Matematika II – 1. přednáška
Polynomiální interpolace, splajny

Michal Bulant

Masarykova univerzita
Fakulta informatiky

17. 9. 2008

Obsah přednášky

Doporučené zdroje

- Martin Panák, Jan Slovák – **Drsná matematika**, *e*-text.
- Roman Hilscher – `MB102`, *e*-text.
- Ivana Horová, Jiří Zelinka – Numerické metody, MU Brno, 2. rozšířené vydání, 2004, 294 s., ISBN 80-210-3317-7.
- Zuzana Došlá, Jaromír Kuben – Diferenciální počet funkcí jedné proměnné, MU Brno, 2003, 215 s., ISBN 80-210-3121-2 (*rovněž na* `http://www.math.muni.cz/~dosla/skript.pdf`).

V tomto semestru budeme budovat nástroje pro modelování závislostí, které nejsou ani lineární ani diskrétní.
Někdy je to přímo záměr či potřeba vzhledem k faktickému stavu zkoumaného problému (třeba u modelů fyzikálních jevů), jindy je to vhodné přiblížení diskrétního modelu (třeba u ekonomických nebo populačních modelů).

Budeme pracovat s funkcemi $\mathbb{R} \to \mathbb{R}$ (**reálné funkce reálné proměnné**) nebo $\mathbb{R} \to \mathbb{C}$ (**komplexní funkce reálné proměnné**), případně s funkcemi $\mathbb{Q} \to \mathbb{Q}$ (funkce jedné racionální proměnné s racionálními hodnotami) apod. Většinou půjdou naše závěry snadno rozšířit na případ s vektorovými hodnotami nad stejnými skaláry, ve výkladu se ale zpravidla omezíme jen na případ reálných čísel, případně komplexních čísel.

Čím větší třídu funkcí připustíme, tím obtížnější bude vybudovat nástroje pro naši práci. Cílem našich prvních dvou kapitol matematické analýzy bude proto explicitně zavést několik typů elementárních funkcí, implicitně popsat více funkcí a vybudovat standardní nástroje pro práci s nimi. Souhrnně se tomu říká diferenciální a integrální počet jedné proměnné.

Polynomy

Skaláry umíme sčítat i násobit a tyto operace splňují řadu vlastností, které jsme vyjmenovali hned na začátku minulého semestru. *Polynomem* nad okruhem skalárů $\mathbb{K}$ rozumíme zobrazení $f : \mathbb{K} \to \mathbb{K}$ dané pro každé $x \in \mathbb{K}$ výrazem

$$x \mapsto f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

kde a_i, $i = 0, \dots, n$, jsou pevně zadané skaláry, násobení je znázorněno prostým zřetězením symbolů a symbol $+$ označuje sčítání. Pokud je $a_n \neq 0$, říkáme, že polynom f je **stupně** n. Stupeň nulového polynomu není definován (někdy se klade roven $- = \infty$).
Skaláry a_i označujeme jako **koeficienty polynomu** f. Polynomy stupně nula jsou konstantní nenulová zobrazení $x \mapsto a_0$.
Algebraicky jsou polynomy definovány jako formální výrazy $f(x)$, tj. jako posloupnosti koeficientů $a_0, a_1, \dots$ s konečně mnoha nenulovými prvky. Jsou tyto přístupy ekvivalentní?

Je snadné ověřit, že polynomy nad okruhem skalárů tvoří opět okruh, kde násobení a sčítání je dáno operacemi v původním okruhu $\mathbb{K}$ pomocí hodnot polynomů. Připomeňte si při této příležitosti vlastnosti skalárů a ověřte!

Lemma

Je-li $\mathbb{K}$ pole s nekonečně mnoha prvky, pak dva polynomy f a g jsou si rovny jako zobrazení, právě když mají shodné koeficienty.

Uvědomme si, že u konečných polí samozřejmě takové tvrzení neplatí. Uvažte např. polynom $x^2 + x$ nad $\mathbb{Z}_2$.

Důkaz.

