Chapter 5
Numerical homogenization beyond scale separation and periodicity

Abstract This chapter presents an alternative numerical approach to homogenization that is not based on the mathematical theory of homogenization but the availability of operator-dependent subspaces with a quasi-local basis and approximation properties independent of oscillations and roughness of the diffusion coefficient.

5.1 An idealized method

In practical applications it is often impossible to model material properties encoded in the coefficient by a locally periodic coefficient of the form $A_\varepsilon(x) = A_1(x, \frac{x}{\varepsilon})$ for some 1-periodic coefficient $A(x,\cdot)$. Often, we are not even able to identify a parameter ε that represents microscopic oscillations. In those cases we are still interested in coarse representations of some rough and highly oscillatory coefficient $A \in \mathcal{M}(D,\alpha,\beta)$ or the corresponding partial differential operator that allows the efficient simulation on some macroscopic scale of interest. The question whether there are stable and accurate methods beyond the strong structural assumptions of analytical homogenization regarding scale separation or even periodicity remained open for a long time. Only recently, the existence of an optimal approximation of the low-regularity solution space by some arbitrarily coarse generalized finite element space (that represents the homogenized problem) was shown in [1] and [3]. However, the constructions therein include prohibitively expensive global solutions of the full fine scale problem or the solution of more involved eigenvalue problems. An efficient and feasible construction, solely based on the solution of localized microscopic cell problems, was given and rigorously justified in [10] and later optimized in [5], generalized in [4, 6] and reinterpreted in [8]. Since then, several other approaches have been developed. One approach with presumably similar properties was suggested in [12] along with the notion of sparse super-localization that reflects the locality of the discrete homogenized operator (similar to the sparsity of standard finite element matrices). Among the latest developments are [13, 9, 7, 11].

By generalizing the one-dimensional derivation of Section 1.3.4, we will now present the approach of [10] to this problem. The method does not rely on symmetry of A. Since some arguments are more illustrative in the symmetric case and to stay as close as possible to the $1d$ template, we will still assume symmetry of A.
Given $\beta \geq \alpha > 0$, some admissible coefficient $A \in \mathcal{M}_{\mathrm{sym}}(D, \alpha, \beta)$, some D convex and polyhedral, and some outer force $f \in L^2(D)$, we wish to approximate the unique function $u \in H^1_0(D) =: V$ such that

$$a(u,v) := \int_D (A\nabla u) \cdot \nabla v \, dx = \int_D f v \, dx =: F(v) \tag{5.1}$$

for all $v \in V$. This is the model problem of Chapter 2 and its well-posedness was already discussed there.
Let $\mathcal{T}_H$ be some regular mesh of D as in Section 4.2 - 4.5. Here the mesh size parameter represents the scale of interest that can be chosen independent of characteristic length scales of A. By V_H, we will refer to the standard finite element space

$$V_H := S^1(\mathcal{T}_H) \cap H^1_0(D),$$

satisfying homogeneous Dirichlet boundary conditions.
As in Section 1.3.4, we shall characterize the functions in V that are not well-captured by finite element shape functions. Note, however, that a characterization by nodal values as in (1.25) is not possible in dimension $d > 1$. This is where the quasi-interpolation operator $I_H : V \to V_H$ of Definition 4.3 that is based on volume averaging comes into play. Define

$$W_H := \{w \in V \mid I_H w = 0\} = \operatorname{kern} I_H, \tag{5.2}$$

the space of (microscopic) fine-scale functions. (Observe that we could have written $W_H = \operatorname{kern} I_H^{\mathrm{nodal}}$ in the one-dimensional case with the nodal interpolation operator I_H^{nodal}.)
The remaining steps of the derivation widely coincide with Section 1.3.4. Observe that the solution space V can be decomposed as

$$V = V_H \oplus W_H \tag{5.3}$$

and, for any $v \in V$, $I_H v \in V_H$ and $(1 - I_H)v \in W_H$ are the unique elements of V_H and W_H such that

$$v = I_H v + (1 - I_H)v.$$

Moreover, by Lemma 4.1, the decomposition is stable in the sense that

$$\|I_H v\| + \|(1 - I_H)v\| \leq 2C_I \|v\|,$$

where $\|\cdot\|$ refers to either the $L^2(D)$ norm or the $H^1(D)$ norm. (In contrast to the $1d$ case with nodal interpolation, I_H is only an oblique projection with respect to both $L^2(D)$ and $H^1_0(D)$.)

The new approach is based on the orthogonalization of (5.3) with respect to the scalar product

$$a(u,v) := \int_D (A\nabla u)\cdot\nabla v\,dx \tag{5.4}$$

associated with the problem (5.1).
Keeping W_H fixed, we characterize a new coarse space $\tilde{V}_H \subset V$ as the subspace that satisfies

$$V = \tilde{V}_H \oplus W_H \quad \text{and} \quad a(\tilde{V}_H, W_H) = 0,$$

i.e.,

$$\tilde{V}_H := \{\tilde{v}_H \in V \mid \forall w \in W_H : a(\tilde{v}_H, w) = 0\}. \tag{5.5}$$

The Galerkin method with subspace $\tilde{V}_H$ applied to (5.1) seeks $\tilde{u}_H \in \tilde{V}_H$ such that

$$a(\tilde{u}_H, \tilde{v}_H) = F(\tilde{v}_H) \tag{5.6}$$

for all $\tilde{v}_H \in \tilde{V}_H$. By Galerkin orthogonality

$$a(u-\tilde{u}_H, \tilde{v}_H) = 0$$

for all $\tilde{v}_H \in \tilde{V}_H$, the error $u-\tilde{u}_H$ of this method is a fine-scale function, i.e.,

$$I_H u = I_H \tilde{u}_H. \tag{5.7}$$

With Lemma 4.1 this readily implies that

$$\begin{aligned}\|u-\tilde{u}_H\|_{L^2(D)} &= \|(1-I_H)(u-\tilde{u}_H)\|_{L^2(D)}\\ &\le C_I H \|\nabla(u-\tilde{u}_H)\|_{L^2(D)}.\end{aligned}$$

Moreover,

$$\begin{aligned}\|\nabla(u-\tilde{u}_H)\|^2_{L^2(D)} &\le \alpha^{-1} a(u-\tilde{u}_H, u-\tilde{u}_H)\\ &= \alpha^{-1} a(u, u-\tilde{u}_H)\\ &= \alpha^{-1}\int_D f(u-\tilde{u}_H)\,dx\\ &\le \alpha^{-1}\|f\|_{L^2(D)}\|u-\tilde{u}_H\|_{L^2(D)}.\end{aligned}$$

The combination of the previous two estimates yields the error bound

$$\|u-\tilde{u}_H\|_{L^2(D)} \le C_I^2 \alpha^{-1} H^2 \|f\|_{L^2(D)} \tag{5.8}$$

for the Galerkin method (5.6) in the space $\tilde{V}_h$. This is in agreement with the $1d$ result (1.27). However, the multidimensional case is structurally very different from the $1d$ case with nodal interpolation when it comes to the practical feasibility of the method. Before we discuss such issues in detail, let us reformulate the method as finite element method with modified bilinear form.
For this purpose, let $-Q_H : V \to W_H$ denote the a-orthogonal projection onto the

closed subspace $W_H \subset V$. We will refer to Q_H as the *correction operator*. Note that its complementary projection

$$(1-(-Q_H)) = (1+Q_H)$$

maps V_H onto $\tilde{V}_H$ and is invertible with inverse I_H. We can, hence, identify any $\tilde{v}_H \in \tilde{V}_H$ with its finite element component $v_H = I_H \tilde{v}_H$ and vice versa $\tilde{v}_H = (1+Q_H)v_H$. This allows us to reformulate the method (5.6) as follows: Find $\bar{u}_H \in V_H$ such that

$$a((1+Q_H)\bar{u}_H, (1+Q_H)v_H) = F((1+Q_H)v_H) \tag{5.9}$$

for all $v_H \in V_H$. Replacing the problem-dependent right-hand side (the evaluation of Q_H requires the solution of variational problems based on the bilinear form a) with $v_H \mapsto F(v_H)$ yields the variant of (5.6) that we will study in more detail in the subsequent sections: Find $u_H \in V_H$ such that

$$a((1+Q_H)u_H, (1+Q_H)v_H) = F(v_H) \tag{5.10}$$

for all $v_H \in V_H$.

Lemma 5.1 (Error of the ideal method). *The discrete problem* (5.10) *admits a unique solution $u_H \in V_H$ for any $F \in H^{-1}(D)$ and the error is bounded by*

$$\|u-u_H\|_{L^2(D)} \leq C\Big(\min_{v_H \in V_H} \|u-v_H\|_{L^2(D)} + H\|f\|_{L^2(D)}\Big).$$

Proof. Observe that the modified bilinear form

$$a((1+Q_H)\bullet, (1+Q_H)\bullet) : V_H \times V_H \to \mathbb{R}$$

satisfies, for any $v_H \in V_H$,

$$\begin{aligned} a((1+Q_H)v_H, (1+Q_H)v_H) \geq \quad & \alpha \|\nabla(1+Q_H)v_H\|^2_{L^2(D)} \\ \geq \quad & \frac{\alpha}{C_I^2}\|\nabla v_H\|^2_{L^2(D)} \end{aligned} \tag{5.11}$$

because of $I_H(1+Q_H)v_H = v_H$ and the $H^1_0(D)$-stability of I_H. This proves well-posedness of (5.10) and, in particular, unique solvability.
To prove the error estimate, observe that the solution $\bar{u}_H$ of (5.9) (which is well-posed by (5.11) as well) is exactly $\bar{u}_H = I_H u = I_H \tilde{u}_H$. This implies, for any $v_H \in V_H$,

$$\begin{aligned} \|u-\bar{u}_H\|_{L^2(D)} = \quad & \|(1-I_H)(u-v_H)\|_{L^2(D)} \\ \leq \quad & C_I \|u-v_H\|_{L^2(D)} \end{aligned} \tag{5.12}$$

The error $(u_H - \bar{u}_H)$ can be estimated with (5.11),

$$\begin{aligned}\frac{\alpha}{C_I^2}\|\nabla(\bar{u}_H - u_H)\|^2_{L^2(D)} &\le a((1+Q_H)(\bar{u}_H - u_H),(1+Q_H)(\bar{u}_H - u_H)) \\ &= -\int_D f \underbrace{Q_H(\bar{u}_H - u_H)}_{=(1-I_H)Q_H(\bar{u}_H - u_H)} dx \\ &\le \|f\|_{L^2(D)} C_I H \|\nabla(\bar{u}_H - u_H)\|_{L^2(D)}.\end{aligned}$$

Hence, Friedrichs' inequality yields

$$\|\bar{u}_H - u_H\|_{L^2(D)} \le \frac{C_F C_I^3}{\alpha} H \|f\|_{L^2(D)}.$$

This, (5.12) and the triangle inequality readily yield the assertion. □

For a slightly sharper version of the previous lemma see [2].

