ICES REPORT 15-03

January 2015

Various Variational Formulations and Closed Range Theorem

by

Leszek Demkowicz

The Institute for Computational Engineering and Sciences
The University of Texas at Austin
Austin, Texas 78712

Reference: Leszek Demkowicz, "Various Variational Formulations and Closed Range Theorem," ICES REPORT 15-03, The Institute for Computational Engineering and Sciences, The University of Texas at Austin, January 2015.

Various Variational Formulations and Closed Range Theorem

Leszek Demkowicz

Institute for Computational Engineering and Sciences (ICES)
The University of Texas at Austin

Abstract

The report contains a set notes that I have prepared for the Functional Analysis class to illuminate the logical connections between different variational formulations, Banach Closed Range Theorems for Continuous and Closed Operators, and the related Babuška-Nečas Theorem. I use the simplest example of classical diffusion-convection-reaction and comment on two other applications: elasticity and Maxwell equations.

1 Introduction

The concept of a variational formulation and the Principle of Virtual Work is usually attributed to brothers Jacob and Johann Bernoulli, and it is related to the classical calculus of variations developed further by Leonhard Euler and Joseph-Louis Lagrange.

Consider the classical one-dimensional problem of calculus of variations:

$$J(u) = \int_a^b F(u'(x), u(x), x)\, dx \rightarrow \min_{u(a)=u_a} \tag{1.1}$$

where we seek a function $u(x),\ x \in [a,b]$, minimizing the cost functional $J(u)$ under the *essential boundary condition (BC):* $u(a) = u_a$. The cost functional is defined using a sufficiently regular integrand $F(u', u, x)$, a function of three variables, and the boundary data u_a is given.

Computing the variation (Gateáux derivative) of J, we get the necessary condition for solution u,

$$\left\{ \begin{array}{l} u(a) = u_a \\ \\ \langle \partial J, v \rangle = \displaystyle\int_a^b \left(\frac{\partial F}{\partial u'} v' + \frac{\partial F}{\partial u} v \right) dx = 0 \quad \forall v \,:\, v(a) = 0 \end{array} \right. \tag{1.2}$$

A few standard comments are in place. As the cost functional is defined by composing values of function $u(x)$ and its derivative $u'(x)$ with integrand F, we customary employ u', u, x to denote the arguments of F. Thus, symbol $\frac{\partial F}{\partial u'}$ denotes simply partial derivative of integrand F with respect to the first argument and nothing more (u' is here a "dummy" argument). Of course, both $\frac{\partial F}{\partial u'}$ and $\frac{\partial F}{\partial u}$ in (1.2) are functions of all three arguments u', u, x and they are evaluated at $u'(x), u(x), x$, respectively. Function[1] v is the test function (virtual displacement in language of mechanics) that must satisfy the homogeneous version of the essential BC. This follows from the computation of ∂J and the fact that $(u + \epsilon v)(a) = u_a$ for any ϵ.

[1] In classical notation v is replaced with δv.

Use of standard Fourier lemma argument leads to the Euler-Lagrange equation or, more precisely, the Euler-Lagrange boundary-value (BV) problem:

$$\begin{cases} u(a) = u_a \\ -\left(\frac{\partial F}{\partial u'}\right)' + \frac{\partial F}{\partial u} = 0 \quad \text{in } (a,b) \\ \frac{\partial F}{\partial u'}(u'(b), u(b), b) = 0\,. \end{cases} \tag{1.3}$$

Variational problem (1.2) and Euler-Lagrange BV problem (1.3) are formally equivalent. By using the word *formally*, we emphasize that we have not paid attention to regularity assumptions for solution u and test function v. Under appropriate assumptions, the two problems are equivalent indeed. If we employ the most general assumptions that make the variational problem well posed, the Euler-Lagrange BV problem has to be understood in the sense of theory of distributions and Sobolev spaces. The variational problem (and, therefore, the equivalent Euler-Lagrange BV problems) constitute, in general, only a necessary condition for the minimization problem (1.1). If integrand $F(u', u, x)$ is convex in u and strictly convex in u', the problems are equivalent.

If F is quadratic in u' and u then both the variational and Euler-Lagrange problems are linear. Consider, for instance, an elastic bar loaded with its own weight, see Fig. 1. The corresponding total potential energy

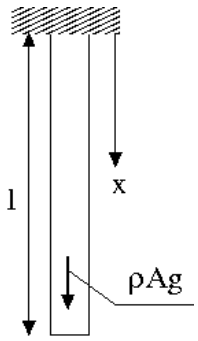

Figure 1: Elastic bar with stiffness EA and mass ρA.

is given by:

$$J(u) = \int_0^l \left[\frac{1}{2} EA(u')^2 - \rho A g u\right] dx \quad \text{where} \quad u(0) = 0\,, \tag{1.4}$$

and the resulted variational problem is,

$$\begin{cases} u(0) = 0 \\ \underbrace{\int_0^l EAu'v'\,dx}_{=:b(u,v)} = \underbrace{g\int_0^1 \rho Av\,dx}_{=:l(v)} \quad \forall v \,:\, v(0) = 0\,. \end{cases} \tag{1.5}$$

The left-hand side of the equation has the algebraic structure of a bilinear form $b(u,v)$ while the right-hand side represents a linear form $l(v)$. We arrive at the *abstract variational problem:*

$$\begin{cases} u \in U \\ b(u,v) = l(v) \quad v \in V \end{cases} \tag{1.6}$$

where, in this case, the *trial space* U and *test space* V are identical,

$$U = V := \{v \in H^1(0,l) \,:\, u(0) = 0\}\,. \tag{1.7}$$

Note that the essential BC has been built into the definition of the space. Above, $H^1(0,1)$ stands for the Sobolev space of order 1 and represents regularity assumptions for u and v.

In general, spaces U and V are different and may be complex-valued. In the complex case we need to decide whether we prefer the dual space to be defined as the space of linear or *antilinear* functionals. If we choose to work with antilinear functionals, form $b(u,v)$ need to be also antilinear in v, we say that b is *sesquilinear* ($1\frac{1}{2}$-linear). On the infinite-dimensional level, the choice is insignificant. Once we start discretizing the variational problem, the different settings may lead to different types of discretizations . For instance, for wave propagation problems, the bilinear setting is more appropriate in context of using the Perfectly Matched Layer (PML). In this exposition, we will stick with the antilinear setting.

It goes without saying that the forms $b(u,v)$ and $l(v)$ must be continuous or, equivalently, there exist constants $M > 0, C > 0$ such that

$$\begin{aligned} |b(u,v)| &\leq M\|u\|_U \,\|v\|_V\,, \\ |l(v)| &\leq C\|v\|_V\,. \end{aligned} \tag{1.8}$$

The continuity assumption and the Cauchy-Schwarz inequality lead usually to the choice of *energy spaces* U, V. For the beam problem, we have,

$$|\int_0^l EAu'v'| \leq \|EA\|_{L^\infty} \left(\int_0^l |u'|^2\right)^{1/2} \left(\int_0^l |v'|^2\right)^{1/2} \leq \|EA\|_{L^\infty}\|u\|_{H^1}\,\|v\|_{H^1}$$

and

$$|g\int_0^l \rho Av| \leq g\|\rho A\|_{L^2}\|v\|_{L^2} \leq g\|\rho A\|_{L^2}\|v\|_{H^1}\,.$$

We have thus estimates,

$$M \leq \|EA\|_{L^\infty}, \quad C \leq g\|\rho A\|_{L^2}\,.$$

Customary, by M, C we mean the *smallest*[2] possible constants.

[2] In the reflexive setting of Babuška-Nečas Theorem, the infimum is indeed attained, the smallest constants exist.

Babuška-Nečas Theorem. The fundamental tool for analyzing the well-posedness of variational problem (1.6) is provided by the following theorem.

