CHAPTER 3

Ellipticity: Laplace, Poisson and diffusion equations

1. Ellipticity

1.1. The notion of ellipticity. Now we say that a linear differential operator \( P \) of order \( k \) defined on an open set \( U \) is elliptic at \( x \in U \) if \( \sigma_d(x, \xi) \neq 0 \) for all \( \xi \in \mathbb{R}^d \setminus \{0\} \) (or for a more complicated geometrical setting for all \( (x, \xi) \in T^*U, \xi \neq 0 \)). We say that it is elliptic on \( U \) if it is elliptic for all \( x \in U \). Equivalently, a linear differential operator \( P \) is elliptic if all hypersurfaces are non-characteristic.

Example 1.1. For a first order operator written as

\[
P = \sum_{i=1}^{\ell} b_i(x) \partial_i + \text{lower order terms}...
\]

the definition of ellipticity can never be satisfied as soon as \( \ell \geq 2 \) since \( \sigma_p(x, \xi) \) cancels on \( b(x)^\perp \) with \( b(x) = (b_1(x), \ldots, b_\ell(x)) \). But \( \ell \geq 2 \) is required for a PDE, unless we consider an ODE.

Example 1.2. Consider now a second order linear operator written as

\[
P = -\sum_{i,j=1}^{\ell} a_{ij}(x) \partial^2_{ij} + \text{lower order terms}...
\]

A first remark is that since \( \partial^2_{ij} = \partial^2_{ji} \) for \( C^2 \) functions we can assume w.l.o.g. that the matrix \( A := (a_{ij})_{ij} \) is symmetric (replacing it by \( \bar{a}_{ij} = (a_{ij} + a_{ji})/2 \) if necessary). Then the ellipticity means that

\[
\forall \xi \in \mathbb{R}^d, \quad \sum_{i,j=1}^{\ell} a_{ij}(x)\xi_i\xi_j = \xi^T A(x)\xi \neq 0
\]

Hence by continuity and connectedness the sign is constant on \( \mathbb{R}^d \setminus \{0\} \), and by convention (changing \( A \) to \(-A\) if necessary) we are thus reduced to the case where \( A(x) \) is a positive definite matrix, i.e. all its eigenvalues are positive. The (opposite of the) Laplacian operator \(-\Delta\) is thus the most simple example of elliptic operator.

The most characteristic feature of ellipticity is so-called elliptic regularity. Let us first do some heuristics. The motivation for this can be found precisely in the above properties of the symbol. In the case of a constant coefficient differential operator of the form

\[
P = \sum_{|\alpha|=k} a_\alpha \partial^\alpha_x
\]
then if \( \sigma_P(x, \xi) = \sigma_P(\xi) = 0 \), then for all \( \lambda \in \mathbb{R} \) we deduce that
\[
u_\lambda(x) = e^{i\lambda x \cdot \xi}
\]
is a solution to \( Pu_\lambda(x) = 0 \). In particular, as \( \lambda \to \infty \), one can construct more and more oscillatory solutions. In the general, non constant-coefficient case, if \( \sigma_P(x_0, \xi) = 0 \) at some point \( x_0 \), then \( u_\lambda(x) \) for large \( \lambda \) can still be viewed as an approximate highly oscillatory solution near \( x_0 \). Thus, operators \( P \) which fail to be elliptic have highly oscillatory solutions (or approximate solutions) of \( Pu = 0 \), and this intuition can be “localised” in a region. It turns out that conversely, the condition of ellipticity “prohibits” highly oscillatory behaviour for solutions of \( Pu = f \) if oscillation are not present on the right hand side. This is at the heart of elliptic regularity.

1.2. Elliptic regularity in the whole Euclidean space. One can distinguish between qualitative and quantitative versions of the elliptic regularity principle. A qualitative version is the statement that if \( f \) is \( C^\infty \) in a neighborhood of a point \( x \in \mathcal{U} \) in an open set \( \mathcal{U} \), and \( u \) satisfies \( Pu = f \), then \( u \) is \( C^\infty \) in a neighbourhood of \( x \) as well. Already, we should have been a bit careful when we said “\( u \) satisfies”. What kind of function are we assuming \( u \) to be a priori? To keep this discussion elementary, let us say that \( u \) is a classical solution, that is to say, it has at least the order of differentiability appearing in the equation, i.e. it is \( C^2 \), and let us first consider \( \mathcal{U} = \mathbb{R}^\ell \). To summarise:

**Proposition 1.3 (Qualitative statement of elliptic regularity).** Let \( u : \mathbb{R}^\ell \to \mathbb{R} \) be \( C^2 \) with \( \Delta u = f \), where \( f \) is \( C^\infty \) on \( \mathbb{R}^\ell \). Then \( u \) is \( C^\infty \) on \( \mathbb{R}^\ell \).

The proof of this proposition makes use of crucial new notions, which are different from the previous chapters, which can be organised along the following method:

1. The use of Sobolev spaces which measure regularity by means of integrals (as opposed to pointwise like \( C^k \) spaces).
2. The proof of a (fundamental!) a priori estimate.
3. The justification of the a priori estimate by a regularization argument (painful technically but absolutely necessary to the rigor of the proof).

1.2.1. Sobolev spaces. Let us first hence review some basics on Sobolev spaces. The proper abstract framework for defining properly these spaces is the theory of distributions of Laurent Schwartz, however they can also be introduced by duality in a more naive way. They appeared first (in this way) in the seminal paper of Leray 1934 on the Navier-Stokes equation, and in then works of Sobolev. “Generalised functions” (which finally lead mathematicians to the notion of distributions) were first motivated by the Heaviside function, the Dirac distribution, and the dipole in physics (each of them being the generalized derivative of the previous one...).

We can define here the Sobolev space \( H^s(\mathbb{R}^\ell) \) on \( \mathbb{R}^\ell \), for \( s \in \mathbb{N} \), as a subset of \( L^2(\mathbb{R}^\ell) \) by defining for \( g \in C^\infty_c(\mathbb{R}^\ell) \) (infinitely differentiable with compact support) the norm
\[
\|g\|_{H^s(\mathbb{R}^\ell)} := \left( \sum_{|\alpha| \leq s} \|\partial_\alpha^\alpha g\|_{L^2(\mathbb{R}^\ell)}^2 \right)^{\frac{1}{2}}
\]
and taking the closure of \( C^\infty_c(\mathbb{R}^\ell) \) in \( L^2(\mathbb{R}^\ell) \) for this norm:
\[
H^s(\mathbb{R}^\ell) = \overline{C^\infty_c(\mathbb{R}^\ell)}^{||.||_{H^s(\mathbb{R}^\ell)}} \subset L^2(\mathbb{R}^\ell).
\]

A second way to define this space is as follows: \( g \in L^2(\mathbb{R}^\ell) \) belongs to \( H^s(\mathbb{R}^\ell) \) iff there is a constant \( C > 0 \) so that
\[
\forall \varphi \in C^\infty_c(\mathbb{R}^\ell), \forall |\alpha| \leq s, \quad \left| \int_{\mathbb{R}^\ell} g(x) \partial_\alpha^x \varphi(x) \, dx \right| \leq C \|\varphi\|_{L^2(\mathbb{R}^\ell)}
\]
and the smaller such constant is precisely the \( H^s(\mathbb{R}^\ell) \) norm of \( g \) (this is the definition used by Leray).

A third definition can be obtained from Fourier analysis: \( g \in H^s(\mathbb{R}^\ell) \subset L^2(\mathbb{R}^\ell) \) iff there is a constant \( C > 0 \) so that
\[
\left( \int_{\mathbb{R}^\ell} |\hat{g}(\xi)|^2 (1 + |\xi|^2)^s \, d\xi \right)^{\frac{1}{2}} \leq C
\]
where \( \hat{g} \) is the Fourier transform of \( g \), and the smaller such constant is precisely the \( H^s(\mathbb{R}^\ell) \) norm of \( g \). This definition allows to consider any \( s \in \mathbb{R}_+ \) (another way to define \( H^s \) for non-integer \( s \) would be to use the interpolation theory, see the mid-term presentations).

**Exercise 25.** Check that all the three previous definitions are equivalent and construct a Hilbert space, which is dense in \( L^2(\mathbb{R}^\ell) \) (hint: use the approximation of the unit for the last point).

A fundamental tool on Sobolev spaces is the family of Sobolev inequalities: they relate the smoothness measured by integrals in the Sobolev spaces \( H^s \) to the usual \( C^k \) spaces where the smoothness is measured in pointwise form:
\[
\forall s \geq \ell/2, \forall \varepsilon > 0, \ \exists C_\varepsilon > 0, \quad \|u\|_{C^{s-\varepsilon/2}(\mathbb{R}^\ell)} \lesssim C_\varepsilon \|u\|_{H^{s+\varepsilon}(\mathbb{R}^\ell)}.
\]
In this last expression, \( C^\theta \) for \( \theta \in \mathbb{R}_+ \) is a subspace \( C[^\theta] \), with \([\theta]\) the integer part of \( \theta \), where the \([\theta]\)-order derivatives are \((\theta-[\theta])\) H"older continuous. In formula this means:
\[
\|u\|_{C^\theta} := \sum_{|\alpha| \leq [\theta]} \|\partial^\alpha u\|_{L^\infty(\mathbb{R}^\ell)} + \sum_{|\alpha| = [\theta]} \sup_{x \neq y \in \mathbb{R}^\ell} \frac{|\partial^\alpha u(x) - \partial^\alpha u(y)|}{|x - y|}.
\]

We can now define the Sobolev homogeneous norm \( \dot{H}^s(\mathbb{R}^\ell) \) for \( s \in \mathbb{N} \) as the subspace of \( L^2(\mathbb{R}^\ell) \) obtained by closing \( C^\infty_c(\mathbb{R}^\ell) \) for the norm
\[
\|g\|_{\dot{H}^s(\mathbb{R}^\ell)} := \left( \sum_{|\alpha| = s} \|\partial^\alpha_x g\|_{L^2(\mathbb{R}^\ell)}^2 \right)^{\frac{1}{2}}
\]
or equivalently
\[
\forall \varphi \in C^\infty_c(\mathbb{R}^\ell), \forall |\alpha| = s, \quad \left| \int_{\mathbb{R}^\ell} g(x) \partial_\alpha^x \varphi(x) \, dx \right| \leq \|g\|_{\dot{H}^s(\mathbb{R}^\ell)} \|\varphi\|_{L^2(\mathbb{R}^\ell)}
\]
or equivalently
\[
\|g\|_{\dot{H}^s(\mathbb{R}^\ell)} := \left( \int_{\mathbb{R}^\ell} |\hat{g}(\xi)|^2 |\xi|^s \, d\xi \right)^{\frac{1}{2}}.
\]
Exercise 26. In general this formula provides only a semi-norm, however since we restrict to functions in \( L^2(\mathbb{R}^\ell) \), prove that it is a norm due to the integrability restriction.

1.2.2. The key a priori estimate. As a second preliminary step, let us now explain the key a priori estimate which is at the core the elliptic regularity principle. Assume that \( \Delta u = f \) is satisfied in \( \mathbb{R}^\ell \), and that \( u, f \in C^\infty_c(\mathbb{R}^\ell) \). This means that we argue a priori by assuming all the necessary regularity and decay at infinity, i.e. decay in both real and Fourier variables.

We now claim that

\[
\forall s \in \mathbb{N}, \quad \|u\|_{H^{s+2}(\mathbb{R}^\ell)} = \|f\|_{H^s(\mathbb{R}^\ell)}.
\]

The proof is elementary but fundamental and is a simple example of “energy method”.

We square and integrate the equation:

\[
\int_{\mathbb{R}^\ell} |f|^2 \, dx = \int_{\mathbb{R}^\ell} (\Delta u)^2 \, dx
\]

and we perform integration by parts on the RHS:

\[
\int_{\mathbb{R}^\ell} (\Delta u)^2 \, dx =_{\text{de}} \int_{\mathbb{R}^\ell} \sum_{i,j=1}^\ell \partial_i^2 u \partial_j^2 u \, dx =_{\text{IBP}} \int_{\mathbb{R}^\ell} \sum_{i,j=1}^\ell \partial_i^2 u \partial_j^2 u \, dx = \|u\|_{H^2(\mathbb{R}^\ell)}
\]

which proves the claim for \( s = 0 \). Now for a general \( s \in \mathbb{N} \) observe that if we differentiate the equation by \( \partial_x^\alpha \) for some \(|\alpha| \leq s\) we have

\[
\Delta(\partial_x^\alpha u) = \partial_x^\alpha f
\]

which means that the function \( \partial_x^\alpha u \) satisfies the same elliptic equation as \( u \). We are using a fundamental fact here:

\[
[\Delta, \partial_x^\alpha] := \Delta \partial_x^\alpha - \partial_x^\alpha \Delta = 0
\]

i.e. the commutation between the operator defining the equation and the operator according to which we want to estimate the regularity. We hence readily deduce by performing the same energy estimate as before on \( \partial_x^\alpha u \) that

\[
\|\partial_x^\alpha u\|_{H^2(\mathbb{R}^\ell)} = \|\partial_x^\alpha f\|_{L^2(\mathbb{R}^\ell)}
\]

and we conclude the proof of the claim for \( s \in \mathbb{N} \) by summing over all \(|\alpha| = s\).

1.2.3. Justification. As a third and last step, we can now complete the proof of the elliptic regularity by justifying and making rigorous the latter a priori estimate. In deriving (1.1), we have used underlying differentiability assumptions, i.e. a qualitative fact. But in fact, these assumptions can be proven from the “quantitative” statement that appears to rely on it. In other words, we shall see a first instance of the important fact that as soon as objects exists in one side of an a priori identity, they will exist in the other side. Roughly speaking, (in most cases) every a priori control of a quantity by a finite estimate will imply that the quantity indeed exists, i.e. “existence follows from the estimate”.

Let us illustrate this idea. For instance, suppose we know only a priori that \( u \) is a classical solution, i.e. \( C^2 \), whereas \( f \) is assumed again to be \( C^\infty_c \). Consider an approximation of the unit, i.e. consider \( \chi \in C^\infty_c(\mathbb{R}^\ell) \), \( 0 \leq \chi \leq 1 \) with compact support...
in $B(0, r)$ and so that $\chi = 1$ on $B(0, 1/4)$ and $\int \chi = 1$, and then scale it as $\chi_\varepsilon(x) := \varepsilon^{-d} \chi(\varepsilon^{-1} x)$ for $\varepsilon > 0$ (meant to be small), and as $\chi_R = \chi(\varepsilon^{-1} x)$, $R > 0$ (meant to be large). Then define

\[(\text{localisation}) \quad \tilde{u} := u \chi_R, \quad \tilde{f} := \Delta \tilde{u}\]

and

\[(\text{regularisation}) \quad u_\varepsilon := \tilde{u} \ast \chi_\varepsilon, \quad f_\varepsilon = \tilde{f} \ast \chi_\varepsilon\]

where $\ast$ denotes convolution, which satisfy

$$\forall \varepsilon > 0, \quad \Delta u_\varepsilon = f_\varepsilon.$$

Observe that

$$\tilde{f} = u(\Delta \chi_R) + (\Delta u) \chi_R + 2 \nabla u \cdot \nabla \chi_R = u(\Delta \chi_R) + f \chi_R + 2 \nabla u \cdot \nabla \chi_R$$

is $H^{s-1}$ (with compact support) as soon as $\tilde{u} \in H^s$.

