

Metoda konečných prvků – učební text

Jan Březina

Jan Stebel

13. října 2016

Obsah

1	Opakování analýzy a lineární algebry	2
2	Opakování plošných a křivkových integrálů	2
2.1	Křivkový integrál 1. druhu	2
2.2	Křivkový integrál 2. druhu	3
2.3	Plošný integrál 1. druhu	3
2.4	Plošný integrál 2. druhu	4
2.5	Integrační věty: Stokesova, Gaussova, Greenova	4
3	Zákony zachování, věta o transportu	5
3.1	Eulerovy rovnice	6
4	Odvození rovnice vedení tepla	6
5	Transportní procesy	6
5.1	Proudění v porézním prostředí	7
5.2	Transport chemických látek	7
6	Vlnová rovnice (akustika)	7
7	Mechanika	8
8	Elektromagnetismus	8
9	Klasifikace PDR	8
9.1	Eliptické rovnice	8
9.2	Parabolické rovnice rovnice	9
9.3	Hyperbolické rovnice	9
10	Odvození slabého řešení	9
11	Úvod do funkcionální analýzy	9
11.1	Normované lineární prostory	10
11.1.1	Množiny v normovaném lineárním prostoru	11
11.1.2	Konvergence	12
11.2	Prostor $L^2(\Omega)$	13
11.3	Prostory $H^1(\Omega)$	14
12	Variační (slabá) formulace okrajové úlohy	15
13	Galerkinova metoda	15
14	Iterační metody řešení soustav lineárních rovnic	16
14.1	Klasické iterační metody	16
14.2	Metody Krylovových podprostorů	18
14.2.1	Metoda sdružených gradientů (CG)	18
14.2.2	Zobecněná metoda minimálních reziduí (GMRES)	19
14.2.3	Metoda bikonjugovaných gradientů (BiCG)	20

1 Opakování analýzy a lineární algebry

2 Opakování plošných a křivkových integrálů

2.1 Křivkový integrál 1. druhu

Jaká je hmotnost vlasu? Představme si natažený vlas a předpokládejme, že takto natažený má konstantní hustotu. Na vlasu si zavedeme souřadnici t , $t = 0$ je začátek vlasu $t = 1$ je konec vlasu. V konkrétním bodě t na vlasu má vlas průřez $S(t)$. Pro malý přírůstek dt je hmotnost kousku vlasu $dm = \rho S(t) dt$. Celková hmotnost pak je:

$$m = \int_0^1 \rho S(t) dt = \int_0^1 \rho_t(t) dt, \quad \rho_t(t) = \frac{dm}{dt} = \rho S(t) \quad (2.1)$$

kde ρ_t je délková hustota. Když vlas pustíme, tak se trochu zkrátí a zkroutí do nějaké křivky v prostoru. Původní bod t má nyní v prostoru polohu $\varphi(t)$. Tím zkroucením se změní průřezy S , ale nezmění se délková hustota, takže hmotnost opět spočteme podle 2.1. Nyní si představme, že vlas je v tíhovém poli $f(\mathbf{x})$, pro jednoduchost si představujeme, že tíha působí pouze v směru z a má skalární velikost f , která se ovšem mění ve všech směrech. Jaká na vlas působí celková síla? Pro natažený vlas podél osy x máme, $dF = f(\mathbf{x}) dm$, a tedy:

$$F = \int_0^1 \rho_t(t) f[(t, 0, 0)] dt$$

a pro zkroucený vlas:

$$F = \int_k f \rho_t dk = \int_0^1 f(\varphi(t)) \rho_t(t) dt$$

Nakonec si představme, že jde o zkamenělý vlas uvnitř skalního bloku, jehož hustota $\rho(\mathbf{x})$ je známá pro každý bod \mathbf{x} . Přírůstek síly působící pouze na ten zkamenělý vlas je $dF = f(\mathbf{x}) \rho(\mathbf{x}) S(t) dl$, kde

$$dl = \sqrt{dx^2 + dy^2 + dz^2} = \sqrt{(\varphi'_x)^2 + (\varphi'_y)^2 + (\varphi'_z)^2} dt = |\varphi'(t)| dt$$

je přírůstek délky křivky pro přírůstek parametru dt . Celková síla působící na vlas pak je:

$$F = \int_k f \rho dk = \int_0^1 f(\varphi(t)) \rho(\varphi(t)) |\varphi'(t)| dt =$$

Derivace $\varphi'(t)$ je tečný vektor ke křivce k a jeho velikost je skutečný přírůstek v prostoru pro přírůstek dt . Tento typ integrálu nazýváme křivkový integrál 1. druhu ze skalárního pole f podél křivky k , která je dána parametricky:

$$k : \{ \varphi(t); t \in (0, 1) \}$$

Integrál je vlastně definován pomocí substituce $\mathbf{x} = \varphi(t)$:

$$\int_k f(\mathbf{x}) dk = \int_0^1 f(\varphi(t)) |\varphi'(t)| dt = \int_0^1 f(\varphi_x, \varphi_y, \varphi_z) \sqrt{(\varphi'_x)^2 + (\varphi'_y)^2 + (\varphi'_z)^2} dt. \quad (2.2)$$

Integrál 1. druhu můžeme aplikovat i na vektorové pole, ale výsledkem pak bude vektor.

Další (fyzikální) příklady použití křivkového integrálu prvního druhu.

- **Moment síly** (vůči počátku), $M(\mathbf{x}) = \mathbf{F}(\mathbf{x}) \times \mathbf{x}$. Celkový moment na ohnutém drátu:

$$M = \int_k \mathbf{F}(\mathbf{x}) \times \mathbf{x} dk = \int_0^1 \mathbf{F}(\varphi(t)) \times \varphi(t) |\varphi'(t)| dt$$

- **Délka křivky** je integrál (1. druhu) ze skalárního pole $f(x, y, z) = 1$, tj.

$$L = \int_\alpha^\beta |\varphi'(t)| dt,$$

- **Průměrná teplota** na poledníku k . Poledník je myšlená křivka na povrchu země a tiše předpokládáme, že je hladká.

$$T = \frac{1}{L} \int_k T(\mathbf{x}) dk = \frac{1}{L} \int_0^1 T(\varphi(t)) |\varphi'(t)| dt,$$

kde L je (skutečná) délka poledníku (viz. bod 2.1).

- **Hmota křivky nebo plochy** je integrál (1. druhu) ze skalárního pole hustoty $\rho(x, y, z)$.

$$M = \int_k \rho(x, y, z) dk$$

- **Souřadnice těžiště** křivky je vektor (T_x, T_y, T_z) integrálů (1. druhu) z vektoru skalárních funkcí $x\rho(x, y, z)$, $y\rho(x, y, z)$, $z\rho(x, y, z)$ dělený celkovou hmotou M . Např. pro plochu S :

$$\mathbf{T} = \frac{1}{M} \int_k \mathbf{x} \rho(\mathbf{x}) dk$$

- **Moment setrvačnosti** vzhledem k ose o je integrál (1. druhu) ze skalární funkce $f(\mathbf{x}) = r^2 \rho(\mathbf{x})$, kde r je vzdálenost bodu \mathbf{x} od osy o . Ideální je transformovat křivku i osu tak aby osa byla jedna ze souřadných os, např. pro o totožnou s osou z je

$$I_z = \frac{1}{M} \int_k (x_x^2 + x_y^2) \rho(\mathbf{x}) dk$$

2.2 Křivkový integrál 2. druhu

Ve vektorovém zápisu je integrál (2. druhu) z vektorového pole \mathbf{F} podél křivky k :

$$\int_k \mathbf{F} \cdot \mathbf{t}_k dk = \int_0^1 \mathbf{F}(\varphi(s)) \cdot \varphi'(s) ds \quad (2.3)$$

Zde je \mathbf{t}_k tečný vektor. Přesněji pokud dk je velikost tečného vektoru jako pro integrál 1. druhu, tak \mathbf{t}_k , je vlastně jednotkový tečný vektor. Ovšem stále je třeba výrazy vlevo v (2.2), (2.3), (2.4), (2.5) jsou pouze symboly (zkratky), pro to co stojí vpravo. Pro některé druhy operací stačí manipulovat se zkratkami, ale někdy je potřeba se ponořit do definice.

Příklady:

- **Práce síly po křivce.** Integrál 2. druhu z vektorové funkce síly.

2.3 Plošný integrál 1. druhu

Podobně jako v případě křivky je plocha dána zobrazením $\varphi(u, v)$ z množiny $M \subset \mathbb{R}^2$ do \mathbb{R}^3 . Normála \mathbf{N} k ploše v bodě daném parametry (u, v) , t.j. v bodě $\varphi(u, v)$ je dána vektorovým součinem tečných vektorů:

$$\mathbf{N} = \mathbf{t}_u \times \mathbf{t}_v, \quad \mathbf{t}_u = \frac{\partial \varphi}{\partial u}, \quad \mathbf{t}_v = \frac{\partial \varphi}{\partial v}.$$

Jednotková normála je pak $\mathbf{n} = \mathbf{N}/|\mathbf{N}|$.

Integrál (1. druhu) ze skalárního pole f přes plochu $S = \{\mathbf{x} = \varphi(u, v), (u, v) \in M\}$ je definován:

$$\int_S f dS = \iint_M f(\varphi(u, v)) |\mathbf{N}(u, v)| du dv = \iint_M f(\varphi_x, \varphi_y, \varphi_z) \sqrt{(N_x)^2 + (N_y)^2 + (N_z)^2} du dv, \quad (2.4)$$

Pozor, pokud plocha není rovina, tak normála a tudíž i její velikost jsou funkcí parametrů u, v , [?] str. 151.

- Velikost povrchu P plochy M je integrál (1. druhu) ze skalárního pole $f(x, y, z) = 1$, tj.

$$P = \int_M 1 dS = \int_M |\mathbf{n}(u, v)| du dv.$$

- **Hmota plochy** je integrál (1. druhu) ze skalárního pole hustoty $\rho(x, y, z)$.

$$M = \int_S \rho(x, y, z) dS$$

Souřadnice těžiště plochy je vektor \mathbf{T} integrálů (1. druhu) z vektoru skalárních funkcí $\mathbf{x}\rho(\mathbf{x})$ dělený celkovou hmotou M . Např. pro plochu S :

$$\mathbf{T} = \frac{1}{M} \int_S \mathbf{x}\rho(\mathbf{x}) dS$$

Moment setrvačnosti vzhledem k ose o je integrál (1. druhu) ze skalární funkce $f(\mathbf{x}) = r^2\rho(\mathbf{x})$, kde r je vzdálenost bodu $[x, y, z]$ od osy o . Pro osu z :

$$I_z = \frac{1}{M} \int_S (x^2 + y^2)\rho(\mathbf{x}) dS$$

Práce síly po křivce. Integrál 2. druhu z vektorové funkce síly. **Tok kapaliny skrze plochu za jednotkový čas.** Integrál 2. druhu z vektorového pole rychlosti.

