SELECTED CHAPTERS FROM THE THEORY OF PARTIAL DIFFERENTIAL EQUATIONS

CTIRAD MATYSKA

CHARLES UNIVERSITY IN PRAGUE
FACULTY OF MATHEMATICS AND PHYSICS
DEPARTMENT OF GEOPHYSICS
2005
Selected chapters from the theory of partial differential equations

Part I. Classical methods

1.9. Introductory notation and definitions

References (books for further reading):


Variational Methods in Mathematics, Science and Engineering; Reidel, Dordrecht - Boston 1976 (in English)


1) (Einstein) Convention: We will use Einstein's summation convention unless stated otherwise.

2) Notation: \( \Omega \subset \mathbb{R}^N \) is an \( N \)-dimensional domain (open connected set in \( \mathbb{R}^N \)).

\( C^k(\Omega) \) is the set of functions \( k \)-times continuously differentiable in \( \Omega \).

\( C^0(\Omega) \) is usually briefly written as \( C(\Omega) \).

\( C^k(\overline{\Omega}) \) is the set of functions \( k \)-times continuously differentiable in \( \overline{\Omega} = \Omega \cup \partial \Omega \), where \( \partial \Omega \) is the boundary.
\( \lambda \) is the multi-index \((k_1, \ldots, k_N)\), \( \lambda_i \geq 0 \) are integers; we sum \( \sum_{i=1}^{N} \lambda_i \) will be called the length of \( \lambda \) and denoted by \( |\lambda| \); we will also write \( D^\lambda \) instead of \( \frac{\partial^{|\lambda|}}{\partial x_1^{\lambda_1} \cdots \partial x_N^{\lambda_N}} \).

3) Definition (PDE):

Let \( a_\lambda(x) \) be real functions defined on \( \Omega \), \( |\lambda| \leq k \), where \( k \) is a natural number. Let \( \forall x \in \Omega \) at least one \( \lambda \) such that \( |\lambda| = k \) and \( a_\lambda(x) \neq 0 \). Then the relation

\[
\sum_{|\lambda| = k} a_\lambda(x) D^\lambda u = f(x)
\]

where \( f \) is an a priori given function, is called linear (partial differential) equation of the \( k \)-th order.

The part

\[
\sum_{|\lambda| = k} a_\lambda(x) D^\lambda u
\]

is called the main part of the equation (1).

4) Definition of the classical solution of the PDE:

A function \( u \) defined on \( \Omega \) is the classical solution of (1) if

i) \( u \in C^k(\Omega) \)

ii) \( u \) satisfies the relation (1) identically (i.e. in each point of \( \Omega \)).

As these requirements are strong, such a solution is also called the strong solution of (1).

5) Remark:

We know from the mathematical physics that we can have also additional requirements, usually in the form of boundary conditions, prescribed on \( \partial \Omega \).
This means that we seek for such a solution from the set of all solutions satisfying (2) (non-uniqueness), which satisfies also the additional requirements.

6) Definition: \text{Function } f: A \rightarrow B \text{, where } A, B \text{ are the Banach spaces (complete normed linear spaces), is called the Lipschitz function, if there exists a real number } \delta \text{, such that}

$$\| f(x_1) - f(y) \|_B \leq \delta \| x_1 - y \|_A \quad \forall x_1, y \in \mathcal{D}(f),$$

where \( \mathcal{D}(f) \) is the definition domain of the function \( f \).

7) Definition (domain with a Lipschitz boundary)

Let \( \Omega \subset \mathbb{R}^N \) be a bounded domain. We say that \( \Omega \) has a Lipschitz boundary \( \partial \Omega \) if there exist real numbers \( \varepsilon > 0, \delta > 0 \) such that for each \( x_0 \in \partial \Omega \), the Cartesian coordinate system can be rotated and translated by \( x_0 \) in such a way that the following statement holds:

Let

$$K_{n-1} = \left\{ (x_1, \ldots, x_{n-1}) \in \mathbb{R}^{n-1} \mid 0 < r < 1 \text{ and } j = 1, \ldots, n-1 \right\}$$

(\( K_{n-1} \) is a \((n-1)\)-dimensional open cube). Then there exists a real function \( a \) defined on \( K_{n-1} \) such that for each point \( (x_1, \ldots, x_{n-1}) \in K_{n-1} \),

$$(x_{n-1}, x_n) \in K_{n-1}, a(x_1, \ldots, x_{n-1}) \in \partial \Omega$$

and \( a \) is the Lipschitz function. Moreover, all the points \( x = (x_1, \ldots, x_{n-1}, x_n) \equiv (x', x_n) \) such that \( x' \in K_{n-1} \) and \( a(x') < x_n < a(x') + \varepsilon \) lie inside \( \Omega \) and all the points \( x = (x'_1, x_n), x' \in K_{n-1}, a(x') - \varepsilon < x_n < a(x') \) lie outside \( \bar{\Omega} \).
Explanation

Roughly speaking:
There are two requirements:

1) The boundary can be locally described as a Lipschitz function.
2) The boundary separates the interior from the exterior in such a sense that there exist two bands of content thickness $\varepsilon$: the first one lies inside $\Sigma$ and the second one lies outside $\Sigma$.

Examples: An $\mathbb{R}^3$ ball, a cube (each corner can be described by a Lipschitz function), or a parallelepiped are domains with a Lipschitz boundary.

The domains do not have Lipschitz boundaries.

Definition: (domain with a continuous boundary)
If the function $u$ from Definition 7 is continuous, we say that $\Sigma$ is the domain with a continuous boundary. Similarly for $u \in C^k(K)$.

Notation: If $\Sigma$ is a domain with a Lipschitz boundary, we write $\Sigma \in C^{0,1}$; if $\Sigma$ is a domain with continuously differentiable boundary up to order $k$, we write $\Sigma \in C^k$.

Remark (for mathematically educated students):
The coordinate systems from the definition of $u$ were introduced to obtain pieces of the boundary by means of the functions $q_i$. These pieces cover the whole $\Sigma$. As $\Sigma$ is bounded, $\partial \Sigma$ is compact and we can choose a finite number of them to cover the whole boundary.
In other words: we need only a finite number of such rotated and translated Cartesian coordinate systems to describe the whole boundary 52.

12) Theorem (Green): postponed after the definition 33 together with Friedrich's inequality and with Poincaré's inequality.

§ 2. Classification of the equations of the second order

13) Motivation: Let us assume that the coefficients of the second order equation

\[ A_{ij}(x) \frac{\partial^2 u}{\partial x_i \partial x_j} + B_i(x) \frac{\partial u}{\partial x_i} + C(x) u = F(x) \]  

are defined on a domain \( \Omega \subset \mathbb{R}^N \) and that the symmetry

\[ A_{ij}(x) = A_{ji}(x), \quad i,j = 1, 2, \ldots, N \]

holds \( \forall x \in \Omega \).

After the substitution \( \xi_k = a_{ki} x_i, \quad \xi_k \in \mathbb{R}^N \), the equation (1) transforms into the form

\[ A_{ij}(\xi) \frac{\partial^2 u}{\partial \xi_i \partial \xi_j} + \ldots \]

Let us consider the quadratic form

\[ A_{ij} \xi_i \xi_j \]

Under the substitution

\[ \xi_i = a_{ki} \xi_k, \]

the form (3) transforms in the same way as the main part of the eqn. (1).
Let us consider any fixed point \( x_0 \in \mathbb{R}^2 \) and put \( A_j = A_j(x_0) \) in the form (3). Then it is known from the theory of quadratic forms that there exists a transformation (4), which converts (3) to the form

\[
\sum_{i=1}^{m} e_i \xi_i^2, \quad m \leq N, \quad e_i = +1 \text{ or } e_i = -1. \tag{5}
\]

Application of this fact to the studied egn. (4):

after employing the transformation, which converts (3) to (5),

the main part of (4) can be written as follows

\[
A_j(x) \frac{\partial^2 u}{\partial x_i \partial y_j} = t \ldots \tag{6}
\]

where in the point \( x_0 \) we get

\[
A_{ij}(x_0) = \pm 1 \quad \text{for } i = j \leq m
\]

\[
A_{ij}(x_0) = 0 \quad \text{for } i \neq j \text{ or } i = j > m
\]

14) Definition:

We call the egn. (6) the canonical form of the egn. (4) in the point \( x_0 \).

15) The equation (6) is called

a) elliptic
b) hyperbolic
c) ultra hyperbolic
d) parabolic in a broader sense
e) parabolic

in a point \( x_0 \), it in the canonical form (6).
a) $m=N$ and all the coefficients $A_{ij}^x$ have the same sign
b) $m=N-4=-11=-11=-11$ with the exception of one coefficient have the same sign
c) $m=N$ and at least two coefficients $A_{ii}^x = 1$ and at least two coefficients $A_{ii}^x = -1$

d) $m < N$
e) $m = N-1$ and the coefficient standing at $\frac{\partial u}{\partial y}$ is 0 at x = 0.

16) Definition: The equation is elliptic (hyperbolic etc...) in a domain $\Omega$ if it is elliptic (hyperbolic etc...) in each $x \in \Omega$.

Examples $\nabla^2 u = 0$ ... elliptic

$\frac{\partial T}{\partial t} = \alpha \nabla^2 T$ ... parabolic

$\frac{\partial^2 u}{\partial t^2} = \frac{1}{c^2} \nabla^2 u$ ... hyperbolic

$\S 3$ Cauchy's problem for a vibrating string

I have chosen this problem to demonstrate the two classical approaches to PDE:

i) analytical approach

ii) the Fourier method

17) Problem (example of analytical considerations):

Let us solve the equation of amplitude or displacement of the string

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}$$

$-\infty < x < \infty$

$t \geq 0$

(velocity of wave propagation is $c$, which can be obtained by scaling the length or time)
with the initial conditions
\[
\begin{align*}
  u(0, x) &= \Phi_0(x), \\
  \frac{\partial u}{\partial t}(0, x) &= \Phi_1(x),
\end{align*}
\]

(initial amplitude or shape of the string)

(\text{initial velocity of the string})

where \( \Phi_0 \) and \( \Phi_1 \) are some "suitable" functions.

