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. Introductory notation and definitions

References (books for further reading):

K. Rektorys: Variační metody v inženýrských problémech a v problémech matematické fyziky, SNTL, Praha 1974, 199? (in Czech)
Variational Methods in Mathematics, Science and Engineering, Reidel, Dordrecht - Boston, 1979 (in English)

K. Rektorys: The Method of Discretization in Time and Partial Differential Equations, Reidel, Dordrecht - Boston - London, 1982 (esp. chapter 2).

M. Křížek, P. Neittaanmäki: Finite Element Approximation of Variational Problems and Applications, Longman, New York, etc. 1990.

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

2) Notation: $\Omega \subset E_N$ is N-dimensional domain (open connected set in N-dimensional Euclidean space E_N)

$C^k(\Omega)$ ($k = 0, 1, 2, \dots$) is the set of functions k-times continuously differentiable in Ω.

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

$C^k(\bar{\Omega})$ is the set of functions k-times continuously differentiable in $\bar{\Omega} = \Omega \cup \partial\Omega$, where $\partial\Omega$ is the boundary

$\bar{\Omega}$ - closure of Ω

α is the multiindex $(\alpha_1, \dots, \alpha_N)$, $\alpha_k \geq 0$ are integers; the sum $\sum_{i=1}^{N} \alpha_i$ will be called the length of α and denoted by $|\alpha|$; we will also write D^α instead of $\dfrac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots \partial x_N^{\alpha_N}}$.

3) Definition (PDE):

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

$$\sum_{|\alpha| \leq k} a_\alpha(x) D^\alpha u = f(x) , \qquad (1)$$

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

The part

$$\sum_{|\alpha| = k} a_\alpha(x) D^\alpha u$$

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

4) Definition of the classical solution of the PDE:

A function u defined on Ω 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 Ω).

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 (1) (non-uniqueness!), which satisfies also the additional requirements.

6) Definition: Function $f: A \mapsto B$, where A, B are the Banach spaces (complete normed linear spaces), is called the Lipschitz function, if there exists a real number δ, that

(complete: each Cauchy sequence is convergent)

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

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

7) Definition (domain with a Lipschitz boundary)

Let $\Omega \subset E_N$ be a bounded domain. We say that Ω 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 to x^0 in such a way that the following statement holds:

Put

$$K_{N-1} = \{ (x_1, \dots, x_{N-1}) \in E_{N-1};\ |x_j| < \delta \text{ for } j = 1, \dots, N-1 \}$$

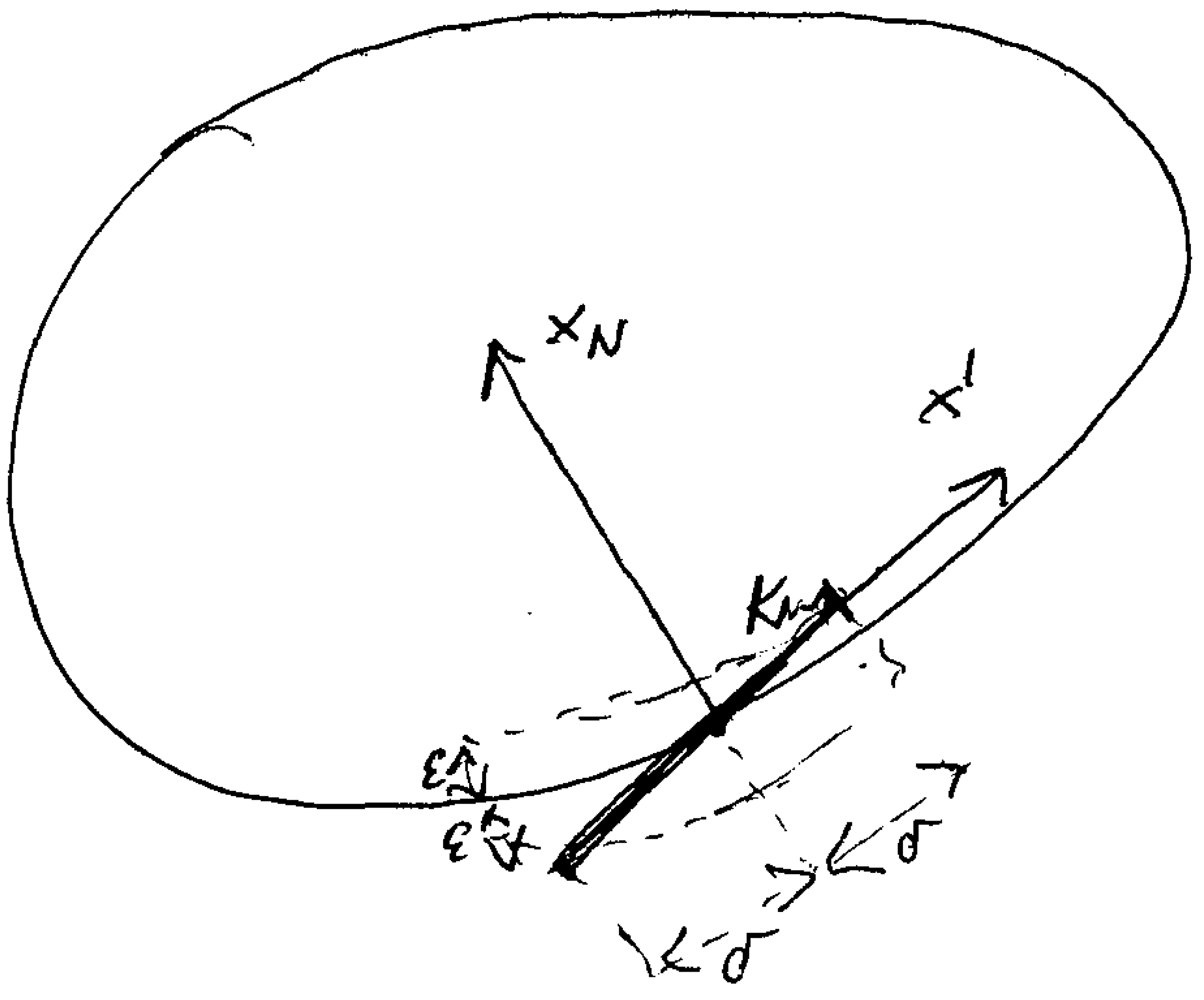
(K_{N-1} is a $(N-1)$-dimensional open cube). Then there exists a real function a defined on K_{N-1} such that for each point $(x_1, \dots, x_{N-1}) \in K_{N-1}$

$$(x_1, \dots, x_{N-1}, a(x_1, \dots, x_{N-1})) \in \partial\Omega$$

and a is the Lipschitz function. Moreover, all the points $x = (x_1, \dots, x_{N-1}, x_N) \equiv (x', x_N)$ such that $x' \in K_{N-1}$ and $a(x') < x_N < a(x') + \varepsilon$ lie inside Ω and all the points $x = (x', x_N)$, $x' \in K_{N-1}$, $a(x') - \varepsilon < x_N < a(x')$ lie outside $\overline{\Omega}$.

x' ≡ x primed

Explanation

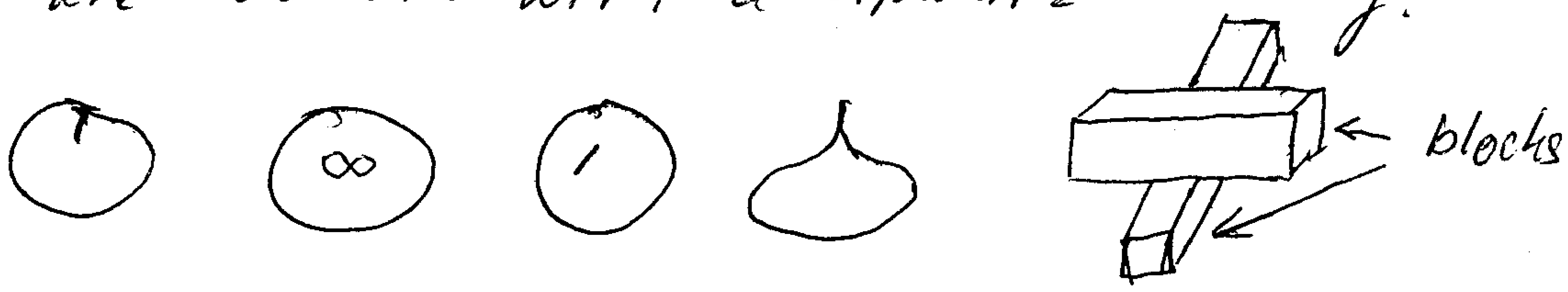


Roughly speaking:
There are two requirements
i) the boundary can be locally described as a Lipschitz function
ii) the boundary separates the interior from the exterior in such a sense that there exist two bands of constant thickness ε: the first one lies inside Ω and the second one lies outside $\bar{\Omega}$.

other examples: a parallelepiped, a square, a rectangle,

8) <u>Examples</u>: The sphere, the cube (each corner can be described by means of a Lipschitz function) or the pyramid are domains with a Lipschitz boundary.

The domains



do not have Lipschitz boundaries.

9) <u>Definition</u>: (domain with a continuous boundary)
If the function a from the definition 7) is continuous, we say that Ω is the domain with a continuous boundary. Similarly for $a \in C^k(K_{N-1})$

10) <u>Notation</u>: If Ω is a domain with a Lipschitz boundary, we write $\Omega \in C^{0,1}$; if Ω is a domain with continuously differentiable boundary up to order k, we write $\Omega \in C^k$.

11) <u>Remark</u> (for mathematically educated students):
The coordinate systems from the definition 7) were introduced to obtain pieces of the boundary by means of the functions a. These pieces cover the whole $\partial\Omega$. As Ω is bounded, $\partial\Omega$ is compact and we can thus 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 $\partial\Omega$.

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

12) $\emptyset$ empty

§2. Classification of the equations of the second order

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

$$A_{ij}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} + B_i(x)\frac{\partial u}{\partial x_i} + C(x)\,u = F(x) \qquad (1)$$

are defined on a domain $\Omega \subset E_N$ and that the symmetry

$A_{ij}(x) = A_{ji}(x)$, $i,j = 1,2,\dots,N$, holds $\forall x \in \Omega$.

After the substitution $\xi_k = a_{ki}x_i$, $a_{ki} \in E_1$, (2).

the eqn. (1) transforms into the form

$$A_{ij}\,a_{ki}\,a_{\ell j}\frac{\partial^2 u}{\partial \xi_k \partial \xi_\ell} + \dots .$$

Let us consider the quadratic form

$$A_{ij}\,y_i\,y_j \qquad (3)$$

Under the substitution

$$y_i = a_{ki}\,\xi_k \,, \qquad (4)$$

the form (3) transforms in the same way as the main part of the eqn. (1).

Let us consider any fixed point $x_0 \in \Omega$ and put $A_{ij} \equiv A_{ij}(x_0)$ 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

$$\sum_{i=1}^{m} c_i \xi_i^2 \quad , \quad m \le N \, , \; c_i = +1 \text{ or } c_i = -1 \, . \quad (5)$$

Application of this fact to the studied eqn. (1):

after employing the transformation, which converts (3) to (5), the ~~main part of~~ equation (1) can be written as follows,

$$A^*_{ij}(x) \frac{\partial^2 u}{\partial \xi_i \partial \xi_j} + \dots \qquad (6)$$

where in the point x_0 we get

$$A^*_{ij}(x_0) = \pm 1 \quad \text{for } i = j \le m$$
$$A^*_{ij}(x_0) = 0 \quad \text{for } i \ne j \text{ or } i = j > m$$

14) Definition:

We call the eqn. (6) the canonical form of the eqn. (1) in the point x_0.