THEOREM 1

(Babuška-Nečas)

Assume spaces U, V are reflexive, and forms $b(u,v), l(v)$ are continuous. Additionally, let form $b(u,v)$ satisfy the inf-sup condition:

$$\inf_{u\neq 0}\sup_{v\neq 0}\frac{|b(u,v)|}{\|u\|_U\|v\|_V} =: \gamma > 0 \quad \Leftrightarrow \quad \sup_{v\neq 0}\frac{|b(u,v)|}{\|v\|_V} \geq \gamma\|u\|_U\,, \tag{1.9}$$

and let $l(v)$ satisfy the compatibility condition:

$$l(v) = 0 \quad v \in V_0 \tag{1.10}$$

where

$$V_0 := \{v \in V \,:\, b(u,v) = 0 \quad \forall u \in U\}\,. \tag{1.11}$$

Then, the variational problem (1.6) is well-posed, i.e. there exists a unique solution u that depends continuously upon the data[3]*:*

$$\|u\|_U \leq \gamma^{-1}\|l\|_{V'} = \gamma^{-1}\sup_{v\neq 0}\frac{|l(v)|}{\|v\|_V}\,. \tag{1.12}$$

∎

Again, it goes without saying that γ is the best constant we can have. The proof of Babuška-Nečas Theorem is a direct reinterpretation of Banach Closed Range Theorem, see [7], Corollary 5.17.1. Indeed, the sesquilinear form $b(u,v)$ induces two linear and continuous operators,

$$B : U \to V', \quad \langle Bu, v\rangle_{V'\times V} = b(u,v)$$

$$B_1 : V \to U', \quad \langle B_1 v, u\rangle_{U'\times U} = \overline{b(u,v)}$$

Space V_0 is the null space of operator B_1. The inf-sup condition says that operator B is bounded below. The transpose operator B' goes from bidual V'' into dual U'. However, in the reflexive setting, spaces V and V'' are isometrically isomorphic and, therefore, operator B_1 can be identified with the transpose of B. Consequently, space V_0 can be identified as the null space of the transpose, and we arrive at the Closed Range Theorem.

We may argue that, from the operator theory (classical Functional Analysis) point of view, the nature of a variational problem lies in the fact that the operator takes values in a dual space (dual to the test space). We may want to mention that neither Closed Range Theorem nor Babuška-Nečas Theorem are constructive. They do not tell how to prove the inf-sup condition (boundedness below of operator B); they only tell you that you must have it.

[3] The problem is stable.

The goal of this note. It may be perhaps surprising that a boundary- or initial-boundary-problem may admit several variational formulations that differ by subtle differences in functional setting (regularity assumptions). Each of these formulations may lead to a different Galerkin, in particular Finite Element (FE) discretizations. The purpose of this exposition is to illustrate this point with a model problem, and show the intristic relations between the different variational formulations stemming from both versions of Closed Range Theorem - for continuous, and for closed operators. A short summary of most important facts about closed operators is provided in the Appendix.

2 Diffusion-Convection-Reaction Problem

Our model problem is the classical diffusion-convection-reaction problem. Given a domain $\Omega \subset \mathbb{R}^N$, $N \geq 1$, we wish to determine $u(x), x \in \overline{\Omega}$, that satisfies the boundary-value problem:

$$\left\{ \begin{array}{rcll} -(a_{ij}u_{,j})_{,i} + (b_i u)_{,i} + cu = & f & \text{in } \Omega \\ u = & 0 & \text{on } \Gamma_1 \\ a_{ij}u_{,j}n_j - b_i n_i u = & 0 & \text{on } \Gamma_2 \,. \end{array} \right. \tag{2.13}$$

Coefficients $a_{ij}(x) = a_{ji}(x), b_i(x), c(x)$ represent (anisotropic) diffusion, advection and reaction, and f stands for a source term. We are using the Einstein summation convention, the simplified, engineering notation for derivatives,

$$u_{,i} \stackrel{'}{=} \frac{\partial u}{\partial x_i} \,,$$

and n_i denote components of the unit outward vector on Γ. For instance, you can think of $u(x)$ being a temperature at point x, and $f(x)$ representing a heat source (sink) at x. Γ_1, Γ_2 represent two disjoint parts of the boundary. For the simplicity of the exposition, we will deal with homogeneous boundary conditions only.

Sobolev spaces

We will need some fundamental facts about two *energy spaces*. The first is the classical H^1 Sobolev space consisting of all L^2-functions whose distributional derivatives are also functions, and they are L^2-integrable as well,

$$H^1(\Omega) := \{u \in L^2(\Omega) \,:\, \frac{\partial u}{\partial x_i} \in L^2(\Omega),\, i = 1, \ldots, N\} \tag{2.14}$$

The space is equipped with the norm,

$$\|u\|^2_{H^1} := \|u\|^2 + \sum_{i=1}^{N} \|\frac{\partial u}{\partial x_i}\|^2$$

where $\|\cdot\|$ denotes the L^2-norm. The second term constitues a seminorm on $H^1(\Omega)$ and will be denoted by

$$|u|^2_{H^1} := \sum_{i=1}^{N} \|\frac{\partial u}{\partial x_i}\|^2 .$$

The second space, $H(\mathrm{div}, \Omega)$, consists of all vector-valued L^2-integrable functions whose distributional divergence is also a function, and it is L^2-integrable,

$$H(\mathrm{div}, \Omega) := \{\sigma = (\sigma_i)_{i=1}^N \in (L^2(\Omega))^N \ : \ \mathrm{div}\ u \in L^2(\Omega)\} . \tag{2.15}$$

The space is equipped with the norm,

$$\|\sigma\|^2_{H(\mathrm{div})} := \|\sigma\|^2 + \|\mathrm{div}\sigma\|^2$$

where the L^2-norm of vector-valued functions is computed componentwise,

$$\|\sigma\|^2 := \sum_{i=1}^{N} \|\sigma_i\|^2 .$$

For both energy spaces, there exist *trace operators* that generalize the classical boundary trace for scalar-valued functions, and boundary normal trace for vector-valued functions,

$$u \to u|_\Gamma, \quad \sigma \to \sigma_n = \sum_{i=1}^{N} \sigma_i|_\Gamma n_i .$$

Above, $u|_\Gamma$ denotes restriction of funtion u to boundary Γ, and n_i is the outward normal unit vector on Γ. The very concept of the trace operator is non-trivial. Recall that elements of $L^2(\Omega)$ are not really functions but *equivalence classes* of functions that are equal *almost everywhere*, meaning everywhere except for a subset of measure zero. Well, the boundary is [4] of measure zero and, therefore, the usual boundary traces may be different for different representatives of u. Speaking of a trace for an L^2 function is mathematically illegal.

The trace operators map *continuously* $H^1(\Omega)$ and $H(\mathrm{div}, \Omega)$ *onto* another energy spaces defined on boundary Γ, fractional Sobolev spaces $H^{1/2}(\Gamma)$ and $H^{-1/2}(\Omega)$ that are in *duality pairing*[5]. In what follows, we shall not introduce any notation for trace operators but it will go without saying that, whenever we speak of u and σ_n on Γ, we understand these objects in the sense of the trace operators. In particular, the classical integration by parts formula generalizes to the energy spaces,

$$(\sigma, \nabla u) = -(\mathrm{div}\ \sigma, u) + \langle \sigma_n, u \rangle, \quad \sigma \in H(\mathrm{div}, \Omega),\ u \in H^1 . \tag{2.16}$$

The brackets denote the duality pairing between $H^{1/2}(\Gamma)$ and $H^{-1/2}(\Gamma)$ that generalizes the usual integral over the boundary.

[4] We are using the N-dimensional measure.

[5] $H^{-1/2}(\Omega)$ is the topological dual of $H^{1/2}(\Gamma)$.

We can use the trace operators now to impose boundary conditions on u and σ. We will need the following subspaces of the energy spaces,

$$\begin{aligned} H^1_{\Gamma_1} &:= \{u \in H^1(\Omega) \,:\, u = 0 \text{ on } \Gamma_1\}\,, \\ H_{\Gamma_2}(\text{div},\Omega) &:= \{\sigma \in H(\text{div},\Omega) \,:\, \sigma_n = 0 \text{ on } \Gamma_2\}\,. \end{aligned} \tag{2.17}$$

As for regular functions, we have

$$(\sigma, \nabla u) = -(\text{div}\,\sigma, u) \quad \sigma \in H_{\Gamma_2}(\text{div},\Omega),\, u \in H^1_{\Gamma_1}\,.$$

Finally, we recall *density results*. For sufficiently regular [6] domain Ω, $C^\infty(\overline{\Omega})$ functions that satisfy the BCs, are *dense* in the energy spaces, i.e. for each function $u \in H^1_{\Gamma_1}$, there exists a sequence of regular functions $u^j \in C^\infty(\overline{\Omega})$, $u^j = 0$ on Γ_1, converging to u in H^1-norm. Similarly, for each $\sigma \in H_{\Gamma_2}(\text{div},\Omega)$, there exists a sequence of regular functions $\sigma^j \in (C^\infty(\overline{\Omega}))^N$, $\sigma_n = 0$ on Γ_2, converging to σ in the $H(\text{div})$-norm.

