2

Basic mathematical models and motivations

Luca Formaggia, Karl Perktold, and Alfio Quarteroni

As explained in Chapter 1, the cardiovascular system is the main responsible of the transport of various chemicals to and from the various organs, enabling their correct functioning and, in fact, life.

The desire to model this system is longstanding. Indeed, the well known Euler equations (nowadays standing at the ground of gas dynamics models) were infact developed by Euler in 1775 with the intent of describing blood flow in the human arteries [139]. We mention also the works by Bernoulli, Poiseuille and Young on this subject.

However, it is only in the past few decades that the application of mathematical models of the cardiovascular system have become widespread within the bioengineering and medical research community. The main reasons are the advancements in the power of modern computers, the progress in imaging and geometry extraction techniques (see Chapter 4) as well as the development of better numerical algorithms (like the ones described in the later chapters of this book).

Nowadays, computer simulations can provide researchers with an invaluable tool for the interpretation and analysis of the circulatory system functionality, in both physiological and pathological situations.

Clearly a main impulse to develop this field of study is the increasing demand from the medical community for scientifically rigorous and quantitative investigations of cardiovascular diseases, which are unfortunately responsible for a large percentage of early mortality in industrialised societies, see for instance [229]. The ageing of the population and the consequent increase of health care costs also call for more effective treatments.

Besides their employment in medical research, numerical models of vascular flows can also provide a virtual experimental platform to be used as training system for new vascular surgeons [494] or anaesthesiologists [91,357]. In perspective, they can give specific design indications for the realisation of surgical operations [331,494] or for the design of better prosthetic devices. For instance, numerical studies have shown how shape optimisation techniques may be used for minimising the downstream vorticity in coronary by-pass

Formaggia L, Quarteroni A, Veneziani A (Eds.): Cardiovascular Mathematics. Modeling and simulation of the circulatory system

grafts [1, 434]. Investigations of this type can help the surgeon in understanding how different surgical solutions may affect blood circulation and guide the choice of the most appropriate procedure for a specific patient or type of patients. Another interesting application of computational haemodynamics to surgical planning is found in [329].

The mathematical modelling of the various functions of the cardiovascular system is, however, still an incredibly challenging problem. The difference in space and time scale of the processes involved, highlighted in the previous chapter, makes the treatment of the system as a whole unfeasible. It is then useful to identify a hierarchy of models, each suited for a different type of investigation or to different parts of the system, see Chapter 10, and possibly devise strategies to couple them, using for instance the multiscale framework that will be illustrated in Chapter 11.

In the following we will focus our attention on models for the main systemic cardiovascular tree, providing for each of them a justification from a physical and computational point of view. A more formal derivation of the equations is postponed to other chapters of the book, in particular Chapters 3 and 7.

Numerical simulations are of course less invasive than *in vivo* investigation, and potentially more accurate and flexible than *in vitro* experiments. Numerical models require patients data: the value of the parameters characterising the properties of blood and possibly the vessel wall, the initial and boundary conditions for the partial differential equations to be solved as well as geometrical data that defines the shape of the computational domain. The latter can be obtained by radiological acquisition through, e.g., computer tomography, magnetic resonance, Doppler anemometry, etc., as will be addressed in Chapter 4.

2.1 Mathematical models for local blood flow dynamics

The mathematical equations of fluid dynamics are the key components of haemodynamics modelling. Rigorously speaking blood is not a fluid but a suspension of particles in the *plasma*, the latter being mainly made of water. As discussed in Chapter 1, the most important blood particles are red cells (erythrocytes), white cells (leukocytes), and platelets (thrombocytes). Being the most numerous, red cells are the main responsible for the special mechanical properties of blood. The prominent macroscopic effect of their presence is that blood is a *shear-thinning*, or thixotropic fluid. A precise definition will be given in Chapter 6, here we just say that a shear-thinning fluid the more it stirs the more it fluidifies (just think to the behaviour of tomato ketchup, another shear-thinning fluid). In other words, its (apparent) viscosity decreases with the increase of the rate of deformation. This effect is stronger in smaller vessels, like the arterioles, venules and the capillaries. Viscoelastic effects can be very important at the fine spatial scale (micro-circulation). Below a critical vessel calibre (about 1 mm), blood viscosity becomes dependent on the vessel radius and decreases very sharply. This is known as Fahraeus-Lindquist

effect (see Chapter 1): red blood cells move to the central part of the capillary, whereas the plasma stays in contact with the vessel wall. This layer of plasma facilitates the movement of the red cells, thus causing a decrease of the apparent viscosity. High shear rate and increased blood cell deformation are further important factors that explain viscoelastic behaviour.

Things get even more complex in the smallest capillaries, since here the size of a red blood cell becomes comparable to that of the vessel and the continuum hypothesis may become questionable.

Therefore, a first separation line between models for blood flow may be drawn: on one side the *Newtonian model* which neglects shear thinning and viscoelastic effects and is suitable in larger vessels or when we are not interested in the finer details of the flow, as non-Newtonian behaviour may affect, for instance, the size of the recirculation area behind a severe stenosis [382]. On the other side, in vessels of diameter, say, less than 1 mm the use of Newtonian models is hardly justifiable. The small velocities and shear stress here involved call for the use of one of the non-Newtonian models described in Chapter 6, see also [348]. Computationwise, non-Newtonian models which just modify the expression for the viscosity by making it dependent on the shear rate would increase the cost of computations of approximately 10 % [382], because of the extra calculations and the increased non-linearity of the problem. Full visco-elastic models may instead be much more costly in terms of computing time.

In the sequel of this section we will focus our investigation on flow in *large and medium sized vessels*. The flow is here governed by the Navier-Stokes equations. If we take $t = 0$ as the initial time of our analysis, we are required to solve for $t > 0$ the following system of partial differential equations,

$$\frac{\partial \boldsymbol{u}}{\partial t} + \rho(\boldsymbol{u} \cdot \boldsymbol{\nabla})\boldsymbol{u} + \boldsymbol{\nabla} P - \mathbf{div}(\mu \boldsymbol{D}(\boldsymbol{u})) = \boldsymbol{f}, \mathbf{div}\boldsymbol{u} = 0, \tag{2.1}$$

in a domain $\Omega \subset \mathbb{R}^3$ representing the lumen of the vessel, or system of vessels, under investigation.

The first equation expresses the conservation of linear momentum. It is a vector equation formed by three differential equations, one for each component of the velocity. The second equation is the continuity equation. The domain Ω is here fixed with time, in Chapter 3 we will discuss the modifications needed for the case of moving computational domain. The viscosity μ is in non-Newtonian models a function of the *strain rate*

$$\boldsymbol{D}(\boldsymbol{u}) = \frac{\boldsymbol{\nabla u} + \boldsymbol{\nabla u}^T}{2}, \tag{2.2}$$

while is kept constant when adopting the hypothesis of a Newtonian behaviour. The principal unknowns are the velocity $\boldsymbol{u}$ and the pressure P, while the density ρ is here constant. The term $\boldsymbol{f}$ in the right hand side accounts for the possible action of external forces, like gravity, and is often taken equal to zero in haemodynamics. We will derive and discuss this system of equations in

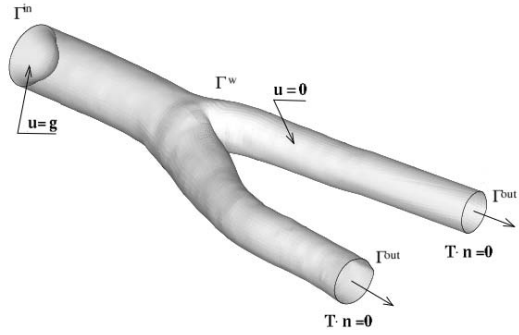

Fig. 2.1. A typical computational domain. Here we have a model of a carotid bifurcation developed by K. Perktold and his research group at Graz University of Technology on the basis of an experimental lumen cast prepared and digitised by D. Liepsch, FH Munich

detail in the next chapter, here we wish only to point out their main characteristics.

The equations have to be supplemented with boundary conditions on $\partial\Omega$. Referring to Fig. 2.1 describing a carotid bifurcation we typically prescribe a velocity profile at the *proximal* boundary Γ_{in}, that is the section closest to the heart along the direction of the mean blood flow, which we will also denote as "inflow" boundary, even if the term "inflow" in not completely correct since in some major vessels we can have flow reversals. We then prescribe zero velocity at the fixed walls and the normal stresses $\mathbf{T} \cdot \mathbf{n}$ at the *distal* boundaries Γ_{out} (also called "outflow" boundaries). Again, the term distal is meant with respect to the heart.

Proximal and distal boundaries are often indicated as *artificial boundaries* since they do not correspond to a physical interface between the fluid and the exterior, but to sections that have been artificially created to separate the region of interest for our investigation from the remaining part of the circulatory system. The set up of boundary conditions on artificial boundaries is an important issue for fluid dynamic computations. Treatments of the boundary data specially suited for haemodynamics will be discussed in Chapter 11.