Nad každým polem skalárů funguje dělení polynomů se zbytkem, tj. pro polynomy f stupně n a g stupně m, existují jednoznačně určené polynomy q a r takové, že stupeň r je menší než m nebo je $r = 0$ a $f = q \cdot g + r$.
Je-li pro nějaký prvek $b \in \mathbb{K}$ hodnota $f(b) = 0$, pak to znamená, že v podílu $f(x) = q(x)(x - b) + r$ musí být $r = 0$. Jinak by totiž nebylo možné dosáhnout $f(b) = q(b) \cdot 0 + r$, kde stupeň r je nulový. Říkáme, že b je **kořen polynomu** f.
Stupeň q je pak právě $n - 1$. Pokud má q opět kořen, můžeme pokračovat a po nejvýše n krocích dojdeme ke konstatnímu polynomu. Dokázali jsme tedy, že každý nenulový polynom nad polem $\mathbb{K}$ má nejvýše tolik kořenů, kolik je jeho stupeň. Odtud již snadno dovodíme i následující pozorování:
Předpokládejme $f = g$, tj. $f - g = 0$ jako zobrazení. Polynom $(f - g)(x)$ tedy má nekonečně mnoho kořenů, což je možné pouze tehdy, je-li nulovým polynomem. $\square$

Interpolační polynom

Častá praktická úloha zní

> *zadejte formuli pro funkci, pro kterou máme zadány hodnoty v předem daných daných bodech* $x_0, \dots, x_n$.

Pokud by šlo o nulové hodnoty, umíme přímo zadat polynom

$$f(x) = (x - x_0)(x - x_1) \dots (x - x_n),$$

který bude mít nulové hodnoty právě v těchto bodech a nikde jinde. To ale není jediná odpověď, protože požadovanou vlastnost má i nulový polynom. Ten je přitom jediný s touto vlastností ve vektorovém prostoru polynomů stupně nejvýše n. Ve skutečnosti jsme dokázali před chvílí, že *nenulový polynom stupně n, nemá nikdy více než n nulových bodů*.

Theorem

Nechť $\mathbb{K}$ je nekonečné pole skalárů, pak pro každnou množinu po dvou různých bodů $x_0, \ldots, x_n \in \mathbb{K}$ a předepsaných hodnot $y_0, \ldots, y_n \in \mathbb{K}$ existuje právě jeden polynom f stupně nejvýše n (případně nulový polynom), pro který platí

$$f(x_i) = y_i, \quad i = 0, \ldots, n.$$

Důkaz.

Protož už víme, že existuje nejvýše jeden polynom stupně n s předepsanými $n+1$ hodnotami y_i v různých bodech $x_0, \ldots, x_n$, stačí sestrojit takový polynom. To ale není těžké, stačí pracovat s polynomy

$$\ell_i(x) = \frac{\prod_{j \neq i}(x - x_j)}{\prod_{j \neq i}(x_i - x_j)}.$$

Hledaný polynom je $f(x) = y_0\ell_0(x) + y_1\ell_1(x) + \cdots + y_n\ell_n(x)$. $\square$

Dosazením požadovaných hodnot do polynomu s neznámými koeficienty dostaneme systém $n+1$ rovnic pro stejný počet neznámých koeficientů a_i

$$\begin{aligned} a_0 + x_0 a_1 + \cdots + (x_0)^n a_n &= y_0 \\ &\vdots \\ a_0 + x_n a_1 + \cdots + (x_n)^n a_n &= y_n. \end{aligned}$$

Jak je dobře známo z lineární algebry, tento systém lineárních rovnic má právě jedno řešení pokud je determinant jeho matice invertibilní skalár, tj. pokud je nenulový. Dokázali jsme proto nenulovost tzv. **Vandermondova determinantu**

$$V(x_0, \dots, x_n) = \det \begin{pmatrix} 1 & x_0 & (x_0)^2 & \dots & (x_0)^n \\ 1 & x_1 & (x_1)^2 & \dots & (x_1)^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & (x_n)^2 & \dots & (x_n)^n \end{pmatrix} = \prod_{i>k=0}^{n-1} (x_i - x_k).$$

