FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

Nonlinear Differential Equations

Lecture notes for the course NMNV406

Scott Congreve

Charles University,
Faculty of Mathematics and Physics,
Department of Numerical Mathematics,
Sokolovská 83, 18675 Prague

Contents

List of Figures

List of Examples

CHAPTER 1

Introduction

We initially look at a brief motivation behind studying nonlinear differential equations as well as some basic results from functional analysis and analysis of partial differential equations.

1.1 Motivation

The studying of the analysis and numerical solution of nonlinear differential equations is important as nonlinear equations appear in modelling of even fairly trivial problems. The following example from Böhmer (2010), which models a fairly simple problem, illustrates the importance of nonlinear differential equations.

Example 1.1 (Bending rod with perpendicular load (Böhmer, 2010, Example 1.1)). Consider a vertical rod of length L which is clamped at the bottom and free to move at the top, with a load P applied perpendicular at the free end; see Figure 1.1(a). For a *small* P the displacement, x, of the end of the rod is proportional to P; i.e., it exhibits a linear relationship

$$x = cP, \tag{1.1}$$

where c is some constant depending on the material properties of the rod. However, a simple experiment with a vertical rod will demonstrate that the displacement behaves non-linearly with respect to P for large P, and may even break under certain loads. As such, it is necessary to derive a nonlinear model to incorporate these effects. We can argue that the (local) strain energy of the rod U at a point with arc length s from the bottom of the rod can be obtained analogously to the kinetic energy of a moving body, see Böhmer (2010, Example

(a) Initial state ($P = 0$)

(b) Bending under load ($P > 0$)

Figure 1.1: Bending rod with perpendicular load

1.1); i.e.,

$$U(s) = \frac{\alpha}{2}\left(\frac{\mathrm{d}^2\varphi}{\mathrm{d}s^2}(s)\right)^2,$$

where $\varphi(s)$ is the angle between the vertical and the tangent to the point s, see Figure 1.1(b), and α is the bending stiffness of the material. We can then calculate the total energy of the deformed rod U_B by simply integrating over the length of the rod L:

$$U_B = \frac{\alpha}{2}\int_0^L \left(\frac{\mathrm{d}^2\varphi}{\mathrm{d}s^2}\right)^2 \mathrm{d}s, \qquad \varphi \in C^1([0, L]). \tag{1.2}$$

Additionally, the potential energy due to moving the top of the rod is given by $-Px(L)$, where $x(s)$ is the displacement at s. By simple trigonometry

$$\frac{\mathrm{d}x}{\mathrm{d}s} = \sin\varphi(s),$$

and, hence, with the fact that $x(0) = 0$ we have that

$$x(L) = \int_0^L \sin\varphi(s)\,\mathrm{d}s. \tag{1.3}$$

We can compute the total potential energy as simply the sum of (1.2) and (1.3)

$$V(\phi) = \frac{\alpha}{2}\int_0^L \left(\frac{\mathrm{d}^2\varphi}{\mathrm{d}s^2}\right)^2 \mathrm{d}s - P\int_0^L \sin\varphi(s)\,\mathrm{d}s. \tag{1.4}$$

One of the fundamental principals of mechanics states that an equilibrium of a system is characterised by a minimum of its potential; i.e., for every small ψ

$$V(\varphi + \psi) \geq V(\varphi).$$

We *claim* the minimum (1.4) is characterised by the nonlinear boundary value problem

$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}s^2} + \lambda\cos\varphi = 0, \qquad \lambda := \frac{P}{\alpha} \tag{1.5}$$

with boundary conditions

$$\varphi(0) = \frac{\mathrm{d}\varphi}{\mathrm{d}s}(L) = 0. \tag{1.6}$$

See Böhmer (2010, Example 1.1) for a demonstration that the solution of the boundary value problem (1.5)–(1.6) is equivalent to the minimisation of V in (1.4). We can show for small φ that the linear and nonlinear models are related. We note that for small angles φ that $\sin\varphi \approx \phi$ and $\cos\varphi \approx 1$; therefore,

$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}s^2} + \lambda\cos\varphi \approx \frac{\mathrm{d}^2\varphi}{\mathrm{d}s^2} + \lambda = 0 \qquad \Longrightarrow \qquad \varphi \approx -\frac{\lambda s^2}{2} + \mu s + \nu$$

where μ and ν are constants. From the boundary conditions (1.6) we can find that $\nu = 0$ and $\mu = L\lambda$. Then,

$$x(L) = \int_0^L \sin\varphi\,\mathrm{d}s \approx \int_0^L \varphi\,\mathrm{d}s = \lambda\left(-\frac{s^3}{6} + L\frac{s^2}{2}\right)\Bigg|_0^L = P\frac{L^3}{3\alpha};$$

hence, setting $c = {}^{L^3}\!/_{3\alpha}$ recovers the linear model equation (1.1).

1.2 Nonlinear (partial) differential equations

We now consider a general initial definition, and classification, of nonlinear differential equations. To this end, we first define some necessary notation.

We denote by $\Omega \subset \mathbb{R}^n$ a bounded measurable domain with Lipschitz continuous boundary. Let $n \in \mathbb{N}$ and define a *multi-index* $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$ with length

$$|\alpha| = \sum_{j=1}^{n} \alpha_j.$$

It is possible to show that for $k \in \mathbb{N}_0$ there exists

$$\kappa = \frac{(n+k)!}{n!k!} \tag{1.7}$$

multi-indices of length $|\alpha| \leq k$. Then, for a function $u(\boldsymbol{x}) : \Omega \to \mathbb{R}$, we define

$$\partial^\alpha u := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}, \tag{1.8}$$

$$\delta_k u := (\partial^\alpha u)_{|\alpha| \leq k} = \left(u, \frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_n}, \frac{\partial^2 u}{\partial x_1^2}, \frac{\partial^2 u}{\partial x_1 \partial x_2}, \ldots, \frac{\partial^k n}{\partial x_n^k} \right), \tag{1.9}$$

$$\widehat{\delta}_k u := (\partial^\alpha u)_{|\alpha| = k} = \left(\frac{\partial^k u}{\partial x_1^k}, \ldots, \frac{\partial^k u}{\partial x_n^k} \right). \tag{1.10}$$

Note, that for simplicity we assume that order of differentiation does not matter for mixed derivatives; e.g., $\frac{\partial^2 u}{\partial x_1 \partial x_2} \equiv \frac{\partial^2 u}{\partial x_2 \partial x_1}$. Therefore, we only include the mixed derivative *once* in $\delta_k u$ and $\widehat{\delta}_k u$. Note that in general we consider so called *weak derivatives*, see Section 1.4.

Definition 1.1. We can write a general nonlinear differential equation of order $k \in \mathbb{N}$ as

$$F(\boldsymbol{x}, \delta_k u(\boldsymbol{x})) = 0, \qquad \text{for } \boldsymbol{x} \in \Omega, \tag{1.11}$$

where $F : \Omega \times \mathbb{R}^\kappa \to \mathbb{R}$ is a function defined over Ω.

We can now classify differential equations in multiple ways according to the type of nonlinearity of the differential equation.

Definition 1.2. A differential equation of order k is said to be

- *linear* if it is linear in the unknown function u and all its derivatives or order less than or equal to k with coefficients depending only on the independent variables ($\boldsymbol{x} \in \Omega$),
- *semilinear* if it is linear in derivatives of order k with coefficients dependent on independent variables only but nonlinear in other terms,
- *quasilinear* if it is linear in derivatives of order k with coefficients depending on independent variables *and* derivatives of order less than k, or
- *(fully) nonlinear* if it is nonlinear in derivatives of order k or has coefficients depending on independent variables *and* derivatives of order k.

In this course we focus mainly on semilinear and quasilinear differential equations.

Definition 1.3 (Divergence form)**.** We can write *linear*, *semilinear*, and *quasilinear* differential equations of order $2k$ in *divergence form*

$$\sum_{|\alpha|\leq k} (-1)^{|\alpha|}\partial^\alpha a_\alpha(\boldsymbol{x}, \delta_k u(\boldsymbol{x})) = f(\boldsymbol{x}), \qquad \text{for } \boldsymbol{x} \in \Omega \tag{1.12}$$

where $k \in \mathbb{N}$, α is an n-dimensional multi-index, $a_\alpha = a_\alpha(x, \xi)$, $|\alpha| \leq k$ is a function of $n + \kappa$ variables over $\boldsymbol{x} \in \Omega$ and $\xi \in \mathbb{R}^\kappa$, and f is a function defined over Ω.

Remark. In general we focus in this course on differential equations of a single real-valued unknown function $u : \Omega \to \mathbb{R}$; however, the above definitions can be extended to systems of differential equations (i.e., $\boldsymbol{u}(\boldsymbol{x}) = (u_1(\boldsymbol{x}), \ldots, u_m(\boldsymbol{x}))$, $m \in \mathbb{N}$) and/or complex valued functions.

Suppose all coefficients a_α in (1.12) have continuous derivatives of order $|\alpha|$, i.e., $a_\alpha \in C^{|\alpha|}(\Omega \times \mathbb{R}^\kappa)$, and $f \in C^0(\Omega)$. Then, $u(\boldsymbol{x})$ over Ω is the *classical solution* of (1.12) if $u \in C^{2k}(\Omega)$ satisfies (1.12).

For second and fourth order partial differential equations we define some common notations. Consider a scalar valued function $u : \Omega \to \mathbb{R}$, a vector-valued function $\boldsymbol{v} : \Omega \to \mathbb{R}^n$, and a matrix-valued function $\underline{\sigma} : \Omega \to \mathbb{R}^{n\times n}$. Then, we define the following operators:

- *gradient operator:*

$$\nabla u := \begin{pmatrix} \dfrac{\partial u}{\partial x_1} \\ \vdots \\ \dfrac{\partial u}{\partial x_n} \end{pmatrix} \qquad \nabla \boldsymbol{v} := \begin{pmatrix} \dfrac{\partial v_1}{\partial x_1} & \cdots & \dfrac{\partial v_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial v_n}{\partial x_1} & \cdots & \dfrac{\partial v_n}{\partial x_n} \end{pmatrix}$$

- *divergence operator:*

$$\nabla \cdot \boldsymbol{v} := \sum_{i=1}^n \frac{\partial v_i}{\partial x_i} \qquad \nabla \cdot \underline{\sigma} := \begin{pmatrix} \displaystyle\sum_{i=1}^n \frac{\partial \sigma_{1i}}{\partial x_i} \\ \vdots \\ \displaystyle\sum_{i=1}^n \frac{\partial \sigma_{ni}}{\partial x_i} \end{pmatrix}$$

- *Laplacian operator:* $\Delta\psi \equiv \nabla \cdot (\nabla\psi)$ for a function ψ
- *biharmonic operator:* $\Delta^2\psi \equiv \Delta(\Delta\psi)$ for a function ψ
- *Hessian operator:* $\nabla^2 u \equiv \nabla(\nabla u)$

Exercise 1.1. Classify the following differential equations as *linear*, *semilinear*, *quasilinear*, or *fully nonlinear*, and justify why.

1. *Poisson equation:* For an unknown function $u : \Omega \to \mathbb{R}$ and known function $f : \Omega \to \mathbb{R}$

$$-\nabla u = f \qquad \in \Omega$$

with suitable boundary conditions.

2. *Chladny sound figures:* Assume a thin flexible square plate fixed at its centre, and uniformly distribute tiny particles (e.g., sand) on the plate. Above the plate a sound wave source, with variable frequency λ is fixed. Pressure from the sound waves induces a load on the plates surfaces, and in the case of resonance vibration of the plate with unmoved nodal lines is induced. The sand collects along these nodal lines. Let $\Omega = [O, L]^2$ be the plate and $u(\boldsymbol{x}) : \Omega \to \mathbb{R}$ be the maximal derivation of vibration of the plate from the fixed horizontal position at each point; then, u satisfies the following (strongly simplified) model, cf. (Böhmer, 2010, Example 1.4):

$$\begin{aligned} \nabla u + \lambda \sin u &= 0 && \text{in } \Omega, \\ \frac{\partial u}{\partial \boldsymbol{n}} &= 0 && \text{on } \partial\Omega. \end{aligned}$$

3. *von Kármán equations:* Consider a plate P under compression ($\Omega \subset \mathbb{R}^2$). It is defined by the Airy stress function $w(\boldsymbol{x}) : \Omega \to \mathbb{R}$ of P at $\boldsymbol{x}$ and the derivation or deflection $u(\boldsymbol{x}) : \Omega \to \mathbb{R}$ of the plate P from the trivial horizontal state. This can be modelled by the von Kármán equations

$$\begin{aligned} D\Delta^2 u - [u, w] &= f, \\ \Delta^2 w + \frac{1}{2}[u, u] &= 0, \end{aligned}$$

where D is a constant, $f : \Omega \to \mathbb{R}$ is a known function, and

$$[u, v] := \frac{\partial^2 u}{\partial x_1^2}\frac{\partial^2 v}{\partial x_2^2} - 2\frac{\partial^2 u}{\partial x_1 \partial x_2}\frac{\partial^2 v}{\partial x_1 \partial x_2} + \frac{\partial^2 u}{\partial x_2^2}\frac{\partial^2 v}{\partial x_1^2}.$$

Note, that this is a system of two equations with two unknown functions.

4. *Non-Newtonian fluid:* Non-Newtonian fluids do not follow Newton's law of viscosity; i.e., they have variable viscosity dependent on stress. In particular, the viscosity of non-Newtonian fluids can change when subject to force; e.g. ketchup becomes runnier when shaken, a suspension of corn starch in water (liquid in normal state) thickens under force such as from low frequency sound waves. One, simple, steady-state model is given by the system

$$\begin{aligned} -\nabla \cdot \{\mu(\boldsymbol{x}, |\underline{e}(\boldsymbol{u})|)\underline{e}(\boldsymbol{u})\} + \nabla p &= \boldsymbol{f} \\ \nabla \cdot \boldsymbol{u} &= 0, \end{aligned}$$

where $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ is a known function, $\mu : \Omega \times \mathbb{R} \to \mathbb{R}$ is a known (potentially nonlinear) function, $\boldsymbol{u} : \Omega \to \mathbb{R}^n$ is the unknown *velocity* vector of the fluid, $p : \Omega \to \mathbb{R}$ is the unknown pressure of the fluid, $\underline{e}(\boldsymbol{u})$ is the symmetric strain tensor defined by

$$e_{ij}(\boldsymbol{u}) := \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right), \qquad i, j = 1, \ldots, n,$$

and $|\cdot|$ denotes the Frobenius norm.

5. *(Steady-state) Navier-Stokes:* The steady-state version of the Navier-Stokes equations is given by

$$\begin{aligned} -\nu\Delta \boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} + \nabla p &= \boldsymbol{f} \\ \nabla\cdot\boldsymbol{u} &= 0, \end{aligned}$$

where $\nu : \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ are known functions, $\boldsymbol{u} : \Omega \to \mathbb{R}^n$ is the unknown *velocity* vector of the fluid, and $p : \Omega \to \mathbb{R}$ is the unknown pressure of the fluid.

6. *Monge-Ampère*

$$\det(\nabla^2 u) - f(\boldsymbol{x}, u, \nabla u) = 0,$$

where $u : \Omega \to \mathbb{R}$ is the unknown function, $f : \Omega \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ is a known function, and $\det(\,\cdot\,)$ is the determinant of the matrix.

1.3 Fundamental results

In order to perform analysis of the proceeding nonlinear differential equations some basic results are required.

Definition 1.4. Let X and Y be normed linear spaces.

1. A subset $K \subset X$ is called *(sequentially) compact* in X if it is closed and every sequence $\{u_n\} \subset K$ has a convergent subsequence in K.

   Equivalently: A subset $K \in X$ is *compact* if for every open covering a finite subcover can be found.

2. A subset $K \subset X$ is called *precompact* (or *relatively compact*) if its closure is compact in X.

3. A general nonlinear operator $A : X \to Y$ is a *compact operator* if the image of the bounded subset $K \subset \mathcal{D}(A)$ of the domain $\mathcal{D}(A)$ of A under A is a precompact subset of Y; i.e., $\overline{A(K)}$ is compact.

   Equivalently: $A : X \to Y$ is *compact* on $\mathcal{D}(A)$ if for any bounded sequence $\{u_n\} \subset \mathcal{D}(A)$ the sequence $\{Au_n\}$ contains a convergent subsequence in Y.

Remark. If $A : X \to Y$ is compact and linear then it is continuous.

We denote by $\mathcal{L}(X, Y)$ the set of continuous linear operators from X to Y and $\mathcal{C}_0(X, Y)$ the set of compact linear operators.

Let X be a normed vector space; then, the set of all linear and continuous functionals on X form the *dual space* X', where for $\ell \in X'$

$$\ell : x \mapsto \langle \ell, x\rangle := \ell(x).$$

The dual space is continuous and we define the space of all continuous linear functions over X' as the *normed bidual space* as $X'' = (X')'$. We note that

$$\langle\,\cdot\,, x\rangle \in X'' \quad \text{for every } x \in X;$$

i.e., it is possible to identify X with a subset of X'', and in some cases with X'' itself.

The norm for the dual space X' is defined as

$$\|\ell\|_{X'} := \sup_{x\in X, \|x\|_X \neq 0} \frac{|\langle \ell, x\rangle|}{\|x\|_X}$$

where $\|\cdot\|_X$ is the norm on X. Additionally, we note that

$$\|\ell\|_{X'} \geq \frac{|\langle \ell, x\rangle|}{\|x\|_X} \quad \forall \ell \in X', v \in X \qquad \implies \qquad |\langle \ell, x\rangle| \leq \|\ell\|_{X'}\|v\|_X \quad \forall \ell \in X', v \in X \tag{1.13}$$

as the case when $\|v\| = 0$ follows trivially.

Definition 1.5. The Banach space X is called *reflexive* if the canonical map $J : X \to X''$ defined by

$$\langle Jx, \ell\rangle = \langle l, x\rangle \qquad \text{for all } \ell \in X', x \in X,$$

is surjective.

Remark. In a reflexive Banach space the canonical map J is linear, isomorphic, and an isometric isomorphism. Additionally, we have that

$$\begin{aligned} X \text{ reflexive} &\implies \text{dual } X' \text{ reflexive} \\ X' \text{ reflexive and complete} &\implies X \text{ reflexive} \end{aligned}$$

In a reflexive Banach space X identifies with X''; hence, we use the notation $X'' = X$ for a reflexive Banach space X.

Definition 1.6. Let X be a normed linear space; then, the sequence $\{u_n\} \subset X$ *converges weakly* (w-converges) to $u \in X$ if for every $\ell \in X'$ it holds that $\langle \ell, u_n\rangle \to \langle \ell, u\rangle$. We denote weak convergence as $u_n \rightharpoonup u$. The sequence $\{\ell_n\} \subset X'$ w^*-converges to $\ell \in X'$ if for every $x \in X$ it holds that $\langle \ell_n, x\rangle \to \langle l, x\rangle$. We denote this as $\ell_n \xrightharpoonup{w^*} \ell$. If X is reflexive then w- and w^*-convergence coincide, and hence we use $\ell_n \rightharpoonup \ell$ for w^*-convergence.

Using the concept of weak convergence we can also talk about *weakly closed* and *weakly compact* sets analogously to *closed* and *compact* sets using weak convergence rather than strong convergence.

Theorem 1.7. *Compact linear operators have the following properties:*

1. *Let X, Y, and Z be Banach spaces, $L_1 \in \mathcal{L}(X, Y)$, and $L_2 \in \mathcal{L}(Y, Z)$; then, $L_2 \circ L_1$ is compact if either L_1 or L_2 is compact.*

2. *A compact $L_1 \in \mathcal{L}(X, Y)$ maps every weakly convergent sequence into a strongly convergent sequence.*

Proposition 1.8. *Let X be a Banach space.*

1. *The sequence $\{u_n\} \subset X$ converges weakly to the point $u \in X$ if the sequence $\{u_n\}$ is bounded and $\lim_{n\to\infty}\langle \ell, u_n\rangle = \langle \ell, u\rangle$ for each continuous linear functional ℓ of some dense subset M of X'.*

2. *If the sequence $\{u_n\}$ weakly converges to u; then,*

$$\|u\| \leq \liminf_{n\to\infty} \|u_n\|.$$

3. *A convex, closed subset $K \subset X$ is* weakly closed*; i.e.,*

$$\{u_n\} \subset K, u_n \rightharpoonup u \qquad \Longrightarrow \qquad u \in K$$

4. *In a* reflexive *Banach space every closed ball is* weakly compact*; i.e., each sequence of elements from this ball contains a weakly convergent subsequence with a limit in this ball (Every bounded sequence contains a weakly convergent subsequence).*

5. *Let $\{u_n\}$ be a bounded sequence in a* reflexive *Banach space X and let all weakly convergent subsequences have u as their (weak) limit; then, the whole sequence converges to u.*

6. *Let K be a convex, bounded, and closed subset in a* reflexive *Banach space X; then, this set is weakly compact.*

7. *Let K be a non-empty, closed, and convex set in a* Hilbert space *H with norm $\|\cdot\|$ and $(\,\cdot\,,\,\cdot\,)$. Then, for every $x \in H$ there exists a $u \in K$ such that*

$$\|x-u\| = \min_{v\in K} \|x-v\|.$$

This element can be characterised as

$$u \in K, (x-u, v-u) \leq 0, \qquad \textit{for all } v \in K.$$

If K is a closed linear subspace of H then u can be characterised by

$$u \in K, (x-u, v-u) = 0, \qquad \textit{for all } v \in K.$$

8. *Let K be a convex closed subset of X. Then, for every $x \notin K$ there exists a functional $\ell \in X'$ such that*

$$\langle \ell, x\rangle > \sup_{y\in K} \langle \ell, y\rangle.$$

9. Uniform boundedness principal: *Let $\mathcal{G} \subset \mathcal{L}(X,Y)$ where Y is a normed linear space; then, the following are equivalent:*
   (a) $\sup\{\|A\| : A \in \mathcal{G}\} < +\infty$,
   (b) $\sup\{\|Ax\| : A \in \mathcal{G}\} < +\infty$ *for every $x \in X$.*

For the Banach spaces X and Y, and $L \in \mathcal{L}(X,Y)$ the *dual operator* $L^d \in \mathcal{L}(Y',X')$ is uniquely determined for every $\ell \in Y'$, $\ell \in Y' \mapsto L^d\ell \in X'$, by

$$\langle \ell, Lx\rangle_{Y'\times Y} = \langle L^d\ell, x\rangle_{X'\times X} \qquad \text{for all } x \in X.$$

L^d is unique, linear, and continuous. Furthermore, for $L, L_1 \in \mathcal{L}(X,Y)$, $L_2 \in \mathcal{L}(Y,Z)$, $\alpha \in \mathbb{C}$,

$$\|L^d\| = \|L\|, \qquad (L_2 \circ L_1)^d = L_1^d \circ L_2^d, \qquad (\alpha L)^d = \overline{\alpha} L^d,$$

and for reflexive Banach spaces $L^{dd} = L$.

Remark. In general we drop the subscript on $\langle \cdot , \cdot \rangle$ unless we need to emphasise the difference.

Lemma 1.9. *An operator $L \in \mathcal{L}(X, Y)$ is compact if and only if the dual operator $L^d \in \mathcal{L}(Y', X')$ is compact.*

In the following we consider X as a Hilbert space and denote the dual as $X^* = X'$.

Theorem 1.10 (Riesz representation theorem). *Let X be a Hilbert space and $\ell \in X' = X^*$; then, there exists a unique $x_\ell \in X$ such that*

$$\langle \ell, x \rangle = \ell(x) = (x, x_\ell), \qquad \text{for all } x \in X$$

and

$$\|x_\ell\|_X = \|\ell\|_{X'}$$

where $(\cdot , \cdot)$ denotes the inner product in the Hilbert space.

Corollary 1.11 (Riesz-isomorphism). *There exists a unique (Riesz-)isomorphism $J_X \in \mathcal{L}(X, X^*)$ such that*

$$J_X x_\ell = \ell, \qquad J_X^{-1}\ell = x_\ell, \qquad \|J_x\| = \|J_x^{-1}\| = 1.$$

Additionally, $X^* = X'$ is a Hilbert space with inner product and norm

$$(x', y')_{X^*} = (J_X^{-1}x', J_X^{-1}y')_X \qquad \text{and} \qquad \|x'\|_{X^*} = \|J_X^{-1}x'\|_X,$$

respectively, for all $x', y' \in X^*$. X and $(X^*)^* = X''$ can be identified by $x_\ell = J_X^{-1}\ell$.

For two Hilbert spaces X and Y with inner products $(\cdot , \cdot)_X$ and $(\cdot , \cdot)_Y$, respectively, let $L \in \mathcal{L}(X, Y)$; then, we can define the *adjoint operator* $L^* \in \mathcal{L}(Y, X)$ for each $y \in Y$ such that

$$(y, Lx)_Y = (L^*y, x)_X \qquad \text{for all } x \in X.$$

The dual and adjoint operators are related:

$$L^* = (J_X)^{-1}L^d J_Y \in \mathcal{L}(Y, X).$$

Only if X and Y are identified with X' and Y', respectively, do we get that $L^* \equiv L^d$. Additionally, $L \in \mathcal{L}(X, X)$ is a *self-adjoint* operator if $L = L^*$ and an *orthogonal operator* if $L = L^2$ and $L = L^*$.

1.4 Sobolev spaces

Given a measurable domain $\mathcal{D} \subset \mathbb{R}^n$, $n \in \mathbb{N}$, we define for $1 \leq p \leq \infty$ the standard Lebesgue spaces

$$L^p(\mathcal{D}) = \{v : \mathcal{D} \to \mathbb{R} : \|v\|_{0,p,\mathcal{D}} < \infty\},$$

where

$$\|v\|_{0,p,\mathcal{D}} := \left(\int_{\mathcal{D}} |v(\boldsymbol{x})|^p \,\mathrm{d}\boldsymbol{x} \right)^{1/p}, \qquad 1 \leq p < \infty, \tag{1.14}$$

$$\|v\|_{0,\infty,\mathcal{D}} := \operatorname*{ess\,sup}_{\boldsymbol{x} \in \mathcal{D}} |v(\boldsymbol{x})|. \tag{1.15}$$

For $p = 2$ we additionally have the inner product

$$(u, v)_{\mathcal{D}} := \int_{\mathcal{D}} uv \, \mathrm{d}\boldsymbol{x}$$

and we often drop the p from the norm subscript; i.e., $\|\cdot\|_{0,\mathcal{D}} \equiv \|\cdot\|_{0,2,\mathcal{D}}$. When $\mathcal{D} = \Omega$ we omit the domain from the subscript of the norm and inner product; i.e., $\|v\|_{0,p}$, $\|v\|_{0,\infty}$, and (u, v).

Lemma 1.12 (Hölder's Inequality). *Let $1 \le p, q \le \infty$ with $1/p + 1/q = 1$; then,*

$$\|fg\|_{0,1,\mathcal{D}} \le \|f\|_{0,p,\mathcal{D}} \|g\|_{0,q,\mathcal{D}}$$

for all $f \in L^p(\mathcal{D})$ and $g \in L^q(\mathcal{D})$.

We use $C^k(\mathcal{D})$ to denote the usual space of k-times differentiable functions, and $C_0^\infty(\mathcal{D}) = \{v \in C^\infty(\mathcal{D}) : \operatorname{supp} v \subset \mathcal{D}\}$ to denote the space of infinitely smooth functions with compact support in $\mathcal{D}$. The definition of the partial derivatives in $\partial^\alpha u$ from Section 1.2 are well-defined for functions $u \in C_0^\infty(\mathcal{D})$

Additionally, we consider $C^{k,1}(\mathcal{D})$ and $C^{k,\mu}(\mathcal{D})$ to denote the space of k-times differentiable functions u where all derivatives $\partial^\alpha u$ of order $|\alpha| \le k$ are *Lipschitz continuous* or *Hölder continuous* with constant $0 < \mu < 1$, respectively; i.e.,

$$|\partial^\alpha u(\boldsymbol{x}) - \partial^\alpha u(\boldsymbol{y})| \le L|\boldsymbol{x} - \boldsymbol{y}|, \qquad \text{for } u \in C^{k,1}(\mathcal{D}),$$
$$|\partial^\alpha u(\boldsymbol{x}) - \partial^\alpha u(\boldsymbol{y})| \le C_H|\boldsymbol{x} - \boldsymbol{y}|^\mu, \qquad \text{for } u \in C^{k,\mu}(\mathcal{D}),$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$, with constants $L > 0$ and $C_H > 0$.

For Sobolev spaces it is necessary to generalise the concept of derivation to *weak derivatives*.

Definition 1.13. For a function $u \in L^p(\mathcal{D})$, $1 \le p \le \infty$, we call a function $v \in L^p(\mathcal{D})$ the α *weak derivative* of u if and only if

$$(w, v)_{\mathcal{D}} = (-1)^{|\alpha|}(\partial^\alpha w, u)_{\mathcal{D}} \quad \text{for all } w \in C_0^\infty(\mathcal{D}).$$

In general we use the notation $\partial^\alpha u$ to denote this weak derivative.