2.4 Plošný integrál 2. druhu

Podobně lze integrál (2. druhu) vektorového pole \mathbf{F} skrze plochu S napsat:

$$\int_S \mathbf{F} \cdot \mathbf{n} dS = \int_M \mathbf{F}(\varphi(u, v)) \cdot (\partial_u \varphi \times \partial_v \varphi) du dv \quad (2.5)$$

- Tento integrál má význam celkového toku pole skrz plochu. Například množství kapaliny, které proteče skrze plochu za jednotkový čas.

2.5 Integrační věty: Stokesova, Gaussova, Greenova

Greenova věta (integrace per partes): Pokud má oblast V hranici S , pak pro hladká skalární pole u a v platí:

$$\int_V \partial_x uv dV = \int_S uv n_x dS - \int_V u \partial_x v dV$$

kde n_x je složka jednotkové normály. Odtud pro hladké vektorové pole \mathbf{v} dostaneme:

$$\int_V (\nabla u) \cdot \mathbf{v} dV = \int_S uv \cdot \mathbf{n} dS - \int_V u \operatorname{div} \mathbf{v} dV$$

Gaussova věta: Pro objem V ohraničený plochou S platí

$$\int_V \operatorname{div} \mathbf{F} dV = \int_S \mathbf{F} \cdot \mathbf{n} dS.$$

Můžeme odvodit z Greenovy věty, použitím $u = 1$ a $\mathbf{v} = \mathbf{F}$:

$$0 = \int_V (\nabla 1) \cdot \mathbf{F} dV = \int_S 1 \mathbf{F} \cdot \mathbf{n} dS - \int_V 1 \operatorname{div} \mathbf{F} dV$$

Stokesova věta: Pro plochu S ohraničenou uzavřenou křivkou k platí

$$\int_S \operatorname{rot} \mathbf{F} \cdot \mathbf{n} dS = \int_k \mathbf{F} \cdot \mathbf{t} dk.$$

Hranici oblasti Ω zapisujeme též jako $\partial\Omega$.

3 Zákony zachování, věta o transportu

Konzervativní veličina.

- Zachování hmoty.
- Zachování hybnosti.
- Zachování momentu hybnosti.
- Zachování energie. Zachování vnitřní energie, tepla.

Natahovací pytlík s vodou se třpytkama. Hustota třpytek v bodě \mathbf{x} v čase t je $\rho(t, \mathbf{x})$. Pytlík v čase t , je oblast (otevřená jednoduše souvislá množina) Ω_t , takže ho můžeme různě deformovat. Počet třpytek v pytlíku je pořád stejný:

$$\frac{d}{dt} \int_{\Omega_t} \rho(t, \mathbf{x}) d\mathbf{x} = 0.$$

Popis deformace v čase. Bod \mathbf{x}_0 v čase 0 je přesunut do bodu \mathbf{x}_t v čase t .

$$\mathbf{x}_t = X(t, \mathbf{x}_0).$$

Rychlostní pole pak je $\mathbf{u}(t, \mathbf{x}_t) = \partial_t \mathbf{X}(t, \mathbf{x}_0)$.

Věta 3.1 (Reynolds transport theorem). *Nechť $q(t, \mathbf{x})$ je hladká skalární funkce na oblasti Ω_t . Oblast Ω_t je dána hladkým zobrazením $X(t, \mathbf{X})$ a počáteční oblastí Ω_0 :*

$$\Omega_t = \{\mathbf{x}_t = \mathbf{X}(t, \mathbf{x}_0); \mathbf{x}_0 \in \Omega_0\}.$$

Pak platí:

$$\frac{d}{dt} \int_{\Omega_t} q(t, \mathbf{x}_t) d\mathbf{x}_t = \int_{\Omega_t} \partial_t q + \operatorname{div}(q\mathbf{u}) d\mathbf{x}_t. \quad (3.1)$$

Důkaz. Nechť χ_0 je hladká "klobouková" funkce nulová mimo Ω_0 a "skoro jednotková" uvnitř Ω_0 :

$$\chi_0(\mathbf{x}) = B(\operatorname{dist}(\mathbf{x}, \partial\Omega_0)),$$

kde vzdálenost dist je kladná uvnitř Ω_0 a záporná vně. Funkce B je nulová na $(-\infty, 0)$, $B = 1$ na (ϵ, ∞) , a je hladká a rostoucí na $(0, \epsilon)$. Tuto funkci necháme "unášet" rychlostním polem \mathbf{u} , takže se v čase t zdeformuje:

$$\chi(t, \mathbf{X}(t, \mathbf{x}_0)) = \chi_0(\mathbf{x}_0).$$

Pro materiálovou derivaci funkce $\chi(t, \mathbf{X})$ platí:

$$\frac{d}{dt} \chi(t, \mathbf{X}) = \frac{d}{dt} \chi(t, \mathbf{X}(t, \mathbf{x}_0)) = \partial_t \chi(t, \mathbf{X}) + \sum_i \partial_{X_i} \chi(t, \mathbf{X}) \partial_t X_i(t, \mathbf{x}_0) = \partial_t \chi + \mathbf{u} \cdot \nabla \chi$$

Nyní spočítáme :

$$\frac{d}{dt} \int_{\Omega_t} q\chi d\mathbf{x} = \frac{d}{dt} \int_{\mathbb{R}^3} q\chi d\mathbf{x} = \int_{\mathbb{R}^3} (\partial_t q)\chi + q(\partial_t \chi) d\mathbf{x} = \int_{\mathbb{R}^3} (\partial_t q)\chi - q(\mathbf{u} \cdot \nabla \chi) d\mathbf{x} = \int_{\Omega_t} [\partial_t q + \operatorname{div}(q\mathbf{u})] \chi d\mathbf{x}$$

Klobouková funkce χ může být libovolně blízko *charakteristické funkci* oblasti Ω_t z čehož plyne důsledek věty. □

Např. zákon zachování hmoty můžeme napsat jako:

$$\frac{d}{dt} \int_{\Omega_t} \rho(t, \mathbf{x}) d\mathbf{x} = \int_{\Omega_t} \partial_t \rho + \operatorname{div}(\rho\mathbf{u}) d\mathbf{x} = 0$$

kde \mathbf{u} je rychlost plynu a ρ jeho hustota.

A jelikož toto platí pro libovolnou Ω_t , pak pro hladké ρ a \mathbf{u} platí:

$$\partial_t \rho + \operatorname{div}(\rho\mathbf{u}) = 0$$

což je *rovnice kontinuity* pro hustotu stlačitelného plynu.

3.1 Eulerovy rovnice

Uvažujme materiál (tekutinu, nebo elastickou pevnou látku) s rychlostním polem u . Ze zákona zachování hybnosti ρu plyne použitím rovnice kontinuity:

$$\partial_t(\rho \mathbf{u}_i) + \operatorname{div}(\rho \mathbf{u}_i \mathbf{u}) = -\nabla P,$$

kde P je tlak, a jeho záporný gradient je hustota síly, která způsobuje změnu hybnosti podle 2. Newtonova zákona. Tato rovnice spolu s rovnicí kontinuity pro plyn:

$$\partial_t \rho + \operatorname{div}(\rho \mathbf{u}) = 0$$

tvoří systém tzv. Eulerových rovnic popisujících proudění neviskózní stlačitelné tekutiny.

4 Odvození rovnice vedení tepla

Rovnice kontinuity platí za předpokladu, že "pohyb veličiny" q je způsoben unášením v rychlostním poli u . Přirozená interpretace je, že se jedná o rychlostní pole média, např. tekutiny. To ovšem obecně neplatí. Například pro koncentraci soli v roztoku platí také zákon zachování a sůl se pohybuje i v (makroskopicky) stacionárním objemu vody pomocí difúze. Je tedy třeba u interpretovat jinak.

Definujeme plošný tok \mathbf{j} veličiny q jako množství veličiny, které projde jednotkovou elementární plochou za jednotku času. Tedy uvažujeme nekonečně malou plošku ΔS v bodě \mathbf{x} s normálou $\mathbf{n} = \mathbf{e}_i$ (bázový vektor) a nekonečně malou změnu času Δt . Pokud mezi časy t a $t + \Delta t$ projde skrz ΔS množství ΔQ veličiny q , platí

$$j_i(t, \mathbf{x}) = \frac{\Delta Q}{\Delta S \Delta t}$$

Pro pevnou oblast Ω je pokles množství veličiny q v Ω roven celkovému toku veličiny ven z Ω přes její hranici:

$$-\frac{d}{dt} \int_{\Omega} q \, d\mathbf{x} = \int_{\partial\Omega} \mathbf{j} \cdot \mathbf{n} \, dS = \int_{\Omega} \operatorname{div} \mathbf{j}$$

Odtud dostaneme bodovou formu obecné rovnice kontinuity:

$$\partial_t q + \operatorname{div} \mathbf{j} = 0.$$

Ke stejnému výsledku dojdeme pokud použijeme Reynoldsovu větu pro rychlostní pole $\mathbf{u} = \mathbf{j}/q$.

Tok $\mathbf{j}(t, \mathbf{x})$ je obecně nějakou funkcí závislou na lokálním chování veličiny q na okolí bodu (t, \mathbf{x}) , může tedy záviset na q , ∇q , na $\partial_t q$ a případně na vyšších prostorových a časových derivacích. Může také záviset na nějakých dalších veličinách, jako například na rychlosti média, viz. případ $\mathbf{j} = q\mathbf{u}$.

Nyní uvažujme specuálně zákon zachování pro energii pevného tělesa. Energie elementárního objemu ΔV je dána jeho teplotou jako:

$$\Delta E = C \rho T \Delta V$$

kde C [$J/K/kg$] je tepelná kapacita a T [K] je teplota. Teplený tok \mathbf{j} [W/m^2] je v nejjednodušší podobě dán Fourierovým zákonem:

$$\mathbf{j} = -k \nabla T$$

přičemž tepelná vodivost k [$W/m/K$] může být případně funkcí teploty $k(T)$. Dostáváme tak rovnici vedení tepla:

$$\partial_t(C\rho T) - \operatorname{div}(k\nabla T) = 0$$

Pokud budou v materiálu nějaké objemové tepelné zdroje f [W/m^3] dostaneme:

$$\partial_t(C\rho T) - \operatorname{div}(k\nabla T) = f$$

Pokud by se jednalo o vedení tepla v kapalině, musíme do \mathbf{j} zahrnout i transport kapalinou:

$$\partial_t(C\rho T) + \operatorname{div}(C\rho T\mathbf{u}) - \operatorname{div}(k\nabla T) = f.$$

5 Transportní procesy

Rovnice vedení tepla odvozená v předchozí kapitole je jedním z příkladů transportních procesů, které popisují transport nějaké veličiny a jsou odvozené ze zákona jejího zachování. Uvedeme pár dalších příkladů.