5.2 Localization of the correctors

In order to transform (5.10) to a feasible numerical method, it is very important to have a sparse (i.e. local) basis representation. For a finite element basis function φ_x, the function $Q_H\varphi_x$ will have global support in general. In this section we justify a localization procedure that leads to a (quasi-)local variant of the method (5.10).
To this end, we introduce the following *element corrector* for any $T \in \mathcal{T}_H$ and unit vector e_i ($i \in \{1,...,d\}$):
Let $w_{T,i} \in W_H$ solve

$$a(w_{T,i}, v) = -\int_T (Ae_i)\cdot\nabla v\, dx \tag{5.13}$$

for all $v \in W_H$. We claim that, for all $v_H \in V_H$,

$$Q_H v_H = \sum_{T\in\mathcal{T}_H} \sum_{i=1}^d \left.\frac{\partial v_H}{\partial x_i}\right|_T w_{T,i}. \tag{5.14}$$

Indeed, we have, for any $v \in W_H$,

$$\begin{aligned}a\Big(\sum_{T\in\mathcal{T}_H}\sum_{i=1}^d \left.\frac{\partial v_H}{\partial x_i}\right|_T w_{T,i}, v\Big) &= \sum_{T\in\mathcal{T}_H}\sum_{i=1}^d \left.\frac{\partial v_H}{\partial x_i}\right|_T a(w_{T,i}, v) \\ &\overset{(5.13)}{=} -\sum_{T\in\mathcal{T}_H}\sum_{i=1}^d \left.\frac{\partial v_H}{\partial x_i}\right|_T \int_T (Ae_i)\cdot\nabla v\, dx \\ &= -a(v_H, v).\end{aligned}$$

and this is exactly the equation to be satisfied by the negative of the Galerkin projection onto W_H. We see from (5.14) that $Q_H v_H$ is built from a sum of contributions that are solutions to (5.13), a problem with local right-hand side. In the following,

we shall prove that $w_{T,i}$ decays exponentially fast away from T. For a proper quantification of the decay we introduce the following element neighborhoods (or patches) for a subdomain $S \subset D$,

$$\begin{aligned}\mathsf{N}(S) &= \bigcup \Big\{ K \in \mathcal{T}_H : K \cap \bar{S} \neq \emptyset \Big\} \\ \mathsf{N}^\ell(S) &= \mathsf{N}(\mathsf{N}^{\ell-1}(S)) \text{ for } \ell \geq 2 \text{ and } \mathsf{N}^1(S) := \mathsf{N}(S).\end{aligned}$$

Theorem 5.1 (Exponential decay). *Let $T \in \mathcal{T}_H$, $i \in \{1,...,d\}$ and let $w_{T,i} \in W_H$ solve (5.13). Then there exists $c = c(\alpha,\beta) > 1$ (independent of T) such that, for any $\ell \geq 1$,*

$$\|\nabla w_{T,i}\|_{L^2(D\setminus \mathsf{N}^\ell(T))} \leq \exp(-c\ell)|T|^{1/2}.$$

Proof. We choose a Lipschitz continuous cutoff function $\eta \in W^{1,\infty}(D;[0,1])$ with

$$\begin{aligned}&\eta \equiv 0 \quad \text{in } \mathsf{N}^{\ell-3}(T) \\ &\eta \equiv 1 \quad \text{in } D \setminus \mathsf{N}^{\ell-2}(T) \\ &\|\nabla \eta\|_{L^\infty(D)} \leq C_\eta H^{-1}\end{aligned} \tag{5.15}$$

and observe that

$$\begin{aligned}\operatorname{supp}(\eta) &= D \setminus \mathsf{N}^{\ell-3}(T) \\ \operatorname{supp}(\nabla\eta) &= \mathsf{N}^{\ell-2}(T) \setminus \mathsf{N}^{\ell-3}(T) =: R.\end{aligned}$$

Abbreviate $w := w_{T,i}$.
We calculate with the product rule

$$\begin{aligned}\|\nabla w\|^2_{L^2(D\setminus \mathsf{N}^\ell(T))} &\lesssim \|A^{1/2}\nabla w\|^2_{L^2(D\setminus \mathsf{N}^\ell(T))} \overset{A \text{ spd}, \eta \geq 0}{\leq} \langle A\nabla w, \eta \nabla w\rangle_{L^2(D)} \\ &\leq \underbrace{|\langle A\nabla w, \nabla(1-I_H)(\eta w)\rangle_{L^2(D)}|}_{=:M_1} + \underbrace{|\langle A\nabla w, \nabla I_H(\eta w))\rangle_{L^2(D)}|}_{=:M_2} + \underbrace{|\langle A\nabla w, w\nabla\eta\rangle_{L^2(D)}|}_{=:M_3}.\end{aligned}$$

We proceed by estimating M_1, M_2, M_3.

$\mathbf{M_1}$ Since $v := (1-I_H)(\eta w) \in W_H$ we have by (5.13)

$$M_1 = \left| \int_T (Ae_i)\cdot \nabla v \, dx \right|,$$

but as the support of v lies outside of T, we may conclude that $M_1 = 0$.

$\mathbf{M_2}$ Since $w \in W_H$, we have $\operatorname{supp}(I_H(\eta w)) \subset \mathsf{N}(R)$. Hence, with $\operatorname{supp}\nabla\eta = R$,

$$M_2 \lesssim C_I \|\nabla w\|_{L^2(\mathsf{N}(R))} \Big(\|\eta \nabla w\|_{L^2(\mathsf{N}^2(R))} + \|\nabla\eta\|_{L^\infty(D)} \|w\|_{L^2(R)} \Big).$$

With the bound (5.15) on $\nabla\eta$ and with

$$\|w\|_{L^2(R)} = \|w - I_H w\|_{L^2(R)} \le C_I H \|\nabla w\|_{L^2(\mathsf{N}(R))},$$

we conclude

$$M_2 \lesssim \|\nabla w\|^2_{L^2(\mathsf{N}^2(R))}.$$

$\mathbf{M_3}$ Similarly, we have with (5.15) that

$$M_3 \lesssim \|\nabla w\|^2_{L^2(\mathsf{N}^2(R))}.$$

Altogether, there is a constant $\tilde{C} > 0$ such that

$$\|\nabla w\|^2_{L^2(D\setminus\mathsf{N}^\ell(T))} \le \tilde{C}\|\nabla w\|^2_{L^2(\mathsf{N}^2(R))}. \tag{5.16}$$

Since

$$\mathsf{N}^2(R) = \mathsf{N}^\ell(T) \setminus \mathsf{N}^{\ell-5}(T),$$

we get

$$\|\nabla w\|^2_{L^2(D\setminus\mathsf{N}^\ell(T))} + \|\nabla w\|^2_{L^2(\mathsf{N}^2(R))} = \|\nabla w\|^2_{L^2(D\setminus\mathsf{N}^{\ell-5}(T))}$$

and with (5.16) it follows that

$$(1+\tilde{C}^{-1})\|\nabla w\|^2_{L^2(D\setminus\mathsf{N}^\ell(T))} \le \|\nabla w\|^2_{L^2(D\setminus\mathsf{N}^{\ell-5}(T))}.$$

A repeated application of this argument with $\gamma = (1+\tilde{C}^{-1})^{-1} < 1$ results in

$$\begin{aligned}
\|\nabla w\|^2_{L^2(D\setminus\mathsf{N}^\ell(T))} &\le \gamma^{\lfloor \ell/5\rfloor}\|\nabla w\|^2_{L^2(D)} \\
&\overset{\text{stability of (5.13)}}{\lesssim} \gamma^{\lfloor \ell/5\rfloor}\|e_i\|^2_{L^2(T)} \\
&\lesssim \gamma^{\lfloor \ell/5\rfloor}|T|.
\end{aligned}$$

Since $\gamma^{\lfloor \ell/5\rfloor} \le \exp(-c\ell)$ for some $c > 0$, this is the assertion. □

The decay motivates a localized version of (5.13). Define the localized form

$$a_{\mathsf{N}^\ell(T)}(v,w) := \int_{D_T^\ell} (A\nabla v)\cdot\nabla w\, dx$$

based on the element patches $\mathsf{N}^\ell(T)$ and define $w_{T,i}^{(\ell)} \in W_H(D_T^\ell)$ as the solution to the 'cell problem'

$$a_{\mathsf{N}^\ell(T)}(w_{T,i}^{(\ell)}, v) = -\int_T (Ae_i)\cdot\nabla v\, dx \tag{5.17}$$

for all $v \in W_H(\mathsf{N}^\ell(T))$, where $W_H(\mathsf{N}^\ell(T))$ is the kernel of I_H when restricted to $H_0^1(\mathsf{N}^\ell(T))$ (with values in the 'restricted' finite element space). We extend these localized correctors by zero to the domain D and define, for $v_H \in V_H$,

$$Q_H^{(\ell)} v_H := \sum_{T\in\mathcal{T}_H} \sum_{i=1}^{d} \left.\frac{\partial v_H}{\partial x_i}\right|_T w_{T,i}^{(\ell)}.$$

Remark 5.1. $Q_H^{(\ell)}\Lambda_z$ has now support in the nodal patch $\mathsf{N}^{\ell+1}(z)$ given by

$$\mathsf{N}^1(z) = \bigcup\Big\{T \in \mathcal{T}_H : x \in T\Big\}$$

and

$$\mathsf{N}^{\ell+1}(z) = \bigcup\Big\{T \in \mathcal{T}_H : T \cap \mathsf{N}^{\ell}(z) \neq \emptyset\Big\}$$

We note the following result and omit the proof.

Corollary 5.1. *There exists a constant $C > 0$ such that for any $T \in \mathcal{T}_H$ and any $i \in \{1,...,d\}$*

$$\|\nabla(w_{T,i} - w_{T,i}^{(\ell)})\|_{L^2(D)} \leq C\exp(-c\ell)|T|^{1/2}$$

where $c = c(\alpha,\beta)$ is the constant from Theorem 5.1.

5.3 The quasi-local method

The results of the previous section motivate the following practical variant of the method (5.10): Find $u_{H,\ell} \in V_H$ such that

$$a((1+Q_H^{(\ell)})u_{H,\ell}, (1+Q_H^{(\ell)})v_H) = F(v_H) \tag{5.18}$$

for all $v_H \in V_H$. We have simply replaced the corrector Q_H by its localized approximation based on the cell problems (5.17). The following theorem states that the results of Lemma 5.1 are widely preserved provided that the *oversampling* or *localization parameter* $\ell \approx |\log H|$.

Theorem 5.2. *The quasi-local numerical homogenization method* (5.18) *admits a unique solution $u_{H,\ell} \in V_H$ (for any $F \in H^{-1}(D)$) and the error is bounded by*

$$\|u - u_{H,\ell}\|_{L^2(D)} \leq C(\alpha,\beta)\Big(\min_{v_H\in V_H} \|u - v_H\|_{L^2(D)} + (H + \ell^d e^{-c\ell})\|f\|_{L^2(D)}\Big). \tag{5.19}$$

Hence, the choice $\ell \approx |\log H|$ recovers the convergence rate of the ideal method.

Proof. **Step 1: Error of localization**
Let $v_H \in V_H$. Then

$$\begin{aligned}
\|\nabla(Q_H v_H - Q_H^{(\ell)} v_H)\|_{L^2(D)} &= \Big\| \sum_{T\in\mathcal{T}_H} \sum_{i=1}^{d} \frac{\partial v_H|_T}{\partial x_i} \nabla(w_{T,i} - w_{T,i}^{(\ell)}) \Big\|_{L^2(D)} \\
&\le \sum_{T\in\mathcal{T}_H} \sum_{i=1}^{d} \Big| \frac{\partial v_H|_T}{\partial x_i} \Big| \|\nabla(w_{T,i} - w_{T,i}^{(\ell)})\|_{L^2(D)} \\
&\lesssim \sum_{T\in\mathcal{T}_H} \big|\nabla v_H|_T\big| e^{-c\ell} \sqrt{|T|} \\
&\lesssim \sum_{T\in\mathcal{T}_H} \|\nabla v_H\|_{L^2(T)} e^{-c\ell} \\
&\lesssim \underbrace{\sqrt{\#\mathcal{T}_H}}_{\approx H^{-d/2}} \Big(\sum_{T\in\mathcal{T}_H} \|\nabla v_H\|_{L^2(T)}^2 \Big)^{1/2} e^{-c\ell} \\
&\lesssim e^{-c\ell} H^{-d/2} \|\nabla v_H\|_{L^2(D)}.
\end{aligned}$$

With more refined arguments it can even be shown that

$$\|\nabla(Q_H - Q_H^{(\ell)}) v_H\|_{L^2(D)} \lesssim \ell^d e^{-c\ell} \|\nabla v_H\|_{L^2(D)}.$$

Step 2: $\mathbf{H_0^1(D)}$ stability of $\mathbf{Q_H^{(\ell)}}$
Step 1 and the stability of Q_H, i.e.,

$$\begin{aligned}
\|\nabla Q_H v_H\|_{L^2(D)} &\le \alpha^{-1/2} \|A^{1/2} \nabla Q_H v_H\|_{L^2(D)} \\
&\le \alpha^{-1/2} \|A^{1/2} \nabla v_H\|_{L^2(D)} \\
&\le \sqrt{\tfrac{\beta}{\alpha}} \|\nabla v_H\|_{L^2(D)}
\end{aligned}$$

readily yield the $H_0^1(D)$ stability of $Q_H^{(\ell)}$,

$$\begin{aligned}
\|\nabla Q_H^{(\ell)} v_H\|_{L^2(D)} &\le \|\nabla(Q_H^{(\ell)} - Q_H) v_H\|_{L^2(D)} + \|\nabla Q_H v_H\|_{L^2(D)} \\
&\lesssim (1 + e^{-c\ell} \ell^d) \|\nabla v_H\|_{L^2(D)}.
\end{aligned}$$

Step 3: Well-posedness of (5.18)
Observe that the bilinear form

$$a((1 + Q_H^{(\ell)})\bullet, (1 + Q_H^{(\ell)})\bullet) : V_H \times V_H \to \mathbb{R}$$

satisfies, for any $v_H \in V_H$,

$$\begin{aligned}
a((1 + Q_H^{(\ell)}) v_H, (1 + Q_H^{(\ell)}) v_H) &\ge \alpha \|\nabla(1 + Q_H^{(\ell)}) v_H\|_{L^2(D)}^2 \\
&\ge \frac{\alpha}{C_I^2} \|\nabla v_H\|_{L^2(D)}^2.
\end{aligned}$$

Here we have used that $v_H = I_H(1 + Q_H^{(\ell)}) v_H$ and

$$\|\nabla v_H\| = \|\nabla I_H(1+Q_H^{(\ell)})v_H\| \le C_I\|\nabla(1+Q_H^{(\ell)})v_H\|.$$

This implies well-posedness.