3 Classical variational formulation.

We will follow the usual strategy by deriving various variational formulations first formally, and only afterward discussing their functional setting and well-posedness. We will make appropriate assumptions on material data a_{ij}, b_i, c and load data f on the fly, as needed.

We multiply equation $(2.13)_1$ with a test function $v(x)$, integrate over domain Ω and integrate diffusion and convection terms by parts to obtain,

$$\int_\Omega (a_{ij}u_{,j}v_{,i} - b_i u v_{,i} + cuv)\,dx \int_\Gamma (-a_{ij}u_{,j}n_j + b_i n_i u)v\,ds = \int_\Omega fv\,dx\,.$$

By virtue of the second boundary condition, the boundary term vanishes on Γ_2. If we choose the test function v to vanish on Γ_1, the boundary term vanishes all-together.

We need to set up the energy spaces now. If we choose to work with a symmetric setting for u and v, the natural choice is the first order Sobolev space with the first boundary condition built in, discussed earlier,

$$U = V := H^1_{\Gamma_1} = \{v \in H^1(\Omega) \,:\, v = 0 \text{ on } \Gamma_1\}\,.$$

We have arrived at the variational formulation:

$$\begin{cases} u \in H^1_{\Gamma_1} \\ \displaystyle\int_\Omega (a_{ij}u_{,j}v_{,i} - b_i u v_{,i} + cuv) = \int_\Omega fv \quad v \in H^1_{\Gamma_1}\,. \end{cases} \tag{3.18}$$

Continuity requirements and Cauchy-Shwarz inequality lead to the following assumptions on the data:

$$a_{ij}, b_i, c \in L^\infty(\Omega), \quad f \in L^2(\Omega)\,. \tag{3.19}$$

[6] Lipshitz domain.

Using integral and discrete Cauchy-Shwarz inequalities, we get,

$$
\begin{aligned}
&|\int_\Omega (a_{ij}u_{,j}v_{,i} - b_i u v_{,i} + cuv)| \\
&\le |\int_\Omega a_{ij}u_{,j}v_{,i}| + |\int_\Omega b_i u v_{,i}| + |\int_\Omega cuv| \\
&\le \|a_{ij}\|_\infty \|u_{,j}\| \|v_{,i}\| + \|b_i\|_\infty \|u\| \|v_{,i}\| + \|c\|_\infty \|u\| \|v\| \\
&\le \max_{i,j} \|a_{ij}\|_\infty N(\sum_i \|u_{,i}\|^2)^{1/2}(\sum_j \|v_{,j}\|^2)^{1/2} + \max_i \|b_i\|_\infty \sqrt{N}\|u\|(\sum_j \|v_{,j}\|^2)^{1/2} + \|c\|_\infty \|u\|\|v\| \\
&\le 3\max\{N \max_{i,j} \|a_{ij}\|_\infty, \sqrt{N} \max_i \|b_i\|_\infty, \|c\|_\infty\} \|u\|_{H^1}\|v\|_{H^1} .
\end{aligned}
$$

The estimate results in an upper bound for the continuity constant M,

$$M \le 3\max\{N \max_{i,j} \|a_{ij}\|_\infty, \sqrt{N} \max_i \|b_i\|_\infty, \|c\|_\infty\} .$$

By no means we claim the upper bound to be optimal.

Similarly, we have,

$$|\int_\Omega fv| \le \|f\|\|v\| \le \|f\|\|v\|_{H^1} ,$$

so $C \le \|f\|$.

Remark 1 Please note that we can accommodate a more irregular source term. For instance, we could add to the right hand-side a term like

$$\int_\Omega g_i v_{,i} .$$

With $g_i \in L^2(\Omega)$, the term is clearly continuous on $H^1(\Omega)$. Integrating by parts, we realize that it may represent a distributional load. Indeed, e.g. with $g_i n_i = 0$ on Γ_2, we have

$$\int_\Omega g_i v_{,i}\, dx = \int_\Omega g_{i,i} v\, dx$$

and divergence $g_{i,i}$ may not be a function. ∎

With a little bit of practice in using the integral and discrete versions of Cauchy-Schwarz inequality, demonstrating continuity of the bilinear (sesquilinear) and linear (antilinear) forms becomes very standard and we will not repeat it again.

Having shown the continuity, we turn to the inf-sup condition now. We will take the standard "easy way out" and make assumptions under which the bilinear form $b(u,v)$ is *V-coercive*, i.e. there exists a positive constant $\alpha > 0$ such that

$$|b(v,v)| \ge \alpha \|v\|^2_{H^1} \quad v \in H^1_{\Gamma_1} .$$

Indeed, coercivity implies automatically the inf-sup condition,

$$\sup_{v\neq 0}\frac{|b(u,v)|}{\|v\|_{H^1}} \geq \frac{|b(u,u)|}{\|u\|_{H^1}} \geq \alpha\frac{\|u\|^2_{H^1}}{\|u\|_{H^1}} = \alpha\|u\|_{H^1}\,.$$

Note that the coercivity condition requires the symmetric energy setting, $U = V$.

The assumptions to guarantee the coercivity of $b(u,v)$ are as follows.

Strong ellipticity assumption for the diffusion coefficients:

$$a_{ij}(x)\xi_i\xi_j \geq \alpha_0\xi_i\xi_i, \qquad \alpha_0 > 0,\, x\in\Omega\,, \tag{3.20}$$

Incompressibility assumption for the advection field:

$$b_{i,i}(x) = 0, \quad x\in\Omega\,,$$

Inflow assumption on Γ_2:

$$b_i n_i \leq 0 \quad \text{on } \Gamma_2\,,$$

Non-negativity of the reaction coefficient:

$$c(x) \geq 0, \quad x\in\Omega\,.$$

With these assumptions in place, we have,

$$b(u,u) = \int_\Omega a_{ij}u_{,j}u_{,i} - \int_\Omega b_i u u_{,i} + \int_\Omega cu^2$$

where

$$\int_\Omega a_{ij}u_{,j}u_{,i} \geq \alpha_0\int_\Omega u_{,j}u_{,j} \geq \alpha_0|u|^2_{H^1}\,,$$

$$-\int_\Omega b_i u u_{,i} = -\int_\Omega b_i\left(\frac{u^2}{2}\right)_{,i} = \frac{1}{2}\int_\Omega b_{i,i}u^2 - \int_\Gamma b_i n_i u^2 = -\int_{\Gamma_2} b_i n_i u^2 \geq 0\,,$$

and,

$$\int_\Omega cu^2 \geq 0\,.$$

We are close but not there yet. In order to show the coercivity, we need one more argument.

Lemma 1

(Poincaré Inequality)

There exists a positive constant $c > 0$ such that

$$c\|u\|^2 \leq |u|^2_{H^1} \tag{3.21}$$

for all $u \in H^1(\Omega)$ *that vanish on* Γ_1. ∎

The Poincaré Inequality implies that seminorm $|u|_{H^1}$ is actually a norm equivalent to $\|u\|_{H^1}$. Indeed,

$$\|u\|^2_{H^1} = \|u\|^2 + |u|^2_{H^1} \leq (c^{-1}+1)|u|^2_{H^1}\,.$$

Coercivity implies also that the transpose operator is injective. Indeed, assume $v \in V_0$. Then

$$b(u,v) = 0 \quad \forall u \in U\,.$$

In particular $b(v,v) = 0$ which implies that $v = 0$. The null space V_0 is thus trivial and, therefore, the compatibility condition is automatically satisfied for any $l \in V'$.

In conclusion, all assumptions of Babuška-Nečas Theorem have been verified, and we know that the variational problem is well posed for *any* $f \in L^2(\Omega)$.

4 First Order Setting. Trivial and Ultraweak Formulations

First order system setting. We introduce an extra variable, the *flux*,

$$\sigma_i = a_{ij}u_{,j} - b_i u\,.$$

The ellipticity assumption implies that a_{ij} is invertible. Introducing the inverse matrix $\alpha_{ij} = (a_{ij})^{-1}$, and multiplying the equation above by the inverse, we obtain,

$$\alpha_{ij}\sigma_j = u_{,i} - \beta_i u\,,$$

where $\beta_i = \alpha_{ij}b_j$.

Our new formulation reads now:

$$\left\{\begin{array}{rcll} \alpha_{ij}\sigma_j - u_{,i} + \beta_i u = & g_i & \text{in } \Omega \\ -\sigma_{i,i} + cu = & f & \text{in } \Omega \\ u = & 0 & \text{on } \Gamma_1 \\ \sigma_i = & 0 & \text{on } \Gamma_2 \end{array}\right. \tag{4.22}$$

where we have thrown in an additional right-hand side g_i in the first equation. For the original problem, $g_i = 0$.