Observe that $u_\varepsilon$ has support in $B(0, R(1 + r) + r \varepsilon)$ and is $C^\infty$ thanks to the convolution, and $f_\varepsilon \in C^\infty$ clearly as well. Observe also that $\tilde{u} \in L^2(\mathbb{R}^d)$ thanks to the compact support and the $L^\infty$ bound on this compact set on $u$. We now apply the previous a priori estimate for $s \in \mathbb{N}$ on differences of two mollified solutions:

$$\|u_{\varepsilon_1} - u_{\varepsilon_2}\|_{H^{s+2}(\mathbb{R}^d)} \leq \|f_{\varepsilon_1} - f_{\varepsilon_2}\|_{H^s(\mathbb{R}^d)}.$$

By induction assume that $\tilde{u} \in H^{s+1}$, then $\tilde{f} \in H^s$. Then let us prove that the RHS is converging to zero. It is enough to prove

$$\lim_{\varepsilon \to 0} \|f_\varepsilon - \tilde{f}\|_{H^s(\mathbb{R}^d)} = 0$$

and by taking $\partial_\alpha^Q$, $|\alpha| = s$, it is enough to prove

$$\lim_{\varepsilon \to 0} \|g_\varepsilon - g\|_{L^2(\mathbb{R}^d)} = 0$$

for a general $g \in L^2(\mathbb{R}^d)$. We then write

$$g_\varepsilon(x) - g(x) = \int_{\mathbb{R}^d} g(x - y) \chi_\varepsilon(y) dy - g(x) = \int_{\mathbb{R}^d} (g(x - y) - g(x))^\varepsilon(x)(y) dy$$

where we have used $\int \chi_\varepsilon = 1$. We then take the $L^2$ norm:

$$\|g_\varepsilon - g\|_{L^2(\mathbb{R}^d)} \leq \sup_{|y| \leq \varepsilon} \|g(x - y) - g\|_{L^2(\mathbb{R}^d)} \left( \int_{\mathbb{R}^d} |\chi_\varepsilon(y)| dy \right) = \sup_{|y| \leq \varepsilon} \|g(x - y) - g\|_{L^2(\mathbb{R}^d)}$$

and we finally use the fundamental fact of the theory of Lebesgue integration (on which all mollification arguments ultimately rely on), which is the continuity of the translation operator in all $L^p$ spaces, $p \in [1, +\infty)$:

$$\sup_{|y| \leq \varepsilon} \|g(x - y) - g\|_{L^2(\mathbb{R}^d)} \xrightarrow{\varepsilon \to 0} 0,$$

which concludes the proof.

We now go back to our original problem. As a consequence of the fact that the RHS goes to zero, we deduce that the sequence $u_\varepsilon$ has the Cauchy property in $H^{s+2}(\mathbb{R}^d)$, and since $\tilde{u} \in L^2(\mathbb{R}^d)$, it is also converging in $L^2(\mathbb{R}^d)$ and we deduce that it has the Cauchy
property in $H^{s+2}(\mathbb{R}^\ell)$. Since we already know that $u_\epsilon \rightarrow \bar{u}$ in $L^2(\mathbb{R}^\ell)$, the uniqueness of the limit almost everywhere shows that $\bar{u} \in H^{s+2}$.

By induction on $s$ we then deduce that $\bar{u} \in \cap_{k \geq 0} H^s$ and finally by Sobolev inequalities, we get that $\bar{u} \in C^\infty$. Finally by relaxing $R \rightarrow \infty$, we deduce that $u \in C^\infty$.

**Remark 1.4.** As one can see, the full argument of justification of the a priori estimate is (1) long (and in particular significantly longer than the proof of the a priori estimate itself!), (2) technically tedious. It is rarely performed in full details in research papers and often postponed when working “heuristically” or “intuitively” on a problem, however it is absolutely compulsory to perfectly understand and check these justification arguments, in order to provide fully rigorous proofs.

### 1.3. Localisation of energy estimates and elliptic regularity in open sets.

We shall now present a very simple by useful refinement of the previous argument: the *localisation* of the a priori estimate in order to obtain the regularity in any neighbourhood.

**Proposition 1.5** (Qualitative local statement of elliptic regularity). Let $\mathcal{U} \subset \mathbb{R}^\ell$ be an open set, and $u : \mathcal{U} \rightarrow \mathbb{R}$ be $C^2$ with $\Delta u = f$ in $\mathcal{U}$, where $f$ is $C^\infty$ on $\mathcal{U}$. Then $u$ is $C^\infty$ on $\mathcal{U}$.

**Remark 1.6.** This statement stresses the fact, in fact already present in the previous proof, that the elliptic regularity is a local phenomenon, in the sense that it is a consequence of the PDE in any neighbourhood, independently of what happens outside of this neighbourhood.

**Proof of Proposition 1.5.** Consider a base point $x_0 \in \mathcal{U}$ and $\eta$ so that $B(x_0, 2\eta) \subset \mathcal{U}$. Then define

\[ \tilde{u}(x) = u(x)\chi_\eta(x - x_0) \]

and

\[ \tilde{f} = \Delta \tilde{u}, \quad u_\epsilon = \tilde{u} \ast \chi_\epsilon, \quad \epsilon \in (0, \eta), \quad f_\epsilon = \tilde{f} \ast \chi_\epsilon, \quad \epsilon \in (0, \eta). \]

Observe that $\tilde{f}$ satisfies the equation

\[ \tilde{f}(x) = \Delta(u(x)\chi_\eta(x - x_0)) = \Delta u(x)\chi_\eta(x - x_0) + u(x)\Delta \chi_\eta(x - x_0) + 2\nabla u(x) \cdot \nabla \chi_\eta(x - x_0) \]

and therefore $\tilde{f} = f$ on $B(x_0, \eta/4)$ since $\chi = 1$ on $B(0, 1/4)$.

Then, upon securing that all functions are defined only within the open set $B(x_0, 2\eta) \subset \mathcal{U}$, we can perform the same argument as above and show that $\tilde{u} \in H^s(B(x_0, 2\eta))$ for any $s \in \mathbb{N}$, which implies that $\tilde{u} \in C^\infty(B(x_0, 2\eta))$. Since the point $x_0$ was arbitrary it shows that $u \in C^\infty(\mathcal{U})$. \hfill \Box

### 1.4. Ill-posedness of the Cauchy problem.

Let us consider the issue of solving the Cauchy problem for the Poisson equation, taking for instance w.l.o.g. the first variable as time. The Cauchy problem then read

\[
\begin{aligned}
\Delta u &= f, \quad x = (x_1, \ldots, x_\ell) \in \mathbb{R}^\ell \\
u(0, x_2, \ldots, x_\ell) &= u_0(x_2, \ldots, x_\ell), \\
\partial_{x_1} u(0, x_2, \ldots, x_\ell) &= u_1(x_2, \ldots, x_\ell)
\end{aligned}
\] (1.2)
where \(u_0\) and \(u_1\) are analytic functions. We know from the Cauchy-Kovalevskaya Theorem that if the data \(u_0\) and \(u_1\) are analytic near the origin, there exists a unique analytic solution \(u\) of the Cauchy problem \((1.2)\) in a neighborhood of the origin. On the other hand, if we replace the analyticity assumption on \(u_0, u_1\), with the assumption that \(u_0 \in C^k, u_1 \in C^{k-1}\), but say \(u_0 \not\in C^{k+1}\), then elliptic regularity immediately tells us that there cannot exist a \(C^2\) solution \(u\) satisfying \((1.2)\). More precisely:

**Proposition 1.7.** Assume that \(u_1 \in C^{k-1}\) and \(u_0\) is \(C^k\) but not \(C^{k+1}\) in a neighbourhood of 0, for some \(k \geq 2\), and \(f\) is a smooth \((C^\infty)\) function in this neighbourhood. Then for no neighbourhood of 0 does there exist a \(C^2\) solution \(u\) of \((1.2)\).

**Proof of Proposition 1.7.** We apply the proposition 1.5: if \(u\) is a \(C^2\) solution to \((1.2)\) in a neighborhood \(U\) of the origin, then \(u \in C^\infty(U)\), which contradicts \(u_0 \not\in C^{k+1}\). \(\square\)

As we shall see, this behaviour is very different from that of the wave equation. The above non-existence result can be thought of as a qualitative statement of ill-posedness, that is to say, saying that there does not exist a solution of a suitable Cauchy problem.

Taken alone, it could be (mis)interpreted as merely saying that one should never consider non-analytic functions. It turns out that like the statement of elliptic regularity, this qualitative statement can be related to a quantitative statement, namely that even if the solution exists, the \(C^k_{loc}\) norm of the solution can grow, and the same holds no matter what other way one tries to measure things, for instance \(H^k\), etc.

**Proposition 1.8.** Given any constant \(B > 0\) and any \(k \geq l \geq 0\), there exists an analytic solution \(u\) to \((1.2)\) with \(f = 0\) such that

\[
\|u(1, \cdot)\|_{C^l} \geq B (\|u(0, \cdot)\|_{C^k} + \|\partial_{x_1} u(0, \cdot)\|_{C^k})
\]

and similarly when \(C^k\), \(C^l\) is replaced by \(H^k(B_R), \ H^l(B_R)\) for some given ball \(B_R = B(0, R)\) with any \(R > 0\).

**Proof of Proposition 1.8.** The proof can be performed going back to the example of Hadamard we already discussed (dropping all variables \(x_3, \ldots, x_\ell\)):

\[
\partial_{x_1}^2 u + \partial_{x_2}^2 u = 0, \quad u(0, x_2) = 0, \quad \partial_{x_1} u(0, x_2) = a_\omega \sin(\omega x_2)
\]

for some parameter \(\omega > 0\), whose solution is explicitly given by

\[
u(x_1, x_2) = a_\omega \sinh(\omega x_1) \cos(\omega x_2)\]

Then if we choose \(a_\omega = \exp(-\sqrt{\omega})\) and \(\omega \to \infty\), we see that the initial data goes to zero in any \(C^k\) or \(H^k\) norm, whereas the solution goes to infinity \(\omega\) in any \(C^k\) or \(H^l\) norm. \(\square\)

**Exercise 27.** Explain how to extend this statement with an analytic right hand side \(f\) in the Poisson equation \(\Delta u = f\).

**Corollary 1.9.** As a consequence, if we define \(D\) the set of analytic \((u_0, u_1)\) for which \(u(1, \cdot)\) is defined, then for all \(k \geq l \geq 0\), the map \((u_0, u_1) \mapsto u(1, \cdot)\) is not continuous from \(D \cap C^k\) to \(C^l\) (or from \(D \cap H^k(B_R)\) to \(H^l(B_R)\), for any given \(R > 0\)).
It was Hadamard who first really understood this and formulated the notion of well-posedness, which states not only that one can solve the initial value problem in suitable family of function classes, but also, that the relevant map be continuous. Note that allowing \( k > l \) gives a little bit of room for well-posedness type statements. For the wave equation, we shall see that in Sobolev spaces we have continuity with \( k = l \), whereas in \( C^k \) spaces we must indeed take \( k > l \). We shall return to this when we study the wave and Schrödinger equations.

One should already view this as a small victory for Mathematical Analysis in illuminating a fundamental physical principle. This difference between elliptic (parabolic equations too will fit this as we shall see) and hyperbolic equations is a deep physical fact with many implications to the notion of time, the admissibility of theories, etc. On the other hand, it is completely swept under the rug when looking only at analytic solutions.

### 2. Solving the Dirichlet problem for the Poisson equation

Since as we have just seen, the Cauchy problem is ill-posed, what is the “correct” way of obtaining solutions to \( \Delta u = f \)? It turns out that the “correct” problem to study is the so-called **Dirichlet problem**, i.e. solving \( \Delta u = f \) on a bounded domain (open) \( \mathcal{U} \) with prescribed boundary values on the boundary \( \partial \mathcal{U} \). A hint that this is correct is provided by the fact that we can estimate a priori all solutions of the Dirichlet problem, as we shall see. Let us first consider the simplest case, when the prescribed boundary values are assumed to vanish. Consider a priori some \( u \) which satisfies

\[
\Delta u = f, \quad x = (x_1, \ldots, x_d) \in \mathcal{U}, \quad u = 0, \quad x \in \partial \mathcal{U}
\]

in a domain \( \mathcal{U} \subseteq \mathbb{R}^d \) such that \( \partial \mathcal{U} \) is sufficiently regular (say \( C^\infty \)). We also assume that \( u \in C^2(\mathcal{U}) \cap C^1(\overline{\mathcal{U}}) \) and \( u = 0 \) on \( \partial \mathcal{U} \).

#### 2.1. The key a priori estimate

Again we shall start by highlighting the key a priori estimate that we can obtain on this problem, and then (shortly) justify rigorously...
the proofs that can be drawn from it. As before we argue \textit{a priori}, assuming that, as we said, \( u \in C^2(\mathcal{U}) \cap C^1(\overline{\mathcal{U}}) \) and \( u = 0 \) on \( \partial \mathcal{U} \). However since we do not prescribe anything on the first derivative\(^1\) it would be a bad idea to try to perform the same \textit{a priori} estimate as we did in the elliptic regularity as it would involve boundary integrals of the gradient on \( \partial \mathcal{U} \ldots \)

We thus here multiply the equation by \( u \) and integrate to obtain

\[
\int_{\mathcal{U}} (\Delta u)u \, dx = \int_{\mathcal{U}} uf \, dx.
\]

Integrating by parts (in view of the boundary conditions\(^2\)), we get

\[
\int_{\mathcal{U}} |\nabla u|^2 \, dx = - \int_{\mathcal{U}} uf \, dx \leq \|u\|_{L^2(\mathcal{U})}\|f\|_{L^2(\mathcal{U})}.
\]

We shall now use the well-known following \textit{functional inequality}

\textbf{Theorem 2.1 (Poincaré’s inequality with Dirichlet conditions).} \textit{Let} \( \mathcal{U} \subset \mathbb{R}^d \) \textit{be an open bounded set such that} \( \partial \mathcal{U} \) \textit{is smooth. Then there exists} \( C_{\mathcal{U}} > 0 \) \textit{(only depending on} \( \mathcal{U} \) \textit{) such that the following holds. Let} \( u \in C^1(\overline{\mathcal{U}}) \) \textit{such that} \( u = 0 \) \textit{on} \( \partial \mathcal{U} \), \textit{then}

\[
\left( \int_{\mathcal{U}} u(x)^2 \, dx \right)^{1/2} \leq C_{\mathcal{U}} \left( \int_{\mathcal{U}} |\nabla u(x)|^2 \, dx \right)^{1/2}.
\]

\( ^1 \)We know from the previous section that it would be a bad idea to prescribe both \( u \) and its normal \( \text{gradient} \) on some Cauchy surface, and the Dirichlet problem consists in removing the condition on the gradient. However there is another important class of boundary conditions, the Neumann boundary conditions, where some orthogonality condition is prescribed on the gradient at \( \partial \mathcal{U} \), but in this case \textit{nothing} is prescribed on the value of \( u \).

\( ^2 \)Note in particular why \( C^1(\overline{\mathcal{U}}) \) is natural.
Proof of Theorem 2.1. Consider any point \( x = (x_1, \ldots, x_\ell) \) in \( U \), and let \( \bar{x}_1 \) be such that \( \bar{x} = (\bar{x}_1, x_2, \ldots, x_\ell) \) (see Figure 2.1).

Then we have
\[
  u(x) = \int_{\bar{x}_1}^{x_1} \partial_{x_1} u(y, x_2, \ldots, x_\ell) \, dy
\]
and we estimate
\[
  \int_{\mathcal{U}} u(x)^2 \, dx = \int_{\mathcal{U}} \left( \int_{\bar{x}_1}^{x_1} \partial_{x_1} u(y, x_2, \ldots, x_\ell) \, dy \right)^2 \, dx
  \leq D_1(\mathcal{U})^2 \int_{\mathcal{U}} |\partial_{x_1} u(x)|^2 \, dx \leq D_1(\mathcal{U})^2 \int_{\mathcal{U}} |\nabla u(x)|^2 \, dx
\]
which concludes the proof. We have denoted here by \( D_1(\mathcal{U}) \) the length of the greatest interval along the first axis included in \( \mathcal{U} \).

\( \square \)

Exercise 28. Show that this inequality remains true under the more general condition that \( \mathcal{U} \) is bounded along one of its coordinate only.