Step 4: Error estimate
To prove the error estimate we shall have a look at $e_\ell := u_H - u_{H,\ell}$ with the solution u_H of (5.10). We have

$$\begin{aligned}
\|\nabla e_\ell\|_{L^2(D)}^2 &\lesssim a((1+Q_H)e_\ell,(1+Q_H)e_\ell)\\
&= \underbrace{a((1+Q_H)u_H,(1+Q_H)e_\ell)}_{\overset{(5.10)}{=}F(e_\ell)\overset{(5.18)}{=}a((1+Q_H^{(\ell)})u_{H,\ell},(1+Q_H^{(\ell)})e_\ell)} - a((1+Q_H)u_{H,\ell},(1+Q_H)e_\ell)\\
&= a((1+Q_H^{(\ell)})u_{H,\ell},(1+Q_H^{(\ell)})e_\ell) - a((1+Q_H)u_{H,\ell},(1+Q_H)e_\ell)\\
&= a((Q_H^{(\ell)}-Q_H)u_{H,\ell},(1+Q_H^{(\ell)})e_\ell) + a((1+Q_H)u_{H,\ell},(Q_H^{(\ell)}-Q_H)e_\ell)\\
&\lesssim \|\nabla(Q_H^{(\ell)}-Q_H)u_{H,\ell}\|_{L^2(D)}\|\nabla(1+Q_H^{(\ell)})e_\ell\|_{L^2(D)}\\
&\quad + \|\nabla(1+Q_H)u_{H,\ell}\|_{L^2(D)}\|\nabla(Q_H^{(\ell)}-Q_H)e_\ell\|_{L^2(D)}\\
&\overset{\text{Step 1,2}}{\lesssim} \ell^d e^{-c\ell}\underbrace{\|\nabla u_{H,\ell}\|_{L^2(D)}}_{\lesssim\|f\|_{L^2(D)}}\|\nabla e_\ell\|_{L^2(D)}.
\end{aligned}$$

This shows that

$$\|\nabla e_\ell\|_{L^2(D)} \lesssim \ell^d e^{-c\ell}\|f\|_{L^2(D)}.$$

Using this and Friedrichs' inequality, we finally get

$$\begin{aligned}
\|u-u_{H,\ell}\|_{L^2(D)} &\le \|u-u_H\|_{L^2(D)} + \|u_H-u_{H,\ell}\|_{L^2(D)}\\
&\le C\Big(\min_{v_H\in V_H}\|u-v_H\|_{L^2(D)} + (H+\ell^d e^{-c\ell})\|f\|_{L^2(D)}\Big). \qquad \square
\end{aligned}$$

The final step towards a fully practical method regards the discretization of the cell problems (5.17). For any $T\in\mathcal{T}_H$ and any $\ell\in\mathbb{N}$, let $\mathcal{T}_h(\mathsf{N}^\ell(T))$ denote a regular mesh of the patch $\mathsf{N}^\ell(T)$ and let $V_h(\mathsf{N}^\ell(T))$ denote the corresponding finite element space that satisfies homogeneous Dirichlet boundary condition on $\partial\mathsf{N}^\ell(T)$. We assume that $\mathcal{T}_h(\mathsf{N}^\ell(T))$ is the result of $\log_2\frac{H}{h}$ uniform refinements of $\mathcal{T}_H(\mathsf{N}^\ell(T))$. The restriction of $V_h(\mathsf{N}^\ell(T))$ to the space of fine scale functions results in the discrete approximation space

$$W_{H,h}(\mathsf{N}^\ell(T)) \subset W_H(\mathsf{N}^\ell(T))$$

for the numerical solution of the cell problems (5.17). Define *approximate localized correctors* $w_{T,i,h}^{(\ell)} \in W_{H,h}(\mathsf{N}^\ell(T))$ as unique solutions to the discrete cell problems

$$a_{\mathsf{N}^\ell(T)}(w_{T,i,h}^{(\ell)},v_h) = -\int_T (Ae_i)\cdot\nabla v_h\,dx, \quad \forall v_h\in W_{H,h}(\mathsf{N}^\ell(T)). \tag{5.20}$$

This leads to a further modification of the correction operator Q_H. For any $v_H \in V_H$, define

$$Q^{(\ell)}_{H,h} v_H := \sum_{T\in\mathcal{T}_h} \sum_{i=1}^{d} \left(\frac{\partial v_H|_T}{\partial x_i} \right) w^{(\ell)}_{T,i,h}, \tag{5.21}$$

where $w^{(\ell)}_{T,i,h}$ has been extended to D by zero outside $\mathsf{N}^\ell(T)$. The practical quasi-local method then seeks $u_{H,\ell,h} \in V_H$ such that

$$a((1+Q^{(\ell)}_{H,h})u_{H,\ell,h}, (1+Q^{(\ell)}_{H,h})v_H) = F(v_H) \tag{5.22}$$

for all $v_H \in V_H$. Under the assumption that there is a global fine mesh $\mathcal{T}_h$ of the whole domain D such that all local meshes $\mathcal{T}_h(\mathsf{N}^\ell(T))$ are submeshes of $\mathcal{T}_h$, Theorem 5.2 remains valid if we replace the solution u by a reference solution $u_h \in V_h$ on the global fine mesh, i.e.,

$$a(u_h, v_h) = F(v_h)$$

for all $v_h \in V_h$. (Note that this reference solution is never computed.) Using standard arguments for Galerkin methods yields an error estimate for the method (5.22).

Theorem 5.3. *The practical quasi-local method* (5.22) *is well-posed and satisfies*

$$\|u_h - u_{H,\ell,h}\|_{L^2(D)} \le C\left(\min_{v_H\in V_H} \|u_h - v_H\|_{L^2(D)} + (H + \ell^d e^{-c\ell})\|f\|_{L^2(D)} \right).$$

Moreover, for $\ell \approx |\log H|$,

$$\|u - u_{H,\ell,h}\|_{L^2(D)} \le C\left(\|u - u_h\|_{L^2(D)} + \min_{v_H\in V_H} \|u - v_H\|_{L^2(D)} + H\|f\|_{L^2(D)} \right).$$

If h is sufficiently small, this yields at least convergence of order $O(H)$.

Remark 5.2 (Reconstruction of an effective diffusion tensor). The modified bilinear form in (5.18) may be re-interpreted as an effective integral operator acting on finite element spaces. Under certain assumptions, it is even possible to link it to a partial differential operator with some effective diffusion tensor that is piecewise constant with respect to the coarse mesh $\mathcal{T}_H$; similar to Section 1.3.3. Unfortunately, a more precise discussion is beyond the scope of this lecture and we refer to [2] for the details.

References

1. I. Babuska and R. Lipton. Optimal local approximation spaces for generalized finite element methods with application to multiscale problems. *Multiscale Model. Simul.*, 9(1):373--406, 2011.
2. D. Gallistl and D. Peterseim. Computation of local and quasi-local effective diffusion tensors in elliptic homogenization. *ArXiv e-prints*, August 2016.

3. L. Grasedyck, I. Greff, and S. Sauter. The al basis for the solution of elliptic problems in heterogeneous media. *Multiscale Model. Simul.*, 10(1):245--258, 2012.
4. P. Henning, A. Målqvist, and D. Peterseim. A localized orthogonal decomposition method for semi-linear elliptic problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, eFirst, 2013.
5. P. Henning and D. Peterseim. Oversampling for the multiscale finite element method. *Multiscale Modeling & Simulation*, 11(4):1149--1175, 2013.
6. Patrick Henning, Philipp Morgenstern, and Daniel Peterseim. Multiscale partition of unity. In Michael Griebel and Marc Alexander Schweitzer, editors, *Meshfree Methods for Partial Differential Equations VII*, volume 100 of *Lecture Notes in Computational Science and Engineering*, pages 185--204. Springer International Publishing, 2015.
7. T. Y. Hou and P. Liu. Optimal Local Multi-scale Basis Functions for Linear Elliptic Equations with Rough Coefficient. *ArXiv e-prints*, August 2015.
8. R. Kornhuber, D. Peterseim, and H. Yserentant. An analysis of a class of variational multiscale methods based on subspace decomposition. November 2016. Submitted for publication.
9. Ralf Kornhuber and Harry Yserentant. Numerical homogenization of elliptic multiscale problems by subspace decomposition. *Multiscale Modeling & Simulation*, 14(3):1017--1036, 2016.
10. A. Målqvist and D. Peterseim. Localization of Elliptic Multiscale Problems. *Mathematics of Computation (in press), also ArXiv e-prints*, 1110.0692, 2011.
11. H. Owhadi. Multigrid with rough coefficients and Multiresolution operator decomposition from Hierarchical Information Games. *ArXiv e-prints*, March 2015.
12. H. Owhadi, L. Zhang, and L. Berlyand. Polyharmonic homogenization, rough polyharmonic splines and sparse super-localization. *ESAIM: Math. Model. Numer. Anal.*, eFirst, 2013.
13. Houman Owhadi. Bayesian numerical homogenization. *Multiscale Modeling & Simulation*, 13(3):812--828, 2015.

Appendix A
Functional analytic preliminaries

The discussion of well-posedness of PDEs as well as the analysis of variational discretization schemes strongly rely on tools from functional analysis. In this chapter, we will briefly recall some of these tools.

A.1 Abstract linear spaces

A.1.1 Normed linear spaces and inner product spaces

Definition A.1 (vector space). A set X together with the mappings $+ : X \times X \to X$ and $\cdot : \mathbb{R} \times X \to X$ is called *(real) vector space*, if the following conditions are satisfied.

1. $(X, +)$ is a commutative group.
2. The scalar-vector-multiplication is associative, i.e., $\alpha(\beta x) = (\alpha\beta)x$ for all $\alpha, \beta \in \mathbb{R}$ and $x \in X$.
3. *Distributivity* holds in the sense that:

$$\alpha(x+y) = \alpha x + \alpha y \quad \text{and} \quad (\alpha+\beta)x = \alpha x + \beta x$$

for all $\alpha, \beta \in \mathbb{R}$ and $x, y \in X$.

Definition A.2 (convexity). A set $M \subseteq X$ is called *convex*, if for every $x, y \in M$ and for all $\lambda \in (0,1)$,

$$\lambda x + (1-\lambda) y \in M.$$

Definition A.3 (subspace). A set $M \subseteq X$ is a *subspace of* X, if it holds:

1. $0 \in M$ and
2. $\forall x, y \in M \; \forall \alpha \in \mathbb{R} \quad \alpha x + y \in M$.

The second property implies that supspaces are convex.

Definition A.4 (scalar/inner product (spaces)). Given a vector space X, a function

$$\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R}$$

is called *scalar product*, if for all $x, y, z \in X$ and for all $\alpha \in \mathbb{R}$

1. $\langle \alpha x + y, z \rangle = \alpha \langle x, z \rangle + \langle y, z \rangle$ (linearity),
2. $\langle x, y \rangle = \langle y, x \rangle$ (symmetry),
3. $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ if and only if $x = 0$ (positive definiteness).

A vector space X with scalar product $\langle \cdot, \cdot \rangle$ is called *pre-Hilbert space* or *inner product space* and is written as $(X, \langle \cdot, \cdot \rangle)$.

Inner product spaces allow a notion of orthogonality.

Definition A.5 (orthogonality). Two vectors $x, y \in X$ are *orthogonal* if

$$\langle x, y \rangle = 0.$$

One also writes $x \perp y$ or even $x \perp_X y$ with emphasis on the inner-product space.

Example A.1 (inner product spaces).

1. $\mathbb{R}^n$ with the standard Euclidean scalar product defined by

$$\langle x, y \rangle := x \cdot y := x^{\mathrm{T}} y \quad \text{for } x, y \in \mathbb{R}^n,$$

2. the space of quadratic summable sequences

$$\ell^2 := \left\{ \left(x_j\right)_{j \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \,\middle|\, \sum_{j=1}^{\infty} x_j^2 < \infty \right\}$$

with scalar product

$$\langle (x_j), (y_j) \rangle := \sum_{j=1}^{\infty} x_j y_j \quad \text{for } (x_j), (y_j) \in \mathbb{R}^{\mathbb{N}},$$

3. The space of quadratic summable functions

$$L^2(D) := \left\{ f : D \to \mathbb{R} \,\middle|\, f \text{ Lebesgue measurable and } \int_D |f|^2 \, dx < \infty \right\}$$

with L^2-scalar product

$$\langle f, g \rangle := \int_D f(x) g(x) \, dx \quad \text{for } f, g \in L^2(D).$$

The spaces L^p or ℓ^p for $p \neq 2$ are not inner product spaces.

