eta}(u_1, u_2) + \lambda(u_1 - u_2) = 0$ are uniquely solvable with respect to u_2. The solutions to these equations will be denoted below respectively by $u_2 = \mu_\eta(u_1)$ and $u_2 = \vartheta_\eta(u_1)$.

Below, $\mathcal{F}_{i\eta,j}$ will denote the partial derivative of $\mathcal{F}_{i\eta}$ with respect to u_j.

LEMMA 1.

a. For all $\eta \in [0, 1]$ and sufficiently large $\lambda > 0$, the system (4) has exactly three solution pairs $(u_1, u_2)(\lambda, \eta)$ such that both u_1 and u_2 belong to the interval $[-\tau, 1 + \tau]$ and such that for $\lambda \to \infty$ they tend to appropriate solutions of Eq. (3).

b. Let V belong to the set $\{0, y_0, 1\}$ of solutions to Eq. (3) and let $F_{,y}(V) = [(F_1 + F_2)(y, y)]_{,y|y=V} > 0$ (< 0) in some open neighbourhood of V in R^1.

Let $(u_1, u_2)(\lambda, \eta)$ be this branch of solutions to (4) which tends to (V, V) as $\lambda \to \infty$. Then, in some open (in R^2) neighbourhood of this solution we have

$$[(\mathcal{F}_{1\eta} + \mathcal{F}_{2\eta})(u_1, u_2)]_{,1} + [(\mathcal{F}_{1\eta} + \mathcal{F}_{2\eta})(u_1, u_2)]_{,2} > 0 \quad (< 0).$$

Below $U_\nu(\lambda, \eta) := (U_{\nu 1}(\lambda, \eta), U_{\nu 2}(\lambda, \eta))$, $\nu \in \{-, 0, +\}$, will denote the solution branch such that $U_-(\lambda, \eta) \to (0, 0)$, $U_0(\lambda, \eta)(y_0, y_0)$ and $U_+(\lambda, \eta) \to (1, 1)$ for $\lambda \to \infty$.

c. For all sufficiently large λ and all $u_1 \in [U_{-1}(\lambda, \eta), U_{+1}(\lambda, \eta)]$, we have $\mu'_\eta(u_1) > 0$, $\vartheta'_\eta(u_1) > 0$. Moreover in all sufficiently small neighbourhoods of the points $U_-(\lambda, \eta)$ and $U_+(\lambda, \eta)$ we have the inequalities, $\mu'_\eta > \vartheta'_\eta$. □

P r o o f. Adding and subtracting both sides of Eqs. (4) we obtain:

$$\begin{aligned} \mathcal{F}_{1\eta} + \mathcal{F}_{2\eta} &= 0, \\ \ell(\mathcal{F}_{1\eta} - \mathcal{F}_{2\eta}) - 2(u_1 - u_2) &= 0, \end{aligned} \tag{4'}$$

where $\ell = \lambda^{-1}$. If $\ell = 0$ and u_1 and u_2 satisfying the second equation are bounded, then they must satisfy the equality $u_2 = u_1$. Putting it into the first equation we infer, according to the definition of $\mathcal{F}_{i\eta}$, that it is equivalent to the equation $F(u_1, u_1) = 0$ i.e. to Eq. (3). Thus, for $\ell = 0$, in the rectangle $[-2\tau, 1 + 2\tau] \times [-2\tau, 1 + 2\tau]$ there are exactly three solutions equal to (V, V), $V \in \{0, y_0, 1\}$. The determinant of the Jacobian of the mapping $(R^2 \to R^2)$ determined by the left-hand sides of (4') for $\ell = 0$ is equal to $2\gamma(\eta)F_{,y}(y, y)|_{y=V}$, $V \in \{0, y_0, 1\}$. So, according to Assumption 3 it is nonzero. Hence point a follows

from the implicit function theorem. The proof of point b follows immediately from the continuity of the partial derivatives of $\mathcal{F}_{i\eta}$.

Now, differentiating both sides of Eqs. (4) with respect to u_1 we obtain the equalities:

$$\mu'_\eta = (-\mathcal{F}_{1\eta,1} + \lambda)(\mathcal{F}_{1\eta,2} + \lambda)^{-1}, \qquad \vartheta'_\eta = (\mathcal{F}_{2\eta,1} + \lambda)(-\mathcal{F}_{2\eta,2} + \lambda)^{-1}.$$

All the terms in the expression for μ'_η are taken at a point $(u_1, \mu_\eta(u_1))$, and in the expression for ϑ'_η at a point $(u_1, \vartheta_\eta(u_1))$. Suppose that $\vartheta'_\eta > \mu'_\eta$ in some neighbourhood of $U_-(\lambda, \eta)$ or $U_+(\lambda, \eta)$. Then, for λ sufficiently large, all brackets in the expressions above are positive and we would have

$$[-\mathcal{F}_{1\eta,1} + \lambda](u_1, \mu_\eta(u_1))[\mathcal{F}_{2\eta,2} + \lambda](u_1, \vartheta_\eta(u_1)) - [\mathcal{F}_{2\eta,1} + \lambda](u_1, \vartheta_\eta(u_1))[\mathcal{F}_{1\eta,2} + \lambda](u_1, \mu_\eta(u_1)) < 0.$$

Sufficiently close to $U_\nu(\lambda, \eta)$ ($\nu = -$ or $\nu = +$) the difference $(\vartheta_\eta(u_1) - \mu_\eta(u_1))$ can be made arbitrarily small. Consequently, for sufficiently large λ, this would imply the inequality

$$-\lambda[(\mathcal{F}_{1\eta} + \mathcal{F}_{2\eta})_{,1} + (\mathcal{F}_{1\eta} + \mathcal{F}_{2\eta})_{,2}](U_\nu(\lambda, \eta)) < 0.$$

However, according to Assumption 3 and point b of this lemma, this would imply, that the left-hand side of the last inequality would be positive, which could not be true. This proves point c. □

LEMMA 2. Let the assumptions of Lemma 1 be fulfilled. Then for λ sufficiently large determinant of the matrix

$$\mathcal{M}(\lambda, \eta, u) := \begin{bmatrix} (\mathcal{F}_{1\eta,1}(u) - \lambda)k & (\mathcal{F}_{1\eta,2}(u) + \lambda)k \\ (\mathcal{F}_{2\eta,1}(u) + \lambda) & (\mathcal{F}_{2\eta,2}(u) - \lambda) \end{bmatrix} k^{-1} \tag{5}$$

has the sign opposite to the sign of the expression

$$[(\mathcal{F}_{1\eta} + \mathcal{F}_{2\eta})(u_1, u_2)]_{,1} + [(\mathcal{F}_{1\eta} + \mathcal{F}_{2\eta})(u_1, u_2)]_{,2}.$$

P r o o f. The determinant of $\mathcal{M}$ is equal to

$$(\mathcal{F}_{1\eta,1}F_{2\eta,2} - \mathcal{F}_{1\eta,2}F_{2\eta,1}) - \lambda(\mathcal{F}_{1\eta,1} + \mathcal{F}_{2\eta,2} + \mathcal{F}_{1\eta,2} + \mathcal{F}_{2\eta,1}).$$

Thus, for λ sufficiently large, we obtain the claim of this lemma. □

4. *A priori* estimates

Global properties of heteroclinic solutions.

According to Lemma 1, especially to the proof of point a of that lemma, for sufficiently large λ, the solutions $U_\nu(\lambda,\eta)$, $\nu \in \{-,0,+\}$ are isolated. To be more precise, there exists a number $\tau^* > 0$ such that in the rectangle

$$R_\eta := \{(u_1,u_2):\ u_i \in [U_{-i}(\lambda,\eta) - \tau^*,\ U_{+i}(\lambda,\eta) + \tau^*], \quad i = 1,2\}$$

there are no other solutions to the system (4).

LEMMA 3. For suffciently large λ there exists a constant $L^* < \infty$ independent of the values of θ, $\eta \in [0,1]$ and λ, such that for all bounded solutions to $(1.\eta)$, for which u_1 and u_2 stay in the rectangle R_η for all ξ we have the estimate:

$$\|u_1 - u_2\|_{C^0} < (\lambda)^{-1} L^*. \quad \square$$

P r o o f. First, suppose that the function $d(\xi) = u_1(\xi) - u_2(\xi)$ achieves a positive maximum (negative minimum) for some $\xi = \zeta \in (-\infty,\infty)$ and that $(u_1(\zeta), u_2(\zeta))$ lies in R_η. Then at this point $d' = 0$ and $d'' \le 0$ (≥ 0). Hence, due to (2),

$$|d| \le \eta k^{-1}(|kF_1 - F_2| + |(c_2 - c_1 k)\theta u_2'|)(1 + k^{-1})^{-1}\lambda^{-1}. \tag{6}$$

As we assume that the solution is bounded (for all times) and $u_2'(\xi) = 0$ for $|\xi| = \infty$, then $\theta u_2'$ must attain the global maximum somewhere. As $d' = 0$ implies $u_1' = u_2' = w'$ at the point of extremum, then by means of the first equation in (2), we can find an upper bound for $|(c_2 - c_1 k)\theta u_2'|$. It is not greater than

$$|(c_2 - c_1 k)((2 - 2\eta)F_s + \eta F_1 + \eta F_2)(c_1 + c_2\eta + c_1 k - c_1\eta k)^{-1}|.$$

After some computations one can prove that the right-hand side of (6) is not greater than $\max\{\eta, (1 + c_1 c_2^{-1}\eta^{-1})^{-1}\} \max_u(2|F_1(u)| + 2|F_2(u)| + 2|F_s(u)|)\lambda^{-1}$, where the maximum is taken over R_η. This expression has a common bound independent of η. $\square$

By means of Lemma 3 we can prove:

LEMMA 4. For sufficiently large λ there exists a number $L < \infty$ independent of $\eta \in [0,1]$, $\theta \in (-\infty,\infty)$ and λ, such that for all solutions to $(1.\eta)$ which (for all ξ) stay in the rectangle R_η the estimates $|u_1'| < L$, $|u_2'| < L$ hold. $\square$

P r o o f. Let us consider an arbitrary solution satisfying the above conditions. Then there is M such that $|\mathcal{F}_{i\eta} + (-1)^i\lambda(u_1 - u_2)| < M$ for $i = 1,2$ and $(u_1,u_2) \in R_\eta$. Let us note, that due to Lemma 3, $|\lambda(u_1 - u_2)| < L^*$, so M *can also be treated as independent of* λ. First, let us examine the case: $(c_1\theta) \ge 1$. Suppose

that for some solution $|u_1'|$ attains a value larger than M. Then u_1' and u_1'' have the same sign and this property is retained for all positive times. Consequently, this solution will grow exponentially, contrary to the boundedness of the solution. Hence, $|u_1'| \leq M$. If $c_1\theta \leq -1$, then changing the direction of "time" we arrive at the equation of the form: $u_1'' = c_1\theta u_1' - [\mathcal{F}_{1\eta} - \lambda(u_1 - u_2)]$, thus for all positive times u_1' and u_1'' have the same sign (as before) and the solution will grow exponentially. So, $|u_1'| \leq M$. In the same way we can prove that $|u_2'| \leq M$, if $|c_{2\eta}\theta| \geq 1$. Now, let us analyze the case $|c_1\theta| \leq 1$. Then for sufficiently large $|u_1'|$, $|u_1''| < |u_1'| + M$ (Remember that $|\mathcal{F}_{i\eta} + (-1)^i\lambda(u_1 - u_2)| < M$). Thus, if $u_1'(0) = L_1 > 0$, then for $\xi > 0$, we would have $u_1'(\xi) > \exp(-\xi)[L_1 - M(\exp(\xi) - 1)]$ independently of the sign of $(c_1\theta)$. Integrating this inequality with respect to ξ over the interval $(0, 1)$ we obtain that $|u_1(1) - u_1(0)| > L_1(1 - e^{-1}) - M$. For L_1 sufficiently large, the right-hand side of this inequality is strictly larger than $(U_{+1}(\lambda, \eta) - U_{-1}(\lambda, \eta) + 2\tau^*)$, which is impossible due to the fact that the solution must lie in R_η. If $L_1 < 0$, the proof is carried out in the same way. Likewise the inequality $|c_{2\eta}\theta| \leq 1$ implies the inequality $|u_2'| < L_2$. Consequently $|u_1'| < L$, $|u_2'| < L$, where $L = \max\{M, L_1, L_2\}$. □

Let $\Gamma = (c_2k^{-1} - c_1)$, $\Gamma_\eta = \eta|\Gamma|$. If $\Gamma \geq 0$, let $m = 2$, $\chi(\eta) = c_1$ and $s = k(1 + k)^{-1}$. If $\Gamma < 0$, then let $m = 1$, $\chi(\eta) = c_{2\eta}k^{-1}$ and $s = (1 + k)^{-1}$. Now, the first equation in (2) can be written as:

$$w'' - \chi_\eta\theta w' - \theta\Gamma_\eta s u_m' + (1 + k)^{-1}\{\mathcal{F}_{1\eta} + \mathcal{F}_{2\eta}\} = 0. \tag{7}$$

The next lemma estimates the "possible" values of θ.

LEMMA 5. There exists $\lambda_0 \in (0, \infty)$ such that for all $\lambda > \lambda_0$, $\eta \in [0, 1]$ the value of θ, for which a heteroclinic solution (with nonnegative derivatives) to system $(1.\eta)$ connecting the points $U_-(\lambda, \eta)$ and $U_+(\lambda, \eta)$ can exist, is positive and bounded uniformly from above and below i.e $\theta \in (\theta_0, \theta_1)$ with $0 < \theta_0 < \theta_1$. □

P r o o f. Suppose that, for some $\theta = \theta(\eta)$, $(u_1(\xi), u_2(\xi))$ satisfies $(1.\eta)$. Then there is no open interval (comprised in $(-\infty, \infty)$) such for $i = 1$ or $i = 2$, $u_i'(\xi) \equiv 0$ for ξ from this interval. For, then $u_i'' \equiv 0$, $u_i = \text{const}$ and $\mathcal{F}_{i\eta} - \lambda(u_1 - u_2)(-1)^{i-1}$ would be equal to 0 in this interval. Due to Lemma 1.c (for λ sufficiently large) the slope of the curve $(\mathcal{F}_{i\eta} - \lambda(u_1 - u_2)(-1)^{i-1}) = 0$ is positive and finite, so this would imply that $u_j = \text{const}$ also, where j is the index complementary to i. Consequently this would be a singular point. But this cannot happen for $|\xi| < \infty$.

Multiplying Eq. (7) by w' and integrating with respect to ξ from $(-\infty)$ to (∞) we obtain:

$$(\chi_\eta + \Gamma_\eta)\theta(\eta) \int_{-\infty}^{\infty} w'^2(\xi)\, d\xi - (1 + k)^{-1} \int_{-\infty}^{\infty} (\mathcal{F}_{1\eta} + \mathcal{F}_{2\eta})(u_1' d\xi + k u_2' d\xi) \geq 0.$$

We claim that for λ sufficiently large, the second term at the left-hand side is, independently of $\eta \in [0, 1]$, positive, say, larger than $8^{-1}I$ (Assumption 4). To

prove this, let us consider for example the integral $\int_{-\infty}^{\infty} (\mathcal{F}_{1\eta} + \mathcal{F}_{2\eta})u_1' d\xi$. It is equal to $\int \{\mathcal{F}_{1\eta}(u_1, u_2(u_1)) + \mathcal{F}_{2\eta}(u_1, u_2(u_1))\} du_1$, where $u_2(u_1) = u_2(\xi(u_1))$, $\xi(u_1)$ is the inverse of the function $u_1(\xi)$ and the integration is made over the interval $[U_{-1}(\lambda, \eta), U_{+1}(\lambda, \eta)]$. Now, using point a of Lemma 1 (and its proof) and Lemma 3 we infer, by means of Taylor expansion with respect to $(u_2(u_1) - u_1)$, that this integral is larger than $(\gamma(\eta) \int_0^1 F(u_1)\, du_1 - C\lambda^{-1})$, where C is a constant independent of λ and η. The second integral can be estimated in the same way, so, we infer that the claim is true. The integral $\int_{-\infty}^{\infty} w'^2(\xi)\, d\xi$ can be written as the integral $\int w'(w)\, dw$ over the interval, which is bounded for every η. As, due to Lemma 4, $|w'(\xi)| < L$ independently of $\xi \in (-\infty, \infty)$ and $\eta \in [0, 1]$, and $(\chi_\eta + \Gamma_\eta) > 0$, then (for all $\eta \in [0, 1]$) we infer that $\theta(\eta) > 0$. Hence $\inf_\eta(\theta(\eta)) > \theta_0 > 0$.

To find an upper bound for $\theta(\eta)$ let us integrate Eq. (7) from $(-\infty)$ to ξ using the fact that $w' \geq 0$. We obtain:

$$2^{-1}w'^2(\xi) \geq \chi\theta(\eta) \int_{-\infty}^{\xi} w'^2(\xi)\, d\xi - (1 + k)^{-1} \int_{-\infty}^{\xi} (\mathcal{F}_{1\eta} + \mathcal{F}_{2\eta})(u_1'\, d\xi + k u_2'\, d\xi).$$

Now, as before, one can easily prove that for λ sufficiently large and all $\eta \in [0, 1]$ there exists $\zeta \in (-\infty, \infty)$ such that for $\xi = \zeta$ the last term of the above inequality (respecting the sign) is positive, say, larger than $(-8^{-1}J)$, where $J = \min_y \int_0^y F(s)\, ds$, where minimum is taken over the interval $[0, 1]$. J is negative due to Assumption 3. Consequently, there is a point on the trajectory, where $2^{-1}w'^2 \geq -8^{-1}J$. Thus, at the point of maximum of w' it follows from (7) that $\theta(\eta) \leq \sup(\mathcal{F}_{1\eta}(u_1, u_2) + \mathcal{F}_{2\eta}(u_1, u_2))2[\chi_\eta(k + 1)\sqrt{J}]^{-1}$, where the supremum is taken over $u \in R_\eta$. The right-hand side of this inequality is bounded from above by a number independent of η, let us denote it by θ_1. Thus, we obtain the claim of the lemma. □

5. Eigenvalues of the linearized system

Below z_1 and z_2 will be variables standing for u_1' and u_2' and $z := (z_1, z_2)$. Equations (1.η) may be written as the first order system:

$$(u_1', u_2', z_1', z_2') = \big(z_1, z_2, c_1\theta z_1 - \mathcal{F}_{1\eta} + \lambda(u_1 - u_2), \\ k^{-1}\{c_{2\eta}\theta z_2 - \mathcal{F}_{2\eta} - \lambda(u_1 - u_2)\}\big).$$

It is seen that the zeros of the right-hand side have the following form:

$$(u_1, u_2, z_1, z_2) = (U(\lambda, \eta), 0, 0),$$

where $U(\lambda, \eta)$ is a solution to system (4). Thus, for (u_1, u_2) in the rectangle $[-\tau, 1+\tau] \times [-\tau, 1+\tau]$ and sufficiently large λ we have exactly three zeros: $(U_\nu(\lambda, \eta), 0, 0)$, $\nu \in \{-, 0, +\}$. Linearizing $(1.\eta)$ around $(U_\nu(\lambda, \eta), 0, 0)$, $\nu \in \{-, 0, +\}$, we obtain the system:

$$(8) \qquad \begin{bmatrix} u_1' \\ u_2' \\ z_1' \\ z_2' \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ & & c_1\theta & 0 \\ -\mathcal{M}_\nu(\lambda, \eta) & & 0 & c_{2\eta}\theta k^{-1} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ z_1 \\ z_2 \end{bmatrix} := M_\nu(\lambda, \eta) \begin{bmatrix} u_1 \\ u_2 \\ z_1 \\ z_2 \end{bmatrix},$$

where $\mathcal{M}_\nu(\lambda, \eta) := \mathcal{M}(\lambda, \eta, U_\nu(\lambda, \eta))$ and $\mathcal{M}(\lambda, \eta, u)$ is defined in Lemma 2 by (5).

LEMMA 6. Let Assumptions 2 and 3 be fulfilled. Then, for all sufficiently large λ, $\theta \in (0, \infty)$ and all $\eta \in [0, 1]$, the matrix $M_\nu(\lambda, \eta)$, $\nu \in \{-, +\}$, has four real eigenvalues. Two of them are positive and two of them are negative. □

The proof of this lemma will be sketched in Appendix B. It is easy to note that the eigenvector corresponding to the eigenvalue q of the matrix $M_\nu(\lambda, \eta)$, $\nu \in \{-, 0, +\}$ has the form $(\xi_1, \xi_2, \zeta_1, \zeta_2)$, where $\zeta_i = q\xi_i$, $i = 1, 2$. (See [1] p. 335.)

LEMMA 7. Let Assumptions 2 and 3 be fulfilled. Then, for all sufficiently large λ, $\theta \in (0, \infty)$ and all $\eta \in [0, 1]$, the matrix $M_0(\lambda, \eta)$ has one negative, one positive eigenvalue and two complex conjugate eigenvalues with positive real parts. The components ξ_1, ξ_2 of the eigenvector corresponding to the negative eigenvalue satisfy the condition $\xi_2 \xi_1^{-1} < 0$. □

The proof of this Lemma 7 will be given in Appendix B.

6. Isolating neighbourhood during continuation

In this section we construct an η-family of compact subsets $N_3(\eta)$, such that for each $\eta \in [0, 1] N_3(\eta)$ is an isolating neighbourhood for the flow generated by the first order system corresponding to $(1.\eta)$. Every $N_3(\eta)$ consists of a parallelepiped

$$N_1 = \left\{(u_1, u_2, z_1, z_2) : U_{-i}(\lambda, \eta) \le u_i \le U_{+i}(\lambda, \eta),\ 0 \le z_i \le L,\ i = 1, 2\right\},$$

plus "small" neighbourhoods of the singular points, which we want to connect $(N(\eta))$, minus a small neighbourhood of the remaining singular point $N_3(0, \kappa, \eta)$. This point can be excised according to Lemma 7 and the Lemma in 4.D in [1].

Let us denote:

$$\mathcal{F}_{1\eta}(u_1, u_2) - \lambda(u_1 - u_2) := \mathcal{H}_1(\lambda, \eta, u_1, u_2),$$
$$\mathcal{F}_{2\eta}(u_1, u_2) + \lambda(u_1 - u_2) := \mathcal{H}_2(\lambda, \eta, u_1, u_2).$$

As the proofs carried out below are the same for all $\eta \in [0,1]$ and all sufficiently large λ, then, *to simplify notation, when there will be no danger of confusion, we will write* $\mathcal{H}_i(\xi)$ *instead of* $\mathcal{H}_i(\lambda, \eta, u_1(\xi), u_2(\xi))$.

According to Lemma 1 for fixed η and λ (sufficiently large), the zero sets of $\mathcal{H}_1$ and $\mathcal{H}_2$ near the points $U_-(\lambda,\eta) := (U_{-1}, U_{-2})(\lambda,\eta)$ and $U_+(\lambda,\eta) := (U_{+1}, U_{+2})(\lambda,\eta)$ have the graph like that in Fig. 1.

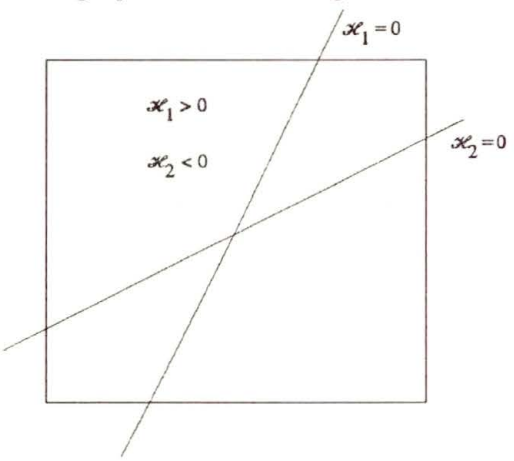

FIG. 1.

LEMMA 8. There exist smooth functions $\varepsilon_{2\nu}(\eta)$, such that for $\eta \in [0,1]$, $\nu \in \{-,+\}$, $\delta \in (0,1]$ and all sufficiently small $\Delta_1 > 0$, the set

$$N(\delta,\eta,\nu,\Delta_1) := \Big\{(u,z) : |u_1 - U_{\nu 1}(\lambda,\eta)| \le \delta\Delta_1,\ |u_2 - U_{\nu 2}(\lambda,\eta)| \le \delta\Delta_{2\nu}(\eta),$$
$$|z_i| < L,\ i = 1,2\Big\},$$

where $\Delta_{2\nu}(\eta) = \Delta_1\varepsilon_{2\nu}(\eta)$ is an isolating neighbourhood. □

By means of Lemma 9 it may be proved that $(U_\nu, 0, 0)$ is the maximal invariant set in $N(\delta,\eta,\nu,\Delta_1)$, but we do not use this fact explicitly belowe.

P r o o f of Lemma 8. According to point c of Lemma 1 we have $\mu'_\eta(u_1) > \vartheta'_\eta(u_1)$ for $u_1 \in \{U_{-1}(\lambda,\eta), U_{+1}(\lambda,\eta)\}$. Thus, for all sufficiently small $\Delta_1 > 0$, we can find a smooth function $\varepsilon_{2\nu}(\eta)$ such that, if $\Delta_{2\nu}(\eta) = \varepsilon_{2\nu}(\eta)\Delta_1$, then the curve $\mathcal{H}_1 = 0$ intersects the upper and lower side of the rectangle $|u_1 - U_{\nu 1}| \le \delta\Delta_1$, $|u_2 - U_{\nu 2}| \le \delta\Delta_{2\nu}(\eta)$ and the curve $\mathcal{H}_2 = 0$ intersects the right and left side of this rectangle. $\mathcal{H}_2 < 0$ (> 0) at the upper (lower) side and $\mathcal{H}_1 < 0$ (> 0) at its right

(left) side. Bounded solutions of the considered family of equations (according to Lemma 4) cannot touch the sets $|z_i| = L$. They can only touch the boundary of this neighbourhood at points whose projection onto the (u_1, u_2)-plane are contained in the sides of the considered rectangle (for fixed δ). However, this is impossible. Suppose for example, that u_2 has a maximum at the upper side of this rectangle. Then $z_2(\zeta) = 0$ and $z_2'(\zeta) \leq 0$ for some $\zeta \in (-\infty, \infty)$. Simultaneously $z_2'(\zeta) = -\mathcal{H}_2(\zeta)$. This contradicts the fact that $\mathcal{H}_2 < 0$ at that side. The remaining cases may be analyzed similarly. □

Let

$$N_1(\eta) := \Big\{(u_1, u_2, z_1, z_2) : \; U_{-i}(\lambda, \eta) \leq u_i \leq U_{+i}(\lambda, \eta), \;\; 0 \leq z_i \leq L, \;\; i = 1, 2\Big\},$$

where L is the number appearing in Lemma 4. Then, let:

$$N(\delta, \eta) := N(\delta, \eta, -, \Delta_1)) \cup N(\delta, \eta, +, \Delta_1),$$

where Δ_1 *is fixed* and so small that $N(\delta, \eta, \nu, \Delta_1)$, $\nu \in \{-, +\}$, is an isolating neighbourhood of the point $(U_\nu(\lambda, \eta), 0, 0)$ and

$$N_3(0, \kappa, \eta) := \Big\{(u_1, u_2, z_1, z_2) : \;\; |u_1 - U_{01}(\lambda, \eta)| + |z_1| < \kappa, \\ |u_2 - U_{02}(\lambda, \eta)| + |z_2| < \kappa\Big\},$$

where κ is a sufficiently small positive number. Finally, let

$$N_2(\delta, \eta) := N_1(\eta) \cup N(\delta, \eta), \qquad N_2(\eta) := N_2(1, \eta),$$
$$N_3(\delta, \eta) := N_2(\delta, \eta) \setminus N_3(0, \kappa, \eta), \qquad N_3(\eta) := N_3(1, \eta).$$

For any compact set Z comprised in the phase space $\mathcal{S}(Z)$ will denote the maximal invariant set comprised in Z.

LEMMA 9. Suppose that Assumptions 1–4 are fulfilled. Then, for sufficiently small $\kappa > 0$ and all $\eta \in [0, 1]$, the set $N_3(\eta)$ is an isolating neighbourhood for the flow determined by (1.η). Furthermore, we have:

1. $\mathcal{S}(N_2(\eta)) = \mathcal{S}(N_1(\eta))$.
2. For any $\theta \in [\theta_0, \theta_1]$

$$\mathcal{S}(N_3(\eta)) = \{\text{singular points} \cup \text{(perhaps) connecting trajectories}\}. \quad \square$$

P r o o f. First, let us note that the following lemma is valid:

LEMMA 10. $\mathcal{S}(N_1(\eta)) \cap \partial N_1(\eta)$ consists only of singular points belonging to N_1. □

The proof of this lemma is given in Appendix C.

Let $N_2(0,\eta) := \cap\{N_2(\delta,\eta) : \delta \in (0,1]\}$. Arguing as in the proof of Lemma Sec. 4. in [1] let us note that

$$N_2(0,\eta)\backslash N_1(\eta) = \Big\{(u_1,u_2,z_1,z_2): \quad (u_1,u_2) = U_\nu(\lambda,\eta),$$
$$\nu \in \{-,+\}, \quad |z_1|, |z_2| < L, \quad z_1 < 0 \text{ or } z_2 < 0\Big\}.$$

Trajectories through the points belonging to this set leave the set $N_2(0)$ in appropriate direction. So, if $\mathcal{S}(N_2(\delta^*,\eta)) \neq \mathcal{S}(N_1(\eta))$ for some $\delta^* \in (0,1]$, then there must exist $\delta \in (0,\delta^*]$ such that $\mathcal{S}(N_2(\delta,\eta)) \cap \partial N_2(\delta,\eta)$ comprises a point not belonging to $\mathcal{S}(N_1(\eta))$. The set $\partial N_2(\delta,\eta)$ may be divided into three parts: $\partial N(\delta,\eta)\backslash N_1(\eta)$, $\partial N_1(\eta)\backslash N(\delta,\eta)$ and $\partial N_1(\eta) \cap \partial N(\delta,\eta)$. In Lemma 8 and the first part of proof of Lemma 10 (Appendix C) we showed that the intersection of $\mathcal{S}(N_2(\delta,\eta))$ with $\partial N(\delta,\eta)\backslash N_1(\eta)$ and $\partial N_1(\eta)\backslash N(\delta,\eta)$ is eiter empty or belongs to N_1. Thus, it suffices to show the following statement:

LEMMA 11. $\mathcal{S}(N_2(\delta,\eta)) \cap \partial N_1(\eta) \cap \partial N(\delta,\eta) = \emptyset$ for all $\delta \in (0,1]$. □

The proof of this lemma is given in Appendix C.