5.1 Proudění v porézním prostředí

Uvažujeme porézní prostředí, kde podíl pórů v referenčním objemu je ν [–]. Tuto bezrozměrnou veličinu nazýváme porozita. Podíl tekutiny (vody) v referenčním objemu θ [–] se nazývá saturace (opět bezrozměrná). Saturace se pohybuje od nějaké minimální (reziduální) saturace θ_r po saturovaný podíl tekutiny θ_s obvykle rovný porozitě ν . Pro tekutinu se zachovává její hmota, resp. hustota v prostoru $\rho_V = \rho\theta$, kde ρ je hustota tekutiny. V nejjednodušším případě uvažujeme nestlačitelnou kapalinu, plně saturované porézní prostředí a uvažujeme malé tlaky. V tom případě je ρ i θ konstanta. Pak z obecné rovnice kontinuity dostaneme:

$$-\operatorname{div} \mathbf{j} = f,$$

kde \mathbf{j} [$kg/m^2/s$] je hustota toku tekutiny a f [$kg/m^3/s$] je hustota objemových zdrojů tekutiny. Podobně jako v případě tepla je nejjednodušší vztah pro \mathbf{j} dán gradientem tlaku p [Pa] = [kgm^2/s] pomocí tzv. Darcyho zákona:

$$\mathbf{j} = -\rho k \mathbb{K} \nabla p.$$

Zde $k = \kappa/\mu$ je hydraulická vodivost daná permeabilitou κ [m^2], která je vlastností porézního média, a viskozitou μ [$Pa.s$], která je vlastností tekutiny. Tenzor \mathbb{K} je jednotkový v případě izotropního prostředí, ale v případě anisotropního prostředí je to obecně symetrický pozitivně definitní tenzor. Pokud má porézní materiál nějak orientované mikroskopické kapiláry, bude v jednom směru mít větší vodivost než ve směrech kolmých. Obecně může mít materiál tři různé vodivosti ve třech různých směrech k_x , k_y , k_z a nakonec tento materiál může být libovolně natočen v prostoru pomocí matice rotace Q :

$$\mathbb{K} = Q^T \mathbb{D} Q, \quad \mathbb{D} = \begin{pmatrix} k_x & 0 & 0 \\ 0 & k_y & 0 \\ 0 & 0 & k_z \end{pmatrix}$$

Vodivosti v hlavních směrech jsou vlastní čísla matice \mathbb{K} , musí být kladné. Matice rotace Q je tvořena (ortogonálními) vlastními vektory. Zde máme příklad anisotropie hydraulické vodivosti. Podobně existují materiály s anisotropní tepelnou vodivostí, nebo anisotropní pevností etc.

Dále můžeme uvažovat stlačitelnou tekutinu, resp. stlačitelný materiál okolo pórů. Pro použití rovnice kontinuity potřebujeme spočítat derivaci hustoty hmoty tekutiny v prostoru podle času:

$$\partial_t \rho_V = \left(\frac{\partial \rho}{\partial p} + \frac{\partial \theta}{\partial p} \right) \partial_t p = S \partial_t p.$$

Velichina S [$kg/m^3/Pa$] se nazývá storativita a zahrnuje jak stlačitelnost tekutiny $\partial_p \rho$ tak stlačitelnost prostředí $\partial_p \theta$. Pro nasycené, stlačitelné porézní prostředí tedy máme rovnici:

$$S \partial_t p - \operatorname{div}(\rho k \mathbb{K} \nabla p) = f$$

Pro nenasyčené prostředí pak dostáváme záporné (sací) tlaky p a pro ně saturaci $\theta_r \leq \theta(p) \leq \theta_s$, která je funkcí tlaku. Navíc i vodivost k , klesá s klesajícím nasycením, je tedy $k(\theta)$ funkcí saturace. Dohromady dostaneme tzv. Richardsovu rovnici:

$$\partial_t \theta(p) - \operatorname{div}(\rho k(\theta(p)) \mathbb{K} \nabla p) = f$$

kde funkce $\theta(p)$ a $k(\theta)$ jsou obecně nelineární a dostáváme tak nelineární parciální diferenciální rovnici.

5.2 Transport chemických látek

$$\partial_t(\rho_i c_i) + \operatorname{div}(\rho_i c_i \frac{\mathbf{u}}{\nu}) - \operatorname{div}(\mathbb{D} \nabla c_i) = r_{ij}. \quad (5.1)$$

6 Vlnová rovnice (akustika)

Odvodíme rovnici pro kmitání struny. Stav struny v čase t a poloze x je dán výchylkou struny $u(t, x)$. Pro zjednodušení si představujeme, že struna může kmitat jen v jednom směru. Na element daný intervalem (a, b) působí síly v koncových bodech:

$$\mathbf{F}(t, a) = -T(t, a) \mathbf{t}(t, a), \quad \mathbf{F}(t, b) = T(t, b) \mathbf{t}(t, b)$$

kde T je napětí ve struně a \mathbf{t} je tečný vektor $\mathbf{t}(t, x) = (1, \partial_x u(t, x))$. Jelikož v horizontálním směru se struna nepohybuje musí být horizontální složka součtu sil rovna nule:

$$T(t, b) - T(t, a) = 0$$

a jelikož jsme body a a b volili libovolně, je napětí ve struně nezávislé na poloze: $T(t, x) = T(t)$. Proto pro vertikální složku síly platí

$$F_y = F_y(t, a) + F_y(t, b) = T(t)(\partial_x u(t, b) - \partial_x u(t, a))$$

Nyní použijeme 2. Newtonův zákon:

$$\frac{d}{dt} \int_a^b \rho(x) \partial_t u(t, x) dx = F_y(t, x) = T(t)(\partial_x u(t, b) - \partial_x u(t, a)) = T(t) \int_a^b \partial_{xx} u(t, x) dx$$

A jelikož a a b jsou libovolné, dostáváme bodovou rovnici:

$$\rho(x) \partial_{tt} u(t, x) = T(t) \partial_{xx} u(t, x)$$

Pokud předpokládáme konstantní hustotu $\rho(x) = \rho_0$ a zanedbáme změnu napětí struny při malé výchylce $T(t) = T_0$ dostaneme vlnovou rovnici ve tvaru:

$$\partial_{tt} u(t, x) = c^2 \partial_{xx} u(t, x)$$

kde

$$c = \sqrt{\frac{T_0}{\rho_0}}$$

je rychlost šíření vlny. Mírně komplikovanější je odvození vlnové rovnice pro změny (akustického) tlaku v prostoru:

$$\partial_{tt} p(t, \mathbf{x}) = c^2 \Delta p(t, \mathbf{x})$$

kde pro rychlost zvuku c platí:

$$c = \sqrt{\frac{B}{\rho_0}}, \quad B = \rho_0 \frac{\partial P}{\partial \rho}$$

přičemž B je objemová stlačitelnost při adiabatické expanzi. Pro vzduch máme $B = 1.45 \times 10^5 \text{ Pa}$ a hustotu $\rho_0 = 1.2 \text{ kg/m}^3$ a dostáváme rychlost zvuku:

$$c = 347 \sqrt{\frac{\text{kg} \cdot \text{m} / \text{s}^2 / \text{m}^2}{\text{kg} / \text{m}^3}} = 1251 \text{ km/h}$$

Tabulková hodnota je 340 m/s .

7 Mechanika

...

8 Elektromagnetismus

9 Klasifikace PDR

9.1 Eliptické rovnice

Základním příkladem je Laplaceova rovnice:

$$\Delta u(\mathbf{x}) = 0$$

respektive Poissonova rovnice

$$\Delta u(\mathbf{x}) = f(\mathbf{x}).$$

Dalšími příklady je stacionární rovnice vedení tepla:

$$\operatorname{div}(k\nabla T(\mathbf{x})) = f(\mathbf{x}),$$

resp. stacionární rovnice Darcyho proudění:

$$\operatorname{div}(\mathbb{K}\nabla p(\mathbf{x})) = f(\mathbf{x}).$$

Obecná rovnici druhého řádu:

$$\operatorname{div}(\mathbb{A}\nabla u(\mathbf{x})) + \mathbf{b}\nabla u(\mathbf{x}) + cu(\mathbf{x}) = f(\mathbf{x})$$

je eliptická, pokud \mathbb{A} je symetrická pozitivně definitní matice.

Pro eliptické rovnice platí (za jistých omezeních pro \mathbf{b} a c) tzv. princip maxima. Pokud u je řešením eliptické rovnice na oblasti Ω pak

$$\max_{\mathbf{x} \in \Omega} u(\mathbf{x}) \leq \max_{\mathbf{x} \in \partial\Omega} u(\mathbf{x}).$$

Podobně pro minimum:

$$\min_{\mathbf{x} \in \Omega} u(\mathbf{x}) \geq \min_{\mathbf{x} \in \partial\Omega} u(\mathbf{x}).$$

9.2 Parabolické rovnice rovnice

Příkladem je nestacionární rovnice vedení tepla:

$$\partial_t T - \operatorname{div}(k\nabla T) = f$$

Vlastnosti řešení:

- I zde platí princip maxima vzhledem k okrajové podmínce.
- Pokles řešení v čase:

$$\max_{\mathbf{x} \in \Omega} u(t, \mathbf{x}) \leq \max_{\mathbf{x} \in \Omega} u(s, \mathbf{x}), \quad \text{pro } t \geq s.$$

Rovnice "zhlazuje" počáteční podmínku. Nekonečná rychlost šíření změn.

9.3 Hyperbolické rovnice

Příklad je vlnová rovnice. Eulerovy rovnice. Neplatí princip maxima. Kvalitativní vlastnosti řešení:

- Konečná rychlost šíření (vln).
- Reverzibilní v čase ($u(-t, \mathbf{x})$ je též řešením).
- Nezhazuje.
- Nesplňuje princip maxima, řešení se může akumulovat v bodě (náraz vlny na pobřeží).

10 Odvození slabého řešení

11 Úvod do funkcionální analýzy

V této kapitole zavedeme prostory funkcí vhodné při studiu slabých řešení. Teorie těchto abstraktních pojmů je poměrně obsáhlá, pro zájemce o hlubší poznatky odkazují na knihu [2].

V dalším textu bude Ω symbol pro otevřenou souvislou množinu v \mathbb{R} , \mathbb{R}^2 nebo \mathbb{R}^3 . Připomínáme, že otevřená souvislá množina se zkráceně nazývá oblast. Pro zjednodušení některých úvah také budeme předpokládat, že Ω je omezená. Hranici Ω budeme značit symbolem $\partial\Omega$ a uzávěr symbolem $\overline{\Omega} := \Omega \cup \partial\Omega$.

