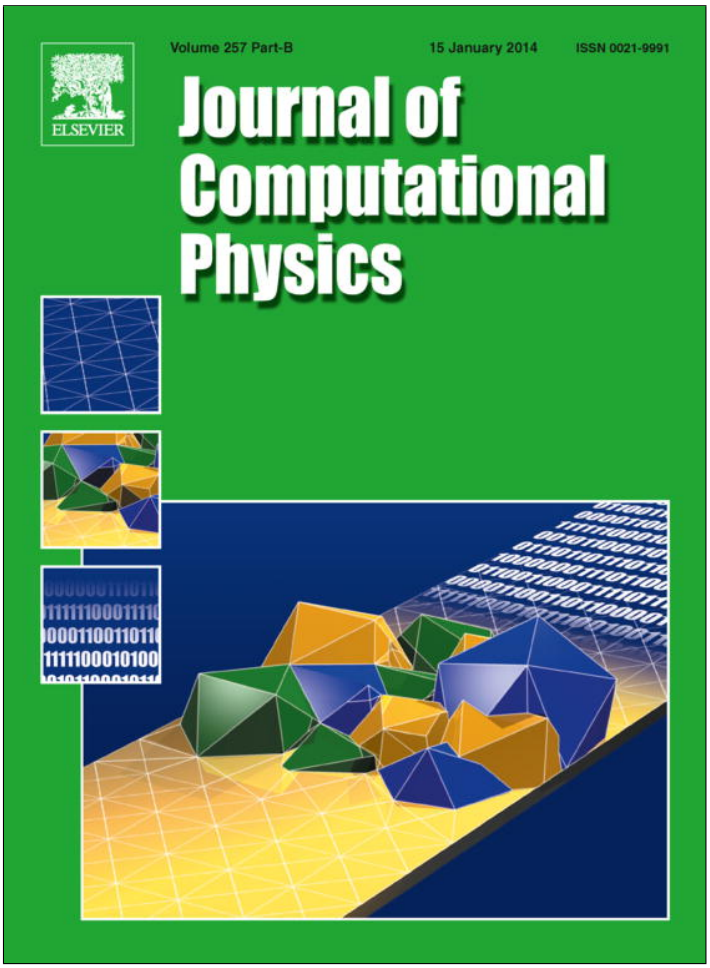


Journal of Computational Physics 257 (2014) 1260–1290

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Why starting from differential equations for computational physics?

Enzo Tonti

University of Trieste, Italy

ARTICLE INFO

Article history:
Received 14 July 2012
Received in revised form 26 June 2013
Accepted 4 August 2013
Available online 29 August 2013

Keywords:
Discrete forms
Algebraic formulation
Cell Method
Finite formulation

ABSTRACT

The computational methods currently used in physics are based on the discretization of differential equations. This is because the computer can only perform algebraic operations. The purpose of this paper is to critically review this practice, showing how to obtain a purely algebraic formulation of physical laws starting directly from experimental measurements. In other words, we can get an algebraic formulation avoiding any arbitrary process of discretization of differential equations. This formulation has the great merit of maintaining close contact between the mathematical description and the physical phenomenon described.

1. Introduction

Computational physics is richer than computational mathematics.

The reason is simple: computational mathematics is concerned with numbers whereas computational physics is concerned with physical quantities. It is true that each of them is represented by one or more numbers, but they also have a physical meaning that mathematics deliberately ignores. A physical variable can describe the source of a field or of a phenomenon, its configuration or its energy. Moreover physical variables are associated with space elements, i.e. with points, lines, surfaces and volumes. In such association the notion of orientations of a space element plays an essential role. There are two types of orientations of a space element, the inner and the outer one. There is also an association with time elements, i.e. time instants and time intervals. The two types of orientation apply to them as well.

On account of their physical meanings, physical variables have a great information content that can and *must* be taken into account doing computational physics. This forbids some choices in the discretization process and, at the same time, suggests other choices. To give an example, the introduction of a pair of dual meshes into computational physics is dictated by physical reasons (with which we will deal later) and also physical reasons oblige to associate each physical variable with a well defined space element, i.e. vertex, or edge, or face or volume. Hence both choices must not be made by trials and errors simply looking for a more stable scheme or a more accurate result.

The computational methods currently used in physics are based on the discretization of the differential equations. Since the physical laws are formulated from experimental measurements, they use “global” variable and this leads directly to a discrete formulation. For historical reasons we are accustomed to introduce the densities of these global variables and to perform the limit process thus obtaining the traditional field functions. Doing so we come to differential equations. Since, in general, we are not able to find an *analytical* solution, we search at first an *approximate* solution. When this is not easily accessible, generally due to the shape of the boundary and the type of boundary conditions, we resort to the *numerical* solution. With the advent of digital computers, numerical methods have proven to be extremely useful.

E-mail address: tonti@units.it.

0021-9991/$ – see front matter
http://dx.doi.org/10.1016/j.jcp.2013.08.016

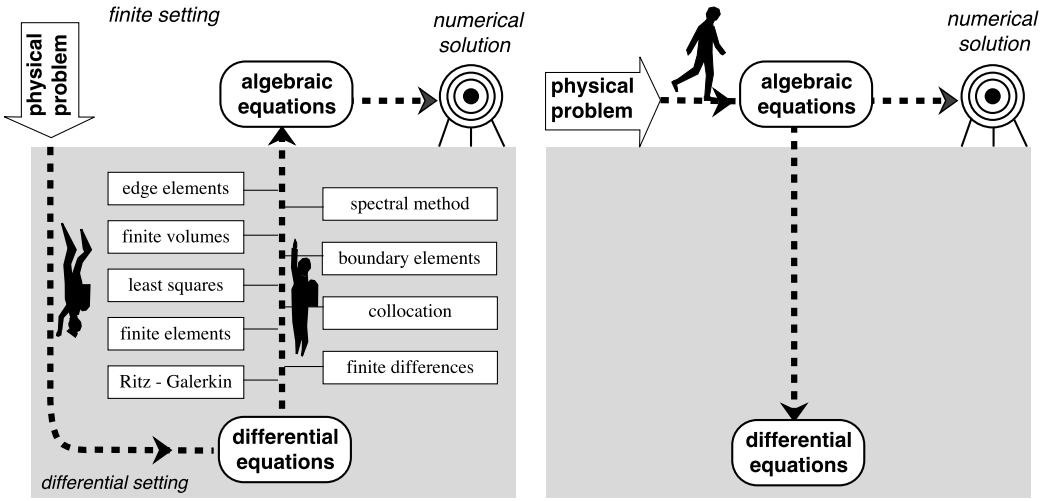


Fig. 1. (left) The tortuous path to obtain a numerical solution to a physical problem; (right) the direct procedure.

Since computers do not have the divine spark of intelligence, peculiar to human beings, they cannot conceive the notion of limit and are therefore confined to use algebra, not the infinitesimal calculus. Hence, to apply the numerical methods we are forced to return to a discrete formulation using one of the many methods of discretization, such as FEM, BEM, FVM, FDM, etc.

If we explained this process to a person who is not familiar with physics, this person might ask: Why do we switch from a discrete formulation derived from experimental measurements to a differential formulation and then go back to a discrete formulation? Why do we follow this convoluted process? Why don't we keep the discrete formulation from the outset? (See Fig. 1.)

A direct discrete approach to physical laws is possible and easy to obtain: but we must change our habits of considering the differential formulation as indispensable or simply privileged.

2. The mathematical description of physics

The peculiarity of the physical science, with respect to other sciences of nature, is that physical systems have many quantitative attributes. These quantitative attributes are responsible for the introduction of physical quantities which allows a quantitative description of the physical phenomena. It follows that the behavior of a phenomenon, i.e. its law, is described by the relationship between the pertinent physical quantities, i.e. by an equation.

With physical quantities and equations, mathematics makes its majestic entrance into physics.

But we must never forget that each measurement process can be performed with a limited accuracy, and this compels us to introduce the concept of *tolerance* of a measurement. Every measuring instrument belongs to a given class of precision. In measurement an "infinite" precision, in the sense of a limit process of mathematics, is meaningless. The same positioning of the measuring probe in a space region implies a tolerance. No measurement of a physical quantity is "exact". The notion of *exactness* is alien to the mind and to the practice of engineers and physicists!

2.1. Exact and approximate solutions

In our culture, trained by three centuries of differential formulation of physical laws, the differential formulation is so familiar that we are led to think that it is the "natural" formulation for physics. Moreover we are convinced that only the solution of a differential equation is the "exact" solution of a physical problem.

However, we know well that only in a few elementary cases, with regions of simple shape, with the most simple boundary conditions and with particular distributions of sources in the region we can obtain a solution in a closed form, i.e. an *analytical* solution. Hence the "exact" solution, that is promised by the differential formulation, is hardly ever attained in practice. By contrast, the great technological progress of our days is made possible by the fact of being able to have an *approximate* solution, in particular a *numerical* solution to the problems posed by the technique. For our culture, formed on the model of infinitesimal analysis, the term "approximate" sounds like "imperfect." However, we must not forget that *the objective of a numerical simulation of physical processes is the agreement with the experimental measurements* and not the convergence to an analytical solution, usually not attainable. In addition, the request of reducing the error of an approximate solution does not mean make the error "arbitrarily small", as required by the process of limit, *but making the error smaller than a preassigned tolerance*.

The notion of precision of a measuring device plays the same role as the notion of tolerance in manufacturing and the notion of error in numerical analysis. In conclusion we cannot deny ourselves the satisfaction of knowing the exact solution of a physical problem in the rare cases in which it is available, what we question is the *need* to refer to an idealized exact solution when this is not available in order to compare a numerical result with experiments.

2.2. Physics and the notion of limit

By reasons of principle, the numerical value of a measurement is a *rational* number which can be expressed, or approximated, by a decimal number with a certain number of significant digits.

One of the main achievements in the history of mathematics is the notion of limit. When this notion is applied to the field of numbers it gives rise to the notion of *real* number. When it is applied to geometrical problems, like the problems of finding the tangent to a curve or the area of a plane region, it gives rise to the *differential calculus*. As it is obvious, the notion of limit belongs to the realm of pure thought and no physical measurement can give rise to a real number.

In our daily life we are accustomed to use the notion of *mean* mass density as the ratio mass/volume. This notion of mean density does not comply with the requisites of the differential formulation which must operate on pointwise variables and, for this reason, we are obliged to define the density as the *limit* of the ratio mass/volume when the volume tends to zero.

The important thing is to remember that with the limit process we perform a purely mental process which takes us away from the physical measurements. In contrast, if we remain in the realm of algebra we are fully adherent to physical measurements.

In order to apply the formalism of differential calculus, such as partial derivatives and the integration over lines, surface and volumes, we need to use the field variables. But the main variables we measure are domain variables, not field variables! So we measure displacement (ruler), angle (goniometer), elongation (strain gauge), force (dynamometer), heat (calorimeter), momentum (ballistic pendulum), mass (balance), mass flow (by changes in content), charge (electrometer), voltage (voltmeter), magnetic flux (fluxmeter), particle content (by simply counting), light intensity (light meter): all of them are *domain variables*.

The domain variables that refer to a time instant, change with time and we can perform on them the time derivative. But since they are not point variables, we cannot perform on them the space derivatives, hence we cannot apply to them the traditional differential operators grad, curl and div.

This poses the following question: where do the field variables originate? The answer is that field variables are mostly obtained forming the densities of domain variables, i.e. line, surface or volume densities, and later performing the limit by contracting the corresponding domain to a point. In this way we come to mass density, charge density, mass current density, electric field strength, linear strain, stress vector, pressure, volume force density, curvature, magnetic flux density, particle current density, etc.

What do we gain by introducing the notion of limit in the mathematical description of physics? The gain lies in easiness of language and of mathematical treatment. We are aware that to understand we must simplify but also that every simplification leads us away from reality. In fact we introduce the ideal categories of *rigid body*, *perfect gas*, *ideal fluid*, *frictionless constraint*, *elastic body*, *adiabatic wall*, *reversible transformation* etc., because these ideal notions make it easier to conceptualize and, consequently, to describe phenomena.

Let us consider the notion of speed. In our everyday experience the speed is simply the ratio of a displacement to its duration: the limit is not involved. The velocity thereby obtained makes reference to a time interval not to an instant (in the mathematical language it is a *mean* velocity). If we want to use the differential formulation we are obliged to reduce velocity to a function of the instant and this requires a limit process. It is much easier to say that the speed is the *derivative* of the function $s(t)$ which gives the arch length as a function of time. The gain lies in the concise statement *the speed is the derivative of the function* $s(t)$ and ease of evaluation *only* when we have at our disposal the analytic form of the function $s(t)$. But if we do not have it, as it happens in physical measurements and (notice the coincidence) also in computational physics, the derivative cannot be performed and we must return to the ratio displacement/duration.

Physics is based on measurements; measurements are necessarily discrete, hence the limit process cannot be applied. The limit process is required only if we insist on applying the differential formulation to physics. To do this we must smear out our discrete measurements by means of interpolation via mathematical functions.

There is no doubt that the infinitesimal analysis has played in the past and will play in future a major role in the mathematical treatment of physics, but we must be aware that its introduction hides some important features of the phenomenon being described, such as the geometrical and topological features. Moreover the limit process introduces limitations by requiring regularity conditions on the field variables. These regularity conditions, in particular the conditions of differentiability, are the price we have to pay to use a formalism which is very advanced and easy to manipulate.

2.3. The current predominance of the differential culture

If you ask a university student what is an irrotational vector field, the answer is, in general, that the *curl* vanishes. If we ask what is a solenoidal vector field, the answer is that it is the one for which the *divergence* vanishes. Moreover, in general, they know the Cartesian form for these two notions and not their intrinsic geometrical meaning.

A proper definition is that an irrotational vector field is one for which the line integral of the vector along any *reducible* closed line vanishes. In the same way a solenoidal vector field is one for which the surface integral of the vector through any *reducible* closed surface vanishes. The corresponding differential expressions $\nabla \times \mathbf{v} = 0$ and $\nabla \cdot \mathbf{v} = 0$ follow from these definitions under regularity conditions.

The differential formulation cannot directly deal with these global definitions. Hence, for the solenoidal vector field we are obliged to consider the volume density of the flux through the closed surface and to perform the limit when the volume contracts to a point. In this fashion we obtain a field function, the divergence of the vector field. This process of forming the pointwise density is not a requirement of physics, but it is imposed by the differential formulation.

3. Tools for a direct discrete formulation

If we want to avoid the differential formulation as a starting point for computational physics, we must introduce some appropriate notions and tools. We quote here some useful notions

1. the notion of global variable as opposed to that of field variable;
2. the distinction between source, configuration and energy variables;
3. the fundamental problem of a physical theory;
4. the notion of space element endowed with inner or outer orientation;
5. the association of global variables with the oriented space elements;
6. the need of a cell complex;
7. the need of the dual cell complex;
8. the notions of faces and cofaces;
9. the notion of incidence numbers;
10. the notion of affine scalar and vector fields.

In what follows we will consider each one of these notions.

3.1. Global variables and field variables

Let us ask ourselves the question: *is the differential formulation the most natural form for the mathematical description of physics?*

Few physical variables arise directly as functions of points and instants but the majority of them are obtained from variables which are associated with extended space elements and to time intervals by calculating the corresponding densities and rates.

Let us look at some examples. The mass of a body makes reference to a body (in the material description) or to a volume (in the spatial description); the internal surface force in continuum mechanics makes reference to a surface; the voltage makes reference to a line. Internal energy, entropy, potential energy and kinetic energy are referred to a system (in the material description) or to a volume (in the spatial description), so they are not point functions, whereas they are functions of a time instant; the number of moles of a substance refers to a body (in the material description) or to a volume (in the spatial description).

The displacement of a particle, the impulse of a force, work and heat are referred to a time interval. The amount of radiation emitted by a specimen of radioactive material refers to the volume of the specimen and to the time interval considered.

The differential formulation uses the *mass density* and not directly mass; uses the *pressure* and not directly the force normal to a plane surface; uses the *strain* and not the extension; uses the *concentration* and not the mole number; uses the *heat current density* and not heat; etc.

These considerations invite us to give a proper status and so a proper name to all physical variables which are not densities.

Definition. We will call *global variable in space* a variable which is neither a line density nor a surface density nor a volume density of another variable.

Since the majority of the measured physical variables are global, in order to obtain the corresponding field functions we are obliged to evaluate their densities and to go to the limit in order to obtain the pointwise densities. Hence, starting from a global variable, we create a field variable: the global variable, when required, is regained with a process of integration.

The integral variables, i.e. those variables which are obtained by line, surface or volume integration of field variables, are also global variables. We may think that the term “global variable” is synonymous of “integral variable”. This is not true: certainly the class of the global variables contains that of integral variables. For example the position vector and the temperature do not arise from a space integration of some field variable, hence they are not integral variables. But since they are not densities of other variables they are global variables in space. This is also the case of the electric potential, the velocity potential in fluid dynamics, the gravitational potential, the total displacement, etc. Fig. 2 shows the relationship between these three classes of variables.

It is important to remark that *the formation of the density of global variables hides the geometric and physical content of the global variables* because we lose information about the association of the variable with space elements. Let us think about the three variables temperature, pressure and mass density which in thermodynamics are all considered intensive variables.

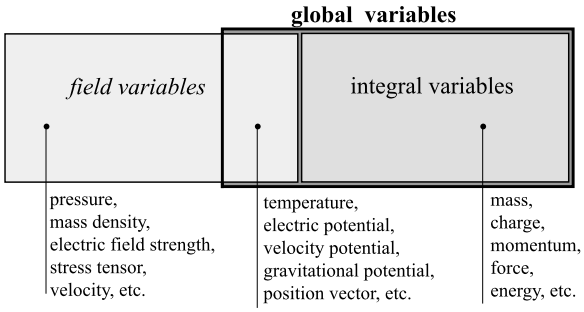


Fig. 2. The relation between global, integral and field variables.

