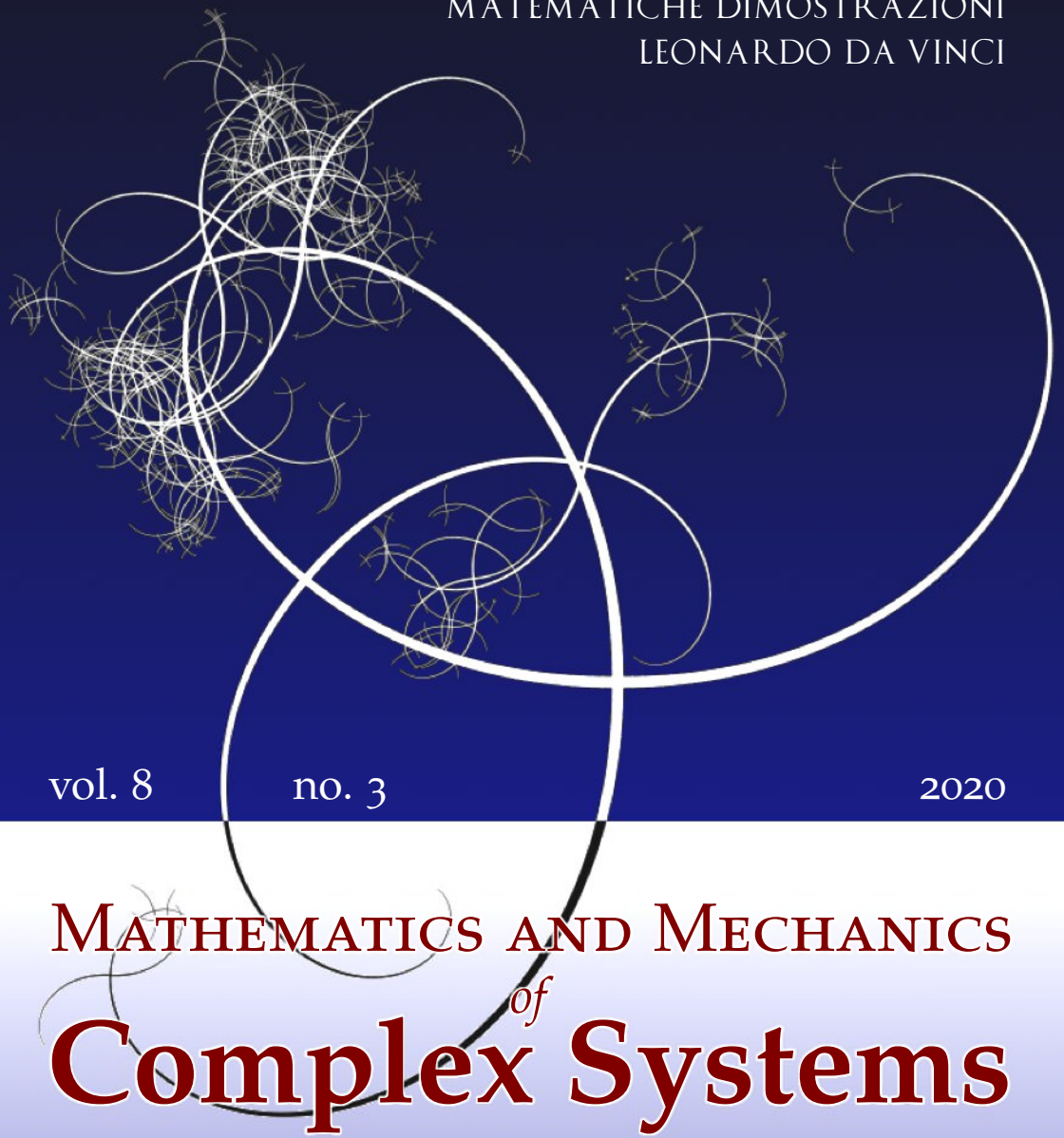
NISSUNA UMANA INVESTIGAZIONE SI PUO DIMANDARE
VERA SCIENZIA S'ESSA NON PASSA PER LE
MATEMATICHE DIMOSTRAZIONI
LEONARDO DA VINCI



vol. 8 no. 3 2020

Mathematics and Mechanics *of* Complex Systems

msp

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS

msp.org/memocs

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

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
Vol. 8, No. 3, 2020
https://doi.org/10.2140/memocs.2020.8.201

ON CLASSICAL CONTINUUM MECHANICS, TWO-SCALE CONTINUA, AND PLASTICITY

GIANPIETRO DEL PIERO

The paper starts with a careful analysis of the kinematics of two-scale continua. The subsequent developments are based on a single additional concept, the concept of energy. Two basic axioms are formulated in energetic terms, and the stress tensors, the constitutive equations, and all other elements required for the formulation of the initial/boundary-value problem are regarded as derived quantities. A comparison with the theory of gradient plasticity shows the innovative aspects of the proposed theory.

1. Introduction

This paper is a revisitation of some ideas developed by the present author in a series of papers [Del Piero 2009; 2014a; 2014b; 2017; 2018a; 2018b; 2019] concerning the principles of classical mechanics, their extension to generalized continua, and their application to plasticity. Over the years, many ideas have evolved and some became obsolete. Here I make an attempt to fix the present state of the art. I chose to restrict my exposition to two-scale continua because, being a first step toward a generalization of classical mechanics, such continua must be well understood before proceeding to further steps.

For the theoretical bases of classical mechanics and for the notation, I follow [Truesdell and Noll 1965; Noll 1973; Truesdell 1991; Gurtin 1981]. For generalized continua, I refer to the book [Gurtin et al. 2010]. I also frequently make reference to my previous paper [Del Piero 2019]. In particular, the following Section 2 on the two-scale geometric representation of deformable bodies is a summary of Chapters 2–4 of [Del Piero 2019], to which the interested reader is referred for further details.

Most treatises on mechanics, including those cited above, can be subdivided into three parts: kinematics, forces or interactions, and constitutive equations. In the passage from the first to the second part, the introduction of objects called *forces*

Communicated by Miroslav Šilhavý.
MSC2010: primary 74AXX, 74A05; secondary 74A60, 74C15, 74C20.
Keywords: foundations of mechanics, two-scale continua, reference frames, gradient plasticity, nonlocal flow rules.

involves an abrupt change of language and rigor. Since their systematic introduction by Newton, to many authors forces appeared as mysterious objects, and the belief in their existence was frequently considered as an act of faith.[1] Regarding "rigor", I reproduce three quotes taken from [Gurtin et al. 2010, §19]:

- "[forces are] obscure and metaphysical beings, capable of nothing but spreading darkness over a science clear by itself" (d'Alembert),
- "we speak of forces only to conceal our ignorance" (Maupertuis),
- "[forces are] an obscure metaphysical notion" (Carnot).

Even Mach's definition[2]

- "force is any circumstance of which the consequence is motion"

does not seem to be illuminating. The main argument in favor of the introduction of forces seems to be that they correctly describe the observed phenomena, which is more or less the same argument of the supporters of the Ptolemaic versus the Copernican cosmology.

For a long time I was puzzled by this difference in the presentations of kinematics and of forces/interactions. Slowly, I started to realize that the difference disappears if energy were taken as a primitive concept. When doing so, the only mathematical object to be added to those of kinematics is a functional involving the deformation measures, local or nonlocal in space, and possibly their past time histories. In this way, the mysterious objects become derived quantities of the energy.

There are historical reasons why this simple and rational approach was not adopted ab initio. Traditionally, energy was conceived as *reversible*, so that an energetic approach was considered appropriate only to *conservative* systems. For the study of irreversible processes, a thermodynamic approach was adopted instead. The energy was subdivided into a recoverable part, the *free energy*, and a part proportional to the *entropy* of the body. In this way the thermal variables, which by their own nature are extraneous to continuum mechanics, became an essential part of the picture. This was mainly due to the ignorance of the concept of *dissipation potential* elaborated by the French school,[3] which renders possible a purely mechanical energetic approach to dissipative systems.[4]

Section 3 is devoted to the *external energy*, that is, to the energy supplied to, or extracted from, the body by the exterior. In Section 3.1 the time derivative of

[1] For a history of the concept of force, see [Jammer 1999]. For earlier criticisms see [Mach 1883].

[2] See [Mach 1883, p. 83].

[3] See [Moreau 1970; 1974; Halphen and Son 1975; Germain et al. 1983].

[4] While the introduction of thermal variables reduces mechanics to a special case of thermodynamics, the approach based on dissipation potentials may lead to the opposite conclusion. For example, in a recently published theory of rigid heat conductors [Del Piero 2020], I consider temperature as a state variable in a purely mechanical context.

the external energy, the *external power*, is assumed to be the sum of elementary powers, consisting of products of partial derivatives of the energy density, the *external actions*, classically subdivided into "distance" and "contact" actions, by the corresponding deformation rates. Section 3.2 contains Noll's deduction of Euler's balance equations of linear and angular momentum from the indifference of the external power to the representation of the body in the Euclidean space [Noll 1963]. In Section 3.3, the introduction of the *cut principle* of Euler and Cauchy leads to the definition of the stress tensor, along the line traced by Cauchy and completed by Noll. Though these are well known subjects, an innovative aspect is that, as already pointed out in [Del Piero 2019], the power associated with the stress is no longer viewed as an internal power, but as an alternative expression of the external power of the contact actions.[5]

Section 4 deals with two postulates, the *conservation principle* and the *dissipation principle*, proposed in [Del Piero 2019] as basic axioms of mechanics in place of Euler's balance laws.[6] The conservation principle states that the external energy supplied to a body is entirely transformed into *internal energy*, and the dissipation principle says that a part of this energy is *dissipated*, that is, cannot be transformed back to external energy. In Section 4.1 the internal energy is assumed to be a function of two tensorial measures of *macroscopic* and *microscopic* deformation. From the theory of *structured deformations* come the results that the internal energy is the sum of two parts, one due to the microscopic deformation and one to the deformation generated by microdefects of various physical nature, collectively labeled as *disarrangements*, and that the latter is measured by the *difference between macroscopic and microscopic deformation*.[7] In Section 4.2 the dissipation principle is used to subdivide the internal energy into two parts, reversible and dissipative. Both are supposed to have a volume density, which is a differentiable function for the reversible part and a dissipation potential for the dissipative part.

The conservation principle imposes the equality between internal and external power. Replacing the deformation rates by *virtual velocities*, the *equation of virtual power* is obtained in Section 4.3. This equation is of fundamental importance in mechanics, and was proposed as a basic postulate in place of Euler's balance laws.[8] In Section 4.4 this equation is subjected to the *Coleman–Noll procedure*. Exploiting the arbitrariness of the virtual velocities, this procedure provides field equations and boundary conditions involving the external actions and the partial

[5] See the comment at the end of Section 3.3 below.

[6] Euler's laws lost the status of postulates after their deduction from the indifference of the external power.

[7] References and more information on structured deformations are given in the same Section 4.1.

[8] See, e.g., [Germain 1973a; 1973b]. In the present context this equation, being a consequence of the conservation principle, cannot be a postulate.

derivatives of the internal energy densities. Finally, in Section 4.5 two types of dissipation potential, given by functions positively homogeneous of order one and of order two, are introduced. They correspond to the dissipative behavior observed in plastic and in viscous materials, respectively.

In Section 5, the field equations and boundary conditions obtained from the Coleman–Noll procedure are used to formulate the initial/boundary-value problem for two-scale continua. This problem takes different forms, depending on the form assumed for the dissipation potential. In the nondissipative case, the problem reduces to a family of independent *equilibrium problems*, one for each instant of time. In Section 5.1, the well-posedness of such problems is investigated for a simple one-dimensional model.

Section 5.2 deals with dissipation potentials of the viscous type. Such potentials are *differentiable* and *rate-dependent*, because the corresponding internal actions involve the time derivatives of the deformation. The initial/boundary-value problem reduces to an *incremental problem*, that is, to a differential problem in which at every instant the current deformation is known from the solution of the same problem at the preceding instants, and the unknown is the deformation rate. An elementary special case of this problem is the Kelvin–Voigt problem of linear viscoelasticity.

Being represented by homogeneous functions of order one, the dissipation potentials of the plastic type are *rate-independent* and *nondifferentiable* at the origin. For such potentials, as shown in Section 5.3, the formulation of the initial/boundary-value problem involves a nonsmooth *flow rule* and an *activation condition* for the plastic strain rate, which in plasticity is the requirement that the stress tensor be located on the boundary of the yield surface.

The final Section 5.4 is devoted to a comparison to Gurtin and Anand's theory of gradient plasticity.[9] The comparison reveals both analogies and basic differences. A major difference is that, while Gurtin and Anand's approach involves additional stress measures,[10] here the only stress measure is the Piola stress of classical continuum mechanics. The role of the additional stress measures is played by *internal actions* which, as said before, are partial derivatives of the internal energy densities which, in turn, are functions of the deformation. This avoids the vicious circle of first inventing new stress measures, and then inventing *constitutive equations* to relate them to the deformation.

[9]See [Gurtin and Anand 2005]. In earlier papers I frequently made comparisons to this theory, which, as far as I know, is the most advanced formulation of gradient plasticity. The theory of gradient plasticity was formulated by Aifantis [1984], and developed by Fleck and Hutchinson [1993] and others [Gudmundson 2004; Fleck and Willis 2009a; 2009b].

[10]The *plastic stress* and the *hyperstress*. Their necessity being considered as self-evident, they are introduced without any motivation or explanation.

Classically, the constitutive equations convey the experimental information on the stress-strain response of specific materials. Here, this information is used to determine the shape of the internal energy. Only a single constitutive equation, relating the Piola stress to a specific combination of partial derivatives of the internal energy density, is required. Once the shape of the internal energy has been fixed, no further constitutive assumption is necessary.

2. Two-scale representations of a physical body

The *physical bodies* are objects located in the *physical space*.[11] In classical mechanics the physical space is represented as a three-dimensional Euclidean point space $\mathscr{E}$, a physical body is represented as a region of $\mathscr{E}$ with the properties of a mathematical continuum, and its elementary constituents, the *material points*, are represented as points of that region. The representation preserves the distances between material points measured by an operator called the *observer*. A *multi-scale* representation involves measurements made at scales of different orders of magnitude, macroscopic, mesoscopic, microscopic, submicroscopic, etc. Here we consider only two scales, *macroscopic* and *microscopic*.[12]

2.1. ***Placements and deformations.*** A second operator, the *placer*, "places" on $\mathscr{E}$ the material points, preserving the distances measured by the observer.[13] Since the observer cannot measure an infinite number of distances, the placer can place only a finite set of material points. An image of the body as a continuum is then constructed by "filling" the interspaces between the placed points.[14] A *macroscopic placement* of the body in $\mathscr{E}$ involves the following sequence of operations.[15]

(1) The observer measures the mutual distances in a selected set Ξ^N made of N material points.

(2) The placer places Ξ^N into a set $\mathfrak{X}^N$ of N points of $\mathscr{E}$, preserving the distances and the orientation.[16]

[11] Physical body and physical space are considered here as primitive, undefined concepts.

[12] In this case it is usual to speak of *two-scale continua*. This is done with some abuse of terminology, since being *two-scale* is a property of the representation of the body on $\mathscr{E}$ and not of the body itself, which by the way could also not be a continuum.

[13] An exact correspondence between material points and points of $\mathscr{E}$ is possible only if the physical space is metrizable with a Euclidean metric. See [Del Piero 2019, §2].

[14] This can be done both if the physical body is itself a continuum, and if it is made of a huge but finite number of material points, like in bodies with an atomic structure. Therefore, the "filling" operation does not imply any assumption about the structure of the physical body.

[15] All operations are discussed in detail in the paper [Del Piero 2019]. The description given here aims to render the present paper as self-contained as possible.

[16] The distance in $\mathfrak{X}^N$ is the Euclidean distance of $\mathscr{E}$.

With the first operation, the observer induces on Ξ^N a *macroscopic distance function* d^N. For the given Ξ^N, there is an infinity of sets $\mathfrak{X}^N$ in $\mathscr{E}$ which preserve the distances and the orientation. Each of them is a *macroscopic placement* of the metric space (Ξ^N, d^N) on $\mathscr{E}$.

The distances d^N may differ from one measurement to another. In this case, we say that the body is *deformable*. To represent a deformable body, it is convenient to fix once and for all a *macroscopic reference placement* $\mathfrak{X}_R^N$ of Ξ^N,[17] and to consider every other placement $\mathfrak{X}^N$ as the composition of the *macroscopic reference map* $\Xi^N \mapsto \mathfrak{X}_R^N$ with a discrete map $f_R^N : \mathfrak{X}_R^N \to \mathfrak{X}^N$. Then the passage to the continuum is made as follows.

(3) A continuous region Ω_R of $\mathscr{E}$ is constructed by "filling" the interspaces between the points of $\mathfrak{X}_R^N$.

(4) The discrete map f_R^N from $\mathfrak{X}_R^N$ is extended to a map f_R from Ω_R.

The region Ω_R is not an exact representative of the body, because only the points of $\mathfrak{X}_R^N$ are images of material points. Nevertheless, Ω_R is taken as the *macroscopic reference placement*, that is, as the *macroscopic representative*, of the body in $\mathscr{E}$. The extended map f_R is a *macroscopic deformation* from Ω_R. This map is largely undetermined, the only condition dictated by physics being that f_R be injective and orientation-preserving.[18] For every f_R, the region $f_R(\Omega_R)$ is a *macroscopic placement* of the body.[19]

For the placement of a body at the microscopic scale, some additional operations are required.

(5) For each point X^α of Ξ^N, the observer selects a set Ξ^{N_α} of N_α material points close to X^α, and measures their mutual distances.[20]

(6) The placer places each Ξ^{N_α} into a set $\mathfrak{X}^{N_\alpha}$ of N_α points of $\mathscr{E}$, preserving the distances and the orientation.

For each Ξ^{N_α}, the distance measurements provide a microscopic distance function d^{N_α}. Then $\mathfrak{X}^{N_\alpha}$ is a placement of the metric space $(\Xi^{N_\alpha}, d^{N\alpha})$ on $\mathscr{E}$, and the set $\{\mathfrak{X}^N, \{\mathfrak{X}^{N_\alpha}\}\}$ made of a placement of Ξ^N and of the N placements of the Ξ^{N_α} is a *two-scale placement* of $\{\Xi^N, \{\Xi^{N_\alpha}\}\}$. Every $\mathfrak{X}^{N_\alpha}$ can be viewed as the composition of a *microscopic reference map* from Ξ^{N_α} to a microscopic reference placement $\mathfrak{X}_R^{N_\alpha}$, followed by a map $f_R^{N_\alpha}$ from $\mathfrak{X}_R^{N_\alpha}$ to $\mathfrak{X}^{N_\alpha}$.

At the microscopic scale, the passage to the continuum is made as follows.

[17] The distances in $\mathfrak{X}_R^N$ may, or may not, correspond to measured distances in Ξ^N.

[18] The first condition excludes interpenetration of matter, and the second excludes placements which are the mirror images of each other.

[19] Different extensions f_R of the same map f_R^N are considered as different macroscopic deformations.

[20] It is assumed that for each α the set Ξ^{N_α} includes the point X^α.

(7) For each X^α, the placer constructs a neighborhood $\mathcal{N}_R^\alpha$ of Ω_R by "filling" the interspaces between the points of the corresponding set $\mathfrak{X}_R^{N_\alpha}$.

(8) They then extend each discrete map $f_R^{N_\alpha}$ from $\mathfrak{X}_R^{N_\alpha}$ to a map f_R^α from $\mathcal{N}_R^\alpha$.[21]

Taking the regions $\mathcal{N}_R^\alpha$ to be sufficiently small, each f_R^α is conveniently described by its gradient ∇f_R^α at x_R^α. Since all f_R^α are injective and orientation-preserving, the tensors

$$F_R^\alpha = \nabla f_R^\alpha(x_R^\alpha), \quad \alpha \in \{1, 2, \ldots, N\}, \tag{2-1}$$

belong to the space Lin^+ of all second-order tensors with positive determinant. The last step is the extension of the discrete set $\{F_R^\alpha\}$ to Ω_R.

(9) The N tensors F_R^α are extended to a continuous field F_R over Ω_R.

The map $F_R : \Omega_R \to \mathrm{Lin}^+$ is a *microscopic deformation* from Ω_R.[22] The pair $(f_R(\Omega_R), F_R(\Omega_R))$ in the product space $\mathscr{E} \times \mathrm{Lin}^+$ is a *two-scale placement* of the body, the pair (f_R, F_R) is a *two-scale deformation* from Ω_R, and the pair $(\nabla f_R(x_R), F_R(x_R))$ is the *local description* of (f_R, F_R) at x_R.

In a two-scale deformation from Ω_R, a point x_R of $\mathcal{N}_R^\alpha$ is mapped into the point $f_R(x_R)$ by the macroscopic deformation and into the point $f_R^\alpha(x_R)$ by the microscopic deformation. Then the vector $e_R = x_R - x_R^\alpha$ is mapped into $\nabla f_R(x_R^\alpha)[e_R]$ and into $F_R(x_R^\alpha)[e_R]$, respectively. The difference $(\nabla f_R(x_R^\alpha) - F_R(x_R^\alpha))$ is a measure of the mismatch between the macroscopic and microscopic images of $\mathcal{N}_R^\alpha$. It will be called the *distortion* of the two-scale deformation at x_R^α.

Two-scale deformations from any two-scale placement $(f_R(\Omega_R), F_R(\Omega_R))$ can be defined using the fact that, due to the injectivity of f_R and F_R, the inverse maps f_R^{-1} and F_R^{-1} exist. Then the inverse of (f_R, F_R) is defined as the map $(f_R, F_R)^{-1} = (f_R^{-1}, F_R^{-1})$ which maps the placement $(f_R(\Omega_R), F_R(\Omega_R))$ back to Ω_R. The composition

$$(f, F) = (f_{R\varnothing}, F_{R\varnothing}) \circ (f_R, F_R)^{-1} \tag{2-2}$$

is a deformation from $(f_R(\Omega_R), F_R(\Omega_R))$ to $(f_{R\varnothing}(\Omega_R), F_{R\varnothing}(\Omega_R))$. This deformation is injective, and its inverse is $(f_R, F_R) \circ (f_{R\varnothing}, F_{R\varnothing})^{-1}$. Moreover, (f, F) is orientation-preserving, because f_R, F_R, $f_{R\varnothing}$, and $F_{R\varnothing}$ are all orientation-preserving. Therefore, the set of all injective and orientation-preserving two-scale deformations (f, F) has the algebraic structure of a *group* with respect to the composition (2-2).