Definition A.6 (norm, normed linear space). Given a vector space X, a function

$$\|\cdot\| : X \to \mathbb{R}$$

is called *norm* if for all $x, y \in X$ and for all $\alpha \in \mathbb{R}$ it holds

1. $\|\alpha x\| = |\alpha| \, \|x\|$;
2. $\|x + y\| \leq \|x\| + \|y\|$;
3. $\|x\| = 0$ implies $x = 0$.

The pair $(X, \|\cdot\|)$ is called *normed linear space* (NLS).

A seminorm is a norm with the property (c) removed.

Remark A.1. For every inner product space $(X, \langle \cdot, \cdot \rangle)$, the function

$$\|\cdot\| : X \to [0, \infty), \quad x \mapsto \sqrt{\langle x, x \rangle}$$

defines a norm. This norm is called *induced by the scalar product* $\langle \cdot, \cdot \rangle$ or *the norm associated to the scalar product* $\langle \cdot, \cdot \rangle$. Hence, every inner product space canonically defines a normed linear space.

Theorem A.1 (Cauchy-Schwarz inequality). *Let $(X, \langle \cdot, \cdot, \rangle)$ be an inner product space and $\|\cdot\|$ be the induced norm. Then, for all $x, y \in X$, the inequality*

$$\langle x, y \rangle \leq \|x\| \, \|y\|$$

holds. Equality,

$$\langle x, y \rangle = \|x\| \, \|y\| ,$$

holds if and only if $x = 0$ or $y = \lambda x$ for some $\lambda \geq 0$.

Theorem A.2 (Parallelogram equality). *Let $(X, \langle \cdot, \cdot, \rangle)$ be an inner product space and $\|\cdot\|$ be the induced norm. For every $x, y \in X$ it holds*

$$\|x + y\|^2 + \|x - y\|^2 = 2 \|x\|^2 + 2 \|y\|^2 . \tag{A.1}$$

Conversely, if $\|\cdot\|$ is a norm that satisfies the parallelogram equality, then there exists a scalar product $\langle \cdot, \cdot \rangle$ which induces $\|\cdot\|$.

Remark A.2 (Inner Product Spaces in $\mathbb{R}^d$.). The space $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ is an inner product space, if and only if there is a symmetric positive definite matrix $A \in \mathbb{R}^{d \times d}$, such that for all $x, y \in \mathbb{R}^d$ it holds

$$\langle x, y \rangle = x^T A y.$$

A.1.2 Hilbert and Banach spaces

Throughout this section, $(X, \|\cdot\|)$ is a normed linear space.

Definition A.7 (Cauchy sequence). A sequence $\left(x_j\right) \in X^{\mathbb{N}}$ is called *Cauchy sequence (CS) in X*, if for every $\varepsilon > 0$ exists an $n_\varepsilon \in \mathbb{N}$, such that for all $j,k \geq n\varepsilon$ it holds

$$\left\|x_j - x_k\right\| < \varepsilon.$$

Definition A.8 (convergence, limit). A sequence $\left(x_j\right) \in X^{\mathbb{N}}$ is called *convergent*, if there exists some $x \in X$, such that for every $\varepsilon > 0$ there exists some $n_\varepsilon \in \mathbb{N}$, such that for all $n \geq n_\varepsilon$ it holds

$$\|x - x_n\| < \varepsilon.$$

In this case, x is called *limit* of (x_j) and written as

$$x = \lim_{j\to\infty} x_j \quad \text{or} \quad (x_j) \to x\,.$$

Remark A.3 (uniqueness of the limit). The limit of a convergent sequence is unique.

Definition A.9 (complete spaces). A normed linear space $(X, \|\cdot\|)$ is *complete* if every Cauchy sequence $(x_j) \in X^{\mathbb{N}}$ in X has a limit in X. A complete inner product space is called *Hilbert space* (HS). A complete normed linear space is called *Banach space* (BS).

Remark A.4. 1. Every Hilbert space is a Banach space.
2. Every inner product space can be completed to a Hilbert space and this extension is unique up to different names.
3. Every normed linear space can be completed to a Banach space and this extension is unique up to different names.
4. Any normed linear space (resp. inner product space) can be considered as dense subspaces of a Hilbert space (resp. Banach space).

Definition A.10 (closed sets). A set $M \subseteq X$ is called *closed*, if the limit of every convergent sequence $\left(x_j\right) \in M^{\mathbb{N}}$ is also in M.

Definition A.11 (complete sets). A set $M \subseteq X$ is called *complete*, if every Cauchy sequence in M is convergent with a limit in M.

Remark A.5. In Banach spaces, every closed set is complete and vice versa.

A.1.3 Best approximation in Hilbert spaces

In this subsection, we will introduce orthogonal projections onto convex subsets in Hilbert spaces. The following concepts and results will be used for studying the errors of finite element approximations in a very abstract form.

Definition A.12 (distance and best approximation). Let $(X, \|\cdot\|)$ be a normed linear space and $K \subset X$ be a nonvoid subset. Then for every $x \in X$ the *distance of x and K* is given by

$$\operatorname{dist}(x,K) := \operatorname{dist}_{\|\cdot\|}(x,K) := \inf_{y\in K} \|x-y\| \,.$$

The (possibly empty) set

$$\mathcal{P}_K(x) := \{y \in K \,|\, \|y-x\| = \operatorname{dist}(x,K)\}$$

is called *set of best approximations of x in K* or *proxima of x in K*.

Theorem A.3. *Given an inner product space $(X,\langle\cdot,\cdot\rangle)$ and a convex, nonvoid subset $K \subset X$, every $x \in X$ and $y \in K$ satisfy*

$$y \in \mathcal{P}_K(x) \quad \Longleftrightarrow \quad \forall z \in K,\ \langle x-y, z-y\rangle \le 0\,.$$

Proof. $\Longrightarrow$) Some elementary algebra with norms and scalar products shows for all $x,y,w \in X$ that

$$\begin{aligned}\|x-w\|^2 - \|x-y\|^2 \\ \|x-y+y-w\|^2 - \|x-y\|^2 &= \|x-y\|^2 + 2\langle x-y, y-w\rangle + \|y-w\|^2 - \|x-y\|^2 \\ &= \|y-w\|^2 - 2\langle x-y, w-y\rangle\,.\end{aligned}$$

For $y \in \mathcal{P}_K(x)$ and $w \in K$, the left-hand side is non-negative. Given any $z \in K$ and $0 < \lambda \le 1$, $w = \lambda z + (1-\lambda)y \in K$. Therefore

$$0 \le \|x - \lambda z - (1-\lambda)y\|^2 - \|x-y\|^2 = -2\lambda\langle x-y, z-y\rangle + \lambda^2 \|y-z\|^2\,.$$

After divison by $\lambda > 0$, this results in

$$\langle x-y, z-y\rangle \le \frac{1}{2}\lambda \|y-z\|^2\,.$$

For $\lambda \searrow 0$ the right-hand side tends to zero. This proves the asserted inequality.
$\Longleftarrow$) For every $z \in K$, a Cauchy inequality shows

$$\begin{aligned}\|x-y\|^2 &= \langle x-y, x-z\rangle + \underbrace{\langle x-y, z-y\rangle}_{\le 0} \\ &\le \|x-y\|\,\|x-z\|\,.\end{aligned}$$

Consequently,

$$\|x-y\| \le \|x-z\| \quad \text{for all } z \in K,$$

and hence $y \in \mathcal{P}_K(x)$.

The following result provides uniqueness of the best approximation in Hilbert spaces.

Theorem A.4 (Chebyshev property). *Let $K \subset X$ be a nonvoid, closed, convex subset in a Hilbert space $(X,\langle\cdot,\cdot\rangle)$. Then, for every $x \in X$, the set of its best approximations*

$$\mathcal{P}_K(x) = \{P_K(x)\}$$

contains exactly one element $P_K(x)$. The hereby defined mapping

$$P_K : X \to K, \quad x \mapsto P_K(x)$$

is idempotent (i.e., $P_K^2 := P_K \circ P_K = P_K$), non-expansive (i.e., P_K is Lipschitz continuous with a Lipschitz constant ≤ 1) and monotone (i.e., $0 \leq \langle P_K x - P_K y, x - y\rangle$ for all $x, y \in X$).

Proof. 1. Proof of uniqueness. According to Theorem A.3 all $x \in X$ and $y_1, y_2 \in \mathcal{P}_K(x)$ satisfy

$$\langle x - y_1, y_2 - y_1\rangle \leq 0 \quad \text{and} \quad \langle x - y_2, y_1 - y_2\rangle \leq 0\,.$$

The sum of these inequalities gives

$$\|y_2 - y_1\|^2 \leq 0\,,$$

hence $y_1 = y_2$.

2. Proof of existence. $d := \operatorname{dist}(x, K)$ is the infimum of the set

$$\{\|x - y\| \mid y \in K\}$$

so there exists a sequence $(y_j)_{j\in\mathbb{N}} \subset K$ such that

$$d^2 \leq \left\|x - y_j\right\|^2 \leq d^2 + 1/j \quad \text{for any } j \in \mathbb{N}\,.$$

According to the parallelogram equality (A.1) on page 75, it holds, for $j, k \in \mathbb{N}$, that

$$\left\|y_j - y_k\right\|^2 + \left\|y_j + y_k - 2x\right\|^2 = 2\left\|y_j - x\right\|^2 + 2\|y_k - x\|^2\,.$$

The right-hand side is

$$\text{RHS} \leq 4d^2 + 2/j + 2/k.$$

Since K is convex, $\frac{1}{2}(y_j + y_k) \in K$, and

$$\text{LHS} = \left\|y_j - y_k\right\|^2 + 4\left\|(y_j + y_k)/2 - x\right\|^2 \geq \left\|y_j - y_k\right\|^2 + 4d^2..$$

The previous estimates prove

$$\left\|y_j - y_k\right\|^2 \leq 2/j + 2/k\,.$$

Hence $(y_j)_{j\in\mathbb{N}}$ is a Cauchy sequence in the closed set K, thus converges towards a limit point $y \in K$ which satisfies

$$d \le \|x-y\| \le \left\|x-y_j\right\| + \left\|y-y_j\right\| \le \sqrt{d^2+1/j} + \left\|y-y_j\right\| .$$

For $j \longrightarrow \infty$, the right-hand side tends to d and shows

$$\|x-y\| = d \text{ and so } y \in \mathcal{P}_K(x).$$

3. Proof of idempotence. Obviously $P_K(x) = x$ for $x \in K$. This proves the claimed $P_K^2(x) = P_K(x)$ for all $x \in X$.

4. Proof of monotonicity. Theorem A.3 shows for every $x, y \in X$ and their best approximations $P_K(x), P_K(y) \in K$ that

$$\begin{aligned} \langle x - P_K(x), P_K(y) - P_K(x) \rangle &\le 0 , \\ \langle y - P_K(y), P_K(x) - P_K(y) \rangle &\le 0 . \end{aligned}$$

The sum of these inequalities gives

$$\begin{aligned} \|P_K(y) - P_K(x)\|^2 &= \langle P_K(y) - P_K(x), P_K(y) - P_K(x) \rangle \\ &\le \langle y - x, P_K(y) - P_K(x) \rangle . \end{aligned} \tag{A.2}$$

This proves monotonicity of P_K.

5. Proof of non-expansiveness. The application of (A.2) and a Cauchy-Schwarz inequality lead to

$$\|P_K(y) - P_K(x)\|^2 \le \|y-x\| \, \|P_K(y) - P_K(x)\| .$$

Consequently,

$$\|P_K(y) - P_K(x)\| \le \|y-x\| ,$$

i.e., P_K is non-expansive.

An immediate consequence of Theorem A.4 is the separation theorem in Hilbert spaces. Notice that this is a very special case of the famous separation principle in Banach spaces which follows from Hahn-Banach extension theorem.

Corollary A.1. *Let K be some closed convex nonvoid set in the Hilbert space X and $x \in X \setminus K$. Then there exist some direction $\ell \in X$ and some real numbers α and β such that*

$$\langle \ell, y \rangle \le \alpha < \beta = \langle \ell, x \rangle \quad \textit{for all } y \in K .$$

Proof. Given the best approximation $z := P_K(x)$ of x in K there holds for any $y \in K$ that

$$\langle x - z, y - z \rangle \le 0 .$$

With $\ell := x - z \in X$ this is equivalent to

$$\langle \ell, y\rangle \le \langle \ell, z\rangle = \langle \ell, x\rangle - \|\ell\|^2 =: \alpha < \beta := \langle \ell, x\rangle$$

Exercise A.1. Prove that closed convex sets in Hilbert spaces are weakly closed.

The following corollary considers the case where the convex subset K is a subspace and the best approximation is characterised by orthogonality.

