Aproximace a interpolace

Matematické algoritmy (11MAG)

Jan Přikryl

12. přednáška 11MAG
pondělí 15. prosince 2014

verze:2014-12-15 11:10

Obsah

1 Numerické úlohy v analýze a aproximace funkcí

Předmětem studia matematické analýzy jsou funkce a operace s nimi. Typickými úlohami matematické analýzy jsou například výpočet integrálu z dané funkce přes daný interval, výpočet hodnoty derivace, řešení diferenciálních rovnic nebo i pouhé stanovení hodnoty funkce v některém bodě. Pro úlohy matematické analýzy je charakteristické, že zpravidla nejde o numerické úlohy, tj. takové úlohy, kde vstupní i výstupní data jsou vektory o konečném počtu složek. Protože na počítači můžeme přímo modelovat (naprogramovat) pouze úlohy numerické (příkladem mohou sloužit úlohy lineární algebry, viz Přednáška 11), bývá při řešení úloh matematické analýzy na počítači nezbytným přípravným krokem přibližné nahrazení (aproximace) dané matematické úlohy úlohou numerickou. Numerické metody matematické analýzy mají proto poněkud jiný charakter než numerické metody algebry.

Obsah této přednášky je podobně, ale mnohem podrobněji vyložen v učebnicích [2], [3] a skriptech [4]. Zájemce najde jiný styl výkladu a další podrobnosti například v Heathově učebnici [1].

1.1 Aproximace funkcí

Jedním ze základních úkolů numerických metod matematické analýzy je studium aproximací funkcí. Při numerickém řešení úloh matematické analýzy totiž často nahrazujeme danou funkci f, vystupující v řešené matematické úloze, jinou funkcí φ, která v nějakém vhodném smyslu napodobuje funkci f a snadno se přitom matematicky zpracovává či modeluje na počítači. Tuto funkci φ pak nazýváme **aproximací (přiblížením)** funkce f.

Oblasti matematiky, v nichž používáme aproximace, jsou značně různorodé. V této přednášce se budeme zabývat pouze těmi úlohami matematické analýzy, jejichž vstupními a výstupními daty jsou reálné funkce jedné reálné proměnné. Již pouhý výpočet funkčních hodnot takových funkcí na počítači se provádí užitím aproximací φ, jejichž hodnoty se dají vypočítat pomocí konečného počtu aritmetických a logických operací. Typicky se zde používají polynomy nebo racionální funkce; to jsou totiž nejobecnější funkce, jejichž hodnoty se dají na počítači přímo vyčíslit pomocí konečného počtu operací. Tyto aproximace jsou ovšem pro řadu funkcí již zabudovány do výpočetního systému a uživatel počítače si často ani neuvědomuje, že píše-li ve svém programu například `y = sin(x)`, nahrazuje výpočet hodnoty funkce $\sin x$ výpočtem hodnoty jistého polynomu, aproximace.

V této přednášce se z časových důvodů omezíme pouze na aproximace pomocí polynomů. Nebudeme se zde zabývat například aproximacemi pomocí racionálních funkcí, jakkoli mají své výhody (a nevýhody) a v posledních letech se jim věnuje v numerické matematice značná pozornost. Podobně vynecháváme aproximace pomocí trigonometrických polynomů a Fourierovu analýzu vůbec.

Typickým příkladem použití aproximací v matematické analýze jsou také numerické metody pro výpočet určitého integrálu z funkce f. Zde nahrazujeme funkci f funkcí φ, která se snadno integruje, například polynomem. Další oblastí numerické matematiky založenou na užití aproximací je zpracování výsledků měření. Hledáme tu zpravidla jednoduchý analytický výraz vyjadřující (přibližně) funkční závislost, zadanou tabulkou naměřených hodnot.

Při použití aproximací tedy místo původní úlohy, ve které vystupovala funkce f, řešíme úlohu, v níž místo f vystupuje její aproximace φ. To má ovšem za následek, že výsledkem výpočtu nebude přesné řešení původní úlohy. Úkolem numerických metod analýzy je proto také zkoumat, co můžeme říci o odchylce získaného přibližného řešení od přesného (teoretického) řešení dané úlohy. Takové odchylce se říká **chyba aproximace**.

1.2 Třídy aproximujících funkcí

V celé této přednášce se budeme zabývat aproximacemi spojitých reálných funkcí jedné reálné proměnné. Při výběru vhodné aproximace postupujeme v numerické analýze tak, že nejprve předem zvolíme tvar aproximující funkce, ve kterém vystupují některé proměnné parametry, a hodnoty těchto parametrů se pak snažíme určit tak, aby získaná aproximace vyhovovala našim konkrétním požadavkům.

Obecně se při hledání vhodné aproximace velmi často postupuje takto následujícím způsobem. Zvolíme nejprve pevně systém jednoduchých **bázových funkcí** $\varphi_0, \varphi_1, \ldots, \varphi_n$ takových, které se snadno matematicky zpracovávají nebo se s nimi dobře pracuje na počítači, a danou funkci f pak aproximujeme lineární kombinací φ těchto základních funkcí. Klademe tedy

$$\varphi(x) = c_0\varphi_0(x) + c_1\varphi_1(x) + \cdots + c_n\varphi_n(x). \tag{1}$$

Otázka volby aproximace k funkci f se tak převádí na určení hodnot parametrů $c_0, c_1, \ldots, c_n$ podle nějakého kritéria vhodného pro tu či onu konkrétní úlohu. Protože aproximace φ daná uvedeným výrazem závisí na parametrech $c_0, c_1, \ldots, c_n$ lineárně, říkáme o ní, že je **lineárního typu**. Při pevných bázových funkcích nazýváme množinu všech možných jejich lineárních kombinací **třídou aproximujících funkcí** (lineárního typu).

Příklady často používaných bázových funkcí jsou

1. $1, x, x^2, \ldots, x^n$,
2. $1, x - x_0, (x - x_0)^2, \ldots, (x - x_0)^n$, x_0 pevné,
3. $1, \cos x, \sin x, \ldots, \cos Lx, \sin Lx \ (n = 2L)$,
4. $1, \mathrm{e}^{\mathrm{i}x}, \mathrm{e}^{\mathrm{i}2x}, \ldots, \mathrm{e}^{\mathrm{i}nx}$ $(\mathrm{i}^2 = -1)$.

Bázové funkce uvedené v bodech 1 a 2 vytvářejí **třídu polynomů** stupně nejvýše n. Pro bázové funkce uvedené v bodech 3 a 4 se aproximující funkce lineárního typu nazývají trigonometrické polynomy.

Příkladem tříd aproximujících funkcí, které nejsou lineárního typu, jsou třídy racionálních funkcí (to jsou podíly dvou polynomů). Díky tomu, že tyto funkce závisejí na svých parametrech nelineárně, je teorie a praxe takových aproximací ve srovnání s aproximacemi lineárního typu složitější. Pro některé úlohy však aproximace nelineárního typu dávají velmi dobré výsledky.

1.3 Výběr aproximující funkce

Definice 1 (Chyba aproximace)**.** Budiž f daná funkce a φ její aproximace na intervalu $[a, b]$. Funkci E danou pro $x \in [a, b]$ vztahem

$$E(x) = f(x) - \varphi(x)$$

budeme nazývat **chybou aproximace**.

Chceme-li vybrat funkci φ z dané třídy aproximujících funkcí tak, aby byla dobrou aproximací funkce f, budeme se jistě snažit o to, aby byla chyba aproximace v nějakém smyslu malá. Ve většině úloh nás přitom prakticky nezajímá znaménko chyby a při posuzování kvality aproximace tak pracujeme pouze s její absolutní hodnotou $|E(x)| = |f(x) - \varphi(x)|$.

Možností, jak posuzovat („měřit") velikost funkce E na intervalu $[a, b]$, je ale celá řada. V první řadě je, podobně jako tomu bylo u vektorů, potřeba přesně říci, co rozumíme „velikostí" funkce. Pojem absolutní hodnoty čísla se tak podobně jako tomu bylo u vektorů rozšiřuje na pojem **normy funkce**. Normu funkce F označujeme $\|F\|$ a některé používané normy zavedeme v příštích odstavcích. Jakou konkrétně zvolíme normu v daném případě závisí na tom, jaké požadavky na velikost chyby v řešené úloze klademe. Můžeme například požadovat, aby maximální hodnota $|E(x)|$ na intervalu $[a, b]$ byla menší než nějaké dané číslo ε. Jindy pro nás může být důležitá velikost $|E(x)|$ pouze v některých bodech intervalu $[a, b]$. Můžeme třeba požadovat, aby byl malý integrál z funkce $|E(x)|$ přes celý interval $[a, b]$. A jsou i další možné přístupy. To, jakým způsobem budeme chybu aproximace měřit, závisí tedy do značné míry na naší libovůli. Volbu vhodného kritéria pro měření velikosti chyby, podle něhož chybu měříme v té či oné normě, pak provádíme s přihlédnutím k řešené matematické úloze a způsobu zadání funkce f.