15) The equation (1) is called

a) elliptic

b) hyperbolic

c) ultrahyperbolic

d) parabolic in a broader sense

e) parabolic

in a point x_0, if in the canonical form (6)

a) $m=N$ and all the coefficients A^*_{ii} have the same sign

b) $m=N$ -"- -"- -"- -"- with the exception of one coefficient have the same sign

c) $m=N$ and at least two coefficients $A^*_{ii}=1$ and at least two coefficients $A^*_{ii}=-1$.

d) $m<N$

e) $m=N-1$ and the coefficient standing at $\frac{\partial u}{\partial \xi_N} \neq 0$ at x_0.

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

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

$$\frac{\partial T}{\partial t} = \varkappa \nabla^2 T \quad \text{... parabolic}$$

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

§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

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

amplitude or displacement of the string

$-\infty < x < \infty$

$t \geq 0$

(the velocity of wave propagation is 1, which can be obtained by scaling the length or time) with the initial conditions

$$u(0,x) = \varphi_0(x), \qquad \text{(initial amplitude or shape of the string)}$$

$$\frac{\partial u}{\partial t}(0,x) = \varphi_1(x), \qquad \text{(initial velocity of the string)}$$

where φ_0 and φ_1 are some "suitable" functions.

Substitution $\xi = x+t$, $\eta = x-t$ transforms the equation to the form

$$\frac{\partial^2 u}{\partial \xi \partial \eta} = 0$$

$$\begin{cases} \dfrac{\partial}{\partial t} = \dfrac{\partial}{\partial \xi} - \dfrac{\partial}{\partial \eta} \Rightarrow \dfrac{\partial^2}{\partial t^2} = \dfrac{\partial^2}{\partial \xi^2} - 2\dfrac{\partial^2}{\partial \xi \partial \eta} + \dfrac{\partial^2}{\partial \eta^2} \\ \dfrac{\partial}{\partial x} = \dfrac{\partial}{\partial \xi} + \dfrac{\partial}{\partial \eta} \Rightarrow \dfrac{\partial^2}{\partial x^2} = \dfrac{\partial^2}{\partial \xi^2} + 2\dfrac{\partial^2}{\partial \xi \partial \eta} + \dfrac{\partial^2}{\partial \eta^2} \end{cases}$$

integration $\Downarrow$

$$\frac{\partial u}{\partial \eta} = F(\eta)$$

integration $\Downarrow$ new variables, old variables

$$u = f(\eta) + g(\xi) \iff u(t,x) = g(x+t) + f(x-t) \qquad (*)$$

To satisfy the initial conditions, it must hold:

$$u(0,x) = \varphi_0(x) = g(x) + f(x)$$

$$\frac{\partial u}{\partial x}(0,x) = \varphi_1(x) = g'(x) - f'(x)$$

(g and f are functions of only one variable!)

$$\varphi_0(x) = g(x) + f(x)$$

$$c + \int_0^x \varphi_1(s) = g(x) - f(x)$$

integration constant

addition and subtraction

$$g(x) = \frac{1}{2}\left[\varphi_0(x) + \int_0^x \varphi_1(s)\,ds + C\right]$$

$$f(x) = \frac{1}{2}\left[\varphi_0(x) - \int_0^x \varphi_1(s)\,ds \mp C\right]$$

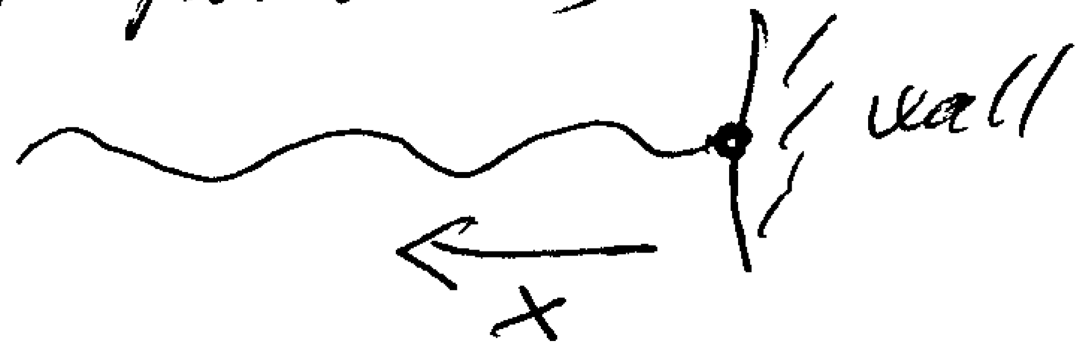
putting this to $(*)$ we obtain

$$u(t,x) = \frac{1}{2}\left[\varphi_0(x+t) + \varphi_0(x-t) + \int_{x-t}^{x+t} \varphi_1(s)\,ds\right]$$

$\boxed{\text{d'Alembert's formula}}$

Reflection: (example of two important tricks which will be needed e.g. in geothermics)

a) on a fixed end:



We should solve the same equation with the same initial conditions but for $x \geq 0$ 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

$$\varphi_0(x) = -\varphi_0(-x)$$
$$\varphi_1(x) = -\varphi_1(-x)$$

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

and employ the solution of the problem studied on $-\infty < x < \infty$.

Indeed, $u(t,0) = \frac{1}{2}\left[\varphi_0(t) + \varphi_0(-t) + \int_{-t}^{t} \varphi_1(\tau)\,d\tau\right] = 0$ and thus the boundary condition is fulfilled.

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 the even extension of initial conditions.

The Fourier method

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

Let us try to find the solution in the form

$$u(t,x)=T(t)\,X(x),$$

after putting into the wave equation it holds

$$T''X-TX''=0, \quad \text{i.e.}$$

$$\frac{T''(t)}{T(t)}=\frac{X''(x)}{X(x)}=-\lambda \qquad \lambda\in\mathbb{R}_1 \quad \text{(constant)}$$

If $\lambda=0 \Rightarrow X=ax+b$ and the boundary conditions yield $a=b=0$ (trivial solution)

If $\lambda<0 \Rightarrow X=a\,\mathrm{ch}\sqrt{-\lambda}\,x+b\,\mathrm{sh}\sqrt{-\lambda}\,x \Rightarrow$ again $a=b=0$

If $\lambda>0$

$$X''+\lambda X=0$$

$$X=a\sin kx+b\cos kx \qquad k=\sqrt{\lambda}$$

The boundary conditions imply $b=0$

$$kl=n\pi \Rightarrow \lambda=\frac{n^2\pi^2}{l^2}$$

$$\Downarrow$$

$$u(t,x)=\sum_{n=1}^{\infty}\sin\frac{n\pi}{l}x\left(A_n\cos\frac{n\pi}{l}t+B_n\sin\frac{n\pi}{l}t\right)$$

The initial conditions:

$$\varphi_0(x)=\sum_{n=1}^{\infty}A_n\sin\frac{n\pi}{l}x \qquad\qquad A_n=\frac{2}{l}\int_0^l\varphi_0(x)\sin\frac{n\pi}{l}x\,dx$$

$$\Rightarrow$$

$$\varphi_1(x)=\sum_{n=1}^{\infty}B_n\frac{n\pi}{l}\sin\frac{n\pi}{l}x \qquad\qquad B_n=\frac{2}{n\pi}\int_0^l\varphi_1(x)\sin\frac{n\pi}{l}x\,dx$$

Here we see why we need two initial conditions (there are two integration constants A_n and B_n).

Part II: Variational (weak) methods

§1. Sobolev spaces

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

$$(f,g) = \int_\Omega \left(fg + \frac{\partial f}{\partial x_i}\frac{\partial g}{\partial x_i}\right)dx, \qquad \|f\| = \sqrt{(f,f)}.$$

The completion of $C^1(\overline{\Omega})$ 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 $\frac{\partial f}{\partial x_i} \in L_2(\Omega)$, $i = 1, 2, \dots, N$. $\int_\Omega \dots dx$ is the Lebesgue integral.

33) ~~Theorem~~ The trace theorem:

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

$$T: W^{1,2}(\Omega) \to L_2(\partial\Omega),$$

which is continuous such that

$$T(u) = u|_{\partial\Omega} \quad \forall u \in C^1(\overline{\Omega}),$$ where

$u|_{\partial\Omega}$ denotes restriction.

Proof: (only for the unit cube $(0,1)$ in E_N):

It is sufficient to show that there exists $c > 0$ such that $\forall u \in C^1(\overline{\Omega})$ it holds:

$$\|T(u)\|_{L_2(\partial\Omega)} \le c\, \|u\|_{W^{1,2}(\Omega)}$$

as ~~the~~ a bounded operator is continuous. The required operator is then easily constructed by the continuous extension from $C^1(\overline{\Omega})$ to $W^{1,2}(\Omega)$. In other words, we will show that the restriction is continuous from $C^1(\overline{\Omega})$ to $L_2(\partial\Omega)$ and then T is just its continuous extension to $W^{1,2}(\overline{\Omega})$.

Let us denote
$$x \equiv (x', x_N).$$

Then for any $f \in C^1(\overline{\Omega})$ we can write
$$f(x', 1) = f(x', \xi) + \int_\xi^1 \frac{\partial f}{\partial x_N}(x', \eta)\, d\eta$$

As $0 \le (a-b)^2 \Leftrightarrow 2ab \le a^2 + b^2 \Rightarrow$

$$\Rightarrow \quad (a+b)^2 \le 2(a^2+b^2) \quad \Rightarrow$$

$$\Rightarrow \quad f^2(x', 1) \le 2\left[f^2(x', \xi) + \left(\int_\xi^1 \frac{\partial f}{\partial x_N}(x', \eta)\, d\eta\right)^2\right] \le$$

$$\le 2\left[f^2(x', \xi) + \int_\xi^1 \left(\frac{\partial f}{\partial x_N}\right)^2 (x', \eta)\, d\eta\right] \le$$

$$\le 2\left[f^2(x', \xi) + \int_0^1 \left(\frac{\partial f}{\partial x_N}\right)^2 (x', \eta)\, d\eta\right]$$

Now we will integrate this inequality over ξ from 0 to 1:

$$f^2(x', 1) \le 2\left[\int_0^1 f^2(x', \xi)\, d\xi + \int_0^1 \left(\frac{\partial f}{\partial x_N}\right)^2 (x', \eta)\, d\eta\right]$$

Let C be the side of the cube given by $x_N = 1 \Rightarrow$

$$\Rightarrow \quad \int_C f^2(x', 1)\, dx' \le 2\left[\int_0^1 f^2(x', \xi)\, d\xi + \int_0^1 \left(\frac{\partial f}{\partial x_N}\right)^2 (x', \eta)\, d\eta\right] \le 2\|f\|^2_{W^{1,2}(\Omega)}$$

this inequality holds for any ~~...~~ sides (there are $2n$ sides) $\Rightarrow$

$$\Rightarrow \quad \|f\|^2_{L_2(\partial\Omega)} \leq 4n \|f\|^2_{W^{1,2}(\Omega)} \qquad \text{q.e.d.}$$

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

35) Remark :

i) The trace theorem demonstrates that the sentence "A function u attains on $\partial\Omega$ a certain value" can be extended from $C^1(\bar{\Omega})$ 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 $T(W^{1,2}(\Omega)) \subset L_2(\partial\Omega)$ but $L_2(\partial\Omega) \setminus T(W^{1,2}(\Omega))$ need not be empty and thus there can exist functions from $L_2(\partial\Omega)$ which are not ~~the~~ a trace of any function from $W^{1,2}(\Omega)$.