We need also to prescribe the initial status of the fluid velocity, for instance

$$\boldsymbol{u}(\boldsymbol{x}, 0) = \boldsymbol{u}_0(\boldsymbol{x}) \qquad \boldsymbol{x} \in \Omega,$$

being $\boldsymbol{u}_0$ a given quantity. We recall that $\boldsymbol{u}_0$ cannot be arbitrary, since it has to satisfy $\mathbf{div}\boldsymbol{u}_0 = 0$ to be admissible.

Unfortunately, in haemodynamics computations usually we do not know a physically relevant "initial condition". Therefore $\boldsymbol{u}_0$ is usually chosen rather arbitrarily, often just equal to zero everywhere. It means that numerical computations may suffer a "false transient" linked to the incorrect initial data. If the boundary conditions are correct, however, it will decay quite rapidly

and after two or three heart beats we may consider that the solution is not anymore influenced by the incorrect initial data. A possibility to get a better guess of the initial data is to solve a stationary Stokes problem for $\boldsymbol{u}_0$ in Ω, that is

$$-\mathbf{div}(\mu \boldsymbol{D}(\boldsymbol{u}_0)) + \boldsymbol{\nabla} P = \boldsymbol{f},$$
$$\mathbf{div}\boldsymbol{u}_0 = 0.$$

where the forcing and boundary terms are those of the original problem at $t = 0$. In this way, $\boldsymbol{u}_0$ is certainly compatible and accounts already of part of the physics of the problem, the only missing terms being those related to fluid inertia. With this choice the "false transient" effect is greatly reduced and in practise one may assume that it has faded away completely after just a few time steps of the solution procedure.

The situation is worsened when the compliance of the wall is taken into account. The continuous exchange of energy between fluid and wall effectively makes the decay slower. In calculations of flow in compliant vessels it is normal practise to wait for at least three cardiac cycles before considering the influence of the initial data negligible.

The solution of the Navier-Stokes equations may develop instabilities, which are normally called *turbulence*. The responsible is the dynamics induced by the non-linear convection term $\rho(\boldsymbol{u} \cdot \boldsymbol{\nabla})\boldsymbol{u}$. It is therefore natural to measure the importance of this term compared with the diffusive part given by $\mathbf{div}(\mu \boldsymbol{D}(\boldsymbol{u}))$. This information is provided by the Reynolds number, defined in (1.1). Typical values of the Reynolds number along the arterial tree are given in Table 1.7. If the Reynolds number is small, say at most of the order of 1000, the flow remains stable, and is called *laminar*. In normal physiological situations, then, the values of the Reynolds number reached in the cardiovascular system do not allow the formation of full scale turbulence. Some flow instabilities may occur only at the exit of the aortic valve and limited to the systolic phase. In this region the Reynolds number may reach the value of few thousands only for the portion of the cardiac cycle corresponding to the peak systolic velocity, however, there is not enough time for a full turbulent flow to develop. When departing from physiological conditions, there are several factors that may induce transition from laminar to turbulent flows. For instance, the increase of flow velocity because of physical exercise, or due to the presence of a stenotic artery or a prosthetic implant such as a shunt, may produce an increase of the Reynolds number and lead to localised turbulence. Smaller values of blood viscosity also raise the Reynolds number; this may happen in the presence of severe anaemia, when the hematocrit drops sharply (and so does the viscosity). More details on the relations among domain geometry, flow characteristics and type of flow regime will be given in Chapter 5, where the definition of other adimensional numbers that characterise blood flow will be given as well.

Knowing the velocity and the pressure fields allows the computation of the *stresses*, in particular the shear stresses to which an arterial wall is subjected

due to the blood movement. Wall shear stresses, whose precise mathematical definition will be given in Chapter 3, are the force per unit area exerted by the fluid tangentially to the wall. We have already mentioned their importance in relation with some vascular diseases, since endothelium cells react to shear stresses. Irregular, and in particular small or "oscillating"[1] shear stresses may cause an alteration in the endothelium covering and induce inflammatory processes. Their calculation require a "point-wise" knowledge of the velocity and pressure field.

To account for the compliance of the vessel wall we need to introduce another unknown, namely the wall displacement $\boldsymbol{\eta}$. The mechanical interaction between the flowing blood and the vessel structure is rather complex. Its general mathematical description will be given in Chapter 3, while more details on the numerical techniques that can be adopted for fluid structure interaction problems are found in Chapter 9. Yet, even with the most advanced techniques available today, accurate computations of fluid structure interaction models of haemodynamics are rather costly. One reason is that the two dynamics (fluid and structure) are here strongly coupled and most of the simplest and cheapest techniques often used in other fields (like aeronautics) simply don't work. Consequently, one of the factors that may affect the choice of a fixed geometry model versus a fluid-structure interaction one is computation time: the latter may be one order of magnitude as expensive as the former.

It is therefore important to appreciate when the approximation of a fixed geometry could be reasonable. It depends on the type of vessels, the type of answers we are seeking and, finally, the type of data available. Smaller vessels experience a smaller relative movement than larger ones, where the change of radius during the heart beat may be of the order of 15 %, like in the aorta. Therefore, the flow in the peripheral vessels, lets say more than two branching levels down from the aorta, can be reasonably modelled using a fixed geometry. An exception being the coronaries, whose movement is however dominated by the heart movement more than the fluid-structure interaction in the vessel. In [396] the effect of heart movement in the shear stress distribution in a coronary artery has been investigated. It has been found that it can be relevant, particularly in vessels with high curvature.

Even in the larger vessels, at least in physiological situations, the main characteristic of the flow are already captured by a fixed geometry model. However, if more details are needed, such as a precise computation of shear stresses or the size of a recirculation region, then compliant models are better suited [250,481]. Furthermore, if it is necessary to have an accurate description of pulse waves, for instance if one wants to investigate altered pressure pattern

[1] Wall shear stress is considered to be oscillating when its component along the main flow direction changes sign during the heart beat. In normal situations the component of wall shear stress along the main flow is always negative. Oscillating shear stresses are usually found in recirculation regions.

possibly caused by anomalous pulse wave reflections, like in the study of aortic aneurysms [285], then compliant models are mandatory. The reason is that fixed geometry models simply cannot describe pulse waves: the propagation speed is here infinite because of the incompressible fluid. It is indeed the mechanical interaction between blood flow and vessel wall deformation that generates the pulse waves.

Lack of sufficient data on the mechanical parameters of the vessel wall may in some cases make compliant models less interesting. This is an important issue in cardiovascular simulations. It is often difficult to obtain accurate values for those parameters for a specific subject. They have often to be inferred from literature data obtained from experiments on animal or human cadaver tissues. More recently, a novel technique called elastography allows to infer some elastic properties from images of the vessel wall movement taken non-invasively. It may then be used to characterise a specific subject. Yet, this technique is still not well widespread. Therefore, sometimes the choice of using a fixed geometry model may be driven by the fact that no data is available to characterise the mechanical property of the vessel under investigation. It must be understood, however, that in large vessels these type of computations may provide qualitative information on the general flow but they may lack precision.

From the mathematical point of view, the analysis of fluid-structure interaction problems is still subject of open research. At the best of our knowledge, a complete mathematical analysis of the coupled fluid-structure problem is not available yet. In the steady case, for small enough applied forces, existence of regular solutions is proved in [201]. In the unsteady case, local solvability in time is proved in the simple case where the structure is a collection of rigid moving bodies in [202]. See also [73, 120]. Formulations based on optimal control on simpler models have been investigated, e.g., in [107, 292, 292, 345, 346, 361]. A overview on the most recent results on the analysis of this type of problems may be found also in Chapter 8.

As for the structural model for the vessel wall, several level of approximation can be considered, depending on the objective of the study and the data available. As we have described in detail in Section 4.2 of Chapter 1, the internal structure of the wall of a blood vessel is rather complex and varies largely with the type of vessel under consideration. The computation of the displacements (and thus the stress) field inside the vessel wall requires to solve the three dimensional equations of elastodynamics which are presented in all generality in the next chapter. Their solution requires to have precise data about the mechanical characteristics of the different layers forming the vessel wall. A piece of information difficult to obtain even by *in vitro* experiments, let alone *in vivo*. Published results like those in [222, 562] may help in the set-up of a proper structural models, yet also in this case we are still far from having the possibility of extracting routinely such type of information for a given person by non invasive techniques. However, technology in this field is progressing fast. As already mentioned, elastographic measurements will

probably become common in the near future, and this will be of great help to the set up of patient-specific models also for the structural part.

Whenever there is little interest in a detailed description of the stress and displacement fields inside the vessel wall while the focus is more on their action on the flow field, it is a common practise to resort to simplified (or reduced) structural models. A first simplification is of course to use uniform "space averaged" parameters to describe the mechanical characteristics of the vessel wall, thus ignoring its internal structure. A further step is to use the so called *shell models*, where the displacement field is defined on the surface described by the lumen-wall interface, as indicated in Fig. 2.2. Shell theory, whose mathematical derivation will be sketched in Chapter 3, is in fact based on the assumption of a thin structure. In the case of a vessel it means that the ratio h/R between the wall thickness and vessel radius should be small. Indeed this assumption is questionable, particularly in arteries, which have normally quite a thick wall (see Chapter 1). Yet, the approximation can still be justified by two empirical observations. The first is that the main responsible of the mechanical strength of the vessel wall (at least in physiological situations) is the elastin, which is mainly present in the media. Thus the "effective" thickness is smaller. The other is that, despite its use beyond the fundamental hypothesis, a shell model has proved to be capable of representing the dynamics of the fluid-wall interface with a sufficient level of accuracy, provided that appropriate averaged values of the mechanical characteristics are given. From the computational point of view, shell models are usually cheaper than three dimensional models, as one has to discretise a surface and not a three dimensional domain, with a reduction of the degrees of freedom required. For this reason they are often used in haemodynamic simulations involving fluid-structure interactions, also on realistic geometries, see for instance [187].