Problémy interpolace

Uvažujme reálné nebo případně racionální polynomy, tj. polynomiálně zadané funkce $\mathbb{R} \to \mathbb{R}$ nebo $\mathbb{Q} \to \mathbb{Q}$. Vypadají jako hezká třída funkcí jedné proměnné, kterou můžeme snadno použít na proložení jakékoliv sady předem zadaných hodnot. Bohužel:

- Vyjádření tzv. **Lagrangeova interpolačního polynomu** je velmi citlivé na nepřesnosti výpočtu při malých rozdílech zadaných hodnot x_i, protože se v ní těmito rozdíly dělí.
- Přímé řešení soustav rovnic by vyžadovalo čas úměrný n^3.
- Interpolační polynomy mají velice špatnou stabilitu hodnot reálných nebo racionálních polynomů při zvětšující se hodnotě proměnné.

Hodnoty polynomů s rostoucí proměnnou rychle míří k nekonečným hodnotám a navíc se uvnitř intervalu daného největší a nejmenší hodnotu z množiny $\{x_0, \dots, x_n\}$ mohou chovat dosti *divoce* . Mohlo by se ale zdát, že podstatně lepší výsledky budeme alespoň mezi body x_i dosahovat, když si budeme kromě hodnot funkce hlídat, jak rychle naše funkce v daných bodech rostou. Zavedeme proto (prozatím spíše intuitivně) pojem **derivace** pro reálné, komplexní nebo racionální polynomy.
Rychlost růstu v bodě $x \in \mathbb{R}$ pro reálný polynom $f(x)$ dobře vyjadřují podíly

$$\frac{f(x + \Delta x) - f(x)}{\Delta x}$$

a protože umíme spočíst (nad libovolným okruhem)

$$(x + \Delta x)^k = x^k + kx^{k-1}\Delta x + \cdots + \binom{k}{l} x^l (\Delta x)^{k-l} + \cdots + (\Delta x)^k,$$

umíme i spočíst zmiňovaný podíl pro $f(x) = a_n x^n + \cdots + a_0$.

$$\frac{f(x+\Delta x)-f(x)}{\Delta x} = a_n\frac{nx^{n-1}\Delta x+\cdots+(\Delta x)^n}{\Delta x}+\cdots+a_1\frac{\Delta x}{\Delta x}$$
$$= na_nx^{n-1}+(n-1)a_{n-1}x^{n-2}+\cdots+a_1+\Delta x(\dots),$$

kde výraz v závorce je polynomiálně závislý na Δx.
Evidentně pro hodnoty Δx velice blízké nule dostaneme hodnotu libovolně blízkou výrazu

$$f'(x) = na_nx^{n-1}+(n-1)a_{n-1}x^{n-2}+\cdots+a_1,$$

který nazýváme **derivace polynomu** $f(x)$ podle proměnné x.

Z definice je jasné, že $f'(x_0)$ dává dobré přiblížení pro chování $f(x)$ v okolí bodu x_0. Přesněji řečeno, přímka

$$y = f'(x_0)(x - x_0) + f(x_0)$$

velice dobře aproximuje přímky procházející body $[x_0, f(x_0)]$ a $[x_0 + \Delta x, f(x_0 + \Delta x)]$ pro malé hodnoty Δx. Hovoříme o **lineárním přiblížení** polynomu f jeho **tečnou**.
Derivace polynomů je lineární zobrazení, které přiřazuje polynomům stupně nejvýše n polynomy stupně nejvýše $n - 1$. Iterací této operace dostáváme druhé derivace f'', třetí derivace $f^{(3)}$ a obecně po k–násobném opakování polynom $f^{(k)}$ stupně nejvýše $n - k$. Po $n + 1$ derivacích je výsledkem nulový polynom.

Hermiteův interpolační problém

Uvažme $m+1$ po dvou různých $x_0, \dots, x_m$ a předepišme hodnoty a derivace $y_i^{(k)}$ pro $k=0$ a $k=1$. Hledáme $f(x) = a_n x^n + \dots a_0$ s těmito hodnotami a derivacemi.