Substitution \( y = t - x \) transforms the equation to the form
\[
\frac{\partial^2 u}{\partial y \partial t} = 0
\]

Integration \( \frac{\partial u}{\partial t} = F(y) \)

Integration with variables \( u = F(y) \rightarrow u(t, x) = g(x - t) + f(t + x) \) \hspace{1cm} \text{(4)}

To satisfy the initial conditions, it must hold:
\[
\begin{align*}
  u(0, x) &= \Phi_0(x) = g(x) + f(x) \\
  \frac{\partial u}{\partial t}(0, x) &= \Phi_1(x) = g'(x) - f'(x)
\end{align*}
\]

\[
\begin{align*}
  \Phi_0(x) &= g(x) + f(x) \\
  \int_0^\infty \Phi_0(0) &\rightarrow g''(x) + \int_0^\infty f'(x) + C
\end{align*}
\]

\[
\begin{align*}
  \frac{\partial u}{\partial t}(0, x) &= g'(x) - f'(x) \\
  \int_0^\infty \Phi_1(0) &\rightarrow g''(x) - \int_0^\infty f'(x) + C
\end{align*}
\]

Integrating constant addition and subtraction

\[
\begin{align*}
  g(x) &= \frac{1}{2} \left[ \int_0^\infty \Phi_0(0) + \int_0^\infty \Phi_1(0) \right] + C \\
  f(x) &= \frac{1}{2} \left[ \int_0^\infty \Phi_0(0) + \int_0^\infty \Phi_1(0) \right] - C
\end{align*}
\]

Putting this \( \Phi_0(x) \) we obtain

\[
\begin{align*}
  u(x, t) &= \frac{1}{2} \left[ \int_0^\infty \Phi_0(x + t) + \int_0^\infty \Phi_0(x - t) \right] + \int_0^\infty \Phi_1(0) \ dx
\end{align*}
\]

\( \alpha \text{ Hentoba's formula} \)
Reflection: (example of two important tricks which will be needed e.g. in geophysics)

a) on a fixed end:

We should solve the same equation, with the same initial conditions but for \( x = 0 \) and, moreover, we should satisfy the boundary condition \( u(t,0) = 0 \), which describes the fact that the string is fixed at \( x = 0 \).

Classical trick: for \( x < 0 \) we define

\[
\begin{align*}
    \psi_0(x) &= -\psi_0(-x) \\
    \psi_1(x) &= -\psi_1(-x)
\end{align*}
\]

(odd continuation extension of initial conditions from a half-line to the whole line)

and employ the solution of the problem shaded on

\[
\begin{align*}
    \Phi(x) &= \int_{-\infty}^{x} \psi_0(x') \, dx' + \int_{0}^{x} \psi_1(x') \, dx' \\
    \text{indeed, } u(t,0) &= \frac{1}{2} \sum (\psi_0(t) + \psi_0(-t)) + \int \psi_1(t) \, dx \implies u = 0 \text{ for } t > 0 \text{ and thus the boundary condition for a fixed end is fulfilled.}
\end{align*}
\]

b) on a free end:

There is no force acting at the string and thus

\[
\frac{\partial u}{\partial x} (t,0) = 0,
\]

which can be satisfied by \( \psi \) even extension of initial conditions.
The Fourier method

Let the string is fixed in two points \( x = 0 \) and \( x = L \), i.e.
\( u(0, t) = u(L, t) = 0 \).

Let us try to find the solution in the form
\[ u(t, x) = T(t) \cdot X(x) \]

Other putting into the wave equation it holds
\[ \frac{\partial^2 X}{\partial t^2} - \frac{1}{c^2} \frac{\partial^2 X}{\partial x^2} = 0 \]

i.e.
\[ \frac{\frac{\partial^2}{\partial t^2}}{I(t)} \cdot X = \frac{\frac{\partial^2}{\partial x^2}}{X} = -\lambda \]

If \( x = 0 \) \( \Rightarrow \) \( X = a \cdot t \) and the boundary conditions yield
\( a = b = 0 \) (trivial solution)

If \( x < 0 \) \( \Rightarrow \) \( X = a \cdot t \cdot Fx + b \cdot t \cdot Fx \) \( \Rightarrow \) again \( a = b = 0 \)

If \( x > 0 \)
\[ X'' + \lambda X = 0 \]

\[ X = a \cdot \sin kx + b \cdot \cos kx \]

\( k = \sqrt{\frac{\lambda}{c^2}} \)

The boundary conditions imply \( b = 0 \)
\[ kL = \pi \Rightarrow \lambda = \frac{\pi^2 c^2}{L^2} \]

\[ X(t, x) = \sum_{n=0}^{\infty} \left( A_n \cos \frac{n \pi x}{L} + B_n \sin \frac{n \pi x}{L} \right) \]

The initial conditions:
\[ u(x, 0) = \sum_{n=0}^{\infty} \left( A_n \sin \frac{n \pi x}{L} + B_n \cos \frac{n \pi x}{L} \right) \]
\[ \left. \frac{\partial u(x, t)}{\partial t} \right|_{t=0} = \sum_{n=0}^{\infty} \left( A_n \cos \frac{n \pi x}{L} \sin \frac{n \pi}{L} x + B_n \sin \frac{n \pi x}{L} \cos \frac{n \pi}{L} x \right) \]

Here we see why we need the initial conditions (there are two integration constants \( A_n \) and \( B_n \)).
Part II: Variational (weak) methods

§ 1. Sobolev spaces

32) Definition: In $C^1(\Omega)$ we define the scalar (inner) product

$$\langle f, g \rangle = \int \left( f g + \frac{\partial f}{\partial x_1} \frac{\partial g}{\partial x_1} \right) dx,$$

$$\|f\| = \sqrt{\langle f, f \rangle}.$$

The completion of $C^1(\Omega)$ in this norm is denoted by $W^{1,2}$ and called the Sobolev space.

Remark: In functional analysis it is possible to prove that if $\Omega \in C^{0,1}$ then $W^{1,2}(\Omega)$ is the same as the space of functions $f \in L^2(\Omega)$ with derivatives

$$\frac{\partial f}{\partial x_1} \in L^2(\Omega), \quad i = 1, \ldots, n,$$

where $\int \cdots dx$ is the Lebesgue integral.

33) Theorem: The Trace Theorem.

Let $\Omega \in C^{0,1}$. Then there exists a uniquely determined linear mapping (operator),

$$T: W^{1,2}(\Omega) \rightarrow L^2(\Omega),$$

which is continuous such that

$$T(u) = u\mid_{\partial \Omega} \quad \forall u \in C^1(\Omega),$$

where

$$u\mid_{\partial \Omega}$$

denotes restriction.

Proof: (only for the unit cube $[0,1]^n$ in $\mathbb{R}^n$).

It is sufficient to show that there exists $c > 0$ such that

$$\forall u \in C^1(\Omega) \text{ it holds:}$$

$$\|T(u)\|_{L^2(\Omega)} \leq c \|u\|_{W^{1,2}(\Omega)}.$$
as a bounded operator is continuous. The required operator is then easily constructed by the continuous extension from \( C^0(S) \) to \( H^1_0(S) \). In other words, we will show that the restriction is continuous from \( C^0(S) \) to \( L^2(\mathbb{R}) \) and that \( T \) is just its continuous extension to \( W^{1,2}(S) \).

Let us write \( x = (x', x_N) \).

Then for any \( f \in C^0(S) \) we can write

\[
f(x', x_N) = f(x'_N) + \int_0^1 \frac{\partial f}{\partial x_N}(x', x_N) \, dx_N
\]

and

\[
0 \leq (x^2 - b^2)^2 \Rightarrow \ l a b \leq a^2 + b^2 \Rightarrow
\]

\[
\Rightarrow \quad (a + b)^2 \leq 2(a^2 + b^2)
\]

\[
\Rightarrow \quad f^2(x', x_N) \leq 2 \left[ f^2(x'_N) + \left( \int_0^1 \frac{\partial f}{\partial x_N}(x', x_N) \, dx_N \right)^2 \right]
\]

\[
\leq 2 \left[ f^2(x'_N) + \left( \int_0^1 \frac{\partial f}{\partial x_N}(x', x_N) \, dx_N \right)^2 \right]
\]

Now we will integrate this inequality over \( f \) from 0 to 1:

\[
f^2(x', x_N) \leq 2 \left[ \int_0^1 f^2(x'_N) \, dx'_N + \int_0^1 \left( \frac{\partial f}{\partial x_N}(x', x_N) \right)^2 \, dx_N \right]
\]

Let \( C \) be the side of the cube given by \( x_N = 1 \) =>

\[
\int_C f^2(x', x_N) \, dx' \leq 2 \left[ \int_0^1 \int_0^1 f^2(x'_N) \, dx'_N + \int_0^1 \int_0^1 \left( \frac{\partial f}{\partial x_N}(x', x_N) \right)^2 \, dx_N \right] \leq 2 \int_C \int_0^1 f^2(x'_N) \, dx'_N \, dx_N
\]
this inequality holds for any \( \delta > 0 \) sides (here \( \delta \) is the side length) so that:

\[
\|u\|_{L^2(\Omega)} \leq C \|u\|_{W^{1,2}(\Omega)} \quad \text{q.e.d.}
\]

34) **Definition:** The operator \( T \) is called the trace operator, \( \mathcal{T}(u) \) is the trace of \( u \in W^{1,2}(\Omega) \) on \( \partial \Omega \).