Further down in the hierarchy of structural models one may find one dimensional models (see Fig. 2.2). These models assume a cylindrical type geometry and therefore are suited only to study a single artery without bifur-

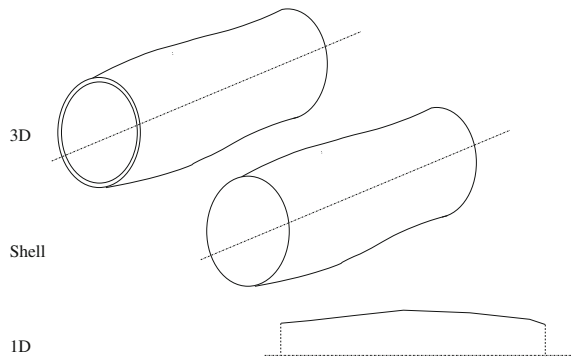

Fig. 2.2. A hierarchy of structural models

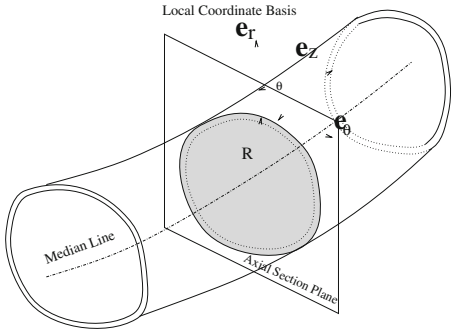

Fig. 2.3. The local coordinate system on the wall surface of a single vessel

cations. They are based on the fact that one can identify on each point of the fluid-structure interface a local system of coordinates $\mathbf{e}_r$, $\mathbf{e}_\theta$, $\mathbf{e}_z$, with $\mathbf{e}_z$ aligned along the vessel axis and $\mathbf{e}_r$ normal to the surface. The corresponding cylindrical coordinates are (r, θ, z), where the z coordinate axis is made to coincide with the centerline of the vessel (see Fig. 2.3).

If one assumes that the circumferential component of the shear stresses at the wall is negligible (this is true for an axial-symmetric geometry and axial-symmetrical deformations) it is possible to write a differential equation in the z variable and time. Often, the additional assumption of only radial displacements, i.e. $\boldsymbol{\eta} = \eta \mathbf{e}_r$ is made as well. No derivatives depending on the circumferential coordinate θ appear in the equations and we may consider each plane θ = const. independently. The resulting displacement field will depend only parametrically on θ. If, in addition, we assume that the problem has an axial symmetry (which implies the further assumption of a straight axis) the dependence on θ is completely neglected. In this case, also the fluid would be described by a 2D axi-symmetric model (see [117]).

Some more details on the derivation of models of this type are given in the Chapter 3. Clearly, we have here quite an important simplification also from the computational point of view, and this explains why these models are widely used to develop and test fluid-structure interaction algorithms. Another reason is that they lend naturally to axi-symmetric formulations [117]. However, their validity in practical computations is limited due to the geometrical restrictions.

Yet, this is not the end of the story. Even simpler structural models may be devised where the normal component of the stress on the fluid-wall interface is directly linked to the normal displacements at the same point by an algebraic relation (or possibly an ordinary differential equation in the time variable). Here the wall mechanics is greatly oversimplified, and these models are indeed used mainly to derive reduced models for blood flow in arteries, like the ones discussed in Chapter 10, even if they have also been used in some early studies of blood flow in simple or 2D geometries [481].

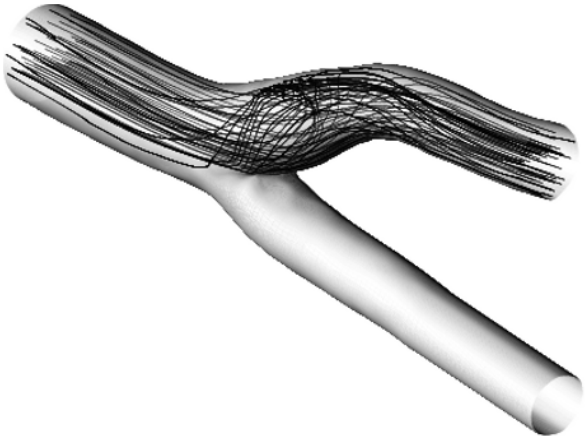

Fig. 2.4. Recirculation in a model of the carotid bifurcation[3]. We show the path of the particles entering the bifurcation. The presence of a recirculation region is evident

A clear major feature of blood flow is its pulsatility. It may induce flow reversal and recirculations near the arterial wall, a phenomenon that can have negative effects on the endothelium and stimulate the deposit of lipids and atherosclerosis. The latter effect is more likely to occur in specific vascular districts, like the carotid bifurcation, see Fig. 2.4.

With some approximation one may think that blood flow is periodic in time. Yet, this can be considered true only for relatively short periods, since the various human activities require to change the amount of blood sent to the various organs. Also the elastic properties of arteries (especially the arterioles) may vary depending on the request of blood by the peripheral organs. Indeed one of the aspects of current research in computational haemodynamics is the interaction between blood flow and the metabolic regulation [108]. It presents several challenges from the mathematical modelling and numerical side. For the sake of space and because only partial results are available so far this aspect has not been extensively covered in this book (see Chapter 10).

In several early studies, however, blood computations were made using steady flow. This can be considered acceptable in peripheral arteries, the capillary bed and in the veins,where the pulsatility of the flow is reduced thanks to the regularising effect of the compliance of the major arteries. In particular, micro-circulation is practically (but not completely) steady. The use of steady computations in larger vessels may again by justified by the lower computational cost. If we eliminate the time derivative in (2.1) we still have a non-linear system of partial differential equations to solve, yet we can make use of acceleration techniques unsuited for unsteady computations because

[3] Computation made by M. Prosi.

they destroy time-accuracy. However, some important quantities (recirculation regions, oscillatory shear stress etc.) cannot be evaluated with high accuracy and only a general qualitative structure of the flow field can be inferred.

We mention that in some particular contexts, for instance in the hyperthermia treatment where some drugs are activated through an artificial localised increase in temperature (see [123, 219]), the variation of blood *temperature* may be relevant. Describing the evolution of temperature requires to introduce another partial differential equation which derives from the principle of energy conservation, and couple it with the Navier-Stokes equations. In large and medium sized vessel the coupling is weak, since here temperature variations have small influence in the flow field. Therefore, one may solve the energy equation (which is instead strongly influenced by the flow field because of the convection term) after having calculated the velocity field. The computational overhead is in this case minimal since we have to solve a single additional equation, moreover of linear type. Things are different in micro-circulation, where the combined effect of temperature on the blood apparent viscosity and on other mechanical properties of the vessel wall makes the situation more complex [178]. However, since in the physiological regime the temperature inside the human body is constant and the situations where temperature variations are relevant are rather special, we will not pursue this topic further in this book.

2.2 Mathematical models for biochemical transport processes

The transport of biochemicals by the arterial blood stream and its interaction with intra-wall transport is of great interest in the vascular physiology and biology. The local mass transfer between the blood and the arterial wall affects the transport of nutrients to the cells, the removal of metabolic wastes from the wall, and the accumulation of potentially atherogenic molecules [175]. The transendothelial mass transfer has been already explained in Section 1.1.3 of Chapter 1. One main aspect of the interest concerns the relation between haemodynamics and molecular transport and the development of pathological vessel alterations.

It has been observed that low density lipoprotein (LDL) accumulation in the intima at zones of low and oscillating wall shear stress is associated with the tendency to intimal thickening and the development of atherosclerotic diseases [64, 278].

The dynamics of dissolved gases (e.g., oxygen or carbon dioxide) and of macromolecules (e.g., lipoprotein or albumin) in arteries and in the arterial wall is strongly related to the flow dynamics of blood. Irregular blood flow patterns with flow stagnation, separation and recirculation, and resulting local low and oscillating wall shear stress, causes local disturbed mass transfer, e.g. [305, 381].

Several mathematical models have been developed for the study of biochemical transport processes in arteries. The simplest model considers solute transport only in the artery lumen and replaces the wall by means of an appropriate boundary condition at the inner surface of the arterial wall (blood-endothelium boundary). This model couples the Navier-Stokes equations with the advection-diffusion equation describing the dynamics of molecules transported in blood.

Improved models account for the arterial wall, where the mass transport in blood and in the wall are described applying physically appropriate laws to model the interaction between the blood flow and the biochemical transport. These models take into account the heterogeneous layers constituting the realistic arterial wall (from inside to outside) the endothelium, the intima, the internal elastic lamina (IEL) and the media. The physical behaviour of the different layers are approximated with the laws of mass transport in porous media (intima and media) and through plasma-permeable membranes (endothelium and IEL), see [251, 381]. In order to simplify the multilayer model the arterial wall can be treated as a single porous layer which is separated from the arterial lumen by a membrane.