We can then define the *Sobolev space* $W^{k,p}(\mathcal{D})$, $k \in \mathbb{N}_0$, $1 \le p \le \infty$, as the set of all functions in $L^p(\mathcal{D})$ with weak derivatives up to order k:

$$W^{k,p}(\mathcal{D}) := \{u \in L^p(\mathcal{D}) : \partial^\alpha u \in L^p(\mathcal{D}) \; \forall \alpha \in \mathbb{N}_0^n, |\alpha| \le k\}.$$

A Sobolev space $W^{k,p}(\mathcal{D})$, $k \in \mathbb{N}_0$, $1 \le p \le \infty$ is a Banach space with norm

$$\|u\|_{k,p,\mathcal{D}} := \left(\sum_{|\alpha| \le k} \|\partial^\alpha u\|_{0,p,\mathcal{D}}^p \right)^{1/p} \tag{1.16}$$

and seminorm

$$|u|_{k,p,\mathcal{D}} := \left(\sum_{|\alpha| = k} \|\partial^\alpha u\|_{0,p,\mathcal{D}}^p \right)^{1/p}. \tag{1.17}$$

In the special case when $p = 2$ we define the space $H^k(\mathcal{D}) \equiv W^{k,2}(\mathcal{D})$ which is a Hilbert space with inner product

$$(u,v)_{k,\mathcal{D}} := \sum_{|\alpha|\leq k} (\partial^\alpha u, \partial^\alpha v)_{\mathcal{D}}$$

and again drop the p from the subscript of the norms and seminorms; i.e., $\|\cdot\|_{k,\mathcal{D}} \equiv \|\cdot\|_{k,2,\mathcal{D}}$ and $|\cdot|_{k,\mathcal{D}} \equiv |\cdot|_{k,2,\mathcal{D}}$. Additionally, when $\mathcal{D} = \Omega$ we omit the domain from the subscript of the norms, seminorms, and inner products.

We define the spaces $W_0^{k,p}(\mathcal{D})$ and $H_0^k(\mathcal{D})$ as the closure of $C_0^\infty(\mathcal{D})$ with respect to the norms $\|\cdot\|_{k,p,\mathcal{D}}$ and $\|\cdot\|_{k,\mathcal{D}}$, respectively; i.e.

$$W_0^{k,p}(\mathcal{D}) := \overline{C_0^\infty(\mathcal{D})}^{\|\cdot\|_{k,p,\mathcal{D}}}, \qquad H_0^k(\mathcal{D}) := \overline{C_0^\infty(\mathcal{D})}^{\|\cdot\|_{k,\mathcal{D}}}.$$

For a bounded domain $\mathcal{D}$ the seminorm $|\cdot|_{k,p,\mathcal{D}}$ is a norm for the space $W_0^{k,p}(\mathcal{D})$ and is equivalent to the norm $\|\cdot\|_{k,p,\mathcal{D}}$ on $W_0^{k,p}(\mathcal{D})$. In particular, for the Poincaré constant $C_P(\mathcal{D})$,

$$|u|_{k,p,\mathcal{D}} \leq \|u\|_{k,p,\mathcal{D}} \leq C_P|u|_{k,p,\mathcal{D}}, \qquad \text{for all } u \in W_0^{k,p}(\mathcal{D}).$$

A similar result holds for $H_0^k(\mathcal{D})$ when $p = 2$.

Remark. Note that $L^p(\mathcal{D}) \equiv W^{0,p}(\mathcal{D})$.

For $W_0^{k,p}(\mathcal{D})$ we define the dual space as $W^{-k,q}(\mathcal{D}) \equiv (W_0^{k,p}(\mathcal{D}))'$, where $1/p + 1/q = 1$, and define the norm as

$$\|u\|_{-k,q,\mathcal{D}} := \sup_{v\in W_0^{k,p}(\mathcal{D}) normv\neq 0} \frac{\langle u,v\rangle}{\|v\|_{k,p,\mathcal{D}}}.$$

For $p = 2$ we define $H^{-k}(\mathcal{D}) \equiv (H^k(\mathcal{D}))'$.

For normed vector spaces $X \subset Y$ we define the *inclusion* or *embedding*

$$\iota : X \to Y, x \mapsto x.$$

We denote this embedding as $X \hookrightarrow Y$. If the mapping is continuous we say that X is *continuous embedded* into Y and there exists a positive constant $C > 0$ such that

$$\|x\|_Y \leq C\|x\|_X, \qquad \text{for all } x \in X.$$

Additionally, if the mapping is also compact ($\iota \in \mathcal{C}_0(X,Y)$) then we say X is *compactly embedded* into Y and each bounded sequence $\{x_n\}$ in X has a convergent subsequence in Y.

Theorem 1.14 (Sobolev embeddings). *Let $\Omega \subset \mathbb{R}^n$ be a bounded Lipschitz domain, $j \in \mathbb{N}_0$, $k \in \mathbb{N}$, $0 \leq j \leq k$, $1 \leq p,q < \infty$, $0 < \mu < 1$ and define*

$$d := \frac{1}{p} - \frac{k-j}{n}.$$

Then, the following embeddings hold.

1. *If $d \leq 1/q$ and $j \leq k$ then*

$$W^{k,p}(\Omega) \hookrightarrow W^{j,q}(\Omega)$$

is continuous. *If $d < 1/q$ and $j < k$ then this embedding is* compact.

2. *If $d < 0$ then*
$$W^{k,p}(\Omega) \hookrightarrow C^j(\overline{\Omega})$$
is compact. *Additionally, for $d + \alpha/n < 0$ or $d < 0$*
$$W^{k,p}(\Omega) \hookrightarrow C^{j,\mu}(\overline{\Omega}) \hookrightarrow C^j(\overline{\Omega}).$$

3. *If $d < 0$ then*
$$W^{k,p}(\mathbb{R}^n) \hookrightarrow C^j(\mathbb{R}^n)$$
is continuous, *and if $d \leq 1/q$ and $j < k$ then*
$$W^{k,p}(\mathbb{R}^n) \hookrightarrow W^{j,q}(\mathbb{R}^n)$$
is continuous.

4. *For $0 \leq p \leq \infty$ the embeddings*
$$C^{0,\mu}(\overline{\Omega}) \hookrightarrow L^p(\Omega) \quad \text{and} \quad C^{0,1}(\overline{\Omega}) \hookrightarrow L^p(\Omega)$$
are continuous.

5. *If $j < k$ the embeddings*
$$C^{k,\mu}(\overline{\Omega}) \hookrightarrow C^k(\Omega) \hookrightarrow C^j(\Omega) \quad \text{and} \quad C^{k,1}(\overline{\Omega}) \hookrightarrow C^k(\Omega) \hookrightarrow C^j(\Omega)$$
are compact.

Corollary 1.15. *Let $\Omega \subset \mathbb{R}^n$ be a bounded Lipschitz domain, $k \in \mathbb{N}$, and $1 \leq p, q < \infty$. Then, the following embeddings are all* continuous*:*

1. *If $kp < n$ and $1 \leq q \leq np/(n-kp)$, then*
$$W^{k,p}(\Omega) \hookrightarrow L^q(\Omega).$$

2. *If $kp = n$ and $1 \leq q < \infty$, then*
$$W^{k,p}(\Omega) \hookrightarrow L^q(\Omega).$$

3. *If $kp > n$, then*
$$W^{k,p}(\Omega) \hookrightarrow C^0(\overline{\Omega}).$$

4. *If $0 \leq k$ and $1 \leq p \leq q \leq \infty$, then*
$$W^{k,p}(\Omega) \hookrightarrow W^{k,q}(\Omega).$$

Theorem 1.16 (Sobolev-Stein extension theorem)**.** *Let $\Omega \subset \mathbb{R}^n$ have Lipschitz boundary, let $k \in \mathbb{N}_0$ and $1 \leq p \leq \infty$. Then, there exists a bounded operator $\mathfrak{E} : W^{k,p}(\Omega) \to W^{k,p}(\mathbb{R}^n)$ and a constant $C > 0$, independent of v such that $\mathfrak{E}v|_\Omega = v$ and*
$$\|\mathfrak{E}v\|_{k,p,\mathbb{R}^n} \leq C\|v\|_{k,p,\Omega} \qquad \text{for all } v \in W^{k,p}(\Omega).$$

1.5 Derivatives in Banach spaces

On Banach spaces we define two types of differentiation.

Definition 1.17. Let X and Y be Banach spaces, $F : \mathcal{D} \subset X \to Y$, and $\mathcal{U} := \mathcal{U}(x_0) \subset \mathcal{D}$ be an open neighbourhood of $x_0 \in \mathcal{U}$.

1. F is called *Fréchet differentiable* in x_0 if there exists a $L_F \in \mathcal{L}(X, Y)$, depending on x_0 but not h, such that

$$F(x_0 + h) - F(x_0) = L_F h + o(h), \qquad \text{for } h \to 0, x_0 + h \in \mathcal{U}. \tag{1.18}$$

Additionally, $F'_F(x_0) := F'(x_0) := L_F$ is called the *Fréchet derivative* of F in x_0.

2. F is called *Gâteaux differentiable* in x_0 if for every h, such that $\|h\| = 1$, there exists a $F'_G(x_0, h) \in Y$ such that

$$F(x_0 + th) - F(x_0) = tF'_G(x_0, h) + o(t), \qquad \text{for } t \to 0, x_0 + th \in \mathcal{U}; \tag{1.19}$$

where $o(t)$ depends on h. If (1.19) is valid only for fixed h then F is called *Gâteaux differentiable* in x_0 *in the direction* h. Furthermore, $F'_G(x_0, h)$ is called the *Gâteaux derivative* of F in x_0 *in the direction* of h.

As a consequence of (1.19) we have that

$$F'_G(x_0, h) = \lim_{t \to 0} \frac{F(x_0 + th) - F(x_0)}{t} = \frac{\mathrm{d}}{\mathrm{d}t} F(x_0 + th)\bigg|_{t=0}. \tag{1.20}$$

In some cases there exists a $L_G \in \mathcal{L}(X, Y)$ such that for all $\|h\| = 1$ $L'_G(x_0, h) = L_G$.

Theorem 1.18. *The following properties and relationships hold for the Fréchet and Gâteaux derivatives.*

1. *The Fréchet and Gâteaux derivatives are uniquely determined.*

2. *If F is Fréchet differentiable in $x_0 \in \mathcal{U}(x_0) \subset \mathcal{D}$ it is continuous in x_0, Gâteaux differentiable, and $F'_G(x_0, h) = L_F h$.*

3. *Let F be Gâteaux differentiable in x_0 and $L_G = L_G(x_0) = F'_G(x_0, h) \in \mathcal{L}(X, Y)$ exist such that*

$$F(x_0 + th) - F(x_0) = tF'_G(x_0)h + o_h(t) \quad \textit{and} \quad o_h(t) = o(t)C(h) \textit{ with } \|C(h)\| \leq C$$

for $\|h\| = 1$. Then, F is Fréchet differentiable in x_0 and for all h with $\|h\| = 1$ we have that $F'_G(x_0, h) = F'_G(x_0)h = L_G h = L_F h$; hence, $L_F = F'_G(x_0)$.

4. *Let F be Gâteaux differentiable in a neighbourhood $\mathcal{U}(x_0)$ of x_0 and*

$$F(x + th) - F(x) = tF'_G(x_0)h + o_h(t)$$

such that $L_G(x) = F'_G(x) \in \mathcal{L}(X, Y)$ is continuous with respect to x for all $x \in \mathcal{U}(x_0)$. Then, F is Fréchet differentiable in $\mathcal{U}(x_0)$ and $L_F(x) = L_G(x)$.

Theorem 1.19 (Mean value theorem). *Let $F : \mathcal{D} \subset X \to Y$ and $a, b \in \mathcal{D}$ be given with $\overline{ab} = \{a + t(b - a) : 0 \le t \le 1\} \subset \mathcal{D}$ such that F is Fréchet differentiable on $\overline{ab}$ and*

$$\|F'(a + t(b - a))\| \le g'(t), \quad 0 \le t \le 1,$$

where g' and $f'(t) = F'(a+t(b-a))(b-a)$ are integrable on $[0, 1]$. Then, the following relationships hold:

$$\begin{aligned} F(b) - F(a) &= \int_0^1 F'(a + t(b - a))\,\mathrm{d}t(b - a), && (1.21)\\ \|F(b) - F(a)\| &\le (g(1) - g(0))\|b - a\|, && (1.22)\\ \|F(b) - F(a)\| &\le \|b - a\| \sup_{x\in\overline{ab}} \|F'(x)\|, && (1.23)\\ \|F(b) - F(a) - F'(x_0)(b - a)\| &\le \|b - a\| \sup_{x\in\overline{ab}} \|F'(x) - F'(x_0)\|. && (1.24) \end{aligned}$$

Remark. Modification to Gâteaux derivatives in the direction $b - a$ follow trivially.

For Cartesian product spaces $X = \prod_{i=1}^n X_i$ and $Y = \prod_{j=1}^m Y_j$, $n, m \in \mathbb{N}$ of Banach spaces X_i, $i = 1, \dots, n$ and Y_j, $j = 1, \dots, m$, respectively, we can define partial Fréchet and Gâteaux derivatives by splitting the variables into components. Let

$$\boldsymbol{F} \coloneqq (F_1(\boldsymbol{x}), \dots, F_m(\boldsymbol{x})) : \mathcal{D} \subset X \to Y,$$

with $\boldsymbol{x} = (x_i, \dots, x_n)$. For Banach spaces X_i, $i = 1, \dots, n$ and Y_j, $j = 1, \dots, m$, and norms $\|\boldsymbol{x}\|_X = \max_{i=1}^n \{\|x_i\|_{X_i}\}$ and $\|\boldsymbol{y}\|_Y = \max_{i=1}^n \{\|y_i\|_{Y_i}\}$ the respective spaces X and Y are also Banach spaces.

Definition 1.20. Choose $\boldsymbol{x}_0 \in \mathcal{U}(\boldsymbol{x}_0) \subset \mathcal{D}$, where $\mathcal{U}(\boldsymbol{x}_0)$ is a neighbourhood of $\boldsymbol{x}_0$. Assume there exists an $L_i \in \mathcal{L}(X_i, Y)$ and $L_i^{(j)} \in \mathcal{L}(X_i, Y_j)$ such that for all small $h_i \in X_i$

$$\begin{aligned} \boldsymbol{F}(\boldsymbol{x}_0 + h_i) - \boldsymbol{F}(\boldsymbol{x}_0) - L_i h_i &= o(h_i),\\ F_j(\boldsymbol{x}_0 + h_i) - F_j(\boldsymbol{x}_0) - L_i^{(j)} h_i &= o(h_i). \end{aligned} \tag{1.25}$$

Then, $\boldsymbol{F}$ and its components F_j are called *partially (Fréchet) differentiable* in $\boldsymbol{x}_0$ with respect to x_i, and $\partial \boldsymbol{F}/\partial x_i(\boldsymbol{x}_0) \coloneqq \partial_{x_i}\boldsymbol{F}(\boldsymbol{x}_0) \coloneqq L_i$ and $L_i^{(j)}$ are called its *partial derivatives*.

Remark. By replacing h_i with th_i and $t \to 0$ we obtain *partial Gâteaux differentiability*.

Remark. For $X_i = \mathbb{R}$, $i = 1, \dots, n$, $Y_j = \mathbb{R}$, $j = 1, \dots, m$, and $h_i = te_i$, where e_i is the i-th unit vector and small $t \in \mathbb{R}$, we recover classical partial derivatives.

CHAPTER 2

Monotone Differential Equations

We want to study the existence of *weak* solutions to nonlinear partial differential equations. We focus on differential equations where the coefficients can be defined as of *monotone* type.

2.1 Monotone operators

We first need to define what is meant by a monotone, and continuous, operator. In general, we consider an operator defined on a reflexive Banach space (or Hilbert space).

Definition 2.1. Let X be a Banach space and $A : X \to X'$ be a (nonlinear) operator; then, we call A

monotone if for all $u, v \in X$

$$\langle Au - Av, u - v \rangle \geq 0,$$

strictly monotone if for all $u, v \in X$, where $u \neq v$,

$$\langle Au - Av, u - v \rangle > 0,$$

uniformly monotone if there exists a continuous and strictly increasing function $a : \mathbb{R}_+ \to \mathbb{R}_+$ with $a(0) = 0$ and $\lim_{t \to +\infty} a(t) = +\infty$ such that for all $u, v \in X$

$$\langle Au - Av, u - v \rangle \geq a(\|u - v\|)\|u - v\|,$$

strongly monotone if there exists a positive constant M such that for all $u, v \in X$

$$\langle Au - Av, u - v \rangle \geq M\|u - v\|^2,$$

(nonlinear coercive) if

$$\lim_{\|u\| \to +\infty} \frac{\langle Au, u \rangle}{\|u\|} = +\infty,$$

weakly coercive if

$$\lim_{\|u\| \to +\infty} \|Au\| = +\infty,$$

Lipschitz continuous if there exists a positive constant L such that for all $u, v \in X$

$$\|Au - Av\| \leq L\|u - v\|,$$

continuous if for a sequence $\{u_n\} \subset X$

$$u_n \to u \in X \quad \Longrightarrow \quad Au_n \to Au,$$

hemicontinuous if the function $t \mapsto \langle A(u+tv), w\rangle$ is continuous on the interval $[0,1]$ for all $u, v, w \in X$, or *equivalently*, if the existence of a sequence $\{t_n\} \subset \mathbb{R}$, $t_n \to 0$ implies that $A(u + t_n v) \rightharpoonup Au$ for all $u, v \in X$,

strongly continuous if for a sequence $\{u_n\} \subset X$

$$u_n \rightharpoonup u \in X \quad \Longrightarrow \quad Au_n \to Au,$$

weakly continuous if for a sequence $\{u_n\} \subset X$

$$u_n \rightharpoonup u \in X \quad \Longrightarrow \quad Au_n \rightharpoonup Au,$$

demicontinuous if for a sequence $\{u_n\} \subset X$

$$u_n \to u \in X \quad \Longrightarrow \quad Au_n \rightharpoonup Au,$$

bounded if A maps bounded sets to bounded sets,

stable if there exists a continuous and strictly increasing function $a : \mathbb{R}_+ \to \mathbb{R}_+$ with $a(0) = 0$ and $\lim_{t\to+\infty} a(t) = +\infty$ such that for all $u, v \in X$

$$\|Au - Av\| \geq a(\|u - v\|).$$

Lemma 2.2. *Let X be a Banach space, and $A : X \to X'$ be a nonlinear operator; then, the following hold:*

$$\begin{aligned}
A \text{ strongly monotone} &\implies A \text{ uniformly monotone}\\
A \text{ uniformly monotone} &\implies A \text{ strictly monotone}\\
A \text{ strictly monotone} &\implies A \text{ monotone}\\
A \text{ uniformly monotone} &\implies A \text{ (nonlinear) coercive}\\
A \text{ uniformly monotone} &\implies A \text{ stable}\\
A \text{ Lipschitz continuous} &\implies A \text{ continuous}\\
A \text{ strongly continuous} &\implies A \text{ continuous}\\
A \text{ strongly continuous} &\implies A \text{ weakly continuous}\\
A \text{ weakly continuous} &\implies A \text{ demicontinuous}\\
A \text{ continuous} &\implies A \text{ demicontinuous}\\
A \text{ demicontinuous} &\implies A \text{ hemicontinuous}
\end{aligned}$$

Exercise 2.1. Prove the statements in Lemma 2.2.

Lemma 2.3. *Let X be a reflexive Banach space and $A : X \to X'$ a* monotone *operator; then,*

$$\begin{aligned}
A \text{ hemicontinuous} &\iff A \text{ demicontinuous}\\
A \text{ linear} &\implies A \text{ continuous}
\end{aligned}$$

Consider a nonlinear form $a(\,\cdot\,,\,\cdot\,) : X \times X \to \mathbb{R}$, which is nonlinear in the first argument and linear in the second argument, such that there exists a function $C : X \to \mathbb{R}$ such that

$$|a(u,v)| \le C(u)\|v\| \qquad \text{for all } u, v \in X. \tag{2.1}$$

In general this form will be the *weak formulation* of a partial differential equation; cf. Section 2.3.2.

Proposition 2.4. *If* (2.1) *holds for a nonlinear form* $a(\,\cdot\,,\,\cdot\,)$ *there exists a (usually nonlinear) operator* $A : X \to X'$ *such that for* $u \in X$

$$\langle Au, v\rangle := a(u,v) \qquad \textit{for all } v \in X, \tag{2.2}$$

on a reflexive Banach space X*. Then, for any* $f \in X'$ *the following are equivalent:*

1. *determine* $u_0 \in X$ *such that* $Au_0 = f \in X'$*,*
2. *determine* $u_0 \in X$ *such that* $a(u_0, v) = \langle f, v\rangle$ *for all* $v \in X'$*,*
3. *determine* $u_0 \in X$ *such that* $\langle Au_0, v\rangle = \langle f, v\rangle$ *for all* $v \in X'$

This proposition can be useful, as we can show that finding the solution of a weak formulation of a partial differential equation $a(u,v) = (f,v)$ for all $v \in X$ is equivalent to finding the solution of an operator equation $Au = f$, where A has various monotone and continuity properties.

2.2 Strongly monotone & Lipschitz continuous

We first consider the existence, and uniqueness, of a solution of $Au = f$ for a strongly monotone and Lipschitz continuous operator $A : X \to X'$ on a Hilbert space X.

In order to prove the existence of this solution we need results from fixed point theory.

Definition 2.5. Let (X, D) be a metric space, with metric $d(\,\cdot\,,\,\cdot\,)$, and mapping $T : X \to X$; then $\overline{x} \in X$ is called a *fixed point* of T if $T(\overline{x}) = \overline{x}$.

Remark. We note that in general we will consider a Hilbert space, which is a metric space with metric $d(u,v) = \|u - v\|$.

Theorem 2.6 (Banach fixed point theorem/Contraction mapping theorem)**.** *Let* (X, d) *be a complete metric space,* $M \subset X$ *be a non-empty closed subset, and* $T : M \to M$ *be* strongly contractive*; i.e., there exists a constant* $k \in (0,1)$ *such that*

$$d(T(x), T(y)) \le k d(x,y) \qquad \textit{for all } x, y \in M.$$

Then,

a) T *has a* unique *fixed point* $\overline{x} \in M$*,*

b) *the fixed point iteration*
$$x_{m+1} = T(x_m), \qquad m \ge 0,$$
converges to $\overline{x}$*, as* $m \to \infty$*, for* any *starting value* $x_0 \in M$*,*

c) the following error bounds holds:

$$d(\overline{x}, x_m) \leq \frac{k^m}{1-k} d(x_0, x_1),$$
$$d(\overline{x}, x_m) \leq \frac{k}{1-k} d(x_m, x_{m-1}).$$

Proof. We prove in three steps.

1. Show convergence of iteration: By recursion

$$d(x_m, x_{m+1}) = d(T(x_{m-1}), T(x_m)) \leq kd(x_{m-1}, x_m) \leq k^m d(x_0, x_1).$$

Then, by the triangle inequality, for $m \geq 1$:

$$\begin{aligned} d(x_m, x_{m+n}) &\leq d(x_m, x_{m+1}) + d(x_{m+1}, d + x + 2) + \cdots + d(x_{n+m-1}, x_{n+m}) \\ &\leq k^m d(x_0, x_1) + k^{m+1} d(x_0, x_1) + \cdots + k^{m+n-1} d(x_0, x_1) \\ &= k^m d(x_0, x_1) \sum_{j=0}^{n-1} k^j \\ &\leq \frac{k^m}{1-k} d(x_0, x_1). \end{aligned}$$

Then, for $n, m \to \infty$, $n \leq m$, we have that $d(x_n, x_m) \to 0$; hence, the sequence $\{x_m\}$ is a Cauchy sequence with a limit $\overline{x} \in X$. As M is closed and $\{x_m\} \subset M$; then, $\overline{x} \in M$.

2. Show that $\overline{x}$ is a fixed point of T: As T is contractive then it is continuous; therefore,

$$\overline{x} = \lim_{m \to \infty} x_{m+1} = \lim_{m \to \infty} T(x_m) = T(\overline{x}).$$

3. Show the fixed point is unique: Suppose that $\overline{x}_1 \neq \overline{x}_2$ are two fixed points in M; then,

$$d(\overline{x}_1, \overline{x}_2) = d(T(\overline{x}_1), T(\overline{x}_2)) \leq kd(\overline{x}_1, \overline{x}_2),$$

but as $k < 1$ and $d(\overline{x}_1, \overline{x}_2) \neq 0$ we have a contradiction; therefore, the fixed point must be unique. □

We can now show the existence of a unique solution of the operator equation $Au = f$ for a strongly monotone and Lipschitz continuous operator A using this theorem.

Lemma 2.7. *Let X be a Hilbert space, $A : X \to X'$ be strongly monotone and Lipschitz continuous, $f \in X'$, and J_X be the Riesz-representation from Corollary 1.11 on X; then, there exists a positive constant ε such that the mapping $T : X \to X$ defined as*

$$T(u) = u - \varepsilon J_X^{-1}(Au - f)$$

is strongly contractive; i.e.,

$$\|T(x) - T(y)\| \leq k\|x - y\|$$

where $k^2 = 1 + \varepsilon^2 L^2 = 2\varepsilon M < 1$, with M and L being the constants from strong monotonicity and Lipschitz continuity, respectively.

Exercise 2.2. Prove Lemma 2.7.

Theorem 2.8 (Zarantonello). *Let X be a Hilbert space, and $A : X \to X'$ be a strongly monotone and Lipschitz continuous operator; then, for each $f \in X'$ the equation $Au = f \in X'$ has a unique solution, depending continuously on f. More precisely, for $Au_1 = f_1$ and $Au_2 = f_2$,*

$$\|u_1 - u_2\| \le M^{-1}\|f_1 - f_2\|,$$

where M is the constant from the strong monotonicity of A.

Proof. From Lemma 2.7 there exists a constant $\varepsilon > 0$ such that

$$T(u) = u - \varepsilon J_X^{-1}(Au - f)$$

is strongly contractive. Then, by Theorem 2.6,

$$\begin{aligned} T \text{ has a unique fixed point } u \quad &\iff \quad u = u - \varepsilon J_X^{-1}(Au - f) \\ &\iff \quad Au - f = 0 \\ &\iff \quad Au = f. \end{aligned}$$

Hence, $Au = f$ has a unique solution $u \in X$. Now, let $Au_1 = f_1$ and $Au_2 = f_2$; then, by the fact that A is strongly monotone and (1.13)

$$M\|u_1 - u_2\|^2 \le \langle Au_1 - Au_2, u_1 - u_2\rangle \le \|Au_1 - Au_2\|\|u_1 - u_2\| = \|f_1 - f_2\|\|u_1 - u_2\|.$$

Therefore,

$$\|u_1 - u_2\| \le M^{-1}\|f_1 - f_2\|,$$

which completes the proof. □

Corollary 2.9. *Let X be a Hilbert space, $A : X \to X'$ be a strongly monotone and Lipschitz continuous operator, and $f \in X'$; then, there exists a positive constant ε such that the sequence $\{u_m\} \subset X$ defined by*

$$u_{m+1} = u_m - \varepsilon J_X^{-1}(Au_m - f),$$

where J_X is the Riesz-representation from Corollary 1.11 on X, converges to the unique solution $u \in X$ of $Au = f$ from Theorem 2.8 for any starting value $x_0 \in X$. Additionally,

$$\|u - u_m\| \le \frac{k^m}{1-k}\|u_0 - u_1\|$$

where $k^2 = 1 + \varepsilon^2 L^2 - 2\varepsilon M < 1$, with M and L being the constants from strong monotonicity and Lipschitz continuity, respectively.

Proof. Follows directly from Lemma 2.7, Theorem 2.8, and Theorem 2.6. □

Exercise 2.3. Compute the optimal value of ε such that the iteration

$$u_{m+1} = u_m - \varepsilon J_X^{-1}(Au_m - f)$$

converges fastest to the *unique* solution of $Au = f$ and compute the contraction constant k. Note that from Corollary 2.9 and Theorem 2.6 the error is given by

$$\|u - u_m\| \le \frac{k^m}{1-k}\|x_0 - x_1\|;$$

hence, the fastest convergence rate is obtained when k is close to zero.

Corollary 2.10. *Let X be a Hilbert space, $A : X \to X'$ be a strongly monotone and Lipschitz continuous operator, and $f \in X'$; then, there exists a positive constant ε such that if*

$$\langle J_X u_{m+1}, v\rangle = \langle J_X u_m, v\rangle - \varepsilon(\langle A u_m, v\rangle - \langle f, v\rangle), \qquad \text{for all } v \in X,$$

the sequence $\{u_m\} \subset X$ converges to the unique solution $u \in X$ of $Au = f$ from Theorem 2.8 for any starting value $x_0 \in X$.

