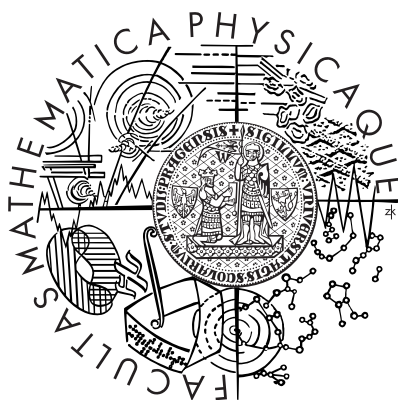


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Pavel Vácha

Registrace obrazu pomocí radiálních bázových funkcí

Ústav teorie informace a automatizace
Akademie věd České republiky

Vedoucí diplomové práce: **Doc. Ing. Jan Flusser, DrSc. ÚTIA AV ČR**

Studijní program: **Informatika, Softwarové systémy**

Studijní plán: **Počítačová grafika**

Tímto bych rád poděkoval svému diplomovému vedoucímu Doc. Ing. Janu Flussrovi, DrSc. za vedení diplomové práce, množství užitečných nápadů, rad a připomínek. Dále bych rád poděkoval Doc. RNDr. Jaromíru Antochovi, CSc. za velmi cenné konzultace. V neposlední řadě bych chtěl poděkovat svým rodičům a blízkým za podporu a trpělivost během celého studia a při psaní této práce.

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne 10. srpna 2003

Pavel Vácha

Obsah

Abstrakt	3
1 Úvod do problému	4
1.1 Registrace obrazu	5
1.2 Konstrukce mapovací funkce	5
1.3 Známé výsledky a motivace	7
1.4 Cíle a výsledky této práce	9
2 Konstrukce mapovací funkce	10
2.1 Interpolační Thin Plate Spline	10
2.2 Aproximační Thin Plate Spline	12
2.3 Řešení mapování pomocí TPS	13
2.3.1 interpolační TPS	13
2.3.2 aproximační TPS	14
2.3.3 shrnutí	15
2.4 Jiné radiální bázové funkce	15
2.4.1 Gaussovy RBF	17
2.5 Jiné modely	17
3 Volba vyhlazovacího parametru	19
3.1 Mean Square Error	20
3.2 Cross Validation	21
3.3 Formulace TPS pomocí hat matrix	22
3.4 Generalized Cross Validation	24
3.5 Robustified Cross Validation	25
3.6 Akaike Information Criterion	26
3.7 Corrected Akaike Information Criterion	27
3.8 Jiné metody	27
4 Navrhovaná metoda	29
4.1 Minimalizace	29
4.2 Mapování	30

5	Experimenty	31
5.1	Interpolace	31
5.2	Aproximace	35
5.3	Kritéria	39
5.3.1	příklad 1	41
5.3.2	příklad 2	45
5.3.3	příklad 3	48
5.3.4	příklad 4	51
5.4	Minimalizace	55
5.5	Mapování	60
5.5.1	příklad 1	60
5.5.2	příklad 2	62
5.5.3	zjištění	63
6	Závěr	64
	Literatura	66

Název práce: Registrace obrazu pomocí radiálních bázových funkcí

Autor: Pavel Vácha

Katedra (ústav): Ústav teorie informace a automatizace, AV ČR

Vedoucí diplomové práce: Doc. Ing. Jan Flusser, DrSc. ÚTIA AV ČR

e-mail vedoucího: flusser@utia.cas.cz

Abstrakt: Tato práce zkoumá registraci obrazu, přesněji konstrukci mapovací funkce, pomocí Gaussových funkcí, které mají výrazný lokální charakter, narozdíl od používané báze Thin Plate Spline (TPS). Dále se práce zabývá automatickou volbou vyhlazovacího parametru. Je zde navržena vlastní metoda pro volbu parametrů, ve které se pomocí zvoleného kritéria určují nejvhodnější hodnoty všech volných parametrů v modelu. Porovnávána byla kritéria: Generalized Cross Validation (GCV), Cross Validation (CV), Robustified Cross Validation (RCV), Akaike Information Criterion (AIC) a Corrected Akaike Information Criterion (cAIC). Nakonec bylo zvoleno GCV kritérium, protože dosahuje stabilně nejlepších výsledků a to i pro malý počet vstupů. Ukázalo se, že mapování založené na Gaussových RBF s navrženou metodou překoná při registraci lokálně deformovaných obrázků dokonce i TPS. Navrhovaná metoda je vhodná pro registraci snímků s lokální deformací, příkladem může být lékařství. Další oblasti využití by mohly být warping a aproximace dvourozměrné funkce.