[21] Like in the macroscopic case, the only condition required for the extended maps f_R^α is to be injective and orientation-preserving.

[22] The tensor field F_R is not in general the gradient of a vector field.

Additional restrictions can be imposed on "admissible" deformations with "nice" regularity properties, such as continuity or differentiability.[23] In general, the restricted set of admissible deformations is required to preserve the group structure. For example, if the admissible deformations are continuous, their inverses must be continuous as well. That is, the admissible deformations must be *homeomorphisms*. If the admissible deformations are differentiable, the admissible deformations must be *diffeomorphisms*, that is, differentiable deformations with differentiable inverses.

2.2. ***Reference systems and reference frames.*** For each set of distance functions on $(\Xi^N, \{\Xi^{N_\alpha}\})$, there is an infinity of two-scale placements $(\mathfrak{X}^N, \mathfrak{X}^{N_\alpha})$ in $\mathscr{E}$ which preserve distances and orientation. They form an equivalence class of two-scale placements called a *two-scale configuration*. A systematic way for selecting a representative element inside each equivalence class is to choose a *reference system* in the physical space, and a *reference frame* in $\mathscr{E}$.

A *reference system* is a discrete set (X^0, X^i) of material points whose mutual distances are supposed to be the same in all distance measurements,[24] and a *reference frame* is a distance- and orientation-preserving image (x^0, x^i) of the reference system on $\mathscr{E}$.[25] If the physical space can be represented as a three-dimensional Euclidean space, the number of the X^i is three. Then in every distance measurement the position x^α in $\mathscr{E}$ of a material point X^α of Ξ^N is determined uniquely by the linear combination

$$x^\alpha - x^0 = \sum_{i=1}^{3} \xi^i e^i, \tag{2-3}$$

with coefficients ξ^i depending on the distances $|x^\alpha - x^i|$ and $|x^\alpha - x^0|$,[26] and with the unit vectors

$$e^i = \frac{x^i - x^0}{|x^i - x^0|} \tag{2-4}$$

representing the *directions* in $\mathscr{E}$ associated with the reference frame.[27]

[23] In fact, regularity is assumed more for computational convenience than for physical reasons. For example, it is physically meaningless to require continuity for the deformations of a body with a discrete atomic structure.

[24] "Like the walls of a laboratory, the fixed stars, or the wooden horses on a merry-go-round" [Truesdell and Noll 1965, §17].

[25] Less precise are the definitions given in the literature. For a comparison with the definitions of Truesdell and Noll [1965], see [Del Piero 2019, footnote 16].

[26] For their explicit form see [Del Piero 2019, §3.3]. An advantage of using a reference system is that, instead of measuring the mutual distances between all points X^α of Ξ^N, it is sufficient to measure the distances of each X^α from the points of the reference system.

[27] It may happen that, like in the case of the "fixed stars", the points X^i are inaccessible to the observer. In this case, the observer is supposed to be able to recognize when three material points

2.3. ***Two-scale evolutions.*** A *two-scale evolution* is a sequence of two-scale placements, ordered according to the flow of the *physical time*. In $\mathscr{E}$, a two-scale evolution is represented by a family $t \mapsto (f_t, F_t)$ of two-scale deformations from a fixed macroscopic reference placement Ω_R. The parameter t, called the "time", need not coincide with the physical time, but must preserve the ordering established by the physical time.

An evolution is constructed starting from a set (f_{t_k}, F_{t_k}) of two-scale deformations from the reference placement, relative to a finite number of instants t_k. With a "filling" procedure in time, this discrete set of deformations is extended to a finite time interval. The family $t \mapsto (f_t, F_t)$ is the representation on $\mathscr{E}$ of a two-scale evolution of the body.[28]

While the reference system is fixed once and for all, the reference frames may vary with t. Then let $t \mapsto (f_t, F_t)$ and $t \mapsto (f_t^*, F_t^*)$ be two representations of the same evolution relative to two families, $t \mapsto \{x_t^0, x_t^i\}$ and $t \mapsto \{x_t^{*0}, x_t^{*i}\}$, of reference frames. For the macroscopic deformations f_t, f_t^*, let

$$x_t^* = f_t^*(x_R), \qquad x_t = f_t(x_R) \tag{2-5}$$

be the trajectories of a point x_R of Ω_R in the two representations. The distances $|x_t^* - x_t^{*0}|$ and $|x_t - x_t^0|$ are equal, because both coincide with the distance of the material points X, X^0 measured at the time t. Then there is a family $t \mapsto Q_t$ of orthogonal tensors such that $(x_t^* - x_t^{*0}) = Q_t[x_t - x_t^0]$, that is,

$$f_t^*(x_R) = x_t^{*0} + Q_t[f_t(x_R) - x_t^0]. \tag{2-6}$$

This is the transformation rule for the macroscopic deformations under a change of reference frame. Moreover, differentiation with respect to x_R provides the transformation rule

$$\nabla f_t^*(x_R) = Q_t \nabla f_t(x_R) \tag{2-7}$$

for the deformation gradients, and differentiation with respect to t provides the transformation rule

$$\dot{f}_t^*(x_R) = \dot{x}_t^{*0} + Q_t[\dot{f}_t(x_R) - \dot{x}_t^0] + \dot{Q}_t[f_t(x_R) - x_t^0] \tag{2-8}$$

for the time derivatives.[29]

---

are *aligned*. Then for each X^i they select an accessible point X^{i*} aligned with X^0 and X^i and such that the distance between X^{i*} and X^0 is the same in all distance measurements, and take (X^0, X^{i*}) as the reference system.

[28] Just like the regularity in space, the assumed regularity in time may have no relation with physical reality.

[29] In the previous versions of the theory the time derivative was denoted by δ, and the superimposed dot was reserved for the derivative with respect to the physical time. Here, to simplify the notation, we use the superimposed dot in both cases.

For the microscopic deformations F_t, F_t^*, at every t the lengths of the vectors $F_t(x_R)[e_R]$ and $F_t^*(x_R)[e_R]$ are the same for all x_R and for all unit vectors e_R. Then at each $x_R \in \Omega_R$ there is an orthogonal tensor $Q_{x_R t}$ such that $F_t^*(x_R) = Q_{x_R t} F_t(x_R)$. By the independence of the distortions on the reference frame, all $Q_{x_R t}$ must be equal to the tensor Q_t in (2-7).[30] Then we have

$$F_t^*(x_R) = Q_t F_t(x_R) \tag{2-9}$$

and, by time differentiation,

$$\dot{F}_t^*(x_R) = Q_t \dot{F}_t(x_R) + \dot{Q}_t F_t(x_R). \tag{2-10}$$

We remark that $\dot{f}_t$ and $\dot{F}_t$ are not physical velocities, because the presence of Q_t, x_t^0, and x_t^{*0} and of their time derivatives makes them dependent on the variable reference frame.

3. The external energy

Classically, the deformations of a body are attributed to the action of physical agents, like contact with other bodies, gravitation, heat, and electric and magnetic fields. These agents are supposed to act through vectorial quantities called *forces* or *actions*, considered as physical entities. Today, a prevailing view is that a body deforms because it receives some energy from, or delivers some energy to, the exterior. In this view, energy is the primitive concept, and forces are derived quantities.

3.1. ***External energy and external power.*** In continuum mechanics it is generally assumed that there are two modes, *at distance* and *by contact*, for interchanging energy between a body and the exterior. In the present theory of two-scale continua, we assume that the external energy depends only on the current value (f, F) of the two-scale deformation

$$\mathscr{E}_{\text{ext}}(\Omega_R, f, F) = \int_{\Omega_R} \beta(f(x_R), F(x_R))\, \mathsf{dv} + \int_{\partial\Omega_R} \sigma(f(x_R), F(x_R))\, \mathsf{da}, \tag{3-1}$$

thereby excluding the dependence on deformation gradients and on the past deformation history. We also assume that the *volume density* β and the *surface density* σ are differentiable functions.[31] In an evolution $t \mapsto (f_t, F_t)$, the *external power*

$$\mathscr{P}_{\text{ext}}(\Omega_R, \dot{f}_t, \dot{F}_t) = \int_{\Omega_R} (b_{Rt} \cdot \dot{f}_t + B_{Rt} \cdot \dot{F}_t)\, \mathsf{dv} + \int_{\partial\Omega_R} (s_{Rt} \cdot \dot{f}_t + S_{Rt} \cdot \dot{F}_t)\, \mathsf{da} \tag{3-2}$$

[30] See [Del Piero 2019, §4.4].

[31] In this way we exclude, for example, the energy due to frictional contact.

is obtained by time differentiation. The *external actions*, that is, the *distance actions* b_{Rt}, B_{Rt} and the *contact actions* s_{Rt}, S_{Rt}, are the partial derivatives of β and σ at (f_t, F_t).[32]

3.2. ***"Indifference principle" and balance laws.*** As already pointed out, $\dot f_t$ and $\dot F_t$ are not physical velocities, because they depend on the representation of the evolution on $\mathscr{E}$. It is just common sense to require that the power be independent of this representation.[33] Therefore, if $t \mapsto (f_t, F_t)$ and $t \mapsto (f_t^*, F_t^*)$ are representations of the same evolution with respect to two different families of reference frames, not only are the derivatives $\dot f_t$, $\dot F_t$ and $\dot f_t^*$, $\dot F_t^*$ different, but also the external actions must transform in such a way that the resulting powers be the same. In particular, for the macroscopic power we have the indifference condition

$$\int_{\Omega_R} b_{Rt}^* \cdot \dot f_t^* \,\mathsf{dv} + \int_{\partial\Omega_R} s_{Rt}^* \cdot \dot f_t^* \,\mathsf{da} = \int_{\Omega_R} b_{Rt} \cdot \dot f_t \,\mathsf{dv} + \int_{\partial\Omega_R} s_{Rt} \cdot \dot f_t \,\mathsf{da}, \tag{3-3}$$

and from the transformation law (2-8) for $\dot f_t$, we get

$$\begin{aligned}\int_{\Omega_R} (b_{Rt}^* \cdot (\dot x_t^{*0} + Q_t[\dot f_t - \dot x_t^0] + \dot Q_t[f_t - x_t^0]) - b_{Rt} \cdot \dot f_t) \,\mathsf{dv} \\ + \int_{\partial\Omega_R} (s_{Rt}^* \cdot (\dot x_t^{*0} + Q_t[\dot f_t - \dot x_t^0] + \dot Q_t[f_t - x_t^0]) - s_{Rt} \cdot \dot f_t) \,\mathsf{da} = 0,\end{aligned} \tag{3-4}$$

for all x_t^{*0}, $\dot x_t^{*0}$, Q_t, $\dot Q_t$ and for all $\dot f_t$.[34] In particular, for $\dot x_t^{*0} = \dot Q_t = 0$, from the arbitrariness of $\dot f_t$ we deduce the transformation laws of the external actions

$$b_{Rt}^* = Q_t b_{Rt}, \qquad s_{Rt}^* = Q_t s_{Rt}. \tag{3-5}$$

Substituting into (3-4), all terms involving $\dot f_t$ cancel. Then after setting $\dot Q_t = Q_t Y_t$ and $a_t = (Q_t^T \dot x_t^{*0} - \dot x_t^0)$, it remains that

$$\int_{\Omega_R} b_{Rt} \cdot (a_t + Y_t[f_t - x_t^0]) \,\mathsf{dv} + \int_{\partial\Omega_R} s_{Rt} \cdot (a_t + Y_t[f_t - x_t^0]) \,\mathsf{da} = 0. \tag{3-6}$$

Moreover, by differentiation of $Q_t^T Q_t = I$ we have the relation

$$\dot Q_t^T Q_t + Q_t^T \dot Q_t = 0, \tag{3-7}$$

[32] Here $b_{Rt} \cdot \dot f_t$ is a simplified notation for $(\beta_f(f_t(x_R), F_t(x_R)) \cdot \dot f_t(x_R))$, and so on.

[33] In the past, this requirement was given the status of a basic principle. For a history of this "principle of material frame indifference" see [Truesdell and Noll 1965, §19A]. See also the remarks in [Murdoch 2012, §12.4]. Only later did Noll recognize the more modest nature of this requirement: "the principle of material frame-indifference is not a law of physics... it is merely a prescription for avoiding nonsense" [Noll 2006].

[34] Indeed, (3-4) holds for all evolutions from the same deformation f_t.

which shows that $Y_t = Q_t^T \dot{Q}_t$ is a skew-symmetric tensor. Since a_t and Y_t are mutually independent, (3-6) splits into the separate conditions[35]

$$\begin{aligned} a \cdot \left(\int_{\Omega_R} b_R \, \mathrm{dv} + \int_{\partial\Omega_R} s_R \, \mathrm{da} \right) = 0, \\ Y \cdot \left(\int_{\Omega_R} b_R \otimes (f - x^0) \, \mathrm{dv} + \int_{\partial\Omega_R} s_R \otimes (f - x^0) \, \mathrm{da} \right) = 0. \end{aligned} \tag{3-8}$$

Denoting by y the vector associated with Y by the identity

$$Y \cdot (a \otimes b) = y \cdot b \times a \quad \text{for all } a, b \in \mathcal{V}, \tag{3-9}$$

and using the arbitrariness of a and y, (3-8) reduces to

$$\begin{aligned} \int_{\Omega_R} b_R \, \mathrm{dv} + \int_{\partial\Omega_R} s_R \, \mathrm{da} = 0, \\ \int_{\Omega_R} (f - x^0) \times b_R \, \mathrm{dv} + \int_{\partial\Omega_R} (f - x^0) \times s_R \, \mathrm{da} = 0. \end{aligned} \tag{3-10}$$

These are the *balance laws of linear and angular momentum*. For a long time, they have been considered the basic axioms of mechanics. In reality, as shown above, they are consequences of the indifference condition (3-3).[36]

For the microscopic external power the indifference condition is

$$\int_{\Omega_R} B^*_{Rt} \cdot \dot{F}^*_t \, \mathrm{dv} + \int_{\partial\Omega_R} S^*_{Rt} \cdot \dot{F}^*_t \, \mathrm{da} = \int_{\Omega_R} B_{Rt} \cdot \dot{F}_t \, \mathrm{dv} + \int_{\partial\Omega_R} S_{Rt} \cdot \dot{F}_t \, \mathrm{da}, \tag{3-11}$$

where, by (2-10), $\dot{F}^*_t = Q_t(\dot{F}_t + Y_t F_t)$, with Q_t and Y_t the same orthogonal and skew-symmetric tensors as above by the independence of the distortions on the reference frame. In particular, for $Y = 0$ we have

$$\int_{\Omega_R} (Q_t^T B^*_{Rt} - B_{Rt}) \cdot \dot{F}_t \, \mathrm{dv} + \int_{\partial\Omega_R} (Q_t^T S^*_{Rt} - S_{Rt}) \cdot \dot{F}_t \, \mathrm{da} = 0, \tag{3-12}$$

and since this must hold for all $\dot{F}_t$, the transformation laws

$$B^*_{Rt} = Q_t B_{Rt}, \qquad S^*_{Rt} = Q_t S_{Rt} \tag{3-13}$$

[35] The powers on the left-hand side are those of the macroscopic rigid translations and of the macroscopic rigid rotations, respectively. For this reason, the indifference principle was also called the *principle of independence of power on superimposed rigid motions* [Green and Rivlin 1957]. Since both equations hold for every evolution and for all times, the subscript t has been omitted.

[36] The balance laws are due to Euler, and their deduction from the indifference of the external power is due to Noll [1963]. It is surprising that such physically relevant conditions come from a requirement on the representation of the evolutions on $\mathcal{E}$.

follow. Substituting into (3-12) we get the condition

$$Y_t \cdot \left(\int_{\Omega_R} B_{Rt} F_t^T \, \mathrm{dv} + \int_{\partial\Omega_R} S_{Rt} F_t^T \, \mathrm{da} \right) = 0, \tag{3-14}$$

which is the microscopic counterpart of the balance law $(3\text{-}10)_2$ of macroscopic angular momentum. Since there is no counterpart to $(3\text{-}10)_1$, this condition does not admit a local form.

3.3. ***The cut principle and the stress tensor.*** It is currently assumed that contact actions $s_{\partial\Pi_R}$ of the same nature as the contact actions s_R act at every internal surface of Ω_R, and that an external power of the form (3-2) holds for every part Π_R of Ω_R. This is the *cut principle of Euler and Cauchy*.[37]

For a fixed t, consider a family $\varepsilon \mapsto \Pi_\varepsilon = \varepsilon\Pi_R$ of homothetic regions, and for each Π_ε write the balance law $(3\text{-}10)_1$. Assuming the uniform boundedness of the distance action b_R, when $\varepsilon \to 0$ the volume integral tends to zero faster than the area of $\partial\Pi_\varepsilon$. Then the area integral also tends to zero faster than $\mathsf{a}(\partial\Pi_\varepsilon)$:

$$\lim_{\varepsilon\to 0^+} \frac{1}{\mathsf{a}(\partial\Pi_\varepsilon)} \int_{\partial\Pi_\varepsilon} s_{\partial\Pi_\varepsilon} \, \mathrm{da} = 0. \tag{3-15}$$

Taking regions Π_R of particular shapes, it has been proved that the contact action $s_{\partial\Pi_R}$ at a boundary point x_R of Π_R depends only on the exterior unit normal n_R to $\partial\Pi_R$ at x_R, and that this dependence is linear. That is, there are a vector-valued map $\tilde{s}$ over $\Omega_R \times \mathscr{V}$ and a tensor field T_R over Ω_R such that

$$s_{\partial\Pi_R}(x_R) = \tilde{s}(x_R, n_R) = T_R(x_R) n_R \tag{3-16}$$

for all regions Π_R for which x_R is a boundary point. The tensor $T_R(x_R)$ is the *Piola stress* at x_R.[38]

From $(3\text{-}10)_1$ and from the divergence theorem we get

$$\int_{\Omega_R} (\operatorname{div} T_R + b_R) \, \mathrm{dv} = 0. \tag{3-17}$$

By the cut principle, this equation holds for every part Π_R of Ω_R. Then due to the arbitrariness of Π_R the punctual equation

$$\operatorname{div} T_R + b_R = 0 \tag{3-18}$$

[37] See, e.g., [Truesdell 1991, p. 154]. This "principle" excludes the possibility that the contact actions at the interior surfaces of the body and those at the physical boundary be of a different nature. In particular, it excludes the presence of *material surfaces*, in the sense of Gurtin and Murdoch [1975], at the body's interior. This implies, for example, the absence of surface tension at the interior surfaces.

[38] The existence of the map $\hat{s}$, formerly a conjecture of Cauchy, was proved later by Noll [1959], and the existence of T_R is proved in Cauchy's *tetrahedron theorem*.

holds at almost every point of Ω_R. This is the *local form* of the balance law of linear momentum. In a similar way, from the divergence theorem applied to the indifference condition (3-8)$_2$ we get the punctual condition

$$Y \cdot T_R \nabla f^T = 0, \tag{3-19}$$

which shows that the local form of the balance law of linear momentum is the requirement that the Cauchy stress tensor $(\det \nabla f^{-1}) T_R \nabla f^T$ be symmetric.

From (3-16) and (3-18) and from the divergence theorem we have

$$\int_{\Pi_R} T_R \cdot \nabla \dot{f} \, \mathrm{dv} = \int_{\Pi_R} b_R \cdot \dot{f} \, \mathrm{dv} + \int_{\partial \Pi_R} s_{\partial \Pi_R} \cdot \dot{f} \, \mathrm{da}, \tag{3-20}$$

and substitution into (3-2) provides the *reduced form* of the external power

$$\mathcal{P}_{\mathrm{ext}}(\Omega_R, \dot{f}, \dot{F}) = \int_{\Omega_R} (T_R \cdot \nabla \dot{f} + B_R \cdot \dot{F}) \, \mathrm{dv} + \int_{\partial \Omega_R} S_{\partial \Pi_R} \cdot \dot{F} \, \mathrm{da}. \tag{3-21}$$

The traditional interpretation of (3-20) is that T_R is an *internal action*, $T_R \cdot \nabla \dot{f}$ is an *internal power*, and the equation is an equality between internal and external power.[39] In the present context this interpretation is incorrect because (3-20), being a consequence of (3-16) and (3-18), *is an identity between two alternative forms of the external power*. Therefore, the stress T_R is not an internal action, but *a tensorial representative of the external contact action* acting on $\partial \Pi_R$.

3.4. ***Inertia.*** When taking energy as a primitive concept, a natural starting point for describing the phenomenon of inertia is to define the kinetic energy. Let ρ be the *mass density* of the body in the reference placement Ω_R.[40] Suppose that the body interacts with a second body, and consider an evolution $t \mapsto f_t$ of the two bodies. The *kinetic energy* at a point x_R of the first body due to the interaction with the second body is an external energy with volume density

$$\beta^{\mathrm{kin}}(f_t(x_R)) = -\tfrac{1}{2} \rho(x_R) |\dot{f}_t(x_R) - \dot{f}_t(x_{\varnothing R})|^2, \tag{3-22}$$

where t is the physical time, $x_{\varnothing R}$ is the position of the center of mass of the second body in the reference placement, and $(\dot{f}_t(x_R) - \dot{f}_t(x_{\varnothing R}))$ is the velocity of the point

[39] See [Germain 1973a; 1973b; Truesdell 1991, §1.15; Gurtin et al. 2010, §19.7.2].

[40] The concept of *mass*, generally considered as a fundamental ingredient for mechanics, is in fact relevant only when inertia is taken into account; see the comment of Noll [2004]. The same author wrote that "the basic concepts of mechanics in general should *not* include items such as momentum, kinetic energy, and angular momentum, because they are relevant only when inertia is important" [Noll 1995].