Proof. Follows directly from Theorem 2.8 and Corollary 2.9 and the equivalence of solutions from Proposition 2.4. □

Example 2.1 (Weak solution of strongly monotone & Lipschitz continuous quasilinear PDE). Consider the boundary value problem in a bounded Lipschitz domain $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$,

$$\begin{aligned} -\nabla \cdot (\mu(\boldsymbol{x}, |\nabla u|)\nabla u) &= f && \text{in } \Omega,\\ u &= 0 && \text{on } \partial\Omega, \end{aligned}$$

where $\mu \in C(\overline{\Omega} \times [0,\infty))$ and there exists positive constants $\alpha_1 \geq \alpha_2 > 0$ such that, for $t \geq s \geq 0$ and $\boldsymbol{x} \in \overline{\Omega}$

$$\alpha_2(t-s) \leq \mu(\boldsymbol{x}, t)t - \mu(\boldsymbol{x}, s)s \leq \alpha_1(t-s).$$

From Liu and Barrett (1994, Lemma 2.1) we can show that for all vector valued functions $\boldsymbol{\sigma}, \boldsymbol{\tau} : \Omega \to \mathbb{R}^n$ that

$$|\mu(\boldsymbol{x}, |\boldsymbol{\tau}|)\boldsymbol{\tau} - \mu(\boldsymbol{x}, |\boldsymbol{\sigma}|)\boldsymbol{\sigma}| \leq \alpha_1 |\boldsymbol{\tau} - \boldsymbol{\sigma}|, \tag{2.3}$$

$$\alpha_2 |\boldsymbol{\tau} - \boldsymbol{\sigma}|^2 \leq (\mu(\boldsymbol{x}, |\boldsymbol{\tau}|)\boldsymbol{\tau} - \mu(\boldsymbol{x}, |\boldsymbol{\sigma}|)\boldsymbol{\sigma}) \cdot (\boldsymbol{\tau} - \boldsymbol{\sigma}). \tag{2.4}$$

We can define the *weak formulation* of this partial differential equation by multiplying by a test function and integrating by parts: Find $u \in X = H_0^1(\Omega)$ such that

$$\langle Au, v\rangle := \underbrace{\int_\Omega \mu(\boldsymbol{x}, |\nabla u|)\nabla u \cdot \nabla v \,\mathrm{d}\boldsymbol{x}}_{a(u,v)} = \int_\Omega f v \,\mathrm{d}\boldsymbol{x} =: \langle F, v\rangle \qquad \text{for all } v \in H_0^1(\Omega). \tag{2.5}$$

We can show that A is strongly monotone, as by (2.4)

$$\begin{aligned} \langle Au - Av, u - v\rangle &= \int_\Omega (\mu(\boldsymbol{x}, |\nabla u|)\nabla u - \mu(\boldsymbol{x}, |\nabla v|)\nabla v) \cdot \nabla (u - v) \,\mathrm{d}\boldsymbol{x}\\ &\geq \int_\Omega \alpha_2 |\nabla (u-v)|^2 \,\mathrm{d}\boldsymbol{x}\\ &= \alpha_2 \|u - v\|_{1,2}^2 \qquad \text{for all } u, v \in H_0^1(\Omega), \end{aligned}$$

and Lipschitz continuous, as by (2.3)

$$\begin{aligned} |\langle Au - Av, w\rangle| &\leq \int_\Omega |\mu(\boldsymbol{x}, |\nabla u|)\nabla u - \mu(\boldsymbol{x}, |\nabla v|)\nabla v||\nabla (u - v)| \,\mathrm{d}\boldsymbol{x}\\ &\leq \int_\Omega \alpha_1 |\nabla (u - v)||\nabla w| \,\mathrm{d}\boldsymbol{x}\\ &\leq \alpha_1 \|u - v\|_{1,2} \|w\|_{1,2} \qquad \text{for all } u, v, w \in H_0^1(\Omega), \end{aligned}$$

and

$$\|Au - Av\|_{-1,2} = \sup_{w \in H_0^1(\Omega)} \frac{|\langle Au - Av, w\rangle|}{\|w\|_{1,2}} \le \sup_{w \in H_0^1(\Omega)} \frac{\alpha_1 \|u - v\|_{1,2} \|w\|_{1,2}}{\|w\|_{1,2}} = \alpha_1 \|u - v\|_{1,2}.$$

Then, by Theorem 2.8 the weak formulation (2.5) has a unique solution $u \in H_0^1(\Omega)$. We note that (2.1) follows similarly to the proof of Lipschitz continuity and, hence, Proposition 2.4 holds.

2.3 Quasilinear elliptic PDEs of order $2k$

We want to consider the existence of *weak* solutions to quasilinear elliptic partial differential in *divergence form* (1.12). To this end, we need the following result for the operator equation $Au = f$.

Theorem 2.11 (Minty-Browder). *Let $A : X \to X'$ be a monotone, coercive, and hemicontinuous operator on a real reflexive Banach space X. Then, for each $f \in X'$ the equation $Au = f$ has at least one solution (A is surjective) and the set of all solutions is bounded, convex, and closed. Additionally, if A is* strictly monotone *then the solution is* unique, *the inverse A^{-1} exists and*

$$\begin{aligned} A \text{ uniformly monotone} &\implies A^{-1} \text{ continuous} \\ A \text{ strongly monotone} &\implies A^{-1} \text{ Lipschitz continuous} \end{aligned}$$

Remark. This theorem is occasionally stated as requiring A demicontinuous instead of hemicontinuous; however, by Lemma 2.3 these are equivalent as A is monotone.

Proof. The proof of this theorem will be shown later; cf., Section 3.3 □

2.3.1 Nemyckii operator

In order to apply Theorem 2.11 to differential equations we require certain properties of the coefficient functions in the divergence form (1.12). To this end, we define the following operator and some required properties on that operator.

Definition 2.12. Let $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$ be non-empty and measurable, $f : \Omega \times \mathbb{R}^m \to \mathbb{R}$, $m \in \mathbb{N}$; then, for a function $\boldsymbol{u} : \Omega \to \mathbb{R}$, $\boldsymbol{u}(\boldsymbol{x}) = (u_1(\boldsymbol{x}), \ldots, u_m(\boldsymbol{x}))$ we define a *Nemyckii operator* $\mathcal{N}$ as

$$(\mathcal{N}\boldsymbol{u})(\boldsymbol{x}) = f(\boldsymbol{x}, u_1(\boldsymbol{x}), \ldots, u_m(\boldsymbol{x}));$$

i.e., replace all variables u_j in $f(\boldsymbol{x}, u_1, \ldots, u_m)$ by $u_j(\boldsymbol{x})$.

We need two conditions for the Nemyckii operators:

(A1) *Carathéodory condition*: Let $\Omega \in \mathbb{R}^n$ be non-empty and measurable, $f : \Omega \times \mathbb{R}^m \to \mathbb{R}$ be a given function, $m \in \mathbb{N}$, and the following hold:

$$\begin{aligned} &\boldsymbol{x} \mapsto f(\boldsymbol{x}, \boldsymbol{u}) \text{ measurable on } \Omega \text{ for all } \boldsymbol{u} \in \mathbb{R}^m, \\ &\boldsymbol{u} \mapsto f(\boldsymbol{x}, \boldsymbol{u}) \text{ continuous on } \mathbb{R}^m \text{ almost everywhere for } \boldsymbol{x} \in \Omega. \end{aligned}$$

(A2) *Growth condition*: For all $(\boldsymbol{x}, \boldsymbol{u}) \in \Omega \times \mathbb{R}^m$

$$|f(\boldsymbol{x}, \boldsymbol{u})| \le g(\boldsymbol{x}) + b \sum_{i=1}^{m} |u_i|^{p_i/q},$$

where $b \ge 0$, $1 \le p_i, q < \infty$, and $g \in L^q(\Omega)$ is non-negative almost everywhere.

Theorem 2.13. *Let $\Omega \subset \mathbb{R}^n$, $k \in \mathbb{N}$ be non-empty and measurable, and the assumptions (A1) and (A2) hold for $f : \Omega \times \mathbb{R}^m \to \mathbb{R}$, $m \in \mathbb{N}$; then, the related Nemyckii operator*

$$\mathcal{N} : \prod_{i=1}^{m} L^{p_i}(\Omega) \to L^q(\Omega)$$

is continuous and bounded such that

$$\|\mathcal{N}\boldsymbol{u}\|_{0,q} \leq C \left(\|g\|_{0,q} + \sum_{i=1}^{j} \|u_j\|_{0,p_i}^{p_i/q} \right) \qquad \text{for all } \boldsymbol{u} \in \prod_{i=1}^{m} L^{p_i}(\Omega),$$

where C is a positive constant.

Proof. We consider the case when $m = 1$ (and set $u_1 = u$, $p_1 = p$); the general case follows analogously.

1. Measurable: Since $u \in L^p(\Omega)$ $x \mapsto u(\boldsymbol{x})$ measurable on Ω and by (A1) $x \mapsto f(\boldsymbol{x}, u)$ measurable on Ω.

2. $\mathcal{N}$ bounded: Let $1 \leq p, q < \infty$; then, using the Minkowski inequality and (A2)

$$\begin{aligned}
\|\mathcal{N}u\|_{0,q} &= \left(\int_\Omega |f(\boldsymbol{x}, u)|^q \, \mathrm{d}\boldsymbol{x} \right)^{1/q} \\
&\leq \left(\int_\Omega |g(\boldsymbol{x}) + b|u|^{p/q}|^q \, \mathrm{d}\boldsymbol{x} \right)^{1/q} \\
&\leq C \left(\|g(\boldsymbol{x})\|_{0,q} + b \left(\int_\Omega \left(|u|^{p/q} \right)^q \, \mathrm{d}\boldsymbol{x} \right)^{1/q} \right) \\
&\leq C \left(\|g(\boldsymbol{x})\|_{0,q} + b\|u\|_{0,p}^{p/q} \right) < \infty;
\end{aligned}$$

hence, $\mathcal{N}$ is a bounded operator from $L^p(\Omega) \to L^q(\Omega)$.

3. $\mathcal{N}$ continuous: Suppose there exists a sequence $\{u_n\} \subset L^p(\Omega)$ such that $u_n \to u \in L^p(\Omega)$. Then, there exists a subsequence $\{u_{n_k}\}$ and $v \in L^p(\Omega)$ such that

$$\begin{aligned}
u_{n_k}(\boldsymbol{x}) &\to u(\boldsymbol{x}) && \text{almost everywhere for } x \in \Omega, \\
|u_{n_k}| &< v(\boldsymbol{x}) && \text{for all } n_k \text{ and almost everywhere for } x \in \Omega.
\end{aligned}$$

Then, by (A2) and the Minkowski inequality

$$\begin{aligned}
\|\mathcal{N}u_{n_k} - \mathcal{N}u\|_{0,q}^q &= \int_\Omega |f(\boldsymbol{x}, u_{n_k}(\boldsymbol{x})) - f(\boldsymbol{x}, u(\boldsymbol{x}))|^q \, \mathrm{d}\boldsymbol{x} \\
&\leq C \left(\int_\Omega |f(\boldsymbol{x}, u_{n_k}(\boldsymbol{x}))| + \int_\Omega |f(\boldsymbol{x}, u(\boldsymbol{x}))|^q \, \mathrm{d}\boldsymbol{x} \right) \\
&\leq C \int_\Omega \left(|g|^q + |v|^q + |u|^q \right) \, \mathrm{d}\boldsymbol{x}.
\end{aligned}$$

By (A1)

$$u \mapsto f(\boldsymbol{x}, u) \text{ continuous almost everywhere for } \boldsymbol{x} \in \Omega$$

$$\begin{aligned}
&\implies && f(\boldsymbol{x}, u_{n_k}) \to f(\boldsymbol{x}, u) \text{ as } u_{n_k} \to u \\
&\implies && f(\boldsymbol{x}, u_{n_k}) - f(\boldsymbol{x}, u) \to 0 \\
&\implies && \|\mathcal{N} u_{n_k} - \mathcal{N} u\|_{0,q} \to 0 \\
&\implies && \mathcal{N} u_{n_k} \to \mathcal{N} u \text{ in } L^q(\Omega).
\end{aligned}$$

As sequences are bounded and all subsequences converge, then, $\mathcal{N} u_n \to \mathcal{N} u$; hence, $\mathcal{N}$ is continuous. □

2.3.2 Weak formulation

We can now study the weak solution of a general quasilinear partial differential equation under certain assumptions on the coefficients, which are Nemyckii operators, using Minty-Browder.

We define for a quasilinear PDE of order $2k$ the *formal differential operator*

$$(\mathcal{A}u)(\boldsymbol{x}) = \sum_{|\alpha| \leq k} (-1)^{\alpha} \partial^{\alpha} a_{\alpha}(\boldsymbol{x}, \delta_k u(\boldsymbol{x})) \tag{2.6}$$

with coefficient functions $a_{\alpha} : \Omega \times \mathbb{R}^{\kappa} \to \mathbb{R}$, for each multi-index $\alpha \in \mathbb{N}_0^n$ with $|\alpha| \leq k$, where κ is defined by (1.7), and consider the boundary value problem

$$\mathcal{A}u = f \quad \text{in } \Omega, \tag{2.7}$$

$$\frac{\partial^i u}{\partial \boldsymbol{n}^i} = 0 \quad \text{on } \partial\Omega, i = 1, \ldots, k-1 \quad \iff \quad \partial^{\beta} u|_{\partial\Omega} = 0 \quad \text{for all } \beta, |\beta| \leq k-1. \tag{2.8}$$

where $\boldsymbol{n}$ is the outward unit normal vector to $\partial\Omega$, on a bounded domain $\Omega \subset \Omega^n$ with Lipschitz continuous boundary. By multiplying by a test function and repeated applications of the Green's formulae we get that for $u, v \in X = W_0^{k,p}(\Omega)$

$$\begin{aligned}
\int_{\Omega} v(\mathcal{A}u) \,\mathrm{d}\boldsymbol{x} &= \int_{\Omega} \sum_{|\alpha| \leq k} v(-1)^{\alpha} \partial^{\alpha} a_{\alpha}(\boldsymbol{x}, \delta_k u(\boldsymbol{x})) \,\mathrm{d}\boldsymbol{x} \\
&= \int_{\Omega} \sum_{|\alpha| \leq k} a_{\alpha}(\boldsymbol{x}, \delta_k u(\boldsymbol{x})) \partial^{\alpha} v \,\mathrm{d}\boldsymbol{x} =: a(u, v).
\end{aligned} \tag{2.9}$$

Then, we can define the *weak formulation* of the partial differential equation (2.7)–(2.8): find $u \in W_0^{k,p}(\Omega)$, $1 < p < \infty$, such that

$$\langle Au, v \rangle := a(u, v) = \int_{\Omega} f v \,\mathrm{d}\boldsymbol{x} =: \langle F, v \rangle \tag{2.10}$$

for all $v \in W_0^{k,p}(\Omega)$, where we assume $f \in L^q(\Omega)$, $1/p + 1/q = 1$. u is called the *weak solution*.

2.3.3 Existence of solution

Note that the above definition (2.10) of the operator $A : W_0^{k,p}(\Omega) \to W^{-k,q}(\Omega)$ assumes that (2.1) holds and, hence, Proposition 2.4 can be applied. Then, finding the weak solution $u \in W_0^{k,p}(\Omega)$ of (2.10) is equivalent to finding the solution of the operator equation $Au = F$, which can be shown by Minty-Browder (Theorem 2.11). Note, that for this to work we require that $F \in W^{-k,q}(\Omega) = X'$, which can be shown trivially:

$$|\langle F, v\rangle| \leq \int_\Omega |fv| \,\mathrm{d}\boldsymbol{x} \leq \|f\|_{0,q}\|v\|_{0,p} \leq \|f\|_{0,q}\|v\|_{k,p}$$

As $v \in W_0^{k,p}(\Omega)$ and $f \in L^q(\Omega)$; then,

$$\|F\|_{-k,q} = \sup_{v \in W_0^{k,p}(\Omega)} \frac{|\langle F, v\rangle|}{\|v\|_{k,p}} < \infty,$$

which implies that $F \in W^{-k,q}(\Omega)$.

Therefore, in order to prove the existence of a weak solution we need to show that (2.1) holds and prove the conditions of Minty-Browder. We note, trivially, that the space $X = W_0^{k,p}(\Omega)$, $1 < p < \infty$ is a reflexive Banach space. In order to prove the rest of the conditions we need to define additional assumptions on the coefficients a_α such that A is monotone, coercive, and hemicontinuous.

To this end, we first define Nemyckii operators for the coefficients a_α as

$$\mathcal{N}_\alpha : W_0^{k,p}(\Omega) \to L^q(\Omega), \qquad (\mathcal{N}_\alpha u)(\boldsymbol{x}) := a_\alpha(\boldsymbol{x}, \delta_k u(\boldsymbol{x})), \tag{2.11}$$

and define modified versions of the Carathéodory (A1) and growth conditions (A2) for $\mathcal{N}_\alpha$:

(B1) *Carathéodory condition* For all α, $|\alpha| \leq k$, $a_\alpha : \Omega \times \mathbb{R}^\kappa \to \mathbb{R}$ has the following properties:

$$\begin{aligned} &\boldsymbol{x} \mapsto a_\alpha(\boldsymbol{x}, \xi) \text{ measurable on } \Omega \text{ for all } \xi \in \mathbb{R}^\kappa, \\ &\xi \mapsto a_\alpha(\boldsymbol{x}, \xi) \text{ continuous on } \mathbb{R}^\kappa \text{ almost everywhere for } \boldsymbol{x} \in \Omega. \end{aligned}$$

(B2) *Growth condition* For $1 < p < \infty$, and all α, $|\alpha| \leq k$

$$|a_\alpha(\boldsymbol{x}, \xi)| \leq C \left(g(\boldsymbol{x}) + \sum_{|\beta| \leq k} |\xi_\beta|^{p-1} \right),$$

where $\beta \in \mathbb{N}_0^n$ is a multi-index, $g \in L^q(\Omega)$ is non-negative almost everywhere, $1/p + 1/q = 1$, $C > 0$ is a positive constant, and $\xi = (\xi_\beta)_{|\beta| \leq k} \in \mathbb{R}^\kappa$, $\xi_\beta \in \mathbb{R}$, $|\beta| \leq k$.

We are now are able to prove that (2.1) holds; hence, Proposition 2.4 holds and (2.10) is valid.

Lemma 2.14. *Let (B1) and (B2) hold for all $\mathcal{N}_\alpha$, $|\alpha| \leq k$, defined by (2.11); then, $\mathcal{N}_\alpha$ is well-defined, bounded, and continuous. Additionally,*

$$|a(u, v)| \leq C \left(\|g\|_{0,q} + \|u\|_{k,p}^{p/q} \right) \|v\|_{k,p} \qquad \forall u, v \in W_0^{k,p}(\Omega),$$

where $g \in L^q(\Omega)$ is non-negative almost everywhere, $1/p + 1/q = 1$, and $C > 0$ is a positive constant, for the nonlinear form $a(\,\cdot\,,\,\cdot\,)$ from (2.9); hence, (2.1) holds.

We can define statements which prove that the operator A is monotone, coercive, and hemicontinuous.

Lemma 2.15. *Let (B1) and (B2) hold for all $\mathcal{N}_\alpha$, $|\alpha| \leq k$, defined by (2.11); then, the operator $A : W_0^{k,p}(\Omega) \to W^{-k,q}(\Omega)$, defined in the weak formulation (2.10), is bounded and demicontinuous.*

Lemma 2.16. *If for the coefficients a_α, $|\alpha| \leq k$, defined in (2.6),*

$$\sum_{|\alpha| \leq k} (a_\alpha(\boldsymbol{x}, \xi) - a_\alpha(\boldsymbol{x}, \eta)) (\xi_\alpha - \eta_\alpha) \geq 0,$$

for all $\xi, \eta \in \mathbb{R}^\kappa$ and almost everywhere for $\boldsymbol{x} \in \Omega$; then, the operator $A : W_0^{k,p}(\Omega) \to W^{-k,q}(\Omega)$, defined in the weak formulation (2.10), monotone. Additionally, if equality of the above condition only holds for

$$\sum_{|\beta| \leq k} |\xi_\beta - \eta_\beta| = 0;$$

then, A is strictly monotone.

Corollary 2.17. *Let (B1) and (B2) hold for all $\mathcal{N}_\alpha$, $|\alpha| \leq k$, defined by (2.11) and the condition from Lemma 2.16 hold. Then, the operator $A : W_0^{k,p}(\Omega) \to W^{-k,q}(\Omega)$, defined in the weak formulation (2.10), is hemicontinuous.*

Proof. By Lemma 2.15 and Lemma 2.16 the operator A is demicontinuous and monotone. Then, by Lemma 2.3, A is hemicontinuous. □

Lemma 2.18. *If there exists a positive constant $C > 0$ and a function $k \in L^1(\Omega)$ such that for the coefficients a_α, $|\alpha| \leq k$, defined in (2.6),*

$$\sum_{|\alpha| \leq k} a_\alpha(\boldsymbol{x}, \xi)\xi_\alpha \geq C \sum_{|\alpha| = k} |\xi_\alpha|^p - h(\boldsymbol{x}),$$

for all $\xi \in \mathbb{R}^\kappa$ and almost everywhere for $\boldsymbol{x} \in \Omega$, where $1 < p < \infty$. Then, the operator $A : W_0^{k,p}(\Omega) \to W^{-k,q}(\Omega)$, defined in the weak formulation (2.10), is (nonlinear) coercive.

Exercise 2.4. Prove Lemmas 2.14–2.18.

We can now combine these results.

Theorem 2.19. *Let the coefficients functions $a_\alpha : \Omega \times \mathbb{R}^\kappa \to \mathbb{R}$, $|\alpha| \leq k$, $k \in \mathbb{N}$, where κ defined by (1.7), from (2.6) satisfy the following conditions for $1 < p < \infty$:*

(B1) Carathéodory condition*: For all α, $|\alpha| \leq k$,*

$$\begin{aligned} &\boldsymbol{x} \mapsto a_\alpha(\boldsymbol{x}, \xi) \text{ measurable on } \Omega \text{ for all } \xi \in \mathbb{R}^\kappa, \\ &\xi \mapsto a_\alpha(\boldsymbol{x}, \xi) \text{ continuous on } \mathbb{R}^\kappa \text{ almost everywhere for } \boldsymbol{x} \in \Omega. \end{aligned}$$

(B2) Growth condition*: For all α, $|\alpha| \leq k$,*

$$|a_\alpha(\boldsymbol{x}, \xi)| \leq C \left(g(\boldsymbol{x}) + \sum_{|\beta| \leq k} |\xi_\beta|^{p-1} \right),$$

where $\beta \in \mathbb{N}_0^n$ is a multi-index, $g \in L^q(\Omega)$ is non-negative almost everywhere, $1/p + 1/q = 1$, $C > 0$ is a positive constant, and $\xi = (\xi_\beta)_{|\beta| \leq k} \in \mathbb{R}^\kappa$, $\xi_\beta \in \mathbb{R}$, $|\beta| \leq k$.

(C1) Monotonicity: *For all* $\xi, \eta \in \mathbb{R}^\kappa$

$$\sum_{|\alpha|\leq k} \left(a_\alpha(\boldsymbol{x}, \xi) - a_\alpha(\boldsymbol{x}, \eta)\right)(\xi_\alpha - \eta_\alpha) \geq 0,$$

almost everywhere for $\boldsymbol{x} \in \Omega$.

(C2) (Nonlinear) coercivity: *There exists a positive constant* $C > 0$ *and a function* $h \in L^1(\Omega)$ *such that*

$$\sum_{|\alpha|\leq k} a_\alpha(\boldsymbol{x}, \xi)\xi_\alpha \geq C \sum_{|\alpha|=k} |\xi_\alpha|^p - h(\boldsymbol{x}),$$

for all $\xi \in \mathbb{R}^\kappa$ *and almost everywhere for* $\boldsymbol{x} \in \Omega$.

Then, for any $f \in L^q(\Omega)$, $1/p + 1/q = 1$, *there exists at least one weak solution* $u \in W_0^{k,p}(\Omega)$ *to the weak formulation* (2.10). *Additionally, let* $a_\alpha : \Omega \times \mathbb{R}^\kappa \to \mathbb{R}$ *satisfy the following condition:*

(D1) Strict Monotonicity: *For all* $\xi, \eta \in \mathbb{R}^\kappa$, *where*

$$\sum_{|\beta|\leq k} |\xi_\beta - \eta_\beta| > 0,$$

it holds that

$$\sum_{|\alpha|\leq k} \left(a_\alpha(\boldsymbol{x}, \xi) - a_\alpha(\boldsymbol{x}, \eta)\right)(\xi_\alpha - \eta_\alpha) > 0,$$

almost everywhere for $\boldsymbol{x} \in \Omega$.

Then, the solution $u \in W_0^{k,p}(\Omega)$ *is unique.*

Proof. From Lemmas Lemma 2.15, Lemma 2.16, and Lemma 2.18 we can show the conditions of Theorem 2.11 are met; therefore, we can show that the equation $Au = F$, for $A : W_0^{k,p}(\Omega) \to W^{-k,q}(\Omega)$ and $F \in W^{-k,q}(\Omega)$ defined by (2.10), has a solution $u \in W_0^{k,p}(\Omega)$. Then, by Lemma 2.14 and Proposition 2.4 this is the weak solution of (2.10). Additionally, if (D1) holds the operator A is strictly monotone due to Lemma 2.16; therefore, by Theorem 2.11 the solution in unique. □

Exercise 2.5. Consider the following boundary value problem problem:

$$\begin{aligned} -\nabla \cdot (\mu(\boldsymbol{x}, \nabla u)\nabla u) + b(\boldsymbol{x}, u) &= f(\boldsymbol{x}) && \text{in } \Omega \subset \mathbb{R}^2 \\ u &= 0 && \text{on } \partial\Omega \end{aligned}$$

where Ω has Lipschitz boundary, $u : \Omega \to \mathbb{R}$ is the unknown function, and $f \in L^2(\Omega)$. Define, for this problem, the coefficient functions $a_\alpha(\boldsymbol{x}, \xi)$, $\xi \in \mathbb{R}^\kappa$, for all multi-indices α, where $|\alpha| \leq 1$, and the weak formulation of the boundary value problem. Additionally, derive conditions for μ and b such that Theorem 2.19 holds; i.e., state conditions such that

1. A is monotone,
2. A is strictly monotone,
3. A is coercive, and

4. a_α satisfies the growth condition (B2).

Exercise 2.6. Define the weak formulation, and use Theorem 2.19 to show that there exists a weak solution, for the following boundary value problem in the bounded Lipschitz domain $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$: For $2 \le p < \infty$, $f \in L^q(\Omega)$, $1/p + 1/q = 1$, and $t \ge 0$, $t \in \mathbb{R}$

$$\begin{aligned} -\sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left(\left| \frac{\partial u}{\partial x_i} \right|^{p-2} \frac{\partial u}{\partial x_i} \right) + tu &= f, && \text{in } \Omega, \\ u &= 0, && \text{on } \partial\Omega. \end{aligned}$$

Additionally, state if the weak solution is unique.

We define two more results on the operator A defined by (2.10) which will be used in Section 2.5.

Lemma 2.20. *Let the coefficients functions $a_\alpha : \Omega \times \mathbb{R}^\kappa \to \mathbb{R}$, $|\alpha| \le k$, $k \in \mathbb{N}$, where κ defined by (1.7), from (2.6) satisfy (B1) and (B2) hold, and additionally only depend on derivatives of up to order $k-1$; i.e., $a_\alpha(\boldsymbol{x}, \delta_k u)$ can be defined instead as $a_\alpha(\boldsymbol{x}, \delta_{k-1} u)$. Then, there exists a unique $A : X \to X'$ such that*

$$\langle Au, v \rangle = a(u, v)$$

which is strongly continuous.

Corollary 2.21. *Lemma 2.20 is valid if the condition (B2) is replaced by the following condition. For all α, $|\alpha| \le k$,*

$$|a_\alpha(\boldsymbol{x}, \xi)| \le C \left(g(\boldsymbol{x}) + \sum_{|\beta| \le k} |\xi_\beta|^{p_\alpha / q_\alpha} \right),$$

where $\beta \in \mathbb{N}_0^n$ is a multi-index, $g \in L^{q_\alpha}(\Omega)$ is non-negative almost everywhere, $C > 0$ is a positive constant, and $1 < p_\alpha, q_\alpha < \infty$ selected such that $1/p_\alpha + 1/q_\alpha = 1$ and

$$\frac{1}{p} - \frac{k - |\alpha|}{n} < \frac{1}{p_\alpha}.$$

2.4 Quasilinear systems

We briefly consider the generalisation of the above case to *systems* of quasilinear partial differential equations. To this end, we consider the Banach space

$$X = [W^{k,p}(\Omega)]^m = W^{k,p}(\Omega, \mathbb{R}^m),$$

which is reflexive for $1 < p < \infty$, where $m \in \mathbb{N}$, $m \ge 2$, is the number of equations and unknowns in the nonlinear system of partial differential equations. We define a function $\boldsymbol{u} \in X$ in the space as $\boldsymbol{u} = (u_1, \ldots, u_n)$, $u_i \in W^{k,p}(\Omega)$, $1 \le i \le m$. We furthermore extend the definition of the Sobolev space to $[W^{k,p}(\Omega)]^m$ via the norm and seminorm

$$\|\boldsymbol{u}\|_{k,p,\Omega} := \left(\sum_{i=1}^{m} \sum_{|\alpha| \le k} \|\partial^\alpha u_i\|_{0,p,\Omega}^p \right)^{1/2} \tag{2.12}$$

$$|\boldsymbol{u}|_{k,p,\Omega} := \left(\sum_{i=1}^{m} \sum_{|\alpha|=k} \|\partial^\alpha u_i\|_{0,p,\Omega}^p \right)^{1/2}. \tag{2.13}$$

We furthermore extend the notation for partial derivatives (1.8)–(1.9) to $\boldsymbol{u} : \Omega \to \mathbb{R}^m$ as

$$\partial^\alpha \boldsymbol{u} = (\partial^\alpha u_i)_{i=1}^m = (\partial^\alpha u_1, \dots, \partial^\alpha u_m),$$

$$\delta_k \boldsymbol{u} = (\partial^\alpha \boldsymbol{u})_{|\alpha| \le k}^m = (\delta_k u_1 \cdots \delta_k u_m) = \begin{pmatrix} u_1 & \dots & u_m \\ \dfrac{\partial u_1}{\partial x_1} & \dots & \dfrac{\partial u_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^k u_1}{\partial x_n^k} & \dots & \dfrac{\partial^k u_m}{\partial x_n^k} \end{pmatrix}$$

where $\partial^\alpha \boldsymbol{u} : \Omega \to \mathbb{R}^m$ and $\delta_k \boldsymbol{u} : \Omega \to \mathbb{R}^{\kappa \times m}$. Note that a row of $\delta_k \boldsymbol{u}$ contains $\partial^\alpha u_i$, $i = 1, \dots, m$, for a fixed multi-index α.