2.2.1 Transport in the arterial lumen

The mathematical description of arterial mass transport requires to augment the Navier-Stokes equations (2.1) with the advection-diffusion-equation for the solute concentration c,

$$\frac{\partial c}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla} c - \boldsymbol{\nabla} \cdot (D \boldsymbol{\nabla} c) = 0 \quad \text{in} \quad \Omega, \quad t > 0. \tag{2.3}$$

The velocity $\boldsymbol{u}$ couples the transport problem to the Navier-Stokes problem. A further coupling of the concentration field to the flow field occurs whenever the diffusivity D of the solute in plasma depends on the strain, see [411] and references therein.

The characterisation of the transport processes uses the Péclet number, defined as $Pe = \frac{UL}{2D}$, which relates the advective transport to the diffusion. Here, L is a typical length scale, for instance the length of the vessel under consideration.

Mass transport processes in medium-sized and large arteries are generally strongly advection dominated, which is reflected in rather large Péclet numbers. The resulting numerical problems are discussed in Chapter 7.

The solution of the time-dependent mass transport problem requires the prescription of an appropriate condition at the initial time $t = 0$, i.e. $c(\boldsymbol{x}, 0) = c_0(\boldsymbol{x})$ for $\boldsymbol{x} \in \Omega$, being c_0 a given function.

For flow domains with "artificial" inflow and outflow boundaries, in most cases a constant concentration profile is prescribed at the inflow cross-section Γ_{in} as Dirichlet boundary condition. At the outflow boundary zero diffusive flux can be assumed, i.e. a homogeneous Neumann condition, i.e. $\partial c / \partial n = 0$.

Appropriate boundary conditions at the inner surface of the arterial wall depends on the molecule size. The transfer of small molecules (dissolved gases) to and into the wall is controlled by diffusion in the boundary layer, as the endothelium is not an significant barrier to the motion of these molecules. Thus, the assumption of a Dirichlet boundary condition at the inner surface is in this case justified. However, in the transfer of macromolecules (LDL) from blood into the arterial wall the main resistance is the endothelial layer. The flux across the endothelium into the arterial wall is determined by the endothelial permeability and by the concentration differential. Therefore, a permeability boundary condition at the blood-endothelium surface Γ_w of the type

$$c\boldsymbol{u}\cdot\boldsymbol{n} - D\frac{\partial c}{\partial \boldsymbol{n}} = \mathcal{P}(c - c_i), \tag{2.4}$$

is appropriate to model the arterial macromolecule transport. $\mathcal{P}$ is the endothelial permeability, c_i is a prescribed concentration in the sub-endothelial intima, and $\boldsymbol{u}\cdot\boldsymbol{n}$ is the normal component of the filtration velocity of plasma at the lumen surface. It is either known or computed using the Darcy model presented in Chapter 7.

Expressing the fact that the endothelium is not a passive barrier to macromolecules, the permeability depends on the local shear stress at the endothelium, i.e $\mathcal{P} = \mathcal{P}(|\boldsymbol{t}^T\boldsymbol{\sigma}\boldsymbol{n}|)$, where the shear stress $\boldsymbol{t}^T\boldsymbol{\sigma}\boldsymbol{n}$ is the tangential component of the Cauchy stress tensor, defined in Chapter 3, equation (3.33).

This model is called *wall-free model* since we are not computing the transport inside the arterial wall. It is suitable to analyse the lumen concentration polarisation effect of large molecules directly at the wall, which happens when the equilibrium concentration at the fluid-endothelium boundary is higher than the concentration in the bulk of the blood stream.

A more realistic model of biochemical transport processes in arteries takes into account the heterogeneous wall, consisting of layers with strongly different thickness and physical properties, as shown in Fig. 1.4. This is called the *multilayer model* and it couples the solute concentration in the blood stream (lumen) and in the intima and media. For a complete derivation of such model, we remand to Chapter 7.

The multilayer model requires to determine a large number of parameters which characterise the physical properties of each layer. The transport parameters of the intima and the media (effective diffusivity, Darcy permeability and porosity) are obtained from the fibre matrix models of the arterial wall tissue. The parameters of the permeable membranes, the endothelium and the IEL (permeability, hydraulic conductivity and reflection coefficients), are calculated from the equations of pore theory. The literature concerning these topics is very extended. Among others, we refer to [11,106,233,234,270]. For the specific case of the multilayer model, we will present in Chapter 7 a brief overview, that is mainly inspired to [251].

2.3 Numerical solution of partial differential equations: a quick review

The mathematical models we have briefly illustrated in the previous sections cannot in general be solved analytically, a part simple cases. Thus we have to resort to numerical techniques to find approximated solutions. It is our intention to give in this section just an introductory glance on this topic. The interested reader can find details in the ample literature available, see for instance [343, 407, 498].

All models just presented are based on partial differential equations (PDEs) for an unknown u (which may be a scalar or a vector field) of the general form

$$\frac{\partial u}{\partial t} + L(u) = f, \quad \text{in } \Omega, \quad 0 < t < T, \tag{2.5}$$

where L indicates a (linear or non-linear) differential operator in the space variable $\boldsymbol{x}$. The former equation will be augmented by proper boundary and inital conditions. In some cases the time derivative is not present (steady problems). Even when the problem is originally set in a semi-infinite time domain, the numerical approximation deals with a bounded time interval, the time T indicating the final time of our simulation.

The most common techniques to solve numerically a PDE are based on a subdivision of the computational domain Ω into a grid, see Fig. 2.5. The solution u is replaced by an approximation u_h which depends on a finite number of parameters, typically (but not necessarily) the values of u_h at the nodes of the grid. The pedix h is here an indication of the grid spacing. Some details on the most common meshing strategies used for cardiovascular geometries are given in Chapter 4.

In the case of time-dependent problems, we will also need to advance the approximation in time, using a so called time-advancing (or time-stepping)

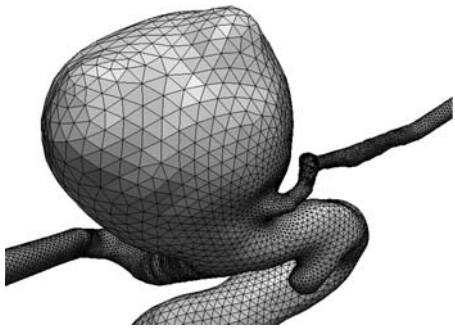

Fig. 2.5. An example of computational grid describing a cerebral artery with an aneurysm. Here only the surface mesh is shown, formed by triangles. The interior is covered by a tetrahedral grid (courtesy of T. Passerini, Aneurisk Project)

scheme. It is typically an iterative method that from the knowledge of the approximation u_h^n at a time t^n, for $n = 0, 1, \ldots, n$ builds the approximation u_h^{n+1} at time $t^{n+1} = t^n + \Delta t$, being $\Delta t > 0$ a chosen *time step*. Most frequently, one-step schemes are adopted, where the computation of u_h^{n+1} will involve just the knowledge of u_h^n.

Let us consider space discretisation first. The most common methods are finite difference, finite volume and finite elements.

2.3.1 Finite difference method (FDM)

When adopting finite differences, the approximated solution u_h is in fact a vector of values $\boldsymbol{u}_h = [u_1, \ldots, u_n]^T$ corresponding to the approximation at the nodes of the computational grid. The differential problem is *collocated* at the grid nodes by replacing the differential operator L with finite differences. For instance, the Laplace operator in two dimensions $\Delta u = \partial^2 u/\partial x^2 + \partial^2 y/\partial y^2$ at node $\boldsymbol{x}_i$ of the regular grid of Fig. 2.6 would be approximated as

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) u(\boldsymbol{x}_i) \simeq \frac{u_e + u_w + u_s + u_n - 4u_i}{h^2},$$

being h the grid spacing in the x and y directions, here taken constant for simplicity.

Historically, finite differencing is probably the first technique adopted for spatial discretisation. Yet, its use is less common in modern solvers. The reason is manifold. The construction of the finite difference operator is rather complicated for grids that are not uniform and not structured (a structured grid is a grid made up by a regular pattern of nodes). Yet, these grids are mandatory to treat complex geometries, particularly in three dimensional problems. The handling of boundary conditions is also not always straightforward with finite differences, in particular boundary conditions of Neumann type which involve the normal derivative of u at the boundary.