Než přistoupíme k zavedení obecných pojmů, objasníme je na příkladu spojitých funkcí. Nechť $C(\overline{\Omega})$ značí vektorový prostor všech spojitých funkcí na $\overline{\Omega}$. Pro funkce $u, v \in C(\overline{\Omega})$ definujeme následující operace:

Definice 11.1. Skalární součin funkcí $u, v \in C(\overline{\Omega})$ je (reálné) číslo

$$(u, v) := \int_{\Omega} u(x)v(x) \, dx.$$

Norma funkce $u \in C(\overline{\Omega})$ je číslo

$$\|u\|_2 := \sqrt{(u, u)} = \sqrt{\int_{\Omega} u^2(x) \, dx}.$$

Skalární součin a normu funkce lze zavést různými způsoby, je však důležité, aby byly splněny jisté vlastnosti, které budou později upřesněny. Na množině spojitých funkcí lze například definovat následující normu:

$$\|u\|_{\infty} := \max_{x \in \overline{\Omega}} |u(x)|.$$

Příklad 11.2. Uvažujme funkci

$$u(x) := \begin{cases} 10 \sin(1000\pi x) & \text{pro } x \in [0, \frac{1}{1000}] \\ 0 & \text{jinak} \end{cases}.$$

Snadno lze spočítat:

$$\begin{aligned} \|u\|_2 &= \sqrt{\int_0^{1/1000} 100 \sin^2(1000\pi x) \, dx} \\ &= \sqrt{100 \left[\frac{x}{2} - \frac{\sin(2000\pi x)}{4000\pi} \right]_{x=0}^{1/1000}} = \frac{1}{2\sqrt{5}} \doteq 0,224, \\ \|u\|_{\infty} &= \max_{x \in [0, 1/1000]} |10 \sin(1000\pi x)| = 10. \end{aligned}$$

Norma $\|\cdot\|_{\infty}$ se zdá být v jistém smyslu přirozenější, neboť měří maximální odchylku hodnot dvou spojitých funkcí. Přesto existují důvody, proč je vhodné používat normu $\|\cdot\|_2$. Předně, $\|\cdot\|_2$ byla zavedena pomocí skalárního součinu. Skalární součin hraje v některých úlohách důležitou roli. Lze ukázat, že na množině spojitých funkcí nelze zavést skalární součin s rozumnými vlastnostmi, který by indukoval normu $\|\cdot\|_{\infty}$. Dalším důvodem je, že v mnoha aplikacích nevystačíme se spojitými funkcemi. Není snadné rozšířit normu $\|\cdot\|_{\infty}$ na obecnější třídu funkcí, zatímco rozšíření normy $\|\cdot\|_2$ na dostatečně obecnou třídu funkcí je velmi jednoduché a vede přirozeně k vytvoření tzv. prostoru L^2 .

11.1 Normované lineární prostory

Pojem norma lze zavést pro různé typy objektů. Vždy očekáváme, že budou splněny některé vlastnosti, např. že norma nulového prvku bude 0 a norma nenulového prvku bude kladná. Proto definujeme abstraktní pojem norma pomocí charakteristických vlastností.

Definice 11.3. Nechť X je vektorový prostor. Funkce $\|\cdot\| : X \rightarrow \mathbb{R}$ se nazývá norma na X , pokud $\forall x, y \in X, \alpha \in \mathbb{R}$:

- (i) $\|x\| = 0 \Leftrightarrow x = \vec{0}$,
- (ii) $\|\alpha x\| = |\alpha| \|x\|$,
- (iii) $\|x + y\| \leq \|x\| + \|y\|$.

Je-li na vektorovém prostoru X definována norma, nazývá se X normovaný lineární prostor.

Z výše uvedených vlastností vyplývá, že norma je nezáporná funkce.

Příklad 11.4. Příklady normovaných lineárních prostorů:

- Množina \mathbb{R} s absolutní hodnotou $\|x\| := |x|$;
- $\mathbb{R}^n, n \in \mathbb{N}$, s normou $\|(x_1, \dots, x_n)\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}, p \in [1, \infty)$, nebo s normou $\|(x_1, \dots, x_n)\|_{\infty} := \max_{i=1, \dots, n} |x_i|$;

- $C(\overline{\Omega})$ s normou $\|\cdot\|_2$;
- $C(\overline{\Omega})$ s normou $\|\cdot\|_\infty$;
- $C^1(\overline{\Omega}) := \{f \in C(\overline{\Omega}); \forall i = 1, \dots, n \frac{\partial f}{\partial x_i} \in C(\overline{\Omega})\}$ s normou $\|f\|_{C^1(\overline{\Omega})} := \|f\|_{\infty, \overline{\Omega}} + \sum_{i=1}^n \|\frac{\partial f}{\partial x_i}\|_{\infty, \overline{\Omega}}$.

Speciálně zde zmíníme ještě normy na prostoru matic.

Definice 11.5. Necht $\|\cdot\|_X$ značí normu na \mathbb{R}^n a $\|\cdot\|_Y$ normu na \mathbb{R}^m . Generovaná norma na prostoru matic $\mathbb{R}^{m \times n}$ je definována vztahem

$$\|\mathbf{A}\|_{XY} := \max_{\vec{x} \in \mathbb{R}^n \setminus \{0\}} \frac{\|\mathbf{A}\vec{x}\|_Y}{\|\vec{x}\|_X} = \max_{\vec{x} \in \mathbb{R}^n, \|\vec{x}\|_X=1} \|\mathbf{A}\vec{x}\|_Y.$$

Generované normy mají následující vlastnosti:

$$\|\mathbf{AB}\|_{XY} \leq \|\mathbf{A}\|_{XY} \|\mathbf{B}\|_{XY}, \quad \rho(\mathbf{A}) \leq \|\mathbf{A}\|_{XY}, \quad \|\mathbf{I}\|_{XY} = 1.$$

Příklad 11.6. Příklady generovaných maticových norem:

- $\|\mathbf{A}\|_1 := \max_{\|\vec{x}\|_1=1} \|\mathbf{A}\vec{x}\|_1 = \max_{j=1, \dots, n} \sum_{i=1}^m |a_{ij}|$;
- $\|\mathbf{A}\|_2 := \max_{\|\vec{x}\|_2=1} \|\mathbf{A}\vec{x}\|_2 = \sqrt{\rho(\mathbf{A}^\top \mathbf{A})}$;
- $\|\mathbf{A}\|_\infty := \max_{\|\vec{x}\|_\infty=1} \|\mathbf{A}\vec{x}\|_\infty = \max_{i=1, \dots, m} \sum_{j=1}^n |a_{ij}|$.

Kromě generovaných norem existuje řada dalších maticových norem. Často používaná je tzv. Frobeniova norma

$$\|\mathbf{A}\|_F := \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}.$$

Dá se ukázat, že Frobeniova norma není generovaná, přesto však je multiplikativní:

$$\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_F.$$

11.1.1 Množiny v normovaném lineárním prostoru

Podobně jako u euklidovské vzdálenosti v \mathbb{R}^n , lze definovat pojmy jako koule, okolí nebo otevřená množina pomocí normy.

Definice 11.7. Necht X je normovaný lineární prostor s normou $\|\cdot\|$.

- Koule se středem $x \in X$ a poloměrem $r > 0$ je množina

$$B_r(x) := \{y \in X; \|x - y\| < r\}.$$

- Množinu $O \subset X$ nazveme okolím bodu x , jestliže existuje poloměr $r > 0$, takže O obsahuje kouli $B_r(x)$.
- Je-li O okolí bodu x , pak množinu $O \setminus \{x\}$ nazýváme prstencové okolí bodu x .
- Množina M se nazývá otevřená, pokud pro každý bod $x \in M$ existuje koule se středem x , která leží v M .
- Množina se nazývá uzavřená, pokud její doplněk v X je otevřený.

Příklad 11.8. Koule v prostoru \mathbb{R}^2 se středem v počátku souřadné soustavy má tvar

- čtverce, jehož vrcholy leží na souřadných osách a těžiště v počátku, uvažujeme-li normu $\|\cdot\|_1$;
- kruhu se středem v počátku, uvažujeme-li euklidovskou normu $\|\cdot\|_2$;
- čtverce, jehož strany jsou rovnoběžné se souřadnými osami a těžiště leží v počátku, uvažujeme-li maximovou normu $\|\cdot\|_\infty$.

Definice 11.9. Necht X je normovaný lineární prostor, $x \in X$ a $M \subset M$.

- Bod x je vnitřním bodem množiny M , pokud existuje poloměr $r > 0$ takový, že $B_r(x) \subset M$. Množinu všech vnitřních bodů M budeme značit $\text{Int } M$.
- Bod x je hraničním bodem množiny M , pokud každé okolí x obsahuje alespoň jeden bod z M a alespoň jeden bod z $X \setminus M$. Množina všech hraničních bodů M se nazývá hranice M a značí se ∂M .
- Uzávěr množiny M je množina $\overline{M} := M \cup \partial M$.
- Bod x je hromadným bodem množiny M , pokud každé jeho prstencové okolí obsahuje nějaký bod M . Množinu všech hromadných bodů M budeme značit $\text{Hr } M$.
- Bod x je izolovaný bod množiny M , pokud $x \in M$, ale x není hromadným bodem M . Množinu všech izolovaných bodů M budeme značit $\text{Iz } M$.

Mezi právě definovanými množinami platí mnoho vztahů. Např.:

$$\begin{aligned}\text{Int } M &\subset M \subset \overline{M}, & \text{Int } M \cap \partial M &= \emptyset, \\ \overline{M} &= \text{Hr } M \cup \text{Iz } M, & \text{Hr } M \cap \text{Iz } M &= \emptyset, \\ \text{Iz } M &\subset \partial M, & \text{Int } M &\subset \text{Hr } M.\end{aligned}$$

11.1.2 Konvergence

Definice 11.10. Necht X je normovaný lineární prostor. Řekneme, že posloupnost $\{x_n\}_{n \in \mathbb{N}}$ prvků z X konverguje k $x \in X$ v normě, jestliže

$$\lim_{n \rightarrow \infty} \|x_n - x\| = 0.$$

Říkáme, že x je limita posloupnosti $\{x_n\}$ a píšeme

$$x = \lim_{n \rightarrow \infty} x_n \text{ v } X, \text{ nebo } x_n \rightarrow x \text{ v } X.$$

Pro limitu v normovaném lineárním prostoru platí obdobná tvrzení jako pro limitu v \mathbb{R}^n známá ze základních kurzů matematiky. Např. každá posloupnost má nejvýše jednu limitu. Pokud posloupnost spojitých funkcí $\{u_n\}$ konverguje k u v normě $\|\cdot\|_2$ nebo $\|\cdot\|_\infty$, pak prvek u se skoro všude shoduje s bodovou limitou, tj.

$$\left(\lim_{n \rightarrow \infty} u_n\right)(x) = \lim_{n \rightarrow \infty} (u_n(x)).$$

Pro zjišťování konvergence posloupnosti funkcí je tedy vhodné nejprve zjistit, zda existuje bodová limita.