Remark. Note that (2.12)–(2.13) is only one choice of possible norm for this space. In general, the norms on this space are defined by defining the standard L^p-norm, cf. (1.14)–(1.15), using any valid norm on $\mathbb{R}^m$ instead of the absolute value $|\cdot|$ of the function and then define the norm and seminorm by (1.16) and (1.17), respectively. For example, for $1 \le p < \infty$, (2.12)–(2.13) are obtained by

$$\|v\|_{0,p,\mathcal{D}} := \left(\int_{\mathcal{D}} |v(\boldsymbol{x})|_p^p \,\mathrm{d}\boldsymbol{x} \right)^{1/p},$$

where $|\cdot|_p$ is the vector p-norm on $\mathbb{R}^m$. Due to equivalence of norms on $\mathbb{R}^m$ we get equivalence of norms in $[W^{k,p}(\Omega)]^m$.

We can now define a general systems of quasilinear elliptic partial differential equations on a bounded Lipschitz domain $\Omega \subset \mathbb{R}^m$, $n \in \mathbb{N}$, in divergence form as

$$\left[\sum_{|\alpha| \le k} (-1)^{|\alpha|} \partial^\alpha a_\alpha^i(\boldsymbol{x}, \delta_k \boldsymbol{u}) \right]_{i=1}^m = \boldsymbol{f}, \qquad \text{in } \Omega, \tag{2.14}$$

$$\boldsymbol{u} = \boldsymbol{0}, \text{on } \partial\Omega, \tag{2.15}$$

where $\boldsymbol{f} \in [L^q(\Omega)]^m$ and $\boldsymbol{a}_\alpha(\boldsymbol{x}, \xi) = (a_\alpha^i(\boldsymbol{x}, \xi))_{i=1}^m : \Omega \times \mathbb{R}^{\kappa \times m} \to \mathbb{R}^m$.

We look for the *weak solution* $\boldsymbol{u} \in [W_0^{k,p}(\Omega)]^m$ such that

$$a(\boldsymbol{u}, \boldsymbol{v}) := \int_\Omega \sum_{|\alpha| \le k} (\boldsymbol{a}_\alpha(x, \delta_k \boldsymbol{u}), \partial^\alpha \boldsymbol{v}) \,\mathrm{d}\boldsymbol{x} = \int_\Omega (\boldsymbol{f}, \boldsymbol{v}) \,\mathrm{d}\boldsymbol{x} \tag{2.16}$$

for all $\boldsymbol{v} \in [W_0^{k,p}(\Omega)]^m$, where $(\,\cdot\,,\,\cdot\,)$ is the Euclidean inner product on $\mathbb{R}^m$. Again, we want the coefficient functions $\boldsymbol{a}_\alpha$ to be Nemyckii operators satisfying Carathéodory and growth conditions; therefore, we define generalisations of the Nemyckii operator and these conditions.

Definition 2.22. Let $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$ be non-empty and measurable, $\boldsymbol{a}_\alpha : \Omega \times \mathbb{R}^{\kappa \times m} \to \mathbb{R}^m$, $m \in \mathbb{N}$; then, we define the *Nemyckii operator* $\mathcal{N}_\alpha : [W_0^{k,p}(\Omega)]^m \to [L^q(\Omega)]^m$ as

$$(\boldsymbol{\mathcal{N}}_\alpha \boldsymbol{u})(\boldsymbol{x}) = \boldsymbol{a}_\alpha(\boldsymbol{x}, \delta_k \boldsymbol{u}(\boldsymbol{x})).$$

Then, we define the following conditions.

(E1) *Carathéodory condition*: For all α, $|\alpha| \leq k$, $\boldsymbol{a}_\alpha : \Omega \times \mathbb{R}^{\kappa \times m} \to \mathbb{R}^m$ has the following properties:

$$\begin{aligned} &\boldsymbol{x} \mapsto a_\alpha(\boldsymbol{x}, \xi) \text{ measurable on } \Omega \text{ for all } \xi \in \mathbb{R}^{\kappa \times m}, \\ &\xi \mapsto a_\alpha(\boldsymbol{x}, \xi) \text{ continuous on } \mathbb{R}^{\kappa \times m} \text{ almost everywhere for } \boldsymbol{x} \in \Omega. \end{aligned}$$

(E2) *Growth condition*: For all α, $|\alpha| \leq k$,

$$|\boldsymbol{a}_\alpha(\boldsymbol{x}, \xi)| \leq C \left(g(\boldsymbol{x}) + \sum_{|\beta| \leq k} |\xi_\beta|^{p-1} \right),$$

where $\beta \in \mathbb{N}_0^n$ is a multi-index, $g \in L^q(\Omega)$ is non-negative almost everywhere, $1/p + 1/q = 1$, $C > 0$ is a positive constant, and $\xi = (\xi_\beta)_{|\beta| \leq k} \in \mathbb{R}^{\kappa \times m}$, $\xi_\beta \in \mathbb{R}^m$, $|\beta| \leq k$ (i.e., ξ_β is a row of the matrix ξ).

Theorem 2.23. *Let $\Omega \subset \mathbb{R}^n$, $k \in \mathbb{N}$ be non-empty and measurable, and the assumptions (E1) and (E2) hold for $|\alpha| \leq k$, $\boldsymbol{a}_\alpha : \Omega \times \mathbb{R}^{\kappa \times m} \to \mathbb{R}^m$; then,*

$$\boldsymbol{\mathcal{N}}_\alpha : [W^{k,p}(\Omega)]^m \to [L^q(\Omega)]^m$$

is continuous and bounded such that

$$\|\boldsymbol{\mathcal{N}}_\alpha \boldsymbol{u}\|_{0,q} \leq C \left(\|g\|_{0,q} + \sum_{|\alpha| \leq k} \|\partial^\alpha \boldsymbol{u}\|_{0,p_i}^{p-1} \right) \qquad \text{for all } [W^{k,p}(\Omega)]^m,$$

where C is a positive constant.

We can now define a result for the existence of the weak solution (2.16).

Theorem 2.24. *Let the coefficients functions $\boldsymbol{a}_\alpha : \Omega \times \mathbb{R}^{\kappa \times m} \to \mathbb{R}^m$, $|\alpha| \leq k$, $k \in \mathbb{N}$, where κ defined by (1.7), from (2.14) satisfy the following conditions for $1 < p < \infty$:*

(E1) Carathéodory condition*: For all α, $|\alpha| \leq k$,*

$$\begin{aligned} &\boldsymbol{x} \mapsto a_\alpha(\boldsymbol{x}, \xi) \textit{ measurable on } \Omega \textit{ for all } \xi \in \mathbb{R}^{\kappa \times m}, \\ &\xi \mapsto a_\alpha(\boldsymbol{x}, \xi) \textit{ continuous on } \mathbb{R}^{\kappa \times m} \textit{ almost everywhere for } \boldsymbol{x} \in \Omega. \end{aligned}$$

(E2) Growth condition*: For all α, $|\alpha| \leq k$,*

$$|\boldsymbol{a}_\alpha(\boldsymbol{x}, \xi)| \leq C \left(g(\boldsymbol{x}) + \sum_{|\beta| \leq k} |\xi_\beta|^{p-1} \right),$$

where $\beta \in \mathbb{N}_0^n$ is a multi-index, $g \in L^q(\Omega)$ is non-negative almost everywhere, $1/p + 1/q = 1$, $C > 0$ is a positive constant, and $\xi = (\xi_\beta)_{|\beta| \leq k} \in \mathbb{R}^{\kappa \times m}$, $\xi_\beta \in \mathbb{R}^m$, $|\beta| \leq k$ (i.e., ξ_β is a row of the matrix ξ).

(F1) Monotonicity: *For all matrices $\xi, \eta \in \mathbb{R}^{\kappa \times m}$, with rows $\xi_\alpha, \eta_\alpha \in \mathbb{R}^m$, $\alpha \leq k$,*

$$\sum_{|\alpha| \leq k} (\boldsymbol{a}_\alpha(\boldsymbol{x}, \xi) - \boldsymbol{a}_\alpha(\boldsymbol{x}, \eta), \xi_\alpha - \eta_\alpha) \geq 0,$$

almost everywhere for $\boldsymbol{x} \in \Omega$.

(F2) (Nonlinear) coercivity: *There exists a positive constant $C > 0$ and a function $h \in L^1(\Omega)$ such that*

$$\sum_{|\alpha| \leq k} (\boldsymbol{a}_\alpha(\boldsymbol{x}, \xi), \xi_\alpha) \geq C \sum_{|\alpha| = k} |\xi_\alpha|^p - h(\boldsymbol{x}),$$

for all matrices $\xi \in \mathbb{R}^{\kappa \times m}$, with rows $\xi_\alpha \in \mathbb{R}^m$, $\alpha \leq k$, and almost everywhere for $\boldsymbol{x} \in \Omega$.

Then, for any $f \in [L^q(\Omega)]^m$, $1/p + 1/q = 1$, there exists exactly one bounded operator $A : [W_0^{k,p}(\Omega)]^m \to [W^{-k,q}(\Omega)]^m$ such that

$$a(\boldsymbol{u}, \boldsymbol{v}) := \langle A\boldsymbol{u}, \boldsymbol{v} \rangle \qquad \text{for all } \boldsymbol{u}, \boldsymbol{v} \in [W_0^{k,p}(\Omega)]^m$$

and

$$\begin{aligned} &\text{find } \boldsymbol{u} \in [W_0^{k,p}(\Omega)]^m : a(\boldsymbol{u}, \boldsymbol{v}) = \int_\Omega (\boldsymbol{f}, \boldsymbol{v}) =: \langle F, \boldsymbol{v} \rangle \quad \forall \boldsymbol{v} \in [W_0^{k,p}(\Omega)]^m \\ \iff \quad &\text{find } \boldsymbol{u} \in [W_0^{k,p}(\Omega)]^m : A\boldsymbol{u} = F. \end{aligned} \tag{2.17}$$

The operator A is monotone, coercive, and hemicontinuous, and the (2.17) has a solution $\boldsymbol{u} \in [W_0^{k,p}(\Omega)]^m$. Additionally, let $\boldsymbol{a}_\alpha : \Omega \times \mathbb{R}^{\kappa \times m} \to \mathbb{R}^m$ satisfy the following condition:

(G1) Strict Monotonicity: *For all $\xi, \eta \in \mathbb{R}^{\kappa \times m}$, where*

$$\sum_{|\beta| \leq k} |\xi_\beta - \eta_\beta| > 0,$$

it holds that

$$\sum_{|\alpha| \leq k} (\boldsymbol{a}_\alpha(\boldsymbol{x}, \xi) - \boldsymbol{a}_\alpha(\boldsymbol{x}, \eta), \xi_\alpha - \eta_\alpha) > 0,$$

almost everywhere for $\boldsymbol{x} \in \Omega$.

Then, the solution $\boldsymbol{u} \in [W_0^{k,p}(\Omega)]^m$ is unique.

2.5 Pseudomonotone operators

We now aim to consider more general quasilinear partial differential equations which contains *lower order* terms which do not satisfy the monotonicity conditions. In general, we consider quasilinear PDEs which can be written as an operator equation of the form

$$A_1 u + A_2 u = f, \qquad u \in X \tag{2.18}$$

where $A_1 + A_2$ is *pseudomonotone*, cf. Definition 2.28, and coercive on a real reflexive Banach space. In general, we will have that

1. $A_1 : X \to X'$ is monotone and hemicontinuous (corresponding to the high order terms of the PDE), and
2. $A_2 : X \to X'$ is strongly continuous (corresponding to the lower order terms of the PDE).

In order to define *pseudomonotone* operators, we first need to define five related conditions for an operator A.

Definition 2.25. Let X be a real reflexive Banach space and $\{u_n\} \subset X$ be a sequence; then, we say that the nonlinear operator $A : X \to X'$ satisfies the conditions (M), $(\mathrm{S})_+$, (S), $(\mathrm{S})_0$, and $(\mathrm{S})_1$ if and only if the following hold.

$$\begin{array}{lrcl}
(\mathrm{M}) & u_n \rightharpoonup u, \quad Au_n \rightharpoonup b, \quad \limsup_{n\to\infty}\langle Au_n, u_n\rangle \le \langle b, u\rangle & \implies & Au = b \\
(\mathrm{S})_+ & u_n \rightharpoonup u, \quad \limsup_{n\to\infty}\langle Au_n - Au, u_n - u\rangle \le 0 & \implies & u_n \to u \\
(\mathrm{S}) & u_n \rightharpoonup u, \quad \lim_{n\to\infty}\langle Au_n - Au, u_n - u\rangle = 0 & \implies & u_n \to u \\
(\mathrm{S})_0 & u_n \rightharpoonup u, \quad Au_n \rightharpoonup b, \quad \lim_{n\to\infty}\langle Au_n, u_n\rangle = \langle b, u\rangle & \implies & u_n \to u \\
(\mathrm{S})_1 & u_n \rightharpoonup u, \quad Au_n \to b & \implies & u_n \to u
\end{array}$$

Additionally,

$$(\mathrm{S})_+ \implies (\mathrm{S}) \implies (\mathrm{S})_0 \implies (\mathrm{S})_1.$$

We can define several properties of operators satisfying the conditions.

Lemma 2.26. *Let X be a real reflexive Banach space, and $A : X \to X'$ be a nonlinear operator; then, the following hold:*

$$\begin{array}{rcl}
A \text{ monotone and hemicontinuous} & \implies & A \text{ satisfies } (\mathrm{M}), \\
A \text{ uniformly monotone} & \implies & A \text{ satisfies } (\mathrm{S})_+.
\end{array}$$

Lemma 2.27. *Let X be a real reflexive Banach space, and $A, B : X \to X'$ be nonlinear operators; then, the following hold:*

$$\begin{array}{rcl}
A \text{ satisfies } (\mathrm{S})_+ \text{ and } B \text{ strongly continuous} & \implies & A + B \text{ satisfies } (\mathrm{S})_+, \\
A \text{ satisfies } (\mathrm{S}) \text{ and } B \text{ strongly continuous} & \implies & A + B \text{ satisfies } (\mathrm{S}), \\
A \text{ satisfies } (\mathrm{M}) \text{ and } B \text{ strongly continuous} & \implies & A + B \text{ satisfies } (\mathrm{M}).
\end{array}$$

Exercise 2.7. Prove Lemma 2.26 and Lemma 2.27.

We can now define a *pseudomonotone* operator.

Definition 2.28. Let X be a real reflexive Banach space, $\{u_n\} \subset X$ be a sequence, and $A : X \to X'$ be a nonlinear operator. Then,

- A is called *pseudomonotone* if and only if

$$u_n \rightharpoonup u, \ \limsup_{n\to\infty}\langle Au_n, u_n - u\rangle \le 0 \quad \implies \quad \langle Au, u - w\rangle \le \liminf_{n\to\infty}\langle Au_n, u_n - w\rangle \ \forall w \in X,$$

- A satisfies the condition (P) if and only if

$$u_n \rightharpoonup u \quad \implies \quad \limsup_{n\to\infty} \langle Au_n, u_n - u\rangle \geq 0.$$

We can define several relationships and properties of pseudomonotone operators.

Lemma 2.29. *Let X be a real reflexive Banach space, and $A, B : X \to X'$ be nonlinear operators; then, the following hold:*

$$\begin{aligned}
A \textit{ monotone and hemicontinuous} \quad &\implies \quad A \textit{ pseudomonotone},\\
A \textit{ strongly continuous} \quad &\implies \quad A \textit{ pseudomonotone},\\
A \textit{ demicontinuous and satisfies } (\mathrm{S})_+ \quad &\implies \quad A \textit{ pseudomonotone},\\
A \textit{ continuous \& } \dim X < \infty \quad &\implies \quad A \textit{ pseudomonotone},\\
A \textit{ pseudomonotone \& } B \textit{ pseudomonotone} \quad &\implies \quad A + B \textit{ pseudomonotone},\\
A \textit{ monotone and hemicontinuous \& } B \textit{ strongly continuous} \quad &\implies \quad A + B \textit{ pseudomonotone},\\
A \textit{ monotone \& } B \textit{ strongly continuous} \quad &\implies \quad A + B \textit{ satisfies } (\mathrm{P}).
\end{aligned}$$

Exercise 2.8. Prove Lemma 2.29, except

$$A \text{ continuous \& } \dim X < \infty \quad \implies \quad A \text{ pseudomonotone.}$$

Note that the last two statements follows trivially from the previous statements.

Lemma 2.30 (Properties of pseudomonotone)**.** *, Let X be a real reflexive Banach space, and $A, B : X \to X'$ be nonlinear operators; then, the following hold:*

$$\begin{aligned}
A \textit{ pseudomonotone} \quad &\implies \quad A \textit{ satisfies } (\mathrm{P}) \textit{ and } (\mathrm{M}),\\
A \textit{ pseudomonotone and locally bounded} \quad &\implies \quad A \textit{ demicontinuous},\\
A \textit{ pseudomonotone \& } B \textit{ monotone and hemicontinuous} \quad &\implies \quad A + B \textit{ pseudomonotone},\\
A \textit{ pseudomonotone \& } B \textit{ strongly continuous} \quad &\implies \quad A + B \textit{ pseudomonotone}.
\end{aligned}$$

Exercise 2.9. Prove the first statement from Lemma 2.30; i.e.,

$$A \text{ pseudomonotone} \quad \implies \quad A \text{ satisfies } (\mathrm{P}) \text{ and } (\mathrm{M}).$$

Note that the last two statements from Lemma 2.30 follows trivially from Lemma 2.29.

We now state a result for the existence of a solution to an operator equation, which we can then apply to (2.18).

Theorem 2.31 (Brézis)**.** *Assume that the nonlinear operator $A : X \to X'$ is pseudomonotone, bounded, and coercive on a real, separable, reflexive Banach space; then, the equation*

$$Au = f,$$

has a solution $u \in X$ for every $f \in X'$.

Proof. The proof of this theorem will be shown later; cf., Section 3.3 □

To apply this result to a quasilinear partial differential equation we shall consider a practical example, which can be extended to more general results.

Example 2.2 (Pseudomonotone operator for quasilinear PDE (Zeidler, 1989b, Section 27.4)). Consider the boundary value problem on a bounded Lipschitz domain $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$,

$$-\sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left(\left|\frac{\partial u}{\partial x_i}\right|^{p-2}\frac{\partial u}{\partial x_i}\right) + g(u) = f \qquad \text{in } \Omega, \tag{2.19}$$

$$u = 0, \qquad \text{on } \partial\Omega, \tag{2.20}$$

where $f \in L^q(\Omega)$ and g satisfies the following conditions:

(H1) *Coerciveness*: $g : \mathbb{R} \to \mathbb{R}$ continuous and

$$\inf_{u\in\mathbb{R}} g(u)u > -\infty;$$

(H2) *Growth condition*: For all $u \in \mathbb{R}$

$$|g(u)| \leq C(1 + |u|^{r-1})$$

where $1 < p, q, r < \infty$, $1/p + 1/q = 1$, and $1/p - 1/n < 1/r$.

We seek $u \in X = W_0^{1,p}(\Omega)$ such that

$$a_1(u, v) + a_2(u, v) = \langle F, v\rangle, \qquad \text{for all } v \in X, \tag{2.21}$$

where

$$a_1(u.v) = \int_\Omega \sum_{i=1}^{n}\left|\frac{\partial u}{\partial x_i}\right|^{p-2}\frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_i}\,\mathrm{d}\boldsymbol{x},$$

$$a_2(u, v) = \int_\Omega g(u)v\,\mathrm{d}\boldsymbol{x},$$

$$\langle F, v\rangle = \int_\Omega fv\,\mathrm{d}\boldsymbol{x}.$$

By Proposition 2.4, Exercise 2.6, and Lemmas 2.14–2.18 we know that there exists a unique monotone, coercive, bounded, and hemicontinuous operator $A_1 : X \to X'$ such that

$$a_1(u, v) = \langle Au_1, v\rangle \qquad \text{for all } u, v \in X.$$

By Lemma 2.20 (for $r = p$) and Corollary 2.21 (otherwise) there exists a strongly continuous $A_2 : X \to X'$ such that

$$a_2(u, v) = \langle Au_2, v\rangle \qquad \text{for all } u, v \in X.$$

it can be shown that $F \in X'$; therefore, (2.21) is equivalent to the operator equation

$$A_1u + A_2u = F \qquad u \in X. \tag{2.22}$$

Set $A = A_1 + A_2$; then, by Lemma 2.29 A is pseudomonotone. Additionally, A_1 is bounded and A_2 is strongly continuous (which implies bounded); therefore, A is bounded. From (H1)

$$\langle A_2, u, u\rangle = \int_\Omega g(u) u \,\mathrm{d}\boldsymbol{x} \geq C \qquad \text{for all } u \in X;$$

hence, $A = A_1 + A_2$ is coercive as A_1 is coercive. We can then apply Theorem 2.31 to show that (2.22) has a solution $u \in X$; which is the weak solution of (2.21).

Remark. We can easily generalise the above to the boundary value problem

$$\begin{aligned} -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}(a_i(\boldsymbol{x}, \nabla u)) + g(u) &= f && \text{in } \Omega, \\ u &= 0, && \text{on } \partial\Omega, \end{aligned}$$

providing that the functions a_i, $i = 1, \ldots, n$ meet the conditions (B1)–(B2) and (C1)–(C2).

2.6 Semimonotone operators

In order to extend the results from the previous section to partial differential equations under certain conditions we consider so-called *semimonotone* operators.

Definition 2.32. Let X be a real, separable, reflexive Banach space and $B : X \times X :\to X'$ be a map such that

$$Au = B(u, u) \qquad \text{for all } u \in X.$$

The operator $A : X \to X'$ is called *semimonotone* if and only if the following hold.

a) For all $u, v \in X$
$$\langle B(u, u) - B(u, v), u - v\rangle \geq 0.$$

b) For each $u \in X$, the operator $v \mapsto B(u, v)$ is hemicontinuous and bounded from X to X', and, for each $v \in X$, the operator $u \mapsto B(u, v)$ is hemicontinuous and bounded from X to X'.

c) If $u_n \rightharpoonup u$ in X and
$$\lim_{n\to\infty} \langle B(u_n, u_n) - B(u_n, u), u_n - u\rangle = 0;$$
then, $B(u_n, v) - B(u, v)$ in X' for all $v \in X$,

d) Let $v \in X$, $u_n \rightharpoonup u$ in X, and $B(u_n, v) \rightharpoonup w$ in X' as $n \to \infty$; then,
$$\lim_{n\to\infty} \langle B(u_n, v), u_n\rangle = \langle w, u\rangle.$$

e) A is bounded.

Lemma 2.33. *Let $A : X \to X'$ be a semimonotone operator on a real, separable, reflexive Banach space X; then, A is pseudomonotone.*

Proof. See Leray and Lions (1965). □

Exercise 2.10. Let $A : X \to X'$ be a semimonotone operator on a real, separable, reflexive Banach space X, and $B : X \times X \to X'$ the associated map. Assume that $u_n \rightharpoonup u$, $B(u_n, u) \rightharpoonup w$ and

$$\limsup_{n\to\infty} \langle B(u_n, u_n), u_n - u\rangle \leq 0.$$

Show that

$$\lim_{n\to\infty} \langle B(u_n, u_n) - B(u_n, u), u_n - u\rangle = 0;$$

i.e, show that Definition 2.32 (c) is satisfied. Hence, show that

$$\langle B(u, u), u - w\rangle \leq \liminf_{n\to\infty} \langle B(u_n, u_n), u_n - u\rangle \qquad \text{for all } w \in X;$$

i.e., A is a pseudo-monotone operator.

As we can show that a semimonotone operator is also a pseudomonotone we use Theorem 2.31 to show the existence of a solution to the operator equation $Au = f$ and, furthermore, a solution to the weak formulation (2.10) for quasilinear partial differential equations can be shown under certain assumptions.

Theorem 2.34 (Leray-Lions)**.** *Let X be a real, separable, reflexive Banach space and $A : X \to X'$ be a semimonotone and coercive operator; then,*

$$Au = f$$

has a solution $u \in X$ for every $f \in X'$.

Proof. By Lemma 2.33 A is pseudomonotone. Then, as A is also bounded (by definition of semimonotone) and coercive there exists a solution to $Au = f$ by Theorem 2.31. □

Theorem 2.35. *Let the coefficients functions $a_\alpha : \Omega \times \mathbb{R}^\kappa \to \mathbb{R}$, $|\alpha| \leq k$, $k \in \mathbb{N}$, where κ defined by (1.7), from (2.6) satisfy, for $1 < p < \infty$, the Carathéodory condition (B1), growth condition (B2), and coercivity condition (C2) from Theorem 2.19, as well as the following conditions:*

- ***(I1)*** *The highest order terms are* strictly monotone *with respect to the highest order derivatives; i.e.,*

$$\sum_{|\alpha|=k} \left(a_\alpha(\boldsymbol{x}, \eta, \xi) - a_\alpha(\boldsymbol{x}, \eta, \widehat{\xi})\right)(\xi_\alpha - \widehat{\xi}_\alpha) > 0,$$

for all $\eta \in \mathbb{R}^{\widetilde{\kappa}}$, $\xi, \widehat{\xi} \in \mathbb{R}^{\widetilde{\kappa}-\kappa}$, where

$$\widetilde{\kappa} = \frac{(n+k-1)!}{n!(k-1)!}$$

is the number of multi-indices of length $|\alpha| \leq k-1$.

- ***(I2)*** *The highest order terms are* coercive *with respect to the highest order derivatives; i.e.,*

$$\lim_{|\xi|\to\infty} \sup_{\eta\in D} \sum_{|\alpha|=k} \frac{a_\alpha(\boldsymbol{x}, \eta, \xi)}{|\xi| + |\xi|^{p-1}} == \infty,$$

for almost all $x \in \Omega$ and bounded sets $D \subset \mathbb{R}^{\widetilde{\kappa}}$.

Then, for any $f \in L^q(\Omega)$, $1/p + 1/q = 1$, *there exists at least one weak solution* $u \in W_0^{k,p}(\Omega)$ *to the weak formulation* (2.10).

Proof. It can be shown that the operator $A : W_0^{k,p}(\Omega) \to W^{-k,q}(\Omega)$, $Au = B(u,u)$, where

$$\langle B(w,u), v\rangle = \int_\Omega \sum_{|\alpha|=k} a_\alpha(\boldsymbol{x}, \delta_{k-1}w(\boldsymbol{x}), \widehat{\delta}_k u(\boldsymbol{x}))\partial^\alpha v \,\mathrm{d}\boldsymbol{x} + \int_\Omega \sum_{|\alpha|\leq k-1} a_\alpha(\boldsymbol{x}, \delta_k w(\boldsymbol{x}))\partial^\alpha v \,\mathrm{d}\boldsymbol{x} \tag{2.23}$$

is semimonotone, coercive, and bounded (Leray and Lions, 1965). Then, by Lemma 2.33 and Theorem 2.34 a solution to $Au = F$, where

$$\langle F, v\rangle := \int_\Omega fv \,\mathrm{d}\boldsymbol{x},$$

exists for every $f \in L^q(\Omega)$; hence, a solution $u \in W_0^{k,p}(\Omega)$ to the weak formulation (2.10) exists. □

Exercise 2.11. Let $A : W_0^{k,p}(\Omega) \to W^{-k,q}(\Omega)$, $Au = B(u,u)$, where B is defined in (2.23). Show that B satisfies Definition 2.32 (a)–(b), and that A is bounded and coercive.

2.7 Locally coercive operators

The finally generalisation we consider in this chapter considers operators of the form

$$A : X \to X^+$$

where $\{X, X^+\}$ are so-called *dual pairs* of Banach spaces.

Definition 2.36. Let X and X^+ be Banach spaces over a field $\mathbb{K} = \mathbb{R}, \mathbb{C}$. Then, $\{X, X^+\}$ is called a *dual pair* if and only if the following holds:

a) There exists a bilinear bounded map $\langle \cdot, \cdot \rangle_X : X^+ \times X \to \mathbb{K}$.

b) If $\langle v, u\rangle_X =$ for all $u \in X$; then, $v = 0$.

c) If $\langle v, u\rangle_X =$ for all $v \in X^+$; then, $u = 0$.