We can multiply now the first equation with a vector-valued test function τ_i, the second equation with a scalar-valued test function v, and integrate over domain Ω. It will be convenient now to switch to the L^2-inner product notation,

$$(u,v) = (u,v)_{L^2} = \int_\Omega uv\,dx\,.$$

We have,

$$(\alpha_{ij}\sigma_j, \tau_i) - (u_{,i}, \tau_i) + (\beta_i u, \tau_i) = (g_i, \tau_i)$$
$$-(\sigma_{i,i}, v) + (cu, v) = (f, v)$$

or, using vector notation,

$$(\alpha\sigma, \tau) - (\nabla u, \tau) + (\beta u, \tau) = (g, \tau)$$
$$-(\text{div } \sigma, v) + (cu, v) = (f, v)$$

Either of the two equations can now be left alone or we can *relax it*, i.e. integrate by parts and build the appropriate BC in. Dependent upon the choice we make, we end up with one of four possible variational formulations. In this section we will discuss two of them, the *trivial formulation* where both equations are left alone, and the *ultraweak formulation* where both equations are relaxed.

4.1 Closed Operators Formalism

The first order system can be studied using the theory of closed operators. The operator governing the problem is defined as,

$$A((\sigma, u)) := (\alpha\sigma - \nabla u + \beta u, -\text{div } \sigma + cu)$$

with the domain $D(A)$ defined the usual way to guarantee that A is well defined,

$$\begin{aligned} D(A) &:= \{(\sigma, u) \in (L^2(\Omega))^N \times L^2(\Omega) \,:\, A(\sigma, u) \in (L^2(\Omega))^N \times L^2(\Omega), \quad \sigma_n = 0 \text{ on } \Gamma_2, \quad u = 0 \text{ on } \Gamma_1\} \\ &= H_{\Gamma_2}(\text{div}, \Omega) \times H^1_{\Gamma_1}(\Omega) \end{aligned}$$

Note that the graph norm

$$\begin{aligned} \|(\sigma, u)\|^2_G &:= \|(\sigma, u)\|^2 + \|A(\sigma, u)\|^2 \\ &= \|\sigma\|^2 + \|u\|^2 + \|\alpha\sigma - \nabla u + \beta u\|^2 + \| - \text{div } \sigma + cu\|^2 \end{aligned} \tag{4.23}$$

and the $H(\text{div}, \Omega) \times H^1(\Omega)$-norm *are equivalent*.

We can use now the Closed Range Theorem for Closed Operators to study the well-posedeness of the first order system. We need to check that A is closed, investigate its boundedness below, and determine the adjoint A^* and its null space.

Showing closedness of A relies on understanding of derivatives in the distributional sense and properties of Sobolev spaces. Given a sequence $(\sigma^j, u^j) \in D(A)$ such that

$$\sigma^j \to \sigma, \quad u^j \to u, \quad \text{and } A(\sigma^j, u^j) =: (g^j, f^j) \to (g, f)\,,$$

we need to conclude that $(\sigma, u) \in D(A)$ and $A(\sigma, u) = (g, f)$. We use upper indices to denote the sequence as we have already used the lower index to denote components of σ and g. All convergence is understood in the L^2 sense.

We start with the first equation. Let $\psi \in (C_0^\infty(\Omega))^N$ be a an arbitrary, vector-valued test function. By the definition of distributional derivatives, we have

$$\int_\Omega \alpha\sigma^j\psi + \int_\Omega u^j \text{div}\ \psi + \int_\Omega u^j \beta\cdot\psi = \int_\Omega g\phi\,.$$

Passing to the limit with $\sigma^j \to \sigma$ and $u^j \to u$, and using the continuity of the L^2 inner product, we get,

$$\int_\Omega \alpha\sigma\psi + \int_\Omega u\text{div}\ \psi + \int_\Omega u\beta\cdot\psi = \int_\Omega g\phi\,.$$

Due to arbitrariness of the test function, this proves that the equation is indeed satisfied in the distributional sense. The same reasoning applies to the second equation. Thus, indeed $A(\sigma,u) = (g,f)$. In order to argue that (σ,u) are in $D(A)$, we need to show yet that they satisfy the appropriate BCs. This relies on the knowledge of Sobolev spaces. As noted above, the convergence in the graph norm (which we have assumed) is equivalent to the convergence of $\sigma^j \to \sigma$ in $H(\text{div},\Omega)$ space, and convergence $u^j \to u$ in the $H^1(\Omega)$ space. The boundary values of all involved functions are understood in the sense of traces, and the trace operators are continuous. Done.

Similarly to the proof of continuity of bilinear and linear forms, the proof of closedness is very standard and in papers it is usually reduced to a short comment.

The procedure to show the boundedness below is as follows. We take $(\sigma,u) \in D(A)$ and consider system (4.22) with g_i, f *being defined* by the equations. We need to show then the existence of $C>0$ such that

$$\|\sigma\|, \|u\| \leq C(\|g\| + \|f\|)\,.$$

The density results for the Sobolev spaces come handy here. We can assume first that σ, u are $C^\infty(\overline{\Omega})$ functions so we can understand all derivatives and boundary values in the classical sense. Once we establish the L^2 bounds for the regular functions, we use the density argument to conclude that they also hold for all σ, u from the domain of the operator.

Using the first equation to eliminate flux σ, we arrive at our original problem with a more general right-hand side,

$$-(a_{ij}u_{,j})_{,i} + (b_i u)_{,i} + cu = f + g_{i,i}\,.$$

Multiplying with a test function v, integrating over Ω, integrating by parts, and using BCs, we realize that u satisfies the classical variational formulation with a linear functional

$$l(v) = \int_\Omega fv + g_i v_{,i} \quad |l(v)| \leq (\|f\| + \|g\|)\|v\|_{H^1}\,.$$

Consequently, if we use the well-posedness of the classical variational formulation, we obtain,

$$\|u\|_{H^1} \leq \gamma^{-1}\|l\|_{V'} \leq \gamma^{-1}(\|f\| + \|g\|)\,.$$

This proves the L^2 bound for u. In order to establish the bound for σ, we need to use the definition of σ,

$$\|\sigma\| \leq C_1|u|_{H^1} + C_2\|u\| \leq (C_1+C_2)\|u\|_{H^1} \leq (C_1+C_2)\gamma^{-1}(\|f\| + \|g\|)$$

where C_1, C_2 depend upon a_{ij} and b_j,

$$C_1 = \sup_x \|a_{ij}(x)\|, \quad C_2 = \max_j \|b_j\|_{L^\infty(\Omega)}$$

with $\|a_{ij}\|$ denoting the norm in space $L(\mathbb{R}^n, \mathbb{R}^n)$ (maximum singular value of the matrix).

In order to conclude the proof of boundedness below, we need to use density results mentioned above. For any $(\sigma, u) \in D(A)$, let (σ^j, u^j) be the sequence of regular functions from the domain of A discussed earlier that converges to (σ, u). We have now,

$$\|\sigma^j\|, \|u^j\| \leq (\|g^j\| + \|f^j\|)$$

where $(g^j, f^j) = A(\sigma^j, u^j)$. Convergence in (σ^j, u^j) in $H(\text{div}, \Omega) \times H^1(\Omega)$ is equivalent to convergence in the graph norm so, passing to the limit in the equation above, we can the required result.

Again, the reasoning is standard but by no means trivial.