$x_t = f_t(x_R)$ relative to the point $x_{\varnothing t} = f_t(x_{\varnothing_R})$.[41] The external power

$$\dot{\beta}^{\text{kin}}(f_t(x_R)) = -\rho(x_R)(\ddot{x}_t - \ddot{x}_{\varnothing t}) \cdot (\dot{x}_t - \dot{x}_{\varnothing t}) \tag{3-23}$$

is the inner product of the *inertial action*

$$b^{\text{kin}}_{Rt}(x_R) = -\rho(x_R)(\ddot{x}_t - \ddot{x}_{\varnothing t}) \tag{3-24}$$

with the relative velocity $(\dot{x}_t - \dot{x}_{\varnothing t})$.[42]

Traditionally, the kinetic energy is written with $\dot{x}_t$ in place of $(\dot{x}_t - \dot{x}_{\varnothing t})$, with the specification that *the inertia law* $b^{\text{kin}}_{Rt} = -\rho\ddot{x}_t$ *holds only for a privileged class of observers, those moving with a uniform rectilinear motion with respect to the fixed stars*.[43] As pointed out in [Truesdell and Noll 1965, §18], this view is based on the Newtonian idea that the physical space is an *absolute space*, in which each body particle occupies a precise *position*, and that the "fixed stars" occupy a fixed position in that space. In particular, the point $x_{\varnothing R}$, which is identified with the center of mass of the universe, is assumed to have a fixed position in the absolute space, so that $\dot{x}_{\varnothing t} = \ddot{x}_{\varnothing t} = 0$.

While the Euclidean point space is indeed an absolute space, in the physical space there are no absolute positions, but only *positions relative to a given reference system*.[44] And since there are no absolute positions, whether or not an observer is "moving", whatever this means, is a false question.[45]

4. The internal energy

4.1. ***The two basic axioms.*** A physical body is supposed to have an *internal energy*. In a two-scale continuum, this energy depends on the local two-scale deformation $(\nabla f, F)$ and possibly on some additional *state variables*.[46] In the simple

[41] It is perfectly conceivable to consider β^{kin} as a function of the two-scale deformation (f_t, F_t), including a *microkinetic energy* depending on F_t, like in [Capriz 1989, §7].

[42] The idea that the "inertia forces" are particular external actions is due to d'Alembert. According to Noll [1963], "…l'on regarde les forces d'inertie comme des forces véritables qui sont les interactions entre les corps dans notre système solaire et la totalité des objets dans le reste de l'univers".

[43] On the contrary, in the kinematics developed in [Del Piero 2019] the only operation made by the observer is the measurement of distances. Then it makes no sense to consider more than one observer or, even worse, to assume there are "privileged" observers.

[44] The conception of the physical space as an absolute space has unpleasant consequences, like the belief in the existence of the "luminiferous aether" and of mysterious "apparent forces". This conception was repeatedly criticized over the centuries; see, e.g., [Mach 1883, p. 543]. For further references see [Del Piero 2019, §8.5]. On the contrary, if the kinetic energy is defined as in (3-22) no *apparent forces* are required because relative instead of absolute velocities are considered.

[45] What are "moving" with respect to each other are the reference frames, and "apparent forces" appear if their relative motions, which are arbitrarily chosen by the placer, are neglected.

[46] For example, thermal, electromagnetic, or chemical variables.

model developed here there are no state variables, and the energy is supposed to be the sum of a *recoverable* and of a *dissipative* part

$$\mathscr{E}_{\text{int}}(\Omega_R, f, F) = \mathscr{E}^{\text{rec}}_{\text{int}}(\Omega_R, f, F) + \mathscr{E}^{\text{dis}}_{\text{int}}(\Omega_R, f, F), \tag{4-1}$$

where the recoverable part is a volume integral with volume density depending on the current values of the local two-scale deformation and of its spatial gradients

$$\mathscr{E}^{\text{rec}}_{\text{int}}(\Omega_R, f, F) = \int_{\Omega_R} \Phi(\nabla f, F, \nabla^2 f, \nabla F)\, \mathsf{dv}, \tag{4-2}$$

and in every deformation process $\tau \mapsto (f_\tau, F_\tau)$ from $\tau = 0$ to $\tau = t$ the dissipative part has the form

$$\mathscr{E}^{\text{dis}}_{\text{int}}(\Omega_R, f_t, F_t) = \mathscr{E}^{\text{dis}}_{\text{int}}(\Omega_R, f_0, F_0) + \int_{\Omega_R} \int_{t_0}^{t} \chi(\nabla \dot f_\tau, \dot F_\tau, \nabla^2 \dot f_\tau, \nabla \dot F_\tau)\, d\tau\, \mathsf{dv}. \tag{4-3}$$

The relation between external and internal energy is subjected to two basic postulates. In classical mechanics, the role of postulates is played by the balance laws of Euler or, alternatively, by the equation of virtual power.[47] In the more general context assumed in [Del Piero 2019], the same role is attributed to two "principles" more in line with general physics, the conservation principle and the dissipation principle.[48]

The *conservation principle* says that to every amount of external energy supplied by, or extracted from, the body by the exterior, corresponds an equal increase, or decrease, of internal energy. This principle is expressed by the *power equation*

$$\mathscr{P}_{\text{ext}}(\Omega_R, \dot f, \dot F) = \mathscr{P}_{\text{int}}(\Omega_R, \dot F, \dot F^d). \tag{4-4}$$

The *dissipation principle* states that in every deformation process the dissipative part of the internal energy is nonnegative. This corresponds to assuming that the *dissipation potential* χ is nonnegative,

$$\chi(\nabla \dot f_\tau, \dot F_\tau, \nabla^2 \dot f_\tau, \nabla \dot F_\tau) \geq 0 \tag{4-5}$$

for whatever values of the variables $\nabla \dot f_\tau, \dot F_\tau, \nabla^2 \dot f_\tau, \nabla \dot F_\tau$.

4.2. ***Disarrangements.*** For the dependence of the energy on two-scale deformations, a particular form based on the additive decomposition of the macroscopic deformation gradient

$$\nabla f = F + F^d, \tag{4-6}$$

[47] For axiomatics founded on the equation of virtual power see [Germain 1973a; 1973b].

[48] The two principles can be regarded as mechanical versions of the two laws of thermodynamics. In fact, there are different opinions on whether the conservation principle was first formulated in mechanics and then extended to other branches of physics or vice versa; see the discussion in [Mach 1896, pp. 295–297].

was assumed in [Del Piero 2018c; 2019]. Though a multiplicative decomposition looks more appropriate to a context of large deformations, a strong motivation in favor of the additive decomposition is provided by the theory of *structured deformations*.[49] In [Del Piero and Owen 1993] it was proved that every structured deformation is the uniform limit of sequences $n \mapsto f_n$ of piecewise continuous one-scale deformations, with a number of discontinuities tending to infinity and amplitudes tending to zero as $n \to \infty$. Subsequently, it was proved in [Del Piero and Owen 1995] that the difference $\nabla f - F$ between the gradient of the limit f of the functions f_n and the limit F of the absolutely continuous parts of the gradients ∇f_n is the volume density of the total deformation due to the singular parts of the ∇f_n, that is, to the microscopic *disarrangements* associated with the limit deformation (f, F).[50]

In the energetics of structured deformations developed by Choksi and Fonseca [1997], the internal energy of a structured deformation was defined as the relaxed limit, that is, as the infimum of the energies of all approximating sequences. The main result of [Choksi and Fonseca 1997] was that the relaxed energy has a volume density Φ depending on the local deformation $(\nabla f, F)$. The paper ended with the conjecture that this energy should be the sum of two parts, the relaxed limits of the bulk and the interfacial part of the energies of the approximating sequences, and that the volume densities of the two parts should depend on F and F^d, respectively:[51]

$$\Phi(\nabla f, F) = \varphi(F) + \psi(F^d). \tag{4-7}$$

Later, this conjecture was proved to be true.[52]

Here, for the volume density Φ of the recoverable part (4-2) we assume the decomposition[53]

$$\Phi(\nabla f, F, \nabla^2 f, \nabla F) = \varphi(F, \nabla F) + \psi(F^d, \nabla F^d), \tag{4-8}$$

[49] For a comparison between additive and multiplicative decompositions in finite plasticity see the paper [Del Piero 2018b]. For the purposes of the present paper, structured deformations can be identified with two-scale deformations.

[50] In crystal plasticity, a phenomenological characterization of $(\nabla f - F)$ based on slip surfaces and dislocation structures has been given by Reina and Conti [2014].

[51] From the mechanical viewpoint this is very reasonable because, due to the indeterminacy inherent to the "filling" procedure, the macroscopic deformation ∇f seems unfit to characterize a strain energy. For the external energy, on the contrary, it seems more appropriate to keep the dependence on the pair $(\nabla f, F)$, whose values are deduced directly from distance measurements.

[52] See [Owen 2000] and the present author's [Del Piero 2001] for the one-dimensional case, and [Šilhavý 2017] for the full three-dimensional case.

[53] In [Barroso et al. 2017], the dependence of φ on the pair $(F, \nabla F)$ was proved for the energy of three-scale deformations, also called *second-order structured deformations*.

and for the dissipation potential (4-3) we assume the form

$$\chi(\nabla \dot{f}_t, \dot{F}_t, \nabla^2 \dot{f}_t, \nabla \dot{F}_t) = \chi(\dot{F}_t^d, \nabla \dot{F}_t^d), \tag{4-9}$$

that is, we assume that the dissipation is entirely due to the disarrangements. Then in every evolution $t \mapsto (f_t, F_t)$ the *internal power*

$$\mathscr{P}_{\text{int}}(\Omega_R, \dot{F}_t, \dot{F}_t^d) = \int_{\Omega_R} (\dot{\varphi}(F_t, \nabla F_t) + \dot{\psi}(F_t^d, \nabla F_t^d) + \chi(\dot{F}_t^d, \nabla \dot{F}_t^d))\, \mathsf{dv} \tag{4-10}$$

is obtained by time differentiation of (4-1).

An energy density is recoverable if it is differentiable.[54] Then we assume that both φ and ψ are differentiable. For the dissipation potential χ, we assume that it is a convex function, positive except at $(\dot{F}_t^d, \nabla \dot{F}_t^d) = (0, 0)$, where $\chi(0, 0) = 0$.[55]

4.3. ***The equation of virtual power.*** By the assumed differentiability of φ and ψ, in the expression (4-10) of the internal power we have

$$\begin{aligned} \dot{\varphi}(F_t, \nabla F_t) &= \varphi_F(F_t, \nabla F_t) \cdot \dot{F}_t + \varphi_{\nabla F}(F_t, \nabla F_t) \cdot \nabla \dot{F}_t, \\ \dot{\psi}(F_t^d, \nabla F_t^d) &= \psi_{F^d}(F_t^d, \nabla F_t^d) \cdot \dot{F}_t^d + \psi_{\nabla F^d}(F_t^d, \nabla F_t^d) \cdot \nabla \dot{F}_t^d, \end{aligned} \tag{4-11}$$

with φ_F, $\varphi_{\nabla F}$ and ψ_{F^d}, $\psi_{\nabla F^d}$ the partial derivatives of φ and ψ. Equating the internal power (4-10) to the reduced external power (3-21) and recalling that $\nabla \dot{f} = \dot{F} + \dot{F}^d$, the power equation (4-4) takes the form[56]

$$\begin{aligned} \int_{\Omega_R} ((T_{Rt} + B_{Rt} - \varphi_{F_t}) \cdot \dot{F}_t - \varphi_{\nabla F_t} \cdot \nabla \dot{F}_t + (T_{Rt} - \psi_{F_t^d}) \cdot \dot{F}_t^d - \psi_{\nabla F_t^d} \cdot \nabla \dot{F}_t^d - \chi(\dot{F}_t^d, \nabla \dot{F}_t^d))\, \mathsf{dv} \\ + \int_{\partial \Omega_R} (S_{Rt} \cdot \dot{F}_t)\, \mathsf{da} = 0, \end{aligned} \tag{4-12}$$

and applying the divergence theorem to the terms involving $\varphi_{\nabla F_t}$ and $\psi_{\nabla F_t^d}$ we get

$$\begin{aligned} &\int_{\Omega_R} ((T_{Rt} + B_{Rt}) \cdot \dot{F}_t + T_{Rt} \cdot \dot{F}_t^d)\, \mathsf{dv} + \int_{\partial \Omega_R} S_{Rt} \cdot \dot{F}_t\, \mathsf{da} \\ &\quad = \int_{\Omega_R} ((\varphi_{F_t} - \operatorname{div} \varphi_{\nabla F_t}) \cdot \dot{F}_t + (\psi_{F_t^d} - \operatorname{div} \psi_{\nabla F_t^d}) \cdot \dot{F}_t^d + \chi(\dot{F}_t^d, \nabla \dot{F}_t^d))\, \mathsf{dv} \\ &\qquad + \int_{\partial \Omega_R} (\varphi_{\nabla F_t} n_R \cdot \dot{F}_t + \psi_{\nabla F_t^d} n_R \cdot \dot{F}_t^d)\, \mathsf{da}. \end{aligned} \tag{4-13}$$

[54] Indeed, if φ is differentiable, from the relation $\varphi(F_t, \nabla F_t) = \varphi(F_0, \nabla F_0) + \varphi_F(F_0, \nabla F_0) \cdot (F_t - F_0) + \varphi_{\nabla F}(F_0, \nabla F_0) \cdot (\nabla F_t - \nabla F_0) + o(|(F_t - F_0)|, |\nabla F_t - \nabla F_0|)$, it follows that $\varphi(F_t, \nabla F_t) = \varphi(F_0, \nabla F_0)$ in every evolution with $(F_t, \nabla F_t) = (F_0, \nabla F_0)$.

[55] This is the simplest way for making the dissipation dependent on the past history of the evolution $t \mapsto (F_t^d, \nabla F_t^d)$.

[56] Here φ_{F_t} is an abbreviation for $\varphi_F(F_t, \nabla F_t)$, and so on.

For a given evolution $t \mapsto (f_t, F_t)$, consider the *perturbed evolution* starting at the instant t_0

$$f_{\varepsilon t} = f_t + \varepsilon(t - t_0)v, \qquad F_{\varepsilon t} = F_t + \varepsilon(t - t_0)V, \tag{4-14}$$

with v and V arbitrary vector and second-order tensor fields. By time differentiation, we get the *perturbed velocities*

$$\dot{F}_{\varepsilon t} = \dot{F}_t + \varepsilon V, \qquad \dot{F}^d_{\varepsilon t} = \dot{F}^d_t + \varepsilon V^d, \tag{4-15}$$

where $V^d = \nabla v - V$. Writing (4-13) for the perturbed evolution and subtracting the same equation for the unperturbed evolution, the *equation of virtual power*

$$\begin{aligned}\int_{\Omega_R} ((T_{Rt} + B_{Rt}) \cdot V + T_{Rt} \cdot V^d)\,\mathrm{dv} + \int_{\partial\Omega_R} S_{Rt} \cdot V\,\mathrm{da}\\ = \int_{\Omega_R} ((\varphi_{F_t} - \operatorname{div}\varphi_{\nabla F_t}) \cdot V + (\psi_{F^d_t} - \operatorname{div}\psi_{\nabla F^d_t}) \cdot V^d + \delta\chi(\dot{F}^d_t, \nabla\dot{F}^d_t, V^d, \nabla V^d))\,\mathrm{dv}\\ + \int_{\partial\Omega_R} (\varphi_{\nabla F_t} n_R \cdot V + \psi_{\nabla F^d_t} n_R \cdot V^d)\,\mathrm{da}\end{aligned} \tag{4-16}$$

is obtained, where

$$\begin{aligned}\delta\chi(\dot{F}^d_t, \nabla\dot{F}^d_t, V^d, \nabla V^d)\\ = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}(\chi(\dot{F}^d_t + \varepsilon V^d, \nabla\dot{F}^d_t + \varepsilon\nabla V^d) - \chi(\dot{F}^d_t, \nabla\dot{F}^d_t)).\end{aligned} \tag{4-17}$$

The equation of virtual power can be viewed as a purely mechanical version of the Clausius–Duhem inequality.[57] In the more general context of thermodynamics, this inequality was exploited by Coleman and Noll to get restrictions on the form of the constitutive equations.[58] On the contrary, in the present purely mechanical context the Coleman–Noll procedure will be used to deduce the field equations and boundary conditions for the initial/boundary-value problem.

4.4. ***The Coleman–Noll procedure.*** Consider first a purely macroscopic perturbation, in which $V^d = 0$ and V is an arbitrary second-order tensor field. Due to the arbitrariness of V, from (4-16) we get the local conditions

$$\begin{aligned}\varphi_{F_t} - \operatorname{div}\varphi_{\nabla F_t} &= T_{Rt} + B_{Rt} \quad &&\text{in } \Omega_R,\\ S_{Rt} &= \varphi_{\nabla F_t} n_R \quad &&\text{on } \partial\Omega_R.\end{aligned} \tag{4-18}$$

[57] Called the *free-energy imbalance* in [Gurtin et al. 2010, §29]. Here the inequality reduces to equality because the whole dissipation is supposed to be due to the dissipation potential χ.

[58] See [Coleman and Noll 1963].

Equation (4-16) then reduces to

$$\int_{\Omega_R} ((T_{Rt} - \psi_{F_t^d} + \operatorname{div} \psi_{\nabla F_t^d}) \cdot V^d - \delta\chi(\dot{F}_t^d, \nabla\dot{F}_t^d, V^d, \nabla V^d)) \,\mathrm{dv} - \int_{\partial\Omega_R} \psi_{\nabla F_t^d} n_R \cdot V^d \,\mathrm{da} = 0. \quad (4\text{-}19)$$

In particular, taking V^d with support on a ball of radius ε centered first at an interior point and then at a boundary point of Ω_R, we get the local conditions

$$\begin{aligned} (T_{Rt} - \psi_{F_t^d} + \operatorname{div} \psi_{\nabla F_t^d}) \cdot V^d - \delta\chi(\dot{F}_t^d, \nabla\dot{F}_t^d, V^d, \nabla V^d) = 0 \quad &\text{in } \Omega_R, \\ (\psi_{\nabla F_t^d} n_R) \cdot V^d = 0 \quad &\text{on } \partial\Omega_R. \end{aligned} \quad (4\text{-}20)$$

These conditions cannot be further refined, as long as the term $\delta\chi$ remains unspecified. The forms taken by $\delta\chi$ for some special forms of the dissipation potential χ are specified in the next section.

4.5. ***Special forms of the dissipation potential.*** According to the definition given in Section 4.2, a dissipation potential is a convex function of the pair $(\dot{F}^d, \nabla\dot{F}^d)$, positive except at $(0, 0)$, and with $\chi(0, 0) = 0$. In the literature we find dissipation potentials of two types, positively homogeneous of order one and of order two.[59] In what follows we deal with such dissipation potentials, and we initially consider functions χ independent of $\nabla\dot{F}^d$.

A positively homogeneous χ of order two and differentiable at $\dot{F}^d = 0$ has the form

$$\chi(\dot{F}^d) = \tfrac{1}{2}\chi_{\dot{F}^d\dot{F}^d}(0)[\dot{F}^d] \cdot \dot{F}^d, \quad (4\text{-}21)$$

and after some routine computation, from (4-18) we get

$$\delta\chi(\dot{F}^d, V^d) = \chi_{\dot{F}^d\dot{F}^d}(0)[\dot{F}^d] \cdot \dot{V}^d. \quad (4\text{-}22)$$

A positively homogeneous χ of order one cannot be differentiable, but only directionally differentiable at $\dot{F}^d = 0$. In this case it has the form

$$\chi(\dot{F}^d) = \breve{\chi}_{\dot{F}^d}(0) \triangleright \dot{F}^d, \quad (4\text{-}23)$$

where $\breve{\chi}_{\dot{F}^d}(0) \triangleright V^d$ denotes the directional derivative of χ at zero in the direction $\dot{F}^d$.[60] From convexity, we have

$$\chi(\tfrac{1}{2}(\dot{F}^d + \varepsilon V^d)) \le \tfrac{1}{2}(\chi(\dot{F}^d) + \chi(\varepsilon V^d)), \quad (4\text{-}24)$$

[59] See, e.g., [Ziegler 1963; Moreau 1970; Hackl and Fischer 2008].

[60] This notation was introduced in [Del Piero 2018c]. An example of a dissipation potential of this form is $\chi(\dot{F}^d) = k|\dot{F}^d|$, with k a positive constant.

with equality for $V^d = \dot{F}^d$, and from homogeneity it follows that

$$\chi(\dot{F}^d + \varepsilon V^d) \le \chi(\dot{F}^d) + \varepsilon\chi(V^d), \tag{4-25}$$

with equality for $V^d = \dot{F}^d$. Then, by (4-18), (4-23), and (4-25),

$$\delta\chi(\dot{F}^d, V^d) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}(\chi(\dot{F}^d + \varepsilon V^d) - \chi(\dot{F}^d)) \le \chi(V^d) = \check{\chi}_{\dot{F}^d}(0) \triangleright V^d, \tag{4-26}$$

with equality for $V^d = \dot{F}^d$.

Let us now consider dissipation potentials depending on both $\dot{F}^d$ and $\nabla\dot{F}^d$, of the special form

$$\chi(\dot{F}^d, \nabla\dot{F}^d) = \chi'(\dot{F}^d) + \chi''(\nabla\dot{F}^d), \tag{4-27}$$

with both χ' and χ'' positive, convex, and differentiable everywhere except at $\dot{F}^d = 0$ and $\nabla\dot{F}^d = 0$, respectively. It can be easily verified that if both χ' and χ'' are positively homogeneous of order two, then

$$\delta\chi(\dot{F}^d, \nabla\dot{F}^d_t, V^d, \nabla V^d) = \chi'_{\dot{F}^d\dot{F}^d}(0)[\dot{F}^d]\cdot V^d + \chi''_{\nabla\dot{F}^d\nabla\dot{F}^d}(0)[\nabla\dot{F}^d]\cdot\nabla V^d, \tag{4-28}$$

and that if both are positively homogeneous of order one, then

$$\delta\chi(\dot{F}^d, \nabla\dot{F}^d_t, V^d, \nabla V^d) \le \check{\chi}'_{\dot{F}^d}(0) \triangleright V^d + \check{\chi}''_{\nabla\dot{F}^d}(0) \triangleright \nabla V^d, \tag{4-29}$$

with equality for $V^d = \dot{F}^d$ and $\nabla V^d = \nabla\dot{F}^d$.