12) Theorem :

Let $\Omega \in \mathcal{C}^{0,1}$ be a bounded domain and let $v, w \in W^{1,2}(\Omega)$.

Then

$$\int_\Omega \frac{\partial v}{\partial x_j} w \, dx \cancel{+ \int_\Omega v \frac{\partial w}{\partial x_j} dx} = \int_{\partial\Omega} v w \, n_j \, dS - \int_\Omega v \frac{\partial w}{\partial x_j} dx,$$

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

$$\|v\|^2_{W^{1,2}(\Omega)} \leq c \left(\int_\Omega \frac{\partial v}{\partial x_j} \frac{\partial v}{\partial x_j} dx + \int_{\partial\Omega} v^2 dS \right) \quad \text{(Friedrichs' inequality)}$$

$$\|v\|^2_{W^{1,2}(\Omega)} \leq C\left(\int_\Omega \frac{\partial v}{\partial x_j}\frac{\partial v}{\partial x_j}\,dx + \left(\int_\Omega v\,dx\right)^2\right) \quad \text{(Poincare's inequality)}$$

<u>Remark:</u> i) The Green theorem holds also for $v, w \in C^1(\bar{\Omega})$; this is the reason why I could have expressed it ~~even~~ already after the item 11).

ii) ~~other forms~~ This is the most general form of the Green theorem. Other forms can be derived from this basic expression, e.g. by summation over j and by putting $w=1$ we obtain

[illegible] $\int_\Omega \nabla \cdot \vec{u}\,dS = \int_{\partial\Omega}$ [illegible] $n_j\,dS$ (Gauss - Ostrogradsky theorem)

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

36) <u>Theorem (Rellich):</u>

Let $\Omega \in C^{0,1}$, then embedding (vnoření) (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 $L^2(\Omega)$.

<u>Proof:</u> (only for functions on interval $\langle a,b\rangle$).

We will employ the modification of Arzel-Ascoli theorem:

Let M be a sequence of functions defined on an interval $\langle a,b\rangle$, which are

i) equally bounded (i.e., $\exists c>0\ \forall f\in M\ \forall x\in\langle a,b\rangle;\ |f(x)|\leq c$)

ii) equally continuous (i.e., $\forall \varepsilon>0\ \exists\delta>0\ \forall f\in M\ \forall x\in\langle a,b\rangle;$ $|x-y|<\delta \Rightarrow |f(x)-f(y)|<\varepsilon$).

Then there is a uniformly con~~tinuous~~vergent subsequence ~~of~~ from M.

$\Rightarrow$ Konec 27.10.!

The application of this theorem:

a) It is sufficient to show that ~~kno~~ if a sequence M is bounded, then the assumptions of this theorem are satisfied.

Then we will choose a uniformy convergent subsequence $f_n \rightrightarrows f$ and ~~we can~~ it holds that

$$\int_a^b |f_n(x)-f(x)|^2 dx \leq (b-a) \sup_{a\leq x\leq b} |f_n(x)-f(x)|^2$$

If $n\to\infty$ then $\sup\limits_{a\leq x\leq b} |f_n(x)-f(x)| \to 0$ and thus

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

b) Proof of equal continuity.

Let us consider a set M of functions f such that $\|f\|_{W^{1,2}} \leq c$ $c\ldots$ constant

We can write

$$(f(y)-f(x))^2 \overset{*}{=} \Big(\int_x^y f'(\xi)\,d\xi\Big)^2 \leq (y-x)\int_x^y (f'(\xi))^2 d\xi \leq$$

* see Remark 37)

$$\leq |y-x| \int_a^b (f'(\xi))^2 d\xi \leq |y-x|\, \|f\|^2_{W^{1,2}} \leq |y-x|\, c^2 \ldots \text{O.K.}$$

c) Equal boundness:

1) By means of a contradiction we will prove a boundness ~~of the in the point a~~ of the set $\{f(a);\ f\in M\}$:

Let ~~functions f are~~ the set $\{f(a),\ f\in M\}$ 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 $\delta>0$ fixed such that $|f_n(x)-f_n(a)|<1$ for $a<x<a+\delta$ and all n,

i.e. $|f_n(x)| \geq n-1$ on $\langle a, a+\delta)$.

Then $\|f_n\|^2_{W^{1,2}} \geq \int_a^b (f_n(x))^2 dx \geq \int_a^{a+\delta} (f_n(x))^2 dx \geq (n-1)^2 \delta$

If $n \to \infty \Rightarrow \|f_n\|^2_{W^{1,2}} \to \infty$, which is the required contradiction. Therefore, there is a constant K ~~an~~ such that $f(a) \leq K$ $\forall f \in M$.

β) We will prove what is required:

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

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

37) Remark:

(*) from 36) holds almost everywhere as it holds everywhere for functions from $C^1(\overline{\Omega})$, which is the dense set in $W^{1,2}$.

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

§2. The Dirichlet problem for an elliptic equation

38) Notation: Let us denote

$L^\infty(\Omega)$ ... all bounded measurable functions defined on Ω

$C_0^1(\overline{\Omega}) = \{v \in C^1(\overline{\Omega});\ v = 0 \text{ on } \partial\Omega\}$

$W_0^{1,2}(\Omega) = \overline{C_0^1(\Omega)}$ (closure with respect to the Sobolev norm; if $\Omega \in C^{0,1}$ then $W_0^{1,2}(\Omega)$ are functions from $W^{1,2}$ with "zero" trace.

39) Formulation of the problem: ξ is any vector from E_N

Let $a_{ij} \in L^\infty(\Omega)$; $a_{ij}^{(x)} \xi_i \xi_j \geq c|\xi|^2$ for a fixed $c>0$, which does not depend on x

ellipticity $\quad$ $|\,|$ is the Euclidean norm in E_N

$b \in L^\infty(\Omega)$; $b \geq 0$

$f \in L^2(\Omega)$

$u_0 \in W^{1,2}(\Omega)$, $\Omega \in C^{0,1}$

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

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

ii) $\forall v \in W_0^{1,2}(\Omega)$ it holds

$$\int_\Omega \left(a_{ij} \frac{\partial v}{\partial x_i} \frac{\partial u}{\partial x_j} + b v u\right) dx = \int_\Omega v f \, dx$$

We will call the functions v "the trial or test functions".

40) Interpretation of the problem:

The condition i) says: "u is equal to u_0 on $\partial\Omega$". We cannot choose $u_0 \in L_2(\partial\Omega)$ and to require u to have the trace equal to $u_0 \in L_2(\partial\Omega)$ as $T(W^{1,2}(\Omega)) \subset L_2(\partial\Omega)$ but, in general, the trace operator

$T(W^{1,2}(\Omega)) \neq L_2(\partial\Omega)$. 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 $C_0^1(\overline{\Omega})$ is dense in $W_0^{1,2}(\Omega)$ it is sufficient to test this integral identity by means of the test functions $v \in C_0^1(\overline{\Omega})$. If the solution u is from the space and $a_{ij} \in C^1(\overline{\Omega})$

$C^2(\overline{\Omega})$, i.e., if it is smoother than required by the condition $u \in W^{1,2}(\Omega)$, we can apply to the left-hand side of the identity the Green theorem and write

$$(*) \qquad \int_{\Omega} \left[-\frac{\partial}{\partial x_i}\left(a_{ij}\frac{\partial u}{\partial x_j}\right) + bu - f\right] v\,dx + \int_{\partial\Omega} a_{ij}\frac{\partial u}{\partial x_j} v\, n_i\, dS = 0 \quad \forall v \in C^1(\overline{\Omega})$$

As $v \in C^1(\overline{\Omega})$ and $v|_{\partial\Omega} = 0$, the surface integral vanishes.

Let us choose now a point $x \in \Omega$. We can choose any small neighbourhood of x and construct such a trial function that its support is inside this neighbourhood, the function v can be oscillating or not whatever we can imagine. This means that integrand must be zero in each $x \in \Omega$, i.e. u satisfies the equation

$$(**) \qquad -\frac{\partial}{\partial x_i}\left(a_{ij}\frac{\partial u}{\partial x_j}\right) + bu = f \quad \text{in } \Omega$$

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

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 classical solution of (**).

This is the reason why the solution of the problem 39) is called "the weak solution" of the equation (**) and the problem 39) is called "the weak formulation of the Dirichlet boundary-value problem for the equation (**).

41) Theorem (Lax-Milgram):

Let H be a Hilbert space, H' be a space of all continuous linear functionals defined on H. Le $A(u,v)$ be a continuous bilinear form defined on $H \times H$. Let there exists a constant $\alpha > 0$ such that $|A(u,u)| \overset{\text{"ellipticity"}}{\geq} \alpha \|u\|^2 \;\; \forall u \in H$ ($\|u\|^2 = (u,u)$, where $(\cdot,\cdot)$ is the scalar (inner) product defined on H).

Then $\forall f \in H' \; \exists! \, u \in H; \; A(v,u) = f(v) \;\; \forall v \in H$ and $\|u\| \leq \frac{1}{\alpha} \|f\|$.

(Then for all $f \in H'$ there exists one and only one $u \in H$ such that ...)

Proof:

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

$$\forall f \in H' \; \exists! \, w \in H; \; (v,w) = f(v) \;\; \forall v \in H \text{ and } \|w\| = \|f\| .$$

The proof of the Lax-Milgram theorem:

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

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

Therefore we defined a linear mapping $P: u \mapsto w$

Moreover, $\|P(u)\| \, \|u\| \overset{\text{the Schwarz inequality}}{\geq} |(u, P(u))| = |A(u,u)| \overset{\text{ellipticity}}{\geq} \alpha \|u\|^2$. (*)

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

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

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

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

H is closed $P^{-1}(P(H)) = H$ and P^{-1} is continuous, therefore $P(H)$ is closed

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

$0 = |(z, P(z))| = |A(z,z)| \geq \alpha \|z\|^2 \neq 0$ which is the contradiction

therefore $P(H) = H$.

And now:

Let $f \in H'$ then according to the Riesz theorem $\exists!\ w \in H$ such that $f(v) = (v, w)$ and $\|f\| = \|w\|$.

As $P(H) = H$ and P^{-1} exists there exists a unique $u \in H$ such that $(v, w) = A(v, u)$ and, moreover, $\|w\| \geq \alpha \|u\|$ q.e.d.

42) Theorem: There exists just one solution of the problem 39.

Proof: Let us denote

$$A(u,v) = \int_{\Omega} \left(a_{ij} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} + bvu\right) dx$$

$$l(v) = \int_{\Omega} f v \, dx$$

and try to find the solution in the form $u = u_0 + w$, $w \in W_0^{1,2}(\Omega)$. It should hold

$$A(u_0 + w, v) = l(v) \quad \forall v \in W_0^{1,2}(\Omega), \text{ i.e.}$$

$$A(w, v) = l(v) - A(u_0, v) \quad \forall v \in W_0^{1,2}(\Omega).$$

It is clear that the linear functional on the right-hand side is continuous on $W_0^{1,2}(\Omega)$. If $A(w,v)$ satisfies on $W_0^{1,2}(\Omega)$ the assumptions of the Lax-Milgram theorem, we get the required statement.

$$A(v,v) \geq \alpha \int_{\Omega} \frac{\partial v}{\partial x_i} \frac{\partial v}{\partial x_i} dx$$ owing to the ellipticity of

a_{ij} and the assumption $b \geq 0$. For $v \in W_0^{1,2}(\Omega)$, Friedrichs' inequality can be written as

$$\|v\|^2_{W^{1,2}} \leq c \int_{\Omega} \frac{\partial v}{\partial x_j} \frac{\partial v}{\partial x_j} \quad \text{as} \int_{\partial\Omega} v^2 dS = 0.$$

Therefore $A(v,v) \geq \frac{\alpha}{c} \|v\|_{W^{1,2}} \quad \forall v \in W_0^{1,2}(\Omega)$. As the continuity of A is clear, there exists just one solution of the

studied Dirichlet problem.