35) **Remark:**

i) The trace theorem demonstrates that the sentence "function \( u \) attains on \( \partial \Omega \) a certain value" can be extended from \( C^1(\Omega) \) to \( W^{1,2}(\Omega) \). If we speak about the value of any function \( u \in W^{1,2}(\Omega) \) on \( \partial \Omega \) we mean the trace of \( u \).

ii) The set of images \( \mathcal{T}(W^{1,2}(\Omega)) \subset L^2(\partial \Omega) \) but \( L^2(\partial \Omega) \setminus \mathcal{T}(W^{1,2}(\Omega)) \) need not be empty, and thus there can exist functions \( u \in L^2(\partial \Omega) \) which are not the trace of any function \( u \in W^{1,2}(\Omega) \).

36) **Theorem:** Let \( \Omega \) be a bounded domain, and let \( \sigma, \mu \in W^{1,2}(\Omega) \).

Then

\[
\int_{\Omega} \sigma \cdot \mu \, dx \pm \int_{\partial \Omega} \sigma \cdot \frac{\partial \mu}{\partial n} \, dS = \int_{\Omega} \sigma \cdot \nu \, dx \pm \int_{\partial \Omega} \sigma \cdot \frac{\partial \nu}{\partial n} \, dS,
\]

where \( n, j = 1, \ldots, N \) are the components of the unit outward normal to \( \partial \Omega \) (the Green theorem)

\[
\|\sigma - \nu\|_{L^2(\Omega)} \leq C \left( \int_{\Omega} \sigma^2 \, dx + \int_{\partial \Omega} \nu^2 \, dS \right) \quad \text{(Friedrichs' inequality)}
\]
\[ \| \nabla u \|_{L^2(\Omega)} \leq C \left( \int_\Omega \frac{\partial u}{\partial x} \, dx + \left( \int_\Omega \frac{\partial u}{\partial y} \right)^2 \right) \] (Poincare's inequality)

Remark: The Green theorem holds also for \( \nabla \cdot \mathbf{F} \in C(\Omega) \); this is the reason why I could have expressed it already after the item ii).

ii) \( \nabla \cdot \mathbf{F} \) is the most general form of the Green theorem. Other forms can be derived from this basic expression, e.g.,

\[ \int_\Omega \mathbf{F} \cdot \mathbf{u} \, dx = \int_{\partial \Omega} \mathbf{F} \cdot \mathbf{n} \, ds \quad \text{(Green-Ostrogradsky theorem)} \]

iii) The Green theorem is a generalization of integration by parts (by parts).

36) Theorem (Kellogg):

Let \( f \in C^1 \), then \( C^2 \)-embedding (i.e., the identical mapping) of

\( W^{1,2}(\Omega) \) into \( L^2(\Omega) \) is compact, i.e., a closed bounded set in \( W^{1,2}(\Omega) \) is compact in \( L^2(\Omega) \). In other words:

from any sequence, which is bounded in \( W^{1,2}(\Omega) \), we can choose a subsequence, which is convergent in \( L^2(\Omega) \).

Proof: (only for functions on interval \( [a,b] \)).

We will employ the modification of Arzel-Ascoli theorem:

Let \( M \) be a sequence of functions defined on an interval \( [a,b] \), which are

i) equally bounded (i.e., \( \exists c > 0 : \forall f \in M \forall x \in [a,b] : |f(x)| \leq c \))

ii) equally continuous (i.e., \( \forall \varepsilon > 0 \exists \delta > 0 : \forall f \in M \forall x, y \in [a,b] : |x-y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon \)).

Then there is a uniformly convergent subsequence \( \{f_n\} \) from \( M \).

- (Hence 37.10.)
The application of this theorem:

a) It is sufficient to show that if a sequence \( f \) is bounded, then the assumptions of this theorem are satisfied.

Then we will choose a uniformly convergent subsequence \( f_n \rightarrow f \) and we can it holds that

\[
\int_a^b \left| f_n(x) - f(x) \right|^2 \, dx \leq (b-a) \sup_{a \leq x \leq b} \left| f_n(x) - f(x) \right|^2
\]

If \( n \to \infty \) then \( \sup_{a \leq x \leq b} \left| f_n(x) - f(x) \right| \to 0 \) and thus

\( f_n(x) \to f(x) \) in \( L_2(a,b) \), q.e.d.

b) Proof of equal continuity.

Let us consider \( f \) functions such that \( \| f \|_{L_2[a,b]} \leq c \) is constant.

We can write

\[
\left| f(y) - f(x) \right|^2 \leq \left( \int_y^x f'(s) \, ds \right)^2 \leq (y-x) \int_y^x \left| f'(s) \right|^2 \, ds
\]

\[
\leq |y-x| \int_y^x \left| f'(s) \right|^2 \, ds \leq |y-x| \| f \|_{L_2[a,b]}^2 \leq |y-x| c^2 \Rightarrow c
\]

c) Equal boundedness:

1) By means of a contradiction, we will prove a boundedness of the set \( \{ f(a) \mid f \in \mathcal{F} \} \).

Let the set \( \{ f(a) \mid f \in \mathcal{F} \} \) is not bounded. Then we can choose a sequence \( f_n \) such that \( \| f_n(a) \| \geq n \). From the paragraph b) we know that there exists \( \varepsilon > 0 \) fixed such that \( |f_n(x) - f_n(a)| < \varepsilon \) for \( a < x < a + \varepsilon \) and all \( n \),

i.e. \( |f_n(x)| \geq n - 1 \) on \( [a, a + \varepsilon) \).
Then \( \| f \|_{W^{1,2}}^2 = \int_a^b (f'(x))^2 dx \geq \int_a^b (f(x))^2 dx \geq (u-1)^2 \delta \) for some \( \delta > 0 \).

If \( n \to \infty \Rightarrow \| f \|_{W^{1,2}} \to \infty \), which is the required contradiction. Therefore, there is a constant \( K \) such that \( f(x) < K \) a.e. in \( M \).

(b) We will prove what is required:

From (a) and equal continuity it is clear that \( |f(x)-f(y)| \leq 1 \) if \( |x-y| \leq \delta \). Hence,

\[ |f(x)| \leq K + \frac{|x-a|}{\delta} \quad \forall x \in M. \]

37) Remark:

(7) from 36 holds almost everywhere as it holds everywhere for functions from \( C^1(S) \), which is dense in \( W^{1,2} \).

Other point of view: it is clear that it is sufficient to prove the Rellich theorem for functions from \( C^1(S) \) and its validity for functions from \( W^{1,2} \) then directly follows from the definition of the Sobolev space \( W^{1,2}(S) \).

38) Notation: Let us define

\[ C^0(S) = \{ f \in \text{all bounded measurable functions on } S \}, \]

\[ C^0_0(S) = \{ f \in C^0(S); f = 0 \text{ on } \partial S \}. \]

\[ W_0^{1,2}(S) = \overline{C^0(S) \text{ w.r.t. Sobolev norm}} \text{ in } W^{1,2}(S). \]

\( W_0^{1,2}(S) \) are functions from \( W^{1,2} \) with "zero" trace.
39) Formulation of the problem: \( \xi \) is any vector from \( \mathbb{R}^n \)

Let \( a_{ij} \in L^\infty(\Omega) \) \( \forall \xi, \eta \geq C \| \xi \|^2 \) for a fixed \( C > 0 \), which does not depend on \( \xi \).

\[ b \in L^\infty(\Omega) \quad \text{and} \quad f \in L^2(\Omega) \]

\( u_0 \in W^{1^2}(\Omega) \), \( \Omega \subset \mathbb{R}^n \)

Find \( u \in W^{1^2}(\Omega) \) such that

i) \( u - u_0 \in W_0^{1^2} \) (boundary condition)

\[ \begin{align*}
\int_{\Omega} \left( a_{ij} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} + b v u \right) \mathrm{d}x = \int_{\Omega} f v \mathrm{d}x
\end{align*} \]

We will call \( W^{1^2}(\Omega) \) the "trial functions".

40) Interpretation of the problem:

The condition i) says: "\( u \) is equal to \( u_0 \) on \( \partial \Omega \)." We cannot choose \( u_0 \in L^2(\Omega) \) and to require \( u \) to have the trace equal to \( u_0 \in L^2(\Omega) \) as \( T(\mathcal{W}^{1^2}(\Omega)) \subset L^2(\Omega) \) but, in general, \( T(\mathcal{W}^{1^2}(\Omega)) \neq L^2(\Omega) \). However, if \( u_0 \in W^{1^2}(\Omega) \) then it has a trace and \( u \) should have the same trace.

The condition ii):

Since \( L^2(\Omega) \) is dense in \( W_0^{1^2}(\Omega) \), it is sufficient to test this integral identity by means of the test functions \( v \in C_0^1(\Omega) \). If the solution \( u \) is from the space
and \( a_{ij} \in C^0(S) \)
\( C^2(S) \) \( i.e. \) if it is smoother than required by the condition \( u \in W^{m,2}(S) \), we can apply to the left-hand side of the identity the Green theorem and write
\[
\int_S \left[ -\frac{\partial}{\partial x_i} \left( a_{ij} \frac{\partial u}{\partial x_j} \right) + bu - f \right] \, dx + \int_{\partial S} a_{ij} \frac{\partial u}{\partial x_j} \, n \, ds \rightarrow 0 \text{ as } \epsilon \rightarrow 0
\]
(\( \epsilon \rightarrow 0 \))

As \( a_{ij} \in C^0(S) \) and \( \frac{\partial}{\partial x_i} = 0 \), the surface integral vanishes.

