Lecture notes on Nonlinear PDEs: I. Elliptic PDEs and Variational Methods

A. N. Yannacopoulos,
Department of Statistics,
Athens University of Economics and Business, ayannaco@aueb.gr

March 29, 2012

Abstract: We provide an introduction to the theory of nonlinear elliptic PDEs using their connection with variational problems.

1 Introduction

Certain elliptic PDEs, linear or nonlinear have a nice interpretation as Euler-Lagrange equations for the minimization of certain functionals. This implies that one may use the powerful tools of the calculus of variations and present elegant proofs concerning the existence of solutions as well as their qualitative properties for such problems. It is the aim of this section to present a brief introduction to this theory [4].

2 Motivation

We will first motivate this approach through some examples.

Example 1 (Laplace equation)**.** Consider the Laplace equation

$$\begin{aligned} \Delta u = f, \;\text{on}\; \mathcal{O}, \\ u = 0, \;\text{on}\; \partial\mathcal{O} \end{aligned} \tag{1}$$

where $\mathcal{O}$ is a domain in $\mathbb{R}^d$ with sufficiently smooth boundary $\partial\mathcal{O}$.

Let us also define the functional $J : \mathbb{X} \to \mathbb{R}$, where $\mathbb{X}$ is a function space to be specified later on, as

$$J(u) := \int_{\mathcal{O}} (|\nabla u|^2 + f\,u)dx \tag{2}$$

for every $u \in \mathbb{X}$ (where $\mathbb{X}$ is selected so that the above integrals make sense.

Assume now that formally we vary u and try to calculate the value of the functional J on a new "position" in $\mathbb{X}$, which corresponds to the function $u + \epsilon v$, where $v \in \mathbb{X}$ and $\epsilon > 0$ is a small real number. A quick calculation yields that

$$\frac{J(u+\epsilon v) - J(u)}{\epsilon} = \left(\int_{\mathcal{O}} \nabla u \cdot \nabla v dx + \int_{\mathcal{O}} f\,v dx \right) + \epsilon \int_{\mathcal{O}} |\nabla v|^2 dx$$

so that the dominant term in this expression as $\epsilon \to 0$ is the first term on the right hand side. That means that the limit of this expression as $\epsilon \to 0$ is

$$DJ(u;v) := \lim_{\epsilon\to 0} \frac{J(u+\epsilon v) - J(u)}{\epsilon} = \left(\int_{\mathcal{O}} \nabla u \cdot \nabla v dx + \int_{\mathcal{O}} f\,v dx \right),$$

where the notation $DJ(u;v)$ implies that we look at the "infinitesimal" variation of the functional J at the position $u \in X$ along the direction $v \in \mathbb{X}$ (if we consider $\mathbb{X}$ as a vector space to start with, this does not lead to any major conceptual problems).

Selecting v so as to vanish on the boundary (i.e. selecting $v \in \mathbb{X} := \mathbb{H}^1_0$) and integrating by parts we obtain,

$$DJ(u;v) := \lim_{\epsilon \to 0} \frac{J(u+\epsilon v) - J(u)}{\epsilon} = \int_{\mathcal{O}} (-\Delta u + f)\, v dx. \tag{3}$$

Intuitively extending our experience from finding critical points for real valued functions, we may assume (we will return to this point later on with the necessary mathematical rigor) that a critical point (we do not need to classify it as maximum, minimum or saddle point yet) for the function J will be given at u, if the value of this functional does not change for small deviations around the point $u \in \mathbb{X}$, that is if $\frac{J(u+\epsilon v)-J(u)}{\epsilon}$ is equal to 0 in the limit as $\epsilon \to 0$. This implies that at a critical point of J, it holds that $DJ(u;v) = 0$. This must be true of every choice of direction $v \in \mathbb{X}$ (or rather for every "acceptable" choice) therefore (3) leads us to characterize the critical point u^* of the functional J as the solution of the inhomogeneous Laplace equation (1). Therefore, if we can say things about the critical points of J then this knowledge transfers to knowledge concerning the solutions of equation (1).

The discussion in Example 1 can be extended to other equations, and importantly to nonlinear elliptic equations. The following example motivates this.

Example 2 (Nonlinear elliptic equations)**.** We now start from the opposite end and consider a functional $J : \mathbb{X} \to \mathbb{R}$ defined as

$$J(u) := \int_{\mathcal{O}} G(x, u, \nabla u) dx \tag{4}$$

for every $u \in \mathbb{X}$, where $G : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is a function assumed to be smooth enough in all its variables. We will use the notation $G(x, u, p)$ to denote the values of the function G, where $x \in \mathcal{O} \subset \mathbb{R}^d$, u is the value of the function u at the selected point $x \in \mathcal{O}$ and $p = \nabla u$, the gradient of u at the selected point $x \in \mathcal{O}$.

We now repeat our formal calculations and try to find a critical point of J. Standard calculations yield

$$\frac{J(u+\epsilon v) - J(u)}{\epsilon} = \int_{\mathcal{O}} \left(\frac{\partial}{\partial u} G\, v + \frac{\partial}{\partial p} G \cdot \nabla v \right) dx + O(\epsilon)$$

where we have used the Taylor expansion of the function G and by $O(\epsilon)$ we denote terms of order ϵ or higher. Note that since p is a vector $\frac{\partial}{\partial p} G$ is a more compact notation for $\nabla_p G(x, u, p)$.

We now integrate by parts and assuming as above that v vanishes on $\partial\mathcal{O}$ we obtain

$$\frac{J(u+\epsilon v) - J(u)}{\epsilon} = \int_{\mathcal{O}} \left[\frac{\partial}{\partial u} G - \nabla \cdot \left(\frac{\partial}{\partial p} G \right) \right] v dx + O(\epsilon)$$

Taking the limit as $\epsilon \to 0$ we formally obtain

$$DJ(u;v) := \int_{\mathcal{O}} \left[\frac{\partial}{\partial u} G - \nabla \cdot \left(\frac{\partial}{\partial p} G \right) \right] v dx. \tag{5}$$

Then, extending our discussion in Example 1, concerning the positions of critical points of the functional J in $\mathbb{X}$ we are able to deduce that a critical point will be positioned at $u^* \in X$ as long as this is chosen so that

$$DJ(u^*; v) = 0, \quad \forall\ v \in \mathbb{X}. \tag{6}$$

A quick comparison of (5) and (6) shows that (under appropriate choice of $\mathbb{X}$ or course) the critical point(s) of J are situated at u^* which solve the Euler-Lagrange equation

$$\frac{\partial}{\partial u}G - \nabla \cdot \left(\frac{\partial}{\partial p}G \right) = 0. \tag{7}$$

Now depending on the choice of the function G we may obtain a large variety of nonlinear PDEs as the Euler-Lagrange equation.

- If $G(x,u,p) = \frac{1}{2}|p|^2 + f\,u$ then we recover the inhomogeneous Laplace equation $\Delta u = f$ with homogeneous Dirichlet boundary conditions.
- If $G(x,u,p) = \frac{1}{2}|p|^2 + F(u)$ where $F : \mathbb{R} \to \mathbb{R}$ is a differentiable function then be recover the semilinear Laplace equation $\Delta u = F^{'}(u)$ with homogeneous Dirichlet boundary conditions.
- If $G(x,u,p) = \frac{1}{r+2}|p|^{r+2} + f\,u$ then we recover the nonlinear Laplacian equation $\nabla \cdot (|\nabla u|^r \nabla u) = f$, which reduces to the standard inhomogeneous Laplace equation for $r = 0$.

Many other choices are possible.

3 Calculus in Banach space

3.1 Gâteaux and Fréchet derivatives

We now try to put on a more rigorous mathematical basis the concepts of derivatives of functionals that were used in Examples 1 and 2 to derive the Euler-Lagrange equation.

Let $\mathbb{X}$ be a Banach space and $F : \mathbb{X} \to \mathbb{R}$ be a functional.

Definition 1 (Directional derivative)**.** The directional derivative of F at $\mathsf{x} \in \mathbb{X}$ along the direction $h \in \mathbb{X}$ is the limit

$$DF(\mathsf{x}; h) = \lim_{\epsilon \to 0} \frac{F(\mathsf{x} + \epsilon\, h) - F(\mathsf{x})}{\epsilon}$$

if it exists.

It is not necessary that the operator defined by $h \mapsto DF(\mathsf{x}; h)$ is a linear operator. If it is then we may talk about the concept of the Gâteaux derivative.

Definition 2 (Gâteaux derivative)**.** A functional F is called weakly (Gâteaux) differentiable at $\mathsf{x} \in \mathbb{X}$ if it is weakly differentiable for any direction $h \in \mathbb{X}$ and the operator $h \mapsto DF(\mathsf{x}; h)$ is linear and continuous (therefore bounded), i.e., if there exists a linear operator $\mathsf{A} : \mathbb{X} \to \mathbb{R}$ such that

$$\lim_{\epsilon \to 0} \frac{|F(\mathsf{x} + \epsilon\, h) - F(\mathsf{x}) - \epsilon \mathsf{A} h|}{\epsilon} = 0, \quad \forall\, h \in \mathbb{X}. \tag{8}$$

In such a case A is the Gâteaux derivative of F at x, denoted by $DF(\mathsf{x})$, and defined by

$$DF(\mathsf{x})\, h := \lim_{\epsilon \to 0} \frac{F(\mathsf{x} + \epsilon\, h) - F(\mathsf{x})}{\epsilon}$$

The Gâteaux derivative if it exists is unique, a result that follows naturaly by the uniqueness of the limit.

Remark 1. If $\mathbb{X}$ is a finite dimensional space, $\mathbb{X} = \mathbb{R}^d$, then the Gâteaux derivative coincides with the gradient, and the Gâteaux derivative along a particular direction $h \in \mathbb{X}$ coincides with the directional derivative $\nabla F \cdot h$.

Remark 2. The Gâteaux derivative may be generalized for functionals $F : \mathbb{X} \to \mathbb{Y}$ where $\mathbb{Y}$ is another Banach space, the definition staying the same, with the only changes being that now (i) the absolute value in (8) has to be replaced by the norm in $\mathbb{Y}$ and that (ii) DF is an operator $DF : \mathbb{X} \to \mathbb{Y}$.

Remark 3. The Gâteaux derivative defines a linear functional $DF(\mathsf{x}) \in \mathbb{X}'$ such that

$$DF(\mathsf{x}, h) = \langle DF(\mathsf{x}), h\rangle_{\mathbb{X}',\mathbb{X}}, \quad \forall h \in \mathbb{X}.$$

This linear functional is called the gradient of F.

The Gâteaux differentiability (weak derivative) is not the only concept of differentiability available in Banach space.

Definition 3 (Fréchet differentiability)**.** The operator $F : \mathbb{X} \to \mathbb{R}$ is called Fréchet (strongly) differentiable at $x \in \mathbb{X}$ if there exists an operator $\mathsf{A} \in \mathcal{L}(\mathbb{X}, \mathbb{R})$ such that

$$\lim_{||h||\to 0} \frac{|F(\mathsf{x}+h) - F(\mathsf{x}) - \mathsf{A}\, h|}{||h||} = 0. \tag{9}$$

The operator A (still denoted by $DF(\mathsf{x})$) is called the Fréchet derivative of F at x.

Remark 4. The Fréchet derivative may be generalized for functionals $F : \mathbb{X} \to \mathbb{Y}$ where $\mathbb{Y}$ is another Banach space, the definition staying the same, with the only changes being that now (i) the absolute value in (10) has to be replaced by the norm in $\mathbb{Y}$ and that (ii) DF is a linear bounded operator $DF : \mathbb{X} \to \mathbb{Y}$.

If a functional is Fréchet differentiable then it is also Gâteaux differentiable but the converse does not necessarily hold. Furthermore, if a functional is Fréchet differentiable then its Fréchet derivative is unique.

Proposition 1. *Consider a subset $D \subset \mathbb{X}$ and let $F : D \to \mathbb{R}$ be strongly differentiable at a point $\mathsf{x} \in int(D)$. Then F is continuous at x.*

Proof. Since $\mathsf{x} \in int(D)$ there exists $\epsilon_1 > 0$ such that $\mathsf{x} + h \in D$ as long as $||h|| \leq \epsilon_1$. Since, by assumption

$$\lim_{||h||\to 0} \frac{|F(\mathsf{x}+h) - F(\mathsf{x}) - DF(\mathsf{x})\, h|}{||h||} = 0.$$

where exists for every $\epsilon > 0$ an $\epsilon_2 > 0$ such that

$$|F(\mathsf{x}+h) - F(\mathsf{x}) - DF(\mathsf{x})\, h| \leq \epsilon\, ||h||, \quad \text{if } ||h|| \leq \epsilon_2.$$

By the triangle inequality

$$\begin{aligned} &|F(\mathsf{x}+h) - F(\mathsf{x})| = |F(\mathsf{x}+h) - F(\mathsf{x}) - DF(\mathsf{x})h + DF(\mathsf{x})h| \leq \\ &|F(\mathsf{x}+h) - F(\mathsf{x}) - DF(\mathsf{x})h| + |DF(\mathsf{x})h| \leq |F(\mathsf{x}+h) - F(\mathsf{x}) - DF(\mathsf{x})h| + ||DF(\mathsf{x})||\, ||h|| \end{aligned}$$

where $||DF(\mathsf{x})||$ is the norm of the operator $DF \in \mathcal{L}(\mathbb{X}, \mathbb{R})$. Choosing $\delta = \min(\epsilon_1, \epsilon_2)$ the above estimate yields

$$|F(\mathsf{x}+h) - F(\mathsf{x})| \leq (\epsilon + ||DF(\mathsf{x})||)\, ||h||,$$

which holds for every $\epsilon > 0$ therefore leading to the conclusion that there exists a constant $C > 0$ such that

$$|F(\mathsf{x}+h) - F(\mathsf{x})| \leq C\, ||h||,$$

from which continuity at x follows. ☐

Remark 5. Clearly, if $F : \mathbb{X} \to \mathbb{Y}$ then Proposition 1 also holds (where of course all absolute values are to be replaced by the norm in $\mathbb{Y}$).