2.3.2 Finite volume method (FVM)

The finite volume method makes use of an integral formulation of the equation. It can be employed whenever the operator L is written in the so called *divergence* or *conservation* form, that is

$$L(u) = \operatorname{div} \boldsymbol{F}(u), \tag{2.6}$$

where $\boldsymbol{F}$ is the so called *flux vector*, which depends on u and on spatial derivatives of u. Let us note that with a few manipulations, the Navier-Stokes equations (2.1) may be written in conservation form, where for the momentum equation

$$\boldsymbol{F}(\boldsymbol{u}, P) = \boldsymbol{u} \otimes \boldsymbol{u} + P\boldsymbol{I} - \mu \boldsymbol{D}(\boldsymbol{u}),$$

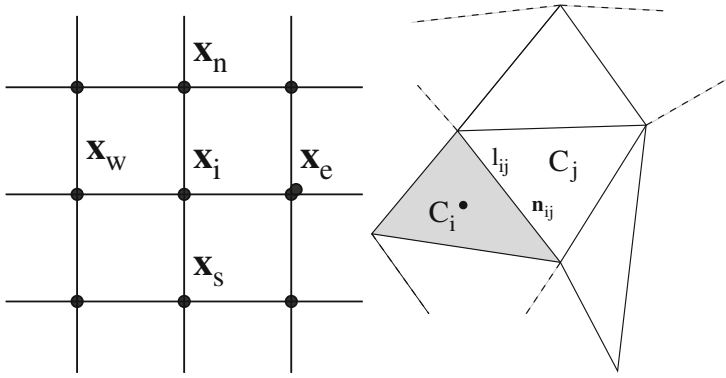

Fig. 2.6. On the left an example of a regular two dimensional grid used for approximation by finite differences. On the right, a generic control volume C_i used in a finite volume scheme. Although we show here a grid made of triangles, finite volumes methods con operate on control volume of general polygonal shape

while for the continuity equation $\boldsymbol{F}(\boldsymbol{u}) = \boldsymbol{u}$. Also equation (2.3) can be recast in conservation form, thanks to the continuity equation $\boldsymbol{u} \cdot \boldsymbol{\nabla} c = \mathrm{div}(c\boldsymbol{u})$.

In general, all differential problems that model a conservation law of physics can be cast in conservation form. We will address in more detail the conservation form of the Navier-Stokes equations in Chapter 3.

For what concerns finite volume methods, once the differential operator is written in conservation form, we integrate the equation on control volumes (usually of polygonal shape) built on the grid, like the one shown in Fig. 2.6. The unknowns are now the approximation of u at each control volume. By applying the divergence theorem we have

$$\int_{C_i} \mathrm{div}\, \boldsymbol{F}(u) d\boldsymbol{x} = \int_{\partial C_i} \boldsymbol{F}(u) \cdot \boldsymbol{n} d\gamma \simeq \sum_j \boldsymbol{F}_{l_{ij}}(u_h) \cdot \boldsymbol{n}_{ij}.$$

Here, $\boldsymbol{F}_{l_{ij}}$ is an approximation, called numerical flux, of the flux vector on the side l_{ij} of the control volume C_i. The latter has normal $\boldsymbol{n}_{ij}$. The numerical flux depends on the numerical solution u_h. In practise, it usually depends on the value of u_h at the control volume C_i and at the adjacent control volumes. In this way the previous expression actually involves a small number of unknowns. By applying it to all control volumes of the grid we can transform (2.6) into system of linear (or non-linear) equations.

Also in this case we can introduce a parameter h which accounts of the size of control volumes. Typically h is the maximum diameter of the control volumes of the grid. The smaller h the finer the approximation and the greater the computational cost (since we have a higher number of control volumes).

The finite volume method is probably nowadays the most used method for computational fluid dynamics (CFD). The reason is that it combines a

high geometrical flexibility (it can operate on arbitrarily complex geometries, even using grids with control volumes of different shapes) with computational efficiency. Furthermore, the construction of the numerical fluxes can be done so that some physical properties (such as local conservation or monotonicity) are preserved also at numerical level. An account of finite volumes for CFD is given in [530].

2.3.3 Finite element method (FEM)

Finite elements are based on a different integral formulation. To describe it let first consider a simple steady problem where L is the Laplace operator with mixed boundary condition, more precisely

$$-\Delta u = f \quad \text{in } \Omega, \tag{2.7}$$

with

$$u = 0, \quad \text{on } \Gamma_D \quad \text{and } \boldsymbol{\nabla} u \cdot \boldsymbol{n} = \frac{\partial u}{\partial n} = g, \quad \text{on } \Gamma_N, \tag{2.8}$$

being Γ_N and Γ_D two parts of the boundary of Ω such that $\Gamma_N \cap \Gamma_D = \emptyset$ and $\Gamma_N \cup \Gamma_D = \partial\Omega$. We will also assume that $\Gamma_D \neq \emptyset$. The function h is a given datum (Neumann boundary condition), while on the Dirichlet boundary Γ_D we have assumed homogeneous conditions only for the sake of simplicity.

The weak formulation

We proceed formally by multiplying both members by a test function $v : \Omega \mapsto \mathbb{R}$, regular enough, integrating over Ω, and using integration by parts (by applying the Green formula) we get

$$\int_\Omega \boldsymbol{\nabla} u \cdot \boldsymbol{\nabla} v d\Omega - \int_{\partial\Omega} v \boldsymbol{\nabla} u \cdot \boldsymbol{n} d\gamma = \int_\Omega f v d\Omega.$$

If v is chosen so that it is zero on Γ_D, by applying the Neumann boundary condition finally we obtain

$$\int_\Omega \boldsymbol{\nabla} u \cdot \boldsymbol{\nabla} v d\Omega = \int_\Omega f v d\Omega + \int_{\Gamma_N} g v d\gamma. \tag{2.9}$$

This statement can be written in the general form $a(u, v) = F(v)$, by setting $a(u, v) = \int_\Omega \boldsymbol{\nabla} u \cdot \boldsymbol{\nabla} v$ and $F(v) = \int_\Omega f v d\Omega + \int_{\Gamma_N} g v d\gamma$.

To give sense to the formal steps made so far, we need to identify the correct functional space for the solution u and the test function v. We will postpone this aspect to Section 2.4. For the time being, we assume that u and v are regular enough so that all the previous steps are well defined and the integrals finite. We can then note that (a) if u is a solution of the original problem (2.7)-(2.8), then it satisfies (2.9); (b) the test function v and the

solution u are subject to the same essential conditions on Γ_D, namely they are both zero.

It comes then natural to introduce the abstract problem

$$\textit{Find } u \in V \textit{ such that} \quad a(u,v) = F(v), \quad \forall v \in V, \tag{2.10}$$

being V a space of function regular enough and null on Γ_D, which we will make more precise later.

Formulation (2.10) is called *weak formulation* . We have obtained it for the Laplace problem (2.7)–(2.8), yet it is a rather general fact that a wide class of partial differential problems can be rewritten in the weak form (2.10), with obviously a different definition of $a(u,v)$ and $F(v)$, and possibly of the space V. The application $F(v)$ returns a real number for each $v \in V$ and is called a *functional*, while $a(u,v)$ returns a real number for each couple of functions u and v in V and is called a *form*. Under certain assumptions on the space V, the form a and the functional F it is possible to prove that the weak formulation is *well-posed*, that is it admits a unique solutions u which depends continuosly on the data of the problem (in our example g and f). This important result takes the name of Lax-Milgram lemma and its statement can be found, for instance, in [408]. Furthermore, it can be proved that regular solution of the weak formulation do indeed satisfy the original differential problem in what is called *strong form* (in contract to the weak form). However, the space of weak solutions is somehow larger than that of the problem in strong form.

Therefore, the weak formulation may be seen as a generalisation of the original problem. It gives a robust mathematical framework to differential problems, encompassing situations (e.g. rather irregular data or domains) which cannot be treated satisfactorily in the classical strong formulation. Even for more complex problems, like the Navier-Stokes equations or the fluid-structure interaction problem, a weak formulation may be found, as it will be illustrated in Chapter 8.

The Galerkin method

Moving from the weak formulation we can replace the space V, which is infinite dimensional, with a finite dimensional subspace V_h, that is we choose a $V_h \subset V$ with $\dim(V_h) = N_h$ and solve the problem:

$$\textit{Find } u_h \in V_h \textit{ such that} \quad a(u_h, v_h) = F(v_h), \quad \forall v_h \in V_h. \tag{2.11}$$

Being V_h finite dimensional the approximate solution u_h may be expanded with respect to a base of V_h as $u_h(\boldsymbol{x}) = \sum_{i=1}^{N_h} u_i \phi_i(\boldsymbol{x})$. In other words, V_h is spanned by the basis $\{\phi_i, i = 1, \ldots, N_h\}$. The coefficients $u_i \in \mathbb{R}$ are called *degrees of freedom* and are indeed the unknowns of the discrete problem. Furthermore, we can choose $v_h = \phi_j$ for $j = 1, \ldots N_h$ in (2.11) to produce linear system of equations $A\boldsymbol{u} = \boldsymbol{b}$, where

$$A_{ij} = a(\phi_j, \phi_i) = \int_\Omega \boldsymbol{\nabla}\phi_j \cdot \boldsymbol{\nabla}\phi_i \, d\Omega, \quad b_j = F(\phi_j) = \int_\Omega f\phi_j d\Omega + \int_{\Gamma_N} g\phi_j d\gamma.$$

Matrix A is traditionally called the *stiffness matrix*. The actual expression for stiffness matrix and source term may be more complex for problems more complicated than Laplace's, yet the way of deriving the discrete system remains fundamentally the same. In the case of a non-linear problem, the resulting system will be non-linear, typically this means that the matrix A will depend on u_h.