Corollary A.2. *Let M be a closed subspace of the Hilbert space $(X, \langle\cdot,\cdot\rangle)$. Then $P = P_M$ is a linear, continuous mapping onto M, and for all $x \in X$ it holds*

$$(x - P(x)) \perp M\,.$$

Furthermore, $Q := 1 - P$ is a linear, continuous mapping onto the orthogonal complement

$$M^\perp := \{x \in X \,|\, x \perp M\} \text{ of } M \text{ in } X\,.$$

For any $x \in X$ there exist unique vectors

$$p_x = P(x) \in M \quad \text{and} \quad q_x = Q(x) \in M^\perp \quad \text{with } x = p_x + q_x\,.$$

Proof. Proof of orthogonality. Given $x \in X$ and $z \in M$ set $p_x := Px$ and consider $w = p_x \pm z \in M$ in the characterising inequality of the best approximation. Then it holds

$$\langle x - p_x, w - p_x\rangle \le 0\,.$$

For $w := p_x + z$ this reads

$$\langle x - p_x, z\rangle \le 0.$$

For $w := p_x - z$ this reads

$$\langle x - p_x, -z\rangle \le 0.$$

Alltogether,

$$\langle x - p_z, z\rangle = 0\,.$$

Since z is arbitrary, $q_x := x - p_x \perp M$.

Proof of linearity. Let $\alpha,\beta \in \mathbb{R}$ and $x, y \in X$ with $x = p_x + q_x$, $y = p_y + q_y$ where $p_x, p_y \in M$ and $q_x, q_y \in M^\perp$. Then we have

$$\alpha x + \beta y - (\alpha p_x + \beta p_y) = \alpha q_x + \beta q_y \in M^\perp\,.$$

Hence the characterisation of Theorem A.3 shows for $\alpha p_x + \beta p_y \in M$ that

$$\alpha p_x + \beta p_y \in \mathcal{P}_M(\alpha x + \beta y) = \{P_M(\alpha x + \beta y)\}\,,$$

which reads

$$\alpha P_M x + \beta P_M y = P_M(\alpha x + \beta y).$$

An important conclusion is that

$$X = M \oplus M^{\perp}$$

for any closed subspace M of X and its orthogonal complement $M^{\perp}$ and this is stable by orthogonality

$$\|x\|^2 = \|Px\|^2 + \|Qx\|^2 .$$

A.1.4 Dual spaces and Riesz representation

Definition A.13 (Dual space). Given a normed linear space X, the vector space

$$X^{\star} := \{F : X \to \mathbb{R} \,|\, F \text{ linear and continuous}\}$$

with canonical addition and (outer) multiplication is called the *(continous) dual space of* X. Any $F \in X^{\star}$ has a norm

$$\|F\|_{X^{\star}} := \sup_{x \in X \setminus \{0\}} \frac{F(x)}{\|x\|_X} .$$

Example A.2. Let $x \in X$ be a vector of an inner product space $(X, \langle \cdot, \cdot \rangle)$ then, by a Cauchy inequality, the mapping

$$\langle x, \cdot \rangle : X \to \mathbb{R}, y \mapsto \langle x, y \rangle$$

is an element of the dual space $X^{\star}$. The Riesz representation theorem states that all linear functionals in a Hilbert space are of that form.

Theorem A.5 (Riesz representation theorem). *Let $(X, \langle \cdot, \cdot \rangle)$ be a Hilbert space. Then the Riesz mapping*

$$\mathcal{R} : X \to X^{\star}, x \mapsto \langle x, \cdot \rangle$$

is a norm isomorphism. In particular, for every $F \in X^{\star}$ there exists a unique $x =: \mathcal{R}^{-1} F \in X$ with $\langle x, \cdot \rangle = F$ and $\|x\|_X = \|F\|_{X^{\star}}$. The vector $x := \mathcal{R}^{-1} F$ is called Riesz representation *of $F \in X^{\star}$.*

Proof. Proof of isometry. The mapping $\mathcal{R}$ inherits its linearity from the scalar product. The Cauchy-Schwarz inequality and the discussion of the equality in there leads to the identity

$$\|\mathcal{R}x\|_{X^{\star}} = \sup_{y \in X, \|y\|_X = 1} |\langle x, y \rangle| = \|x\|_X \qquad \text{for all } x \in X.$$

Hence $\|\mathcal{R}\| = 1$ and $\mathcal{R}$ is an isometry.

Proof of injectivity. Since $\|\mathcal{R}x\|_{X^{\star}} = \|x\|_X$ for all x, $\mathcal{R}x = 0$ implies $x = 0$. Hence, $\mathcal{R}$ is injective.

Proof of surjectivity. Let $F \in X^\star \setminus \{0\}$ be a non-vanishing element of the dual space. Then the closed subspace $M = \ker(F)$ is a genuine subset of X and hence $M^\perp \setminus \{0\} \neq \emptyset$. For a $z \in M^\perp \setminus \{0\}$ define

$$x = F(z)\, z / \|z\|^2 \in X\,.$$

Due to the linearity of F, every $y \in X$ satisfies

$$F(z)y - F(y)z \in M\,.$$

Finally the identity

$$\langle x, y\rangle \|z\|^2 = \langle F(z)z, y\rangle = \langle z, F(z)y\rangle = \langle z, F(y)z\rangle = F(y)\|z\|^2$$

proves

$$F = \langle x, \cdot\rangle.$$

A.2 Lebesgue spaces and test functions

The content of this subsection on higher analysis is partly copied from the book of Evans [3] to which we refer for details, references, and proofs.

Lebesgue's measure theory provides a powerful integration theory in $\mathbb{R}^d$ and is preferred over the Riemann integral, since it provides certain "completeness" properties, i.e., appropriate limits of integrable functions are integrable, a property that the Riemann integral does not have. We recall a few basic facts and definitions.

Definition A.14 (measurable sets). The measurable subsets of $\mathbb{R}^d$ form the smallest countable additive σ algebra that includes all open and closed sets. The Lebesgue measure $|M|$ of a measurable set $M \subset \mathbb{R}^d$ extends the d-dimensional volume of balls and each subset of a measurable set of measure zero is measurable and of measure zero.

Definition A.15 (measurable functions). A function $f : \mathbb{R}^d \to \mathbb{R}$ is called a measurable function if

$$f^{-1}(\omega) \quad \text{is a measurable set}$$

for every open subset $\omega \subset \mathbb{R}$.

The following result illustrates that measurable functions are continuous up to small sets.

Theorem A.6 (Theorem of Lusin). *Given a Lebesgue measurable set $D \subset \mathbb{R}^d$ with Lebesgue measure $|D| < \infty$ and a bounded and measurable function $f : D \to \mathbb{R}$, and $\varepsilon > 0$, there exists a compact subset $K \subset D$ with $|D \setminus K| < \varepsilon$ such that $f|_K \in C(K)$.*

Definition A.16 (summable function). A measurable function is summable in D, written $f \in L^1(D)$, if

$$\int_D |f|\,dx < \infty.$$

A measurable function is locally summable, written as $f \in L^1_{\mathrm{loc}}(D)$, if it is summable on all compact subsets $\omega \subset\subset D$, i.e., $f|_\omega \in L^1(\omega)$.

Definition A.17 (almost everywhere (a.e.)). Two functions $f, g : D \to \mathbb{R}$ are said to be equal almost everywhere, written $f = g$ a.e., if the set $\{f \neq g\} := \{x \in D \,|\, f(x) \neq g(x)\}$ has measure zero, i.e.,

$$f = g \text{ a.e. if } |\{f \neq g\}| = 0\,.$$

In the context of Lebesgue functions, f and g are identified if they coincide almost everywhere.

We identify two functions f and g that satisfy $\|f - g\|_{L^p(D)} = 0$ and say $f = g$ almost everywhere (a.e.). For example, take $n = 1$, $D = (-1, 1)$ and functions

$$f(x) = \begin{cases} 1 \text{ for } x \geq 0, \\ 0 \text{ for } x < 0, \end{cases} \quad \text{and} \quad g(x) = \begin{cases} 1 \text{ for } x > 0, \\ 0 \text{ for } x \leq 0. \end{cases}$$

Theorem A.7 (Lebesgue differentiation theorem). *Let $f : \mathbb{R}^d \to \mathbb{R}$ be locally summable.*

1. Then for a.e. point $x_0 \in \mathbb{R}^d$,

$$\lim_{r \to 0} |B(x_0, r)|^{-1} \int_{B(x_0,r)} f\,dx = f(x_0)\,.$$

2. In fact, almost every point $x_0 \in \mathbb{R}^d$ is a Lebesgue point of f, i.e.,

$$\lim_{r \to 0} |B(x_0, r)|^{-1} \int_{B(x_0,r)} |f(x) - f(x_0)|\,dx = 0.$$

Given an open subset $D \subset \mathbb{R}^d$ and $1 \leq p < \infty$, define

$$\|f\|_{L^p(D)} := \left(\int_D |f|^p \right)^{1/p},$$

and, for $p = \infty$, define

$$\begin{aligned} \|f\|_{L^\infty(D)} &:= \operatorname{ess\,sup} f \\ &:= \inf \{\eta > 0 \,|\, |\{x \in D \,|\, |f(x)| > \eta\}| = 0\}\,. \end{aligned}$$

Then, for $p \in \mathbb{N} \cup \{\infty\}$, $\|\cdot\|_{L^p(D)}$ defines a semi-norm on the space

$$\widehat{L^p}(D) := \{f : D \to \mathbb{R} \text{ measurable } \mid \|f\|_{L^p(D)} < \infty\},$$

in particular it fulfills the triangle inequality.

Theorem A.8 (Minkowski inequality). *Assume* $1 \le p \le \infty$ *and* $u, v \in \widehat{L}^p(D)$*. Then it holds*

$$\|u+v\|_{L^p(D)} \le \|u\|_{L^p(D)} + \|v\|_{L^p(D)}\,.$$

To obtain a normed vector space we factorise $\widehat{L}(D)$ by the kernel of $\|\cdot\|_{L^p(D)}$,

$$\begin{aligned}\ker(\|\cdot\|_{L^p(D)}) &:= \{u \in \widehat{L}^p(D) \,|\, \|u\|_{L^p(D)} = 0\}\\ &= \{u \in \widehat{L}^p(D) \,|\, u = 0 \text{ a.e. in } D\},\end{aligned}$$

and define the space of Lebesgue functions as equivalence classes almost everywhere,

$$L^p(D) := \widehat{L}^p(D)/\ker(\|\cdot\|_{L^p(D)})\,.$$

Two Lebesgue functions coincide (i.e., they belong to the same class of Lebesgue functions) if they are equal almost everywhere.

Theorem A.9 (Hölder inequality). *Assume* $1 \le p, q \le \infty$*,* $1/p + 1/q = 1$*. Then for* $u \in L^p(D), v \in L^q(D)$*, it holds*

$$\|uv\|_{L^1(D)} \le \|u\|_{L^p(D)}\,\|v\|_{L^q(D)}\,.$$

If $p = q = 2$, the Hölder inequality is known as *Schwarz inequality* or *Cauchy-Schwarz inequality*.

A very important fact is the following theorem.

Theorem A.10. *For* $D \subseteq \mathbb{R}^d$ *open and* $1 \le p \le \infty$*,* $L^p(D)$ *is a Banach space.*

Proof. A proof employs the dominated convergence theorem and is left as an exercise.

Given any non-empty open set $D \subset \mathbb{R}^d$, recall

$$C^\infty(D) := \bigcap_{k \in \mathbb{N}o} C^k(D)$$

and let

$$\mathrm{d}(D) := C_c^\infty(D) := \{f \in C^\infty(\mathbb{R}^d) : \operatorname{supp} f \subset\subset D\}$$

denote the space of test functions. The support of f is

$$\operatorname{supp} f = \overline{\{x \in \mathbb{R}^d \,|\, f(x) \neq 0\}} \tag{A.3}$$

and $\subset\subset$ denotes a compact subset. This means that $f \in \mathrm{d}(D)$ vanishes outside a big ball and also in a neighbourhood of the boundary.

Theorem A.11. *For* $1 \le p < \infty$ *it holds that*

$$\mathcal{D}(D) \text{ is dense in } L^p(D).$$

Proof. For a proof we refer to [7, Ch. 3, Thm. 3.14, p. 69].

Remark A.6. 1. $\mathcal{D}(D)\backslash\{0\}$ does not include any complex differentiable functions as they would be bounded and entire. The theorem of Liouville implies that this function is constant which leads to a contradiction.
2. For every domain $D \subset \mathbb{R}^d$ it holds $\mathcal{D}(D) \subset \mathcal{D}(\mathbb{R}^d)$.
3. For every domain $D \subset \mathbb{R}^d$, Theorem A.11 states that $\mathcal{D}(D)$ is dense in $L^2(D)$. Hence, L^2-functions have no boundary data.