Remark 6. In contrast to the assurance of Proposition 1 the Gâteaux differentiability of a functional at a point does not guarantee continuity at this point for the functional, but rather a weaker property called hemicontinuity. In particular, if $F : D \to \mathbb{Y}$ (where of course the choice $\mathbb{Y} = \mathbb{R}$ is possible) is Gâteaux (weakly) differentiable at $\mathsf{x} \in int(D)$ then we can only guarantee that

$$\lim_{\epsilon \to 0} F(\mathsf{x} + \epsilon h) = F(\mathsf{x}), \quad \forall\, h \in \mathbb{X},$$

i.e. that F is hemicontinuous at x. Put differently, hemicontinuity of F is equivalent to the continuity of the real valued function $\phi(\epsilon) = F(\mathsf{x} + \epsilon\, h)$ for every $h \in \mathbb{X}$. Since Gâteaux differentiability of F at x is equivalent to the differentiability of the real value function ϕ with respect to its argument, continuity of phi follows and hence the hemicontinuity of F. Clearly, a hemicontinuous functional at $\mathsf{x} \in \mathbb{X}$ needs not be continuous at x. An example is e.g., the function $F : \mathbb{R}^2 \to \mathbb{R}$, defined by

$$F(x_1, x_2) = \begin{cases} \left(\frac{x_2}{x_1}\right)(x_1^2 + x_2^2) & x_1 \neq x_2 \\ 0 & x_1 = 0. \end{cases}$$

which is hemicontinuous and Gâteaux differntiable at $\mathsf{x} = (x_1, x_2) = (0, 0)$, but not continuous (see [1]).

Example 3. Let $\mathbb{X}$ be a Hilbert space, $a : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ a bilinear form and $L : \mathbb{X} \to \mathbb{R}$ a linear form and define $F(\mathsf{x}) = a(\mathsf{x}, \mathsf{x}) + L(\mathsf{x})$.

Then, F is Gâteaux differentiable at every point $\mathsf{x} \in \mathbb{X}$ and every direction $h \in \mathbb{X}$ and

$$DF(\mathsf{x}, h) = a(\mathsf{x}, h) + L(h), \quad \forall\, h \in \mathbb{X}.$$

If a is (bi)-continuous and L is continuous then F is also Fréchet differentiable and $DF(\mathsf{x})$ is defined by

$$DF(\mathsf{x})h := a(\mathsf{x}, h) + L(h).$$

Example 4. Let $\mathcal{O} \subset \mathbb{R}^d$ (open set) and $\mathbb{X} = L^p(\mathcal{O})$, $p \geq 1$.

Let g be a C^1 function $g : \mathbb{R} \to \mathbb{R}$. Under standard assumptions on g, for every $\mathsf{x} \in \mathbb{X}$

$$\mathsf{x} \mapsto \int_{\mathcal{O}} g(\mathsf{x}(x))dx$$

defines a functional $F : \mathbb{X} \to \mathbb{R}$.

This functional is Gâteaux differentiable at all $\mathsf{x} \in \mathbb{X}$ and for all directions $h \in \mathbb{X}$ and

$$DF(\mathsf{x}, h) = \int_{\mathcal{O}} g'(\mathsf{x}(x))h(x)dx.$$

3.2 Higher derivatives

Higher order derivatives may be defined in a standard fashion.

Definition 4. Let $F : \mathbb{X} \to \mathbb{Y}$. This is twice Gâteaux differentiable at point $\mathsf{x} \in \mathbb{X}$ in the directions $h, j \in \mathbb{X}$ if the operator $DF(\mathsf{x}, h)$ is once Gâteaux differentiable at point x in the direction j. The second derivative is denoted by $D^2(\mathsf{x}, h, j)$ (this is an element of $\mathbb{Y}$),

$$\lim_{\epsilon \to 0} \frac{DF(\mathsf{x} + \epsilon\, j, h) - DF(\mathsf{x}, h)}{\epsilon}.$$

Example 5. If $F(x) = a(\mathsf{x},\mathsf{x}) + L(\mathsf{x})$ then

$$D^2F(\mathsf{x}, h, j) = a(h, j) + a(j, h).$$

Remark 7. The second Gâteaux derivative defines an operator $D^2F(x) \in \mathcal{L}(\mathbb{X}, \mathbb{X}')$ (equivalently a bilinear form) such that

$$D^2F(\mathsf{x}, h, j) = \langle D^2F(\mathsf{x})\, h, j\rangle_{\mathbb{X}',\mathbb{X}}, \ \ \forall\, h, j \in \mathbb{X}.$$

This operator is called the Hessian.

Proposition 2.

(i) If F is once Gâteaux differentiable then there exists an $s \in (0, 1)$ such that

$$F(\mathsf{x} + h) = F(\mathsf{x}) + DF(\mathsf{x} + s\, h; h)$$

or equivalently

$$F(\mathsf{x} + h) = F(\mathsf{x}) + \langle DF(\mathsf{x} + s\, h), h\rangle.$$

(ii) If F is twice Gâteaux differentiable then there exists an $s \in (0, 1)$ such that

$$F(\mathsf{x} + h) = F(\mathsf{x}) + DF(\mathsf{x}, h) + \frac{1}{2}D^2(\mathsf{x} + s\, h, h, h)$$

, or equivalently

$$F(\mathsf{x} + h) = F(\mathsf{x}) + \langle DF(\mathsf{x}), h\rangle + \frac{1}{2}\langle D^2(\mathsf{x} + s\, h)h, h\rangle.$$

Proof. (i) If F is once Gâteaux differentiable then the real valued function

$$t \mapsto \phi(t) := F(\mathsf{x} + t\, h), \ \ \forall\, h \in \mathbb{X},$$

is differentiable. Then application of the mean value theorem on the function ϕ yields the stated result.
(ii) Similarly, by application of the Taylor formula for the fuction ϕ. □

3.3 Convexity and differentiability

Convexity plays a very important role in optimization.

Definition 5. A subset $U \subset \mathbb{X}$ is convex if $\forall\, \mathsf{x}, \mathsf{y} \in U$, if holds that $\lambda\, \mathsf{x} + (1 - \lambda)\mathsf{y} \in U$ for all $\lambda \in (0, 1)$.

Definition 6. A functional $F : U \subset \mathbb{X} \to \mathbb{R}$ is convex if $\forall\, \mathsf{x}, \mathsf{y} \in U$, if holds that

$$F(\lambda\, \mathsf{x} + (1 - \lambda)\mathsf{y}) \leq \lambda\, F(\mathsf{x}) + (1 - \lambda)F(\mathsf{y}),$$

for all $\lambda \in (0, 1)$.

Convexity is related with differentiability.

Theorem 1. *Let $F : \mathbb{X} \to \mathbb{R}$ be Gâteaux differentiable in a convex open subset $U \subset \mathbb{X}$. Then F is convex if and only if either of the following holds:*

(i)

$$F(\mathsf{y}) - F(\mathsf{x}) \geq DF(\mathsf{x}; \mathsf{y} - \mathsf{x}), \ \ \forall \, \mathsf{x}, \mathsf{y} \in U$$

or equivalently, interpreting the Gâteaux derivative as a map $DF(x) : \mathbb{X} \to \mathbb{R}$ (i.e., as an element of $\mathbb{X}'$)

$$F(\mathsf{y}) - F(\mathsf{x}) \geq \langle DF(\mathsf{x}), \mathsf{y} - \mathsf{x} \rangle, \ \ \forall \, \mathsf{x}, \mathsf{y} \in U. \tag{10}$$

(ii) The Gâteaux derivative $DF(\mathsf{x}) \in \mathbb{X}'$ is a monotone operator, i.e.,

$$\langle DF(\mathsf{y}) - DF(\mathsf{x}), \mathsf{y} - \mathsf{x} \rangle \geq 0, \ \ \forall \, \mathsf{x}, \mathsf{y} \in U. \tag{11}$$

Proof. (i) Consider two points $\mathsf{x}, \mathsf{y} \in U$ and take the convex combination $(1-\epsilon)\mathsf{x} + \lambda \mathsf{y} = \mathsf{x} + \epsilon(\mathsf{y} - \mathsf{x})$. Convexity of F implies that

$$F(\mathsf{x} + \epsilon(\mathsf{y} - \mathsf{x})) \leq (1-\epsilon) F(\mathsf{x}) + \epsilon F(\mathsf{y}) = F(\mathsf{x}) + \epsilon (F(\mathsf{y}) - F(\mathsf{x})), \ \ \forall \, x, y \in U, \ \ \epsilon \in (0,1)$$

which leads upon rearrangement to

$$\frac{F(\mathsf{x} + \epsilon(\mathsf{y} - \mathsf{x})) - F(\mathsf{x})}{\epsilon} \leq F(\mathsf{y}) - F(\mathsf{x}).$$

Since F is Gâteaux differentiable at $\mathsf{x} \in \mathbb{X}$ we may pass to the limit $\epsilon \to 0$ and interpreting $DF(\mathsf{x})$ as an element of $\mathbb{X}'$ this leads to

$$\langle DF(x), \mathsf{y} - \mathsf{x} \rangle \leq F(\mathsf{y}) - F(\mathsf{x}).$$

To prove the converse, suppose (10) holds for every pair $(\mathsf{x}, \mathsf{y}) \in U \times U$. It then holds for the pair $(\mathsf{x}, \mathsf{x} + \epsilon(\mathsf{y} - \mathsf{x}))$ as well as for the pair $(\mathsf{y}, \mathsf{x} + \epsilon(\mathsf{y} - \mathsf{x}))$, for every $\epsilon \in (0,1)$. This leads to the inequalities

$$\begin{aligned} F(\mathsf{x}) &\geq F(\mathsf{x} + \epsilon(\mathsf{y} - \mathsf{x})) - \epsilon \, \langle DF(\mathsf{x} + \epsilon(\mathsf{y} - \mathsf{x})), \mathsf{y} - \mathsf{x} \rangle, \\ F(\mathsf{y}) &\geq F(\mathsf{x} + \epsilon(\mathsf{y} - \mathsf{x})) + (1-\epsilon) \, \langle DF(\mathsf{x} + \epsilon(\mathsf{y} - \mathsf{x})), \mathsf{y} - \mathsf{x} \rangle. \end{aligned}$$

We multiply the first by $(1-\epsilon)$ and the second by ϵ and add to obtain convexity.
(ii) Assume convexity of F. Write (10) twice, interchanging x and y. Adding, yields (11).

Conversely, let (11) hold. An application of the mean value formula (see Proposition 2) implies that forall $\mathsf{x}, \mathsf{y} \in U$ there exists $s \in (0,1)$ such that

$$F(\mathsf{y}) - F(\mathsf{x}) = \langle DF(\mathsf{x} + s(\mathsf{y} - \mathsf{x})) - DF(\mathsf{x}), \mathsf{y} - \mathsf{x} \rangle + \langle DF(\mathsf{x}), \mathsf{y} - \mathsf{x} \rangle.$$

We now apply (11) for the pair $(\mathsf{x} + s(\mathsf{y} - \mathsf{x}), \mathsf{x}) \in U \times U$ we obtain

$$\langle DF(\mathsf{x} + s(\mathsf{y} - \mathsf{x})) - DF(\mathsf{x}), \mathsf{y} - \mathsf{x} \rangle \geq 0, \ \ \forall \, s \in (0,1).$$

Combining these two inequalities we obtain (10), therefore F is convex. □

Remark 8. If F is strictly convex the inequality (10) or (11) is strict.

Theorem 2. *Let $U \subset \mathbb{X}$ be a convex and open set. If $F : U \subset \mathbb{X} \to \mathbb{R}$ is twice Gâteaux differentiable in all directions in U then $D^2 F$ defines a positive definite form, i.e.,*

$$D^2 F(\mathsf{x}, h, h) \geq 0, \ \ \forall \, \mathsf{x} \in \mathbb{X}, \ h \in U, \ \ h \neq 0,$$

or equivalently,

$$\langle D^2 F(\mathsf{x}) h, h \rangle \geq 0, \ \ \ \forall \, \mathsf{x} \in U, \ \ h \in \mathbb{X}. \tag{12}$$

Proof. Follows immediately by the Taylor expansion formula of Proposition 2. □

Remark 9. If F is strictly convex the inequality (12) is strict.