While the mass density arises from the ratio mass/volume, the pressure is force/area and the temperature is not a density at all. Hence the mass is *associated* with a volume; the pressure is *associated* with a surface whereas the temperature is *associated* with a point. These three variables, even if are united by the fact of being intensive, have a different genealogy. This proves that the formation of densities hides the origin of the physical quantities. At first glance this seems unimportant, but it is not. In fact, *it is exactly the link between global physical variables and space and time elements what will allow us to obtain a rationale for computational physics*.

Some readers may think that we are suggesting the abandonment of the field functions and the differential formulation of physics: far from it! We are simply stating that the global physical variables contain some geometrical information which field variables do not contain and we will see that these geometrical information allows a proper assignment of physical variables to the corresponding space elements. This association plays a fundamental role in computational physics, even if this role, up to now, has not been dealt with in literature. Consequently, we suggest the development of the algebraic formulation of physics *before* the differential formulation and not vice versa.

3.2. Source, configuration and energy variables

At the beginning of the mathematical description of a physical theory it is wise to make an important distinction between the causes of a phenomenon, such as the source of a field and the effects on a system, described by the configuration of the system or by the potentials of the field. This distinction allows us to separate the variables into three classes: the source variables, the configuration variables and the energy variables.

Source variables. So the forces are the sources of motion and of deformation; the electric charges are the sources of the electric field; the electric currents are the sources of the magnetic field; heat is the source of the thermal field; the gravitational masses are the sources of the gravitational field. Each of these variables gives rise to other variables of the same type performing derivatives or integrals or obtaining a density, without the introduction of physical constants.

Configuration variables. In mechanics the Lagrangian coordinates of a system describe the configuration of the system; in mechanics of deformable solids the total displacement describes the configuration of the solid; in fluid dynamics the distribution of velocities describes the configuration of the fluid; in an electric field the electric potential describes the configuration of the field; in the gravitational field the gravitational potential describes the configuration of the field, etc. Each of these variables gives rise to other variables of the same type by integration, derivation, the appropriate combinations of partial derivatives, the formation of densities, etc. without the intervention of physical constants. The same happens for the strain tensor, for the electric field strength, for the velocity of deformation, for the gravitational acceleration, etc.

Energy variables. The original variable of this type is the work: force $\times$ displacement, pressure $\times$ volume. Others are the kinetic energy (momentum $\times$ velocity); the potential energy (gravitational, electric, magnetic, of deformation) and all types of energy (internal energy, Helmholtz free energy, Gibbs energy, enthalpy) and the action.

This distinction between these three types of variables is not commonly taken into account. However, this distinction plays a significant role in the conceptual framework of physics, in particular in computational physics, because there is a strict relationship between the two types of variables and the two types of orientation of the associated space element.

3.3. The fundamental problem of a physical theory

The fundamental problem of a theory is to find the configuration of a system or of a field once it is assigned the spatial and temporal distribution of the sources with their intensity. More in detail:

- *Given the space and time distributions of the sources acting on a system or in a space region*:
- *given the shape and the dimensions of the system or of the region*;

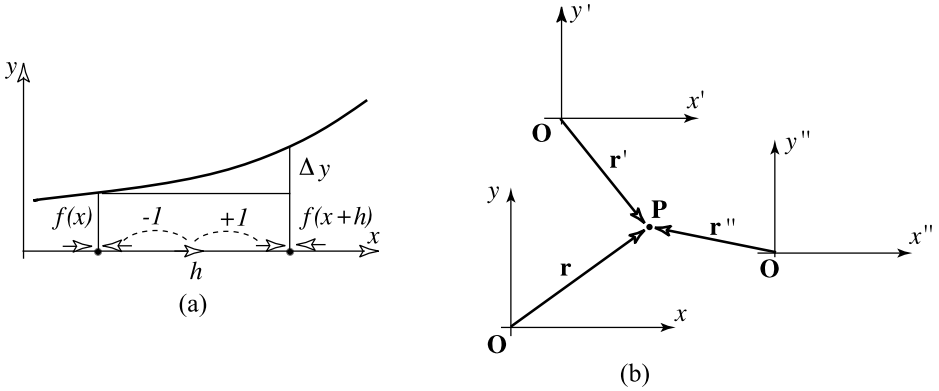


Fig. 3. (a) The definition of the increment of a function implies that points have been oriented as sinks; (b) the radius vector is always oriented from the origin of a reference system towards the point and hence it is oriented as a sink.

- *given the nature of the materials which compose the system or which fill the region*;
- *given the boundary conditions, which summarize the action of the external sources on the system or on the region*;
- *given the initial conditions* (*for time variable fields*);
- *find the configuration of the system or of the field at every instant.*

3.4. Space elements and the two types of orientation

With the locution *space elements* we mean points, lines, surfaces and volumes: these will be denoted with their initial capital letter, i.e. **P** for points, **L** for lines, **S** for surfaces, **V** for volumes.

We have chosen capital letters in order to render the notation consistent with the common practice of denoting points with capital letters, such as A, B, C, P, Q and we have chosen boldface type because we usually denote a point function by $f(t, \mathbf{x})$ or $f(t, \mathbf{P})$. Another reason for using boldface type lies in the fact that we find it natural to use the corresponding letters L, S, V, in plain text, to denote the extension of the corresponding space elements, i.e. *length, area* and *volume*.

What is meant by "inner" and "outer" orientation of a space element? Let us consider a surface: if we draw a curvilinear triangle on it and consider an order of its vertices we have fixed an inner orientation of the triangle. The same orientation can be propagated to all adjacent triangles on the surface. Excluding the case of two-sided surfaces, like Möbius band and Klein bottle, which are amazing objects but that have nothing to do with physical applications, we say that the whole surface is endowed with inner orientation. The term "inner" refers to the fact that this orientation requires to consider only points of the surface. In contrast, an "outer" orientation requires us to consider both sides of the surface, i.e. to pass from one side to the other crossing the surface and this implies that we must go out from the surface. This justifies the term "outer" for this orientation.

The notion of inner and outer orientation can be applied also to points, even if some books of topology do not consider this possible! (See [10, p. 53].) In particular, we will show that *points are implicitly oriented as sinks*, as shown in Fig. 3. This is never explicitly stated, but it can be inferred from the fact that the usual definition of the increment $\Delta f(x) \stackrel{\text{def}}{=} f(x+h) - f(x)$ can be reinterpreted as the following sum

$$\Delta f = (+1) f(x+h) + (-1) f(x). \tag{1}$$

In fact, if we consider that all intervals of the time axis inherited the orientation of the x-axis, we see that the plus sign in front of the term $f(x+h)$ and the minus sign in front of the term $f(x)$ can be interpreted as *incidence numbers* between the oriented intervals and the two oriented bounding points, as shown in Fig. 3(a). Hence *the same definition of increment of a function deals implicitly with the notion of orientation of points.*

Another reason for considering points oriented as sinks is the choice of the radius vector **r** in geometry and physics which starts at the origin of a coordinate system and ends at the point, not vice versa, as shown in Fig. 3(b). A point endowed with inner orientation will be denoted by placing a bar over the symbol, i.e. $\bar{\mathbf{P}}$.

There are physical variables referred to points with inner orientation, such as the *electric potential* and the *velocity potential*, whereas others are referred to points endowed with outer orientation, such as the *scalar magnetic potential* and the *stream function* in fluid dynamics. Fig. 4 shows that the two types of orientation of a point can be useful in fluid dynamics. In particular the points **B** and **R** of the right side of Fig. 4, conceived as points of the plane, are the traces of segments in 3D, orthogonal to the plane and endowed with outer orientation, i.e. by a direction of rotation around the segment. The lines connecting **B** with **R** are the traces of surfaces attached to these segments as the sails of a boat with its trees.

While the notion of inner orientation of a manifold does not depend on the dimension of the embedding space, the notion of outer orientation depends on the dimension of the embedding space. So the stream function in 2D, which is a pseudoscalar, is the component of the stream vector in 3D which is orthogonal to the plane of motion. The stream vector, which is a pseudovector, is associated with a line endowed with outer orientation in 3D. Table 1 shows the inner and the

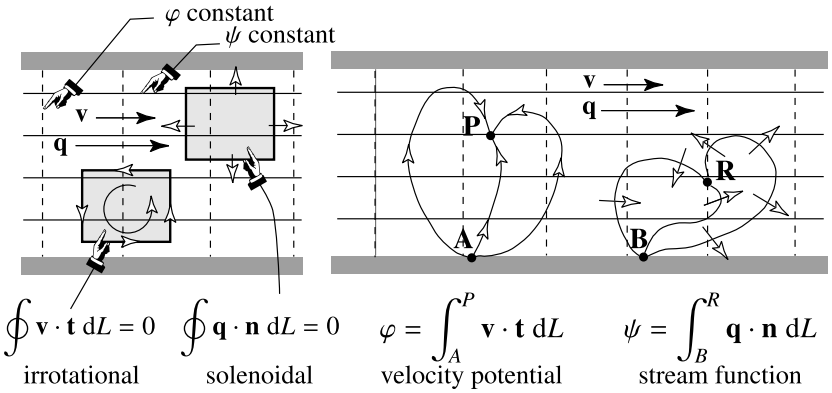


Fig. 4. A fluid flow in a plane motion.

Table 1
The eight oriented space elements. The filled circles indicate the most common orientations of lines, surfaces and volumes used in physics.

inner orientation		outer orientation	
$\overline{\mathbf{P}}$	inner orientation of a point: a positive orientation is the one of sink	outer orientation of a volume: a positive orientation is the one with outwards normals.	$\widetilde{\mathbf{V}}$
$\overline{\mathbf{L}}$	inner orientation of a line: this is the basic notion used to give a meaning to all other space elements.	outer orientation of a surface: this is an inner orientation of a line crossing the surface.	$\widetilde{\mathbf{S}}$
$\overline{\mathbf{S}}$	inner orientation of a surface: this is a compatible orientation of its edges, i.e. a direction along its boundary.	outer orientation of a line: this is an inner orientation of a surface crossing the line.	$\widetilde{\mathbf{L}}$
$\overline{\mathbf{V}}$	inner orientation of a volume: this is a compatible orientation of its faces. It is equivalent to the screw rule.	outer orientation of a point: this is an inner orientation of the volume containing the point.	$\widetilde{\mathbf{P}}$

outer orientation of the four space elements. We will denote the space elements endowed with inner orientation by placing a bar over the boldface, uppercase letters, i.e. $\overline{\mathbf{P}}, \overline{\mathbf{L}}, \overline{\mathbf{S}}, \overline{\mathbf{V}}$ and the space elements endowed with outer orientation by placing a tilde over the boldface, uppercase letters, i.e. $\widetilde{\mathbf{P}}, \widetilde{\mathbf{L}}, \widetilde{\mathbf{S}}, \widetilde{\mathbf{V}}$.

3.5. Global variables are associated with oriented space elements

We have said that global variables in space are "associated" with space elements: What do we mean with the term *associated*? To explain this, let us consider the notions of "flux" and "flow": every time we talk about flux or flow, e.g. magnetic flux Φ, electric flux Ψ, vortex flux W, energy flow E^{f}, entropy flow S^{f}, mass flow M^{f}, charge flow Q^{f}, we refer to a surface. Wanting to highlight this attribute of the variable, we will say that every flux is *associated* with a surface.

Also the variables which are obtained by a line integral of a vector, such as the voltage E, the magnetomotive force F_{m}, the line integral of the velocity Γ in fluid dynamics, the work of a force W along a line in a force field, since they involve a line in their definition, we say that they are *associated* with a line.

Similarly, the concept of "content" of a variable, which is typical of the spatial description, e.g. the mass content M^{c}, the energy content E^{c}, the entropy content S^{c}, the momentum content $\mathbf{P}^{\mathrm{c}}$, refers to a volume. We express this fact by saying that these variables are *associated* with a volume.

There are also variables that are referred to points *and that are not densities of other variables*: these include temperature T, electric potential ϕ, total displacement $\mathbf{u}$ in continuum mechanics. We will say that these variables are *associated* with points.

To highlight this association, we place the corresponding space element enclosed within square brackets by also indicating the nature of its orientation. Table 2 provides examples of this nomenclature. This association with the space element highlights a characteristic aspect of the variable, a sort of "DNA".

Some fluxes are associated with surfaces endowed with inner orientation, whereas other fluxes are associated with surfaces endowed with outer orientation.

For example, the vortex flux W and the magnetic flux Φ are associated with surfaces $\overline{\mathbf{S}}$ endowed with inner orientations.

Table 2
Association of some physical variables with space elements.

$M^{\mathrm{c}}[\widetilde{\mathbf{V}}]$	mass content	$\mathbf{F}^{\mathrm{v}}[\widetilde{\mathbf{V}}]$	volume force	$V^{\mathrm{c}}[\widetilde{\mathbf{V}}]$	pot. energy content
$Q[\widetilde{\mathbf{S}}]$	heat	$\mathbf{T}[\widetilde{\mathbf{S}}]$	int. surface force	$\Phi[\bar{\mathbf{S}}]$	magnetic flux
$E[\bar{\mathbf{L}}]$	voltage	$e[\bar{\mathbf{L}}]$	extension	$F_{\mathrm{m}}[\widetilde{\mathbf{L}}]$	magnetomotive force
$T[\bar{\mathbf{P}}]$	temperature	$\mathbf{u}[\bar{\mathbf{P}}]$	total displacement	$\check{\phi}_{\mathrm{m}}[\widetilde{\mathbf{P}}]$	scalar magn. potential

Table 3
Global variables in space.

	inner orientation	*outer orientation*	
$\bar{\mathbf{P}}$	$\phi[\bar{\mathbf{P}}], U[\bar{\mathbf{P}}], \phi[\bar{\mathbf{P}}], \mathbf{u}[\bar{\mathbf{P}}]$	$M^{\mathrm{c}}[\widetilde{\mathbf{V}}], E^{\mathrm{c}}[\widetilde{\mathbf{V}}], Q^{\mathrm{c}}[\widetilde{\mathbf{V}}], S^{\mathrm{c}}[\widetilde{\mathbf{V}}], \mathbf{P}^{\mathrm{c}}[\widetilde{\mathbf{V}}], \check{\mathbf{L}}^{\mathrm{c}}[\widetilde{\mathbf{V}}]$	$\widetilde{\mathbf{V}}$
$\bar{\mathbf{L}}$	$E[\bar{\mathbf{L}}], \Gamma[\bar{\mathbf{L}}], W[\bar{\mathbf{L}}]$	$M^{\mathrm{f}}[\widetilde{\mathbf{S}}], E^{\mathrm{f}}[\widetilde{\mathbf{S}}], Q^{\mathrm{f}}[\widetilde{\mathbf{S}}], S^{\mathrm{f}}[\widetilde{\mathbf{S}}], \mathbf{P}^{\mathrm{f}}[\widetilde{\mathbf{S}}], Q[\widetilde{\mathbf{S}}]$	$\widetilde{\mathbf{S}}$
$\bar{\mathbf{S}}$	$\Phi[\bar{\mathbf{S}}], W[\bar{\mathbf{S}}]$	$F_{\mathrm{m}}[\widetilde{\mathbf{L}}]$	$\widetilde{\mathbf{L}}$
$\bar{\mathbf{V}}$	$\check{G}[\bar{\mathbf{V}}]$	$\check{\phi}_{\mathrm{m}}[\widetilde{\mathbf{S}}], \check{\psi}[\widetilde{\mathbf{P}}]$	$\widetilde{\mathbf{P}}$

In fact the vortex flux "across" a surface is defined as the velocity circulation along the line forming the boundary of the surface. The inner orientation of the boundary line defines an inner orientation of the surface. In a similar way the magnetic flux "across" a surface is defined as the time integral of the circulation of the electric vector $\mathbf{E}$ along the line forming the boundary of the surface (impulse of the electromotive force, in volt $\times$ second $=$ weber). The inner orientation of the boundary line defines an inner orientation to the surface. Note that nothing "crosses" the surface.

In contrast, the electric flux Ψ, mass flow M^{f}, energy flow E^{f}, charge flow Q^{f}, entropy flow S^{f}, momentum flow $\mathbf{P}^{\mathrm{f}}$, heat Q are associated with a surfaces $\widetilde{\mathbf{S}}$ endowed with outer orientation. In fact in these cases something really crosses the surface: it can be electric charge, mass, energy, momentum, heat, etc. Hence the outer orientation of the surface is involved.

Similarly, the voltage E, the line integral of the velocity Γ and the work W of a force in a force field are associated with a line $\bar{\mathbf{L}}$ endowed with inner orientation, whereas the magnetomotive force F_{m} is associated with a line $\widetilde{\mathbf{L}}$ with outer orientation. In fact the magnetomotive force along the axis of a solenoid depends on the direction of the current inside the solenoid, i.e. to the outer orientation of the axis.