5. Initial/boundary-value problems

Equations (4-18) and (4-20) are the result of the Coleman–Noll procedure applied to the present model. On them and on the indifference conditions (3-10) and (3-14) is based the formulation of the initial/boundary-value problems for two-scale continua. To begin, we consider the case without dissipation.

5.1. ***The nondissipative problem.*** If there is no dissipation, then $\chi = \delta\chi = 0$ and, since V^d is allowed to be any second-order tensor, conditions (4-20) reduce to

$$\begin{aligned} \psi_{F^d_t} - \operatorname{div}\psi_{\nabla F^d_t} &= T_{Rt} \quad \text{in } \Omega_R, \\ \psi_{\nabla F^d_t} n_R &= 0 \quad \text{on } \partial\Omega_R. \end{aligned} \tag{5-1}$$

Together with (3-18) and $(4\text{-}18)_1$, $(5\text{-}1)_1$ forms the differential system

$$\operatorname{div} T_R + b_R = 0, \qquad \varphi_F - \operatorname{div}\varphi_{\nabla F} = T_R + B_R, \qquad \psi_{F^d} - \operatorname{div}\psi_{\nabla F^d} = T_R, \tag{5-2}$$

in which only the current values of the variables at a fixed instant are involved. Therefore, equations relative to different t are independent.[61] By consequence, the deformation process $t \mapsto (f_t, F_t)$ due to given external actions $t \mapsto (b_{Rt}, B_{Rt})$

[61] For this reason, the subscript t has been omitted.

and to given boundary data and initial conditions is obtained solving a family of boundary value problems, one for each t. No initial conditions are required, and the boundary conditions can be either *boundary conditions of place*

$$f = \hat{f}, \qquad F = \hat{F}, \qquad F^d = 0, \tag{5-3}$$

or *boundary conditions of traction*

$$T_R n_R = \hat{s}_R, \qquad \varphi_{\nabla F} n_R = \hat{S}_R, \qquad \psi_{\nabla F^d} n_R = 0. \tag{5-4}$$

Equations (5-2)–(5-4) form the equilibrium problem which determines the deformation (f, F) as a function of the external actions (b_R, B_R) and of the boundary data $(\hat{f}, \hat{F})$ or $(\hat{s}_R, \hat{S}_R)$.

To investigate the well-posedness of this problem, it may be helpful to consider the elementary example of a one-dimensional body

$$\Omega_R = \{x_R = x_0 + \lambda(x_l - x_0) \mid \lambda \in (0, 1)\}, \tag{5-5}$$

subject to null external actions, $b_R = B_R = 0$, and with the quadratic energies

$$\varphi(F) = \tfrac{1}{2}k|F|^2 + \tfrac{1}{2}K|\nabla F|^2, \qquad \psi(F^d) = \tfrac{1}{2}k_d|F^d|^2 + \tfrac{1}{2}K_d|\nabla F^d|^2, \tag{5-6}$$

with k, K and k_d, K_d given positive constants. Since in one dimension both gradient and divergence coincide with the ordinary derivative, (5-2) reduces to

$$\nabla T_R + b_R = 0, \qquad kF - K\nabla^2 F = T_R + B_R, \qquad k_d F^d - K_d \nabla^2 F^d = T_R, \tag{5-7}$$

where ∇ now denotes the ordinary derivative. From $(5\text{-}7)_1$ integrated over (x_0, x_R) we have the equation

$$T_R(x_R) = T_R(x_0) - \int_{x_0}^{x_R} b_R(\xi_R)\, d\xi_R, \tag{5-8}$$

which determines T_R to within the constant $T_R(x_0)$. The two remaining equations (5-7) are second-order differential equations with constant coefficients. They admit the closed-form solutions

$$\begin{aligned} F(x_R) &= A\exp(\kappa^2 x_R) + B\exp(-\kappa^2 x_R) + \bar{F}(x_R), & \kappa^2 &= k/K, \\ F^d(x_R) &= A^d\exp(\kappa_d^2 x_R) + B^d\exp(-\kappa_d^2 x_R) + \bar{F}^d(x_R), & \kappa_d^2 &= k_d/K_d, \end{aligned} \tag{5-9}$$

where $\bar{F}$ and $\bar{F}^d$ are particular solutions of $(5\text{-}7)_1$ and $(5\text{-}7)_2$, respectively.

For example, if b_R and B_R are the constant fields $\hat{b}_R$ and $\hat{B}_R$, for $x_0 = 0$ and $x_l - x_0 = l$, we have

$$k\bar{F}(x_R) = T_R(0) + \hat{B}_R + \hat{b}_R x_R, \qquad k_d\bar{F}^d(x_R) = T_R(0) + \hat{b}_R x_R. \tag{5-10}$$

The constants A, B, A^d, B^d, and $T_R(0)$ are determined by the boundary conditions (5-3) and (5-4). For example, for the boundary conditions of place

$$f(0)=\hat{f}_0, \quad f(l)=\hat{f}_l, \quad F(0)=\hat{F}_0, \quad F(l)=\hat{F}_l, \quad F^d(0)=F^d(l)=0, \tag{5-11}$$

the five constants are the solution of the linear system

$$\begin{aligned}
A+B+k^{-1}(T_R(0)+\hat{B}_R) &= \hat{F}_0,\\
A\exp(\kappa^2 l)+B\exp(-\kappa^2 l)+k^{-1}(T_R(0)+\hat{B}_R+\hat{b}_R l) &= \hat{F}_l,\\
A^d+B^d+k_d^{-1}T_R(0) &= 0,\\
A^d\exp(\kappa_d^2 l)+B^d\exp(-\kappa_d^2 l)+k_d^{-1}(T_R(0)+\hat{b}_R l) &= 0,\\
A\kappa^{-2}(\exp(\kappa^2 l)-1)-B\kappa^{-2}(\exp(-\kappa^2 l)-1)&\\
+A^d\kappa_d^{-2}(\exp(\kappa_d^2 l)-1)-B^d\kappa_d^{-2}(\exp(-\kappa_d^2 l)-1)&\\
+k^{-1}((T_R(0)+\hat{B}_R)l+\hat{b}_R l^2/2)+k_d^{-1}(T_R(0)l+\hat{b}_R l^2/2) &= \hat{f}_l-\hat{f}_0,
\end{aligned} \tag{5-12}$$

in which the last equation comes from the integration over $(0,l)$ of the relation $F+F^d=\nabla f$.

5.2. ***The rate-dependent dissipative problem.*** Consider a dissipation potential of the form (4-27), positively homogeneous of order two:

$$\begin{aligned}
\chi(\dot{F}^d,\nabla\dot{F}^d) &= \tfrac{1}{2}\mathbb{H}[\dot{F}^d]\cdot\dot{F}^d+\tfrac{1}{2}\mathbb{K}[\nabla\dot{F}^d]\cdot\nabla\dot{F}^d,\\
\mathbb{H}=\chi'_{\dot{F}^d\dot{F}^d}(0), &\qquad \mathbb{K}=\chi''_{\nabla\dot{F}^d\nabla\dot{F}^d}(0).
\end{aligned} \tag{5-13}$$

Because it describes the dissipation exhibited by viscous materials, this χ can be called a *dissipation potential of the viscous type*, and since in the power $\chi(\dot{F}^d,\nabla\dot{F}^d)$ the internal actions $\mathbb{H}[\dot{F}^d]$, $\mathbb{K}[\nabla\dot{F}^d]$ depend on the time derivative of F^d, this χ is *rate-dependent*. For $\delta\chi$, from (4-28) we have

$$\delta\chi(\dot{F}^d,\nabla\dot{F}^d,V^d,\nabla V^d)=\mathbb{H}[\dot{F}^d]\cdot V^d+\mathbb{K}[\nabla\dot{F}^d]\cdot\nabla V^d, \tag{5-14}$$

and if V^d is allowed to be any second-order tensor field, from (4-19) we get

$$\begin{aligned}
T_{Rt}-\psi_{F_t^d}+\operatorname{div}\psi_{\nabla F_t^d}-\mathbb{H}[\dot{F}_t^d]+\operatorname{div}(\mathbb{K}[\nabla\dot{F}_t^d])=0 \quad &\text{in } \Omega_R,\\
(\psi_{\nabla F_t^d}+\mathbb{K}[\nabla\dot{F}_t^d])n_R=0 \quad &\text{on } \partial\Omega_R.
\end{aligned} \tag{5-15}$$

Due to the presence of $\dot{F}^d$ and $\nabla\dot{F}^d$, there is a relation between equations written at different times. The family of equilibrium problems (5-2)–(5-4) then merges into a single initial/boundary-value problem, consisting of (3-18), $(4\text{-}18)_1$, $(5\text{-}15)_1$, plus initial conditions $(\hat{f}_0,\hat{F}_0)$ over Ω_R and a family $t\mapsto(\hat{f}_t,\hat{F}_t)$ or $t\mapsto(\hat{s}_{Rt},\hat{S}_{Rt})$ of boundary conditions. For this problem, a standard solution strategy is to transform it into a family of *incremental problems*, one for each t, in which the deformation

(f_t, F_t) and the stress T_{Rt}, both compatible with (3-18) and $(4\text{-}18)_1$, are supposed to be known at some instant t, and (5-15) is used to determine the derivative $\dot{F}^d_t$.

For example, in the case of φ, ψ, χ independent of the gradients ∇F, ∇F^d, $\nabla \dot{F}^d$, the tensor $\mathbb{K}$ is zero, $\dot{F}^d_t$ is determined by the *flow rule*[62]

$$\dot{F}^d_t = \mathbb{H}^{-1}[T_{Rt} - \psi_{F^d}(F^d_t)], \tag{5-16}$$

and $(\dot{f}_t, \dot{F}_t)$ and $\dot{T}_{Rt}$ are determined solving the incremental problem formed by the incremental versions of (3-18) and $(4\text{-}18)_1$,

$$\operatorname{div} \dot{T}_{Rt} + \dot{b}_{Rt} = 0, \qquad \dot{T}_{Rt} = \varphi_{FF}(F_t)[\dot{F}_t] - \dot{B}_{Rt}, \tag{5-17}$$

and by the corresponding boundary conditions.

An even more elementary example is the following. Assume that the body is a set of three points, with quadratic energies independent of the gradient,

$$\varphi(F) = \tfrac{1}{2}c_1|F|^2, \qquad \psi(F^d) = \tfrac{1}{2}c_2|F^d|^2, \qquad \chi(\dot{F}^d) = \tfrac{1}{2}c_3|\dot{F}^d|^2, \tag{5-18}$$

and that $b_{Rt} = B_{Rt} = 0$. Then from $(4\text{-}18)_1$ and $(5\text{-}15)_1$ we have

$$T_{Rt} = c_1 F_t, \qquad T_{Rt} = c_2 F^d_t + c_3 \dot{F}^d_t. \tag{5-19}$$

These are the equations of the Kelvin–Voigt model, which consists of a spring of stiffness c_1, in series with a system made of a second spring of stiffness c_2 and a piston of stiffness c_3, set in parallel. The elongation of the first spring is F_t, and F^d_t is the elongation of both the second spring and the piston. The corresponding internal actions are $c_1 F_t$, $c_2 F^d_t$, $c_3 \dot{F}^d_t$, the external action is T_{Rt}, and (5-19) are the balance equations between internal and external actions. By elimination of T_{Rt}, we get the differential equation

$$c_3 \dot{F}^d_t + c_2 F^d_t = c_1 F_t. \tag{5-20}$$

The boundary condition of place consists of prescribing the total elongation

$$\hat{F}^{\mathrm{tot}}_t = F_t + F^d_t \tag{5-21}$$

at all t, and the initial conditions consist of prescribing F_t and F^d_t at $t = 0$. However, due to the compatibility condition $F_0 + F^d_0 = \hat{F}^{\mathrm{tot}}_0$ at $t = 0$, only one of the conditions $F_0 = \hat{F}_0$ or $F^d_0 = \hat{F}^d_0$ is required. Thus, we have the differential problem

$$\dot{F}^d_t + k F^d_t = c_1 c_3^{-1} \hat{F}^{\mathrm{tot}}_t, \quad k = (c_1 + c_2)c_3^{-1}, \tag{5-22}$$

with $\hat{F}^{\mathrm{tot}}_t$ given and with, for example, $F_0 = \hat{F}_0$. The solution is

$$F^d_t = A \exp(-kt) + \bar{F}_t, \tag{5-23}$$

[62] The convexity of χ' at $F^d = 0$ requires that the fourth-order tensor $\mathbb{H}$ be positive definite. We assume that it is in fact positive definite.

where $\bar{F}_t$ is a solution of the nonhomogeneous equation (5-22), and the constant A is determined by the initial condition

$$\hat{F}_0^{\text{tot}} = \hat{F}_0 + A + \bar{F}_0, \tag{5-24}$$

which is (5-21) at $t = 0$.

5.3. ***The rate-independent dissipative problem.*** For a dissipation potential positively homogeneous of order one of the form (4-27), from (4-23) we have

$$\chi(\dot{F}^d, \nabla\dot{F}^d) = \chi'_{\dot{F}^d}(0) \triangleright \dot{F}^d + \chi''_{\nabla\dot{F}^d}(0) \triangleright \nabla\dot{F}^d. \tag{5-25}$$

Since it describes the dissipation exhibited by plastic materials, this χ can be called a *dissipation potential of the plastic type*, and because in the power $\chi(\dot{F}^d, \nabla\dot{F}^d)$ the internal actions $\chi'_{\dot{F}^d}(0)$, $\chi''_{\nabla\dot{F}^d}(0)$ do not depend on time derivatives, this χ is *rate-independent*. Then from (4-19) and from the upper bound (4-29) on $\delta\chi$ we have

$$\int_{\Omega_R} \big((T_{Rt} - \psi_{F_t^d} + \operatorname{div}\psi_{\nabla F_t^d}) \cdot V^d - \breve{\chi}'_{\dot{F}^d}(0) \triangleright V^d - \breve{\chi}''_{\nabla\dot{F}^d}(0) \triangleright \nabla V^d\big)\,\mathsf{dv} - \int_{\partial\Omega_R} \psi_{\nabla F_t^d} \cdot (V^d \otimes n_R)\,\mathsf{da} \le 0, \tag{5-26}$$

and from the divergence formula (A-4) in the Appendix we get

$$\int_{\Omega_R} \big((T_{Rt} - \psi_{F_t^d} + \operatorname{div}\psi_{\nabla F_t^d}) \cdot V^d - (\breve{\chi}'_{\dot{F}_t^d}(0) - \operatorname{div}\breve{\chi}''_{\nabla\dot{F}_t^d}(0)) \triangleright V^d\big)\,\mathsf{dv} - \int_{\partial\Omega_R} \big(\psi_{\nabla F_t^d} \cdot (V^d \otimes n_R) + \breve{\chi}''_{\nabla\dot{F}^d}(0) \triangleright (V^d \otimes n_R)\big)\,\mathsf{da} \le 0. \tag{5-27}$$

For $V^d = \dot{F}_t^d$ the inequality turns to equality, and the punctual conditions

$$\begin{aligned} (T_{Rt} - \psi_{F_t^d} + \operatorname{div}\psi_{\nabla F_t^d}) \cdot \dot{F}_t^d - (\breve{\chi}'_{\dot{F}_t^d}(0) - \operatorname{div}\breve{\chi}''_{\nabla\dot{F}_t^d}(0)) \triangleright \dot{F}_t^d = 0 \quad &\text{in } \Omega_R, \\ \psi_{\nabla F_t^d} n_R \cdot \dot{F}_t^d + \breve{\chi}''_{\nabla\dot{F}_t^d}(0) \triangleright (\dot{F}_t^d \otimes n_R) = 0 \quad &\text{on } \partial\Omega_R \end{aligned} \tag{5-28}$$

follow. Calling N_t the direction $\dot{F}_t^d / |\dot{F}_t^d|$ of $\dot{F}_t^d$, the first equation can be given the form

$$|\dot{F}_t^d| ((T_{Rt} - \psi_{F_t^d} + \operatorname{div}\psi_{\nabla F_t^d}) \cdot N_t - (\breve{\chi}'_{\dot{F}_t^d}(0) - \operatorname{div}\breve{\chi}''_{\nabla\dot{F}_t^d}(0)) \triangleright N_t) = 0. \tag{5-29}$$

It tells us that a nonnull deformation $\dot{F}_t^d$ can occur only if there is a direction N_t for which the term between parentheses is zero. This is the *activation condition* for the strain rate $\dot{F}_t^d$, and

$$\dot{F}_t^d = \begin{cases} |\dot{F}_t^d| N_t & \text{if } N_t \text{ is a solution of problem (5-28),} \\ 0 & \text{otherwise} \end{cases} \tag{5-30}$$

is the corresponding *flow rule*. The determination of N_t requires the solution of the differential problem (5-28). Once again we have an initial/boundary-value problem, which can be reduced to a family of incremental problems. Each problem is formulated like in the rate-dependent case, with the *nonlocal* flow rule (5-30) in place of the local flow rule (5-16).[63]

For $\chi'' = 0$ and ψ independent of ∇F^d, which is the case of classical plasticity, the local form of inequality (5-27) is

$$(T_{Rt} - \psi_{F^d}(F_t^d)) \cdot V^d - \breve{\chi}_{\dot{F}^d}(0) \triangleright V^d \leq 0, \tag{5-31}$$

(5-29) reduces to

$$|\dot{F}_t^d|((T_{Rt} - \psi_{F^d}(F_t^d)) \cdot N_t - \breve{\chi}_{\dot{F}^d}(0) \triangleright N_t) = 0, \tag{5-32}$$

the flow rule (5-30) reduces to

$$\dot{F}_t^d = \begin{cases} |\dot{F}_t^d| N_t & \text{if } (T_{Rt} - \psi_{F^d}(F_t^d))N_t = \breve{\chi}_{\dot{F}^d}(0) \triangleright N_t, \\ 0 & \text{if } (T_{Rt} - \psi_{F^d}(F_t^d))N_t < \breve{\chi}_{\dot{F}^d}(0) \triangleright N_t, \end{cases} \tag{5-33}$$

and the differential problem (5-28) reduces to the determination of a unit tensor N_t for which the term between parentheses in (5-32) is zero.[64] The initial/boundary-value problem can still be reduced to a family of incremental problems, whose field equations are the incremental equations (5-17) and the flow rule (5-33).

5.4. ***A comparison with Gurtin and Anand's theory.*** It is instructive to compare the present theory with the theory of gradient plasticity of Gurtin and Anand.[65] The two theories differ in the forms assumed for the equation of virtual power. The form assumed by Gurtin and Anand is[66]

$$\int_{\Omega_R} b_R \cdot v \, \mathrm{dv} + \int_{\partial\Omega_R} (s_R \cdot v + S_R \cdot V^d) \, \mathrm{da} = \int_{\Omega_R} (T_R \cdot V + T_R^d \cdot V^d + \mathbb{T}_R \cdot \nabla V^d) \, \mathrm{dv}. \tag{5-34}$$

[63] Note that if $\dot{F}_t^d$ is a solution to problem (5-28), then $\lambda \dot{F}_t^d$ is a solution for all $\lambda \geq 0$. Correspondingly, the flow rule (5-30) specifies the direction, but not the amplitude, of the strain rate $\dot{F}_t^d$.

[64] In the paper [Del Piero 2018c] it has been shown that the set of all T_{Rt} for which this parenthesis is nonpositive for all V^d is a convex set in the space of the second-order tensors. This set is the *elastic range*, and the tensor ψ_{F^d} is the *backstress tensor* T_{Bt} of kinematic plasticity. A plastic deformation can be activated only if the tensor $(T_{Rt} - T_{Bt})$ is located on the *yield surface*, which is the boundary of the elastic range. In this case, N_t belongs to the normal cone to the yield surface at $(T_{Rt} - T_{Bt})$. This is the *normality law* (5-33). As pointed out in [Del Piero 2018c], what distinguishes the present approach from the traditional ones is that the activation condition, elastic range, backstress tensor, and normality law are not assumed a priori, but systematically deduced as consequences of a specific choice of the dissipation potential.

[65] See [Gurtin and Anand 2005]. Here and in the following I make reference to the version given in [Gurtin et al. 2010, §90]. To make the comparison easier, wherever possible I translated the notations of [Gurtin et al. 2010] into the ones of the present paper.

[66] See [Gurtin et al. 2010, (90.16)]. This form of the *principle of virtual power* was proposed by Fleck and Hutchinson [1993] with a scalar measure of plastic strain, and extended later to a tensorial measure by Gudmundson [2004]. See also [Fleck and Willis 2009a; 2009b].

Comparing with (4-16) above, we see that in the external power the action B_R is missing, and that the velocity associated with the contact action S_R is V^d instead of V. In the internal power, the internal actions are the stress tensors T_R, T_R^d, $\mathbb{T}_R$. No motivations are given for neglecting B_R, for taking V^d as independent virtual velocity, and for introducing the stress tensors T_R^d and $\mathbb{T}_R$.