§3. Variational approach

43) <u>Definition:</u> Let X, 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 ~~...~~ function ~~...~~ $\delta F(x_0, \cdot): X \to Y$ such that for $\delta F(x_0, h)$ there exists a ~~...~~ limit

$$\lim_{t\to 0} \left\| \frac{F(x_0+th)-F(x_0)}{t} - \delta F(x_0,h)\right\|_Y = 0$$ 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 Gâteaux differential and denote it

$$DF(x_0, h), \quad (\text{where } DF(x_0, \cdot): X \to Y).$$

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 Fréchet (total) differential.

44) <u>Notation:</u> We already introduced

$$A(v,u) = \int_\Omega \left(a_{ij} \frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_j} + b v u\right) dx$$

$$(v, f) = \int_\Omega v f \, dx$$

45) <u>Definition:</u> Let H be a Hilbert space. We say that a sequence $\{u_n\}_{n=1}^{\infty}$ converges weakly to u; $u_n \in H, u \in H$, if

$F(u_n) \to F(u) \quad \forall F \in H'$ (H' is a space of all ~~~~ continuous linear functionals defined on H)
(H'... dual space to H)

Let ϕ be a functional defined on H. We say that ϕ is
i) coercive
ii) weakly lower semi-continuous

if ~~...~~

i) $\lim\limits_{\|u\| \to \infty} \phi(u) = \infty$

ii) $u_n \rightharpoonup u \Rightarrow \liminf \phi(u_n) \geq \phi(u)$ ... $\lim\limits_{n\to\infty} (\inf\limits_{k \geq n} \phi(u_k))$

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 $\forall u, h \in H$
i) $D\phi(u+h,h) - D\phi(u,h) \geq 0$
ii) $D\phi(u+h,h) - D\phi(u,h) > 0$
iii) $D\phi(u+h,h) - D\phi(u,h) \geq c\|h\|^2$
where $c > 0$ is a fixed constant

Remark:
We could also write
$D\phi(u_1, u_2 - u_1) - D\phi(u_2, u_2$

We will call u a critical point of the functional ϕ if

$$D\phi(u,v) = 0 \quad \forall v \in H$$

46) <u>Definition:</u> If, moreover, $a_{ij} = a_{ji}$ in the ~~studied~~ Dirichlet problem 39) and $u_0 = 0$ (i.e., the problem is homogeneous as to the boundary condition) we call the functional

$$\phi(u) = \frac{1}{2}\int_{\Omega} \left(a_{ij}\frac{\partial u}{\partial x_i}\frac{\partial u}{\partial x_j} + bu^2 - 2fu\right)dx$$ the functional of potential energy.

47) Remark:

i) We have already seen that a non-homogeneous problem can be ~~...~~ converted 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 $H = W_0^{1,2}(\Omega)$ and study the properties of ϕ on $W_0^{1,2}(\Omega)$ only without any loss of generality.

ii) However, if there is a symmetry $a_{ij} = a_{ji}$ then $D\phi(u,v) \overset{(*)}{=} A(u,v) - (v,f)$ (* see the proof below) and thus the critical point of ϕ is the solution of the Dirichlet problem. In other words, <u>if the coefficients a_{ij} are symmetric then the solution can be found by minimizing a suitable functional.</u>

<u>Proof of (*):</u>

$$D\phi(u,v) = \frac{d}{dt}\left[\int_\Omega \left(a_{ij} \cdot \frac{\partial(u+tv)}{\partial x_i} \frac{\partial(u+tv)}{\partial x_j} + b(u+tv)^2 - 2f(u+tv)\right)dx\right]\Bigg|_t$$

$$= \frac{1}{2}\int_\Omega \left(a_{ij} \frac{\partial v}{\partial x_i}\frac{\partial u}{\partial x_j} + a_{ij} \frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_j} + 2buv - 2fv\right)dx =$$

$$= \int_\Omega \left(a_{ij} \frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_j} + buv - fv\right)dx \equiv A(u,v) - (v,f) \qquad \text{q.e.d.}$$

48) Lemma: Let a functional ϕ defined on a Hilbert space H has the Gâteaux differential $D\phi(u,h)$, which is continuous also in the variable u ($D\phi(\cdot,h): H \to H'$ is continuous) and which is, moreover, monotone. Then ϕ is weakly lower semi-continuous.

Proof: Let us introduce $\varphi(t) = \phi(u + t(u_n - u))$ where $u_n \rightharpoonup u$.

here we need the continuity of $D\phi(\cdot,h)$

Then $\phi(u_n) - \phi(u) = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)\,dt = \int_0^1 D\phi(u + t(u_n - u), u_n - u)\,dt$

$= \int_0^1 [D\phi(u + t(u_n - u), u_n - u) - D\phi(u, u_n - u)]\,dt + D\phi(u, u_n - u) =$

here we need the monotonicity

$= \int_0^1 \frac{1}{t}[D\phi(u + t(u_n - u), t(u_n - u)) - D\phi(u, t(u_n - u))]\,dt + D\phi(u, u_n - u) \geq$

$\geq D\phi(u, u_n - u) \to 0$ as $u_n \rightharpoonup u$ and $D\phi(u,\cdot) \in H'$.

49) Lemma:

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

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

$$0 = D\phi(u_2, u_2 - u_1) - D\phi(u_1, u_2 - u_1) \geq C\|u_2 - u_1\|^2 \Rightarrow u_1 = u_2.$$

50) Lemma:

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

Proof: Let R be a sufficiently great number and $m = \inf_{u \in H} \phi(u)$. Then $\forall u \in H$ such that $\|u\| > R$ it holds $\phi(u) > m$ owing to the coerciveness of ϕ. If there exists a minimizing point, it must lay inside the ball of radius R.

13

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 x_n; $\|x_n\| \in R$ 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 $u_{n_k} \to u$.

Weak lower semi-continuity yields $m = \lim_{k\to\infty} \phi(u_{n_k}) \geq \phi(u) \geq m$, i.e. $\phi(u) = m$ q.e.d.

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

<u>Proof:</u> We will demonstrate that the assumptions of lemmas 48) and 49) are satisfied.

$$D\phi(u+h,h) - D\phi(u,h) = A(u+h,h) - (h,f) - A(u,h) + (h,f) =$$

$$= A(h,h) \geq \alpha \|h\|^2_{W^{1,2}}.$$

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

It remains to prove the coerciveness

$$\phi(u) = \frac{1}{2}A(u,u) - (u,f) \geq \frac{\alpha}{2}\|u\|^2_{W^{1,2}} - |(u,f)| \geq \frac{\alpha}{2}\|u\|^2_{W^{1,2}} - \|u\|_{L^2}\|f\|_{L^2} \geq$$

$$\geq \frac{\alpha}{2}\|u\|^2_{W^{1,2}} - \|u\|_{W^{1,2}}\|f\|_{L^2} = \|u\|_{W^{1,2}}\left(\frac{\alpha}{2}\|u\|_{W^{1,2}} - \|f\|_{L^2}\right).$$

As f is fixed we have that $\phi(u) \to \infty$ if $\|u\|_{W^{1,2}} \to \infty$.

Lemma 50) thus ensures the existence of a minimum and the lemma 49) yields uniqueness.

§4. Generalized problem for an elliptic equation

52) Notation: $\mathcal{M} = \{v \in C^1(\bar{\Omega});\ v = 0 \text{ on } \Gamma \subset \partial\Omega\}$

$V = \overline{\mathcal{M}}$ in $W^{1,2}(\Omega)$, i.e. $W_0^{1,2}(\Omega) \subset V \subset W^{1,2}(\Omega)$

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

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

53) Formulation of the problem:

Let $a_{ij} \in L^\infty(\Omega)$, $a_{ij}\xi_i\xi_j \geq \alpha|\xi|^2$, $\alpha > 0$, $\Omega \in C^{0,1}$.

$b \in L^\infty(\Omega)$, $b \geq 0$

$\sigma \in L^\infty(\partial\Omega)$, $\sigma \geq 0$

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

$g \in L^2(\partial\Omega)$ (~~Neumann~~ Newton 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) $((v,u)) = \int_\Omega v f \, dx + \int_{\partial\Omega} v g \, dS \quad \forall v \in V \quad (1)$

54) Interpretation of the problem:

Let us apply the Green theorem to (1), which is possible if all functions are smooth eno[ugh]

$$0 = \int_\Omega \left(-\frac{\partial}{\partial x_i}\left(a_{ij}\frac{\partial u}{\partial x_j}\right) + bu - f\right) v \, dx + \int_{\partial\Omega} \left(a_{ij}\frac{\partial u}{\partial x_j} n_i + \sigma u - g\right) v \, dS \quad (\dots$$

As v is any function from V and $W_0^{1,2}(\Omega) \subset V$, let us first consider the situation, when the trial function is any function from $W_0^{1,2}$. The same considerations as in the interpretation 40) 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

$$-\frac{\partial}{\partial x_i}\left(a_{ij}\frac{\partial u}{\partial x_j}\right) + bu = f .$$

Now let us take into account also the trial functions $v \in V \setminus W_0^{1,2}(\Omega)$

To satisfy (2), it must hold

$$\boxed{a_{ij}\frac{\partial u}{\partial x_j} n_i + \sigma u = g} \quad \text{on } \partial\Omega \setminus \Gamma,$$ where the trace of v need not be zero

this is called the Newton condition

If $\sigma = 0$ we get

$$\boxed{a_{ij}\cdot\frac{\partial u}{\partial x_j} n_i = g \quad \text{on } \partial\Omega\setminus\Gamma},$$

which is called the Neumann condition.

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

$$u = u_0 \quad \text{on } \Gamma .$$

55) <u>Theorem:</u> Let us consider the problem 53).

If $(N-1)$-dimensional measure of Γ is positive or $b>0$ on a set of positive N-dimensional ~~positive~~ measure or $\sigma>0$ on a set of positive $(N-1)$-dimensional measure, then there exists one and only on solution of the problem 53).

<u>Proof:</u> We will prove that assumptions of the Lax-Milgram theorem are satisfied. As the continuity of $((u,v))$ is clear, we need to prove the ellipticity of $((u,v))$, i.e. we need to prove that

$$((v,v)) \geq c \|v\|^2_{W^{1,2}} \quad \forall v \in V, \ c>0.$$

We will use the following contradiction:

Let the ellipticity does not hold. Then for each $c_n = \frac{1}{n}$ there exists v_n such that $((v_n,v_n)) < \frac{1}{n}\|v_n\|^2_{W^{1,2}}$, i.e.

$$\left(\left(\frac{v_n}{\|v_n\|}, \frac{v_n}{\|v_n\|}\right)\right) < \frac{1}{n}.$$

We have thus constructed the sequence $\{v_n'\}_{n=1}^{\infty}$, $v_n' \equiv \frac{v_n}{\|v_n\|}$ i.e. $\|v_n'\| = 1$ such that

$$((v_n', v_n')) < \frac{1}{n}. \qquad (*)$$

Now (n is not summing index in what follows):

$$((v_n', v_n')) = \int_\Omega \left(a_{ij} \frac{\partial v_n'}{\partial x_i}\frac{\partial v_n'}{\partial x_j} + b v_n'^2\right)dx + \int_{\partial\Omega} \sigma v_n'^2 \, dS \geq$$

$$\geq \int_\Omega \left(\alpha \frac{\partial v_n'}{\partial x_i}\frac{\partial v_n'}{\partial x_i} + b v_n'^2\right)dx + \int_{\partial\Omega} \sigma v_n'^2 \, dS$$

Now it is clear that $\frac{\partial V_n'}{\partial x_i} \to 0$ in $L_2(\Omega)$ $\forall i \in 1,2,\dots,N$ as (*) must be satisfied, i.e. $V_n' \to$ constant in L_2.