Example A.3. Define functions f and g by

$$f(x) = \begin{cases} \exp(-1/x) & \text{for } x > 0, \\ 0 & \text{for } x \leq 0 \end{cases}$$

and

$$g(x) = f(x)\, f(1-x).$$

Then it holds $\operatorname{supp}(g) = [0,1]$ and hence $g \in \mathcal{D}(-1,2)$.

Definition A.18. For each $\varepsilon > 0$, define the standard mollifier η_ε by

$$\eta_\varepsilon := C(d)/\varepsilon^d \begin{cases} \exp\left(\varepsilon^2/(|x|^2 - \varepsilon^2)\right) & \text{if } |x| < \varepsilon, \\ 0 & \text{if } |x| \geq \varepsilon. \end{cases} \tag{A.4}$$

The functions η_ε are C^∞ and satisfy

$$\int_{\mathbb{R}^d} \eta_\varepsilon \, dx = 1 \quad \text{and} \quad \operatorname{supp} \eta_\varepsilon = \overline{B(0,\varepsilon)}.$$

We set

$$L^p_{\mathrm{loc}}(D) = \{f : D \to \mathbb{R} \text{ measurable, such that } f|_\omega \in L^p(\omega) \text{ for all } K \subset\subset D\}.$$

Definition A.19. If $f : D \to \mathbb{R}$ is locally integrable, i.e., $f \in L^1_{\mathrm{loc}}(D)$, define its mollification

$$f^\varepsilon := \eta_\varepsilon * f \quad \text{in} \quad D_\varepsilon := \{x \in D \,|\, \operatorname{dist}(x, \partial D) > \varepsilon\} \tag{A.5}$$

for any $x \in D_\varepsilon$, hence $B(x,\varepsilon) \subseteq D$, by

$$f^\varepsilon(x) = \int_D \eta_\varepsilon(x-y) f(y)\, dy = \int_{B(0,\varepsilon)} \eta_\varepsilon(y) f(x-y)\, dy.$$

Theorem A.12 (Properties of mollifiers). *It holds*

1. If f is locally integrable then $f^\varepsilon \in C^\infty(D_\varepsilon)$
2. If f is in $L^1(D)$ then $f^\varepsilon \to f$ a.e. as $\varepsilon \to 0$

3. *If $f \in C(D)$, then $f^\varepsilon \to f$ uniformly on compact subsets of D.*
4. *If $1 \le p < \infty$ and $f \in L^p_{loc}(D)$, then $f^\varepsilon \to f$ in $L^p_{loc}(D)$.*
5. *If $1 \le p < \infty, k \in \mathbb{N}_0$ and $f \in W^{k,p}_{loc}(D)$, then $f^\varepsilon \to f$ in $W^{k,p}_{loc}(D)$.*
6. *If $D^\alpha f$ is locally integrable for some derivative D^α of f then*

$$\eta_\varepsilon * (D^\alpha f) = D^\alpha(\eta_\varepsilon * f).$$

7. *If $f : (a,b) \to \mathbb{R}$ is monoton, so is $\eta_\varepsilon * f$.*
8. *If $f : D \to \mathbb{R}$ is convex, so is $\eta_\varepsilon * f$.*
9. *If $f : B(0,1) \to \mathbb{R}$ is asymmetric, namely $f(x) = -f(-x)$, then $\eta_\varepsilon * f$ is asymmetric in $B(0, 1-\varepsilon)$.*

A.3 Sobolev spaces

A.3.1 Weak derivatives and Sobolev functions

Definition A.20 (Weak derivative). Suppose that $D \subseteq \mathbb{R}^d$ is open and $f \in L^1_{\ell oc}(D)$. A function $g_j \in L^1_{\text{loc}}(D)$ is called *weak derivative* of f with respect to x_j for $j \in \{1,...,d\}$, if for all $\varphi \in \mathcal{D}(D)$ there holds

$$\int_D f \frac{\partial \varphi}{\partial x_j}\, dx = -\int_D g_j \varphi\, dx. \tag{A.6}$$

In this case we say that f is *weakly differentiable* with respect to x_j and set

$$\frac{\partial f}{\partial x_j} = g_j.$$

If all weak derivatives $\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_d}$ exist, then we say that f is weakly differentiable and define

$$\nabla f = \Big(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_d}\Big),$$

the weak derivative of f.

Example A.1 *Let $D = (-1,1)$ and $f(x) := |x|$ for $x \in D$. Then f is weakly differentiable with*

$$\nabla f = \begin{cases} +1 \text{ for } x > 0, \\ -1 \text{ for } x < 0. \end{cases}$$

Proof. Let $\varphi \in D(D)$. There holds by integration by parts

$$\begin{aligned}\int_D |x|\varphi'(x)\,dx &= -\int_{(-1,0)} x\varphi'(x)\,dx + \int_{(0,1)} x\varphi'(x)\,dx \\ &= \int_{(-1,0)} \varphi(x)\,dx - [x\varphi(x)]_{-1}^{0} - \int_{(0,1)} \varphi(x)\,dx + [x\varphi(x)]_{0}^{1} \\ &= \int_{(-1,0)} \varphi(x)\,dx - \int_{(0,1)} \varphi(x)\,dx \\ &= -\int_{(-1,1)} \operatorname{sgn}(x)\varphi(x)\,dx,\end{aligned}$$

where we used $\varphi(-1) = \varphi(1) = 0$.

Lemma A.1 (Uniquness of the weak derivative). *The weak derivative is (up to sets of measure zero) uniquely defined.*

Proof. If g_j and h_j are weak partial derivatives of $f \in L^1_{\mathrm{loc}}(D)$ with respect to x_j then by (A.6) we get

$$\int_D (g_j - h_j)\varphi\,dx = 0 \quad \text{for all } \varphi \in D(D).$$

Owing to Theorem A.11 there holds $g_j = h_j$ almost everywhere in D.

Lemma A.2. *For $f \in C^1(\overline{D})$, the classical (strong) derivative and the weak derivative coincide (almost everywhere).*

Proof. Let us assume that ∂D is sufficient regular so that Gauss' theorem holds, i.e.,

$$\int_D F\,dx = \int_{\partial D} F \cdot n\,ds \quad \text{for all } F \in C^1(\overline{D})^d.$$

Set $F = \varphi f e_j$, where $(e_j : j = 1,2,...,d)$ is the canonical basis of $\mathbb{R}^d$, i.e., the j-th component of e_j equals 1 and and all other components are equal to 0. Since $\varphi|_{\partial D} = 0$, we have $\operatorname{div} F = \frac{\partial}{\partial x_j}(f\varphi) = \frac{\partial f}{\partial x_j}\varphi + f\frac{\partial \varphi}{\partial x_j}$ and $F|_{\partial D} = 0$. Therefore, we get

$$\int_D \frac{\partial f}{\partial x_j}\varphi + f\frac{\partial \varphi}{\partial x_j}\,dx = 0 \quad \text{for all } \varphi \in \mathcal{D}(D). \tag{A.7}$$

Hence $\frac{\partial f}{\partial x_j}$ is the weak partial derivative of f with respect to x_j.

Remark A.7. Suppose that $D \subseteq \mathbb{R}^d$ is open and connected, and $f \in L^1_{\mathrm{loc}}(D)$ is weakly differentiable with $\nabla f = 0$. Then f is constant.

Remark A.8. Lipschitz continuous functions are weakly differentiable.

Example A.2 *The function $\log\big|\log|x|\big|$, $x \in B_{1/2}(0) \subseteq \mathbb{R}^d$, $d = 2,3$, has a singularity at $x = 0$, but is weakly differentiable.*

Definition A.21. Let $D \subseteq \mathbb{R}^d$ be open. A function $f : D \to \mathbb{R}$ is called *Sobolev function* if f is weakly differentiable and if there exists p with $1 \le p \le \infty$ such that $f \in L^p(D)$ and $\nabla f \in L^p(D)^d$.

We will use multiindeces to describe partial derivatives of any order. In more details, $\alpha = (\alpha_1, \ldots, \alpha_d), \alpha \in \mathbb{N}_0^d$ for $\alpha_1, \ldots, \alpha_d \in \mathbb{N}_0$, and

$$|\alpha| = \alpha_1 + \cdots + \alpha_d, \quad \alpha! = \alpha_1! \cdots \alpha_d!, \quad D^\alpha := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}.$$

The summation $\sum_{|\alpha|=0}^m$ means the sum over all such mutliindices with $|\alpha| = 0, 1, 2, \ldots, m$. For $m = 0$ this is only one, $\alpha = (0,0)$; for $m = 1$ this is $\alpha = (0,0), (1,0), (0,1)$ and for $m = 2$ this is $\alpha = (0,0), (1,0), (0,1), (2,0), (1,1), (0,2)$. Compare the related notation for the functional matrix D and the Hessian D^2.

Definition A.22 (Higher weak derivative). Given a multiindex $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d$, we define

$$D^\alpha \varphi = \frac{\partial^{|\alpha|} \varphi}{\partial x^\alpha} := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} ... \partial x_d^{\alpha_d}}$$

for $\varphi \in C^{|\alpha|}(D)$. We say that $f \in L^1_{\ell oc}$ possesses the *weak partial derivative* $\frac{\partial^\alpha}{\partial x^\alpha} f$, if there exists a function $g \in L^1_{\text{loc}}(D)$ such that

$$\int_D g\, \varphi \, dx = (-1)^{|\alpha|} \int_D f \, \frac{\partial^{|\alpha|} \varphi}{\partial x^\alpha} \, dx \quad \text{for all } \varphi \in \mathcal{D}(D).$$

In this case we set $\frac{\partial^\alpha f}{\partial x^\alpha} = g_\alpha$.

A.3.2 Sobolev spaces

Definition A.23. Let $D \subseteq \mathbb{R}^d$ be open, k a non-negative integer, and $f \in L^1_{\text{loc}}(D)$. Suppose that f possesses weak partial derivatives $\frac{\partial^\alpha f}{\partial x^\alpha}$ for all $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \le k$. We define the *Sobolev norm*

$$\|f\|_{W^{k,p}(D)} = \Big(\sum_{\alpha \in \mathbb{N}_0^d, |\alpha| \le k} \Big\| \frac{\partial^{|\alpha|} f}{\partial x^\alpha} \Big\|^p_{L^p(D)} \Big)^{1/p} \quad \text{for } 1 \le p < \infty$$

and

$$\|f\|_{W^{k,p}(D)} = \max_{\alpha \in \mathbb{N}_0^d, |\alpha| \le k} \Big\| \frac{\partial^{|\alpha|} f}{\partial x^\alpha} \Big\|_{L^p(D)} \quad \text{for } p = \infty.$$

In either case we define the *Sobolev space* $W^{k,p}(D)$ as

$$W^{k,p}(D) = \{ f \in L^1_{\text{loc}}(D) : \|f\|_{W^{k,p}(D)} < \infty \}$$

and the *periodic Sobolev space*

$$W^{k,p}_{\#}(D) = \{f \in L^1_{\mathrm{loc}}(\mathbb{R}^d) : \ f \in W^{k,p}(D),\ f \text{ periodic w.r.t } D\}$$

Theorem A.13. *For $D \subseteq \mathbb{R}^d$ open, $k \in \mathbb{N}_0$, and $1 \le p \le \infty$ the Sobolev space $W^{k,p}(D)$ is a Banach space.*

Proof. Exercise.

Remark A.9 (Notation). For $k \in \mathbb{N}$, it is customary to write

$$H^k(D) = W^{k,2}(D).$$

Theorem A.14. *For $k \in \mathbb{N}$, the bilinear form $\langle \cdot, \cdot \rangle_{H^k(D)} : H^k(D) \times H^k(D) \to \mathbb{R}$ given by*

$$\langle u, v \rangle_{H^k(D)} := \sum_{\alpha \in \mathbb{N}_0^d, |\alpha| \le k} \left\langle \frac{\partial^{|\alpha|} u}{\partial x^\alpha}, \frac{\partial^{|\alpha|} v}{\partial x^\alpha} \right\rangle$$

for $u, v \in H^k(D)$ defines a scalar product. The spaces $H^k(D)$ are a Hilbert spaces.

Proof. Exercise.

The following theorem is due to Meyers and allows for an alternative definition of $W^{k,p}(D)$ in case $1 \le p < \infty$ (cf. Remark A.15 below).

Theorem A.15. *Let $D \subseteq \mathbb{R}^d$ be an open set and $1 \le p < \infty$. Then $C^\infty(D) \cap W^{k,p}(D)$ is dense in $W^{k,p}(D)$, i.e., given any $f \in W^{k,p}(D)$ and $\varepsilon > 0$, there exists $g \in C^\infty(D)$ such that*

$$\|f - g\|_{W^{k,p}(D)} < \varepsilon.$$

Proof. See [3].