In view of this lemma, point 1 of Lemma 9 is proved.

Now, due to Assumption 3, Lemma 2 and Lemma 7 we infer that the set of points on trajectories comprised in $\mathcal{S}(N_2(\delta,\eta)) = \mathcal{S}(N_1(\eta))$ tending to the point $U_0(\lambda,\eta)$ as $\xi \to \infty$ is empty. Thus, according to Lemma in Sec. 4.D, for κ sufficiently small $N_3(0,\kappa,\eta)$ can be excised from $N_2(\delta,\eta)$, $\delta \in (0,1]$, $\eta \in [0,1]$. It follows that $N_3(\eta)$ is a good isolating neighbourhood. Point 2 of Lemma 9 follows straightforwardly from the definition of $N_1(\eta)$. □

7. Connection index for $\eta = 0$ and existence proof

Now, for $\eta \in [0,1]$, let

$$S'_\eta := (U_-(\lambda,\eta),0,0) \times [\theta_0,\theta_1], \qquad S''_\eta := (U_+(\lambda,\eta),0,0) \times [\theta_0,\theta_1].$$

Let S_η denote the maximal invariant set in the set $N_3(\eta) \times [\theta_0,\theta_1]$ with respect to the flow generated by $(1.\eta)$ together with the equation $\theta' = 0$. Due to the results of the above sections, the connection triples (S'_0, S''_0, S_0) and (S'_1, S''_1, S_1) are related by continuation. By Theorem in Sec. 2.D of [1] these triples have the same (homotopic) connection indices. According to the definition (see [1] and the Appendix A), the connection index of the triple (S'_0, S''_0, S_0) is the Conley index of $N_3(0) \times [\theta_0,\theta_1]$ with respect to the flow generated by Eqs. (1.0) (by which we mean $(1.\eta)$ with $\eta = 0$) written as a first order system, i.e. the system:

$$\begin{aligned} u_1' &= z_1, \\ u_2' &= z_2, \\ z_1' - c_1\theta z_1 + F_s - \lambda(u_1 - u_2) &= 0, \\ kz_2' - c_1 k\theta z_2 + F_s + \lambda(u_1 - u_2) &= 0, \end{aligned}$$

together with the equation

$$\theta' = \beta\phi(u, u')(\theta - 2^{-1}(\theta_0 + \theta_1)),$$

where β is a sufficiently small positive parameter. Let U' and U'' denote open neighbourhoods in $R^4 \times (\theta_0 - \varepsilon, \theta_1 + \varepsilon)$ of $S'(\theta_0) \cup S'(\theta_1)$ and $S''(\theta_0) \cup S''(\theta_1)$, respectively, having disjoint closures. The real-valued continuous function ϕ is arbitrary except for the fact that it is positive in U' and negative in U'' (see Definition A.4 of the Appendix A).

To analyze the connection index for the above system it is convenient to change the dependent variables, namely to consider the system:

$$w' = z'_w, \qquad z'_w - c_1\theta z_w + 2(1+k)^{-1}F_s(w) = 0; \tag{9.a}$$

$$\Delta' = z_\Delta, \qquad z'_\Delta - c_1\theta z_\Delta - (1 + k^{-1})\lambda\Delta = 0; \tag{9.b}$$

$$\theta' = \beta\phi_\Delta\left(\theta - 2^{-1}(\theta_0 + \theta_1)\right), \tag{9.c}$$

where

$$(w, \Delta, z_w, z_\Delta) = \left[(1+k)^{-1}(u_1 + ku_2),\ u_1 - u_2,\ (1+k)^{-1}(z_1 + kz_2),\ z_1 - z_2\right]$$

and

$$\phi_\Delta(w, \Delta, w', \Delta') = \phi(u(w, \Delta), u'(z_w, z_\Delta)).$$

The transformation $(u_1, u_2, z_1, z_2, \theta) \to (w, \Delta, z_w, z_\Delta, \theta)$ is a linear homeomorphism which transforms $N_3(0) \times [\theta_0, \theta_1]$ to a compact set; let us write it as $N_{w\Delta} \times [\theta_0, \theta_1]$. The set of exit points are transformed into the set of exit points, so the invariant set comprised in $N_{w\Delta} \times [\theta_0, \theta_1]$ has the same Conley index as the invariant set contained in $N_3(0) \times [\theta_0, \theta_1]$. Let us denote:

$$P := \{(w, \Delta, z_w, z_\Delta) : \quad \Delta = 0, \quad z_\Delta = 0\}.$$

LEMMA 12. For $\lambda > 0$ the set $S(N_{w\Delta} \times [\theta_0, \theta_1])$ is comprised in the set $P \times [\theta_0, \theta_1]$ □

P r o o f. For any finite values of θ all nonconstant trajectories of solutions to system (9.b) lie either on stable or unstable manifold of the singular point $(0, 0, 0, 0)$, so it leaves $N_{w\Delta}$ in positive or negative "time" direction. □

So, in variables $(w, \Delta, z_w, z_\Delta, \theta)$ the sets S_0' and S_0'' are contained in $P \times [\theta_0, \theta_1]$. Moreover, without losing generality we may assume that the function $\phi_\Delta(w, \Delta, z_w, z_\Delta)$ is constant with respect to (Δ, z_Δ) in some open neighbourhood of the plane P. It is clear that the set $[(\bigcup_{x \in N_{w\Delta} \cap P} Z_{x\varepsilon}) \cap N_{w\Delta}] \times [\theta_0, \theta_1]$, where $Z_{x\varepsilon}$ is the set of points of a plane perpendicular to P at x (in $(w, \Delta, z_w, z_\Delta)$-space) and whose distance from P is not greater than ε, is a good isolating neighbourhood of the considered invariant set. Moreover, according to the robustness of the Conley index theory for sufficiently small (positive) ε, the set

$$(\bigcup_{x \in \mathcal{J}} Z_{x\varepsilon}) \times [\theta_0, \theta_1],$$

where $\mathcal{J}$ is the subset of $N_{w\Delta} \cap P$, such that $Z_{x\varepsilon}$ is completely comprised in $N_{w\Delta}$ for $x \in \mathcal{J}$, retains this property *only if ε is taken sufficiently small.*

Obviously this set can be written as a Cartesian product

$$N^* := N_w \times N_{\Delta\varepsilon} \times [\theta_0, \theta_1] := N_w \times \{(\Delta, z_\Delta) : \ \operatorname{dist}[(\Delta, z_\Delta), (0,0)] \leq \varepsilon\} \times [\theta_0, \theta_1],$$

where N_w is equivalent to $\mathcal{J}$ (defined in terms of w and z_w). Note, that the system consisting of (9.a) and (9.c) does not depend, for sufficiently small ε, on Δ. According to this fact the trajectories belonging to the invariant set do not change, if the second equation in (9.b) is replaced by any of the family of equations $z'_\Delta - \alpha c_1 \theta z_\Delta - (1 + k^{-1})\lambda\Delta = 0$, where $\alpha \in [0, 1]$. Thus, for all $\alpha \in [0, 1]$, the set N^* is a good isolating neighbourhood and we can replace (9.b) by

$$\Delta' = z_\Delta, \qquad z'_\Delta - (1 + k^{-1})\lambda\Delta = 0. \tag{10}$$

In this way the system for (Δ, z_Δ) is completely decoupled from the rest of the system as the equations of (9.a) and (9.c) do not depend on (Δ, z_Δ). Due to the known properties, the Conley index of N^* is homotopic to $h_\Delta \wedge h_{w\theta}$, where h_Δ is the Conley index of $N_{\Delta\varepsilon}$ with respect to (10) and $h_{w\theta}$ is the Conley index of $N_w \times [\theta_0, \theta_1]$ with respect to the flow generated by (9.a), (9.c).

Now, according to Assumptions 2–4 there exists $\theta^s \in (\theta_0, \theta_1)$ such that (9.a) has a heteroclinic solution connecting the points $(w, z_w) = (0, 0)$ and $(w, z_w) = (1, 0)$. Let T_1 denote the trajectory of (9.a) for $\theta = \theta^s$, crossing the z_w axis at a point, say $(0, 1)$. Let $\theta_c \in (0, \theta_0)$ be so small that for $\theta = \theta_c$ the eigenvalues of the linearization matrix of the sytem (9.a) at $(w_0, 0)$ are complex conjugate (and have positive real part). Let T_2 denote a (connected) segment of the spiral trajectory of (9.a) with $\theta = \theta_c$ which lies in the halfplane $z \geq 0$ sufficiently close to $(w_0, 0)$. One can see that (without changing the Conley index) $N_w \times [\theta_0, \theta_1]$ can be deformed to the region bounded by T_1, T_2, the lines $w = -\omega$, $w = 1 + \omega$, $\omega > 0$ small, the boundaries of small diamonds consisting of the points $(1, 0)$ and $(0, 0)$ and the line $z_w = 0$ as it is done in [1]. (During the deformation the invariant trajectory, if it exists, does not touch the boundary of the deformed region). Thus

the Conley index of $N_w \times [\theta_0, \theta_1]$ with respect to the flow generated by (9.a), (9.c) can be computed to be homotopic to $\bar{0}$. Consequently, the connection index of N^* i.e. $h_\Delta \wedge \bar{0} \cong \Sigma^1 \wedge \bar{0} \cong \bar{0}$. On the other hand, according to the results of Sec. 5, for any $\eta \in [0,1]$, the singular points $(U_+(\lambda,\eta),0,0)$ and $(U_-(\lambda,\eta),0,0)$ are isolated invariant sets and the Conley index of them is homotopic to Σ^2. As $(\Sigma^1 \wedge \Sigma^2) \vee \Sigma^2 = \Sigma^3 \vee \Sigma^2$ is not in the homotopy class of $\bar{0}$, then according to Theorem in Ses. 2.F of [1], it follows that $S_1' \cup S_1' \neq S_1$. Consequently in view of Lemma 9 point 2 we infer (by letting $\beta \to 0$) that the following theorem is true:

THEOREM. *Let Assumptions 1 – 4 be satisfied. Then there exists* $\theta^* \in [\theta_0, \theta_1]$ *such that for* $\theta = \theta^*$ *and all sufficiently large* $\lambda > 0$, *there exists a heteroclinic solution to system* (1.1) *connecting the constant states* $U_-(\lambda,1)$ *and* $U_+(\lambda,1)$. □

8. Discussion

It is possible to estimate the minimal value of λ which is sufficient to prove existence of a heteroclinic solution to the system (1.1), which was rather impossible in the method chosen in [5]. An example of such an estimation will be given below. It is worthwhile to note, that this value of λ depends only on the functions F_i and their first derivatives. Especially, as one could foresee, this value does not depend on the other coefficients i.e. k, c_1, c_2. Finally, let us stress that from the mathematical point of view Assumption 4 is not necessary. This condition, which reflects the physical situation described by the system (0), was assumed only for definiteness.

To see, how the minimal value of λ can be estimated, let us take for example a quite realistic situation, when $F_2 \equiv 0$ and $F_1 = F(u_1)$ (which corresponds to the assumption that the energy is gained and radiated out only by the electron component).

LEMMA 13. For $F_2 \equiv 0$ the solutions to system (4) are independent of η. □

P r o o f. The system (4) takes the form:

$$-\lambda\Delta + \eta F(u_1) + (1-\eta)F_s(w) = 0,$$
$$\lambda\Delta + (1-\eta)F_s(w) = 0,$$

where $\Delta = (u_1 - u_2)$ and $w = (u_1 + ku_2)(1+k)^{-1}$. We have $F(u_1) = F_s(w) + F'(w^*)k(k+1)^{-1}\Delta$, where $w^* \in [u_1, w]$. Suppose that, for a fixed $\eta \in [0,1]$, this system has a solution (u_1, u_2), for which $\Delta \neq 0$. Multiplying the second equation by $(1-\eta F'(w^*)k(k+1)^{-1}\lambda^{-1})$ and adding it to the first one we obtain an equation $[1 + (1-\eta)(1-\eta F'(w^*)k(k+1)^{-1}\lambda^{-1}]F_s = 0$. Consequently, for λ sufficiently large $F_s = 0$, and from the second equation we infer that $\Delta = 0$. □

To find the estimation we will verify in turn all the conditions imposed on λ in the text above.

First, the positiveness of the determinant at U_ν, $\nu \in \{, -, +\}$, its negativeness at U_0 and the conditions $(-\lambda(u_1 - u_2) + \eta F + (1 - \eta)F_s)_{,2} > 0$, $(\lambda(u_1 - u_2) + k(1 - \eta)F_s)_{,1} > 0$ (which were necessary in the proof of Lemmas 7 and 10) are guaranteed by $\lambda > \widehat{F} := \max_{u\in[0,1]} |F'(u)|$.

Now, according to the proof of Lemma 3, $|u_1 - u_2|$ is *a priori* smaller than

$$d = 6\lambda^{-1}(\max_{u\in[0,1]} |F(u)|) := 6\lambda^{-1}F_m\,.$$

Let

$$\int_0^1 F(u)\,du = I > 0, \qquad \min_{u\in[0,1]} \int_0^u F(u)\,du = J = \int_0^j F(u)\,du < 0.$$

We have

$$\begin{aligned}\mathcal{I} &:= (1 + k)^{-1} \int_{-\infty}^{\infty} (\mathcal{F}_{1\eta} + \mathcal{F}_{2\eta})(u_1'\,d\xi + ku_2'\,d\xi)\\ &= (1 + k)^{-1} \int_{-\infty}^{\infty} [\eta F(u_1) + 2(1 - \eta)F_s(w)]\,(u_1'\,d\xi + ku_2'\,d\xi)\\ &= 2(1 - \eta) \int_0^1 F(w)\,dw + (1 + k)^{-1}\eta \int_0^1 F(u_1)\,du_1\\ &\quad +(1 + k)^{-1}\eta \left\{ \int_{-\infty}^{\infty} F(u_2)ku_2'\,d\xi + \int_{-\infty}^{\infty} F'(u^*)dku_2'\,d\xi \right\},\end{aligned}$$

where $u^*(\xi) \in [u_1(\xi), u_2(\xi)]$.

The sum of the first three terms is equal to $[2 - \eta]I$ and the module of last term is estimated by the number $\widehat{F}dk = 6\lambda^{-1}\widehat{F}F_m$. Thus, for $\lambda > 6\widehat{F}F_m I^{-1}$ the integral $\mathcal{I}$ is larger than 0. Likewise, we can *a priori* estimate the minimum over ξ of the integral

$$\mathcal{J}(\xi) := (1 + k)^{-1} \int_{-\infty}^{\xi} (\mathcal{F}_{1\eta} + \mathcal{F}_{2\eta})(u_1'\,d\xi + ku_2'\,d\xi).$$

So, acting as before we can write $\mathcal{J}$ as:

$$\begin{aligned}&2(1 - \eta)) \int_0^{w(\xi)} F(w)\,dw + (1 + k)^{-1}\eta \int_0^{u_1(\xi)} F(s)\,ds\\ &\quad +(1 + k)^{-1}\eta \left\{ \int_0^{u_2(\xi)} F(u_2)ku_2'\,d\xi + \int_0^{u_2(\xi)} F'(u^*)\,[u_2(\xi) - u_1(\xi)]\,ku_2'\,d\xi \right\},\end{aligned}$$

where $u^*(\xi) \in [u_1(\xi), u_2(\xi)]$.

Let us choose ξ in such a way that $u_1(\xi) = j$. Then, $\mathcal{J}$ can be estimated from above by:

$$2(1-\eta)J+\eta(1+k)^{-1}(1+k)J+2(1-\eta)dF_m+\eta(1+k)^{-1}k\left\{F_m d+\hat{F}d(j+d)\right\}$$
$$=[2-\eta]J+[2(1-\eta)+\eta(1+k)^{-1}k]F_m d+\hat{F}dk\eta(1+k)^{-1}(j+d).$$

As the trajectory must stay in the rectangle $[0,1]\times[0,1]$, then $(j+d)\leq 1$ and $\mathcal{J}$ is smaller than zero if

$$6\lambda^{-1}F_m < d < (2-\eta)|J|\left[2(1-\eta)F_m+\eta(1+k)^{-1}kF_m+\eta k(1+k)^{-1}\hat{F}\right]^{-1}.$$

This condition is satisfied for

$$\lambda > 6F_m\left[2(1-\eta)F_m+\eta(1+k)^{-1}k\left(F_m+\hat{F}\right)\right](2-\eta)^{-1}|J|^{-1}.$$

The right-hand side of this inequality is smaller than $6F_m(F_m+\hat{F})|J|^{-1}$, independently of η and k. Putting everything together we can say that the heteroclinic trajectory for some finite $\theta=\theta^*>0$ exists if only

$$\lambda > \max\left\{\hat{F}, 6F_m(F_m+\hat{F})|J|^{-1}, 6F_m\hat{F}I^{-1}\right\}.$$

As $|J|<F_m$, $I<F_m$, then

$$\lambda > 6F_m(F_m+\hat{F})\left(\min\{|J|,I\}\right)^{-1}.$$

In a general case the evaluation can be carried out in principle in the same way, though it would be a little bit more laborious.

Appendix A

Let us recall the basic facts of the connection triple theory taken from [1] (see also [2, 3]), which are used to prove existence of heteroclinic orbits. Suppose that we are given a system of n first order ordinary differential equations (in R^n) parametrized (continuously) by a parameter θ belonging to some nonempty closed interval $[\theta_0,\theta_1]$. Let $X=R^n\times[\theta_0,\theta_1]$. Let S', S'' and S be isolated invariant sets for the flow on X determined by this family of equations and let $S'(\theta)$, $S''(\theta)$, $S(\theta)$ be the set of points in S', S'', S with parameter value θ.

DEFINITION A.1. *The triple S', S'', S is called a connection triple if the following conditions are satisfied:*

a. $S'\cup S''\subset S$,

b. $S'\cap S''=\emptyset$,

c. *for* $\theta=\theta_0$ *and* $\theta=\theta_1$, $S(\theta)=S'(\theta)\cup S''(\theta)$. □

Now, suppose that we are given a family of local flows on a space X parametrized in a continuous way by a parameter $\eta \in [0,1]$. Suppose that S_0 and S_1 are isolated invariant sets for the flows on X corresponding to $\eta = 0$ and $\eta = 1$.

DEFINITION A.2. *We say that S_0 and S_1 are related by continuation, if there exists a compact set N in the space $X \times [0,1]$ such that N_η (i.e. the set of points in N with parameter value η) is an isolating neighbourhood and such that N_0 and N_1 are isolating neighbourhoods for S_0 and S_1, respectively.* □

DEFINITON A.3. *Suppose that for each $\eta \in [0,1]$ there exist compact sets N', N'', N such that N'_η, N''_η, N_η are isolating neighbourhoods for the isolated invariant sets S'_η, S''_η and S_η, respectively. Suppose then, that for each $\eta \in [0,1]$ $(S'_\eta, S''_\eta, S_\eta)$ is a connection triple. Then, we say that the triples (S'_0, S''_0, S_0) and (S'_1, S''_1, S_1) are related by continuation.* □

With a connection triple an index $h(S', S'', S)$ may be connected. Its definition may be found for instance in [1] (see Lemma, p. 325).

DEFINITION A.4. *Let (S', S'', S) be a connection triple for a family of differential equations on R^n parametrized by θ in the interval $[\theta_0, \theta_1]$. Assume the equations are defined for $\theta \in (\theta_0 - \varepsilon, \theta_1 + \varepsilon)$ for some $\varepsilon > 0$ (this is no real restriction – they can be extended to such an interval). Let U' and U'' be open neighbourhoods in $R^n \times (\theta_0 - \varepsilon, \theta_1 + \varepsilon)$ of $S'(\theta_0) \cup S'(\theta_1)$ and $S''(\theta_0) \cup S''(\theta_1)$ (respectively); choose these to have disjoint closures. Let ϕ be a continuous real-valued function on R^n which is positive on U' and negative U''. Append to the given family of equations the equation $\theta' = \mu\phi(x)[\theta - 2^{-1}(\theta_0 + \theta_1)]$, where μ is a small positive parameter. Let N be a compact neighbourhood in $R^n \times (\theta_0 - \varepsilon, \theta_1 + \varepsilon)$ such that $N(\theta)$ is an isolating neighbourhood of $S(\theta)$ for each θ. Then, there is a $\mu_0 > 0$ such that if $\mu \in (0, \mu_0)$ then N is an isolating neighbourhood for the "enlarged system". Let h_μ be the index of $S(N)$, $\mu \in (0, \mu_0)$. Then h_μ is independent of μ, and in fact depends only on the triple (S', S'', S).* □

Now, let us assume (as it is in our case) that for every $\theta \in [0,1]$ $S'(\theta)$, $S''(\theta)$ are fixed hyperbolic singular points and their indices are constant. Let us denote them by h' and h'', respectively. Now, if there was no connection between S' and S'', then due to point b of Definition A.1, and Definition A.4 we would have $h(S', S'', S) = (\Sigma^1 \wedge h') \vee h''$, where h' and h'' are the Conley indices of $S'(\theta)$ and $S''(\theta)$, $\theta \in [\theta_0, \theta_1]$. So, if $h(S', S'', S) \neq (\Sigma^1 \wedge h') \vee h''$, then $S \neq S' \cup S''$.

The final theorem necessary for our proof is stated in [1] Section D p. 326.

PROPOSITION. The index of a connection triple is constant on equivalence classes under the continuation relation. □

Appendix B

In this appendix we prove Lemmas 6 and 7 concerning the eigenvalues of the linearized system.

Let us fix λ, η and ν and denote for simplicity:

$$\mathcal{M} := \mathcal{M}_\nu(\lambda,\eta) := \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad M := M_\nu(\lambda,\eta), \qquad t := c_1\theta, \qquad Tt := c_{2\eta}k^{-1}\theta.$$

Let $(\xi_1,\xi_2,\zeta_1,\zeta_2)$ be the eigenvector of M corresponding to an eigenvalue q of M. As we mentioned in Sec. 5, it follows from the structure of $M(\lambda,\eta)$ and considerations in [1] p. 335 that $(\xi_1,\xi_2,\zeta_1,\zeta_2)$ and q are coupled simultaneously by the following three relations:

$$\begin{aligned} \zeta_i &= q\xi_i, \qquad i = 1,2, \\ q &= 2^{-1}t \pm \sqrt{4^{-1}t^2 - (a + b\xi_2\xi_1^{-1})}, \\ q &= 2^{-1}Tt \pm \sqrt{4^{-1}(Tt)^2 - (d + c\xi_2^{-1}\xi_1)}. \end{aligned} \tag{B.1}$$

The eigenvalues of $M(\lambda,\eta)$ are the roots of the equation:

$$\det(\mathcal{M}) + q(-dt - atT) + q^2(a + d + t^2T) + q^3(-t - tT) + q^4 = 0. \tag{B.2}$$

Using this fact we can prove the following lemma.

LEMMA B.1. Suppose that $\det\mathcal{M} \neq 0$, $a < 0$, $d < 0$, $t > 0$, $tT > 0$. Then the real part of the eigenvalues of M is different from 0. □

P r o o f. As we have noticed, these eigenvalues are given by the roots of Eq. (B.2). It is obvious that 0 is not a solution of it. So, suppose that there is a pair of eigenvalues q_1, q_2 such that $q_1 = iL$, $q_2 = -iL$, $L \neq 0$. Substituting in (B.2) first $q = iL$ and next $q = -iL$ and subtracting the obtained equations, we arrive at the equation $2iLt(-aT - d + (1 + T)L^2) = 0$. However, according to the assumptions of the lemma we infer that the expression in the bracket is not equal to zero. □

Now, let us note that according to Lemma 2 and Assumption 3 for sufficiently large λ the following conditions are fulfilled (independently of $\eta \in [0,1]$):

$$\det(\mathcal{M}) \neq 0, \qquad t > 0, \qquad tT > 0, \qquad a < 0, \qquad d < 0, \qquad bc > 0. \tag{B.3}$$

P r o o f of Lemma 6. Let us note that, if we put $\nu \in \{-,+\}$ in the definition of $\mathcal{M}$, then, due to Assumption 3, Lemma 1 and Lemma 2, $\det(\mathcal{M}) > 0$ for sufficiently large λ (independently of $\eta \in [0,1]$). Lemma 6 follows straightforwardly from the following more general lemma.

LEMMA B.2 Assume (B.3) and that $\det(\mathcal{M}) > 0$. Then the matrix $M(\lambda,\eta)$ has four real eigenvalues. Two of them are positive and two are negative. Moreover, $\xi_2\xi_1^{-1} > 0$ only for one of the positive and one of the negative eigenvalues. □

P r o o f. The condition $\det(\mathcal{M}) > 0$ implies that either $a^2 - b^2 > 0$ or $d^2 - c^2 > 0$ (or both). However, as one can easily see by renumbering the components of u, we may always assume that the first one is satisfied, i.e. $(a^2 - b^2) > 0$.

As a starting point of our analysis we will take the situation characterized by the equalities: $T = 1$, $a = d$, $b = c$. In this case (B.2) takes the form:

$$(-a - b - q^2 + qt)(-a + b - q^2 + qt) = 0.$$

As $(a^2 - b^2) > 0$, then this equation has four real solutions:

$$q_{1\pm} = 2^{-1}\left(t \pm \sqrt{-4a - 4b + t^2}\right), \qquad q_{2\pm} = 2^{-1}\left(t \pm \sqrt{-4a + 4b + t^2}\right).$$

Two of them are positive and two are negative.

For $\varrho \in [0, 1]$ let us make the following deformation:

$$\begin{aligned} &a(\varrho) = a, \qquad d(\varrho) = a + (d - a)\varrho, \qquad b(\varrho) = b, \\ &c(\varrho) = b + (c - b)\varrho, \qquad T(\varrho) = 1 + (T - 1)\varrho. \end{aligned} \tag{B.4}$$

Then we have:

$$\det(\mathcal{M}(\varrho)) = (1 - \varrho)\det(\mathcal{M}(0)) + \varrho \det(\mathcal{M}(1)). \tag{B.5}$$

Thus, it follows that the deformed coefficients retain assumptions (B.3). For simplicity we will omit the explicit dependence of c, d and T on ϱ, if it does not cause confusion.

Now, as b and c $(= c(\varrho))$ are both nonzero, then, according to relations (B.1) and to the structure of eigenvectors, we can assume that ξ_1 is nonzero (the ratio $\xi_2\xi_1^{-1}$ must stay finite). Thus without losing generality we can assume for definiteness that $\xi_1 = 1$ and the eigenvector corresponding to an eigenvalue q has the form $(1, \xi_2, q, q\xi_2)$. So (B.1) can be written as:

$$\begin{aligned} q &= 2^{-1}t \pm \sqrt{4^{-1}t^2 - (a + b\xi_2)}, \\ q &= 2^{-1}Tt \pm \sqrt{4^{-1}(Tt)^2 - (d + c\xi_2^{-1})}. \end{aligned} \tag{B.6}$$

Now, it can be easily proved that during the above deformation (with respect to ϱ) the eigenvalues of M stay real (and according to Lemma 8 two of them are positive). For $\varrho = 0$ the eigenvectors of M take the form $\{1, 1, q_{1\pm}, q_{1\pm}\}$, $\{1, -1, q_{2\pm}, q_{2\pm}\}$, where $q_{1\pm} = 2^{-1}(t \pm \sqrt{-4a - 4b + t^2})$ and $q_{2\pm} = 2^{-1}(t \pm \sqrt{-4a + 4b + t^2})$. (Remind that we have set $\xi_1 = 1$ for definiteness).

As ξ_2 never becomes 0, then its sign will not change during the deformation. The proof of Lemmat B.2 is thus completed. □

P r o o f of Lemma 7. Let us note, that, if we put $\nu = 0$ in the definition of $\mathcal{M}$, then, due to Assumption 3, Lemma 1 and Lemma 2, $\det(\mathcal{M}) < 0$ for

sufficiently large λ (independently of $\eta \in [0,1]$). In this proof we will also use the fact that for sufficiently large λ both b and c are positive. Note, that during the deformation (B.4), the condition $\det(\mathcal{M}(\varrho)) < 0$ retains its validity according to (B.5). Using arguments as in the proof of Lemma 6 we may assume without losing generality that $(a^2 - b^2) < 0$. According to Lemma B.1, the eigenvalues of M cannnot cross the imaginary axis. The number of eigenvalues with positive or negative real parts is constant during the deformation. As $(a+b)(a-b) < 0$, then one of these factors is positive and the other is negative. Thus for $\varrho = 0$ there exists only one eigenvalue with negative real part and three ones with $\operatorname{Re} q > 0$. As b and c are positive, then the sign of ξ_2 for negative eigenvalue is the same as its sign for $\varrho = 0$, as it cannot become 0 (see the proof of Lemma 6). As $(a-b) < (a+b)$ and $(a-b) < 0$, then, for $\varrho = 0$ the negative eigenvalue is equal to $q = 2^{-1}(t - \sqrt{4a + 4b + t^2})$. Comparing it with $(B.6)_1$ we obtain the claim of the lemma. □

Appendix C

P r o o f of Lemma 10. To prove Lemma 10 we will show that a trajectory in the closure of N_1 cannot touch ∂N_1 and then return to its interior unless at singular points. As, according to Lemma 4, all bounded solutions of our system have its derivatives estimated in their absolute value by a common finite constant, then it suffices to examine the following cases:

1. $u_i(\zeta) = U_{-i}(\lambda, \eta)$ or $u_i(\zeta) = U_{+i}(\lambda, \eta)$ for some $\zeta \in (-\infty, \infty)$.

a. Let $z_i(\zeta) \neq 0$. Then the trajectory leaves N_1 immediately.

b. Let $z_i(\zeta) = 0$. Due to Lemma 1, μ'_η and ϑ'_η cannot achieve nonpositive values, so the lines $\mathcal{H}_1 = 0$, $\mathcal{H}_2 = 0$ cannot intersect the sides $N_1(\eta) \cap \{(u_1, u_2)\}$ except at the singular points. Moreover, at the upper (lower) side of this rectangle we have $\mathcal{H}_2 < 0$ (> 0) and at the right (left) side $\mathcal{H}_1 < 0$ (> 0), except for the singular points. The proof that such a trajectory leaves $N_1(\eta)$ if it does not reach singular points is carried out as in Lemma 8.