All the contents, e.g. mass content M^{c}, energy content E^{c}, charge content Q^{c}, entropy content S^{c}, momentum content $\mathbf{P}^{\mathrm{c}}$, angular momentum content $\check{\mathbf{L}}^{\mathrm{c}}$, are associated with volumes $\widetilde{\mathbf{V}}$ with outer orientation: only the hypothetical magnetic charge $\check{G}$ is associated with volumes $\bar{\mathbf{V}}$ endowed with inner orientation. We use the "check" symbol over the letter to denote its "pseudo" character, i.e. axial vector and axial scalar. With regard to the variables associated with points, some of them, such as the electric potential ϕ, the gravitational potential U_{g}, the velocity potential ϕ, the total displacement $\mathbf{u}$ are associated with points $\bar{\mathbf{P}}$ endowed with inner orientation, whereas others, such as, for example, the magnetic scalar potential $\check{\phi}_{\mathrm{m}}$ and the stream function $\check{\psi}$ (in 2D), are associated with points $\widetilde{\mathbf{P}}$ endowed with outer orientation. Table 3 shows the variables which we have mentioned.

We raise the question: is the marriage between physics and geometry so strict that the notion of inner and outer orientation is really necessary? Our answer is categorically: yes. Can this association of global physical variables with oriented space elements suggest the introduction of a new classification of global physical variables? The answer is affirmative as it is shown in the book [10]. We can state the following

Association principle. *In physical theories, in the* spatial *description, the global physical variables have a natural association with one of the four space elements* $\mathbf{P}$, $\mathbf{L}$, $\mathbf{S}$, $\mathbf{V}$.

The field variables which are deduced from global variables performing their density or their rate, inherit the association with the same space element of the corresponding global variable.

How can we decide if a global physical variable is associated with a space element endowed with inner or outer orientation? The main criterion is offered by the following oddness principle: [10]

Oddness principle. *A global physical variable associated with an oriented space or time element changes sign when the orientation of the element is inverted.*

This principle is obvious for those variables which are associated with lines and surfaces endowed with inner or outer orientation, but it is difficult to accept when applied to variables associated with points and volumes.

A partial reason for this difficulty lies in our lack of familiarity with the orientation of points, a notion not expressly dealt with in the literature; moreover as far as concerns the variables associated with volumes endowed with outer orientation, the difficulty lies in the custom of not considering the inversion of the role active $\leftrightarrow$ passive. So we are well aware that the weight of our body, which is described by a force directed downwards, is a volume force acting on the body due to the gravitational attraction of the Earth. On account of the action and reaction principle, we may consider the attraction that our body exerts on Earth, this force goes upward. Hence there is a change in sign of the force when we consider our body as the active subject and the Earth as the passive one.

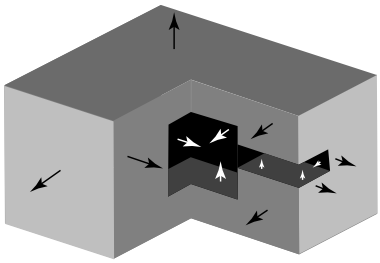


Fig. 5. The outward normals of a volume become inward normals for a hole in the volume.

Also the number of particles contained within a volume can be considered negative: this happens when we consider the volume as a hole in a region and hence with inwards normals, as seen in Fig. 5. The oddness principle is implicitly included in the additivity of physical variables and is required for the validity of the Green's and Stokes' theorems.

Empirical rule. *Configuration variables are associated with space elements endowed with inner orientation, hence* $\bar{\mathbf{P}}, \bar{\mathbf{L}}, \bar{\mathbf{S}}, \bar{\mathbf{V}}$ *whereas source variables and energy variables are associated with space elements endowed with outer orientation, hence* $\tilde{\mathbf{P}}, \tilde{\mathbf{L}}, \tilde{\mathbf{S}}, \tilde{\mathbf{V}}$.

The author is unable to give a justification of this empirical rule that reveals an unexpected marriage between physics and geometry. Table 4 provides an intuitive insight into this association for four physical theories. This subject is extensively dealt with in the book [10] where about 180 physical variables of different theories have been analyzed and classified.

In physics the distinction between the two types of orientation of a space element is not usually considered. Nobody talks about points endowed with inner or outer orientation and volumes endowed with inner orientation. Rarely in physics we talk about surfaces endowed with inner orientation and almost never we talk about lines with outer orientation.

In contrast, the distinction between inner and outer orientation plays a pivotal role in the present analysis, because it corresponds to the distinction between configuration and source variables which are mapped respectively on the left and on the right part of the classification diagram of physical variables (see later).

In turn, this distinction allows to highlight the location of the constitutive relationships in the classification diagram, as a link between the variables associated with space elements endowed with inner orientation and the variables associated with space elements endowed with outer orientation. We think that this association

global physical variable $\rightarrow$ *oriented space element*

must have a fundamental role in computational physics, even if, up to now, it has been ignored.

3.6. The birth of cell complexes

"*In* 1752 *the great mathematician Leonhard Euler discovered a simple geometrical fact that had escaped notice by geometers for two thousand years*." [6, p. 241]. Counting the number of vertices V, of edges E and of faces F of a polyhedron, Euler found the simple arithmetical rule $V - E + F = 2$. This is true *whatever* the number of faces of the polyhedron, *whatever* the number of edges of every face and it is valid even if the polygons are not regular. This shows that this is a topological property rather than a geometrical property. The quantity $\chi = V - E + F$ is called *Euler characteristic*.

At first glance, it is not surprising that this rule holds good even if the faces and their edges are curved, but this seemingly innocent structure opens the door to a branch of mathematics: algebraic topology.

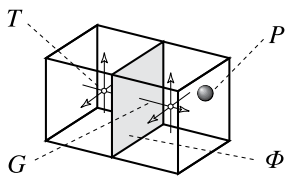
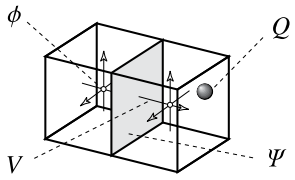
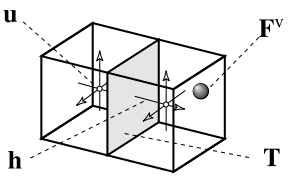
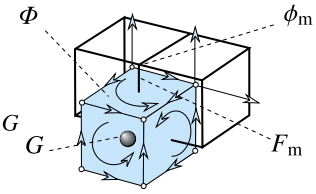
To this end, consider an octahedron, as shown on the left of Fig. 6. Imagining that it is made of a transparent plastic and it is contained inside a glass sphere, if it is inflated via a straw, like a soap bubble, as shown in the center of Fig. 6, we can deform the octahedron to become a spherical surface. In this process the number of vertices, edges and faces is the same as the original octahedron. What we obtain is like drawing the lines on a globe, i.e. the equator, the Greenwich meridian and another meridian at 90° degrees to it. This shows that, instead of starting by deforming a polyhedron, we can start directly with a surface, e.g. a sphere, drawing on it a network of lines thus dividing the surface into many pieces. For a closed surface, like a sphere, the Euler characteristic remains 2.

When we consider a triangulation of a flat surface the Euler's characteristic becomes one instead of two: $\chi = 1$ and when we consider a polyhedron with a hole, torus-like, the characteristic becomes zero: $\chi = 0$ as shown in Fig. 7.

Topologists, being interested in characterizing the properties in the large of manifolds, found in the Euler characteristic what they were searching. We can cover the manifold (here we consider only 2D and 3D regions) with a *cell complex*, what topologists call CW-complex. When the cells of greater dimension are simplexes, the cell complex becomes a *simplicial complex*, the most used by topologists.

A physicist and an engineer, whose main interest is to describe physical phenomena and solve physical problems, after realizing that physical variables are associated with space elements, can see in a cell complex what they seek for computation. Such cell complexes are the familiar *meshes*.

Table 4
Analogies between four physical fields.

temperature T refers to vertices of the primal cell complex temperature difference G refers to edges of the primal cell complex	**thermal field** 	heat generation rate P refers to cells of the dual cell complex heat current Φ refers to faces of the dual cell complex
electric potential ϕ refers to vertices of the primal cell complex voltage V refers to edges of the primal cell complex	**electric field** 	electric charge Q refers to cells of the dual cell complex electric flux Ψ refers to faces of the dual cell complex
total displacement $\mathbf{u}$ refers to vertices of the primal cell complex relative displacement $\mathbf{h}$ refers to edges of the primal cell complex	**elastic field** 	volume force $\mathbf{F}^{\mathrm{v}}$ refers to cells of the dual cell complex surface force $\mathbf{T}$ refers to faces of the dual cell complex
magnetic flux Φ refers to faces of the primal cell complex hypotetical magnetic charge G refers to cells of the primal cell complex	**magnetic field** 	scalar magnetic potential ϕ_{m} refers to vertices of the dual cell complex magnetom. force F_{m} refers to edges of the dual cell complex

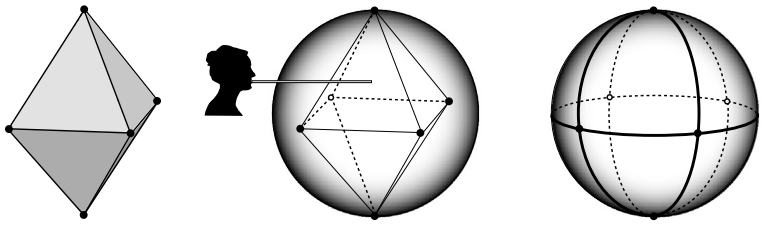


Fig. 6. From a polyhedron to a sphere performing a continuous deformation.

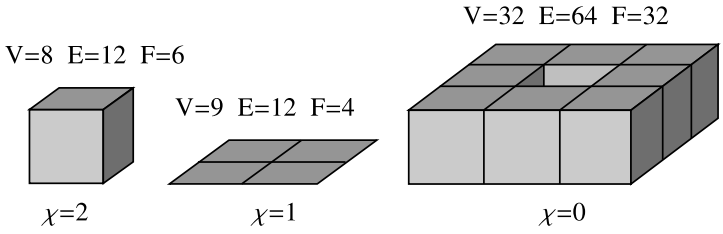


Fig. 7. The Euler characteristics is a global property of a manifold.

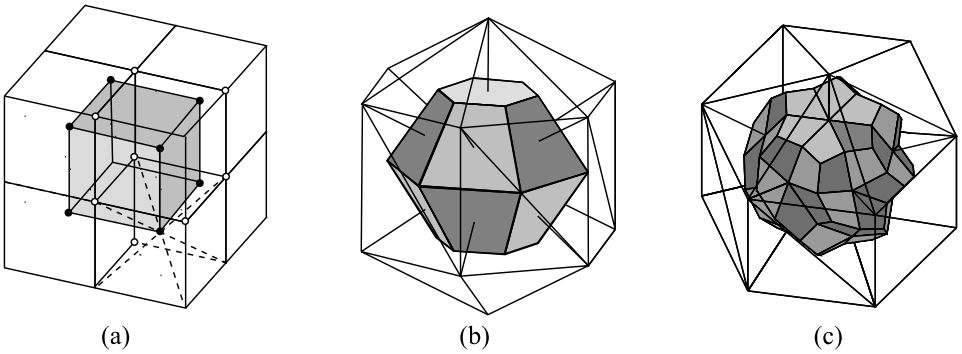


Fig. 8. Dual cells. (a) Cartesian cell complex; (b) Voronoi dual in a simplicial complex; (c) barycentric dual in a simplicial complex.

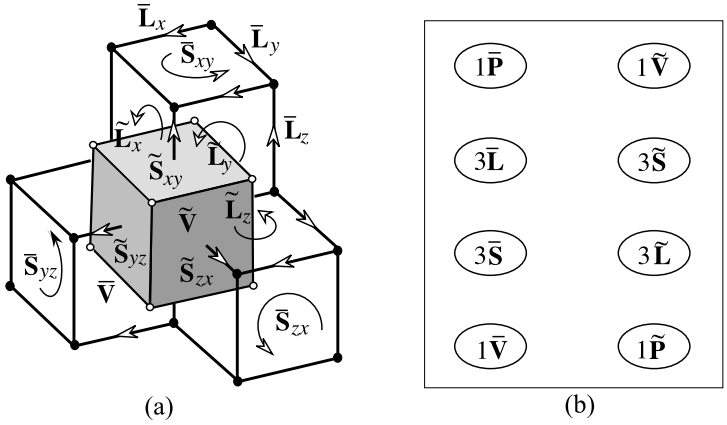


Fig. 9. (a) The outer orientation of the dual complex is the one induced by the inner orientation of the primal complex. (b) The eight oriented space elements can be disposed in a table.

Table 5
The correspondence between physical, geometrical and analytical properties.

configuration variables	source variables
inner orientation	outer orientation
primal cell complex	dual cell complex
(even differential forms)	(odd differential forms)

3.7. The need of a dual cell complex

We have said that configuration variables are naturally associated with space elements endowed with inner orientation whereas source variables are associated with outer orientation. It is spontaneous to assign an inner orientation to the elements of all dimensions of a cell complex, i.e. to vertices (points $\bar{\mathbf{P}}$, also 0-cells), edges (lines $\bar{\mathbf{L}}$, also 1-cells), faces (surfaces $\bar{\mathbf{S}}$, also 2-cells) and three-dimensional cells (volumes $\bar{\mathbf{V}}$, also 3-cells). At this point we need an analogous cell complex whose cells of all dimensions are endowed with an *outer* orientation to associate the source variables with the space elements. How do we accomplish this? The simplest way is to consider another cell complex, dual of the first. This means that to every p-cell of the first complex, from now on called *primal complex*, corresponds a $(3-p)$-cell of the second complex, which will be called *dual complex*. This is achieved by introducing a mesh staggered, as shown in Fig. 8. In this way we can assign to every cell of the dual complex the outer orientation induced by the inner orientation of the corresponding primal cell, as shown in Fig. 9(a). Hence we are lead to consider the dual volumes $\tilde{\mathbf{V}}$, the dual faces $\tilde{\mathbf{S}}$, the dual edges $\tilde{\mathbf{L}}$ and the dual vertices $\tilde{\mathbf{P}}$. The $4 \times 2 = 8$ oriented space elements can be classified as shown in Fig. 9(b).

In computational literature the dual complex is commonly called “staggered”. In 1966 Yee [18] introduced the staggered meshes to deal with electromagnetic field in order to avoid spurious solutions, i.e. to avoid numerical drawbacks. Even today we do the same, as it is evident in the book of Wesseling [17]. The present analysis gives a conceptual justification of the staggered mesh, which coincides with the dual cell complex, because it is required by the source variables which need the outer orientation of p-cells. *It is the fundamental distinction between the two types of physical variables, the configuration and the source variables, which requires two cell complexes. The reason is that the two types of variables require the two types of orientations, the inner and the outer one*. This double distinction is neglected in literature. Table 5 shows the correspondence between the notions discussed up to now.

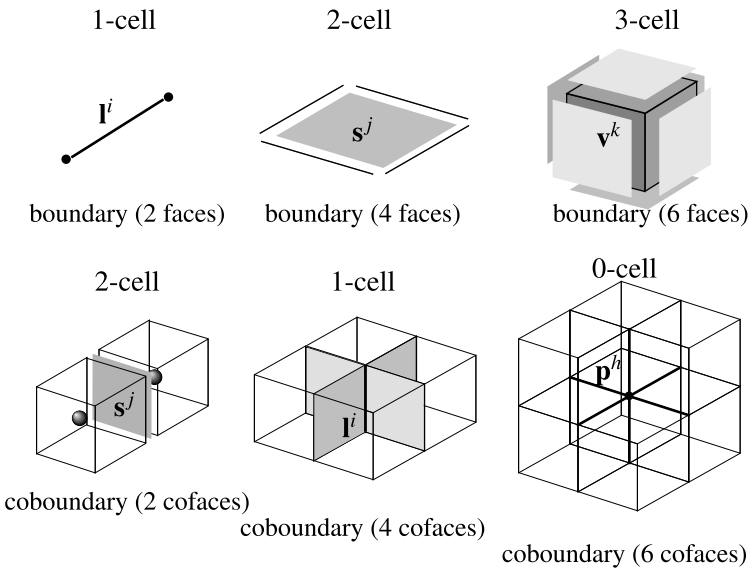


Fig. 10. The faces and cofaces of 0, 1, 2, 3-cells, respectively, in a Cartesian complex.

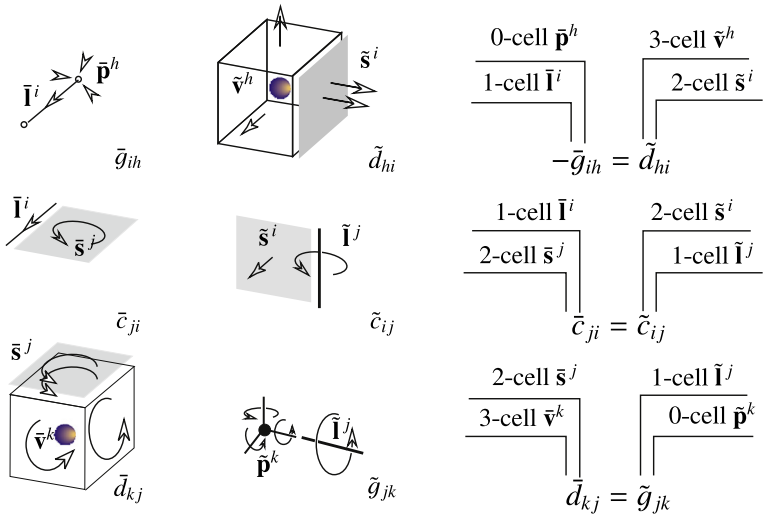


Fig. 11. The incidence numbers between the cells of a cell complex and its dual.

3.8. Faces and cofaces of a p-cell

The *faces* of a p-cell are those $(p-1)$-cells which are incident to the p-cell, the set of all faces is the *boundary* of the p-cell. The *cofaces* of a p-cell are those $(p+1)$-cells which are incident to the p-cell; the set of all cofaces is called the *coboundary* of the p-cell.