Remark 2.2. Note that there are also Poincaré’s inequalities in the whole space, provided the reference measure \( \gamma \) has some strong decay (essentially at least exponential) and regularity properties. The most famous example is the gaussian case \( \gamma(x) = e^{-|x|^2} \): 
\[
  \left( \int_{\mathbb{R}^\ell} \left| u(x) - \int_{\mathbb{R}^\ell} u(y) \gamma(y) \, dy \right|^2 \, dx \right)^{1/2} \leq C_\gamma \left( \int_{\mathbb{R}^\ell} |\nabla u(x)|^2 \gamma(x) \, dx \right)^{1/2}.
\]
The proof is more involved than the one above, see for instance the 2011 exam paper of the course on kinetic theory for intermediate steps.

Going back to our a priori estimate, we can apply the Poincaré inequality and deduce
\[
  \|u\|_{L^2(\mathcal{U})}^2 \leq C_\mathcal{U}^2 \int_{\mathcal{U}} |\nabla u|^2 \, dx \leq C_\mathcal{U}^2 \|u\|_{L^2(\mathcal{U})} \|f\|_{L^2(\mathcal{U})}
\]
which implies
\[
  \|u\|_{L^2(\mathcal{U})} \leq C_\mathcal{U}^2 \|f\|_{L^2(\mathcal{U})}.
\]

But now we can build up on this first estimate, and obtain more: by boostraping the information on the \( L^2 \) norm of \( u \) into the first a priori estimate we obtain
\[
  \int_{\mathcal{U}} |\nabla u|^2 \, dx \leq \|u\|_{L^2(\mathcal{U})} \|f\|_{L^2(\mathcal{U})} \leq C_\mathcal{U}^2 \|f\|_{L^2(\mathcal{U})}^2.
\]
Combining the two last inequalities we can write
\[
  \|u\|^2_{H^1(\mathcal{U})} \leq (C_\mathcal{U}^4 + C_\mathcal{U}^2) \|f\|^2_{L^2(\mathcal{U})}.
\]
As a consequence of this discussion we can obtain immediately

Proposition 2.3. Suppose \( u, v \in C^2(\mathcal{U}) \cap C^1(\overline{\mathcal{U}}) \) satisfy \( \Delta u = \Delta v = f \) on \( \mathcal{U} \) with \( u = v = g \) on \( \partial \mathcal{U} \). Then \( u = v \).
2. SOLVING THE DIRICHLET PROBLEM FOR THE POISSON EQUATION

Proof of Proposition 2.3. The proof is obvious and follows from the application of the previous estimate to the solution \( w = u - v \) which solves \( \Delta w = 0 \) on \( \mathcal{U} \) with \( w = 0 \) on \( \partial \mathcal{U} \). Hence

\[
\|w\|_{H^1(\mathcal{U})}^2 \leq 2C_u \|0\|_{L^2(\mathcal{U})}^2 = 0
\]

from our previous estimate, which shows by continuity that \( w = 0 \) everywhere. \( \square \)

This last proposition solves the problem of uniqueness, but leaves open that of existence and continuity according to the data, which are the object of the next subsections.

2.2. Hilbertian and Sobolevian preliminaries. Weak formulations are an important tool for the analysis of PDEs that permit the transfer of concepts of linear algebra to solve the problems. In a weak formulation, an equation is no longer required to hold absolutely (and this is not even well defined) and has instead weak solutions only with respect to certain “test vectors” or “test functions”. This is equivalent to formulating the problem to require a solution in the sense of a distribution. We introduce a formulation for weak solutions for the Poisson equation and show how to construct solutions using Riesz representation Theorem. (We will then explain how to construct solutions to more general elliptic problems with the help of the Lax-Milgram theorem.) We shall need some preliminary recalls here, on Hilbert spaces, on the space \( H^1_0(\mathcal{U}) \), and on the Riesz representation theorem.

2.2.1. Recalls on Hilbert spaces. A normed vector space \( H \) needs two (or three depending on the convention) properties to be a Hilbert space

- The norm should satisfy the parallelogram identity

\[
\forall x, y \in H, \quad \frac{\|x + y\|^2 + \|x - y\|^2}{2} = \|x\|^2 + \|y\|^2.
\]

Exercise 29. Show that this identity is equivalent to the existence of a scalar product \( \langle \cdot, \cdot \rangle \) compatible with the norm. A vector space \( H \) having the property but not the completeness is sometimes called a pre-Hilbert space.

Hint: Use the so-called polarization identity which is for real Hilbert space

\[
\langle x, y \rangle = \frac{1}{4} \left( \|x + y\|^2 - \|x - y\|^2 \right)
\]

and for complex Hilbert spaces

\[
\langle x, y \rangle = \frac{1}{4} \left( \|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2 \right).
\]

- The space should be complete when endowed with the distance given by the norm.
- (This last property may or may not be included in the standard definition)

The space should be separable, i.e. there should be a countable subset \( S \subset H \) which is dense in \( H \).

Assuming these three properties, one can speak of the (complex or real) Hilbert space, as all such spaces possess a Hilbert basis, i.e. a countable orthonormal family \( (e_n)_{n \geq 0} \) so that \( \text{Span}\{e_n, n \geq 0\} \) is dense in \( H \), and are therefore isomorphic to \( \ell^2(\mathbb{R}) \) (or \( \ell^2(\mathbb{C}) \)).

Exercise 30. Show the last claim.
Exercise 31. Show with the following example how to construct a non-separable Hilbert space: consider $H$, the subset of $L^2_{\text{loc}}(\mathbb{R})$ (square-integrable on any compact subset) with finite norm for the scalar product

$$\langle f, g \rangle := \lim_{R \to \infty} \frac{1}{R} \int_{-R}^{R} f(x)g(x) \, dx.$$ 

Then $H$ has a Hilbert space structure, but one can check that for any $a \in \mathbb{R}$, the function $S_a : x \mapsto \sin(ax)$ belongs to $H$ (with norm 1) and that

$$\forall a, a' \in \mathbb{R}, \quad \langle S_a, S_{a'} \rangle = \delta_{a=a'}.$$ 

This produces a non-countable orthonormal family.

Some history [Adapted from wikipedia, whose source is Bourbaki 1987]: Prior to the development of Hilbert spaces, other generalizations of Euclidean spaces were known to mathematicians and physicists. In particular, the idea of an abstract linear space had gained some traction towards the end of the 19th century. In the first decade of the 20th century, parallel developments led to the introduction of Hilbert spaces. The first of these was the observation, which arose during David Hilbert and Erhard Schmidt’s study of integral equations, that square-integrable real-valued functions $f$ and $g$ on an interval $[a, b]$ form a vector space with a scalar product and a corresponding notion of “orthogonality”. The second development was triggered by the Lebesgue integration theory, and in 1907, F. Riesz and E. S. Fischer independently proved that the space $L^2$ of square Lebesgue-integrable functions is a complete metric space. As a consequence of the interplay between geometry and completeness, the 19th century results of J. Fourier, F. Bessel and M.-A. Parseval on trigonometric series easily carried over to these more general spaces, resulting in a geometrical and analytical apparatus now usually known as the Riesz-Fischer theorem. Further basic results were proved in the early 20th century. For example, the Riesz representation theorem was independently established by M. Fréchet and F. Riesz in 1907. J. von Neumann coined the term abstract Hilbert space in his work on unbounded Hermitian operators. Although other mathematicians such as H. Weyl and N. Wiener had already studied particular Hilbert spaces in great detail, often from a physically motivated point of view, von Neumann gave the first complete and axiomatic treatment of them. Von Neumann later used them in his seminal work on the foundations of quantum mechanics, and in his continued work with E. Wigner. The name “Hilbert space” was soon adopted by others, for example by H. Weyl (a former student of Hilbert) in his book on quantum mechanics and the theory of groups.

2.2.2. The $H^1(\mathcal{U})$ and $H^1_0(\mathcal{U})$ spaces. We consider $\mathcal{U}$ a bounded and open set of $\mathbb{R}^d$ with smooth boundary $\partial \mathcal{U}$. Then the Sobolev space $H^1(\mathcal{U})$ in the open set $\mathcal{U}$ is defined as the closure in $L^2(\mathcal{U})$ of $C^\infty(\mathcal{U})$ endowed with the norm

$$\|u\|_{H^1(\mathcal{U})} := \left( \|u\|_{L^2(\mathcal{U})}^2 + \|\nabla u\|_{L^2(\mathcal{U})}^2 \right)^{1/2}.$$ 

We then introduce the Sobolev space $H^1_0(\mathcal{U})$ as the closure of $C^\infty_c(\mathcal{U})$ (this time with compact support included in $\mathcal{U}$) in $L^2(\mathcal{U})$ for the norm $H^1(\mathcal{U})$ again. Observe that $H^1_0(\mathcal{U}) \subset H^1(\mathcal{U})$ since the compact support condition restricts the functions considered.
Exercise 32.  

1. Show that $H^1_0(U)$ is a Hilbert space, included in the Hilbert space $H^1(U)$.

2. Show that on this subspace the homogeneous semi-norm $\dot{H}^1$ is a norm thanks to the Poincaré inequality.

3. Show also that for any $u, v \in H^1_0(U)$ and any first-order partial derivative $\partial$ we have

$$\int_U (\partial u)v \, dx = -\int_U u(\partial v) \, dx.$$

(Actually check that it is enough that only one of the two functions $u$ and $v$ is in $H^1_0(U)$, while the other one can be merely in $H^1(U)$.)

Hint: Use an approximation argument by $C^1_c(U)$ functions.

Remark 2.4. As is well known, functions in $L^2$ are only defined up to redefinition on a set of measure 0. Thus, the restriction of an $L^2$ function to a point or any hypersurface with non-zero codimension is meaningless. When we now consider solutions in spaces of generalised functions, this is an important thing not to be forgotten: the space should have enough regularity for the boundary (and/or initial) conditions to make sense.

On the one hand, we have seen with the Sobolev inequalities that the $H^s$ regularity implies the $C^{s-\ell/2-0}$ regularity. Obviously our boundary conditions make sense for continuous functions, therefore the regularity $H^s$ with $s > \ell/2$ would clearly guarantee that these boundary conditions make sense. However in dimension $\ell \geq 2$ this seems to prevent using $H^1$...

On the other hand, another family of inequality, the so-called trace inequalities, allows one to restrict $H^s$ functions to $H^{s-d/2}$ functions on a codimension-$d$ hypersurface. We can thus characterize $H^1_0(U)$ as the subset of $H^1(U)$ consisting of functions such that, extending to an $H^1(\mathbb{R}^\ell)$, which is always possible, the trace on $\partial U$ vanishes as an $L^2$ function. More generally trace inequalities allow one to restrict $H^s$ functions to $H^{s-d/2}$ functions on a codimension-$d$ hypersurface. When $d = \ell$ we recover the numerology of Sobolev inequalities: being continuous is similar to the fact that restriction to points (full codimension subset) is well-defined.

2.2.3. The Riesz (Hilbert space) representation Theorem. Warning: there are two main theorems usually called “Riesz representation Theorem” that should be not confused, the other one being that of representation of positive linear functionals on $C_c(X)$ (X locally compact Hausdorff – i.e. separated– space) by regular Borel measures.

Let us state and prove two important theorems generalising the intuition of Euclidean spaces:

Proposition 2.5 (Projection on a closed convex set). Consider $(H, \langle \cdot, \cdot \rangle)$ a Hilbert space, $C \subset H$ a closed non-empty convex subset of $H$, and $x \in H$. Then there exists a unique $P_C(x) \in C$ which realizes the following minimization problem

$$\| P_C(x) - x \| = \min_{y \in C} \| y - x \|.$$

Moreover the application $P_C : H \to C$ is 1-Lipschitz.
Figure 2.3. Geometric summary of the proof.

Proof of Proposition 2.5. Set $x = 0$ w.l.o.g. If $0 \in C$ then we are done. If not we denote the minimum value by $I$ consider a minimizing sequence $y_n \in C$, $n \in \mathbb{N}$, so that $\|y_n\| \to I \geq 0$.

The sequence is Cauchy:

$$\|y_m - y_n\|^2 = 2\|y_m\|^2 + 2\|y_n\|^2 - \|y_m + y_n\|^2 = 2\|y_m\|^2 + 2\|y_n\|^2 - 4 \left\|\frac{y_m + y_n}{2}\right\|^2$$

and since $(y_m + y_n)/2 \in C$ by convexity, we have

$$\left\|\frac{y_m + y_n}{2}\right\|^2 \geq I^2$$

and

$$0 \leq \|y_m - y_n\|^2 = 2\|y_m\|^2 + 2\|y_n\|^2 - 4 \left\|\frac{y_m + y_n}{2}\right\|^2$$

$$\leq 2\|y_m\|^2 + 2\|y_n\|^2 - 4I^2 \xrightarrow{m,n \to \infty} 0.$$  

This show the existence of the limit $y_n \to y_\infty$ by using the completeness of the space. By closedness of $C$ then we have $y_\infty \in C$, which implies the existence of a minimizer and also that $I > 0$ since $0 \notin C$.

Let us show that the minimizer is unique. The intuitive reason is that a scalar product structure implies the strict convexity of the unit ball of the space, and therefore there are no non-trivial segment at constant distance from the origin. Let us argue by contradiction and suppose that there are two distinct minimizers $y_1, y_\infty \in C$.

First proof: consider the mid-point $y_3 = (y_1 + y_\infty)/2$ and write (by expanding)

$$I^2 \leq \|y_3\|^2 = \frac{I^2}{2} + \frac{1}{2} \langle y_\infty, y_\infty \rangle \implies I^2 \leq \langle y_\infty, y_\infty \rangle.$$
This implies since $I^2 = \|y_1\|^2 = \|y_2\|^2$ that

$$
\langle y_1^1, y_2^1 - y_1^1 \rangle \geq 0, \quad \langle y_2^1, y_2^2 - y_2^1 \rangle \geq 0, \quad \implies \quad \|y_1^1 - y_2^1\|^2 = 0.
$$

Second proof: We have for any $t \in [0, 1]$ that $ty_1^1 + (1-t)y_2^2 = C$ and therefore

$$
I^2 \leq \|ty_1^1 + (1-t)y_2^2\|^2 = \|y_2^2\|^2 = 2t \langle y_2^2, y_2^1 - y_2^2 \rangle + t^2 \|y_1^1 - y_2^1\|^2.
$$

Using that $\|y_2^2\|^2 = I^2$ we deduce that

$$
\forall t \in [0, 1], \quad 2t \langle y_2^2, y_1^1 - y_2^1 \rangle + t^2 \|y_1^1 - y_2^1\|^2 \geq 0.
$$

Taking $t \to 0^+$ this imposes $\langle y_2^2, y_1^1 - y_2^1 \rangle \geq 0$. But a symmetric argument would yield $\langle y_1^1, y_2^2 - y_1^1 \rangle \geq 0$. Summing both inequalities gives

$$
0 \leq \langle y_2^2, y_1^1 - y_2^1 \rangle + \langle y_1^1, y_2^2 - y_1^1 \rangle = \langle y_2^2 - y_1^1, y_2^2 - y_1^1 \rangle = -\|y_1^1 - y_2^1\|^2
$$

which shows uniqueness.