2. $z_i(\zeta) = 0$ for some $\zeta \in (\infty, \infty)$.

a. Let $z'_i(\zeta) \neq 0$. Then the trajectory leaves N_1 immediately.

b. Let $z'_i(\zeta) = 0$. Then also $[-\mathcal{F}_{i\eta} + \lambda(u_1 - u_2)(-1)^{i-1}](\zeta) = 0$. Let j denote the index complementary to i. Then one obtains by differentiation:

$$z''_i(\zeta) = \{-(\mathcal{F}_{i\eta,j})(\zeta) - \lambda\} z_j(\zeta).$$

Thus, if $\lambda > 0$ is sufficiently large (larger than nondiagonal entries of the matrix $\mathcal{F}_{i\eta,j}$) and $z_j(\zeta) > 0$, then $z''_i(\zeta) < 0$, so that near this point $z_i < 0$ and the trajectory lies outside N_1. Now, let us assume that $z_j(\zeta) = 0$. The trajectory leaves N_1 (in appropriate direction), unless $z'_j(\zeta) = 0$. Then, however, we would

have also $[-\mathcal{F}_{j\eta} + \lambda(u_1 - u_2)(-1)^{j-1}](\zeta) = 0$, so this point would be a singular point. □

P r o o f of Lemma 11. First, let us note that $\partial N_1(\eta)$ consists of the following sets:

$$\{z_i = 0\} \cap N_1(\eta), \qquad \{z_i = L\} \cap N_1(\eta),$$
$$\{u_i = U_{-i}(\lambda, \eta)\} \cap N_1(\eta), \qquad \{u_i = U_{+i}(\lambda, \eta)\} \cap N_1(\eta),$$

$i \in \{1, 2\}$. The second pair of sets cannot comprise points lying on bounded trajectories. The first one intersecting $\partial N(\delta, \eta)$ gives us eight sets, namely:

$$\{z_i = 0,\ u_1 - U_{-1}(\lambda, \eta) = \delta\Delta_1,\ 0 \le z_j < L\},$$
$$\{z_i = 0,\ u_2 - U_{-2}(\lambda, \eta) = \delta\Delta_{-2}(\eta),\ 0 \le z_j < L\}$$
$$\{z_i = 0,\ u_1 - U_{+1}(\lambda, \eta) = \delta\Delta_1,\ 0 \le z_j < L\},$$
$$\{z_i = 0,\ u_2 - U_{+2}(\lambda, \eta) = \delta\Delta_{+2}(\eta),\ 0 \le z_j < L\},$$

where $i \in \{1, 2\}$ and j is the index complementary to i and $\nu \in \{-, +\}$. Below, we will show that a trajectory touching one of the above sets cannot belong to $\mathcal{S}(N_2(\delta, \eta))$, i.e. it leaves $N_2(\delta, \eta)$ when continued in appropriate direction. Let us consider particular cases.

1. $u_1(\zeta) - U_{+1} = -\delta\Delta_1$ for some $\zeta \in (-\infty, \infty)$. If $z_1(\zeta) = 0$, then we arrive at the case analyzed in Lemma 8. So, let us suppose that $z_1(\zeta) > 0$ and $z_2(\zeta) = 0$. We can distinguish the three possibilities:

a. $\mathcal{H}_2(\zeta) < 0$, $z_2'(\zeta) = -\mathcal{H}_2(\zeta) > 0$. Then, for decreasing "times" the trajectory leaves $N(\delta, \eta)$ (as u_1 decreases) and $N_1(\eta)$ (as z_2 becomes negative).

b. $\mathcal{H}_2(\zeta) > 0$, $z_2'(\zeta) = -\mathcal{H}_2(\zeta) < 0$, so u_2 achieves a maximum. Consider increasing "times". Then z_2 becomes negative and never achieves the value 0 again while staying in $N(\delta, \eta)$. For, suppose to the contrary, that there exists $\zeta_1 \in (\zeta, \infty]$, such that $z_2(\zeta_1) = 0$ and $z_2(\xi) < 0$ for $\xi \in (\zeta, \zeta_1)$. Then $z_2'(\zeta_1) \ge 0$. But, simultaneously $z_2'(\zeta_1) = -\mathcal{H}_2(\zeta_1) < 0$, as the curve $\mathcal{H}_2 = 0$ lies above the starting point P_1 (see Fig. 2) and it has positive slope in $N(\delta, \eta)$. The trajectory can reach the curve $\mathcal{H}_2 = 0$ only outside $N(\delta, \eta)$. But leaving $N(\delta, \eta)$ would imply leaving also $N_1(\eta)$, as $z_2(\xi) < 0$ for $\xi < \zeta$.

c. $\mathcal{H}_2(\zeta) = 0$. Then $z_2'(\zeta) = -\mathcal{H}_{2,1}(\zeta)z_1 < 0$. Thus, for increasing "times" this case is the same as case b.

2. $u_1(\zeta) - U_{-1} = \delta\Delta_1$ for some $\zeta \in (-\infty, \infty)$. If $z_1(\zeta) = 0$, then we arrive at the case considered in Lemma 8. So, let us suppose that $z_1(\zeta) > 0$ and $z_2(\zeta) = 0$. As before, some particular cases are to be distinguished:

a. $\mathcal{H}_2(\zeta) > 0$, $z_2'(\zeta) = -\mathcal{H}_2(\zeta) < 0$. Then for increasing "times" z_2 becomes negative and the trajectory "immediately" leaves $N(\delta, \eta)$ (u_1 grows) and $N_1(\eta)$ (z_2 decreases).

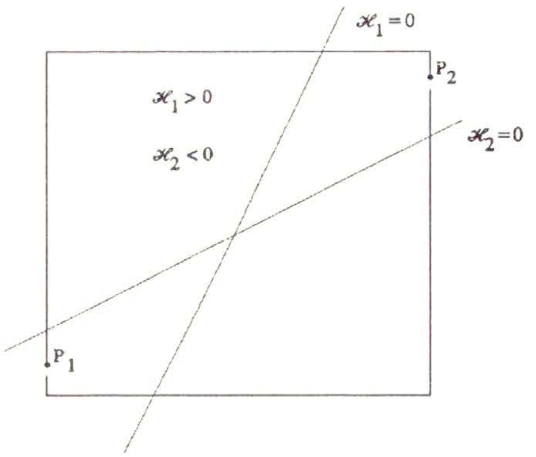

FIG. 2.

b. $\mathcal{H}_2(\zeta) < 0$, $z_2'(\zeta) = -\mathcal{H}_2(\zeta) > 0$, so u_2 achieves a minimum. Consider decreasing "times". Then z_2 becomes negative and never achieves the value 0 again while staying in $N(\delta, \eta)$. For, suppose to the contrary, that there exists $\zeta_1 \in [-\infty, \zeta)$, such that $z_2(\zeta_1) = 0$ and $z_2(\xi) < 0$ for $\xi \in (\zeta_1, \zeta)$. Then $z_2'(\zeta_1) \leq 0$. But, simultaneously $z_2'(\zeta_1) = -\mathcal{H}_2(\zeta_1) > 0$, as the curve $\mathcal{H}_2 = 0$ lies below the starting point P_2 (see Fig. 2) and it has positive slope in $N(\delta, \eta)$. The trajectory can reach the curve $\mathcal{H}_2 = 0$ only outside $N(\delta, \eta)$. But leaving $N(\delta, \eta)$ would imply leaving also $N_1(\eta)$, as $z_2(\xi) < 0$ for $\xi < \zeta$.

c. $\mathcal{H}_2(\zeta) = 0$, $z_2'(\zeta) = -\mathcal{H}_{2,1} z_1 < 0$. Thus for increasing "times" this case is the same as case a.

3. The remaining cases are considered similarly.

Now, the intersection of the sets: $\{u_i = U_{-i}(\lambda, \eta)\} \cap N_1(\eta)$, $\{u_i = U_{+i}(\lambda, \eta)\} \cap N_1(\eta)$ with $\partial N(\delta, \eta)$ gives us the following sets:

$$\{u_1 - U_{-1}(\lambda, \eta) = \delta\Delta_1,\ u_2 = U_{-2}(\lambda, \eta),\ 0 \leq z_k < L,\ k = 1, 2\},$$
$$\{u_2 - U_{-2}(\lambda, \eta) = \delta\Delta_{-2},\ u_1 = U_{-1}(\lambda, \eta),\ 0 \leq z_k < L,\ k = 1, 2\},$$
$$\{u_1 - U_{+1}(\lambda, \eta) = -\delta\Delta_1,\ u_2 = U_{+2}(\lambda, \eta),\ 0 \leq z_k < L,\ k = 1, 2\},$$
$$\{u_2 - U_{+2}(\lambda, \eta) = -\delta\Delta_{+2},\ u_1 = U_{+1}(\lambda, \eta),\ 0 \leq z_k < L,\ k = 1, 2\}.$$

Let us take, for example, the first set. Let us look at the projection of the trajectory onto the (u_1, u_2)-space. This projection starts at the point

$$X_s = (U_{-1}(\lambda, \eta) + \delta\Delta_1,\ U_{-2}(\lambda, \eta)) := (u_1(\xi_s), u_2(\xi_s)).$$

As $\mathcal{H}_1(\xi_s) > 0$, then, $z_1(\xi) > 0$ for all $\xi < \xi_s$ sufficiently close to ξ_s. (If $z_1(\xi_s) = 0$, then $z_1'(\xi_s) < 0$.) So that the backward trajectory could stay in the set $N_-(\delta, \eta) \cup N_1(\eta)$, for sufficiently small $\xi < \xi_s$ we should have $z_1'(\xi) < 0$ i.e. $z_1(\xi^*) = 0$ and

$z_1(\xi^*) = -\mathcal{H}_1(\xi^*) \geq 0$ for some $\xi^* < \xi_s$. Such a situation could happen only below the curve $\mathcal{H}_1 = 0$ (or just on it). However, the trajectory arriving at the curve $\mathcal{H}_1 = 0$ must come below the curve $\mathcal{H}_2 = 0$, first. But, below that curve we would have $z_2 \geq 0$, due to the fact that $z_2 = 0$ implies $z_2' = -\mathcal{H}_2 < 0$ (we consider the backward trajectory). Consequently, the projection of the backward trajectory must cross the boundary of $N_-(\delta, \eta)$ at the point not belonging to $\partial N_1(\eta)$. Thus this trajectory does not stay in $N_2(\delta, \eta)$ (see Fig. 1).

The proof that the trajectory (in appropriate time direction) starting at a point belonging to the other three of the sets written down below does not stay in the set $N_2(\delta, \eta)$, is carried out almost verbatim in the same way as above. So, the proof of Lemma 11 is completed. □

Acknowledgments

The author wishes to express his gratitude to Prof. Z. PERADZYŃSKI for pointing out the problem. This work was supported by grant KBN No 7T07A00109.

References

1. C. CONLEY and R. GARDNER, *An application of the generalized Morse index to travelling wave solutions of a competitive reaction diffusion model*, Indiana University Mathematical J., **33**, 3, 1984.
2. C. CONLEY and J. SMOLLER, *Isolated invariant sets of parametrized systems of differential equations*, [in:] The Structure of Attractors in Dynamical Systems, Lecture Notes in Math., **668**, Springer-Verlag, Berlin 1978.
3. C. CONLEY, *A qualitative singular perturbation theorem*, [in:] Global Theory of Dynamical Systems, Lecture Notes in Math., **819**, Springer-Verlag, Berlin 1981.
4. R. GARDNER, *Existence of travelling wave solutions of predator-prey systems via the connection index*, SIAM J. Appl. Math., **44**, 1, 1984.
5. B. KAŹMIERCZAK and Z. PERADZYŃSKI, *Heteroclinic solutions for a system of strongly coupled ODE's*, Math. Meth. in Appl. Scien., **19**, 6, pp. 451-461.

POLISH ACADEMY OF SCIENCES
INSTITUTE OF FUNDAMENTAL TECHNOLOGICAL RESEARCH

Received February 16, 1996.

Arch. Mech., 48, 5, pp. 893–910, Warszawa 1996

Constitutive relations and internal equilibrium condition for fluid-saturated porous solids Nonlinear theory

M. CIESZKO and J. KUBIK (POZNAŃ)

NONLINEAR CONSTITUTIVE relations for the fluid-saturated porous elastic solid with isotropic pore structure undergoing pure mechanical large deformations are developed. The fluid-solid composition is considered as the immiscible mixture consisting of physically identifiable constituents preserving its own individual, physical properties during deformation process. Considerations are based on the balance equation for the internal energy of the whole composition which is required to be satisfied identically by the internal energy constitutive functions postulated for particular components independently. Constitutive relations for partial stresses of particular constituents are obtained, and the internal equilibrium condition for the whole composition is established. These relations for the media with incompressible matrix material and for nonsaturated porous skeleton are discussed.

1. Introduction

MACRO-CONTINUUM constitutive modelling of fluid-saturated porous solids has been a subject of wide discussion through the last decades. Nonlinear models of such materials are based mostly upon the fundamental notions of the Classical Mixture Theory, [5, 23], and its reformulated form – the Theory of Interacting Continua, [9, 10]. Classical mixtures are considered to be composed of miscible constituents (miscible mixture) and within that theory, a fluid-filled porous medium is treated as the superposition of two continua (solid and fluid) characterized by two independent velocity fields. In such approach, the microstructure of solid-fluid composition is not taken into account in formulation of the balance equations and constitutive relations. At the same time, constitutive theories of classical mixtures quickly become complex and unwieldy, even for the simplest constitutive assumptions, when they are based on the principle of equipresence [23], which assumes that each constitutive quantity of a particular component depends on a set of independent variables for the whole solid-fluid composition (see e.g. [1, 2, 7, 11, 12]).

It is evident, however, that such materials as saturated sands, soils, porous rocks, sintered metals, sponges etc. consist of physically identifiable solid matrix and a fluid filling its pores that retain their material integrity, and thus their individual physical properties, during a deformation process. Therefore, porous materials filled with fluid, contrary to the classical mixtures, have internal geometrical structure reflecting the fact of immiscibility of constituents and characteristics of this structure play important role in both transport phenomena and

constitutive modelling of such materials. This also proves that in the local sense, each constituent will obey the constitutive relations for that constituent alone.

Among many works developing the macro-continuum constitutive modelling of solid-fluid mixture there are papers that regard the immiscibility effect by incorporating in the description the parameter of volume porosity characterizing the volume fractions of the constituents (see for example [4, 6, 8, 13, 19–22]). Most of these papers have in common the fact that they apply the principle of equipresence in formulation of constitutive relations for partial quantities of the individual components of the solid-fluid mixture ([6, 8, 19–22]), or for the energy constitutive functions concerning the whole aggregate ([4, 13]). Such approach does not prove to be self-consistent in treating the immiscibility effect as the immanent feature of the porous solid-fluid composition.

The extensive literature concerning the different descriptions of immiscible and structured mixtures can be found in the review paper [3].

The purpose of this study is to develop, within the macro-continuum description, the nonlinear constitutive relations for fluid-saturated porous solids undergoing pure mechanical deformations where the main consequences of the immiscibility, i.e. the skeleton pore structure characteristics and mutual independence of mechanical properties of individual constituents are taken into account. The components are assumed to be elastic and the pore structure has isotropic and homogeneous properties in the macroscopic (averaged) sense.

Considerations are based on the balance equation for the internal energy of the whole composition, which is required to be satisfied identically by the internal energy constitutive functions postulated for particular components, the functional forms of which reflect their individual features.

This enables one to obtain two nonlinear constitutive relations for stresses (one for porous skeleton and the other for pore fluid), and the relation for interface interaction force. Moreover, the additional relation is derived which is the condition of internal equilibrium for the solid-fluid composition. It relates the pore fluid pressure with independent variables describing the deformation state of porous skeleton.

In the paper three particular cases of constitutive relations are also considered. They concern the porous solids with isotropic mechanical properties, with incompressible material of the skeleton and the case when porous solid is not saturated with fluid.

2. Balance equations for mass, linear momentum and internal energy

In our considerations we use the macroscopic continuum description of fluid-saturated porous solid, the pore structure of which is isotropic and characterized by two parameters: the volume porosity f_v and the structural permeability λ (or equivalently $\kappa = \lambda/f_v < 1$). The quantity f_v represents the fluid volume fraction

and λ is the measure of inhomogeneity of the fluid micro-velocity field in its flow relative to the skeleton, caused by the pore structure [14–16]. The characteristic feature of this theory is that the description of kinematics and dynamics of the porous solid-fluid mixture is referred to the so-called virtual components while the description of constitutive properties of the mixture is formulated for its physical constituents.

The physical constituents are: the fluid ($|^f$) and the porous skeleton ($|^s$), which are chemicaly inert and their mass is conserved. Therefore the appropriate continuity equations have the classical form used within the mixture theory [5]

$$\frac{\partial \overline{\varrho}^s}{\partial t} + \operatorname{div}(\overline{\varrho}^s \mathbf{v}^s) = 0,$$
$$\frac{\partial \overline{\varrho}^f}{\partial t} + \operatorname{div}(\overline{\varrho}^f \mathbf{v}^f) = 0,$$

where $\overline{\varrho}^s$ and $\overline{\varrho}^f$ are the partial densities of porous solid and fluid, respectively, and $\mathbf{v}^s$ and $\mathbf{v}^f$ stand for the mass average velocities of the constituents.

The virtual constituents are formed by the porous skeleton and fluid associated with it – the first virtual constituent ($|^1$) moving at the skeleton velocity $\overset{1}{\mathbf{v}}$, and the free fluid – the second virtual constituent ($|^2$) moving at its own velocity $\overset{2}{\mathbf{v}}$. These velocities are related to velocities of the physical constituents as follows, [14, 15]:

$$\overset{1}{\mathbf{v}} = \mathbf{v}^s,$$
$$\overset{2}{\mathbf{v}} = \mathbf{v}^s + \frac{1}{\kappa}(\mathbf{v}^f - \mathbf{v}^s).$$

The virtual constituents in the macroscopic description result from the requirement that the whole linear momentum and kinetic energy of the particular constituents considered within the Elementary Volume Element and described by the quantities defined at the micro-level (pore, grain level) should be fully represented by the macroscopic (averaged) quantities at the macro-level.

Since during a deformation process the amount of associated fluid can change, the virtual constituents form systems interchanging their masses and the corresponding continuity equations have the following form, [15],

$$\frac{\partial \overset{1}{\varrho}}{\partial t} + \operatorname{div}(\overset{1}{\varrho}\,\overset{1}{\mathbf{v}}) = g,$$
$$\frac{\partial \overset{2}{\varrho}}{\partial t} + \operatorname{div}(\overset{2}{\varrho}\,\overset{2}{\mathbf{v}}) = -g. \tag{2.1}$$

The function g is the mass exchange intensity between the free and associated

fluid and is given explicitly by the expression

$$g = \overline{\varrho}^s \frac{\overset{1}{D}}{Dt}\left[(1-\kappa)\frac{\overline{\varrho}^f}{\overline{\varrho}^s}\right].$$

The motion equations of the virtual constituents are, [14–16],

$$\begin{aligned} \overset{1}{\varrho}\frac{\overset{1}{D}\overset{1}{\mathbf{v}}}{dt} &= \operatorname{div}\overset{1}{\mathbf{T}} + \overset{1}{\varrho}\mathbf{b} + \overset{1}{\pi} + \frac{1}{2}g(\overset{2}{\mathbf{v}} - \overset{1}{\mathbf{v}}), \\ \overset{2}{\varrho}\frac{\overset{2}{D}\overset{2}{\mathbf{v}}}{dt} &= \operatorname{div}\overset{2}{\mathbf{T}} + \overset{2}{\varrho}\mathbf{b} + \overset{2}{\pi} + \frac{1}{2}g(\overset{2}{\mathbf{v}} - \overset{1}{\mathbf{v}}); \end{aligned} \tag{2.2}$$

$$\frac{D^\alpha(\,)}{Dt} = \frac{\partial(\,)}{\partial t} + \overset{\alpha}{\mathbf{v}} \cdot \operatorname{grad}(\,), \qquad \alpha = 1,2,$$

where $\overset{1}{\pi}$ and $\overset{2}{\pi}$ ($\overset{1}{\pi} = \pi = -\overset{2}{\pi}$) stand for the internal interaction forces between constituents and $\mathbf{b}$ is the external body force per unit mass. The last terms on the RHS of Eqs. (2.2) represent the coupling between virtual constituents caused by the linear momentum exchange accompanying their mass exchange.

The partial densities $\overset{1}{\varrho}$, $\overset{2}{\varrho}$ and partial Cauchy stresses $\overset{1}{\mathbf{T}}$, $\overset{2}{\mathbf{T}}$ of the virtual constituents are related to the partial densities $\overline{\varrho}^s$, $\overline{\varrho}^f$ and partial Cauchy stresses $\mathbf{T}^s$, $\mathbf{T}^f$ of the physical constituents through the following equations, [14]

$$\overset{1}{\varrho} = \overline{\varrho}^s + (1-\kappa)\overline{\varrho}^f, \qquad \overset{2}{\varrho} = \kappa\overline{\varrho}^f,, \tag{2.3}$$

$$\overset{1}{\mathbf{T}} = \mathbf{T}^s + (1-\kappa)\mathbf{T}^f, \qquad \overset{2}{\mathbf{T}} = \kappa\mathbf{T}^f, \tag{2.4}$$

where

$$\overline{\varrho}^f = f_v\varrho^f, \qquad \overline{\varrho}^s = (1-f_v)\varrho^s$$

and ϱ^f, ϱ^s stand for the effective density of fluid and porous skeleton, respectively. Stress tensors $\mathbf{T}^s$ and $\mathbf{T}^f$ are assumed to be symmetric, so that stresses $\overset{1}{\mathbf{T}}$ and $\overset{2}{\mathbf{T}}$ are also symmetric.

The local form of the internal energy balance equation can be formulated both for the individual components or for the whole solid-fluid composition. In our case we use the second one that allows us to avoid the specification of the terms describing the interchange of energy between constituents.

Accounting for the immiscibility of the physical constituents, the internal energy of the porous solid-fluid composition is considered as the sum of the internal energies of these constituents. When thermal efects are disregarded, its form is

as follows, [16, 17],

$$\overset{1}{\varrho}{}^{s}\frac{\overset{1}{D}e^{s}}{Dt}+(1-\kappa)\overline{\varrho}^{f}\frac{\overset{1}{D}e^{f}}{Dt}+\kappa\overline{\varrho}^{f}\frac{\overset{2}{D}e^{f}}{Dt} = \pi\cdot(\overset{2}{\mathbf{v}}-\overset{1}{\mathbf{v}})+\operatorname{tr}(\overset{1}{\mathbf{T}}{}^{T}\overset{1}{\mathbf{L}})+\operatorname{tr}(\overset{2}{\mathbf{T}}{}^{T}\overset{2}{\mathbf{L}}), \tag{2.5}$$

where e^f and e^s are the internal energies per unit mass for the fluid and the solid skeleton, respectively, and the tensors

$$\overset{1}{\mathbf{L}}=\operatorname{grad}\overset{1}{\mathbf{v}}, \qquad \overset{2}{\mathbf{L}}=\operatorname{grad}\overset{2}{\mathbf{v}}$$

are the velocity gradients of the virtual constituents. The subscript T stands for the transposition of the tensor.

The particular terms of the LHS of (2.5) describe the rate of changes of the internal energy in the matrix material, associated fluid and free fluid, respectively, which are balanced by the rate of work of volume and surface forces represented by the RHS terms of (2.5). Equation (2.5) will be used in further part of this work to derive the necessary constitutive relations.

3. Constitutive relations for the elastic porous solid filled with barotropic fluid. The internal equilibrium of the system

In this section we formulate constitutive relations for porous solid filled with fluid undergoing large elastic deformations. It is assumed that both, porous skeleton and fluid filling pores have elastic properties, and mutual solid-fluid interaction on the interface is that of mechanical type only. We disregard the viscous effects of the fluid confining our considerations to the pure elastic interactions.

Under the above assumptions, the fluid-saturated porous solid forms the non-dissipative system of two immiscible constituents, each of which preserves its own physical properties during a deformation process. The mechanical behaviour of such system is entirely described by the mass and linear momentum balance equations of virtual constituents (2.1) and (2.2), respectively, and appropriate constitutive relations which have to be formulated for the physical components. At the same time, the balance equation (2.5) for the internal energy of the system must be identically satisfied by constitutive relations for an arbitrary mechanical process.

We apply the internal energy balance equation (2.5) to obtain nonlinear constitutive equations for the elastic porous solid filled with fluid. Their forms will be derived from Eq. (2.5) which has to be identically satisfied by the postulated functions for the internal energies of individual physical constituents of the solid-fluid system.

Such method of derivation of the constitutive relations is analogous to the clasical approach used for the hyperelastic medium.

3.1. Constitutive postulates for the fluid and porous skeleton internal energies

The essence of the immiscibility is the fact that the physical constituents of porous solid-fluid mixture remain separated during a deformation process and then, in the local sense, each constituent shall obey the constitutive relations for that constituent alone. Therefore it is reasonable to define the internal energy for each physical constituent independently by the field quantities describing its own state of deformation.

In the case of the elastic (barotropic) fluid filling pores of the skeleton, its local state is defined by the effective fluid mass density ϱ^f. Thus, the constitutive function for the fluid internal energy can be written as follows:

$$e^f = \hat{e}^f(\varrho^f). \tag{3.1}$$

The local deformation state of the elastic porous skeleton filled with fluid, contrary to a non-porous material, is characterized by two kinds of independent variables describing, say, the internal and external skeleton deformations. The internal deformation of the skeleton is connected with a change of its geometrical pore structure and is measured by variations of the pore structure parameters: f_v and λ (or equivalently κ). Both the pore structure parameters will be used in the description as the internal state variables.

The external skeleton deformation (bulk deformation of a porous sample) is defined by the deformation gradient

$$\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}} = \frac{\partial}{\partial \mathbf{X}} \chi_k(\mathbf{X}, t), \tag{3.2}$$

where

$$\mathbf{x} = \chi_k(\mathbf{X}, t) \equiv \chi(k^{-1}(\mathbf{X}), t)$$

is the deformation function of the porous body $\mathcal{B}$ which relates the position $\mathbf{X}$ of the skeleton particle (macroscopic particle) $X \in \mathcal{B}$ in the reference configuration

$$\mathbf{X} = \mathrm{k}(X)$$

to its position $\mathbf{x}$ in the current configuration

$$\mathbf{x} = \chi(X, t).$$

The derivative in (3.2) is defined by the identity, [18],

$$\frac{\partial \chi_k}{\partial \mathbf{X}} \mathbf{D} \equiv \frac{\partial}{\partial h} \chi_k(\mathbf{X} + h\mathbf{D}, t)|_{h=0},$$

where $\mathbf{D}$ is an arbitrary vector quantity.

The deformation process in which the pore structure parameters of porous skeleton changes while the deformation gradient $\mathbf{F}$ is constant and equal to the

identity tensor is called the pure internal deformation, whereas the case when the pore parameters are constant during the deformation measured by the gradient **F** is called the pure external deformation.

Taking the above into account, the constitutive relation for the internal energy of porous elastic solid can be proposed in the following form

$$e^s = e^s_k(\mathbf{F}, f_v, \kappa). \tag{3.3}$$

Because of dependence of the deformation gradient **F** on the choice of reference configuration, the function e^s_k must also depend on the reference configuration to ensure the value of internal energy e^s to be insensitive to changes of this configuration.

It is commonly accepted that each constitutive relation should satisfy the principle of material objectivity, that is to be independent of the choice of reference frame.

The relation (3.1) as the scalar-valued scalar function satisfies this principle automatically whereas the objectivity condition for the relation (3.3) takes the form of the following identity, [18],

$$e^s_k(\mathbf{QF}, f_v, \kappa) \equiv e^s_k(\mathbf{F}, f_v, \kappa), \tag{3.4}$$

that has to be satisfied for every orthogonal transformation **Q** ($\mathbf{Q}^T = \mathbf{Q}^{-1}$) and for arbitrary values of independent variables **F**, f_v and κ.

The condition (3.4) when applied to (3.3) yields

$$e^s = \widehat{e}^s_k(\mathbf{C}, f_v, \kappa), \tag{3.5}$$

where

$$\mathbf{C} = \mathbf{F}^T\mathbf{F} \tag{3.6}$$

is the right Cauchy–Green deformation tensor.

Accounting for the fact that the independent variables **C**, f_v and κ are objective quantities, the scalar-valued function (3.5) satisfies the objectivity principle for arbitrary form of $\widehat{e}^s_k$.

The representation (3.5) is the general (nonlinear) constitutive relation for the internal energy of porous solid of anisotropic elastic properties and the isotropic pore structure.

The relation (3.5) is not the only form that represents the internal energy of porous skeleton. We can derive two other, equivalent forms replacing the current volume porosity f_v with the porosity f^L_v (say Lagrangean porosity) defined by

$$f^L_v = f_v J, \tag{3.7}$$

where

$$J = \det(\mathbf{F}) = [\det(\mathbf{C})]^{1/2}, \tag{3.8}$$

or with the skeleton density ϱ^s related to the porosity f_v by the skeleton continuity equation

$$\varrho^s(1 - f_v)J = \varrho_0^s(1 - f_v^0) = \overline{\varrho}_0^s . \tag{3.9}$$

In (3.9) ϱ_0^s and f_v^0 stand for the values of ϱ^s and f_v, respectively, in the reference configuration.

The porosity f_v^L is the ratio of the pore volume contained in the Elementary Volume Element (macro-particle) of the porous body in the current configuration to the total volume of the same Volume Element in the reference configuration. Variation of this quantity, contrary to variation of f_v, is the local absolute measure of the change of a pore volume during a deformation process.

Similarly, the skeleton mass density variation may be considered as the local absolute measure of the change of a skeleton volume. In such a case the density ϱ^s, similarly to the porosity f_v, plays the role of internal state variable.

These are good points in application of f_v^L and ϱ^s to the skeleton internal energy formulation. We obtain

$$e^s = \hat{e}_k^{sf}(\mathbf{C}, f_v^L, \kappa), \tag{3.10}$$

$$e^s = \hat{e}_k^{s\varrho}(\mathbf{C}, \varrho^s, \kappa), \tag{3.11}$$

where the constitutive functions $\hat{e}_k^{sf}$ and $\hat{e}_k^{s\varrho}$ are defined by the following identities

$$\hat{e}_k^{sf}(\mathbf{C}, f_v^L, \kappa) \equiv \hat{e}_k^s(\mathbf{C}, f_v^L/J, \kappa), \tag{3.12}$$

$$\hat{e}_k^{s\varrho}(\mathbf{C}, \varrho^s, \kappa) \equiv \hat{e}_k^s(\mathbf{C}, 1 - \overline{\varrho}_0^s/\varrho^s J, \kappa). \tag{3.13}$$

The constitutive functions for the fluid (3.1), and for the porous skeleton (3.5) or its alternative forms (3.10) and (3.11) entirely describe the energetic state of elastic porous solid filled with fluid and undergoing finite deformations.