Example 2.3 (Examples of dual pairs). The following examples demonstrate that the standard dual of a Banach space forms a dual pair, but dual pairs are not necessary duals.

a) Let X be a Banach space and X' its dual. Then, $\{X, X'\}$ is a dual pair if we use the usual dual operator

$$\langle v, u\rangle_X = v(u) \qquad \text{for all } v \in X', u \in X.$$

b) Let $X = C(\overline{\Omega})$, $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$, bounded domain and set

$$\langle v, u\rangle_X = \int_\Omega vu \,\mathrm{d}\boldsymbol{x} \qquad \text{for all } u, v \in X; \tag{2.24}$$

then, $\{X, X\}$ is a dual pair.

c) Let $X = W^{k,2}(\Omega)$ and $X^+ \in L^2(\Omega)$, with $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$, bounded domain; then, $\{X, X^+\}$ is a dual pair with respect to (2.24).

We now define a result which gives a solution of an operator equation $Au = f$ for *only some* right hand sides f.

Theorem 2.37 (Hess-Kato)**.** *Let the following conditions holds:*

(J1) Dual pairs*: Let $\{X, X^+\}$ and $\{Y, Y^+\}$ be dual pairs where X, X^+, Y, Y^+ are Banach spaces with bilinear forms $\langle \cdot\,,\,\cdot \rangle_X$ and $\langle \cdot\,,\,\cdot \rangle_Y$, respectively, with continuous embeddings*

$$Y \hookrightarrow X, \qquad \textit{and} \qquad X^+ \hookrightarrow Y^+$$

which are compatible; i.e.,

$$\langle v, u\rangle_X = \langle v, u\rangle_Y \qquad \textit{for all } v \in X^+, y \in Y.$$

Moreover, X is reflexive and separable, and Y is reflexive.

(J2) Operator A*: $A : \mathcal{D}(A) \subset X \to Y^+$ be a given operator and $K \subset X$ a bounded, closed, convex set containing zero and $K \cap Y \subset \mathcal{D}(A)$.*

(J3) Local coerciveness*: There exists a constant $\alpha > 0$ such that*

$$\langle Av, v\rangle_Y \geq \alpha, \qquad \textit{for all } v \in Y \cap \partial K.$$

(J4) Continuity*: For each finite dimensional subspace Y_0 of the Banach space Y, the mapping*

$$w \mapsto \langle Aw, v\rangle_Y,$$

is continuous on $K \cap Y_0$ for all $v \in Y_0$.

(J5) Generalised condition (M)*: Let $\{u_n\}$ be a sequence in $Y \cap K$ and let $b \in X^+$; then,*

$$\begin{aligned} &u_n \rightharpoonup u \in X,\ \langle Au_n, v\rangle_Y \to \langle b, v\rangle_Y \ \ \forall v \in Y, \\ &\qquad\qquad\qquad\qquad \limsup_{n\to\infty} \langle Au_n, u_n\rangle_Y \leq \langle b, u\rangle_X \qquad \Longrightarrow \qquad Au = b. \end{aligned}$$

(J6) Quasi-boundedness*: Let $\{u_n\}$ be a sequence in $Y \cap K$; then, for a constant $C > 0$,*

$$u_n \rightharpoonup u \in X,\ \langle Au_n, u_n\rangle \leq C\|u_n\|_X \ \forall n \in \mathbb{N} \qquad \Longrightarrow \qquad \{Au_n\} \subset Y^+ \textit{ bounded}.$$

Then, for each $b \in X^+$ with $\langle b, v\rangle_X \leq \alpha$ for all $v \in K \cap Y$,

$$Au = b, \qquad u \in \mathcal{D}(A),$$

has a solution u. Additionally, this solution is unique if

(J7) Local strict monotonicity*: There exists a dual pair $\{Z, Z^+\}$ of Banach space Z and Z^+ with continuous embedding*

$$X \hookrightarrow Z \qquad \textit{and} \qquad Y^+ \hookrightarrow Z^+$$

such that

$$\langle Au - Av, u - v\rangle_Z > 0, \qquad \textit{for all } u, v \in \mathcal{D}(A) \textit{ with } u \neq v.$$

The solution has continuous dependence on the data if the following holds; i.e., for $Au_1 = b_1$ and $Au_2 = b_2$, $b_1, b_2 \in X^+$ it holds that

$$\|u_1 - u_2\|_Z \leq C\|b_1 - b_2\|_{X^+}.$$

(J8) Local strong monotonicity: *(J7) holds and there exists a constant $d > 0$ such that*

$$\langle Au - Av, u - v\rangle_Z \geq d\|u - v\|_Z^2, \qquad \text{for all } u, v \in \mathcal{D}(A).$$

Remark. Consider Banach space X and Y over a field $\mathbb{K}$ with continuous dense embedding $Y \hookrightarrow X$ (i.e., $X = Y$); then, $X' \hookrightarrow Y'$ is a continuous embedding in the sense that a linear continuous functional $b : X \to \mathbb{K}$ is also a linear continuous functional $b : Y \to \mathbb{K}$ by restricting b to Y. In this case, we can set $X^+ = X'$ and $Y^+ = Y$ which define dual pairs, and set K to be a closed ball in X.

Corollary 2.38 (Special case for balls)**.** *Suppose (J1)–(J6) holds, and let K° denote the polar set of K; i.e.,*

$$K^\circ = \{b \in X^+ : \langle b, v\rangle_X \leq 1 \ \forall v \in K\}.$$

Then, $\alpha K^\circ \subset \mathcal{R}(A)$; i.e., $Au = b$ has a solution u for each $b \in \alpha K^\circ$. In particular, if K is a ball of radius $R > 0$, i.e.

$$K = \{v \in X : \|v\|_X \leq R\},$$

and if

$$\langle b, v\rangle_X \leq C\|b\|_{X^+}\|v\|_X \qquad \text{for all } b \in X^+, v \in X,$$

and fixed $C > 0$; then,

$$\left\{b \in X^+ : \|b\|_{X^+} \leq {}^1\!/_{CR}\right\} \subset K^\circ,$$

i.e. $Au = b$ has a solution u for each $b \in X^+$ with $\|b\|_{X^+} \leq {}^\alpha\!/_{CR}$.

Corollary 2.39 (Global coerciveness)**.** *Suppose (J1)–(J6) holds for all balls K in X and suppose the global coerciveness condition*

$$\lim_{\|v\|_X \to \infty} \frac{\langle Av, v\rangle_Y}{\|v\|_X} = +\infty, \qquad v \in Y,$$

is satisfied. Then, $X^+ \subset \mathcal{R}(A)$; i.e., $Au = b$ has a solution u for each $b \in X^+$.

Example 2.4 (Strongly nonlinear semilinear PDE (Zeidler, 1989b, Section 27.8)). Consider the boundary value problem

$$-\Delta u + g(u) = f \qquad \text{in } \Omega, \tag{2.25}$$
$$u = 0, \qquad \text{on } \partial\Omega. \tag{2.26}$$

Here, we assume no growth condition on g; therefore, we talk about a *strong nonlinear* problem, such as $g(u) = \mathrm{e}^u$. We assume the following conditions:

(K1) $g : \mathbb{R} \to \mathbb{R}$ continuous with $(g(u) - a)u \geq 0$ for all $u \in \mathbb{R}$ with fixed $a \in \mathbb{R}$.

(K2) Ω bounded domain in $\mathbb{R}^n$ with piecewise smooth boundary.

If $g : \mathbb{R} \to \mathbb{R}$ is continuous and monotone (i.e., $g(u) = \mathrm{e}^u$), then (K1) is satisfied and $a = g(0)$. We reduce the problem to an operator equation

$$Au = b, \qquad u \in \mathcal{D}(A) := \{u \in X : h \in L^1(\Omega)\},$$

where $h(\boldsymbol{x}) = (g(u(\boldsymbol{x})) - a)u(\boldsymbol{x})$ and $X = W_0^{1,2}(\Omega)$. Setting $Y = W^{k,2}(\Omega) \cap X$, $k > n/2$, and $\|u\|_X = \|u\|_{k,2}$. Since $k > n/2$, $Y \hookrightarrow C(\overline{\Omega})$ continuously by Sobolev embeddings. Note, that $x \mapsto g(u(\boldsymbol{x})) \in L^1(\Omega)$.

Let $g \in L^2(\Omega)$; then, we seek $u \in \mathcal{D}(A)$ such that

$$a_1(u, v) + a_2(u, v) = b(v), \qquad \text{for all } v \in Y, \tag{2.27}$$

where

$$\begin{aligned} a_1(u, v) &= \int_\Omega \nabla u \cdot \nabla v \,\mathrm{d}\boldsymbol{x}, \\ a_2(u, v) &= \int_\Omega g(u(\boldsymbol{x}))v(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}, \\ b(v) &= \int_\Omega f v \,\mathrm{d}\boldsymbol{x}. \end{aligned}$$

Instead of (2.25) we can consider

$$-\Delta u + (g(u) - a) = f - a \qquad \text{in } \Omega.$$

For simplicity we assume $a = 0$.

We aim to apply Theorem 2.37 to a sufficiently large ball K in X. Set $X^+ = X'$ and $Y^+ = Y'$. Then we need to show the various requirements of the theorem:

1. $|b(v)| \le C\|f\|_{0,2}\|v\|_X$ for all $v \in X$; hence, $b \in X'$.

2. $|a_1(u, v)| \le C\|u\|_X\|v\|_Y$ for all $u \in X$, $v \in Y$; hence, there exists a unique $A_1 : X \to Y'$ such that
$$\langle A_1 u, v\rangle_Y = a_1(u, v), \qquad \text{for all } u \in X, v \in Y.$$

3. By Sobolev embeddings
$$|a_2(u, v)| \le \int_\Omega |g(u)| \,\mathrm{d}\boldsymbol{x} \|v\|_{C(\overline{\Omega})} \le C\|v\|_Y \qquad \text{for all } v \in Y; \tag{2.28}$$
hence, there exists a unique $A_2 : \mathcal{D}(A) \subset X \to Y'$ such that
$$\langle A_2 u, v\rangle_Y = a_2(u, v), \qquad \text{for all } u \in \mathcal{D}(A) \subset X, v \in Y.$$

4. Define $A = A_1 + A_2$; then, (2.27) is equivalent to
$$Au = b \qquad u \in \mathcal{D}(A).$$
Note that $Y \subset C(\overline{\Omega}) \subset \mathcal{D}(A) \subset X$.

5. *Global coerciveness of Y*: $g(u)u \geq 0$; therefore, by Poincaré-Friedrich

$$\langle Av, v\rangle_Y = a_1(v,v) + a_2(v,v) \geq a_1(v,v) \geq d\|v\|_X^2 \qquad \text{for all } v \in Y.$$

Hence, for $v \in Y$,

$$\lim_{\|v\|_X\to\infty} \frac{\langle Av, v\rangle_Y}{\|v\|_X} \geq \lim_{\|v\|_X\to\infty} d\|v\|_X = +\infty.$$

6. *Generalised condition* (M): Let $b \in Y'$, $\{u_n\} \subset Y$ be a sequence such that $u_n \rightharpoonup u$ in X, and

$$\langle Au_n, v\rangle_Y \to b(v) \qquad \text{for all } v \in Y, \tag{2.29}$$

$$\limsup_{n\to\infty} \langle Au_n, u_n\rangle_Y \leq b(u); \tag{2.30}$$

then, we need to show that $Au = b$.

- $A_1 : X \to Y'$ is linear and continuous; therefore, it is also weakly continuous:

$$\langle A_1 u_n, v\rangle_Y \to \langle A_1 u, v\rangle_Y \qquad \text{for all } v \in Y.$$

- If we can show

$$\langle A_2 u_n, v\rangle_Y \to \langle A_2 u, v\rangle_Y \qquad \text{for all } v \in Y,$$

and $u \in \mathcal{D}(A)$; then, by (2.30) we complete this proof. As $U \subset C(\overline{\Omega})$ and

$$\langle A2u, v\rangle = \int_\Omega g(u)v \, \mathrm{d}\boldsymbol{x};$$

then, by (2.28) it is sufficient to show that $g(u_n(\boldsymbol{x})) \to g(u(\boldsymbol{x}))$ in $L^1(\Omega)$ for a subsequence of u_n. This requires additional results (Vitali convergence theorem and Fatou lemma), hence, we skip this prove here.

7. *Quasi-boundedness of A*: Let $\{u_n\}$ be a sequence in Y with $u_n \rightharpoonup u$ in X, and suppose that

$$\langle Au_n, u_n\rangle_Y \leq C\|u_n\|_X \qquad \text{for all } n \in \mathbb{N};$$

then, we need to show that $\{Au_n\}$ is bounded in Y'. The boundedness of $\{u_n\}$ in X, implies that there exists a constant $C > 0$ such that

$$\limsup_{n\to\infty} \langle Au_n, u_n\rangle_Y \leq C.$$

Suppose that $\{Au_n\}$ is unbounded in Y'; then, there exists a subsequence $\{u_n\}$ such that

$$\|Au_n\|_{Y'} \to \infty.$$

Similarly to step 6 above we can show that

$$\langle Au_n, v\rangle_Y \to \langle Au, v\rangle_Y \qquad \text{for all } v \in Y$$

for the subsequence. By the *uniform boundedness principle*, Proposition 1.8 (9), the sequence $\{Au_n\}$ is bounded; which is a contradiction.

By Corollary 2.39 $Au = b$ has a solution $u \in \mathcal{D}(A)$ if $b \in X'$.

CHAPTER 3

Finite Dimensional Approximation

We want to study the operator equations $Au = b$ and nonlinear partial differential equations when we have a finite dimensional Banach space $X_n \subset X$. There are two main reasons for studying finite dimensional spaces:

- The proofs of several of the theorems from Chapter 2 are based on finding the solution of an operator equation on a sequence of finite dimensional spaces which tend to the solution on the infinite dimensional space in the limit.
- Numerical approximation of the solution on a finite dimensional space; e.g., by the finite element method.

3.1 Galerkin approximation

We first need to define the approximation to the operator equation $Au = b$ on a finite dimensional subspace.

Definition 3.1 (Galerkin approximation)**.** Let X_n be a finite dimensional subspace of a *separable* Banach space X. Then, the *Galerkin approximation* of the operator equation $Au = b$ on X_n is: Find $u_n \in X_n$ such that

$$\langle Au_n, v\rangle = \langle b, v\rangle \qquad \text{for all } v \in X_n. \tag{3.1}$$

This can be understood as

$$A_n u_n = b_n$$

where $A_n : X_n \to X_n'$ is the restriction of A to X_n and $A_n u_n$, b_n are functionals on X restricted to X_n.

Let $X_n = \operatorname{span}\{v_i, i = 1, \ldots, n\}$ be a finite dimensional subspace of a Banach space X. We can define P_n as the projection from X to X_n (a continuous linear operator) and construct the dual operator projection $P_n^d : X_n' \to X'$ by

$$\langle f_n, P_n x\rangle_{X_n'\times X_n} = \langle P_n^d f_n, x\rangle_{X'\times X} \qquad \text{for all } f_n \in X_n', x \in X;$$

cf. Section 1.3. Let $A_n = P_n^d A P_n$ and $b_n = P_n^d b$; then, $A_n u_n = b_n$ is the Galerkin approximation to $Au = b$ on X_n. By definition of the Galerkin approximation the following are equivalent:

- Find $u_n \in X_n$ such that $\langle Au_n, v\rangle = \langle b, v\rangle$ for all $v \in X_n$.

- Find $u_n \in X_n$ such that $\langle P_n^d A P_n u_n, v\rangle = \langle P_n^d b, v\rangle$ for all $v \in X_n$.
- Find $u_n \in X_n$ such that $A_n u_n = b_n$.

For the basis v_i, $i = 1, \ldots, n$ of X_n we construct the dual basis $\{v_i'\}$, $v_i' \in X_n$, $i = 1, \ldots, n$; i.e.,

$$\langle v_j', v_i\rangle = \delta_{ij} \qquad \text{for } i, j = 1, \ldots, n.$$

We also denote by v_i' the continuous linear extension of $v_i' \in X_n'$ to X (cf. Hahn-Banach theorem). We then define

$$P_n x = \sum_{i=1}^{n} \langle v_i', x\rangle v_i, \qquad x \in X.$$

Clear P_n is a projection in the sense that $P_n x = x$ for all $x \in X_n$, and is a linear and continuous operator from X to X_n.

Lemma 3.2. *Let $X_n \subset X$ be a finite dimensional subspace of a Banach space, $A : X \to X'$ and $A_n : X_n \to X_n'$ with $A_n = P_n^d A P_n$ for the projection $P_n : X \to X_n$. If A is (strictly/uniformly/strongly) monotone, (strongly/weakly/Lipschitz/hemi-/demi-) continuous, or coercive; then, A_n is also (strictly/uniformly/strongly) monotone, (strongly/weakly/Lipschitz/hemi-/demi-) continuous, or coercive, respectively.*

Remark. In general, the constants for strongly monotone and Lipschitz continuity will be the same.

3.2 Iterative Galerkin for strongly monotone

We first consider the Galerkin approximation of the operator equation $Au = f$ for a strongly monotone and Lipschitz operator on a Hilbert space X and its application to the numerical approximation of a quasilinear elliptic partial differential equation; cf. Section 2.2. By Theorem 2.8 and Corollary 2.9 we know that there exists a unique $u \in X$ such that $Au = f$ for all $f \in X'$ and the iteration

$$u_{m+1} = u_m - \varepsilon J_X^{-1}(Au_m - f)$$

converges to u.

Let $X_n \subset X$ be a finite dimensional subspace of X with $\dim X_n < \infty$ and approximate $Au = f$ in X_n.

Theorem 3.3 (Zarantonello). *Let A be a strongly monotone and Lipschitz continuous operator, $f \in X'$, and $X_n \subset X$ be a finite dimensional subspace, $A_n = P_n^d A P_n$, $f_n = P_n^d f$, where $P_n : X \to X_n$ is the projection; then, $A_n u = f_n$ has a unique solution $u_n \in X_n$ and the fixed point iteration*

$$u_n^{(m+1)} = u_n^{(m)} - \varepsilon J_{X_n}^{-1}(A_n u_n^{(m)} - f_n) \qquad m \geq 0 \tag{3.2}$$

converges to u_n for any initial $u_n^{(0)} \in X_n$ with contraction factor $k = (1 + \varepsilon^2 L^2 - 2\varepsilon M)^{1/2}$ if ε is selected such that $k^2 \leq 1$, where M and L are the constants from strong monotonicity and Lipschitz continuity. Furthermore, given the solution $u \in X$ to $Au = f$ from Theorem 2.8,

$$\|u - u_n\|_X \leq \frac{L}{M} \inf_{v \in X_n} \|u - v\|_X$$

where

$$\inf_{v \in X_n} \|u - v\|_X \to 0 \qquad \text{as } n \to \infty.$$

Proof. By Lemma 3.2 A_n is strongly monotone and Lipschitz continuous with constants M and L, respectively. Theorem 2.8 and Corollary 2.9 proves the existence of the solution $u_n \in X_n$ and convergence of the iteration. Furthermore,

$$\begin{aligned} \langle Au_n, v\rangle = \langle A_n u_n, v\rangle = \langle f, v\rangle = \langle f, v\rangle &\qquad \text{for all } v \in X_n, \\ \langle Au, v\rangle = \langle f, v\rangle &\qquad \text{for all } v \in X \supset X_n, \\ \Longrightarrow \qquad \langle Au_n - Au, v\rangle = 0 &\qquad \text{for all } v \in X_n. \end{aligned}$$

Therefore, by strongly monotone and Lipschitz continuity, for all $z \in X_n$,

$$\begin{aligned} M\|u - u_n\|_X^2 &\leq \langle Au - Au_n, u - u_n\rangle \\ &\leq \langle Au - Au_n, u - z\rangle \\ &\leq \|Au - Au_n\|_{X'}\|u - z\|_X \\ &\leq L\|u - u_n\|_X\|u - z\|_X. \end{aligned}$$

Hence,

$$\|u - u_n\|_X \leq \frac{L}{M}\|u - z\|_X \qquad \text{for all } z \in X_n,$$

and, as X_n is closed,

$$\|u - u_n\|_X \leq \frac{L}{M}\inf_{v \in X_n}\|u - v\|_X.$$

□

Theorem 3.4. *Let $u \in X$ be the solution from Theorem 2.8 to $Au = f$ on the Banach space X, $u_n^0 \in X_n$ be an initial guess $u_n^m \in X_n$, $m \geq 1$ be the approximation on the finite dimensional subspace $X_n \subset X$ after the m-th iteration of (3.2), and*

$$\varepsilon = \frac{M}{L^2}, \qquad k = \left(1 - \frac{M^2}{L^2}\right)^{1/2};$$

then,

$$\|u - u_n^{(m)}\|_X \leq \frac{L}{M}\inf_{v \in X_m}\|u - v\| + \frac{2L^2}{M^2}\left(1 - \frac{M^2}{L^2}\right)^{n/2}\|u_n^{(0)} - u_n^{(1)}\|_X.$$

Proof. By the triangle inequality, Theorem 3.3, Theorem 2.8, and the fact that $0 < k < 1$,

$$\begin{aligned} \|u - u_n^{(m)}\|_X &\leq \|u - u_n\|_X + \|u_n - u_n^{(m)}\|_X \\ &\leq \frac{L}{M}\inf_{v \in X_n}\|u - v\|_X + \frac{k^n}{1-k}\|u_n^{(0)} - u_n^{(1)}\|_X \\ &= \frac{L}{M}\inf_{v \in X_n}\|u - v\|_X + \frac{k^n(1+k)}{1-k^2}\|u_n^{(0)} - u_n^{(1)}\|_X \\ &\leq \frac{L}{M}\inf_{v \in X_n}\|u - v\|_X + \frac{2k^n}{1-k^2}\|u_n^{(0)} - u_n^{(1)}\|_X. \end{aligned}$$

Replacing k with its definition completes the proof. □

Remark (Practical Implementation). Choose the basis $\{v_1, \ldots, v_n\}$ of X_n and define

$$u_n = \sum_{i=1}^{n} \alpha_i v_i$$

for some unknown vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^\top \in \mathbb{R}^n$. Then, we can write the Galerkin approximation (3.1) as

$$\begin{aligned} && \langle Au_n, v_i \rangle &= \langle f, v_i \rangle, && i = 1, \dots, n, \\ \Longrightarrow && \left\langle A\left(\sum_{j=1}^{n} \alpha_j v_j\right), v_i \right\rangle &= \langle f, v_i \rangle, && i = 1, \dots, n. \end{aligned}$$

This can be written as a (nonlinear) algebraic system

$$\boldsymbol{F}(\boldsymbol{\alpha}) = \boldsymbol{\ell},$$

where $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^n$ is defined as

$$\boldsymbol{F}(\boldsymbol{\alpha}) = \left(\left\langle A\left(\sum_{j=1}^{n} \alpha_j v_j\right), v_i \right\rangle\right)_{i=1}^{m}$$

and $\boldsymbol{\ell} = (\langle f, v_i \rangle)_{i=1}^n$. We can define the fixed point iteration (3.2) as

$$(u_n^{(m+1)}, v_i)_X = (u_n^{(m)}, v_i)_X - \varepsilon\langle A_n u_n^m - f_n, v_i \rangle, \qquad i = 1, \dots, n.$$

Let $\alpha^{(m+1)}, \alpha^{(m)} \in \mathbb{R}^n$ be the vectors corresponding to $u_n^{(m+1)}, u_n^{(m)}$, respectively; then

$$\sum_{j=1}^{n} \alpha_j^{(m+1)} (v_j, v_i)_X = \sum_{j=1}^{n} \alpha_j^{(m)} (v_j, v_i)_X - \varepsilon(F_i(\boldsymbol{\alpha}^{(m)}) - \ell_i), \qquad i = 1, \dots, n, m \geq 0.$$

Define the mass matrix $\mathbb{M} \in \mathbb{R}^{n \times n}$ as $M_{ij} = (v_j, v_i)_X$; then we get the following linear algebraic system for each iteration:

$$\mathbb{M}\boldsymbol{\alpha}^{(m+1)} = \mathbb{M}\boldsymbol{\alpha}^{(m)} - \varepsilon(\boldsymbol{F}(\boldsymbol{\alpha}^{(m)}) - \boldsymbol{\ell}), \qquad m \geq 0.$$

Remark. $\mathbb{M}$ is symmetric positive definite.

Example 3.1 (Iterative Galerkin method for strongly monotone & Lipschitz continuous PDE). Consider the boundary value problem in a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$,

$$\begin{aligned} -\nabla \cdot (\mu(\boldsymbol{x}, |\nabla u|)\nabla u) &= f && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega, \end{aligned}$$

where $\mu \in C(\overline{\Omega} \times [0, \infty))$ and there exists positive constants $\alpha_1 \geq \alpha_2 > 0$ such that, for $t \geq s \geq 0$ and $\boldsymbol{x} \in \overline{\Omega}$

$$\alpha_2(t - s) \leq \mu(\boldsymbol{x}, t)t - \mu(\boldsymbol{x}, s)s \leq \alpha_1(t - s).$$

This is the Example 2.1; hence, we know there exists a unique weak solution $u \in X = H_0^1(\Omega)$ such that

$$a(u, v) := \int_\Omega \mu(\boldsymbol{x}, |\nabla u|)\nabla u \cdot \nabla v \, \mathrm{d}\boldsymbol{x} = \int_\Omega f v \, \mathrm{d}\boldsymbol{x} =: \langle F, v \rangle \qquad \text{for all } v \in H_0^1(\Omega). \tag{3.3}$$

We consider the following finite element discretisation:

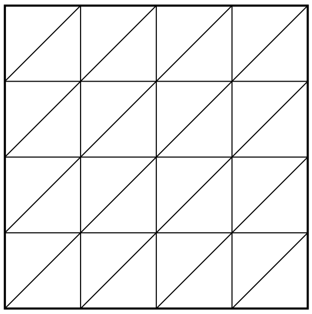

Figure 3.1: Example finite element mesh for $\Omega = (0,1)^2$

- Partition Ω into non-overlapping simplices (elements) T such that $\Omega = \bigcup_{T \in \mathcal{T}_h} T$ to create the mesh $\mathcal{T}_h$. We assume standard finite element assumptions on the mesh (shape regularity, etc.).
- Denote by h_T the diameter of T and set $h = \max_{h \in \mathcal{T}_h} h_T$.
- On each element approximate by polynomials of total order $p \geq 1$ and impose continuity of the approximation over element intersections (edges/faces).
- Define the finite dimensional space

  $$X_h := \{v \in H_0^1(\Omega) : v|_T \in P_p(T), \forall T \in \mathcal{T}_h\} \subset X = H_0^1(\Omega),$$

  where $P_p(T)$ is the space of polynomials of total order p on $T \in \mathcal{T}_h$.

We can now define the *iterative Galerkin* finite element method: Given an initial guess $u_h^{(0)} \in X_h$ we iterate for $m = 1, 2, \ldots$ and find $u_h^{(m+1)} \in X_h$ such that

$$(u_h^{(m+1)}, v_h)_X = (u_h^{(m)}, v_h)_X = \frac{\alpha_2}{\alpha_1^2}\left(a(u_h^{(m)}, v_h) - \langle F, v_h\rangle\right) \qquad \text{for all } v_h \in X_h, \tag{3.4}$$

where

$$(u, v)_X = \int_\Omega \nabla u \cdot \nabla v \, \mathrm{d}\boldsymbol{x}.$$

By Theorem 3.3 this converges to the solution $u_h \in X_h$ with contraction factor

$$k = \left(1 - \frac{\alpha_2^2}{\alpha_1^2}\right)^{1/2} < 1.$$

Let $u \in H^{s+1}(\Omega) \cap H_0^1(\Omega)$, $s \geq 1$ be the weak solution given by (3.3), $u_h^{(0)} \in X_h$ be any initial guess, and $u_h^{(m)} \in X_h$ be the numerical solution after m steps of the iteration (3.4); then, for $m \geq 1$,

$$\|u - u_h^m\| \leq C\frac{\alpha_1}{\alpha_2} h^{\min(p,s)} \|u\|_{s+1,2} + \frac{2\alpha_1^2}{\alpha_2^2}\left(1 - \frac{\alpha_2^2}{\alpha_1^2}\right)^{n/2} \|u_h^{(0)} - u_h^{(1)}\|_{1,2},$$

where $C > 0$ is a constant independent of h, α_1 and α_2. This result follows from Theorem 3.4 and standard finite element results to bound $\inf_{v \in X_h} \|u - v\|$.

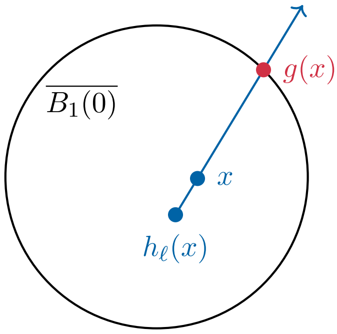

Figure 3.2: Definition of $g(x)$ for proof of Brouwer fixed point theorem

3.3 Minty-Browder & Brézis

We can also use the Galerkin approximation to prove both the Minty-Browder (Theorem 2.11) and Brézis (Theorem 2.31) theorems from Chapter 2. In order to prove Theorem 2.11 we require the *Brouwer fixed point theorem*, which we can state as three related theorems.