This technique of building the discrete problem by projection on a subspace and searching the solution in the same subspace takes the name of *Galerkin method*.

The finite element space

A last (but not the least) step is to choose the way of building V_h. To this aim, various different techniques are possible. In the classical spectral element method for periodic solutions a truncated Fourier series is used on the whole Ω. It is a valid method if the solution is very smooth and on very simple geometries (typically cubic or cylindrical domains). The other methods do require to subdivide the domain into a grid $\mathcal{T}_h$ of polygonal elements (or at least elements that can be mapped by simple transformations into a polygon), in a way similar to what is done in the finite volume method. Given the grid, the space V_h may be expressed by piecewise polynomial functions, for instance we could consider the space

$$X_h^r(\mathcal{T}_h) = \{v_h \in C^0(\overline{\Omega}), : \quad v_h|_K \in \mathbb{P}^r, \quad K \in \mathcal{T}_h\}$$

of piecewise polynomials of degree r on each grid element K. The space V_h is then taken as the subspace of $X_h^r(\mathcal{T}_h)$ that accounts for the constraints at the Dirichlet boundary. Methods of that sort are the *finite element method* (FEM) and the *spectral element method* (SEM). They differ on the choice of the basis for V_h and on the fact that the SEM adopts high order polynomials, while in the FEM rarely r exceeds 3 (and often is equal to 1, i.e linear finite elements). Another characteristic of the SEM method is the use of special quadrature rule to approximate the integrals which guarantee high convergence rate (for smooth solutions) while keeping the computational cost of building the linear system reasonably small. Thus, they are quite interesting when dealing with smooth solutions. However, are much less used than FEM in standard solvers, since often the solution is not as regular as needed to benefit from the method and their implementation on complex geometries may be rather complex. Indeed, SEM methods are usually (although not always) implemented on grids whose elements are hexahedra (quadrilaterals in 2D), while finite elements are also implemented on tetrahedral (triangular) grids, which are much more flexible. For more details on the spectral element method the reader may consult [60, 253].

The usual (though not exclusive) choice for the basis function ϕ_i in the FEM is such that the degrees of freedom u_i do correspond to the value of

u_h at some points of the grid called nodes. An important characteristic of the FEM is that the basis function ϕ_i has small support, i.e. is different from zero only on a limited connected portion of Ω. The main consequence is that the stiffness matrix A is *sparse*, since $A_{ij} = 0$ whenever the support of ϕ_i and ϕ_j has zero intersection, and this happens whenever nodes $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are not shared by any element of the mesh. This fact is very relevant in applications since sparse matrices have a smaller memory requirement than their full counterpart. The computer implementation is also rather efficient, since the matrix and the right hand side can be built by looping over the mesh elements and performing local operations on each element.

The FEM is probably one of the most adopted numerical method for partial differential equations, especially in the field of structural mechanics, although, as already mentioned, in the field of CFD finite volume methods are probably more popular. Yet, finite elements are gaining grounds also in this field, particularly in the case of incompressible flow. The richness (and flexibility) of their mathematical formulation allows in fact to develop a large variety of numerical schemes, basically changing the discrete spaces where the solution is sought and the test functions chosen.

The parameter h in the finite element method is identified as the maximum diameter of the finite elements in the given mesh. The smaller h, the higher the number of elements necessary to cover Ω and the higher the number of degrees of freedom (and the dimension of the linear system).

In the case of non linear problems, like the Navier-Stokes equation, the application of the method leads to a non-linear system of equation, where the stiffness matrix is itself function of the discrete solution.

Convergence

In all methods shown so far we expect that the approximation u_h becomes more precise as h goes to zero. Indeed, a request we make is that a discretisation method be *convergent*, that is the error should go to zero as $h \to 0$. Furthermore, we say that the method converges with order p if it exists a constant $C = C(u) > 0$ so that, for h sufficiently small, the following inequality applies

$$||u - u_h|| \leq C\, h^p.$$

The error is measured in a suitable norm, here indicated by $||\cdot||$. The order of convergence typically depends on the particular method chosen, the regularity of the exact solution and the selected norm.

For instance, a standard FEM discretisation of the Laplace problem satisfies

$$||u - u_h||_{H^1(\Omega)} \leq C\|u\|_{H^s(\Omega)} h^p,$$

for a constant $C > 0$, where $p = \min(r, s-1)$, being r the degree of the finite element, and s a measure of the regularity of the solution u. More precisely, the error estimate is true whenever $u \in H^s(\Omega)$. The definition of the Sobolev

spaces $H^k(\Omega)$ as well as that of the H^k norm $||\cdot||_{H^k(\Omega)}$ will be given in Section 2.4. Therefore, if we use linear finite elements, we have a method whose convergence is linear with respect to h when using the H^1 norm to measure the error.

We mention also the possibility of generating another class of schemes by replacing the original form $a(u_h, v_h)$ with an approximation $a_h(u_h, v_h)$. The simplest way by which this is done is by performing numerical quadratures to approximate the integrals which define the form. Methods of this sort are usually called *generalised Galerkin methods*. We give in Fig. 2.7 a synthesis of the main numerical techniques for PDEs which includes those mentioned here.

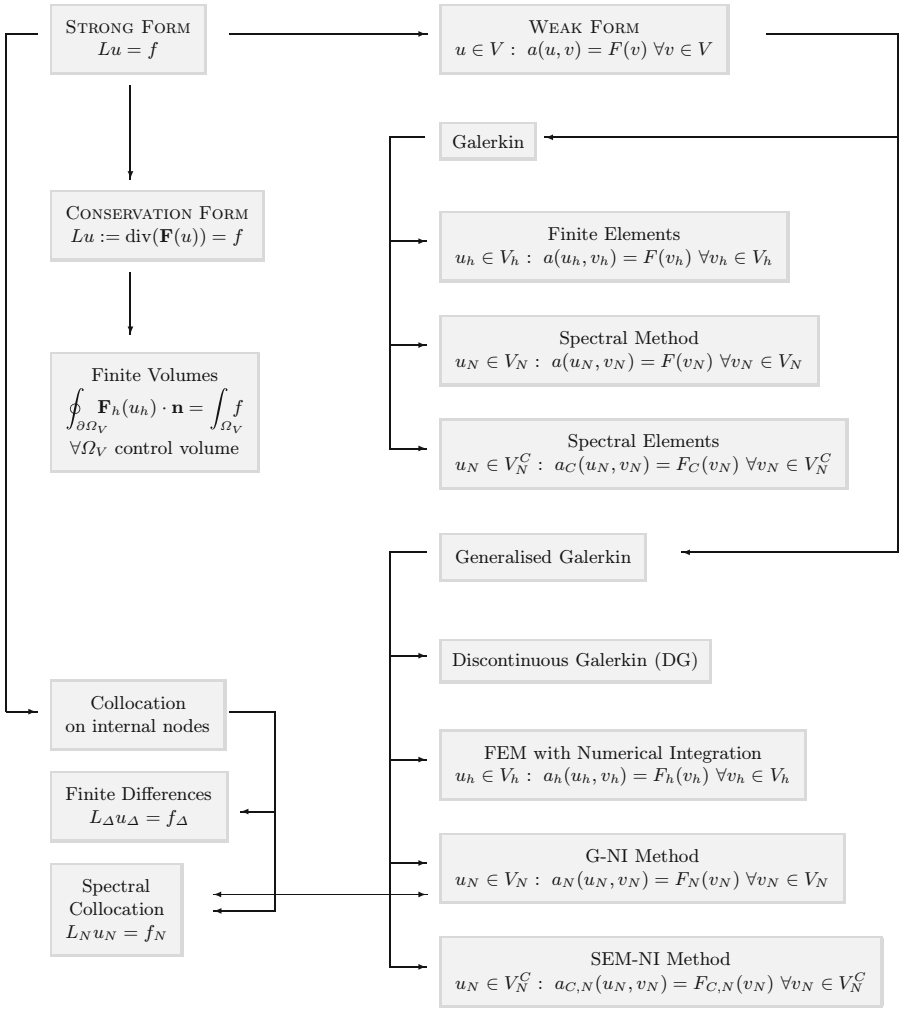

Fig. 2.7. A flow-chart of the main discretisation methods for PDEs

2.3.4 Time advancing schemes

In the previous paragraphs we have considered the discretisation with respect to the space variables. If we have a transient problem we need to consider the time as well. What is usually done in this case is to first discretise in space and then consider the time evolution. The reason why the two variables are treated differently is due to their different nature. The differential problems we will consider here are usually boundary value problems with respect to space, that is conditions are set on the whole boundary of the spatial domain Ω. On the contrary, they require a condition only at one end of the time axis, the initial condition. It is natural, then, to think of using a different numerical techniques to treat the time variable.