Předpokládejme nyní na okamžik, že jsme již zvolili nějakou vhodnou normu pro měření chyby aproximace dané funkce f, takže platí $\|E\| = \|f - \varphi\| \geq 0$ (to je jedna ze základních vlastností všech norem). Chceme tedy zvolit φ z předem vybrané třídy aproximací tak, aby byla tato norma pokud možno malá. Je pak nasnadě otázka, zda v dané třídě aproximací existuje **nejlepší aproximace**, tj. aproximace φ^* taková, že pro všechny aproximace φ z uvažované třídy platí

$$0 \leq \|f - \varphi^*\| \leq \|f - \varphi\|.$$

Dále nás bude zajímat, zda taková aproximace φ^* (pokud existuje) je jediná a jak ji efektivním způsobem sestrojit, tj. jak stanovit hodnoty koeficientů $c_0, c_1, \ldots, c_n$ ve vztahu (1).

O nejlepších aproximacích pojednává celá jedna oblast matematiky, teorie aproximací. Některé příklady nejlepších aproximací polynomy uvedeme i zde. Je ale třeba mít na paměti i to, že v praxi je zpravidla primárním požadavkem, aby velikost chyby aproximace měřená ve zvolené normě byla menší než nějaké předem dané číslo (požadovaná přesnost). Ke splnění tohoto požadavku není vždy nutné používat nejlepší aproximace, někdy je to dokonce neefektivní nebo nemožné.

V dalším výkladu budeme rozlišovat dva základní typy kritérií pro měření velikosti chyby aproximace podle toho, zda kvalitu aproximace posuzujeme na základě hodnot chyby E ve *všech* bodech $x \in [a, b]$ (v takovém případě budeme používat termín **spojitý případ**), nebo jen podle hodnot ve **vybrané konečné soustavě** bodů $x_i \in [a, b], i = 0, 1, \ldots, m$ (*diskrétní případ*).

1.4 Měření chyby aproximace: spojitý případ

Jak jsme již uvedli v úvodu této přednášky, zabýváme se zde funkcemi f, které jsou definovány a spojité na uzavřeném intervalu $[a, b]$, Takové funkce jsou na tomto intervalu konečné a nabývají na něm svého maxima a minima. Pro takové funkce rovněž existují a jsou konečné všechny integrály, s nimž budeme dále pracovat.

Normy, ve kterých budeme měřit chybu aproximace ve spojitém případě, budou všechny – podobně jako tomu bylo u vektorových norem v Přednášce 11 – speciálními případy tzv. *p*-

normy nebo také L_p-*normy*

$$\|F\|_p = \left(\int_a^b |F(x)|^p \,\mathrm{d}x\right)^{1/p}, \quad p \geq 1.$$

Z těchto norem se v praxi velmi často používají především dvě následující normy pro $p = 2$ a $p = \infty$. První z nich, nazývaná obvykle L_2-*norma*, je tedy

$$\|F\|_2 = \left(\int_a^b |F(x)|^2 \,\mathrm{d}x\right)^{1/2}.$$

Všimněme si, že tato norma reprezentuje jakýsi průměr funkce F přes celý interval $[a, b]$. Další hojně používanou normou je **nekonečno-norma**, které se také říká **Čebyševova norma** a která je pro spojité funkce vlastně limitou p-norem při $p \to \infty$. Je dána vztahem

$$\|F\|_\infty = \max_{x \in [a,b]} |F(x)|.$$

Jako měřítko kvality aproximace se tedy ve spojitém případě často užívá například veličina

$$\|E\|_\infty = \|f - \varphi\|_\infty = \max_{x \in [a,b]} |f(x) - \varphi(x)|.$$

Typickým příkladem úloh numerické analýzy, při nichž se používají aproximace vybrané na základě tohoto kritéria, je přibližný výpočet hodnot funkcí, zadaných různými analytickými výrazy, na počítači. Nahrazujeme-li na počítači výpočet hodnot takových funkcí na intervalu $[a, b]$ hodnotami jejich aproximací, je přirozené požadovat, aby $|E(x)|$ byla malá pro všechna $x \in [a, b]$ stejnoměrně. Tomuto požadavku právě odpovídá použití Čebyševovy normy pro měření chyby a nejlepší aproximace ve smyslu této normy se často nazývá **Čebyševova aproximace**.

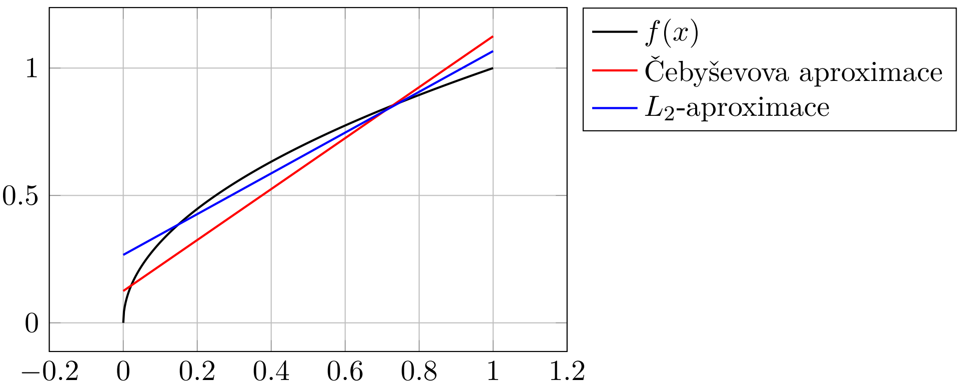
Pokud jde o L_2-normu chyby, je dána vztahem

$$\|E\|_2 = \|f - \varphi\|_2 = \left(\int_a^b |f(x) - \varphi(x)|^2 \,\mathrm{d}x\right)^{1/2}.$$

Tato norma nijak nezdůrazňuje maximální absolutní hodnotu chyby aproximace a dá se o ní říci, že vyjadřuje jakousi střední hodnotu chyby.

Měření chyby aproximace způsoby popsanými v tomto odstavci, tedy užití kritérií spojitého typu, se uplatní zejména při práci s funkcemi zadanými analytickým výrazem. Při zpracování výsledků měření a v řadě numerických metod matematické analýzy používáme jejich diskrétní obdoby, které popíšeme v následujícím odstavci. Poznamenáváme ještě na závěr, že všechny zde uvedené normy mají stejné základní vlastnosti, jako měly vektorové normy uvedené v Přednášce 11, s tím rozdílem, že funkční normy nejsou pro měření velikosti ekvivalentní tak, jako jsme se o tom zmiňovali u vektorových norem. Důvodem je tu mimo jiné skutečnost, že prostory obecných spojitých funkcí nemají konečnou dimenzi n.

Příklad 2 (Nejlepší aproximace: spojitý případ). Ukážeme, že nejlepší aproximace se mohou pro různé normy použité k měření chyby aproximace podstatně lišit. Postup odvození nejlepších aproximací jsme zde nuceni vynechat, uvádíme pouze výsledky. Uvedeme zde dvě různé nejlepší aproximace funkce $f(x) = \sqrt{x}$ na intervalu $[0, 1]$ polynomem prvního stupně. Jako bázové funkce tedy volíme $\varphi_0(x) = 1$, $\varphi_1(x) = x$ a nejlepší aproximace hledáme ve třídě funkcí tvaru $\varphi(x) = c_0 + c_1 x$.



Obrázek 1: Původní funkce $f(x) = \sqrt{x}$ a její nejlepší Čebyševova a L_2-aproximace

Nejlepší Čebyševova aproximace φ^*_∞ má tvar $\varphi^*_\infty(x) = x + \frac{1}{8}$ a Čebyševova norma její chyby je $\|E\|_\infty = 0{,}125$. Nejlepší L_2-aproximace φ^*_L má tvar $\varphi^*_L(x) = \frac{4}{5}x + \frac{4}{15}$ a L_2-norma její chyby je 0,0471. Může být zajímavé nakreslit si graf aproximované funkce i obou nejlepších aproximací – pro kontrolu jej k nahlédnutí uvádíme na Obrázku 1.