Let us choose now a point \( x \in S \). We can choose any small neighborhood of \( x \) and construct such a trial function that its support is inside this neighborhood; the function \( \phi \) can be oscillating or not whatever we can imagine. This means that integral must be zero in each \( x \in S \), i.e. \( u \) satisfies the equation,
\[
\left( x, y \right) \frac{\partial}{\partial x_i} \left( a_{ij} \frac{\partial u}{\partial x_j} \right) + bu = f \text{ in } S
\]

Summary: If there exists a classical solution of our problem equation \( (x,y) \) then it corresponds to the solution of the problem 39) because I can do all the steps also in the opposite order starting with \( (x,y) \)

However there can exist a solution of the problem (we will demonstrate that there is a unique solution of the problem 39) which is not smooth enough and thus there is no solution of \( a_{ij} \).

This is the reason why the solution of the problem 39) is called "the weak solution" of the equation \( (x,y) \) and the problem 39) is called "the weak formulation of the Dirichlet boundary-value problem for the equation \( (x,y) \)."
Theorem (Lax-Milgram):

Let $H$ be a Hilbert space, $H'$ be a space of all continuous linear functionals defined on $H$. Let $A(u,v)$ be a continuous bilinear form defined on $H \times H$. Let there exist a constant $\lambda > 0$ such that

$$|A(u,v)| \leq \lambda \|u\|^2 \quad \forall u \in H \quad (\|u\|^2 = (u,u), \text{where} \ (\cdot, \cdot) \text{is the scalar (inner) product defined on } H).$$

Then $\forall v \in H$ there exists one and only one $u \in H$ such that...

Proof:

We will start from the Riesz representation theorem, which sounds:

$\forall v \in H \exists! u \in H \quad (u, v) = f(v) \quad \forall v \in H$ and $\|u\| = \|f\|.$

The proof of the Lax-Milgram theorem:

Let $u \in H$ fixed and thus $u$ generates the continuous linear functional $A(u,v)$. According to the Riesz theorem, there exists a unique $w$

$$A(u,v) = (u,v) \quad \forall v \in H.$$

Therefore we defined a linear mapping $P : H \to H^*$.

Moreover, $\|P(w)\| \geq |(w, P(w))| = |A(w,w)| \geq \lambda \|w\|^2$. (X)

This means that $P$ is injective as $P(w) = 0 \Rightarrow w = 0$.

We can thus define the inverse operator $P^{-1}$ on the set of images $P(H)$.

It follows from (X) that $\|P^{-1}(w)\| \leq \frac{1}{\lambda} \|w\|$ and thus $P$ is continuous.

By means of a contradiction we will prove that $P(H) = H$.

If $P(H) \neq H$ then $\exists z \neq 0$ such that $(z, P(z)) = 0 \quad \forall z \in H$ and thus

$$0 = |(z, P(z))| = |A(z, z)| \geq \lambda \|z\|^2 \neq 0$$

which is the contradiction. Therefore $P(H) = H$. 

And now: according to the Rice theorem
Let \( \mathbf{H} \) be such that \( \mathbf{H} (\mathbf{w}) = (0, \mathbf{w}) \) and \( \mathbf{H} \mathbf{H} = \mathbf{H} \).

As \( \mathbf{P} (\mathbf{H}) = \mathbf{H} \) and \( \mathbf{P}^T \) exists there exists a unique \( \mathbf{u} \in \mathbf{H} \) such that
\( (\mathbf{u}, \mathbf{w}) = \mathbf{A} (\mathbf{0}, \mathbf{w}) \) and, moreover, \( \mathbf{H} \mathbf{H} \mathbf{H} = \mathbf{H} \mathbf{H} \mathbf{H} \), q.e.d.

**42) Theorem:** There exists just one solution of the problem \( \mathbf{B} \).

**Proof:** Let us denote
\[
\mathbf{A} (\mathbf{w}, \mathbf{v}) = \sum_{\alpha \beta} \alpha \beta \mathbf{a}_{\alpha \beta} \mathbf{v}_{\alpha} \mathbf{v}_{\beta},
\]
\[
\mathbf{c} (\mathbf{w}) = \sum_{\alpha} \mathbf{b}_{\alpha} \mathbf{v}_{\alpha},
\]
and try to find the solution in the form \( \mathbf{v} = \mathbf{v}_0 + \mathbf{w} \), \( \mathbf{w} \in \mathbf{W}^{2,2} (\Omega) \).

It should hold
\[
\mathbf{A} (\mathbf{w}_0 + \mathbf{w}, \mathbf{v}) = \mathbf{c} (\mathbf{v}) + \mathbf{w} \in \mathbf{W}^{2,2} (\Omega), \text{ i.e.}
\]
\[
\mathbf{A} (\mathbf{w}_0, \mathbf{v}) = \mathbf{c} (\mathbf{v}) + \mathbf{A} (\mathbf{w}_0, \mathbf{v}) + \mathbf{w} \in \mathbf{W}^{2,2} (\Omega).
\]

It is clear that the linear functional on the right-hand side is continuous on \( \mathbf{W}^{2,2} (\Omega) \). If \( \mathbf{A} (\mathbf{w}, \mathbf{v}) \) satisfies the assumptions of the Lax-Milgram theorem, we get the required statement.

\[
\mathbf{A} (\mathbf{w}, \mathbf{v}) = \int_{\Omega} \sum_{\alpha \beta} \mathbf{a}_{\alpha \beta} \frac{\partial \mathbf{v}_{\alpha}}{\partial x_{\alpha}} \frac{\partial \mathbf{v}_{\beta}}{\partial x_{\beta}} \, dx \text{ owing to the ellipticity of}
\]
\[
\mathbf{a}_{\alpha \beta} \text{ and the assumption } b(\Omega).
\]

For \( \mathbf{v} \in \mathbf{W}^{2,2} (\Omega) \), the Friedrichs' inequality can be written as
\[
\| \mathbf{v} \|^2 \mathbf{W}^{2,2} \leq \int_{\Omega} \sum_{\alpha \beta} \mathbf{a}_{\alpha \beta} \frac{\partial \mathbf{v}_{\alpha}}{\partial x_{\alpha}} \frac{\partial \mathbf{v}_{\beta}}{\partial x_{\beta}} \, dx \leq 2 \int_{\Omega} \mathbf{v}^2 \, dx \leq 0.
\]

Then, since \( \mathbf{A} (\mathbf{w}, \mathbf{v}) \geq \int_{\Omega} \sum_{\alpha \beta} \mathbf{a}_{\alpha \beta} \frac{\partial \mathbf{v}_{\alpha}}{\partial x_{\alpha}} \frac{\partial \mathbf{v}_{\beta}}{\partial x_{\beta}} \, dx \geq 0 \), it is clear there exists just one solution of the
§ 3. Variational approach

43) Definition: Let \( X \) and \( Y \) be Banach spaces, let \( M \subset X \) be an open set, \( x_0 \in M \), \( F : M \to Y \) (\( F \) is a mapping).

a) Let \( h \in X \). If there exists a \( (\text{bounded}) \) function \( \delta F(x_0, \cdot) : X \to Y \) such that for \( \delta F(x_0, h) \) there exists a \( \dot{\text{limit}} \)

\[
\lim_{t \to 0} \frac{||F(x_0 + th) - F(x_0) - \delta F(x_0, h)||_Y}{t} = 0
\]

we call \( \delta F(x_0, h) \) the directional derivative of \( F \) at the point \( x_0 \) (for a fixed direction \( h \)).

b) If the directional derivative exists for all \( h \in X \) and is linear and continuous with respect to the variable \( h \), we call it the \( \text{Gâteaux differential} \) and denote it \( DF(x_0, h) \), (where \( DF(x_0, \cdot) : X \to Y \)).

c) If, moreover, \( h \) is not fixed and

\[
\lim_{||h||_X \to 0} \frac{||F(x_0 + h) - F(x_0) - DF(x_0, h)||_Y}{||h||_X} = 0
\]

\( DF(x_0, \cdot) \) is called the \( \text{Frechet (total) differential} \).

44) Example: We already introduced

\[
A(u, v) = \int_\Omega \frac{\partial^2 u}{\partial x_i^2} \frac{\partial^2 v}{\partial x_j^2} + b(x, u, v) \, dx
\]

\[
\langle v, f \rangle = \int_\Omega vf \, dx
\]
45) Definition: Let $H$ be a Hilbert space. We say that a sequence $(u_n)_{n=1}^{\infty}$ converges weakly to $u$ if $u_n \to u$ in $H'$ (if $H$ is a space of all $\mathcal{C}$-continuous linear functionals defined on $H$) 
($H'$... dual space to $H$)

Let $\Phi$ be a functional defined on $H$. We say that $\Phi$ is

i) coercive

ii) weakly lower semi-continuous

if

i) \[ \lim_{\|u\| \to \infty} \Phi(u) = \infty \]

ii) \[ \Phi(u_n) \to \Phi(u) \quad \text{as} \quad n \to \infty \]

\[ \lim_{n \to \infty} \inf \Phi(u_n) \geq \Phi(u) \]

If $D\Phi(u, h)$ exists for all $u \in H$, we say that it is

i) monotone

ii) strictly monotone

iii) strongly monotone

If $u, u', h \in H$

i) $D\Phi(u+u', h) - D\Phi(u', h) \geq 0$

ii) $D\Phi(u, u' + h) - D\Phi(u, h) > 0$

iii) $D\Phi(u, u') - D\Phi(u', h) \geq \epsilon \|u\|^{\|h\|^{2}}$

where $\epsilon > 0$ is a fixed constant

We will call $u$ a critical point of the functional $\Phi$ if

$D\Phi(u, u') = 0 \quad \forall u' \in H$

46) Definition: If, moreover, any $a_j$ in the Dirichlet problem 39) and $u_0 = 0$ (i.e., the problem is homogeneous)

we call the functional

$\phi(u) = \frac{1}{2} \sum \int a_{ij} \frac{\partial u}{\partial x_i} \frac{\partial u}{\partial x_j} + b u^2 - 2f u \, dx$

the functional of potential energy

\[ u(x) = \int _{-\infty}^{+\infty} \frac{e^{i k x}}{2 \sqrt{k^2 - \alpha^2}} \left( \sqrt{\frac{A}{\alpha}} - i \frac{B}{\sqrt{\alpha}} \right) \, \cos \omega t \, dt \]
Remark:

We have already seen that a non-homogeneous problem can be transformed to the homogeneous problem by seeking the solution in the form \( w = w_0 + w \) and considering

where \( w \) is the solution of the homogeneous problem with a new right-hand side. This is the reason why we can put \( H = W_0^2(\mathbb{R}^2) \) and study the properties of \( \phi \) on \( W_0^2(\mathbb{R}^2) \) only without any loss of generality.