If we consider the general case of equation (2.5), after space discretisation with finite differences or finite volumes we obtain a system of ordinary differential equations on each node i, of the form

$$\frac{du_i}{dt}(t) + \boldsymbol{s}_i^T \boldsymbol{u}(t) = f_i(t), \quad 0 < t < T, \tag{2.12}$$

where $\boldsymbol{u}(t) = [u_1(t), \ldots, u_{N_h}(t)]^T$ are the degrees of freedom (here equal to the approximated solution at the nodes or at the control volumes), which are now function of the time. Furthermore, $f_i(t) = f(\boldsymbol{x}_i, t)$ and $\boldsymbol{s}_i$ is here a vector of coefficients. For instance, for the Laplace equation, finite differences on a the regular grid of Fig. 2.6 would give at any internal node i

$$\frac{du_i}{dt}(t) + \frac{4u_i(t) - u_e(t) - u_w(t) - u_s(t) - u_n(t)}{h^2} = f_i(t).$$

In matrix form

$$\frac{d\boldsymbol{u}}{dt}(t) + A\boldsymbol{u}(t) = \boldsymbol{f}(t),$$

where A is the matrix with the $\boldsymbol{s}_i$ as rows.

Schemes for this system of ordinary differential equations require to choose a time step Δt and solve a problem for $\boldsymbol{u}^n$, approximation of $\boldsymbol{u}(t^n)$, being $t^n = n\Delta t$. The most common are

- the *explicit Euler* method, also called *forward Euler* method

$$\boldsymbol{u}^{n+1} = \boldsymbol{u}^n + \Delta t \left[A\boldsymbol{u}^n + \boldsymbol{f}(t^n)\right]; \tag{2.13}$$

- the *implicit Euler* method, also called *backward Euler* method

$$(\boldsymbol{I} + \Delta t A)\boldsymbol{u}^{n+1} = \boldsymbol{u}^n + \Delta t \boldsymbol{f}(t^{n+1}); \tag{2.14}$$

- the *Crank-Nicolson* method,

$$(\boldsymbol{I} + \frac{1}{2}\Delta t A)\boldsymbol{u}^{n+1} = \boldsymbol{u}^n + \frac{\Delta t}{2}\left[A\boldsymbol{u}^n + \boldsymbol{f}(t^n) + \boldsymbol{f}(t^{n+1})\right]. \tag{2.15}$$

All methods compute in successive steps the sequence $\boldsymbol{u}^1, \boldsymbol{u}^2, \ldots$, being $\boldsymbol{u}^0$ known from the initial condition of the differential problem.

The main difference among (2.13) and (2.14) or (2.15) is that the former allows to compute the approximation at time step t^{n+1} without the need to solve a linear system. This is why is called explicit. Crank-Nicolson scheme is more accurate than the other two since it a second order method with respect to time. It means that assuming that the space operator is exact the error between the approximated solution and the exact one goes to zero as $(\Delta t)^2$ when $\Delta t \to 0$. The other two methods are just first order. Explicit methods are always subject to a stability condition, that is they provide a reasonable approximation only if $\Delta t \leq \widehat{\Delta t}$, where the *critical time step* $\widehat{\Delta t}$ depends on the particular method chosen and on the eigenvalues of the matrix A. The latter is normally a decreasing function of h. Thus, typically the finer the grid, the smaller the time step we have to adopt with an explicit scheme to satisfy the stability condition. The type of problems we have to face to solve haemodynamic applications usually exhibit a second order differential operator L and, in general, for this class of problems

$$\widehat{\Delta t} = Ch^{-2},$$

which penalises explicit schemes strongly. Therefore, despite their higher computational complexity, implicit and unconditionally stable schemes are often preferred for this class of problems. For instance, both implicit Euler and Crank-Nicolson schemes fall in this category.

Were the original differential problem non-linear, implicit schemes lend to non-linear problems, to be solved at each time step. A possibility is to resort to a Newton iteration or other fixed point strategies, see for instance [257,403].

When dealing with finite element computations, the ordinary differential equations stemming from space discretisation are in fact of the form

$$M\frac{du_i}{dt}(t) + S\boldsymbol{u}(t) = f_i(t),$$

where M is the so called *mass matrix*, of elements $M_{ij} = \int_\Omega \phi_i \phi_j \, d\Omega$. Consequently, we end up with a non trivial linear system even when adopting explicit schemes. Yet, often the mass matrix M can be replaced with a diagonal matrix, called *lumped mass matrix*, and we are able to write the differential system in the same form as in (2.12).

We mention that several other time advancing schemes are possible. In particular, for the class of problems we deal in this book, methods based on backward difference formulae (BDF) are quite interesting since they couple good stability and convergence properties with an acceptable computational cost. Interested readers can refer, for instance, to [403].

2.4 Some elements of functional analysis

We give here some elements of functional analysis to help the reader through some of the next chapters. Because of the scope of the book and the sake of space we will be very brief and rather informal. A more complete introduction in the context of partial differential equations is [424], while among the more advanced books on the subject we mention [552] and [438].

First of all with *functional space* we denote a *linear* space of functions $\Omega \subset \mathbb{R}^d \mapsto \mathbb{R}^n$, where n is typically 1 (scalar functions) or either 2 or 3 (vector functions), as well as d. We assume that the domain Ω is open and bounded. A common example of functional space is the space of continuous real functions on an interval $\Omega \subset \mathbb{R}$, usually denoted by $C^0(\Omega)$.

A norm is an application $V \mapsto \mathbb{R}$ such that

$$||v|| \geq 0, \quad \text{and} \quad ||v|| = 0 \quad \text{iff} \quad v = 0, \tag{2.16}$$

$$||v + w|| \leq ||v|| + ||w||, \qquad \forall v, w \in V, \tag{2.17}$$

$$||\alpha v|| = |\alpha| ||v||, \quad \forall v \in V, \quad \forall \alpha \in \mathbb{R}.$$

We recall that a normed and complete linear space V is also called a *Banach* space.

2.4.1 Functionals and bilinear forms

Given a functional space V an application

$$F: \quad V \mapsto \mathbb{R}$$

is called a *functional* on V. A functional is linear if $F(\alpha v + \beta w) = \alpha F(v) + \beta F(w)$, for all real numbers α and β and all v and w in V. A linear functional on a normed space V is *continuous* if and only if it is bounded, i.e. $\exists C > 0$ such that

$$|F(v)| \leq C||v||_V, \quad \forall v \in V.$$

We have indicated by $||\cdot||_V$ the norm of V. The space of linear and continuous functionals on V is itself a normed space, called the dual space V'. The norm of a functional is in fact the smallest constant C in the previous inequality, or equivalently

$$||F||_{V'} = \sup_{\substack{v \in V \\ v \neq 0}} \frac{|F(v)|}{||v||_V}.$$

A linear and continuous functional applied to an element $v \in V$ is often indicated using the *crochet* symbol, that is $F(v)$ can be written alternatively as ${}_{V'}\langle F, v \rangle_V$ (or simply $\langle F, v \rangle$ whenever there is no ambiguity). This notation puts into evidence the duality between the two spaces.

An application

$$a : V \times V \mapsto \mathbb{R},$$

that maps two elements of V to a real number is called a form. It is a *bilinear* form if linear with respect to each argument (taken singularly), i.e

$$a(\lambda u + \mu w, v) = \lambda a(u, v) + \mu a(w, v), \quad \forall \lambda, \mu \in \mathbb{R}, \forall u, v, w \in V,$$

$$a(u, \lambda w + \mu v) = \lambda a(u, v) + \mu a(u, w), \quad \forall \lambda, \mu \in \mathbb{R}, \forall u, v, w \in V.$$

A bilinear form is continuous if there exists a constant $M > 0$ such that

$$|a(u, v)| \leq M \|u\|_V \|v\|_V, \quad \forall u, v \in V,$$

and is *coercive* if $\exists \alpha > 0$ such that

$$a(u, u) \geq \alpha \|u\|_V^2, \quad \forall u \in V.$$

A scalar product (u, v) of a space V is an application $V \times V \mapsto \mathbb{R}$ such that it is bilinear with respect to each argument, $||v|| = \sqrt{(u, u)}$ is a norm on V (called normed induced by the scalar product) and in addition the Cauchy-Schwarz inequality holds:

$$(u, v) \leq \|u\| \|v\|, \quad \forall u, v \in V.$$

In other words, the scalar product is a continuous bilinear form with respect to the induced norm, with continuity constant equal to 1. A complete functional space V equipped with a scalar product and the induced norm is called *a Hilbert space*. Hilbert spaces play a fundamental role in the analysis of partial differential equations.

2.4.2 Support of a function

The support of a function f is the closure of the subset of Ω where $f \neq 0$. A function is said to have compact support in Ω if its support is contained in a closed and bounded subset of Ω. In particular, if f has compact support in Ω is zero on the boundary of Ω.

2.4.3 Sobolev spaces

The space $L^2(\Omega)$ is the space of square integrable functions, that is

$$L^2(\Omega) = \{v : \Omega \mapsto \mathbb{R}, \quad \int_\Omega v^2 d\Omega < +\infty\}.$$

It is a Hilbert space with scalar product $(u, v)_{L^2(\Omega)} = \int_\Omega u, v \, d\Omega$ and norm $\|v\|_{L^2(\Omega)} = \left(\int_\Omega u, v \, d\Omega\right)^{1/2}$. Often, the L^2 scalar product is simply indicated

as (u, v). To the sake of precision, we mention that the integral in the definition of $L^2(\Omega)$ (and of all the other functional spaces introduced in this section) is a Lebesgue integral. Lebesgue integration is a mathematical construction that extends the classical integral due to Riemann to a wider class of functions. From the practical point view there is little difference in using Lebesgue integrals, since bounded functions integrable in the classical sense are also Lebesgue integrable and the two integrals coincide.