Theorem 3.5 (Brouwer fixed point theorem (unit ball)). *Let $\overline{B_1(0)} = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ (unit ball) and $f : \overline{B_1(0)} \to \overline{B_1(0)}$ be continuous. Then, f has a fixed point $\overline{x}$ in $\overline{B_1(0)}$; i.e., $f(\overline{x}) = \overline{x}$.*

Proof. $f : \overline{B_1(0)} \to \overline{B_1(0)}$ is continuous; then, by Weierstrass approximation theorem there exists a sequence of polynomials

$$p_\ell : \overline{B_1(0)} \to \mathbb{R}^n, \qquad \ell \geq 1,$$

such that

$$\sup_{x \in \overline{B_1(0)}} |f(x) - p_\ell(x)| \leq \frac{1}{\ell}, \qquad \ell \geq 1.$$

Then, for $x \in \overline{B_1(0)}$,

$$|p_\ell(x)| \leq |f(x)| + |p_\ell(x) - f(x)| \leq 1 + \frac{1}{\ell}.$$

We can then scale $p_\ell(x)$ to define $h_\ell : \overline{B_1(0)} \to \overline{B_1(0)}$ as

$$h_\ell(x) = \left(1 + \frac{1}{\ell}\right)^{-1} p_\ell(x),$$

and we have that

$$|f(x) - h_\ell(x)| \leq |f(x) - p_\ell(x)| + |p_\ell(x) - h_\ell(x)| \leq \frac{1}{\ell} + \left(1 - \left(1 + \frac{1}{\ell}\right)^{-1}\right) |p_\ell(x)| \to 0 \quad (3.5)$$

as $\ell \to \infty$. We suppose that h_ℓ does not have a fixed point; therefore, for any $x \in \overline{B_1(0)}$, $h_\ell(x) \neq x$. We define $g(x)$ as the intersection of a line starting at the point $h_\ell(x)$ and passing through x with the boundary $\partial B_1(0)$; cf. Figure 3.2. We note that $g : \overline{B_1(0)} \to \partial B_1(0)$ is a C^1 function and $g(x) = x$ for all $x \in \partial B_1(0)$. The *retraction principal*, however, states that there exists no C^1-mapping $F : \overline{B_1(0)} \to \partial B_1(0)$ with $F(x) = x$ for all $x \in \partial B_1(0)$; hence, we have a contradiction, meaning that $h_\ell(x)$ has a fixed point $\xi_\ell \in \overline{B_1(0)}$ for all $\ell \geq 1$. The sequence $\{\xi_\ell\}_{\ell\geq 1} \subset \overline{B_1(0)}$ is a bounded sequence and, thus, has a convergent subsequence $\xi_{\ell'} \to \overline{\xi} \in \overline{B_1(0)}$. Then, by (3.5)

$$|f(\overline{\xi}) - \overline{\xi}| = \lim_{\ell'\to\infty} |f(\xi_{\ell'}) - \xi_{\ell'}| = \lim_{\ell'\to\infty} |f(\xi_{\ell'}) - h_{\ell'}(\xi_{\ell'})| = 0.$$

Hence $f(\overline{\xi}) = \overline{\xi}$, and $\overline{\xi}$ is a fixed point of f. □

Theorem 3.6 (Brouwer fixed point theorem (subset $\mathbb{R}^n$))**.** *Let $K \subset \mathbb{R}^n$, $n \geq 1$, be a non-empty, closed, convex, and bounded set, and $f : K \to K$ be a continuous mapping. Then, f has a fixed point $\overline{x} \in K$.*

Proof. We only briefly outline the proof. Let $z \in \operatorname{int} K$, $K_1 = \{-z + x, x \in K\} \equiv -z + K$, and $T(x) = -z + x$, $x \in K$, $T : K \to K_1$. Define

$$f_1(y) = T \circ f \circ T^{-1}y, \qquad y \in K_1.$$

If $\overline{y} \in K_1$ is a fixed point of f_1 then $\overline{x} = T^{-1}\overline{y}$ is a fixed point of f; therefore, it is sufficient to show a fixed point of f_1 exists. It is possible to define a continuous mapping $h : K_1 \to \overline{B_1(0)}$ and show that h is a homeomorphism (which we skip, as the definition requires more results). Then, we define $g = h \circ f_1 \circ h^{-1}$, where $g : \overline{B_1(0)} \to \overline{B_1(0)}$ is continuous, and there exists a fixed point $\overline{w} \in \overline{B_1(0)}$ by Theorem 3.5 such that

$$\overline{w} = g(\overline{w}) = (h \circ f_1 \circ h^{-1})(\overline{w}) \qquad \iff \qquad h^{-1}(\overline{w}) = f(h^{-1}(\overline{w})).$$

Therefore, $\overline{y} = h^{-1}(\overline{w}) \in K$ is a fixed point of f_1 on K_1. □

Theorem 3.7 (Brouwer fixed point theorem (finite dimensional linear space))**.** *Let X be a finite dimensional linear space, $K \subset X$ a closed, convex, and bounded subset, and $f : K \to K$ be a continuous mapping. Then, f has a fixed point $\overline{x} \in K$.*

Proof. Define

$$x = \sum_{i=1}^{n} \alpha_i x_i,$$

where $x_1, \ldots, x_n$ are the basis for X, $\dim X = n$, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$. Then, defining $T : X \to \mathbb{R}^n$ as $T(x) = \boldsymbol{\alpha}$, it is sufficient to show that T is a homeomorphism such that Theorem 3.6 can be applied to show $g = T \circ f \circ T^{-1}$ has a fixed point. □

Exercise 3.1. Prove Theorem 3.7.

Remark. We note several points about Brouwer fixed point theorem:

1. In the one-dimensional case, let $f : [a, b] \to [a, b]$. Then, defining $g(x) = f(x) - x$, we have that

$$\begin{aligned} & g(a) = f(a) - a \geq a - a = 0 \\ & g(b) = f(b) - b \leq b - b = 0 \\ \implies \quad & g(b) \leq 0 \leq g(a). \end{aligned}$$

   The intermediate value theorem states that there exists a $\xi \in [a, b]$ with $g(\xi) = 0 \iff f(\xi) = \xi$. Hence, Brouwer fixed point theorem is equivalent to the intermediate value theorem.

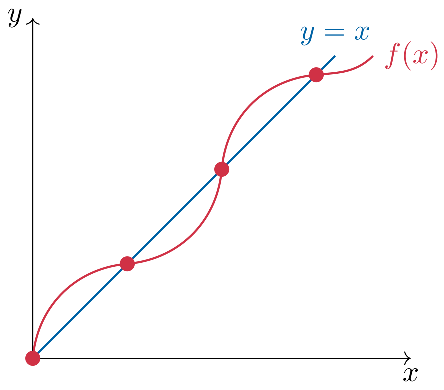

Figure 3.3: Simple proof that fixed point for Brouwer is not unique — fixed points of $f(x)$ occur when $f(x)$ intersects with $y = x$

2. The fixed point may not necessarily be unique, cf. Figure 3.3.
3. The fixed point iteration $x_{n+1} = f(x_n)$ may not necessarily converge.
4. The theorem does not hold for infinite dimensional spaces, see *Kakutani's counter-example*.

To prove Minty-Browder we use a corollary of the Brouwer fixed point theorem.

Corollary 3.8. *Let $\overline{B_R(0)} = \{x \in \mathbb{R}^n : \|x\| \leq R\}$ for fixed $R > 0$, the functions $g_i : \overline{B_R(0)} \to \mathbb{R}$, $i = 1, \ldots, n$ be continuous, and*

$$\sum_{i=1}^{n} g_i(x)x_i \geq 0 \qquad \text{for all } x : \|x\| \leq R,$$

where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Then,

$$g_i(x) = 0, \qquad i = 1, \ldots, n,$$

has a solution $x \in \mathbb{R}^n$ such that $\|x\| \leq R$.

Proof. Set $g(x) = (g_1(x), \ldots, g_n(x))$ and suppose that $g(x) \neq 0$ for all $x \in \overline{B_R(0)}$. Then, define

$$f(x) = \frac{-Rg(x)}{\|g(x)\|},$$

and by Theorem 3.6 there exists a fixed point $\overline{x}$ such that $\overline{x} = f(\overline{x})$ and $\|\overline{x}\| = \|f(\overline{x})\| = R$. Furthermore,

$$\sum_{i=1}^{n} g_i(\overline{x})\overline{x}_i = -\frac{1}{R}\|g(\overline{x})\| \sum_{i=1}^{n} f_i(\overline{x})\overline{x}_i^2 < 0,$$

which contradictions the assumption on g_i. Therefore, there must exist a $x \in \overline{B_R(0)}$ such that $g(x) = 0$. $\square$

We can now prove Theorem 2.11, but we first recall the theorem along with a result about the existence and convergence of the Galerkin approximation which we prove at the same time.

Theorem 2.11 (Minty-Browder)**.** *Let $A : X \to X'$ be a monotone, coercive, and hemicontinuous operator on a real reflexive Banach space X. Then, for each $f \in X'$ the equation $Au = f$ has at least one solution (A is surjective) and the set of all solutions is bounded, convex, and closed. Additionally, if A is* strictly monotone *then the solution is* unique, *the inverse A^{-1} exists and*

$$\begin{aligned} A \text{ uniformly monotone} &\implies A^{-1} \text{ continuous} \\ A \text{ strongly monotone} &\implies A^{-1} \text{ Lipschitz continuous} \end{aligned}$$

Corollary 3.9. *Let $\dim X = \infty$ and X be separable in Theorem 2.11, and $X_n = \operatorname{span}\{v_1, \dots, v_n\} \subset X$ be a finite dimensional subspace; then, the Galerkin equation*

$$\langle A_n u_n, v_i \rangle = \langle f, v_i \rangle \qquad i = 1, \dots, n \tag{3.6}$$

has a solution

$$u_n = \sum_{i=1}^{n} c_{in} v_i,$$

where $c_{in} \in \mathbb{R}$, $i = 1, \dots, n$, and the sequence $\{u_n\}$ has a weakly convergent subsequence $u_{n'} \rightharpoonup u$ in X as $n \to \infty$, where $u \in X$ is the solution to $Au = f$ from Theorem 2.11.

If A is strictly monotone then the sequence $u_n \rightharpoonup u$ and if A is uniformly monotone then $u_n \to u$.

Proof of Theorem 2.11 and Corollary 3.9. Note that we only prove for X separable. We start by proving the existence of the solution to the Galerkin equation (3.6) and that the sequence of Galerkin approximations converges to the solution of $Au = f$ in X.

1. Set $g(u) = \langle Au - f, u \rangle$ and $g_k = \langle Au - f, v_k \rangle$, $k = 1, \dots, n$. As A is coercive

   $$\frac{g(u)}{\|u\|} \to +\infty \qquad \text{as} \|u\| \to \infty;$$

   therefore, there exists a constant R such that

   $$g(u) > 0 \qquad \text{for all } \|u\| > R. \tag{3.7}$$

   We can write the Galerkin equation (3.6) as

   $$g_k(u_n) = 0, \qquad u_n \in X_n, \quad k = 1, \dots, n, \tag{3.8}$$

   where

   $$u_n = \sum_{k=1}^{n} c_{kn} v_k.$$

   Hence, (3.8) is a nonlinear system of real-valued unknowns $c_{1n}, \dots, c_{nn} \in \mathbb{R}$. As A is demicontinuous (see Lemma 2.3), $u \mapsto g(x)$ is a continuous map on X. Then, for all $u_n \in X_n$ with $\|u_n\| = R$ we have from (3.7)

   $$\sum_{i=1}^{n} g_k(u_n) c_{kn} = \left\langle Au_n - f, \sum_{i=1}^{n} c_{kn} v_k \right\rangle = g(u_n) > 0;$$

   hence, by Corollary 3.8 we have a solution u_n to (3.8).

2. Let u_n be a solution of (3.8), then $g(u_n) = 0$; hence, from (3.7)

$$\|u_n\| \leq R \qquad \text{for all } n.$$

Hence, the sequence $\{u_n\}$ is bounded. If u is a solution of $Au = f$ then $g(u) = 0$; hence,

$$\|u\| \leq R.$$

3. As A is monotone it is also *locally bounded*; i.e., there exists $r, \delta \in \mathbb{R}$ such that

$$\|v\| \leq r \qquad \Longrightarrow \qquad \|Av\| \leq \delta.$$

A is monotone; then

$$\langle Au_n - Av, u_n - v\rangle \geq 0 \qquad \Longrightarrow \qquad \langle Au_n, v\rangle \leq \langle Au_n, u_n\rangle - \langle Av, u_n - v\rangle.$$

By the Galerkin approximation (3.1)

$$\langle Au_n, u_n\rangle = \langle f, u_n\rangle \qquad \text{for all } n;$$

hence,

$$|\langle Au_n, u_n\rangle| \leq \|f\|\|u_n\| \leq R\|f\| \qquad \text{for all } n.$$

By the definition of the norm on the dual space

$$\begin{aligned}
\|Au_n\| &= \sup_{\|v\|=r} \frac{1}{r}\langle Au_n, v\rangle \\
&\leq \sup_{\|v\|=r} \frac{1}{r}\left(\langle Av, v\rangle + \langle Au_n, u_n\rangle - \langle Av, u_n\rangle\right) \\
&\leq \sup_{\|v\|=r} \frac{1}{r}\left(\delta r + R\|f\| + \delta R\right).
\end{aligned}$$

Hence, the sequence $\{Au_n\}$ is bounded.

4. As X is a reflexive Banach space the bounded sequence $\{u_n\}$ has a weakly convergent subsequence (see Proposition 1.8 (4)) $\{u_n\}$; i.e., $u_n \rightharpoonup u$ in X as $n \to \infty$. From the Galerkin approximation

$$\lim_{n\to\infty} \langle Au_n, w\rangle = \langle f, w\rangle, \qquad \text{for all } w \in \bigcup_{n=1}^{\infty} X_n.$$

As $\bigcup_{n=1}^{\infty} X_n$ is dense in X and $\{Au_n\}$ is bounded in X'; then, the above holds for all $x \in X$; i.e., $Au_n \rightharpoonup f$ in X' as $n \to \infty$. Furthermore,

$$\lim_{n\to\infty} \langle Au_n, u_n\rangle = \langle f, u\rangle.$$

By the monotonicity of A

$$\langle Au_n, u_n\rangle - \langle Av, u_n\rangle - \langle Au_n - Av, v\rangle = \langle Au_n - Av, u_n - v\rangle \geq 0.$$

Let $n \to \infty$; then,

$$\begin{aligned} & \langle f, u\rangle - \langle Av, u\rangle - \langle f - Av, v\rangle \geq 0, \\ \implies \qquad & \langle f - Av, u - v\rangle \geq 0 \qquad \text{for all } v \in X. \end{aligned}$$

Let $v = u - tw$, $t > 0$; then, $\langle f - A(u - tw), w\rangle \geq 0$. As A is hemicontinuous, let $t \to 0$; then,

$$\langle f - Au, w\rangle \geq 0 \quad \text{for all } w \in X \qquad \implies \qquad Au = f.$$

Note that in general for A hemicontinuous

$$\langle f - Av, u - v\rangle \geq 0 \quad \text{for all } v \in X \qquad \implies \qquad Au = f. \tag{3.9}$$

So the limit u of the weakly convergent subsequence $\{u_n\}$ is a solution of $Au = f$.

This completes the proof of the existence of the solution to the Galerkin approximations, the existence of the solution to $Au = f$ and the weak convergence of a subsequence of the Galerkin approximations to u.

We now proof the properties of the set of all solutions $\mathcal{S}$ to $Au = f$ for a fixed $f \in X'$. We note that $\mathcal{S}$ is non-empty due to step 4 above. Furthermore, we can show that

1. $\mathcal{S}$ is bounded by step 2 above.
2. $\mathcal{S}$ is convex: Let $u_1, u_2 \in \mathcal{S}$; i.e., $Au_1 = f$, $Au_2 = f$. Define $u = t_1u_1 + t_2u_2$, $0 \leq t_1, t_2 \leq 1$, $t_1 + t_2 = 1$; then,

$$\begin{aligned} \langle f - Av, u - v\rangle &= t_1\langle f - Av, u_1 - v\rangle + t_2 b - Av, u_2 - v \\ &= t_1\langle Au_1 - Av, u_1 - v\rangle + t_2 Au_2 - Av, u_2 - v \\ &\geq 0 \end{aligned}$$

   for all $v \in X$; hence, $Au = f$ by (3.9), which implies $u \in \mathcal{S}$.
3. $\mathcal{S}$ is closed: Let $Av_n = f$ for all n, $v_n \to u$; then,

$$\langle f - Av, u - v\rangle = \lim_{n\to\infty} \langle Av_n - Av, v_n - v\rangle \geq 0 \qquad \text{for all } v \in X.$$

   By (3.9), $Au = f$; hence, $u \in \mathcal{S}$ and $\mathcal{S}$ is closed.

Now assume that A is strictly monotone and $Au_1 = f$, $Au_2 = f$, $u_1 \neq u_2$; then,

$$\begin{aligned} \langle Au_1 - Au_2, u_1 - u_2\rangle &> 0, \\ \langle f - f, u_1 - u_2\rangle &> 0, \\ \langle 0, u_1 - u_2\rangle &> 0. \end{aligned}$$

This is a contradiction; hence, $u_1 = u_2$ must be true and the solution to $Au = f$ is unique. The rest of the proof is left as an exercise. □

We also can prove Theorem 2.31 for existence of solution to $Au = f$ for pseudomonotone operators. To this end, we first study a result about convergence of the Galerkin approximations for operators satisfying (M) or $(\mathrm{S})_0$.

Proposition 3.10. *Let $A : X \to X'$ be a bounded operator satisfying* (M) *on a real, separable, reflexive, and infinite dimensional Banach space X, and $f \in X'$. Let $\{v_1, v_2, \dots\}$ be the basis of X and there exist $R > 0$ and $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$*

$$\langle Au_n - f, v_k \rangle = 0, \qquad u_n \in X_n, \quad k = 1, \dots, n,$$

where $X_n = \operatorname{span}\{v_1, \dots, v_n\}$, has a solution u_n with $\|u_n\| \leq R$. Then,

a) there exists a subsequence $\{u_{n'}\}$ with $u_{n'} \rightharpoonup u$ as $n \to \infty$ such that $u \in X$ is a solution of $Au = f$,

b) if $Au = f$ has a unique solution $u \in X$ then $u_n \rightharpoonup u$ as $n \to \infty$, and

c) if A satisfies $(\mathrm{S})_0$ and is demicontinuous instead of satisfying (M) *then $u_{n'} \to u$ and $u_n \to u$ in steps (a) and (b), respectively.*

Exercise 3.2. Prove step (a) of Proposition 3.10 and the first part of step (c); i.e., $u_{n'} \to u$.

We can now prove Theorem 2.31, but we first recall the theorem along with a result about the existence and convergence of the Galerkin approximation which we prove at the same time.

Theorem 2.31 (Brézis). *Assume that the nonlinear operator $A : X \to X'$ is pseudomonotone, bounded, and coercive on a real, separable, reflexive Banach space; then, the equation*

$$Au = f,$$

has a solution $u \in X$ for every $f \in X'$.

Corollary 3.11. *For a fixed $f \in X'$ and each $n \in \mathbb{N}$ the Galerkin approximation*

$$\langle Au_n - f, v_k \rangle = 0, \qquad u_n \in X_n, \quad k = 1, \dots, n,$$

where $X_n = \operatorname{span}\{v_1, \dots, v_n\} \subset X$ and $A : X \to X'$ satisfies Theorem 2.31, has a solution $u_n \in X_n$. Additionally, there exist a subsequence $\{u_{n'}\}$ such that $u_{n'} \rightharpoonup u$, where $u \in X$ is a solution of $Au = f$.

If A satisfies $(\mathrm{S})_+$, then $u_{n'} \to u$, and if $Au = f$ has a unique solution $u \in X$, then u_n converges (weakly or strongly) to u.

Proof of Theorem 2.31 and Corollary 3.11. By Lemma 2.30 A is demicontinuous and satisfies (M). By identical proof to the steps 1–2 of the proof of Theorem 2.11/Corollary 3.9 the Galerkin approximation has a solution u_n such that $\|u_n\| \leq R$ for all n and fixed $R > 0$. Hence, the conditions of Proposition 3.10 are met, which completes the proof. □

Example 3.2 (Galerkin method for the p-Laplacian). Consider the boundary value problem, in the bounded Lipschitz domain $\Omega \in \mathbb{R}^n$, $n \in \mathbb{N}$,

$$\begin{aligned} -\nabla \cdot (|\nabla u|^{p-2} \nabla u) &= f && \text{in } \Omega \\ u &= 0 && \text{on } \partial\Omega, \end{aligned}$$

where $p \in (1, \infty)$, $f \in L^q(\Omega)$, $1/p + 1/q = 1$. This is similar to Exercise 2.6 and, hence similarly, there exists a unique weak solution $u \in W_0^{k,p}(\Omega)$ such that

$$\langle Au, v \rangle = a(u, v) := \int_\Omega |\nabla u|^{p-2} \nabla u \cdot \nabla v \, \mathrm{d}\boldsymbol{x} = \int_\Omega f v \, \mathrm{d}\boldsymbol{x} =: \langle F, v \rangle$$

for all $v \in W_0^{k,p}(\Omega)$. As per Example 3.1 we consider a finite element method (Galerkin approximation) by partitioning Ω into non-overlapping simplices (elements) T such that $\Omega = \bigcup_{T \in \mathcal{T}_h} T$ to create the mesh $\mathcal{T}_h$, with standard assumptions on the mesh from the theory of finite element methods (e.g., shape regularity). Defining the finite element space

$$X_h := \{v \in C(\overline{\Omega}) : v|_T \in P_1(T) \ \forall T \in \mathcal{T}_h\} \subset W_0^{k,p}(\Omega) =: X$$

(i.e., space of continuous piecewise linear functions) we can define the finite element approximation: find $u_h \in X_h$ such that

$$a(u_h, v_h) = \langle F, v_h \rangle \qquad \text{for all } v_h \in X_h. \tag{3.10}$$

Define the basis of X_h as $\{\phi_i\}_{i=1}^N$, $N = \dim X_h$, (e.g., the hat functions at each interior vertex of the mesh) and

$$u_h = \sum_{j=1}^N \alpha_j \phi_j, \qquad \alpha_j \in \mathbb{R}, j = 1, \ldots, N;$$

then (3.10) is equivalent to

$$\begin{aligned} & a\left(\sum_{j=1}^N \alpha_j \phi_j, \phi_i\right) = \langle F, \phi_j \rangle && i = 1, \ldots, N, \\ \Longrightarrow \quad & \sum_{j=1}^N \alpha_j \int_\Omega \left| \sum_{k=1}^N \alpha_k \phi_k \right|^{p-2} \nabla \phi_j \cdot \nabla \phi_i \, \mathrm{d}\boldsymbol{x} = \int_\Omega f \phi_j && i = 1, \ldots, N. \end{aligned}$$

hence, we need to solve a nonlinear system of algebraic equations for the unknown $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N) \in \mathbb{R}^N$ using nonlinear solvers or linearization of the PDE; cf. Chapter 4. Note that as $h \to 0$, $N \to \infty$ and $u_h \rightharpoonup u$ (as A is strictly monotone). See Barrett and Liu (1993) for more details of the method, including error analysis.

3.4 Potential operator

We want to study another finite dimensional approximation. To this end, we first study potential operators, which are required for the approximation, and provide further useful information for certain problems.

Definition 3.12. Let X be a Banach space; then, the operator $A : X \to X'$ is a *potential operator* if there exists a functional $F \in X'$ such that at each $x \in X$ there exists a Gâteaux derivative F'_G such that

$$\langle Ax, y \rangle = F'_G(x, y) = \lim_{t \to 0} \frac{F(x + ty) - F(x)}{t}$$

for all $x, y \in X$. The functional F is called the *potential* of A. We also denote the Gâteaux derivative by $F' : X \to X'$, where $A = F'$; i.e., $\langle F'x, y \rangle = F'_G(x, y)$.

In order to show properties of potential operators we require an additional definition of continuity.

Definition 3.13. Let X be a Banach space and $A : X \to X'$; then, A is *radially continuous* if the function

$$\varphi_{u,v}(t) := \langle A(u + tv), v\rangle$$

is continuous on $[0, 1]$ for all $u, v \in X$.

Lemma 3.14. *If $A : X \to X'$ is a monotone operator on a Banach space X; then, the following are equivalent:*

- *A is hemicontinuous,*
- *A is demicontinuous,*
- *A is radially continuous,*
- *$\langle f - Av, u - v\rangle \geq 0$ for all $v \in X \implies Au = f$, $f \in X'$,*
- *A satisfies* (M).

We now state various properties about a potential operator $A : X \to X'$ on a real Banach space X, and the solution of $Au = f$, $f \in X'$.

Lemma 3.15. *If $A : X \in X'$ is a radially continuous potential operator with potential F; then, for any $x \in X$,*

$$F(x) = F(0) + \int_0^1 \langle Atx, x\rangle \,\mathrm{d}t.$$

Proof. Choose $x \in X$ and define $\varphi(t) = F(tx)$ for $t \in [0, 1]$; then,

$$\varphi'(t) = \lim_{s\to 0} \frac{1}{s}(F(tx + sx) - F(tx)) = \langle Atx, x\rangle.$$

As A is radially continuous then $\langle Atx, x\rangle$ continuous on $[0, 1]$; hence,

$$F(x) - F(0) = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t) \,\mathrm{d}t = \int_0^1 \langle Atx, x\rangle \,\mathrm{d}t. \qquad \square$$

Theorem 3.16. *Let X be a real Banach space and F a smooth functional (there exists a Gâteaux derivative derivative F' at each point $x \in X$). Then, the potential operator $A = F'$ is monotone if and only if,*

$$F(y) \geq F(x) + \langle Ax, y - x\rangle,$$

for any $x, y \in X$.

Lemma 3.17. *Let $A : X \to X'$ be a potential operator with potential F. Then, for $u \in X$ to be a solution of $Au = f$, $f \in X'$ it is sufficient for*

$$F(u) - \langle f, u\rangle = \min_{v\in X}(F(v) - \langle f, v\rangle), \qquad f \in X'$$

to be fulfilled. If A is monotone; then this condition is necessary.

Proof. Let the functional $g(v) = F(v) - \langle f, v\rangle$, $v \in X$ achieve its minimum at $u \in X$ and for every $v \in X$ we have the Gâteaux derivative $g'(v)$ which satisfies

$$\langle g'(v), h\rangle = \langle F'(v), h\rangle - \langle f, h\rangle \qquad \Longleftrightarrow \qquad g'(v) = F'(v) - f.$$

Then, $g'(u) = 0$ (as its a minimum), and, therefore, for any $h \in X$

$$0 = \langle F'(v), h\rangle - \langle f, h\rangle = \langle Au - f, h\rangle;$$

hence, $Au = f$.

Suppose that A is monotone and $Au = f$. Then, for any $v \in X$, by Theorem 3.16,

$$F(v) \geq F(u) + \langle Au, v - u\rangle;$$

hence,

$$(F(v) - \langle f, v\rangle) - (F(u) - \langle f, u\rangle) = F(v) - F(u) - \langle Au, v - u\rangle \geq 0.$$

Therefore, $u \in X$ gives the minimum of $F(v) - \langle f, v\rangle$. □

Lemma 3.18. *Every monotone potential operator is demicontinuous.*

Lemma 3.19. *Let $A : X \to X'$ be a monotone potential operator; then, $u \in X'$ is a solution to $Au = f$, $f \in X'$, if and only if*

$$\int_0^1 \langle Atu, u\rangle \,\mathrm{d}t - \langle f, u\rangle = \min_{v\in X}\left(\int_0^1 \langle Atv, v\rangle \,\mathrm{d}t - \langle f, v\rangle\right).$$

Theorem 3.20. *Let $A : X \to X'$ be a monotone, coercive, potential operator. Then $Au = f$ has a solution for every right-hand side $f \in X'$. If A is strictly monotone then the solution is unique.*

Proof. This follows from Lemma 3.18 and Minty-Browder (Theorem 2.11) on a Banach space X. □

Corollary 3.21. *The potential*

$$F(x) = \int_0^1 \langle Atx, x\rangle \,\mathrm{d}t, \qquad x \in X,$$

of a monotone, coercive, potential operator $A : X \to X'$ is bounded from below.

Proof. By Theorem 3.20 there exists a solution $u \in X$ to $Au = 0$; then, by Lemma 3.17 $F(u) \leq F(v)$, for all $v \in X$. □

Definition 3.22. Let F be a non-trivial functional on a reflexive Banach space X; then,

$$F^*(x^*) = \sup_{x\in X}(\langle x^*, x\rangle - F(x)), \qquad x \in X',$$

is called the *dual functional* (or *associated*, *adjoint*) to F.