Opět obdržíme pro neznámé koeficienty a_i systém rovnic

$$
\begin{aligned}
a_0 + x_0 a_1 + \cdots + (x_0)^n a_n &= y_0 \\
&\vdots \\
a_0 + x_m a_1 + \cdots + (x_m)^n a_n &= y_m \\
a_1 + 2x_0 a_2 + \cdots + n(x_0)^{n-1} a_n &= y_0' \\
&\vdots \\
a_1 + 2x_m a_2 + \cdots + n(x_m)^{n-1} a_n &= y_m'.
\end{aligned}
$$

Při volbě $n = 2m+1$ bude determinant tohoto systému rovnic nenulový. Opět lze také zkonstruovat takový polynom f přímo. Nazýváme jej **Hermiteův interpolační polynom**.

Úplně nejjednodušší případ je zadání hodnoty a derivace v jediném bodě. Tím určíme beze zbytku polynom stupně 1

$$f(x) = f(x_0) + f'(x_0)(x - x_0)$$

tj. právě rovnici přímky zadané hodnotou a směrnicí v bodě x_0. Když zadáme hodnotu a derivaci ve dvou bodech, tj. $y_0 = f(x_0)$, $y_0' = f'(x_0)$, $y_1 = f(x_1)$, $y_1' = f'(x_1)$ pro dva různé body x_i, dostaneme ještě pořád poměrně snadno řešitelný problém.

Ukažme si jej v zjednodušeném provedení, kdy $x_0 = 0$, $x_1 = 1$. Pak matice systému a její inverze budou

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 3 & 2 & 1 & 0 \end{pmatrix}, \; A^{-1} = \begin{pmatrix} 2 & -2 & 1 & 1 \\ -3 & 3 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Přímým vynásobením $A \cdot (y_0, y_1, y_0', y_1')^T$ pak vyjde vektor $(a_3, a_2, a_1, a_0)^T$ koeficientů polynomu f, tj.

$$f(x) = (2y_0 - 2y_1 + y_0' + y_1')x^3 + (-3y_0 + 3y_1 - 2y_0' - y_1')x^2 + y_0' x + y_0.$$

Formuli lze snadno upravit pro libovolné body x_0, x_1 a lze tak počítat aproximace funkcí *po kouskách*.
Obdobně lze předepisovat libovolný konečný počet derivací v jednotlivých bodech a vhodnou volbou stupně polynomu obdržíme vždy jednoznačné interpolace.

I u polynomiálních interpolací s derivacemi pořád zůstávají problémy zmíněné už v případě jednoduchých interpolací hodnot – složitost výpočtů a nestabilita. Navíc přibývá problém spojený s odhadem derivací pokud je zadána pouze množina hodnot.
Použití derivací však podbízí jednoduché vylepšení metodiky – **splajny**.
Nabízí se tedy využití malých polynomiálních kousků, které ale musíme umět rozumně navazovat.
Nejjednodušší je propojení vždy dvou sousedních bodů lineárním polynomem. Tak se nejčastěji zobrazují data. Z pohledu derivací to znamená, že budou na jednotlivých úsecích konstantní a pak se skokem změní.

O něco sofistikovanější možností je předepsat v každém bodě hodnotu a derivaci, tj. pro dva body budeme mít 4 hodnoty a jednoznačně tím určíme Hermiteův polynom 3. stupně, viz výše. Tento polynom pak můžeme použít pro všechny hodnoty nezávislé proměnné mezi krajními hodnotami $x_0 < x_1$. Hovoříme o **intervalu** $[x_0, x_1]$. Takové polynomiální přiblížení po kouskách už bude mít tu vlastnost, že derivace na sebe budou navazovat.

V praxi ale není pouhé navazování první derivace dostatečné (viz třeba koleje tramvají) a navíc při naměřených datech nemíváme hodnoty derivací k dispozici. Přímo se proto vnucuje pokus využívat pouze zadané hodnoty ve dvou sousedních bodech, ale požadovat zároveň rovnost prvních i druhých derivací u sousedních kousků polynomů třetího stupně!