The existence of the L^2 scalar product is based on the fact that if u and v are in $L^2(\Omega)$, then the integral $\int_\Omega uv\, d\Omega$ exists and is finite.

The space $H^s(\Omega)$ is defined as the space of function of $L^2(\Omega)$ such that all derivatives (partial derivatives if Ω is multidimensional) of order up to s belong to $L^2(\Omega)$ as well. For instance, in the case $\Omega \subset \mathbb{R}$,

$$H^1(\Omega) = \{v \in L^2(\Omega) : \quad \frac{dv}{dx} \in L^2(\Omega)\}.$$

The derivative in the definition has to be intended as generalised derivative (also called "distributional derivative"). In this context, $dv/dx \in L^2(\Omega)$ actually means that there exists a $g \in L^2(\Omega)$ such that for all functions $w \in C^\infty(\Omega)$ with compact support in Ω the following equality holds, i.e.

$$-\int_\Omega g w d\Omega = \int_\Omega v \frac{dw}{dx} d\Omega.$$

We will identify g with $\frac{dw}{dx}$. The notion of generalised derivative effectively extends the concept of derivative to non-differentiable functions in the classical sense. However, the two derivatives coincide for regular functions.

The Sobolev space $H^s(\Omega)$, with s a positive integer and $\Omega \subset \mathbb{R}^d$, is a Hilbert space when endowed with the scalar product

$$(u, v)_{H^s(\Omega)} = \sum_{|\boldsymbol{\alpha}| \le s} \int_\Omega \frac{\partial^{|\boldsymbol{\alpha}|} u}{\partial x^{\alpha_1} \dots \partial x^{\alpha_d}} \frac{\partial^{|\boldsymbol{\alpha}|} v}{\partial x^{\alpha_1} \dots \partial x^{\alpha_d}} d\Omega,$$

and the corresponding norm

$$\|u\|_{H^s(\Omega)} = \left[\sum_{|\boldsymbol{\alpha}| \le s} \int_\Omega \left(\frac{\partial^{|\boldsymbol{\alpha}|} u}{\partial x^{\alpha_1} \dots \partial x^{\alpha_d}} \right)^2 d\Omega \right]^{1/2}$$

Here, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_d]$ is a *multi-index* of non negative integers and $|\boldsymbol{\alpha}| = \alpha_1 + \dots + \alpha_n$, and we have also adopted the convention that $\frac{\partial^{|\boldsymbol{\alpha}|} u}{\partial x^{\alpha_1} \dots \partial x^{\alpha_d}} = u$ whenever $|\boldsymbol{\alpha}| = 0$. As a consequence of the definition, $\|u\|_{H^s(\Omega)} \le \|u\|_{L^2(\Omega)}$ and $H^s(\Omega) \subset L^2(\Omega)$, for all s. We can also conventionally set $H^0(\Omega) \equiv L^2(\Omega)$, so all previous definitions and properties extend trivially to the case $s = 0$ as well.

A most important space for the differential problems of our interest is $H^1(\Omega)$, where

$$(u,v)_{H^1(\Omega)} = \int_\Omega (uv + \boldsymbol{\nabla} u \cdot \boldsymbol{\nabla} v)\, d\Omega, \quad \|u\|_{H^1(\Omega)} = \sqrt{\int_\Omega (u^2 + ||\boldsymbol{\nabla} u||^2)\, d\Omega}.$$

Here, $||\boldsymbol{\nabla} u|| = \sqrt{\sum_{i=1}^3 (\partial u/\partial x_i)^2}$ indicates the Euclidean norm of the gradient. The H^s *seminorm*[4] is defined as

$$|u|_{H^s(\Omega)} = \sqrt{\sum_{|\boldsymbol{\alpha}|=s} \int_\Omega \left(\frac{\partial^{|\boldsymbol{\alpha}|} u}{\partial x^{\alpha_1} \dots \partial x^{\alpha_d}} \right)^2 d\Omega}.$$

We have that $H^{k+1}(\Omega) \subset H^k(\Omega)$ for $k = 0, 1, \dots$ with continuous injections, indeed if $u \in H^{k+1}(\Omega)$ then $\|u\|_{H^k(\Omega)} \leq \|u\|_{H^{k+1}(\Omega)}$.

We mention that in the next chapters, the symbol $[H^k(\Omega)]^3$ will be used to indicate the space of vector functions whose components belong to $H^k(\Omega)$, i.e. $[H^k(\Omega)]^3 = H^k(\Omega) \times H^k(\Omega) \times H^k(\Omega)$.

2.4.4 Traces

Let us first notice that two square integrable functions u_1 and u_2 which differ only on a set of zero measure identify in fact the same member of $L^2(\Omega)$, as $\|u_1 - u_2\|_{L^2(\Omega)}$. Being the boundary of Ω of zero measure it is clear that we cannot in general give a meaning to the value on $\partial\Omega$ of a function of $L^2(\Omega)$. Yet, what about a function belonging to $H^s(\Omega)$ with $s \geq 1$?

A major result is that if Ω is sufficiently regular, for instance polygonal or having a C^1 boundary (more details in the cited bibliography) there exists a linear and continuous application

$$\gamma_0 : H^s(\Omega) \mapsto \mathrm{L}^2(\partial\Omega),$$

such that $\gamma_0 v = v|_{\partial\Omega}$, $\forall v \in H^s(\Omega) \cap C^0(\overline{\Omega})$. The application $\gamma_0 v$ is called *trace* of v on $\partial\Omega$.

The fact that γ_0 is linear and continuous implies that $\exists C > 0$ so that

$$\|\gamma_0 v\|_{L^2(\partial\Omega)} \leq C \|v\|_{H^s(\Omega)}.$$

The result can be extended to the case of $\gamma_\Gamma : H^s(\Omega) \mapsto L^2(\Gamma)$ where $\Gamma \subset \partial\Omega$ is sufficiently regular and of non-null $d-1$ measure[5].

[4] A seminorm enjoys all properties of a norm a part that it can be zero when its argument is different from zero.

[5] $\partial\Omega$ is of zero d-measure, i.e. when considered as immersed in the space R^d, while in general has a non-zero $d-1$ measure. For instance the surface of a sphere has zero volume (3-measure) but a certain area (2-measure).

This result allows to give sense to the Dirichlet conditions whenever we seek a solution to a differential problem in $H^s(\Omega)$. We wish to point out that the operator γ_Γ is not onto $L^2(\Gamma)$. In particular, the functions of $L^2(\Gamma)$ which are traces of functions of $H^1(\Omega)$ is a subspace of $L^2(\Gamma)$ denoted as $H^{1/2}(\Gamma)$.

We can extend to functions in $H^s(\Omega)$ with $\Omega \subset \mathbb{R}^d$ and $d = 1, 2$ or 3 the well known Green integration and, consequently, the divergence theorem of classical vector calculus. In the following of the book for the sake of simplicity we will indicate $\gamma_0 u$ with $u_{|\partial\Omega}$, using the same notation adopted for a continuous function.

The space $H_0^1(\Omega)$

We can define the space $H_0^1(\Omega)$ as the space of functions with null trace on $\partial\Omega$

$$H_0^1(\Omega) = \{v \in H^1(\Omega) : \ \gamma_0 v = 0\}.$$

In fact, it is possible to define $H_0^1(\Omega)$ for arbitrary bounded domains Ω, using technicalities we prefer to avoid in this sketchy notes.

It is also possible to define $H_\Gamma^1(\Omega)$ as the space of function with null trace on $\Gamma \subset \partial\Omega$,

$$H_\Gamma^1(\Omega) = \{v \in H^1(\Omega) : \ \gamma_\Gamma v = 0\}.$$

An important result for what concerns the analysis of partial differential problems is the Poincaré inequality, which states that there exists a constant C_Ω such that

$$\|v\|_{L^2(\Omega)} \leq C_\Omega \|v\|_{H^1(\Omega)} \qquad \forall v \in H_0^1(\Omega).$$

2.4.5 Back to the weak formulation

We are now in the position of making expression (2.10) more precise. Indeed, by inspecting all the integrals that make up the bilinear form and the functional, we can note that the requirements we have to make so that all integrals exists and are finite is that $V = H_{\Gamma_D}^1$, provided that $f \in L^2(\Omega)$ and $h \in L^2(\Gamma_N)$. Actually, the conditions on the data indicated here are not the most general possible, yet are already quite broad to demonstrate the generality of the weak formulation.

We mention that with this choice problem (2.10) is well posed.

2.5 Conclusions

The objective of this chapter was twofold. On the one hand, we gave an overview of some basic mathematical models governing haemodynamics, with a greater emphasis on their physical significance and applicability rather than

on a rigorous formal derivation. To the latter are dedicated some of next chapters.

One the other hand, we provided some basic notions on methods for their numerical solution. We have also introduced some of the notation that will be used throughout the book. The notions given here are only elementary and have only the aim of introducing the occasional reader to the subject. The bibliography provided could however serve as a complement.