Příklad 11.11. Uvažujme posloupnost funkcí $\{u_n\}$,

$$u_n(x) := \begin{cases} 10 \sin(n\pi x) & \text{pro } x \in [0, \frac{1}{n}] \\ 0 & \text{jinak} \end{cases}$$

v prostoru $C([0, 1])$. Pro ověření, zda má daná posloupnost limitu, nejprve potřebujeme vhodného “kandidáta”. Spočteme proto nejprve bodovou limitu. Zřejmě $\lim u_n(0) = 0$. Je-li $x \in (0, 1]$, pak lze najít číslo $n_0 \in \mathbb{N}$ takové, že $x > \frac{1}{n_0}$, takže pro $n \geq n_0$ platí $u_n(x) = 0$, a proto musí být $\lim u_n(x) = 0$. Bodová limita posloupnosti je tedy nulová funkce. Lze ukázat, že

$$\|u_n - 0\|_2 = \sqrt{\frac{50}{n}}, \text{ takže } \lim \|u_n - 0\|_2 = 0,$$

a tedy

$$\lim u_n = 0 \text{ v prostoru } C([0, 1]) \text{ s normou } \|\cdot\|_2.$$

Dále platí

$$\|u_n - 0\|_\infty = 10,$$

z čehož plyne, že v prostoru $C([0, 1])$ s normou $\|\cdot\|_\infty$ není nulová funkce limitou posloupnosti $\{u_n\}$ (ve skutečnosti posloupnost není v tomto prostoru konvergentní).

Uvedený příklad poukazuje na to, že existence limity v metrickém prostoru závisí na tom, jakou uvažujeme metriku.

Definice 11.12. Necht $\| \cdot \|_A$ a $\| \cdot \|_B$ jsou normy ve vektorovém prostoru X . Jestliže existují konstanty $\alpha, \beta > 0$ takové, že pro každé $x \in X$ platí

$$\alpha \|x\|_A \leq \|x\|_B \leq \beta \|x\|_A,$$

pak říkáme, že $\| \cdot \|_A$ a $\| \cdot \|_B$ jsou na X ekvivalentní.

Jsou-li normy $\| \cdot \|_A$ a $\| \cdot \|_B$ ekvivalentní, pak platí

$$x_n \rightarrow x \text{ v } (X, \| \cdot \|_A) \Leftrightarrow x_n \rightarrow x \text{ v } (X, \| \cdot \|_B).$$

Ekvivalentní normy také generují stejné otevřené a uzavřené množiny. Normy $\| \cdot \|_p$ a $\| \cdot \|_\infty$ obecně nejsou ekvivalentní.

11.2 Prostor $L^2(\Omega)$

Definice 11.13. Prostorem $L^2(\Omega)$ rozumíme množinu funkcí

$$L^2(\Omega) := \left\{ u : \Omega \rightarrow \mathbb{R}; \left| \int_{\Omega} u(x) dx \right| < \infty, \|u\|_2 < \infty \right\}.$$

Spolu s normou $\| \cdot \|_2$ tvoří $L^2(\Omega)$ normovaný lineární prostor.

Z definice plyne, že každá spojitá funkce v $\overline{\Omega}$ patří do prostorů $L^2(\Omega)$. Do těchto prostorů ovšem patří i mnoho dalších funkcí, které mohou být nespojité nebo neomezené. Poznamenejme, že pro správnost některých tvrzení je třeba uvažovat integrály v Definici 11.13 v tzv. Lebesgueově smyslu.

Příklad 11.14. Uvažujme funkce

$$u(x) := \frac{1}{\sqrt{x}} \quad a \quad v(x) := \frac{1}{\sqrt[3]{x}}$$

na intervalu $\Omega := (0, 1)$. Platí:

$$\begin{aligned} \|u\|_2 &= \sqrt{\int_0^1 \frac{1}{x} dx} = +\infty, \\ \|v\|_2^2 &= \int_0^1 |v(x)|^2 dx = \int_0^1 \frac{1}{\sqrt[3]{x^2}} dx = 3. \end{aligned}$$

Proto $u \notin L^2(0, 1)$ a $v \in L^2(0, 1)$.

Příklad 11.15. Funkce

$$\operatorname{sgn} x := \begin{cases} -1 & \text{pro } x < 0 \\ 0 & \text{pro } x = 0 \\ 1 & \text{pro } x > 0 \end{cases}$$

je prvkem prostoru $L^2(-1, 1)$, neboť

$$\int_{-1}^1 |\operatorname{sgn} x|^2 dx = \int_{-1}^0 |\operatorname{sgn} x|^2 dx + \int_0^1 |\operatorname{sgn} x|^2 dx = \int_{-1}^0 1 dx + \int_0^1 1 dx = 2,$$

a tedy $\|\operatorname{sgn} x\|_2 = \sqrt{2}$. Podobně funkce

$$u(x) := \begin{cases} 0 & \text{pro } x \neq 0 \\ 10 & \text{pro } x = 0 \end{cases}$$

patří do $L^2(-1, 1)$ a její norma je $\|u\|_2 = 0$. Vidíme, že norma nezávisí na hodnotě funkce v bodě $x = 0$. Dokonce není nutné, aby byla funkce v bodě 0 definována.

Do prostoru $L^2(\Omega)$ tedy patří i některé nespojité a neomezené funkce. Funkce, která je rovna nule všude až na hodnotu v jednom bodě, má nulovou normu a je v jistém smyslu ekvivalentní s nulovou funkcí. Obecněji postačí, když je funkce nulová všude v Ω mimo množinu míry nula. Mezi množiny s nulovou mírou patří např. všechny konečné a spočetné množiny.

Definice 11.16. *Nechť funkce $u, v \in L^2(\Omega)$ jsou si v oblasti Ω rovny skoro všude, tj. všude mimo množinu míry nula (kde se buď jejich hodnoty liší nebo některá z funkcí není definována). Pak řekneme, že u a v jsou v prostoru $L^2(\Omega)$ ekvivalentní. Píšeme $u = v$ v $L^2(\Omega)$.*

Funkce u a v jsou tedy v tomto prostoru pokládány za sobě rovné. Dvě funkce u, v ekvivalentní v prostoru $L^2(\Omega)$ jsou charakterizovány vlastností

$$\int_{\Omega} |u(x) - v(x)|^2 dx = 0.$$

V prostoru $L^2(\Omega)$ je zaveden skalární součin stejným způsobem jako v Definici 11.1:

$$(u, v) := \int_{\Omega} u(x)v(x) dx, \quad u, v \in L^2(\Omega).$$

To, že skalární součin je konečný pro libovolné $u, v \in L^2(\Omega)$, je důsledkem tzv. Schwarzovy nerovnosti:

$$\forall u, v \in L^2(\Omega) : |(u, v)| \leq \|u\|_2 \|v\|_2.$$

Na omezené oblasti Ω platí následující vztah mezi limitami.

Věta 11.17. *Nechť Ω je omezená oblast v \mathbb{R}^d , $d \in \{1, 2, 3\}$, a $\{u_n\}$ je posloupnost funkcí. Pak platí:*

- (i) *Jestliže $u_n \rightarrow u$ v $\|\cdot\|_{\infty}$, pak $u_n \rightarrow u$ také v $\|\cdot\|_2$.*
- (ii) *Jestliže $u_n \rightarrow u$ v $L^2(\Omega)$, pak $u_n(x) \rightarrow u(x)$ pro skoro všechna $x \in \Omega$.*

11.3 Prostory $H^1(\Omega)$

Pro funkce z prostoru $L^2(\Omega)$ lze zavést pojem derivace. Uvažujme nejprve funkci $u \in C^1([0, 1])$. Je-li $v \in C^1([0, 1])$, $v(0) = v(1) = 0$, pak díky pravidlu per partes platí

$$\int_0^1 u'(x)v(x) dx = [u(x)v(x)]_{x=0}^1 - \int_0^1 u(x)v'(x) dx = - \int_0^1 u(x)v'(x) dx.$$

Zde vidíme, že zatímco pro integrál nalevo je třeba, aby existovala u' , výraz na pravé straně je definován i pro $u \in L^1(0, 1)$. To vede k pojmu *zobecněná derivace*.

Definice 11.18. *Nechť $u \in L^2(\Omega)$. Funkce $g \in L^2(\Omega)$ se nazývá zobecněná parciální derivace funkce u podle i -té proměnné, pokud pro každé $v \in C^1(\bar{\Omega})$, $v|_{\partial\Omega} = 0$, platí:*

$$\int_{\Omega} g(x)v(x) dx = - \int_{\Omega} u(x) \frac{\partial v}{\partial x_i}(x) dx.$$

Píšeme $g = \frac{\partial u}{\partial x_i}$ v $L^2(\Omega)$.

Příklad 11.19. *Spočtěme zobecněnou derivaci funkce $u(x) := |x|$ na intervalu $(-1, 1)$. Pro $v \in C^1([-1, 1])$, $v(-1) = v(1) = 0$ platí:*

$$\begin{aligned} - \int_{-1}^1 |x|v'(x) dx &= - \int_{-1}^0 (-x)v'(x) dx - \int_0^1 xv'(x) dx \\ &= [xv(x)]_{x=-1}^0 - \int_{-1}^0 v(x) dx - [xv(x)]_{x=0}^1 + \int_0^1 v(x) dx = \int_{-1}^1 \operatorname{sgn} x v(x) dx. \end{aligned}$$

Je tedy $u' = \operatorname{sgn} v$ v $L^2(-1, 1)$.

Příklad 11.20. *Uvažujme funkci $u(x) = \operatorname{sgn} x$. Pro $v \in C^1([-1, 1])$, $v(-1) = v(1) = 0$ platí:*

$$\begin{aligned} - \int_{-1}^1 u(x)v'(x) dx &= - \int_{-1}^0 (-v'(x)) dx - \int_0^1 v'(x) dx \\ &= [v(x)]_{x=-1}^0 - [v(x)]_{x=0}^1 = 2v(0) = 2 \int_{-1}^1 \delta_0(x)v(x) dx, \end{aligned}$$

kde δ_0 se nazývá Diracova δ -funkce (ve skutečnosti to není funkce, ale tzv. distribuce). V jistém smyslu tedy platí $\operatorname{sgn}' = 2\delta_0$, nicméně $\delta_0 \notin L^2(\Omega)$.

Ne každá funkce z $L^2(\Omega)$ tedy má zobecněnou derivaci v $L^2(\Omega)$.