3.2. Constitutive relations for stresses. The condition of mechanical internal equilibrium

To establish constitutive stress-strain relations for each constituent of the fluid-porous solid immiscible mixture, and relations describing mutual solid-fluid interaction, we apply the approach characteristic for the hyperelastic medium. We introduce the relations (3.1) and (3.5) to the energy balance equation (2.5) for the whole porous solid-fluid mixture which has to be identically satisfied for

an arbitrary mechanical process. Using, moreover, the continuity equation $(2.1)_2$, Eq. (2.5) can be written as follows

$$\text{(3.14)}\qquad \operatorname{tr}\left\{\left[2\overline{\varrho}^s\mathbf{F}\frac{\partial\hat{e}^s_k}{\partial\mathbf{C}}\mathbf{F}^T-\frac{d\hat{e}^f}{d\varrho^f}\left(\varrho^f\right)^2 f_v(1-\kappa)\mathbf{I}-\overset{1}{\mathbf{T}}\right]\overset{1}{\mathbf{L}}\right\}$$

$$-\operatorname{tr}\left\{\left[\frac{d\hat{e}^f}{d\varrho^f}\left(\varrho^f\right)^2\kappa f_v\mathbf{I}+\overset{2}{\mathbf{T}}\right]\overset{2}{\mathbf{L}}\right\}+\left[\overline{\varrho}^s\frac{\partial\hat{e}^s_k}{\partial f_v}-\frac{d\hat{e}^f}{d\varrho^f}\left(\varrho^f\right)^2\right]\frac{\overset{1}{D}f_v}{Dt}$$

$$-\left[\frac{d\hat{e}^f}{d\varrho^f}\left(\varrho^f\right)^2\operatorname{grad}(\lambda)+\boldsymbol{\pi}\right]\cdot(\overset{1}{\mathbf{v}}-\overset{2}{\mathbf{v}})+\frac{\partial\hat{e}^s_k}{\partial\kappa}\frac{\overset{1}{D}\kappa}{Dt}=0.$$

Equation (3.14) is the linear function of the independent quantities

$$\overset{1}{\mathbf{L}},\ \overset{2}{\mathbf{L}},\ (\overset{1}{\mathbf{v}}-\overset{2}{\mathbf{v}}),\ \frac{\overset{1}{D}f_v}{Dt},\ \frac{\overset{1}{D}\kappa}{Dt},$$

for an arbitrary mechanical process in the body. Since these quantities can assume arbitrary values, equation (3.14) will be identically satisfied if the corresponding coefficients are equal to zero. Defining the quantity

$$\text{(3.15)}\qquad p^f=\left(\varrho^f\right)^2\frac{d\hat{e}^f}{d\varrho^f}$$

which is considered as the effective pore pressure, from (3.14) we have

$$\text{(3.16)}\qquad \overset{1}{\mathbf{T}}+(1-\kappa)f_v p^f\mathbf{I}=2\overline{\varrho}^s\mathbf{F}\frac{\partial\hat{e}^s_k}{\partial\mathbf{C}}\mathbf{F}^T,$$

$$\text{(3.17)}\qquad \overset{2}{\mathbf{T}}=-\kappa f_v p^f\mathbf{I},$$

$$\text{(3.18)}\qquad p^f=\overline{\varrho}^s\frac{\partial\hat{e}^s_k}{\partial f_v},$$

$$\text{(3.19)}\qquad \frac{\partial\hat{e}^s_k}{\partial\kappa}=0,$$

$$\text{(3.20)}\qquad \boldsymbol{\pi}=-p^f\operatorname{grad}(\lambda).$$

Condition (3.20) describes the force exerted on the solid skeleton by fluid filling its pores. From (3.20) it is seen that, despite the lack of fluid viscosity effects in the considerations, the solid-fluid interface interaction force does exist due to the nonhomogeneity of the skeleton pore structure.

The expression (3.19) is the necessary condition for minimum of the skeleton internal energy function at constant strain tensor $\mathbf{C}$ and volume porosity f_v. If, additionally, the sufficient condition is satisfied, i.e.

$$\frac{\partial^2\hat{e}^s_k}{\partial\kappa^2}>0,$$

then Eq. (3.19) indicates that the saturated porous body undergoes a deformation process in such a way that the parameter κ takes values for which the skeleton internal energy has a minimum, (Fig. 1). Therefore, the condition (3.19) may be treated as the implicit equation of variation of the κ-parameter during a deformation process.

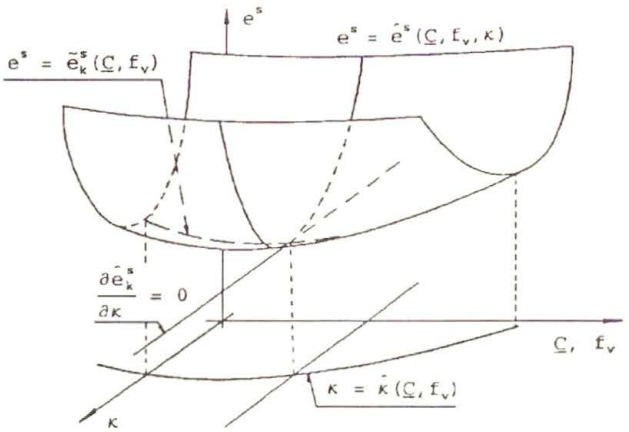

FIG. 1. Illustration of the changes of the pore structure parameter κ during the deformation process.

Assuming that (3.19) is the smooth function of κ, it can be rewritten (at least locally) in the explicit form

$$\kappa = \hat{\kappa}(\mathbf{C}, f_v), \tag{3.21}$$

that has to satisfy the condition

$$\frac{\partial \hat{e}^s_k}{\partial \kappa}(\mathbf{C}, f_v, \kappa)\Big|_{\kappa=\hat{\kappa}(\mathbf{C},f_v)} \equiv 0.$$

Equation (3.21) offers the possibility of exclusion of the κ-parameter from the set of independent variables defining the skeleton internal energy function. In such a case, the constitutive functions (3.5), (3.10) and (3.11) take the form

$$\tilde{e}^s_k(\mathbf{C}, f_v) \equiv \hat{e}^s_k(\mathbf{C}, f_v, \hat{\kappa}(\mathbf{C}, f_v)), \tag{3.22}$$

$$\tilde{e}^{sf}_k(\mathbf{C}, f^L_v) \equiv \hat{e}^{sf}_k(\mathbf{C}, f^L_v, \hat{\kappa}^L(\mathbf{C}, f^L_v)), \tag{3.23}$$

$$\tilde{e}^{s\varrho}_k(\mathbf{C}, \varrho^s) \equiv \hat{e}^{s\varrho}_k(\mathbf{C}, \varrho^s, \hat{\kappa}^\varrho(\mathbf{C}, \varrho^s)), \tag{3.24}$$

where

$$\hat{\kappa}^L(\mathbf{C}, f^L_v) \equiv \hat{\kappa}(\mathbf{C}, f^L_v/J), \tag{3.25}$$

$$\hat{\kappa}^\varrho(\mathbf{C}, \varrho^s) \equiv \hat{\kappa}(\mathbf{C}, 1 - \overline{\varrho}^s_0/\varrho^s J). \tag{3.26}$$

From the condition (3.19), due to identities (3.12) and (3.13), we obtain

$$\frac{\partial \widehat{e}_k^{sf}}{\partial \kappa} = \frac{\partial \widehat{e}_k^{s\varrho}}{\partial \kappa} = 0,$$

and consequently (3.12) and (3.13) reduce to the following identities, respectively

$$\widetilde{e}_k^{sf}(\mathbf{C}, f_v^L) \equiv \widetilde{e}_k^s(\mathbf{C}, f_v^L/J), \tag{3.27}$$

$$\widetilde{e}_k^{s\varrho}(\mathbf{C}, \varrho^s) \equiv \widetilde{e}_k^s(\mathbf{C}, 1 - \overline{\varrho}_0^s/\varrho^s J). \tag{3.28}$$

From the above consideration it is seen that the condition (3.19) should be treated not only as the condition of independence of the skeleton internal energy of the parameter κ, but also as the equation of changes of this parameter. Such interpretation of the condition (3.19) is supported by the analysis given in Sec. 4.3 where the skeleton mass density changes of non-saturated porous material is obtained as a particular case of the internal equilibrium condition (for $p^f = 0$) in the same form as the condition (3.19).

Now, taking into account relations (3.22), (3.19) and (2.4), Eqs. (3.16) – (3.18) can be written in the following form

$$\mathbf{T}^s = 2\overline{\varrho}^s \mathbf{F} \frac{\partial \widetilde{e}_k^s}{\partial \mathbf{C}} \mathbf{F}^T, \tag{3.29}$$

$$\mathbf{T}^f = -f_v p^f \mathbf{I}, \tag{3.30}$$

$$p^f = \overline{\varrho}^s \frac{\partial \widetilde{e}_k^s}{\partial f_v}. \tag{3.31}$$

Equations (3.29) and (3.30) (together with (3.15)) are the constitutive relations for the partial stresses of an elastic porous skeleton and of a barotropic fluid filling its pores, respectively. Equation (3.31) relates the pore fluid pressure p^f with the deformation tensor $\mathbf{C}$ and the volume porosity f_v; quantities which define the state of deformation of the porous skeleton. It is the condition for internal mechanical equilibrium between porous skeleton and fluid filling its pores. As will be shown in Sec. 4.2, this equation does not appear in the description in the case when the skeleton material is incompressible.

The condition (3.31) and the skeleton continuity equation (3.9) define (optionally) changes of the two internal parameters: the volume porosity f_v and the skeleton mass density ϱ^s.

From the above considerations it is seen that the constitutive functions

$$\frac{\partial \widetilde{e}_k^s}{\partial \mathbf{C}}; \qquad \frac{\partial \widetilde{e}_k^s}{\partial f_v}$$

in relations (3.29) and (3.31) are defined by the mechanical properties of porous skeleton and do not depend on the properties of the fluid filling its pores. In

such constitutive formulation the mechanical coupling between fluid and porous skeleton appears only in the internal equilibrium condition (3.31) where the fluid pore pressure p^f is present.

It is worth to note that relation similar to (3.31) was considered by KENYON in his paper [13] on the equilibrium theory of the solid-fluid mixture. He introduced constitutive postulate relating the volume porosity f_v to the fluid bulk density $\overline{\varrho}^f$ and $J = \det(\mathbf{F})$, without any physical motivation. Such relation can be considered as a particular case of the equation (3.31) that has the resonable physical interpretation.

The form of the constitutive equation (3.29) and the internal equilibrium condition (3.31) will change if the Lagrangean porosity f_v^L or the skeleton mass density ϱ^s, instead of the current volume porosity f_v, is used in the expression for skeleton internal energy. In the first case, after differentiation of (3.27) with respect to $\mathbf{C}$ and f_v^L, we obtain

$$\frac{\partial \widetilde{e}_k^{sf}}{\partial \mathbf{C}} = \frac{\partial \widetilde{e}_k^{s}}{\partial \mathbf{C}} - \frac{1}{2} f_v \frac{\partial \widetilde{e}_k^{s}}{\partial f_v} \mathbf{C}^{-T},$$

$$\frac{\partial \widetilde{e}_k^{sf}}{\partial f_v^L} = \frac{1}{J} \frac{\partial \widetilde{e}_k^{s}}{\partial f_v}.$$

The above relations, when applied to (3.29) and (3.31) yield

$$\mathbf{T}^s - f_v p^f \mathbf{I} = 2\overline{\varrho}^s \mathbf{F} \frac{\partial \widetilde{e}_k^{sf}}{\partial \mathbf{C}} \mathbf{F}^T, \tag{3.32}$$

$$p^f = \overline{\varrho}_0^s \frac{\partial \widetilde{e}_k^{sf}}{\partial f_v^L}. \tag{3.33}$$

Taking Eq. (3.30) into account, we can conclude that Eq. (3.32) is the constitutive relation for the total stress

$$\mathbf{T} = \mathbf{T}^s + \mathbf{T}^f$$

in the solid-fluid composition.

Equations (3.32) and (3.33) coincide with the equations derived in another way by BIOT [4].

In the second case, after differentiation of (3.28) with respect to $\mathbf{C}$ and ϱ^s we have

$$\frac{\partial \widetilde{e}_k^{s\varrho}}{\partial \mathbf{C}} = \frac{\partial \widetilde{e}_k^{s}}{\partial \mathbf{C}} - \frac{1}{2}(1 - f_v) \frac{\partial \widetilde{e}_k^{s}}{\partial f_v} \mathbf{C}^{-T},$$

$$\frac{\partial \widetilde{e}_k^{s\varrho}}{\partial \varrho^s} = \frac{\overline{\varrho}_0^s}{(\varrho^s)^2 J} \frac{\partial \widetilde{e}_k^{s}}{\partial f_v}.$$

These relations, when applied to (3.29) and (3.31), give

$$\mathbf{T}^s = -(1 - f_v)p^f \mathbf{I} + 2\overline{\varrho}^s \mathbf{F} \frac{\partial \widetilde{e}_k^{s\varrho}}{\partial \mathbf{C}} \mathbf{F}^T, \tag{3.34}$$

$$\frac{p^f}{(\varrho^s)^2} = \frac{\partial \widetilde{e}_k^{s\varrho}}{\partial \varrho^s}. \tag{3.35}$$

From (3.34) it is seen that the stress in the skeleton is composed of two parts; the first part

$$-(1 - f_v)p^f \mathbf{I}$$

is due to the presence of the pore fluid in the skeleton pores, and the second part

$$2\overline{\varrho}^s \mathbf{F} \frac{\partial \widetilde{e}_k^{s\varrho}}{\partial \mathbf{C}} \mathbf{F}^T$$

is due to the deformation of porous solid.

Introducing the effective stress tensor in the skeleton by the definition

$$\mathbf{T}^{*s} = \mathbf{T}^s/(1 - f_v), \tag{3.36}$$

from (3.34) we obtain the constitutive equation for the effective stresses in the following form

$$\mathbf{T}^{*s} = -p^f \mathbf{I} + 2\varrho^s \mathbf{F} \frac{\partial \widetilde{e}_k^{s\varrho}}{\partial \mathbf{C}} \mathbf{F}^T. \tag{3.37}$$

It should be noted that the constitutive relation (3.37) as well as the internal equilibrium condition (3.35) and the constitutive equation (3.15) for the effective pore fluid pressure, do not depend explicitly on the volume porosity f_v. It limits the number of the quantities appearing in these equations thus simplifying their forms.

In such a case the condition (3.35) can be considered as the equation describing variations of the skeleton mass density ϱ^s (particularly in the case of non-saturated pores ($p^f = 0$); see Sec. 4.3). Then the skeleton continuity equation (3.12) plays the role of equation for the volume porosity changes.

The constitutive relations (3.15) and (3.37), the condition (3.35) and Eq. (3.21) form a complete set of the constitutive equations for the fluid-saturated porous solid of elastic mechanical properties (in general anisotropic) and of the isotropic, initially homogeneous pore structure. This set of equations is supplemented by the relation (3.20) describing the solid-fluid interface interaction force.

4. Constitutive relations. Special cases

In this section the nonlinear constitutive relations for the practically important fluid-saturated porous media of simplified mechanical properties are analysed. We

consider three particular cases that concern porous solids: of isotropic mechanical properties, of an incompressible skeleton material and the case when porous solid is not saturated with fluid.

4.1. Saturated porous solid with isotropic properties of skeleton

The constitutive relations (3.29), (3.31) and (3.21) will describe elastic properties of the isotropic porous skeleton filled with fluid if all of them are isotropic relations. This can be achieved by imposing the isotropy condition on the relation (3.4) for the skeleton internal energy.

This condition takes the following form, [18],

$$\hat{e}_k^s(\mathbf{Q}\,\mathbf{C}\,\mathbf{Q}^T, f_v, \kappa) \equiv \hat{e}_k^s(\mathbf{C}, f_v, \kappa), \tag{4.1}$$

which has to be fulfilled for all orthogonal transformations $\mathbf{Q}$ ($\mathbf{Q}^T = \mathbf{Q}^{-1}$) and all values of independent variables $\mathbf{C}$, f_v, κ. Such requirement imposed on the fluid internal energy (3.1) is satisfied identically.

The condition (4.1) shows that the skeleton internal energy is an invariant of the deformation tensor $\mathbf{C}$ and thus, it can be considered as a function of the invariants of $\mathbf{C}$.

In such case Eq. (3.5) becomes

$$e^s = \hat{e}_k^s(\mathbf{C}, f_v, \kappa) \equiv \overline{e}_k^s(I_1^C, I_2^C, I_3^C, f_v, \kappa), \tag{4.2}$$

where

$$I_1^C = \mathrm{tr}\,(\mathbf{C}), \qquad I_2^C = \frac{1}{2}(\mathrm{tr}^2(\mathbf{C}) - \mathrm{tr}\,(\mathbf{C}^2)), \qquad I_3^C = \det(\mathbf{C}),$$

are the principal invariants of the tensor $\mathbf{C}$.

Relation (4.2) is a general form of the constitutive equation for the irternal energy of the skeleton with isotropic mechanical properties.

Isotropy of (4.2), due to the relation (3.19), results in isotropy of Eq. (3.21) and consequently, due to the identity (3.22), leads to the isotropy of the stress-strain relation (3.29) and of the internal equilibrium condition (3.31).

We obtain

$$\kappa = \overline{\kappa}(I_1^B, I_2^B, I_3^B, f_v), \tag{4.3}$$

$$\mathbf{T}^s = 2\overline{\varrho}^s \left\{ I_3^B E_3^B \mathbf{I} + (E_1^B + I_1^B E_2^B)\mathbf{B} - E_2^B \mathbf{B}^2 \right\}, \tag{4.4}$$

$$p^f = \overline{\varrho}^s \frac{\partial \overline{\overline{e}}_k^s}{\partial f_v}, \tag{4.5}$$

where

$$\overline{\overline{e}}_k^s(I_1^B, I_2^B, I_3^B, f_v) \equiv \overline{e}_k^s(I_1^B, I_2^B, I_3^B, f_v, \overline{\kappa}(I_1^B, I_2^B, I_3^B, f_v)) \tag{4.6}$$

and the right Cauchy–Green deformation tensor C is replaced with the left deformation tensor

$$\mathbf{B} = \mathbf{F}\mathbf{F}^T,$$

the invariants of which are identical

$$I_\alpha^B = I_\alpha^C, \qquad \alpha = 1,2,3.$$

The quantity E_α^B stands for

$$E_\alpha^B = \frac{\partial \bar{\bar{e}}_k^s}{\partial I_\alpha^B}, \qquad \alpha = 1,2,3.$$

Relations (3.15), (4.3)–(4.5) form the set of the constitutive equations describing the mechanical behaviour of the isotropic, elastic porous solid-fluid composition.

Using the relations (3.7) and (3.9) we can derive two other, equivalent sets of constitutive relations in which, instead of the volume porosity f_v, the mass density ϱ^s or Lagrangean porosity f_v^L are used as idependent constitutive variable.

4.2. Saturated porous medium with incompressible skeleton material

Incompressibility of the porous skeleton material is defined by the condition

$$\varrho^s = \varrho_0^s. \tag{4.7}$$

It is a kinematic constraint confining the skeleton motion during its deformation. This condition, at the same time, is the special (trivial) case of the equation describing changes of the skeleton mass density and replaces in this role the condition of internal equilibrium for the solid-fluid composition.

Taking (4.7) into account, the skeleton continuity equation (3.9) reduces to the relation

$$f_v = 1 - (1 - f_v^0)/J \tag{4.8}$$

that uniquely defines the volume porosity changes by means of the skeleton deformation gradient F.

The assumption (4.7) and relation (4.8) eliminate the density ϱ^s and porosity f_v or f_v^L from the set of independent variables describing the internal energy of the skeleton. We have

$$e^s = \overset{*}{e}{}^s_k(\mathbf{C}, \kappa). \tag{4.9}$$

Now, requiring the balance equation of the internal energy (2.5) to be identically satisfied by relations (3.1) and (4.9), one can find the constitutive equations for the interface interaction force and the effective fluid stresses identical with the relations (3.20) and (3.15), respectively. At the same time, the function of the

κ-parameter variation and the constitutive relation for the skeleton effective stress take forms similar to (3.19) and (3.37), respectively, i.e.

$$\frac{\partial \overset{*}{e}{}^{s}_{k}}{\partial \kappa} = 0, \tag{4.10}$$

$$\mathbf{T}^{*s} = -p^{f}\mathbf{I} + 2\varrho_0^s \mathbf{F}\frac{\partial \widetilde{\overset{*}{e}}{}^{s}_{k}}{\partial \mathbf{C}}\mathbf{F}^{T}, \tag{4.11}$$

where

$$\widetilde{\overset{*}{e}}{}^{s}_{k}(\mathbf{C}) = \overset{*}{e}{}^{s}_{k}(\mathbf{C}, \widetilde{\kappa}(\mathbf{C}))$$

and

$$\kappa = \widetilde{\kappa}(\mathbf{C})$$

is the explicit form of the relation (4.10) defined by the identity

$$\frac{\partial \overset{*}{e}{}^{s}_{k}}{\partial \kappa}(\mathbf{C}, \kappa)\Big|_{\kappa=\widetilde{\kappa}(\mathbf{C})} \equiv 0.$$

The term in (4.11), related to the fluid pressure p^f represents the stresses in the skeleton caused by the presence of the fluid in pores. These stresses contribute to the pore fluid energy during the skeleton deformation. However, due to incompressibility of the skeleton material, they do not influence the skeleton internal energy.

It should be pointed out that the set of constitutive relations mentioned above does not contain the condition of internal, mechanical equilibrium for the considered solid-fluid composition. This results from the fact that the skeleton mass density has been excluded from the set of independent variables.

4.3. The non-saturated porous solid

Constitutive description of non-saturated, elastic porous solids can be obtained from the constitutive relation (3.37) and the internal equilibrium condition (3.35) through the assumption that the effective fluid pressure is equal to zero ($p^f = 0$). In such case, we have

$$\mathbf{T}^{*s} = 2\varrho^s \mathbf{F}\frac{\partial \widetilde{e}^{s\varrho}_{k}}{\partial \mathbf{C}}\mathbf{F}^{t}, \tag{4.12}$$

$$\frac{\partial \widetilde{e}^{s\varrho}_{k}}{\partial \varrho^s} = 0. \tag{4.13}$$

The form of Eq. (4.13) is similar to that of (3.19). Therefore, we conclude that during a deformation process of elastic, non-saturated porous solid, the skeleton mass density ϱ^s takes values for which the skeleton internal energy has a minimum. This additionally justifies our interpretation of (3.31) as the internal equilibrium condition between the pore fluid and skeleton.

If $\tilde{e}_k^{s\varrho}$ is a smooth function, equation (4.13) can be rewritten in the form

$$\varrho^s = \tilde{\varrho}^s(\mathbf{C}) \tag{4.14}$$

that explicitly describes the skeleton mass density changes. Then the identity

$$\frac{\partial \tilde{e}_k^{s\varrho}}{\partial \varrho^s}(\mathbf{C}, \varrho^s)\Big|_{\varrho^s = \tilde{\varrho}^s(\mathbf{C})} \equiv 0$$

is satisfied.

Introducing (4.14) into the skeleton the mass continuity equation (3.9) we obtain the equation for the changes of volume porosity

$$f_v = 1 - \overline{\varrho}_0^s / J\tilde{\varrho}^s(\mathbf{C}). \tag{4.15}$$

Moreover, the equation (4.14) eliminates the skeleton mass density from the set of independent variables describing the internal energy of the skeleton. We have

$$e^s = \tilde{e}_k^s(\mathbf{C}) \equiv \tilde{e}_k^{s\varrho}(\mathbf{C}, \tilde{\varrho}^s(\mathbf{C})). \tag{4.16}$$

Thus, the constitutive relation (4.12), when (4.13) is taken into account, reduces to the form

$$\mathbf{T}^{*s} = 2\varrho^s \mathbf{F} \frac{\partial \tilde{e}_k^s}{\partial \mathbf{C}} \mathbf{F}^T, \tag{4.17}$$

similar to that for a non-porous solid.

5. Final remarks

Nonlinear constitutive relations for the fluid-saturated porous solid immiscible mixture undergoing pure mechanical large deformations have been developed.

Considerations have been based on the balance equation for the internal energy of the whole composition which was required to be satisfied identically by the internal energy constitutive functions postulated independently for individual components.

General constitutive relations for partial stresses in an anisotropic, elastic skeleton and barotropic fluid have been formulated and the internal equilibrium condition for the composition has been established. This condition relates the pore fluid pressure to independent variables describing the state of porous skeleton and it does not appear in the constitutive description when the skeleton material is incompressible. Also the constitutive relations for the medium with simplified physical properties have been discussed.

References

1. R.J. ATKIN, *Constitutive theory for a mixture of isotropic elastic solid and non-Newtonian fluid*, ZAMP, **18**, 803–825, 1967.
2. E. AIFANTIS, W. GEBERICH and D. BESKOS, *Diffusion equation for mixtures of an elastic fluid and elastic solid*, Arch. Mech., **29**, 2, 1977.
3. D.E. BEDFORD and D.S. DRUMHELLER, *Theories of immiscible and structured mixtures*, Int. J. Engng. Sci., **21**, 8, 863–960, 1983.
4. M.A. BIOT, *Theory of finite deformations of porous solids*, Ind. Univ. Math. J., **21**, 7, 597–620, 1972.
5. R.M. BOWEN, *Mixtures and EM field theories*, [in:] Continuum Physics, Vol. III, A.C. ERINGEN [Ed.], Academic Press, 1–127, 1976.
6. R.M. BOWEN, *Compressible porous media models by use of the theory of mixtures*, Int. J. Engng. Sci., **20**, 697–735, 1982.
7. M.J. CROCHET and P.M. NAGHDI, *On constitutive equations for flow of fluid through an elastic solid*, Int. J. Engng. Sci., **4**, 383–401, 1966.
8. D.S. DRUMHELLER, *The theoretical treatment of a porous solid using a mixture theory*, Int. J. Solids Structures, **14**, 441–456, 1978.
9. A.E. GREEN and P.M. NAGHDI, *A dynamical theory of interacting continua*, Int. J. Engng. Sci., **3**, 231–241, 1965.
10. A.E. GREEN and P.M. NAGHDI, *On basic equations for mixtures*, Quart. J. Appl. Math., **22**, 427–438, 1969.
11. A.E. GREEN and P.M. NAGHDI, *The flow of fluid through an elastic solid*, Acta Mech., **9**, 329, 1970.
12. D.E. KENYON, *Thermostatics of solid-fluid mixtures*, Arch. Rat. Mech. Analysis, **62**, 117–130, 1976.
13. D.E. KENYON, *The theory of incompressible solid-fluid mixture*, Arch. Rat. Mech. Analysis, **62**, 131–147, 1976.
14. J. KUBIK, *On internal coupling in dynamic equations of fluid-saturated porous solid*, Int. J. Engng. Sci., **24**, 981–989, 1986.
15. J. KUBIK and M. CIESZKO, *On internal forces in a fluid-saturated porous medium* [in Polish], Engng. Trans., **35**, 1, 55–70, 1987.
16. J. KUBIK, *Pore structure in dynamic behaviour of saturated materials*, Transport in Porous Media, **9**, 15–24, 1992.
17. J. KUBIK, *On the linear momentum balance for immiscible fluid-porous solid mixture*, Studia Geotechnica et Mechanica, **16**, 1-2, 31–41, 1994.
18. D.C. LEIGH, *Nonlinear continuum mechanics*, McGraw-Hill Book Company, New York 1968.
19. L.W. MORLAND, *A simple constitutive theory for a fluid-saturated porous solid*, J. Geophys. Res., **77**, 890–900, 1972.
20. L.W. MORLAND, *Flow of viscous fluids through a porous deformable matrix*, Surveys in Geophysics, **13**, 209–268, 1992.
21. J.H. PREVOST, *Mechanics of continuous porous media*, Int. J. Engng. Sci., **18**, 787–800, 1980.
22. G. SZEFER, *Nonlinear problems of consolidation theory*, Problems de Reologie, Symposium Franco-Polonaise, Kraków 1977, PWN, Warszawa 1978.
23. C. TRUESDEL and R.A. TOUPIN, *The classical field theories*, [in:] Handbuch der Physik, S. FLÜGGE [Ed.], III/1, Springer-Verlag, Berlin 1960.

POLISH ACADEMY OF SCIENCES
INSTITUTE OF FUNDAMENTAL TECHNOLOGICAL RESEARCH

Received March 14, 1996.

Arch. Mech., 48, 5, pp. 911–923, Warszawa 1996

Constitutive relations and internal equilibrium condition for fluid-saturated porous solids Linear description

M. CIESZKO and J. KUBIK (POZNAŃ)

USING THE NONLINEAR THEORY established in the paper [5], the constitutive relations for small deformations of the fluid-saturated porous solid are derived. It is assumed that the elastic properties of porous skeleton are non-isotropic while the skeleton pore structure is isotropic. Fluid filling pores is assumed to be barotropic. Such approach made it possible to construct the consistent linear description of elastic behaviour of porous medium in which all material constants are precisely defined and represent mechanical properties of individual constituents. It is shown that the pure elastic properties of fluid-filled anisotropic skeleton are characterized by 36 material constants and reduces to 7 constants for the isotropic case, and to 4 constants when the skeleton is isotropic and its material is incompressible. In each considered case, the only one material constant characterizes mechanical properties of the pore fluid whereas the remaining constants characterize elastic properties of porous skeleton.

1. Introduction

The purpose of this paper is to formulate the linear constitutive theory for fluid-saturated porous elastic solid using as a starting point the results of nonlinear theory established in [5], where the special attention was paid to the consequences of the constituent immiscibility in such a medium.

The elastic properties of porous skeleton are assumed to be anisotropic while its pore structure is isotropic and is described by two scalar parameters: volume porosity f_v and the structural permeability λ (or, equivalently, by parameter $\kappa = \lambda/f_v$), [10].

The comprehensive constitutive macro-description of mechanical behaviour of fluid-saturated porous solids during a deformation process – also within the linear theory – should include all characteristic features resulting from the fact of immiscibility of physical constituents. Therefore, in the case of a solid-fluid elastic system, apart from the constitutive relations for the skeleton stresses and the pore fluid pressure, the additional relations for pore structure parameters and effective skeleton mass density changes must be established. The formulation of such relations should provide clear physically motivated interpretation of interactions between the porous skeleton and pore fluid and to give precisely defined material constants with clear physical interpretation.