Za zmínku stojí ještě to, že L_2-norma chyby Čebyševovy aproximace je 0,0854 a Čebyševova norma nejlepší L_2-aproximace je 0,267. Vidíme tedy, že při hledání optimální aproximace zřetelně záleží na našem hledisku, tj. na požadavcích, které na hledanou aproximaci klademe.

1.5 Měření chyby aproximace: diskrétní případ

Funkce f, kterou chceme aproximovat, může být zadána různými způsoby. V praxi je velmi častý případ, kdy je funkce f dána tabulkou

$$\{(x_i, f(x_i)), i = 0, 1, \ldots, m\}$$

$m+1$ funkčních hodnot $f(x_i)$ v bodech $x_i \in [a, b]$. V takových případech měříme chybu aproximace diskrétními obdobami kritérií popsaných pro spojitý případ, v nichž vystupuje pouze daný konečný počet hodnot funkce f. Spojité p-normy používané ve spojitém případě tak přecházejí v *diskrétní p-normy* dané jako

$$\|F\|_p^m = \left(\sum_{i=0}^{m} |F(x_i)|^p \right)^{1/p}, \quad p \geq 1.$$

Z těchto norem se v praxi opět velmi často používají především dvě následující diskrétní normy pro $p = 2$ a $p = \infty$. První z nich, nazývaná obvykle opět *L_2-norma*, je tedy

$$\|F\|_2^m = \left(\sum_{i=0}^{m} |F(x)|^2 \right)^{1/2}.$$

Další hojně používanou normou je *diskrétní nekonečno-norma*, které se také říká *diskrétní Čebyševova norma* a která je vlastně limitou diskrétních p-norem při $p \to \infty$. Je dána vztahem

$$\|F\|_\infty^m = \max_{i=0,1,\ldots,m} |F(x_i)|.$$

Jako měřítko kvality aproximace se tedy v diskrétním případě může užít například veličina

$$\|E\|_\infty^m = \|f-\varphi\|_\infty^m = \max_{i=0,1,\ldots,m} |f(x_i) - \varphi(x_i)|.$$

Nejlepší aproximace ve smyslu této normy se často nazývá *diskrétní Čebyševova aproximace.* V diskrétním případě se Čebyševova norma příliš často nepoužívá. Podle našeho názoru se v diskrétním případě čtenář častěji setká s měřením chyby v L_2-normě, která je zde dána vztahem

$$\|E\|_2^m = \|f-\varphi\|_2^m = \left(\sum_{i=0}^{m} |f(x_i) - \varphi(x_i)|^2\right)^{1/2}.$$

Tato norma nijak nezdůrazňuje maximální absolutní hodnotu chyby aproximace a dá se o ní opět říci, že vyjadřuje jakousi střední hodnotu chyby. Nejlepší aproximaci ve smyslu této normy se říká také **aproximace metodou nejmenších čtverců**.

Poznamenáváme ještě na závěr, že všechny zde uvedené normy mají opět stejné základní vlastnosti, jako měly vektorové normy uvedené v Přednášce 11. Diskrétní normy jsou totiž vlastně vektorové normy vektorů funkčních hodnot v tabulkových bodech. Přesto, že vektorové normy jsou ve smyslu uvedeném v Přednášce 11 ekvivalentní, nejsou nejlepší aproximace sestrojené v různých normách obecně identické.

Jak ještě uvidíme, ve speciálním případě $n = m$ může být diskrétní norma chyby aproximace nulová, aniž by nutně platila rovnost $\varphi = f$ jako funkcí na celém intervalu $[a, b]$. Nulovost diskrétní p-normy chyby ovšem znamená nulovost všech sčítanců, a v takovém případě tedy platí $\varphi(x_i) = f(x_i), i = 0, 1, \ldots, m$. Interpolace s touto vlastností se pak nazývá **interpolační aproximace**. Poznamenáváme, že o chování interpolační aproximace mezi tabulkovými body však samotná nulovost diskrétní normy chyby nic neříká.

Příklad 3 (Nejlepší aproximace: diskrétní případ)*.* Budeme opět vyšetřovat aproximaci funkce $f(x) = \sqrt{x}$ na intervalu $[0, 1]$ polynomem prvního stupně. Funkce f nechť je nyní zadána tabulkou.

Máme-li funkci f zadánu tabulkou hodnot v bodech $x_0 = 0$ a $x_1 = 1$, takže $f(x_0) = 0, f(x_1) = 1$, má nejlepší diskrétní L_2-aproximace φ_L^* tvar $\varphi_L^*(x) = x$.

Že je φ_L^* nejlepší aproximace, je ihned vidět z toho, že pro její chybu E platí

$$\|E\|_2^1 = \|f-\varphi_L^*|_2^1 = \left(\sum_{i=0}^{1} |f(x_i) - \varphi_L^*(x_i)|^2\right)^{1/2} = 0.$$

Jde tedy o interpolační aproximaci.

Bude-li funkce f zadána tabulkou hodnot ve třech bodech $x_0 = 0, x_1 = \frac{1}{2}, x_2 = 1$, bude mít nejlepší diskrétní L_2-aproximace φ_L^* tvar $\varphi_L^*(x) = x + \frac{1}{6}(\sqrt{2} - 1)$. Diskrétní norma chyby této aproximace jž bude nenulová (kladná). Aproximace není interpolační.

2 Aproximace interpolačním polynomem

Jako aproximace funkcí se v numerické matematice velmi často používají polynomy. Důvodů pro používání polynomiálních aproximací je celá řada [1]. Patří sem především to, že polynomy se dají plně popsat konečným počtem údajů (stupeň, koeficienty) a že se jejich hodnoty dají bez problémů počítat konečným počtem aritmetických operací. Dalším důvodem pro užívání polynomů je to, že se s nimi snadno matematicky pracuje (derivování, integrování). Navíc, jak jsme již uvedli v Přednášce 10 o numerické integraci, platí pro polynomy Weierstrassova věta, která v podstatě říká, že pomocí vhodně vybraných polynomů můžeme spojitou funkci aproximovat s libovolně vysokou přesností. Upozorňujeme ještě na to, že mnohé výsledky, které uvádíme v této kapitole pro interpolační aproximace pomocí polynomů, si zachovávají platnost i pro jiné třídy interpolačních aproximací lineárního typu (např. trigonometrické polynomy).

2.1 Obecně o interpolační aproximaci

Představme si, že máme dánu následující tabulku dat:

t	1,0	2,0	3,0	4,0	5,0	6,0
y	1,9	2,7	4,8	5,3	7,1	9,4

Taková data mohou být výsledkem laboratorních měření, přičemž t by mohlo představovat čas nebo teplotu a y by mohlo představovat vzdálenost nebo tlak. Nebo by uvedená data mohla představovat měření nějakého přírodního jevu, jako je třeba populace nějakého ohroženého druhu či křivku záření supernovy v čase. Nebo by data mohla představovat ceny akcií v různých dnech nebo úhrny prodejů za určitá období. A data by mohla také představovat hodnoty nějaké matematické funkce pro různé argumenty.

U takových dat existuje řada věcí, které bychom s nimi mohli chtít dělat. Mohli bychom chtít vynést data do grafu a nakreslit hladkou křivku procházející tabulkovými body. Mohli bychom chtít nějak odhadnou hodnoty dat mezi tabulkovými body nebo předpovědět hodnoty pro t ležící vně tabulky. Pokud data reprezentují nějaký fyzikální jev, třeba populaci, mohli bychom chtít stanovit některé důležité parametry, třeba tedy růst počtu narozených a zemřelých jedinců. Pokud tabulková data reprezentují hodnoty jisté matematické funkce, mohli bychom chtít najít přibližné hodnoty derivace nebo integrálu, popřípadě rychle vypočítat hodnotu funkce pro daný argument.

V této přednášce se budeme zabývat tím, jak taková diskrétní data vyjádřit pomocí poměrně jednoduchých funkcí, s nimiž se pak dá snadno manipulovat. Budeme se přitom nejprve snažit o to, aby tyto jednoduché funkce nejen aproximovaly celkový trend tabulkových dat, ale aby přímo procházely tabulkovými body, což znamená, že aproximující funkce musí nabývat hodnot daných tabulkou přesně. Znamená to, že diskrétní normy chyby aproximace brané přes tabulkové body budou rovny nule a budeme tedy hledat nejlepší aproximace ve smyslu těchto norem. Takovým funkcím se říká **interpolanty**. S interpolantami jsme se vlastně již v minulých přednáškách setkali. Tak třeba v metodě sečen pro řešení nelineárních rovnic jsme prokládali přímku dvěma funkčními hodnotami. Nebo, i když jsme to podrobně nepopisovali, je Simpsonovo pravidlo pro numerický výpočet integrálu výsledkem toho, že třemi hodnotami integrované funkce proložíme kvadratický polynom. Pokusíme se nyní podat zde obecnější a systematický pohled na problematiku interpolace.