3.4 Convexity and continuity

Convexity guarantees some rather useful continuity properties. We start by recalling two important notions of continuity.

Definition 7 (Lower semicontinuity)**.** A functional is called lower semicontinuous if for any sequence $\{\mathsf{x}_n\} \subset \mathbb{X}$ such that $\mathsf{x}_n \to \mathsf{x}$ in $\mathbb{X}$ it holds that

$$\liminf_n F(\mathsf{x}_n) \geq F(\mathsf{x}).$$

Definition 8 (Weak lower semicontinuity)**.** A functional is called weakly lower semicontinuous if for any sequence $\{\mathsf{x}_n\} \subset \mathbb{X}$ such that $\mathsf{x}_n \rightharpoonup x$ in $\mathbb{X}$ it holds that

$$\liminf_n F(\mathsf{x}_n) \geq F(\mathsf{x}).$$

For convex functionals and convex sets these two notions are related.

We recall an important result from functional analysis, Mazur's lemma that will help us in this direction.

Lemma 1 (Mazur)**.** *Let $\mathbb{X}$ be a Banach space and consider a sequence $\{\mathsf{x}_n\} \subset \mathbb{X}$ such that $\mathsf{x}_n \rightharpoonup \mathsf{x}$ in $\mathbb{X}$. Then, for every $n \in \mathbb{N}$ there exists a $N(n)$ and a sequence of sets of real numbers $\{a(n)_k\}_{k=n}^{N(n)}$ with the properties $a(n)_k \in (0,1)$ for every k and $\sum_{k=n}^{N(n)} a(n)_k = 1$, such that the sequence*

$$\bar{\mathsf{x}}_n := \sum_{k=n}^{N(n)} a(n)_k \, \mathsf{x}_k,$$

has the property $\bar{\mathsf{x}}_n \to \mathsf{x}$ (where now the convergence is strong).

Mazur's lemma allows us to turn the weak convergence into strong for the proper convex combination of terms of the original sequence.

Let $\mathsf{x}_n \rightharpoonup \mathsf{x}$ in $\mathbb{X}$. Then, using Mazur's lemma we may construct the sequence $\bar{\mathsf{x}}_n \to \mathsf{x}$ in $\mathbb{X}$. Since for each n, $\bar{\mathsf{x}}_n$ is a convex combination of elements of the original sequence, if F is convex we have that

$$F(\bar{\mathsf{x}}_n) = F\left(\sum_{k=n}^{N(n)} a(n)_k \, \mathsf{x}_k\right) \leq \sum_{k=n}^{N(n)} a(n)_k F(\mathsf{x}_k)$$

Weak lower semicontinuity is a stronger property than strong lower semicontinuity, in the sense that any strongly lower semicontinuous functional is not weakly lower semicontinuous, while the contrary always holds. However, for convex functionals, the two concepts are equivalent.

Proposition 3. *Let $F : U \subset \mathbb{X} \to \mathbb{R}$ be a convex functional. Then F is weakly lower semicontinuous if and only if its is strongly lower semicontinuous.*

Proof. It is straightforward to check that if F is weakly lower semicontinuous then it is also strongly lower semicontinuous (this does not require convexity). Assume now that F is strongly lower semicontinuous and convex. Consider any sequence x_n with the property $\mathsf{x}_n \rightharpoonup \mathsf{x}$ and construct the sequence $\bar{\mathsf{x}}_n$ the existence of which is guaranteed by Mazur's lemma such that $\bar{\mathsf{x}}_n \to \mathsf{x}$. However, by the convexity of F,

$$F(\bar{\mathsf{x}}_n) = F\left(\sum_{k=n}^{N(n)} a(n)_k \, \mathsf{x}_k\right) \leq \sum_{k=n}^{N(n)} a(n)_k F(\mathsf{x}_k), \tag{13}$$

for any $k = n, \cdots, N(n)$. By the definition of $\liminf F(\mathsf{x}_n)$ there exists a subsequence $F(\mathsf{x}_{n_r})$ such that $\lim_r F(\mathsf{x}_{n_r}) = \liminf F(\mathsf{x}_n)$. Therefore for every $\epsilon > 0$, there exists an N such that $F(\mathsf{x}_{n_r}) < \liminf F(\mathsf{x}_n) + \epsilon$ for $n > N$. Then, (13), applied for the chosen subsequence, implies that for large enough r,

$$F(\bar{\mathsf{x}}_{n_r}) < \liminf F(\mathsf{x}_n) + \epsilon$$

therefore,

$$\liminf F(\bar{\mathsf{x}}_n) < \liminf F(\mathsf{x}_n) + \epsilon, \ \ \forall \epsilon > 0.$$

Strong lower semicontinuity implies that

$$F(\mathsf{x}) \leq \liminf_n F(\bar{\mathsf{x}}_n)$$

which yields that

$$F(\mathsf{x}) < \liminf F(\mathsf{x}_n) + \epsilon, \ \ \forall \epsilon > 0,$$

therefore,

$$F(\mathsf{x}) \leq \liminf F(\mathsf{x}_n)$$

and F is weakly lower semicontinuous. □

Remark 10. Note that this result can also arise from the fact that if the epigraph of a convex functional is a strongly closed subset of $\mathbb{X}$ is also weakly closed (a fact that also comes from Mazur's lemma).

Proposition 4. *If a functional $F : \mathbb{X} \to \mathbb{R}$ is convex and is Gâteaux differentiable at $\mathsf{x} \in \mathbb{X}$ then F is weakly lower semicontinuous at x.*

Proof. Consider a sequence $\{\mathsf{x}_n\} \subset \mathbb{X}$, such that $\mathsf{x}_n \rightharpoonup \mathsf{x}$ in X. Since F is convex and Gâteaux differentiable at x, apply (11) for the choice $\mathsf{y} = \mathsf{x}$ and $\mathsf{x} = \mathsf{x}_n$ to obtain

$$F(\mathsf{x}_n) - F(\mathsf{x}) \geq \langle DF(\mathsf{x}), \mathsf{x}_n - \mathsf{x} \rangle, \ \ n \in \mathbb{N}. \tag{14}$$

Since $\mathsf{x}_n \rightharpoonup \mathsf{x}$ in $\mathbb{X}$, for every $v \in \mathbb{X}'$, it holds that $\langle v, \mathsf{x}_n - \mathsf{x} \rangle \to 0$ as $n \to \infty$. Choose $v = DF(\mathsf{x}) \in \mathbb{X}'$ and go to the limit as $n \to \infty$ in (14) to obtain that

$$\liminf_n F(\mathsf{x}_n) \geq F(\mathsf{x})$$

which guarantees the weak lower semicontinuity of F. □

Example 6. Consider the quadratic functional $F : \mathbb{X} \to \mathbb{R}$, $F(\mathsf{x}) = a(\mathsf{x}, \mathsf{x}) + L(\mathsf{x})$.

Convexity of the functional is related to coercivity of the bilinear form, i.e., the existence of $C > 0$ such that

$$|a(\mathsf{x}, \mathsf{x})| \geq C ||\mathsf{x}||^2_{\mathbb{X}}, \ \ \forall \mathsf{x} \in \mathbb{X}$$

It the above hold, then F is weakly lower semicontinuous.

Remark 11. Coercivity is not a very easy condition to be satisfied at least in the whole of $\mathbb{X}$. It may be satisfied in subsets of $\mathbb{X}$, which are compactly embedded in $\mathbb{X}$. An example may be functionals defined in Sobolev spaces.

4 Optimization in Banach space

4.1 Optimization in vector space

Theorem 3 (Weierstrass)**.** *Let $U \subset \mathbb{X}$ be a bounded and weakly closed subset of a reflexive Banach space $\mathbb{X}$. Let $F : U \subset \mathbb{X} \to \mathbb{R}$ be a weakly lower semicontinuous functional. Then F admits a minimum in U.*

Proof. Let $\{\mathsf{x}_n\} \subset U$ be a minimizing sequence, i.e., a sequence such that $F(\mathsf{x}_n) \to m$ where $m = \inf_{\mathsf{x}\in U} F(\mathsf{x})$. This sequence has a weak limit (it is bounded and $\mathbb{X}$ is reflexive) so there exists $\mathsf{x} \in \mathbb{X}$ such that $\mathsf{x}_n \rightharpoonup \mathsf{x}$ in $\mathbb{X}$. We will show that this element x is such that $F(\mathsf{x}) = m$ i.e., x is the minimizer.

Indeed, by the weak lower semicontinuity $\liminf_n F(\mathsf{x}_n) \geq F(\mathsf{x})$ and since $\{\mathsf{x}_n\}$ is a minimizing sequence it follows that $F(\mathsf{x}) = m$. □

Remark 12. We may substitute the boundedness of U by the condition

$$\lim_{||\mathsf{x}||\to\infty} F(\mathsf{x}) = \infty.$$

Theorem 4 (First order conditions)**.** *Let $F : U \subset \mathbb{X} \to \mathbb{R}$ have a local minimum at $\mathsf{x} \in \mathbb{X}$ and suppose that F is Gâteaux differentiable in U (where U is an open set). Then, the first order condition $DF(\mathsf{x}) = 0$ holds.*

Proof. Since F has a local minimum at x, for every direction $h \in X$ it holds that

$$F(\mathsf{x}) \leq F(\mathsf{x} + \epsilon\, h), \quad \forall\, h \in \mathbb{X},$$

for small enough ϵ. A simple manipulation leads to

$$\frac{F(\mathsf{x} + \epsilon\, h) - F(\mathsf{x})}{\epsilon} \geq 0, \quad \forall\, h \in \mathbb{X},$$

and since F is Gâteaux differentiable at x we have that

$$\langle DF(\mathsf{x}), h\rangle \geq 0, \quad \forall\, h \in \mathbb{X}.$$

Since $\mathbb{X}$ is a vector space we may repeat the above procedure for $-h \in \mathbb{X}$ so that we finally obtain that at the local minimum x it holds that

$$\langle DF(\mathsf{x}), h\rangle = 0, \quad \forall\, h \in \mathbb{X}.$$

□

Remark 13. The first order condition is to be understood as an equality in the dual space $\mathbb{X}'$, i.e., that

$$\langle DF(\mathsf{x}), h\rangle_{\mathbb{X}',\mathbb{X}} = 0, \ \ \forall\, h \in \mathbb{X}.$$

If considered as an operator equation this is an equation in weak form.

Remark 14. An alternative formulation of the first order condition is in terms of a variational inequality

$$\langle DF(\mathsf{x}), \mathsf{y} - \mathsf{x}\rangle_{\mathbb{X}',\mathbb{X}} \geq 0, \ \ \forall\, \mathsf{y} \in \mathbb{X}.$$

(y must be admissible in the sense that $x + \epsilon_n(\mathsf{y} - \mathsf{x}) \in U$ for every member of a real sequence ϵ_n such that $\epsilon_n \to 0$)

4.2 Optimization and convexity

Convexity leads to some interesting properties as far as minimization is concerned.

Proposition 5.

(i) A local minimum for a convex functional defined in a convex set is a global minimum.

(ii) If a functional is strictly convex that a minimum is unique.

Proof. (i) Assume that $\mathsf{x} \in \mathbb{X}$ is a local minimum, i.e. $F(\mathsf{x}) \leq F(\mathsf{x}')$ for every $\mathsf{x}' \in V$ where V is a small enough neighbourghood around x. For any $\mathsf{y} \in U$ take the convex combination $(1-\epsilon)\mathsf{x} + \epsilon\,\mathsf{y} = \mathsf{x} + \epsilon(\mathsf{y}-\mathsf{x}) \in U$, $\epsilon \in [0,1]$. For small enough values of ϵ, $\mathsf{x}+\epsilon(\mathsf{y}-\mathsf{x}) \in V$ and since x is a local minimum

$$F(\mathsf{x}) \leq F(\mathsf{x} + \epsilon(\mathsf{y}-\mathsf{x}))$$

and by convexity of F it follows that

$$F(\mathsf{x} + \epsilon(\mathsf{y}-\mathsf{x})) \leq (1-\epsilon)F(\mathsf{x}) + \epsilon\, F(\mathsf{y}) = F(\mathsf{x}) + \epsilon(F(\mathsf{y}) - F(\mathsf{x})),$$

for all $\epsilon \geq 0$, small enough. Combining the above we obtain for all positive and small enough ϵ that

$$F(\mathsf{x}) \leq F(\mathsf{x}) + \epsilon\,(F(\mathsf{y}) - F(\mathsf{x})),$$

or equivalently that $F(\mathsf{x}) \leq F(\mathsf{y})$ for every $\mathsf{y} \in U \subset \mathbb{X}$, therefore, x is a global minimum.
(ii) Let $\mathsf{x}_1, \mathsf{x}_2 \in \mathbb{X}$ be two global minima of F such that $\mathsf{x}_1 \neq \mathsf{x}_2$. Consider the point $\mathsf{x} = \frac{1}{2}\mathsf{x}_1 + \frac{1}{2}\mathsf{x}_2 \in \mathbb{X}$. By strict convexity of F it follows that $F(\mathsf{x}) < F(\mathsf{x}_1) = F(\mathsf{x}_2)$ which leads to contradiction. □