However, if there is a symmetric \( a_{ij} = a_{ji} \) then

\[ D\phi(w, \nu) = A(w, \nu) - (\nu, \nu) \]

and thus the critical point of \( \phi \) is the solution of the Dirichlet problem. In other words, if the coefficients \( a_{ij} \) are symmetric then

the solution can be found by minimizing a suitable functional.

Proof of (\( \ast \)): \n
\[ D\phi(w, \nu) = \frac{\partial}{\partial t} \left[ \sum_{i,j} \left( a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + b \left( u + t \right) - 2 f(u + t) \right) \right] \]

\[ = \sum_{i,j} \left( a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + b \frac{\partial u}{\partial x_i} \frac{\partial u}{\partial x_j} + 2 b u - 2 f(u) \right) d\tau = \]

\[ = \int_{\mathbb{R}^2} \left( a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + 6 u - f(u) \right) d\tau = A(u, \nu) - (\nu, \nu) \quad q.e.d. \]
Lemma: Let a functional \( \Phi \) defined on a Hilbert space \( H \) have the Gateaux differential \( D\Phi(u, h) \), which is continuous in the variable \( u \) (\( D\Phi(\cdot, h): H \to H' \) is continuous) and which is monotone. Then \( \Phi \) is weakly lower semi-continuous.

Proof: Let us introduce \( \Phi(u) = \Phi(1/2(u + u_0)) \) where \( u_0 \to u \). Then \( \Phi(u_0) - \Phi(u) = \int_0^1 \Phi'(t(u + u_0)) dt \leq \int_0^1 D\Phi(t(u + u_0), u_0 - u) dt \)

\[ = \int_0^1 D\Phi(t(u_0 - u), u_0 - u) dt + D\Phi(u, u_0 - u) \]

\[ = \int_0^1 D\Phi(t(u_0 - u), u_0 - u) - D\Phi(u, u_0 - u) dt + D\Phi(u, u_0 - u) \geq 0 \]

\[ \geq D\Phi(u, u_0 - u) \to 0 \text{ as } u_0 \to u \text{ and } D\Phi(u, \cdot) \in H' \]

**49) Lemma:**

Let \( D\Phi \) is strongly monotone, then the critical point of \( \Phi \) is only one (if it exists).

Proof: Let \( u_1, u_2 \) be two critical points. Then

\[ 0 = D\Phi(u_1, u_2 - u_1) - D\Phi(u_2, u_1 - u_1) \geq C\|u_2 - u_1\|^2 \Rightarrow u_1 = u_2 \]

**50) Lemma:**

Let a functional \( \Phi \) defined on a Hilbert space \( H \) be coercive and weakly lower semi-continuous. Then there exists a point \( u \in H \) where \( \Phi \) attains its minimum.

Proof: Let \( R \) be a sufficiently great number and \( u = \inf_{u \in H} \Phi(u) \). Then \( \exists u_0 \text{ such that } \forall u \geq R \text{ it holds } \Phi(u) \geq u_0 \) owing to the coerciveness of \( \Phi \). If there exists a minimizing point, it must lay inside the ball of radius \( R \).
From the definition of infimum it follows that there exists a sequence \( u_n \) such that \( \phi(u_n) \to m \). The ball is weakly compact (if it were not weakly compact there would exist a continuous functional \( f \) and a sequence \( u_n \) such that it would not be possible to choose a subsequence, the images of which are convergent, i.e. \( f \) would map the ball onto an interval of infinite length, which contradicts the continuity of \( f \) and thus we can choose \( u_n \to u \).

With lower semi-continuity yields \( m = \limsup_{n \to \infty} \phi(u_n) \geq \phi(u) \geq m \), i.e. \( \phi(u) = m \). Q.E.D.

51) Theorem: There exists one and only one minimum of the functional of potential energy.

Proof: We will demonstrate that the assumptions of lemmas 48 and 49 are satisfied.

\[
\begin{align*}
D\phi(u+h, b) &- D\phi(u, b) = A(u+h, b) - (b, f) - A(u, b) + (b, f) = \\
&= A(h, b) \geq \lambda \|h\|_{W_2}^2.
\end{align*}
\]

\( D\phi(u, b) \) is continuous in both variables.

It remains to prove the convexity:

\[
\phi(u) = \frac{1}{2} A(u, u) - (u, f) \geq \frac{1}{2} \|u\|_{W_2}^2 - 1 (u, f) \geq \frac{1}{2} \|u\|_{W_2}^2 - \|u\|_{L_2}.
\]

\[
\|u\|_{L_2} \geq \frac{1}{2} \|u\|_{W_2}^2 - \|u\|_{H_2} H_{L_2} = \|u\|_{W_2} \left( \frac{1}{2} \|u\|_{W_2} - H_{L_2} \right).
\]

As \( f \) is fixed we have that \( \phi(u) \to \infty \) if \( \|u\|_{W_2} \to \infty \).

Lemma 50) thus ensures the existence of a minimum and the lemma 49) yields uniqueness.
§4. Generalized problem for an elliptic equation

62) Notation: \( \eta = \{ v \in C(\overline{\Omega}) ; \eta = 0 \text{ on } \Gamma \subset \partial \Omega \} \)

\[ V = \overline{\eta} \text{ in } W^{1,2}(\Omega), \text{ i.e. } W^{1,2}(\Omega) \subset V \subset W^{2,2}(\Omega) \]

Let \( \sigma \in L^\infty(\partial \Omega), \sigma \geq 0 \); we define a generalized bilinear form

\[ ((v, u)) = A(v, u) + \int_{\partial \Omega} \sigma v u \, dS \]

63) Formulation of the problem:

Let \( a_{ij} \in L^\infty(\Omega), \sigma \geq 0 \) \( L^2(\Omega), b > 0 \)

\( \sigma \in C(\Omega), c \geq 0 \)

\( u_0 \in W^{1,2}(\Omega) \) (Dirichlet or stable boundary condition)

\( g \in L^2(\partial \Omega) \) (Neumann or unstable boundary condition)

\( f \in L^2(\Omega) \) (right-hand side)

Find \( u \in W^{1,2}(\Omega) \) such that

i) \( u - u_0 \in V \)

ii) \( ((v, u)) = \int_\Omega \sigma v f + \int_{\partial \Omega} \sigma v g \, dS + k e V \) \( \text{ (1) } \)

54) Interpretation of the problem:

Let us apply the Green theorem to (1), which is possible if all functions are smooth:

\[ 0 = \int_\Omega \left( -\frac{\partial}{\partial x_i} (a_{ij} \frac{\partial u}{\partial y_j}) + bu - f \right) \, dx + \int_{\partial \Omega} \left( a_{ij} \frac{\partial u}{\partial y_j} m_i + \sigma v g \right) \, dS \]

\( i \leq j \leq N \) for \( i, j = 1, \ldots, N \)
As \( v \) is any function from \( V \) and \( W_0^{2,2}((D) \subseteq V \), let us consider the situation when the trial function is any function from \( W_0^{2,2} \). The same considerations as in the interpretation (2) of the Dirichlet problem lead to the conclusion that we deal with the same equation as before, i.e., that our problem is a weak formulation of the equation

\[
-\frac{\partial}{\partial x_i} (a_{ij} \frac{\partial u}{\partial x_j}) + bu = f.
\]

Now let us take into account also the trial functions \( v \in V \setminus W_0^{2,2} \).

To satisfy (2), it must hold

\[
a_{ij} \frac{\partial u}{\partial x_j}. n_i + b u = g \quad \text{on } \partial \Omega \setminus \Gamma,
\]

where the trace of \( v \) need not be zero.

This is called the Neumann condition.

If \( b = 0 \) we get

\[
a_{ij} \frac{\partial u}{\partial x_j}. n_i = g \quad \text{on } \partial \Omega \setminus \Gamma,
\]

which is called the Neumann condition.

From the requirement \( u - u_0 \in V \) it is clear that

\[
u = u_0 \quad \text{on } \Gamma.
\]
Theorem: Let us consider the problem (53). If \((N-1)\)-dimensional measure of \(P\) is positive or \(b > 0\) on a set of \(N\)-dimensional positive measure or \(b > 0\) on a set of positive \((N-1)\)-dimensional measure, then there exists one and only one solution of the problem (53).