Remark 2 We could have reproduced the reasoning based on the coercivity for the classical variational formulation to establish the bound. The point in referring to the stability of the classical variational formulation directly, is to realize the relation between the inf-sup constants. More importantly, if the classical variational formulation is stable under weaker assumptions on coefficients, the result will automatically imply that operator A is bounded below as well. ∎

Once we have proved that the operator is closed and bounded below, we proceed with the determination of the adjoint. A direct integration by parts leads to the formula:

$$\begin{aligned}(A(\sigma, u), (\tau, v) &= (\alpha\sigma, \tau) - (\nabla u, \tau) + (\beta u, \tau) - (\text{div } \sigma, v) + (cu, v) \\ &= (\sigma, \alpha\tau + \nabla v) + (u, \text{div } \tau + \beta \cdot \tau + cv) - \int_\Gamma \sigma_n v - \int_\Gamma u\tau_n\end{aligned}$$

Due to boundary conditions on (σ, u), the boundary terms reduce to

$$-\int_{\Gamma_1} \sigma_n v - \int_{\Gamma_2} u\tau_n \,.$$

Due to arbitrariness of (σ, u), this implies BCs for the adjoint that have to be incorporated into the definition of its domain,

$$\begin{aligned}D(A^*) &= \{(\tau, v) \in (L^2(\Omega))^N \times L^2(\Omega) \,:\, A^*(\tau, u) \in (L^2(\Omega))^N \times L^2(\Omega) \text{ and } \tau_n = 0 \text{ on } \Gamma_2,\, v = 0 \text{ on } \Gamma_1\} \\ &= H_{\Gamma_2}(\text{div}, \Omega) \times H^1_{\Gamma_1}(\Omega) \\ A^*(\tau, v) &= (\alpha\tau + \nabla v, \text{div } \tau + \beta \cdot \tau + cv)\,.\end{aligned} \tag{4.24}$$

We proceed now with the determination of the null space of the adjoint. Assuming again regularity for granted, we eliminate τ to obtain a single equation on v,

$$-\text{div}\,(a\nabla v) - b\cdot\nabla v + cv = 0$$

Except for the different sign in front of the convection term, this is our original diffusion-convection-reaction problem. We now repeat our reasoning in the proof of coercivity of the bilinear form for the classical variational formulation to learn that v must vanish. The null space of the adjoint is thus trivial.

Summing up, we have used the closed operators theory to show that the first order system has a unique solution for any right hand side g, f. Boundedness below of A implies the stability of the solution in the L^2-norm,

$$\|\sigma\|^2 + \|u\|^2 \leq c^2(\|g\|^2 + \|f\|^2)\,.$$

Relation between different stability constants. The first order formulation was strongly advocated by Kurt Otto Friedrichs, one of the co-founders of the Courant Institute. For that reason, I will call the L^2 boundedness constant c above the *Friedrichs constant*.

The first order problem can now be easily interpreted using theory of continuous operators. The modification is very simple. We equip $D(A)$ with the graph norm and identify it as a new energy space and the domain of the continuous operator.

$$X = D(A), \quad \|u\|^2 := \|u\|^2 + \|Au\|^2$$

We keep the L^2 space for Y. The operator $A : X \to Y$ is then trivially continuous with continuity constant equal one. The continuous operator A is bounded below as well. Adding side-wise,

$$c^{-2}\|Au\|^2 \geq \|u\|^2, \quad \|Au\|^2 = \|Au\|^2$$

we obtain[7],

$$\text{(new) } c \overset{'}{=} (c^{-2}+1)^{-1/2}\,. \tag{4.25}$$

4.2 Trivial variational formulation

We leave both equations alone. The variational problem looks as follows.

$$\begin{cases} \sigma \in H_{\Gamma_2}(\text{div},\Omega), u \in H^1_{\Gamma_1}(\Omega) \\ (\alpha\sigma,\tau) - (\nabla u,\tau) + (\beta u,\tau) = (g,\tau) \quad \tau \in (L^2(\Omega))^N \\ -(\text{div}\,\sigma, v) + (cu, v) = (f, v) \quad v \in L^2(\Omega) \end{cases} \tag{4.26}$$

[7] If Friedrichs constant for the closed operator A is realized by a function u, the same function realizes the Friedrichs constant for the corresponding continuous operator.

In order to fit the problem into our abstract setting, we have to introduce group variables,

$$u \overset{\prime}{=} (\sigma, u), \quad v \overset{\prime}{=} (\tau, v)\,.$$

The accent over the equality sign indicates a metalanguage. As we are running out of letters, we *overload*[8] symbols u, v which now have a different meaning dependent upon the context. Life is tough...

The energy spaces are thus,

$$U := H_{\Gamma_2}(\text{div}, \Omega) \times H^1_{\Gamma_1}(\Omega)$$
$$V := (L^2(\Omega))^N \times L^2(\Omega)\,.$$

Finally, the bilinear and linear forms can be obtained by simply summing up the two individual equations,

$$b(u, v) \overset{\prime}{=} b((\sigma, u), (\tau, v)) = (\alpha\sigma, \tau) - (\nabla u, \tau) + (\beta u, \tau) + -(\text{div}\ \sigma, v) + (cu, v)$$
$$l(v) \overset{\prime}{=} l((\tau, v)) = (g, \tau) + (f, v)\,.$$

The reason we termed the variational formulation to be trivial is that it is fully equivalent to the strong, continuous operator setting. Indeed, the variational statement

$$f \in L^2(\Omega), \quad (f, v) = 0 \quad \forall v \in L^2(\Omega)$$

is equivalent to the strong statement,

$$f \in L^2(\Omega), \quad f = 0$$

where the equality is understood in the L^2-sense, i.e. almost everywhere. The inf-sup constant for the bilinear form is equal to the Friedrichs constant for the continuous operator if we use the graph norm for the trial space. Indeed, $b(u, v) = (Au, v)$ implies

$$\sup_{v \neq 0} \frac{|b(u, v)|}{\|v\|} = \sup_{v \neq 0} \frac{|(Au, v)|}{\|v\|} = \|Au\| \geq c\|u\|_G$$

If we trade the graph norm for the Sobolev norms, the norm equivalence constants will come in.

4.3 Ultraweak Variational Formulation

If we relax (integrate by parts) both equations in the first order system, we obtain the co-called *ultraweak variational formulation*. In the language of the adjoints (for closed operators), we obtain,

$$\begin{cases} u \in L^2 \\ (u, A^* v) = 0 \quad \forall v \in D(A^*)\,. \end{cases}$$

[8] A term used also by programmers.

We have thus for our model problem,

$$\begin{aligned} U &= (L^2(\Omega))^N \times L^2(\Omega) \\ V &= D(A^*) = H_{\Gamma_2}(\text{div}, \Omega) \times H^1_{\Gamma_1}(\Omega) \\ b((\sigma, u), (\tau, v)) &= (u, A^* v) = (\sigma, \alpha\tau + \nabla v) + (u, \text{div}\, \tau + \beta \cdot \tau + cv) \\ l(\tau, v)) &= (g, \tau) + (f, v)\,. \end{aligned} \tag{4.27}$$

And here is the exciting news. We do not need to prove anything new. The operator B corresponding to ultraweak variational formulation goes from $U \to V'$. Its transpose $B' : V \to U' \sim L^2$ corresponds to the trivial variational formulation for the adjoint operator and it was already shown to be bounded below. By the Closed Range Theorem for continuous operators, the inf-sup constant for the ultraweak formulation equals the inf-sup constant for operator B'. This one in turn can be expressed in terms of the Friedrichs constant for A^* using formula (4.25). However, by the Closed Range Theorem for Closed Operators, the Friedrichs constants for A and A^* are equal. Consequently, inf sup constants for the trivial and ultraweak variational formulations are identical. The trivial and ultraweak variational formulations are simultaneously well or ill-posed.

Remark 3 In our case, A^* is injective. The conclusions above hold as well if A^* is not injective. One has then to switch in the reasoning to quotient spaces. ∎

5 Mixed formulations

We discuss now the two remaining variational formulations where only one of the equations is relaxed.

Classical Variational Formulation Revisited

What will happen if we relax only the second equation in (4.22) ? After integration by parts the second equation turns into:

$$(\sigma, \nabla v) - \int_\Gamma \sigma_n v + (cu, v) = (f, v)$$

If we take into account the BC for σ_n and assume $v = 0$ on Γ_1, the boundary term disappears. As the first equation is equivalent to the strong form, we can use it to eliminate σ, and we arrive precisely at the classical variational formulation. In this sense, the classical variational formulation presents itself as one of the four natural choices that we have.

Equivalently, if we stick with the first order formalism, we have,

$$\begin{aligned} U = V &= \{(\sigma, u) \in (L^2(\Omega)^N) \times H^1(\Omega) \,:\, u = 0 \text{ on } \Gamma_1\} \\ &= (L^2(\Omega))^N \times H^1_{\Gamma_1}(\Omega) \\ b((\sigma,u),(\tau,v)) &= (\alpha\sigma,\tau) = (\nabla u,\tau) + (\beta u,\tau) + (\sigma,\nabla v) + (cu,v) \\ l((\tau,v)) &= (g,\tau) + (f,v) \end{aligned} \tag{5.28}$$

So, how do we show the inf-sup condition now ?