Lemma 3.23. *Let $A : X \to X'$ and let there exist the inverse $A^{-1} : X' \to X$; then,*

$$A \text{ monotone potential operator} \quad \Longleftrightarrow \quad A^{-1} \text{ monotone potential operator}.$$

Theorem 3.24. *Let $A : X \to X'$ be a strictly monotone, coercive, potential operator on a reflexive Banach space X. Then, there exists an inverse $A^{-1} : X' \to X$ which is a strictly monotone potential operator. The functional*

$$F(x) = \int_0^1 \langle Atx, x\rangle \,\mathrm{d}t, \qquad x \in X,$$

is the potential of A and, for any $x \in X$ and $x^* \in X'$,

$$\begin{aligned} F^*(x^*) &= F^*(0) + \int_0^1 \langle x^*, A^{-1}tx^*\rangle \,\mathrm{d}t, \\ F^*(0) &= -F(A^{-1}0), \\ F(x) + F^*(x^*) - \langle x^*, x\rangle &\geq 0, \\ F(x) + F^*(x^*) - \langle Ax, x\rangle &= 0, \end{aligned}$$

where F^* is the potential of A^{-1}.

Corollary 3.25. *Let $A : X \to X'$ be a strictly monotone, coercive, potential operator with potential F. For any $f \in X'$ there exists a unique solution $u \in X$ of $Au = f$ which minimises the potential of the problem $G = F - f$ and*

$$\begin{aligned} G(u) \equiv F(u) - \langle f, u\rangle &= \min_{v\in X} \left(\int_0^1 \langle Atv, v\rangle \,\mathrm{d}t = \langle f, v\rangle \right) \\ &= -\int_0^1 \langle f, A^{-1}tf\rangle \,\mathrm{d}t + \int_0^1 \langle AtA^{-1}0, A^{-1}0\rangle \,\mathrm{d}t. \end{aligned}$$

3.5 Ritz method

We now look at another finite dimensional approximation of an operator $A : X \to X'$ on a real reflexive Banach space X.

Definition 3.26. Let $A : X \to X'$ be a potential operator with potential F, and $\{h_i\}_{i\geq 1}$ be a dense set in X; then $u_n \in X_n$ is called the n-th *Ritz approximation* of $Au = f$, $f \in X'$, if it holds that

$$F(u_n) - \langle f, u_n\rangle = \min_{v\in X_n} (F(v) - \langle f, v\rangle),$$

where $X_n = \operatorname{span}\{h_i, i = 1, \ldots, n\}$.

From the theory of linear operator equations it is known that the Galerkin and Ritz approximation for symmetric non-negative operators given the same approximation under certain conditions. Can we show a similar result for nonlinear operator?

Theorem 3.27. *Let $A : X \in X'$ be a monotone potential operator with potential F. Then, $u_n \in X_n$ is the Ritz approximation of $Au = f$, $f \in X'$, if and only if u_n is the Galerkin approximation in X_n; i.e, $A_n u_n = f_n$, $n \in \mathbb{N}$, where f_n is the projection of f onto X_n.*

Proof. Let u_n be the Ritz approximation. Then, for any $t \in \mathbb{R}$ and $h \in X_n$, set $v = u_n + th$ and by the definition of the Ritz approximation (minimisation)

$$F(u_n + th) - F(u_n) \geq t\langle f, h\rangle;$$

hence, for any $h \in X_n$,

$$0 = \lim_{t \to 0} \frac{1}{t}(F(u_n + th) - F(u_n)) - \langle f, h \rangle = \langle Au_n - f, h \rangle.$$

This implies that $\langle Au_n, h \rangle = \langle f, h \rangle$ for all $h \in X_n$, which is the Galerkin approximation.

Let $u_n \in X_n$ be the Galerkin approximation. Then, as A is monotone by Theorem 3.16

$$F(y) \geq F(x) + \langle Ax, y - x \rangle \qquad \text{for all } x, y \in X;$$

hence, for any $v \in X_n$, setting $y = v$ and $x = u_n$,

$$\begin{aligned} & (F(v) - \langle f, v \rangle) - (F(u_n) - \langle f, u_n \rangle) \geq \langle Au_n - f, v - u_n \rangle = 0; \\ \implies \qquad & F(u_n) - \langle f, u_n \rangle \leq F(v) - \langle f, v \rangle \qquad \forall v \in X_n. \end{aligned}$$

So, $u_n \in X_n$ minimises $F(v) - \langle f, v \rangle$ and, hence, is the Ritz approximation. □

Theorem 3.28. *Let $A : X \to X'$ be a strictly monotone, coercive, potential operator which satisfies* (S). *Then, $Au = f$ has a unique solution for every right hand side $f \in X'$ to which the sequence $\{u_n\}$ of Ritz approximations converges.*

3.6 Variational problems & quasilinear PDEs

We now look at why the potential operator and the Ritz approximation are useful. Consider the variational problem

$$\min_{u \in X} F(u), \qquad F(u) := \int_\Omega L(\boldsymbol{x}, \delta_k u(\boldsymbol{x}))\, \mathrm{d}\boldsymbol{x},$$

where

$$\begin{aligned} & X := \{u \in C^{2k}(\overline{\Omega}) : \partial^\alpha u = 0 \text{ on } \partial\Omega\ \forall \alpha, |\alpha| \leq k - 1\}, \\ & L : \Omega \times \mathbb{R}^\kappa \to \mathbb{R}; \end{aligned}$$

see, for example, Example 1.1.

We can show that for $u \in X$ to be a solution it is necessary for u to satisfy the Euler equation:

$$\begin{aligned} \sum_{|\alpha|} (-1)^{|\alpha|} \partial^\alpha L_\alpha(\boldsymbol{x}, \delta_k u(\boldsymbol{x})) &= 0 && \text{in } \Omega, \\ \partial^\alpha u &= 0 && \text{on } \partial\Omega \text{ for all } \alpha, |\alpha| \leq k - 1, \end{aligned}$$

where $L_\alpha : \Omega \times \mathbb{R}^\kappa \to \mathbb{R}$ is the derivative of L with respect to $\partial^\alpha u$. This is the divergence form of a quasilinear PDE.

Example 3.3 (Variational problem as second-order quasilinear PDE). Consider, for $\Omega \in \mathbb{R}^N$, $p \geq 2$, the variational problem

$$\min_u \int_\Omega \left(p^{-1} \sum_{i=1}^N \left| \frac{\partial u}{\partial x_i} \right|^p + g(u) - f(\boldsymbol{x}) \right) \mathrm{d}\boldsymbol{x}, \qquad \text{such that } u = 0 \text{ on } \partial\Omega. \tag{3.11}$$

Let $u \in C^2(\overline{\Omega})$ be the solution of (3.11); then, define the boundary value problem

$$-\sum_{i=1}^{N} \frac{\partial}{\partial x_i}\left|\frac{\partial u}{\partial x_i}\right|^{p-2}\frac{\partial u}{\partial x_i} + g'(u) = f \qquad \text{in } \Omega, \tag{3.12}$$

$$u = 0, \qquad \text{on } \partial\Omega. \tag{3.13}$$

Let

$$F_1(u) = \int_\Omega p^{-1}\sum_{i=1}^{N}\left|\frac{\partial u}{\partial x_i}\right|^p \mathrm{d}\boldsymbol{x}, \qquad F_2(u) = \int_\Omega g(u)\,\mathrm{d}\boldsymbol{x},$$

$X = W_0^{k,p}(\Omega)$, $f \in L^q(\Omega)$, $1/p + 1/q = 1$. We know that there exists a $b \in X'$ such that

$$\langle b, u\rangle = \int_\Omega f u\,\mathrm{d}\boldsymbol{x} \qquad \text{for all } u \in X.$$

We can write (3.11) as

$$\min_{u\in X}(F(u) - \langle b, u\rangle),$$

where $F = F_1 + F_2$. We have that

$$\langle F'u, h\rangle := \int_\Omega \left(\sum_{i=1}^{N}\left|\frac{\partial u}{\partial x_i}\right|^{p-2}\frac{\partial u}{\partial x_i}\frac{\partial h}{\partial x_i} + g'(u)h\right)\mathrm{d}\boldsymbol{x};$$

hence, (3.12)–(3.13) corresponds to the weak formulation

$$\langle F'u, h\rangle = \langle b, h\rangle \qquad \text{for all } h \in X,$$

with Ritz approximation

$$\langle F'u_n, v_i\rangle = \langle b, v_i\rangle, \qquad i = 1, \dots, n,$$

for $u_n \in X_n = \operatorname{span}\{v_1, \dots, v_n\}$, where $v_1, v_2, \dots$ form a basis of X. Assuming that $g \in C^1(\mathbb{R})$ satisfies, for all $u \in \mathbb{R}$,

$$g(u) \ge -C_1 u - C_2,$$
$$|g(u)| \le C_3(1 + |u|^p),$$
$$|g'(u)| \le C_4(1 + |u|^{p-1}),$$

where C_1, C_2, C_3, and C_4 are constants. Then, a solution $u \in X$ to (3.11) exists and satisfies (3.12)–(3.13).

We look at the general variational problem,

$$\min_u \left(\int_\Omega L(\boldsymbol{x}, \delta_k u(\boldsymbol{x}))\,\mathrm{d}\boldsymbol{x} - \int_\Omega f u\,\mathrm{d}\boldsymbol{x}\right) \qquad \partial^\alpha u = 0 \text{ on } \partial\Omega, \forall\alpha, |\alpha| \le k-1,$$

which can be considered as the boundary value problem

$$\sum_{|\alpha|}(-1)^{|\alpha|}\partial^\alpha a_\alpha(\boldsymbol{x}, \delta_k u(\boldsymbol{x})) = f(\boldsymbol{x}) \qquad \text{in } \Omega, \tag{3.14}$$

$$\partial^\alpha u = 0 \qquad \text{on } \partial\Omega \text{ for all } \alpha, |\alpha| \le k-1, \tag{3.15}$$

with $a_\alpha = L_\alpha$ for all α, $|\alpha| \le k$.

This is identical to the problems studied in Chapter 2; so when are the problems we studied in that chapter variational problems? We assume that for a smooth L that $L_{\alpha\beta} = L_{\beta\alpha}$ for all multi-indices α, β, where $L_{\alpha\beta}$ is the derivative of L with respect to $\partial^\alpha u$ and then $\partial^\beta u$. Therefore, it follows that

$$\frac{\partial a_\alpha(\boldsymbol{x}, \xi)}{\partial \xi_\beta} = \frac{\partial a_\beta(\boldsymbol{x}, \xi)}{\partial \xi_\alpha} \qquad \text{for all } |\alpha|, |\beta| \le k. \tag{3.16}$$

So, if $a_\alpha \in C^1(()\overline{\Omega} \times \mathbb{R}^\kappa)$ and (3.16) holds, then there exists a functional $L : \Omega \times \mathbb{R}^\kappa \to \mathbb{R}$ such that $a_\alpha = L_\alpha$, for all α, $|\alpha| \le k$ on $\overline{\Omega} \times \mathbb{R}^\kappa$ if Ω is a simply connected domain in $\mathbb{R}^\kappa$. Therefore, (3.14)–(3.15) is a variational problem if (3.16) holds.

In order to consider in a similar way to Example 3.3 we assume that we can write

$$L(\boldsymbol{x}, \xi) = L^{(1)}(\boldsymbol{x}, \xi) + L^{(2)}(\boldsymbol{x}, \xi)$$

where the following assumptions hold for all $(\boldsymbol{x}, \xi) \in \Omega \times \mathbb{R}^\kappa$:

(L1) Ω is a bounded region on $\mathbb{R}^n$, $n \ge 1$, $1 < p < \infty$, $1/p + 1/q = 1$, $k \ge 1$, $X = W_0^{k,p}(\Omega)$ and $f \in L^q(\Omega)$.

(L2) *Growth condition*: $L \in C(\overline{\Omega} \times \mathbb{R}^\kappa)$ and

$$|L(\boldsymbol{x}, \xi)| \le C_1 \left(|a_1(\boldsymbol{x})| + \sum_{|\alpha| \le k} |\xi_\alpha|^p \right),$$

where $C_1 > 0$ constant and $a_1 \in L^1(\Omega)$. Additionally, this should hold for $L^{(1)}$ and $L^{(2)}$.

(L3) *Coerciveness condition*:

$$L(\boldsymbol{x}, \xi) \ge C_2 \sum_{|\alpha| \le k} |\xi_\alpha|^p - C_3 \xi_0 - a_2(\boldsymbol{x}),$$

where $C_2, C_3 > 0$ constant and $a_2 \in L^1(\Omega)$.

(L4) *Convexity*: $\xi \mapsto L^{(1)}(\boldsymbol{x}, \xi)$ convex on $\mathbb{R}^\kappa$ for all $x \in \Omega$.

(L5) *Degenerate perturbation*: $L^{(2)}$ depends only on $\boldsymbol{x}$ and partial derivatives of order less than or equal to $k - 1$.

(L6) *Growth condition*:

$$|L_\alpha(\boldsymbol{x}, \xi)| \le C_4 \left(|b(\boldsymbol{x})| + \sum_{|\alpha| \le k} |\xi_\alpha|^{p-1} \right),$$

where $C_4 > 0$ constant and $b \in L^q(\Omega)$.

(L7) *Uniform monotonicity condition*:

$$\sum_{|\alpha|\le k}\left(L^{(1)}_\alpha(\boldsymbol{x},\xi)-L^{(1)}_\alpha(\boldsymbol{x},\eta)\right)(\xi_\alpha-\eta_\alpha)\ge C_5\sum_{|\alpha|=k}|\xi_\alpha-\eta_\alpha|^p,$$

for all $\xi,\eta\in\mathbb{R}^\kappa$, $\boldsymbol{x}\in\Omega$, with constant $C_5>0$.

We note these are similar to the assumptions from Chapter 2.

Lemma 3.29. *We note that (L4) can be simplified according to the properties of $L^{(1)}$.*

1. *If $L^{(1)}\in C^1(\overline{\Omega}\times\mathbb{R}^\kappa)$ then (L4) is equivalent to*

$$\sum_{|\alpha|\le k}\left(L^{(1)}_\alpha(\boldsymbol{x},\xi)-L^{(1)}_\alpha(\boldsymbol{x},\eta)\right)(\xi_\alpha-\eta_\alpha)\ge 0,$$

for all $\xi,\eta\in\mathbb{R}^\kappa$, $\boldsymbol{x}\in\Omega$; i.e., the same as the monotonicity condition (C1) from Chapter 2.

2. *If $L^{(1)}\in C^2(\overline{\Omega}\times\mathbb{R}^\kappa)$ then (L4) is equivalent to*

$$\sum_{|\alpha|\le k}\sum_{|\beta|\le k}L^{(1)}_{\alpha\beta}(\boldsymbol{x},\xi)\eta_\alpha\eta_\beta\ge 0,$$

for all $\xi,\eta\in\mathbb{R}^\kappa$, $\boldsymbol{x}\in\Omega$; i.e., the eigenvalues of the symmetric Hessian matrix $(L^{(1)}_{\alpha\beta})$ are non-negative for all $(\boldsymbol{x},\xi)\in\Omega\times\mathbb{R}^\kappa$.

We can define

$$F_j(u)=\int_\Omega L^{(j)}(\boldsymbol{x},\delta_k u(\boldsymbol{x}))\,\mathrm{d}\boldsymbol{x},\qquad F=F_1+F_2,\qquad \langle b,u\rangle=\int_\Omega f u\,\mathrm{d}\boldsymbol{x},$$

and consider the generalised variational problem

$$\min_{u\in X}(F(u)-\langle b,u\rangle)\tag{3.17}$$

which can be considered as the solution of the boundary value problem

$$\langle F'(u),h\rangle=\langle b,h\rangle\qquad\text{for all } h\in X,\tag{3.18}$$

where

$$\langle F'(u),h\rangle=\int_\Omega\sum_{|\alpha|\le k}L_\alpha(\boldsymbol{x},\delta_k u(\boldsymbol{x}))\partial^\alpha h\,\mathrm{d}\boldsymbol{x}.$$

Theorem 3.30. *If (L1)–(L5) hold; then, the variational problem (3.17) has a solution $u\in X$. Additionally, if (L6) holds, then the continuous derivatives F_1', F_2', and F' exist and $u\in X$ satisfies (3.18). Furthermore, F_1' is monotone amd F_2' is strongly continuous. If $L^{(2)}\equiv 0$; then, (3.17) and (3.18) are equivalent.*

Corollary 3.31. *If (L1)–(L7) hold and $L^{(2)} \equiv 0$; then, (3.17) and (3.18) has a unique solution $u \in X$ and the sequence $\{u_n\}$ obtained by the Ritz approximation*

$$\langle F'(u_n), v_i \rangle = \langle b, v_i \rangle, \qquad i = 1, \ldots, n,$$

for $u_n \in X_n = \operatorname{span}\{v_1, \ldots, v_n\}$, with $v_1, v_2, \ldots$ forming a basis of X, converges strongly to $u \in X$. Furthermore, F' is uniformly monotone in the sense that there exists a constant $C > 0$ such that

$$\langle F'(u) - F'(v), u - v \rangle \geq C \|u - v\|^p$$

for all $u, v \in X$.

Remark. We note that the definition of uniformly monotone in Corollary 3.31 is the same as in Definition 2.1 with $a(\|u - v\|) = C\|u - v\|^{p-1}$, as $p > 1$.

CHAPTER 4

Linearisation & Iterative Methods

In order to find a solution, or a Galerkin approximation of the solution, to a nonlinear partial differential equation it is necessary to consider linearised versions of the the problem, and then use iterative methods to converge the nonlinear solution; see, for example, the Section 3.2.

4.1 Kačanov method

We first look at a classical linearisation technique by means of an example.

Example 4.1 (Kačanov method for conservation law (Zeidler, 1989a, Section 25.13)). Consider the conservation law equation

$$\begin{aligned} -\nabla(\mu(|\nabla u|^2)\nabla u) &= f && \text{in } \Omega, \\ u &= g && \text{on } \Gamma_D, \\ -\mu(|\nabla u|^2)\frac{\partial u}{\partial \boldsymbol{n}} &= h && \text{on } \Gamma_N, \end{aligned}$$

where $\partial\Omega = \overline{\Gamma}_D \cup \overline{\Gamma}_N$. On the 1950s engineers started using the following iteration method: Given an initial guess $u^{(0)}$, such that $u^{(0)} = g$ on Γ_D, solve for $m = 0, 1, \ldots$

$$\begin{aligned} -\nabla(\mu(|\nabla u^{(m)}|^2)\nabla u^{(m+1)}) &= f && \text{in } \Omega, \\ u &= g && \text{on } \Gamma_D, \\ -\mu(|\nabla u^{(m)}|^2)\frac{\partial u^{(m+1)}}{\partial \boldsymbol{n}} &= h && \text{on } \Gamma_N. \end{aligned}$$

This is called a *Kačanov iteration*. This is a linear partial differential equation in the unknown $u^{(m)}$. We note that this has a simple physical interpretation which justifies the linearisation: if u represents temperature; then

$$j = -\mu(|\nabla u|^2)\nabla u$$

is the current density vector of the heat flow and μ is the heat conductivity of the material. Hence, the unknown approximation $u^{(m+1)}$ is determined by the heat conductivity $\mu(|\nabla u^{(m)}|^2)$ corresponding to the known approximation $u^{(m)}$. This method corresponds to a variational problem

$$\min_u \left(\int_\Omega (\beta(|\nabla u|) - fu)\, \mathrm{d}\boldsymbol{x} + \int_{\Gamma_N} hu \,\mathrm{d}s \right), \qquad u = g \text{ on } \Gamma_D,$$

and the iteration method corresponds to the quadratic variational problem for $u^{(m+1)}$

$$\min_{u^{(m+1)}} \left(\int_\Omega \left(\frac{1}{2}\mu(|\nabla u^{(m)}|^2)|\nabla u^{(m+1)}| - f u^{(m+1)} \right) \mathrm{d}\boldsymbol{x} + \int_{\Gamma_N} h u \,\mathrm{d}s \right), \qquad u^{(m+1)} = g \text{ on } \Gamma_D.$$

Set $X = \{u \in W^{1,2}(\Omega) : u = g \text{ on } \Gamma_D\}$; then, these can be written as

$$\min_{u \in X} \left(F(u) - b(u)\right), \tag{4.1}$$

$$\min_{u^{(m+1)} \in X} \left(a(u^{(m)}; u^{(m+1)}, u^{(m+1)}) - b(u^{(m+1)}) \right), \tag{4.2}$$

respectively, where

$$\begin{aligned} b(u) &= \int_\Omega f u \,\mathrm{d}\boldsymbol{x} - \int_{\Gamma_N} h u \,\mathrm{d}s, \\ \beta(s) &= \frac{1}{2}\int_0^{s^2} \mu(t)\,\mathrm{d}t, \\ F(u) &= \int_\Omega \beta(|\nabla u|)\,\mathrm{d}\boldsymbol{x}, \\ a(u; v, w) &= \int_\Omega \mu(|\nabla u|^2)\nabla v \cdot \nabla w \,\mathrm{d}\boldsymbol{x}. \end{aligned}$$

We note that a is (bi)linear in the final two arguments. We assume that Ω is a bounded region in $\mathbb{R}^n$ such that$\partial\Omega = \overline{\Gamma}_D \cup \overline{\Gamma}_N$, $\Gamma_D \neq \emptyset$, $f \in \left[W^{1,2}(()\Omega)\right]'$, $g \in \left[W^{1/2,2}(\Gamma_D)\right]'$, $h \in \left[W^{1/2,2}(\Gamma_N)\right]'$ (e.g., $f \in L^2(\Omega)$, $g \in W^{1,2}(\Omega)$, $h \in L^2(\Gamma_N)$), $\mu : \mathbb{R}_+ \to \mathbb{R}_+$ is C^1, and there exists positive constants a, c, d such that

$$a \le \mu(s) \le c, \qquad \mu'(s) \ge 0, \qquad \beta''(s) = \mu'(s^2)s^2 + \mu(s^2) \ge d, \qquad \text{for all } s \ge 0.$$

Then, (4.1) has a unique solution; cf., Zeidler (1989a, Section 25.13). If $u^{(0)} \in X$ is given, then for $m = 0, 1, \dots$ (4.2) has a unique solution $u^{(m+1)}$ and the Kačanov iteration converges:

$$u^{(m)} \to u \in W^{1,2}(\Omega) \qquad \text{as } n \to \infty.$$

Let $\Gamma_N = \emptyset$ and $g = 0$ for simplicity; then, we can define the weak formulation of the method and iteration:

$$\begin{aligned} &\text{Find } u \in X \text{ such that} & \langle Au, v\rangle &= \langle b, v\rangle & &\text{for all } v \in X, \\ &\text{Find } u^{(m+1)} \in X \text{ such that} & \langle L[u^{(m)}]u^{(m+1)}, v\rangle &= \langle b, v\rangle & &\text{for all } v \in X, \end{aligned}$$

where $A : X \to X'$ and $L[\overline{u}] : X \to X'$ are defined by

$$\begin{aligned} \langle Au, v\rangle &= a(u; u, v), \\ \langle L[\overline{u}]u, v\rangle &= a(\overline{u}; u, v), \end{aligned}$$

for all $\overline{u}, u, v \in X$. Note that $L[\overline{u}]$ is a linearisation of A at $\overline{u}$ such that $L[u]u = Au$. Similarly we can compute the Galerkin approximations on a finite dimensional subspace $X_n \subset X$:

$$\begin{aligned} &\text{Find } u_n \in X \text{ such that} & \langle Au_n, v\rangle &= \langle b, v\rangle & &\text{for all } v \in X_n, \\ &\text{Find } u_n^{(m+1)} \in X \text{ such that} & \langle L[u_n^{(m)}]u_n^{(m+1)}, v\rangle &= \langle b, v\rangle & &\text{for all } v \in X_n. \end{aligned}$$

Prove that u_n converges to u and $u_n^{(m)}$ converges to $u^{(m)}$ follow from standard nonlinear and linear results.

4.2 Newton method

Let X and Y be two Banach spaces. Given an open subset $\Omega \subset X$ and a nonlinear operator $A : \Omega \to Y$ we consider the nonlinear operator equation $Au = 0$ for some unknown solutions $u \in \Omega$. Let A' be the Fréchet derivative of A in Ω (or a suitable subset of Ω); then, the classical *Newton method* is given by the following: For an initial guess $u^{(0)} \in X$ the we define the iteration

$$u^{(m+1)} = u^{(m)} + \delta^{(m)}, \qquad m \geq 0, \tag{4.3}$$

where the update $\delta^{(m)}$ is given as the solution of the linear equation

$$A'(u^{(m)})\delta^{(m)} = -Au^{(m)}, \qquad m \geq 0.$$

For this to be well-defined $A'(u^{(m)})$ must be invertible for all $m \geq 0$ and $\{u^{(m)}\}_{m\geq 0} \subset \Omega$.

Let $B_r(x) := \{z \in X : \|x - z\|_X < r\}$ be the open ball in X with centre x and radius r. Then, we state a *restricted* convergence result for the Newton method.

Theorem 4.1 (Kantorovich). *Let $\Omega_0 \subset \Omega$ be an open convex subset, A be Fréchet differentiable with*

$$\|A'(x) - A'(y)\| \leq L\|x - y\| \qquad x, y \in \Omega_0.$$

Assume that $u^{(0)} \in \Omega_0$ is such that

1. $\left[A'(u^{(0)})\right]^{-1}$ *exists and there exists constants β, η such that*

$$\left\|\left[A'(u^{(0)})\right]^{-1}\right\| \leq \beta,$$
$$\left\|\left[A'(u^{(0)})\right]^{-1}(Au^{(0)})\right\| \leq \eta,$$
$$h = BL\eta \leq \frac{1}{2}$$

2. $B_r(u^{(0)}) \subset \Omega_0$ *where*

$$r = \left(\frac{1 - \sqrt{1 - 2h}}{h}\right)\eta.$$

Then, the Newton iterates $u^{(m)}$ defined by (4.3) is well-defined, $u^{(m)} \in B_r(u^{(0)})$ for $m \geq 0$, and (4.3) converges to $u^* \in \overline{B_r(u^{(0)})}$ such that $Au^* = 0$. Furthermore,

$$\|u^* - u_n\| \leq \frac{\eta}{h}\left(\frac{(1 + \sqrt{1 - 2h})^{2^m}}{2^m}\right), \qquad m \geq 0.$$

Providing $A'(u)$ is invertible on a suitable subset of Ω we can define the *Newton transform* by

$$u \mapsto N(u) = -\left[A'(u)\right]^{-1}(Au)$$

.

To improve the reliability of the method when the initial guess $u^{(0)}$ is far-away from the root u^* of $Au^* = 0$ we can introduce a damping parameter $\varepsilon_m \in (0, 1]$ that may be adjusted adaptively at each iteration step to give the *damped Newton* iteration:

$$u^{(m+1)} = u^{(m)} + \varepsilon_m N(u^{(m)}), \qquad m \geq 0. \tag{4.4}$$

The trick is to find ε_m at each step of the iteration such that $\|A(u^{(m)} + \varepsilon_m N(u^{(m)})\|$ is sufficiently smaller than $\|Au^{(m)}\|$. We can guarantee this if $A \in C^1(X)$, as then we have that

$$A(u^{(m)} + \varepsilon_m N(u^{(m)})) \approx (1 - \varepsilon_m)Au^{(m)} + o(\varepsilon_m)$$

for sufficiently small ε_m. Determining the step size ε_m is called the *damping process*. We consider a simple strategy for this process.

Algorithm 4.1 (Incremental damping)**.** For parameters $0 < I < 1, 0 < \alpha < 1, 0 < M < 1$, and $0 < \varepsilon_0 \leq 1$ we define the damped Newton method with incremental damping as follows:

Start the Newton iteration with initial guess $u^{(0)} \in \Omega$
for $m = 0, 1, \ldots$ **do**
  Compute $u^{(m+1)}$ based on the damped Newton iteration (4.4)
  while $\|Au^{(m+1)}\| > \|Au^{(m)}\|$ **do**
    $\varepsilon_m \leftarrow \alpha\varepsilon_m$
    if $\varepsilon_m < M$ **then**
      $\varepsilon_m \leftarrow M$
      break
    end if
    Compute $u^{(m+1)}$ based on the damped Newton iteration (4.4)
  end while
  $\varepsilon_{m+1} \leftarrow \min(1, \varepsilon_m + I)$
end for

Remark. The parameters in this algorithm have the following meaning:

α : Value to multiply the damping parameter by to *decrease* the damping
M : The minimum value for the damping parameter
I : The amount to increase the damping parameter by at each iteration

Remark. In the algorithm we set ε_m to the minimum of 1 and the computed value. This ensures that ε_m is chosen as 1 wherever possible — giving the standard (un-damped) Newton method (4.3). This ensures optimal convergence close to a simple root.

Remark. In the above algorithm we have iterated for $m = 0, 1, \ldots$. In practice, we can use the fact that $Au^{(m)} \approx 0$ when $u^{(m)}$ is close to a root and instead use a stopping condition

$$\|Au^{(m)}\| < \texttt{TOL},$$

where `TOL` is a desired tolerance. Additionally, a maximum number of iterations is often specified in case of no convergence.

We note other methods for adaptively setting ε_m exist; see, for example, Amrein and Wihler (2014, Algorithm 2.1) and Amrein and Wihler (2015, Algorithm 2.4).