Klíčová slova: registrace obrazu, Gaussovy radiální bázové funkce, volba vyhlazovacího parametru, Generalized Cross Validation

Title: Image registration by means radial basis functions

Author: Pavel Vácha

Department: Institute of Information Theory and Automation, AS CR

Supervisor: Doc. Ing. Jan Flusser, DrSc. ÚTIA AV ČR

Supervisor's e-mail address: flusser@utia.cas.cz

Abstract: This work studies image registration, in particular the construction of the mapping. For this purpose, we will use a Gaussian function, which has intrinsic localized behavior, in contrary to the commonly used Thin Plate Spline basis. We will focus on automatic smoothing parameter selection and introduce a new method in which we determine the best values for all free parameters according to the chosen criterion. We consider the following criteria: Generalized Cross Validation (GCV), Cross Validation (CV), Robustified Cross Validation (RCV), Akaike Information Criterion (AIC), and Corrected Akaike Information Criterion (cAIC). Finally, we choose GCV criterion since it yields the best results even for small datasets. It turns out that mapping based on Gaussian functions with proposed method outperforms the TPS in the registration of locally deformed images. The proposed method is suitable for image registration with local deformation, e.g. in medical applications. It could be also used for warping and approximation of two-dimensional functions. Keywords: image registration, Gaussian radial basis function, smoothing parameter selection, Generalized Cross Validation

Kapitola 1

Úvod do problému

Registrace obrazu jako vědní disciplína patří do oblasti zpracování obrazu. Přestože se již poměrně dlouhou dobu uplatňuje v celé řadě aplikací, neustále se vyvíjí a hledají se nové přístupy pro její zdokonalení.

Tato diplomové práce se zaměřuje na jeden ze čtyř kroků registrace obrazu, a to na konstrukci mapovací funkce. Konkrétně se snaží prozkoumat jiné možnosti konstrukce mapování než jsou používané Thin Plate Spline (TPS), jedná se o Gaussovy radiální funkce, které by měli přinést některé odlišné vlastnosti.

Struktura práce je následující. V úvodu jsou nejprve vysvětleny základní pojmy registrace obrazu a je zde popsáno, kde se při registraci radiální báze funkce využívají. Dále jsou naznačeny možné problémy a ukázána motivace této práce; úvod uzavírají cíle a výsledky.

První část zabývající se známou teorií obsahuje dvě kapitoly. Kapitola 2 popisuje konstrukci mapování pomocí radiálních báze funkcí a uvádí s tím související teorii včetně odkazů na podrobnější informace, jiné aplikace a možnosti řešení mapování. Kapitola 3 se zabývá možnostmi volby vyhlazovacího parametru a vysvětluje kritéria, která je možné pro automatickou volbu tohoto parametru použít.

Druhá část obsahující výsledky se skládá také ze dvou kapitol. Kapitola 4 popisuje navrhovanou metodu více kriteriální optimalizace pro Gaussovy radiální báze funkce. Kapitola 5 předvádí chování mapování pomocí TPS a Gaussových RBF, porovnává vhodnost kritérií pro volbu vyhlazovacího parametru a ukazuje výsledky navrhované metody.

Poslední kapitola shrnuje dosažené výsledky a naznačuje možné oblasti využití. Součástí práce je CD s experimenty a zdrojovými soubory v prostředí MATLAB.

1.1 Registrace obrazu

Co skrývá pojem registrace obrazových dat? Jedná se o nalezení geometrické transformace pro následné použití s referenčním snímkem. V úloze o registraci máme dva snímky. První - referenční, který se nemění, a druhý, který chceme transformovat tak, aby jej bylo možné složit dohromady s referenčním snímkem.

Mějme například snímky zemského povrchu, ty jsou při pořízení neúmyslně zdeformovány (zakřivením zemského povrchu, nerovnoměrným letem letadla a pod.). Potřebujeme tuto deformaci odstranit, abychom mohli snímky složit do jednoho většího a zhotovit například mapu. Referenčním snímkem v tomto případě může být stará mapa nebo nedeformovaný snímek.

Jinými příklady jsou registrace fotografií stejného místa, ale pořízené s jistým časovým odstupem, což umožňuje sledovat vzniklé změny. Také se registrují snímky pořízené v různých částech spektra nebo výstupy z rozdílných metod.