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

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

5b) The Neumann problem:

If the assumptions of the theorem 5b) are not satisfied, we have the Neumann problem:

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

$$\int_\Omega a_{ij} \cdot \frac{\partial v}{\partial x_i} \frac{\partial u}{\partial x_j}\, dx = \int_\Omega v f\, dx + \int_{\partial\Omega} v g\, dS \quad \forall v \in W^{1,2}(\Omega).$$

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:

$$\int_\Omega f\, dx + \int_{\partial\Omega} g\, dS = 0 .$$

57) <u>Theorem</u>: 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.

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

$$u = u_0 + \frac{\int_\Omega u\,dx}{\mu(\Omega)} \quad , \text{where } \mu(\Omega) \text{ is}$$

the N-dimensional Lebesgue measure of Ω. This means that

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

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

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

Let us consider first the situation, where the test functions are also only from Q (restriction of the problem to the subspace $Q \subset W^{1,2}(\Omega)$), i.e. we try to find $u \in Q$ such that

$$\int_\Omega a_{ij} \frac{\partial v}{\partial x_i} \frac{\partial u}{\partial x_j}\,dx = \int_\Omega v f\,dx + \int_{\partial\Omega} v g\,dS \quad \forall v \in Q .$$

On the product $Q \times Q$ the bilinear form $((u,v))$ ~~satisfies~~ is elliptic since

$$((v,v)) \underset{\substack{\uparrow \\ \text{ellipticity} \\ \text{of } a_{ij}}}{\geq} c \int_\Omega \frac{\partial v}{\partial x_i}\frac{\partial v}{\partial x_i}\,dx \overset{\substack{\text{Poincaré's inequality} \\ \downarrow}}{\geq} c_2 \|v\|^2_{W^{1,2}},$$

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

$$\|v\|^2_{W^{1,2}} \leq C\left(\int_\Omega \frac{\partial v}{\partial x_i}\frac{\partial v}{\partial x_i}\,dx + \left(\int_\Omega v\,dx\right)^2\right)$$

and the fact that $v \in Q$, i.e. $\int_\Omega v\,dx = 0$. Therefore, there exists a solution $\tilde{u}$ of the Neumann problem restricted to 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

$$v = v_0 + \frac{\int_\Omega v\,dx}{\mu(\Omega)} \equiv v_0 + k_v.$$

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

$$\int_\Omega a_{ij}\frac{\partial v}{\partial x_i}\frac{\partial \tilde{u}}{\partial x_j}\,dx = \int_\Omega v_0 f\,dx + \int_{\partial\Omega} v_0 g\,dS. \qquad (1)$$

If the equilibrium condition is fulfilled, we have

$$0 \overset{(2)}{=} \int_\Omega k_v f\,dx + \int_{\partial\Omega} k_v g\,dS \quad \left[= k_v\left(\int_\Omega f\,dx + \int_{\partial\Omega} g\,dS\right)\right]$$

By adding (1) and (2) we get that $\tilde{u}$ is also the solution of the Neumann problem on $W^{1,2}$, q.e.d.

58) <u>Theorem:</u> Let u be a weak solution of the problem 53.

Then $\|u\|_{W^{1,2}(\Omega)} \le c\left(\|u_0\|_{W^{1,2}(\Omega)} + \|f\|_{L^2(\Omega)} + \|g\|_{L^2(\partial\Omega)}\right)$,

where $c>0$ is a fixed constant.

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

$$((v,w)) = \int_\Omega v f\,dx + \int_{\partial\Omega} v g\,dS - ((v,u_0)) \equiv F(v)$$

The Lax-Milgram theorem yields

$$\|w\|_{W^{1,2}} \le \frac{1}{\alpha}\|F\| = \frac{1}{\alpha}\sup_{\|v\|_{W^{1,2}}\le 1} F(v)$$

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

1. $$\sup_{\|v\|_{W^{1,2}}\le 1}\int_\Omega v f\,dx \le \sup_{\|v\|_{W^{1,2}}\le 1}\left|\int_\Omega v f\,dx\right| \le \left(\int_\Omega f^2\,dx\right)^{1/2}\sup_{\|v\|_{W^{1,2}}\le 1}\left(\int_\Omega v^2\,dx\right)^{1/2} \le$$

$$\le \left(\int_\Omega f^2\,dx\right)^{1/2} = \|f\|_{L^2(\Omega)}$$

2. $$\sup_{\|v\|_{W^{1,2}}\le 1}\int_{\partial\Omega} v g\,dS \le \left(\int_{\partial\Omega} g^2\,dS\right)^{1/2}\sup_{\|v\|_{W^{1,2}}\le 1}\left(\int_{\partial\Omega} v^2\,dS\right)^{1/2} \le$$ the trace theorem

$$\le \|g\|_{L^2(\partial\Omega)}\sup_{\|v\|_{W^{1,2}}\le 1}\left(c_1\|v\|_{W^{1,2}(\Omega)}\right) = c_1\|g\|_{L^2(\partial\Omega)}$$

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

$$((v,u_0)) \le c_2\|v\|_{W^{1,2}}\|u_0\|_{W^{1,2}} \quad \text{and thus} \quad (c_2>0)$$

$$\sup_{\|v\|_{W^{1,2}}\le 1}((v,u_0)) \le c_2\|u_0\|_{W^{1,2}(\Omega)}$$

<u>Summary:</u> The generalized problem is correct according to Hadamard as
- i) the solution exists,
- ii) the solution is unique,
- iii) the solution is stable.

59) Theorem: Let $\tilde{u}$ be a solution of the Neumann problem 56), $\tilde{u} \in Q$.
Then $\|\tilde{u}\|_{W^{1,2}(\Omega)} \leq c \left(\|f\|_{L^2(\Omega)} + \|g\|_{L^2(\Omega)} \right)$, $c > 0$ is a constant.

Proof: As $\tilde{u} \in Q$ and $((u, v))$ satisfies the assumptions of the Lax-Milgram theorem on $Q \times Q$ and

$((\tilde{u}, v)) = \int_\Omega f v \, dx + \int_{\partial\Omega} g v \, ds \quad \forall v \in Q$, 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:

Seek for $u \in V$; $W_0^{1,2}(\Omega) \subset V \subset W^{1,2}(\Omega)$ such that

$$((u, v)) = F(v) \quad \forall v \in V, \tag{1}$$

where $((\ ,\ ))$ is bounded, elliptic bilinear form and $F \in (W^{1,2}(\Omega))'$.

Galerkin method:

Let $V_h \subset V$; $\dim V_h < \infty$; find $u_h \in V_h$ such that

$$((u_h, v_h)) = F(v_h) \quad \forall v_h \in V_h \tag{2}$$

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

Ritz method.

If $((v, w)) = ((w, v)) \quad \forall v, w \in V$, then (1) is equivalent to minimizing the function

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

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

Evidently $J(u) \le J(u_h)$ 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 ~~to~~ 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 numerical approach:

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

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

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

$$((\sum_{j=1}^m c_j \cdot v^j, v^i)) = F(v^i) \quad \forall\ i = 1,2,\dots,m ,$$

i.e.

$$\sum_{j=1}^m \underbrace{((v^j, v^i))}_{\text{stiffness matrix } A} \overset{\text{vector of unknowns (parameters)}}{c_j} = \underbrace{F(v^i)}_{\text{right-hand side or load vector}} \quad \forall i = 1,2,\dots,m$$

Regularity of the stiffness matrix:

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

$$(A\xi, \xi) = \sum_{i,j} ((v^i, v^j))\, \xi_i \xi_j = ((\sum_j \xi_j v^j, \sum_i \xi_i v^i)) \equiv ((v, v)) \geq$$

$$\geq \alpha \|v\|^2_{W^{1,2}} > 0 \quad \text{as } \alpha > 0 \text{ and } \xi \neq 0 .$$

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 $\overline{\Omega}$,

(ii) the functions $v_h \in V_h$ 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: $\{v^j\}_{j=1}^m$ is formed by basis functions of usual Fourier series (trigonometric functions, spherical harmonics), which leads to spectral methods.

Example: Prof. Martinec employs triangle ~~FE~~ 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 $\overline{\Omega}$:

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

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

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

(iii) $\forall\, K_1, K_2 \in \mathcal{T}_h$, $K_1 \neq K_2$, $K_1^0 \cap K_2^0 = \emptyset$,

(iv) the boundary ∂K is the Lipschitz one $\forall\, K \in \mathcal{T}_h$.

The <u>discretization (triangulation) parameter h</u> is the maximum diameter of all $K \in \mathcal{T}_h$.

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

$$P_K = \{ v_h|_K \; ; \; v_h \in V_h \}$$

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, Σ), where:

(i) K is a closed subset of 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) Σ is a finite set of linearly independent linear forms Φ_i, $1 \le i \le N$, defined over the space P (or over a space which contains P).

The set Σ is said to be <u>P-unisolvent</u> if for any real scalars α_i, $1 \le i \le N$, there exists a unique function $p \in P$ which satisfies

$$\Phi_i(p) = \alpha_i \quad , \; i = 1, 2, \dots, N$$

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

$$\Phi_j(p_i) = \delta_{ij} \, , \quad j = 1, 2, \dots, N,$$

where δ_{ij} is Kronecker's symbol. ~~Since~~

~~$\forall p \in P$~~

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

Consequence of P-unisolvency:

If Σ 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, $\phi_j\left(\sum_{i=1}^{N} \phi_i(p)\, p_i\right) = \sum_{i=1}^{N} \phi_i(p)\, \phi_j(p_i) = \sum_{i=1}^{N} \phi_i(p)\, \delta_{ij} =$

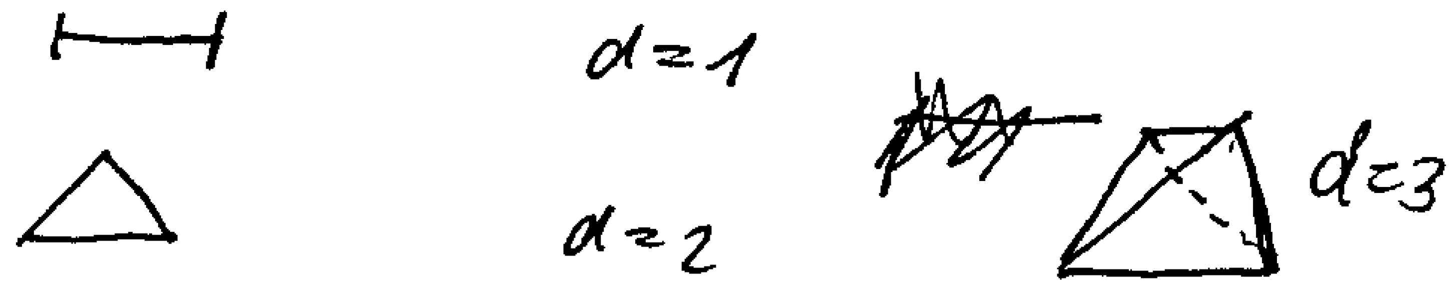
$= \phi_j(p)$.