With the aid of the divergence theorem and of the relation $V = \nabla v - V^d$, (5-34) takes the form

$$\int_{\Omega_R} ((\operatorname{div} T_R + b_R)\cdot v + (T_R - T_R^d + \operatorname{div}\mathbb{T}_R)\cdot V^d)\,\mathsf{dv} + \int_{\partial\Omega_R} ((s_R - T_R n_R)\cdot v + (S_R - \mathbb{T}_R n_R)\cdot V^d)\,\mathsf{da} = 0, \tag{5-35}$$

and from the arbitrariness of v and V^d the local equations[67]

$$\begin{aligned} \operatorname{div} T_R + b_R = 0, \qquad & T_R - T_R^d + \operatorname{div}\mathbb{T}_R = 0 && \text{in } \Omega_R, \\ s_R = T_R n_R, \qquad & S_R = \mathbb{T}_R n_R && \text{in } \partial\Omega_R \end{aligned} \tag{5-36}$$

follow. Moreover, T_R^d and $\mathbb{T}_R$ are split into the sum of an *energetic* and a *dissipative part*. For the energetic parts, a comparison between $(5\text{-}36)_1$ and (4-19) suggests the identifications

$$T_{R/\mathrm{en}}^d = \psi_{F^d}(F^d, \nabla F^d), \qquad \mathbb{T}_{R/\mathrm{en}} = \psi_{\nabla F^d}(F^d, \nabla F^d). \tag{5-37}$$

As for the dissipative parts, from (4-19) we have

$$T_{R/\mathrm{dis}}^d \cdot \dot{F}^d + \mathbb{T}_{R/\mathrm{dis}} \cdot \nabla\dot{F}^d = -\delta\chi(\dot{F}^d, \nabla\dot{F}^d, \dot{F}^d, \nabla\dot{F}^d), \tag{5-38}$$

and from (4-28) and (4-29) it follows that

$$T_{R/\mathrm{dis}}^d = \chi'_{\dot{F}^d \dot{F}^d}(0)[\dot{F}^d], \qquad \mathbb{T}_{R/\mathrm{dis}} = \chi''_{\nabla\dot{F}^d \nabla\dot{F}^d}(0)[\nabla\dot{F}^d] \tag{5-39}$$

for rate-dependent dissipation potentials, and

$$T_{R/\mathrm{dis}}^d = \check{\chi}'_{\dot{F}^d}(0), \qquad \mathbb{T}_{R/\mathrm{dis}} = \check{\chi}''_{\nabla\dot{F}^d}(0) \tag{5-40}$$

for rate-independent dissipation potentials.[68]

In [Gurtin et al. 2010], $(5\text{-}36)_{1,2}$ are the *macroscopic and microscopic force balances*, while (5-37), (5-39), and (5-40) are *constitutive equations* relating the

[67] See equations (90.18), (90.23), (90.17), and (90.24) of [Gurtin et al. 2010], respectively.

[68] In [Gurtin et al. 2010] the stress T_R^d is assumed to be totally dissipative, with constitutive equation (90.54). For $\mathbb{T}_R$ the constitutive equations for the energetic and dissipative parts are (90.47) and $(90.59)_2$, respectively. For $\mathbb{T}_{R/\mathrm{dis}}$ the identifications $(5\text{-}39)_2$ and $(5\text{-}40)_2$ are only approximative, because in [Gurtin et al. 2010] all constitutive functions are differentiable and there are no dissipation potentials.

forces T_R, T_R^d, $\mathbb{T}_R$ with the deformations F, F^d, and the flow rules (5-16) and (5-30) roughly correspond to (90.65) of [Gurtin et al. 2010].[69]

In the present theory there is only one *force*, the Piola stress T_R, and only one constitutive equation, given by $(4\text{-}18)_1$:

$$T_R + B_R = \varphi_F - \operatorname{div} \varphi_{\nabla F}. \tag{5-41}$$

The stress tensors T_R^d and $\mathbb{T}_R$ are artificial entities, not really required by the theory. Indeed, in the reduced equation of virtual power (4-19) the internal actions are the partial derivatives of φ, ψ, χ specified in the right-hand sides of (5-37)–(5-39). These derivatives contain all available information about the nature of the material response.

Since the internal actions are already expressed in terms of deformation, there is no need to invent stress tensors and constitutive equations relating them to the deformation. On the contrary, such a redundant procedure could be misleading because, while a constitutive equation is restricted only by indifference conditions, the internal actions may be interrelated by the fact of being partial derivatives of the same energy densities.

Appendix: The divergence of gradient-dependent positively homogeneous functions of order one

Let $\chi = \chi(\nabla \dot{F}^d)$ be a function with the same properties as the function χ'' in Section 5.3: convex, differentiable everywhere except at $\nabla \dot{F}^d = 0$, and positively homogeneous of order one. For such functions the directional derivatives in the direction ∇V^d are the same at zero and at ∇V^d:

$$\check{\chi}_{\nabla \dot{F}^d}(0) \triangleright \nabla V^d = \chi_{\nabla \dot{F}^d}(\nabla V^d) \cdot \nabla V^d. \tag{A-1}$$

Then, by the divergence theorem,

$$\begin{aligned} \int_{\Omega_R} \check{\chi}_{\nabla \dot{F}^d}(0) \triangleright \nabla V^d \, \mathsf{dv} &= \int_{\Omega_R} \chi_{\nabla \dot{F}^d}(\nabla V^d) \cdot \nabla V^d \, \mathsf{dv} \\ &= -\int_{\Omega_R} \operatorname{div} \chi_{\nabla \dot{F}^d}(\nabla V^d) \cdot V^d \, \mathsf{dv} + \int_{\partial \Omega_R} \chi_{\nabla \dot{F}^d}(\nabla V^d) n_R \cdot V^d \, \mathsf{da}. \end{aligned} \tag{A-2}$$

[69] Almost all current theories of generalized continua include one or more equations of microscopic force balance [Capriz 1989; Germain 1973b; Fleck and Hutchinson 1993; Gurtin et al. 2010]. The nature of such equations is far from being clear. It was just *believed* that the microscopic forces should be somehow balanced. In the more fundamental approaches [Gurtin and Martins 1976; Noll 1973; Šilhavý 2008], the macroscopic balance of linear momentum (3-18) was deduced from a boundedness assumption on the system of contact actions. In the microscopic scale, this was the approach followed in my previous papers [Del Piero 2009; 2014a; 2014b; 2017; 2018a; 2019], in which equations of the form $(5\text{-}36)_2$ were called *pseudobalance equations* because they were not expressing the balance of any physical quantity. Here, $(5\text{-}36)_2$ is a direct consequence of energy conservation. Therefore, it does not need any further physical characterization.

The divergence and boundary values of $\chi_{\nabla\dot{F}^d}$ can be extended to the singular point $\nabla V^d = 0$ by defining

$$\begin{aligned}(\operatorname{div} \breve{\chi}_{\nabla\dot{F}^d}(0)) \triangleright V^d &= \operatorname{div} \chi_{\nabla\dot{F}^d}(\nabla V^d) \cdot V^d,\\ \breve{\chi}_{\nabla F^d}(0) \triangleright (V^d \otimes n_R) &= \chi_{\nabla\dot{F}^d}(\nabla V^d) \cdot (V^d \otimes n_R).\end{aligned} \tag{A-3}$$

Then at $\nabla V^d = 0$ we have

$$\int_{\Omega_R} (\breve{\chi}_{\nabla\dot{F}^d}(0) \triangleright \nabla V^d + (\operatorname{div} \breve{\chi}_{\nabla\dot{F}^d})(0) \triangleright V^d)\, \mathrm{dv} = \int_{\partial\Omega_R} \breve{\chi}_{\nabla\dot{F}^d}(0) \triangleright (V^d \otimes n_R)\, \mathrm{da}. \tag{A-4}$$

This can be regarded as a generalized version of the divergence theorem at the singular point $\nabla V^d = 0$.

References

[Aifantis 1984] E. C. Aifantis, "On the microstructural origin of certain inelastic models", *ASME J. Eng. Mater. Technol.* **106**:4 (1984), 326–330.

[Barroso et al. 2017] A. C. Barroso, J. Matias, M. Morandotti, and D. R. Owen, "Second-order structured deformations: relaxation, integral representation and applications", *Arch. Ration. Mech. Anal.* **225**:3 (2017), 1025–1072.

[Capriz 1989] G. Capriz, *Continua with microstructure*, Springer Tracts in Natural Philosophy **35**, Springer, 1989.

[Choksi and Fonseca 1997] R. Choksi and I. Fonseca, "Bulk and interfacial energy densities for structured deformations of continua", *Arch. Rational Mech. Anal.* **138**:1 (1997), 37–103.

[Coleman and Noll 1963] B. D. Coleman and W. Noll, "The thermodynamics of elastic materials with heat conduction and viscosity", *Arch. Rational Mech. Anal.* **13** (1963), 167–178.

[Del Piero 2001] G. Del Piero, "The energy of a one-dimensional structured deformation", *Math. Mech. Solids* **6**:4 (2001), 387–408.

[Del Piero 2009] G. Del Piero, "On the method of virtual power in continuum mechanics", *J. Mech. Mater. Struct.* **4**:2 (2009), 281–292.

[Del Piero 2014a] G. Del Piero, "Nonclassical continua, pseudobalance, and the law of action and reaction", *Math. Mech. Complex Syst.* **2**:1 (2014), 71–107.

[Del Piero 2014b] G. Del Piero, "On the method of virtual power in the mechanics of non-classical continua", pp. 29–58 in *Multiscale modeling of complex materials: phenomenological, theoretical and computational aspects*, edited by T. Sadowski and P. Trovalusci, CISM Courses and Lectures **556**, Springer, 2014.

[Del Piero 2017] G. Del Piero, "Une approche rationnelle des milieux continus avec microstructure", pp. 11–114 in *Mécanique des milieux continus généralisés*, Cépaduès, Toulouse, France, 2017.

[Del Piero 2018a] G. Del Piero, "An axiomatic framework for the mechanics of generalized continua", *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.* **29**:1 (2018), 31–61.

[Del Piero 2018b] G. Del Piero, "On the decomposition of the deformation gradient in plasticity", *J. Elasticity* **131**:1 (2018), 111–124.

[Del Piero 2018c] G. Del Piero, "The variational structure of classical plasticity", *Math. Mech. Complex Syst.* **6**:3 (2018), 137–180.

[Del Piero 2019] G. Del Piero, "'Reality' and representation in mechanics: the legacy of Walter Noll", *J. Elasticity* **135**:1–2 (2019), 117–148.

[Del Piero 2020] G. Del Piero, "A mechanical model for heat conduction", *Contin. Mech. Thermodyn.* **32**:4 (2020), 1159–1172.

[Del Piero and Owen 1993] G. Del Piero and D. R. Owen, "Structured deformations of continua", *Arch. Rational Mech. Anal.* **124**:2 (1993), 99–155.

[Del Piero and Owen 1995] G. Del Piero and D. R. Owen, "Integral-gradient formulae for structured deformations", *Arch. Rational Mech. Anal.* **131**:2 (1995), 121–138.

[Fleck and Hutchinson 1993] N. A. Fleck and J. W. Hutchinson, "A phenomenological theory for strain gradient effects in plasticity", *J. Mech. Phys. Solids* **41**:12 (1993), 1825–1857.

[Fleck and Willis 2009a] N. A. Fleck and J. R. Willis, "A mathematical basis for strain-gradient plasticity theory, I: Scalar plastic multiplier", *J. Mech. Phys. Solids* **57**:1 (2009), 161–177.

[Fleck and Willis 2009b] N. A. Fleck and J. R. Willis, "A mathematical basis for strain-gradient plasticity theory, II: Tensorial plastic multiplier", *J. Mech. Phys. Solids* **57**:7 (2009), 1045–1057.

[Germain 1973a] P. Germain, "La méthode des puissances virtuelles en mécanique des milieux continus, I: Théorie du second gradient", *J. Mécanique* **12** (1973), 235–274.

[Germain 1973b] P. Germain, "The method of virtual power in continuum mechanics, 2: Microstructure", *SIAM J. Appl. Math.* **25**:3 (1973), 556–575.

[Germain et al. 1983] P. Germain, Q. S. Nguyen, and P. Suquet, "Continuum thermodynamics", *ASME J. Appl. Mech.* **50**:4b (1983), 1010–1020.

[Green and Rivlin 1957] A. E. Green and R. S. Rivlin, "The mechanics of non-linear materials with memory", *Arch. Rational Mech. Anal.* **1** (1957), 1–21.

[Gudmundson 2004] P. Gudmundson, "A unified treatment of strain gradient plasticity", *J. Mech. Phys. Solids* **52**:6 (2004), 1379–1406.

[Gurtin 1981] M. E. Gurtin, *An introduction to continuum mechanics*, Mathematics in Science and Engineering **158**, Academic Press, New York, 1981.

[Gurtin and Anand 2005] M. E. Gurtin and L. Anand, "A theory of strain-gradient plasticity for isotropic, plastically irrotational materials, II: Finite deformations", *Int J. Plasticity* **21**:12 (2005), 2297–2318.

[Gurtin and Martins 1976] M. E. Gurtin and L. C. Martins, "Cauchy's theorem in classical physics", *Arch. Rational Mech. Anal.* **60**:4 (1976), 305–324.

[Gurtin and Murdoch 1975] M. E. Gurtin and A. I. Murdoch, "A continuum theory of elastic material surfaces", *Arch. Rational Mech. Anal.* **57** (1975), 291–323.

[Gurtin et al. 2010] M. E. Gurtin, E. Fried, and L. Anand, *The mechanics and thermodynamics of continua*, Cambridge University Press, 2010.

[Hackl and Fischer 2008] K. Hackl and F. D. Fischer, "On the relation between the principle of maximum dissipation and inelastic evolution given by dissipation potentials", *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **464**:2089 (2008), 117–132.

[Halphen and Son 1975] B. Halphen and N. Q. Son, "Sur les matériaux standards généralisés", *J. Mécanique* **14** (1975), 39–63.

[Jammer 1999] M. Jammer, *Concepts of force*, Dover, Mineola, NY, 1999.

[Mach 1883] E. Mach, *Die Mechanik in ihrer Entwickelung: historisch-kritisch dargestellt*, Brockhaus, Leipzig, 1883. Translated as *The science of mechanics: a critical and historical account of its development*, 2nd ed., Open Court, Chicago, 1919.

[Mach 1896] E. Mach, *Die Principien der Wärmelehre*, Barth, Leipzig, 1896. Translated as *Principles of the theory of heat*, 2nd ed., Open Court, Chicago, 1904.

[Moreau 1970] J. J. Moreau, "Sur les lois de frottement, de plasticité et de viscosité", *C. R. Acad. Sci. Paris Sér. A* **271** (1970), 608–611.

[Moreau 1974] J.-J. Moreau, "On unilateral constraints, friction and plasticity", pp. 171–322 in *New variational techniques in mathematical physics* (Bressanone, Italy, 1973), edited by G. Capriz and G. Stampacchia, Cremonese, Rome, 1974.

[Murdoch 2012] A. I. Murdoch, *Physical foundations of continuum mechanics*, Cambridge University Press, 2012.

[Noll 1959] W. Noll, "The foundations of classical mechanics in the light of recent advances in continuum mechanics", pp. 266–281 in *The axiomatic method: with special reference to geometry and physics*, edited by L. Henkin et al., North-Holland, Amsterdam, 1959. Reprinted in [Noll 1974, pp. 31–47].

[Noll 1963] W. Noll, "La mécanique classique basée sur un axiome d'objectivité", pp. 47–56 in *La méthode axiomatique dans les mécaniques classiques et nouvelles*, edited by A. Chatelet, Les Grands Problèmes **13**, Gauthier-Villars, Paris, 1963. Reprinted in [Noll 1974, pp. 135–144].

[Noll 1973] W. Noll, "Lectures on the foundations of continuum mechanics and thermodynamics", *Arch. Rational Mech. Anal.* **52** (1973), 62–92. Reprinted in [Noll 1974, pp. 293–324].

[Noll 1974] W. Noll, *The foundations of mechanics and thermodynamics: selected papers*, Springer, 1974.

[Noll 1995] W. Noll, "On material frame-indifference", research report 95-NA-022, Carnegie Mellon University, 1995.

[Noll 2004] W. Noll, "Updating *The Non-Linear Field Theories of Mechanics*", pp. 23–39 in *Five contributions to natural philosophy*, 2004.

[Noll 2006] W. Noll, "On the past and future of natural philosophy", *J. Elasticity* **84**:1 (2006), 1–11.

[Owen 2000] D. R. Owen, "Structured deformations, II", lecture notes 58, Ravello Summer School of the Gruppo Nazionale di Fisica Matematica, 2000, Available at http://www.math.cmu.edu/users/nw0z/publications/99-CNA-014/99-CNA-014/.

[Reina and Conti 2014] C. Reina and S. Conti, "Kinematic description of crystal plasticity in the finite kinematic framework: a micromechanical understanding of $\mathbf{F} = \mathbf{F}^e\mathbf{F}^p$", *J. Mech. Phys. Solids* **67** (2014), 40–61.

[Truesdell 1991] C. Truesdell, *A first course in rational continuum mechanics*, vol. 1: General concepts, 2nd ed., Pure and Applied Mathematics **71**, Academic Press, Boston, 1991.

[Truesdell and Noll 1965] C. Truesdell and W. Noll, *The non-linear field theories of mechanics*, edited by S. Flügge, Handbuch der Physik **III-3**, Springer, 1965.

[Šilhavý 2008] M. Šilhavý, "Cauchy's stress theorem for stresses represented by measures", *Contin. Mech. Thermodyn.* **20**:2 (2008), 75–96.

[Šilhavý 2017] M. Šilhavý, "The general form of the relaxation of a purely interfacial energy for structured deformations", *Math. Mech. Complex Syst.* **5**:2 (2017), 191–215.

[Ziegler 1963] H. Ziegler, "Some extremum principles in irreversible thermodynamics with application to continuum mechanics", pp. 91–193 in *Progress in solid mechanics*, vol. IV, edited by I. N. Sneddon and R. Hill, North-Holland, Amsterdam, 1963.

Received 27 Jan 2020. Revised 31 May 2020. Accepted 8 Jul 2020.

GIANPIETRO DEL PIERO: dlpgpt@unife.it
Dipartimento di Ingegneria, Università di Ferrara, Ferrara, Italy

msp

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
Vol. 8, No. 3, 2020
https://doi.org/10.2140/memocs.2020.8.233

A NEW VIRUS-CENTRIC EPIDEMIC MODELING APPROACH 1. GENERAL THEORY AND MACHINE LEARNING SIMULATION OF 2020 SARS COV 2 (COVID-19) FOR BELGIUM, FRANCE, ITALY, AND SPAIN

JEAN RÉMOND AND YVES RÉMOND

We are trying to test the capacity of a simplified macroscopic virus-centric model to simulate the evolution of the SARS CoV 2 epidemic (COVID 19) at the level of a country or a geographical entity, provided that the evolution of the conditions of its development (behaviors, containment policies) are sufficiently homogeneous on the considered territory. For example, a uniformly deployed lockdown on the territory, or a sufficiently uniform overall crisis management. The virus-centric approach means that we favor to model the population dynamic of the virus rather than the evolution of the human cases. In other words, we model the interactions between the virus and its environment – for instance a specific human population with a specific behavior on a territory, instead of modeling the interactions between individuals. Therefore, our approach assumes that an epidemic can be analyzed as the combination of several elementary epidemics which represent a different part of the population with different behaviors through time. The modeling proposed here is based on the finite superposition of Verhulst equations commonly known as logistic functions and used in dynamics of population. Modelling the lockdown effect at the macroscopic level is therefore possible. Our model has parameters with a clear epidemiological interpretation, therefore the evolution of the epidemic can be discussed and compared among four countries: Belgium, France, Italy, and Spain. Parameter optimization is carried out by a classical machine learning process. We present the number of infected patients with SARS-CoV-2 and a comparison between data from the European Centre for Disease Prevention and Control and the modeling. In a general formulation, the model is applicable to any country with similar epidemic management characteristics. These results show that a simple two epidemics decomposition is sufficient to simulate with accuracy the effect of a lockdown on the evolution of the COVID-19 cases.

Communicated by Francesco dell'Isola.
MSC2020: 34C60, 68T20, 92-10, 92D25, 92D30.
Keywords: SARS-CoV-2, COVID-19, epidemic, modeling, simulation, machine learning, infected cases, logistic function.

1. Introduction

Significant progress has been made in epidemic modeling thanks to new capabilities of numerical simulation, improved mathematical modeling as well as artificial intelligence techniques. These modeling also benefit of continuous improvement in data quality. It is impossible to quote all the previous works done, and this is not the purpose of this article. However, the authors suggest consulting [Wiemken and Kelley April 2020] or [Colizza et al. December 2006], and among the most recent publications [Jia et al. 2020] or [Caccavo 2020] and the interesting state of the art concerning SIR and SEIR modelling published on the CNRS web site [Bayette and Monticelli 2020; Gonçalves 2020; Perra and Gonçalves 2015], as well as sites that allow interactive interfacing [Github]. The SIR model and their various developments are probably the cleverest approach at the micro and meso scale and permit to model a large range of macroscopic phenomena. Only a few specific studies used a single logistic function for modeling epidemics, especially for plant diseases and with interesting results, such as [Moral and Trapero 2009; Mesha and Hau 2008; Holb et al. 2005]. Some others used a double logistic curve for modeling HIV, including [Mahiane et al. 2017], or an r-hybrid model for the same virus [Eaton et al. 2019]. Another promising way consists in using PGD-like model reduction, as in [Chinesta and Cueto 2014], for analyzing epidemic kinetics by parametric optimization.

A large part of the existing approaches tries to model the epidemic at the individual scale, i.e. the microscopic scale, and consider the interactions between each individual. Then, it induces a theoretical epidemic evolution at the macroscopic level. A lot of contributions can be found in the literature using that method. This study takes the reverse way and try to find interesting conclusions depending on the microscopic scale, using a macroscopic modeling based on a generalization of the logistic function. It is a common approach developed in theoretical or applied mechanics or physics to use this type of homogenization methods to go from the macroscopic to the microscopic scale; see [Allaire 2001; Oleinik et al. 1992; Sanchez-Palencia 1980; Suquet 1987; Rémond et al. 2016]. Our simplified virus-centric macroscopic modeling is coupled with an automatic parameters optimization by machine learning and gives interesting results for the SARS-CoV-2 early 2020 pandemic. However, predicting the outcome of the epidemic across countries seems to be a lucky guess considering the variability of the containment policies through time and countries. Therefore, readers must be warned that our predictions, *mutatis mutandis*, cannot consider subsequent events such as a possible second epidemic, that could appear after the end of the lockdown, or other unexpected effects. However, the results obtained by this new and simplified approach seemed to us instructive enough to have explained it here.