Complement: arguing similarly we can now prove the Lipschitz regularity. Consider $t \in (0, 1)$, two points $x_1, x_2 \in H$ and their projections $P_C(x_1), P_C(x_2) \in C$, and we write down the two inequalities

$$
\|P_C(x_2) - x_2\|^2 \leq \|tP_C(x_1) + (1-t)P_C(x_2) - x_2\|^2
$$

$$
= \|P_C(x_2) - x_2\|^2 + t^2 \|P_C(x_2) - P_C(x_1)\|^2 + 2t \langle P_C(x_2) - x_2, P_C(x_1) - P_C(x_2) \rangle
$$

$$
\|P_C(x_1) - x_1\|^2 \leq \|tP_C(x_2) + (1-t)P_C(x_1) - x_1\|^2
$$

$$
= \|P_C(x_1) - x_1\|^2 + t^2 \|P_C(x_2) - P_C(x_1)\|^2 + 2t \langle P_C(x_1) - x_1, P_C(x_2) - P_C(x_1) \rangle
$$

which use the fact that $tP_C(x_1) + (1-t)P_C(x_2) \in C$, $tP_C(x_2) + (1-t)P_C(x_1) \in C$. We deduce

$$
t^2 \|P_C(x_2) - P_C(x_1)\|^2 + 2t \langle P_C(x_2) - x_2, P_C(x_1) - P_C(x_2) \rangle \geq 0
$$

$$
t^2 \|P_C(x_2) - P_C(x_1)\|^2 + 2t \langle P_C(x_1) - x_1, P_C(x_2) - P_C(x_1) \rangle \geq 0
$$

and by summing the two inequalities

$$
2t^2 \|P_C(x_2) - P_C(x_1)\|^2 \geq 2t \langle P_C(x_2) - x_2 + x_1 - P_C(x_1), P_C(x_2) - P_C(x_1) \rangle
$$

which implies

$$
\|P_C(x_2) - P_C(x_1)\|^2 \leq \frac{1}{1-t} \|x_2 - x_1\|^2
$$

and we finally let $t \to 0$. \hfill \Box

Let us now prove the Riesz representation theorem (in the case of a real Hilbert space, the complex case being similar).

**Theorem 2.6 (Riesz representation Theorem).** Consider a Hilbert space $(H, \langle \cdot, \cdot \rangle)$ and a continuous linear form $g : H \to \mathbb{R}$. Then there is a unique $y \in H$ such that

$$
\forall x \in H, \quad g(x) = \langle x, y \rangle.
$$

Moreover the map $g \in H' \to y \in H$ is linear and continuous (in fact isometric)

$$
\|g\|_{\mathcal{L}(H, \mathbb{R})} = \|y\|_H
$$
where the LHS in the last equation denotes the canonical operator norm
\[ \|g\|_{\mathcal{L}(H, \mathbb{R})} := \sup_{\|x\|_H = 1} |g(x)|. \]

**Proof of Theorem 3.2.** If \( g = 0 \) then \( y = 0 \) is the only solution and we are done. If \( g \) is non-zero, then the vector structure implies that \( g(H) = \mathbb{R} \), and there is \( y \in H \) so that \( g(y) = 1 \). Consider the set
\[ C_g = \{ y \in H : g(y) = 1 \}. \]
It is convex by linearity of \( g \), and closed by continuity of \( g \). We then define \( y^* = P_C(0) \) from the previous proposition, which satisfies
\[ \|y^*\| = \min_{y \in C_g} \|y\|. \]
Since \( y^* \in C_g \) we have \( y^* \neq 0 \). Let us prove that \( y^* \perp \text{Null}(g) \). Consider \( x \in \text{Null}(g) \) and \( t \in \mathbb{R} \), then \( g(y^* + tx) = 1 \) hence \( y^* + tx \in C_g \). Therefore
\[ \forall t \in \mathbb{R}, \quad \|y^* + tx\|^2 \geq \|y^*\|^2 \]
which implies
\[ \forall t \in \mathbb{R}, \quad 2t \langle y^*, x \rangle + t^2 \|x\|^2 \geq 0 \]
and therefore \( \langle y^*, x \rangle = 0 \) by taking \( t \to 0^- \) or \( t \to 0^+ \).
Finally for any \( x \in H \) we have the decomposition
\[ \forall x \in H, \quad (x - g(x)y^*) \in \text{Null}(g) \subset (y^*)^\perp \]
and therefore \( g(x) = \langle x, y^* \rangle / \|y^*\|^2 \), which conclude the proof with \( y^{**} := y^* / \|y^*\|^2 \). The identity on the norm is trivial. \( \square \)

### 2.3. Existence of weak (generalised) solutions.

Let us first define the notion of weak solutions. Assume that \( \Delta u = f \) with \( u \in C^2(\mathcal{U}) \cap C^1(\overline{\mathcal{U}}) \) and \( u = 0 \) on \( \partial \mathcal{U} \) then for any \( v \in C^2(\mathcal{U}) \cap C^1(\overline{\mathcal{U}}) \) with \( v = 0 \) on \( \partial \mathcal{U} \) we have
\[ \langle \langle u, v \rangle \rangle := \int_{\mathcal{U}} \nabla u \cdot \nabla v \, dx = -\int_{\mathcal{U}} (\Delta u)v \, dx = -\int_{\mathcal{U}} fv \, dx \]
where we have denoted by \( \langle \langle \cdot, \cdot \rangle \rangle \) the scalar product associated with the homogeneous norm \( H^1(\mathcal{U}) \), keeping the notation \( \langle \cdot, \cdot \rangle \) for the usual \( L^2(\mathcal{U}) \) scalar product. Observe crucially that all the objects in this statement still make sense as soon as \( u, v \in H^1_0(\mathcal{U}) \).

**Definition 2.7.** We call generalised (or weak) solution a function \( u \in H^1_0(\mathcal{U}) \) such that
\[ \forall v \in H^1_0(\mathcal{U}), \quad \langle \langle u, v \rangle \rangle = -\langle f, v \rangle. \]
Note that we have an equivalent characterization that \( u \in H^1_0(\mathcal{U}) \) and
\[ \langle u, \Delta v \rangle = \langle f, v \rangle \]
for all \( v \in C^\infty_c(\mathcal{U}) \) smooth of compact support in \( \mathcal{U} \). This latter equality is the statement that \( u \) is a distributional solution of \( \Delta u = f \).

**Exercise 3.3.** Prove the equivalence in the definition.

**Remark 2.8.** Note the important idea behind this reformulation: the boundary conditions have been enforced-encoded in the functional space itself.
We can now state and prove the existence theorem:

**Theorem 2.9.** Let $\mathcal{U} \subset \mathbb{R}^\ell$ and open set, and $f \in L^2(\mathcal{U})$. Then there exists a unique $u \in H^1_0(\mathcal{U})$ such that $u$ is a weak solution of $\Delta u = f$, in the sense defined above.

**Proof of Theorem 2.9.** The proof is a straightforward application of the Riesz representation Theorem: we consider the following linear form on $H^1_0(\mathcal{U})$:

$$\forall v \in H^1_0(\mathcal{U}), \quad g(v) := -\langle f, v \rangle$$

which is continuous by Cauchy-Schwarz and Poincaré’s inequalities

$$\forall v \in H^1_0(\mathcal{U}), \quad |g(v)| = |\langle f, v \rangle| \leq \|f\|_{L^2(\mathcal{U})}\|v\|_{L^2(\mathcal{U})} \lesssim \|f\|_{L^2(\mathcal{U})}\|v\|_{H^1(\mathcal{U})}.$$ 

Then the Riesz representation theorem applied in the Hilbert space $H^1_0(\mathcal{U})$ endowed with the norm $H^1(\mathcal{U})$, shows that there is a unique $u \in H^1_0(\mathcal{U})$ so that

$$\forall v \in H^1_0(\mathbb{R}^\ell), \quad g(v) = \langle \langle u, v \rangle \rangle$$

which concludes the proof. $\square$

**Remark 2.10.** Observe moreover that in the previous statement the solution map $S : f \mapsto u$ is continuous from $L^2(\mathcal{U})$ to $H^1_0(\mathcal{U})$ since

$$\|g\|_{H^1_0(\mathcal{U})} = \sup_{\|v\|_{H^1(\mathcal{U})} = 1} |g(v)| \lesssim \|f\|_{L^2(\mathcal{U})},$$

and $g \mapsto u$ is an isometry in the representation theorem.

**Exercise 34.** In fact prove that $S$ is even continuous from $H^{-1}(\mathcal{U})$, the dual of $H^1_0(\mathcal{U})$ for the $L^2(\mathcal{U})$ scalar product $\langle \cdot, \cdot \rangle$:

$$\|f\|_{H^{-1}(\mathcal{U})} := \sup_{\|v\|_{H^1_0(\mathcal{U})} = 1} \langle f, v \rangle_{L^2(\mathcal{U})},$$

to $H^1_0(\mathcal{U})$.

### 3. General elliptic equations and Lax-Milgram Theorem

Let us consider now the more general boundary-value problem

(3.1)

$$\begin{cases}
\mathcal{P}u = f & \text{in } \mathcal{U} \\
u = g & \text{on } \partial \mathcal{U}.
\end{cases}$$

where $\mathcal{U} \subset \mathbb{R}^\ell$ is open, bounded and smooth, $\mathcal{P}$ denotes a second-order linear elliptic partial operator, and $g$ is smooth enough to be the trace of an $H^1(\mathcal{U})$ function.

#### 3.1. Divergence and non-divergence forms.

The operator $\mathcal{P}$ can be given in two different forms. The first one is

(3.2)

$$\mathcal{P}u := -\sum_{i,j=1}^\ell \partial_j (a_{ij}(x)\partial_i u) + \sum_{i=1}^\ell b_i(x)\partial_i u + c(x)u$$

and is called the divergence form, as the leading-order part of the operator writes $\nabla \cdot U$ for some $U(x) \in \mathbb{R}^\ell$. 

The second formulation is
\[ \mathcal{P}u = - \sum_{i,j=1}^{\ell} a_{ij}(x) \partial_{ij}^2 u + \sum_{i=1}^{\ell} b_i(x) \partial_i u + c(x) u \]
and is called the non-divergence form for obvious reasons.

**Remark 3.1.** Note the minus sign in front of the second-order term, which is convenient for reminding that the ellipticity is related to coercivity (positivity) of the Fourier symbol.

**Exercise 35.** Assume that the coefficients \(a_{ij}, b_i, c\) are \(C^1\) and show that an operator given in divergence form can be rewritten in non-divergence form, and vice-versa.

Both forms are useful. The divergence form is usually better adapted to energy estimates which uses integration by parts, Hilbertian methods, and more generally integral arguments. The non-divergence form is better adapted to pointwise arguments and maximum principles.

### 3.2. Ellipticity and reduction of the problem.

We consider in the rest of this section the operator \(\mathcal{P}\) as above in divergence form and we assume (1) w.l.o.g. the symmetry condition
\[ \forall i,j \in \{1, \ldots, \ell\}, \quad a_{ij} = a_{ji} \]
and (2) the uniform ellipticity condition
\[ \forall x \in U, \quad A(x) \geq \alpha > 0, \quad A(x) = (a_{ij})_{1 \leq i,j \leq \ell}. \]
We also assume that the coefficients satisfy \(a_{ij}, b_i, c \in L^{\infty}(U)\). Let us denote \(B(x) = (b_1(x), \ldots, b_\ell(x))^\perp\). We consider the equation \(\mathcal{P}u = f\) in \(U\) with some \(f \in L^2(U)\), together with the boundary condition \(u = g\) on \(\partial U\) with some \(g \in C^0(\partial U)\).

Observe that we can reduce to the case of zero boundary conditions: consider \(w \in H^1(\mathbb{R}^\ell)\) so that \(w\) restricts to \(g\) on \(\partial U\). Then \(\tilde{u} = u - w\) solves a similar elliptic PDE for the same matrix \(A(x)\) and some other RHS \(\tilde{f}\). We therefore assume that \(g = 0\) in the sequel.

### 3.3. The key a priori estimate.

Assuming that \(u \in C^2(U) \cap C^1(\bar{U})\), we multiply the equation by \(u\) and integrate it over \(U\):
\[ \int_U (\nabla u)^T A(x) (\nabla u) \, dx + \int_U u(x) B(x) \cdot \nabla u \, dx + \int_U cu^2 \, dx = \int_U f u \, dx \]
and we deduce that
\[ \alpha \int_U |\nabla u|^2 \, dx \leq \|B\|_{\infty} \|u\|_{L^2(\bar{U})} \|\nabla u\|_{L^2(\bar{U})} + \|c\|_{\infty} \|u\|_{L^2(\bar{U})}^2 + \|u\|_{L^2(\bar{U})} \|f\|_{L^2(\bar{U})} \]
\[ \leq \frac{\alpha}{2} \|\nabla u\|_{L^2(\bar{U})}^2 + \|f\|_{L^2(\bar{U})}^2 + C \|u\|_{L^2(\bar{U})}^2 \]
for some constant \(C > 0\) (possibly large). We deduce that
\[ \|\nabla u\|_{L^2(\bar{U})}^2 \leq \frac{\alpha}{2} \|f\|_{L^2(\bar{U})}^2 + \|u\|_{L^2(\bar{U})}^2, \]

(Recall that the sign \( \preceq \) means that the inequality holds with a constant unimportant for the argument and not depending on the quantities in the inequality). Hence we see again a gain of regularity, however now the uniqueness does not follow immediately from the a priori estimate. We shall first deal with the case where the coercivity is recovered in the sense that one can establish the a priori estimate \( \|\nabla u\|^2_{L^2(U)} \preceq \|f\|^2_{L^2(U)}. \) This first case will be solved with the help of Lax-Milgram Theorem. We will then consider the general case thanks to the Fredholm theory.

### 3.4. The Lax-Milgram Theorem

Consider a linear form \( \mathcal{L} \) and a bilinear form (not necessarily symmetric!) \( \mathfrak{B} \) on \( H_0^1(U) \). Then we consider the problem of finding \( u \in H_0^1(U) \) so that

\[
\forall v \in H_0^1(U), \quad \mathfrak{B}(u, v) = \mathcal{L}(v). \tag{3.4}
\]

**Theorem 3.2 (Lax-Milgram).** Assume in (3.4) that \( \mathcal{L} \) and \( \mathfrak{B} \) are bounded as linear (resp. bilinear) forms on \( H_0^1(U) \). Assume moreover that \( \mathfrak{B} \) is coercive: there is \( \lambda > 0 \) so that

\[
\forall v \in H_0^1(U), \quad B(v, v) \geq \lambda \|v\|^2_{H^1(U)}.
\]

Then there exists a unique solution \( u \in H_0^1(U) \) to (3.4). The solution map \( \mathfrak{S} : \mathcal{L} \mapsto u \) is moreover continuous from \( H_0^{-1}(U) \) to \( H_0^1(U) \).

**Proof of Theorem 3.2.** For any given \( u \in H_0^1(U) \), the linear form \( v \in H_0^1(U) \mapsto \mathfrak{B}(u, v) \in \mathbb{R} \) is continuous, and therefore the Riesz representation theorem shows that there is a (unique) \( u^* \) so that

\[
\forall v \in H_0^1(U), \quad \mathfrak{B}(u, v) = \langle \langle u^*, v \rangle \rangle.
\]

Moreover since from the theorem the map \( u \mapsto u^* \) is linear and continuous we can represent it as \( u^* = \mathfrak{T}u \) for some bounded operator \( \mathfrak{T} \) on \( H_0^1(U) \). We also represent \( \mathcal{L} \) by \( l \in H_0^1(U) \) by the same theorem

\[
\forall v \in H_0^1(U), \quad \mathcal{L}(v) = \langle \langle l, v \rangle \rangle
\]

and therefore the equation on \( u \) rewrites as

\[
\forall v \in H_0^1(U), \quad \langle \langle \mathfrak{T}u, v \rangle \rangle = \langle \langle l, v \rangle \rangle
\]

which is equivalent to \( \mathfrak{T}u = l \in H_0^1(U) \). It remains to prove that \( \mathfrak{T} \) is invertible in \( H_0^1(U) \) to conclude the proof (defining finally \( u = \mathfrak{T}^{-1}l \)).