In the commonly used linear theory of fluid-saturated porous solids developed by BIOT [1–3], the problem of changes of pore structure parameters and skeleton mass density do not appear. The Biot constitutive relations derived from

the internal energy function postulated for the whole aggregate does not provide simple interpretation of mechanical couplings between constituents, and the corresponding material constants characterizing these couplings are complex [3, 7, 8]. The above difficulties are unfortunately not overcome in works in which the linear constitutive relations are obtained from their nonlinear form formulated with the use of the principle of equipresence (see e.g. [4, 6, 9, 12]).

In our analysis of the porous solid deformation process the notions of the external (bulk) deformation defined by the right Cauchy-Green deformation tensor $\mathbf{C}$ (the infinitesimal strain tensor $\mathbf{E}$ in the linear case) and of the internal deformation measured by the change of the effective skeleton mass density ϱ^s (or, equivalently, volume porosity f_v) are used. Such approach enables one to obtain the linear constitutive description of elastic solid-fluid composition in which the mechanical coupling between the deformable skeleton and pore fluid appearing in constitutive relations, and the corresponding material constants are well defined and have clear physical meaning.

In the paper, it is shown that the elastic properties of fluid-filled anisotropic skeleton with isotropic pore structure are characterized by 36 material constants and reduces to 7 constants for the isotropic case and to 4 constants when the skeleton is isotropic and its material is incompressible. It should be pointed out that in each considered case, only one material constant characterizes the mechanical properties of the pore fluid whereas the remaining constants characterize elastic properties of the porous skeleton.

2. Initial set of constitutive relations for an elastic porous skeleton filled with barotropic fluid

The starting point for our considerations is the macroscopic nonlinear constitutive description of an elastic porous skeleton filled with barotropic fluid, formulated in the former paper [5]. It is assumed that the skeleton pore structure is isotropic and characterized by two scalar parameters: the volume porosity f_v and structural permeability λ (or, equivalently, parameter $\kappa = \lambda / f_v$). From different forms of the constitutive relations derived for the elastic porous skeleton in this discussion we use that one in which the independent variables are the effective mass density ϱ^s and the right Cauchy-Green deformation tensor

$$\mathbf{C} = \mathbf{F}^T \mathbf{F},$$

where $\mathbf{F}$ is the porous solid deformation gradient and the superscript T stands for transposition of the tensor.

In such a case the complete set of constitutive equations comprises:

- the constitutive stress-strain relation for the porous skeleton

$$\mathbf{T}^{*s} = -p^f \mathbf{I} + 2\varrho^s \mathbf{F} \frac{\partial \tilde{e}^s}{\partial \mathbf{C}} \mathbf{F}^T; \tag{2.1}$$

• the internal, mechanical equilibrium condition for the porous solid-fluid aggregate

$$\frac{p^f}{(\varrho^s)^2} = \frac{\partial \tilde{e}^s}{\partial \varrho^s}; \tag{2.2}$$

• the equation for the κ-parameter variation

$$\kappa = \tilde{\kappa}(\mathbf{C}, \varrho^s); \tag{2.3}$$

• the constitutive relation for the barotropic fluid

$$\frac{p^f}{(\varrho^f)^2} = \frac{\partial \tilde{e}^f}{\partial \varrho^f}. \tag{2.4}$$

In the above equations, the constitutive relations

$$e^s = \tilde{e}^s(\mathbf{C}, \varrho^s), \qquad e^f = \tilde{e}^f(\varrho^f)$$

represent the internal energies of the porous skeleton and fluid, respectively, and $\mathbf{T}^{*s}$ is the effective Cauchy stress tensor related to the partial stress tensor $\mathbf{T}^s$ by expression

$$\mathbf{T}^s = (1 - f_v)\mathbf{T}^{*s}.$$

The quantities p^f and ϱ^f stand for the fluid pore pressure and its mass density, respectively.

The derivative in (2.1) is defined by the identity, [11]

$$\frac{\partial \tilde{e}^s}{\partial \mathbf{C}} \cdot \mathbf{D} = \frac{\partial}{\partial h} \tilde{e}^s(\mathbf{C} + h\mathbf{D}, \varrho^s)\Big|_{h=0}, \tag{2.5}$$

where $\mathbf{D}$ is an arbitrary second order symmetric tensor.

Equations (2.1)–(2.4) have been derived from the internal energy balance equation of porous solid-fluid aggregate which was required to be identically satisfied by the independent internal energy functions postulated for the physical constituents and an arbitrary nondissipative mechanical process. Such approach takes into account the fact of immiscibility of the physical components that provides preservation of their individual physical properties during a deformation process.

The constitutive functions in Eqs. (2.1)–(2.3) are related to the elastic properties of the porous skeleton and do not depend explicitly on the volume porosity f_v. It reduces the number of quantities appearing in these equations simplifying their forms. Therefore the internal equilibrium condition (2.2), that relates the quantities p^f, $\mathbf{C}$ and ϱ^s, may be considered as the equation describing variations

of the skeleton mass density ϱ^s during a deformation process. Consequently, variations of the volume porosity parameter f_v are defined by the continuity equation for the skeleton

$$(1 - f_v)\varrho^s \det(\mathbf{F}) = (1 - f_v^0)\varrho_0^s, \tag{2.6}$$

where quantities f_v^0 and ϱ_0^s are the values of f_v and ϱ^s, respectively, in the reference configuration.

All the three quantities: ϱ^s, f_v and κ can not be controlled directly by the boundary conditions and in this sense they play the role of internal parameters.

3. Linear constitutive relations for elastic fluid-filled porous medium

We are interested in the linear constitutive description of elastic porous solid filled with barotropic fluid undergoing small deformations. We consider deformations around the equilibrium state of the medium that is assumed to be its reference configuration. The linear constitutive relations are derived by linearization of the general nonlinear equations (2.1)–(2.4).

Since the fluid does not have the natural stress-free states, both physical constituents (fluid and porous solid) are in some initial stress state (in any arbitrary reference configuration). Assuming that the medium in the reference configuration is homogeneous, its initial state will be characterized by the following set of quantities:

$$\mathbf{T}_0^{*s}, \quad \varrho_0^s, \quad f_v^0, \quad \kappa_0, \quad p_0^f, \quad \varrho_0^f,$$

the values of which, due to (2.1)–(2.4), are related to each other by

$$\mathbf{T}_0^{*s} = -p_0^f \mathbf{I} + 2\varrho_0^s \frac{\partial \tilde{e}^s}{\partial \mathbf{C}}\Big|^0, \tag{3.1}$$

$$\frac{p_0^f}{(\varrho_0^s)^2} = \frac{\partial \tilde{e}^s}{\partial \varrho^s}\Big|^0, \tag{3.2}$$

$$\kappa = \tilde{\kappa}|^0, \tag{3.3}$$

$$\frac{p_0^f}{(\varrho_0^f)^2} = \frac{\partial \tilde{e}^f}{\partial \varrho^f}\Big|^0, \tag{3.4}$$

where

$$\alpha|^0 = \alpha(\mathbf{C}_0, \varrho_0^s), \qquad \frac{\partial \tilde{e}^f}{\partial \varrho^f}\Big|^0 = \frac{\partial \tilde{e}^f}{\partial \varrho^f}(\varrho_0^f)$$

for

$$\alpha = \frac{\partial \tilde{e}^s}{\partial \mathbf{C}}, \quad \frac{\partial \tilde{e}^s}{\partial \varrho^s}, \quad \tilde{\kappa}$$

and

$$\mathbf{C}_0 = \mathbf{F}_0^T \mathbf{F}_0 = \mathbf{I}.$$

For further discussion we introduce the solid displacement gradient $\mathbf{H}$

$$\mathbf{H} = \mathbf{F} - \mathbf{I} \tag{3.5}$$

and the Lagrange strain tensor $\mathbf{E}$

$$2\mathbf{E} = \mathbf{C} - \mathbf{I}$$

that are linked by the geometrical relation

$$2\mathbf{E} = \mathbf{H} + \mathbf{H} + \mathbf{H}^T\mathbf{H}. \tag{3.6}$$

Then, at small values of the displacement gradient $\mathbf{H}$, from (3.6) we obtain

$$\mathbf{E} \simeq (\mathbf{H} + \mathbf{H})/2 = \tilde{\mathbf{E}}$$

and the right Cauchy-Green deformation tensor $\mathbf{C}$ can be expressed as follows

$$\mathbf{C} \simeq \mathbf{I} + 2\tilde{\mathbf{E}}, \tag{3.7}$$

where $\tilde{\mathbf{E}}$ is the infinitesimal strain tensor of the skeleton. The quantity $2\tilde{\mathbf{E}}$ is the linear increment of the deformation tensor $\mathbf{C}$.

To obtain linear constitutive relations from (2.1)-(2.4) we introduce the incremental form of quantities $\mathbf{T}^{*s}$, p^f, ϱ^s, ϱ^f, and κ

$$\begin{aligned} \mathbf{T}^{*s} &= \mathbf{T}_0^{*s} + \Delta\mathbf{T}^{*s}, \qquad & p^f &= p_0^f + \Delta p^f, \\ \varrho^s &= \varrho_0^s + \Delta\varrho^s, & \varrho^f &= \varrho_0^f + \Delta\varrho^f, \\ & & \kappa &= \kappa_0 + \Delta\kappa. \end{aligned} \tag{3.8}$$

Then, using expressions (3.5), (3.7) and (3.8) in the constitutive relation (2.1), after expansion of the internal energy function we can write the effective stresses in the skeleton as follows

$$\begin{aligned} \mathbf{T}_0^{*s} + \Delta\mathbf{T}^{*s} &= -(p_0^f + \Delta p^f)\mathbf{I} \\ &+2(\varrho_0^s + \Delta\varrho^s)(\mathbf{H} + \mathbf{I})\left[\frac{\partial\tilde{e}^s}{\partial\mathbf{C}}\bigg|^0 + 2\frac{\partial^2\tilde{e}^s}{\partial\mathbf{C}^2}\bigg|^0 \cdot \tilde{\mathbf{E}} + \frac{\partial^2\tilde{e}^s}{\partial\varrho^s\partial\mathbf{C}}\bigg|^0 \Delta\varrho^s + \ldots\right](\mathbf{H} + \mathbf{I})^T. \end{aligned} \tag{3.9}$$

The above relation, when the condition (3.1) is taken into account and all the nonlinear terms are neglected, assumes the form

$$\begin{aligned} \Delta\mathbf{T}^{*s} + \Delta p^f\mathbf{I} &= \mathbb{C}^* \cdot \tilde{\mathbf{E}} + \mathbb{K}^*\frac{\Delta\varrho^s}{\varrho_0^s} \\ &+(\mathbf{T}_0^{*s} + p_0^f\mathbf{I})\frac{\Delta\varrho^s}{\varrho_0^s} + 2p_0^f\tilde{\mathbf{E}} + \mathbf{H}\,\mathbf{T}_0^{*s} + \mathbf{T}_0^{*s}\mathbf{H}^T, \end{aligned} \tag{3.10}$$

where quantities

$$\mathbb{C}^* = 4\varrho_0^s \frac{\partial^2 \tilde{e}^s}{\partial \mathbf{C}^2}\Big|^0, \qquad \mathbb{K}^* = 2(\varrho_0^s)^2 \frac{\partial^2 \tilde{e}^s}{\partial \varrho^s \partial \mathbf{C}}\Big|^0 \tag{3.11}$$

are the effective material constants of the porous skeleton which have tensorial character; $\mathbb{C}^*$ is the fourth order tensor and $\mathbb{K}^*$ is the second order tensor.

In a similar way we can obtain the linear form of Eqs. (2.2) and (2.3). They are

$$\Delta p^f = \mathbb{K}^* \cdot \tilde{\mathbf{E}} + (K_c^* + 2p_0^f)\frac{\Delta \varrho^s}{\varrho_0^s}, \tag{3.12}$$

$$\Delta \kappa = \nu_c^* \frac{\Delta \varrho^s}{\varrho_0^s} + \mathbb{P}^* \cdot \tilde{\mathbf{E}}, \tag{3.13}$$

where

$$K_c^* = (\varrho_0^s)^2 \frac{\partial^2 \tilde{e}^s}{\partial (\varrho^s)^2}\Big|^0, \tag{3.14}$$

$$\nu_c^* = \varrho_0^s \frac{\partial \tilde{\kappa}}{\partial \varrho^s}, \qquad \mathbb{P}^* = \frac{\partial \tilde{\kappa}}{\partial \mathbf{C}}. \tag{3.15}$$

In derivation of (3.12) the commutative law of differentiation

$$\mathbb{K}^* = 2(\varrho_0^s)^2 \frac{\partial^2 \tilde{e}^s}{\partial \varrho^s \partial \mathbf{C}}\Big|^0 = 2(\varrho_0^s)^2 \frac{\partial^2 \tilde{e}^s}{\partial \mathbf{C} \partial \varrho^s}\Big|^0 \tag{3.16}$$

was used.

Equations (3.10), (3.12) and (3.13) are the set of linear constitutive relations for fluid-saturated porous solid of an anisotropic elastic properties and the isotropic pore structure. From the definitions (3.11), (3.14) and (3.15) it is seen that the material constants $\mathbb{C}^*$, $\mathbb{K}^*$, K_c^*, ν_c^* and $\mathbb{P}^*$ characterize the mechanical properties of porous skeleton only and depend on the chosen initial state of the porous solid. The fourth order tensor $\mathbb{C}^*$ in (3.10) is the tensor of elastic constants for the porous skeleton undergoing small external deformations at constant effective skeleton mass density ϱ^s. The material constant K_c^* in (3.12) represents the volumetric modulus of elasticity of the skeleton material corresponding to the pure internal deformation caused by the change of the pore pressure p^f at constant deformation tensor $\mathbf{C}$ ($\tilde{\mathbf{E}} = 0$, external deformation does not exist).

The second order tensor $\mathbb{K}^*$, as it is seen from the definition (3.16) and relations (3.10) and (3.12), is the tensor characterizing the coupling between two independent kinds of deformations measured by tensor $\mathbf{C}$ and the increment of ϱ^s. Due to the symmetry of tensor $\mathbf{C}$, the tensor $\mathbb{K}^*$ is also symmetric.

Combination of the tensor $\mathbb{K}^*$ with the constant K_c^* in the form

$$\mathbb{V}_p = \mathbb{K}^*/(K_c^* + 2p_0^f) \tag{3.17}$$

characterizes the volume changes of the skeleton material caused by external deformation of the porous solid at constant pore pressure ($p^f = p_0^f$). In such a case from (3.12) we have

$$\frac{\Delta\varrho^s}{\varrho_0^s} = -\mathbb{V}_p \cdot \tilde{\mathbf{E}}. \tag{3.18}$$

On the other hand, in the case when the external deformation does not exist ($\tilde{\mathbf{E}} = 0$), from equations (3.10) and (3.12) we obtain the relation

$$\Delta\mathbf{T}^{*s} = -\Delta p^f \mathbf{I} + \mathbb{V}_E \Delta p^f, \tag{3.19}$$

where the tensor $\mathbb{V}_E$ is expressed by tensor $\mathbb{K}^*$and quantities K_c^*, $\mathbf{T}_0^{*s}$ in the following way

$$\mathbb{V}_E = \mathbb{V}_p + (\mathbf{T}_0^{*s} + p_0^f \mathbf{I})/(K_c^* + 2p_0^f). \tag{3.20}$$

It characterizes the change of the skeleton stresses resulting from the internal solid deformation caused by the change of the fluid pore pressure.

Coefficients appearing in equation (3.13) describe the changes of the pore structure parameter κ during the deformation process. The scalar coefficient ν_c^* defined by $(3.15)_1$ characterizes the change of κ as a result of the change of the effective skeleton density at the constant deformation tensor $\mathbf{C}$, whereas the coefficient $\mathbb{P}^*$ is the symmetric second order tensor characterizing the changes of κ caused by the external skeleton deformation defined by $\tilde{\mathbf{E}}$ at constant effective density ϱ^s.

The changes of the second pore parameter, i.e. the volume porosity, are characterized by the skeleton mass continuity equation (2.6). Its linear form is

$$\frac{\Delta f_v}{1 - f_v^0} = \frac{\Delta\varrho^s}{\varrho_0^s} + \operatorname{tr}(\tilde{\mathbf{E}}). \tag{3.21}$$

To complete the linear constitutive description of an elastic porous solid-fluid composition it is necessery to linearize the constitutive relation (2.4) for fluid. We have

$$\Delta p^f = a_0^2 \Delta\varrho^f, \tag{3.22}$$

where

$$a_0 = \left(2\varrho_0^f \frac{d\tilde{e}^f}{d\varrho^f}\bigg|^0 + (\varrho_0^f)^2 \frac{d^2\tilde{e}^f}{d(\varrho^f)^2}\bigg|^0\right)^{1/2} \tag{3.23}$$

is the velocity of the wave-front propagation in a bulk fluid.

Constitutive stress-strain relations of the porous solid (3.10), the internal equilibrium equation (3.12), the constitutive relation for the barotropic fluid (3.22) and the equation of changes of the pore parameter κ (3.13) form the complete set of the linear constitutive equations for the elastic fluid-saturated porous solid of an anisotropic mechanical properties and isotropic pore structure.

These equations contain six material constants. Three of them are scalar coefficients (K_c^*, a_0, ν_c^*) and three other are tensorial coefficients of the fourth order $(\mathbb{C}^*)$ and second order $(\mathbb{K}^*, \mathbb{P}^*)$.

It should be mentiond that in the above description the velocity a_0 or, equivalently, the fluid volume compressibility K^f

$$K^f = \varrho_0 a_0^2 \tag{3.24}$$

is the only material constant characterizing the fluid properties while the remaining parameters characterize the skeleton properties.

Regarding the symmetry of tensorial coefficients we have, in general, $3 + 21 + 6 + 6 = 36$ scalar quantities that have to be determined experimentally. Moreover, in solving any mathematical problem it is necessary to know the quantities $\mathbf{T}_0^{*s}$, ϱ_0^s, κ, f_v^0 and p_0^f characterizing the state of saturated solid in its reference configuration.

4. Linear constitutive relations. Special cases

The obtained constitutive equations of an anisotropic fluid-saturated porous solid are a good basis for derivation of constitutive relations for elastic porous solids with high symmetry of mechanical properties or reduced physical properties. In this section we consider elastic behaviour of the fluid-porous solid composition with skeleton of isotropic mechanical properties, the case when the skeleton material is incompressible and the case when the porous medium is unsaturated. Constitutive relations for porous materials of such reduced properties have simple form and are important in practical applications.

4.1. Porous medium with isotropic skeleton

The constitutive relations (3.10), (3.12), (3.13) and (3.22) will describe the isotropic properties of a porous body if their form is invariant under any orthogonal transformation of the dependent and independent variables

$$\left\{\Delta p^f,\ \Delta\varrho^f,\ \Delta\varrho^s,\ \Delta\kappa,\ \Delta\mathbf{T}^{*s},\ \tilde{\mathbf{E}},\ \mathbf{H}\right\}.$$

Taking the orthogonal transformations of these variables, i.e.

$$\begin{aligned} &'\Delta p^f = \Delta p^f, \quad '\Delta\varrho^f = \Delta\varrho^f, \quad '\Delta\varrho^s = \Delta\varrho^s, \quad '\Delta\kappa = \Delta\kappa, \\ &'\Delta\mathbf{T}^{*s} = \mathbf{Q}\,\Delta\mathbf{T}^{*s}\mathbf{Q}^T, \quad '\tilde{\mathbf{E}} = \mathbf{Q}\tilde{\mathbf{E}}\,\mathbf{Q}^T, \quad '\mathbf{H} = \mathbf{Q}\,\mathbf{H}\,\mathbf{Q}^T, \end{aligned} \tag{4.1}$$

the constitutive relations (3.10), (3.12) and (3.13) assume the form

$$\mathbf{Q}\,\Delta\mathbf{T}^{*s}\mathbf{Q}^T + \Delta p^f\mathbf{I} = \mathbb{C}^* \cdot (\mathbf{Q}\,\tilde{\mathbf{E}}\,\mathbf{Q}^T) + \mathbb{K}^*\frac{\Delta\varrho^s}{\varrho_0^s} + (\mathbf{T}_0^{*s} + p_0^f\mathbf{I})\frac{\Delta\varrho^s}{\varrho_0^s} + 2p_0^f(\mathbf{Q}\,\tilde{\mathbf{E}}\,\mathbf{Q}^T) + (\mathbf{Q}\,\mathbf{H}\,\mathbf{Q}^T)\mathbf{T}_0^{*s} + \mathbf{T}_0^{*s}(\mathbf{Q}\,\mathbf{H}\,Q^T)^T, \tag{4.2}$$

$$p^f = \mathbb{K}^* \cdot (\mathbf{Q}\,\tilde{\mathbf{E}}\,\mathbf{Q}^T) + K_c^*\frac{\Delta\varrho^s}{\varrho_0^s} + 2p_0^f\frac{\Delta\varrho^s}{\varrho_0^s}, \tag{4.3}$$

$$\Delta\kappa = \nu_c^*\frac{\Delta\varrho^s}{\varrho_0^s} + \mathbb{P}^* \cdot (\mathbf{Q}\,\tilde{\mathbf{E}}\,\mathbf{Q}^T), \tag{4.4}$$

where $\mathbf{Q}$ ($\mathbf{Q}\,\mathbf{Q}^T = \mathbf{I}$) is the orthogonal tensor.

Equations (4.2) – (4.4) will be identical with the corresponding equations (3.10), (3.12) and (3.13) for arbitrary values of variables Δp^f, $\Delta\varrho^s$, $\Delta\kappa$, $\Delta\mathbf{T}^{*s}$, $\tilde{\mathbf{E}}$, $\mathbf{H}$ and any orthogonal tensor $\mathbf{Q}$ if the following conditions are satisfied

$$\begin{aligned} \mathbf{Q} * \mathbb{C}^* &= \mathbb{C}^*, \\ \mathbf{Q}\,\mathbb{K}^*\,\mathbf{Q}^T &= \mathbb{K}^*, \qquad \mathbf{Q}\,\mathbb{P}^*\mathbf{Q}^T = \mathbb{P}^*, \\ \mathbf{Q}\,\mathbf{T}_0^{*s}\mathbf{Q}^T &= \mathbf{T}_0^{*s}, \end{aligned} \tag{4.5}$$

where $\mathbf{Q}*$ is a linear operator defined by the equation

$$\mathbf{Q} * (\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \mathbf{v}_3 \otimes \mathbf{v}_4) = \mathbf{Q}\,\mathbf{v}_1 \otimes \mathbf{Q}\,\mathbf{v}_2 \otimes \mathbf{Q}\,\mathbf{v}_3 \otimes \mathbf{Q}\,\mathbf{v}_4$$

and $\otimes$ denotes the tensorial product of vectors.

It follows from (4.5) that the isotropy conditions for the constitutive relations are equivalent to the requirement of isotropy of tensorial material constants $\mathbb{C}^*$, $\mathbb{K}^*$ and $\mathbb{P}^*$ and, additionally, the isotropy of the skeleton stress state $\mathbf{T}_0^{*s}$ in the reference configuration.

The isotropy conditions (4.5) reduce the quantities $\mathbb{C}^*$, $\mathbb{K}^*$ and $\mathbb{P}^*$ and $\mathbf{T}_0^{*s}$ to the following form

$$\begin{aligned} \mathbb{C}^* &= \lambda_\varrho^*\mathbf{I} \otimes \mathbf{I} + 2\mu_\varrho^*\mathbb{J}, \\ \mathbb{K}^* &= K^*\mathbf{I}, \qquad \mathbb{P}^* = \nu_\varrho^*\mathbf{I}, \\ \mathbf{T}_0^{*s} &= -p_0^s\mathbf{I}, \end{aligned} \tag{4.6}$$

where $\mathbb{J}$ is the fourth order unit tensor defined as the identity operator for the second order tensors $\mathbf{A}$ ($\mathbb{J} \cdot \mathbf{A} = \mathbf{A}$). The quantities λ_ϱ^* and μ_ϱ^* are the effective Lamé constants of porous skeleton measured at the constant effective mass density of the skeleton material, and p_0^s is the initial stress in the skeleton.

Using (4.6) in Eqs. (3.10), (3.12) and (3.13) we obtain

$$\Delta \mathbf{T}^{*s} + \Delta p^f \mathbf{I} = 2(\mu_\varrho^* + p_0^f - p_0^s)\tilde{\mathbf{E}} + \left(\lambda_\varrho^* \operatorname{tr}(\tilde{\mathbf{E}}) + (K^* + p_0^f - p_0^s)\frac{\Delta \varrho^s}{\varrho_0^s}\right)\mathbf{I}, \tag{4.7}$$

$$\Delta p^f = K^* \operatorname{tr}(\tilde{\mathbf{E}}) + (K_c^* + 2p_0^s)\frac{\Delta \varrho^s}{\varrho_0^s}, \tag{4.8}$$

$$\Delta \kappa = \nu_\varrho^* \operatorname{tr}(\tilde{\mathbf{E}}) + \nu_c^* \frac{\Delta \varrho^s}{\varrho_0^s}. \tag{4.9}$$

In the case when the initial stress in the porous skeleton is equal to the intial pore fluid pressure

$$p_0^s = p_0^f,$$

Eq. (4.7) takes the reduced form

$$\Delta \mathbf{T}^{*s} + \Delta p^f \mathbf{I} = 2\mu_\varrho^* \tilde{\mathbf{E}} + \left(\lambda_\varrho^* \operatorname{tr}(\tilde{\mathbf{E}}) + K^* \frac{\Delta \varrho^s}{\varrho_0^s}\right)\mathbf{I}. \tag{4.10}$$

Equations (3.22), (4.7) (or (4.10)), (4.8) and (4.9) form the complete set of the linear constitutive relations for fully isotropic porous solid filled with fluid.

Seven material constants

$$\mu_\varrho^*, \quad \lambda_\varrho^*, \quad K^*, \quad K_c^*, \quad \nu_\varrho^*, \quad \nu_c^*, \quad K^f$$

are involved in the description, where the first six constants characterize elastic properties of the porous skeleton and one constant describes the mechanical fuid property.

Methods of determination of these material constants will be discussed n a seperate paper.

4.2. Saturated porous medium with incompressible matrix material

In the analysis of deformation processes of fluid-saturated porous media there are many physical situations in which the skeleton material can be considered as incompressible. The incompressibility condition takes the form

$$\varrho^s = \varrho_0^s \tag{4.11}$$

and is the kinematic constraint that confines the skeleton motion during its deformation.

In such a case the macroscopic volume deformations of porous skeleton arise at the cost of the change of pore volume. This is evidently seen in the skeleon continuity equation (3.21) that has the form

$$\frac{\Delta f_v}{1 - f_v^0} = \operatorname{tr}(\tilde{\mathbf{E}}). \tag{4.12}$$

The incompressibility condition (4.11) is, at the same time, a particular case of the equation defining changes of the effective skeleton mass density and, as it was shown in [5], it replaces the internal equilibrium condition (in our case, Eqs. (3.12) and (4.8)).

The skeleton material incompressibility has no influence on the form of constitutive relation (3.22) for the fluid pressure, however, it substantially simplifies the form of two other relations (3.10) and (3.13) reducing the number of material constants. We have

$$\Delta \mathbf{T}^{*s} + \Delta p^f \mathbf{I} = (\mathbb{C}^* + 2p_0^f \mathbb{J}) \cdot \widetilde{\mathbf{E}} + \mathbf{H}\,\mathbf{T}_0^{*s} + \mathbf{T}_0^{*s}\mathbf{H}^T, \tag{4.13}$$

$$\Delta \kappa = \nu_\varrho^* \operatorname{tr}(\widetilde{\mathbf{E}}). \tag{4.14}$$

In this case the increment of fluid pressure Δp^f is the part of the skeleton stresses that during the skeleton deformation does the work over the pore fluid but does not change the energetic state of the skeleton due to its material incompressibility. Equations (4.13) and (4.14) for the fully isotropic porous solid, according to the analysis done in Sec. 4.1 assume the form

$$\Delta \mathbf{T}^{*s} + \Delta p^f \mathbf{I} = 2(\mu_\varrho^* + p_0^f - p_0^s)\widetilde{\mathbf{E}} + \lambda_\varrho^* \operatorname{tr}(\widetilde{\mathbf{E}})\mathbf{I}, \tag{4.15}$$

$$\Delta \kappa = \nu_\varrho^* \operatorname{tr}(\widetilde{\mathbf{E}}). \tag{4.16}$$

The above equations form, together with (3.22), the set of three constitutive relations defining the mechanical behaviour of fluid-saturated, isotropic porous solid with incompressible skeleton material. Such porous medium is characterized by four material constants:

$$\mu_\varrho^*, \quad \lambda_\varrho^*, \quad \nu_\varrho^*, \quad K^f.$$

The first three constants describe mechanical properties of porous skeleton and the last one describes the pore fluid.

4.3. Non-saturated porous solid

To obtain the constitutive relations describing the elastic behaviour of an anisotropic porous solid not saturated with fluid, one can assume in the equations (3.10) and (3.12) that the pore fluid pressure p^f is equal to zero ($p^f = 0$). Therefore, these equations get the form

$$\Delta \mathbf{T}^{*s} = \mathbb{C}^* \cdot \widetilde{\mathbf{E}} + (\mathbb{K}^* + \mathbf{T}_0^{*s})\frac{\Delta \varrho^s}{\varrho_0^s} + \mathbf{H}\,\mathbf{T}_0^{*s} + \mathbf{T}_0^{*s}\mathbf{H}^T, \tag{4.17}$$

$$0 = \mathbb{K}^* \cdot \widetilde{\mathbf{E}} + K_c^* \frac{\Delta \varrho^s}{\varrho_0^s}, \tag{4.18}$$

while the equation (3.13) is not changed.

Equation (4.17) will be simplified, if the skeleton reference configuration is its natural configuration, i.e. $\mathbf{T}_0^{*s} = 0$.

We have

$$\Delta \mathbf{T}^{*s} = \mathbb{C}^* \cdot \tilde{\mathbf{E}} + \mathbb{K}^* \frac{\Delta \varrho^s}{\varrho_0^s}. \tag{4.19}$$

From the internal equilibrium equation (4.18) it follows that for fluid-free porous skeleton, the density change of the skeleton material is uniquely defined by the porous solid strain tensor $\tilde{\mathbf{E}}$. Therefore the constitutive relation (4.19) can be written in the form

$$\mathbf{T}^{*s} = \mathbb{C}_z^* \cdot \tilde{\mathbf{E}} \tag{4.20}$$

which is analogous to that of non-porous elastic solid.