Example 4.2 (Newton method for singularly perturbed PDE (Amrein and Wihler, 2015)). We consider a singularly perturbed semilinear elliptic PDE: Find $u : \Omega \to \mathbb{R}$ which satisfies

$$\begin{aligned} -\varepsilon\Delta u &= f(u) && \text{in } \Omega \\ u &= 0 && \text{on } \partial\Omega \end{aligned}$$

where $\Omega \subset \mathbb{R}^d$ is an open bounded Lipschitz domain, $\epsilon > 0$ (potentially $\epsilon \ll 1$) is a fixed parameter. and $f : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function. We suppose that there exists a (non-unique) solution $u \in X := H_0^1(\Omega)$ and define $A_\epsilon : X \to X'$ as

$$\langle A_\epsilon u, v\rangle := \int_\Omega (\varepsilon \nabla u \cdot \nabla v)\,\mathrm{d}\boldsymbol{x} \qquad \text{for all } v \in X.$$

We can write the problem as a nonlinear operator equation in X': Find $u \in X$ such that $A_\varepsilon u = 0$. We note that the Fréchet/Gâteaux derivative (there are the same) of A_ε at $u \in X$ is defined by

$$\langle A'_\varepsilon(u)w, v\rangle := \int_\Omega (\varepsilon \nabla w \cdot \nabla v - f'(u)wv)\,\mathrm{d}\boldsymbol{x} \qquad \text{for } v, w \in X.$$

It is possible to show that $A_\varepsilon u$ is a well-defined linear and bounded map from X to X'; cf. Amrein and Wihler (2015). Hence, given an initial guess $u^{(0)} \in X$ we an define the damped Newton method

$$A'_\varepsilon(u^{(m)})u^{(m+1)} = A'_\varepsilon(u^{(m)})u^{(m)} - \varepsilon_m A_\varepsilon u^{(m)}$$

in X'. Equivalently,

$$a_\varepsilon(u^{(m)}; u^{(m+1)}, u^{(m)}) = a_\varepsilon(u^{(m)}; u^{(m)}, u^{(m)}) - \varepsilon_m \ell_\varepsilon(u^{(m)}; v) \tag{4.5}$$

for all $v \in X$, where for fixed $u \in X$

$$a_\varepsilon(u; w, v) = \int_\Omega (\varepsilon \nabla w \cdot \nabla v - f'(u)wv)\,\mathrm{d}\boldsymbol{x}$$
$$\ell_\varepsilon(u; v) = \int_\Omega (\varepsilon \nabla u \cdot \nabla v - f(u)v)\,\mathrm{d}\boldsymbol{x}$$

are bilinear and linear forms on $X \times X$ and X, respectively.

If there exists positive constants $\underline{\lambda}, \overline{\lambda}$ with $\varepsilon C_P^{-2} > \overline{\lambda}$, such that $-\underline{\lambda} \leq f'(u) \leq \overline{\lambda}$ for all $u \in \mathbb{R}$, where C_P is the Poincáre constant ($\|w\|_{0,2} \leq C_P \|\nabla w\|_{0,2}$); then, for any fixed $u^{(m)} \in X$ it is possible to show that $a_\varepsilon(u^{(m)}; \cdot, \cdot)$ is a continuous, coercive, bilinear form on $X \times X$ and $a_\varepsilon(u^{(m)}; u^{(m)}, \cdot) - \varepsilon_m \ell_\varepsilon(u^{(m)}; \cdot)$ is a continuous linear form on X. Hence, by Lax-Milgram the damped Newton method (4.5) has a unique solution $u^{(m+1)} \in X$.

We now construct a finite element discretisation of this Newton iteration to get a Newton-Galerkin approximation scheme. To this end, we partition the domain Ω into intervals/triangles T such that $\Omega = \bigcup_{T \in \mathcal{T}_h} T$ to create the mesh $\mathcal{T}_h$, and define the finite element space

$$X_h := \{v \in H_0^1(\Omega) : v|_T \in P_1(T)\ \forall K \in \mathcal{T}_h\} \subset X.$$

Then, we consider the finite element formulation: Find $u_h^{(m+1)} \in X_h$ such that

$$a_\varepsilon(u_h^{(m)}; u_h^{(m+1)}, v_h) = a_\varepsilon(u_h^{(m)}; u_h^{(m)}, v_h) - \varepsilon_m \ell_\varepsilon(u_h^{(m)}; v_h) \tag{4.6}$$

for all $v_h \in X_h$, with given initial guess $u_h^{(0)} \in X_h$. We can change this into a linear system by defining

$$u_h^{(m)} = \sum_{j=1}^N \alpha_j^{(m)} \phi_j, \qquad u_h^{(m+1)} = \sum_{j=1}^N \alpha_j^{(m+1)} \phi_j,$$

where ϕ_j, $j = 1, \dots, N = \dim X_h$, are the basis functions of X_h (more exactly, hat functions at each interior vertex of the mesh). Then, (4.6) can be written as

$$\sum_j^N a_\varepsilon(u_h^{(m)}; \phi_j, \phi_i)\alpha_j^{(m+1)} = \sum_j^N a_\varepsilon(u_h^{(m)}; \phi_j, \phi_i)\alpha_j^{(m)} - \varepsilon_m \ell_\varepsilon(u_h^{(m)}; \phi_i) \qquad \text{for } i = 1, \dots, N.$$

This can be written as

$$\mathbb{J}^{(m)}\boldsymbol{\alpha}^{(m+1)} = \mathbb{J}^{(m)}\boldsymbol{\alpha}^{(m)} - \varepsilon_m \mathcal{R}^{(m)}, \tag{4.7}$$

where the *Jacobian* matrix $\mathbb{J}^{(m)} \in \mathbb{R}^{N\times N}$ and *residual vector* $\mathcal{R}^{(m)} \in \mathbb{R}^N$, dependent on $u_h^{(m)}$, are defined as

$$\begin{aligned} \mathbb{J}_{ij}^{(m)} &\coloneqq a_\varepsilon(u_h^{(m)}; \phi_j, \phi_i), && i, j = 1, \dots, N, \\ \mathcal{R}_i^{(m)} &\coloneqq \ell_\varepsilon(u_h^{(m)}; \phi_i), && i = 1, \dots, N. \end{aligned}$$

Hence, the iteration (4.7) can be changed to

$$\boldsymbol{\alpha}^{(m+1)} = \boldsymbol{\alpha}^{(m)} + \varepsilon_m \boldsymbol{\delta}^{(m)}, \tag{4.8}$$

where $\boldsymbol{\delta}^{(m)} \in \mathbb{R}^N$ is the solution the linear system

$$\mathbb{J}^{(m)}\boldsymbol{\delta}^{(m)} = -\mathcal{R}^{(m)}. \tag{4.9}$$

We now consider the iterative process of the damped Newton method, for example in Algorithm 4.1. At each step m it is only necessary to construct the matrix $\mathbb{J}^{(m)}$ and residual vector $\mathcal{R}^{(m)}$, and compute the solution of the linear system (4.9), only once and then computing $u_h^{(m+1)}$ only requires evaluating (4.8). Additionally, evaluating the norm $\|A_\varepsilon u_h^{(m)}\|$ can be replaced by evaluating $\|\mathcal{R}^{(m)}\|$ in some vector norm (such as the vector 2-norm) as

$$\|A_\varepsilon u_h^{(m)}\| = \sup_{v\in X_h} \frac{\langle A_\varepsilon u_h^{(m)}, v\rangle}{\|v\|} \qquad \text{and} \qquad \mathcal{R}^{(m)} = \Big(\ell_\varepsilon(u_h^{(m)}; \phi_i)\Big)_{i=1}^N = \Big(\langle A_\varepsilon u_h^{(m)}, \phi_i\rangle\Big)_{i=1}^N.$$

4.3 Iterative linearised Galerkin method

We now consider a more generalised iterative method. Consider the operator equation $Au = f$ where $A : X \to X'$, $u \in X$, and $f \in X'$; then, we construct a general fixed point iterative as follows. For a fixed $v \in X$, define a *linear* invertible operator $L[v] : X \to X'$, and the fixed point equation

$$u = u - L[u]^{-1}(Au - f).$$

Then, for an initial guess $u^{(0)} \in X$, we can propose the iterative scheme

$$u^{(m+1)} = u^{(m)} - L[u^{(m)}]^{-1}(Au^{(m)} - f), \qquad m \geq 0.$$

Equivalently,

$$L[u^{(m)}]u^{(m+1)} = L[u^{(m)}]u^{(m)} - (Au^{(m)} - f), \qquad m \geq 0. \tag{4.10}$$

This is a linear scheme for finding $u^{(m+1)}$; i.e., we call (4.10) a *linear iterisation scheme* for $Au = f$. Defining $G : X \to X'$ as

$$G(u) = L[u]u - (Au - f),$$

we can write (4.10) as

$$L[u^{(m)}]u^{(m+1)} = G(u^{(m)}), \qquad m \geq 0, \tag{4.11}$$

or

$$u^{(m+1)} = L[u^{(m)}]^{-1}G(u^{(m)}) \qquad m \geq 0. \tag{4.12}$$

Remark. This later form is in the form of a *Picard iteration*

$$u^{(m+1)} = T(u^{(m)}),$$

where $T : X \to X$ is an operator.

We assume there exists a *bilinear form* for a fixed $u \in X$

$$a(u; v, w) = \langle L[u]v, w\rangle \qquad v, w \in X;$$

e.g., from the weak formulation of (linearised) partial differential equation. Then, for a given $u^{(m)} \in X$ the solution $u^{(m+1)} \in X$ of (4.11) can be obtained from the weak formulation

$$a(u^{(m)}; u^{(m+1)}, v) = \langle G(u^{(m)}), v\rangle, \qquad \text{for all } v \in X. \tag{4.13}$$

We assume that the bilinear form $a(u; \cdot, \cdot)$ is uniformly coercive and bounded; i.e., there exists positive constants α and β independent of u such that

$$a(u; v, v) \geq \alpha \|v\|_X^2 \qquad \text{for all } v \in X, \tag{4.14}$$

$$a(u; v, w) \leq \beta \|v\|_X \|w\|_X \qquad \text{for all } v, w \in X. \tag{4.15}$$

In particular, these properties imply the well-posedness of the solution $u^{(m+1)} \in X$ of (4.13) for any $u^{(m)} \in X$ by Lax-Milgram.

We consider this general form as it allows us to consider several different linearisation schemes. We note, for example, that the following linearisation schemes fit this form.

Zarantonello For X a Hilbert space, a *Zarantonello iteration* is given by

$$(u^{(m+1)}, v)_X = (u^{(m)}, v)_X - \varepsilon\langle Au^{(m)} - f, v\rangle \qquad \text{for all } v \in X, m \geq 0,$$

for $\varepsilon > 0$ sufficiently small; see Corollary 2.10 and (3.2) for strongly monotone and Lipschitz continuous A. This is equivalent to (4.13) if

$$L[u] := \frac{1}{\varepsilon} J_X.$$

Kačanov Defining $L[u]$ such that $L[u]u = Au$; then, $G(u) = L[u]u - (Au - f) = f$, which gives a Kačanov iteration.

(Damped) Newton Set $L[u] = \varepsilon_m^{-1} A'(u)$, where A' is the Gâteaux derivative of A; then, (4.13) is equivalent to (4.5).

We now consider a finite dimensional subspace $X_n \subset X$ and consider the galerkin approximation: Find $u_n \in X_n$ such that

$$\langle Au_n, v\rangle = \langle f, v\rangle \qquad \text{for all } v \in X_n. \tag{4.16}$$

Then, the *iterative linearised Galerkin (ILG)* (cf. Heid and Wihler (2020)) scheme is to consider a Galerkin approximation of the linearised weak formulation (4.13): Given an initial guess $u_n^{(0)} \in X_n$, find $u_n^{(m+1)} \in X$ such that

$$a(u_n^{(m)}; u_n^{(m+1)}, v) = \langle G(u_n^{(m)}), v\rangle \qquad \text{for all } v \in X_n, m \geq 0. \tag{4.17}$$

We shall now only consider an operator A which meets the following assumptions:

(M1) A is Lipschitz continuous; i.e., there exists a constant $L > 0$ such that

$$|\langle Au - Av, w\rangle| \leq L\|u - v\|\|w\|.$$

We note this is different to the previous definition, but that it follows from the previous definition and the definition of the dual norm $\|\cdot\|_{X'}$.

(M2) A is strongly monotone; i.e., there exists a constant $M > 0$ such that

$$\langle Au - Av, u - v\rangle \geq M\|u - v\|^2.$$

Lemma 4.2. *Consider the sequence* $\{u^{(m)}\}_{m\geq 0} \subset X$ *generated by* (4.13)*. Let* A *satisfy (M1)–(M2) and* $a(u; \cdot, \cdot)$ *satisfy* (4.14) *and* (4.15) *for fixed* $u \in X$*; then, it holds that*

$$\|u - u^{(m)}\| \leq \left(1 + \frac{\beta}{M}\right)\|u^{(m)} - u^{(m-1)}\|, \qquad m \geq 1,$$

where $u \in X$ *is the solution of* $Au = f$*.*

Proof. From (4.13) and definition of $a(\cdot; \cdot, \cdot)$ we have for all $v \in X$ that

$$a(u^{(m-1)}; u^{(m)}, v) = \langle G(u^{(m-1)}), v\rangle = \langle L[u^{(m-1)}]u^{(m-1)}, v\rangle - \langle Au^{(m-1)} - f, v\rangle;$$

hence, for all $v \in X$

$$\langle Au^{(m-1)} - f, v\rangle = a(u^{(m-1)}; u^{(m-1)}, v) - a(u^{(m-1)}; u^{(m)}, v) = a(u^{(m-1)}; u^{(m-1)} - u^{(m)}, v).$$

The, by (M2) and (4.15), as $u^{(m-1)} - u \in X$,

$$\begin{aligned} M\|u - u^{(m-1)}\|^2 &\leq \langle Au - Au^{(m-1)}, u - u^{(m-1)}\rangle \\ &= \langle f - Au^{(m-1)}, u - u^{(m-1)}\rangle \\ &= a(u^{(m-1)}; u^{(m-1)} - u^{(m)}, u^{(m-1)} - u) \\ &\leq \beta\|u^{(m-1)} - u^{(m)}\|\|u^{(m-1)} - u\|. \end{aligned}$$

This yields that

$$\|u - u^{(m-1)}\| \leq \frac{\beta}{M}\|u^{(m-1)} - u^{(m)}\|;$$

then, by the triangle inequality,

$$\|u - u^{(m)}\| \leq \|u - u^{(m-1)}\| + \|u^{(m-1)} - u^{(m)}\| \leq \left(1 + \frac{\beta}{M}\right)\|u^{(m-1)} - u^{(m)}\|. \qquad \square$$

Remark. The above lemma also holds for $u_n, u_n^{(m)} \in X_n \subset X$ from (4.16) and (4.17), respectively, instead of $u, u^{(m)}$.

We now consider a sequence of hierarchical finite dimensional subspaces X_n, $n \geq 0$; i.e.,

$$X_0 \subset X_1 \subset X_2 \subset \cdots \subset X.$$

Remark. For example, X_0 could be a finite element space on a coarse mesh $\mathcal{T}_0$; then, X_n, $n \geq 1$ is the finite element space on a mesh $\mathcal{T}_n$ which is a refinement (not necessarily uniform) of $\mathcal{T}_{n-1}$.

We suppose that there exists a *computable* error estimator $\eta_n : X_n \to \mathbb{R}_+$, $n \geq 0$, and constants $C_S, C_E > 0$ independent of n such that

(N1) $|\eta_n(u) - \eta_n(v)| \leq C_S \|u - v\|$ for all $u, v \in X_n$,

(N2) the error of the Galerkin approximation $u_n \in X_n$ from (4.16) is controlled by the *a posteriori* error bound
$$\|u - u_n\| \leq C_E \eta_n(u_n)$$
where $u \in X$ is the exact solution of $Au = f$.

We can show that $\eta_n(u_n)$ and $\eta_n(u_n^{(m)})$ are equivalent (in a norm equivalence-like sense) once the linearisation error is sufficiently small.

Lemma 4.3. *Suppose A satisfies (M1)–(M2), η_n satisfies (N1), and for $m \geq 1$*

$$\|u_n^{(m)} - u_n^{(m-1)}\| \leq \lambda \eta_n(u_n^{(m)}),$$

with $\lambda \in (0, C_\lambda^{-1})$ where $C_\lambda = (1 + \beta/M) C_S$; then,

$$\|u - u_n^{(m)}\| \leq \lambda \left(1 + \frac{\beta}{M}\right) \min \left(\eta_n(u_n^{(m)}), (1 - \lambda C_\lambda)^{-1} \eta_n(u_n)\right).$$

Moreover,
$$(1 - \lambda C_\lambda) \eta_n(u_n^{(m)}) \leq \eta_n(u_n) \leq (1 + \lambda C_\lambda) \eta_n(u_n^{(m)}).$$

Remark. Note that the assumption

$$\|u_n^{(m)} - u_n^{(m-1)}\| \leq \lambda \eta_n(u_n^{(m)}),$$

essentially requires that the linearisation error, cf, Lemma 4.2 is less than the *discretisation* or *Galerkin approximation* error from η_n. See Algorithm 4.2 for a practical example of ensuring this assumption is met.

Proof. By Lemma 4.2 and (N1)

$$\begin{aligned}\|u_n - u_n^{(m)}\| \leq \left(1 + \frac{\beta}{M}\right) \|u_n^{(m)} - u_n^{(m-1)}\| &\leq \lambda \left(1 + \frac{\beta}{M}\right) \eta_n(u_n^{(m)}) \\ &\leq \lambda \left(1 + \frac{\beta}{M}\right) \left(\eta_n(u^{(m)}) + C_S \|u_n - u_n^{(m)}\|\right).\end{aligned} \tag{4.18}$$

Hence,

$$\|u_n - u_n^{(m)}\| \le \frac{\lambda\left(1+\frac{\beta}{M}\right)}{1-\lambda\left(1+\frac{\beta}{M}\right)C_S}\eta_n(u_n) = \lambda\left(1+\frac{\beta}{M}\right)(1-\lambda C_\lambda)^{-1}\eta_n(u_n).$$

Combining this result with (4.18) proves the first result. Moreover, by this result and (N1),

$$\eta(u_n) \le \eta_n(u_n^{(m)}) + C_S\|u_n - u^{(m)}\| \le (1+\lambda C_\lambda)\,\eta_n(u_n^{(m)}),$$
$$\eta(u_n^{(m)}) \le \eta_n(u_n) + C_S\|u_n - u^{(m)}\| \le \left(1+\frac{\lambda C_\lambda}{1-\lambda C_\lambda}\right)\eta_n(u_n) = (1-\lambda C_\lambda)^{-1}\eta_n(u_n). \qquad \square$$

Using these assumptions we can derive an iterative algorithm which only performs enough iterations on each Galerkin space X_n to reduce the error from the linearisation to be less than the Galerkin approximation error rather than continuing until an arbitrary tolerance is reached.

Algorithm 4.2 (Heid and Wihler (2020))**.** For a prescribed tolerance `TOL` and $\lambda > 0$, set $n = 0$, and start with an initial Galerkin approximation space X_0 with initial guess $u_n^{(0)} \in X_0$.

while $\eta_n(u_n^{(0)}) > \texttt{TOL}$ **do**
  $m \leftarrow 1$
  Compute $u_n^{(1)}$ from single step of (4.17)
  while $\|u_n^{(m)} - u_n^{(m-1)}\| > \lambda\eta_n(u_n^{(m)})$ **do**
    Compute $u_n^{(m+1)}$ from single step of (4.17)
    $m \leftarrow m+1$
  end while
  $u_n^F \leftarrow u^{(m)} \in X$
  Enrich X_n based on $\eta_n(u_n^F)$ to obtain X_{n+1}
  $u_{n+1}^{(0)} \leftarrow u_n^F$ by inclusion ($X_n \hookrightarrow X_{n+1}$)
  $n \leftarrow n+1$
end while
Output the sequence $\{u_n^F\}_{n\ge 0}$

Remark. Note, we have no guarantee that the inner loop actually terminates; depends on the convergence of the selected ILG method.

Remark. We note that in the case of finite element spaces the *a posteriori* error estimate η_n can be often split into element-wise contributions, indicating the elements which need refining to create an enhanced space X_{n+1} from X_n.

We want to show convergence of this algorithm. To this end, we need extra assumptions on the sequence of Galerkin approximations, which we note can be shown for finite element spaces under certain conditions on mesh refinement; cf. Gantner et al. (2017).

Proposition 4.4. *Let (M1)–(M2) and (N1) hold, and $\lambda \in (0, C_\lambda)$ be given. Moreover, for $n \ge 0$ assume that the inner while loop of Algorithm 4.2 terminates; i.e., there exists a sequence $\{u_n^F\}_{n\ge 0}$. Furthermore, suppose that there exists constants $0 < q < 1$ and $C > 0$ such that*

$$\eta_{n+1}(u_{n+1})^2 \le q\eta_n(u_n)^2 + C\|u_{n+1} - u_n\|^2, \qquad \textit{for all } n \ge 0,$$

where $u_n \in X_n$ is the unique solution of the Galerkin approximation (4.16). Then, $\eta_n(u_n^F) \to 0$ as $n \to \infty$.

Proof. See, for example, Gantner et al. (2017) for the case of finite element spaces. □

Corollary 4.5. *Let the assumptions in Proposition 4.4 and (N2) hold; then, $u_n^F \to u$ as $n \to \infty$, where the sequence $\{u_n^F\}$ is generated by Algorithm 4.2 and $u \in X$ is the unique solution of $Au = f$.*

Proof. Let $u_n^F = u_n^{(m+1)}$, $m \geq 1$, be the output of Algorithm 4.2 on X_n. By Lemma 4.3 and (N2)

$$\begin{aligned}
\|u - u_n^F\| &\leq \|u_n - u_n^{(m)}\| + \|u - u_n\| \\
&\leq \lambda \left(1 + \frac{\beta}{M}\right) \eta_n(u^{(m)}) + C_E \eta_n(u_n) \\
&\leq \left(\lambda \left(1 + \frac{\beta}{M}\right) \eta_n(u^{(m)}) + C_E(1 + \lambda C_\lambda)\right) \eta_n(u_n^F).
\end{aligned}$$

Hence, by Proposition 4.4, $\|u - u_n^F\| \to 0$ as $n \to \infty$, which completes the proof. □

Example 4.3 (ILG form strongly monotone & Lipschitz continuous PDE). We again revisit Example 2.1 & Example 3.1. Consider

$$\begin{aligned}
-\nabla \cdot (\mu(\boldsymbol{x}, |\nabla u|)\nabla u) &= f && \text{in } \Omega, \\
u &= 0 && \text{on } \partial\Omega,
\end{aligned}$$

where there exists positive constants $\alpha_1 \geq \alpha_2 > 0$ such that, for $t \geq s \geq 0$ and $\boldsymbol{x} \in \overline{\Omega}$

$$\alpha_2(t - s) \leq \mu(\boldsymbol{x}, t)t - \mu(\boldsymbol{x}, s)s \leq \alpha_1(t - s).$$

Let $X = H_0^1(\Omega)$, we can define $A : X \to X'$ and $F \in X'$ as

$$\begin{aligned}
\langle Au, w \rangle &= \int_\Omega \mu(\boldsymbol{x}, |\nabla u|)\nabla u \cdot \nabla w \, \mathrm{d}\boldsymbol{x} \\
\langle F, w \rangle &= \int_\Omega f v \, \mathrm{d}\boldsymbol{x}.
\end{aligned}$$

The selection of $a(\cdot; \cdot, \cdot)$ depends on the linearisation method:

Zarantonello: For $\varepsilon > 0$, see Example 3.1,

$$a_Z(u; v, w) = \frac{1}{\varepsilon} \int_\Omega \nabla u \cdot \nabla v \, \mathrm{d}\boldsymbol{x}$$

Kačanov:

$$a_K(u; v, w) = \int_\Omega \mu(\boldsymbol{x}, |\nabla u|)\nabla v \cdot \nabla w \, \mathrm{d}\boldsymbol{x}$$

Damped Newton: For damping parameter $\varepsilon_m \in (0, 1]$, $m \geq 0$,

$$a_N(u; v, w) = \frac{1}{\varepsilon_m} \int_\Omega \left(\mu'(\boldsymbol{x}, |\nabla u|)\frac{\nabla u \cdot \nabla v}{|\nabla u|}\nabla u \cdot \nabla w + \mu(\boldsymbol{x}|\nabla u|)\nabla v \cdot \nabla w\right) \, \mathrm{d}\boldsymbol{x}$$

where $\mu'(\boldsymbol{x}, t)$ denotes the derivative of $\mu(\boldsymbol{x}, t)$ with respect to t.

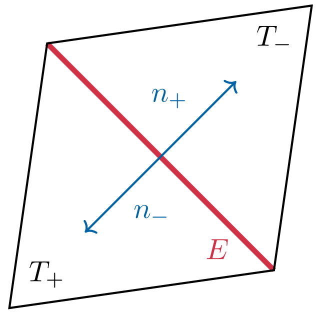

Figure 4.1: Edge E between mesh elements $T_+, T_- \in \mathcal{T}_n$

We can define a sequence of finite element meshes $\mathcal{T}_n$, $n \geq 0$, where $\mathcal{T}_{n+1}$ is a refinement of $\mathcal{T}_n$ and let

$$X_n = \{v \in H_0^1(\Omega) : v|_T \in P_p(T) \; \forall T \in \mathcal{T}_n\} \subset X, \qquad n \geq 0;$$

hence, $X_0 \subset X_1 \subset X_2, \subset \ldots$. We note that by standard *a posteriori* error bounds (Congreve and Wihler, 2017, Gantner et al., 2017, Heid and Wihler, 2020) we can show that

$$\|\nabla u - \nabla u_n\|^2 \leq C\eta_n(u_n)^2,$$

where

$$\eta_n(u)^2 := \sum_{T \in \mathcal{T}_n} \eta_{n,T}(u)^2,$$
$$\eta_{n,T}(u)^2 := h_T^2 \|f + \nabla \cdot (\mu(\boldsymbol{x}, |\nabla u|)\nabla u)\|_{0,2,T}^2 + h_T \|[\![\mu(\boldsymbol{x}, |\nabla u|)\nabla u]\!]\|_{0,2,\partial T \setminus \partial\Omega}^2$$

where h_T is the diameter of $T \in \mathcal{T}_n$ and on an edge $E = \partial T_+ \cap \partial T^-$ shared between two neighbouring elements $T_+, T_- \in \mathcal{T}_n$, see Figure 4.1,

$$[\![\boldsymbol{v}]\!] = \boldsymbol{v}|_{T_+} \cdot \boldsymbol{n}_+ + \boldsymbol{v}|_{T_-} \cdot \boldsymbol{n}_-.$$

Bibliography

M. Amrein and T. P. Wihler. An adaptive Newton-method based on a dynamical systems approach. *Commun. Nonlinear Sci.*, 19:2958–2973, 2014. DOI: 10.1016/j.cnsns.2014.02.010.

M. Amrein and T. P. Wihler. Fully adaptive Newton-Galerkin methods for semilinear elliptic partial differential equations. *SIAM J. Sci. Comput.*, 37:A1637–A1657, 2015. DOI: 10.1137/140983537.

J. Barrett and W. Liu. Finite element approximation of the p-Laplacian. *Math. Comp.*, 61(204): 523–537, 1993. DOI: 10.1090/S0025-5718-1993-1192966-4.

K. Böhmer. *Numerical Methods for Nonlinear Elliptic Differential Equations: A Synopsis*. Oxford University Press, Oxford, 2010. DOI: 10.1093/acprof:oso/9780199577040.001.0001.

S. Congreve and T. P. Wihler. Iterative Galerkin discretizations for strongly monotone problems. *J. Comput. Appl. Math.*, 311:457–472, 2017. DOI: 10.1016/j.cam.2016.08.014.

V. Dolejší and K. Najzar. *Nelineární Funkcionální Analýza*. Matfyzpress, Prague, 2010.

L. C. Evans. *Partial Differential Equations*. AMS, Providence, RI, 1998.

G. Gantner, A. Haberl, D. Praetorius, and B. Stiftner. Rate optimal adaptive FEM with inexact solver for nonlinear operators. *IMA J. Numer. Anal.*, 38(4):1797–1831, 09 2017. DOI: 10.1093/imanum/drx050.

P. Heid and T. P. Wihler. On the convergence of adaptive iterative linearized Galerkin methods. *Calcolo*, 57(3):24, 2020. DOI: 10.1007/s10092-020-00368-4.

J. Leray and J.-L. Lions. Quelques résultats de Višik sur les problèmes elliptiques non linéaires par les méthodes de Minty-Browder. *Bull. Soc. Math. France*, 93:97–107, 1965. DOI: 10.24033/bsmf.1617.

W. Liu and J. Barrett. Quasi-norm error bounds for the finite element approximation of some degenerate quasilinear elliptic equations and variational inequalities. *RAIRO Modél. Math. Anal Numér*, 28(6):725–744, 1994. DOI: 10.1007/s10915-011-9463-9.

E. Zeidler. *Nonlinear Functional Analysis and its Applications. III: Variational Methods and Optimization*. Springer, New York, 1984.

E. Zeidler. *Nonlinear Functional Analysis and its Applications, I: Fixed-Point Theorems*. Springer, New York, 1985.

E. Zeidler. *Nonlinear Functional Analysis and its Applications. II/A: Linear Monotone Operators*. Springer, New York, 1989a.

E. Zeidler. *Nonlinear Functional Analysis and its Applications. II/B: Nonlinear Monotone Operators*. Springer, New York, 1989b.