Strategy I: We follow the same steps as in the proof of the boundedness below for the strong form of the operator. For any $(\sigma, u) \in U$, we define,

$$\begin{aligned} \alpha\sigma - \nabla u + \beta u =: &\quad g \\ (\sigma, \nabla v) + (cu, v) =: &\quad \langle f, v\rangle \end{aligned}$$

and we need to find a constant c such that

$$\|\sigma\|^2 + \|u\|^2_{H^1} \leq c\left(\|g\|^2 + \|f\|^2_*\right)$$

Note that above g is a function but f is a functional and it is measured in the dual norm,

$$\|f\|_* := \|f\|_{(H^1_{\Gamma_1}(\Omega))'} = \sup_{v \in H^1_{\Gamma_1}(\Omega)} \frac{|\langle f, v\rangle|}{\|v\|_{H^1}}\,.$$

Using the first equation to eliminate σ we arrive at the old friend:

$$(a\nabla u, \nabla v) - (bu, \nabla v) + (cu, v) = \langle f, v\rangle - (ag, \nabla v)$$

Using the inf-sup condition for the classical variational formulation we get,

$$\|u\|_{H^1} \leq \gamma^{-1}(\|f\|_* + \|ag\|)$$

where γ is the inf-sup constant for the classical variational formulation. The estimate for σ follows now easily from the first equation. The inf-sup constant can now easily be estimated in terms of γ.

Strategy II. Once we know that the ultraweak variational formulation is well posed, we can use a simpler reasoning. The first, strong equation can be interpreted in the weak sense. Multiplying it with a test function $\tau \in H(\text{div}, \Omega), \tau_n = 0$ on Γ_2, and integrating by parts, we obtain,

$$(\alpha\sigma, \tau) + (u, \text{div}\,\tau) + (\beta u, \tau) = (g, \tau)\,.$$

Thus (σ, u) is a solution to the ultraweak variational formulation as well, and we can conclude the bound in the L^2-norm,

$$\|\sigma\|^2 + \|u\|^2 \leq c(\|g\|^2 + \|f\|_*)\,.$$

But once we have established the bound for the L^2-norms of σ and u, we can use the first equation to establish also a bound for the L^2-norm of ∇u. Indeed,

$$\|\nabla u\| = \|\alpha\sigma + \beta u + g\| \leq C_1\|\sigma\| + C_2\|u\| + \|g\|$$

where constants C_1, C_2 depend upon α and β, respectively.

The conjugate operator acts from V to U', has a similar structure, is injective and, by the Closed Range Theorem for Continuous Operators, has the same inf-sup constant.

Mixed formulation

The final, so-called *mixed formulation*, is obtained by relaxing the first equation and keeping the second one in the strong form. We obtain,

$$\begin{aligned} U = V &= \{(\sigma, u) \in H(\text{div}, \Omega) \times L^2(\Omega) \,:\, \sigma_n = 0 \text{ on } \Gamma_2\} \\ &= H_{\Gamma_2}(\text{div}, \Omega) \times L^2(\Omega) \\ b((\sigma, u), (\tau, v)) &= (\alpha\sigma, \tau) + (u, \text{div}\, \tau) + (\beta u, \tau) - (\text{div}\, \sigma, v) + (cu, v) \\ l((\tau, v) &= (g, \tau) + (f, v) \end{aligned} \tag{5.29}$$

The inf-sup condition is shown in a similar way as in the last case. We take $(\sigma, u) \in U$ and define:

$$\begin{aligned} (\alpha\sigma, \tau) + (u, \text{div}\, \tau) + (\beta u, \tau) =: &\quad \langle g, \tau \rangle \\ -\text{div}\, \sigma + cu =: &\quad f\,. \end{aligned}$$

We need to show now the estimate:

$$\|\sigma\|^2_{H(\text{div},\Omega)} + \|u\|^2 \leq c(\|g\|^2_* + \|f\|^2)$$

where the functional g is measured in the dual norm:

$$\|g\|_* = \sup_{\tau \in H_{\Gamma_2}(\text{div},\Omega)} \frac{|\langle g, \tau \rangle|}{\|\tau\|_{H(\text{div})}}\,.$$

We use now the well-posedness of the ultra-weak variational formulation to establish first L^2 bounds, and then the second equation in the strong form to obtain the bound for the divergence.

Remark 4 If coefficient $c \neq 0$ we can use the second equation to eliminate u and obtain a variational formulation in terms of flux σ only. More precisely, if we want to stay within the current functional setting, wee need c to be bounded below: $c(x) \geq c_0 > 0$. ∎

6 Other Examples

The logic of deriving different variational formulations and the relation with Closed Range Theorems generalizes to many other common problems. We comment on a couple of them.

Linear elasticity

We look for elastic displacements $u_i(x)$, strains $\epsilon_{ij}(x)$ and stresses $\sigma_{ij}(x) = \sigma_{ji}(x)$ that satisfy the following equations.

Cauchy geometric relations:

$$\epsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i})\,, \tag{6.30}$$

constitutive equations (Hooke's law):

$$\sigma_{ij} = E_{ijkl}\epsilon_{kl}\,, \tag{6.31}$$

linear momentum equations:

$$-\sigma_{ij,j} - \rho\omega^2 u_i = f_i\,, \tag{6.32}$$

displacement boundary conditions:

$$u_i = 0 \quad \text{on } \Gamma_1\,, \tag{6.33}$$

traction boundary conditions:

$$t_i = \sigma_{ij}n_j = 0 \quad \text{on } \Gamma_2\,. \tag{6.34}$$

Here $E_{ijkl} = E_{ijkl}(x)$ denote the elasticities tensor, $\rho = \rho(x)$ is the density of the body, ω denotes the angular frequency, t_i is the stress vector (traction), and Γ_1 and Γ_2 denote two disjoint parts of the boundary. All unknowns and tractions are complex-valued functions. If we resort ourselves to the static case, inertial term $\rho\omega^2 u_i$ disappears (momentum equations reduce to equilibrium equations), and all unknowns become real-valued.

The elasticities tensor satisfies the following symmetry assumptions:

$E_{ijkl} = E_{jikl}, \quad E_{ijkl} = E_{ijlk}$ (minor symmetries related to symmetry of stresses and strains)

$E_{ijkl} = E_{klij}$ (major symmetry)

and is positive definite,

$$E_{ijkl}\xi_{ij}\xi_{kl} \geq \alpha_0\xi_{ij}\xi_{ij} \quad \xi_{ij} = \xi_{ji}, \quad \alpha_0 > 0\,. \tag{6.35}$$

The major symmetry and positive definitness of the elasticitities result from thermodynamic considerations. In the case of an isotropic solid, the elasticities are expressed in terms of just two Lamé constants λ, μ,

$$E_{ijkl} = \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta jk) + \lambda\delta_{ij}\delta_{kl}\,.$$

The elasticities tensor can be viewed as a map from symmetric tensors into themselves. The positive-definitness implies that it can be inverted,

$$\epsilon_{ij} = C_{ijkl}\sigma_{kl}\,.$$

The inverse C_{ijkl} is known as the *compliance tensor*. For the isotropic case, we get,

$$\epsilon_{ij} = \frac{1}{2\mu}\sigma_{ij} - \frac{\lambda}{2\mu(2\mu+N\lambda)}\delta_{ij}\sigma_{kk}\,.$$

In the incompressible limit, $\lambda \to \infty$, the strain is related only to the deviatoric part of the stress,

$$\epsilon_{ij} = \frac{1}{2\mu}\left(\sigma_{ij} - \delta_{ij}\frac{1}{N}\sigma_{kk}\right)\,.$$

The geometric and constitutive equations can be combined into one system of equations,

$$\sigma_{ij} = E_{ijkl}\frac{1}{2}(u_{i,j}+u_{j,i}) = E_{ijkl}u_{k,l}$$

or, in the inverse form,

$$C_{ijkl}\epsilon_{kl} = \frac{1}{2}(u_{i,j}+u_{j,i}) = u_{i,j} - \underbrace{\frac{1}{2}(u_{i,j}-u_{j,i})}_{\omega_{ij}}$$

where the antisymmetric part ω_{ij} of the gradient is identified as the *infinitesimal rotation tensor*.