Theorem 5. *Let $U \subset \mathbb{X}$, convex and $F : U \subset \mathbb{X} \to \mathbb{R}$ Gâteaux differentiable in all directions and convex. Then $\mathsf{x} \in U$ is a minimum if and only if $DF(\mathsf{x};\mathsf{y}-\mathsf{x}) \geq 0$, $\forall\, \mathsf{y} \in U$, or equivalently, $\langle DF(\mathsf{x}), \mathsf{y}-\mathsf{x}\rangle \geq 0$, $\forall\, \mathsf{y} \in U$.*

Proof. Assume $\mathsf{x} \in U$ is a minimum. Then, $F(\mathsf{x}) \leq F(\mathsf{z})$, for every $\mathsf{z} \in U$. For any $\mathsf{x}, \mathsf{y} \in U$, set $\mathsf{z} = (1-\epsilon)\mathsf{x} + \epsilon\,\mathsf{y} = \mathsf{x} + \epsilon\,(\mathsf{y}-\mathsf{x}) \in U$ for $\epsilon \in (0,1)$, and apply this inequality for obtain

$$F(\mathsf{x}) \leq F(\mathsf{x} + \epsilon\,(\mathsf{y}-\mathsf{x})), \;\; \forall\, \mathsf{y} \in U, \;\; \epsilon > 0,$$

which yields

$$\frac{F(\mathsf{x} + \epsilon\,(\mathsf{y}-\mathsf{x})) - F(\mathsf{x})}{\epsilon} \geq 0, \quad \epsilon > 0,$$

and going to the limit as $\epsilon \to 0$,

$$\langle DF(\mathsf{x}), \mathsf{y}-\mathsf{x}\rangle \geq 0, \;\; \forall\, \mathsf{y} \in U.$$

For the converse, since F is convex and Gâteaux differentiable we have that

$$F(\mathsf{y}) - F(\mathsf{x}) \geq \langle DF(\mathsf{x}), \mathsf{y}-\mathsf{x}\rangle, \;\; \forall\, \mathsf{x}, \mathsf{y} \in U.$$

Since for $\mathsf{x} \in U$ it holds that $\langle DF(\mathsf{x}), \mathsf{y}-\mathsf{x}\rangle \geq 0$, $\forall\, \mathsf{y} \in U$, we find that $F(\mathsf{x}) \leq F(\mathsf{y})$ for all $\mathsf{y} \in U$ so that x is a local minimum. □

4.3 Projections

We start by a fundamental result, the projection theorem on a convex closed subset.

Theorem 6 (Projection theorem)**.** *Let $\mathbb{X}$ be a Hilbert space and $K \subset \mathbb{X}$ be closed and convex. Then, for any $\mathsf{x} \in \mathbb{X}$ the minimization problem*

$$\min_{\mathsf{z}\in K} ||\mathsf{x}-\mathsf{z}||$$

has a unique solution, x^*, defining a contraction operator $\Pi_K : \mathbb{X} \to K$, by $\Pi_K \mathsf{x} := \mathsf{x}^*$. Furthermore, x^* is characterized by the solution of the inequality

$$\langle \mathsf{x}-\mathsf{x}^*, \mathsf{y}-\mathsf{x}^*\rangle \leq 0, \ \ \forall \ \ \mathsf{y} \in K. \tag{15}$$

Proof. Define $F(\mathsf{z}) := ||\mathsf{x}-\mathsf{z}||^2$. The existence follows by taking a minimizing sequence $\{z_n\} \subset K$, i.e. a sequence such that $F(\mathsf{z}_n) \to m$ where $m = \inf_{z\in K} ||\mathsf{x}-\mathsf{z}||^2$. This sequence is bounded so it is weakly convergent, i.e. there exists $x^* \in \mathbb{X}$ such that $\mathsf{z}_n \rightharpoonup \mathsf{x}^*$. By the closedness[1] of K we know that $x^* \in K$. This limit is the required minimizer. This follows easily since the norm is a weakly lower semicontinuous function (the norm is a strictly convex function). Uniqueness follows from strict convexity.

To show that the minimizer satisfies the inequality (15) observe that $F(\mathsf{z}) = \langle \mathsf{x}-\mathsf{z}, \mathsf{x}-\mathsf{z}\rangle$. If x^* is the element of K that minimizes the distance then $F(\mathsf{x}^*) \leq F(\mathsf{z})$ for all $\mathsf{z} \in K$. That implies that x^* is the solution of the inequality

$$\langle \mathsf{x}-\mathsf{x}^*, \mathsf{x}-\mathsf{x}^*\rangle \leq \langle \mathsf{x}-\mathsf{z}, \mathsf{x}-\mathsf{z}\rangle, \ \ \forall, \ \ \mathsf{z} \in K.$$

Let y be any element of K and take $\mathsf{z} = (1-\epsilon)\mathsf{x}^* + \epsilon \mathsf{y} = \mathsf{x}^* + \epsilon(\mathsf{y}-\mathsf{x}^*)$, $\epsilon \in (0,1)$. Since $\mathsf{z} \in K$ we have that

$$\langle \mathsf{x}-\mathsf{x}^*, \mathsf{x}-\mathsf{x}^*\rangle \leq \langle (\mathsf{x}-\mathsf{x}^*) - \epsilon(\mathsf{y}-\mathsf{x}^*), (\mathsf{x}-\mathsf{x}^*) - \epsilon(\mathsf{y}-\mathsf{x}^*)\rangle, \ \ \forall, \ \ \mathsf{y} \in K, \ \ \epsilon \in (0,1).$$

Using the properties of the inner product this leads to

$$0 \leq -\langle \mathsf{x}-\mathsf{x}^*, \mathsf{y}-\mathsf{x}^*\rangle + \epsilon \, ||\mathsf{y}-\mathsf{x}^*||^2, \ \ \forall, \ \ \mathsf{y} \in K, \ \ \epsilon \in (0,1).$$

and going to the limit as $\epsilon \to 0$ leads to the inequality,

$$\langle \mathsf{x}-\mathsf{x}^*, \mathsf{y}-\mathsf{x}^*\rangle \leq 0, \ \ \forall \ \ \mathsf{y} \in K.$$

□

5 Variational inequalities

5.1 Bilinear forms

Let $\mathbb{X}$ be a Hilbert space.

Definition 9. A mapping $a : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ which is linear in both variables, i.e., such that

$$a(\lambda_1 \mathsf{x}_1 + \lambda_2 \mathsf{x}_2, \mathsf{y}) = \lambda_1 a(\mathsf{x}_1, \mathsf{y}) + \lambda_2 a(\mathsf{x}_2, \mathsf{y}), \ \ \forall \, \lambda_1, \lambda_2 \in \mathbb{R}, \ \ \mathsf{x}_1, \mathsf{x}_2, \mathsf{y} \in \mathbb{X},$$
$$a(\mathsf{x}, \lambda_1 \mathsf{y}_1 + \lambda_2 \mathsf{y}_2) = \lambda_1 a(\mathsf{x}, \mathsf{y}_1) + \lambda_2 a(\mathsf{x}, \mathsf{y}_2), \ \ \forall \, \lambda_1, \lambda_2 \in \mathbb{R}, \ \ \mathsf{x}, \mathsf{y}_1, \mathsf{y}_2 \in \mathbb{X}$$

is called a bilinear form on $\mathbb{X}$.

[1] Since K is convex, if it is strongly closed it is also weakly closed

Definition 10. A bilinear form is called symmetric if $a(\mathsf{x},\mathsf{y}) = a(\mathsf{y},\mathsf{x})$ for all $\mathsf{x},\mathsf{y} \in \mathbb{X}$.

Definition 11. A bilinear form is called continuous if there exists a constant $C > 0$ such that

$$|a(\mathsf{x},\mathsf{y})| \leq C\,||\mathsf{x}||\,||\mathsf{y}||,\ \ \forall\,\mathsf{x},\mathsf{y} \in \mathbb{X}.$$

Definition 12. A bilinear form is called coercive if there exists a constant $\alpha > 0$ such that

$$\alpha\,||\mathsf{x}||^2 \leq\ |a(\mathsf{x},\mathsf{x})|,\ \ \forall\,\mathsf{x} \in \mathbb{X}.$$

Example 7. Let $\mathcal{O} \subset \mathbb{R}^d$ and $\mathbb{X} = W_0^{1,2}(\mathcal{O})$, the Sobolev space of functions that have generalized derivatives of first order that are defined in the $L^2(\mathcal{O})$ sense and whose trace vanishes at $\partial\mathcal{O}$. An element $\mathsf{x} \in \mathbb{X}$ is considered as a function $u : \mathcal{O} \to \mathbb{R}$ such that ∇u is defined as an element of $L^2(\mathcal{O})$ with the property $u(x) = 0$ (in the sense of traces) for $x \in \partial\mathcal{O}$. The space $\mathbb{X} = W_0^{1,2}(\mathcal{O})$ is a Hilbert space when endowed with the norm

$$||u||^2_{W^{1,2}(\mathcal{O})} = ||u||^2_{L^2(\mathcal{O})} + ||\nabla u||^2_{L^2(\mathcal{O})},$$

which is generated by the inner product

$$\langle u, v\rangle = \int_{\mathcal{O}} (uv + \nabla u \cdot \nabla v)dx.$$

Recall the famous Poincaré inequality, that yields

$$||\nabla u||_{L^2(\mathcal{O})} \geq C\,||u||_{L^2(\mathcal{O})},\ \ \forall\,u \in W_0^{1,2}(\mathcal{O}).$$

This means that $||\nabla u||_{L^2(\mathcal{O})}$ is an equivalent norm for $W_0^{1,2}(\mathcal{O})$. In what follows, we endow $\mathbb{X}$ with this norm, that will be denoted by $||\cdot||$.

Consider the bilinear form $a : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$, defined by

$$a(u, v) := \int_{\mathcal{O}} \nabla u \cdot \nabla v\,dx.$$

Clearly, this bilinear form is symmetric.

The Cauchy-Schwarz inequality gives

$$|a(u, v)| \leq \int_{\mathcal{O}} |\nabla u \cdot \nabla v|\,dx \leq \left\{\int_{\mathcal{O}} |\nabla u|^2 dx\right\}^{1/2} \left\{\int_{\mathcal{O}} |\nabla v|^2 dx\right\}^{1/2} = ||u||\,||v||$$

which guarantees that a is continuous.

Furthermore

$$a(u, u) = \int_{\mathcal{O}} |\nabla u|^2 dx = ||u||^2$$

which guarantees that $a : \mathbb{X} \to \mathbb{X} \to \mathbb{R}$ is coercive.

5.2 The abstract theory

We now present an introduction to the abstract theory of variational inequalities (see e.g. [2] or [5]).

Let $\mathbb{X}$ be a Hilbert space and $F : \mathbb{X} \to \mathbb{R}$, $F(\mathsf{x}) = \frac{1}{2}a(\mathsf{x},\mathsf{x}) + L(\mathsf{x})$ where a is a continuous, symmetric and coercive bilinear form and L is a continuous linear form.

Theorem 7. *Let $K \subset \mathbb{X}$ convex and closed. The problem $\min_{\mathsf{x}\in K} F(\mathsf{x})$ has a unique solution which is equivalent to the solution of the variational inequality*

$$\textit{find } \mathsf{x} \in K,\ \textit{such that } a(\mathsf{x}, \mathsf{y} - \mathsf{x}) \geq L(\mathsf{y} - \mathsf{x}),\ \forall\,\mathsf{y} \in K.$$

Let $F : \mathbb{X} \to \mathbb{R}$, $F(\mathsf{x}) = \frac{1}{2}a(\mathsf{x},\mathsf{x}) + L(\mathsf{x})$ where a is a continuous and coercive bilinear form, and let $K \subset \mathbb{X}$ closed and convex.

Theorem 8 (Lax-Milgram-Stampacchia)**.**

(i) For any $L \in \mathbb{X}'$ the variational inequality

$$\textit{find } \mathsf{x} \in K, \ \textit{such that } a(\mathsf{x}, \mathsf{y}-\mathsf{x}) \geq L(\mathsf{y}-\mathsf{x}), \ \forall\, \mathsf{y} \in K,$$

or equivalently

$$\textit{find } \mathsf{x} \in K, \ \textit{such that } \langle \mathsf{Ax}, \mathsf{y}-\mathsf{x}\rangle \geq \langle f, \mathsf{y}-\mathsf{x}\rangle, \ \forall\, \mathsf{y} \in K, \ f \in \mathbb{X}', \tag{16}$$

has a unique solution.

(ii) If furthermore, a is symmetric that x is the unique minimizer on K of the functional

$$F(\mathsf{x}) = \frac{1}{2}a(\mathsf{x},\mathsf{x}) - \langle f, \mathsf{x}\rangle.$$

Proof. (i) By the continuity of the bilinear form a, for fixed $\mathsf{y} \in \mathbb{X}$ by the Riesz representation there exists a unique $\mathsf{z} \in \mathbb{X}$ such that $a(\mathsf{x},\mathsf{y}) = \langle \mathsf{z}, \mathsf{y}\rangle$. We will denote $\mathsf{z} = \mathsf{Ax}$ and this will define an operator $\mathsf{A} : \mathbb{X} \to \mathbb{X}'$ (we may of course consider $\mathbb{X}' \simeq \mathbb{X}$ by the Riesz isometry). The operator A is bounded, since

$$|\langle \mathsf{Ax}, \mathsf{y}\rangle| = |a(\mathsf{x},\mathsf{y})| \leq C\,||\mathsf{x}||\,||\mathsf{y}||, \ \ \forall\, \mathsf{x} \in \mathbb{X}, \ \mathsf{y} \in \mathbb{X}'$$

which leads to $||\mathsf{A}||_{\mathcal{L}(\mathbb{X},\mathbb{X}')} \leq C$.