2. General theory

2.1. ***Introduction to logistic function or sigmoid function.*** On a macroscopic scale, the elementary logistic function law, known as Verhulst's law [Verhulst 1838; Daley and Gani 1999], has been known for a long time (1838) as a law useful for classical modeling of epidemics at the macroscopic scale. It was first implemented in population dynamics. We can consider in a first approach that a population $y(t)$ - a read-valued function of time - of individuals evolves according to a very simple ordinary differential equation: $\frac{dy}{dt} = y(N(y) - M(y))$ where $N(y)$ is the birth rate and $M(y)$ represents the death rate. If $N(y)$ and $M(y)$ are linear functions, this equation can be written $\frac{dy}{dt} = ry\left(1 - \frac{y}{K}\right)$, a and K being strictly positive real numbers. K is conventionally called the carrying capacity in population dynamics theory, r is the growth rate which leads to an increase of population if $y(0) = y_0 < K$, to a decrease if $y_0 > K$ and is stable if $y_0 = K$.

The resolution of this simple ordinary differential equation allows to define the logistic function

$$y(t) = K\left[1 + \left(\frac{K}{y_0} - 1\right)e^{-rt}\right]^{-1}, \quad \lim_{t \to +\infty} y(t) = K. \tag{1}$$

This function is the solution defined on $[0, +\infty[$, of the system constituted by $y(0) = y_0$ and $y' = ry\left(1 - \frac{y}{K}\right)$.

2.2. ***General and macroscopic epidemic modeling.*** In this paper, we consider a virus centric epidemic modeling, which is the modeling of the evolution of virus through time in a specific environment. For this modeling we assume that the number cases of people can be assimilated to the number of virus. The human population of a given territory is the environment the virus has to survive into. The environment is more or less welcoming depending of the human behavior, the territory density and obviously the considered human sub-population. If we consider the whole human population, the number of virus is assimilated to the number of human COVID 19 cases. If we consider a more viable environment for the virus, for instance the sub-population of human susceptible enough to be hospitalized, the number of virus is assimilated of the number of hospitalizations. The same logical thinking could be applied for the number of intensive cares, death and recovery cases.

Therefore, we assume that the number of daily or cumulative cases of people is characterized by a function $E^k(t)$ of $\mathbb{R}$ in $\mathbb{R}$. $E^k(t)$ is defined over a given geographical territory with k the studied phenomena (infection, hospitalization, intensive care, death, recovery). Geographic territories should be chosen as territories in which we notice a similar epidemic management (for instance, territories with a similar lockdown intensity, as countries or other administrative entities, etc.).

This function $E^k(t)$ is itself the sum of continuous functions $E_i^k(t)$ of $\mathbb{R}$ in $\mathbb{R}$ with $i \in \{1, \ .., i, .., P\}$, characterizing the effect of the epidemic on P different populations belonging to the same geographic territory:

$$E^k(t) = \sum_{i=1}^{i=P} E_i^k(t)$$

Finally, each population may have a series of C_i different behaviors over time, and therefore

$$E_i^k(t) = \sum_{j=1}^{j=c_i} q_{ij}(t) E_{ij}^k(t),$$

$$E^k(t) = \sum_{i=1}^{i=P} \sum_{j=1}^{j=Ci} q_{ij}(t) E_{ij}^k(t) = q_{ij}(t) E_{ij}^k(t),$$

where the Einstein summation convention is in effect for i and j, and where $\sum_{j=1}^{j=C} q_{ij}(t) = 1$.

In the case of two different behaviors of a unique population i, the transition function $q(t)$ can be written as follow:

$$E_i^k(t) = q(t) E_{i1}^k(t) + (1 - q(t)) E_{i2}^k(t) \tag{2}$$

with $q(t)$ a monotonically increasing function defined from $[0, +\infty[$ to $]0, 1[$. Many functions may be suitable. We will define a specific one in the application paragraph.

3. Basic application to the 2020 Covid-19 epidemic

In the case of the Covid-19 epidemic which particularly occurred in Europe at the beginning of 2020, an interesting way to test the validity of such a model consists in taking particularly simplified hypothesis: one population per country ($i = 1$), two behaviors ($j = 2$), with a continuous passage from one to the other. Those two behaviors can be identified as the transition from a first behavior of the population before the containment measures to a second behavior considering the containment measures, especially the lockdown. Other subsequent behaviors could have been considered and identified, such as the end of containment measures, or a less rigorous behavior through time, lockdown, etc.

In our case, we define the elementary epidemic function $E^k(t)$ with an equation similar to (1):

$$E^k(t) = L_k[1 + e^{-r_k(t - t_{Sk})}]^{-1} \tag{3}$$

$E^k(t)$ thus, represents the cumulative number of cases of k (infected, hospitalized, in intensive care, deceased, cured) at time t. We will note L_k the final number of

cases of k in one epidemic, r_k the characteristic of the epidemic kinetics for the criterium k and t_S the time taken to reach the peak of the epidemic, described on a daily basis of new cases.

To model it, the following assumptions and development have been made:

H1: We assume that the global behavior of the population is given by the combination of two functions $E_1^k(t)$ and $E_2^k(t)$, as mention by equation (2), which characterize the difference of behavior, before and after the lockdown. As we will consider the unique criterium "number of infected cases" we will not use the exponent k anymore. Consequently, we have

$$E_1(t) = L_1[1 + e^{-r_1(t-t_{S1})}] \quad \text{and} \quad E_2(t) = L_2[1 + e^{-r_2(t-t_{S2})}].$$

The two functions $E_i(t)$ are distinguished by their respective coefficient r_i.

H2: Let us assume that the studied epidemic has a classical sigmoid / gaussian behavior. We have then the total number of cumulative case L and a "new infected case" peak t_s. We assume furthermore that the studied epidemy can be described as the combination of two epidemics with the same total number of cumulative cases L and the same peak t_s. Therefore, we keep the same value $L_i = L$ and we keep the same time $t_{Si} = t_S$ for the two functions $E_i(t)$.

H3: We choose the transition function $q(t)$ as $q(t) = \frac{1}{2}[\tanh(\alpha_c(t - t_i) + 1]$, defined on $[0, +\infty[$.

We obviously have $0 \leq q(t) \leq 1$. The coefficient α_c, positive number, represents the efficiency of the lockdown. t_i represents the duration between the start of the epidemic and the date of lockdown increased by the duration of appearance of its first effects. Due to the asymptotic variation of $q(t)$, $q(0) = \varepsilon \ll 1$ for the used values of t_i, which does not influence the results of the modeling.

With these assumptions, the evolution of the new infected cases of the epidemic is given by:

$$E(t) = q(t)E_1(t) + (1 - q(t))E_2(t)$$

The model is therefore defined by the six simple parameters given in Table 1.

4. Machine learning identification algorithm

To obtain the parameters of the model by supervised machine learning, a python code was developed using gradient descents conventionally used in machine learning. In our case, the Levenberg–Marquardt algorithm [Marquardt 1963; Levenberg 1944], was chosen to optimize the mean squared error function.

The learning datasets for each country have been compiled using official European information from the European Center for Disease Prevention and Control or

Parameters	Definition	Unit
L	Total number of cases	-
k_1	Characteristic of the epidemic without lockdown	Day^{-1}
k_2	Characteristic of the epidemic with lockdown	Day^{-1}
t_i	Duration between the start of the epidemic and the start of the lockdown + time of appearance of the first effects of the lockdown	Day
t_S	Duration from start of epidemic to date of inflection point*	Day
α_c	Lockdown efficiency coefficient	Day^{-1}

Table 1. List and definition of the model parameters. * The inflection point of the sigmoid corresponds to the maximum value of its derivative, which is often called the peak of the epidemic.

Parameters	Input value in the algorithm
L	$0.0031 \times$ Population of the country
k_1	0.3
k_2	0.1
t_i	Time from the start of the epidemic to the start of lockdown
t_S	Time between the start of the epidemic and the appearance of the inflection point + 15 days
α_c	0.1

Table 2. input values of the model parameters, i.e., before supervised learning.

ECDC [ECDC]. Considering the small amount of data, it was not possible to build a test data set.

To achieve the learning optimizations for each country, the following initial values have been set:

These values have been chosen through data analysis to be consistent with our model and the available experimental data.

For L, k_1, k_2, t_i, t_S, α_c, the input values influences the speed of the convergence of the algorithm and also ensure the bypassing of local minima:

L: the input value in the algorithm for the total number of people reached by Covid-19 was chosen to be 0.31% of the total population, following several data analysis.

k_1, k_2: their input values are linked to the analysis of the evolution of the epidemic as well as the analysis of experimental data.

t_i: the start dates of containment of the targeted countries are obviously available on government websites. They are as follows:

Country	Lockdown dates
Belgium	18 Mar 2020
France	17 Mar 2020
Italy	09 Mar 2020
Spain	14 Mar 2020

Ten days are added on those values, reflecting the delay needed to see the results of the lockdown. This duration, like the other parameters, is optimized by the algorithm.

t_S: the initial value for the inflection date t_S is chosen lockdown so t_S is greater than the lockdown date. 15 days were added on top of that after careful data analysis. α_c: the initial value was also set after data analysis and after the analysis of simulation curves.

Note that the total population of each country is also provided by the ECDC (https://www.ecdc.europa.eu/en); this information is reported to be from the World Bank Group (https://www.worldbank.org).

5. Results

5.1. ***The data.*** We took data from a single, reliable official source so that it could be compared across countries [ECDC]. It is obvious that regarding the detection of cases of infected persons, these data are highly dependent on the number of tests carried out and the identification protocols. In addition, it is regularly the case that the data is corrected later following updates. The figures used for Belgium, France, Italy, and Spain are shown in Figure 1. We can see that the raw data is unsurprisingly very noisy. We will not comment here on the roots of this fact. To increase the ability of the algorithm to quickly converge on a solution, we smoothed the raw data over several days (3 days, 5 days, 7 days). So, for a smoothing over three days, we have: $C_{i\,(\text{smoothed})} = \frac{1}{3}(C_{i-1} + C_i + C_{i+1})$. To minimize the noise effects and obtain the best possible convergence, we decided to use only the values smoothed over 7 days. The last data used dates from 21 April 2020. The smoothing over 7 days implies that the last valid dates for modeling correspond to 18 April 2020. We see on the following figures the raw data and the smoothed data for each country.

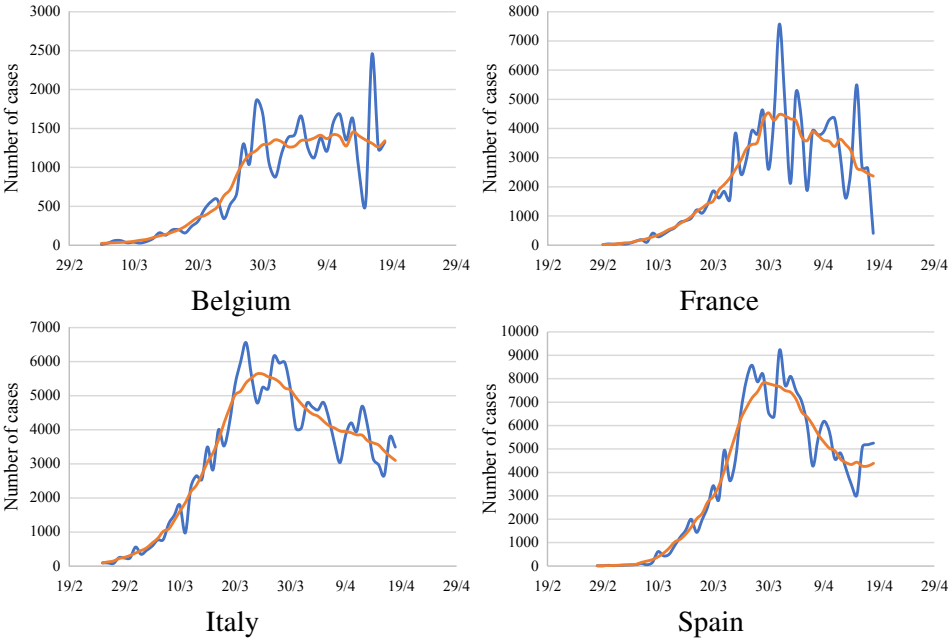


Figure 1. Number of detected cases of COVID-19 in 2020 in the four countries under consideration: raw data and data smoothed over 7 days. Values taken from [ECDC].

5.2. ***Comparison of the basic functions of the model by country.*** We see in Figure 2 comparisons of the graphs of the two epidemic functions $E_1(t)$ and $E_2(t)$ which frame the behavior of the epidemic between its standard evolution and its evolution with a lockdown from the start. A marked difference between these two graphs indicates a more significant effect of the lockdown. We could relate it to the efficiency of this lockdown in the considered territory.

The transition from the function $E_1(t)$ to the function $E_2(t)$ is done by the function $q(t)$ (Fig. 3). q(t) has been created to be a smoothed Heaviside step like function. The intensity of the slope is causally linked to the efficiency of the lockdown effect as the slope depends of the parameter α_c

5.3. ***Simulation results for past and current data.*** After the machine learning optimization of the parameters, the basic functions have been presented for each country. We now compare the smoothed data and the model in Figure 4.

5.4. ***Model predictions.*** So far, we presented the results up to the date of 18 April, i.e., up to the available data with which we trained the model. Obviously, one the main question is the quality of the predictions beyond this date. We present the results up to 18 May in Figure 5. Therefore, the reader can observe that the model

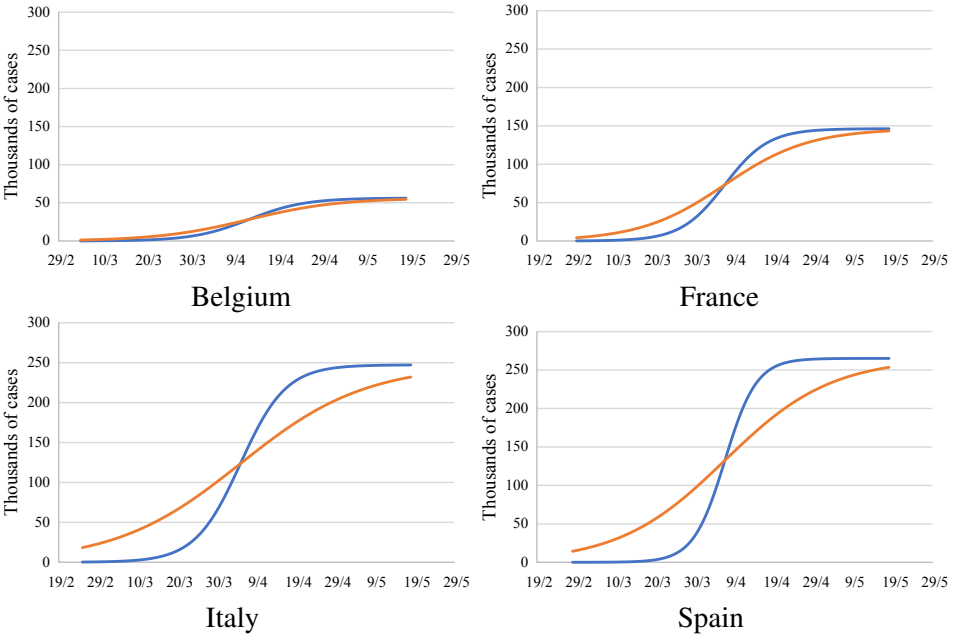


Figure 2. Graph of $E_1(t)$ (no lockdown, blue curve) and $E_2(t)$ (lockdown from the start, orange curve) for the four countries under consideration.

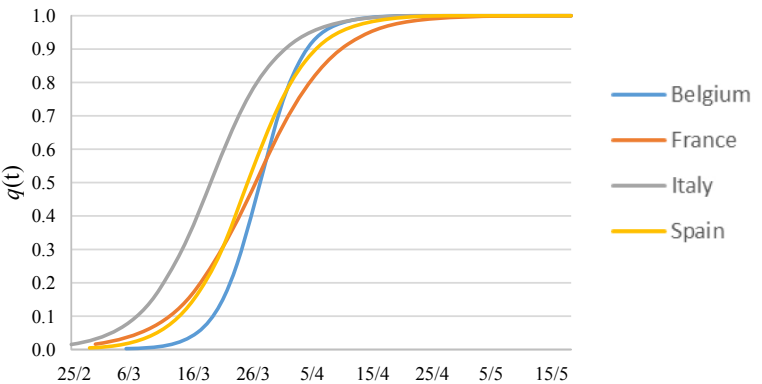


Figure 3. Comparison of the functions $q(t)$ for the four countries. The parameters were optimized by machine learning.

is based on data available up to 21 April (18 April once smoothed), we also draw the available data at the time of the publication, that is to say up to 3 May (30 April once smoothed). This allow the reader to assess the quality of the prediction.

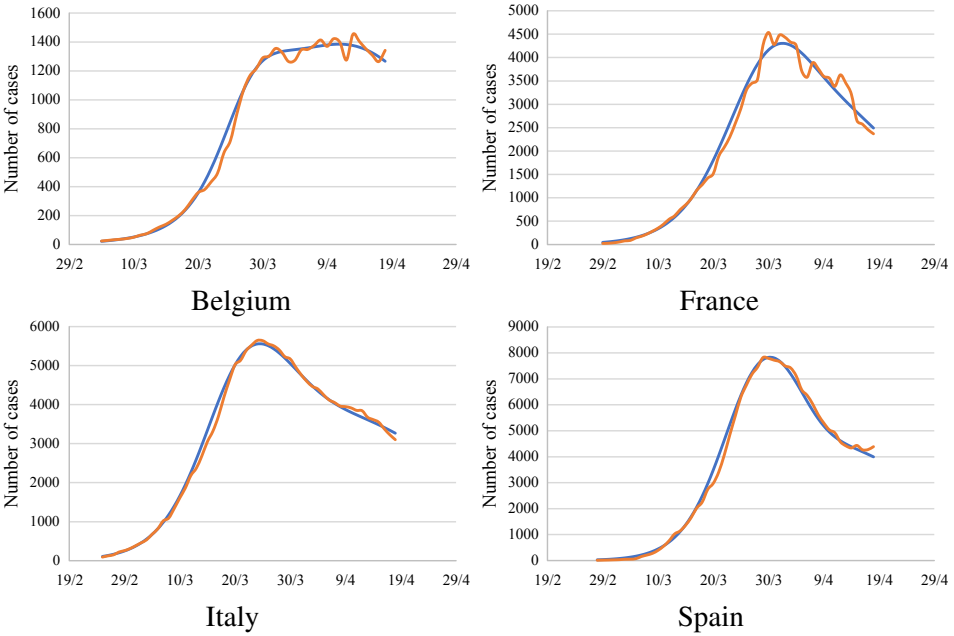


Figure 4. Comparison of our model's results with the smoothed data of the number of detected cases of COVID-19.

Final Value	t_i	α_c	L	k_2	t_S	k_1
Belgium	22.31	0.140	56271	0.098	38.23	0.158
France	26.53	0.076	146277	0.093	37.47	0.181
Italy	23.12	0.089	247302	0.063	39.95	0.174
Spain	26.27	0.096	265181	0.074	38.53	0.250
Error	t_i	α_c	L	k_2	t_S	k_1
Belgium	0.22	0.01	2089	0.00	0.77	0.00
France	1.26	0.01	11025	0.01	1.39	0.01
Italy	0.27	0.00	10356	0.00	1.22	0.01
Spain	0.78	0.00	17900	0.01	1.51	0.02

Table 3. Final values of the model parameters for the four countries, and analysis of the corresponding errors (standard deviation error) - t_i and t_S in days.

The reader will find below the corresponding value of the optimized parameters for each targeted country. Interestingly, those values are close for all countries, except for the total number of cases L, which obviously relies on country population.

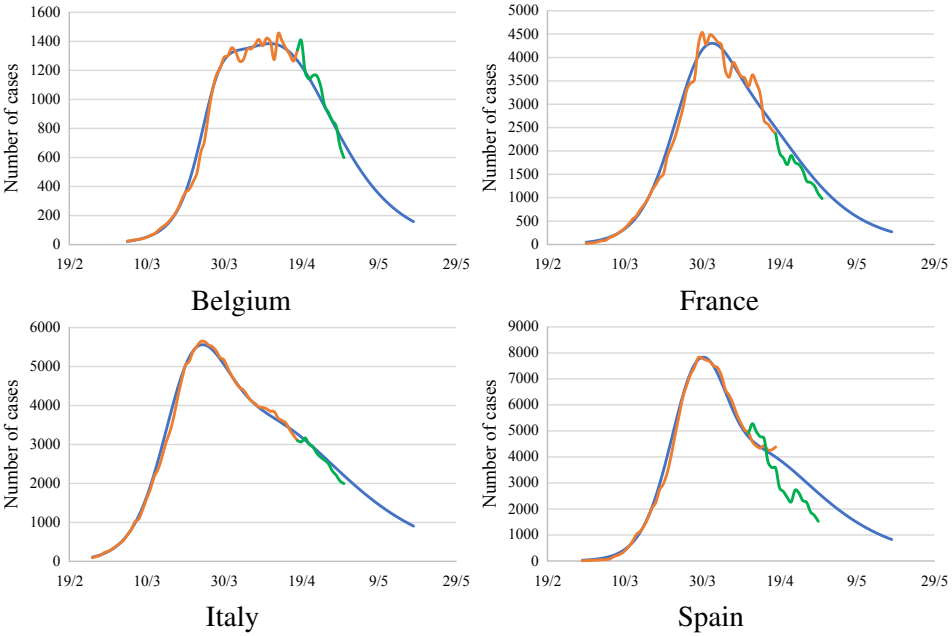


Figure 5. Comparison between the smoothed data of the number of detected cases of COVID-19 (orange: smoothed data available up to 18 April, green: up to 30 April) and our model (blue curve) based on the smoothed data of 18 April. (As discussed later, the data quality for Spain seems debatable since on the dataset up to 30 April, the raw data has been changed from that of 15 April with variation up to 160%; even stranger is that the number of cases on 19 April is negative (-1400 cases).