Kubický interpolační splajn

Definition

Nechť $x_0 < x_1 < \cdots < x_n$ jsou reálné (nebo racionální) hodnoty, ve kterých jsou zadány požadované hodnoty $y_0, \ldots, y_n$. **Kubickým interpolačním splajnem** pro toto zadání je funkce $S : \mathbb{R} \to \mathbb{R}$ (nebot $S : \mathbb{Q} \to \mathbb{Q}$), která splňuje následující podmínky:

- zúžení S na interval $[x_{i-1}, x_i]$ je polynom S_i třetího stupně, $i = 1, \ldots, n$
- $S_i(x_{i-1}) = y_{i-1}$ a $S_i(x_i) = y_i$ pro všechny $i = 1, \ldots n$,
- $S_i'(x_i) = S_{i+1}'(x_i)$ pro všechny $i = 1, \ldots, n-1$,
- $S_i''(x_i) = S_{i+1}''(x_i)$ pro všechny $i = 1, \ldots, n-1$.

Kubický splajn pro $n+1$ bodů sestává z n kubických polynomů, tj. máme k dispozici $4n$ volných parametrů (první definiční podmínka). Další podmínky přitom zadávají $2n+(n-1)+(n-1)$ rovností, tj. dva parametry zůstávají volné. Při praktickém použití se dodávají předpisy pro první derivace v krajních bodech (tzv. **úplný splajn**) nebo jsou druhé derivace zadány jako nula (tzv. **přirozený splajn**). Výpočet celého splajnu už není bohužel tak jednoduchý jako u nezávislých výpočtů Hermiteových polynomů třetího stupně, protože data se prolínají vždy mezi sousedními intervaly. Výpočty splajnů jsou však základem takřka všech grafických balíčků pracujících s křivkami, proto je pochopení principu jejich fungování velmi důležité. Ti z vás, kteří tíhnout k počítačové grafice, se s tímto pojmem určitě ještě setkají.

Aproximace je narozdíl od interpolace postup, který bere ohled na to, že pracujeme s potenciálně nepřesnými vstupními daty, a nesnaží se proto *trefit* přesně do zadaných bodů, ale výstupem je funkce, která má ze zadané třídy funkcí (ve vhodném smyslu) nejmenší vzdálenost od zadaných bodů. Častým případem je rovněž situace, kdy řešíme tzv. **přeurčenou soustavu rovnic**, tj. máme více rovnic než neznámých (např. z výše uvedených důvodů nechceme aproximovat $n+1$ daných bodů hodnotami polynomu stupně n ale stupně nižšího).

Metoda nejmenších čtverců

Metoda nejmenších čtverců je založena na tom, že hledáme funkci z dané množiny (např. lineární polynomy, kvadratické polynomy, polynomy stupně nejvýše n, ale i mnohé jiné funkce v závislosti na zvoleném modelu), jejíž hodnoty v daných bodech $x_1, \ldots, x_n$ mají nejmenší **součet druhých mocnin vzdáleností** od zadaných hodnot $y_1, \ldots, y_n$.
Tato metoda se velmi často objevuje ve zejména ve statistice (regresní analýza).
Ukažme si použití této metody v nejjednodušším případě, kdy máme dáno n bodů $([x_1, y_1], \ldots, [x_n, y_n])$ a hledáme přímku, která nejlépe *vystihuje* rozložení těchto bodů.

Hledáme tedy funkci tvaru $f(x) = a \cdot x + b$ s neznámými $a, b \in \mathbb{R}$ tak, aby hodnota

$$\sum_{i=1}^{n} (f(x_i) - y_i)^2$$

byla minimální.

S pomocí odhadů nebo základních metod diferenciálního počtu (toho budeme schopni za několik týdnů) lze snadno odvodit následující tvrzení.

Věta

Mezi přímkami tvaru $f(x) = a \cdot x + b$ má nejmenší součet čtverců vzdáleností funkčních hodnot v bodech $x_1, \ldots, x_n$ od hodnot y_i funkce splňující

$$a \sum x_i^2 + b \sum x_i = \sum x_i y_i$$
$$a \sum x_i + b \cdot n = \sum y_i$$