Let $\Pi_K : \mathbb{X} \to K$ be the projection mapping from $\mathbb{X}$ to the closed convex $K \subset \mathbb{X}$, defined by the solution of the problem

$$\Pi_K \mathsf{x} = \arg\min_{\mathsf{z} \in K} ||\mathsf{x} - \mathsf{z}||.$$

As is well known, this map is a contraction i.e. $||K\mathsf{x}|| \leq ||\mathsf{x}||$, for all $\mathsf{x} \in \mathbb{X}$.

Define now the family of maps $R_t : \mathbb{X} \to \mathbb{X}$, $t \in \mathbb{R}_+$, such that

$$R_t\mathsf{x} := \mathsf{x} - t\,(\mathsf{Ax} - f).$$

For the right choice of t this map is a contraction; let $\mathsf{x}_1, \mathsf{x}_2 \in \mathbb{X}$,

$$R_t\mathsf{x}_1 - R_t\mathsf{x}_2 = \mathsf{x}_1 - \mathsf{x}_2 - t\mathsf{A}(\mathsf{x}_1 - \mathsf{x}_2) = (I - t\mathsf{A})(\mathsf{x}_1 - \mathsf{x}_2)$$

where we have used the linearity of A. We estimate $||R_t\mathsf{x}_1 - R_t\mathsf{x}_2||$ by

$$\begin{aligned} ||R_t\mathsf{x}_1 - R_t\mathsf{x}_2||^2 &= ||(I - t\mathsf{A})(\mathsf{x}_1 - \mathsf{x}_2)||^2 = \langle (I - t\mathsf{A})(\mathsf{x}_1 - \mathsf{x}_2), (I - t\mathsf{A})(\mathsf{x}_1 - \mathsf{x}_2)\rangle \\ &= ||\mathsf{x}_1 - \mathsf{x}_2||^2 - 2t\langle \mathsf{A}(\mathsf{x}_1 - \mathsf{x}_2), \mathsf{x}_1 - \mathsf{x}_2\rangle + t^2||\mathsf{A}(\mathsf{x}_1 - \mathsf{x}_2)||^2. \end{aligned}$$

The coercivity of the bilinear form implies that $\langle \mathsf{Ax}, \mathsf{x}\rangle \geq \alpha||\mathsf{x}||^2$ for some $\alpha > 0$. This combined with the boundedness of A lead to the estimate

$$||R_t\mathsf{x}_1 - R_t\mathsf{x}_2||^2 \leq \ (1 - 2t\,\alpha + Ct^2)\,||\mathsf{x}_1 - \mathsf{x}_2||^2.$$

For t small enough $1 - 2t\,\alpha + Ct^2 < 1$. Choosing a t^* with this property we observe that $R_{t^*} : \mathbb{X} \to \mathbb{X}$ is a contraction. Furthermore, since $\Pi_K : \mathbb{X} \to K$ is a contraction, the composition $\Pi_K R_{t^*}$ is a contraction as well. Finally if we define $\Gamma_K : K \to K$ as the restriction of $\Pi_K R_*$ on K, i.e. $\Gamma_K := \Pi_K R_*|_K$ this will be a contraction map as well. By the Banach contraction map theorem, it has a unique fixed point; i.e., there exists a unique $\mathsf{x}^* \in K$ such that $\Gamma_K \mathsf{x}^* = \mathsf{x}^*$ or equivalently

$$\mathsf{x}^* = \Pi_K \left(\mathsf{x}^* - t^*(\mathsf{A}\mathsf{x}^* - f)\right).$$

By the definition of the projection operator, x^* is the element in K that minimizes the distance from the element $\mathsf{x}^* - t^*(\mathsf{A}\mathsf{x}^* - f) \in \mathbb{X}$. The conditions for the projection imply that this is the required solution. Indeed, let $\mathsf{x} = \mathsf{x}^* - t^*(\mathsf{A}\mathsf{x}^* - f) \in \mathbb{X}$, and apply the condition for the projection to obtain

$$\langle \mathsf{x}^* - t^*(\mathsf{A}\mathsf{x}^* - f) - \mathsf{x}^*, \mathsf{y} - \mathsf{x}^* \rangle \leq 0, \quad \forall \mathsf{y} \in K,$$

which lead to the desired inequality.

(ii) Let x be such that

$$a(\mathsf{x}, \mathsf{y} - \mathsf{x}) \geq \langle f, \mathsf{y} - \mathsf{x} \rangle, \quad \forall \mathsf{y} \in K.$$

The symmetry of a implies that

$$a(\mathsf{y} - \mathsf{x}, \mathsf{x}) \geq \langle f, \mathsf{y} - \mathsf{x} \rangle, \quad \forall \mathsf{y} \in K. \tag{17}$$

For all $\mathsf{y} \in K$,

$$\begin{aligned} F(\mathsf{y}) = \quad & F(\mathsf{x} + (\mathsf{y} - \mathsf{x})) = \frac{1}{2} a(\mathsf{x} + (\mathsf{y} - \mathsf{x}), \mathsf{x} + (\mathsf{y} - \mathsf{x})) - \langle f, \mathsf{x} + (\mathsf{y} - \mathsf{x}) ra \\ = \quad & \frac{1}{2}(\mathsf{x}, \mathsf{x}) - \langle f, \mathsf{x} \rangle + a(\mathsf{y} - \mathsf{x}, \mathsf{x}) + \frac{1}{2} a(\mathsf{y} - \mathsf{x}, \mathsf{y} - \mathsf{x}) - \langle f, \mathsf{y} - \mathsf{x} \rangle \\ = \quad & J(x) + \frac{1}{2} a(\mathsf{y} - \mathsf{x}, \mathsf{y} - \mathsf{x}) + a(\mathsf{y} - \mathsf{x}, \mathsf{x}) - \langle f, \mathsf{y} - \mathsf{x} \rangle. \end{aligned} \tag{18}$$

By coercivity of a it follows that $a(\mathsf{y} - \mathsf{x}, \mathsf{y} - \mathsf{x}) \geq 0$, and (17) implies that $a(\mathsf{y} - \mathsf{x}, \mathsf{x}) - \langle f, \mathsf{y} - \mathsf{x} \rangle \geq 0$. Therefore, (18) implies that

$$F(x) \leq F(y), \quad \forall y \in K$$

so that x is the minimizer of F. ∎

Theorem 9 (Minty). *The variational inequality is equivalent to*

$$\textit{find } \mathsf{x} \in K \textit{ such that } a(\mathsf{y}, \mathsf{y} - \mathsf{x}) \geq \langle f, \mathsf{y} - \mathsf{x} \rangle, \quad \forall \mathsf{y} \in K. \tag{19}$$

Proof. Suppose $\mathsf{x}^* \in K$ solves (16). Then, for any $\mathsf{y} \in K$

$$a(\mathsf{y}, \mathsf{y} - \mathsf{x}^*) = a(\mathsf{y} - \mathsf{x}^*, \mathsf{y} - \mathsf{x}^*) + a(\mathsf{x}^*, \mathsf{y} - \mathsf{x}^*) \geq a(\mathsf{x}^*, \mathsf{y} - \mathsf{x}^*) \geq \langle f, \mathsf{y} - \mathsf{x}^* \rangle,$$

where we used (i) the linearity of a with respect to the first argument and (ii) the coercivity to deduce that $a(\mathsf{y} - \mathsf{x}^*, \mathsf{y} - \mathsf{x}^*) > 0$. Therefore, a solution of (16) is a solution of (19).

Suppose now that $\mathsf{x}^* \in K$ solver (19). For any $\mathsf{y} \in K$ and $\epsilon \in (0, 1)$, define $\mathsf{z} = \mathsf{x}^* + \epsilon(\mathsf{y} - \mathsf{x}^*) \in K$ and apply (19) for the pair $(\mathsf{x}^*, \mathsf{z}) \in K \times K$. This yields,

$$a(\mathsf{x}^* + \epsilon(\mathsf{y} - \mathsf{x}^*), \epsilon(\mathsf{y} - \mathsf{x}^*)) \geq \langle f, \epsilon(\mathsf{y} - \mathsf{x}^*) \rangle, \quad \forall \mathsf{y} \in K$$

which in the limit as $\epsilon \to 0$ leads to (16). ∎

5.3 Application: Free boundary value problems

Let $\mathcal{O} \subset \mathbb{R}^d$ with sufficiently smooth boundary $\partial\mathcal{O}$. Consider now the problem of minimization of the functional

$$J(v) = \frac{1}{2}\int_{\mathcal{O}} |\nabla v(x)|^2 dx$$

over all the functions $v : \mathcal{O} \to \mathbb{R}$ such that $v \geq \phi$ in $\mathcal{O}$, where $\phi; \mathcal{O} \to \mathbb{R}$ is a smooth given function. We will show that this problem is a particular form of the general class of variational inequalities encountered in Section 5, and then through the general solvability results of this section show the existence for a particular PDE problem.

This PDE problem is in the form of a differential inequality,

$$-\Delta u = 0, \text{ if } u > \phi,$$
$$-\Delta u > 0 \text{ if } u = \phi.$$

therefore the unknown function u satisfies the Laplace equation for the points in $\mathcal{O}$ such that $u(x) > \phi(x)$, and the inequality $-\Delta u > 0$ on the points where u coincides with the obstacle, i.e. on the points in $\mathcal{O}$ such that $u(x) = \phi(x)$. We will call this latter set, the coincidence set

$$\mathcal{C} := \{x \in \mathcal{O} \,:\, u(x) = \phi(x)\}.$$

Since u is unknown, the coincidence set $\mathcal{C}$ as well as its boundary is an unknown of the problem, therefore, this type of problems is called a free boundary value problem. This particular class of problems we treat here, is also called an obstacle problem. Obstacle problems find interesting applications in mechanics, probability and mathematical finance.

To write this problem in terms of the abstract formulation needed, we will choose $\mathbb{X} = W_0^{1,2}(\mathcal{O})$. This is a Sobolev space, and any $\mathsf{x} \in \mathbb{X}$ is considered as $\mathsf{x} = u$, a function $u : \mathcal{O} \to \mathbb{R}$, whose trace vanishes on $\mathcal{O}$ and that possess generalized derivatives in the $L^2(\mathcal{O})$ sense. By the Poincaré inequality the quantity $||\nabla u||_{L^2(\mathcal{O})}$ is an equivalent norm for this space, which in fact is a Hilbert space, when endowed with it. This remark, guarantees that the functional $J : \mathbb{X} \to \mathbb{R}$ is coercive, for this choice of functional setting, and continuous.

Lemma 2. *The set $K := \{v \in W_0^{1,2}(\mathcal{O}) \,:\, v(x) \geq \phi(x) \text{ a.e. } x \in \mathcal{O}\}$ is a convex and closed subset of $W_0^{1,2}(\mathcal{O})$.*

Proof. Convexity is immediate, since if $v_1, v_2 \in K$ then $\lambda v_1 + (1-\lambda)v_2 \in K$, for all $\lambda \in (0,1)$.

We will only check the K is closed. To this end, consider a sequence $\{v_n\} \subset K$, such that $v_n \to v$ in $W_0^{1,2}(\mathcal{O})$. We will show that $v \in W_0^{1,2}(\mathcal{O})$.

Since $v_n \to v$ in $W_0^{1,2}(\mathcal{O})$, by the compact embedding $W_0^{1,2}(\mathcal{O}) \hookrightarrow L^2(\mathcal{O})$ it follows that there exists a subsequence $v_{n_r} \to v$ in $L^2(\mathcal{O})$, therefore, there exists a subsequence (denoted the same) such that $v_{n_r} \to v$ a.e. in $\mathcal{O}$. Since $v_{n_r} \in k$ it follows that $v_{n_r}(x) \geq \phi(x)$, a.e. in $\mathcal{O}$ and going to the limit of this subsequence the inequality remains valid, so that $v(x) \geq \phi(x)$, a.e. in $\mathcal{O}$. Therefore $v \in K$ and K is closed. □

Remark 15. In fact, the above argument implies that K is weakly closed; if $v_n \rightharpoonup v$ in $W_0^{1,2}(\mathcal{O})$, the compact embedding $W_0^{1,2}(\mathcal{O}) \hookrightarrow L^2(\mathcal{O})$ implies that for a subsequence $v_n \to v$ in $L^2(\mathcal{O})$ and the result follows.