6. Discussion

We see that a simple macroscopic modeling of the Covid-19 epidemic in 2020, built with only two basic functions which permit to consider lockdown and its effects, allows to correctly model the evolution of the cases of people infected by the virus until the dates for which data are available. Simulation of other characteristics of the epidemic, such as the follow-up of hospitalizations or resuscitation, have also been carried out and will be published later [Rémond and Rémond]. They use the same model and the same optimization algorithm. For the forecasts, they are always to be taken with caution because on the one hand, unexpected events having biological, human causes or of management of the epidemic can occur which modify the form of it in a significant way, on the other hand the assumptions used may seem too summary to assign a level of probability to them. In

peculiarly, the unlockdown time can be considered as a new behavior for a given population and modeled with a new elementary function in addition to the two functions used. This unlockdown time is not considered in the simulations. The results are however interesting and show how a learning algorithm can allow a simple model to correspond well to the macroscopic effects of the epidemic.

It will also be noted that the inflection point of the t_S sigmoid occurs on average 38.65 days (from 37.47 to 39.95) after the start of the epidemic (Table 3) for the four countries. This means that as of this date, half of the people who will be affected by the epidemic have been infected. Analysis of the values of t_i (duration entered at the start of the epidemic and the date of lockdown + duration of appearance of the first effects of lockdown) shows that the duration of appearance of the effects of lockdown is very similar for the four countries, of the order of 10.30 days (between 9.53 for France and 11.27 for Spain). It should also be noted that the number of infected cases is half lower in France than in Italy and then in Spain.

The parameters α_c are associated to the velocity of changing of behavior after the lockdown. France, Italy, and Spain, with α_c between 0.076 and 0.096, have similar reactions for this evolution. By the way, to appreciate the step between the two behaviors, we must analyze the step between k_1 and k_2 for each country. The ratio k_1/k_2 represents the intensity of this change. In that case, Belgium appears to have the smallest change of behavior with a ratio of 1.62. France a ratio of 1.92 is better. Italy has a high ratio of 2.76 and Spain has strongly changed its behavior with a ratio of 3.36.

It is interesting to remember that the equation of the elementary basic functions $E_k(t)$ of the epidemic, given by the equation (2), or its initial form given in equation (1), are the solution of a elementary differential equation detailed in 2.1. This differential equation is solved for the boundary condition $y(0) = y_0$. The boundary condition y_0 characterizes the number of cases at time t_0. We show in the table 4 the values of y_0 given by the raw data and the smoothed data. They can be compared with the values of y_0 obtained by the modeling after the machine learning process. The differences of these values for raw and smoothed data are the intensity of the noise associated to these data. On the opposite, the differences between the values of y_0 for the data and the modeling are interesting. It shows that the number of cases at time t_0 were strongly underestimated by a factor 5 in Italy and in Belgium, and by a factor 10 for France and Spain. The more precise measurement of these boundary conditions at t_0 could have helped to better appreciate the intensity of this epidemic in a short time.

Note the extremely high value of y_0 for Belgium compared with the number of the inhabitants of this country, six times lower than the three other countries.

The coefficients R^2 of the model are given in the table below and by country, with smoothed and non-smoothed data. We see that the simulation is particularly

Country	y_0 **– Raw data** (Sum of new cases from $t = -3$ to $t = 0$)	y_0 **– Smoothed data** (Value of new cases at $t = 0$)	y_0 **– Given by the model**
Belgium	22	24	135
France	45	24	237
Italy	226	92	524
Spain	23	12	109

Table 4. Values of the number of cases y_0 at t_0 measured with the raw and smoothed data, comparing to the number of cases y_0 at t_0 given by the model.

Country	**Smoothed data**	**Raw data – 18th April**
Belgium	0.993	0.733
France	0.986	0.648
Italy	0.996	0.928
Spain	0.995	0.928

Table 5. R^2 coefficient of quality of simulations compared to raw data and smoothed data.

good for smoothed data. For raw data, the case of France is special given the positive data jumps that were recorded on certain days for administrative reasons. For Spain, despite of a good R^2 coefficient which show a good correlation between the data and the model, the official data given by this country after the identification of the model have changed strongly after 15 April. Then there is no peculiar significance on the analysis of this variation.

Therefore, this global modeling of the COVID-19 epidemic seems to be understandable as a sum of only two different elementary basis functions including the effects of the lockdown, and the development of such analysis will probably permit to analyze the specific behavior of population, in complement of the classical approaches by micro-macro analysis.

7. Conclusion

We have created a particularly simple virus-centric model of the Covid-19 epidemic, based on a decomposition in generic basic functions adaptable to all countries and to all the characteristic criteria of its development. Using a simple machine learning process, we show that only two basic elementary functions were sufficient to simulate the epidemic evolution for four European countries applying a lockdown, with accuracy. The results permit to quantify the difference of behavior

before and after the lockdown for these countries as well as the velocity of change and the intensity of change. We focused here on the model's ability to simulate the numbers of new cases reported in the Covid-19 epidemic over time. However, this model could be used for modeling hospitalizations, intensive care, and death. The prediction of a unlockdown effects should be also possible with this model, by adding a third elementary basic function describing the specific behavior of population after this event. The presented simulations are relevant and clearly show the effects of the various lockdowns carried out. The analysis of basic functions used in this decomposition could in any case allow us to have a macroscopic analysis of how the lockdowns were respected. Other characteristics and simulations of the 2020 SARS CoV 2 epidemic for other countries will be given in a sequel to this paper [Rémond and Rémond].

References

[Allaire 2001] G. Allaire, *Shape optimization by the homogenization method*, Springer, 2001.

[Bayette and Monticelli 2020] C. Bayette and M. Monticelli, "Modélisation d'une épidémie, 1", 2020, available at https://images.math.cnrs.fr/Modelisation-d-une-epidemie-partie-1.html.

[Caccavo 2020] D. Caccavo, "Chinese and Italian COVID-19 outbreaks can be correctly described by a modified SIRD model", 2020, available at https://tinyurl.com/caccavo-sird.

[Chinesta and Cueto 2014] F. Chinesta and E. Cueto, *PGD-based modeling of materials, structures and processes*, ESAform Series on Materials Forming, Springer, 2014.

[Colizza et al. December 2006] V. Colizza, A. Barrat, M. Barthelemy, and A. Vespignani, "The modeling of global epidemics: stochastic dynamics and predictability", *Bulletin of Mathematical Biology* **68** (December 2006), 1893–1921.

[Daley and Gani 1999] D. J. Daley and J. Gani, *Epidemic modelling*, Cambridge University Press, 1999.

[Eaton et al. 2019] J. W. Eaton et al., "The estimation and projection package age-sex model and r-hybrid model: new tools for estimating HIV incidence trends in sub-Saharian Africa", *AIDS* **33**:Suppl. 3 (2019), S235–S244.

[ECDC] ECDC, available at https://www.ecdc.europa.eu/en.

[Github] Github, available at https://github.com/uiuc-covid19-modeling/pydemic.

[Gonçalves 2020] B. Gonçalves, "Epidemic Modeling 102: all CoVID-19 models are wrong, but some are useful", 2020, available at https://tinyurl.com/goncalves-covid.

[Holb et al. 2005] I. J. Holb et al., "Analysis of summer epidemic progress of apple scab at different apple production systems in the Netherlands and Hungary", *J. of Phytopathology* **95**:9 (2005), 1001–1020.

[Jia et al. 2020] L. Jia, K. Li, Y. Jiang, X. Guo, and T. Zhao, "Prediction and analysis of coronavirus disease 2019", 2020, available at arxiv.org/abs/2003.05447v2.

[Levenberg 1944] K. Levenberg, "A method for the solution of certain problems in least squares", *Quart. Appl. Math* **2** (1944), 164–168.

[Mahiane et al. 2017] S. G. Mahiane et al., "Improvements in Spectrum's fit to program data tool", *AIDS* **31**:Suppl. 1 (2017), S23–S30.

[Marquardt 1963] D. W. Marquardt, "An algorithm for least-squares estimation of non linear parameters", *SIAM J. Appl. Math* **11** (1963), 431–441.

[Mesha and Hau 2008] Z. Mesha and B. Hau, "Effects of bean rust (*Uromyces appendiculatus*) epidemics on host dynamics of common bean", *Plant Pathology* **57** (2008), 674–686.

[Moral and Trapero 2009] J. Moral and A. Trapero, "Assessing the susceptibility of olive cultivars to anthracnose caused by *Colletottrichum acutatum*", *Plant Disease* **93**:10 (2009), 1028–1036.

[Oleinik et al. 1992] O. Oleinik, A. Shamaev, and G. Yosifian, *Mathematical problems in elasticity and homogenization*, Studies in Mathematics and its Application **26**, Elsevier, Amsterdam, 1992.

[Perra and Gonçalves 2015] N. Perra and B. Gonçalves, *Modeling and predicting human infectious diseases, social phenomena*, Springer, 2015.

[Rémond and Rémond] J. Rémond and Y. Rémond, "On a new virus-centric epidemic modeling, 2: Simulation of deaths from 2020 SARS CoV 2 in several countries", *Mathematics and Mechanics of Complex System*.

[Rémond et al. 2016] Y. Rémond, S. Ahzi, M. Baniassadi, and H. Garmestani, *Applied RVE reconstruction of heterogeneous materials*, Wiley-ISTE, 2016.

[Sanchez-Palencia 1980] E. Sanchez-Palencia, *Non-homogeneous media and vibration theory*, Lecture Notes in Physics **129**, Springer, 1980.

[Suquet 1987] P. M. Suquet, "Elements of homogenization for inelastic solid mechanics", pp. 193–278 in *Homogenization techniques for composite media*, Lecture Notes in Physics **105**, Springer, 1987.

[Verhulst 1838] P.-F. Verhulst, "Notice sur la loi que la population poursuit dans son accroissement", *Correspondance mathématique et physique* **10** (1838), 113–121.

[Wiemken and Kelley April 2020] T. L. Wiemken and R. R. Kelley, "Machine learning in epidemiology and health outcomes research", *Annual Review of Public Health* **41** (April 2020), 21–36.

Received 5 May 2020. Revised 12 May 2020. Accepted 16 Jun 2020.

JEAN RÉMOND: j.remond@stanwell.fr
Stanwell Consulting, Paris, France

YVES RÉMOND: remond@unistra.fr
ECPM - ICUBE Laboratory, University of Strasbourg / CNRS, Strasbourg, France

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
Vol. 8, No. 3, 2020
https://doi.org/10.2140/memocs.2020.8.249

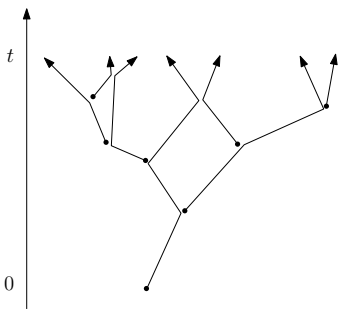
A KINETIC MODEL FOR EPIDEMIC SPREAD

MARIO PULVIRENTI AND SERGIO SIMONELLA

We present a Boltzmann equation for mixtures of three species of particles reducing to the Kermack–McKendrick (SIR) equations for the time evolution of the density of infected agents in an isolated population. The kinetic model is potentially more detailed and might provide information on space mixing of the agents.

1. Boltzmann–SIR equations

Consider a population of identical individuals (particles) moving in physical space and interacting upon contact. One (or several) of the individuals, say particle 1, has an infected status at time zero. As the dynamics runs, the infection can be transmitted, at the interaction times, to the individuals entering in contact with 1 or with the newly infected individuals. A cluster $\{i_1, i_2, \dots\}$ of infection grows in time, determined by the particle evolution: an individual is potentially infected at time $t > 0$ if it is involved, directly or indirectly, in the forward-in-time dynamics of 1. The "forward cluster of particle 1" (in the terminology of [Aoki et al. 2015; Pulvirenti and Simonella 2020b]) is represented symbolically in the picture below:



For concreteness, we may want to fix an idealized mechanical setting. Let us then proceed, as is customary in kinetic theory, by looking at N hard spheres of unit mass and diameter $\varepsilon > 0$. The balls move in $\Lambda \subset \mathbb{R}^d$, $d = 2, 3$, and interact

Communicated by Raffaele Esposito.
MSC2020: 00A71.
Keywords: Boltzmann equation, SIR model, low-density limit, stochastic particle system, forward cluster.

through elastic collisions. Each particle flies freely with constant velocity, and when two hard spheres collide with positions x, x_* at distance ε and incoming velocities v, v_*, the latter are instantaneously transformed to outgoing velocities v', v'_* by the relations

$$\begin{cases} v' = v - \omega[\omega \cdot (v - v_*)], \\ v'_* = v_* + \omega[\omega \cdot (v - v_*)], \end{cases} \tag{1-1}$$

where ω is the normalized relative distance $\omega = (x - x_*)/|x - x_*| = (x - x_*)/\varepsilon \in \mathbb{S}^{d-1}$.

We shall mimic the basic model in the mathematical theory of epidemics [Kermack and McKendrick 1927], by means of several assumptions. There are three different species of particles, S, I, and R, which stand for susceptible, infected, and recovered, respectively. Upon collision between a particle of type S and a particle of type I, the reaction

$$S + I \to I + I$$

occurs instantaneously with rate $\beta \in [0, 1]$. All the other collisions do not change the particle type, but in addition, a decay

$$I \to R$$

occurs with rate $\gamma \in [0, 1]$. Note that the population size is fixed (no deaths) and that the infection implies complete immunity. Finally for simplicity, we shall assume that β and γ are constants (they do not depend on time).

We are relying on the idea that the details of the interactions should not be of crucial importance (see [Stevens 2020] for a recent popular article simulating a similar system of particles). The main features are instead the following:

- The interactions are binary and localized.
- The number of interactions per unit time is expected to be finite.
- The qualitative behavior is independent of the number of particles N, provided that this is large in a suitable scaling limit.
- A statistical description is appropriate.

Under these assumptions, the Boltzmann equation for rarefied gases provides a tool of investigation.

Let us perform the so-called Boltzmann–Grad limit [Grad 1949] on the hard-sphere system under consideration. Denoting the one-particle distribution functions by

$$f_S = f_S(t, x, v),$$
$$f_I = f_I(t, x, v),$$
$$f_R = f_R(t, x, v)$$

for the three species of particles, we obtain the set of equations

$$\begin{cases} (\partial_t + v\cdot\nabla_x) f_S = Q(f_S, f_S) + Q(f_S, f_R) + (1-\beta) Q(f_S, f_I) - \beta Q_-(f_S, f_I), \\ (\partial_t + v\cdot\nabla_x) f_I = Q(f_I, f) + \beta Q_+(f_S, f_I) - \gamma f_I, \\ (\partial_t + v\cdot\nabla_x) f_R = Q(f_R, f) + \gamma f_I, \end{cases} \tag{1-2}$$

where

$$f = f_S + f_I + f_R$$

and Q is Boltzmann's operator (expressed in asymmetric form)

$$\begin{aligned} Q &= Q_+ - Q_-, \\ Q_+(f,g)(v) &:= \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B(\omega; v - v_*) f(v') g(v'_*)\, d\omega\, dv_*, \\ Q_-(f,g)(v) &:= f(v) \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B(\omega; v - v_*) g(v_*)\, d\omega\, dv_*. \end{aligned}$$

Note that the sum $f = f_S + f_I + f_R$ satisfies the classical Boltzmann equation

$$(\partial_t + v\cdot\nabla_x) f = Q(f, f).$$

Here we chose $B(\omega; v - v_*) = (\omega\cdot(v - v_*))\mathbb{1}(\omega\cdot(v - v_*) \geq 0)$, corresponding to the hard-sphere cross section. However, as said above, conclusions drawn from the kinetic model should not be very sensitive to the interaction rule; e.g., we shall consider as well different kernels $B(\omega; v - v_*) \geq 0$ such that

$$\int_{\mathbb{S}^{d-1}} B(\omega; v - v_*)\, d\omega = |v - v_*|^b,$$

for some $b \geq 0$.

In the second part of this exposition we will give more details on the passage from the particle dynamics to (1-2). Before that, we make a few elementary remarks on the equations themselves.

1.1. ***Maxwell collisions: Kermack–McKendrick equations.*** Averaging (1-2) over velocities, the Q operators vanish (because $\int Q_+ = \int Q_-$), and in the spatially homogeneous case (no dependence on x), the expected fractions of individuals of the species $A \in \{S, I, R\}$, $A(t) = \int f_A(t, v)\, dv$, satisfy the equations

$$\begin{cases} \dot S = -\beta \int |v - v_*|^b f_S(v) f_I(v_*)\, dv\, dv_*, \\ \dot I = \beta \int |v - v_*|^b f_S(v) f_I(v_*)\, dv\, dv_* - \gamma I, \\ \dot R = \gamma I. \end{cases}$$

These equations are not closed, except when dealing with "Maxwellian molecules" (case $b = 0$ [Bobylev 1988]), for which we get

$$\begin{cases} \dot{S} = -\beta I S, \\ \dot{I} = \beta S I - \gamma I, \\ \dot{R} = \gamma I, \end{cases} \tag{1-3}$$

namely the epidemiology model of [Kermack and McKendrick 1927] in the case of time-independent rates. This model has been analyzed and used extensively, and several generalizations have been conceived; see, e.g., [Anderson and May 1979; Brauer and Castillo-Chávez 2001; Murray 2002; Harko et al. 2014]. The kinetic equation (1-2) stands as an extension accounting for dependence on space and velocity of the individuals.

To remind the reader of the original motivations for such SIR models [Ross 1916; Kermack and McKendrick 1927], we recall that an epidemic is not necessarily terminated by the exhaustion of the susceptible individuals, nor by the extinction of the virulence. This is apparent from (1-3), over a threshold value of the density. Setting indeed $A_\infty := \lim_{t\to\infty} A(t)$, $A_0 = A(0)$, and $R(t) = R_0 + \gamma \int_0^t I(\tau)\, d\tau$ (showing that $I(t) \to 0$ as $t \to \infty$), one has that $\frac{dS}{dR} = -(\beta/\gamma)S$ and hence (by $R_\infty + S_\infty = 1$ and the assumption $R_0 = 0$) $S_\infty = S_0 e^{-(\beta/\gamma)(1-S_\infty)}$, or

$$e^{-(\beta/\gamma)S_\infty} \frac{\beta}{\gamma} S_\infty = S_0 \frac{\beta}{\gamma} e^{-\beta/\gamma}. \tag{1-4}$$

Since $\max y e^{-y} = 1/e$, given a value of β/γ one can find nonvanishing solutions for S_∞.

1.2. ***Confinement.*** The model can be easily adapted to investigate several different situations. Examples might be boundary conditions or external potentials, imposing internal spatial constraints or local enhancing of density. There has been recent intense interest in the effects of isolation of individuals, and of the reduction of social mixing, by means of physical distancing measures [Li et al. 2020; Prem et al. 2020]. At the level of (1-2), the energy can be used as a simple parameter regulating the interaction rate.

Here we give an example of one adaptation of (1-2), intended to model a confinement effect. Following [Stevens 2020] we assume that, for each species, there are two types of particles: wandering and confined. We denote by g_A, $A \in \{S, I, R\}$, the distribution of confined particles, while we maintain the notation f_A for the wandering particles. The distribution of the species A is $h_A := f_A + g_A$ and $f = \sum_A h_A$. Wandering particles have mass $m_w = 1$, while confined particles have mass $m_c = +\infty$ and zero velocity. The distribution g_A is proportional to a Dirac delta in velocity. Confined particles are frozen, and their total distribution is

stationary:

$$g_S(t,x)+g_I(t,x)+g_R(t,x)=\text{const.}\quad\text{for all } t.$$

The collision law becomes

$$\begin{cases} v' = v-(2m_*/(m+m_*))\omega[\omega\cdot(v-v_*)],\\ v'_* = v_*+(2m/(m+m_*))\omega[\omega\cdot(v-v_*)], \end{cases}$$

where m, m_* are the masses of the incoming particles, and (1-2) is replaced by

$$\begin{cases} (\partial_t+v\cdot\nabla_x)f_S = Q(f_S,h_S)+Q(f_S,h_R)+(1-\beta)Q(f_S,h_I)-\beta Q_-(f_S,h_I),\\ (\partial_t+v\cdot\nabla_x)f_I = Q(f_I,f)+\beta Q_+(f_S,h_I)-\gamma f_I,\\ (\partial_t+v\cdot\nabla_x)f_R = Q(f_R,f)+\gamma f_I,\\ \dot{g_S} = -\beta Q_-(g_S,f_I),\\ \dot{g_I} = \beta Q_+(g_S,f_I)-\gamma g_I,\\ \dot{g_R} = \gamma g_I. \end{cases} \tag{1-5}$$

In the spatially homogeneous case, integrating (1-5) in v, calling $A_w=\int f_A\,dv$ and $A_c=\int g_A\,dv$, $A=S,I,R$, we obtain

$$\begin{cases} \dot{S}_w = -\beta\int|v-v_*|^b f_S(v)h_I(v_*)\,dv\,dv_*,\\ \dot{I}_w = \beta\int|v-v_*|^b f_S(v)h_I(v_*)\,dv\,dv_*-\gamma I_w,\\ \dot{R}_w = \gamma I_w,\\ \dot{S}_c = -S_c\int|v_*|^b f_I(v_*)\,dv_*,\\ \dot{I}_c = S_c\int|v_*|^b f_I(v_*)\,dv_*-\gamma I_c,\\ \dot{R}_c = \gamma I_c. \end{cases}$$

Again, the above equations reduce to a standard SIR model in the case of Maxwellian molecules:

$$\begin{cases} \dot{S}_w = -\beta S_w(I_w+I_c),\\ \dot{I}_w = \beta S_w(I_w+I_c)-\gamma I_w,\\ \dot{R}_w = \gamma I_w,\\ \dot{S}_c = -S_c I_w,\\ \dot{I}_c = S_c I_w-\gamma I_c,\\ \dot{R}_c = \gamma I_c. \end{cases} \tag{1-6}$$

1.3. ***Related problems.*** The kinetic model presented above should be interpreted as a remark in the vein of mathematical physics: we do not pretend that it can be of use in epidemiology. It is more detailed than the classical SIR, insofar as it includes space and velocities of the agents. Presumably, its main potential interest in applications is the identification of spatial patterns having an impact on the history of epidemics. Moreover, a dynamical representation in terms of forward (or backward) clusters would provide information on the tracing of the infection. We comment next on a few other problems arising naturally.