Proof: We will prove that assumptions of the Takahashi theorem are satisfied. As the continuity of \((u, v)\) is clear, we need to prove the ellipticity of \((u, v)\), i.e.,

\[
((u, v)) \geq c \|u\|^2 \|v\|^2, \quad \forall v \in V, \quad c > 0.
\]

We will use the following contradiction:

Let the ellipticity does not hold. Then for each \(e_n \rightarrow 0\),

there exist \(v_n\) such that \((u_n, v_n) \leq \frac{1}{n} \|u\|^2 \|v_n\|^2\), i.e.,

\[
\left( \frac{\|v_n\|^2}{\|u_n\|^2} \right) \leq \frac{1}{n}.
\]

We have thus constructed the sequence \(\left( v_n \right)_{n=1}^{\infty}\) such that

\[
(C(u_n, v_n)) \leq \frac{1}{n}. \quad (5)
\]

Now \((u, v)\) is not summing index in what follows:

\[
(C(u_n', v_n')) = \int \left( a_{ij} \frac{\partial u_n'}{\partial x_i} \frac{\partial v_n'}{\partial x_j} + b u_n'^2 \right) dx + \int_{\partial \Omega} u_n'^2 \, ds \geq \\
\geq \int \left( a \frac{\partial u_n'}{\partial x_i} \frac{\partial v_n'}{\partial x_i} + b v_n'^2 \right) dx + \int_{\partial \Omega} v_n'^2 \, ds
\]
Now it is clear that $V_m' \to 0$ in $L_2(\Omega)$. Yes (c) must be satisfied, i.e. $\frac{\partial V_m}{\partial x_i} \to \text{constant in } L_2$.

d) If $b > 0$ on a set of positive $N$-dimensional measure or $a > 0$ on a set of positive $(N-1)$-dimensional measure then $V_m' \to 0$ in $L_2(\Omega)$ and thus $\|V_m'\|_{W^{1,2}} \to 0$, which is the contradiction as $\|V_m'\|_{W^{1,2}} = 1$.

b) If the trace of $V_m' = 0$ on a set of positive $(N-1)$-dimensional measure, then again $V_m' \to 0$ as $V_m' \to \text{const.} \, \text{almost everywhere}$, and we have the same contradiction as in the case d).

56) The Neumann problem:

If the assumptions of the theorem 56) are not satisfied, we have the Neumann problem:

Find $u \in W^{1,2}(\Omega)$ such that

$$\int_{\Omega} \nabla \cdot \nu \left( \frac{\partial u}{\partial x_i}, \frac{\partial u}{\partial x_j} \right) \, dx = \int_{\Omega} f \, dx + \int_{\Omega} g \, \nu \, dx \quad \forall \nu \in W^{1,2}(\Omega).$$

This means that the Neumann boundary condition must be satisfied on the whole boundary $\partial \Omega$.

Moreover, by putting $V = 1$, we see that the necessary condition to the existence of the Neumann problem is the so-called equilibrium condition:

$$\int_{\Omega} f \, dx + \int_{\partial \Omega} g \, df = 0.$$
**Theorem:** There exist solutions of the Neumann problem if and only if the equilibrium condition is satisfied. If $u_1$ and $u_2$ are two solutions, then $u_1 - u_2$ is constant almost everywhere.

**Proof:** If $u$ is a solution, the $u + k$, where $k$ is a constant, is also the solution. Any solution can thus be decomposed into the two terms:

$$u = u_0 + \frac{1}{J(u)} \int_{\Omega} u \, dx$$

where $J(u)$ is the $N$-dimensional Lebesgue measure of $\Omega$. This means that

$$\int_{\Omega} u_0 \, dx = 0.$$

Therefore, it is sufficient to prove that there exists a unique solution on the space

$$Q = \{ u \in W^{1,2}(\Omega), \int_{\Omega} u \, dx = 0 \}.$$

Let us consider first the situation where the test functions are also only from $Q$ (restriction of the problem to the subspace $C^{0}(\Omega)$), i.e., try to find $u \in Q$ such that

$$\int_{\Omega} \sum_{i,j} \frac{\partial^2 u}{\partial x_i \partial x_j} \, dx = \int_{\Omega} f \, dx + \int_{\Omega} g \, ds + \int_{\partial \Omega} \nu \cdot \vec{n} \, v \cdot d\sigma.$$
on the product $Q \times Q$ the bilinear form $(\alpha y, v_1)\times$ is elliptic since

$$((v, w)) \geq \int \left( \sum_{i=1}^{n} \frac{\partial v}{\partial x_i} \frac{\partial w}{\partial x_i} \right) dx \geq C^2 \| v \|_{W^{2,1}}^2,$$

where we have employed Poincaré's inequality (see 12).

$$U_0 U_{\infty} \leq C \left( \int \frac{\partial v}{\partial x_1} \frac{\partial w}{\partial x_1} \right) + \left( \int v w \right)^2.$$

and the fact that $v \in Q$, i.e. $\int v \partial x = 0$. Therefore, there exists a solution $v$ of the Neumann problem $\text{div} Q$.

We need, however, to satisfy the integral identity in (56) for all trial functions $v \in W^{1,2}(\Omega)$ and not only for $v \in Q$. Let us also decompose a general trial function

$$v \in W^{1,2}(\Omega)$$

into

$$v = v_0 + \frac{1}{\Omega} \int_{\partial \Omega} v_{\partial} dS \equiv v_0 + k_v.$$

As \( \frac{\partial v}{\partial x_i} = \frac{\partial v_0}{\partial x_i} \), we can write

$$\int_{\Omega} \left( \sum_{i,j} \alpha_{ij} \frac{\partial v}{\partial x_i} \frac{\partial w}{\partial x_j} \right) dx = \int_{\Omega} v_0 f dx + \int_{\Omega} v_0 g dx. \quad (1)$$

If the equilibrium condition is fulfilled, we have

$$0 = \int_{\Omega} v_0 f dx + \int_{\partial \Omega} k_v g dS \quad \text{subject to} \quad \| v \|_{W^{2,1}}^2 \leq v_{\partial} (f dx + \beta g dS). \quad (2)$$

By adding (1) and (2), we get that it is also the solution of the Neumann problem on $W^{1,2}(\Omega)$. \( \text{g.c.o.l.} \).
**Theorem:** Let $u$ be a weak solution of the problem 53.

Then $\|u\|_{W^{1,2}(\Omega)} \leq c \left( \|u_0\|_{W^{1,2}(\Omega)} + \|f\|_{L^2(\Omega)} + \|g\|_{L^2(\Omega)} \right),$ where $c > 0$ is a linear constant.

**Proof:** Let us write $u$ in the form $u = u_0 + w$, then

$$\left( (u, v) \right) = \int_{\Omega} w f dx + \int_{\partial \Omega} \varphi v ds - \left( (u_0, v) \right) \in F(v)$$

The Lax-Milgram theorem yields

$$\|u\|_{W^{1,2}(\Omega)} \leq \frac{1}{c} \|F\| = \frac{1}{c} \sup_{v \neq 0} \frac{\|F(v)\|}{\|v\|_{W^{1,2}(\Omega)}}$$

We will deal with each of the terms forming $F(v)$:

1. $\sup_{v \neq 0} \int_{\Omega} w f dx \leq \sup_{v \neq 0} \left( \int_{\Omega} f^2 dx \right)^{1/2} \sup_{v \neq 0} \left( \int_{\Omega} v^2 dx \right)^{1/2} \leq \left( \int_{\Omega} f^2 dx \right)^{1/2} = \|f\|_{L^2(\Omega)}$.

2. $\sup_{v \neq 0} \int_{\partial \Omega} \varphi v ds \leq \left( \int_{\partial \Omega} \varphi^2 ds \right)^{1/2} \sup_{v \neq 0} \left( \int_{\partial \Omega} v^2 ds \right)^{1/2} \leq \left( \int_{\partial \Omega} \varphi^2 ds \right)^{1/2} \sup_{v \neq 0} \left( \int_{\partial \Omega} v^2 ds \right)^{1/2} \leq \|\varphi\|_{L^2(\partial \Omega)}.$

3. It follows from the fact that coefficients $b_{ij} \in L^\infty(\Omega)$ $\in L^\infty(\Omega)$ and $\sigma \in L^\infty(\Omega)$ that

$$\left( (v, \mu_0) \right) \leq c_2 \|w\|_{W^{1,2}(\Omega)} \|v\|_{W^{1,2}(\Omega)}$$

and thus ($c_2 > 0$)

$$\sup_{v \neq 0} \left( (v, \mu_0) \right) \leq c_2 \|w\|_{W^{1,2}(\Omega)} \|v\|_{W^{1,2}(\Omega)}$$

**Summary:** The generalized problem is correct according to Hadamard as:

1. The solution exists,
2. The solution is unique,
3. The solution is stable.
59) Theorem: Let \( \tilde{u} \) be a solution of the Neumann problem 56), \( \tilde{u} \in \Omega \).

Then \( \| \tilde{u} \|_{L^2(\Omega)} \leq c (\| f \|_{L^2(\Sigma)} + \| g \|_{L^2(\Sigma)}) \), \( c > 0 \) is a constant.

Proof: As \( \tilde{u} \in \Omega \) and \( (u, \mu) \) satisfies the assumptions of the Lax–Milgram theorem on \( \Omega \times \Omega \) and

\[
(\tilde{u}, \mu) = \int_{\Omega} \int_{\Omega} f(x) g(y) \, d\mu(x, y),
\]

we can repeat the considerations from the items 1. and 2. in the proof of the theorem 58 to obtain what is required.
Finite element method (FEM)

§ 1. Main ideas

Let us consider the problem:

Find \( u \in V \subset W_0^{1,2}(\Omega) \subset C(\overline{\Omega}) \) such that

\[
( ( u, v ) ) = F ( v ) \quad \forall \ v \in V ,
\]

where \( ( , ) \) is bounded, elliptic bilinear form and \( F \in \left( W_0^{1,2}(\Omega) \right)^* \).

Galerkin method:

Let \( V_h \subset V \) with \( \dim V_h < \infty \); find \( u_h \in V_h \) such that

\[
( ( u_h, v_h ) ) = F ( v_h ) \quad \forall \ v_h \in V_h ,
\]

The solution \( u_h \in V_h \), which exists according to the Lax-Milgram theorem, will be called a discrete solution.