The operator \( \mathfrak{T} \) satisfies the following

\[
\forall v \in H_0^1(U), \quad \|v\|^2_{H^1(U)} \leq \lambda \mathfrak{B}(v, v) \leq \lambda \|\mathfrak{T}v\|_{H^1(U)} \|v\|_{H^1(U)}
\]

which implies \( \|\mathfrak{T}v\|_{H^1(U)} \geq \lambda^{-1} \|v\|_{H^1(U)} \) for any non-zero \( v \in H_0^1(U) \). This proves the injectivity. This also proves that the range of \( \mathfrak{T} \) is closed: if \( \mathfrak{T}v_n \) is converging it is Cauchy and therefore \( v_n \) is Cauchy and converging to \( v_\infty \) since the space is complete, and by continuity of \( \mathfrak{T} \), \( \mathfrak{T}v_n \rightarrow \mathfrak{T}v_\infty \). Finally if the range of \( \mathfrak{T} \) is not \( H_0^1(U) \), because of its closedness there is a non-zero \( w \perp \text{Range}(\mathfrak{T}) \) which contradicts \( \langle \langle \mathfrak{T}w, w \rangle \rangle = \mathfrak{B}(w, w) > 0. \)
Remark 3.3. Note that if $\mathfrak{B}$ is symmetric: $\mathfrak{B}(u, v) = \mathfrak{B}(v, u)$, then a simpler proof can be devised: defining a scalar product out of $\mathfrak{B}$ one observes that it endows $H_0^1(\mathcal{U})$ with a norm equivalent to $H^1(\mathcal{U})$, for which Riesz representation theorem can be immediately applied. Lax-Milgram theorem is really meaningful only in the non-symmetric case.

3.5. Existence of weak solutions assuming global coercivity. We go back to the problem and assume the operator $\mathfrak{P}$ is so that the following stronger a priori estimate yields

$$
\|u\|_{H^1(\mathcal{U})} \lesssim \|u\|_{L^2(\mathcal{U})}.
$$

Exercise 36. Check that is always possible to satisfy this condition by taking the coefficients $b_i$ and $c$ small enough in terms of the constant $\lambda$ so that $A \geq \lambda$.

Then we want to apply Lax-Milgram theorem with

$$
\mathfrak{B}(u, v) := \int_{\mathcal{U}} (\nabla u)^T A(x)(\nabla v) \, dx + \int_{\mathcal{U}} vB(x) \cdot \nabla u \, dx + \int_{\mathcal{U}} cuv \, dx,
$$

$$
\mathfrak{L} v := \langle f, v \rangle.
$$

The definition of $\mathfrak{B}$ corresponds to $\langle \mathfrak{P}u, v \rangle$ after integration by parts. The fact that $\mathfrak{B}$ is bounded in $H^1(\mathcal{U})$ is clear from $a_{ij} \in L^\infty(\mathcal{U})$. The fact that $\mathfrak{L}$ is bounded is clear from $f \in L^2(\mathcal{U})$ and Poincaré’s inequality (note here that in the definition of $\mathfrak{L}$ we use the standard $L^2(\mathcal{U})$ scalar product).

The coercivity assumption means in a more precise form that

$$
\forall v \in H_0^1(\mathcal{U}), \quad \mathfrak{B}(v, v) \geq \lambda \|v\|_{H^1(\mathcal{U})}^2
$$

for some constant $\lambda > 0$. We can then apply the Lax-Milgram theorem and conclude the proof.

3.6. Existence of weak solutions without the global coercivity. We now consider the same framework but only assume the (strict uniform) ellipticity

$$
A(x) \geq \lambda_c > 0
$$

on the second-order part of the operator $\mathfrak{P}$, but no more global coercivity on $\mathfrak{P}$. Let us denote

$$
\mathfrak{P}_e = \sum_{i,j=1}^\ell a_{ij}(x) \partial^2_{ij}
$$

the second-order part of the operator $P$.

3.6.1. A first attempt. The previous study show that the equation

$$
\begin{cases}
\mathfrak{P}_e u = f & \text{in } \mathcal{U} \\
u = 0 & \text{on } \partial \mathcal{U}.
\end{cases}
$$

admits a unique solution $u \in H^1_0(\mathcal{U})$, which depends continuously on the linear form $v \mapsto (f, v)$ on $H^1_0(\mathcal{U})$. The norm of this linear form is the $H^{-1}_0(\mathcal{U})$ norm of $f$. We have hence inverted the operator $\mathfrak{P}_e$ (with the Dirichlet conditions) and shown that

$$
\|\mathfrak{P}_e^{-1} f\|_{H^1_0(\mathcal{U})} \lesssim \|f\|_{H^{-1}_0(\mathcal{U})}.
$$
However the complete operator $\mathcal{P}$ does not have the necessary coercivity for inverting it in a similar manner. A first idea could be to factorize $\mathcal{P}$ by $\mathcal{P}_c$:

$$\mathcal{P} = \mathcal{P}_c (\text{Id} + \mathcal{P}_c^{-1} \mathcal{B} \cdot \nabla + \mathcal{P}_c^{-1} c)$$

and use compactness properties of $\mathcal{R} := \mathcal{P}_c^{-1} \mathcal{B} \cdot \nabla + \mathcal{P}_c^{-1} c$. This would require more regularity on the coefficients and we shall follow a simpler and more optimal approach.

3.6.2. A simpler and more optimal approach. The operator $\mathcal{P}$ can be made coercive by a zero-order simple modification, and we can then factorize by the modified operator.

**Proposition 3.4.** There is $\lambda_0 > 0$ so that $\mathcal{P} + \lambda_0 \text{Id}$ is coercive in the sense

$$\forall v \in H^1_0(\mathcal{U}), \quad \langle (P + \lambda_0 \text{Id}) v, v \rangle \geq \bar{\lambda} \|v\|_{H^1(\mathcal{U})}^2$$

for some $\bar{\lambda} > 0$.

**Proof of Proposition 3.4.** We perform the same a priori energy estimate

$$\langle (P + \lambda_0 \text{Id}) v, v \rangle \geq \lambda_e \|v\|_{H^1(\mathcal{U})}^2 - \langle B \cdot \nabla v, v \rangle - \langle cv, v \rangle + \lambda_0 \|v\|_{L^2(\mathcal{U})}^2$$

$$\geq \lambda_e \|v\|_{H^1(\mathcal{U})}^2 - 2 \|B\|_{L^\infty} \|v\|_{H^1(\mathcal{U})} \|v\|_{L^2(\mathcal{U})} - \frac{\lambda_e}{2} \|v\|_{L^2(\mathcal{U})}^2 + \lambda_0 \|v\|_{L^2(\mathcal{U})}^2$$

$$\geq \frac{\lambda_e}{2} \|v\|_{H^1(\mathcal{U})}^2 + \left(\lambda_0 - \frac{\|B\|_{L^\infty}}{2\lambda_e} - \|c\|_{L^\infty} \right) \|v\|_{L^2(\mathcal{U})}^2$$

for some $\bar{\lambda} > 0$, when choosing $\lambda_0$ large enough, which concludes the proof. □

Therefore from the previous subsection we can invert $(\mathcal{P} + \lambda_0 \text{Id})$ with the Dirichlet conditions. We now write the following factorisation

$$\mathcal{P} = (\mathcal{P} + \lambda_0 \text{Id}) - \lambda_0 \text{Id} = (\mathcal{P} + \lambda_0) \left[ \text{Id} - \lambda_0 (\mathcal{P} + \lambda_0 \text{Id})^{-1} \right] = (\mathcal{P} + \lambda_0) \left[ \text{Id} - \mathcal{R} \right]$$

where we denote $\mathcal{R} := -\lambda_0 (\mathcal{P} + \lambda_0 \text{Id})^{-1}$.

We shall show later that the operator $\mathcal{R}$ maps $L^2(\mathcal{U})$ to $H^1_0(\mathcal{U})$ and is a compact operator in $L^2(\mathcal{U})$, i.e. it maps the closed unit ball inside a compact set. Let us assume this for now and let us prove a criterion for the existence and uniqueness of solutions.

Observe that the equation is now equivalent to finding $u \in L^2(\mathcal{U})$ so that

$$u + \mathcal{R}u = F, \quad F := (\mathcal{P} + \lambda_0 \text{Id})^{-1} f \in H^1_0(\mathcal{U}).$$

Indeed any such $u \in L^2(\mathcal{U})$ will in fact be in $H^1_0(\mathcal{U})$ since both $\mathcal{R}u$ and $F$ belong to $H^1_0(\mathcal{U})$. Let us analyse the operators of the form $\text{Id} + \mathcal{R}$ with $\mathcal{R}$ compact.

**Proposition 3.5.** Given a compact operator $\mathcal{R}$ on a Hilbert space $H$, then $\text{Id} + \mathcal{R}$ has closed range, finite dimensional kernel and finite dimensional cokernel (the cokernel is the orthogonal complement of the range).

**Proof.** First, the null space of $\text{Id} + \mathcal{R}$ has a compact unit ball by the compactness of $\mathcal{R}$: any sequence $g_n$ with $\|g_n\| \leq 1$ and $g_n + \mathcal{R}g_n = 0$ satisfies $\mathcal{R}g_n$ has a
converging subsequence by the compactness of $\mathcal{R}$, which implies that $g_n$ has a converging subsequence. This implies by Riesz theorem that $\text{Ker}(\text{Id} + \mathcal{R})$ is finite dimensional.

Second, let us prove that $(\text{Id} + \mathcal{R})$ is coercive on the orthogonal of its kernel: there is $\lambda_{\mathcal{R}} > 0$ so that $\|g + \mathcal{R}g\| \geq \lambda_{\mathcal{R}}\|g\|$ for $g \in \text{Ker}(\text{Id} + \mathcal{R})^\perp$. It is proved by contradiction: consider $g_n \in \text{Ker}(\text{Id} + \mathcal{R})^\perp$ a sequence so that $\|g_n\| = 1$ and $\|g_n + \mathcal{R}(g_n)\| \to 0$ as $n$ goes to infinity. Then using on the one hand the weak compactness of the unit ball and on the other hand the compactness of the operator $\mathcal{R}$, there is a subsequence $g_{\varphi(n)}$ so that $g_{\varphi(n)} \to g$ (weak convergence) and $\mathcal{R}g_{\varphi(n)}$ is strongly converging. Since the weak convergence of $g_{\varphi(n)}$ implies $\mathcal{R}g_{\varphi(n)} \to \mathcal{R}g$ and the weak and strong limits are the same, we deduce that $\mathcal{R}g_{\varphi(n)} \to \mathcal{R}g = -g$ (strong convergence) and then finally $g_{\varphi(n)} \to g \in \text{Ker}(\text{Id} + \mathcal{R})$ (strong convergence). But by weak limit we also have $g \in \text{Ker}(\text{Id} + \mathcal{R})^\perp$, which implies that $g = 0$. Moreover the strong convergence implies $\|g\| = 1$, which yields the desired contradiction.

Third let us prove that the range of $\text{Id} + \mathcal{R}$ is closed. Consider a sequence $g_n + \mathcal{R}g_n \to h$. We can decompose $g_n = g^1_n + g^2_n \in \text{Ker}(\text{Id} + \mathcal{R})^\perp + \text{Ker}(\text{Id} + \mathcal{R})$. From the previous point, we have that $g^2_n$ is a bounded sequence, with $g^1_n + \mathcal{R}g^1_n \to h$. There is a subsequence $g_{\varphi(n)}$ so that $\mathcal{R}g_{\varphi(n)}$ is converging (in the strong topology), which implies that $g_{\varphi(n)}$ is strongly converging as well, say to some $g$. Finally by continuity of $\mathcal{R}$ we deduce $g + \mathcal{R}g = h \in \text{Range}(\text{Id} + \mathcal{R})$.

Fourth let us prove that the adjoint of a compact operator is also compact. Let us recall that the adjoint operator $\mathcal{R}^*$ of $\mathcal{R}$ is defined by the formula

$$\forall f, g \in H, \quad \langle \mathcal{R}f, g \rangle = \langle f, \mathcal{R}^*g \rangle$$

and Riesz’ representation theorem. Consider a bounded sequence $g_n \in H$. Then $\mathcal{R}^*g_n$ is bounded and there is a subsequence $g_{\varphi(n)}$ so that $g_{\varphi(n)} \to g \in H$ and $\mathcal{R}^*g_{\varphi(n)} \to \mathcal{R}^*g \in H$. Then

$$\|\mathcal{R}^*g_{\varphi(n)} - h\|^2 = \langle \mathcal{R}^2(g_{\varphi(n)} - g), g_{\varphi(n)} - g \rangle.$$ 

Then by the compactness of $\mathcal{R}$, we have $\mathcal{R}^2(g_{\varphi(n)} - g) \to 0$ (strong convergence), which implies that $\|\mathcal{R}^*g_{\varphi(n)} - h\| \to 0$ and concludes the proof of the fact that $\mathcal{R}^*$ is compact.

Then applying the points 1-2-3 above to $(\text{Id} + \mathcal{R})^* = \text{Id} + \mathcal{R}^*$, we deduce that the cokernel is also finite dimensional, since we recall that

$$\text{Range}(\text{Id} + \mathcal{R})^\perp = \text{Ker}(\text{Id} + \mathcal{R}^*), \quad \text{Range}(\text{Id} + \mathcal{R})^\perp = \text{Ker}(\text{Id} + \mathcal{R}).$$

This concludes the proof. Observe also that $H_0^1(U)$ is the direct sum of $\text{Ker}(\text{Id} + \mathcal{R}^*)$ and $\text{Range}(\text{Id} + \mathcal{R})$ (the latter being closed).

Let us now study the dimensions of $\text{Ker}(\text{Id} + \mathcal{R})$ and $\text{Coker}(\text{Id} + \mathcal{R})$ and prove that they are the same.

**Proposition 3.6.** Given a compact operator $\mathcal{R}$ on a Hilbert space $H$, then one has

$$\dim \text{Ker}(\text{Id} + \mathcal{R}) = \dim \text{Coker}(\text{Id} + \mathcal{R}) < +\infty.$$ 

**Proof.** The proof relies on a continuation argument and the following more general claim: if a bounded operator $T$ has closed range, finite dimensional kernel and cokernel, and satisfies $\dim(\text{Ker}(T)) = \dim(\text{Coker}(T))$, then all these properties are preserved under small enough perturbation (for the norm of bounded operators).
The proof of this claim is a reduction to the finite dimensional case. We decompose the space as \( C \oplus \text{Ker}(\mathcal{T}) \) and \( \text{Range}(\mathcal{T}) \oplus D \) with \( \dim(\text{Ker}(\mathcal{T})) = \dim(D) \). The operator \( \mathcal{T} \) writes

\[
\mathcal{T} = \begin{pmatrix}
\tilde{\mathcal{T}} & 0 \\
0 & 0
\end{pmatrix}
\]

in this decomposition, and we consider a small perturbation \( P \) which writes

\[
P = \begin{pmatrix}
P_{11} & P_{12} \\
P_{21} & P_{22}
\end{pmatrix}
\]

in this decomposition. Then we define the two invertible operators

\[
G = \begin{pmatrix}
\text{Id} & (\tilde{\mathcal{T}} + P_{11})^{-1}P_{12} \\
P_{21} & \text{Id}
\end{pmatrix},
\]

\[
H = \begin{pmatrix}
\text{Id} & 0 \\
-P_{21}(\tilde{\mathcal{T}} + P_{11})^{-1} & \text{Id}
\end{pmatrix}
\]

and we write

\[
H(\mathcal{T} + P)G = \begin{pmatrix}
\tilde{\mathcal{T}} + P_{11} & 0 \\
0 & -P_{21}(\tilde{\mathcal{T}} + P_{11})^{-1}P_{12} + P_{22}
\end{pmatrix}.
\]

Let us denote \( \mathcal{A} := -P_{21}(\tilde{\mathcal{T}} + P_{11})^{-1}P_{12} + P_{22} \) which goes from \( \text{Ker}(\mathcal{T}) \) to \( D \). Since these two spaces have the same (finite) dimension we have from the rank nullity theorem

\[
\dim(\text{Ker}(\mathcal{A})) = \dim(\text{Coker}(\mathcal{A})).
\]

Since \( G, H \) are invertible, and, for \( P \) small enough, \( \tilde{\mathcal{T}} + P_{11} \) is invertible, we deduce that

\[
\dim(\text{Ker}(\mathcal{T})) = \dim(\text{Ker}(\mathcal{A})), \quad \dim(\text{Coker}(\mathcal{T})) = \dim(\text{Coker}(\mathcal{A}))
\]

which concludes the proof of the claim.