The typical question concerning SIR equations is determining the long-time behavior in relation with the parameters β, γ and its dependence on local characteristics of the initial data. We are interested in masses but also in local densities in the presence of spatial inhomogeneities. From the mathematical side, little can be done, but the problem is suited to numerical investigation. In analogy to gas dynamics, it is natural to use stochastic methods, as we will discuss in the next section.

At the theoretical level, it would be interesting to detect large-scale limits and derive, starting from (1-2), equations for locally conserved quantities. Equation (1-2) can be useful in fact for limited amounts of time. Preliminarily, one should characterize the equilibria. Let $F_A = \lim_{t\to 0} f_A$ be the asymptotic distributions. Then we expect $F_I = 0$, and the other two distributions should satisfy

$$\begin{cases} Q(F_S, F_S) + Q(F_S, F_R) = 0, \\ Q(F_R, F_R) + Q(F_R, F_S) = 0. \end{cases}$$

The latter equation is satisfied if both F_S and F_R are Maxwellians

$$F_A = A_\infty \frac{e^{-(v-u)^2/(2\sigma^2)}}{(2\pi\sigma^2)^{d/2}}$$

for some constants S_∞ and R_∞, with σ and u determined by the initial conditions. A_∞ would be obtained as in (1-4). Notice that, when $f = f_S + f_I + f_R$ is a global equilibrium, a solution (f_S, f_I, f_R) of (1-2) for $b = 0$ is given by the same global equilibrium with densities $S(t), I(t), R(t)$ driven by (1-3).

2. Particle systems

2.1. ***Stochastic particle system.*** In this section we introduce a particle system yielding, in a suitable scaling limit, kinetic equations of type (1-2). The interest of this dynamics is twofold. First, it can be considered a microscopic model to be accepted through phenomenology, covering a large variety of kernels B. It would be somewhat funny to believe that the laws of Newton can be used to efficiently describe the interaction among individuals. On the other hand, we do not know so much concerning the details of such interactions; thus, a stochastic collision appears to be more robust than a deterministic one. Secondly, the particle scheme corresponds numerically to the direct simulation Monte Carlo method, widely used to approximate rarefied gas dynamics. There are several variants of such methods [Bird 1994; Rjasanow and Wagner 2005]. Below, we will deal with an inhomogeneous Kac model [1956] for three species with reactions.

We start by regularizing the collision operator (1-2). The strictly local interaction is smeared as

$$Q^h = Q^h_+ - Q^h_-,$$
$$Q^h_+(f,g)(x,v) := \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{S}^{d-1}} B(\omega; v-v_*)h(|x-y|)f(x,v')g(y,v'_*)\,d\omega\,dv_*\,dy,$$
$$Q^h_-(f,g)(x,v) := \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{S}^{d-1}} B(\omega; v-v_*)h(|x-y|)f(x,v)g(y,v_*)\,d\omega\,dv_*\,dy,$$

where $h : \mathbb{R}^+ \to \mathbb{R}^+$ is a smooth approximation of the delta function.

To simplify the notation, we limit ourselves to the case of (1-2) with $\beta = 1$, with the more general cases being a trivial extension. We therefore consider

$$\begin{cases} (\partial_t + v\cdot\nabla_x) f_S = Q^h(f_S, f_S) + Q^h(f_S, f_R) - Q^h_-(f_S, f_I), \\ (\partial_t + v\cdot\nabla_x) f_I = Q^h(f_I, f) + Q^h_+(f_S, f_I) - \gamma f_I, \\ (\partial_t + v\cdot\nabla_x) f_R = Q^h(f_R, f) + \gamma f_I. \end{cases} \tag{2-1}$$

We can pass to the limit $Q^h \to Q$ inside (2-1), whenever we have a smooth solution of the initial value problem.

We shall indicate by $\mathcal{A} = \mathcal{S}, \mathcal{I}, \mathcal{R} \subset \{1, 2, \dots, N\}$ the (random) disjoint sets of particles of types $A = S, I, R$, respectively. They form a partition of $\{1, 2, \dots, N\}$, so that the process $Z_N : \mathbb{R}^+ \to \mathcal{X}$, $Z_N = Z_N(t) = (z_1(t), \dots, z_N(t))$, $z_i = (x_i, v_i)$, takes values in

$$\mathcal{X} = \bigcup_{\mathcal{S},\mathcal{I},\mathcal{R}} \mathcal{X}(\mathcal{S},\mathcal{I},\mathcal{R}), \quad \mathcal{X}(\mathcal{S},\mathcal{I},\mathcal{R}) = \{(Z_{\mathcal{S}}, Z_{\mathcal{I}}, Z_{\mathcal{R}})\},$$

with

$$|\mathcal{S}| + |\mathcal{I}| + |\mathcal{R}| = N$$

and $z_i \in \Lambda \times \mathbb{R}^d$. Here $|\mathcal{A}|$ denotes the cardinality of the set $\mathcal{A}$. The configurations of particles in the three species are $Z_{\mathcal{S}} = (z_{s_1}, z_{s_2}, \dots)$, $Z_{\mathcal{I}} = (z_{i_1}, z_{i_2}, \dots)$, and $Z_{\mathcal{R}} = (z_{r_1}, z_{r_2}, \dots)$, respectively.

Le us define the time evolution. Particles move freely for a random time, exponentially distributed with intensity scaling like N. Then two particles are randomly chosen, say particles j and k, according to $\int B(\omega; v_j - v_k)h(|x_j - x_k|)\,d\omega$, and their velocities are updated as in (1-1) with $\omega \sim B(\cdot\,; v_j - v_k)$. If the pair of colliding particles is of type (A, A) or (S, R) or (I, R), the particles do not change their species. If the pair is of type (S, I), then the outgoing pair is of type (I, I). We abbreviate from now on $h_{j,k} = h(|x_j - x_k|)$, and we denote by J_{jk} the linear operator transforming the velocities j and k to a postcollisional pair with scattering vector ω. The generator of the process reads

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_i + \mathcal{L}_d$$

where $\mathscr{L}_0 = \sum v_i \cdot \nabla_{x_i}$ is the generator of the free motion,

$$\begin{aligned}\mathscr{L}_i\phi(Z_N) = \frac{1}{N}\sum_{j\in\mathscr{S}}\sum_{k\in\mathscr{I}}\int B(\omega; v_j - v_k)h_{j,k} \\ \times (J_{jk}\phi(Z_{\mathscr{S}\setminus\{j\}}, Z_{\mathscr{I}\cup\{j\}}, Z_{\mathscr{R}}) - \phi(Z_N))\,d\omega \\ + \frac{1}{N}\left(\sum_{j\in\mathscr{S}}\sum_{k\in\mathscr{R}} + \sum_{j\in\mathscr{I}}\sum_{k\in\mathscr{R}}\right)\int B(\omega; v_j - v_k)h_{j,k}(J_{jk}\phi(Z_N) - \phi(Z_N))\,d\omega \\ + \frac{1}{2N}\sum_{\mathscr{A}=\mathscr{S},\mathscr{I},\mathscr{R}}\sum_{\substack{j,k\in\mathscr{A}\\ j\neq k}}\int B(\omega; v_j - v_k)h_{j,k}(J_{jk}\phi(Z_N) - \phi(Z_N))\,d\omega, \qquad (2\text{-}2)\end{aligned}$$

and

$$\mathscr{L}_d\phi(Z_N) = \gamma\sum_{i\in\mathscr{I}}(\phi(Z_{\mathscr{S}}, Z_{\mathscr{I}\setminus\{i\}}, Z_{\mathscr{R}\cup\{i\}}) - \phi(Z_N)). \qquad (2\text{-}3)$$

We choose now test functions of the form

$$\phi_A(Z_N) = \frac{1}{N}\sum_{\ell\in\mathscr{A}}\varphi(z_\ell)$$

and focus, for instance, on the case $\mathscr{A} = \mathscr{S}$. We have that $\mathscr{L}_d\phi_S = 0$. Evaluating (2-2) in ϕ_S we notice that, given j and k, all the terms with $\ell \neq j, k$ cancel out. In the second line of (2-2) we find

$$\sum_{\substack{\ell\in\mathscr{S}\\ \ell\neq j}} J_{jk}\varphi(z_\ell) - \sum_{\ell\in\mathscr{S}}\varphi(z_\ell) = -\varphi(z_j).$$

Therefore,

$$\begin{aligned}\mathscr{L}_i\phi_S(Z_N) = -\frac{1}{N^2}\sum_{j\in\mathscr{S}}\sum_{k\in\mathscr{I}}\int B(\omega; v_j - v_k)h_{j,k}\varphi(z_j)\,d\omega \\ + \frac{1}{N^2}\sum_{j\in\mathscr{S}}\sum_{k\in\mathscr{R}}\int B(\omega; v_j - v_k)h_{j,k}(\varphi(x_j, v_j') - \varphi(z_j))\,d\omega \\ + \frac{1}{2N^2}\sum_{\substack{j,k\in\mathscr{S}\\ j\neq k}}\int B(\omega; v_j - v_k)h_{j,k}(\varphi(x_j, v_j') + \varphi(x_k, v_k') - \varphi(z_j) - \varphi(z_k))\,d\omega. \qquad (2\text{-}4)\end{aligned}$$

Next, we introduce a probability measure with density $W^N : \mathscr{X} \to \mathbb{R}^+$, assumed to be symmetric in the exchange of the particle labels within each one of the species. An example is provided by the fully factorized (chaotic) state, which we shall assume, to fix ideas, as the initial distribution of the particle process: $W^N(0) = f_0^{\otimes N}$ with $f^0 = \sum_A f_A^0$, $A = (S, I, R)$, where f_A^0 are the initial data for (2-1). We further

denote by $f_A^N = f_A^N(z)$ the one-particle marginals of W^N, defined as

$$\int f_A^N(z)\varphi(z)\,dz = \int W^N(Z_N)\phi_A(Z_N)\,dZ_N.$$

It is the probability density of finding a particle of type A in z. Similarly, $f^N_{A_1,A_2} = f^N_{A_1,A_2}(z_1, z_2)$ denotes the two-particle marginal, namely the probability density of finding two particles of type A_1 and A_2 in z_1 and z_2:

$$\int f^N_{A_1,A_2}(z_1, z_2)\varphi(z_1, z_2)\,dz_1\,dz_2 = \int W^N(Z_N)\phi_{A_1,A_2}(Z_N)\,dZ_N$$

for

$$\phi_{A_1,A_2}(Z_N) = \frac{1}{N(N-1)} \sum_{j\in\mathcal{A}_1} \sum_{\substack{k\in\mathcal{A}_2\\ k\neq j}} \varphi(z_j, z_k).$$

Even though the initial measure is factorized, the time-evolved density $W^N(t)$ is not, due to correlations generated by the dynamics. The factorization is however recovered in the limit $N \to \infty$ and

$$f^N_{A_1,A_2}(z_1, z_2) \approx f^N_{A_1}(z_1) f^N_{A_2}(z_2). \tag{2-5}$$

We are ready to compute

$$\frac{d}{dt}\int W^N(t)\phi_S = \int W^N(t)\mathcal{L}\phi_S.$$

Using (2-4), the definition of marginal, and (2-5), we deduce that, as $N \to \infty$,

$$\frac{d}{dt}\int f_S^N(t)\varphi \approx \int f_S^N(v\cdot\nabla_x\varphi) + \int Q^h(f_S^N, f_S^N)\varphi + \int Q^h(f_S^N, f_R^N)\varphi - \int Q^h_-(f_S^N, f_I^N)\varphi,$$

that is, the first equation of (2-1) in weak formulation.

The other two equations can be recovered similarly. For $A = I$, (2-3) yields

$$\mathcal{L}_d\phi_I(Z_N) = \frac{\gamma}{N}\sum_{i\in\mathcal{I}}\left(\sum_{\ell\in\mathcal{I}\setminus\{i\}}\varphi(z_\ell) - \sum_{\ell\in\mathcal{I}}\varphi(z_\ell)\right) = -\frac{\gamma}{N}\sum_{i\in\mathcal{I}}\varphi(z_i),$$

while in the second line of (2-2) we find

$$\sum_{\ell\in\mathcal{I}\cup\{j\}} J_{jk}\varphi(z_\ell) - \sum_{\ell\in\mathcal{I}}\varphi(z_\ell) = J_{j,k}\varphi(z_j) + (J_{j,k}\varphi(z_k) - \varphi(z_k))$$

so that

$$\frac{d}{dt}\int f_I^N(t) \approx \int f_I^N (v\cdot\nabla_x\varphi) + \int Q^h\Big(f_I^N, \sum_A f_A^N\Big)\varphi + \int Q_+^h(f_S^N, f_I^N)\varphi - \gamma\int f_I^N\varphi,$$

which is the second equation of (2-1).

2.2. ***Mechanical system.*** We briefly come back to the deterministic particle model, which was our starting point, that is, N hard spheres of diameter ε moving in physical space and colliding elastically, with reactions simulating infection and recovery. We call this system "mechanical" as the interaction is deterministic. Clearly there is still stochasticity in the reactions and, strictly speaking, we are dealing again with a stochastic process.

We can easily adapt to this case the formal arguments of the previous section. The process Z_N still takes values in $\mathscr{X}$, but in addition the strict exclusion $\min_{i\neq j}|x_i - x_j| > \varepsilon$ is imposed. In the generator (2-2), $1/N$ is replaced by ε^{d-1}, B is the hard-sphere kernel $(\omega\cdot(v_j - v_k))\mathbb{1}(\omega\cdot(v_j - v_k)\geq 0)$, $h_{j,k}$ is absent, and the operator $(J_{j,k} - 1)$ is replaced by $(\delta(x_k - x_j - \omega\varepsilon)J_{j,k} - \delta(x_k - x_j + \omega\varepsilon))$. Following [Pulvirenti and Simonella 2020a, §2.1] and assuming the chaos property (2-5), (1-2) is obtained in the limit $N\to\infty$, $\varepsilon\to 0$ with $\varepsilon^{d-1}N = 1$.

2.3. ***Rigorous results.*** We have formally derived the kinetic equations under proper scaling limits, presenting only the basic ideas. A rigorous approach is possible, based on existing literature. In the case of the stochastic system, one can apply martingale techniques as in [Wagner 1992], or the hierarchy of equations for the family of the marginals [Pulvirenti et al. 1994], or coupling techniques [Graham and Méléard 1997]. In the case of the mechanical model, one can resort to the validity techniques for the Boltzmann equation, leading to a short-time result; see [Lanford 1975] and subsequent works [Illner and Pulvirenti 1989; Spohn 1991; Cercignani et al. 1994; Gallagher et al. 2013; Pulvirenti et al. 2014; Pulvirenti and Simonella 2017; Denlinger 2018].

Acknowledgments

We are indebted to Grégory Miermont for enlightening comments inspiring this work. We thank Nicola Cotugno for advice on the bibliography and Robert Patterson for sharing numerical simulations and for useful comments on the manuscript.

References

[Anderson and May 1979] R. M. Anderson and R. M. May, "Population biology of infectious diseases, I", *Nature* **280** (1979), 361–367.

[Aoki et al. 2015] K. Aoki, M. Pulvirenti, S. Simonella, and T. Tsuji, "Backward clusters, hierarchy and wild sums for a hard sphere system in a low-density regime", *Math. Models Methods Appl. Sci.* **25**:5 (2015), 995–1010.

[Bird 1994] G. A. Bird, *Molecular gas dynamics and the direct simulation of gas flows*, Oxford Engineering Science Series **42**, Clarendon, Oxford, 1994.

[Bobylev 1988] A. V. Bobylev, "The theory of the nonlinear spatially uniform Boltzmann equation for Maxwell molecules", *Soviet Sci. Rev. C Math. Phys. Rev.* **7** (1988), 111–233.

[Brauer and Castillo-Chávez 2001] F. Brauer and C. Castillo-Chávez, *Mathematical models in population biology and epidemiology*, Texts in Applied Mathematics **40**, Springer, 2001.

[Cercignani et al. 1994] C. Cercignani, R. Illner, and M. Pulvirenti, *The mathematical theory of dilute gases*, Applied Mathematical Sciences **106**, Springer, 1994.

[Denlinger 2018] R. Denlinger, "The propagation of chaos for a rarefied gas of hard spheres in the whole space", *Arch. Ration. Mech. Anal.* **229**:2 (2018), 885–952.

[Gallagher et al. 2013] I. Gallagher, L. Saint-Raymond, and B. Texier, *From Newton to Boltzmann: hard spheres and short-range potentials*, Zurich Lectures in Advanced Mathematics **18**, European Mathematical Society, Zürich, 2013.

[Grad 1949] H. Grad, "On the kinetic theory of rarefied gases", *Comm. Pure Appl. Math.* **2** (1949), 331–407.

[Graham and Méléard 1997] C. Graham and S. Méléard, "Stochastic particle approximations for generalized Boltzmann models and convergence estimates", *Ann. Probab.* **25**:1 (1997), 115–132.

[Harko et al. 2014] T. Harko, F. S. N. Lobo, and M. K. Mak, "Exact analytical solutions of the susceptible-infected-recovered (SIR) epidemic model and of the SIR model with equal death and birth rates", *Appl. Math. Comput.* **236** (2014), 184–194.

[Illner and Pulvirenti 1989] R. Illner and M. Pulvirenti, "Global validity of the Boltzmann equation for two- and three-dimensional rare gas in vacuum: erratum and improved result", *Comm. Math. Phys.* **121**:1 (1989), 143–146.

[Kac 1956] M. Kac, "Foundations of kinetic theory", pp. 171–197 in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (Berkeley, CA, 1954–1955), vol. III: Contributions to astronomy and physics, edited by J. Neyman, University of California, Berkeley, CA, 1956.

[Kermack and McKendrick 1927] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics", *Proc. Roy. Soc. Lond. A* **115**:772 (1927), 700–721.

[Lanford 1975] O. E. Lanford, III, "Time evolution of large classical systems", pp. 1–111 in *Dynamical systems, theory and applications* (Seattle, 1974), edited by J. Moser, Lecture Notes in Physcs **38**, Springer, 1975.

[Li et al. 2020] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. M. Leung, E. H. Y. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Y. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, and Z. Feng, "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia", *New Engl. J. Med.* **382**:13 (2020), 1199–1207.

[Murray 2002] J. D. Murray, *Mathematical biology*, vol. I: An introduction, 3rd ed., Interdisciplinary Applied Mathematics **17**, Springer, 2002.

[Prem et al. 2020] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, M. Jit, and P.

Klepac, "The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study", *Lancet Publ. Health* **5**:5 (2020), E261–E270.

[Pulvirenti and Simonella 2017] M. Pulvirenti and S. Simonella, "The Boltzmann–Grad limit of a hard sphere system: analysis of the correlation error", *Invent. Math.* **207**:3 (2017), 1135–1237.

[Pulvirenti and Simonella 2020a] M. Pulvirenti and S. Simonella, "A brief introduction to the scaling limits and effective equations in kinetic theory", preprint, 2020. arXiv

[Pulvirenti and Simonella 2020b] M. Pulvirenti and S. Simonella, "On the cardinality of collisional clusters for hard spheres at low density", preprint, 2020. arXiv

[Pulvirenti et al. 1994] M. Pulvirenti, W. Wagner, and M. B. Zavelani Rossi, "Convergence of particle schemes for the Boltzmann equation", *European J. Mech. B Fluids* **13**:3 (1994), 339–351.

[Pulvirenti et al. 2014] M. Pulvirenti, C. Saffirio, and S. Simonella, "On the validity of the Boltzmann equation for short range potentials", *Rev. Math. Phys.* **26**:2 (2014), art. id. 1450001.

[Rjasanow and Wagner 2005] S. Rjasanow and W. Wagner, *Stochastic numerics for the Boltzmann equation*, Springer Series in Computational Mathematics **37**, Springer, 2005.

[Ross 1916] R. Ross, "An application of the theory of probabilities to the study of a priori pathometry, I", *Proc. Roy. Soc. Lond. A* **92**:638 (1916), 204–230.

[Spohn 1991] H. Spohn, *Large scale dynamics of interacting particles*, Springer, 1991.

[Stevens 2020] H. Stevens, "Why outbreaks like coronavirus spread exponentially, and how to 'flatten the curve'", *The Washington Post* (March 14, 2020).

[Wagner 1992] W. Wagner, "A convergence proof for Bird's direct simulation Monte Carlo method for the Boltzmann equation", *J. Statist. Phys.* **66**:3–4 (1992), 1011–1044.

Received 16 May 2020. Accepted 26 Jun 2020.

MARIO PULVIRENTI: pulvirenti@mat.uniroma1.it
Dipartimento di Matematica, Università di Roma La Sapienza, Rome, Italy

SERGIO SIMONELLA: sergio.simonella@ens-lyon.fr
Unité de Mathématiques Pures et Appliquées, CNRS UMR 5669, École Normale Supérieure de Lyon, Lyon, France

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the submission page.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in MEMOCS are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and a Mathematics Subject Classification or a Physics and Astronomy Classification Scheme code for the article, and, for each author, postal address, affiliation (if appropriate), and email address if available. A home-page URL is optional.

Format. Authors are encouraged to use LaTeX and the standard amsart class, but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should normally be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages — Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc. — allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with as many details as you can about how your graphics were generated.

Bundle your figure files into a single archive (using zip, tar, rar or other format of your choice) and upload on the link you been provided at acceptance time. Each figure should be captioned and numbered so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript with all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables.

White Space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Mathematics and Mechanics of Complex Systems
vol. 8 no. 3 2020

MEMOCS is a journal of the International Research Center for the Mathematics and Mechanics of Complex Systems at the Università dell'Aquila, Italy.