Examples of finite elements:

Lagrange finite elements are those with the degrees of freedom of the form $p \longmapsto p(A)$, $A \in K$, where A is called the node.

a) <u>Linear simplicial element</u>:

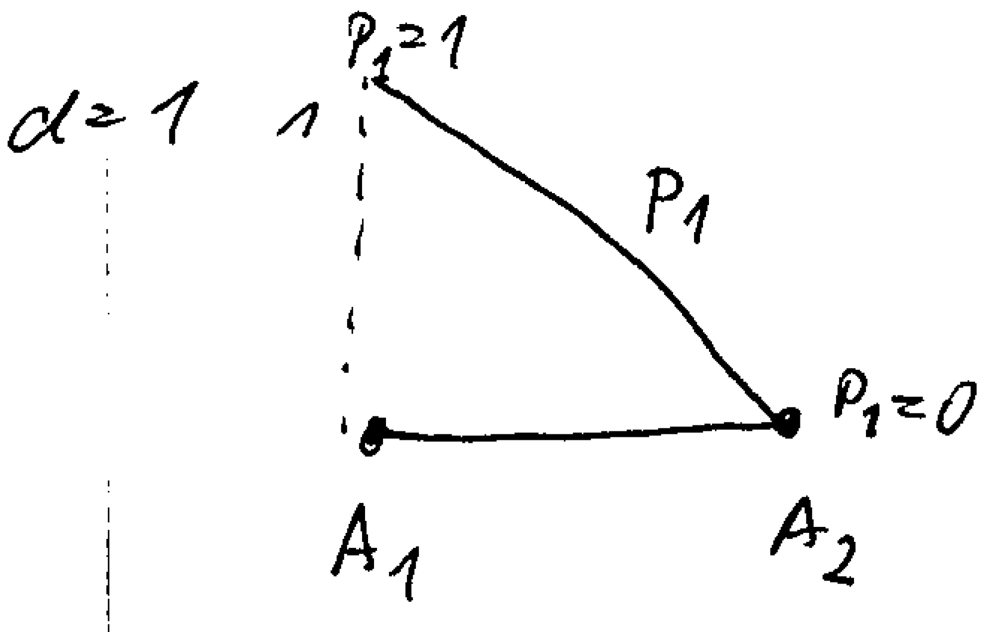
The set K is a d-simplex

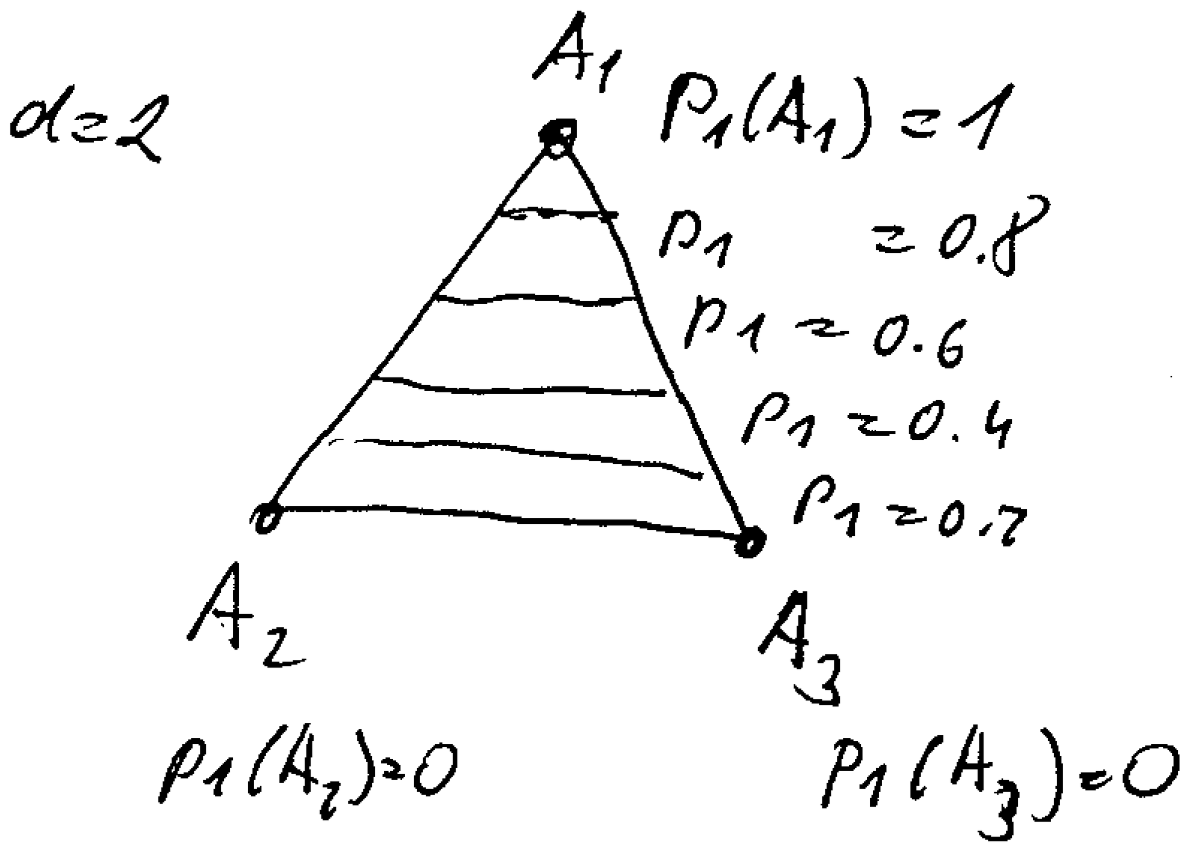


$P_K = \{$ space of linear functions $p = \gamma_0 + \gamma_1 x_1 + \dots + \gamma_d x_d \; ; \; \gamma_i \in E_1 \}$

$\dim P_K = d+1$

$\Sigma_K = \{ \phi_i(p) = p(A_i) \,,\; i = 1, 2, \dots, d+1 \,,$ where A_i are the vertices of $K \}$





b) <u>Quadratic simplicial element</u>

$K \; \dots \; d$-simplex

$$P_K = \left\{ p = \gamma_0 + \sum_{i=1}^{d} \gamma_i x_i + \sum_{\substack{1 \le i \le d \\ 1 \le j \le d}} \gamma_{ij} x_i x_j \; ; \; \gamma_i, \gamma_{ij} \in E_1 \right\}$$

As $x_i x_j = x_j x_i$ for $i \neq j$ the number of independent γ_{ij} is

$$\binom{d}{2} + d \quad \Rightarrow \quad \dim P_K = \binom{d}{2} + d + d + 1 = \frac{1}{2}(d+1)(d+2)$$

$$\Sigma_K = \{ p(A_i) \,,\; i = 1, 2, \dots, d+1 ; \; p(A_{ij}) \,,\; i, j = 1, 2, \dots, d+1 \,, \text{ where}$$

symbolic notation of $\phi_i(p) = p(A_i)$

A_i are the vertices and

$A_{ij} = \frac{1}{2}(A_i + A_j)$ - ~~middle~~ midpoint of edges

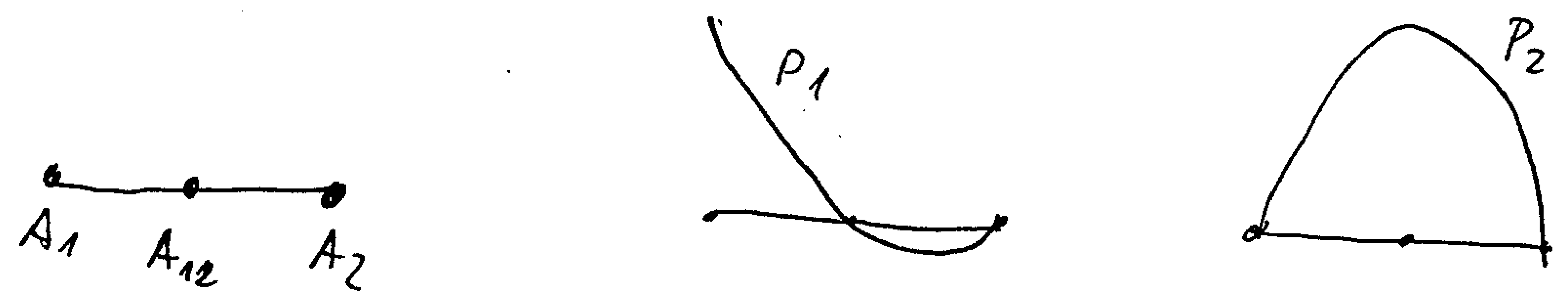
number of vertices .... $d+1$

number of midpoints ... $\binom{d+1}{2}$

$$\Rightarrow \binom{d+1}{2} + d + 1 = \frac{1}{2}(d+1)d + d + 1 =$$

$$= (d+1)\left(\frac{d}{2} + 1\right) = \frac{1}{2}(d+1)(d+2)$$

$d=1$



c) Bilinear and trilinear rectangular elements

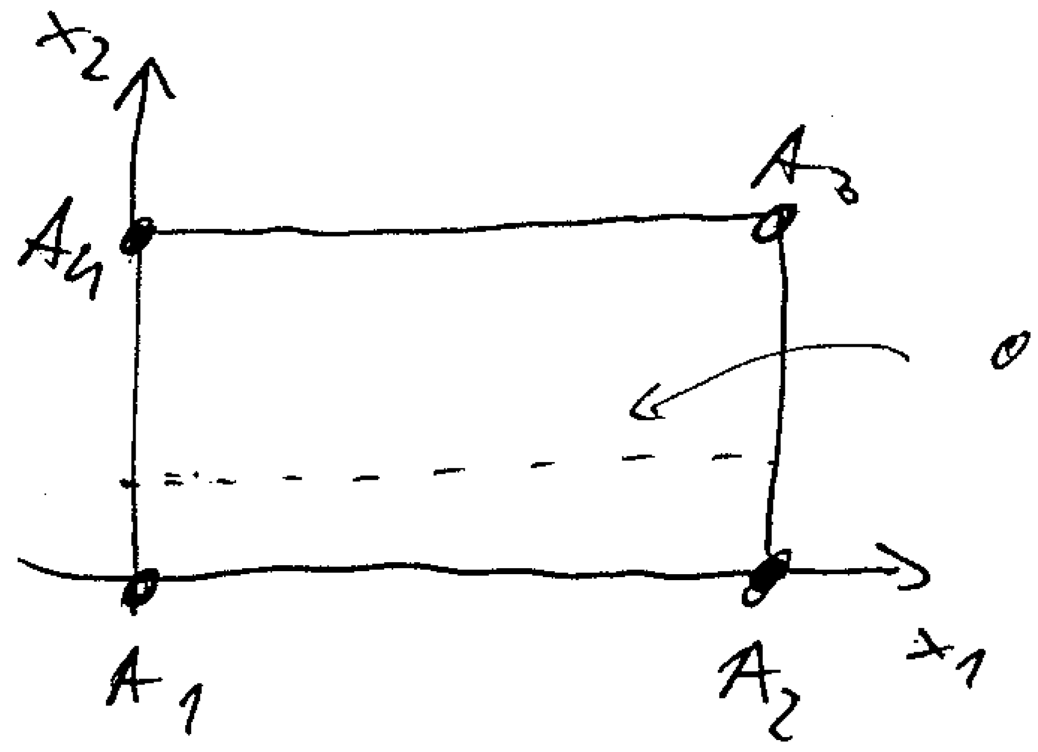
K ... d-rectangle

$$P_K = \left\{ p = \sum_{i,j=0}^{1} \gamma_{ij}\, x_1^i x_2^j \right\}, \quad d=2$$

$$P_K = \left\{ p = \sum_{i,j,k=0}^{1} \gamma_{ijk}\, x_1^i x_2^j x_3^k \right\}, \quad d=3$$

$$\Sigma_K = \left\{ p(A_i),\ i=1,2,\dots,2^d \right\}$$

$d=2$



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:

K ... triangle

$$P_K = \{ p = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_1^2 + \gamma_4 x_1 x_2 + \gamma_5 x_2^2 + \gamma_6 x_1^3 + \gamma_7 x_1^2 x_2 + \gamma_8 x_1 x_2^2 + \gamma_9 x_2^3, \ \gamma_i \in E_1 \} \ ; \ \dim P = 10$$

$$\Sigma_K = \{ p(A_i), \frac{\partial p}{\partial x_1}(A_i), \frac{\partial p}{\partial x_2}(A_i), p(G); \ A_i, \ i = 1,2,3, \text{ are vertices and } G \text{ is its centre of gravity} \}$$

<u>other types of degrees of freedom</u>, which are used:

$\phi(p) = \frac{\partial p}{\partial n}(A)$ — normal derivative in a node on a side

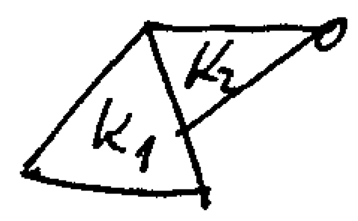
$\phi(p) = (D^i p)(A), \ |i| \geq 1$ — higher order derivative

$\phi(p) = \int_K p(x)\, dx$ — integral over the element

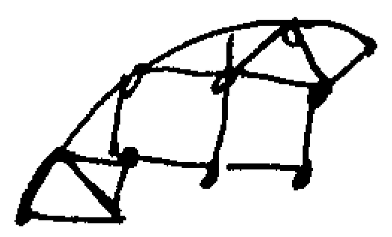
etc.