While the notion of "face" is a familiar one, that of "coface" is almost completely unknown even in books of algebraic topology. Fig. 10 shows the cofaces of vertices, edges and faces in a Cartesian cell complex.

Let us consider two rooms separated by a wall: the wall is the face common to both rooms and the two rooms are the cofaces of the wall. While the notion of *face* is meaningful for a single cell, the notion of *coface* is meaningful only when the cell is considered as a member of a cell complex.

3.9. The incidence numbers

Fig. 11 shows the incidence numbers of a 3-cell complex and its dual. For historical reasons points have been oriented as *sinks* whereas volumes are oriented with *outward* normals. This explains the minus sign in the relationship between the incidence numbers edges-vertices $\bar{g}_{ih}$ and the incidence numbers cells-faces $\tilde{d}_{hi}$.

3.10. Affine scalar and vector fields

In the study of scalar and vector fields it is highly opportune to treat first the special case of the *affine* fields for two reasons: (1) every scalar or vector field, in a region of regularity and in the neighborhood of every point can be

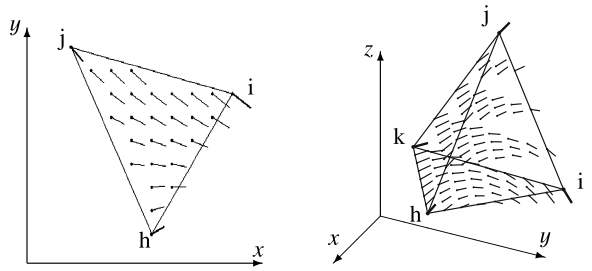


Fig. 12. (left) A triangle in 2D and the vector field approximated by affine fields induced by the vectors at the three vertices. (right) The same for a tetrahedron in 3D.

approximated with an affine field; (2) an affine field benefits from simple properties which make evident the introduction of the typical operators such as gradient, curl and divergence. An affine scalar field is one that, in the Cartesian coordinates, has the form

$$\phi(x, y, z) = a + h_x x + h_y y + h_z z \quad \longrightarrow \quad \phi = a + \mathbf{h} \cdot \mathbf{r} \tag{2}$$

whereas an affine vector field, in the Cartesian coordinates, has the form

$$\begin{cases} v_x = a_x + h_{xx}x + h_{xy}y + h_{xz}z \\ v_y = a_y + h_{yx}x + h_{yy}y + h_{yz}z \\ v_z = a_z + h_{zx}x + h_{zy}y + h_{zz}z \end{cases} \quad \longrightarrow \quad \mathbf{v} = \mathbf{a} + \mathbf{H}\mathbf{r} \tag{3}$$

i.e. each Cartesian component is an affine scalar function. *The affine behavior is compatible with simplexes* (triangles in 2D and tetrahedra in 3D) not with cells of arbitrary number of vertices. Hence for an affine scalar field in 2D we have 3 unknowns (a, h_x, h_y) which can be expressed in terms of the values of the scalar function at the 3 vertices of the triangle. In an affine scalar field in 3D we have 4 unknowns (a, h_x, h_y, h_z) which can be expressed in terms of the values of the scalar function at the 4 vertices of the tetrahedron. For an affine vector field in 2D we have 6 unknowns $(a_x, a_y, h_{xx}, h_{xy}, h_{yx}, h_{yy})$ which can be expressed in terms of the 6 components of the 3 vectors at the vertices of the triangle. In an analogous way in an affine vector field in 3D we have 12 unknowns for every tetrahedron and these can be determined in terms of the 3 components of the 4 vectors at the 4 vertices of the tetrahedron. This is the reason why we will refer only to simplicial complexes (Fig. 12).

The matrix **H** in Eq. (3) is called *gradient* of the vector **v**. Its elements can be written as $h_{hk} = \partial v_h / \partial x_k$. Every field which is *regular* in a region (i.e. is continuous and with continuous variations) can be approximated in the neighborhood of a point with an affine field. This is the natural extension of the property that a regular line, can be locally approximated by its tangent. Hence an affine function is obtained from the development of a function in a Taylor series, ignoring the terms higher than the first order. An affine field is the sum of a *uniform* field (i.e. invariant under translation) and a *linear* field.

The interest of affine scalar and vector fields for computational physics lies in the fact that many properties, which in the realm of differential formulation are valid at the limit, are also valid in the small where the field can be considered affine. In practice, instead of saying "*ignoring infinitesimals of higher order*", we can say "*supposing that the field is locally affine*."

We list here some basic properties of affine fields which are useful for computational physics. We list them without proofs (see, for example, [10, p. 465] or [14, p. 22]).

Theorem 1. *The value of an affine function in the midpoint of a segment is the mean of the values of the function at the two extremes.*

This property is easily proved.

Theorem 2. *In an affine vector field the vector in the midpoint of a segment is the mean of the values of the vector at the two extremes.*

This property is evident from Fig. 13.

Theorem 3. *In an affine vector field the line integral of the vector along a straight line segment is equal to the scalar product of the vector evaluated in the midpoint of the segment for the vector which describes the oriented segment. In short*

$$\int_{\mathbf{P}}^{\mathbf{Q}} \mathbf{v} \cdot \mathbf{t}\, \mathrm{d}L = \mathbf{v}(\mathbf{M}) \cdot (\mathbf{Q} - \mathbf{P}). \tag{4}$$

Theorem 4. *The flux of an affine vector field through a plane surface of area A and unit normal* **n** *is equal to the scalar product of the vector evaluated in the centroid of the plane surface for the vector* $\mathbf{n}A$ *which describes the surface* (Fig. 14).

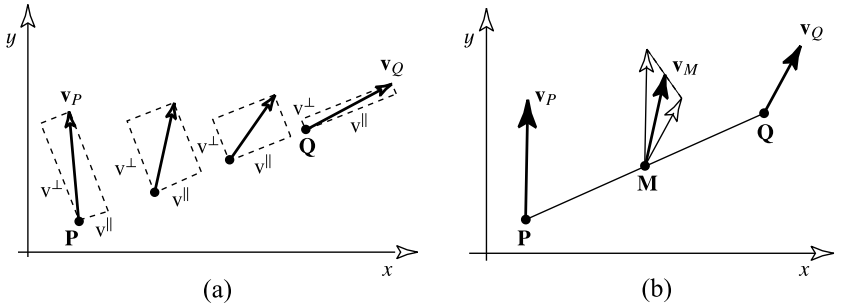


Fig. 13. The property of the midpoint.

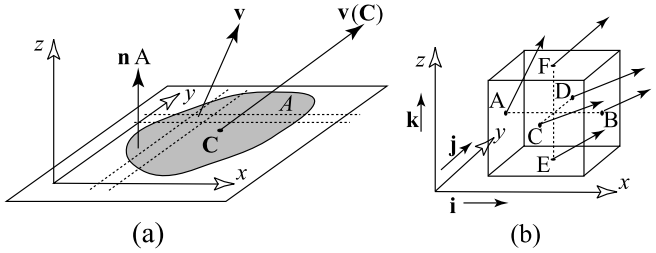


Fig. 14. (a) The flux across a plane surface; (b) flux across the boundary of a cube.

Theorem 5. *In an affine vector field, the circulation of the vector along the boundary of a plane surface is proportional to the area of the surface*

$$\Gamma \stackrel{\text{def}}{=} \int_{\partial \mathbf{S}} \mathbf{v} \cdot \mathbf{t} \, \mathrm{d}L \propto A. \tag{5}$$

Theorem 6. *In an affine vector field, the flux of the vector on a closed surface is proportional to the volume enclosed by the surface*

$$\Phi \stackrel{\text{def}}{=} \int_{\partial \mathbf{V}} \mathbf{v} \cdot \mathbf{n} \, \mathrm{d}S \propto V. \tag{6}$$

Theorem 7. *The gradient of an affine scalar* (*resp. vector*) *field is a uniform vector* (*resp. tensor*) *field.*

Theorem 8. *The curl of an affine vector field is a uniform vector field.*

Theorem 9. *The divergence of an affine vector field is a uniform scalar field.*

4. Computational physics PDE free

The use of global variables as starting point in the description of physics leads us to an *algebraic* or *discrete* or *direct* or *finite* formulation of physical theories, which is an alternative to the *differential formulation*, including the use of the *exterior differential forms*. Starting from the algebraic formulation, we may easily deduce the differential formulation. Presenting physical theories by using the algebraic formulation is very useful in teaching because it makes evident the geometrical concepts involved, it is adherent to physical measurements and avoids the premature recourse to the differential operators whose symbolism is often too abstract for the average university student. Let us consider, for example, the Gauss law of electrostatics, which states that the electric charge Ψ collected on the boundary of a volume (on a metallic shell) is equal to the charge contained inside the volume. This law can be grasped more easily in the global form than in the integral or in the differential form ($\mathbf{V}$ is a space region, $\partial\mathbf{V}$ is its boundary)

$$\underbrace{\Psi[\partial \mathbf{V}] = Q[\mathbf{V}]}_{\text{global formulation}} \qquad \underbrace{\int_{\partial \mathbf{V}} \mathbf{D} \cdot \mathbf{n} \, \mathrm{d}S = \int_{\mathbf{V}} \rho \, \mathrm{d}V}_{\text{integral formulation}} \qquad \underbrace{\nabla \cdot \mathbf{D} = \rho}_{\text{differential formulation}} \quad . \tag{7}$$

The integral formulation sounds more familiar to us only because we are accustomed to use the electric displacement $\mathbf{D}$ and the density ρ. In contrast, the global formulation can be grasped more easily because it refers directly to the measured variables, i.e. avoiding the introduction of the field functions and their integration.

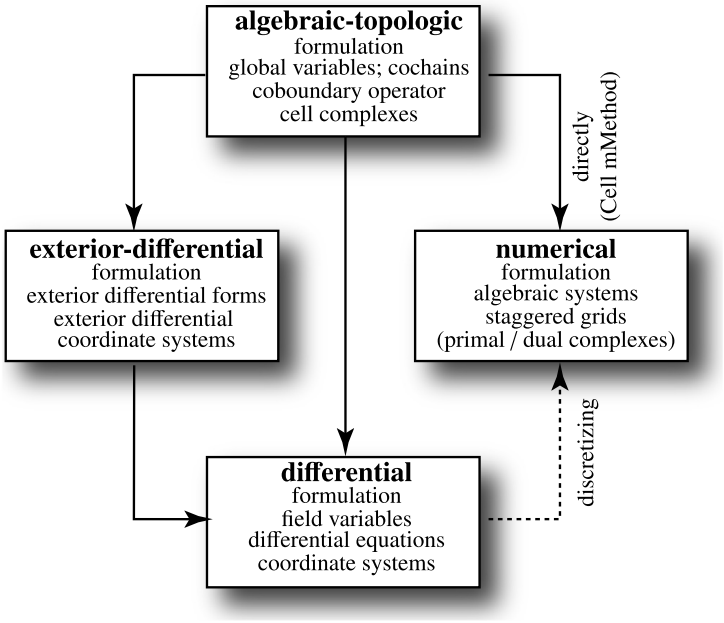


Fig. 15. The four mathematical formulations of physics.

Since the global formulation does not require an integration process, it is of algebraic nature and, as such, can be directly used for digital computation by using a cell complex and its dual in the computational domain. This formulation does not need the discretization of the differential equations, which is necessary in the existing numerical methods, because it uses global variables and balance equations in a global form. So, from a computational point of view, the direct algebraic formulation is free from the inherent difficulties of the differentiability such as: (1) the splitting of physical laws into differential equations in the regions of regularity and jump conditions across the surfaces of discontinuity; (2) the introduction of the generalized functions (e.g. Dirac delta function, etc.); (3) the recourse to the weak formulation.

A numerical method which is based on the direct algebraic formulation is the *Cell Method*. Papers dealing with this method are listed in [15] (Fig. 15).

4.1. Discrete forms (cochains)

The fact that global physical variables have a natural association with the oriented space elements leads us to introduce in the region a cell complex and its dual and to associate the global variables to the cells of the corresponding dimensions. The resulting framework is the following: instead of considering the points of the space region and the field variables, i.e. the point functions, we divide the region into cells, i.e. we build a cell complex. With the cells of the various dimensions (0-, 1-, 2-, 3-cells) we associate the corresponding value of the global physical variables. In this way all global variables of a physical field are fragmented in the region.

Examples. In a magnetic field we associate the magnetic flux Φ (weber) with every 2-cell of the primal complex; in an electric field we associate the voltage E (volt) with every 1-cell of the primal complex; in a fluid field we associate the line integral of velocity Γ (m^2/s) with every 1-cell of the primal complex. In continuum mechanics we associate the total displacement **u** (meter) with every primal 0-cell; the relative displacement of two points **h** (meter) with every primal 1-cell; the surface force **T** (newton) with every dual 2-cell; the volume force $\mathbf{F}^v$ (newton) with every dual 3-cell.

Around 1930 the algebraic topologists have conceived the idea of associating each p-cell, with an element of a group $\mathcal{G}$, calling the resulting collection a *p-cochain*. This unusual term arises because a collection of p-cells of a complex was previously called *p-chain*.

Much later it was realized that cochains are the discrete analogue of exterior differential forms. This suggested as alternative to the term “cochain” the term *discrete differential form* [1,2]. Since the juxtaposition of two opposite terms *discrete/differential* sounds contradictory, we prefer to use the simplest locution *discrete forms*, as some other authors do [3,5]. Hence, when we associate a global physical variable (or an element of a group $\mathcal{G}$) with every p-cell of a cell complex, we say that we have defined a discrete form of *degree p* (note: a cochain has a *dimension*, whereas a discrete form has a *degree*).

4.2. The astonishing coboundary process

Let us start by considering a balance law, such as the balance of electric charge. Usually we consider, in a Cartesian coordinate system, an infinitesimal "cube" delimited by the coordinate surfaces xy, yz, zx. We consider the electric charge Q associated with the cube and the electric fluxes Ψ associated with each of the six faces of the cube and we sum all these fluxes, each being multiplied by $+1$ or -1 according to the mutual incidence between the orientation of the face and the orientation of the cube. The faces have an outer orientation, usually represented by an arrow crossing the face and oriented in the positive direction of the coordinate axes. The cube is oriented with outward normals and since this is an outer orientation, it follows that the 3-cell belongs to a dual complex in harmony with the fact that the electric charges and fluxes are source variables. The numbers $+1$ and -1 are then the incidence numbers between the 3-cells and the 2-cells. Denoting an arbitrary 3-cell of the dual complex with $\tilde{\mathbf{v}}^h$, an arbitrary 2-cell with $\tilde{\mathbf{s}}^i$ and with $\tilde{d}_{hi}$ the incidence numbers, the balance law can be written

$$\boxed{\sum_i \tilde{d}_{hi}\Psi_i = Q_h} \tag{8}$$

instead of the more sophisticated form

$$-D(x,y,z)\,\mathrm{d}y\,\mathrm{d}z + D(x+\mathrm{d}x,y,z)\,\mathrm{d}y\,\mathrm{d}z + \cdots = \rho\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \tag{9}$$

which culminates in the equation

$$\operatorname{div}\mathbf{D} = \rho. \tag{10}$$

Let us compare Eq. (8) with Eq. (10). The latter form requires the introduction of the density ρ and of the electric displacement $\mathbf{D}$, whereas Eq. (8) uses only the global variables Q and Ψ, *the ones we can measure directly*. Eq. (10) presupposes that the electric displacement is a differentiable function, whereas Eq. (8) does not require this mathematical property. Eq. (10) has been deduced with recourse to an infinitesimal volume, hence ignoring the infinitesimal of order greater than one, whereas Eq. (8) has been applied to a 3-cell of whatever shape which belongs to a dual cell complex. This means that we are free to choose *at will* the shape of the cells, hence to adapt their shape to the shape of the region in which the field is defined, without any constraint due to the coordinate system used. But what exceeds all expectations is that Eq. (8) can be directly used for numerical analysis, whereas Eq. (10) needs a discretization.

Which is preferable between these two equations?

But there is much more.

Balance equations. The electric flux Ψ_i, involved in the balance for the 3-cell $\tilde{\mathbf{v}}^h$, is also involved in the balance of the 3-cell adjacent to it. In each coface of a 2-cell, the amount Ψ_i enters multiplied by the incidence number between the 2-cell and the 3-cell. This fact strongly suggests that we consider a simple two-steps process to be performed on the fluxes

- the first step consists in transferring the fluxes associated with all 2-cells to their cofaces, each multiplied by the relative incidence number;
- the second step is to perform, for each cell, the sum of the fluxes that come to them from the first step.

With this two-step process we start from a discrete form of degree 2 and we obtain a discrete form of degree 3. Since the fluxes are transferred to the cofaces of every 2-cell, hence to their coboundary, the process is rightly called *coboundary process*. This sum is equaled to the global variable associated with the 3-cell, in the case of electrostatics to the electric charge contained. This process is illustrated in the bottom part of Table 8.

Circuital equations. Let us examine the introduction of the vorticity. In a fluid flow usually we consider circuits formed by coordinate lines. So in a Cartesian coordinate system we consider an infinitesimal circuit on a coordinate surface, say xy, to write

$$w_z\,\mathrm{d}x\,\mathrm{d}y = v_x(x,y,z)\,\mathrm{d}x + v_y(x+\mathrm{d}x,y,z)\,\mathrm{d}y - v_x(x,y+\mathrm{d}y,z)\,\mathrm{d}x - v_y(x,y,z)\,\mathrm{d}y \tag{11}$$

and the same for the faces yz and zx. In this way we have defined the three variables w_x, w_y, w_z, i.e. the components of the vortex pseudovector $\mathbf{w}$. In the end we come to the relation

$$\mathbf{w} \stackrel{\text{def}}{=} \operatorname{curl}\mathbf{v}. \tag{12}$$

Let us see how to obtain the vortex flux using a simplicial complex. Let us consider the line integral Γ of the velocity along the edges of a primal face $\bar{\mathbf{s}}^j$. We denote by $\bar{\mathbf{l}}^i$ an arbitrary 1-cell and by Γ_i the corresponding elementary line integral. We

Table 6
The discrete versions of the three differential operators.