Registrace se skládá z následujících částí:

1. *Výběr řídicích bodů*

Vybírají se význačné body, linie, plochy a to v referenčním i v připojovaném snímku. Proto by tyto objekty měly existovat v obou snímcích a být tedy například stabilní v čase, případně se projevovat ve velké části spektra. Záleží na tom, jaké změny se mezi snímky očekávají. Výběr bývá realizován buď ručně nebo algoritmičky.

2. *Stanovení korespondence bodů*

Stanoví se korespondence řídicích bodů z referenčního snímku s body z připojovaného snímku. Využívá se míry podobnosti, prostorových vztahů mezi body, případně symbolických popisů objektů.

3. *Konstrukce mapovací funkce*

Na základě korespondence řídicích bodů se konstruuje mapovací funkce. Typicky se zvolí model funkce a volné parametry se dopočítají z korespondence bodů.

4. *Převzorkování a transformace připojovaného obrazu*

Pomocí funkce zkonstruované v předchozím bodě se provede transformace připojovaného obrazu. Při převzorkování může dojít k interpolaci, případně k průměrování.

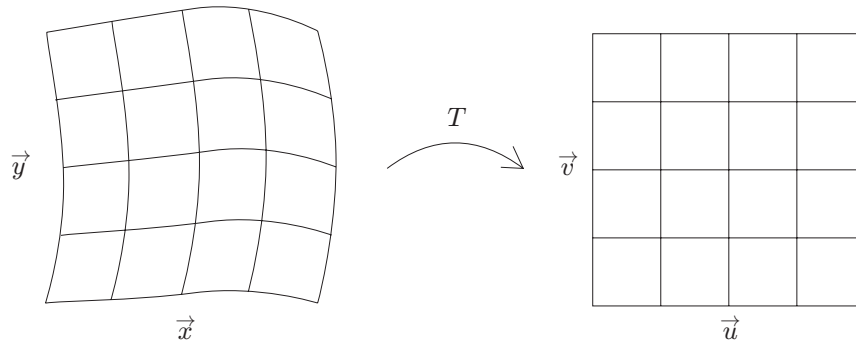
1.2 Konstrukce mapovací funkce

Vstupem této úlohy jsou korespondující body z referenčního a z přidávaného snímku. Mějme dvě souřadné soustavy, v referenčním snímku to bude soustava určená vektory (\vec{u}, \vec{v}) , v přidávaném snímku potom soustava (\vec{x}, \vec{y}) .

Korespondující body pak vytvoří páry (x_i, y_i) a (u_i, v_i) pro $i = 1, 2, \dots, N$. Úloha tedy zní sestavit zobrazení $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ splňující interpolační podmínku:

$$T(x_i, y_i) = (u_i, v_i) \quad \forall i = 1, 2, \dots, N \quad (1.1)$$

Tato funkce by měla být v určitém smyslu rozumná, například spojitá, hladká, bez náhlých výkyvů a přitom dobře vystihovat lokální vlastnosti, záleží však na konkrétním případě. V případě šumu nebo nevěrohodných dat se uvolňuje interpolační podmínka, přidává se regularizační člen a hledá se funkce aproximující.



Obrázek 1.1: Náčrt mapovací funkce.

Typicky se zvolí nějaký model mapovací funkce, a potom se na základě korespondence bodů dopočítají parametry modelu. Model musíme určit a priori na základě znalosti problému, podle toho jakou transformaci očekáváme. Modely se dělí na globální a lokální. Globální model má parametry platné pro celý snímek, narozdíl od lokálního modelu, kde se mapovací funkce určuje pro každou část zvlášť. Potom se tyto lokální funkce napojují, aby utvořily celkové mapování. Nastává zde však problém se spojitostí na okrajích lokálních mapování.

Příklady mapovacích funkcí jsou:

- jednoduché modely jako otočení, posunutí, zvětšení a zmenšení
- afinní a projektivní transformace
- lokální model založený na triangulaci, mapování se vyřeší pro trojúhelníčky, které se potom lineárně nebo kubicky spojují
- Thine Plane Spline (TPS), spline funkce napodobující chování tenkého plátu

Pro vystihnutí složitějších transformací se často používají právě TPS. Protože vstupní řídicí body jsou, jako i jiná data, obvykle zatíženy chybou (způsobenou například šumem), uvolňuje se podmínka interpolace a hledá se aproximující TPS. Mapovací funkce potom nemusí přesně splňovat rovnici (1.1).