§3. Spaces of Lagrangian finite elements

Let us consider a triangulation $\mathcal{T}_h$ such that the sides of the elements are either parts of the boundary $\partial\Omega$ 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 $\mathcal{N}_K$. Let $\mathcal{N}_h = \bigcup_{K\in\mathcal{T}_h} \mathcal{N}_K$.

The space of finite elements:

$$X_h = \{ v_h \in C(\overline{\Omega});\ v_h|_K \in P_K \ \ \forall K \in \mathcal{T}_h \},$$

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

$$\Sigma_h = \{ v_h(A),\ A \in \mathcal{N}_h \}$$

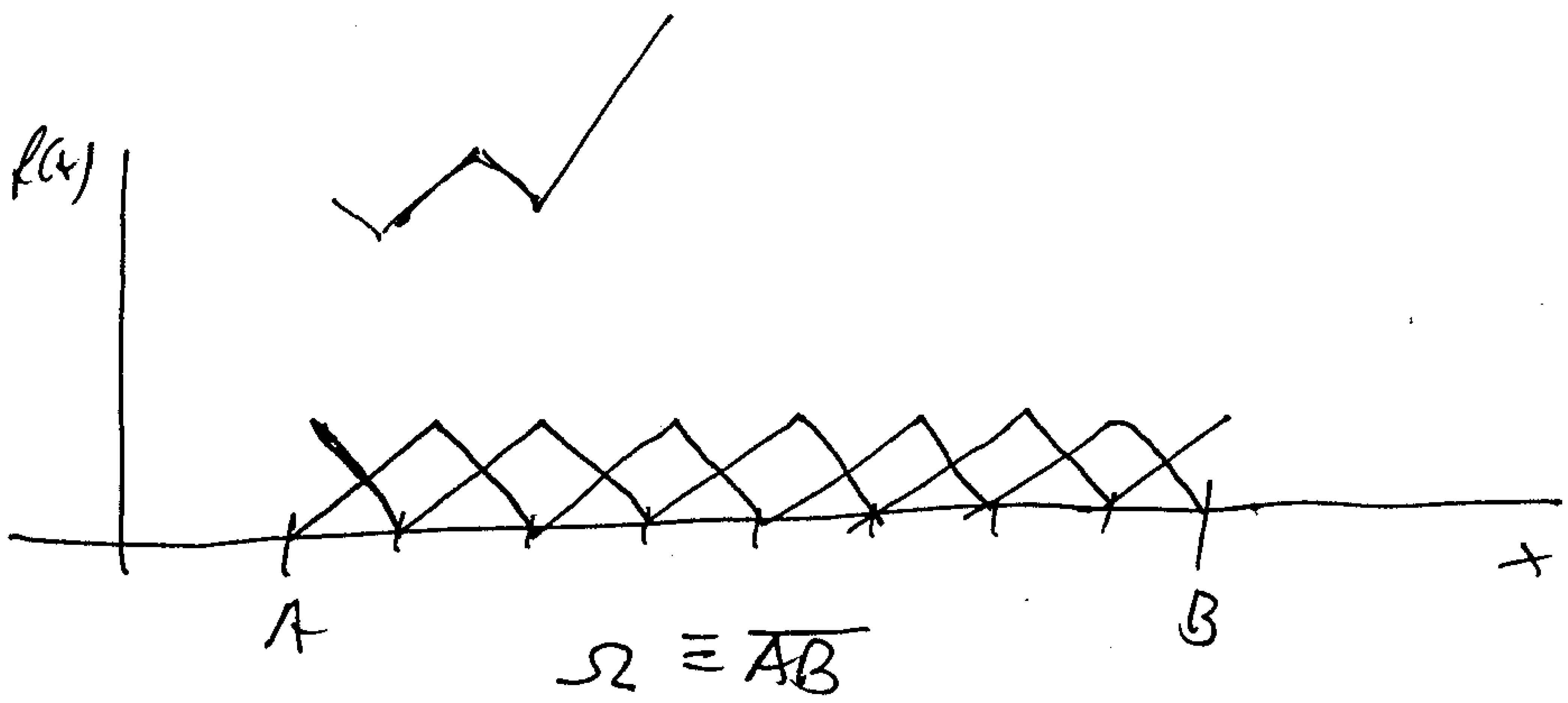
which is the called by the set of degrees of freedom of the finite element space.

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

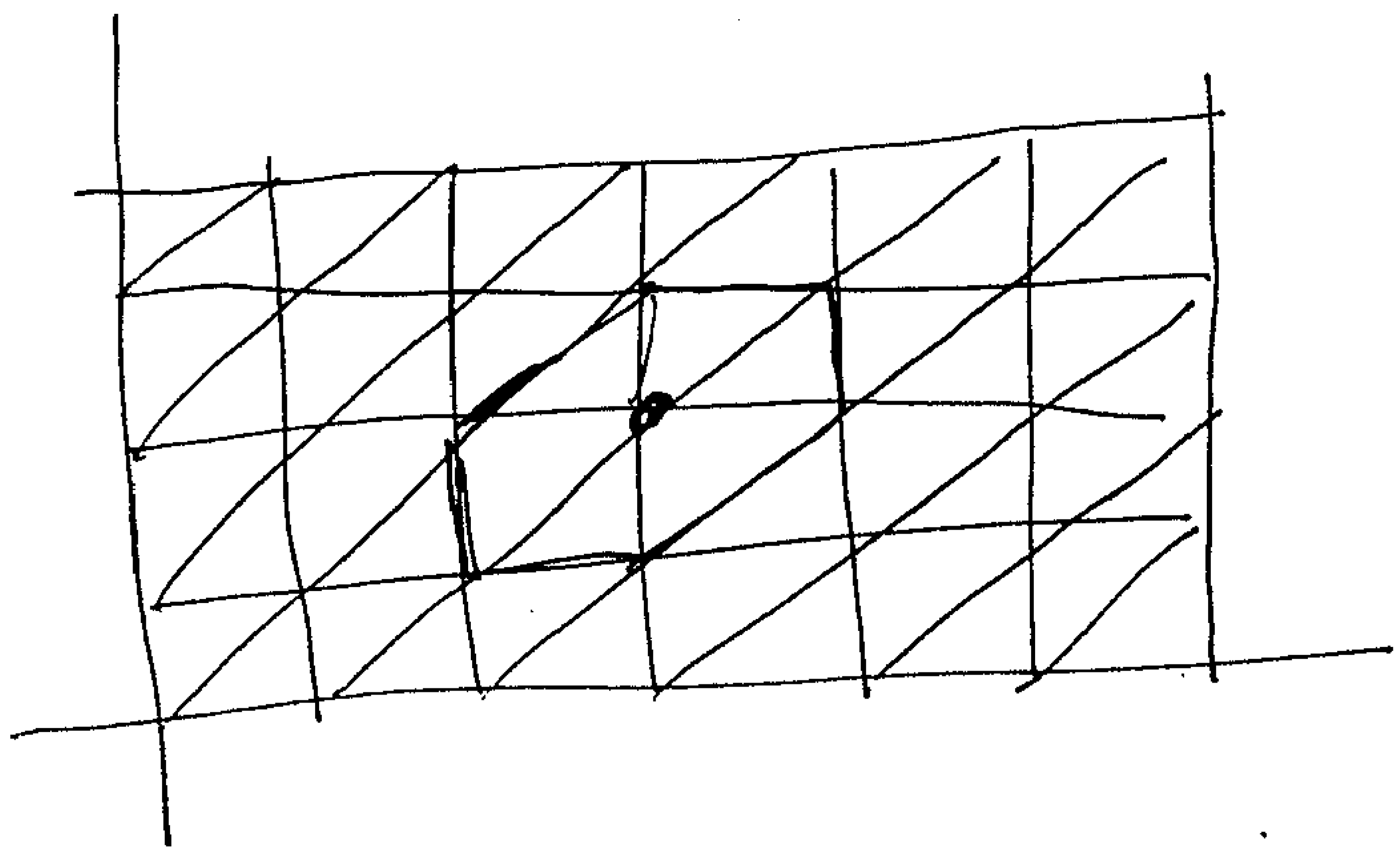
$$\{ v_h^i \in X_h \; ; \; v_h^i(A_j) = \delta_{ij} \; , \; i,j = 1,2,\dots, \dim(X_h) \}$$

Examples:

$d = 1$

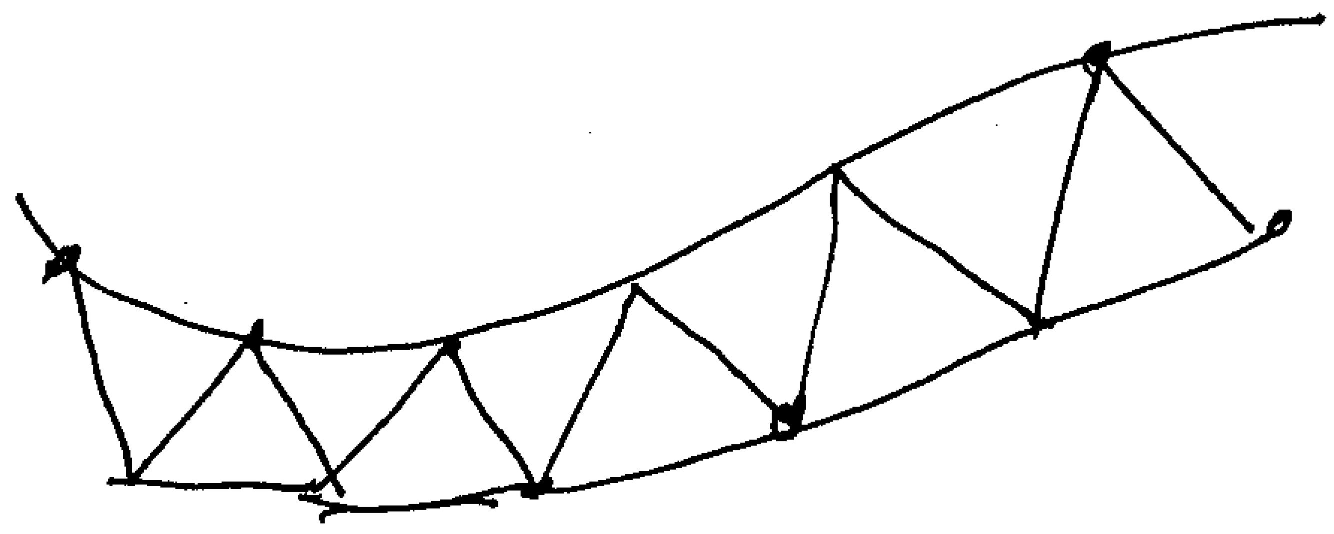


$d = 2$



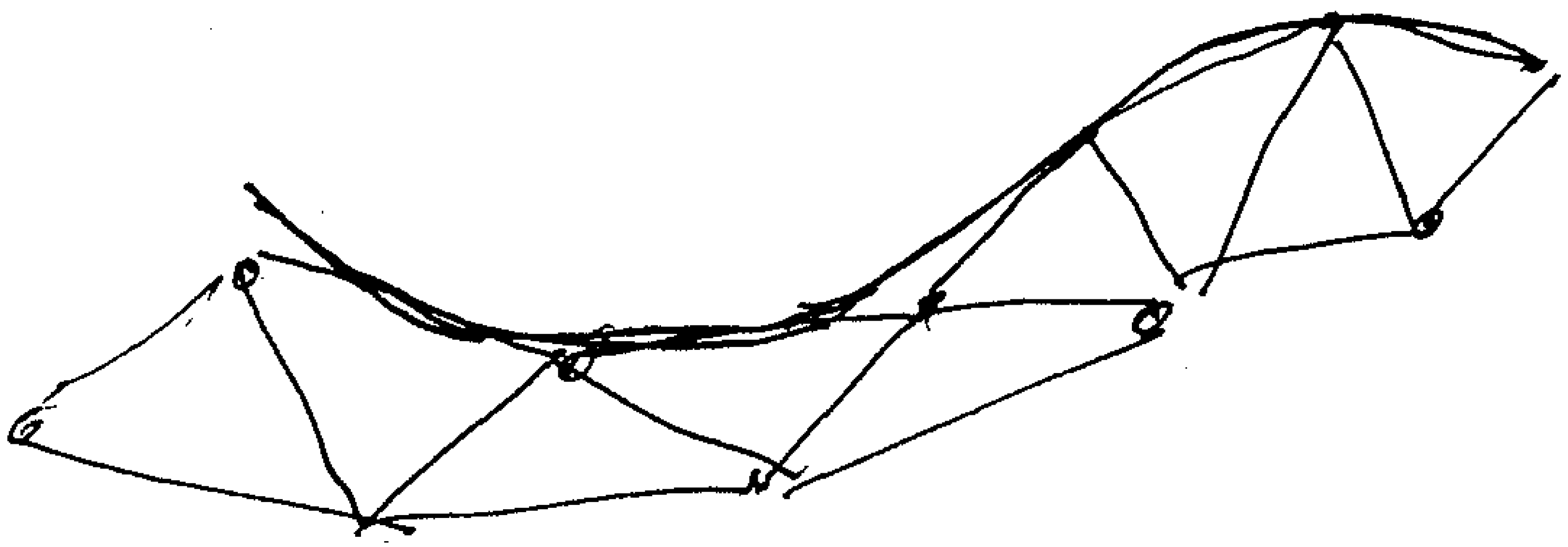
Curved boundaries

- isoparametric (curved) elements



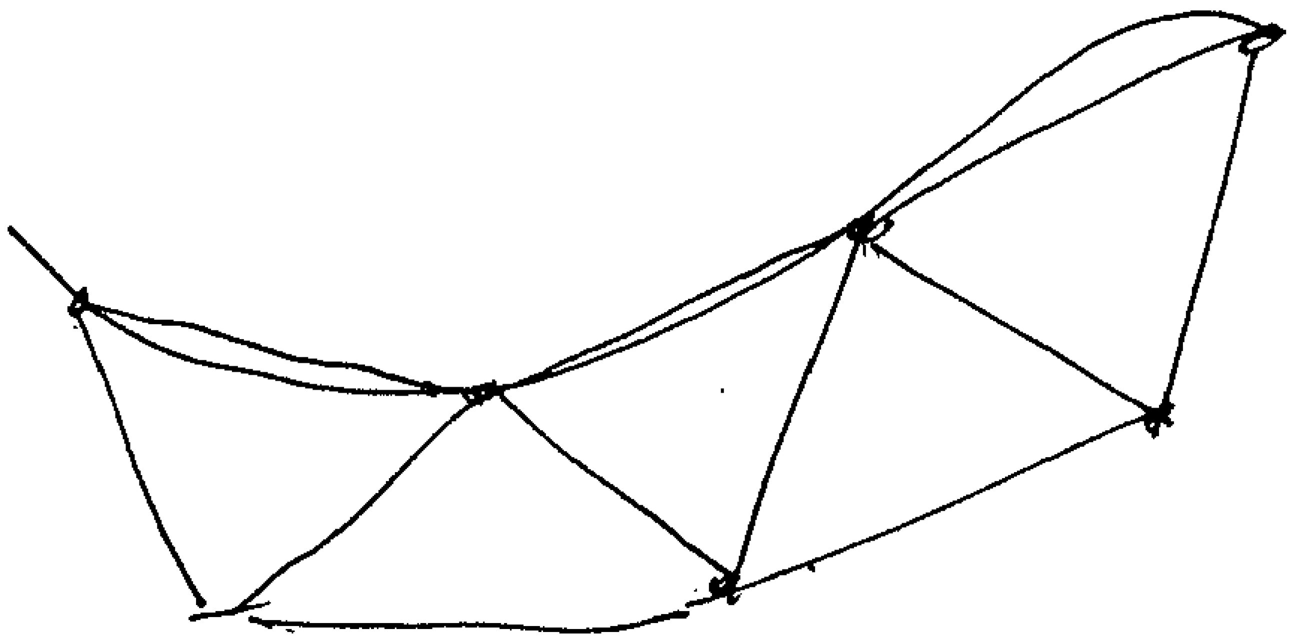
K need not be convex and P_K may be formed, e.g. by rational functiones

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



- nonconforming methods ($V_h \not\subset V$)

$\Omega_h \not\subset \Omega \Rightarrow V_h \not\subset V$



§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 subspace $\{V_h\}$, $V_h \subset V$ and the set $\{u_h\}$ of the solutions of the problem
$$((u_h, v_h)) = F(v_h) \quad \forall v_h \in V_h$$
such that
$$\lim_{h \to 0} \|u - u_h\|_V = 0 .$$
Then we say that the associated family of discrete problems is convergent.

Theorem (Cea's lemma): There exists a constant $c > 0$ independent of the subspace V_h such that
$$\|u - u_h\|_V \le C \inf_{v_h \in V_h} \|u - v_h\|_V .$$
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 $\|\cdot\|_V$-nor.

Proof: Let w_h be an arbitrary element in V_h, evidently $((u-u_h, w_h)) = 0$; the V-ellipticity and continuity of $((\cdot,\cdot))$ yield
$$c_2 \|u - u_h\|_V^2 \le ((u-u_h, u-u_h)) = ((u-u_h, u-u_h)) + ((u-u_h, \underbrace{u_h - v_h}_{w_h \in V_h})) \overset{0}{}$$

v_h is arbitrary
$$= ((u-u_h, u - v_h)) \le c_1 \|u - u_h\|_V \|u - v_h\|_V$$
Therefore, $\|u - u_h\|_V \le \dfrac{c_1}{c_2} \|u - v_h\|_V \quad \forall v_h \in V_h$

§5. Solution of a discrete problem

§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 $\Gamma \subset \partial\Omega$. Let $W^{1,2}((0,T),V)$ denotes the mapping $t \in (0,T) \longmapsto v(t) \in V$ such that the function $t \longmapsto \|v(t)\|_V$ is from $W^{1,2}(0,T)$.

Formulation of the problem:

Find $u(t,x) \in W^{1,2}((0,T),V)$ such that

$$\underbrace{\left(\frac{\partial u}{\partial t}, v\right)}_{\int_\Omega \frac{\partial u}{\partial t} v \, dx} + \underbrace{((u,v))}_{\int_\Omega \left(a_{ij}\frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_j} + bvu\right)dx + \int_{\partial\Omega} \sigma v u \, dS} = F(v) \quad \forall v \in V \text{ and allmost all } t \in (0,T)$$

$$F(v) = \int_\Omega v f \, dx + \int_{\partial\Omega} v g \, dS$$

$$u(0,x) = \varphi(x) \in W^{1,2}(\Omega) \cap C(\overline{\Omega})$$

Galerkin approximation:

Let $V_h \subset V$ be of finite dimensions; find $u_h(t,x) \in V_h$ $\forall t \in (0,T)$ such that

$$\left(\frac{\partial u_h}{\partial t}, v_h\right) + ((u_h, v_h)) = F(v_h) \qquad \forall v_h \in V_h .$$

Let us try to find u_h in the form

$$u_h(t,x) = \sum_{i=1}^{m} U_i(t)\, v^i(x) \quad ; \quad \{v^i\}_{i=1}^{m} \text{ is a basis in } V_h .$$

Then $$\sum_{i=1}^{m} \frac{\partial U_i}{\partial t}(v^i, v^j) + \sum_{i=1}^{m} U_i ((v^i, v^j)) = F(v^j) \quad j = 1,2,\dots,m$$

If $U = (u_1, \dots, U_m)^T$ is the vector of unknowns, we get

$$MU' + AU = f, \quad \text{where} \quad \begin{aligned} M_{ij} &= (v^i, v^j) \\ A_{ij} &= ((v^i, v^j)) \\ f &= (F(v^1), \dots, F(v^m))^T \end{aligned}$$

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:

$$M \frac{U^{k+1} - U^k}{\Delta t} + A \frac{U^{k+1} + U^k}{2} = \frac{f^{k+1} + f^k}{2}, \quad \text{i.e.}$$

we must solve the system

$$\left(M + \frac{\Delta t A}{2}\right) U^{k+1} = \left(M - \frac{\Delta t A}{2}\right) U^k + \Delta t \, \frac{f^{k+1} + f^k}{2}$$

$$U^0 = (U_1(0), \dots, U_m(0))^T$$

Euler method: (fully implicit method)

$$M \frac{U^{k+1} - U^k}{\Delta t} + A U^{k+1} = f^{k+1}$$

$$U^0 = U(0)$$

Explicit methods:

Euler explicit method: $M \dfrac{U^{k+1} - U^k}{\Delta t} + AU^k = f^k$, i.e.

$$MU^{k+1} = MU^k + \Delta t \, (f^k - AU^k)$$

(unstable, problems with accuracy in periodic or curved problems)

Higher-order methods

They start from the fact that

$$U' = M^{-1}(f - AU)$$

represent the system of ordinary differential ~~problem~~ equations 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

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

let us look at the solution in a form

$$u = e^{-\frac{t}{T}} \sin kx, \text{ i.e.}$$

$$-T^{-1} = -\varkappa k^2 \Rightarrow T = \frac{1}{\varkappa k^2}.$$

Maximal resolved k under the discretization with the discretization parameter h is $k \sim \frac{1}{h}$, i.e. minimal characteristic time is

$$T_{min} \sim \frac{h^2}{\varkappa}.$$

To be able to "catch" such a T_{min} it is clear that

$$\Delta t \lesssim T_{min} \sim \frac{h^2}{\varkappa}.$$

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