$\gamma_i \stackrel{\text{def}}{=} \sum_i \bar{g}_{ih}\,\phi_h$	$W_j \stackrel{\text{def}}{=} \sum_j \bar{c}_{ji}\,\Gamma_i$	$Q_k \stackrel{\text{law}}{=} \sum_j \tilde{d}_{kj}\Psi_j$
$\mathbf{g} \stackrel{\text{def}}{=} \operatorname{grad}\phi$	$\mathbf{w} \stackrel{\text{def}}{=} \operatorname{curl}\mathbf{v}$	$\rho \stackrel{\text{law}}{=} \operatorname{div}\mathbf{D}$
gradient forming equation	circuital equation	balance equation

denote by W_j the vortex flux associated with the 2-cell $\bar{\mathbf{s}}^j$ defined by the relation

$$W_j \stackrel{\text{def}}{=} \sum_j \bar{c}_{ji}\,\Gamma_i. \tag{13}$$

Written in this way the relationship can be applied to every face of a primal cell complex, not necessarily to a piece of coordinate surfaces of a coordinate system.

The process we have followed can be divided into two steps. In the first step we consider all 1-cells of the primal complex and the corresponding Γ_i. The amount Γ_i associated with the 1-cell $\bar{\mathbf{l}}^i$ is transferred to the cofaces of the 1-cell by pre-multiplication by the incidence number between the 1-cell and the coface. Once this is done, the second step consists of summing, for every 2-cell $\bar{\mathbf{s}}^j$, the amounts coming from its faces to obtain the value W_j. The process is described in the middle part of Table 7.

Definition of the gradient. The process of the formation of the gradients presupposes, as a first stage, the formation of the difference of a function ϕ between two points. Given two points $\bar{\mathbf{p}}^h$ and $\bar{\mathbf{p}}^k$ the difference between the values of a function ϕ_h and ϕ_k will be denoted γ_i, i.e.

$$\gamma_i \stackrel{\text{def}}{=} \sum_i \bar{g}_{ih}\,\phi_h. \tag{14}$$

Once more this process can be described in two steps as follows. Referring to the upper part of Table 7, the first step is to transfer the value ϕ_h from the point $\bar{\mathbf{p}}^h$ to its cofaces, each value being multiplied by the incidence number between the 0-cell and the 1-cell. The second step is to sum, for every 1-cell, the values coming from its faces. The sum thereby obtained is the value which must be associated with the 1-cell. When this process is performed in the realm of the differential formulation and in Cartesian coordinates it gives rise to the quantities

$$\mathrm{d}_x\phi = \partial_x\phi\,\mathrm{d}x \qquad \mathrm{d}_y\phi = \partial_y\phi\,\mathrm{d}y \qquad \mathrm{d}_z\phi = \partial_z\phi\,\mathrm{d}z \tag{15}$$

which can be collected in the single vector

$$\mathbf{g} \stackrel{\text{def}}{=} \operatorname{grad}\phi. \tag{16}$$

We can summarize the three processes in Table 6.

5. The unifying role of the coboundary process

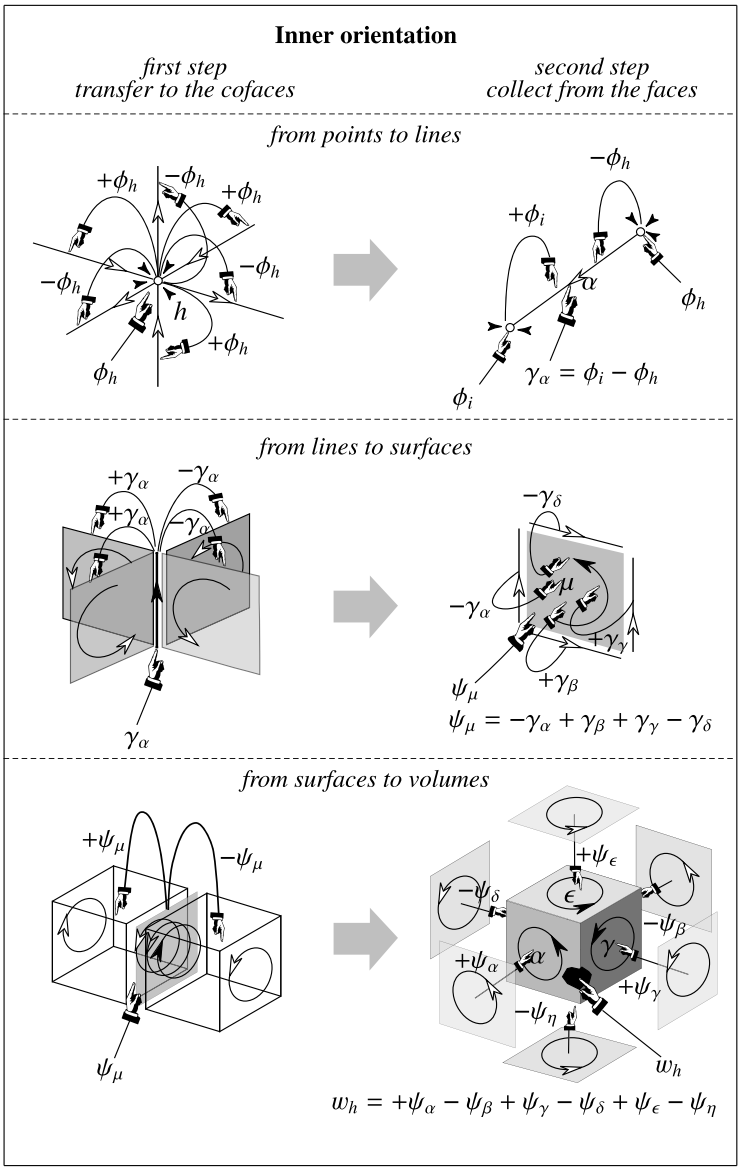
As we can see from the three cases examined, i.e. balance equations, circuital equations and gradient forming equations, the discrete formulation of some basic equations of physics can be described by a unified process consisting of two steps. We start with a discrete form of degree p, i.e. with the assignment of the value of a global physical variable to every p-cell of a cell complex, primal or dual. Hence

1. we transfer the values associated with each p-dimensional cell to its coboundary (i.e. to all its cofaces);
2. for every $(p+1)$-cell of the complex we add the values so transferred.

In this way we obtain a discrete form of degree $(p+1)$. These two simple operations form are called *coboundary process*. This two-step process is illustrated in Table 7 for the primal complex and in Table 8 for the dual complex.

Remark. The author was unable to find any illustration of the coboundary process in books of algebraic topology. He was shocked by the Lagrange's famous statement in the preface of the Mécanique Analytique "*The reader will find no figures in this work.*" which is repeated in the vaunted absence of figures in the excellent books of the Bourbaki group. In contrast Hilbert, in the preface of *Geometry and the Imagination* wrote: "*… it is still true today as it ever was that* intuitive *understanding plays a major role in geometry. And such concrete intuition is of great value not only for the research worker, but also for anyone who wishes to study and appreciate the results of research in geometry.*" There can be no doubts that in mathematics there are different schools of thought just as in politics.

Table 7
The two steps of the coboundary process performed on the primal cell complex.



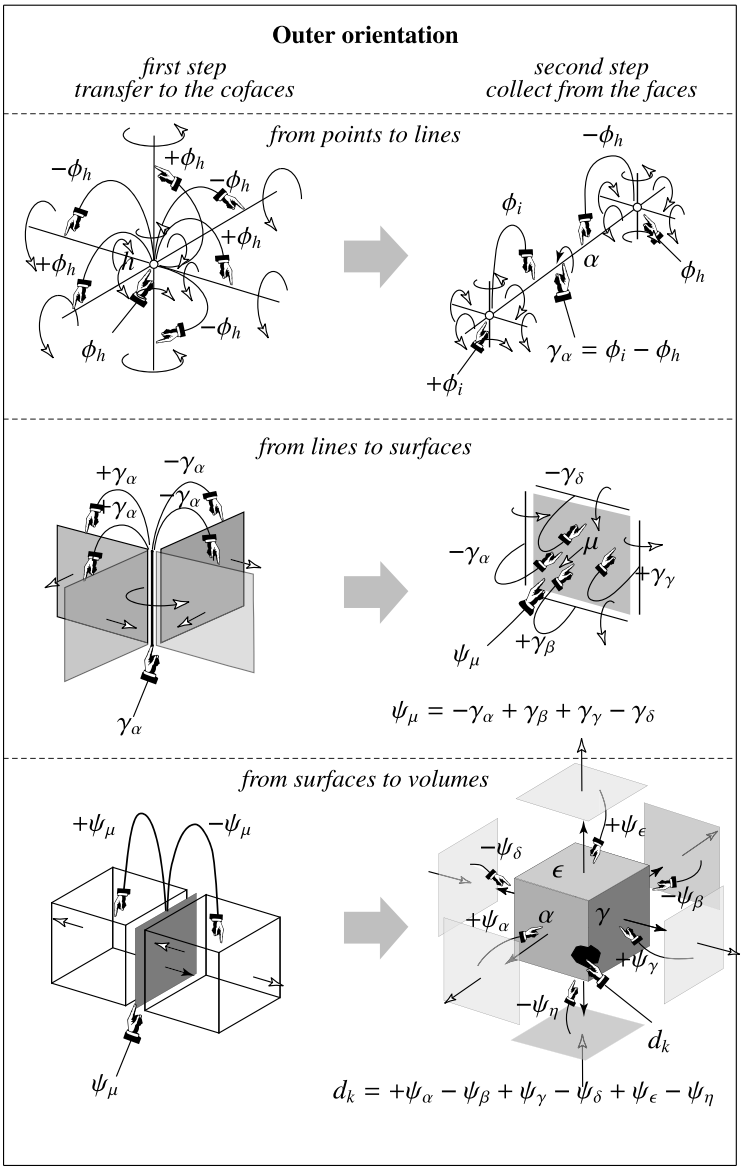
5.1. How to operate in practice

Let us consider, first of all, a two-dimensional region decomposed with a simplicial complex. Let us start by considering a scalar field with the values of the function in the nodes as the main unknowns. To fix our ideas, let us consider the stationary thermal conduction in solids summarized in Table 9. We consider the two main dual complexes.

Voronoi dual complex. This is shown on the left of Fig. 16. The orthogonality of the faces of the Voronoi dual to the edges of the primal complex, is much agreeable because, at least in the most common cases in which the material is isotropic and the constitutive equations are linear, it makes immediate the application of the constitutive law in its discrete form. The constitutive equation is applied under the approximation that inside a region which contains the primal edge $\bar{\mathbf{l}}^i$ and the dual face $\tilde{\mathbf{s}}^i$ the field can be considered uniform, whereas the evaluation of the source term is a bit more elaborate because the area of the dual cell is unknown a priori. Moreover when the two adjacent cells belong to different materials, as in the case of the separation surface between two media, to write the constitutive equation we must perform a further approximation.

The whole procedure is described in Table 9. This procedure is of no use when we have a vector field with vectors associated with the nodes, as in continuum mechanics, because in this case we have to deal with tensor fields and the constitutive relationship loses the simplicity of the scalar theories, such as heat conduction, electrostatics, acoustics in fluids, etc.

Table 8
The two steps of the coboundary process performed on the dual cell complex.



Barycentric dual complex. This is shown on the right of Fig. 16. In what follows we will make reference to the barycentric dual. This scheme is more general than Voronoi's and can be easily applied to continuum mechanics and to fluid dynamics. It is convenient to make the approximation of considering the scalar/vector field to be affine inside every primal (simplicial) cell which implies that the vectorial/tensorial gradient, the vectorial/tensorial curl and the scalar/vectorial divergence are uniform fields, as stated in Theorems 7–9. The calculation of the source term is immediate because the area of the dual cell is always 1/3 of the area of the whole 2-cell in 2D and 1/4 of the volume of the whole 3-cell in 3D.

In the case of a scalar field, such as the thermal field, the gradient $\mathbf{g}$ is a uniform vector field inside every primal cell. Denoting with c a cell, we denote $\mathbf{g}_c$ the gradient inside the cell and with $\mathbf{q}_c$ the corresponding heat current density vector.

If the material inside each primal cell is homogeneous and if the constitutive equation is linear, then the vector $\mathbf{q}$ is also uniform, hence it has vanishing divergence.

To write the balance equation we will use the barycentric dual (shaded area) and evaluate the heat production rate P_h generated inside the dual cell h. It is important to note that if there are also concentrated sources, such as laser spots, they can be added to the distributed sources without complications. Here the "lever rule" of FEM is not required because the heat produced by a concentrated source is entirely assigned to the dual cell in which it is produced, in full respect of the physics of the problem.

We come now to the heat current flowing through the boundary of the dual cell. Referring to the right side of Fig. 16, the heat current flowing through the piece of boundary ABC is the same as the heat current flowing through the line AC because the vector field $\mathbf{q}$ is uniform in the cell, hence it is divergence-free. This means that the total flux leaving the region $ABCA$ vanishes. This implies that to evaluate the heat current crossing the piece of the boundary contained in the primal

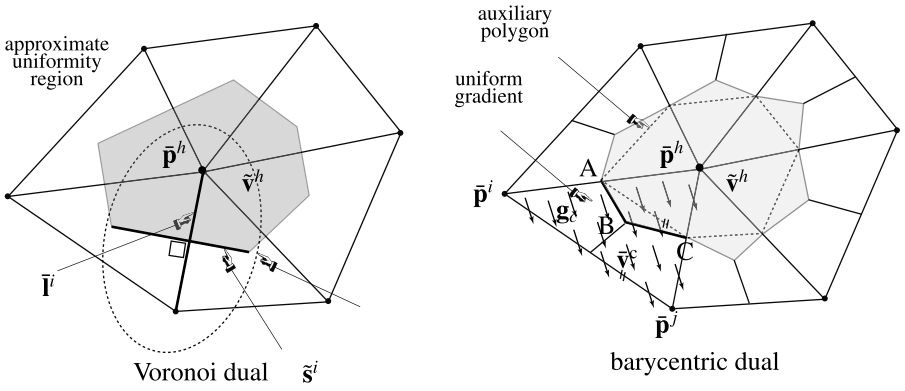
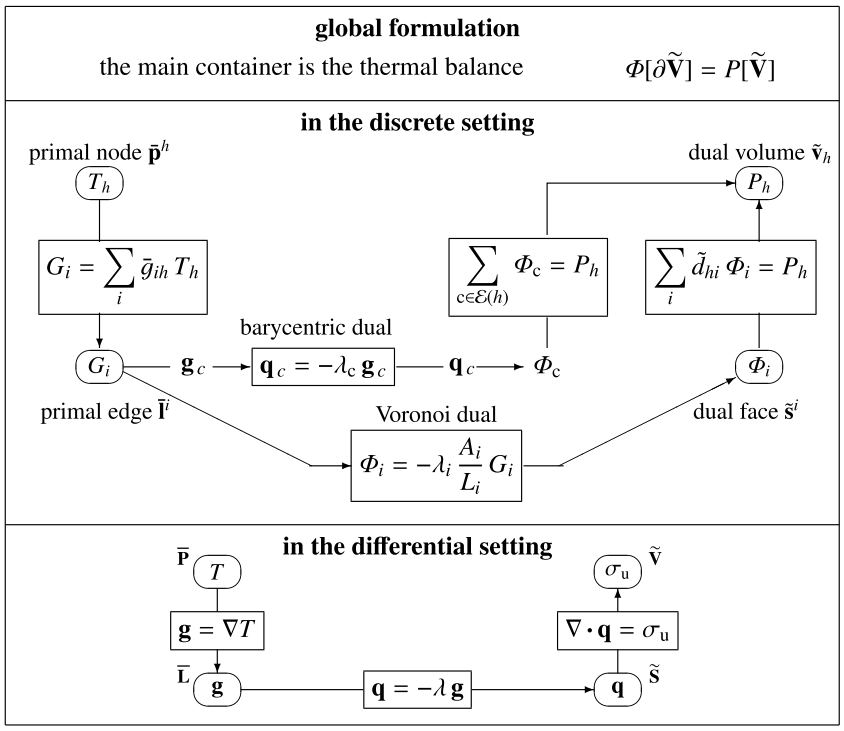


Fig. 16. A simplicial complex and a dual cell.

Table 9
Discrete formulation for stationary thermal conduction.



cell $\bar{\mathbf{v}}^c$ we can calculate the heat current crossing the side AC of the dashed polygon, which we call the *auxiliary polygon*. We must express the uniform vector $\mathbf{g}_c$ in terms of the three nodal temperatures. Subsequently, taking into account the constitutive relationship $\mathbf{q}_c = -\lambda_c \mathbf{g}_c$ we find the heat current vector $\mathbf{q}_c$. Since the material is supposed to be homogeneous inside every primal cell, also the vector $\mathbf{q}_c$ is uniform. This allows an easy calculation of the heat current Φ_c which *leaves* the dual cell through the face AC. The reader can see the details of this process for 2D and 3D problems in the paper [11]. In the same paper it is shown how to extend the affine approximation to gain the fourth order of convergence in 2D. Analogous procedure was used in [9].