Definice 11.21. Prostor $H^1(\Omega)$ je množina

$$H^1(\Omega) := \left\{ u \in L^2(\Omega); \forall i = 1, \dots, n : \frac{\partial u}{\partial x_i} \in L^2(\Omega) \right\}.$$

Norma na tomto prostoru je definována výrazem

$$\|u\|_{1,2} := \left(\|u\|_2^2 + \sum_{i=1}^n \left\| \frac{\partial u}{\partial x_i} \right\|_2^2 \right)^{1/2}$$

a skalární součin výrazem

$$((u, v)) := (u, v) + \sum_{i=1}^n \left(\frac{\partial u}{\partial x_i}, \frac{\partial v}{\partial x_i} \right).$$

Poznamenejme, že skalární součin obecně je charakterizován následujícími vlastnostmi:

Definice 11.22. Nechť X je (reálný) vektorový prostor. Zobrazení $(\cdot, \cdot)_X : X \times X \rightarrow \mathbb{R}$ se nazývá skalární součin, pokud pro každé $x, y, z \in X$ a $\alpha \in \mathbb{R}$ platí

$$(i) \quad (x, x)_X \geq 0, \quad (x, x)_X = 0 \Leftrightarrow x = \vec{0},$$

$$(ii) \quad (x, y)_X = (y, x)_X,$$

$$(iii) \quad (\alpha x, y)_X = \alpha(x, y)_X,$$

$$(iv) \quad (x + y, z)_X = (x, z)_X + (y, z)_X.$$

Je-li na vektorovém prostoru definován skalární součin, nazývá se X prostor se skalárním součinem.

Skalární součin indukuje normu $\|x\|_X := \sqrt{(x, x)_X}$, pro niž platí tzv. *Cauchyova-Schwarzova nerovnost*:

$$\forall x, y \in X : |(x, y)_X| \leq \|x\|_X \cdot \|y\|_X.$$

Všechny uvedené vlastnosti jsou splněny pro skalární součiny v $L^2(\Omega)$ i $H^1(\Omega)$.

12 Variační (slabá) formulace okrajové úlohy

13 Galerkinova metoda

Na závěr ještě uvedeme příklad, jak lze formulovat okrajovou úlohu s nespojitou pravou stranou a její aproximaci.

Je dána okrajová úloha

$$-u'' + u = f \text{ v } (0, 1), \quad u(0) = u(1) = 0.$$

Je-li $f \in C[0, 1]$, pak má smysl hledat klasické řešení, tj. funkci $u \in C^2(0, 1) \cap C[0, 1]$ takovou, že uvedené rovnosti platí v celém intervalu $(0, 1)$. Pro méně regulární pravou stranu ale klasické řešení nemusí existovat. Ukážeme odvození definice tzv. slabého (zobecněného) řešení.

Předpokládejme, že u je klasické řešení. Pak pro každé $v \in V := \{v \in C^1[0, 1]; v(0) = v(1) = 0\}$ platí

$$(-u'' + u, v) = (f, v).$$

Integrací per partes dostaneme

$$\begin{aligned} (-u'' + u, v) &= \int_0^1 (-u''(x) + u(x))v(x) \, dx \\ &= [-u'(x)v(x)]_{x=0}^1 + \int_0^1 u'(x)v'(x) + u(x)v(x) \, dx = ((u, v)). \end{aligned}$$

Místo klasického řešení můžeme tedy hledat funkci u takovou, aby

$$((u, v)) = (f, v)$$

pro všechna $v \in V$. Protože ale V není úplný prostor v normě $H^1(0, 1)$, nehodí se pro definici zobecněného řešení u . Zúplněním V v normě $H^1(0, 1)$ dostaneme prostor

$$H_0^1(0, 1) := \{v \in H^1(0, 1); v(0) = v(1) = 0\}.$$

Slabé (zobecněné) řešení okrajové úlohy tedy lze definovat jako funkci $u \in H_0^1(0, 1)$, která splňuje

$$\forall v \in H_0^1(0, 1) : ((u, v)) = (f, v).$$

Všimněme si, že tato formulace okrajové úlohy má smysl pro $f \in L^2(0, 1)$.

Nechť $\{v_i\}_{i=1}^\infty$ je nějaká báze prostoru $H_0^1(0, 1)$. Galerkinova aproximace slabého řešení je definována jako funkce

$$u^n(x) := \sum_{i=1}^n \alpha_i^n v_i(x),$$

která splňuje

$$\forall j = 1, \dots, n : ((u^n, v_j)) = (f, v_j).$$

Dosazením za u^n dostaneme soustavu lineárních algebraických rovnic

$$\sum_{i=1}^n \alpha_i^n ((v_i, v_j)) = (f, v_j), \quad j = 1, \dots, n,$$

pro koeficienty $\vec{u} := (\alpha_1^n, \dots, \alpha_n^n)^\top$. Definujme-li matici $\mathbf{A} = (a_{ij})_{i,j=1}^n$, kde $a_{ij} := ((v_j, v_i))$, a vektor $\vec{b} = (b_i)_{i=1}^n$, kde $b_i := (f, v_i)$, pak lze tuto soustavu zapsat zkráceně

$$\mathbf{A}\vec{u} = \vec{b}.$$

Díky vlastnostem skalárního součinu je matice \mathbf{A} symetrická pozitivně definitní, soustava má proto pro libovolné $\vec{b} \in \mathbb{R}^n$ právě jedno řešení.

Lze také ukázat, že posloupnost $\{u^n\}$ je v jistém smyslu konvergentní a její limita je slabé řešení u .

14 Iterační metody řešení soustav lineárních rovnic

V této kapitole se budeme zabývat numerickým řešením soustavy

$$\mathbf{A}\vec{x} = \vec{b}.$$

Předpokládáme, že čtenáři je známa Gaussova eliminační metoda, která je příkladem tzv. přímých metod. Její výhodou je univerzálnost — metoda vyřeší v přesné aritmetice soustavu s libovolnou regulární maticí. Nevýhodou je její neefektivita pro velké matice a také to, že v průběhu výpočtu nemá uživatel žádnou informaci o výsledku. Pro úlohy s velkou řídkou maticí \mathbf{A} , se kterými se setkáváme v mnoha praktických problémech, nebo pro úlohy, kde matice není dána explicitně nebo je drahé ji sestavit, může být výhodou použít iterační metody. Tyto metody v zásadě používají jen násobení matic a v průběhu výpočtu postupně vylepšují aproximaci přesného řešení. Konvergence iteračního procesu může být asymptotická nebo v konečném počtu iterací.

Pro detaily ohledně odvození a dalších vlastností iteračních metod odkazujeme na knihu [1].

14.1 Klasické iterační metody

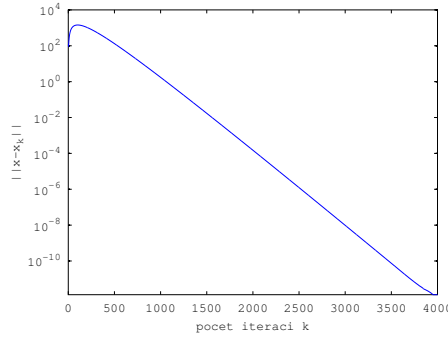
Klasické iterační metody jsou založeny na štěpení matice soustavy $\mathbf{A} = \mathbf{M} + \mathbf{N}$ takovém, že matice \mathbf{M} je regulární a snadno invertovatelná a \mathbf{M} a \mathbf{N} jsou zvoleny nějakým vhodným způsobem. Dosazením tohoto štěpení do vztahu $\mathbf{A}\vec{x} = \vec{b}$ dostáváme $(\mathbf{M} + \mathbf{N})\vec{x} = \vec{b}$ a odtud

$$\vec{x} = \mathbf{M}^{-1}(\vec{b} - \mathbf{N}\vec{x}) = \mathbf{M}^{-1}(\vec{b} + \mathbf{M}\vec{x} - \mathbf{A}\vec{x}) = \vec{x} + \mathbf{M}^{-1}(\vec{b} - \mathbf{A}\vec{x}).$$

Je-li dána počáteční aproximace řešení \vec{x}_0 , můžeme definovat iterační proces následovně:

$$\vec{x}_k = \vec{x}_{k-1} + \mathbf{M}^{-1}(\vec{b} - \mathbf{A}\vec{x}_{k-1}) = (\mathbf{I} - \mathbf{M}^{-1}\mathbf{A})\vec{x}_{k-1} + \mathbf{M}^{-1}\vec{b}.$$

¹Výraz $v(0)$, resp. $v(1)$ zde značí tzv. stopu funkce v .



Obrázek 1: Přejchodový jev u klasické iterační metody.

Lze ukázat, že pro chybu aproximace platí odhad

$$\frac{\|\vec{x} - \vec{x}_k\|}{\|\vec{x} - \vec{x}_0\|} \leq \|(\mathbf{I} - \mathbf{M}^{-1}\mathbf{A})^k\| \leq \|\mathbf{I} - \mathbf{M}^{-1}\mathbf{A}\|^k,$$

přičemž pro velká k je $\|(\mathbf{I} - \mathbf{M}^{-1}\mathbf{A})^k\| \approx \rho(\mathbf{I} - \mathbf{M}^{-1}\mathbf{A})^k$ (symbolem $\rho(\mathbf{A})$ značíme tzv. spektrální poloměr, který je definován jako $\max\{|\lambda|; \lambda \in \sigma(\mathbf{A})\}$). Vidíme tedy, že metody konvergují k přesnému řešení, pokud $\rho(\mathbf{I} - \mathbf{M}^{-1}\mathbf{A}) < 1$. I v případě, že tato podmínka je splněna, však může být $\|\mathbf{I} - \mathbf{M}^{-1}\mathbf{A}\|^k > 1$ a dochází pak k tzv. přechodovému jevu, kdy chyba aproximace nejprve roste a teprve pak začne klesat (viz obr. 1).

Příklady klasických iteračních metod. Následující metody jsou založeny na štěpení $\mathbf{A} = \mathbf{D} - \mathbf{L} - \mathbf{U}$, kde \mathbf{D} je hlavní diagonála, $-\mathbf{L}$ je striktně dolní trojúhelník matice \mathbf{A} a $-\mathbf{U}$ je striktně horní trojúhelník. Z rovnice

$$(\mathbf{D} - \mathbf{L} - \mathbf{U})\vec{x} = \vec{b}$$

pak lze odvodit jednotlivé metody.