Uvidíme, že interpolační polynom stupně nejvýše n se dá k tabulkovým bodům jednoznačně sestrojit tehdy a jen tehdy, pokud je těchto bodů právě $n+1$. Budeme tedy předpokládat, že funkce f, kterou chceme aproximovat, je dána pouze svými hodnotami v $n+1$ bodech $x_0, x_1, \ldots, x_n$ a že chceme najít aproximaci φ takovou, která bude co nejlépe napodobovat chování funkce f v okolí všech těchto bodů. O tabulkových bodech $x_i, i = 0, 1, \ldots, n$, v nichž jsou zadány (tabelovány) hodnoty $f(x_i)$, zde budeme předpokládat, že jsou navzájem *různé*, a budeme jim říkat **uzly interpolace** nebo **interpolační uzly**. O **interpolaci** tedy hovoříme tehdy, jestliže se aproximující funkci φ snažíme volit tak, aby platilo

$$\varphi(x_i) = f(x_i), \quad i = 0, 1, \ldots, n,$$

a aby tedy přesně nabývala zadaných funkčních hodnot. Uvedeným podmínkám se říká **interpolační podmínky**.

Pokud se nechceme omezit pouze na interpolační polynomy, můžeme úlohu o interpolaci chápat i obecněji. Postupujeme tak, že nejprve stanovíme — s přihlédnutím k předpokládaným vlastnostem tabelované funkce f — vhodnou třídu aproximujících funkcí, v níž budeme aproximaci φ hledat. Prakticky to znamená, že vybereme vhodnou soustavu bázových funkcí. Pro polynomiální interpolaci jsou často používány jako bázové funkce **monomy**, tedy funkce $1, x, x^2, \ldots, x^n$, i když to není jediná možná volba. Konkrétní aproximující funkci z vybrané třídy, tedy vlastně koeficienty v jejím vyjádření jako lineární kombinace bázových funkcí, pak vybereme na základě vstupních dat naší úlohy, tedy pomocí interpolačních podmínek. Protože vstupní data naší úlohu jsou dána konečnou tabulkou a výstupní data jsou dána konečným počtem koeficientů $c_0, c_1, \ldots, c_n$, je úloha o interpolaci numerickou úlohou. Uvidíme, že interpolační aproximace se při pevně daných bázových funkcích dá najít v konečném počtu kroků.

Jak jsme již poznamenali, interpolační aproximace φ minimalizuje v uvažované třídě aproximujících funkcí generované n bázovými funkcemi diskrétní normy $\|E\|_p^m$ s $m = n$ pro libovolné $p \geq 1$, neboť tyto diskrétní normy jsou u interpolačních aproximací rovny nule. K otázce existence a jednoznačnosti interpolačních aproximací se za okamžik vrátíme, nyní zde ale učiníme ještě několik poznámek.

Tak především jsme v Příkladu 3 vlastně pro $n = 1, x_0 = 0, x_1 = 1$ uvedli interpolační aproximaci funkce $f(x) = \sqrt{x}$. Skutečně, nejlepší diskrétní L_2-aproximace $\varphi_L^*(x) = x$ z tohoto příkladu splňuje interpolační podmínky $x_i = \sqrt{x_i}, i = 0, 1$, a je tedy současně interpolační aproximací ze třídy polynomů nejvýše prvního stupně.

Ačkoliv je interpolace mocným nástrojem výpočtové matematiky, není požadavek přesného splnění interpolačních podmínek vždy oprávněný. Tak například, pokud jsou vstupní data zatížena chybami měření nebo jiným významným zdrojem chyb, je obvykle namístě dát přednost vyhlazení takového možného šumu postupem, který se používá v metodě nejmenších čtverců, tedy aproximovat $n + 1$ tabulových dat aproximující funkcí, která má méně než $n + 1$ volných parametrů c_i (tedy například polynomem stupně $m < n$). Chyba nejlepší aproximace je zde pak obecně nenulová, kladná. Příklad takové nejlepší L_2-aproximace funkce $\sqrt{x}$ jsme uvedli opět v Příkladu 3 pro $m = 2$ a $n = 1$.

Podobně se interpolační aproximace obecně nepoužívají v softwarových knihovnách pro výpočet elementárních a speciálních funkcí, které jsou zahrnuty do většiny programovacích jazyků (včetně Matlabu). V takové situaci je totiž důležité, aby aproximující funkce byla blízká aproximované funkci ve všech bodech intervalu $[a, b]$, a není přitom podstatné, aby se této funkci v několika vybraných bodech přímo přesně rovnala. Používá se zde tedy spíše spojitá Čebyševova aproximace.

Je dobré si uvědomit, že ve většině interpolačních problémů je jistá libovůle, protože k daném souboru tabulkových dat může existovat mnoho interpolant. I když vyžadujeme splnění interpolačních podmínek, zůstává zde ještě řada otevřených otázek:

- Jak má vůbec interpolanta vypadat, tj. jakou třídu aproximujících funkcí máme použít? Tady nám leckdy mohou pomoci vhodné matematické nebo fyzikální úvahy.

- Jak by se měla interpolanta chovat mezi tabulkovými body?
- Měla by interpolanta zachovávat některé vlastnosti dat, třeba jejich monotónii, konvexnost nebo periodicitu?
- Zajímají nás primárně hodnoty koeficientů c_i, které při zvolených bázových funkcích definují interpolační aproximaci, nebo nám jde spíše o výhodný výpočet hodnot interpolanty pro její vykreslování či podobné účely?
- Pokud se tabulková data a průběh interpolanty vynesou do grafu, má být tento graf vizuálně uspokojivý?

Volba třídy aproximujících funkcí tak při interpolaci závisí nejen na datech, která chceme popsat, ale i na odpovědích na takové otázky. Výběr interpolační aproximace je obvykle založen na tom,

- jak snadno se s interpolantou pracuje (jde tu o určení koeficientů z daných tabulkových dat, výpočet hodnot mimo tabulkové body, derivování nebo integrace interpolanty atd.),
- jak dobře souhlasí vlastnosti interpolanty s vlastnostmi tabulkových dat (hladkost, monotónie, konvexita, periodicita atd.).

Některé běžné třídy aproximujících funkcí, které se k interpolaci využívají, jsou

- polynomy,
- funkce, které jsou po částech polynomy (splajny),
- trigonometrické funkce,
- exponenciální funkce,
- racionální funkce.

Jak jsme již uvedli, v této přednášce se kromě určitých úvah na obecné úrovni soustředíme výhradně na interpolaci jednorozměrných dat prostřednictvím polynomů. Jiné možnosti jsou k nalezení v literatuře [2, 3, 4, 1].

2.2 Existence, jednoznačnost a podmíněnost

Otázka existence a jednoznačnosti interpolační aproximace se, jak uvidíme, vlastně redukuje na to, zda počet volných parametrů v obecné funkci z dané uvažované třídy aproximací souhlasí s počtem uzlů interpolace. Není-li tomu tak, pak buď interpolanta obecně neexistuje (parametrů méně než uzlů) nebo není určena jednoznačně (parametrů více než uzlů). Omezíme se proto v dalším na situaci, kdy počet uzlů i volných parametrů je stejný, tedy $n + 1$.

Jak jsme uvedli již v úvodu, při daném souboru tabulkových dat $\{(x_i, f(x_i)), i = 0, 1, \ldots, n\}$ (v souladu s předchozím bereme nyní $m = n$) vybíráme interpolantu z třídy funkcí, které jsou lineárními kombinacemi daných **bázových funkcí** $\varphi_0, \varphi_1, \ldots, \varphi_n$. Hledaná interpolační aproximace φ tedy bude mít tvar

$$f(x) = \sum_{i=0}^{n} c_i \varphi_i(x),$$

kde c_i jsou volné parametry, které je třeba určit tak, aby byly splněny interpolační podmínky. Pro daná tabulková data tedy požadujeme, aby platilo

$$\varphi(x_i) = \sum_{i=0}^{n} c_i \varphi_i(x_j) = f(x_j) = y_j, \quad j = 0, 1, \ldots, n,$$

což je ale soustava $n+1$ lineárních algebraických rovnic pro $n+1$ neznámých $c_i, i = 0, 1, \dots, n$, která se dá v maticovém tvaru zapsat jako

$$\mathbf{Ac} = \mathbf{y}, \tag{2}$$

kde prvky čtvercové matice $\mathbf{A}$ řádu $n+1$ jsou dány jako $a_{ij} = \varphi_{j-1}(x_{i-1})$ (tj. a_{ij} je hodnota $(j-1)$ bázové funkce v $(i-1)$ uzlu), složky vektoru pravých stran $\mathbf{y}$ jsou známá data $y_j = f(x_{j-1})$, a složky vektoru neznámých $\mathbf{c}$ jsou hledané koeficienty c_i.

