

SEMESTRÁLNÍ PRÁCE X.

Aproximace křivek

Numerické vyhlazování

Praha, leden 1999

Velice častým případem v experimentální chemii je sledování určitého jevu formou měření hodnoty jedné veličiny (Y) v závislosti na nastavované hodnotě jiné veličiny (X). Pokud chceme zjistit hodnotu měřené veličiny Y odpovídající nastavené hodnotě X jiné, než bylo experimentálně změřeno, je nutno použít interpolačních technik. Tyto techniky se rovněž používají při numerické derivaci a integraci.

Chceme-li eliminovat náhodné chyby v daném souboru dat, lze použít aproximačních technik a to i např. při numerické derivaci a integraci zatížené experimentálními chybami. Při zpracování rozsáhlých souborů dat lze experimentální diskrétní data nahradit hladkými matematicky definovanými funkcemi zachovávajícími informace obsažené v původních experimentálních datech. Tyto funkce potom značně zmenšují nároky na kapacitu paměti počítačů. Aproximačních technik lze využít při tvorbě speciálních empirických modelů, které nelze jednoduše zkonstruovat pomocí regresních metod.

Grafickou aplikací aproximačních a vyhlazovacích technik je např. prokládání křivky experimentálními body pro grafickou ilustraci daného jevu.

Jako příklad bude níže zpracováno UV spektrum neznámého analytu, které bylo změřeno experimentálně chromatografickým detektorem s diodovým polem při jeho HPLC separaci (viz. úloha 5)

Společným cílem těchto metod je proložení vhodné funkce zadanými body v prostoru.

Při interpolaci hledáme hodnotu Y odpovídající nastavené hodnotě X jiné, než byla experimentálně stanovena (tabelována), přičemž nastavená hodnota X leží v intervalu ohraničeném známými (např. experimentálně změřenými, tabelovanými) hodnotami X,Y. Předpokládá se, že hledané hodnoty Y jsou nenáhodné veličiny.

Jsou-li dány body v prostoru (X,Y) a cílem metody je nahradit tyto body v celém rozsahu vhodnou funkcí $F(X)$, nazývá se metodika aproximace. Jednou z oblastí

využití aproximujících funkcí je numerického vyhlazování, které umožňuje odstranit náhodné šumy experimentálně zkonstruované závislosti. Cílem není nalezení přesně definované matematické funkce popisující chování systému, jde o nalezení nejtěsnější bezšumové rekonstrukce experimentální závislosti.

Neparametrická regrese je metoda, která nahrazuje aproximační postupy v případech, že je znám tvar funkce, která má odpovídat daným (naměřeným) hodnotám.

Úloha 3

Kubický spline

Interpolační i aproximační spline-techniky jsou vhodné pro popis dějů, které vykazují proměnlivé chování v průběhu sledovaného intervalu nastavované hodnoty X . Sledovaný interval je tak rozdělen na části, uvnitř kterých odpovídá chování systému jedinému modelu (funkci), zatímco při přechodu z jednoho intervalu do druhého se parametry modelu (funkce) mění a to mnohdy radikálně. V jednotlivých částech měřeného intervalu jsou tak definovány lokální funkce (např. polynommické), které jsou v místech styku sousedních částí spojitě ve funkčních hodnotách a v hodnotách zadaných derivací. Tak vznikne při proložení danými body hladká a spojitá křivka.

Kubický spline je funkcí třídy $C^m(a,b)$, která je definovaná na pěti uzlových bodech jako úseky popsané polynomem třetího řádu, které jsou spojitě v těchto uzlových bodech ve funkčních hodnotách a v hodnotách prvních dvou derivací.