Now going back to the operator \( \text{Id} + \mathcal{R} \) we consider the path \( \text{Id} + t\mathcal{R}, t \in [0, 1] \) and perform a continuation argument. The interval \( [0, t_0) \), \( t_0 > 0 \), for which the properties above are satisfied is non-empty and open from the previous claim. But since \( \text{Id} + t_0\mathcal{R} \) still has finite range and finite dimensional kernel and cokernel (as \( t_0\mathcal{R} \) is compact), the same argument could be performed around \( t_0 \) as well, which shows that \( t_0 = 1 \), and therefore concludes the proof.

\[\Box\]

### 3.7. Compactness tools in PDEs

To finish the proof above, it remains to prove the compactness of \( K = -\lambda_0(P + \lambda_0\text{Id})^{-1} \) on \( L^2(U) \).

Let us first give some recalls. We consider a separable Hilbert space \( \mathcal{H} \).

**Definition 3.7.** *The weak compactness of a sequence \( u_n \in \mathcal{H} \) in a Hilbert space \( \mathcal{H} \) means that there is a subsequence \( u_{\varphi(n)} \) so that

\[
\forall v \in \mathcal{H}, \quad \langle u_{\varphi(n)}, v \rangle \to \langle u_\infty, v \rangle
\]

for some \( u_\infty \in \mathcal{H} \).*

*The strong compactness of a sequence \( u_n \in \mathcal{H} \) in a Hilbert space \( \mathcal{H} \) means that there is a subsequence \( u_{\varphi(n)} \) so that

\[
\| u_{\varphi(n)} - u_\infty \|_\mathcal{H} \to 0
\]

for some \( u_\infty \in \mathcal{H} \).*

*The weak topology on \( \mathcal{H} \) is the topology induced by continuous linear forms on \( \mathcal{H} \), i.e. the dual \( \mathcal{H}^* \), which turns out to isometric to \( \mathcal{H} \) by Riesz representation theorem.*

Exercise 37. The unit ball of $\mathcal{H}$ is strongly compact iff $\mathcal{H}$ is finite dimensional. Show that in this case the weak and strong topologies coincide.

Hint: If $\mathcal{H}$ is finite dimensional the implication is Bolzano-Weierstrass. Else argue by contradiction and construct a sequence with no cluster points (with distances uniformly bounded from below between all pairs).

Exercise 38. Prove however that the unit ball is always weakly compact.

Hint: consider a Hilbertian basis and perform a diagonal process.

Exercise 39. Show that the weak topology cannot be associated to any metric on the whole space $\mathcal{H}$. Show however that the unit ball of $\mathcal{H}$, endowed with the weak topology, can be endowed with a metric.

Hint for the first part: Show that an open set for the weak topology always contains a whole line.

In a non-reflexive Banach space, the situation is more complicated, there are two non-strong topologies: the weak and weak-* topologies, and the Banach-Alaoglu theorem gives compactness of the unit ball of $\mathcal{H}^*$ for weak-* topology. The most general form relies upon Tychonov’s theorem and the axiom of choice.

Exercise 40. Show that the Banach-Alaoglu theorem for the dual of a separable space can be proved without using the axiom of choice, thanks to a countable diagonal argument.

As an important remark in PDEs we almost always have spaces with the separability property and we thus almost never need the axiom of choice. The space of Schwartz’ distributions may suggest the opposite, but any concrete problem lives in a subset of the set of distributions with countable basis of neighborhoods (e.g. tempered distributions).

Exercise 41. What are the conditions for a sequence to be weakly-* compact in $L^p(\mathbb{R})$, $p \in (1, \infty)$? in $L^1(\mathbb{R})$?

Let us now turn to the question of strong compactness. Here the “root” of all theorems is Arzelà-Ascoli’s Theorem:

**Theorem 3.8 (Arzelà-Ascoli’s Theorem).** A sequence $f_n$ of continuous real functions on a compact Hausdorff space $X$ is relatively compact in the topology induced by the uniform norm if and only if it is equicontinuous: for any $\varepsilon > 0$ and $x \in X$, there is $V$ neighborhood of $x$ so that

$$\forall y \in V, \quad |f(x) - f(y)| \leq \varepsilon,$$

and pointwise bounded: for any $x \in X$

$$\sup_{n \geq 0} |f_n(x)| < +\infty.$$

**Remark 3.9.** Observe that the boundedness assumption is reminiscent of weak compactness statements. However the equicontinuity is an assumption of uniform regularity along the sequence that proscribes the possibility of oscillations. As a general principle, the main obstacle to weak compactness is the divergence, and the main obstacle to a weak compactness being strong is oscillations (i.e. divergence in Fourier variable) which can be ruled out by uniform regularity assumption. The Arzelà-Ascoli Theorem is in fact the origin of most compactness Theorem in analysis.
Proof of Theorem 3.8. Construct a countable dense subset $Y$ of $X$, then show convergence of a subsequence of $f_n(x)$ for any $x \in Y$, and find a subsequence convergence for all points of $Y$ by a diagonal argument. This defines a limit $f$ on $Y$. This equicontinuity implies continuity of this limit. It then can extended by density to a continuous function on $X$. The uniform convergence follows from the equicontinuity and the pointwise convergence on $Y$. □

Exercise 42. Give example of sequences converging weakly but not strongly. Show that the weak limit of a product is not in general the product of the weak limits. Show that weak convergence plus convergence of the norm implies strong convergence.

Exercise 43. Search in textbooks some sufficient conditions to strong compactness of a set of functions of $L^p(\mathbb{R})$, $p \in (1, +\infty)$?

We shall conclude this subsection with the following fundamental compactness theorem:

Theorem 3.10 (Rellich-Kondrachov Compactness Theorem). Assume $\mathcal{U} \subset \mathbb{R}^\ell$ is a bounded open set with $\partial \mathcal{U}$ smooth ($C^1$ is enough), then

$$H^1(\mathcal{U}) \subset L^2(\mathcal{U})$$

which means that the canonical inclusion is compact.

Remark 3.11. If one restricts to $H^1_0(\mathcal{U})$ then the $C^1$ smoothness on the boundary can be relaxed.

Proof of Theorem 3.10. Let us only sketch the proof as it requires tools from Sobolev inequalities.

In dimension $\ell = 1$, we know from Morrey’s inequality (cf. mid-term assignments) that $H^1(\mathcal{U}) \subset C^{1/2-0}(\mathcal{U})$. Then the classical Arzelà-Ascoli Theorem on the compact set $\mathcal{U}$ concludes the proof, as the Hölder regularity implies the equicontinuity, and $L^\infty$ bound implies the required boundedness.

In higher dimension $\ell \geq 2$, the Sobolev inequalities show the inclusion and only the compactness of the inclusion remains to be proved. Then one considers a sequence $u_n \in H^1(\mathcal{U})$ uniformly bounded, extends the function to $\mathbb{R}^\ell$ with a common support $K$. Next one shows that a uniformly bounded sequence in $H^1(\mathbb{R}^\ell)$ can be uniformly approximated in $L^2(\mathcal{U})$ by the convolution with an approximation of the unit. Finally apply Arzelà-Ascoli Theorem on $K$ on each approximation to show its compactness in $L^2(K)$, and use a standard diagonal argument. □

3.8. Back to the end of the proof. We consider a sequence $u_n \in L^2(\mathcal{U})$ with $\|u_n\|_{L^2(\mathcal{U})} \leq 1$, $n \geq 0$. Let us denote $v_n := \mathcal{R}u_n \in H^1_0(\mathcal{U})$. Then we have

$$\mathcal{P} + \lambda_0 \text{Id})v_n = -\lambda_0 u_n$$

and we can perform the following estimate, which takes advantage of the coercivity estimate on $\mathcal{P} + \lambda_0 \text{Id})$:

$$\|v_n\|_{H^1_0(\mathcal{U})} \lesssim \|u_n\|_{L^2(\mathcal{U})}.$$
Theorem 3.12 (Weak solutions to the general Dirichlet problem). We assume that the second-order elliptic operator \( \mathcal{P} \) has coefficients \( a_{ij}, b_i, c \in L^\infty(\mathcal{U}) \) in divergence form, and satisfies the uniform ellipticity condition on \( \mathcal{U} \). We consider the problem
\[
\begin{align*}
\mathcal{P} u &= f \quad \text{in } \mathcal{U} \\
u &= 0 \quad \text{on } \partial\mathcal{U}.
\end{align*}
\] where \( f \in H_0^{-1}(\mathbb{R}^d) \).

Then there are two exclusive possible cases:

\begin{enumerate}
\item Either there is a unique weak solution \( u \in H_0^1(\mathcal{U}) \) to this elliptic problem \( (\dim \ker(\text{Id} + \mathcal{R})) = 0 \) above,
\item or else there is at least one non-zero weak solution \( u \in H_0^1(\mathcal{U}) \) to the homogeneous problem \( (\dim \ker(\text{Id} + \mathcal{R})) \neq 0 \) above
\end{enumerate}

Remark 3.13. As said before one could consider the more general Dirichlet condition \( u = g \) on \( \partial\mathcal{U} \) where \( g \) is the restriction of an \( H^1(\mathbb{R}^d) \) function to \( \partial\mathcal{U} \).

Remark 3.14. In the situation (II) the dimension of the subspace \( N \) of weak homogeneous solutions to (3.6) is the same as the dimension of the subspace \( N^\ast \) of weak homogeneous solutions to the adjoint problem
\[
\begin{align*}
\mathcal{P}^\ast u &= 0 \quad \text{in } \mathcal{U} \\
u &= 0 \quad \text{on } \partial\mathcal{U}.
\end{align*}
\]

Remark 3.15. In the situation (II), the necessary and sufficient condition for the existence of a solution to the original problem (3.5) is
\[
\forall v \in N^\ast \in H_0^1(\mathcal{U}), \quad \langle f, v \rangle = 0.
\]

Remark 3.16. Observe finally that the reduction to the homogeneous problem is not simply a consequence of linearity due to the boundary conditions.

4. Regularity study of elliptic equations

We now address the question as to whether a weak solution to \( \mathcal{P} u = f \) in \( \mathcal{U} \) is smooth. This is the regularity problem for weak solutions, which we had already encountered for classical solutions in the whole space. We must distinguish between interior regularity and boundary regularity. We shall start with a warm-up on the key a priori estimate and some comments on the question of the approximation argument.

4.1. A priori estimate and approximation argument. Let us first consider the Poisson equation, and note that our energy estimate holds for weak solutions. For, choosing \( v = u \) in the definition of a weak solution, we have
\[
\|u\|_{H^1(\mathcal{U})}^2 = \langle f, u \rangle \leq \|f\|_{L^2(\mathcal{U})} \|u\|_{L^2(\mathcal{U})} \lesssim \|f\|_{L^2(\mathcal{U})} \|u\|_{H^1(\mathcal{U})}.
\]
where we have used the Poincaré inequality, and therefore by dividing by the $H^1(U)$ of $u$ we get 

$$\|u\|_{H^1(U)}^2 \lesssim \|f\|_{L^2(U)}^2.$$ 

But if we remember the previous a priori estimate we used in the regularity study, it is reasonable to expect that, away from boundary, we would have the stronger estimate 

$$\|u\|_{H^2(V)}^2 \lesssim \|f\|_{L^2(U)}^2$$ 

where $V$ is an open set with $V \subset U$. This leaves open the question of what happens at the boundary. However we shall see that the control of tangential derivatives combined with the differential relation provided by the PDE and the fact that $\partial U$ is non-characteristic will allow to control the tangential derivative. Then our previous estimate indeed extends up to the boundary, and we have 

$$\|u\|_{H^2(U)}^2 \lesssim \|f\|_{L^2(U)}^2.$$ 

Finally it is not hard to convince oneself formally that one should expect the same a priori estimate to hold at any order of regularity: 

$$\forall s \in \mathbb{N}, \quad \|u\|_{H^{s+2}(U)}^2 \lesssim \|f\|_{H^s(U)}^2.$$ 

Let us now consider the more general case of an elliptic equation in the form (3.5) with a second-order elliptic operator $\mathcal{P}$ with coefficients $a_{ij}, b_i, c \in L^\infty(U)$ in divergence form, and which satisfies the uniform ellipticity condition on $U$. We assume that we are in the situation (I) in the theorem 3.12 above, i.e. there is no non-zero solution to the homogeneous problem. Then the previous proof shows that the solution map $S : L^2(U) \rightarrow H^1_0(U)$ is well-defined, linear and continuous. In particular the $L^2(U)$ norm of $u$ is controlled by that of $f$. This allows to "fix" the following incomplete a priori estimate that we have already seen: 

$$\|u\|^2_{H^1(U)} = \langle (u, u) \rangle \leq \langle \mathcal{P}u, u \rangle + \|u\|^2_{L^2(U)} \lesssim C_U \|f\|_{L^2(U)} \|u\|_{L^2(U)} + \|u\|_{L^2(U)}^2 \lesssim \|f\|_{L^2(U)}^2 + \|u\|_{L^2(U)}^2 \lesssim \|f\|_{L^2(U)}^2$$ 

in the sense that the RHS is now fulling under control (in the last inequality we use the continuity of the solution map) and one can use the estimate to gain further knowledge on the higher regularity of the solution. In other words the inductive structure of the a priori estimate is still preserved.

**Remark 4.1.** Actually the discussion on the last equation is slightly artificial since the solution map already provides the $H^1_0(U)$ regularity. However the idea is important for higher regularity, and leads to the next a priori estimate.

Going back the full elliptic regularity estimate again now, we expect that, away from boundary, we would have the stronger estimate 

$$\|u\|_{H^2(V)}^2 \lesssim \|f\|_{L^2(U)}^2 + \|u\|_{H^1(U)}^2 \lesssim \|f\|_{L^2(U)}^2 + \|u\|_{L^2(U)}^2.$$
where $\mathcal{V}$ is an open set with $\overline{\mathcal{V}} \subset \mathcal{U}$, and we have used the previous estimate in the last inequality. Using the fact that the boundary is non-characteristic for an elliptic PDE we would then expect formally to extend it to

$$\|u\|_{H^2(\mathcal{U})}^2 \lesssim \|f\|_{L^2(\mathcal{U})}^2 + \|u\|_{L^2(\mathcal{U})}^2.$$  

Finally it is again not hard to convince oneself formally that one should expect the same a priori estimate to hold at any order of regularity:

$$\forall s \in \mathbb{N}, \quad \|u\|_{H^{s+2}(\mathcal{U})}^2 \lesssim \|f\|_{H^s(\mathcal{U})}^2 + \|u\|_{H^s(\mathcal{U})}^2 \lesssim \|f\|_{H^s(\mathcal{U})}^2 + \|u\|_{L^2(\mathcal{U})}^2 \lesssim \|f\|_{H^s(\mathcal{U})}^2.$$  

This “starting point”, i.e. solving the equation in some “ground functional space” was provided by the careful study of the kernel of a Fredholm operator. Once obtained however the a priori estimate can now be used again, as it shows a gain of two derivative and can be used “inductively”.

Now the question is again: how to transform these formal arguments into rigorous ones? Observe that the previous calculations do not constitute a proof, as they assume the regularity of $u$. We need again some approximation argument. The point is to derive analytic estimates from the structural, algebraic assumption of ellipticity.

We already, in the beginning of this chapter, presented a classical such approximation argument based on the convolution by a mollifier. The proofs in the rest of this section could be performed with this approximation, which is a good exercise. We shall present another interesting approximation argument, based on the discretisation of the differentiation process.