Existence a jednoznačnost řešení dané soustavy nyní závisí pouze na tom, zda je matice soustavy $\mathbf{A}$ regulární. V takovém případě zřejmě interpolační aproximace existuje a je jediná, jakkoli její konkrétní tvar závisí na tom, jak jsme konkrétně zvolili bázové funkce. A podobně, citlivost řešení c na poruchy v datech závisí na podmíněnosti matice $\mathbf{A}$. Uvažujeme-li nějakou předem danou třídu funkcí (třeba polynomy stupně nejvýše n), pak pro ní může existovat i řada různých systémů bázových funkcí. Konkrétní volba bázových funkcí pak ovlivňuje nejen podmíněnost soustavy $\mathbf{Ac} = \mathbf{y}$, ale také objem práce potřebné k jejímu vyřešení a snadnost výpočtu hodnot interpolanty či jiných manipulací s ní. Nyní již se budeme zabývat pouze konkrétně polynomiální interpolací. Jakkoli se zde přitom jsme nuceni omezit pouze na dva případy systému bázových funkcí, upozorňujeme, že existují a používají se i jiné přístupy, s nimž se lze seznámit v literatuře [2, 3, 4, 1].

2.3 Monomy jako interpolační bázové funkce

Protože interpolační podmínky představují $n+1$ podmínek na aproximující funkci, můžeme očekávat, že jednou z vhodných tříd aproximujících funkcí bude třída polynomů stupně nejvýše n, které mají $n+1$ koeficientů. Nejpřirozenější systém bázových funkcí pro tuto třídu tvoří prvních $n+1$ monomů, tedy funkcí

$$\varphi_i(x) = x^i, \quad i = 0, 1, \dots, n.$$

Každá lineární kombinace těchto bázových funkcí je polynom

$$p_n(x) = c_0 + c_1 x + \dots + c_n x^n$$

stupně nejvýše n. Zapíšeme-li nyní interpolační podmínky v maticovém tvaru tak, jak jsme to udělali v minulém odstavci, zjistíme, že vektor $\mathbf{c}$ koeficientů polynomu, který interpoluje tabulková data $\{(x_i, f(x_i)), i = 0, 1, \dots, n\}$, je řešením soustavy $n+1$ rovnic o $n+1$ neznámých

$$\mathbf{Ac} = \begin{bmatrix} 1 & x_0 & \cdots & x_0^n \\ 1 & x_1 & \cdots & x_1^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^n \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} f(x_0) \\ f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} = \mathbf{y}.$$

S maticí tohoto tvaru jsme se již setkali v Odstavci 1.4 Přednášky 10, kde jsme se zabývali numerickými metodami pro přibližný výpočet integrálů; nazývá se **Vandermondova matice** a dá se o ní celkem snadno ukázat, že pokud jsou uzly kvadratury navzájem různé, je regulární. Pokud by tomu totiž tak nebylo, měla by homogenní soustava rovnic $\mathbf{Ac} = \mathbf{o}$ netriviální řešení a tedy by existoval nenulový polynom stupně nejvýše n, který by měl $n+1$ kořenů. To ale není podle toho, co víme z algebry, možné.

Z regularity Vandermondovy matice pak plyne, že pro každá námi popsaná tabulková data existuje právě jeden interpolační polynom stupně nejvýše n. Čtenáře, který již třeba netrpělivě listoval některou učebnicí numerické matematiky, by tady mohlo zmást, že se setkal s interpolačními polynomy různých názvů (třeba Newtonův nebo Lagrangeův interpolační polynom). Nejde přitom o nic jiného než o různé způsoby zápisu téhož jediného interpolačního polynomu, jehož koeficienty jsou řešením výše uvedené soustavy s Vandermondovou maticí. Různé zápisy se mohou ukázat užitečnými v různých praktických situacích.

Všimněme si na tomto místě ještě dvou věcí, o kterých jsme se možná mohli zmínit již dříve. Především mohou být uzly interpolace (tabulkové body) rozloženy zcela libovolně, nemusí být ekvidistantní. Jediný požadavek je zde, aby byly vzájemně různé. Pro jistotu také vysvětlíme, proč zde všude hovoříme o polynomech stupně *nejvýše* n a ne přímo o polynomech stupně n. Je to proto, že některé z koeficientů c_i interpolačního polynomu nám mohou vyjít nulové a interpolační polynom pro našich $n+1$ dat může mít stupeň *menší* než n. Stačí si představit třeba 10 tabulkových dat ležících na přímce. Ta budou interpolována vždy polynomem prvního stupně (lineárním polynomem), ať už interpolační polynom hledáme třeba ve třídě polynomů stupně 9.

Příklad 4 (Interpolace polynomem). Jako ukázku sestrojení interpolačního polynomu pomocí monomiální báze odvodíme interpolační polynom druhého stupně, který bude interpolovat data $(-2,-27),(0,-1),(1,0)$. Víme již z předchozího, že existuje právě jeden kvadratický polynom

$$p_2(x) = c_0 + c_1x + c_2x^2,$$

který interpoluje tři body $(x_0,y_0),(x_1,y_1),(x_2,y_2)$. Použijeme-li monomiální bázi, pak jsou koeficienty tohoto polynomu řešením soustavy rovnic

$$\mathbf{Ac} = \begin{bmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix} = \mathbf{y}.$$

Dosadíme-li sem naše konkrétní data, dostaneme soustavu

$$\begin{bmatrix} 1 & -2 & 4 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} -27 \\ -1 \\ 0 \end{bmatrix}.$$

Vyřešíme-li tuto soustavu (například Gaussovou eliminační metodou), dostaneme jako její řešení vektor $\mathbf{c} = [-1,5,-4]^{\mathrm{T}}$, takže hledaný interpolační polynom je

$$p_2(x) = -1 + 5x - 4x^2.$$

Řešení soustavy $\mathbf{Ac} = \mathbf{y}$ pomocí některého standardního řešiče (v Matlabu třeba `c = A\y`) vyžaduje provést řádově n^3 aritmetických operací. Vzniká tak otázka, zda při vhodné volbě bázových funkcí bychom nemohli dospět k soustavě s nějakou maticí speciálního tvaru. Ukazuje se, že je to možné a uvidíme to například v příštím odstavci. Při použití monomiální báze se navíc ukazuje, že Vandermondovy matice jsou pro větší n často špatně podmíněné. Závisí to jistě také na rozložení uzlů interpolace, ale ve většině případů se ukazuje, že s rostoucím počtem tabulkových bodů číslo podmíněnosti Vandermondovy matice vzrůstá přinejmenším exponenciálním způsobem. Jakkoli tedy teoreticky je Vandermondova matice regulární, stává se při rostoucím n stále hůře podmíněnou a pro dostatečně velká n se na počítači v jeho konečné aritmetice může jevit jako singulární.

Poznamenáváme ale, že přes eventuální špatnou podmíněnost a z ní plynoucí nepřesné stanovení koeficientů interpolačního polynomu bude i tento polynom dobře splňovat interpolační podmínky. Plyne to z toho, že — jak jsme již uvedli u Gaussovy eliminace — Gaussova eliminační metoda s částečným výběrem hlavních prvků dává při řešení soustav lineárních rovnic vždy malá rezidua, nezávisle na podmíněnosti soustavy. Přitom ovšem hodnoty koeficientů polynomu samotné mohou být vypočítány s velkou chybou.

