3.2 Finite Element Methods

In order to find a numerical estimate to the solution of a partial differential equation we need a way to approximate the operators involved. And while there are many different ideas of how to do so the one we have chosen to employ is a finite element approach, as they have shown to be one of the most powerful and versatile methodologies for the problem at hand [7]. The purpose of the subsequent section is not to establish the whole finite element framework from scratch but rather to provide the introduction of a unified notation that will be referred to throughout this thesis and a recollection of the most important properties needed. Anything else would be far beyond the scope this thesis, as a full description of the underlying mathematical constructions can quickly become rather involved but we would like to refer to [20] or *[good finite element specific source?!]* for a comprehensive *discussion* of the topic.

3.2.1 General Setting

The foundation of every finite element formulation is finding an appropriate weak formulation which includes the choice of suitable trial and solution spaces. This is especially applicable in the case of a least squares approach and will be discussed in further detail in section [...].
Given Banach spaces X and Y, a bounded linear operator $\mathcal{A} : X \to Y$, $f \in Y$, we consider the problem:

$$\text{Find } u \in X \text{ such that} \quad \mathcal{A}u = f \text{ in } Y. \tag{3.8}$$

We are interested in the case where $\mathcal{A}$ represents a partial differential operator. As mentioned before the process of discretisation begins with turning (3.6) into a suitable variational equation which is defined in terms of two Hilbert spaces V and W, a continuous bilinear form $a(\cdot,\cdot) : V \times W \to \mathbb{R}$, and a bounded linear functional $L_f(\cdot) : W \to \mathbb{R}$ and is given by

$$\text{Find } u \text{ in } V \text{ such that: } a(u,v) = L_f(v) \quad \forall v \in V \tag{3.9}$$

An operator equation such as (3.6) may be reformulated into several different variational equations. We can see that we were originally seeking for a solution u in the space X whereas in the weak formulation one attempts to find a solution in the space V, and which generally doesnt lie in X, and is therefore often referred to as a weak solution. Hence the relationship between the spaces X, Y and V, W, and the operator $\mathcal{A}$ and the bilinear form $a(\cdot,\cdot)$ are of great importance, and while one generally wants the solution of the variational formulation (3.7) to be a "good" representation of the solution of the original problem (3.6), the definition of what that exactly means varies and usually depends on the nature of the problem and often some practicality issues. One possibility could be ... or too much? Therefore we have denoted them by the same letter but to be precise the solution u appearing in the subsequent paragraphs will always be referring to $u \in V$, because our aim now is to solve the variational formulation.
So let us assume for now that we have found a suitable weak formulation of the operator equation where trial and test space are equal, that is $V = W$. In addition to $a(\cdot,\cdot)$ being linear and bounded, which is equivalent to the continuity, we will also require it to be symmetric, hence we have more specifically that

$a(v_1, v_2) = a(v_2, v_1)$ for all $v_1, v_2 \in V$ (symmetry)
$a(v_1, v_2) \leq \beta ||v_1||_V \cdot ||v_2||_V$, for all $v_1, v_2 \in V$ and $\beta > 0$ (boundedness)
$a(v_1, v_1) \geq \alpha ||v_1||_V^2$, for all $v_1 \in V$ and $\alpha > 0$ (coercivity)

and $f \in V^*$, the dual space of V. Furthermore let us have homogeneous Dirichlet boundary conditions, that is $u = v = 0$ on $\partial\Omega$. Then by *Riesz representation theorem/Lax-Milgram* we obtain that there exists a unique solution $u \in V$ that solves (2.2). And additionally the existence of an operator $\tilde{\mathcal{A}} : V \to V^*$ given by

$$a(u,v) = \langle \tilde{\mathcal{A}}u, v\rangle_{V^*,V} \quad \forall u, v \in V \tag{3.10}$$

where $\langle \cdot, \cdot \rangle$ denotes the duality pairing (more...) between V and its dual space $V*$. Likewise we obtain for $L_f(\cdot)$ the existence of an unique (!) element $\tilde{f}$ through the relation

$$L_f(v) = \langle \tilde{f}, v\rangle_{V^*,V} \quad \forall v \in V \tag{3.11}$$

The variational formulation is therefore equivalent to the problem

$$\text{Find } u \in V \text{ such that} \quad \tilde{\mathcal{A}}u = \tilde{f} \quad \text{in } W^* \tag{3.12}$$

In the special case that $X = U$ and $Y = W^*$ we have that $\mathcal{A} = \tilde{\mathcal{A}}$ and $f = \tilde{f}$ but this is generally not the case.