**Definition 4.2.** We introduce the so-called difference quotients. For each coordinate $i$, and each $h \in \mathbb{R} \setminus \{0\}$, we define

$$D^h_i u := \frac{u(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_L) - u(x_1, \ldots, x_L)}{h}.$$

Let us prove that (1) if $u \in H^s(\mathbb{R}^L)$, then so does $D^h_i u$, (2) conversely if $D^h_i u$ is uniformly bounded in $L^2(\mathbb{R}^L)$ as $h \to 0$ then $\partial_i u \in L^2(\mathbb{R}^L)$. And similarly for functions which are supported inside some open set $\mathcal{V}$ whose closure is contained in $\mathcal{U}$, as long as $h$ is sufficiently small. Since classical derivatives commute with the difference quotients, it is enough to prove the result for $s = 0$.

**Proposition 4.3** (Manipulation of the difference quotient). First for $u \in L^2(\mathcal{V})$ with support included in $\mathcal{V}$, there is $C_1 > 0$ so that

$$\forall |h| \in (0, h_0), \quad \left\| D^h_i u \right\|_{L^2(\mathcal{V})} \leq C_1 \| \partial_i u \|_{L^2(\mathcal{U})}$$

where $h_0 < \text{dist}(\overline{\mathcal{V}}, \partial \mathcal{U})$ is so that $D^h_i u$ can be defined on $\mathcal{V}$ from the values of $u$ in $\mathcal{U}$.

Second assume that for $u \in L^2(\mathcal{V})$ with support included in $\mathcal{V}$ there is a constant $C_2 > 0$ so that

$$\forall |h| \in (0, h_0), \quad \left\| D^h_i u \right\|_{L^2(\mathcal{V})} \leq C_2,$$

then $u \in H^1(\mathcal{V})$ with

$$\| \partial_i u \|_{L^2(\mathcal{V})} \leq C_2.$$
Proof of Proposition 4.3. The first part relies on the Taylor integral formula:

\[ u(x + he_i) - u(x) = \int_0^1 \partial_i u(x + t(e_i)h) dt \]

and therefore

\[ \int_V \left| D_i^h \right|^2 dx \lesssim \left| h_0 \right| \int_0^1 \left( \int_V \left| \partial_i u(x + te_i) \right|^2 dx \right) dt \lesssim \int_U \left| \partial_i u(y) \right|^2 dy. \]

For the second part, we first note the following discrete integration-by-parts formula:

for \( u \in L^2(V) \) with support included in \( V \) and \( \varphi \in C_c^\infty(V) \)

\[ \int_V u(x) \left[ \frac{\varphi(x + he_i) - \varphi(x)}{h} \right] dx = -\int_V \left[ \frac{u(x - he_i) - u(x)}{h} \right] \varphi(x) dx \]

which means

\[ \int_V u(x) D_i^h \varphi(x) dx = -\int_V D_i^{-h} u(x) \varphi(x) dx \]

The assumption means that

\[ \sup_{|h| \in (0, h_0)} \left\| D_i^{-h} u \right\|_{L^2(V)} < \infty \]

and therefore (weak compactness of the unit ball) there is a subsequence \( h_k \to 0 \) so that \( D_i^{-h_k} u \to v_i \) for some \( v_i \in L^2(V) \) with

\[ \left\| v_i \right\|_{L^2(V)} \leq \sup_{|h| \in (0, h_0)} \left\| D_i^{-h} u \right\|_{L^2(V)} \]

But then if \( \varphi \in C_c^\infty(V) \) we obviously have \( D_i^h \varphi \to \partial_i \varphi \) with uniform (and \( L^2 \)) convergence, and therefore

\[ \int_V u \partial_i \varphi dx = \int_U u \partial_i \varphi dx = \lim_{h_k \to 0} \int_U u D_i^{h_k} \varphi dx. \]

But the RHS is also

\[ \int_U u D_i^{h_k} \varphi dx = -\int_U D_i^{-h_k} u \varphi dx \xrightarrow{h_k \to 0} -\int_U v_i \varphi dx \]

from the weak convergence. This implies that

\[ \forall \varphi \in C_c^\infty(V), \quad \int_V u \partial_i \varphi dx = -\int_U v_i \varphi dx. \]

This shows from the definition of generalised derivatives that the generalised derivative \( \partial_i u = v_i \) exists and is \( L^2(V) \), which concludes the proof. \( \square \)
4.2. Interior regularity. We first consider the case $s = 0$ in the gain of regularity, with $f \in L^2(\mathcal{U})$ and the usual other assumptions, plus $a_{ij}, b_i, c \in C^1(\mathcal{U}) \cap L^\infty(\mathcal{U})$. Consider the situation (I) and a solution $u$ to the elliptic problem $\mathcal{P}u = f$ in $H^1(\mathcal{U})$.

Let us show that $u \in H^2_{\text{loc}}(\mathcal{U})$. This is the the space of locally $H^2$ functions, i.e. that are $H^2$ on any compact included in $\mathcal{U}$. Therefore we want to show that for any open subset $\mathcal{V} \subset \subset \mathcal{U}$ (i.e. $\mathcal{V} \subset \subset \mathcal{U}$) we have the estimate

$$\|u\|^2_{H^2(\mathcal{V})} \leq C \left( \|f\|^2_{L^2(\mathcal{U})} + \|u\|^2_{L^2(\mathcal{U})} \right)$$

where the constant $C > 0$ depends on $\mathcal{U}$, $\mathcal{V}$, and the coefficients of $L$.

Remark 4.4. The boundary conditions are not required in the proof here, as we shall stay away from the boundary, i.e. we do not require $u \in H^1_0(\mathcal{U})$ here but only $u \in H^1(\mathcal{U})$ and solves the PDE in weak sense: it is enough that $u$ is a (we do not care about uniqueness here) solution to $\mathcal{P}u = f$ in the open set $\mathcal{U}$ in weak sense, and that $a_{ij}, b_i, c \in C^1(\mathcal{U})$ (no $L^\infty(\mathcal{U})$ is required on the coefficients in the regularity study of the interior, unless one is interested in how the estimates “degenerate” at the boundary; similarly for higher regularity we will only need $a_{ij}, b_i, c \in C^k(\mathcal{U})$).

Remark 4.5. Observe also that it allows to make sense of the PDE in the almost everywhere sense, as it shows that in the open set $\mathcal{U}$, a distributional second-order derivative exists in $L^2_{\text{loc}}(\mathcal{U})$, and by integration by parts one can recover $\mathcal{P}u = f$ almost everywhere from the weak form.

Choose an open set $\mathcal{W}$ with $\mathcal{V} \subset \subset \mathcal{W} \subset \subset \mathcal{U}$. Define a localisation smooth function $\zeta \in C^\infty_c(\mathcal{V})$ with support in $\mathcal{U}$ and which is one on $\mathcal{V}$ and zero outside $\mathcal{W} \subset \subset \mathcal{U}$, and the following test function

$$v := -D_i^{-h}(\zeta^2 D_i^h u)$$

for any index $i$. We then use the weak formulation

$$\int_{\mathcal{U}} (\nabla u)^T A(x)(\nabla v) \, dx = \int_{\mathcal{U}} \tilde{f} v \, dx, \quad \tilde{f}(x) := f(x) - \nabla \cdot (B(x)u) - c(x)u(x).$$

From the construction of weak solution in $H^1_0(\mathcal{U})$ and the assumption on the coefficients $b_i, c$, we know that $\tilde{f} \in L^2(\mathcal{V})$. We know calculate for the LHS

$$\int_{\mathcal{U}} (\nabla u)^T A(x)(\nabla v) \, dx = -\int_{\mathcal{U}} (\nabla u)^T A(x)(\nabla D_i^{-h}(\zeta^2 D_i^h u)) \, dx$$

$$= -\int_{\mathcal{U}} (\nabla u)^T A(x)(D_i^{-h}\nabla(\zeta^2 D_i^h u)) \, dx$$

$$= \int_{\mathcal{U}} (D_i^h \nabla u)^T A(x)(\nabla(\zeta^2 D_i^h u)) \, dx + \int_{\mathcal{U}} (\nabla u)^T D_i^h A(x)(\nabla(\zeta^2 D_i^h u)) \, dx$$

$$= \int_{\mathcal{U}} \zeta^2 (D_i^h \nabla u)^T A(x)(\nabla(\zeta D_i^h u)) \, dx + \int_{\mathcal{U}} 2(D_i^h \nabla u)^T A(x)(\nabla \zeta) \zeta (D_i^h u) \, dx$$

$$+ \int_{\mathcal{U}} \zeta^2 (\nabla u)^T D_i^h A(x)(\nabla(\zeta D_i^h u)) \, dx + 2 \int_{\mathcal{U}} (\nabla u)^T D_i^h A(x)(\nabla \zeta) \zeta D_i^h u \, dx$$

$$=: I_1 + I_2 + I_3 + I_4$$
where we have used the commutation $\nabla D_i^h = D_i^h \nabla$, and the distributivity $D_i^h(uv) = (D_i^h u)v + u(D_i^h v)$.

We calculate for the first term, using the ellipticity assumption,

$$I_1 \geq \lambda e \int_U \zeta^2 \left| D_i^h \nabla u \right|^2 \, dx.$$

We calculate for the second term, using Cauchy-Schwarz’s inequality and the bounds on $\zeta$,

$$I_2 \geq -\varepsilon \int_U \zeta^2 \left| D_i^h \nabla u \right|^2 \, dx - C \varepsilon \int_U \left| \nabla \zeta \right|^2 \left| D_i^h u \right|^2 \, dx,$$

$$\geq -\varepsilon \int_U \zeta^2 \left| D_i^h \nabla u \right|^2 \, dx - \frac{C}{\varepsilon} \|u\|_{H^1(W)}^2,$$

for any $\varepsilon > 0$ and some constant $C > 0$. We calculate similarly for the third term, using Cauchy-Schwarz’s inequality,

$$I_3 \geq -\varepsilon \int_U \zeta^2 \left| D_i^h \nabla u \right|^2 \, dx - C \varepsilon \int_U \zeta^2 \left| D_i^h u \right|^2 \, dx,$$

$$\geq -\varepsilon \int_U \zeta^2 \left| D_i^h \nabla u \right|^2 \, dx - \frac{C}{\varepsilon} \|u\|_{H^1(W)}^2.$$

Finally we calculate for the last term, using the bounds on $\zeta$,

$$I_4 \geq -C\|u\|_{H^1(U)}^2.$$

Choosing $\varepsilon = \lambda e / 3$ we deduce that

$$I_1 + I_2 + I_3 + I_4 \geq \frac{\lambda e}{3} \int_U \zeta^2 \left| D_i^h \nabla u \right|^2 \, dx - C'\|u\|_{H^1(W)}^2$$

for some other constant $C' > 0$.

Going back to the weak formulation, we now estimate the RHS:

$$\int_U \hat{f} \nabla v \, dx \lesssim \left\| \hat{f} \right\|_{L^2(W)} \left\| v \right\|_{L^2(W)}$$

$$\lesssim \left\| \hat{f} \right\|_{L^2(W)} \left\| \nabla^2 u \right\|_{L^2(W)}$$

$$\leq C_\varepsilon \|f\|_{L^2(W)}^2 + C_\varepsilon \|u\|_{H^1(W)} + \varepsilon \|\nabla^2 u\|_{L^2(W)}^2$$

for any $\varepsilon > 0$ and some corresponding constant $C_\varepsilon > 0$. We deduce that

$$\int_U \left| D_i^h \nabla u \right|^2 \, dx \leq \int_U \zeta^2 \left| D_i^h \nabla u \right|^2 \, dx \leq C' \|\hat{f}\|_{L^2(W)}^2 + C' \|u\|_{H^1(W)} + \varepsilon \|\nabla^2 u\|_{L^2(W)}^2.$$

Finally using the property of the difference quotient we deduce

$$\|\partial_i \nabla u\|_{L^2(W)}^2 \leq C' \|\hat{f}\|_{L^2(W)}^2 + C' \|u\|_{H^1(W)} + \varepsilon \|\nabla^2 u\|_{L^2(W)}^2.$$

and since it is true for any $i = 1, \ldots, \ell$,

$$\|\nabla^2 u\|_{L^2(W)}^2 \leq C' \|\hat{f}\|_{L^2(W)}^2 + C' \|u\|_{H^1(W)} + \varepsilon \|\nabla^2 u\|_{L^2(W)}^2.$$
By taking \( \varepsilon \) small enough it concludes the proof of

\[
\| \nabla^2 u \|^2_{L^2(\Omega)} \lesssim \| f \|^2_{L^2(\Omega)} + \| u \|^2_{H^1(\Omega)}.
\]

Let us now explain how to prove higher-order interior regularity estimates. We argue by induction with the induction assumption

\[
(\mathcal{H}_m) \quad \begin{cases} a_{ij}, b_i, c \in C^{m+1}(\Omega) \\ f \in H^m_{\text{loc}}(\Omega) \\ \Psi u = f \text{ in any } \Omega \subset \subset \Omega \end{cases} \implies u \in H^{m+2}_{\text{loc}}(\Omega)
\]

on \( m \in \mathbb{N} \).

The initialisation of the induction (\( \mathcal{H}_0 \)) was proven above. Assume \( (\mathcal{H}_m) \) holds with \( m \geq 1 \), and assume that \( a_{ij}, b_i, c \in C^{m+2}_B(\Omega) \) and \( f \in H^{m+1}_{\text{loc}}(\Omega) \). Let us prove that \( u \in H^{m+3}_{\text{loc}}(\Omega) \). Consider any first order derivative \( \partial_i \), and estimate the weak formulation with \( -\partial_i \bar{v}, \tilde{v} \in C^\infty_c(\Omega) \) (with support included in \( \Omega \subset \subset \Omega \)).

Let us recall the weak formulation

\[
\forall v \in H^1_0(\Omega), \quad \Psi(u, v) = \mathcal{L}v
\]

with

\[
\begin{align*}
\Psi(u, v) &= \int_{\Omega} (\nabla u)^T A(x) (\nabla v) \, dx + \int_{\Omega} v B(x) \cdot \nabla u \, dx + \int_{\Omega} c uv \, dx, \\
\mathcal{L}v &= (f, v).
\end{align*}
\]

and let us also recall that is enough to check this formulation for \( v \in C^\infty_c(\Omega) \) by approximation of \( H^1_0(\Omega) \) functions.

An easy calculation by integration by parts shows that

\[
\begin{align*}
\int_{\Omega} (\nabla \partial_i u)^T A(x) (\nabla \tilde{v}) \, dx + \int_{\Omega} \tilde{v} B(x) \cdot \nabla \partial_i u \, dx + \int_{\Omega} c \partial_i w \tilde{v} \, dx \\
&= \int_{\Omega} (f - \partial_i cu - \partial_i B(x) \cdot \nabla u) \tilde{v} \, dx
\end{align*}
\]

which means that \( \tilde{u} = \partial_i u \) satisfies

\[
\begin{align*}
\Psi(\tilde{u}, \tilde{v}) &= \tilde{\mathcal{L}} \tilde{v}, \\
\tilde{\mathcal{L}} v := \int_{\Omega} \tilde{f} v \, dx, \\
\tilde{f} := f - \partial_i cu - \partial_i B(x) \cdot \nabla u.
\end{align*}
\]

Since \( m \geq 1 \), we know that \( u \in H^3(\Omega) \), and therefore the equation translates into

\[
\Psi \tilde{u} = \tilde{f}
\]

in weak sense on \( \Omega \). Since \( \tilde{f} \in H^m(\Omega) \) we deduce from the previous study that \( \tilde{u} = \partial_i u \in H^{m+2}(\Omega) \). Since it is true for any first-order partial derivative \( \partial_i \) and any \( \Omega \subset \subset \Omega \) we deduce that \( u \in H^{m+3}_{\text{loc}}(\Omega) \). This concludes the proof of the induction.