Na závěr v tomto odstavci ještě připomeneme jeden užitečný způsob výpočtu hodnot polynomu, který je vyjádřen pomocí monomiální báze, tedy polynomu tvaru

$$p_n(x) = c_0 + c_1 x + \cdots + c_n x^n.$$

Celý postup, kterému se říká **Hornerovo schéma** nebo někdy také **syntetické dělení**, spočívá vlastně pouze ve vhodném uzávorkování. Zapíšeme

$$p_n(x) = c_0 + x(c_1 + x(c_2 + x(\cdots(c_{n-1} + xc_n)\cdots)))$$

a vidíme, že výpočet jedné hodnoty polynomu tímto způsobem vyžaduje pouze n sčítání a n násobení. Podobný princip se používá i při sestavování Vandermondovy matice. Místo aby se explicitně počítaly mocniny uzlů, pracujeme se vztahem

$$a_{ij} = \varphi_{j-1}(x_{i-1}) = x_{i-1}^{j-1} = x_{i-1}\varphi_{j-2}(x_{i-1}) = x_{i-1}a_{i,j-1}, \quad j = 2, \ldots, n.$$

V Matlabu je pro výpočet hodnot polynomů k dispozici funkce `polyval`. Koeficienty interpolačního polynomu v monomiální bázi lze stanovit pomocí funkce `polyfit` (ta umožňuje pracovat i s metodou nejmenších čtverců, tedy prokládat daty polynomy, které jsou nižšího stupně než odpovídá počtu tabulkových dat) nebo pomocí funkce `interp1` (ta umožňuje provádět interpolaci i s jinými bázovými funkcemi než jsou polynomy).

2.4 Lagrangeův interpolační polynom

Jak jsme viděli již dříve, matice soustavy interpolačních podmínek (2) je dána hodnotami bázových funkcí v uzlech interpolace. Vzniká tedy otázka, jak vhodně volit polynomiální bázové funkce tak, aby matice **A** soustavy (2) měla jednoduchý tvar a tato soustava se snadno řešila. Nejjednodušší matici dostaneme postupem, který pochází od klasického francouzského matematika Lagrange. Tato matice **A** bude *jednotková*, nebude tedy co řešit a koeficienty u Lagrangeových bázových funkcí budou přímo funkční hodnoty.

Aby matice **A** vyšla jednotková, potřebujeme tedy zvolit bázové funkce $\varphi_i, i = 0, 1, \ldots, n$, tak, aby platilo $\varphi_i(x_i) = 1$ a $\varphi_i(x_k) = 0$ pro $i \neq k$. Snadno se ověří, že těmto podmínkám budou vyhovovat bázové funkce tvaru

$$\varphi_i(x) = \ell_i(x) = \frac{(x - x_0)(x - x_1)\cdots(x - x_{i-1})(x - x_{i+1})\cdots(x - x_n)}{(x_i - x_0)(x_i - x_1)\cdots(x_i - x_{i-1})(x_i - x_{i+1})\cdots(x_i - x_n)}, \tag{3}$$

kterým se říká **Lagrangeovy bázové polynomy** nebo někdy také **fundamentální polynomy**. Interpolační polynom zapsaný v těchto bázových funkcích se nazývá **Lagrangeův interpolační polynom** a podle toho, co jsme výše řekli o matici **A**, má tvar

$$L_n(x) = \sum_{i=0}^{n} f(x_i)\ell_i(x).$$

Je to tedy opět polynom stupně nejvýše n a vzhledem k jednoznačnosti interpolačního polynomu je pouze jiným zápisem polynomu, který bychom dostali s monomiální bází ve tvaru, na který jsme běžně zvyklí.

Použijeme-li Lagrangeovy bázové polynomy, je tedy snadné pro daná tabulková data data zapsat interpolační polynom a koeficienty v jeho vyjádření jsou naprosto dobře podmíněné. Ale hodnoty interpolačního polynomu v Lagrangeově tvaru je pracnější počítat než je tomu u polynomu v monomiální bázi, kde se dá použít Hornerovo schéma. Lagrangeovy polynomy se také hůře derivují, integrují atd. Z těchto důvodů se z Lagrangeových interpolačních polynomů používají spíše polynomy nižších stupňů.

Příklad 5 (Interpolace Lagrangeovým polynomem). Nechť je funkce f dána svými hodnotami ve čtyřech bodech $x_0 = 0, x_1 = 1, x_2 = -1, x_3 = 3$ a nechť funkční hodnoty v těchto bodech jsou $f(0) = 1, f(1) = 2, f(-1) = 2, f(3) = 0$. Vypočítáme přibližnou hodnotu $f(2)$ pomocí Lagrangeova interpolačního polynomu L_3, který při těchto datech bude třetího stupně ($n = 3$),

$$L_3(x) = f(x_0)\ell_0(x) + f(x_1)\ell_1(x) + f(x_2)\ell_2(x) + f(x_3)\ell_3(x).$$

Podle (3) vyjde první báze

$$\ell_0(x) = \frac{(x-1)(x+1)(x-3)}{(0-1)(0+1)(0-3)} = \tfrac{1}{3}(x-1)(x+1)(x-3)$$

a dále podobně obdržíme

$$\begin{aligned}\ell_1(x) &= -\tfrac{1}{4}x(x+1)(x-3),\\ \ell_2(x) &= -\tfrac{1}{8}x(x-1)(x-3),\\ \ell_3(x) &= \tfrac{1}{24}x(x-1)(x+1).\end{aligned}$$

Je $\ell_0(2) = -1, \ell_1(2) = \frac{3}{2}, \ell_2(2) = \frac{1}{4}, \ell_3(2) = \frac{1}{4}$, a tedy

$$L_3(2) = 1 \cdot (-1) + 2 \cdot \tfrac{3}{2} + 2 \cdot \tfrac{1}{4} + 0 \cdot \tfrac{1}{4} = \tfrac{5}{2} \approx f(2).$$

Samotný polynom L_3 má tvar

$$\begin{aligned}L_3(x) &= 1 \cdot \ell_0(x) + 2 \cdot \ell_1(x) + 2 \cdot \ell_2(x) + 0 \cdot \ell_3(x) =\\ &= \tfrac{1}{3}(x-1)(x+1)(x-3) - \tfrac{1}{2}x(x+1)(x-3) - \tfrac{1}{4}x(x-1)(x-3) =\\ &= -\tfrac{5}{12}x^3 + x^2 + \tfrac{5}{12}x + 1,\end{aligned}$$

kde poslední řádek jsme získali roznásobením a je identický s tvarem polynomu v monomiální bázi.

2.4.1 Lineární a kvadratická interpolace

Interpolujeme-li funkci f na základě jejích hodnot ve dvou bodech, aproximujeme ji polynomem prvního stupně ($n = 1$) a hovoříme o *lineární interpolaci*. Interpolační polynom prvního stupně L_1 má pak tvar

$$L_1(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1). \tag{4}$$

Právě lineární interpolace se občas používá při práci s tabulkami funkcí ve střední škole.

Je-li $n = 2$, hovoříme o *kvadratické interpolaci*. Interpolační polynom L_2 druhého stupně má tvar

$$L_2(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} f(x_0) + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} f(x_1) + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} f(x_2). \tag{5}$$

Příklad 6 (Lineární interpolace). Máme dánu tabulku hodnot funkce $f(x) = \sin x$ na 5 desetinných míst:

x	$20°$	$21°$	$22°$
$\sin x$	0,34202	0,35837	0,37461

Vypočítáme přibližnou hodnotu $\sin 20°18'$, tj. $\sin 20{,}3°$, lineární interpolací z funkčních hodnot v bodech $x_0 = 20°$ a $x_1 = 21°$. Dostaneme

$$\sin 20°18' \approx \frac{20{,}3 - 21}{20 - 21} \cdot 0{,}34202 + \frac{20{,}3 - 20}{21 - 20} \cdot 0{,}35837 \approx 0{,}34693.$$

Přesná hodnota $\sin 20°18'$ zaokrouhlená na 5 desetinných míst je 0,34694. V tomto případě je tedy lineární interpolace zcela postačující, bylo by zbytečně pracné volit $n > 1$.

2.5 Chyba aproximace interpolačním polynomem

Z toho, že při interpolaci pro chybu aproximace $E = f - p_n$ polynomem stupně n platí $E(x_i) = 0, i = 0, 1, \ldots, n$ (protože interpolační polynom splňuje interpolační podmínky), nemůžeme ještě dělat přímé závěry o hodnotě chyby aproximace mimo tabulkové body. Popravdě řečeno samotná tabulka o chování aproximované funkce mimo tabulkové body nic neříká a pro $x \neq x_i$ se aproximovaná funkce může chovat zcela libovolně. Abychom tedy mohli o chování chyby aproximace mimo tabulkové body vůbec něco říci, musíme mít k dispozici více informací o funkci, která má být tabulkou reprezentována. Příkladem takových informací je například informace o tom, že aproximovaná funkce f je na intervalu $[a, b]$ dostatečně „hladká", tj. má tam určitý počet spojitých derivací.