Obecně lze kubický spline popsat funkcí (vztah 9.31 v povinné literatuře):

$$P_j(x) = c_0 + \sum_{k=1}^3 c_k \cdot (x - \xi_i)^k$$

kde c_k označuje koeficienty, které lze určit pomocí nezávislých podmínek (podmínka interpolace, podmínky spojitosti a okrajové podmínky)

ξ_i definuje uzlové body, tj. body, ve kterých dochází ke změně charakteru polynomu

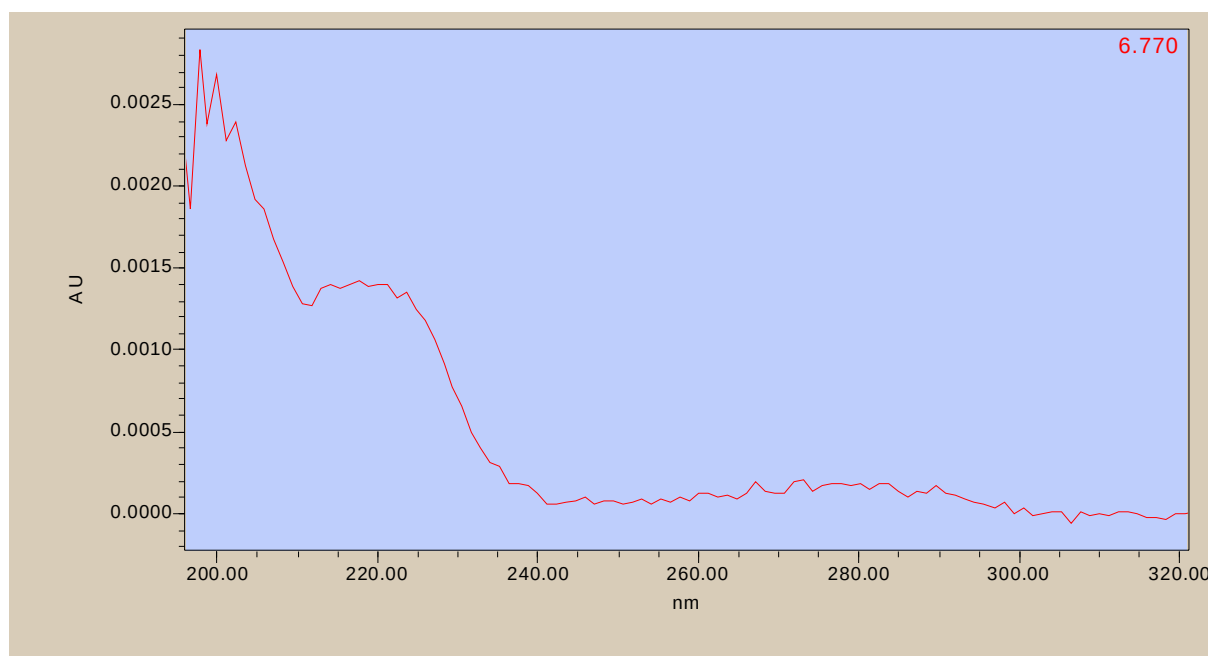
Kubický spline lze nejlépe vypočítat jako (vztah 9.39 v povinné literatuře):

$$S_3(x) = \sum_{j=1}^{n+4} c_j \cdot B_{4,j}(x)$$

Interpolační i aproximační spline-techniky jsou vhodné pro popis dějů, které vykazují proměnlivé chování v průběhu sledovaného intervalu nastavované hodnoty X . Sledovaný interval je tak rozdělen na části, uvnitř kterých odpovídá chování systému jedinému modelu (funkci), zatímco při přechodu z jednoho intervalu do druhého se parametry modelu (funkce) mění. V jednotlivých částech měřeného intervalu jsou nejčastěji definovány lokální polynomické funkce, které jsou v místech styku sousedních částí spojité ve funkčních hodnotách a v hodnotách zadaných derivací. Tyto polynomické funkce jsou označovány jako funkce třídy $C^m(a,b)$. Vlastnosti těchto funkcí závisí na řádu polynomu, počtu a polohách uzlů a případných defektech v uzlových bodech. Konkrétním příkladem takové funkce je kubický spline popsáný v příkladu 3.