We deduce from this subsection the following property, similar to that we have proved for classical solutions in the whole plane: if \( a_{ij}, b_i, c, f \in C^\infty(\Omega) \) and \( u \in H^1_0(\Omega) \) is a weak solution to \( Pu = f \), with the uniform ellipticity assumption, then \( u \in C^\infty(\Omega) \). Observe that it says nothing so far about the possible singularities of \( u \) at the boundary, only that any such possible singularities do not propagate into the interior.
4.3. Boundary regularity. We first consider again the gain of $H^2(U)$ regularity, but in the whole domain, i.e. including the boundary. We consider $a_{ij}, b_i, c \in C^1(\overline{U})$, and assume that $\partial U$ is $C^2$. As became clear from the previous subsection, the regularity study is an a priori study. This notion is simple but extremely important: this means that the argument applies to any solutions that satisfies a minimal set of assumptions.

To be concrete we proved the interior regularity as soon as $a_{ij}, b_i, c, f \in C^\infty(U)$ and $u \in H^1(U)$ solves the weak sense $\mathfrak{P}u = f$ in the open set $U$.

In the case of the boundary regularity we now need to impose, on the contrary to the previous case, some boundary conditions on $u$, and we shall therefore assume that $u \in H^1_0(U)$. We also need now some estimates on the coefficients and the forcing term $f$ that are uniform when approaching the boundary. Recall that if, furthermore, we assume that $u$ is the solution to the Dirichlet problem given by the situation $(I)$ in the study we made before, then we have the additional information

$$\|u\|_{H^1_0(U)} \lesssim \|f\|_{L^2(U)}.$$ 

We now want to show that $u \in H^2(U)$ with the precise estimate

$$\|u\|_{H^2(U)} \lesssim \|f\|_{L^2(U)} + \|u\|_{H^1(U)}.$$ 

From the previous interior regularity study it is clear that it is enough to prove the result in neighborhoods of the boundary.

We first consider the case where $U = B(0,1) \cap \mathbb{R}^\ell_+$ is a half-ball, with $\mathbb{R}^\ell_+ = \{x_\ell \geq 0\}$. Consider the subset $\mathcal{V} = B(0,1/2) \cap \mathbb{R}^\ell_+$ and some smooth localisation function $\zeta$ that is $1$ on $\mathcal{V}$ and zero outside $B(0,1)$, with $0 \leq \zeta \leq 1$.

For any coordinate $i = \{1, \ldots, \ell - 1\}$ that is tangential to the boundary, we consider the test function $v := -D_i^{-h}(\zeta^2 D_i^h u)$, which belongs to $H^1_0(U)$. Indeed heuristically

$$v(x) = -\frac{1}{h^2} (\zeta^2 (x - he_i)(u(x) - u(x - he_k)) - \zeta^2(x)(u(x + he_k) - u(x))) .$$

and the RHS cancels at $x_\ell = 0$. Rigorously we must prove that there is $v_n \to v$ with $v_n \in C_c^\infty(U)$: indeed a sequence $u_n \to u$, $u_n \in C_c^\infty(U)$ certainly exists for $u \in H^1_0(U)$, and then one checks that

$$v_n(x) := -\frac{1}{h^2} (\zeta^2 (x - he_i)(u_n(x) - u_n(x - he_k)) - \zeta^2(x)(u_n(x + he_k) - u_n(x)))$$

belongs to $C_c^\infty(U)$ with $v_n \to v$ in $H^1(U)$.

Then reproducing the same key calculation performed for the interior regularity (relying on the uniform ellipticity estimate and the bound on the coefficients) we easily get

$$\int_{\mathcal{V}} \left| D_i^{-h} \nabla u \right|^2 \, dx \leq \int_{U} \zeta^2 \left| D_i^{-h} \nabla u \right|^2 \, dx \lesssim \|f\|_{L^2(U)}^2 + \|u\|_{H^1(U)}^2,$$

(observe that here we use bounds on the coefficients up to the boundary) which shows that

$$\forall i \in \{1, \ldots, \ell - 1\}, \quad \|\partial_i u\|_{L^2(\mathcal{V})}^2 \lesssim \|f\|_{L^2(\mathcal{V})}^2 + \|u\|_{H^1(\mathcal{V})}^2.$$

We still need to estimate the derivative along the last coordinate $\partial_\ell$. But then observe that the uniform ellipticity $A(x) \geq \lambda_\ell > 0$ on the symmetric matrix $A(x)$ implies that
(consider the quadratic form at the vector $e_{\ell}$) that $a_{\ell\ell} \geq \lambda_{e} > 0$. Hence using the following equality on $V$

$$a_{\ell\ell} \partial_{\ell}^2 u = - \sum_{j=1}^{\ell-1} \sum_{i=1}^{\ell} \partial_j (a_{ij}(x) \partial_i u(x)) - (\partial_l a_{\ell\ell}) \partial_l u - B(x) \cdot \nabla u - cu$$

between $L^2(V)$ functions, we deduce that $\partial_{\ell}^2 u \in L^2(V)$ (using the bounds on the coefficients), which finally shows that

$$\|u\|_{H^2(V)} \lesssim \|f\|_{L^2(\mathcal{U})} + \|u\|_{H^1(\mathcal{U})}.$$  

Let us consider the case of a general boundary shape. For the sake of simplicity we now rewrite the PDE in non-divergence form

$$\sum_{i,j=1}^{\ell} a_{ij}(x) \partial_{ij}^2 u + \sum_{i=1}^{\ell} b_i(x) \partial_i u + cu = f,$$

which changes only the coefficients $b_i, c$, still denoted with the same name by a slight abuse of notation. Consider a base point $x_0 \in \partial \mathcal{U}$ on the boundary. Using the regularity assumption on the boundary we have (for $r$ small enough so that any sphere with radius with $0 < r' \leq r$ intersects $\partial \mathcal{U}$ as a connected $(\ell - 2)$-hypersurface)

$$\mathcal{U} \cap B(x_0, r) = \{ x_{\ell} > \varphi(x_1, \ldots, x_{\ell-1}) \} \cap B(x_0, r)$$

where $\varphi : \mathbb{R}^{\ell-1} \to \mathbb{R}$ is $C^2$. We now perform the same kind of change of variable as in the proof of the Cauchy-Kovalevskaya theorem (however with less regularity): $\Phi \in C^2$ maps $\mathcal{U} \cap B(x_0, r)$ to

$$\tilde{\mathcal{U}}_{x_0} := \{ y \in B(0, \tilde{r}) : y_{\ell} > 0 \}$$

where furthermore the normal vectors to $\partial \mathcal{U}$ are mapped to $e_{\ell}$, and we write $\Psi$ for the inverse of $\Phi$. This maps the PDE to a new PDE

$$\tilde{\mathcal{P}} \tilde{u} = \tilde{f}$$

for a second-order linear operator $\tilde{\mathcal{P}}$ of the same form, with $\tilde{u}(y) = u(\Psi(y)), \tilde{f} = f(\Psi(y)), \tilde{c}(y) = c(\Psi(y))$, and (using the chain rule)

$$\tilde{a}_{ki}(y) := \sum_{i,j=1}^{\ell} a_{ij}(\Psi(y))(\partial_i \Phi)_k(\Psi(y))(\partial_j \Phi)_l(\Psi(y)),$$

$$\tilde{b}_k(y) := \sum_{i=1}^{\ell} b_i(\Psi(y))(\partial_i \Phi)_k(\Psi(y)).$$

Exercise 44. Check the previous transformation and prove that it preserves the regularity and pointwise bounds from above on the coefficients.

Let us now prove that the new operator $\tilde{\mathcal{P}}$ is uniformly elliptic. Consider any point $y \in \tilde{\mathcal{U}}_{x_0}$, and any $\xi \in \mathbb{R}^\ell$ and calculate

$$\tilde{a}_{\ell\ell}(x, \xi) = \sum_{k,l=1}^{\ell} \tilde{a}_{ki}(y) \xi_k \xi_l = \sum_{k,l=1}^{\ell} \sum_{i,j=1}^{\ell} a_{ij}(\Psi(y))(\partial_i \Phi)_k(\Psi(y))(\partial_j \Phi)_l(\Psi(y)) \xi_k \xi_l.$$
5. MAXIMUM PRINCIPLES FOR ELLIPTIC EQUATIONS

\[ = \sum_{i,j=1}^{\ell} a_{ij}(\Psi(y)) (\Xi D\Phi(\Psi(y)))_{i} (\Xi D\Phi(\Psi(y)))_{j} \]

where we denoted \( D\Phi(x) = ((\partial_{i}\Phi)_{j})_{i,j} \) the \( \ell \times \ell \) usual jacobian matrix, and \( \Xi = (\xi_{1}, \ldots, \xi_{\ell}) \). Let us denote \( \Theta = \Xi D\Phi(\Psi(y)) \). This relation can be inverted as \( \Xi = \Theta D\Phi(y) \). Now using the uniform bounds on the differential we have \( |\Theta|^2 \geq C_{\psi,\Phi} |\Xi|^2 \) and we deduce

\[ \hat{\sigma}_{p}(x, \xi) \geq \lambda_{c} |\Theta|^2 \geq \lambda_{c} C_{\psi,\Phi} |\Xi|^2 \]

which allows to apply the previous study of the half-ball and concludes the proof of the regularity in \( V \) by using the transformation \( \Psi \).

Finally since \( \partial U \) is compact we can use a finite covering by small enough balls, and taking the worst constant, it concludes the proof.

**Exercise 45.** Prove the higher-order boundary regularity statement: for \( a_{ij}, b_{i}, c_{2} \in C^{m+1}(U) \), a \( C^{m+2} \) boundary \( \partial U \) and \( f \in H^{m}(U) \), any weak solution \( u \in H^{1}_{0}(U) \) to \( Pu = f \) is \( H^{m+2}(U) \) with

\[ \|u\|_{H^{m+2}(U)} \leq \|f\|_{H^{m}(U)} + \|u\|_{L^{2}(U)}. \]

5. Maximum principles for elliptic equations

We now see a last fundamental property of elliptic (and parabolic!) equations, that of maximum principles. The idea is now to quantify in a pointwise (instead of integral as in the energy estimates) manner the “negative” aspect of the operator. To be more precise we want to exploit the following basis remark: if \( u \in C^{2} \) attains a maximum \( x_{0} \in U \) in the interior of the domain of definition of \( u \), then \( \nabla u(x_{0}) = 0 \) and \( \nabla^{2} u(x_{0}) \leq 0 \) (i.e. the symmetric matrix is matrix \( \nabla^{2} u \) is non-positive at \( x_{0} \)). It is clear that we will need then that \( u \) is a classical solution in \( C^{2} \) to exploit this idea. We know however from the previous study than any weak solution is indeed in \( C^{2} \) for a right-hand side \( f \) regular enough.

We consider a second-order differential operator in non-divergence form

\[ Pu = - \sum_{i,j=1}^{\ell} a_{ij}(x) \partial_{i,j}^{2} u(x) + B(x) \cdot \nabla u(x) + c(x) u(x) \]

and we assume the uniform ellipticity condition, as well as \( a_{ij}, b_{i}, c \in C^{0}(\overline{U}) \) (and the symmetry w.l.o.g. of \( A = (a_{ij}) \)).

5.1. Weak maximum principle with no zero-order term.

**Proposition 5.1.** Let us first assume that \( c = 0 \). Then the weak version of the maximum principle states that for any \( u \in C^{2}(U) \cap C^{0}(\overline{U}) \) so that \( Pu \leq 0 \) (resp. \( Pu \geq 0 \) in \( U \), then the maximum (resp. minimum) of \( u \) on \( \overline{U} \) is attained at the boundary \( \partial U \).

**Remark 5.2.** Such functions are called subsolutions (resp. supersolutions), just like for ordinary differential inequalities.
Let us prove this property. We obviously only need to study subsolutions thanks to the transformation \( u \to -u \). We shall reduce to the case where \( Pu < 0 \) on \( U \) by an important lifting argument (see later). So let us now assume that \( u \in C^2(U) \cap C^0(\overline{U}) \) with \( Pu < 0 \) on \( U \).

In a first stage, to see the naked idea of the proof, simplify a bit further and assume that \( Pu < 0 \) on \( U \). Then argue by contradiction: assume there is a point \( x_0 \in U \) so that \( u \) is maximum at \( x_0 \). Basic calculus of variation then shows that \( u(x_0) \leq 0 \) and \( Pu(x_0) = 0 \) which implies that \( Pu(x_0) \geq 0 \) and therefore a contradiction. In the general setting the matrix \( A(x_0) \) can be diagonalised by some orthogonal \( \ell \times \ell \)-matrix \( Q \):

\[
A(x_0) = Q^T \text{Diag}(\lambda_1, \ldots, \lambda_\ell)Q
\]

with \( \lambda_i \geq \lambda_e \) for any \( i \in \{1, \ldots, \ell\} \). Define the new variable \( y = x_0 + Q(x - x_0) \) by a “rotation” around \( x_0 \). Then at \( x_0 \):

\[
\sum_{i,j=1}^{\ell} a_{ij}(x) \partial^2_{x_i x_j} u(x_0) = \sum_{i,j=1}^{\ell} \sum_{k,l=1}^{\ell} a_{ij}(x) Q_{ki} Q_{lj} \partial^2_{y_k y_l} u(x_0) = \sum_{k=1}^{\ell} \lambda_k \partial^2_{y_k} u(x_0) \leq 0
\]

and we deduce again that \( Pu(x_0) > 0 \), which contradicts the assumption.

Let us now consider the lifting argument. This is a further instance of approximation: the key thing is that the property we are searching for is stable by uniform convergence approximation, whereas the closure of our assumption \( Pu < 0 \) under this convergence includes \( Pu \leq 0 \).

More precisely we define \( u_\varepsilon(x) = u(x) + \varepsilon \kappa x_1 \) with \( \varepsilon > 0 \) meant to be small, and \( \kappa \) meant to be large. Then prove that \( Pu_\varepsilon < 0 \) for \( \kappa \) large enough, uniformly as \( \varepsilon \to 0 \):

\[
-Pu_\varepsilon(x) + \varepsilon \kappa x_1 (a_{11}(x) + b_1(x)) \leq \varepsilon \kappa \varepsilon x_1 (-\kappa \lambda_e + \|b_1\|_\infty)
\]

which yields the required control by choosing \( \kappa \geq 2\|b_1\|_\infty / \lambda_e \).

Then we deduce, using the previous step, that

\[
\forall \varepsilon > 0, \quad \max_{\overline{U}} u_\varepsilon = \max \partial u_\varepsilon
\]

and pass to the limit \( \varepsilon \to 0 \) to conclude the proof.

5.2. Weak maximum principle with a signed zero-order term. Now let us consider the case where we include a zero-order term \( c \) in the formula for \( P \). In order to fit with the idea that the operator should push “up” (resp. down) when at a maximum (resp. minimum), we assume that \( c \geq 0 \), as the maximum (resp. minimum) will only be considered for non-negative (resp. non-positive) values.
Proposition 5.3. Assume \( Pu \leq 0 \) (resp. \( Pu \geq 0 \)) and \( c \geq 0 \) in \( U \), then

\[
\max_u u \leq \max_{\partial U} u_+ \quad \text{(resp.} \quad \min_u u \geq -\max_{\partial U} u_-)\]

where \( u_+ \geq 0 \) (resp. \( u_- \geq 0 \)) denotes the non-negative (resp. non-positive) part of \( u \).

The proof is straightforward by considering the region where \( u \geq 0 \). Consider w.l.o.g. the case of a subsolution \( Pu \leq 0 \) and define the open set \( V \subset U \) by

\[
V := \{ x \in U : u(x) > 0 \}.
\]

If \( V = \emptyset \) we are done, else observe that in \( V \) we have \( \tilde{P} u := Pu - cu \leq 0 \) and \( \tilde{P} \) has no zero-order term. We can therefore apply the previous result to conclude.