V dalším výkladu budeme symbolem $\mathrm{int}(x_0, x_1, \ldots, x_n, x)$ označovat nejmenší otevřený interval takový, že uvedené body $x_0, x_1, \ldots, x_n, x$ leží uvnitř tohoto intervalu nebo na jeho hranici. V učebnicích numerické matematiky se pak dokazuje o chybě aproximace interpolačním polynomem následující tvrzení, zde uváděné bez důkazu.

Věta 7 (Chyba interpolační aproximace)**.** *Nechť* $[a, b]$ *je* libovolný *interval, který obsahuje všech* $n + 1$ *interpolačních uzlů* $x_0, x_1, \ldots, x_n$. *Nechť* $f, f', \ldots, f^{(n)}$ *existují a jsou spojité na* $[a, b]$ *a nechť pro všechna* $x \in (a, b)$ *existuje derivace* $f^{(n+1)}(x)$.

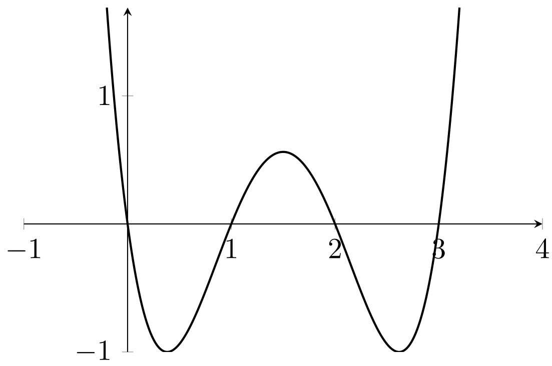
Pak pro chybu E *aproximace funkce* f *interpolačním polynomem* p_n *platí na intervalu* $[a, b]$ *vzorec*

$$E(x) = f(x) - p_n(x) = (x - x_0)(x - x_1) \cdots (x - x_n) \frac{f^{(n+1)}(\xi)}{(n + 1)!}, \tag{6}$$

kde ξ *je nějaké (blíže neurčené) číslo z intervalu* $\mathrm{int}(x_0, x_1, \ldots, x_n, x)$*, které tedy závisí na konkrétní hodnotě* x.

Vidíme, že pokud se $f^{(n+1)}(x)$ na intervalu (a, b) příliš nemění, je pro průběh chyby aproximace rozhodující průběh polynomu $\omega_n(x) = (x - x_0)(x - x_1) \cdots (x - x_n)$, Tento polynom nezávisí na interpolované funkci, ale pouze na bodech x_n. Vhodnou volbou uzlů $x_i, i = 0, 1, \ldots, n$, tedy můžeme chybu aproximace zmenšit. Bližší podrobnosti v tomto směru lze najít v literatuře [2, 3, 4, 1]. Zde pouze poznamenáváme, že optimální volbou uzlů interpolace jsou kořeny Čebyševových ortogonálních polynomů (viz [1]) a ne tedy rovnoměrně rozložené uzly.

Příklad 8. Pro $n = 3$ a $x_i = i$, $i = 0, 1, 2, 3$ je průběh polynomu ω_3 znázorněn na obrázku 2. Z obrázku vidíme, že se dá čekat, že vně intervalu $\mathrm{int}(x_0, x_1, \ldots, x_n)$ chyba aproximace interpolačním polynomem prudce poroste.



Obrázek 2: Průběh polynomu ω_3 pro uzly $x_i = i$.

2.5.1 Odhad chyby aproximace

Vzorec (6) pro chybu aproximace se pro přímé praktické použití příliš nehodí, neznáme totiž předem hodnotu argumentu ξ, který ve vzorci vystupuje. Ze vzorce je ovšem vidět, že chyba aproximace je v interpolačních uzlech nulová, neboť to jsou kořeny polynomu $\omega_n(x)$. Dále ze vzorce také vidíme, že je-li aproximovaná funkce f sama polynomem nejvýše stupně n, aproximuje ji interpolační polynom $p_n(x)$ stupně n s nulovou chybou, přesně, a je tedy přímo $p_n(x) = f(x)$. To plyne ze skutečnosti, že v takovém případě je derivace funkce f řádu $n+1$ (a všechny vyšší derivace) identická nula pro všechny hodnoty argumentu.

Ze vzorce (6) ale plyne prakticky použitelný odhad chyby, alespoň v případech, kdy dokážeme hodnotu v něm vystupující derivace na intervalu $[a, b]$ nějak odhadnout. Pokud se nám totiž podaří určit konstantu M takovou, že pro všechna $x \in [a, b]$ je $|f^{(n+1)}(x)| \leq M$, pak je

$$|E(x)| \leq \frac{M}{(n+1)!}|\omega_n(x)| \leq \frac{M}{(n+1)!} \max_{x\in[a,b]} |\omega_n(x)$$

pro všechna $x \in [a, b]$.

Příklad 9 (Odhad chyby). Odhadneme chybu aproximace, které jsme se dopustili při lineární interpolaci v příkladu 6.

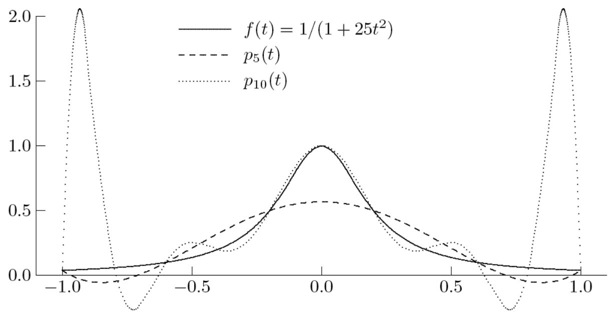
Je $n = 1$ a $f''(x) = -\sin x$. Takto počítaná derivace se ale odvozuje pro x brané v obloukové míře, kde $360° = 2\pi$, takže musíme celý odhad provést na intervalu $[\pi/9, 7\pi/60]$, který v obloukové míře odpovídá intervalu $[20°, 21°]$. Z tabulky v příkladu 6 a na základě toho, co víme o průběhu funkce $\sin x$, vidíme, že na tomto intervalu máme na 5 platných číslic $|f''(x)| = |\sin x| \leq 0{,}35837$, a tedy (do $\omega_1(x)$ dosazujeme opět v obloukové míře!)

$$|E(20°18'')| \leq \frac{1}{2} \cdot 0{,}35837 \cdot \left(\frac{18}{60} \cdot \frac{\pi}{180}\right) \cdot \left(\frac{42}{60} \cdot \frac{\pi}{180}\right) \approx 1{,}1 \times 10^{-5}.$$

Skutečná chyba aproximace je (viz údaje v příkladu 6) přibližně rovna 1×10^{-5}. Všimněme si ještě toho, že pokud bychom užili k odhadnutí druhé derivace pouze známý vztah $|\sin x| \leq 1$, dostali bychom odhad chyby aproximace zbytečně pesimistický.

2.5.2 Extrapolace

Používáme-li interpolační polynom k výpočtu přibližných hodnot interpolované funkce vně intervalu $\mathrm{int}(x_0, x_1, \ldots, x_n)$, říkáme, že provádíme **extrapolaci**. Jak je víceméně vidět z obrázku 2,



Obrázek 3: Interpolace Rungovy funkce – ekvidistantní uzly.

chyba aproximace může být ve větších vzdálenostech od koncových bodů interpolačního intervalu velká. Pro hodnoty x blízké koncovým bodům intervalu $\mathrm{int}(x_0, x_1, \ldots, x_n)$ lze však přesto dostat dobré výsledky.

Příklad 10 (Lineární extrapolace). Lineární interpolace z hodnot $\sin 20°$ a $\sin 21°$ uvedených v tabulce v příkladu 6 dává podle vzorce (4) pro $\sin 21°18'$ přibližnou hodnotu 0,36328 s chybou asi 3×10^{-5}, což je zcela přijatelný výsledek. Lineární interpolace s uzly $x_0 = 21°$ a $x_1 = 22°$ by nám dala jako výsledek $\sin 21°18' \approx 0{,}36324$ s chybou přibližně 1×10^{-5}.