Tensor

$$\mathbb{C}_z^* = \mathbb{C}^* - (\mathbb{K}^* \otimes \mathbb{K}^*)/K_c^* \tag{4.21}$$

is the equivalent elasticity tensor of the effective elastic constants of a porous skeleton.

In the isotropic case relation (4.20) is

$$\mathbf{T}^{*s} = 2\mu_z^* \tilde{\mathbf{E}} + \lambda_z^* \operatorname{tr}(\tilde{\mathbf{E}})\mathbf{I}, \tag{4.22}$$

where

$$\mu_z^* = \mu_\varrho^*, \qquad \lambda_z^* = \lambda_\varrho^* - (\nu_\varrho^*)^2/K_c^*.$$

If, additionally, the incompressibility of the skeleton material is assumed, the stress in the skeleton can be written as

$$\mathbf{T}^{*s} = 2\mu_\varrho^* \tilde{\mathbf{E}} + \lambda_\varrho^* \operatorname{tr}(\tilde{\mathbf{E}})\mathbf{I}. \tag{4.23}$$

The material coefficients appearing in relations (4.20), (4.22) and (4.23) play an analogous role as those in the classical linear elasticity of solids, and their measurement can be done in the classical way. These material constants completely assure the determination of stress and strain state in the porous skeleton. For description of the change of the skeleton mass density ϱ^s, or the change of the pore structure parameter κ it is necessary to evaluate additional coefficients appearing in Eqs. (4.18) and (3.13) or in their reduced forms (4.9) and (4.8). Measurement of these coefficients requires some new methods to be proposed.

5. Final remarks

The complete set of constitutive relations for a fluid-saturated porous solid with anisotropic properties of elastic skeleton and isotropic pore structure characterized by two parameters have been formulated in the paper. It comprises: the constitutive relations for the effective skeleton stresses and the pore fluid pressure, the internal mechanical equilibrium condition and the equation of changes of the pore structure parameter κ. These relations are supplemented with the skeleton continuity equation which describe the changes of porosity f_v.

Considerations have been based on the nonlinear constitutive relations of such medium obtained in the paper [5], where the consequences of the constituent immiscibility for these relations have been analysed.

Such approach made it possible to construct the consistent linear description of elastic behaviour of porous skeleton filled with barotropic fluid in which all material constants are precisely defined and have clear physical meaning. Also the character of couplings appearing in the constitutive relations and their interpretation are simpler.

References

1. M.A. BIOT, *Theory of propagation of elastic waves in a fluid-saturated porous solid*, JASA, **28**, 2, 168–191, 1956.
2. M.A. BIOT, *Mechanics of deformation and acoustic propagation in porous media*, J. Appl. Phys., **33**, 4, 1482–1498, 1962.
3. M.A. BIOT and D.G. WILLIS, *The elastic coefficients of theory of consolidation*, J. Appl. Mech., **24**, 594–601, 1957.
4. R.M. BOWEN, *Compressible porous media models by use of the theory of mixtures*, Int. J. Engng. Sci., **3**, 231–241, 1965.
5. M. CIESZKO and J. KUBIK, *Constitutive relations and internal equilibrium condition for fluid-saturated porous solid*, Arch. Mech., **48**, 5, 893–910, 1996.
6. M.J. CROCHET and P.M. NAGHDI, *On constitutive equations for flow of fluid through an elastic solid*, Int. J. Engng. Sci., **4**, 383–401, 1966.
7. M. KACZMAREK and J. KUBIK, *Determination of material constants of fluid-saturated porous solid* [in Polish], Engng. Trans., **33**, 4, 589–609, 1985.
8. H.B. KINSBURY, *Determination of material parameters of poroelastic media*, [in:] Fundamentals of Transport Phenomena in Porous Media, J. BEAR, M. CARAPCIOGLU [Eds.], Martinus Nijhof, 574–615, 1984.
9. J. KUBIK, *Large elastic deformations of fluid-saturated porous solid*, J. de Mecanique Theorique et Applique, 203–218, 1982.
10. J. KUBIK, *Pore structure in dynamic behaviour of saturated materials*, Transport in Porous Media, **9**, 15–24, 1992.
11. D.C. LEIGH, *Nonlinear continuum mechanics*, McGraw-Hill Book Company, New York 1968.
12. G. SZEFER, *Nonlinear problems of consolidation theory*, Problems de Reologie, Symposium Franco-Polonaise, Kraków 1977, PWN, Warszawa 1978.

POLISH ACADEMY OF SCIENCES
INSTITUTE OF FUNDAMENTAL TECHNOLOGICAL RESEARCH

Received March 14, 1996.

Arch. Mech., 48, 5, pp. 925–931, Warszawa 1996

On existence theorems of periodic traveling wave solution to the generalized forced Kadomtsev–Petviashvili equation

Y. CHEN (FAYETTEVILLE)

THIS PAPER is concerned with periodic traveling wave solutions of the generalized forced Kadomtsev-Petviashvili equation in the form $(u_t + [f(u)]_x + \alpha u_{xxx})_x + \beta u_{yy} = h_0$. The basic approach to this problem is to establish an equivalence relationship between a periodic boundary value problem and nonlinear integral equations with symmetric kernels by using the Green's function method. The integral representations generate compact operators in a Banach space of real-valued continuous periodic functions with a given period $2T$. Schauder's fixed point theorem is then used to prove the existence of nonconstant periodic traveling wave solutions.

1. Introduction

THE KORTEWEG–DE VRIES EQUATION (KdV equation for short) is a nonlinear evolution equation governing long one-dimensional, small amplitude, surface gravity waves propagating in a shallow channel of water [1]. It has many applications in the study of other physical problems, such as plasma waves, lattice waves, and waves in elastic rods, etc. A two-dimensional generalization of the KdV equation is the Kadomtsev–Petviashvili equation (referred to as KP equation henceforth), which was obtained in 1970 in the study of plasma [2]. The evolution described by the KP equation is weakly nonlinear, weakly dispersive, and weakly two-dimensional, with all the three effects being of the same order. The KP equation has also been proposed as a model for the surface waves and internal waves in channels of varying depth and width [3].

Twenty years ago, in an impressive survey on the KdV equation, MIURA listed seven open problems of the KdV equation [2]. The seventh open problem concerns the forced KdV equation. At that time the physical basis for the forced KdV equation was not clear. PATOINE and WARN were the first two who used the forced KdV equation as a physical model equation in 1982 [5]. However, it was not until 1984 that AKYLAS first systematically derived the forced KdV equation from the model of long nonlinear water wave forced by a moving pressure [6]. After that, WU [7] and SHEN [8] also derived the forced KdV equation in the study of long water waves in a two-dimensional channel forced by a bottom topography and/or an external pressure applied on the free surface. In a recent paper [9], SHEN derived the one-dimensional stationary forced KdV equation of the form $\lambda u_t + \alpha u u_x + \beta u_{xxx} = h_x$ for the long nonlinear water waves flowing over long bumps, and proved the existence of positive solitary wave solutions to the stationary forced KdV equation with the boundary value conditions $u(\pm\infty) = u'(\pm\infty) = 0$.

In this paper the author considers the generalized forced KP equation of the form

$$(u_t + [f(u)]_x + \alpha u_{xxx})_x + \beta u_{yy} = h_0, \tag{1.1}$$

where α and β are positive numbers, and f is a C^2 function of its argument. When $f(u) = u^2/2$, $\alpha = 1$, and $\beta = 3$, Eq. (1.1) reduces to the two-dimensional forced KP equation of the form

$$(u_t + uu_x + u_{xxx})_x + 3u_{yy} = h_0, \tag{1.2}$$

which is a two-dimensional generalization of the equation obtained by Akylas, Wu, and Shen. The author will prove an existence theorem of nonconstant periodic traveling wave solution to the generalized forced KP equation following the idea of LIU and PAO [10].

The author applies the Green's function method to derive nonlinear integral equations which are equivalent to the generalized forced KP equation with periodic boundary conditions. Imposing suitable conditions, the author establishes the existence of solutions to the integral equations, and hence proves the existence of periodic traveling wave solutions to Eq. (1.1). Furthermore, we note that the nonconstant periodic traveling wave solutions are infinitely differentiable.

The content of the paper is arranged as follows. In Sec. 2, the author converts the generalized forced KP equation into nonlinear integral equations using the Green's function method. Section 3 contains the proof of the existence theorem for these integral equations.

2. Formulation of the problem

We start from the generalized forced KP equation

$$(u_t + [f(u)]_x + \alpha u_{xxx})_x + \beta u_{yy} = h_0, \tag{2.1}$$

where f is a C^2 function of its argument and h_0 is a nonconstant function of x, y and t. We are interested in the periodic traveling wave solutions of the form $U(z) = u(x, y, t)$, where $z = ax + by - \omega t$ with a, b, and ω being real constants. Without any loss of generality we assume $a > 0$. Consider the case that $h_0(x, y, t) = a^2 h(z)$ is a $2T$-periodic continuous function of z, where T is a preassigned positive number. Substitution of the $U(z)$ into Eq. (2.1) leads then to the fourth-order nonlinear ordinary differential equation

$$U^{(4)}(z) = \frac{C}{\alpha a^2} U''(z) - \frac{1}{\alpha a^2}\left[f(U(z))(U'(z))^2 + f'(U(z))U''(z)\right] + \frac{1}{\alpha a^2} h(z), \tag{2.2}$$

where $C = (\omega a - \beta b^2)/a^2$. We impose the following periodic boundary conditions

$$U^{(n)}(0) = U^{(n)}(2T), \qquad n = 0, 1, 2, 3. \tag{2.3}$$

In addition, in order to rule out non-zero constant solutions, another condition is introduced

$$\int_0^{2T} U(z)\,dz = 0. \tag{2.4}$$

Thus, any solution of the boundary value problem consisting of Eqs. (2.2) – (2.4) can be extended to a $2T$-periodic traveling wave solution to Eq. (2.1).

Integrating both sides of Eq. (2.2) with respect to z twice and using Eqs. (2.3), (2.4), we obtain

$$U''(z) - \frac{C}{\alpha a^2}U(z) = E - \frac{1}{\alpha a^2}[f(U(z)) - H(z)], \tag{2.5}$$

$$U^{(n)}(0) = U^{(n)}(2T), \qquad n = 0, 1, \tag{2.6}$$

where

$$E = \frac{1}{2T} \cdot \frac{1}{\alpha a^2} \int_0^{2T} [f(U(z)) - H(z)]\,dz,$$

and $H(z)$ is a $2T$-periodic function of z such that $H''(z) = h(z)$. Conversely, integrating both sides of Eq. (2.5) from 0 to $2T$ and using Eqs. (2.6) we are led to Eq. (2.4), and direct differentiations of Eq. (2.5) will give us Eqs. (2.2), (2.3). Therefore, we have proved the following theorem by noting from Eq. (2.5) that $U \in C^2[0, 2T]$ implies $U \in C^4[0, 2T]$ since f is a C^2 function of its argument.

THEOREM 1. *Suppose that* $C \neq 0$; *a function* $U(z)$ *is a solution of the boundary value problem* Eqs. (2.2) – (2.4) *if and only if it is a solution of the boundary value problem* Eqs. (2.5) *and* (2.6).

From now on we consider only the two cases: 1. $C > 0$, and 2. $C < 0$ but $-C/(\alpha a^2) \neq (k\pi/T)^2$ with k being any integer.

Denote the function $f(U(z)) - H(z)$ on the right-hand side of Eq. (2.5) by $F(U(z))$. Treating the right-hand side of Eq. (2.5) as a forcing term and using the Green's function method [11], the boundary value problem Eqs. (2.5), (2.6) can be converted to an integral equation

$$U(z) = \frac{1}{\alpha a^2} \int_0^{2T} K_i(z, s) F(U(s))\,ds, \tag{2.7}$$

where the kernels K_i, $i = 1, 2$, are defined as follows:

1. When $C > 0$, let $\lambda_1 = \sqrt{C/(\alpha a^2)}$; then

$$K_1(z,s) = \frac{\cosh \lambda_1(T - |z-s|)}{2\lambda_1 \sinh \lambda_1 T} - \frac{1}{2\lambda_1^2 T}, \qquad \forall z, s \in [0, 2T]. \tag{2.8}$$

2. When $C < 0$ but $-C/(\alpha a^2) \neq (k\pi/T)^2$ with k being any integer, let $\lambda_2 = \sqrt{-C/(\alpha a^2)}$; then

$$K_2(z,s) = \frac{\cos \lambda_2(T - |z-s|)}{2\lambda_2 \sin \lambda_2 T} - \frac{1}{2\lambda_2^2 T}, \qquad \forall z, s \in [0, 2T]. \tag{2.9}$$

LEMMA 1. The kernels K_1 and K_2 have the following properties:

$$K_i(0,s) = K_i(2T,s), \qquad \forall s \in [0,2T], \qquad i = 1,2,$$
$$K_i(z,2T-s) = K_i(2T-z,s), \qquad \forall s \in [0,2T], \qquad i = 1,2.$$

P r o o f. Straightforward computations follow from the definitions of the kernels K_i, $i = 1, 2$, given in Eqs. (2.8), (2.9). □

THEOREM 2. *A function $U(z)$ is a solution of the boundary value problem* Eqs. (2.5), (2.6) *if and only if it is a solution of the integral equation* (2.7).

P r o o f. The "if" part can be proved by direct differentiations of Eq. (2.7) and the "only if" part is based on the Green's function method by treating the right-hand side of Eq. (2.5) as a nonhomogeneous term. □

3. Existence theorem

To show the existence of $2T$-periodic traveling wave solutions to Eq. (2.1) it is suffucient to show that solutions to the Eq. (2.7) exist.

To this end we define C_{2T} as a collection of real-valued continuous functions, $v(z)$, on $[0, 2T]$ such that $v(0) = v(2T)$. Equip C_{2T} with the sup norm $\|\cdot\|$ as $\|v\| = \sup_{0 \le z \le 2T} |v(z)|$, for each $v \in C_{2T}$. Then $(C_{2T}, \|\cdot\|)$ is a Banach space.

We now define operators $\mathcal{A}_i$, $i = 1, 2$, on C_{2T} as

$$\mathcal{A}_i v(z) = \frac{1}{\alpha a^2} \int_0^{2T} K_i(z,s) F(v(s))\, ds, \qquad \forall v \in C_{2T}, \tag{3.1}$$

where the kernels K_i, $i = 1, 2$, are given in Eqs. (2.8), (2.9). We shall demonstrate that there exist functions v in C_{2T} such that $v = \mathcal{A}_i v$, $i = 1, 2$, and hence, prove that there exist solutions to Eq. (2.7).

Let

$$Q_i \geq \max_{0\leq z\leq 2T} \int_0^{2T} |K_i(z,s)|\,ds, \qquad i = 1,2, \tag{3.2}$$

$$\tau_1 = 1, \qquad \tau_2 = |\sin \lambda_2 T|. \tag{3.3}$$

A consequence of Lemma 1 can now be stated.

LEMMA 2. Let v be an element of C_{2T}. If $v(z) = v(2T - z)$ for $z \in [0, 2T]$, then $\mathcal{A}_i v(z) = \mathcal{A}_i v(2T - z)$, $i = 1, 2$.

We now define $B(0, r)$ to be a closed ball in C_{2T} and let $M = \sup[\|F(v)\| : v \in B(0, r)]$. We then have the following existence theorem.

THEOREM 3. *$\mathcal{A}_i$, $i = 1, 2$, is a compact operator from C_{2T} into C_{2T}. In particular, if $Q_i M/(\alpha a^2) \leq r$, $i = 1, 2$, then $\mathcal{A}_i$ maps $B(0, r)$ into itself. Hence, the integral equation* (2.7) *has at least one solution in $B(0, r)$.*

P r o o f. First we show $\mathcal{A}_i : C_{2T} \to C_{2T}$, $i = 1, 2$. Since it is obvious from Lemma 1 that $\mathcal{A}_i v(0) = \mathcal{A}_i v(2T)$ for each $v \in C_{2T}$, $i = 1, 2$, it suffices to show that $\mathcal{A}_i v$, $i = 1, 2$, is continuous on $[0, 2T]$.

Let v be an arbitrary function in C_{2T}; we have then

$$\frac{d\mathcal{A}_1 v(z)}{dz} = \frac{-1}{2\alpha a^2 \sinh \lambda_1 T} \int_0^z \sinh \lambda_1 (T - z + s) F(v(s))\,ds + \frac{1}{2\alpha a^2 \sinh \lambda_1 T} \int_z^{2T} \sinh \lambda_1 (T + z - s) F(v(s))\,ds, \tag{3.4}$$

$$\frac{d\mathcal{A}_2 v(z)}{dz} = \frac{1}{2\alpha a^2 \sin \lambda_2 T} \int_0^z \sin \lambda_2 (T - z + s) F(v(s))\,ds + \frac{-1}{2\alpha a^2 \sin \lambda_2 T} \int_z^{2T} \sin \lambda_2 (T + z - s) F(v(s))\,ds. \tag{3.5}$$

The existence of $d\mathcal{A}_1 v/dz$ and $d\mathcal{A}_2 v/dz$ implies that both $\mathcal{A}_1 v$ and $\mathcal{A}_2 v$ are continuous on $[0, 2T]$, and hence, $\mathcal{A}_i : C_{2T} \to C_{2T}$, $i = 1, 2$.

Let $\mathbf{S}$ be any bounded subset of C_{2T}, i.e., there exists an $L_0 > 0$ such that $\|v\| < L_0$ for all $v \in \mathbf{S}$. Then there must be an $M_0 > 0$ such that

$$\|F(v)\| = \sup_{0\leq z\leq 2T} |F(v(z))| \leq \sup_{-L_0\leq w\leq L_0} |F(w)| \leq M_0, \qquad \forall v \in \mathbf{S}.$$

Thus from Eqs. (3.1), (3.4), (3.5) we shall have

$$\|\mathcal{A}_i v\| \leq \frac{1}{\alpha a^2} Q_i M_0, \qquad \forall v \in \mathbf{S}, \quad i = 1, 2,$$
$$\|d\mathcal{A}_i v/dz\| \leq \frac{T}{\alpha a^2 \tau_i} M_0, \qquad \forall v \in \mathbf{S}, \quad i = 1, 2.$$

Therefore, $\mathcal{A}_i\mathbf{S}$, $i = 1, 2$, is uniformly bounded and equi-continuous, and by the Ascoli-Arzela Theorem both $\mathcal{A}_1$ and $\mathcal{A}_2$ are compact.

To show that $\mathcal{A}_i$, $i = 1, 2$, has a fixed point in $B(0, r)$ when $Q_i M/(\alpha a^2) \leq r$, $i = 1, 2$, we write

$$\begin{aligned} |\mathcal{A}_i v(z)| &= \frac{1}{\alpha a^2} \left| \int_0^{2T} K_i(z, s) F(v(s))\, ds \right| \\ &\leq \frac{1}{\alpha a^2} \int_0^{2T} |K_i(z, s)|\, |F(v(s))| ds \\ &\leq \frac{Q_i M}{\alpha a^2} \leq r, \qquad \forall v \in B(0, r). \end{aligned}$$

This implies that $\|\mathcal{A}_i v\| \leq r$ for all $v \in B(0, r)$, $i = 1, 2$, and hence, $\mathcal{A}_i$, $i = 1, 2$, maps $B(0, r)$ into itself. Therefore, by the Schauder's fixed point theorem we proved that $\mathcal{A}_i$ has a fixed point in $B(0, r)$ for each $i = 1, 2$. And hence, Eq. (2.7) has a solution for each case of $C > 0$ and $C < 0$ with $-C/(\alpha a^2) \neq (k\pi/T)^2$. □

It is worth noting that as long as $\int_0^{2T} K_i(z, s) H(s)\, ds \neq 0$, $i = 1, 2$, by Theorem 3, there exists a nonconstant function $v(z)$ on $[0, 2T]$ such that $v = \mathcal{A}_i v$, $i = 1, 2$, which implies that $v(z)$ is infinitely differentiable on $[0, 2T]$ since $\mathcal{A}_i v$ is differentiable on $[0, 2T]$. The extension of the $v(z)$ to a $2T$-periodic function $V(z)$ provides an infinitely differentiable $2T$-periodic traveling wave solution to the generalized forced KP equation.

References

1. D.J. KORTEWEG and G. DE VRIES, *On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves*, Philosophical Magazine, **39**, 422–443, 1895.
2. R.M. MIURA, *The Korteweg-de Vries equation; a survey of results*, SIAM Rev., **18**, 412–459, 1976.
3. B.B. KADOMTSEV and V.I. PETVIASHVILI, *On the stability of solitary waves in weak dispersing media*, Sov. Phys. Dokl., **15**, 6, 539–541, 1970.
4. P.M. SANTINI, *On the evolution of two-dimensional packets of water waves over an uneven bottom*, Lett. Nuovo Cimento, **30**, 236–240, 1981.
5. A. PATOINE and T. WARN, *The interaction of long quasi-stationary baroclinic waves with topography*, J. Atmos. Sci., **39**, 1018–1025, 1982.

6. T.R. AKYLAS, *On the excitation of long nonlinear water waves by a moving disturbance*, J. Fluid. Mech., **141**, 455–466, 1984.
7. T.Y. WU, *Generation of upstream-advancing solitons by moving disturbance*, J. Fluid Mech., **184**, 75–99, 1987.
8. S.S.P. SHEN, *Disturbed critical surface waves in a channel of arbitrary cross-section*, J. Appl. Math. Phys., **40**, 216–229, 1989.
9. S.S.P. SHEN, *On the accuracy of the stationary forced Korteweg-de Vries equation as a model equation for flows over bump*, Quar. Appl. Math., **53**, 4, 701–719, 1995.
10. BAO-PING LIU and C.V. PAO, *Green's function method for periodic traveling wave solutions of the KdV equation*, Appl. Anal., **14**, 292–301, 1983.
11. I. STAKGOLD, *Green's functions and boundary value problems*, John Wiley and Sons, New York 1979.
12. L. WANG, *Mathematics handbook*, People's Education Publisher, Beijing 1977.

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
FAYETTEVILLE STATE UNIVERSITY, FAYETTEVILLE, NORTH CAROLINA, USA.

Received March 18, 1996.

Arch. Mech., 48, 5, pp. 933–946, Warszawa 1996

The stationary transverse Euler and Stokes gas flows through a cylindrical region with large variations of density and viscosity coefficient

Z. PŁOCHOCKI and B. KAŹMIERCZAK (WARSZAWA)

THE FLOW of a gas in space, which encounters a cylindrical region, where the density of the gas (and its viscosity coefficient) changes abruptly, is considered both in the Euler and the Stokes approximations. The flow is homogeneous at infinity. Density and viscosity coefficients of the gas are assumed to be constants, which are different outside and inside the cylinder. The analytical solutions of the problem are found in both cases. These solutions may be useful for building the models of flow in flames or laser-sustained (or generated) plasmas.

1. Introduction

MODEL EXAMPLES of a stationary gas flow through a region with large variation of density (and viscosity coefficient) may be useful for constructing the simple hydraulic models of gas flow in systems with large heat perturbation, as for example – in flames or laser-generated or sustained plasmas. The idea of such models depends on the assumption, that the constant density of a gas inside the region is small as compared to (also constant) density outside the region. Such a density distribution is thought to be generated by a suitable temperature field, therefore in fact the viscosity coefficient of the gas should also be assumed to vary in a similar way.

The first such a hydraulic model was proposed in [1] for a spherical region in the Euler approximation. Numerical solution of the Navier–Stokes equations for such a flow was presented in [2]. The Stokes approximation of a gas flow through a spherical region was analyzed in [3]. The stationary transverse gas flow through a cylindrical region both in the Euler and the Stokes approximations is examined in the present paper.

2. General assumptions

Let us consider a stationary and homogeneous at infinity, transverse gas flow through a cylinder of radius R. The z-axis of the Cartesian coordinate system is the symmetry axis of the cylinder. At infinity the gas flows along the x-axis toward the cylinder. The flow is assumed to be plane in the sense, that the z-coordinate of the velocity is identically equal to zero. The gas density and the shear viscosity

coefficient are assumed in the form:

$$\begin{aligned} \overline{\varrho} &:= \frac{\varrho}{\varrho_\infty} = \varepsilon_\varrho + (1-\varepsilon_\varrho) H(\overline{r}-1), \\ \overline{\eta} &:= \frac{\eta}{\eta_\infty} = \frac{1}{\varepsilon_\eta} - \frac{1-\varepsilon_\eta}{\varepsilon_\eta} H(\overline{r}-1), \end{aligned} \tag{2.1}$$

$$\begin{aligned} \varepsilon_\varrho &:= \frac{\varrho_{\text{int}}}{\varrho_\infty}, \\ \varepsilon_\eta &:= \frac{\eta_\infty}{\eta_{\text{int}}}, \end{aligned} \tag{2.2}$$

where ϱ_∞, ϱ_{int}, η_∞, η_{int} stand for constant density and shear viscosity coefficient outside and inside the cylinder, respectively, $H(x-x_0)$ stands for the Heaviside function, $\overline{r} = r/R$ is dimensionless r-coordinate, and cylindrical coordinate system r, φ, z is used. Let us note that because the assumed distributions of ϱ and η may be thought to be generated by a suitable temperature field T, therefore the quantities ε_ϱ and ε_η are interrelated. In the case of an ideal gas ($\varrho \propto 1/T$, $\eta \propto \sqrt{T}$) this relationship has the form:

$$\varepsilon_\eta = \sqrt{\varepsilon_\varrho}\,. \tag{2.3}$$

The solution of the governing equations, which describe the velocity and pressure fields, will be looked for separately outside and inside the cylinder, and next these external and internal solutions will be matched using the continuity conditions for the mass and momentum flux densities at the surface of the cylinder.

3. The Euler approximation

3.1. Formulation of the problem

According to the assumptions adopted, the governing equations in the cylindrical coordinate system both outside ($\overline{r} > 1$) and inside the cylinder ($\overline{r} < 1$) can be written in the following dimensionless form:

$$\begin{aligned} \frac{1}{\overline{r}} \frac{\partial}{\partial \overline{r}} (\overline{r}\,\overline{v}_r) + \frac{1}{\overline{r}} \frac{\partial}{\partial \varphi} \overline{v}_\varphi &= 0, \\ \overline{v}_r \frac{\partial \overline{v}_r}{\partial \overline{r}} + \frac{\overline{v}_\varphi}{\overline{r}} \frac{\partial \overline{v}_r}{\partial \varphi} - \frac{\overline{v}_\varphi^2}{\overline{r}} + \frac{1}{2\overline{\varrho}} \frac{\partial \overline{\Delta p}}{\partial \overline{r}} &= 0, \\ \overline{v}_r \frac{\partial \overline{v}_\varphi}{\partial \overline{r}} + \frac{\overline{v}_\varphi}{\overline{r}} \frac{\partial \overline{v}_\varphi}{\partial \varphi} + \frac{\overline{v}_r \overline{v}_\varphi}{\overline{r}} + \frac{1}{2\overline{\varrho}} \frac{1}{\overline{r}} \frac{\partial \overline{\Delta p}}{\partial \varphi} &= 0, \end{aligned} \tag{3.1}$$

where

$$\overline{v}_\alpha := \frac{v_\alpha}{v_\infty}, \qquad \alpha = r, \varphi, \qquad \overline{\Delta p} := 2\frac{p - p_\infty}{\varrho_\infty v_\infty^2},$$

where, in turn, v_∞ and p_∞ stand for the velocity modulus and pressure at $\bar{r} = \infty$, respectively.

The boundary and matching conditions are:

$$
\begin{aligned}
&\text{at} \quad \bar{r} = \infty : \begin{cases} \bar{v}_r = \cos\varphi, \\ \bar{v}_\varphi = -\sin\varphi, \\ \overline{\Delta p} = 0; \end{cases} \\
&\text{at} \quad \bar{r} = 0 : |\bar{v}_r|,\ |\bar{v}_\varphi|,\ |\overline{\Delta p}| < \infty; \\
&\text{at} \quad \bar{r} = 1 : \begin{cases} [\![\ \bar{\varrho}\,\bar{v}_r\ ]\!] = 0, \\ [\![\ \tfrac{1}{2}\overline{\Delta p} + \bar{\varrho}\,\bar{v}_r^2\ ]\!] = 0, \\ [\![\ \bar{\varrho}\,\bar{v}_r\bar{v}_\varphi\ ]\!] = 0; \end{cases}
\end{aligned}
\tag{3.2}
$$

where

$$
[\![\ \psi\ ]\!] := \psi(\bar{r} = 1 + 0) - \psi(\bar{r} = 1 - 0) =: \psi^{\text{ext}}(\bar{r} = 1) - \psi^{\text{int}}(\bar{r} = 1).
$$

Because all the considerations will run in terms of the dimensionless variables introduced only, therefore from now on, all the bars will be ignored.

3.2. The solution

The velocity field is looked for in the form:

$$
\begin{aligned}
v_r &= f(r)\cos\varphi, \\
v_\varphi &= -g(r)\sin\varphi.
\end{aligned}
\tag{3.3}
$$

Substituting Eqs. (3.3) into Eq. $(3.1)_1$ one may obtain the following relationship between the functions f and g:

$$
g = (rf)' = rf' + f, \tag{3.4}
$$

where prime denotes the derivative with respect to r. Substituting Eqs. (3.3) into Eqs. $(3.1)_{2,3}$ and using Eq. (3.4) one may obtain:

$$
\begin{aligned}
\frac{1}{2\varrho}\frac{\partial \Delta p}{\partial r} &= r(f')^2 + ff' - \left\{r(f')^2 + 2ff'\right\}\cos^2\varphi, \\
\frac{1}{2\varrho}\frac{\partial \Delta p}{\partial \varphi} &= \left\{r^2 ff'' + rff' - r^2(f')^2\right\}\sin\varphi\cos\varphi.
\end{aligned}
\tag{3.5}
$$

Integrating Eq. $(3.5)_2$ one may obtain:

$$
\begin{aligned}
\frac{1}{2\varrho}\Delta p &= \phi(r) - \frac{1}{2}\chi(r)\cos^2\varphi, \\
\chi(r) &:= r^2 ff'' + rff' - r^2(f')^2.
\end{aligned}
\tag{3.6}
$$

Comparison of Eqs. (3.6) and $(3.5)_1$ gives, after some algebra:

$$\phi' = \frac{1}{2}(\chi - f^2)', \qquad \left\{\frac{f''}{f} + \frac{3}{r}\frac{f'}{f}\right\}' = 0. \tag{3.7}$$

From Eq. $(3.7)_2$ one has immediately:

$$f'' + \frac{3}{r} f' = \beta f, \tag{3.8}$$

where β stands for an integration constant. If $\beta = 0$, then Eq. (3.8) gives

$$f_0 = C_1 + \frac{C_2}{r^2}, \tag{3.9}$$

where C_1 and C_2 stand for integration constants.