We will frequently make use of the following *semi-norms*: For $f \in W^{k,p}(D)$ set

$$|f|_{W^{k,p}(D)} = \Big(\sum_{\alpha \in \mathbb{N}_0^d, |\alpha| = k} \Big\| \frac{\partial^{|\alpha|} f}{\partial x^\alpha} \Big\|^p_{L^p(D)} \Big)^{1/p} \quad \text{for } 1 \le p < \infty$$

and

$$|f|_{W^{k,p}(D)} = \max_{\alpha \in \mathbb{N}_0^d, |\alpha| = k} \Big\| \frac{\partial^{|\alpha|} f}{\partial x^\alpha} \Big\|_{L^p(D)} \quad \text{for } p = \infty.$$

A.3.3 Lipschitz domains and integration by parts

Definition A.24. A set $D \subseteq \mathbb{R}^d$ is called *Lipschitz domain*, if it is open and connected and if for each $x \in \partial D$ there exists a coordinate transformation $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ (i.e.,

$\Phi(y) = Ay + z$ with $A \in \mathbb{R}^{d\times d}$ orthogonal and $z \in \mathbb{R}^d$), some parameter $\delta > 0$, and a Lipschitz continuous function $\eta : [-\delta,\delta]^{d-1} \to \mathbb{R}$ such that

$$\begin{aligned} D \cap B_\delta(x) &= \Phi(\{(y_1,...,y_d) \in \mathbb{R}^d : \eta(y_1,...,y_{d-1}) > y_d\}) \cap B_\delta(x), \\ \partial D \cap B_\delta(x) &= \Phi(\{(y_1,...,y_d) \in \mathbb{R}^d : \eta(y_1,...,y_{d-1}) = y_d\}) \cap B_\delta(x), \\ B_\delta \backslash \overline{D}(x) &= \Phi(\{(y_1,...,y_d) \in \mathbb{R}^d : \eta(y_1,...,y_{d-1}) < y_d\}) \cap B_\delta(x). \end{aligned}$$

Roughly speaking, Lipschitz domains are open and connected sets, whose boundary is locally parameterized by a Lipschitz continuous function and which lies on one side of its boundary.
Lipschitz domains allow for the following integration by parts formula with boundary terms:

Theorem A.16 (Integration by parts). *For a bounded Lipschitz domain $D \subseteq \mathbb{R}^d$ and functions $f, g \in C^1(D) \cap C(\overline{D})$ there holds*

$$\int_D \Big(\frac{\partial f}{\partial x_j} g + f \frac{\partial g}{\partial x_j}\Big) dx = \int_{\partial D} f g n_j \, ds \quad \text{for } 1 \le j \le d. \tag{A.8}$$

Here n_j is the j-th component of the outer unit normal n to ∂D.

Proof. We refer to standard literature, e.g. [5, Theorem3.34], for a proof of Theorem A.16 and only mention that it is based on the one-dimensional integration by parts formula.

Remark A.10. The outer unit normal n on ∂D is only defined almost everywhere on ∂D (with respect to the surface measure). This however is sufficient to define the integral on the right-hand side of (A.8).

Remark A.11. Suppose that $D \subseteq \mathbb{R}^d$ is a bounded Lipschitz domain. Then $W^{1,\infty}(D) = Lip(D) = \{f \in C(D) : f \text{ is Lipschitz continuous in } D\}$ in the sense that $f \in W^{1,\infty}(D)$ if and only if there exists $f^* \in Lip(D)$ such that $f = f^*$ almost everywhere.

A.3.4 Traces of Sobolev functions

We want to extend the integration by parts formula (A.8) to functions $f \in W^{1,p}(D)$. Therefore, we need to understand in which sense Sobolev functions have well defined boundary values. Recall that it does not make sense to talk about boundary values for functions in $L^p(D)$.

Throughout this subsection, we assume that D is a bounded Lipschitz domain in $\mathbb{R}^d$.

Definition A.25 (Trace operator). Suppose $1 \le p \le \infty$. The *trace operator* γ is defined for $f \in W^{1,p}(D)$ and $x \in \partial D$ as

$$(\gamma f)(x) := \begin{cases} \lim_{r\to 0} |D \cap B_r(x)|^{-1} \int_{D\cap B_r(x)} f(y)\,dy & \text{if this limit exists,} \\ 0 & \text{otherwise.} \end{cases} \tag{A.9}$$

Remark A.12.

1. We interpret the trace γf of f as the boundary values of f. For $f \in C(\overline{D}) \cap W^{1,p}(D)$ and $x \in \partial D$, there holds $(\gamma f)(x) = f(x)$.
2. The trace operator extends the restriction $\cdot|_{\partial D} : f \mapsto f|_{\partial D}$ from $C(\overline{D}) \cap W^{1,p}(D)$ to $W^{1,p}(D)$.

Theorem A.17 (Bounded traces). *For $1 \le p \le \infty$ the trace operator γ defines a bounded, linear mapping $\gamma : W^{1,p}(D) \to L^p(D)$, i.e., γ is linear and $\|\gamma f\|_{L^p(\partial D)} \le C_\gamma \|f\|_{W^{1,p}(D)}$ for $C_\gamma > 0$ and all $f \in W^{1,p}(D)$.*

Remark A.13. The traces of $H^1(D)$-functions are dense in $L^2(\partial D)$.

Theorem A.18 (Generalized integration by parts). *Suppose $f \in C^1(D) \cap C(\overline{D})$ and $u \in W^{1,p}(D)$ for $1 \le p \le \infty$. Then*

$$\int_D u \frac{\partial f}{\partial x_j}\,dx + \int_D \frac{\partial u}{\partial x_j} f\,dx = \int_{\partial D} f n_j \gamma(u)\,ds. \tag{A.10}$$

If $F \in C^1(D)^d \cap C(\overline{D})^d$ and $v \in W^{1,p}(D)$

$$\int_D v \operatorname{div} F\,dx + \int_D \nabla v \cdot F\,dx = \int_{\partial D} F \dot{n} \gamma(v)\,ds. \tag{A.11}$$

In order to deal with Dirichlet type boundary conditions on some (closed) part Γ_0 of ∂D in the Poisson problem, it is useful to define subspaces of Sobolev spaces for which functions vanish on Γ_0.

Definition A.26. Let Γ_0 be a closed subset of ∂D. Define

$$W_D^{1,p}(D) = \{f \in W^{1,p}(D) : \gamma f|_{\Gamma_0} = 0\}.$$

If $p = 2$ we may write $H_D^1(D) = W_D^{1,2}(D)$.

Remark A.14 (Notation).

1. If $\Gamma_0 = \partial D$ we also write $W_0^{1,p}(D)$ instead of $W_D^{1,p}(D)$.
2. We sometimes only write $f|_{\Gamma_0}$ instead of $(\gamma f)|_{\Gamma_0}$.

Remark A.15 (Sobolev Spaces by Density). An alternative way of defining the Sobolev spaces reads as follows. Let

$$W^{m,p}(\mathbb{R}^d) := \overline{\mathcal{D}(\mathbb{R}^d)}^{\|\cdot\|_{W^{m,p}(\mathbb{R}^d)}}$$

denote the completion of the normed linear space $\mathcal{D}(\mathbb{R}^d)$ endowed with the Sobolev norm. The restriction to D leads to an equivalent definition of Sobolev spaces

$$W^{m,p}(D) = \{f|_D \,|\, f \in W^{m,p}(\mathbb{R}^d)\}.$$

The completion of $\mathcal{D}(D)$ with respect to Sobolex norms yields spaces

$$W_0^{m,p}(D) := \overline{\mathcal{D}(D)}^{\|\cdot\|_{W^{m,p}(D)}}. \tag{A.12}$$

Clearly, $W_0^{m,p}(D) \subseteq W^{m,p}(D)$. Note that the definitions of $W_0^{1,p}(D)$ in (A.12) and Definition A.26 for $\Gamma_0 = \partial D$ are equivalent.

A.3.5 Important theorems

We have seen that Sobolev functions are special Lebesgue functions, at least, they are not more general measures or other strange objects. They can be identified by a function, namely the precise representation. This lecture series will make use of a few properties of Sobolev functions only and leaves the technical proofs to the PDE literature.

Throughout this section, we assume that D is a bounded Lipschitz domain in $\mathbb{R}^d$.

If a Sobolev function is sufficiently often weakly differentiable then it equals a continuous function almost everywhere.

Theorem A.19 (Sobolev embeddings). *Let $D \subset \mathbb{R}^d$ be a Lipschitz domain and $k, p, d \in \mathbb{N}$. If $kp > d$, then there exists a continuous embedding $W^{k,p}(D) \hookrightarrow C(\overline{D})$. If $kp < d$, then there exists a continuous embedding $W^{k,p}(D) \hookrightarrow L^{dp/(d-p)}(\overline{D})$.*

For detailled proofs of these embeddings, we refer to [4, Sec. 7.7.].

Theorem A.20 (Stein Extension Theorem). *Given a bounded Lipschitz domain $D \subset\subset \widehat{D}$ compactly included in a bounded open set $\widehat{D}$, there exists a bounded linear extension operator*

$$E : W^{1,p}(D) \to W^{1,p}(\widehat{D})$$

such that, for each $u \in W^{1,p}(D)$,

1. *$Eu = u$ a.e. in D;*
2. *$\operatorname{supp}(Eu) \subset \widehat{D}$;*
3. *$\|Eu\|_{W^{1,p}(\mathbb{R}^d)} \le C(p, D, \widehat{D}) \|u\|_{W^{1,p}(D)}$.*

Proof. We refer to [8] (or [3]) for a proof.

Theorem A.21 (Rellich-Kondrachov Compactness Theorem). *Assume D is a bounded open subset of $\mathbb{R}^d$, with a Lipschitz boundary ∂D. Suppose $1 \le p < n$. Then*

$$W^{1,p}(D) \overset{c}{\hookrightarrow} L^q(D)$$

for each $1 \le q < p^* := pd/(d-p)$.

Proof. For a proof we refer to [3].

This compactness result gives rise to the following inequality.

Theorem A.22 (Poincaré inequality). *Given a bounded, connected and open subset of $D \subset \mathbb{R}^d$ with a Lipschitz boundary ∂D, and $1 \leq p \leq \infty$. Then there exists a constant $C(d,p,D) < \infty$ with*

$$\|f - |D|^{-1} \int_D f\,dx\|_{L^p(D)} \leq C_P(d,p,D)\,\|Df\|_{L^p(D)}$$

for every $f \in W^{1,p}(D)$.

Proof. For a proof we refer to [3].

A similar result can be established for function that vanishes at some part of the boundary; the L^p-norm of functions in $W^{1,p}_D(D)$ can be bounded by the L^p-norm of their weak gradients.

Theorem A.23 (Friedrichs inequality). *Given a bounded, connected and open subset of $D \subset \mathbb{R}^d$ with a Lipschitz boundary ∂D, and $\Gamma_0 \subset \partial D$ with positive surface measure $|\Gamma_0|$ and $1 \leq p \leq \infty$. Then there exists a constant $C(d,p,\Gamma_0,D) < \infty$ with*

$$\|f\|_{L^p(D)} \leq C_F(d,p,\Gamma_0,D)\,\|Df\|_{L^p(D)}$$

for every $f \in W^{1,p}_{\Gamma_0}(D) := \{u \in W^{1,p}(D) \,|\, u = 0 \text{ on } \Gamma_0\}$.

Proof. For a proof we refer to [2] where this inequality is called Poincaré inequality.

In two particular situations for $p = 2$, the constants C_P and C_F can be estimated exampleicitly.

Theorem A.24 (Payne-Weinberger). *Given a convex bounded open set $D \subseteq \mathbb{R}^d$ of diameter* $\operatorname{diam}(D) := \sup\{|x-y| \,|\, x,y \in D\}$ *it holds*

$$C_P(n,2,D) \leq \operatorname{diam}(D)/\pi.$$

In other words,

$$\left\| f - |D|^{-1} \int_D f\,dx \right\|_{L^2(D)} \leq \operatorname{diam}(D)/\pi\,\|Df\|_{L^2(D)}$$

for any $f \in H^1(D)$.

Proof. The original proof in [6] relies on a weighted one-dimensional estimate plus a nice intersection argument. The given application for $d \geq 3$ contains some mistake which can be removed [1]. The assumption holds for all $d \geq 1$ and is sharp in the sense that the constant cannot be better under the assumption to have D arbitrary convex.