3 Dodatky

3.1 Konvergence interpolačního procesu

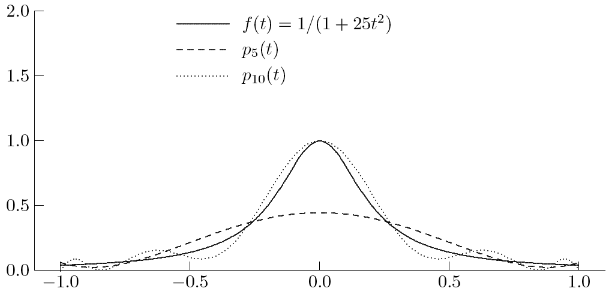
Podobně jak jsme se o tom už zmiňovali v Přednášce 10 u numerických kvadratur, má smysl i zde zabývat se otázkou, zda při interpolaci funkce f na intervalu $[a, b]$ s uzly $x_0, x_1, \ldots, x_n$ povede zvyšování počtu uzlů (a tedy aproximace interpolačními polynomy stále vyšších stupňů) ke stále lepší aproximaci funkce f. Odpověď na tuto otázku je v jistém smyslu negativní. Ať už totiž volíme uzly interpolace jakkoli, vždy se dá najít spojitá funkce taková, že námi generovaná posloupnost interpolačních polynomů k ní nebude konvergovat *stejnoměrně* na celém $[a, b]$.

Volíme-li uzly interpolace při tomto procesu ekvidistantně, dělají potíže už poměrně jednoduché spojité funkce jako $\sqrt{x}$ na $[0, 1]$ nebo $1/(1 + x^2)$ na intervalu $[-1, 1]$. Jako příklad problematické interpolace s ekvidistantními uzly uvádíme na obrázku 3 interpolační polynomy stupně 5 a 10 pro Rungovu funkci $f(t) = 1/(1 + 25t^2)$ na intervalu $[-1, 1]$. Vidíme, že interpolace polynomem vyššího stupně zlepšuje kvalitu aproximace kolem středu intervalu, ale výrazně ji zhoršuje u krajů intervalu.

3.2 Volba uzlů interpolace

Jak jsme právě viděli, může interpolace spojitých funkcí při použití ekvidistantních uzlů dávat problematické výsledky. Není tomu tak naštěstí vždy, v teorii interpolace se ukazuje, že interpolační proces s ekvidistantními uzly konverguje k aproximované funkci f stejnoměrně na intervalu $[a, b]$, je-li f celistvá analytická funkce (jde o pojem z teorie funkcí komplexní proměnné, v podstatě jde o funkci mající všechny derivace, která se dá rozvinout v mocninnou řadu). Příkladem takových funkcí jsou například $\sin x, \cos x$ nebo e^x.

Interpolace ekvidistantních dat polynomy vyšších stupňů je v některých případech také špatně podmíněná úloha, tj. malé nepřesnosti ve vstupních datech mohou působit velké chyby v hod-



Obrázek 4: Interpolace Rungovy funkce – Čebyševovy uzly.

notách interpolačního polynomu. Jak jsme už viděli na obrázku, potíže vzikají především při aproximování funkce poblíž konců intervalu vytvářeného uzly interpolace, kde se pak navíc interpolační polynomy vyšších stupňů zpravidla značně „vlní“.

Chceme-li aproximovat nějakou funkci polynomem na celém intervalu a můžeme-li si vybrat body, v nichž počítáme nebo měříme funkční hodnoty, je proto vždy rozumnější *nevolit* uzly x_i ekvidistantně. Dobrá strategie je volit tabulkové body tak, aby byly rozloženy obdobně jako *kořeny Čebyševových polynomů*. O těchto ortogonálních polynomech, které v numerické matematice hrají roli i jinde než při interpolaci, se lze poučit v literatuře [1, 2, 3, 4]. Takovou volbou uzlů také minimalizujeme hodnotu $\max |\omega_n(x)|$ v odhadu chyby interpolace.

Jako informaci pro čtenáře uvádíme, že pracujeme-li s intervalem $[a, b] = [-1, 1]$, volíme při práci s kořeny Čebyševových polynomů jako uzly

$$x_i = \cos\left(\frac{2i+1}{n+1} \cdot \frac{\pi}{2}\right), \quad i = 0, 1, \ldots, n.$$

Na obecný interval $[a, b]$ tyto uzly zobrazíme jednoduchou lineární transformací

$$z = \tfrac{1}{2}[(b-a)x + (b+a)], \quad x \in [-1, 1], \quad z \in [a, b].$$

Volíme-li uzly interpolace tímto způsobem, má interpolační proces při $n \to \infty$ tu vlastnost, že vzniklé interpolační polynomy konvergují na $[a, b]$ stejnoměrně k aproximované funkci už při velmi mírných požadavcích na tuto funkci. Postačující podmínkou pro stejnoměrnou konvergenci je tu například již pouhá existence spojité první derivace f' na $[a, b]$. Na závěr tohoto odstavce ještě na obrázku ukážeme, jak vypadají interpolační polynomy stupňů 5 a 10 pro Rungovu funkci při Čebyševově rozložení uzlů interpolace.

3.3 Metoda nejmenších čtverců

Je-li naším úkolem aproximovat data

$$\{(x_i, f(x_i)), i = 0, 1, \ldots, m\}$$

a jsou li hodnoty $f(x_i)$ zatíženy chybami (jde například o naměřené veličiny), není namístě požadovat splnění interpolačních podmínek a hledat jako aproximující funkcí polynom stupně $n = m$, který tabulkovými daty přesně prochází. Místo toho spíše volíme jako aproximující funkci polynom p_n nižšího stupně $n < m$ a jeho koeficienty se snažíme volit tak, abychom minimalizovali střední kvadratickou odchylku od tabulkových dat, tedy diskrétní L_2-normu chyby

$$\|E\|_2^m = \|f - p_n\|_2^m = \left(\sum_{i=0}^{m} |f(x_i) - p_n(x_i)|^2\right)^{1/2}.$$

Tak například můžeme prokládat tabulkou dvaceti naměřených hodnot přímku $p_1(x) = ax + b$. V tomto případě předpokládáme, že charakter měřené závislosti je možno s dostatečnou přesností vystihnout právě polynomem prvního stupně. Obecně se dá říci, že situace, kdy se charakter většího počtu naměřených dat dá dobře vystihnout polynomem poměrně nízkého stupně, jsou v praxi dost běžné. Popsanému přístupu k aproximaci dat zatížených chybami se říká *metoda nejmenších čtverců*.

Aproximace získané metodou nejmenších čtverců mají jisté užitečné statistické vlastnosti a vyrovnávají vliv náhodných chyb v zadaných (naměřených) hodnotách. Aproximace získané metodou nejmenších čtverců jsou tedy také nejlepší diskrétní L_2-aproximace při $n < m$. Čtenáře možná nepřekvapí, že pro $n = m$ dává aproximace metodou nejmenších čtverců interpolační polynom stupně n. Metodě nejmenších čtverců a její teorii i praxi je v numerické matematice věnována značná pozornost. Bohužel nám tento text nedává možnost se o ní podrobněji rozepsat. Zájemce tak odkazujeme především na [2], [3] a [4]; jiný pohled a nejnovější výsledky lze najít v [1].

V Matlabu umožňuje práci s metodou nejmenších čtverců již zde zmiňovaná funkce `polyfit`, kde jako jeden z parametrů můžeme volit stupeň aproximujícího polynomu. Pokud jde o vhodnou volbu stupně aproximujícího polynomu, doporučují statistické úvahy konstruovat metodou nejmenších čtverců postupně polynomy stupně $n = 0, 1, 2, \ldots$, počítat navíc pro každý polynom p_n hodnotu veličiny

$$\frac{(\|E\|_2^m)^2}{m - n}$$

a pokračovat se zvyšováním stupně tak dlouho, dokud tato veličina s rostoucím n významně klesá.

I v metodě nejmenších čtverců se stejně jako při interpolaci může ukázat, že se naměřené hodnoty pro aproximaci polynomem nehodí. V takovém případě je třeba přejít k jiným systémům základních funkcí nebo hledat aproximaci nelineárního typu. Zde jsme již ale opět nuceni odkázat případného zájemce na literaturu.

Reference

[1] Michael T. Heath. *Scientific computing: an introductory survey.* McGraw-Hill, Boston, 2 edition, 2002.

[2] Petr Přikryl. *Numerické metody matematické analýzy.* MVŠT. SNTL, Praha, 1985.

[3] Petr Přikryl. *Numerické metody matematické analýzy.* MVŠT. SNTL, Praha, 2., opr. a dopl. edition, 1988.

[4] Petr Přikryl and Marek Brandner. *Numerické metody II.* FAV ZČU, Plzeň, 2001.