If $\beta = -c^2 < 0$, then after substituting:

$$f_- = \frac{\psi_-(\zeta)}{r}, \qquad \zeta = c\,r,$$

Eq. (3.8) is transformed to the Bessel equation of the first order, therefore:

$$f_- = \frac{1}{r}\{C_3 J_1(c\,r) + C_4 Y_1(c\,r)\}, \tag{3.10}$$

where C_3 and C_4 stand for integration constants, J_1 – for the Bessel function of the first kind and the first order, and Y_1 – for the Bessel function of the second kind (the Neumann or the Weber function) and of the first order.

If $\beta = \hat{c}^2 > 0$, then in the same way one may obtain:

$$f_+ = \frac{1}{r}\{C_5 I_1(\hat{c}\,r) + C_6 K_1(\hat{c}\,r)\}, \tag{3.11}$$

where C_5 and C_6 are integration constants, I_1 is the modified Bessel function of the first kind and the first order, and K_1 is the modified Beessel function of the first kind (the MacDonald function) and the first order.

The boundary condition at infinity can be fulfilled only by the function given by Eq. (3.9) with $C_1 = 1$. The boundary condition at $r = 0$ can be satisfied only by the functions given by Eqs. (3.10) and (3.11) with $C_4 = 0 = C_6$. The matching conditions at $r = 1$ can be satisfied only by the pair: f_0 as an external solution (outside the cylinder) and f_- as an internal one (inside the cylinder). Thus, we obtain:

$$\begin{aligned} f^{\text{ext}} &= 1 - 2\frac{a}{r^2}, \qquad r > 1, \\ f^{\text{int}} &= b\frac{J_1(c\,r)}{r}, \qquad r < 1, \end{aligned} \tag{3.12}$$

where the superscripts ext and int refer to the external and to the internal region of the cylinder, respectively, and a, b, c stand for constants (which have to be determined from the matching conditions at $r = 1$).

Thus, the velocity and pressure fields outside and inside the cylinder, which satisfy the boundary conditions, may be written in the form:

$$
\begin{aligned}
r > 1: \qquad v_r^{\mathrm{ext}} &= \left(1 - \frac{2a}{r^2}\right)\cos\varphi, \\
v_\varphi^{\mathrm{ext}} &= -\left(1 + \frac{2a}{r^2}\right)\sin\varphi, \\
\Delta p^{\mathrm{ext}} &= -\frac{4a}{r^2}\left(1 + \frac{a}{r^2}\right) + \frac{8a}{r^2}\cos^2\varphi, \\
r < 1: \qquad v_r^{\mathrm{int}} &= b\frac{J_1}{r}\cos\varphi, \\
v_\varphi^{\mathrm{int}} &= -b\left(cJ_0 - \frac{J_1}{r}\right)\sin\varphi, \\
\Delta p^{\mathrm{int}} &= d - \varepsilon_\varrho b^2\left\{\left(cJ_0 - \frac{J_1}{r}\right)^2 + J_1^2c^2\right\} \\
&\quad + \varepsilon_\varrho b^2\left\{(J_0^2 + J_1^2)c^2 - 2\frac{J_0J_1}{r}c\right\}\cos^2\varphi,
\end{aligned}
\tag{3.13}
$$

where the abbreviation $J_n = J_n(c\,r)$, $n = 0, 1$ was used.

The constants: a, b, c, d have to be determined from the matching conditions at the cylinder surface (Eqs. $(3.2)_{5-7}$). In fact, using these conditions one may obtain, after some algebra, the following set for these constants:

$$
\begin{aligned}
a &= \frac{1}{2M}\{ch_0 - h_1 - \varepsilon_\varrho h_1\}, \\
b &= \frac{2}{M}, \\
&(h_0^2 + h_1^2)\varepsilon_\varrho c^2 = M^2, \\
d &= -4a(1 + a) + \varepsilon_\varrho b^2\{(h_0c - h_1)^2 + c^2h_1^2\} \\
&= 2 - 4a(1 + a) - \varepsilon_\varrho b^2 h_1^2,
\end{aligned}
\tag{3.14}
$$

where

$$M = ch_0 - h_1 + \varepsilon_\varrho h_1;$$

the second formula for d may be obtained, after some algebra, from the first one using properties of Eqs. $(3.14)_{1-3}$; and, for distinguishing, the abbreviation $h_n = J_n(c)$, $n = 0, 1$ was used.

The scheme of calculations is as follows. First, the third equation is solved with respect to $\varepsilon_\varrho(c)$, and next the inverse function $c(\varepsilon_\varrho)$ is numerically calculated.

Then from the first and the second equations the quantities $a(\varepsilon_\varrho)$ and $b(\varepsilon_\varrho)$ are obtained. Finally, from the fourth equation the quantity $d(\varepsilon_\varrho)$ is calculated. In this way all the constants considered are obtained (in numerical way) as the functions of ε_ϱ:

$$
\begin{aligned}
&a = a(\varepsilon_\varrho) \cong a_0 - \alpha_1\sqrt{\varepsilon_\varrho}, && a_0 = 0.5, && \alpha_1 = 0.4773,\\
&b = b(\varepsilon_\varrho) \cong b_0 + \alpha_2\frac{1}{\sqrt{\varepsilon_\varrho}}, && b_0 = 0.3276, && \alpha_2 = 1.6405,\\
&c = c(\varepsilon_\varrho) \cong c_0 - \alpha_3\sqrt{\varepsilon_\varrho}, && c_0 = 1.8412, && \alpha_3 = 1.6141,\\
&d = d(\varepsilon_\varrho) \cong d_0 - \alpha_4\sqrt{\varepsilon_\varrho}, && d_0 = 0.0888, && \alpha_4 = 0.3639,
\end{aligned}
\tag{3.15}
$$

where the approximate relationships represent the asymptotical behaviour of these constants as $\varepsilon_\varrho \to 0$. Substituting the constants calculated into Eqs. (3.13) we obtain the final solution of the problem examined.

The asymptotical behaviour of the flow functions outside and inside the cylinder at small ε_ϱ is, according to the structure of the solution, completely determined by the asymptotical behaviour of the functions f^{ext} and f^{int}, which are given by the formulae:

$$
\begin{aligned}
&f^{\text{ext}} \cong 1 - \frac{1}{r^2}\left(1 - 2\alpha_1\sqrt{\varepsilon_\varrho}\right),\\
&f^{\text{int}} \cong -\alpha_5 J_0(c_0 r) + \frac{J_1(c_0 r)}{r}\left(\alpha_6 + \frac{\alpha_7}{\sqrt{\varepsilon_\varrho}}\right),\\
&\alpha_5 = 2.6480, \qquad \alpha_6 = 1.7658, \qquad \alpha_7 = 1.6405,
\end{aligned}
$$

where α_1 is given by Eq. $(3.15)_{1/3}$, and c_0 – by Eq. $(3.15)_{3/2}$.

3.3. Results

From the results given in the previous subsection one may obtain all the information about the flow examined. Examples of two types of such an information will be present.

The information of the first type concerns the flow fields at a given ε_ϱ. The example value $\varepsilon_\varrho = 2.5 \times 10^{-2}$ is assumed. Thus, the lower half of Fig. 1. presents the streamlines picture. Figure 2 presents the dimensionless x-coordinate of velocity:

$$v_x = v_r \cos\varphi - v_\varphi \sin\varphi$$

at the flow symmetry plane ($\varphi = 0, \pi$, respectively) as a function of dimensionless x-coordinate (as referred to the cylinder radius). Figure 3 presents the dependence of the dimensionless pressure difference Δp on the dimensionless x-coordinate at the flow symmetry plane.

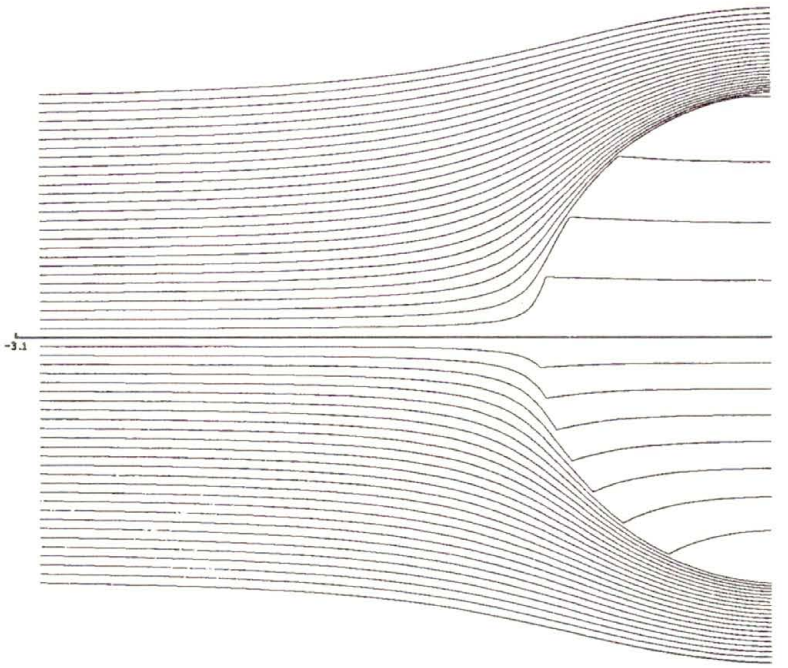

FIG. 1. Streamline pictures for the flow through the sphere in the Euler (the lower half) and Stokes (the upper half) approximations under the assumptions: $\varepsilon_\eta = \sqrt{\varepsilon_\varrho}$, $\varepsilon_\varrho = 2.5 \times 10^{-2}$.

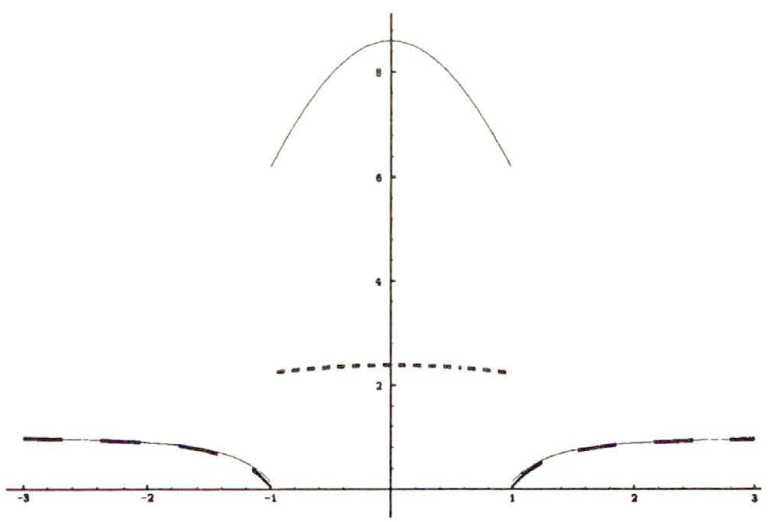

FIG. 2. Dimensionless velocity (as referred to v_∞) at the flow symmetry axis as a function of the dimensionless z-coordinate (as referred to R) under the same assumptions about ε_η and ε_ϱ as in the case of Fig. 1, in the Euler (solid line) and Stokes (dashed line) approximations.

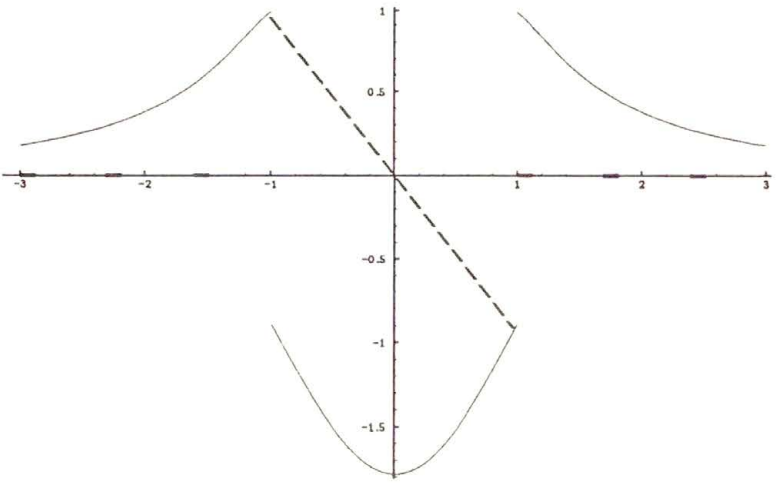

FIG. 3. Scaled relative pressure at the flow symmetry axis for $\varepsilon_\eta = \sqrt{\varepsilon_\varrho}$, $\varepsilon_\varrho = 2.5 \times 10^{-2}$;
solid line – the Euler approximation: $2(p - p_\infty)/(\varrho_\infty v_\infty^2)$,
dashed line – the Stokes approximation: $2(p - p_\infty)/(\varrho_\infty v_\infty^2)(\mathrm{Re})/(20)$.

The information of the second type concerns the characteristics of the flow considered as functions of ε_ϱ, as for example: velocity and pressure at the symmetry plane at the center and the boundary of the cylinder (Fig. 4 a, Fig. 5 a) ([1]):

$$
\begin{aligned}
& v_x^{\mathrm{ext}}(1) = 1 - 2a && \cong 2\alpha_1 \sqrt{\varepsilon_\varrho}, && \\
& v_x^{\mathrm{int}}(1) = bJ_1(c) && \cong \alpha_8 + \frac{\alpha_9}{\sqrt{\varepsilon_\varrho}}, && \alpha_8 = 0.1906,\ \alpha_9 = 0.9546, \\
& v_x^{\mathrm{int}}(0) = \tfrac{1}{2}bc && \cong -\alpha_{10} + \frac{\alpha_{11}}{\sqrt{\varepsilon_\varrho}}, && \alpha_{10} = 1.0224,\ \alpha_{11} = 1.5102, \\
& [\![v_x]\!] = 1 - 2a - bJ_1(c) && \cong -v_x^{\mathrm{int}}(1), && \\
& \Delta p^{\mathrm{ext}}(1) = 4a(1-a) && \cong 1 - \alpha_{12}\varepsilon_\varrho, && \alpha_{12} = 0.9112, \\
& \Delta p^{\mathrm{int}}(1) = d - \varepsilon_\varrho b^2 J_1^2(c) && \cong -\alpha_{13} - \alpha_{14}\sqrt{\varepsilon_\varrho}, && \alpha_{13} = 0.8223,\ \alpha_{14} = 0.7278, \\
& \Delta p^{\mathrm{int}}(0) = d - \frac{1}{4}\varepsilon_\varrho b^2 c^2 && \cong -\alpha_{15} + \alpha_{16}\sqrt{\varepsilon_\varrho}, && \alpha_{15} = 2.1920,\ \alpha_{16} = 2.7242, \\
& [\![\Delta p]\!] = 4a(1-a) - d && && \\
& \qquad + \varepsilon_\varrho b^2 J_1^2(c) && \cong \alpha_{17} + \alpha_{14}\sqrt{\varepsilon_\varrho}, && \alpha_{17} = 1.8223,
\end{aligned}
\tag{3.16}
$$

([1]) Note, that the part of the gas flux flowing through the cylinder (per unit of its length) as referred to the flux incoming from infinity is given by $v_x^{\mathrm{ext}}(1)$ (if follows from an immediate calculation and application of Eqs. (3.4), $(3.2)_5$ and $(3.16)_1$).

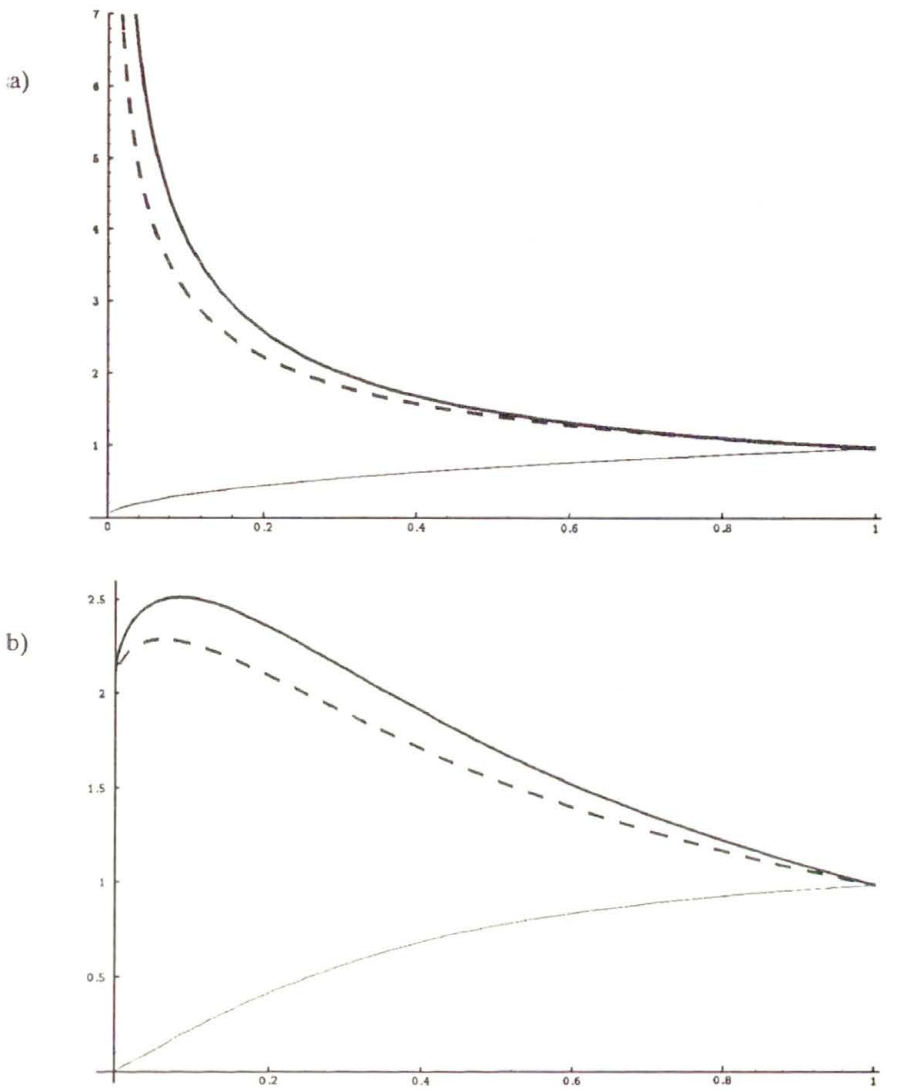

FIG. 4. Dependence of $\overline{v}_z^{\text{ext}}(1)$ (solid line), $\overline{v}_z^{\text{int}}(1)$ (dashed line) and $\overline{v}_z^{\text{int}}(0)$ (bold line) on ε_ϱ for the flow through the cylinder in the Euler (a) and Stokes (b) approximation under the assumption: $\varepsilon_\eta = \sqrt{\varepsilon_\varrho}$.

where the first column represents the exact formulae, the second one – the asymptotic formulae for small ε_ϱ; α_1 is given by Eq. $(3.15)_{1/3}$;

$$\psi(1) := \psi(\varphi = \pi, r = 1),$$

$$\psi(0) := \psi(\varphi = \pi, r = 0),$$

and $[\![\psi]\!]$ is defined by the equation following Eqs. (3.2).

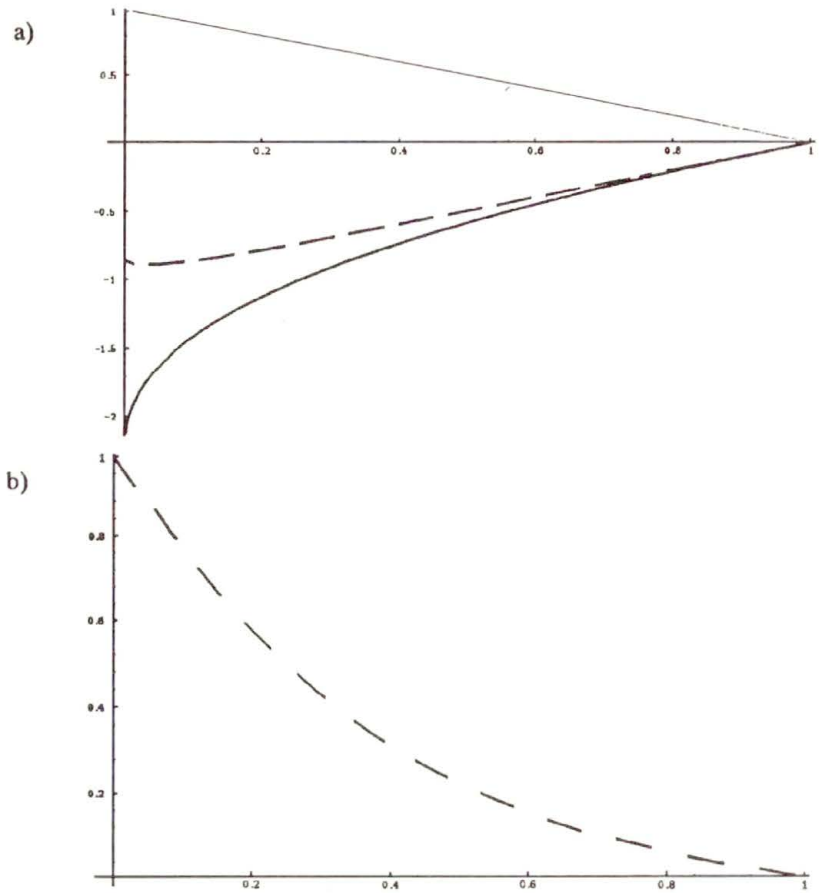

FIG. 5. Dependence of $\overline{\Delta p}^{\text{ext}}(1)$ (solid line), $\overline{\Delta p}^{\text{int}}(1)$ (dashed line) and $\overline{\Delta p}^{\text{int}}(0)$ (bold line) on ε_ϱ for the flow through the cylinder in the Euler approximation (a) and $\frac{Re}{20}\overline{\Delta p}^{\text{int}}(1)$ in the Stokes approximation (b) under the same assumptions as in the case of Fig. 3.

4. The Stokes approximation

4.1. The problem

The governing equations in this case may be written in the form (in terms of the same dimensionless variables as previously):

$$
\begin{aligned}
&\frac{1}{r}\frac{\partial}{\partial r}(r\,v_r) + \frac{1}{r}\frac{\partial v_\varphi}{\partial \varphi} = 0,\\
&\frac{1}{2}\mathrm{Re}\,\frac{1}{\eta}\frac{\partial \Delta p}{\partial r} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial v_r}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 v_r}{\partial \varphi^2} - \frac{2}{r^2}\frac{\partial v_\varphi}{\partial \varphi} - \frac{v_r}{r^2},\\
&\frac{1}{2}\mathrm{Re}\,\frac{1}{\eta}\frac{1}{r}\frac{\partial \Delta p}{\partial \varphi} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial v_\varphi}{r}\right) + \frac{1}{r^2}\frac{\partial^2 v_\varphi}{\partial \varphi^2} + \frac{2}{r^2}\frac{\partial v_r}{\partial \varphi} - \frac{v_\varphi}{r^2},
\end{aligned}
\tag{4.1}
$$

where the Reynolds number

$$\mathrm{Re} = \frac{\varrho_\infty v_\infty R}{\eta_\infty}$$

plays the role of the scale factor only, and η stands for the dimensionless shear viscosity coefficient (dimension coefficient as referred to η_∞).

The boundary conditions at $r = \infty$ and $r = 0$ are the same, as in the Euler approximation (Eqs. $(3.2)_{1-4}$), the matching conditions are:

$$\text{at } r = 1: \qquad \begin{aligned} &[\![\varrho v_r]\!] = 0, \\ &\left[\!\!\left[\frac{1}{2} \mathrm{Re}\, \Delta p - 2\eta \frac{\partial v_r}{\partial r} \right]\!\!\right], \\ &\left[\!\!\left[\eta \left(\frac{\partial v_\varphi}{\partial r} - \frac{v_\varphi}{r} + \frac{1}{r} \frac{\partial v_r}{\partial \varphi} \right) \right]\!\!\right]. \end{aligned} \tag{4.2}$$

4.2. The solution

Applying the same procedure as in the previous case, one may find the general solution of the problem outside and inside the cylinder, which satisfies the boundary conditions, namely:

$$\begin{aligned} r > 1: \qquad v_r^{\mathrm{ext}} &= \left(1 - \frac{2\tilde{a}}{r^2}\right) \cos\varphi, \\ v_\varphi^{\mathrm{ext}} &= -\left(1 + \frac{2\tilde{a}}{r^2}\right) \sin\varphi, \\ \mathrm{Re}\, \Delta p^{\mathrm{ext}} &= 0, \\ r < 1: \qquad v_r^{\mathrm{int}} &= (\tilde{b} + \tilde{c} r^2) \cos\varphi, \\ v_\varphi^{\mathrm{int}} &= -(\tilde{b} + 3\tilde{c} r^2) \sin\varphi, \\ \mathrm{Re}\, \Delta p^{\mathrm{int}} &= \tilde{d} + \frac{16\tilde{c}}{\varepsilon_\eta} r \cos\varphi. \end{aligned} \tag{4.3}$$

The constants $\tilde{a}$, $\tilde{b}$, $\tilde{c}$, $\tilde{d}$ have to be determined from the matching conditions at the cylider surface. In fact, substituting Eqs. (4.3) into Eqs. (4.2) we obtain the following set of equations for the constants considered:

$$\begin{aligned} 1 - 2\tilde{a} &= \varepsilon_\varrho (\tilde{b} + \tilde{c}), \\ 2\tilde{a}\varepsilon_\eta &= -c, \\ \tilde{d} &= 0. \end{aligned} \tag{4.4}$$

It is seen that we have two equations for three constants: $\tilde{a}$, $\tilde{b}$ and $\tilde{c}$.

Thus, in order to obtain a unique solution we should adopt an additional condition, and the continuity condition for the tangent component of velocity at

the cylinder surface is assumed (2):

$$[\![v_\varphi]\!] = 0, \tag{4.5}$$

which leads to the following additional equation:

$$1 + 2\tilde{a} = \tilde{b} + 3\tilde{c}. \tag{4.6}$$

Now, solving Eqs. (4.4) and Eq. (4.6) we obtain:

$$\begin{aligned} \tilde{a} &= \frac{1}{2}\frac{1-\varepsilon_\varrho}{1+\varepsilon_\varrho(1+2\varepsilon_\eta)} \cong \frac{1}{2} - \varepsilon_\varrho, \\ \tilde{b} &= \frac{2+3\varepsilon_\eta - \varepsilon_\varrho\varepsilon_\eta}{1+\varepsilon_\varrho(1+2\varepsilon_\eta)} \cong 2 + 3\sqrt{\varepsilon_\varrho}, \\ \tilde{c} &= -\frac{(1-\varepsilon_\varrho)\varepsilon_\eta}{1+\varepsilon_\varrho(1+2\varepsilon_\eta)} \cong -\sqrt{\varepsilon_\varrho}, \\ \tilde{d} &= 0, \end{aligned} \tag{4.7}$$

where the first equation in a given line represents the exact relationship, and the second one – the asymptotical expression as $\varepsilon_\varrho \to 0$ (under the assumption $\varepsilon_\eta = \sqrt{\varepsilon_\varrho}$).

Thus, Eqs. (4.3) with Eqs. (4.7) represent the solution of the problem as expressed by Eqs. (4.1), Eqs. $(3.2)_{1-4}$ and Eqs. (4.2), which is unique in the class of functions specified by Eqs. (3.3) (and under the assumption expressed by Eq. (4.5)).

4.3. Results

Similarly to the case of the Euler approximation, two types of information, which is contained in the formulae given in the previous subsection, will be presented.

The information of the first type concerns the flow fields at a given ε_ϱ. The example value $\varepsilon_\varrho = 2.5 \times 10^{-2}$ is adopted, and ε_η as given by Eq. (2.3) is assumed. Thus, the upper half of Fig. 1 presents the streamlines picture. Figure 2 presents the dimensionless x-coordinate of velocity (see the formula given at the beginning of Subsec. 3.3.) at the flow symmetry plane ($\varphi = \pi, 0$, respectively). Figure 3 presents the dependence of the dimensionless pressure difference on the dimensionless x-coordinate at the symmetry plane.

The information of the second type concerns, as previously, the characteristics of the flow considered as functions of ε_ϱ (under the same assumption about ε_η

(2) For comments on this assumption – see [3].

as above), namely – velocity and pressure at the symmetry plane at the center and at the boundary of the cylinder, in the same convention as in the case of the Euler approximation (Eqs. (3.16)) (Fig. 4 b, Fig. 5 b) ([3]):

$$
\begin{aligned}
& v_x^{\mathrm{ext}}(1) && = 1 - 2\tilde{a} && \cong 2\varepsilon_\varrho, \\
& v_x^{\mathrm{int}}(1) && = \tilde{b} + \tilde{c} && \cong 2 + 2\sqrt{\varepsilon_\varrho}, \\
& v_x^{\mathrm{int}}(0) && = \tilde{b} && \cong 2 + 3\sqrt{\varepsilon_\varrho}, \\
& [\![v_x]\!] && = 1 - 2\tilde{a} - \tilde{b} - \tilde{c} && \cong -v_x^{\mathrm{int}}(1), \\
& \mathrm{Re}\, \Delta p^{\mathrm{ext}}(1) && = 0, && \\
& \mathrm{Re}\, \Delta p^{\mathrm{int}}(1) && = -16\frac{\tilde{c}}{\varepsilon_\eta} && \cong 16 - 32\varepsilon_\varrho, \\
& \mathrm{Re}\, \Delta p^{\mathrm{int}}(0) && = 0, && \\
& \mathrm{Re}\, [\![\Delta p]\!] && = -\mathrm{Re}\, \Delta p^{\mathrm{int}}(1). &&
\end{aligned}
\tag{4.8}
$$

5. Conclusions

Comparing the results obtained for the cylindrical case (in particular – the asymptotic relationships) with those for the spherical case (see [1] and [3]) one may conclude, that:

1. The velocity and the pressure fields and their dependence on ε_ϱ in the Euler approximation are very similar in both flow geometries; there occur only relatively small quantitative differences; the influence of low density region on the flow fields is, in general, greater in the case of cylinder as compared to that in the case of a sphere;

2. The same concerns the flow through the cylinder as compared to that through the sphere in the Stokes approximation;

3. The similarities and differences between the flow through the cylinder in the Euler and in the Stokes approximations are, generally, the same as in the case of flow through the sphere (see discussion in [3]).