We remark that using the barycentric dual the balance law becomes

$$\sum_{c\in\mathcal{E}(h)} \Phi_c = P_h \quad \text{instead of} \quad \sum_i \tilde{d}_{hi}\Phi_i = P_h \tag{17}$$

where $\mathcal{E}(h)$ denotes the set of primal 2-cells (in 2D) or of the primal 3-cells (in 3D) which share the vertex $\bar{\mathbf{p}}^h$. This means that when using the Voronoi dual we can strictly use the coboundary process to express the balance; whereas with the barycentric dual we are lead to use the affine approximation in every primal cell and then to write the balance without the incidence numbers. In both cases the physical meaning of a balance is clearly readable. Remember that, in a stationary state, the balance states that the total heat current that leaves a region is equal to the heat production rate in the region.

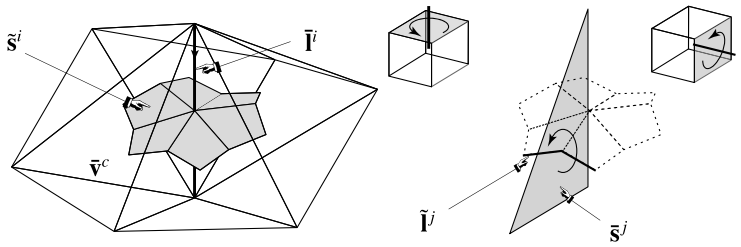


Fig. 17. Barycentric dual of a simplicial complex: (left) a primal edge and its dual face; (right) a primal face and its dual edge. The cubes show the analogous elements in a structured cell complex.

The procedure described for heat conduction is valid for all static and stationary fields whose potential is a function associated with points. Hence for electrostatics, diffusion, acoustics in fluids, irrotational fluid motion, gravitation. It is not valid for the magnetostatics whose potential is a vector **A**.

5.2. Magnetostatics

We now show how to obtain a direct algebraic formulation of magnetostatics. Since the magnetic vector potential **A** is associated with lines, as shown in Diagram ELE3, and since the corresponding global variable a_i is the line integral of **A** along the edge $\bar{\mathbf{l}}^i$ of a cell complex, we consider the scalars a_i as unknowns. The number of these unknowns is equal to the number of primal edges. The sources of a magnetic field are electric currents I_i passing through the dual cells $\tilde{\mathbf{s}}^i$. The a_i and the I_i are global variables in space. The fundamental problem of magnetostatics becomes the following: to find the a_i once the electric currents I_i are given.

To solve the problem we have at our disposal two topological laws and one constitutive law.

A first law of the magnetic field states that the total magnetic flux associated with the *boundary* of a volume, i.e. with whatever closed surface, vanishes. The law, when applied to every primal cell of a simplicial complex, states that the sum of the fluxes associated with the four faces of each simplex vanishes.

The second law states that the magnetomotive force associated with the *boundary* of any surface is equal to the electric current across the surface. We must apply this law to every face $\tilde{\mathbf{s}}^i$ of the dual complex. On account of the one-to-one correspondence between a cell $\bar{\mathbf{l}}^i$ and its dual $\tilde{\mathbf{s}}^i$, the number of resulting equations is equal to the number of the unknowns, as shown in Fig. 17.

The constitutive law in a non-magnetic medium is $\check{\mathbf{H}} = \nu\check{\mathbf{B}}$.

The procedure is the following. Starting with the variables a_i for each edge $\bar{\mathbf{l}}^i$ of the primal complex, we perform the algebraic sum of the a_i for every face, as shown in the first equation of Eq. (18). This sum gives the magnetic fluxes Φ_j associated with the primal faces $\bar{\mathbf{s}}^j$.

Remembering that the faces of the primal complex are endowed with inner orientation, denoting with A_j the area of the face $\bar{\mathbf{s}}^j$, using the screw rule, we can denote each face by a pseudovector $\check{\mathbf{A}}_j$ normal to the face.

Once the four fluxes are computed we can search for a uniform pseudovector field $\check{\mathbf{B}}_c$ inside the primal cell $\bar{\mathbf{v}}^c$ such that its fluxes $\check{\mathbf{B}}_c \cdot \check{\mathbf{A}}_j$, associated with everyone of the four 2-cells, are equal to the scalars Φ_j

$$\Phi_j = \sum_i \bar{c}_{ki} a_i, \qquad \check{\mathbf{B}}_c \cdot \check{\mathbf{A}}_j = \Phi_j. \tag{18}$$

The second equation in Eq. (18) represents four equations for three unknowns. On account of the fact that the boundary is a closed surface, we have the relation $\sum_j \check{\mathbf{A}}_j = 0$ from which we deduce $\sum_j \Phi_j = 0$. Hence we have three independent equations which allow us to obtain the pseudovector $\check{\mathbf{B}}_c$ in terms of the a_i. Inside every 3-cell $\bar{\mathbf{v}}^c$ we can introduce the vector $\check{\mathbf{H}}_c = \nu_c\check{\mathbf{B}}_c$. Hence we can evaluate the magnetomotive force F_j for each piece of the boundary of $\tilde{\mathbf{s}}^i$ which is composed of two pieces one on the left and the other on the right of the face, as shown in the right part of Fig. 17.

To do this let us consider Fig. 18. Starting with a dual 2-cell $\bar{\mathbf{s}}^i$ we localize its two cofaces using the incidence matrix $\tilde{c}_{ji}$. Let them be $\bar{\mathbf{v}}^c$ and $\bar{\mathbf{v}}^d$. With reference to Fig. 18 we must evaluate the electromotive force F_j on the two pieces forming the dual cell $\tilde{\mathbf{l}}^j$. The whole process is summarized in Table 10.

The total electromotive forces on the boundary of the dual cells are thus expressed in terms of the quantities a_k. Since the number of such circuits of the dual complex is equal to the number of primal edges of the primal complex, we have obtained a number of equations equal to the number of unknowns.

5.3. Elastostatics

Table 11 shows the procedure to obtain a direct algebraic formulation of elasticity. The reader can see the whole procedure in detail in [12] and, for elastodynamics, in [13].

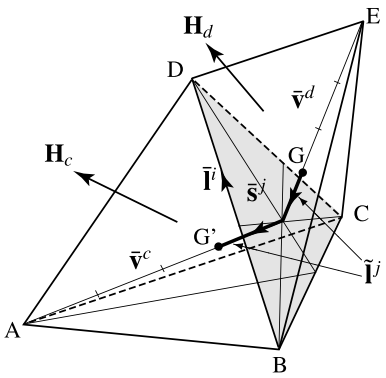
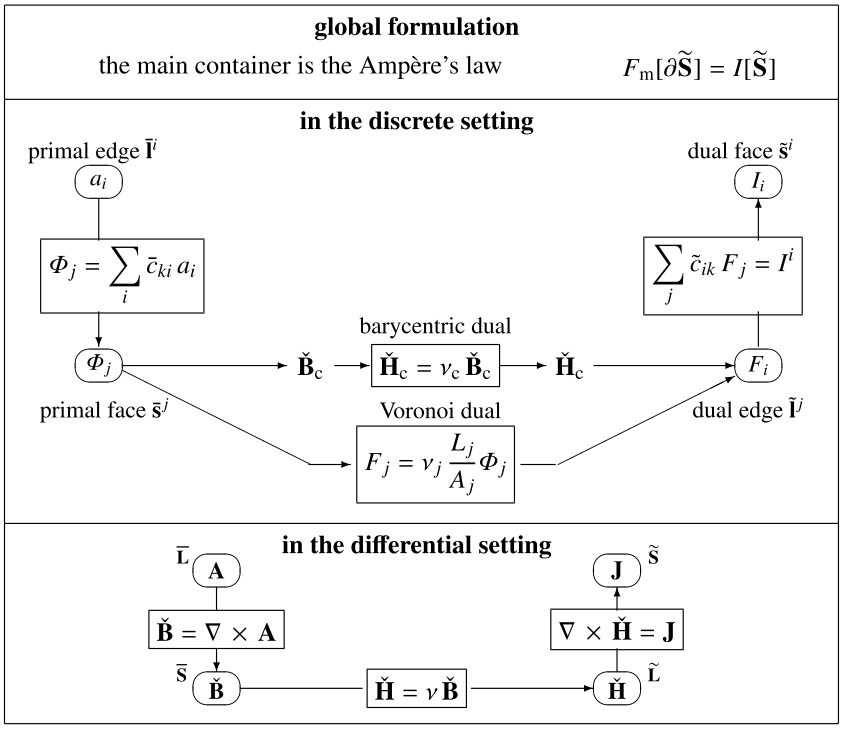


Fig. 18. The magnetomotive force evaluated on the dual edges.

Table 10
Discrete formulation for the magnetostatic field.



6. Time elements

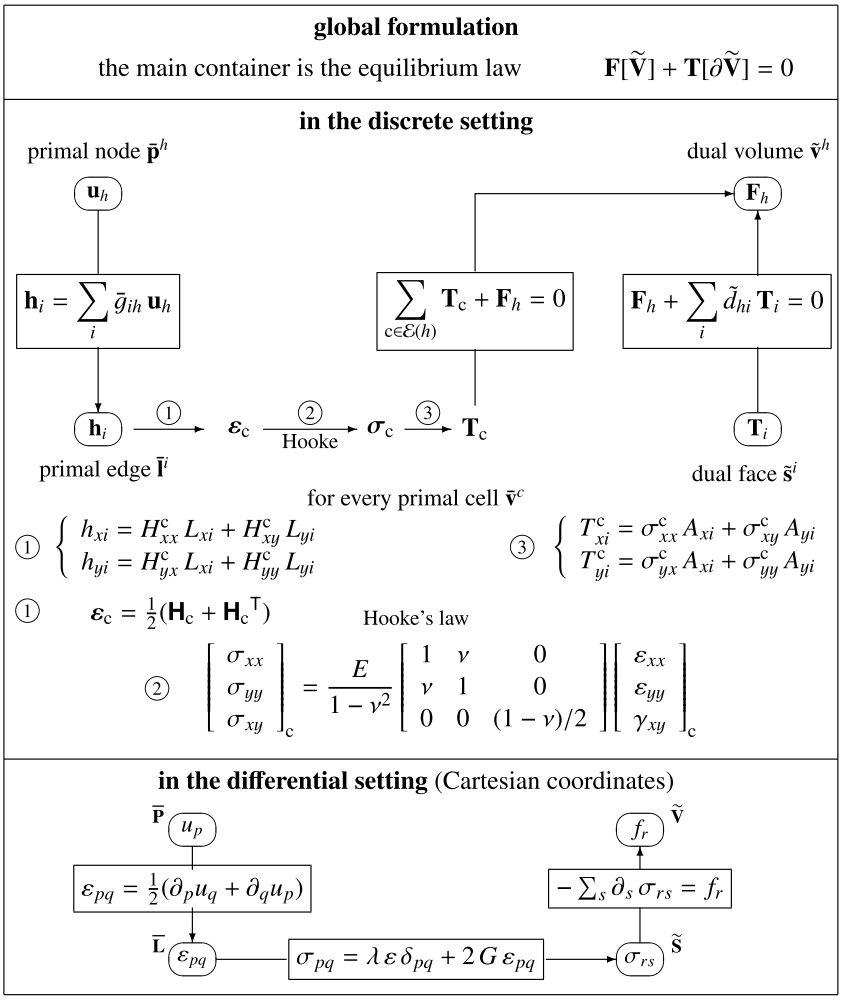
Global variables do not only refer to space elements, but also to time elements which are time instants and time intervals. In order to analyze the time elements a geometrical representation of time is appropriate. This allows us to apply geometrical concepts to time elements, in particular the notions of inner and outer orientation. If we subdivide the time axis into intervals, as shown in Fig. 19, we obtain a *cell complex in time*. This cell complex will be called *primal* and its elements will be denoted by $\dot{\mathbf{I}}$ and $\bar{\mathbf{T}}$ respectively.

With reference to Fig. 19, the *inner* orientation of the primal time intervals is the same as the orientation of the time axis: each time interval is oriented from the preceding time instant to the following one in the natural time sequence. The arrow under each primal interval is a geometrical representation of the progression of time. The inner orientation of time instants is borrowed from the inner orientation of space points, i.e. they are oriented as sinks. The *motion reversal* reverses the inner orientation of time intervals, leaving unvaried the inner orientation of time instants.

6.1. Dual time complex

By means of the geometrical representation of time, we introduce the notion of *dual complex* in time. The intermediate time instant of each interval, see Fig. 20, will be called *dual instant* and a time interval between dual consecutive time instants will be called *dual interval*. In this way we have constructed a second cell complex on the time axis which can be called the *dual* of the first: its elements are denoted by placing a tilde over the symbol, i.e. $\widetilde{\mathbf{I}}$ and $\widetilde{\mathbf{T}}$.

Table 11
Discrete formulation for mechanics of solids. (For the notation of incidence numbers see Fig. 11.)



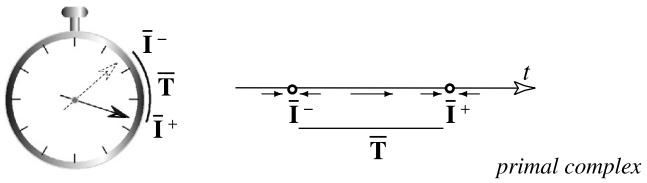


Fig. 19. Two time elements of a primal cell complex in time.

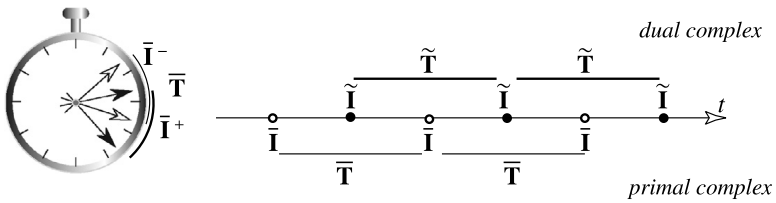


Fig. 20. A primal and a dual cell complex on the time axis: four time elements.

We are not obliged to choose the dual time instants in the middle of the primal intervals: any time instant which belongs to a primal time interval is a possible candidate for dual instant.

The reason for introducing the dual time elements becomes clear when we examine the notion of acceleration of a particle. In order to define acceleration, let us consider three time instants t_1, t_2, t_3 with the same duration τ of the time intervals between them, i.e. $\tau = t_3 - t_2 = t_2 - t_1$. By denoting the three abscissae x_1, x_2, x_3, the mean velocity and the mean

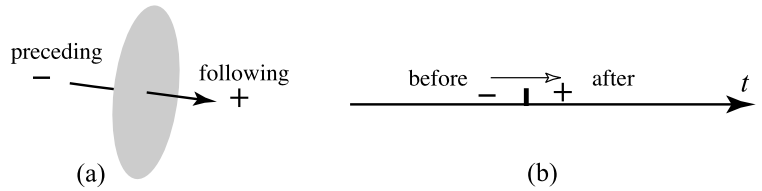


Fig. 21. (a) The outer orientation of a surface in space compared with (b) the outer orientation of a time interval.

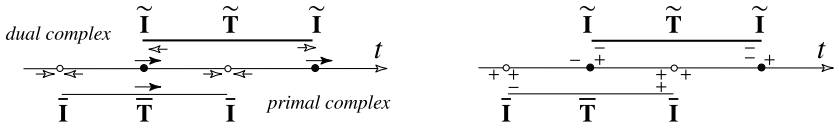


Fig. 22. The two types of orientation of time elements (a) using arrows and (b) using signs.

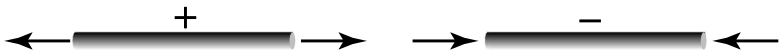


Fig. 23. This is the convention used in the books on strength of materials.

acceleration are given by

$$v_{\mathrm{m}} = \frac{x_2 - x_1}{\tau} \qquad v'_{\mathrm{m}} = \frac{x_3 - x_2}{\tau} \qquad a_{\mathrm{m}} = \frac{v'_{\mathrm{m}} - v_{\mathrm{m}}}{\tau} = \frac{x_3 - 2x_2 + x_1}{\tau^2}, \tag{19}$$

respectively. From these formulas we see that whereas velocities change sign under motion reversal, acceleration does not change sign. In fact, by inverting the time sequence t_1, t_2, t_3, we have the sequence t_3, t_2, t_1 and the mean acceleration remains the same despite the change $x_1 \Leftrightarrow x_3$.

The fact that acceleration does not change its sign under motion reversal implies that it cannot be associated with primal time intervals which are endowed with inner orientation, otherwise the *oddness principle* would be violated. This leads us to consider intervals not endowed with inner orientation. Considering that the mean acceleration makes use of the middle time instants of two consecutive intervals, we see that the time intervals with which it is associated are the dual ones.

6.2. Outer orientation of dual time elements

We have seen that, in a cell complex in space the inner orientation of a p-dimensional cell of the primal complex induces an outer orientation on its dual cell. It is therefore natural to define as the outer orientation of the dual time elements the one induced by the inner orientation of the primal time elements, as shown in Fig. 22.

Let us examine the meaning of the arrow which goes across the dual instant $\widetilde{\mathbf{I}}$. Referring to Fig. 21 let us compare the outer orientation of a surface embedded in space and the dual instant on the time axis. What does *outer orientation* mean? In the case of a surface it means the choice of one of the two sides of the surface as the preceding one and the other as the following one. It is natural to indicate the preceding part with a minus sign and the following part with a plus sign. It is also natural to use an arrow which goes from minus to plus. The arrow crosses the surface.

In the time axis it is natural to consider the instants which precede and those which follow a given dual instant and consider these two sets in this order, i.e. to talk about *before* and *after* the instant. It is also natural to assign a minus sign to *before* and a plus sign to *after*. In accordance with the rule used up to now we can draw an arrow from the minus sign to the plus sign, i.e. from *before* to *after*. So the arrow which goes across the dual instant simply means the order from before to after or from after to before.