Consequently, the elasticity problem can be reduced to the system:

$$\begin{aligned} C_{ijkl}\sigma_{k,l} - u_{i,j} + \omega_{ij} &= 0\,, \\ -\sigma_{ij,j} - \rho\omega^2 u_i &= f_i\,. \end{aligned} \tag{6.36}$$

If we multiply the first set equations with conjugated test stresses $\tau_{ij} = \tau_{ji}$, and the second one with conjugated test displacements v_i, and integrate over Ω, we obtain, using the vector notation,

$$\begin{aligned} (C\sigma,\tau) - (\nabla u,\tau) &= 0\,, \\ -(\text{div }\sigma, v) - \omega^2(\rho u, v) &= (f,v)\,. \end{aligned} \tag{6.37}$$

We have now the same four choices as in the case of the diffusion-convection-reaction problem. We can choose to relax either set of equations, i.e. integrate it by parts and build the BCs in, or we can leave it in a strong form. If we relax only the momentum equations, and use the strong form of the first set of equations to eliminate the stresses, we obtain the classical *Principle of Virtual Work*:

$$\begin{cases} u \in H^1_{\Gamma_1}(\Omega) \\ (E_{ijkl}u_{k,l}, v_{i,j}) - \omega^2(u_i, v_i) = (f_i, v_i) \quad v \in H^1_{\Gamma_1}(\Omega) \end{cases} \tag{6.38}$$

or, using vector notation,

$$\begin{cases} u \in H^1_{\Gamma_1}(\Omega) \\ (E\nabla u, \nabla v) - \omega^2(u,v) = (f,v) \quad v \in H^1_{\Gamma_1}(\Omega)\,. \end{cases} \tag{6.39}$$

The energy space is again the usual Sobolev space of order one, with the kinematic BCs built in,

$$H^1_{\Gamma_1}(\Omega) := \{u = (u_i)_{i=1}^N \in (H^1(\Omega))^N \ : \ u = 0 \text{ on } \Gamma_1\}\,. \tag{6.40}$$

The energy space for stresses, needed for the other formulations, incorporates additionally the symmetry assumption,

$$H^{sym}_{\Gamma_2}(\text{div}, \Omega) := \{\sigma = (\sigma_{ij}) \in (L^2(\Omega))^{N^2} \ : \ \sigma_{ij} = \sigma_{ji} \text{ and } \sigma_{ij} n_j = 0 \text{ on } \Gamma_2\}\,. \tag{6.41}$$

In the analysis, on top of the tools discussed in the previous sections, one needs the Korn Inequality.

Remark 5 In all four formulations, we can pass to the limit with $\lambda \to \infty$ to arrive at formulations for the Stokes problem (incompressible elasticity). Note that the hydrostatic pressure is not present explicitly in the formulation but it can be recovered a-posteriori by computing the stress axiator,

$$p = -\frac{1}{N}\sigma_{ii}\,.$$

In the incompressible limit, we cannot use the strong form of the first set of equations to eliminate stresses as only the deviatoric part of the stress is related to the gradient of the displacement. If we insist on eliminating the stresses nevertheless, we need to introduce the pressure as an additional unknown. The resulting classical variational formulation involves then the displacement and the pressure. ∎

Maxwell's equations

The time-harmonic Maxwell equations read as follows. We seek electric field $E = (E_i)_{i=1}^N$ and magnetic field $H = (H_i) i = 1^N$ that satisfy the following equations.

Faraday Law :

$$\frac{1}{\mu}\nabla \times E + i\omega H = 0\,, \tag{6.42}$$

Ampère Law:

$$\nabla \times H - (\sigma + i\omega\epsilon)E = J^{imp}\,, \tag{6.43}$$

Perfectly Conducting Boundary (PEC) boundary condition on Γ_1:

$$n \times E = 0\,, \tag{6.44}$$

Vanishing surface current boundary condition on Γ_2:

$$n \times H = 0\,. \tag{6.45}$$

Here ϵ, μ, σ are the material constants: permittivity, permeability and conductivity, and J^{imp} denotes a prescribed *impressed current* (the load).

If the two previous examples involved the operators of gradient and divergence, this one is all about the curl. It leads to a new energy space,

$$H(\text{curl}, \Omega) := \{E \in (L^2(\Omega))^N \ : \ \nabla \times E \in (L^2(\Omega))^N\} \tag{6.46}$$

and its subspaces incorporating BCs on Γ_1 and Γ_2. New facts from Sobolev spaces concerning the new energy space are needed, but the overall logic of four possible variational formulations remains exactly the same.

7 Concluding Remarks

I would like to finish by making a few comments. The first one deals with non-homogeneous boundary-conditions. We cannot build the non-homogeneous BCs directly into the definition of the energy space as we have done it here for the homogeneous case. The reason is very simple: the subset of the energy space that consists of functions satisfying any non-homogeneous conditions, does not have the algebraic structure of a subspace. Consequently, we cannot tab directly to the Closed Range and Babuška-Nečas Theorems that use the formalism of vector spaces. So what to we need to do ?

I know two solutions. The first consists of the usual trick of introducing *a particular solution* that satisfies the non-homogeneous BC. The ultimate solution is then sought as a sum of the particular solution and a *general solution* that satisfies the homogeneous BC. The particular solution *lifts* BC data and results in additional loads for the general solution. This is the quickest way to solve the analysis problem but not necessarily the best practical approach to solve the problem numerically. For instance in Finite Elements, we prefer to interpolate the BC data into the FE space and use the FE shape functions to construct the lift. The advantage of such an approach is that the ultimate solution - the sum of the particular and general FE solutions, depends upon the interpolation procedure but it is independent of the choice of the lift.

The second approach to analyze the problem is to extend the definition of the operator to include also the boundary operators. This was the approach taken e.g. in [8] to study the Stokes problem where a non-homogeneous BC was a must. All essential points discussed in this report go through but the exposition is more technical and it involves more facts from Sobolev spaces. This report has aimed at students who meet with the concept of the variational formulation for the first time and accounting for the non-homogeneous BC here might be too difficult.

The second comment deals with finite elements and the use of mesh-dependent *broken test spaces*. In each of the presented formulations, I can "break" test functions, i.e. test with a larger set of functions that do not involve any conformity assumptions across interelement boundaries. The price we pay is the introduction of additional unknowns that live on the union of interelement boundaries - the *mesh skeleton*. Things get even more technical. On the positive side, handling non-homogeneous BCs is not longer a problem. The

interesting fact is that I can take *any* of the variational formulations discussed here and "break" the test functions. If the original formulation is well posed, so is the one with broken space functions and the inf-sup constant remains of the same order [4].

Finally, I would like to comment on the assumptions that have lead to the coercivity. If they do not hold, we need to resort to more advanced analysis tools. One of the classical tools is the *Fredholm Alternative*. If we can demonstrate that the operator governing our problem is a *compact perturbation* of a self-adjoint, coercive problem, then the Fredholm alternative applies and, similarly to finite-dimensional problems, uniqueness implies that the operator is bounded below. We need to demonstrate uniqueness then which may be easier.

For a limited class of self-adjoint problems, we can compute the inf-sup constants in terms of eigenvalues, see e.g. [2, 3], and use of Helmholtz decomposition comes handy for Maxwell problems[1]. With the logical connections explained in this report, once we can prove the well-posedness of the classical variational formulation, we should be able to show it also for the remaining ones.

References

[1] A. Buffa. Remarks on the discretization of some non-positive operator with application to heterogeneous Maxwell problems. *SIAM J. Numer. Anal.*, 43(1):1–118, 2005.

[2] L. Demkowicz. Asymptotic convergence in finite and boundary element methods. Part 1: Theoretical results. *Comput. Math. Appl.*, 27(12):69–84, 1994.

[3] L. Demkowicz. Asymptotic convergence in finite and boundary element methods. Part 2: The lbb constant for rigid and elastic problems. *Comput. Math. Appl.*, 28(6):93–109, 1994.

[4] L. Demkowicz, C. Carstensen, and J. Gopalakrishnan. The paradigm of broken test functions in DPG discretizations of elliptic second–order PDEs. 2014. in preparation.

[5] J. T. Joichi. On closed operators with closed range. *Proc. Amer. Math. Soc.*, 11(1):80–83, 1960.

[6] T. Kato. Estimation of iterated matrices, with application to the von Neumann condition. *Numer. Math.*, 2:22–29, 1960.

[7] J.T. Oden and L.F. Demkowicz. *Applied Functional Analysis for Science and Engineering*. Chapman & Hall/CRC Press, Boca Raton, 2010. Second edition.

[8] N. Roberts, Tan Bui-Thanh B., and L. Demkowicz. The DPG method for the Stokes problem. *Comput. Math. Appl.*, 67(4):966–995, 2014.

[9] K. Yosida. *Functional Analysis*. Springer-Verlag, New York, 4 edition, 1974.

Exercises

1. Recall the classical *Fourier lemma*: Let f be a continuous function defined on closed interval $[a, b]$ such that
$$\int_a^b f\phi = 0 \quad \forall \phi \in C[a, b] \,:\, \phi(a) = \phi(b) = 0\,.$$
Then f must vanish, $f = 0$.

   Use the Fourier lemma to demonstrate that the variational problem (1.2) is formally equivalent with the Euler-Lagrange problem (1.3).