Ritz method:

Define

\[
( ( v, \nu ) ) = ( ( \nu, \nu ) ) + \nu \in V \text{ then } (v) \text{ is equivalent to minimizing the functional}
\]

\[
J ( v ) = \frac{1}{2} ( ( v, v ) ) - F ( v ) \quad \text{on } V , \quad \text{then}
\]

\[
J ( u_h ) = \inf_{v_h \in V_h} J ( v_h ) .
\]

Evidently \( J ( u ) \leq J ( u_h ) \) because \( V_h \subset V \).

Orthogonality of an error:

(1) must hold for all \( v_h \in V_h \), as \( V_h \subset V \), i.e. the solution \( u \) of the problem (1) satisfies the considered relation also for all trial functions from \( V_h \). By subtracting (2) from (1) we obtain

\[
( ( u - u_h, v_h ) ) = 0 \quad \forall \ v_h \in V_h .
\]
i.e., the error $u - u_h$ is orthogonal to the subspace $V_h$ in the sense of the norm $\|w\| = \sqrt{(w, w)}$

**Principle of a numerical approach:**

Let $(v_i^m)_{i=1}^m$ be a basis in $V_h$. We shall look for the discrete solution $u_h$ as a linear combination of the basis functions

$$ u_h = \sum_{j=1}^m c_j v_j. $$

The relation (2) is satisfied for each trial function if it is satisfied for all basis functions, i.e., if we hold that

$$ \left( \sum_{j=1}^m c_j (v_j, v_i^m) \right) = F(v_i^m) \quad \forall \ i = 1, 2, \ldots, m, $$

i.e.,

$$ \sum_{j=1}^m \left( (v_j, v_i^m) \right) c_j = F(v_i^m) \quad \forall \ i = 1, 2, \ldots, m $$(\text{Stiffness matrix} A \text{ right-hand side or load vector})

**Regularity of the stiffness matrix:**

Let $\xi \in E_m$, $\xi \neq 0$ and $(\cdot, \cdot)$ be the scalar product in $E_m$. Then

$$ (A \xi, \xi) = \sum_{ij} ((v_i, v_j)) \xi_i \xi_j = \sum_{ij} A_{ij} \xi_i \xi_j = ((\sum_{i,j} A_{ij} v_i v_j), v_i v_j) \equiv ((v, v)) \geq $n \|v\|_{W_2^m}^2 > 0 \quad \text{as} \quad \lambda > 0 \quad \text{and} \quad \xi \neq 0. $$

If there existed a non-trivial solution of the equation $A \xi = 0$, we would get $(A \xi, \xi) = 0$, which is the contradiction, i.e., the matrix $A$ is regular.

We can clearly see that the regularity of $A$ is a direct consequence of ellipticity of the bilinear form $(\cdot, \cdot)$. 
The main idea of the FEM

The FEM in its simplest setting is a Galerkin method characterized by the three basic aspects in the construction of the space $V_h$:

(i) a triangulation $\mathcal{T}_h$ is established over the set $\Omega$,
(ii) the functions $V_0 \in V_h$ are piecewise polynomials,
(iii) there exists a basis in the space $V_h$ whose functions have small supports $\Rightarrow$ $A$ is a sparse matrix and we can employ special numerical methods to deal with the corresponding system of linear equations.

Other methods: $V_0^H, \ldots$ is formed by basis functions of Fourier series (trigonometric functions, spherical harmonics), which leads to spectral methods.

Example: Prof. Havlin uses triangle finite elements in radial direction and spherical harmonics in angular direction when he solves some problems on a sphere approximating the Earth.

Triangulation $\mathcal{T}_h$ over the set $\Omega$:

We subdivide the set $\Omega$ into a finite number of subsets $K$ (called elements) in such a way that the following properties hold

(i) $\Omega = \bigcup_{K \in \mathcal{T}_h} K$,

(ii) for each $K \in \mathcal{T}_h$, the set $K$ is closed and its interior $K^0$ is non-empty,

(iii) $\forall K_1, K_2 \in \mathcal{T}_h$, $K_1 \neq K_2$, $K_1 \cap K_2 = \emptyset$,

(iv) the boundary $\partial K$ is the Lipschitz one $\forall K \in \mathcal{T}_h$. 
The discretization (triangulation) parameter $h$ is the maximum diameter of all $K \in \mathcal{T}_h$.

**Theorem:** Let $\mathcal{T}_h$ be a triangulation of $\Omega$ formed by convex elements. Let $V_h$ be a subspace of $L^2(\Omega)$ such that the space

$$P_k = \{ V_h | V_h \in V_h \}$$

consists of polynomial functions for any $K \in \mathcal{T}_h$.

Then $V_h \subset W^{1,2}(\Omega)$ if and only if $V_h \subset C(\overline{\Omega})$, i.e., a piecewise polynomial function is from $W^{1,2}(\Omega)$ if and only if it is continuous.

§2. Finite elements

The finite element is a triple $(K, P, \Sigma)$, where:

(i) $K$ is a closed subset of $\mathbb{E}^d$ with a non-empty interior and a Lipschitz boundary,

(ii) $P$ is a space of real-valued functions defined over the set $K$,

(iii) $\Sigma$ is a finite set of linearly independent linear forms $\Phi_i$, $i \leq i \leq N$ defined over the space $P$ (or over a space which contains $P$).

The set $\Sigma$ is said to be $P$-unisolvent if for any real scalars $\lambda_i$, $1 \leq i \leq N$, there exists a unique function $\phi \in P$ which satisfies

$$\Phi_i(\phi) = \lambda_i, \quad i = 1, 2, \ldots, N$$

consequently, if $\Sigma$ is $P$-unisolvent then there exist functions $p_i \in P$, $i = 1, 2, \ldots, N$, which satisfy

$$\Phi_i(p_i) = \delta_{ij}, \quad j = 1, 2, \ldots, N,$$
Where $\delta_{ij}$ is Kronecker's symbol.

The linear forms $\phi_i$, $i=1,2,\ldots,N$ are called the degrees of freedom of the finite element, and the functions $p_i$, $i=1,2,\ldots,N$, are called the basis functions of the finite element.

**Consequence of $P$-unisolvency:**

If $\Sigma$ is $P$-unisolvent, then

$$P = \sum_{i=1}^{N} \phi_i(p) p_i, \quad \forall p \in P.$$

**Proof:** It is sufficient to show that the degrees of freedom map the left-hand side as well as the right-hand side of this relation to the same $N$-dimensional vector of real numbers. The definition of $P$-unisolvency then yields that both sides are equal.

Really,

$$\sum_{i=1}^{N} \phi_i(p) p_i = \sum_{i=1}^{N} \phi_i(p) \delta_{ij} p_i = \sum_{i=1}^{N} \delta_{ij} \phi_i(p) = \phi_j(p).$$

**Examples of finite elements:**

Lagrange finite elements are those with the degrees of freedom of the form $p \mapsto p(A)$, $A \in K$, where $A$ is called the node.
a) Linear simplicial element:

The set $K$ is a $d$-simplex

$$K = \{ \text{space of linear functions } \mathbf{p} = \mathbf{p}_0 + \mathbf{p}_1 x_1 + \ldots + \mathbf{p}_d x_d ; \mathbf{p}_i \in \mathbb{R}^d \}$$

$$\dim K = d+1$$

$$\sum_i = \{ \varphi_i(\mathbf{p}) = \mathbf{p}(A_i) ; i = 1, \ldots, d+1 \}$$ where $A_i$ are the vertices of $K$

b) Quadratic simplicial element

$$K = \ldots d\text{-simplex}$$

$$K = \{ \mathbf{p} = \mathbf{p}_0 + \sum_{i=1}^{d} \mathbf{p}_i x_i + \sum_{1 \leq i < j \leq d} \mathbf{p}_{ij} x_i x_j ; \mathbf{p}_i, \mathbf{p}_{ij} \in \mathbb{R}^d \}$$

$$\sum_i = \{ p(\mathbf{A}_i) ; i = 1, \ldots, d+1 \} \mathbf{p}(\mathbf{A}_{ij}) ; i, j = 1, \ldots, d+1$$ where symbol notation of $\varphi_i(\mathbf{p}) = \mathbf{p}(\mathbf{A}_i)$ $A_i$ are the vertices and $A_{ij} = \frac{1}{2} (A_i + A_j)$ - midpoint of edges

number of vertices .... $d+1$

number of midpoints .... $\binom{d+1}{2}$
c) Bilinear and trilinear rectangular elements

\( K \quad \alpha \)-rectangle

\[ P_K = \left\{ p = \sum_{i=0}^{\alpha} \sum_{j=0}^{\alpha} \sum_{k=0}^{\alpha} a_{ijk} x_i^i x_j^j x_k^k \right\}, \quad \alpha = 2 \]

\[ P_K = \left\{ p = \sum_{i=0}^{\alpha} \sum_{j=0}^{\alpha} \sum_{k=0}^{\alpha} a_{ijk} x_i^i x_j^j x_k^k \right\}, \quad \alpha = 3 \]

\[ \sum_{P_K} = \left\{ p(A_i) \mid i = 1, 2, \ldots, 2^\alpha \right\} \]

On such a cross-section, which is parallel to an axis, each basis function is linear. 

\( p_1 \) has some positive value on the left side and is zero on the right side of the rectangle.