3.2.2 Discretisation

A key element to actually finding a good approximation u^h of u is to choose a suitable finite dimensional (sub)space V_h where we search for the solution. We will consider a *Galerkin approach*, where we indeed have $V_h \subset V$, which itself is again a Hilbert space and therefore the projected finite dimensional problem called Galerkin equation looks as follows

$$\text{Find } u_h \text{ in } V_h \text{ such that: } a(u_h, v_h) = L_f(v_h) \quad \forall v_h \in V_h \tag{3.13}$$

and has a unique solution itself. Since (2.2) holds for all $v \in V$ it also holds for all $v \in V_h$, and hence $a(u - u_h, v_h) = 0$, a key property known as Galerkin orthogonality. With respect to the energy norm induced by $a(\cdot, \cdot)$, u_h is a best approximation to u, in the sense that

$$\begin{aligned} ||u - u_h||_a^2 &= a(u - u_h, u - u_h) = a(u - u_h, u) + a(u - u_h, v_h) \\ &\le ||u - u_h||_a \cdot ||u - v_h||_a \quad \forall v_h \in V_h. \end{aligned} \tag{3.14}$$

We derive the third term from the second by using the Galerkin orthogonality. If we now divide both sides by $||u - u_h||_a$, we obtain that $||u - u_h||_a \le ||u - v_h||_a$ for all $v_h \in V_h$. We also have an estimate on $u - u_h$ in terms of the norm $||\cdot||_V$. Using the coercivity constant α and the bound from above β, we see that

$$\begin{aligned} \alpha||u - u_h||_V^2 &\le a(u - u_h, u - u_h) = a(u - u_h, u - u_h) = a(u - u_h, u + v_h - v_h - u_h) \\ &= a(u - u_h, u - v_h) + a(u - u_h, v_h - u_h) = a(u - u_h, u - v_h) \\ &\le \beta||u - u_h||_V \cdot ||u - v_h||_V \quad \forall v_h \in V_h. \end{aligned} \tag{3.15}$$

Dividing by $\alpha||u - u_h||$ we have shown *Céa's lemma*, which states that (accuracy ... constant thing):

$$||u - u_h||_V \le \inf_{v_h \in V_h} \frac{\beta}{\alpha}||u - v_h||_V, \quad u \in V, u_h \in V_h \tag{3.16}$$

where u is the solution to (2.2) and u_h to the corresponding finite dimensional problem (2.3). Hence accuracy of our approximation depends in this case on the constants α and β.

If we assume that we have a discretisation Ω_h of our domain Ω, where $h > 0$ is a parameter depending on the mesh size. We furthemore want to assume that as h tends to zero this implies that $dim(V_h) \Rightarrow \infty$. Additionally let $\{V_h : h > 0\}$ denote a family of finite dimensional subspaces of V, for which we assume that

$$\forall v \in V : \quad \inf_{v_h \in V_h} ||v - v_h||_V \to 0 \text{ as } h \to 0. \tag{3.17}$$

That is with a mesh size tending to zero there exist increasingly precise approximations for every $v \in V$, whose infimum tends to zero as the mesh size does. But then we can also conclude by the beforementioned properties (3.24) and (3.25) that $||u - u_h||_V \to 0$ as $h \to 0$. Hence our approximate solution u_h will converge to the weak solution u.

3.2.3 Matrix Formulation

After establishing these theoretical properties our aim is now to construct a linear system of equations that can be solved efficiently. Since V_h is a finite dimensional Hilbertspace, it has a countable basis $\{\phi_1, \phi_2, ..., \phi_n\}$ and we can write every element in V_h as a linear combination of such, that is we also have $u_h = \sum_{j=1}^{n} u_j \phi_j$, where $u_1, ..., u_n$ are constant coefficients. Writing (3.21) in terms of the basis we obtain by linearity

$$a(\sum_{j=1}^{n} u_j \phi_j, \phi_i) = \sum_{j=1}^{n} u_j a(\phi_j, \phi_i) = L_f(\phi_i) \quad \forall \phi_i, \ i = 1, 2, ..., n \tag{3.18}$$

If we now write this as a system of the form $A_h u_h = L_h$ with entries entries $(A_h)_{ij} = a(\phi_j, \phi_i)$, $(L_h)_i = L_f(\phi_i)$, then this becomes a linear system of equations which we can solve for an unknown vector u_h, where each matrix entry represents the evaluation of an integral expression. The question of how to choose favorable subspaces V_h, and a suitable basis for it has no trivial answer and depends on many factors and goes hand in hand with the question of how to best discretise the domain. Generally it seems like a sensible aim to opt for easily computable integrals giving rise to a linear system that is in turn as easy as possible to solve. Hence one objective might be to choose the basis $\{\phi_1, ..., \phi_n\}$ such that $supp(\phi_i) \cap supp(\phi_j) = \emptyset$ for as many pairs (i, j) as possible. Since this would ideally give rise to a sparse system of equations. It is also worth noting that due to the symmetry of $a(\cdot, \cdot)$, we have that $a_{ij} = a_{ji}$.

Depending on the operator $\mathcal{A}$, there is not necessarily a straight forward way to translate a strong formulation, that is a problem of the type (...), into a symmetric variational formulation, that is a symmetric bilinear form $a(\cdot, \cdot)$, which can subsequently be restricted to finite-dimensional subspaces and where we search for approximate solutions. However one possibility is through the differentiation of certain energy functionals, because we know by the theorem of Schwarz that order of differentiation with respect to partial derivatives is interchangeable and therefore leads to symmetry. How to construct these functional to be related to particular differential equations will be discussed in the following section.

3.3 Least Squares Finite Element Methods

In this section which is based on ([7], mainly ch. 2.1) we would like to introduce least squares finite element methods (LSFEMs), a class of methods for finding the numerical solution of partial differential equations that incorporates two main ideas concepts; the concept of finite elements and optimisation problems. They are based on the minimisation of functionals which are constructed from residual equations. Historically finite element methods were first developed and analysed for problems like linear elasticity whose solutions describe minimisers for convex, quadratic functionals over infinite dimensional Hilbert spaces and therefore actually emerged in