Theorem A.25 (Friedrichs inequality in $H^1_0(D)$). *For $\Gamma_0 = \partial D$ it holds $C(d,2,\partial D,D) \leq \operatorname{width}(D)/\pi$ for the size*

$$\operatorname{width}(D) := L := \beta - \alpha$$

defined as the smallest length $L := \beta - \alpha$ such that the open set D lies between two parallel hyperplanes $\{x \cdot \nu = \alpha\}$ and $\{x \cdot \nu = \beta\}$ of distance L for some unit vector $\nu \in \mathbb{R}^d$. In other words,

$$\|f\|_{L^2(D)} \leq \operatorname{width}(D)/\pi \|Df\|_{L^2(D)}$$

for all $f \in H^1_0(D)$.

Proof. Without loss of generality, let the coordinate system be with $\nu = (0,\dots,0,1)$ and

$$D \subseteq \widehat{D} := \{(\xi, x_n) \in \mathbb{R}^d \,|\, 0 < x_n < L\}.$$

Any test function $f \in \mathcal{D}(D)$ is extended by zero to $\widehat{D}$. For any $\xi \in \mathbb{R}^{d-1}$, the partial function

$$f(\xi,\cdot) : (0,L) \to \mathbb{R}$$

belongs to $\mathcal{D}(0,L) \subseteq H^1_0(0,L)$ and the one-dimensional Friedrichs inequality results in

$$\int_0^L |f(\xi,x_n)|^2 \, dx_n \leq (L/\pi)^2 \int_0^L |\frac{\partial f}{\partial x_n}(\xi,x_n)|^2 \, dx_d.$$

An integration with respect to $\xi \in \mathbb{R}^d$ and $\left|\frac{\partial f}{\partial x_n}\right| \leq |Df|$ lead to

$$\int_{\widehat{D}} |f(x)|^2 \, dx \leq (L/\pi)^2 \int_{\widehat{D}} |Df(x)|^2 \, dx$$

for any $f \in \mathcal{D}(D)$. A density argument with $\overline{\mathcal{D}(D)}^{\|\cdot\|_{H^1(D)}}$ proves the assumption.

A.4 Well-posedness of linear problems

The analysis and the computation of PDEs is based on a weak (or ultra-weak) form which involves some bilinear form

$$b : X \times Y \to \mathbb{R} \tag{A.13}$$

on some real-valued Sobolev spaces X and Y which are reflexive Banach spaces or even Hilbert spaces. Recall that a Banach space is reflexive if it can be identified with its bidual $X^{**} := (X^*)^*$ (Hilbert spaces are reflexive by Riesz' representation theorem A.5).

While $X = Y$ for many simple elliptic second order PDEs (e.g. the Poission problem), the Banach spaces X and Y may be very different in other circumstances (e.g. for ultra weak formulations).

This section discusses the general well-posedness of linear problems of the primal form

$$b(x,\cdot) = F \tag{A.14}$$

or the dual form

$$b(\cdot,y) = G. \tag{A.15}$$

The point is that, given $F \in Y^*$ in the dual Y^* of Y (resp. given $G \in X^*$ in the dual of X) there exists some unique solution $x \in X$ (resp. $y \in Y$) of the primal problem (A.14) (resp. the dual problem (A.15)).

Besides the unique solvability of the two problems (A.14) and (A.15), the perturbation analysis is relevant. Well-posedness means that the solution x of (A.14) (resp. y of (A.15)) depends continuously on the right-hand side $F \in Y^*$ (resp. $G \in X^*$). It will be a consequence of the fundamental properties of linear operators between Banach spaces that unique solvability of (A.14) (resp. (A.15)) readily implies the well-posedness, the unique solvability of the primal problem is equivalent to the unique sovability of the dual problem, and all this is equivalent to the inf-sup conditions

$$0 < \alpha := \inf_{x\in X\setminus\{0\}} \frac{\|b(x,\cdot)\|_{Y^*}}{\|x\|_X} = \inf_{y\in Y\setminus\{0\}} \frac{\|b(\cdot,y)\|_{X^*}}{\|y\|_Y}. \tag{A.16}$$

To illustrate this important condition, suppose for the moment that the bounded linear operator

$$B_1 : X \to Y^*, \quad x \mapsto b(x,\cdot) \tag{A.17}$$

is continously invertable. In other words, the linear operator

$$B_1^{-1} : Y^* \to X$$

is bounded. The operator norm of B_1^{-1} reads

$$\|B_1^{-1}\| = \sup_{F\in Y^*\setminus\{0\}} \frac{\|B_1^{-1}F\|_X}{\|F\|_{Y^*}}.$$

Given the solution $x = B_1^{-1}F$ of (A.14),

$$\|B_1^{-1}F\|_X = \|x\|_X \ \text{ and } \ \|F\|_{Y^*} = \|B_1 x\|_{Y^*}.$$

Since all $F \in Y^*$ can be written in this form, it follows

$$\frac{1}{\|B_1^{-1}\|} = \inf_{F\in Y^*\setminus\{0\}} \frac{\|F\|_{Y^*}}{\|B_1 F\|_X} = \inf_{x\in X\setminus\{0\}} \frac{\|B_1 x\|_{Y^*}}{\|x\|_X} = \alpha.$$

In other words, the inf-sup constant (A.16) equals the reciprocal of the norm of B_1^{-1}. The linear operator

$$B_2 : Y \to X^*, \quad y \mapsto b(\cdot,y) \tag{A.18}$$

is the dual of B_1 for reflexive Banach spaces where X and Y are identified with their respective bidual spaces X^{**} and Y^{**}.

In fact, the dual operator $B_1^* : Y^{**} \to X^*$ of B_1 is defined by

$$B_1^* : Y^{**} \to X^*, \quad \Lambda \mapsto \Lambda \circ B_1$$

via the composition $\Lambda \circ B_1 : X \to \mathbb{R}$, $x \mapsto \Lambda(B_1 x)$ which maps any $\Lambda \in Y^{**}$ (this is a bounded linear functional $\Lambda : Y^* \to \mathbb{R}$) onto its value at $B_1 x = b(x, \cdot) \in Y^*$. The identification of Y with Y^{**} can be written as the evaluation functional δ_y at some $y \in Y$, i.e.,

$$\Lambda(F) \equiv \delta_y(F) := F(y) \quad \text{for any } F \in Y^{**}.$$

For any $y \in Y$, δ_y belongs to Y^{**}. For a reflexive Banach space Y, those evaluation functionals describe all elements in Y^{**}, i.e., the mapping

$$\delta : Y \to Y^{**}, \quad y \mapsto \delta_y.$$

is surjective. This implies, for all $x \in X$, that

$$(B_1^*(\delta_y))(x) = \delta_y(B_1 x) = \delta_y(b(x, \cdot)) = b(x, y) = (B_2(y))(x).$$

Since $x \in X$ is arbitrary, this reads $B_1^* \delta_y = B_2 y$. The identification $Y = Y^{**}$ and the aforementioned calculations allow the notation

$$B_1^* : Y \to X^*, \quad y \mapsto b(y, \cdot)$$

and hence $B_1^* = B_2$. The same argument for $X = X^{**}$ shows $B_2^* = B_1$. This is behind the equality in (A.16), namely

$$\|B_1^{-1}\| = \|(B_1^*)^{-1}\| = \|B_2^{-1}\| = \alpha^{-1}.$$

We summarize the previous discussion in the subsequent theorem.

Theorem A.26 (). *Let X and Y reflexive Banach spaces and let $b : X \times Y \to \mathbb{R}$ be a bounded bilinear form with X and Y as above. Then the following conditions are pairwise equivalent:*

(a) $\forall F \in Y^* \exists! x \in X,\ b(x, \cdot) = F$;
(b) $\forall G \in X^* \exists! y \in Y,\ b(\cdot, y) = F$;
(c) The infsup conditions (A.16) *are satisfied.*

For a complete proof of the theorem, we refer to classical textbooks in Functional Analysis. An important special case for the present lecture is when $X = Y$ and for some Hilbert space X.

Definition A.27 (Ellipticity). Some bilinear form $a : X \times X \to \mathbb{R}$ is called X-elliptic if there exists $\alpha > 0$ such that, for all $v \in X$, it holds

$$\alpha \|v\|_X^2 \leq a(x, x).$$

For an X-elliptic bilinear form a, Theorem A.26 readily yields a famous result due to Lax and Milgram which plays a dominant role in the existence theory of elliptic PDEs.

Corollary A.3 (Lax-Milgram theorem). *Suppose X is a Hilbert space, $F \in X^*$ and $a : X \times X \to \mathbb{R}$ is an elliptic continuous bilinear form. Then there is a unique $u \in X$ such that*

$$a(u,v) = F(v) \quad \textit{for all } v \in X. \tag{A.19}$$

Moreover,

$$\|u\|_X \leq \frac{1}{\alpha} \|F\|_{X^*}.$$

The difference between the Lax-Milgram theorem and Riesz' representation theorem is that a does not need to be symmetric in the Lax-Milgram theorem. For the sake of completeness, we present a proof below.

Proof. For all $v \in X$ we know that $a(v,\cdot) \in X^*$. Hence, by Riesz' representation theorem, there exists $Av = \mathcal{R}^{-1}a(v,\cdot) \in X$ for all $v \in X$ such that

$$\langle Av, w\rangle_X = a(v,w) \quad \text{for all } w \in X.$$

Moreover, there exists $f = \mathcal{R}^{-1}F \in X$ such that

$$F(w) = \langle f, w\rangle_X \quad \text{for all } w \in X.$$

The mapping $A : v \to Av$ is linear and continuous with

$$\|Av\|_X = \|Ra(v,\cdot)\|_X = \|a(v,\cdot)\|_{X^*} = \sup_{0\neq w\in X} \frac{a(v,w)}{\|w\|_X} \leq \sup_{0\neq w\in X} \frac{\beta\|v\|_X\|w\|_X}{\|w\|_X} = \beta\|v\|_X \tag{A.20}$$

where we used that the operator R defined in Theorem A.5 is an isometry. With this notation (A.19) is equivalent to finding $u \in X$ such that

$$Au = f.$$

We want to show that the mapping

$$T_\delta : X \longrightarrow X, \quad v \mapsto v - \delta(Av - f)$$

is a contraction for an appropriate $\delta > 0$, i.e., T_δ satisfies $\|T_\delta v - T_\delta w\|_X \leq q\|v - w\|_X$ with some $0 < q < 1$ for all $v, w \in X$. Given $v, w \in X$, set $e = v - w$. Then

$$\begin{aligned}
\|T_\delta v - T_\delta w\|_X^2 &= \|v - \delta(Av - f) - w + \delta(Aw - f)\|_X^2 \\
&= \|v - w - \delta(Av - Aw)\|_X^2 \\
&= \|v - w - \delta A(v - w)\|_X^2 \\
&= \|e - \delta Ae\|_X^2 \\
&= \|e\|_X^2 - 2\delta \underbrace{\langle e, Ae\rangle_X}_{=\langle Ae,e\rangle_X = a(e,e)} + \delta^2 \|Ae\|_X^2 \\
&= \|e\|_X^2 - 2\delta \underbrace{a(e,e)}_{\geq \alpha\|e\|_X^2} + \delta^2 \underbrace{\|Ae\|_X^2}_{\leq \beta^2\|e\|_X^2 \text{ by (A.20)}} \\
&\leq \|e\|_X^2 - 2\delta\alpha\|e\|_X^2 + \delta^2\beta^2\|e\|_X^2 \\
&= (1 - \delta\alpha + \delta^2\beta^2)\|v - w\|_X^2.
\end{aligned}$$

For δ such that

$$0 < q^2 = 1 - 2\alpha\delta + \delta^2\beta^2 < 1,$$

e.g. $\delta = \frac{3\alpha}{2\beta^2}$ (recall that $\alpha \leq \beta$), the operator T_δ is a contraction. By Banach's fixed point theorem, there exists a unique fixed point $u \in X$. For this particular u, we have

$$u = T_\delta u = u - \delta(Au - f)$$

and, hence, $Au = f$ (or equivalently $a(u, v) = F(v)$ for all $v \in X$). Choosing $v = u$, we verify that

$$\alpha\|u\|_X^2 \leq a(u,u) = F(u) \leq \|F\|_{X^*}\|u\|_X,$$

which finishes the proof. □

References

1. M. Bebendorf. A note on the Poincaré inequality for convex domains. *J. Anal. Appl.*, 22:751--756, 2003.
2. Susanne C. Brenner and L. Ridgway Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.
3. Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.
4. D. Gilbarg and N.S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Springer-Verlag, 1983.
5. William McLean. *Strongly elliptic systems and boundary integral equations*. Cambridge University Press, Cambridge, 2000.
6. L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Arch. Rat. Mech. Anal.*, 5:286--292, 1960.
7. Walter Rudin. *Real and complex analysis, 3rd ed.* McGraw-Hill, Inc., New York, NY, USA, 1987.
8. E. M. Stein. *Singular integrals and differentiability properties of functions*. Princeton Mathematical Series, No. 30. Princeton University Press, Princeton, N.J., 1970.