What type of orientation can be assigned to the dual intervals? The inner orientation of primal time intervals induces an outer orientation on the dual time instants $\widetilde{\mathbf{I}}$, that can be represented by an arrow which "crosses" the instant, see Fig. 21(b). Instead of the arrows we can use the "+" and "−" signs.

Let us raise the question: can we call a time interval endowed with arrows directed outside or inside the interval *oriented*? Let us remember that in the theory of bars we may consider positive the traction and negative the compression, as shown in Fig. 23. The notion of "orientation" arises whenever one attribute has two opposite states [10].

The outer orientation of dual intervals, induced by the inner orientation of primal time instants, is represented by arrows pointing toward the interior of the interval, as a bar under compression. We stated that the outer orientation of a volume induced by the inner orientation of a point (sink) was opposite to the orientation currently used in physics, i.e. the orientation with outwards normal. Hence, to agree with existing conventions, we take as outer orientation of volumes the *opposite* of the orientation induced by the primal points.

Therefore we will do the same here for outer orientation of intervals: whereas primal time instants induce an orientation on the dual time intervals with the arrows directed inwards, we choose to direct the arrows outwards, i.e. opposite to the

orientation induced by the primal time instants, as in Fig. 22. This inversion of sign is in agreement with the formula which defines the acceleration.

When time instants and time intervals are represented geometrically, as in Fig. 22 the arrows give a geometrical meaning to the time orientation. In fact in Eq. (19) we have used the relation

$$a_{\mathrm{m}} = \frac{(+1)v'_{\mathrm{m}} + (-1)v_{\mathrm{m}}}{\tau} \tag{20}$$

in which the coefficients $(+1)$ and (-1) can be interpreted as incidence numbers between the time interval with outgoing arrows and the time instants with arrows crossing them, as shown in Fig. 22.

6.3. Global variables in time

By a careful examination of the time dependence of physical variables we are led to extend the notion of *association* of a physical variable also to oriented time elements, i.e. with one of the four elements $\dot{\mathbf{I}}, \bar{\mathbf{T}}, \widetilde{\mathbf{I}}, \widetilde{\mathbf{T}}$. To do this we begin by noting that every physical variable refers to either an instant or an interval.

The majority of physical variables used in practice are referred to the instants, i.e. they are *function of* instants: this is the case of temperature, velocity, pressure, mass density, momentum, electric potential, electric field vector, electric current, force, stress, strain, stream function, power emitted, etc. This is what we usually mean when we write $T(t)$, $\mathbf{v}(t)$, $p(t)$, $\rho(t)$, $\mathbf{p}(t)$, etc.

A smaller number of variables makes reference to the intervals: this is the case of the displacement of a particle, of the work given to a system, the heat exchanged between two systems, the electric charge stored in a battery, the impulse given to a body, the production of entropy, flows of energy, entropy, mass, etc.

This suggests that the global physical variables can be divided into two classes: those associated with instants and those associated with time intervals. This is indeed the case, but we must make an important remark: *among the physical variables which refer to instants, there are some which require a time interval for their definition and their measurements*. This happens with those variables which are *rates* of variables referred to intervals.

So the mass content M^{c} in a given volume and the power P emitted by an energy generator are both referred to a time instant. In spite of this, in order to define and to measure the power we need a small time interval to measure the work given by the generator, whereas the measurement of the mass content does not involve a time interval. For this reason power makes reference to an instant after performing a rate (work/duration) and a limit (duration $\to$ 0), whereas mass content is attributed directly to an instant. This shows that although both variables refer to instants, power needs a time interval for its definition and for its measurement. For this reason we will say that power is *associated* with a time interval, like the work from which it derives. This distinction prompts us to introduce the following

Definition. We call *global variable in time* every variable which is not the rate of another variable.

6.4. Time association of physical variables

The impulse of a force $\mathbf{J}$ makes reference to time intervals. Since the motion reversal does not invert the impulse (think about the impulse that a ball gives to a wall on a rebound) [10], hence it cannot be associated with primal intervals. It must be associated with the dual time intervals, i.e. $\mathbf{J}[\widetilde{\mathbf{T}}]$. It follows that, since the force is the rate of the impulse, it *inherits* the association with dual time intervals, $\mathbf{F}[\widetilde{\mathbf{T}}]$.

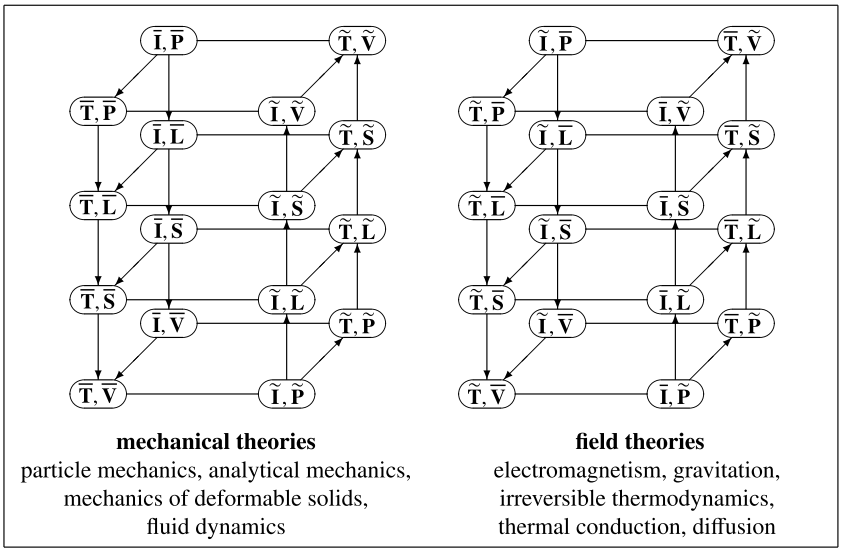
In contrast the displacement $\mathbf{u}$, which makes reference to a time interval, changes sign under motion reversal and for this reason it is associated with primal time intervals, i.e. $\mathbf{u}[\bar{\mathbf{T}}]$. The velocity, rate of displacement, inherits this inversion of sign, hence $\mathbf{v}[\bar{\mathbf{T}}]$. The momentum $\mathbf{p}(t)$ makes reference to instants, it is not the rate of another variable and changes sign under motion reversal, hence $\mathbf{p}[\widetilde{\mathbf{I}}]$.

These associations permeate all mechanical theories, such as particle mechanics, analytical mechanics, mechanics of solids, fluid mechanics. The result is that the configuration variables of mechanics are associated with time elements endowed with inner orientation whereas the source variables are associated with outer orientation, as it happens for the space association.

In contrast, in field theories, like electromagnetism, gravitation, irreversible thermodynamics, heat conduction and diffusion, the association with time elements is inverted: the configuration variables are associated with time elements endowed with *outer* orientation and source variables with time elements endowed with *inner* orientation. We can see this starting with the electric field strength $\mathbf{E}(t)$ which is a configuration variable defined as the ratio force/charge, $\mathbf{E} \stackrel{\mathrm{def}}{=} \mathbf{F}[\widetilde{\mathbf{T}}]/q$. Since force and charge do not change sign under motion reversal, it follows that $\mathbf{E}$ is also invariant, hence $\mathbf{E}[\widetilde{\mathbf{T}}]$. The same association is valid for the electromotive force $E[\widetilde{\mathbf{T}}]$, which is the line integral of $\mathbf{E}$, and for the electric potential $\phi[\widetilde{\mathbf{T}}]$. On the contrary, the electric charge flow Q^{f}, a source variable, changes sign on motion reversal and is associated with time intervals, hence $Q^{\mathrm{f}}[\bar{\mathbf{T}}]$. The electric current density inherits this behavior, hence $\mathbf{J}[\bar{\mathbf{T}}]$. The electric flux Ψ, like the electric charge content Q^{c} are associated with instants and do not change sign under motion reversal, hence $Q^{\mathrm{c}}[\dot{\mathbf{I}}]$.

The association of global physical variables with primal and dual time elements leads in a natural way to the leap-frog algorithm for the time integration.

Table 12
The two classification diagrams for physical variables.



mechanical theories
particle mechanics, analytical mechanics, mechanics of deformable solids, fluid dynamics

field theories
electromagnetism, gravitation, irreversible thermodynamics, thermal conduction, diffusion

6.5. A classification diagram for physical variables

The $(2+2)$ time elements and the $(4+4)$ space elements can be combined to give $(2+2)\times(4+4)=32$ space–time elements. These space–time elements can be divided into two classes of 16 elements, as shown in the classification diagram of Table 12. The important result is that *of about* 180 *physical variables of various physical theories we have examined, all of them are associated with one of these space–time elements* [10]. This gives rise to a new classification of physical variables, both global variables and field functions. Such classification can be exhibited in a diagram, called *classification diagram of physical variables*.

The explicit use of the dual cell complex in time, required by the fact that the source variables are associated with the elements of one complex in time whereas the configuration variables are associated with the elements of the corresponding dual complex. This permits to avoid the cumbersome notation originating from the notation $A^{n+1/2}$ which gives rise, in many cases, to formulas which are really ... monstrous! We suggest to consider physical variables, not equations, as the main pillars of a computational procedure and to decompose the fundamental equation into its original constituents. These are: two topological equations and one constitutive equation. In some cases the constitutive equations are two or at most three, as the diagram of particle mechanics shows (see the diagram PAR1 in [15]).

7. Fluid dynamics

7.1. Some peculiarities of the fluid field

A first peculiarity lies in the fact that in some cases the velocity is associated with dual instants and primal lines, i.e. $\mathbf{v}[\widetilde{\mathbf{I}},\bar{\mathbf{L}}]$, as shown in Tables [FLU3] whereas in other cases it is associated with primal time intervals and primal points, i.e. $\mathbf{v}[\bar{\mathbf{T}},\bar{\mathbf{P}}]$, as shown in Tables [FLU6]. This double association of the fluid velocity makes the velocity unique among the approximately 180 physical variables of the many physical theories which are classified in the book [10]. How to explain this ambiguity?

Well, this double association seems to be the consequence of the double description used in fluid dynamics. When we consider a fluid particle $\mathcal{P}$ in the material description, the velocity $\mathbf{v}(t,\mathcal{P})$ is associated with a time interval, hence $\mathbf{v}[\bar{\mathbf{T}},\mathcal{P}]$. In the transition to the spatial description, we decide to consider the point $\mathbf{P}$ through which the particle $\mathcal{P}$ goes at the instant t and to assign its velocity to the point. This shift *particle* $\rightarrow$ *point* implies forgetting the history of the particle motion and considering the velocity as a field vector applied to the point $\mathbf{P}$ at the instant t instead of to the particle $\mathcal{P}$. The association with time is changed and the velocity is associated with the instant considered.

However, since velocity, on account of its physical meaning, must change sign under motion reversal, it follows that the instant must be the dual. Hence the passage from the material to the spatial description implies not only the passage $\mathcal{P}\rightarrow\mathbf{P}$ but also $\mathbf{v}[\bar{\mathbf{T}}]\rightarrow\mathbf{v}[\widetilde{\mathbf{I}}]$.

Once the velocity is considered as a vector *field* it is natural to perform the integral along lines. The same happens to the electric field vector $\mathbf{E}$, and to the gravitational acceleration $\mathbf{g}$, to a force $\mathbf{F}$ in a force field. It follows that, in the spatial description, velocity refers to lines, i.e. $\mathbf{v}[\widetilde{\mathbf{I}},\bar{\mathbf{L}}]$. Hence velocity has two possible space and time associations (Table 13).

There is another feature of fluid motion which is linked to the passage between the two descriptions. This feature can be grasped by addressing the following question: *What is the meaning of the product* $\rho\mathbf{v}$? We can see that there are two possible answers to this question

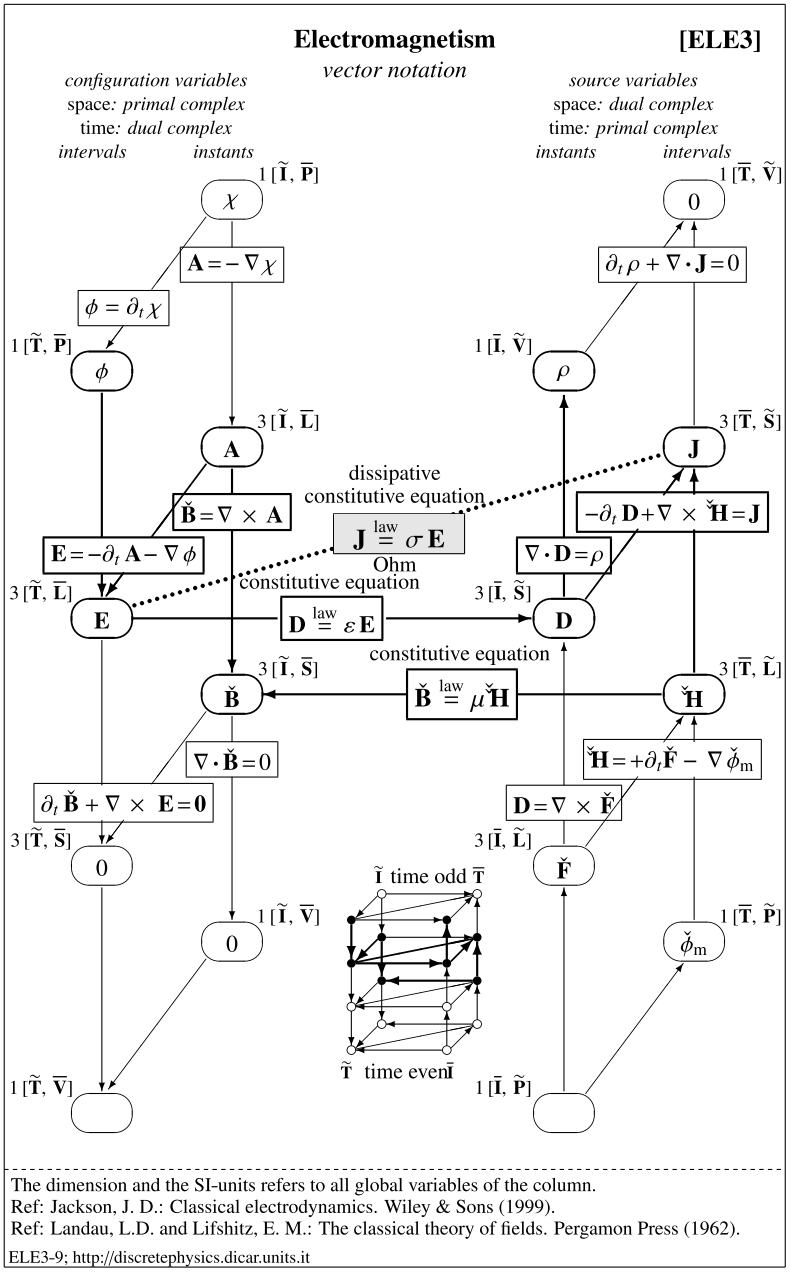
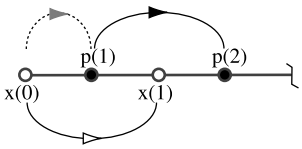


Table 13
The leap-frog algorithm for the harmonic oscillator.

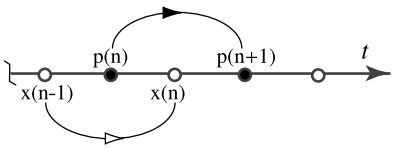


assigned $x(0)$ and $p(1)$
assigned m, k, τ

$$x(1) = x(0) + \tau\left[\frac{1}{m}\, p(1)\right]$$

$$p(2) = p(1) + \tau\left[-k\, x(1)\right]$$

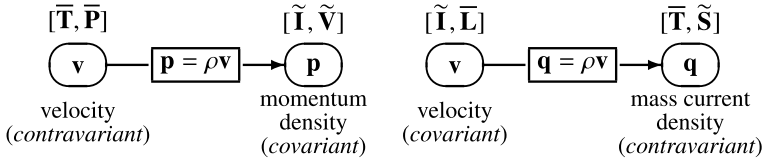
obtained $x(1)$ and $p(2)$
and so on ...



assigned $t(0), x(0), p(1), m, k, \tau, N$

for $n = 1$ to $(N-1)$
$t(n) = t(n-1) + \tau$
% backward difference
$$x(n) = x(n-1) + \tau\left[\frac{1}{m}\, p(n)\right]$$
% forward difference
$p(n+1) = p(n) + \tau\left[-k\, x(n)\right]$
end for

Table 14
The two possible associations of velocity in fluid dynamics.



- it can be the *momentum density*, denoted by $\mathbf{p} = \rho\mathbf{v}$. In this case it is associated with volumes and time instants: $\mathbf{p}[\widetilde{\mathbf{I}}, \widetilde{\mathbf{V}}]$. The association with $\widetilde{\mathbf{I}}$ is required by the fact that momentum, like velocity, changes sign under motion reversal; the association with $\widetilde{\mathbf{V}}$ is due to the fact that we must consider the momentum of the fluid contained in a volume endowed with outer orientation.
- it can be the *mass current density*, denoted by $\mathbf{q} = \rho\mathbf{v}$. In this case it is associated with surfaces and time intervals: $\mathbf{q}[\overline{\mathbf{T}}, \widetilde{\mathbf{S}}]$. The association with $\overline{\mathbf{T}}$ is required by the fact that mass current density, like velocity, changes sign under motion reversal; the association with the surfaces endowed with outer orientation is due to the fact that we must consider the mass current through a surface, hence a surface endowed with outer orientation.