Proposition 6. *Let $\phi \in W_0^{1,2}(\mathcal{O})$ such that $\phi \leq 0$ on $\partial\mathcal{O}$.*

(i) Then, there exists a unique solution of the minimization problem

$$\min_{v \in K} J(v).$$

This minimizer $u \in k$ is characterized as the solution of the variational inequality

$$\int_{\mathcal{O}} \nabla u \cdot \nabla (v-u)(x)dx \geq 0 \;\; \forall\, v \in K \tag{20}$$

(ii) If $u-\phi$ is continuous on $\mathcal{O}$, the minimizer is the solution of the free boundary value problem

$$-\Delta u \geq 0, \;\; \textit{on} \;\; \mathcal{O},$$
$$-\Delta u = 0 \;\; \textit{on} \;\; \mathcal{O} \setminus \mathcal{C}$$

where $\mathcal{C}$ is the coincidence set

$$\mathcal{C} := \{x \in \mathcal{O} \,:\, u(x) = \phi(x)\}.$$

Proof. (i) By the convexity amd closedness of K, as well as by the continuity and coercivity of J the result follows by a straightforward application of the general theorem.

(ii) Let $\psi \in C_0^\infty(\mathcal{O})$ be a test function, such that $\psi \geq 0$. Then, if $u \in K$ it is clear that $v = u + \psi \in K$. Inserting that in the variational inequality (20) we obtain

$$\int_{\mathcal{O}} \nabla u \cdot \nabla \psi dx \geq 0$$

which is the weak form for the inequality $-\Delta u \geq 0$ (as a simple integration by parts argument, plus a density argument of $C_0^\infty(\mathcal{O})$ in $W_0^{1,2}(\mathcal{O})$ shows).

Let us now assume a test function $\psi \in C_0^\infty(\mathcal{O} \setminus \mathcal{C})$. By continuity, if $\epsilon > 0$ is small enough we have that both $u + \epsilon\psi - \phi > 0$ and $u - \epsilon\psi - \phi > 0$. That means $u + \epsilon\psi \in K$, so inserting that into (20) we get

$$\int_{\mathcal{O}} \nabla u \cdot \nabla \psi dx \geq 0,$$

and that $u - \epsilon\psi \in K$, so inserting that into (20) we get

$$\int_{\mathcal{O}} \nabla u \cdot \nabla \psi dx \leq 0$$

so that

$$\int_{\mathcal{O}} \nabla u \cdot \nabla \psi dx \leq 0, \quad \forall\, \psi \in C_0^\infty(\mathcal{O} \setminus \mathcal{C})$$

so that we have

$$-\Delta u = 0 \;\; x \in \mathcal{O} \setminus \mathcal{C}$$

in the weak sense. □

6 The calculus of variations

6.1 A simpler problem

Assumption 1. Let $G : \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d$ be such that there exist costants $\alpha > 0$ and $\beta \geq 0$ such that

$$G(x, u, P) \geq \alpha\, |P|^p - \beta, \;\; \forall\, P \in \mathbb{R}^d, \;\; u \in \mathbb{R}, \;\; x \in \mathcal{O}$$

for some integer $p \in (1, \infty)$.

Assumption 2. The mapping $P \mapsto G(x,u,P)$ is convex and C^1 for all $x \in \mathcal{O}$, $u \in \mathbb{R}$.

Our aim is to find a minimizer for the functional $J : W^{1,p}(\mathcal{O}) \to \mathbb{R}$, defined by

$$J(u) := \int_{\mathcal{O}} G(x, u(x), \nabla u(x))dx.$$

This is a functional defined on the Banach space $\mathbb{X} := W^{1,p}(\mathcal{O})$ consisting of functions $u : \mathcal{O} \to \mathbb{R}$ that are once differentiable in the sense of distributions and the weak derivatives can be understood as elements of $L^p(\mathcal{O})$. Note that now we do not impose any boundary conditions (in the sense of traces) on $\partial\mathcal{O}$.

Assumption 1 on the function G implies some coercivity on the functional J. In particular for J it holds that

$$J(u) \geq \alpha \, ||\nabla u||^p_{L^p(\mathcal{O})} - \beta \, \mu(\mathcal{O}).$$

This is a coercivity assumption that guarantees that as long as $J(u)$ is bounded then $||\nabla u||_{L^p(\mathcal{O})}$ is also bounded. We will need this condition in order to ensure the weak convergence of the minimizing sequence in $W^{1,p}(\mathcal{O})$.

Theorem 10. *Under Assumptions 1 and 2 there exists a minimizer of J in $W^{1,p}(\mathcal{O})$.*

Proof. We will first show that J is weakly lower semicontinuous. Let $\{u_n\} \subset W^{1,p}(\mathcal{O})$ be a sequence such that $u_n \rightharpoonup u$ in $W^{1,p}(\mathcal{O})$. Define $L := \liminf J(u_n)$. There exists a subsequence $\{J(u_{n_k})\} \subset \mathbb{R}$ such that $\lim J(u_{n_k}) = L$.

Since $u_{n_k} \rightharpoonup u$ in $W^{1,p}(\mathcal{O})$, by the compact embedding $W^{1,p}(\mathcal{O}) \hookrightarrow L^p(\mathcal{O})$ we know that there exists a subsequence $\{u_{n_r}\}$ such that $u_{n_r} \to u$ in $L^p(\mathcal{O})$.

However, not much can be said for the sequence $\{\nabla u_n\}$ with respect to convergence strongly or a.e. to ∇u (even in terms of subsequences). This makes it difficult to estimate $\liminf J(u_n)$, and compare it with $J(u)$ so as to check weak lower semicontinuity. Furthermore, another complication is that we assume convexity only in P and not jointly on (u, P).

To remedy this situation, we first of all notice that $u_{n_r} \to u$ in $L^p(\mathcal{O})$ implies that there exists a subsequence (denoted the same for simplicity) such that $u_{n_r} \to u$ a.e. in $\mathcal{O}$. We will use Egoroff's theorem and for every $\epsilon > 0$ we will choose $U_\epsilon \subset \mathcal{O}$ such that $\mu(\mathcal{O} \setminus U_\epsilon) < \epsilon$ and with the property that $u_{n_r} \to u$ uniformly in $\mathcal{O}_\epsilon$. Since $||u||_{L^p(\mathcal{O})}$ and $||\nabla u||_{L^p(\mathcal{O})}$ are bounded, for any $\epsilon > 0$ it holds that

$$\mu\left(\left\{x \in \mathcal{O} \, : \, |u(x)| + |\nabla u(x)| \geq \frac{1}{\epsilon}\right\}\right) \to 0, \text{ for } \epsilon \to 0.$$

Let $B_\epsilon := \left(x \in \mathcal{O} \, : \, |u(x)| + |\nabla u(x)| \leq \frac{1}{\epsilon}\right)$ and define $\mathcal{O}_\epsilon = D_\epsilon \cap B_\epsilon$. When $x \in \mathcal{O}_\epsilon$ we have at the same time that $u_{n_r} \to u$ uniformly and that both u and ∇u are bounded by $\frac{1}{\epsilon}$.

Assume without loss of generality that $G \geq 0$ (or else work with $G + \beta$ which is positive, by Assumption 1). Since $\mathcal{O}_\epsilon \subset \mathcal{O}$ and $G \geq 0$ we have that

$$J(u_n) = \int_{\mathcal{O}} G(x, un, \nabla u_n)dx \geq \int_{\mathcal{O}_\epsilon} G(x, u_n, \nabla u_n)dx.$$

where to simplify notation we denote the chosen subsequence $\{u_{n_r}\}$ by $\{u_n\}$. But $P \mapsto G(x,u,P)$ is convex for all $(x,u) \in \mathcal{O} \times \mathbb{R}$ therefore by the standard convexity inequality

$$G(x, u_n, \nabla u_n) \geq G(x, u_n, \nabla u) + \frac{\partial G}{\partial p}(x, u_n, \nabla u) \cdot (\nabla u_n - \nabla u), \text{ a.e.}$$

This lead to the inequality

$$J(u_n) \geq \int_{\mathcal{O}_\epsilon} G(x, u_n, \nabla u)dx + \int_{\mathcal{O}_\epsilon} \frac{\partial G}{\partial p}(x, u_n, \nabla u) \cdot (\nabla u_n - \nabla u)dx. \tag{21}$$

Now take the limit as $r \to \infty$. Since in $\mathcal{O}_\epsilon$ we have uniform convergence of u_{n_r} to u for the first term we have

$$\int_{\mathcal{O}_\epsilon} G(x, u_n, \nabla u)dx \to \int_{\mathcal{O}_\epsilon} G(x, u, \nabla u)dx$$

Now, for the second term we have that $\nabla u_n - \nabla u \rightharpoonup 0$ in $W^{1,p}(\mathcal{O})$ and by the properties of G, $\frac{\partial G}{\partial p}(x, u_n, \nabla u) \to \frac{\partial G}{\partial p}(x, u, \nabla u)$ uniformly in $\mathcal{O}_\epsilon$. Therefore,

$$\int_{\mathcal{O}_\epsilon} \frac{\partial G}{\partial p}(x, u_n, \nabla u) \cdot (\nabla u_n - \nabla u)dx \to 0.$$

With this information, we now go to the limit as $n \to \infty$ in (21). This leads to

$$L \geq \int_{\mathcal{O}_\epsilon} G(x, u, \nabla u)dx, \ \ \forall \, \epsilon > 0. \tag{22}$$

We are close to obtaining the result we desire, but our result is still on the approximation set $\mathcal{O}_\epsilon$. However, since $G > 0$ and $\mu(\mathcal{O}_\epsilon) \to \mu(\mathcal{O})$ as $\epsilon \to 0$, the monotone convergence theorem guarantees that

$$\lim_{\epsilon \to 0} \int_{\mathcal{O}_\epsilon} G(x, u, \nabla u)dx = \lim_{\mathcal{O}} G(x, u, \nabla u)dx = J(u)$$

Therefore, going to the limit as $\epsilon \to 0$ in (22) we finally obtain $L \geq J(u)$ and since $L = \liminf J(u_n)$ we have the weak lower semicontinuity.

Now, let $\{u_n\} \subset W^{1,p}(\mathcal{O})$ be a minimizing sequence, i.e., a sequence such that $\lim J(u_n) = m$, where $m = \inf_{u \in W^{1,p}(\mathcal{O})} J(u)$. By the coercivity inequality it is clear that

$$\int_{\mathcal{O}} |\nabla u_n|^p dx \leq C\, m + \epsilon, \ \ \forall \, \epsilon > 0,$$

so that

$$||\nabla u_n||_{L^p(\mathcal{O})} < C, \tag{23}$$

for an appropriate constant. If we were working in $W_0^{1,p}(\mathcal{O})$ this would be an equivalent norm (by Poincaré's inequality) and then we would automatically have a weakly convergent subsequence of the minimizing sequence in this space. However, in the whole of $W^{1,p}(\mathcal{O})$ we need a slight modification of this argument. We write any element u of $W^{1,p}(\mathcal{O})$ as the sum $U + w$ where $w \in W_0^{1,p}(\mathcal{O})$ and U is the boundary condition (in the sense of traces). This leads to

$$||u_n||_{L^p(\mathcal{O})} = ||u_n - U + U||_{L^p(\mathcal{O})} \leq ||u_n - U||_{L^p(\mathcal{O})} + ||U||_{L^p(\mathcal{O})}$$

and we then can apply Poincaré inequality on w, to finally obtain

$$||u_n||_{L^p(\mathcal{O})} \leq C||\nabla u_n - \nabla U||_{L^p(\mathcal{O})} + ||U||_{L^p(\mathcal{O})} < C \tag{24}$$

since ∇U and U are bounded in $L^p(\mathcal{O})$ (because $U \in W^{1,p}(\mathcal{O})$). Combinining (23) and (24) we get that

$$||u_n||_{W^{1,p}(\mathcal{O})} < C,$$

and by the reflexivity of $W^{1,p}(\mathcal{O})$ there exists a weakly convergent subsequence of the minimizing subsequence in this space. Then, employing the weak lower semicontinuity property of J, we obtain the existence of the minimizer. □

Remark 16. The proof would have been a lot simpler if we have assumed that $(u, P) \mapsto G(x, u, P)$ is convex for every $x \in \mathcal{O}$. Then, as long as G is C^1 with respect to (u, P) the convexity inequality leads directly to the weak lower semicontinuity without the need to resort to the extraction of the subsequence of u_n that converges uniformly to u, and without the need to use the approximation of the sequence on $\mathcal{O}_\epsilon$.

The uniqueness of the minimizer requires stronger conditions on the functional. For example if $(u, P) \mapsto G(x, u, P)$ is strictly convex for all $x \in \mathcal{O}$, then the functional J is strictly convex and the minimizer is unique. Alternatively, we may assume $G = G(x, P)$ (independent of u) and C^2 satisfying the condition $D_P^2 G(x, P)\xi \cdot \xi \geq \theta |\xi|^2$, for some $\theta > 0$, and for all $P, \xi \in \mathbb{R}^d$, $x \in \mathcal{O}$. This is a uniform convexity assumption, related to ellipticity of the differential operator which is related to the functional.