Při HPLC analýze směsi bylo experimentálně změřeno chromatografickým detektorem UV spektrum neznámého analytu v oblasti vlnových délek 195 nm až 312 nm. Pro identifikaci tohoto analytu je podstatný tvar spektra a lokální maxima absorbance v daném intervalu vlnových délek.

Sledovaný analyt se ve směsi vyskytuje ve velmi nízké koncentraci. Jeho spektrum je tak zřetelně ovlivněno experimentálními šумы, jak je patrné z následujícího obrázku:



Pro vyhlazení tvaru spektra a následnou derivaci je k dispozici software Adstat 1.25, který nabízí tři možnosti spline vyhlazování.

1. Vyhlazování kubickým spline podle Reinschova postupu

Matematický popis tohoto postupu lze nalézt v povinné literatuře, str. 760-772.

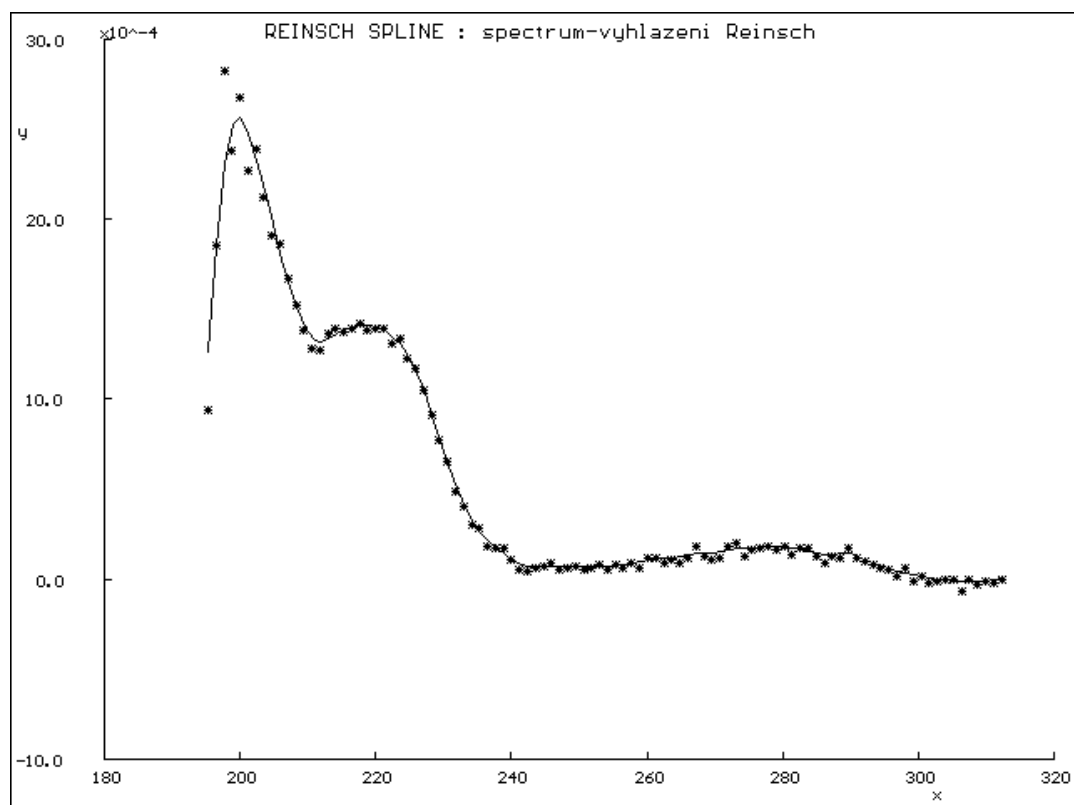
Při praktické aplikaci metody softwarem Adstat 1.25 je třeba najít vhodnou hodnotu parametru S, který určuje poměr mezi hladkostí a jejím přiblížením k experimentálním bodům. Vhodným kritériem pro těsnost proložení je střední kvadratická chyba predikce:

hodnota parametru S	průměr residuí	průměr absolutních hodnot residuí	reziduální rozptyl	vizuální hodnocení
10^{-8}	3.7E-21	4.5E-6	1E-10	příliš kopíruje body
10^{-3}	-7.0E-21	3.4E-4	1.8E-7	proložena přímka
10^{-5}	-2.2E-20	2.4E-4	1.0E-7	proložena jednoduchá křivka
10^{-7}	-2.3E-21	1.4E-5	1.0E-9	lepší proložení, stále příliš kopíruje body
10^{-6}	4.4E-21	3.9E-5	1.0E-8	hladké proložení, ale neodpovídající celkové maximum
$5 \cdot 10^{-7}$	-2.5E-21	3.0E-5	5.0E-9	ještě těsnější hladké proložení
$3 \cdot 10^{-7}$	3.5E-21	2.5E-05	3.0E-9	ještě těsnější, ale již méně hladké proložení

Jako vhodné se jeví dvě hodnoty parametru S. Pomocí numerického výstupu z Adstatu lze numericky vyjádřit polohu lokálních maxim, z grafu 1. derivace potom graficky odečíst tyto hodnoty:

hodnota P	lokální maximum: numericky (nm)			lokální maximum: z 1/2. derivace		
$5 \cdot 10^{-7}$	199.7			200nm	220nm	280nm
$3 \cdot 10^{-7}$	198.4	217.6	245.6	200nm	220nm	280nm
	267.3	272.8				

Následující graf prezentuje nejlepší vyhlazení spektra metodou podle Rheinsche:



2. Vyhlazování kubickým spline podle Spaethova postupu

Matematický popis tohoto postupu lze nalézt v povinné literatuře, str. 760-772.

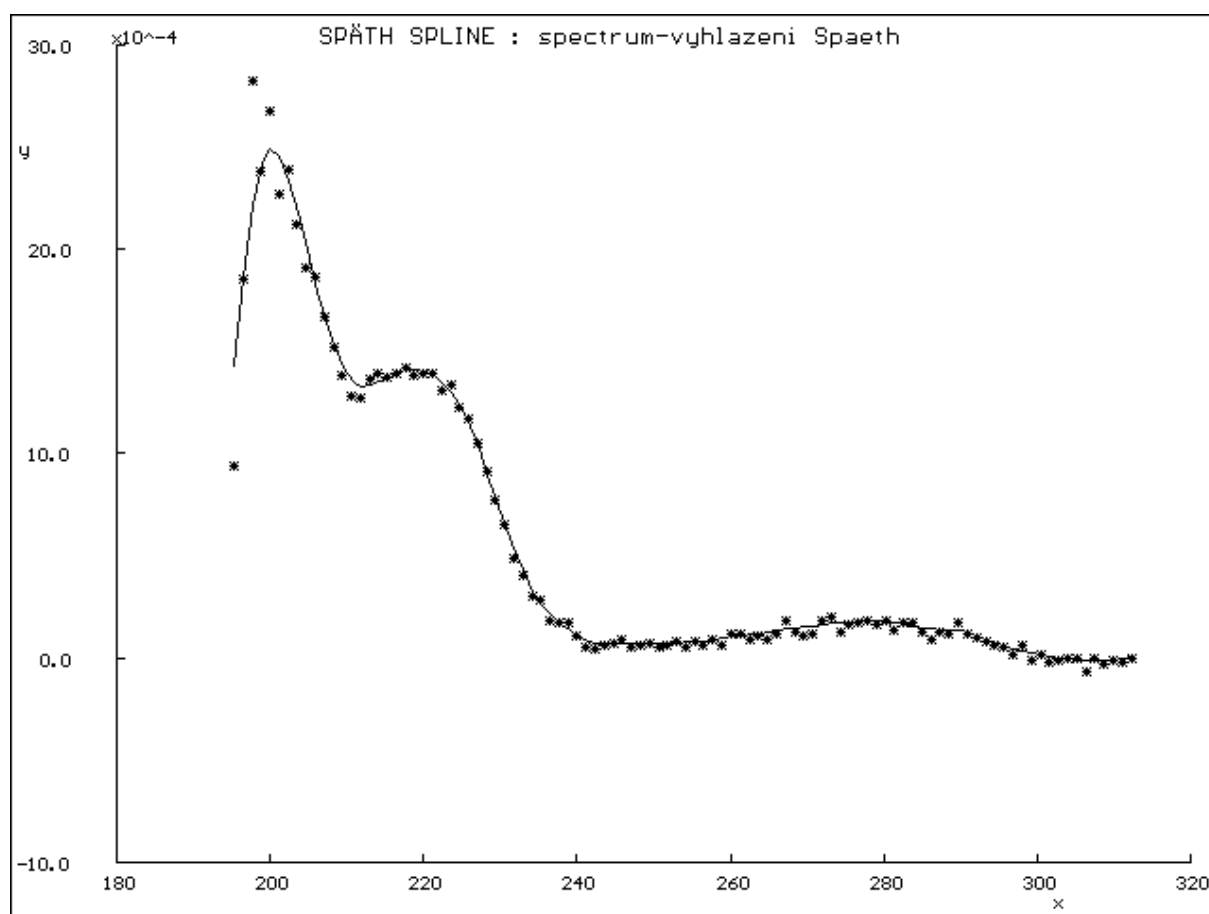
Při praktické aplikaci metody softwarem Adstat 1.25 je třeba najít vhodnou hodnotu parametru P, který určuje poměr mezi hladkostí a jejím přiblížením k experimentálním bodům. Vhodným kritériem pro těsnost proložení je střední kvadratická chyba predikce:

hodnota parametru P	průměr residuí	průměr hodnot residuí	absolutních reziduální rozptyl	vizuální hodnocení
10^{-2}	6.8E-23	5.8E-5	2.0E-8	křivka hladká, nerespektuje však dostatečně všechny body
1	3.8E-21	2.5E-5	3.0E-9	křivka dobře kopíruje průběh bodů, není dostatečně vyhlazena
10^{-1}	3.5E-21	3.4E-5	7.6E-9	křivka hladká, dobře odpovídá průběhu bodů
$5 \cdot 10^{-1}$				křivka lépe kopíruje průběh bodů než při $P=10^{-1}$, hůře vyhlazena než při $P=10^{-1}$

Jako vhodné se jeví dvě hodnoty parametru P. Pomocí numerického výstupu z Adstatu lze numericky vyjádřit polohu lokálních maxim, z grafu 1. derivace potom graficky odečíst tyto hodnoty:

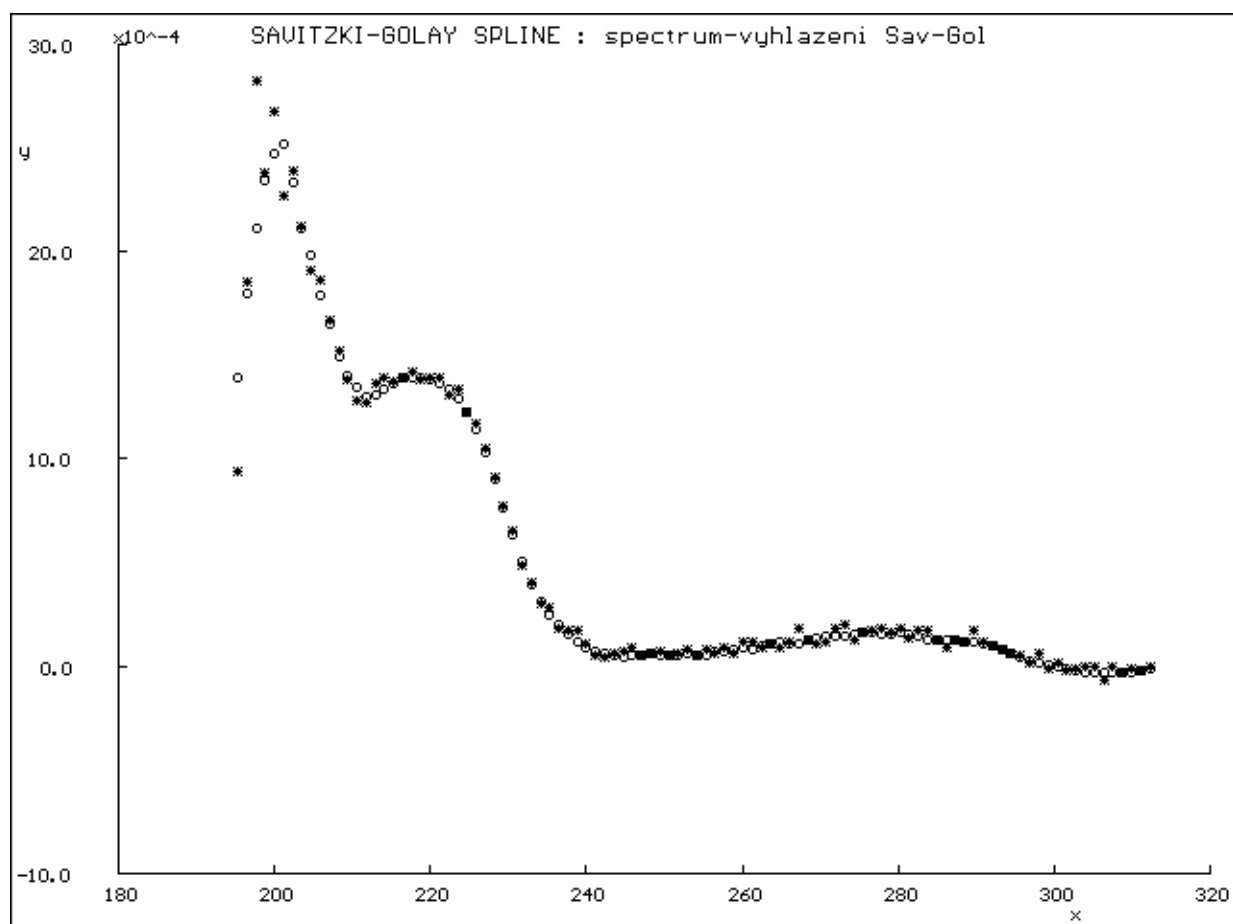
hodnota P	lokální maximum: numericky (nm)			lokální maximum: z 1/2. derivace		
10^{-1}	200.1			200nm	220nm	280nm
$5 \cdot 10^{-1}$	194.5	273.2	289.5	200nm	220nm	280nm

Následující graf prezentuje nejlepší vyhlazení spektra metodou podle Spaetha:



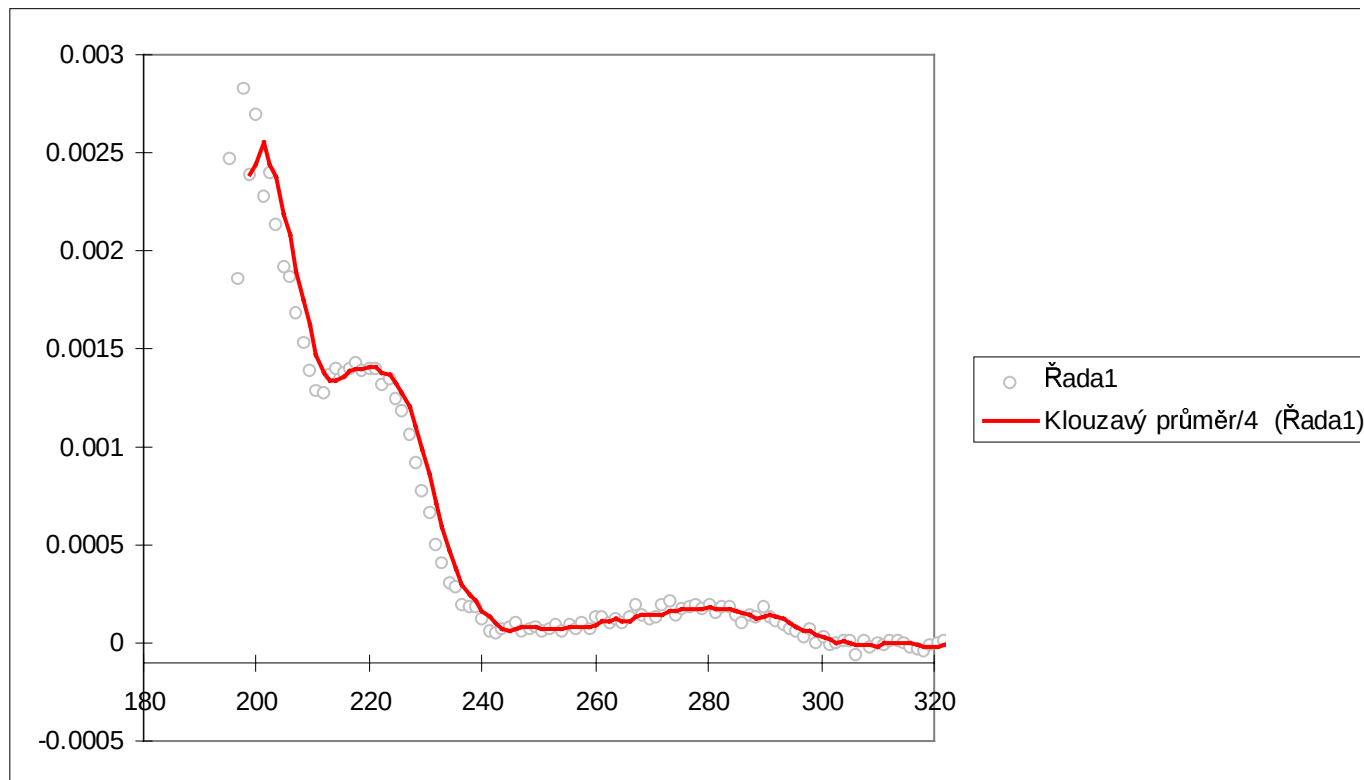
3. Vyhlazování kubickým spline podle postupu Savitzki-Golay

Nejlepšího výsledku bylo dosaženo při zadání parametru M=5:



4. Vyhlazování pomocí software Excel

Pro porovnání s vyhlazením pomocí Adstatu 1.25 následuje použití nejdostupnějšího nástroje, kterým je tabulkový procesor Excel (použita verze 7.0). Jako nejlepší se jeví proložení křivky metodou klouzavého průměru



5. Závěr

Všechny tři použité vyhlazovací techniky softwaru Adstat 1.25 poskytují srovnatelné výsledky. Vyhlazené spektrum lze spolehlivě porovnat se spektry uloženými ve spektrální knihovně a pokusit se tak identifikovat neznámý analyt. Spektrum vykazuje absorpční maximuma při 200 nm, 217nm a 280 nm.

Software Excel, který je rozšířenější, poskytuje větší komfort při aproximačních postupech, výsledek aproximace však není uspokojivý. Chybí podrobnější numerický popis aproximačního postupu.

Název souboru: SEMINA10
Adresář: E:\Pom\approx
Šablona: D:\Program Files\Microsoft Office\Sablony\Normal.dot
Název: Práce pro Prof. Milírkého
Předmět:
Autor: Plynový chromatograf Hewlett-Packard
Klíčová slova:
Komentáře:
Datum vytvoření: 13.01.99 13:33
Číslo revize: 36
Poslední uložení: 15.01.99 09:44
Uložil: Plynový chromatograf Hewlett-Packard
Celková doba úprav: 303 min.
Poslední tisk: 15.09.00 08:27
Jako poslední úplný tisk
Počet stránek: 11
Počet slov: 1 334 (přibližně)
Počet znaků: 7 607 (přibližně)