Jacobiova metoda je definována iterací

$$\mathbf{D}\vec{x}_k = \mathbf{L}\vec{x}_{k-1} + \mathbf{U}\vec{x}_{k-1} + \vec{b}.$$

Rozepíšeme-li tento vzorec po složkách (x_i^k značí i -tou složku vektoru \vec{x}_k), dostaneme pro $i = 1, \dots, n$:

$$x_i^k = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1, j \neq i}^n a_{ij} x_j^{k-1} \right).$$

Nevýhodou této metody může být, že v průběhu výpočtu je třeba uchovávat dvě posobě jdoucí aproximace řešení \vec{x}_k, \vec{x}_{k-1} . Metoda **Gauss-Seidelova** se od předchozí liší v tom, že ihned využívá již spočtené složky vektoru \vec{x}_k , tj. po složkách počítá

$$x_i^k = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^k - \sum_{j=i+1}^n a_{ij} x_j^{k-1} \right).$$

Spočtené složky aproximace řešení je tedy možné ihned přepisovat. Maticově lze tuto iteraci zapsat jako

$$\mathbf{D}\vec{x}_k = \mathbf{L}\vec{x}_k + \mathbf{U}\vec{x}_{k-1} + \vec{b}.$$

Z Gauss-Seidelovy metody je odvozena **Superrelaxační metoda** (SOR, successive over-relaxation). Pracuje s relaxačním parametrem $\omega \in [0, 2]$ a je definována vztahem

$$\mathbf{D}\vec{x}_k = \omega(\mathbf{L}\vec{x}_k + \mathbf{U}\vec{x}_{k-1} + \vec{b}) + (1 - \omega)\mathbf{D}\vec{x}_{k-1},$$

tj. kombinuje Gauss-Seidelovu metodu s předchozí iterací.

Příklad 14.1. Uvažujme matici

$$\mathbf{A} = \begin{pmatrix} 0.01 & -0.4 \\ 0 & 0.01 \end{pmatrix}.$$

Konvergence Jacobiovy metody pro tuto matici závisí na vlastnostech matice

$$\mathbf{I} - \mathbf{D}^{-1}\mathbf{A} = \begin{pmatrix} 0 & -40 \\ 0 & 0 \end{pmatrix}.$$

Jelikož tato matice má pouze nulové vlastní číslo, její spektrální poloměr je 0, a tedy podmínka $\rho(\mathbf{I} - \mathbf{D}^{-1}\mathbf{A}) < 1$ je splněna, neboli metoda konverguje. Na druhé straně $\|\mathbf{I} - \mathbf{D}^{-1}\mathbf{A}\| = 40 > 1$, při výpočtu proto lze pozorovat přechodový jev.

14.2 Metody Krylovových podprostorů

Důležitá třída iteračních metod je založena na myšlence projektovat soustavu $\mathbf{A}\vec{x} = \vec{b}$ na posloupnost tzv. Krylovových prostorů a tím získávat postupně aproximace řešení.

Definice 14.2. Nechť $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\vec{v} \in \mathbb{R}^n$ a $k \leq n$. k -tým Krylovovým prostorem nazýváme podprostor

$$\mathcal{K}_k(\mathbf{A}, \vec{v}) := \langle \vec{v}, \mathbf{A}\vec{v}, \mathbf{A}^2\vec{v}, \dots, \mathbf{A}^{k-1}\vec{v} \rangle.$$

Metody, které zmíníme v následující části, mají společnou vlastnost tzv. projekčních metod, tj. hledají aproximace ve tvaru

$$\vec{x}_k \in \vec{x}_0 + \mathcal{S}_k, \quad \vec{r}_k \perp \mathcal{C}_k,$$

kde $\vec{r}_k := \vec{b} - \mathbf{A}\vec{x}_k$ je tzv. residuum a \mathcal{S}_k a \mathcal{C}_k jsou vhodné podprostory. Prostor \mathcal{S}_k je obvykle roven Krylovovu podprostoru $\mathcal{K}_k(\mathbf{A}, \vec{r}_0)$, ale jsou možné i jiné volby, např. $\mathbf{A}\mathcal{K}_k(\mathbf{A}, \vec{r}_0)$. Volbou prostoru \mathcal{C}_k lze docílit optimality aproximace řešení v tom smyslu, že chyba aproximace $\vec{x} - \vec{x}_k$ je v nějaké normě minimální. Pokud dimenze podprostorů $\mathcal{S}_k, \mathcal{C}_k$ roste, pak pro $k = n$ dostáváme $\mathcal{C}_n = \mathbb{R}^n$ a z podmínky $\vec{r}_k \perp \mathbb{R}^n$ plyne $\vec{r}_n = \vec{0}$, tedy $\vec{x}_n = \vec{x}$ je přesné řešení. Jinak řečeno, rostou-li dimenze prostorů $\mathcal{S}_k, \mathcal{C}_k$, pak projekční metody najdou řešení systému $\mathbf{A}\vec{x} = \vec{b}$ nejvýše v n krocích.

14.2.1 Metoda sdružených gradientů (CG)

Tato metoda (stručně ji budeme označovat symbolem CG — z anglického *Conjugate gradients*) je určena pro symetrické pozitivně definitní matice.

Definice 14.3. Matice $\mathbf{A} \in \mathbb{R}^{n \times n}$ je pozitivně definitní, pokud pro každý nenulový vektor $\vec{x} \in \mathbb{R}^n$ platí:

$$\mathbf{A}\vec{x} \cdot \vec{x} > 0.$$

Výraz

$$\|\vec{x}\|_{\mathbf{A}} := \sqrt{\vec{x} \cdot \mathbf{A}\vec{x}}$$

se nazývá energetická norma nebo také \mathbf{A} -norma. Řekneme, že vektory $\vec{u}, \vec{v} \in \mathbb{R}^n$ jsou navzájem \mathbf{A} -ortogonální, jestliže

$$\vec{u} \cdot \mathbf{A}\vec{v} = 0.$$

Aproximace řešení je v metodě CG konstruována podle vzorce

$$\vec{x}_k := \vec{x}_{k-1} + \gamma_{k-1}\vec{p}_{k-1},$$

kde \vec{p}_{k-1} je směrový vektor a γ_{k-1} je délka kroku. Tyto parametry se určí následujícím způsobem:

- \vec{p}_k volíme ve tvaru $\vec{p}_k := \vec{r}_k + \delta_k\vec{p}_{k-1}$ tak, aby byl \mathbf{A} -ortogonální na \vec{p}_{k-1} , tj. $\vec{p}_k \cdot \mathbf{A}\vec{p}_{k-1} = 0$. Toho docílíme pro

$$\delta_k := \frac{\vec{r}_k \cdot \vec{r}_k}{\vec{r}_{k-1} \cdot \vec{r}_{k-1}}.$$

- γ_{k-1} volíme takové, aby byla minimální energetická norma $\|\vec{x} - \vec{x}_k\|_{\mathbf{A}}$. To nastane právě tehdy, když

$$\gamma_{k-1} := \frac{\vec{r}_{k-1} \cdot \vec{r}_{k-1}}{\vec{p}_{k-1} \cdot \mathbf{A}\vec{p}_{k-1}}.$$

Na základě předchozích vztahů lze ukázat že CG patří mezi Krylovovské metody, neboť platí:

$$\vec{x}_k \in \vec{x}_0 + \mathcal{K}_k(\mathbf{A}, \vec{r}_0), \quad \vec{r}_k \perp \mathcal{K}_k(\mathbf{A}, \vec{r}_0).$$

Na metodu sdružených gradientů lze také nahlížet jako na metodu, která hledá minimum kvadratického funkcionálu $\frac{1}{2}\vec{x} \cdot \mathbf{A}\vec{x} - \vec{x} \cdot \vec{b}$. Následující algoritmus reprezentuje standardní implementaci metody CG.

Algoritmus A1 Metoda sdružených gradientů

```

input  $\mathbf{A}, \vec{b}, \vec{x}_0$ 
 $\vec{r}_0 := \vec{b} - \mathbf{A}\vec{x}_0$ 
 $\vec{p}_0 := \vec{r}_0$ 
for  $k = 1, 2, \dots$ 
     $\gamma_{k-1} := \frac{\vec{r}_{k-1} \cdot \vec{r}_{k-1}}{\vec{p}_{k-1} \cdot \mathbf{A}\vec{p}_{k-1}}$ 
     $\vec{x}_k := \vec{x}_{k-1} + \gamma_{k-1}\vec{p}_{k-1}$ 
     $\vec{r}_k := \vec{r}_{k-1} - \gamma_{k-1}\mathbf{A}\vec{p}_{k-1}$ 
     $\delta_k := \frac{\vec{r}_k \cdot \vec{r}_k}{\vec{r}_{k-1} \cdot \vec{r}_{k-1}}$ 
     $\vec{p}_k := \vec{r}_k + \delta_k\vec{p}_{k-1}$ 
end

```

Vidíme, že v každé iteraci je třeba provést 1 násobení matice \mathbf{A} s vektorem a v průběhu výpočtu je třeba uchovávat pouze 4 vektory. Metoda CG je tedy velmi efektivní zejména pro velké řídké matice. Je-li matice symetrická pozitivně definitní, pak v přesné aritmetice algoritmus nalezne řešení nejvýše po n iteracích. V praxi ovšem kvůli zaokrouhlovacím chybám dochází ke ztrátě \mathbf{A} -ortogonalit vektorů $\{\vec{p}_k\}$ (resp. ortogonalit vektorů $\{\vec{r}_k\}$), což způsobuje zpoždění konvergence, tedy že i po n krocích je $\vec{x}_n \neq \vec{x}$. Tento nedostatek se někdy odstraňuje tak, že se vektor \vec{r}_k ortogonalizuje proti všem předchozím $\{\vec{r}_i\}_{i=0}^{k-1}$ a proces ortogonalizace se zopakuje vícekrát (obvykle stačí dvakrát).

Nyní si uvedeme, co je známo o rychlosti konvergence metody CG. K tomu potřebujeme znát pojem *číslo podmíněnosti*.

Definice 14.4. *Nechť \mathbf{A} je symetrická pozitivně definitní matice. Číslo podmíněnosti matice \mathbf{A} je definováno předpisem*

$$\kappa(\mathbf{A}) := \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})},$$

kde $\lambda_{\max}(\mathbf{A})$, $\lambda_{\min}(\mathbf{A})$ značí největší, resp. nejmenší vlastní číslo matice \mathbf{A} .

Označíme-li $\vec{e}_k := \vec{x}_k - \vec{x}$ chybu k -té aproximace řešení, pak platí následující odhad chyby:

$$\frac{\|\vec{e}_k\|_{\mathbf{A}}}{\|\vec{e}_0\|_{\mathbf{A}}} \leq 2 \left(\frac{\sqrt{\kappa(\mathbf{A})} - 1}{\sqrt{\kappa(\mathbf{A})} + 1} \right)^k.$$

Všimněme si, že číslo v závorce v předchozí nerovnosti je vždy menší než 1. Pokud je $\kappa(\mathbf{A})$ blízké 1, pak odhad chyby říká, že chyba klesá velmi rychle. Pro špatně podmíněné matice (tj. je-li $\kappa(\mathbf{A})$ velké) je číslo v závorce blízké jedné a odhad často nadhodnocuje skutečnou velikost \mathbf{A} -normy chyby. Špatná podmíněnost matice přesto může mít za následek pomalou konvergenci metody. Tuto skutečnost lze řešit pomocí tzv. předpodmínění, které spočívá v tom, že původní soustava $\mathbf{A}\vec{x} = \vec{b}$ se nahradí ekvivalentní soustavou $\hat{\mathbf{A}}\hat{\vec{x}} = \hat{\vec{b}}$ s maticí $\hat{\mathbf{A}}$, která má menší číslo podmíněnosti než \mathbf{A} .