2. Prove that minimization problem (1.1) and variational problem (1.2) are equivalent if the integrand $F(u', u', x)$ is strictly convex in u' and convex in u.

3. Load the bar in Fig.1 with an additional concentrated force F at $x = l$ acting downwards. Modify the linear form accordingly. Explain why this fully legitimate variational problem does not fall into the area of classical variational calculus.

4. Specialize the discussion on the various variational formulations for the diffusion-convection-reaction problem to the one-dimensional case. Assume $\Omega = (0, l)$ and $\Gamma_1 = \{0\}, \Gamma_2 = \{l\}$. For the 1D case, you do not need to be an expert in Sobolev spaces. Furnish all technical details concerning traces (all in 1D) and the 1D version of the Poincaré lemma.

5. Use elementary means to prove the Poincaré lemma for a square domain $\Omega = (0, 1)^2 \in \mathbb{R}^2$, with $\Gamma_1 = \{x = (0, x_2) \,:\, x_2 \in (0, 1)\}$.

6. In the discussion on the classical variational formulation for the diffusion-convection-problem, we have integrated the advection term by parts. Under the incompressibility assumption on b_i, we can write the equation in the form:
$$-(a_{ij}u_{,j})_{,i} + b_i u_{,i} + cu = f$$
   Replace the second boundary condition with
$$a_{ij}u_{,j}n_i = 0 \text{ on } \Gamma_2$$
   and repeat the whole analysis *without* integrating the advection term by parts.

7. In the case outlined above, it is more natural to use the viscous flux $\sigma_i = a_{ij}u_{,j}$. Repeat the main points of the analysis for this case, and discuss differences with the case discussed in the text.

8. Prove that the graph norm (4.23) and $H(\mathrm{div}, \Omega) \times H^1(\Omega)$-norm are indeed equivalent. Show estimates for the equivalence constants in terms of material data.

9. Study the trivial and ultraweak variational formulations for the simplest example of a boundary value problem

$$u' = f \quad \text{in } (0, l)$$

with each of the possible boundary conditions:

(i) Inflow BC: $u(0) = 0$.

(ii) Outflow BC: $u(l) = 0$.

(iii) Inflow and outflow BC together: $u(0) = 0, u(l) = 0$.

(iv) No BCs at all.

In each case, discuss existence, uniqueness, possible compatibility conditions on f, and the relation between Friedrichs and inf-sup constants.

10. Attempt to show that the mixed formulation is well posed directly, i.e. w/o using the connection with the ultraweak formulation. Consult literature, if necessary.

11. Why did the infinitesimal rotation tensor ω_{ij} disappear between 6.36 and (6.37) ?

12. Attempt to generalize the presented analysis to the elasticity example in the static case.

13. Expand Remark 5. Introduce the pressure and eliminate the deviatoric stresses to arrive at the classical variational formulation.

14. Consider the elastodynamic case. Assume that the classical variational formulation is well posed and attempt to fill in the rest of the analysis.

15. Repeat the exercise above for the Maxwell problem.

16. Boundary conditions studied in this report are by no means exhaustive. Below are a few more examples of additional boundary conditions that you may want to study.

- For diffusion-convection-reaction:

$$a_{ij} u_{,j} n_i + \beta u = 0 \,.$$

- For elastictity:

$$\sigma_{ij} n_j + \beta_i u_i = 0 \,.$$

- For elastictity:

$$\sigma_{ij} n_j n_i = 0 \quad \text{and} \quad u_t = 0$$

where u_t is the tangential component of the displacement vector: $u_t = u - (u \cdot n)n$.

- For Maxwell:

$$n \times H + \gamma E_t = 0$$

where E_t denotes the tangential component of electric field E, $E_t = E - (E \cdot n)n$.

Appendix: A Crash Course on Closed Operators

Closed operators are a generalization of continuous operators. Let X, Y be two normed spaces. A linear operator $T : X \supset D(X) \to Y$ is *closed* if its graph is a closed subspace of $X \times Y$. In our discussion $X = Y = L^2(\Omega), \Omega \subset \mathbb{R}^N$, and this is indeed the most common case in applications. If we use a curvilinear system of coordinates for Ω, we may need a weighted L^2-space. $X \times Y$ is a normed space, and in a normed space, a set is closed if and only if it is sequentially closed. Hence, a working condition to check whether the operator is closed is:

$$D(T) \ni x_n \to x, Tx_n \to y \quad \stackrel{?}{\Longrightarrow} \quad x \in D(T) \text{ and } Tx = y\,.$$

Every continuous operator is automatically closed. The simplest examples of closed operators that are not continuous are differential operators including those discussed in this report. If X, Y are Banach (complete) and $D(T) = X$ then, by the Closed Graph Theorem, operator T is continuous. Consequently, a non-trivial closed operator is always defined on a proper subspace of X. Note that, in contrary, in studying continuous operators, we can always assume that they are defined on the *whole space*. Indeed, if T is continuous then it admits a unique continuous extension to closure $\overline{D(T)}$ which, as a closed subspace of Banach space, is itself Banach as well. Thus, we can always replace X with $\overline{D(T)}$ and restrict ourselves to studying continuous operators defined on the whole space only.

If the domain of operator T is *dense* in X, $\overline{D(T)}^X = X$, we can define its transpose T'. This is not a trivial concept. As in the purely algebraic case, the transpose is defined on elements of dual Y',

$$T'y' := y' \circ T$$

There two problems to begin with. First of all, $y' \circ T$ is defined only on $D(T)$ and not on the whole X. Secondly, even if we consider the topological dual of $D(T)$ only, the composition $y' \circ T$ may not be continuous since operator T is only closed. We circumvent the problem by restricting the domain of T' only to those functionals y' for which the composition *is continuous* i.e. we land in $D(T)'$. The domain of the transpose $D(T')$ is thus, in general only a proper subspace of Y'. Finally, if $D(T)$ is dense in X then the continuous functional $y' \circ T$ admits a unique extension to the whole X with the same norm. This extension is identified as the ultimate value of the transpose $T' : Y' \supset D(T') \to X'$. By definition, domain $D(T')$ is *maximal*, i.e. it contains *all functionals* y' for which the composition $y' \circ T$ is continuous.

For Hilbert spaces, including the L^2-space, the duals are identified with the spaces themselves. For the L^2-space discussed here, $X = X' = Y = Y' = L^2(\Omega)$.

The Closed Range Theorem tells us that closed operator T and its transpose T' are intimately related.

THEOREM 2

(Closed Range Theorem for Closed Operators)

Let X, Y be Banach spaces, and $T : X \supset D(T) \to Y$ a closed operator with the domain $D(T)$ dense in X. The following conditions are equivalent to each other.

(i) Range $\mathcal{R}(T)$ is closed in Y.

(ii) Quotient operator corresponding to T, is bounded below, i.e. there exists a constant $\gamma > 0$ such that,

$$\|Tx\|_Y \geq \gamma \inf_{x_0 \in \mathcal{N}(T)} \|x + x_0\|_X \qquad \forall\, x \in D(T)\,.$$

(iii) Range $\mathcal{R}(T)$ forms the orthogonal component to the null space of the transpose,

$$\mathcal{R}(T) = \mathcal{N}(T')^{\perp} := \{y \in Y \,:\, \langle y', y\rangle = 0 \quad \forall\, y' \in \mathcal{N}(T')\}\,.$$

(iv) Range $\mathcal{R}(T')$ is closed in X'.

(v) Quotient operator corresponding to T' is bounded below, i.e. there exists a constant $\gamma' > 0$ such that,

$$\|T'y'\|_{X'} \geq \gamma \inf_{y_0' \in \mathcal{N}(T')} \|y' + y_0'\|_{Y'} \qquad \forall\, y' \in D(T')\,.$$

(vi) Range $\mathcal{R}(T')$ forms the orthogonal component to the null space of operator T,

$$\mathcal{R}(T') = \mathcal{N}(T)^{\perp} := \{x' \in X' \,:\, \langle x', x\rangle = 0 \quad \forall\, x \in \mathcal{N}(T)\}\,.$$

Moreover, constants γ and $\gamma\prime$ are equal. It goes w/o saying that we are talking about the best constants (sup) we can have. ∎

Yosida attributes the theorem to Banach and reduces its proof to the case of continuous operators [9]. I could not reproduce his arguments which seem to me to be incorrect. A beautiful proof was provided by Kato [6] and, independently (using different arguments), by Joichi [5].

Note the full symmetry in the relation between the operator and its adjoint. It is remarkable that the theorem holds for non-reflexive spaces as well.