6.2 Generalizations

We now consider the functional $J : \mathbb{X} := W^{1,p}(\mathcal{O}) \to \mathbb{R}$ defined by

$$J(u) = \int_{\mathcal{O}} G(x, u(x), \nabla u(x)) dx. \tag{25}$$

Our aim is to impose such conditions on the function $G : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d$ so that the functional J admits a minimum, through application of the general theory developed above. Then, using the relevant Euler-Lagrange equation we will derive results for a class of elliptic nonlinear PDEs related to the functional. We will impose more general conditions than those of the last section (see [3] for more details).

We will first consider the concept of Caratheodory functions.

Definition 13. A function $f : \mathcal{O} \to \mathbb{R}^m$ is called Caratheodory if

(i) $P \mapsto f(x, P)$ is continuous a.e. in x,

(ii) $\mathsf{x} \mapsto f(x, P)$ is measurable for every $P \in \mathbb{R}^m$.

Caratheodory functions may be approximated by continuous functions.

Theorem 11. *Assume that $\mathcal{O} \subset \mathbb{R}^d$ is bounded and measurable and $S \subset \mathbb{R}^m$ compact. If f is a Caratheodory function then for every $\epsilon > 0$ there exists a compact set $K_\epsilon \subset \mathcal{O}$ such that $\mu(\mathcal{O} \setminus K_\epsilon) < \epsilon$ and $f : K_\epsilon \times S \to \mathbb{R}$ is continuous.*

Our stading assumption in this section if that G is a Caratheodory function. We will further assume that:

Assumption 3 (Assuptions on G)**.**

(i) $P \mapsto G(x, u, P)$ is a convex function for every $(x, u) \in \mathcal{O} \times \mathbb{R}$

(ii) There exist $p > q \geq 1$ and $C_1 > 0$, $C_2, C_3 \in \mathbb{R}$ such that

$$G(x, u, P) \geq C_1 |P|^p + C_2 |u|^q + C_3, \;\; \forall\, (x, u, P) \in \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d$$

We will need to following simpler proposition.

Proposition 7. *Let $\mathcal{O} \subset \mathbb{R}^d$ be an open set, $q \geq 1$ and let q' be the conjugate exponent of q, $1/q' + 1/q = 1$. Let $G : \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d$ be a Caratheodory function such that*

$$G(x, P) \geq a(x) \cdot P + b(x), \ \ \forall \ P \in \mathbb{R} \times \mathbb{R}^d, \ x \in \mathcal{O},$$

where $a \in L^{q'}(\mathcal{O}; \mathbb{R}^d)$, $b \in L^1(\mathcal{O})$ and $\cdot$ is the scalar product in $\mathbb{R}$.

If $P \mapsto f(x, P)$ is convex, then the functional $J : L^q(\mathcal{O}; \mathbb{R}^d) \to \mathbb{R}$ defined by

$$J(P) := \int_{\mathcal{O}} G(x, P(x))dx$$

enjoys the following lower semicontinuity property:

If

$$P_n \rightharpoonup P \quad in \ L^q(\mathcal{O})$$

then

$$\liminf_n J(P_n) \geq J(P).$$

Proof. Consider a sequence $\{P_n\} \subset L^q(\mathcal{O}; \mathbb{R}^d)$ such that $P_n \rightharpoonup P$ in $L^q(\mathcal{O}; \mathbb{R}^d)$. Without loss of generality let us assume that $G \geq 0$.

We will first show that J is strongly lower semicontinuous. To this end let $P_n \to P$ in $L^q(\mathcal{O}; \mathbb{R}^d)$. Then there exists a subsequence $\{P_{n_k}\}$ such that $P_{n_k} \to P$ a.e. Since $G \geq 0$ we may apply Fatou's lemma to obtain

$$\liminf_n \int_{\mathcal{O}} G(x, P_n(x))dx \geq \int_{\mathcal{O}} \liminf_n G(x, P_n(x))dx.$$

Since G is convex in P is is lower semicontinuous with respect to that variable so that

$$\liminf_n G(x, P_n(x)) \geq G(x, P(x))$$

from which it follows that

$$\liminf_n \int_{\mathcal{O}} G(x, P_n(x))dx \geq \int_{\mathcal{O}} G(x, P(x))dx$$

which is the strong lower semicontinuity property for J.

Since $P \mapsto G(x, P)$ is convex for every $x \in \mathcal{O}$, the functional J is convex and therefore strong lower semicontinuity guarantees weak lower semicontinuity as well.

If $G < 0$ then by assumption $G(x, P) - a(x) \cdot P - b(x) \geq 0$ and this is weakly lower semicontinuous. If $P_n \rightharpoonup P$ then $a(x) \cdot P_n + b(x) \to a(x) \cdot P + b(x)$ so that $G(x, P)$ is weakly lower semicontinuous. □

Theorem 12. *Let $\mathcal{O} \subset \mathbb{R}^d$ be an open set, $p, q \geq 1$ and let q' be the conjugate exponent of q, $1/q' + 1/q = 1$. Let $G : \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d$ be a Caratheodory function such that*

$$G(x, u, P) \geq a(x) \cdot P + b(x) + c|u|^p, \ \ \forall \ (u, P) \in \mathbb{R} \times \mathbb{R}^d, \ x \in \mathcal{O},$$

where $a \in L^{q'}(\mathcal{O}; \mathbb{R}^d)$, $b \in L^1(\mathcal{O})$ and $\cdot$ is the scalar product in $\mathbb{R}$.

If $P \mapsto f(x, u, P)$ is convex, then the functional $J : L^p(\mathcal{O}) \times L^q(\mathcal{O}; \mathbb{R}^d) \to \mathbb{R}$ defined by

$$J(u, P) := \int_{\mathcal{O}} G(x, u(x), P(x))dx$$

enjoys the following lower semicontinuity property: If

$$u_n \to u \quad in \ L^p(\mathcal{O}) \quad and P_n \rightharpoonup P \quad in \ L^q(\mathcal{O})$$

then

$$\liminf_n J(u_n, P_n) \geq J(u, P).$$

Proof. Assume without loss of generality that $G \geq 0$.

We will show (see Lemma 3) that for every $\epsilon > 0$ there exists a measurable set $\mathcal{O}_\epsilon \subset \mathcal{O}$ and a subsequence $n_k \to \infty$ such that $\mu(\mathcal{O} \setminus \mathcal{O}_\epsilon) < \epsilon$ and

$$\int_{\mathcal{O}_\epsilon} |G(x, u_{n_k}(x), P_{n_k}(x)) - G(x, u(x), P_{n_k}(x)| dx < \epsilon \, \mu(\mathcal{O}) \tag{26}$$

where μ is the Lebesgue measure.

Now define $\bar{G}(x, P) = \mathbb{1}_{[\mathcal{O}_\epsilon]}(x) G(x, u(x), P)$. The function $P \mapsto \bar{G}(x, P)$ is convex and we may invoke Proposition 7 to show that the functional $\bar{J} : L^q(\mathcal{O}; \mathbb{R}^d) \to \mathbb{R}$ defined by

$$\bar{J}(P) := \int_{\mathcal{O}} \bar{G}(x, P(x)) dx$$

is weakly lower semicontinuous. Since $P_{n_k} \rightharpoonup P$ in $L^q(\mathcal{O}; \mathbb{R}^d)$ we have that

$$\liminf_{n_k} \bar{J}(P_{n_k}) \geq \bar{J}(P)$$

so that

$$\liminf_{n_k} \int_{\mathcal{O}} \mathbb{1}_{[\mathcal{O}_\epsilon]}(x) G(x, u(x), P_{n_k}(x)) dx \geq \int_{\mathcal{O}} \mathbb{1}_{[\mathcal{O}_\epsilon]}(x) G(x, u(x), P(x)) dx$$

This holds for u fixed at the limit of the sequence u_n. Ideally we would like this inequality to hold when on the left hand side of the above inequality we substitute u_{n_k} instead of u. This is where we need the approximation result (28) of Lemma 3. Since

$$\int_{\mathcal{O}_\epsilon} \left(G(x, u(x), P_{n_k}(x)) - G(x, u_{n_k}(x), P_{n_k}(x) \right) dx \leq \int_{\mathcal{O}_\epsilon} |G(x, u_{n_k}(x), P_{n_k}(x)) - G(x, u(x), P_{n_k}(x)| dx$$

it holds that

$$\int_{\mathcal{O}_\epsilon} G(x, u(x), P_{n_k}(x)) dx - \int_{\mathcal{O}_\epsilon} |G(x, u_{n_k}(x), P_{n_k}(x)) - G(x, u(x), P_{n_k}(x)| dx \leq \int_{\mathcal{O}_\epsilon} G(x, u_{n_k}(x), P_{n_k}(x) dx$$

and by (28)

$$\int_{\mathcal{O}_\epsilon} G(x, u(x), P_{n_k}(x)) dx - \epsilon \, \mu(\mathcal{O}) \leq \int_{\mathcal{O}_\epsilon} G(x, u_{n_k}(x), P_{n_k}(x) dx.$$

Since $\mathcal{O}_\epsilon \subset \mathcal{O}$ and $G \geq 0$ it holds that

$$\int_{\mathcal{O}_\epsilon} G(x, u_{n_k}(x), P_{n_k}(x) dx \leq \int_{\mathcal{O}} G(x, u_{n_k}(x), P_{n_k}(x) dx,$$

so that finally,

$$\int_{\mathcal{O}_\epsilon} G(x, u(x), P_{n_k}(x)) dx - \epsilon \, \mu(\mathcal{O}) \leq \int_{\mathcal{O}} G(x, u_{n_k}(x), P_{n_k}(x) dx$$

or equivalently

$$\bar{J}(P_{n_k}) - \epsilon \, \mu(\mathcal{O}) \leq J(u_{n_k}, P_{n_k}).$$

We take the limit as $n_k \to \infty$, in the above inequality to obtain that

$$\liminf_{n_k} \bar{J}(P_{n_k}) - \epsilon\, \mu(\mathcal{O}) \leq \liminf_{n_k} J(u_{n_k}, P_{n_k})$$

and recall the weak lower semicontinuity of $\bar{J}$ to obtain

$$\bar{J}(P) - \epsilon\, \mu(\mathcal{O}) \leq \liminf_{n_k} J(u_{n_k}, P_{n_k}).$$

By the definition of $\bar{J}(P)$ this inequality yields

$$\int_{\mathcal{O}} \mathbb{1}_{[\mathcal{O}_\epsilon]} G(x, u(x), P(x))dx - \epsilon\, \mu(\mathcal{O}) \leq \liminf_{n_k} J(u_{n_k}, P_{n_k}).$$

This holds for every $\epsilon > 0$, so we take the limit as $\epsilon \to 0$. Since $G \geq 0$ the monotone convergence theorem can be used to show that the integral on the left hand side converges to $J(u, P)$ thus leading to

$$J(u, P) \leq \liminf_{n_k} J(u_{n_k}, P_{n_k})$$

which is the weak lower semicontinuity property. □

Lemma 3. *If $\mathcal{O}$ is bounded, for every $\epsilon > 0$ there exists a measurable set $\mathcal{O}_\epsilon \subset \mathcal{O}$ and a subsequence $n_k \to \infty$ such that $\mu(\mathcal{O} \setminus \mathcal{O}_\epsilon) < \epsilon$ and*

$$\int_{\mathcal{O}_\epsilon} |G(x, u_{n_k}(x), P_{n_k}(x)) - G(x, u(x), P_{n_k}(x)|dx < \epsilon\, \mu(\mathcal{O}) \tag{27}$$

where μ is the Lebesgue measure.

Theorem 13. *Let $\mathcal{O} \subset \mathbb{R}^d$ be bounded with Lipschitz boundary. Suppose that G satisfies*

$$G(x, u, P) \geq a(x) \cdot P + b(x) + c|u|^r, \;\; \forall\ P \in \mathbb{R} \times \mathbb{R}^d,\;\; x \in \mathcal{O},$$

with $a \in L^{p'}(\mathcal{O})$, ($p'$ is the conjugate exponent of p, $p \geq 1$), $b \in L^1(\mathcal{O})$, $c \in \mathbb{R}$ and $r \in [1, \frac{dp}{d-p}$ if $p < d$ and $r \in [1, \infty)$ if $p \geq d$. If $P \mapsto G(x, u, P)$ is convex, then, $J : \mathbb{X} := W^{1,p}(\mathcal{O}) \to \mathbb{R}$ defined by

$$J(u) = \int_{\mathcal{O}} G(x, u(x), \nabla u(x))dx,$$

is weakly lower semicontinuous.

Proof. Consider a sequence $\{u_n\} \subset W^{1,p}(\mathcal{O})$ such that $u_n \rightharpoonup u$ in $\mathbb{X} := W^{1,p}(\mathcal{O})$. Then, $\nabla u_n \rightharpoonup \nabla u$ in $L^p(\mathcal{O})$ and by the Rellich-Kondrachov compact embeddings (see Theorem 15) $u_n \to u$ in $L^r(\mathcal{O})$ (for a subsequence). Applying Theorem 12 for u and $P = \nabla u$ we have the stated result. □

The weak lower semicontinuity result now allows us to guarantee the existence of a minimizer for the functional J.

7 Connection with nonlinear PDEs: The Euler-Lagrange equation

Under certain conditions, we will show that the functional J is Gâteaux differentiable, therefore, the first order condition leads to a nonlinear PDE which is the Euler-Lagrange equation for J. It is important to realize that this is true under certain restrictions on G. One may construct examples in which this is not true even in one spatial dimension.