14.2.2 Zobecněná metoda minimálních reziduí (GMRES)

Metodu GMRES (generalized minimal residual method) lze charakterizovat ve smyslu projekčních metod pomocí vztahů

$$\vec{x}_k \in \vec{x}_0 + \mathcal{K}_k(\mathbf{A}, \vec{r}_0), \quad \vec{r}_k \perp \mathbf{A}\mathcal{K}_k(\mathbf{A}, \vec{r}_0).$$

Jak napovídá název, její vlastností je, že v každé iteraci minimalizuje normu rezidua $\|\vec{r}_k\|$. To vede na úlohu nejmenších čtverců, jejíž efektivní implementace je poměrně technicky obtížná. Proto zde její algoritmus neuvádíme. Nepříjemnou vlastností metody GMRES je, že produkuje posloupnost ortogonálních vektorů $\{\vec{v}_k\}$, které je třeba uchovávat, (říkáme, že metoda generuje dlouhé rekurence) a to klade vysoké nároky na paměť. Za tuto cenu ovšem metoda dokáže řešit soustavu s libovolnou regulární maticí.

Stejně jako u metody CG, vlivem zaokrouhlovacích chyb dochází ke zpomalení konvergence kvůli ztrátě ortogonalit systému $\{\vec{v}_k\}$. I u GMRES tedy obvykle provádíme vícenásobnou ortogonalizaci. Problém s pamětovou náročností se obvykle řeší pomocí tzv. restartu — program uchovává místo celé posloupnosti jen posledních m vektorů $\{\vec{v}_i\}_{i=k-m+1}^k$.

14.2.3 Metoda bikonjugovaných gradientů (BiCG)

Posledním a často používaným příkladem krylovovské metody je metoda BiCG, jež na rozdíl od předchozích dvou řeší zároveň dvě soustavy $\mathbf{A}\vec{x} = \vec{b}$ a $\mathbf{A}^\top \vec{y} = \vec{c}$. Označíme-li $\vec{s}_k := \vec{c} - \mathbf{A}^\top \vec{y}_k$, pak je metoda BiCG charakterizována vztahy

$$\begin{aligned}\vec{x}_k &\in \vec{x}_0 + \mathcal{K}_k(\mathbf{A}, \vec{r}_0), & \vec{r}_k &\perp \mathcal{K}_k(\mathbf{A}^\top, \vec{s}_0), \\ \vec{y}_k &\in \vec{y}_0 + \mathcal{K}_k(\mathbf{A}^\top, \vec{s}_0), & \vec{s}_k &\perp \mathcal{K}_k(\mathbf{A}, \vec{r}_0).\end{aligned}$$

Vektory $\{\vec{r}_k\}$ a $\{\vec{s}_k\}$ jsou navzájem biortogonální: $\vec{s}_i \cdot \vec{r}_j = 0$ pro $i \neq j$.

Algoritmus A2 Metoda bikonjugovaných gradientů (BiCG)

```
input  $\mathbf{A}, \vec{b}, \vec{c}, \vec{x}_0, \vec{y}_0$ 
 $\vec{r}_0 := \vec{p}_0 := \vec{b} - \mathbf{A}\vec{x}_0$ 
 $\vec{s}_0 := \vec{q}_0 := \vec{c} - \mathbf{A}^\top \vec{y}_0$ 
for  $k = 1, 2, \dots$ 
   $\gamma_{k-1} := \frac{\vec{s}_{k-1} \cdot \vec{r}_{k-1}}{\vec{q}_{k-1} \cdot \mathbf{A}\vec{p}_{k-1}}$ 
   $\vec{x}_k := \vec{x}_{k-1} + \gamma_{k-1} \vec{p}_{k-1}$ 
   $\vec{r}_k := \vec{r}_{k-1} - \gamma_{k-1} \mathbf{A}\vec{p}_{k-1}$ 
   $\vec{y}_k := \vec{y}_{k-1} + \gamma_{k-1} \vec{q}_{k-1}$ 
   $\vec{r}_k := \vec{s}_{k-1} - \gamma_{k-1} \mathbf{A}^\top \vec{q}_{k-1}$ 
   $\delta_k := \frac{\vec{s}_k \cdot \vec{r}_k}{\vec{s}_{k-1} \cdot \vec{r}_{k-1}}$ 
   $\vec{p}_k := \vec{r}_k + \delta_k \vec{p}_{k-1}$ 
   $\vec{q}_k := \vec{s}_k + \delta_k \vec{q}_{k-1}$ 
end
```

Metoda generuje krátké rekurence, je tedy paměťově úsporná, a lze ji použít na obecné regulární matice. Na rozdíl od CG a GMRES však není zaručena konvergence BiCG. Je-li totiž matice \mathbf{A} nesymetrická, může dojít k předčasnému zastavení, když $\vec{r}_k \cdot \vec{s}_k = 0$.

14.3 Předpodmínění

Jak již bylo zmíněno v odst. 14.2.1, konvergence krylovovských metod úzce souvisí s číslem podmíněnosti matice \mathbf{A} . Ukážeme si myšlenku předpodmínění pro metodu CG (u jiných metod lze postupovat obdobně). Nechť \mathbf{C} je libovolná regulární matice. Potom lze soustavu $\mathbf{A}\vec{x} = \vec{b}$ se symetrickou pozitivně definitní maticí zapsat ve tvaru

$$(\mathbf{C}^{-1}\mathbf{A}\mathbf{C}^{-\top})(\mathbf{C}^\top \vec{x}) = \mathbf{C}^{-1}\vec{b}.$$

Označíme-li $\hat{\mathbf{A}} := \mathbf{C}^{-1}\mathbf{A}\mathbf{C}^{-\top}$, $\hat{\vec{x}} := \mathbf{C}^\top \vec{x}$ a $\hat{\vec{b}} := \mathbf{C}^{-1}\vec{b}$, pak novou soustavu můžeme zapsat jako $\hat{\mathbf{A}}\hat{\vec{x}} = \hat{\vec{b}}$, přičemž $\hat{\mathbf{A}}$ je opět symetrická pozitivně definitní. Tuto soustavu lze řešit metodou CG a mezi aproximacemi řešení nové a původní soustavy platí vztah $\vec{x}_k = \mathbf{C}^{-\top} \hat{\vec{x}}_k$. Pro úplnost zde uvádíme algoritmus předpodmíněné metody CG:

Algoritmus A3 Předpodmíněná metoda sdružených gradientů (PCG)

```
input  $\mathbf{A}, \vec{b}, \vec{x}_0$ 
 $\vec{r}_0 := \vec{b} - \mathbf{A}\vec{x}_0$ 
 $\vec{z}_0 := \mathbf{C}^{-\top} \mathbf{C}^{-1} \vec{r}_0$ 
 $\vec{p}_0 := \vec{z}_0$ 
for  $k = 1, 2, \dots$ 
   $\hat{\gamma}_{k-1} := \frac{\vec{z}_{k-1} \cdot \vec{r}_{k-1}}{\vec{p}_{k-1} \cdot \mathbf{A}\vec{p}_{k-1}}$ 
   $\vec{x}_k := \vec{x}_{k-1} + \hat{\gamma}_{k-1} \vec{p}_{k-1}$ 
   $\vec{r}_k := \vec{r}_{k-1} - \hat{\gamma}_{k-1} \mathbf{A}\vec{p}_{k-1}$ 
   $\vec{z}_k := \mathbf{C}^{-\top} \mathbf{C}^{-1} \vec{r}_k$ 
   $\hat{\delta}_k := \frac{\vec{z}_k \cdot \vec{r}_k}{\vec{z}_{k-1} \cdot \vec{r}_{k-1}}$ 
   $\vec{p}_k := \vec{z}_k + \hat{\delta}_k \vec{p}_{k-1}$ 
end
```

Poznamenejme, že v algoritmu nikdy nepočítáme inverzní matici \mathbf{C}^{-1} , ale operaci $\vec{z}_k := \mathbf{C}^{-\top} \mathbf{C}^{-1} \vec{r}_k$ převedeme na řešení dvou soustav

$$\mathbf{C}\vec{y} = \vec{r}_k, \quad \mathbf{C}^\top \vec{z}_k = \vec{y}.$$

Aby bylo řešení nové soustavy efektivnější než řešení soustavy původní, je třeba zvolit matici \mathbf{C} podle následujících požadavků:

- Matici \mathbf{C} volíme tak, aby metoda CG konvergovala co nejrychleji. Ideálně $\hat{\mathbf{A}} = \mathbf{C}^{-1}\mathbf{A}\mathbf{C}^{-\top} \approx \mathbf{I}$.
- Aby nedošlo k výraznému zvýšení náročnosti výpočtu, je potřeba, aby soustavy $\mathbf{C}\vec{y} = \vec{r}_k$ a $\mathbf{C}^{\top}\vec{z}_k = \vec{y}$ byly rychle řešitelné.
- Pokud je matice \mathbf{A} řídká, pak by i \mathbf{C} měla být řídká. Jinak výrazně vzrostou paměťové i výpočetní nároky.

Efektivní volba předpodmiňovací matice často vychází z daného (např. fyzikálního) problému nebo z konkrétní struktury matice \mathbf{A} . Mezi používané obecné předpodmiňovací strategie patří např.:

- neúplný Choleského rozklad, který konstruuje dolní trojúhelníkovou matici \mathbf{C} tak, aby $\mathbf{A} \approx \mathbf{C}\mathbf{C}^{\top}$,
- neúplný LU rozklad: $\mathbf{A} \approx \mathbf{L}\mathbf{U}$, kde \mathbf{L} je dolní trojúhelníková a \mathbf{U} je horní trojúhelníková matice. Předpodmiňovaná soustava pak má tvar

$$(\mathbf{L}^{-1}\mathbf{A}\mathbf{U}^{-1})(\mathbf{U}\vec{x}) = \mathbf{L}^{-1}\vec{b}.$$

Reference

- [1] J. Duintjer Tebbens, I. Hnětynková, M. Plešinger, Z. Strakoš, and P. Tichý. *Analýza metod pro maticové výpočty. Základní metody*. Matfyzpress, 2012. ISBN 978-80-7378-201-6.
- [2] K. Rektorys. *Variační metody v inženýrských problémech a v problémech matematické fyziky*. SNTL, Praha, 1974.