**Hermite finite elements:** At least one directional derivative occurs as a degree of freedom.

d) **Hermite cubic element in 2-D:**

\( \Omega \quad \text{triangle} \)
\[ P_k = \{ p = p_j + p_{i1} x_1 + p_{i2} x_2 + p_{i3} x_3^2 + p_{i4} x_1 x_2 + p_{i5} x_1 x_3 + p_{i6} x_2 x_3 + p_{i7} x_1^2 + p_{i8} x_2^2 + p_{i9} x_3^2 \} \quad p_j \in \mathbb{R}^3, \quad \dim P = 10 \]

\[ \sum_k = \{ p(A_i), \frac{\partial p}{\partial x_1}(A_i), \frac{\partial p}{\partial x_2}(A_i), p(G) \} \quad A_i, i = 1, 2, 3, \text{are vertices and } G \text{ is its centre of gravity} \]

Other types of degrees of freedom, which are used:

\[ \phi(p) = \frac{\partial p}{\partial n}(A) \quad - \text{normal derivative in a node on a side} \]

\[ \phi(p) = (D^l p)(A), \quad l \geq 1 \quad - \text{higher order derivative} \]

\[ \phi(p) = \int_K p(x) \, dx \quad - \text{integral over the element} \]

etc.
§3. Spaces of Lagrangian finite elements

Let us consider a triangulation $\mathcal{T}$ such that the sides of the elements are either parts of the boundary $\Gamma$ or are common to two elements, i.e. the situations like this

are excluded by changes of the type of the elements like this

are allowed.

Let us take into account the Lagrangian elements defined by means of the nodes $N_k$. Let $N_h = \bigcup_{k \in T_h} N_k$.

The space of finite elements:

$$X_h = \{ v_h \in C^0(\Omega) ; \quad v_h |_{P_k} \in P_k \quad \forall k \in T_h \},$$

as these elements are Lagrangian, each function $v_h \in X_h$ is uniquely determined by its value in the nodes, i.e. by the set

$$\Sigma_h = \{ v_h(A) \quad A \in N_h \},$$

which is the called by the set of degrees of freedom of the finite element space.
Remark: The basis in $X_h$ cannot be just the union of all bases of individual elements because $V_h$ must be a continuous function. For Lagrangian elements $\dim X_h$ corresponds to the number of nodes and thus the natural choice of the basis is

$$\{ v^i \in X_h : v^i(N_j) = \delta_{ij}, \; i,j = 1, 2, \ldots, \dim(X_h) \}$$

Examples:

$d = 1$

\[ \text{Graph} \]

\[ \Omega \equiv AB \]

$d = 2$

\[ \text{Grid} \]
Curved boundary

- Isoparametric (curved) elements

\( \Omega \) need not be convex and \( P_k \) may be formed, e.g. by rational functions

- Approximation of \( \Omega \) by a polygon (polyhedral) domain \( \Omega_h \subset \Omega \) and to extend finite element function from \( \Omega_h \) to the whole \( \Omega \) in an appropriate manner

- Nonconforming methods \( (V_h \not\subset V) \)

\( \Omega_h \subset \Omega \Rightarrow V_h \not\subset V \)
§4. Convergence of the Finite Element Method

Definition: Consider the problem. Find \( u \in V \) such that
\[
(u, v) = F(v) \quad \forall v \in V
\]
Consider also the set of subspaces \( \{ V_h \} \) fulfills \( V_h \subseteq V \) and
the set \( \{ u_h \} \) of the solutions of the problem
\[
(u, v_h) = F(v_h) \quad \forall v_h \in V_h
\]
such that
\[
\lim_{h \to 0} ||u - u_h||_V = 0.
\]
Then we say that the associated family of discrete problems is convergent.

Theorem (Ca's lemma): There exists a constant \( C > 0 \) independent of the subspaces \( V_h \) such that
\[
||u - u_h||_V \leq C \inf_{v_h \in V_h} ||u - v_h||_V,
\]
consequently, a sufficient condition for convergence is
that there exists a family \( \{ V_h \} \) of subspaces of the space \( V \)
such that, for each \( v \in V \)
\[
\lim_{h \to 0} \inf_{v_h \in V_h} ||v - v_h||_V = 0
\]
(i.e., the union \( \bigcup_{h > 0} V_h \) is dense in \( V \) with respect to the \( ||\cdot||_V \)-norm.

Proof: Let \( u_h \) be an arbitrary element in \( V_h \), evidently \( (u - u_h, u_h) = 0 \); the \( V \)-ellipticity and continuity of \( (\cdot, \cdot) \) yield
\[
c_2 ||u - u_h||_V^2 \leq (u - u_h, u - u_h) = ((u - u_h, u - u_h)) + ((u - u_h, u_h - v_h))
\]
\[
= (u - u_h, u - v_h) \leq c_1 ||u - u_h||_V ||u - v_h||_V
\]
\[
\text{Therefore, } ||u - u_h||_V \leq \frac{c_1}{c_2} ||u - v_h||_V \quad \forall v_h \in V_h
\]
SOLUTION OF A DISCRETE PROBLEM

5.6. APPROXIMATION OF A PARABOLIC PROBLEM

With homogeneous Dirichlet's boundary conditions.

Let us consider the time interval $(0, T)$, $V \subset W^{1,2}(\Omega)$ of functions with zero trace on $\Gamma_0 \partial \Omega$. Let $W^{1,2}(0, T), V)$ denotes the mapping $t \in (0, T) \mapsto V(t) \in V$ such that the function $t \mapsto u(t, \cdot)$ is from $W^{1,2}(0, T)$.

**Formulation of the problem**: Find $u(t, x) \in W^{1,2}(0, T), V)$ such that

$$\begin{align*}
(\frac{\partial u}{\partial t}, v) + ((u, v)) & = F(v) \quad \forall v \in V \text{ and almost all } t \in (0, T) \\
\int_{\Omega} \frac{\partial u}{\partial t} v \, dx & = \int_{\Omega} (a \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_i} + b(v)) \, dx + \int_{\Gamma_0} v \, ds + \int_{\Omega} g \, ds \\
u(0, x) & = u_0(x) \in W^{1,2}(\Omega) \cap C(\bar{\Omega})
\end{align*}$$

**Galerkin approximation**: Let $V_n \subset V$ be of finite dimensions; find $u_n(t, x) \in V_n \forall t \in (0, T)$ such that

$$\begin{align*}
(\frac{\partial u}{\partial t}, v_n) + (((u, v_n)) & = F(v_n) \quad \forall v_n \in V_n.
\end{align*}$$

Let us try to find $u_n$ in the form

$$u_n(t, x) = \sum_{i=1}^{n} u_i(t) v_i(x); \quad \{v_i\}_{i=1}^{n} \text{ is a basis in } V_n.$$

Then

$$\begin{align*}
\sum_{i=1}^{n} \frac{\partial u_i}{\partial t} (v_i, v_i) + \sum_{i=1}^{n} V_i ((v_i, v_i)) & = F(v_i) \quad j = 1, \ldots, m.
\end{align*}$$
If \( \mathbf{U} = (u_1, \ldots, u_m)^T \) is the vector of unknowns, we get
\[
\mathbf{M} \mathbf{U}' + \mathbf{A} \mathbf{U} = \mathbf{f},
\]
where
\[
A_{ij} = (u_i, u_j), \\
\mathbf{f} = (f(u_1), \ldots, f(u_m))^T
\]

**Discretization in time:**

**Implicit methods:** in each new time-level it is necessary to solve a system of linear equations.

Consider two time levels \( k \)-th and \( (k+1) \)-th with \( \Delta t = t^{k+1} - t^k \)

Crank–Nicolson scheme:
\[
\frac{\mathbf{M} \mathbf{U}^{k+1} - \mathbf{U}^k}{\Delta t} + \frac{\mathbf{A} \mathbf{U}^{k+1} + \mathbf{A} \mathbf{U}^k}{2} = \frac{\mathbf{f}^{k+1} + \mathbf{f}^k}{2}, \ 	ext{i.e.}
\]
\[
(M + \frac{\Delta t \mathbf{A}}{2}) \mathbf{U}^{k+1} = (M - \frac{\Delta t \mathbf{A}}{2}) \mathbf{U}^k + \Delta t \left( \frac{\mathbf{f}^{k+1} + \mathbf{f}^k}{2} \right)
\]
\[
\mathbf{U}^0 = (u_1(0), \ldots, u_m(0))^T
\]

**Euler method:** (fully implicit method)
\[
\frac{\mathbf{M} \mathbf{U}^{k+1} - \mathbf{U}^k}{\Delta t} + \mathbf{A} \mathbf{U}^{k+1} = \mathbf{f}^{k+1}
\]
\[
\mathbf{U}^0 = \mathbf{U}(0)
\]

**Explicit methods:**

**Euler explicit method:**
\[
\frac{\mathbf{M} \mathbf{U}^{k+1} - \mathbf{U}^k}{\Delta t} + \mathbf{A} \mathbf{U}^k = \mathbf{f}^k, \ 	ext{i.e.}
\]
\[
\mathbf{M} \mathbf{U}^{k+1} = \mathbf{M} \mathbf{U}^k + \Delta t (\mathbf{f}^k - \mathbf{A} \mathbf{U}^k)
\]

(if unstable, problems with accuracy in periodic or curved problems)
Higher-order methods

They start from the fact that

\[ U^1 = \Phi^{-1} (I - \Phi \Delta t) \]

represent the system of ordinary differential equation resolved with respect to the time-derivative; that is why we can use, in principle, any higher-order method designed for ordinary differential equations.

Example: Runge-Kutta methods, method of predictor-corrector, etc.

Problem of the explicit schemes: instability of time-integration if the time-stepping is larger than a certain criterion.

Example: Consider the equation

\[ \frac{\partial u}{\partial t} = x \frac{\partial^2 u}{\partial x^2} \quad \text{on a line} \]

let us look at the solution in a form

\[ u = e^{-\frac{x}{T}} \sin 4x, \text{ i.e.} \]

\[ T^{-1} = -x e^{T} \Rightarrow T = \frac{1}{x e^{T}} \]

minimal resolved \( t \) under the discretization with the discretization parameter \( h \) is \( t < \frac{1}{h} \), i.e. minimal characteristic time is

\[ t_{\text{min}} = \frac{b^2}{x} \]

To be able to "catch" such a \( t_{\text{min}} \), it is clear that

\[ t < t_{\text{min}} = \frac{b^2}{x} \]

Therefore, the problem is that for a fine discretization, the time-stepping must be very small.