As we see, also the variable $\rho\mathbf{v}$ has two different interpretations: this ambiguity is linked to the ambiguity of the velocity of which we have just dealt with, as shown in Table 14. In the first case the dual of $[\widetilde{\mathbf{I}}, \widetilde{\mathbf{V}}]$ is $[\overline{\mathbf{T}}, \overline{\mathbf{P}}]$, hence $\mathbf{v}[\overline{\mathbf{T}}, \overline{\mathbf{P}}]$. In the second case the dual of $[\overline{\mathbf{T}}, \widetilde{\mathbf{S}}]$ is $[\widetilde{\mathbf{I}}, \overline{\mathbf{L}}]$, hence $\mathbf{v}[\widetilde{\mathbf{I}}, \overline{\mathbf{L}}]$. It is interesting to remark that whereas momentum is a covariant vector, velocity must be a contravariant vector. In the second case, since the mass current density is a contravariant vector, the velocity must be a covariant vector.

Since the velocity is the rate of a displacement, which is a contravariant vector, it follows that velocity is also a contravariant vector. In order to reduce the velocity to a covariant vector we must make use of the metric tensor: $v_h = g_{hk} v^k$.

In the first case, in which $\mathbf{v}[\overline{\mathbf{T}}, \overline{\mathbf{P}}]$, the velocity is the time rate of a global variable in time and is a global variable in space; in the second case, in which $\mathbf{v}[\widetilde{\mathbf{I}}, \overline{\mathbf{L}}]$, it is a global variable in time and a line density in space.

When velocity is a covariant vector, as in Table [FLU3], we may perform its line integral along a line, like the vector $\mathbf{E}$ of electromagnetism: this testifies that velocity is associated with lines. The velocity line integral is homologous to the voltage in electrostatics. Moreover, when the velocity line integral along all reducible closed lines vanishes we may introduce the velocity potential ϕ; the homologous variable in electrostatics is the electric potential ϕ. We have in both theories $\phi[\widetilde{\mathbf{I}}, \overline{\mathbf{P}}]$: compare Table [FLU3] with Table [ELE3].

8. Conclusions

The tradition to describe mathematically physical theories in terms of differential equations, which lasted for three centuries, hinders the consideration of other possible descriptions. The addiction is such that, since the digital computer requires an algebraic formulation of physical laws, it was preferred to discretize the differential equations, rather than considering other more convenient tools for the mathematical description of physics.

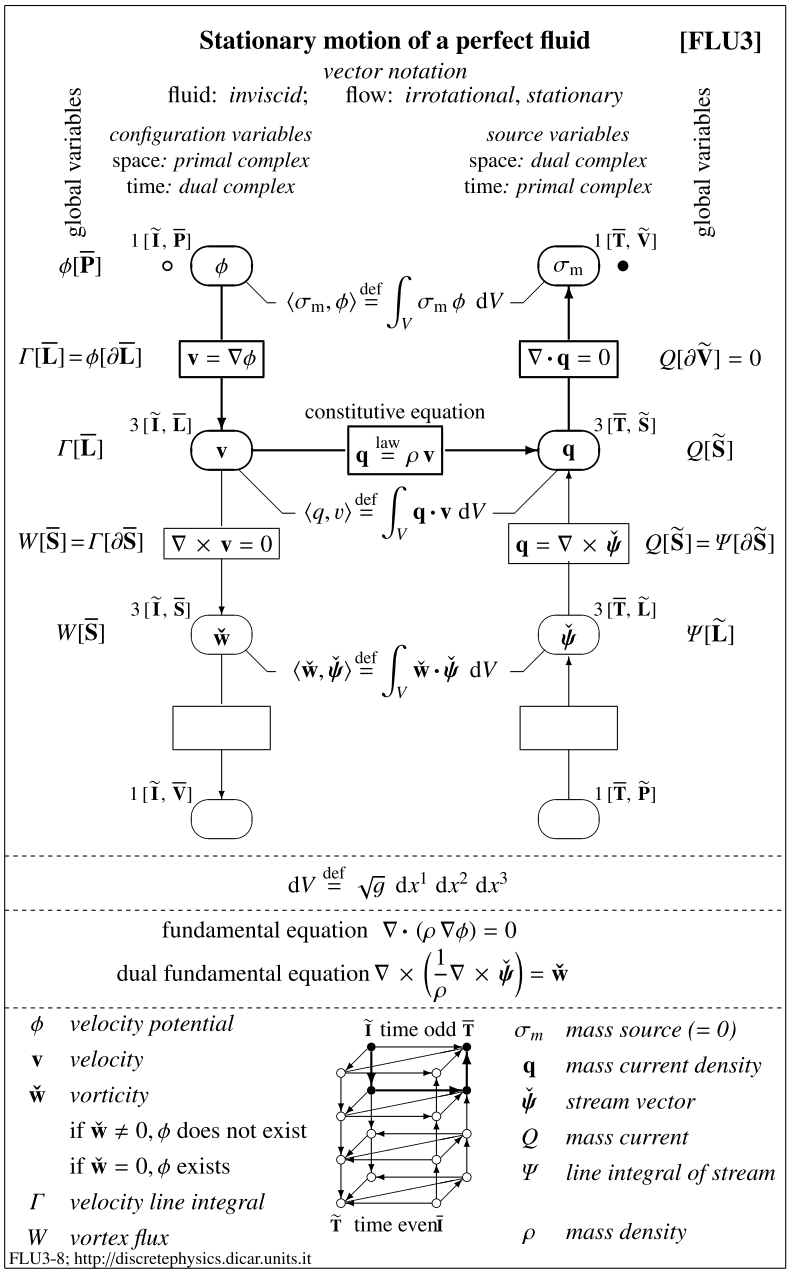
In a interesting paper by Desbrun, Hirani, Leok and Marsden [1], it is stated:

> "*Our vision of this theory is that it should proceed* ab initio *as a discrete theory that parallels the continuous one.... if one takes the point of view that the discrete theory can, and indeed should, stand in its own right, then the range of application areas naturally is enriched and increases. Convergence and consistency considerations alone are inadequate to discriminate between the various choices of discretization available to the numerical analyst, and only by requiring, when appropriate, that the discretization exhibits discrete analogues of continuous properties of interest can we begin to address the question of what makes a discrete theory a canonical discretization of a continuous one*."

We have shown that it is possible to give *ab initio* an algebraic description of physics by using the natural association of global physical variables with the oriented space elements and by using the rudiments of algebraic topology [4,7].

Computational physics has evolved under the pressing demand to improve the accuracy and the stability of the solution, the convergence to the exact value, etc. So, in the course of history, *staggered grids* and *leap-frog algorithms*, were introduced.

In this development the physical aspect has often been ignored. The role of global variables, so common in physical measurements, does not seem to be considered in the discrete formulation, where it is preferred to obtain the integral variables by means of integration of the field functions. The notion of orientation, in its two forms, inner and outer, is not used in numerical methods simply because it is not commonly used in geometry and topology! At such a point, that in the theory of exterior differential forms, some authors are disturbed by the presence of *twisted* differential forms whereas they arise for those variables which are associated with space elements endowed with *outer* orientation, i.e. for source variables. We are indebted to Schouten [8] for his systematic presentation of the two orientations. We mention the paper of Weiland [16] on the *Finite Integration Theory* (FIT), dealing with the electromagnetic field, in which are used both the integral variables and the distinction between the inner and outer orientation of the space elements.



The natural association of global physical variables with oriented spatial elements appears not to be discussed in literature. The same is for the distinction between configuration and source variables, with the corresponding association with spatial elements endowed with inner and outer orientations. *This lack of distinction has prevented to put into evidence the role of the dual cell complex which, on the contrary, appears to be a basic requirement in the mathematical description of physics*. Hence the staggered grids, introduced for purely mathematical reasons, play a fundamental role in the same mathematical description of physics. All these notions show that computational physics is richer than computational mathematics (Table 15).

In recent decades it has been realized that the theory of exterior differential forms offers a more complete description of some physical theories because it recovers the geometrical content which the traditional differential formulation hides.

In spite of this progress the theory of exterior differential forms maintains the inconveniences linked to its differential nature. This suggests that instead of discretizing the exterior differential forms by introducing the *discrete forms*, i.e. *cochains*, it is much simpler to start with the *discrete forms* that assign an amount of each global physical variable with every p-cell of a cell complex.

We state that are the global physical variables, not the differential equations, that must be the pillars of computational physics. The use of global variables allows us to avoid the integration of field variables on lines, surfaces and volumes. This is possible only if we strictly follow the association of the configuration and source variables with the space elements endowed with inner and outer orientation respectively.

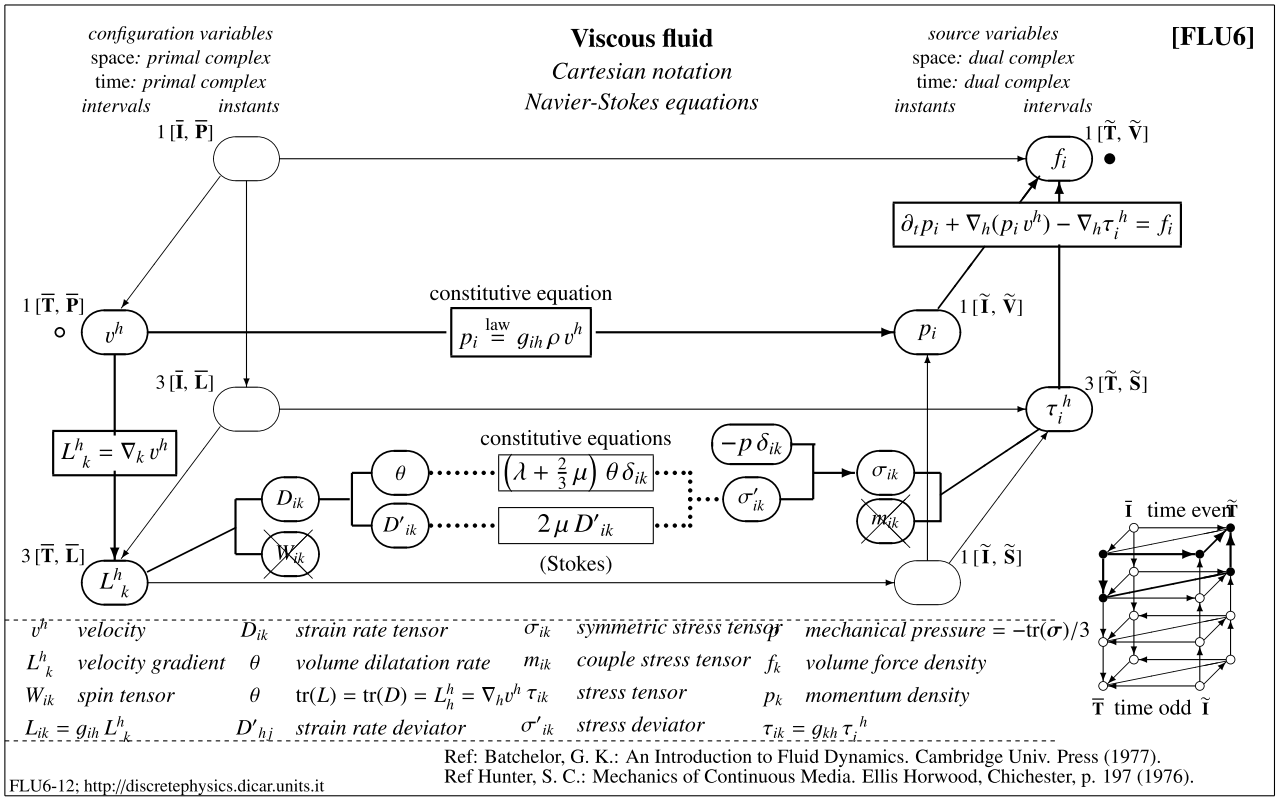


Table 15
The topological laws linking the source variables of various theories.

Electromagnetism	
Magnetic Gauss' law	$\Phi[\widetilde{\mathbf{I}}, \partial\overline{\mathbf{V}}] \overset{\text{law}}{=} 0$
Electric Gauss' law	$\Psi[\overline{\mathbf{I}}, \partial\widetilde{\mathbf{V}}] \overset{\text{law}}{=} Q^{\mathrm{c}}[\overline{\mathbf{I}}, \widetilde{\mathbf{V}}]$
Faraday's law	$\underbrace{\mathcal{E}[\widetilde{\mathbf{T}}, \partial\overline{\mathbf{S}}]}_{\text{electrom. force}} + \underbrace{\left\{\Phi[\widetilde{\mathbf{I}}^+, \overline{\mathbf{S}}] - \Phi[\widetilde{\mathbf{I}}^-, \overline{\mathbf{S}}]\right\}}_{\text{magnetic flux variation}} \overset{\text{law}}{=} 0$
Ampère-Maxwell's law	$\underbrace{\mathcal{F}_{\mathrm{m}}[\overline{\mathbf{T}}, \partial\widetilde{\mathbf{S}}]}_{\text{magnetom. force}} - \underbrace{\left\{\Psi[\overline{\mathbf{I}}^+, \widetilde{\mathbf{S}}] - \Psi[\overline{\mathbf{I}}^-, \widetilde{\mathbf{S}}]\right\}}_{\text{electric flux variation}} \overset{\text{law}}{=} \underbrace{Q^{\mathrm{f}}[\overline{\mathbf{T}}, \widetilde{\mathbf{S}}]}_{\text{electric charge flow}}$
Continuum mechanics	
Mass conservation law	$\underbrace{M^{\mathrm{f}}[\overline{\mathbf{T}}, \partial\widetilde{\mathbf{V}}]}_{\text{mass outflow}} + \underbrace{M^{\mathrm{c}}[\overline{\mathbf{I}}^+, \widetilde{\mathbf{V}}] - M^{\mathrm{c}}[\overline{\mathbf{I}}^-, \widetilde{\mathbf{V}}]}_{\text{mass storage}} \overset{\text{law}}{=} 0$
Momentum balance law	$\underbrace{\mathbf{P}^{\mathrm{f}}[\widetilde{\mathbf{T}}, \partial\widetilde{\mathbf{V}}]}_{\text{momentum outflow}} + \underbrace{\mathbf{P}^{\mathrm{c}}[\widetilde{\mathbf{I}}^+, \widetilde{\mathbf{V}}] - \mathbf{P}^{\mathrm{c}}[\widetilde{\mathbf{I}}^-, \widetilde{\mathbf{V}}]}_{\text{momentum storage}} \overset{\text{law}}{=} \underbrace{\mathbf{J}^{\mathrm{s}}[\widetilde{\mathbf{T}}, \partial\widetilde{\mathbf{V}}]}_{\text{surface impulse}} + \underbrace{\mathbf{J}^{\mathrm{v}}[\widetilde{\mathbf{T}}, \widetilde{\mathbf{V}}]}_{\text{volume impulse}}$
Particle mechanics	
Momentum balance	$\mathbf{p}[\widetilde{\mathbf{I}}^+] - \mathbf{p}[\widetilde{\mathbf{I}}^-] \overset{\text{law}}{=} \mathbf{J}[\widetilde{\mathbf{T}}]$

References

[1] M. Desbrun, A.N. Hirani, M. Leok, J.E. Marsden, Discrete exterior calculus, http://arxiv.org/abs/math/0508341, 2005.
[2] M. Desbrun, E. Kanso, Y. Tong, Discrete differential forms for computational modeling, in: I. Bobenko, et al. (Eds.), Discrete Differential Geometry, Birkhäuser, 2008, pp. 287–323.
[3] J. Frauendiener, Discrete differential forms in general relativity, Class. Quantum Gravity 23 (2006) S369–S385.
[4] L.J. Grady, J.R. Polimeni, Discrete Calculus, Applied Analysis on Graphs for Computational Science, Springer, London, 2010.
[5] A.N. Hirani, Discrete exterior calculus, Dissertation (Ph.D.) Caltech, 2003, http://thesis.library.caltech.edu/1885/.
[6] J.G. Hocking, G.S. Young, Topology, Addison-Wesley, Reading, MA, 1961.

[7] C. Mattiussi, An analysis of finite volume, finite element, and finite difference methods using some concepts from algebraic topology, J. Comput. Phys. 133 (1997) 289–309.
[8] J.A. Schouten, Tensor Calculus for Physicists, 2nd ed., Dover, New York, 1989.
[9] V. Subramanian, J.B. Perot, Higher-order mimetic methods for unstructured meshes, JCP 219 (1) (2006).
[10] E. Tonti, The Mathematical Structure of Classical and Relativistic Physics, Birkhäuser-Springer, 2013, in press.
[11] E. Tonti, A direct discrete formulation of field laws: the cell method, CMES Comput. Model. Eng. Sci. 2 (2) (2001) 237–258.
[12] E. Tonti, F. Zarantonello, Algebraic formulation of elastostatics: the cell method, CMES Comput. Model. Eng. Sci. 39 (3) (2009) 201–236.
[13] E. Tonti, F. Zarantonello, Algebraic formulation of elastodynamics: The cell method, CMES Comput. Model. Eng. Sci. 64 (1) (2010) 37–70.
[14] E. Tonti, E. Nuzzo, Gradiente Rotore Divergenza, Pitagora Editrice, Bologna, 2007. The book is under translation into English.
[15] E. Tonti, http://discretephysics.dica.units.it/.
[16] T. Weiland, Time domain electromagnetic field computation with finite difference methods, Int. J. Numer. Model. 9 (1996) 295–319.
[17] P. Wesseling, Principles of Computational Fluid Dynamics, Springer, 2001.
[18] K.S. Yee, Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, IEEE Trans. A.P. Ap-14 (3) (1966).