Assumption 4 (Growth condition)**.** Assume that the functions $\frac{\partial}{\partial u}G$, $\frac{\partial}{\partial P}G$ are Caratheodory functions and satisfy the growth conditions

$$\left|\frac{\partial}{\partial u}G\right| < a(x) + \beta\,(|u|^r + |P|^r),$$
$$\left|\frac{\partial}{\partial P}G\right| < a(x) + \beta\,(|u|^r + |P|^r),$$

where $a \in L^{p/r}(\mathcal{O})$ and $\beta > 0$.

Remark 17. The choice of r in the above condition can be either $r = p$ or $r = p - 1$.

Theorem 14. *Under the growth assumption 4 for the choice $r = p$, the minimizer satisfies the Euler-Lagrange equation in weak form for test functions $v \in C_0^\infty(\mathcal{O})$.*

Proof. The strategy of proof is to show that J is Gâteaux differentiable and recognize the Euler-Lagrange equation as the first order condition for the minimum.

The key to the existence of the Gâteaux derivative is the existence of the limit

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}\int_{\mathcal{O}}\left(G(x,u(x)+\epsilon v(x),\nabla u(x)+\epsilon\nabla v(x)) - G(x,u(x),\nabla u(x))\right)dx.$$

Since G is differentiable we use the obvious formula

$$\int_0^\epsilon \frac{d}{ds}G(x,u(x)+s\,v(x),\nabla u(x)+s\,\nabla v(x))ds =$$
$$G(x,u(x)+\epsilon v(x),\nabla u(x)+\epsilon\nabla v(x)) - G(x,u(x),\nabla u(x))$$

which by a simple change of variable of integration $s = \epsilon t$ yields

$$\int_0^1 \frac{d}{dt}G(x,u(x)+\epsilon\,t\,v(x),\nabla u(x)+\epsilon\,t\;\nabla v(x))dt$$
$$= G(x,u(x)+\epsilon v(x),\nabla u(x)+\epsilon\nabla v(x)) - G(x,u(x),\nabla u(x)),$$

so that

$$\frac{1}{\epsilon}(J(u+\epsilon\,v)-J(u)) = \frac{1}{\epsilon}\int_{\mathcal{O}}\left(\int_0^1\frac{d}{dt}G(x,u(x)+\epsilon\,t\,v(x),\nabla u(x)+\epsilon\,t\,\nabla v(x))dt\right)dx.$$

But

$$\frac{d}{dt}G(x,u(x)+\epsilon\,t\,v(x),\nabla u(x)+\epsilon\,t\;\nabla v(x)) =$$
$$\frac{\partial}{\partial u}G(x,u(x)+\epsilon\,t\,v(x),\nabla u(x)+\epsilon\,t\;\nabla v(x))\epsilon v(x) +$$
$$\frac{\partial}{\partial P}G(x,u(x)+\epsilon\,t\,v(x),\nabla u(x)+\epsilon\,t\;\nabla v(x))\cdot\epsilon\nabla v(x)$$

so that

$$\begin{aligned}\frac{1}{\epsilon}(J(u+\epsilon\,v)-J(u)) \;=\; &\int_{\mathcal{O}}\int_0^1\left(\frac{\partial}{\partial u}G(x,u(x)+\epsilon\,t\,v(x),\nabla u(x)+\epsilon\,t\;\nabla v(x))v(x)\right.\\ &\left.+\frac{\partial}{\partial P}G(x,u(x)+\epsilon\,t\,v(x),\nabla u(x)+\epsilon\,t\;\nabla v(x))\cdot\nabla v(x)\right)dt\,dx.\end{aligned}$$

Consider the sequence of functions $\{f^\epsilon\}$ defined by

$$\begin{aligned} f^\epsilon(x) \;&:=\; \int_0^1 \Big(\frac{\partial}{\partial u} G(x, u(x) + \epsilon\, t\, v(x), \nabla u(x) + \epsilon\, t\; \nabla v(x)) v(x) + \\ &\quad \frac{\partial}{\partial P} G(x, u(x) + \epsilon\, t\, v(x), \nabla u(x) + \epsilon\, t\; \nabla v(x)) \cdot \nabla v(x) \Big)\, dt. \end{aligned}$$

Clearly, $f^\epsilon(x) \to f$ a.e in $\mathcal{O}$ where

$$f(x) := \frac{\partial}{\partial u} G(x, u(x), \nabla u(x)) v(x) + \frac{\partial}{\partial P} G(x, u(x), \nabla u(x)) \cdot \nabla v(x).$$

By the assumptions on G it follows that $|f^\epsilon(x)| < \rho(x)$ for every $\epsilon > 0$ where $\rho \in L^1(\mathcal{O})$. Indeed, for any $v \in C_0^\infty(\mathcal{O})$, it follows that

$$\left| \frac{\partial}{\partial u} G(x, u(x) + \epsilon\, t\, v(x), \nabla u(x) + \epsilon\, t\; \nabla v(x)) v(x) \right| \le \\ (\alpha(x) + \beta\, (|u(x) + \epsilon\, t\, v(x)|^p + |\nabla u(x) + \epsilon\, t\; \nabla v(x)|^p)\; |v(x)|$$

and

$$\left| \frac{\partial}{\partial P} G(x, u(x) + \epsilon\, t\, v(x), \nabla u(x) + \epsilon\, t\; \nabla v(x)) \nabla v(x) \right| \le \\ (\alpha(x) + \beta\, (|u(x) + \epsilon\, t\, v(x)|^p + |\nabla u(x) + \epsilon\, t\; \nabla v(x)|^p)\; |\nabla v(x)|.$$

We add these two inequalities and note that for all ϵ and t the functions $|u(x) + \epsilon\, t\, v(x)|^p$ and $|\nabla u(x) + \epsilon\, t\; \nabla v(x)|^p$ are in $L^1(\mathcal{O})$ (since $u \in W^{1,p}(\mathcal{O})$ and $v \in C_0^\infty(\mathcal{O})$). We then take the supremum on the right hand side over all ϵ and t and prove the claim that for every $\epsilon > 0$, f^ϵ is bounded by an integrable function. Therefore using the Lebesgue dominated convergence theorem it follows that

$$DJ(u) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} (J(u + \epsilon\, v) - J(u)) = \lim_{\epsilon \to 0} \int_{\mathcal{O}} f^\epsilon(x) dx = \int_{\mathcal{O}} f(x) dx$$

so that

$$\langle DJ(u), v \rangle = \int_{\mathcal{O}} \left(\frac{\partial}{\partial u} G(x, u(x), \nabla u(x)) v(x) + \frac{\partial}{\partial P} G(x, u(x), \nabla u(x)) \cdot \nabla v(x) \right) dx, \;\; v \in C_0^\infty(\mathcal{O}).$$

Therefore, if u is a minimizer then it must hold that

$$0 = \langle DJ(u), v \rangle = \int_{\mathcal{O}} \left(\frac{\partial}{\partial u} G(x, u(x), \nabla u(x)) v(x) + \frac{\partial}{\partial P} G(x, u(x), \nabla u(x)) \cdot \nabla v(x) \right) dx, \;\; \forall\, v \in C_0^\infty(\mathcal{O}).$$

This equation is a weak form of the Euler-Lagrange equation. □

Remark 18. For the choice $r = p - 1$, the minimizer satisfies the same Euler-Lagrange equation but for a wider class of test functions, ii.e. $v \in W_0^{1,p}(\mathcal{O})$. This follows from the fact that then we may derive an L^1 bound for the sequence $\{f^\epsilon\}$, for $v \in W_0^{1,p}(\mathcal{O})$ rather than $v \in C_0^\infty$ (by applications of Hölder inequality). This, in turn, allows the use of the Lebesgue dominated convergence theorem to derive the weak form of the Euler-Lagrange equation.

8 Appendix: Useful results

Theorem 15 (Rellich-Kontrachov). *Suppose that $\mathcal{O} \subset \mathbb{R}^d$ is bounded and of class C^1. Then,*

(i) $W^{1,p}(\mathcal{O}) \hookrightarrow L^q(\mathcal{O})$, $q \in [1, p')$, $\frac{1}{p'} = \frac{1}{p} - \frac{1}{d}$, *if* $p < d$

(ii) $W^{1,p}(\mathcal{O}) \hookrightarrow L^q(\mathcal{O})$, $q \in (p, \infty)$, *if* $p = d$,

(iii) $W^{1,p}(\mathcal{O}) \hookrightarrow C(\bar{\mathcal{O}})$, *if* $p > d$,

the embeddings being compact.

Theorem 16 (Egoroff). *Assume $\mathcal{O}$ is such that $\mu(\mathcal{O}) < \infty$. Let $\{f_n\}$ be a sequence of measurable functions, such that $f_n \to f$ a.e. in $\mathcal{O}$. Then, for every $\epsilon > 0$, there exists an $\mathcal{O}_\epsilon \subset \mathcal{O}$ such that $\mu(\mathcal{O} \setminus \mathcal{O}_\epsilon) \le \epsilon$ and $f_n \to f$, uniformly in $\mathcal{O}_\epsilon$.*

Lemma 4. *If $\mathcal{O}$ is bounded, for every $\epsilon > 0$ there exists a measurable set $\mathcal{O}_\epsilon \subset \mathcal{O}$ and a subsequence $n_k \to \infty$ such that $\mu(\mathcal{O} \setminus \mathcal{O}_\epsilon) < \epsilon$ and*

$$\int_{\mathcal{O}_\epsilon} |G(x, u_{n_k}(x), P_{n_k}(x)) - G(x, u(x), P_{n_k}(x)|dx < \epsilon\, \mu(\mathcal{O}) \tag{28}$$

where μ is the Lebesgue measure.

Proof. We provide a constructive proof. Since u_n is strongly convergent in $L^p(\mathcal{O})$ and P_n is weakly convergent in $L^q(\mathcal{O}; \mathbb{R}^d)$ they are both bounded in the corresponding norms. Therefore, for every $\epsilon > 0$ one may find a constant $C(\epsilon)$ such that the sets

$$B_{n,\epsilon} := \{x \in \mathcal{O} : |u_n(x)| < C(\epsilon), \text{ and } |P_n(x)| < C(\epsilon)\}$$

have measure $\mu(B_n) \ge 1 - \epsilon/3$.

Since G is a Caratheodory function we may find a compact subset of $B^c_{n,\epsilon}$, that will be called $D_{n,\epsilon}$ such that $D_{n,\epsilon} \subset B^c_{n,\epsilon}$ and $\mu(B^c_{n,\epsilon} \setminus D_{n,\epsilon}) < \epsilon/3$ and when G is restricted to $D_{n,\epsilon} \times \{u \in \mathbb{R} : |u| < C_\epsilon\} \times \{P \in \mathbb{R}^d, : |P| < C_\epsilon\}$ it is continuous. The continuity of the restriction of G implies that for every $\epsilon > 0$ we may find a δ depending on ϵ, denoted therefore by $\delta(\epsilon)$ such that $|G(x, v_1, P) - G(x, v_2, P)| < \epsilon$ as long as $|v_1 - v_2| < \delta$ and $x \in D_{n,\epsilon}$, $|v_1| < C(\epsilon)$, $|v_2| < C(\epsilon)$ and $|P| < C(\epsilon)$.

But we know that $u_n \to u$ in $L^p(\mathcal{O})$. This implies that for the chosen $\delta(\epsilon)$ we may find $N = N(\delta)$ such that $E_{n,\delta} = \{x \in \mathcal{O} : |u_n(x) - u(x)| < \delta(\epsilon)\}$ has measure $\mu(E_{n,\delta}) > 1 - \epsilon/3$ as long as $n > N(\delta)$.

Define $\mathcal{O}_{n,\epsilon} = D_{n,\epsilon} \cap E_{n,\delta}$. On this set

$$\int_{\mathcal{O}_{n,\epsilon}} |G(x, u_n(x), P_n(x)) - G(x, u(x), P_n(x)|dx < \epsilon\, \mu(\mathcal{O})$$

as long as $n > N(\delta)$ (where $N(\delta)$ ultimately depends on ϵ) and $\mu(\mathcal{O} \setminus \mathcal{O}_{n,\epsilon}) < \epsilon$.

Since this holds for all $\epsilon > 0$ we take the sequence $\epsilon_k = \epsilon/2^k$. Choose the sequence $n_k \to \infty$ such that the above inequality holds for every one of the ϵ in the chosen sequence (i.e. for every k) and define $\mathcal{O}_\epsilon = \bigcap_k \mathcal{O}_{n_k, \epsilon_k}$. Then, this is the required subset of $\mathcal{O}$. □

References

[1] J.P. Aubin and I. Ekeland. *Applied nonlinear analysis.* Wiley New York, 1984.

[2] M. Chipot. *Elements of nonlinear analysis.* Birkhäuser, 2000.

[3] B. Dacorogna. *Direct methods in the calculus of variations*, volume 78. Springer Verlag, 2008.

[4] LC Evans. *Partial Differential Equations.* 1998.

[5] D. Kinderlehrer and G. Stampacchia. *An introduction to variational inequalities.* Academic Press, New York, 1981.