Acknowledgments

The authors are deeply grateful to Prof. Z. PERADZYŃSKI for valuable comments.

([3]) The part of the gas flux flowing through the cylinder (per unit of its length) as referred to the flux incoming from infinity is given by $v_x^{\mathrm{ext}}(1)$.

References

1. K.G. GUS'KOV, YU.P. RAYZER and S.T. SURZHIKOV, *On an observed speed of the optical discharge slow propagation* [in Russian], Kvantovaya Elektronika, **17**, 7, 937–942, 1990.
2. Z. PERADZYŃSKI and E. ZAWISTOWSKA, *Flow in a region with large density variations*, Arch. Mech., **46**, 6, 881–891, 1984.
3. Z. PŁOCHOCKI, B. KAŹMIERCZAK and Z. PERADZYŃSKI, *The stationary Stokes flow through a spherical region with large variations of density and viscosity coefficient*, Arch. Mech., **48**, 4, 765–778, 1996.

POLISH ACADEMY OF SCIENCES
INSTITUTE OF FUNDAMENTAL TECHNOLOGICAL RESEARCH

Received March 22, 1996.

Archi. Mech., 48, 5, pp. 947–953, Warszawa 1996

BRIEF NOTES

A note on the hyperelastic constitutive equation for rotated Biot stress

K. WIŚNIEWSKI and E. TURSKA (WARSZAWA)

THE FORWARD-ROTATED BIOT STRESS and the right stretch strain are defined, and the virtual work of the rotated stress is found. It is shown that it involves a corotational variation of the Green-McInnis-Naghdi type. For the strain energy assumed in terms of principal invariants of the right stretching tensor, a constitutive equation and a constitutive (4th rank) operator for the Biot stress is derived. Subsequently, they are subjected to the rotate-forward operation, and it is demonstrated how their structure is carried over to the rotated measures.

1. Introduction

THE CO-ROTATIONAL FORMULATIONS are applied to many problems of mechanics, ranging from finite strain plasticity to large rotation shells, mostly due to relative simplicity of manipulating on orthogonal rotation tensors.

In finite strain plasticity, see e.g. DIENES [3] and JOHNSON, BAMMANN [4], the so-called rotated description is based on a back-rotated Kirchhoff stress $\Sigma = \mathbf{Q}^T \tau \mathbf{Q}$ and a back-rotated spatial rate of deformation $\mathbf{D} = \mathbf{Q}^T \mathbf{d}\, \mathbf{Q}$, where $\mathbf{d} = \mathrm{sym}(\dot{\mathbf{F}}\, \mathbf{F}^{-1})$. The rotated measures are exploited to define a constitutive equation, which later is converted to $\overset{\circ}{\tau}$ and $\mathbf{d}$, where $\overset{\circ}{\tau}$ is the Green–McInnis–Naghdi objective stress rate.

It was noticed by several authors, e.g. see the introduction to CRISFIELD [2], that nonlinearities resulting from large rotations of beams or shells can be eliminated if corotational local frames are introduced. Among recent works using the corotational frames, we would like to mention contributions of RANKIN, BROGAN [6], SIMO [7], SIMO, VU-QUOC [9], and CRISFIELD [2]. In RANKIN, BROGAN [6] a general framework to handle large rotations has been constructed, in which already existing linear finite elements can be embedded. In [7] and [9] a finite strain/rotation beam model for dynamics has been consistently derived from three-dimensional equations. In [2] an issue of symmetry of the tangent operator for the finite rotation beam has been undertaken. In all these papers separation of frame rotations simplified the equations.

In the present note we extend the concept of the corotational frame used for beams and shells and introduce a forward-rotated description: the rotated

stress and strain measures, and the corotational variation. We address in detail an issue of a hyperelastic constitutive equation and a constitutive operator for the rotated measures as derived from the constitutive relations for the Biot stress. The forward-rotated description, as a general concept, can be found convenient in problems involving independent rotation fields, not only in beam or shell theories but also in three-dimensional elasticity formulated as e.g. in SIMO, FOX, HUGHES [8].

Notation

Small letters – vectors, capital letters – 2nd rank tensors, capital letters with a superscribed digit 4 – 4th rank tensors, dots $\cdot$ – scalar products, colons : – contractions of a 4th and a 2nd rank tensors yielding a 2nd rank tensor, $\otimes$ – tensorial products.

2. Rotated stress and strain

In this section the rotated strain and stress measures are introduced and a corresponding form of the virtual work of stress is given.

The Cauchy (true) stress, $\mathbf{T}$, can be expressed in terms of other stress measures as follows, see e.g. OGDEN [5],

$$\mathbf{T} = J^{-1}\boldsymbol{\tau} = J^{-1}\mathbf{P}\mathbf{F}^T = J^{-1}\mathbf{F}\mathbf{S}\mathbf{F}^T, \tag{2.1}$$

where $\boldsymbol{\tau}$ is the Kirchhoff stress, $\mathbf{P}$ is the 1st Piola–Kirchhoff stress, (its transpose is a nominal stress), $\mathbf{S}$ is the 2nd Piola–Kirchhoff stress. Besides, $\mathbf{F}$ denotes the gradient of deformation, and $J = \det \mathbf{F}$.

Let us introduce a symmetric Biot stress tensor, $\mathbf{T}^B \equiv \mathrm{sym}\,(\mathbf{Q}^T\mathbf{P})$. The rotation tensor $\mathbf{Q} \in SO(3)$ is obtained from the polar decomposition of the deformation gradient. The Biot stress $\mathbf{T}^B$ and the right stretch strain $\mathbf{E}$ are work conjugates because the virtual work of stress can be expressed as follows

$$\mathbf{P} \cdot \delta\mathbf{F} = \mathbf{T}^B \cdot \delta\mathbf{E}, \tag{2.2}$$

where $\mathbf{E} = \mathbf{U} - \mathbf{I}$ is the right stretch strain, and $\mathbf{U} \equiv (\mathbf{F}^T\mathbf{F})^{1/2}$ is the right stretching tensor. This tensor appears also in the (right) polar decomposition of the deformation gradient, $\mathbf{F} = \mathbf{Q}\mathbf{U}$. On the basis of Eq. (2.1), the Biot stress is related to other stress measures in the following way

$$\mathbf{T}^B \equiv \mathrm{sym}\,(\mathbf{Q}^T\mathbf{P}) = \mathrm{sym}\,(\mathbf{Q}^T\boldsymbol{\tau}\mathbf{F}^{-T}) = \mathrm{sym}\,(\mathbf{U}\mathbf{S}). \tag{2.3}$$

The Biot stress tensor $\mathbf{T}^B$ and the right stretch strain $\mathbf{E}$ can be used to introduce a set of rotated measures defined as follows

$$\mathbf{T}^* \equiv \mathbf{Q}\mathbf{T}^B\mathbf{Q}^T, \qquad \mathbf{E}^* \equiv \mathbf{Q}\mathbf{E}\mathbf{Q}^T, \tag{2.4}$$

for which the virtual work of stress (2.2) yields

$$\mathbf{T}^B \cdot \delta\mathbf{E} = \mathbf{T}^* \cdot \overset{\circ}{\delta}\mathbf{E}^*, \tag{2.5}$$

where

$$\overset{\circ}{\delta}\mathbf{E}^* \equiv \mathbf{Q}\,\delta\mathbf{E}\,\mathbf{Q}^T = \mathbf{Q}\,\delta(\mathbf{Q}^T\mathbf{E}^*\mathbf{Q})\,\mathbf{Q}^T.$$

The above corotational variation corresponds with the Green–McInnis–Naghdi objective time derivative, and consists of the rotate-back, take a variation and rotate-forward operations. The definition (2.4) yields

$$\mathbf{E}^* = \mathbf{V} - \mathbf{I}, \tag{2.6}$$

where $\mathbf{V} = \mathbf{Q}\,\mathbf{U}\,\mathbf{Q}^T$ is a left stretching tensor defined as $\mathbf{V} \equiv (\mathbf{F}\,\mathbf{F}^T)^{1/2}$. Hence, $\mathbf{E}^*$ is the left stretch strain. The rotated Biot stress is related to other stress measures as follows

$$\mathbf{T}^* = \operatorname{sym}(\mathbf{P}\,\mathbf{Q}^T) = \operatorname{sym}(\boldsymbol{\tau}\,\mathbf{V}^{-1}) = \operatorname{sym}(\mathbf{F}\,\mathbf{S}\,\mathbf{V}^{-1}). \tag{2.7}$$

We can see that $\mathbf{T}^*$ is different than other spatial stress measures, such as Cauchy stress $\mathbf{T}$ or Kirchhoff stress $\boldsymbol{\tau}$.

3. Constitutive equation for rotated measures

In this section a constitutive equation and a constitutive operator for $\mathbf{T}^B$ and $\mathbf{U}$ are introduced for an isotropic hyperelastic (Green) material. Next, the same constitutive equation and the constitutive operator are expressed in terms of the rotated tensors, $\mathbf{T}^*$ and $\mathbf{V}$.

Let us assume the existence of a strain energy function $W(\mathbf{U})$. On arguments discussed e.g. in OGDEN [5], a strain energy given in terms of $\mathbf{U}$ is objective, and provides a response function, which is invariant under an observer transformation. On the basis of the representation theorem for isotropic functions, we can write

$$W(\mathbf{U}) = \tilde{W}(I_1(\mathbf{U}), I_2(\mathbf{U}), I_3(\mathbf{U})), \tag{3.1}$$

where the principal invariants of $\mathbf{U}$ are defined as follows

$$I_1(\mathbf{U}) = \operatorname{tr}\mathbf{U}, \qquad I_2(\mathbf{U}) = \frac{1}{2}\left[(\operatorname{tr}\mathbf{U})^2 - \operatorname{tr}\mathbf{U}^2\right], \qquad I_3(\mathbf{U}) = \det\mathbf{U}. \tag{3.2}$$

A constitutive equation for the Biot stress tensor is defined as

$$\mathbf{T}^B \equiv \frac{\partial W(\mathbf{U})}{\partial \mathbf{U}} = \frac{\partial \tilde{W}(I_1(\mathbf{U}), I_2(\mathbf{U}), I_3(\mathbf{U}))}{\partial \mathbf{U}}. \tag{3.3}$$

From the chain rule of differentiation we obtain

$$\frac{\partial \tilde{W}}{\partial \mathbf{U}} = \frac{\partial \tilde{W}}{\partial I_1}\frac{\partial I_1}{\partial \mathbf{U}} + \frac{\partial \tilde{W}}{\partial I_2}\frac{\partial I_2}{\partial \mathbf{U}} + \frac{\partial \tilde{W}}{\partial I_3}\frac{\partial I_3}{\partial \mathbf{U}}. \tag{3.4}$$

Taking into account that

$$\frac{\partial I_1}{\partial \mathbf{U}} = \mathbf{I}, \qquad \frac{\partial I_2}{\partial \mathbf{U}} = I_1\mathbf{I} - \mathbf{U}, \qquad \frac{\partial I_3}{\partial \mathbf{U}} = I_3\mathbf{U}^{-1}, \tag{3.5}$$

the constitutive equation can be rewritten as a polynomial of $\mathbf{U}$

$$\mathbf{T}^B = \beta_0\mathbf{I} + \beta_1\mathbf{U} + \beta_2\mathbf{U}^{-1}, \tag{3.6}$$

where β_0, β_1 and β_2 are scalar coefficients depending on the invariants and derivatives of $\tilde{W}$ with respect to the invariants. Note that using the Cayley–Hamilton theorem, the above equation can be converted to a second order polynomial of $\mathbf{U}$. A variation of stress with respect to the strain can be written as

$$\delta\mathbf{T}^B = \frac{\partial \mathbf{T}^B}{\partial \mathbf{U}} : \delta\mathbf{U} = \overset{4}{\mathbf{C}} : \delta\mathbf{U}, \tag{3.7}$$

where the constitutive operator (elasticity tensor) can be defined as a 4-th rank tensor

$$\overset{4}{\mathbf{C}} \equiv \frac{\partial \mathbf{T}^B}{\partial \mathbf{U}} = \frac{\partial^2 W(\mathbf{U})}{\partial \mathbf{U}\partial \mathbf{U}} = \frac{\partial^2 \tilde{W}(I_1(\mathbf{U}), I_2(\mathbf{U}), I_3(\mathbf{U}))}{\partial \mathbf{U}\partial \mathbf{U}}. \tag{3.8}$$

Hence, from the formula for the derivative of the product of a scalar and a second rank tensor we have

$$\begin{aligned}\overset{4}{\mathbf{C}} = \frac{\partial \mathbf{T}^B}{\partial \mathbf{U}} = \mathbf{I}\otimes\frac{\partial \beta_0}{\partial \mathbf{U}} + \beta_0\frac{\partial \mathbf{I}}{\partial \mathbf{U}} \\ + \mathbf{U}\otimes\frac{\partial \beta_1}{\partial \mathbf{U}} + \beta_1\frac{\partial \mathbf{U}}{\partial \mathbf{U}} + \mathbf{U}^{-1}\otimes\frac{\partial \beta_2}{\partial \mathbf{U}} + \beta_2\frac{\partial \mathbf{U}^{-1}}{\partial \mathbf{U}},\end{aligned} \tag{3.9}$$

where

$$\frac{\partial \beta_i}{\partial \mathbf{U}} = \frac{\partial \beta_i}{\partial I_k}\frac{\partial I_k}{\partial \mathbf{U}} \qquad \text{for} \quad i = 0,1,2 \quad \text{and} \quad k = 1,2,3 \tag{3.10}$$

due to the chain rule of differentiation. We can say that in Eq. (3.9) the 1st, 3rd and 5th components are expressed in terms of nine tensorial products, provided by all combinations of $\mathbf{I}$, $\mathbf{U}$ and $\mathbf{U}^{-1}$. Furthermore, for the 2nd, 4th and 6th components (and a symmetric $\mathbf{U}$), we have

$$\begin{aligned}\frac{\partial \mathbf{I}}{\partial \mathbf{U}} &= \overset{4}{\mathbf{0}}, \qquad \frac{\partial \mathbf{U}}{\partial \mathbf{U}} = \frac{1}{2}(\overset{4}{\mathbf{I}}_a + \overset{4}{\mathbf{I}}_c), \\ \frac{\partial \mathbf{U}^{-1}}{\partial \mathbf{U}} &= -\frac{1}{2}\left\{\mathbf{U}^{-1}(\mathbf{e}_i\otimes\mathbf{e}_j)\mathbf{U}^{-1}\right\}\otimes\{\mathbf{e}_i\otimes\mathbf{e}_j + \mathbf{e}_j\otimes\mathbf{e}_i\},\end{aligned} \tag{3.11}$$

where $\mathbf{e}_i$ are vectors of an orthonormal frame. The 4th rank invariants used here are $\overset{4}{\mathbf{I}}_a = \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_i \otimes \mathbf{e}_j$ and $\overset{4}{\mathbf{I}}_c = \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_j \otimes \mathbf{e}_i$, and operate on an arbitrary 2nd rank $\mathbf{A}$ as follows: $\overset{4}{\mathbf{I}}_a\mathbf{A} = \mathbf{A}$ and $\overset{4}{\mathbf{I}}_c\mathbf{A} = \mathbf{A}^T$, see [1].

The derivation of $\partial \mathbf{U}^{-1}/\partial \mathbf{U}$, being more complicated, is described below. Consider $\mathbf{I} = \mathbf{U}\,\mathbf{U}^{-1}$ as a tensor-valued function of a tensor argument. As $\mathbf{U}$ is symmetric, it may be replaced by $\frac{1}{2}[\mathbf{U} + \mathbf{U}^T]$, and thus $\mathbf{U}^{-1}$ can also be considered as a function of $\frac{1}{2}[\mathbf{U} + \mathbf{U}^T]$. A directional derivative of $\mathbf{I}$ at $\mathbf{U}$ in direction $\mathbf{A}$ yields

$$\frac{\partial \mathbf{I}}{\partial \mathbf{U}} : \mathbf{A} \equiv \left[\frac{d}{d\varepsilon}\mathbf{I}(\mathbf{U} + \varepsilon\mathbf{A})\right]_{\varepsilon=0} = \mathbf{0}, \tag{3.12}$$

where $\mathbf{A}$ is an arbitrary 2nd rank tensor. After straightforward calculations, from (3.12) we obtain

$$\frac{\partial \mathbf{U}^{-1}}{\partial \mathbf{U}} : \mathbf{A} = -\frac{1}{2}\mathbf{U}^{-1}(\mathbf{A} + \mathbf{A}^T)\mathbf{U}^{-1}. \tag{3.13}$$

To introduce a constitutive operator, we have to rewrite the above equation as a contraction of a fourth rank tensor and a second rank tensor $\mathbf{A}$. Introducing the 4th rank invariants we have

$$(\mathbf{A} + \mathbf{A}^T) = (\overset{4}{\mathbf{I}}_a + \overset{4}{\mathbf{I}}_c) : \mathbf{A} = \mathbf{e}_i \otimes \mathbf{e}_j \{[\mathbf{e}_i \otimes \mathbf{e}_j + \mathbf{e}_j \otimes \mathbf{e}_i] \cdot \mathbf{A}\}, \tag{3.14}$$

where the identity $(\mathbf{T} \otimes \mathbf{S}) : \mathbf{Q} = \mathbf{T}(\mathbf{S} \cdot \mathbf{Q})$ is used. Note that the product in the parentheses is a scalar. Substituting Eq. (3.14) into Eq. (3.13), and recovering the 4th rank tensor, we obtain

$$\frac{\partial \mathbf{U}^{-1}}{\partial \mathbf{U}} : \mathbf{A} = \left[-\frac{1}{2}\left\{\mathbf{U}^{-1}(\mathbf{e}_i \otimes \mathbf{e}_j)\mathbf{U}^{-1}\right\} \otimes \{\mathbf{e}_i \otimes \mathbf{e}_j + \mathbf{e}_j \otimes \mathbf{e}_i\}\right] : \mathbf{A}, \tag{3.15}$$

where the 4th rank tensor given by Eq. (3.11) can be easily identified. □

Having derived the constitutive equation (3.6) and the elasticity tensor (3.7) for the Biot stress $\mathbf{T}^B$, we can find the respective equations for the rotated stress $\mathbf{T}^*$. For $\mathbf{T}^B$ given by Eq. (3.6) we obtain

$$\mathbf{T}^* \equiv \mathbf{Q}\,\mathbf{T}^B\,\mathbf{Q}^T = \mathbf{Q}\left(\beta_0\mathbf{I} + \beta_1\mathbf{U} + \beta_2\mathbf{U}^{-1}\right)\mathbf{Q}^T. \tag{3.16}$$

On the basis of identities

$$\mathbf{Q}\,\mathbf{I}\,\mathbf{Q}^T = \mathbf{I}, \qquad \mathbf{Q}\,\mathbf{U}\,\mathbf{Q}^T = \mathbf{V}, \qquad \mathbf{Q}\,\mathbf{U}^{-1}\mathbf{Q}^T = \mathbf{V}^{-1} \tag{3.17}$$

we have

$$\mathbf{T}^* = \beta_0\mathbf{I} + \beta_1\mathbf{V} + \beta_2\mathbf{V}^{-1}, \tag{3.18}$$

which is a polynomial of the left stretching tensor $\mathbf{V}$.

Next, we find the elasticity tensor for the rotated stress $\mathbf{T}^*$,

$$\overset{\circ}{\delta}\mathbf{T}^* \equiv \mathbf{Q}\,\delta\mathbf{T}^B\,\mathbf{Q}^T = \mathbf{Q}\left[\overset{4}{\mathbf{C}} : \frac{1}{2}\delta\mathbf{U}\right]\mathbf{Q}^T = \mathbf{Q}\left[\frac{\partial\mathbf{T}^B}{\partial\mathbf{U}} : \delta\mathbf{U}\right]\mathbf{Q}^T, \tag{3.19}$$

where the expression for $\partial\mathbf{T}^B/\partial\mathbf{U}$ is given by Eq. (3.9). Consider the 1st, 3rd and 5th component of this equation contracted with $\delta\mathbf{U}$. As mentioned earlier, these components contain nine tensorial products, and the contraction can be written as $(\mathbf{A}_i \otimes \mathbf{A}_j) : \delta\mathbf{U}$, where $\mathbf{A}_i, \mathbf{A}_j \in \{\mathbf{I}, \mathbf{U}, \mathbf{U}^{-1}\}$. Furthermore, $(\mathbf{A}_i \otimes \mathbf{A}_j) : \delta\mathbf{U} = \mathbf{A}_i(\mathbf{A}_j \cdot \delta\mathbf{U})$, where in the parentheses we have a scalar. Hence,

$$\mathbf{Q}\left[(\mathbf{A}_i \otimes \mathbf{A}_j) : \delta\mathbf{U}\right]\mathbf{Q}^T = \left[\mathbf{Q}\,\mathbf{A}_i\mathbf{Q}^T\right](\mathbf{A}_j \cdot \delta\mathbf{U}) = \mathbf{B}_i(\mathbf{A}_j \cdot \delta\mathbf{U}) \tag{3.20}$$

where $\mathbf{Q}\,\mathbf{A}_i\mathbf{Q}^T = \mathbf{B}_i$ and $\mathbf{B}_i \in \{\mathbf{I}, \mathbf{V}, \mathbf{V}^{-1}\}$ in accordance with Eq. (3.17). Besides, for the scalar product we have

$$\begin{aligned}\mathbf{A}_j \cdot \delta\mathbf{U} = \operatorname{tr}(\mathbf{A}_j\delta\mathbf{U}) = \operatorname{tr}\left(\mathbf{Q}^T(\mathbf{Q}\,\mathbf{A}_j\mathbf{Q}^T)(\mathbf{Q}\delta\mathbf{U}\,\mathbf{Q}^T)\mathbf{Q}\right)\\ = \operatorname{tr}(\mathbf{B}_j\,\overset{\circ}{\delta}\mathbf{V}) = \mathbf{B}_j \cdot \overset{\circ}{\delta}\mathbf{V},\end{aligned} \tag{3.21}$$

where $\mathbf{Q}\,\mathbf{A}_j\mathbf{Q}^T = \mathbf{B}_j$. Hence

$$\mathbf{Q}\left[(\mathbf{A}_i \otimes \mathbf{A}_j) : \delta\mathbf{U}\right]\mathbf{Q}^T = (\mathbf{B}_i \otimes \mathbf{B}_j) : \overset{\circ}{\delta}\mathbf{V}. \tag{3.22}$$

For the 2nd component of Eq. (3.9) we have $\partial\mathbf{I}/\partial\mathbf{U} = \overset{4}{\mathbf{0}}$ and the respective term does not need to be considered. For the 4th and 6th component we have

$$\begin{aligned}\frac{\partial\mathbf{U}}{\partial\mathbf{U}} : \delta\mathbf{U} &= \frac{1}{2}(\overset{4}{\mathbf{I}}_a + \overset{4}{\mathbf{I}}_c) : \delta\mathbf{U} = \frac{1}{2}(\delta\mathbf{U} + \delta\mathbf{U}^T),\\ \frac{\partial\mathbf{U}^{-1}}{\partial\mathbf{U}} : \delta\mathbf{U} &= -\frac{1}{2}\mathbf{U}^{-1}(\delta\mathbf{U} + \delta\mathbf{U}^T)\mathbf{U}^{-1},\end{aligned} \tag{3.23}$$

where Eq. (3.13) was used to derive the second equation. Applying the rotation operations to both of these equations we obtain

$$\begin{aligned}\mathbf{Q}\left(\frac{\partial\mathbf{U}}{\partial\mathbf{U}} : \delta\mathbf{U}\right)\mathbf{Q}^T &= \frac{1}{2}\mathbf{Q}(\delta\mathbf{U} + \delta\mathbf{U}^T)\mathbf{Q}^T = \frac{1}{2}(\overset{\circ}{\delta}\mathbf{V} + \overset{\circ}{\delta}\mathbf{V}^T),\\ \mathbf{Q}\left(\frac{\partial\mathbf{U}^{-1}}{\partial\mathbf{U}} : \delta\mathbf{U}\right)\mathbf{Q}^T &= -\frac{1}{2}\mathbf{Q}\,\mathbf{U}^{-1}(\delta\mathbf{U} + \delta\mathbf{U}^T)\mathbf{U}^{-1}\mathbf{Q}^T\\ &= -\frac{1}{2}\mathbf{V}^{-1}(\overset{\circ}{\delta}\mathbf{V} + \overset{\circ}{\delta}\mathbf{V}^T)\mathbf{V}^{-1}.\end{aligned} \tag{3.24}$$

Note that as a result of the rotate-forward operation in the above formulas, $\mathbf{U}$ is replaced by $\mathbf{V}$, $\mathbf{U}^{-1}$ by $\mathbf{V}^{-1}$, and $\delta\mathbf{U}$ by $\overset{\circ}{\delta}\mathbf{V}$. Hence, we may introduce an elasticity tensor $\overset{4}{\mathbf{C}}{}^{*}$ relating $\overset{\circ}{\delta}\mathbf{T}^{*}$ with $\overset{\circ}{\delta}\mathbf{V}$,

$$\overset{\circ}{\delta}\mathbf{T}^{*} = \mathbf{Q}\left[\overset{4}{\mathbf{C}} : \delta\mathbf{U}\right]\mathbf{Q}^{T} \equiv \overset{4}{\mathbf{C}}{}^{*} : \overset{\circ}{\delta}\mathbf{V} \tag{3.25}$$

of the same structure as $\overset{4}{\mathbf{C}}$.

For an infinitesimal deformation, when $\mathbf{F} \approx \mathbf{I}$, we have

$$\mathbf{U} = \mathbf{V} = \mathbf{I}, \qquad \frac{\partial \mathbf{U}^{-1}}{\partial \mathbf{U}} = -\frac{1}{2}(\overset{4}{\mathbf{I}}_{a} + \overset{4}{\mathbf{I}}_{c}), \qquad \frac{\partial \mathbf{V}^{-1}}{\partial \mathbf{V}} = -\frac{1}{2}(\overset{4}{\mathbf{I}}_{a} + \overset{4}{\mathbf{I}}_{c}) \tag{3.26}$$

and therefore the linearized elasticity tensors $\overset{4}{\mathbf{C}}$ and $\overset{4}{\mathbf{C}}{}^{*}$ are identical.

4. Conclusion

We have shown that under the rotate-forward operation, the structure of a general hyper-elastic constitutive equation and the respective constitutive operator for the Biot stress is carried over to the respective relations for the rotated Biot stress, with $\mathbf{U}$ replaced by $\mathbf{V}$, and $\delta\mathbf{U}$ by $\overset{\circ}{\delta}\mathbf{V}$, where the corotational variation is of the Green–McInnis–Naghdi type.

References

1. R. DE BOER, *Vector- und Tensorrechnung für Ingenieure*, Springer-Verlag, 1982.
2. M.A. CRISFIELD, *A consistent co-rotational formulation for nonlinear, three-dimensional, beam-elements*, Comput. Methods Appl. Mech. Engng., **81**, 131–150, 1990.
3. J.K. DIENES, *On the analysis of rotation and stress rate in deforming bodies*, Acta Mech., **32**, 217–232, 1979.
4. G.C. JOHNSON and D.J. BAMMANN, *A discussion of stress rates in finite deformation problems*, Int. J. Solids Structures, **20**, 725–737, 1984.
5. R. OGDEN, *Nonlinear elastic deformations*, Ellis Horwood, Chichester, UK 1984.
6. C.C. RANKIN and F.A. BROGAN, *An element-independent corotational procedure for treatment of large rotations*, [in:] Collapse Analysis of Structures, L.H. SOBEL, K. THOMAS [Eds.], 85–100, ASME, New York 1984.
7. J. SIMO, *A finite strain beam formulation. The three-dimensional dynamic problem*, Comput. Methods Appl. Mech. Engng., **49**, 55–70, 1985.
8. J.C. SIMO, D.D. FOX and T.J.R. HUGHES, *Formulations of finite elasticity with independent rotations*, Comput. Methods Appl. Mech. Engng., **95**, 227–288, 1992.
9. J. SIMO and L. VU-QUOC, *A three-dimensional finite strain rod model. The three-dimensional dynamic problem*, Comput. Methods Appl. Mech. Engng., **58**, 79–116, 1986.

POLISH ACADEMY OF SCIENCES
INSTITUTE OF FUNDAMENTAL TECHNOLOGICAL RESEARCH

Received March 6, 1996.

Announcement and call for papers

3rd EUROMECH SOLID MECHANICS CONFERENCE

Royal Institute of Technology, Stockholm, Sweden

18 – 22 August 1997

The 3rd EUROMECH Solid Mechanics Conference, organised by the European Mechanics Society (EUROMECH), will be held August 18 – 22 1997 at the Royal Institute of Technology, Stockholm, Sweden. The conference will cover the entire field of theoretical and applied solid mechanics and the scientific program consists of one General Opening Lecture, eight invited Plenary Lectures and parallel Technical Sessions.

EUROMECH Solid Mechanics Conference Committee (ESMCC)

A. Del Grosso, Genua
J. Engelbrecht, Tallinn
N. Jones (Chairman), Liverpool
M. Kleiber, Warsaw
J. Lemaitre, Paris
B. Lundberg, Uppsala
G. Maier, Milan
K. Pfeiffer, Munich
W. Schiehlen, Stuttgart
F. Sidoroff, Lyon
L. Slepyan, Tel Aviv
B. Storåkers (Chairman, Local Organizing Committee), Stockholm

The General Opening Lecture will be given by

V. Tvergaard, Lyngby, Denmark

and the invited Plenary Lectures are

E. Benvenuto, Genova, Italy
R. Billardon, Cachan, France
N. A. Fleck, Cambridge, U.K.
B. Fredriksson, Linköping, Sweden
E. van der Giessen, Delft, The Netherlands
W. Gutkowski, Warsaw, Poland
H. M. Hiller, Duisburg, Germany
E. Kreuzer, Hamburg-Harburg, Germany

Authors are invited to submit papers for the technical sessions. A one page abstract of approximately 500 words is requested before January 15, 1997. Those who wish to receive further information about the conference are invited to contact

Dr. Per-Lennart Larsson
Secretary General, LOC
Dept. of Solid Mechanics
Royal Institute of Technology
S-100 44, Stockholm
SWEDEN

Tel: +46-8-790 7540
Fax: +46-8-411 24 18
e-mail: 3esmc@hallf.kth.se