1.3 Známé výsledky a motivace

Při používání TPS se objevují následující nevýhody. Je potřebné znát poměrně hodně řídicích bodů, neboť při malém počtu řídicích bodů může posunutí jednoho z nich ovlivnit i dosti vzdálené oblasti. Tento problém popisují ukazují Arad a spol. v článku [2], podle tohoto článku jsou také sestrojeny obrázky 1.2, které tyto vlastnosti TPS demonstrují.

Jak je také ukázáno v článku [2], při použití jiné báze než TPS, konkrétně Gaussových funkcí tento problém nenastává. Tyto funkce jsou totiž, narozdíl od báze TPS, hodně lokální, což může být někdy výhodné. Rozdílné chování báze z Gaussových funkcí ukazuje dvojice obrázků 1.3.

Tímto vznikla motivace prozkoumat jiné radiální bázové funkce než TPS, obzvláště ty, které mají více lokální charakter. Tyto báze by měly odstranit potřebu zbytečně mnoha řídicích bodů, nutných pro TPS. Zadávání těchto bodů je nejen obtížné, ale také zvyšuje výpočetní nároky.

Nyní je potřeba naznačit něco z teorie stojící za TPS, podrobný popis teorie TPS následuje v kapitole 2. Mapovací funkce se u TPS hledá pomocí dvou funkcí, první z nich minimalizuje výraz

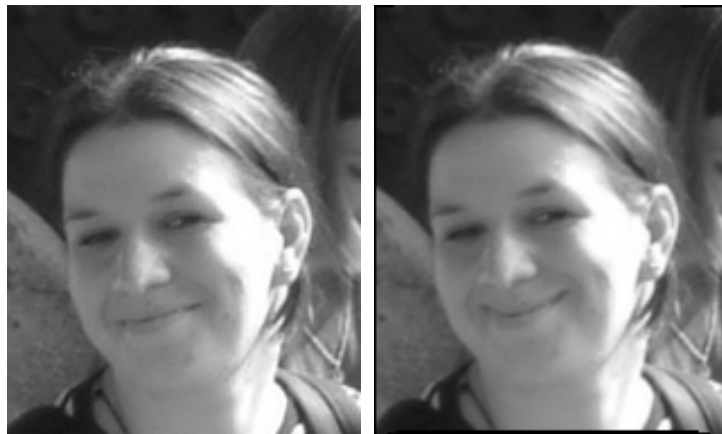
$$\tilde{J}_\lambda(f) = \sum_{i=1}^N [f(x_i, y_i) - u_i]^2 + \lambda\Theta(f), \quad (1.2)$$

druhá potom výraz analogický (v_i místo u_i). Přitom funkcionál $\Theta(f)$ popisuje míru křivosti funkce f . Nezáporný parametr λ se nazývá vyhlazovací parametr a určuje, zda klademe důraz na přesnost aproximace nebo na míru křivosti funkce. Vhodnou volbou tohoto parametru lze odstranit šum v datech, a přitom nezkruslit mapovací funkci. Vyhlazovací parametr se zkouší ručně nebo se automaticky volí z dat, k čemu se běžně používá metoda nazývaná Generalized Cross Validation (GCV).

Nevýhodou, která se někdy u GCV uvádí, je, že při malé změně vstupních dat může dojít k poměrně velké změně hodnoty λ , kterou metoda označí za optimální; tato situace může nastat obzvláště při menším počtu vstupních dat. Dále, jak popisují Robinson a Moyeed v článku [20], GCV v některých případech volí λ menší než by bylo optimální, opět spíš při menším počtu vstupních dat. Proto vznikla druhá motivace, vyzkoušet jiné metody, které se používají ve statistice pro volbu vyhlazovacího parametru, jestli nejsou pro konstrukci mapovací funkce vhodnější než běžně používaná GCV.



Obrázek 1.2: Ukázka mapování pomocí TPS – vidíme nepřírozené nafouknutí dolní části obličeje, řídicí body byly umístěny do rohů, dále středu a kraji úst, změněna byla pouze poloha bodů na ústech.



Obrázek 1.3: Ukázka mapování pomocí Gaussových funkcí – působí přirozeněji než TPS, řídicí body byly umístěny do rohů, dále středu a kraji úst, změněna byla pouze poloha bodů na ústech.

1.4 Cíle a výsledky této práce

Jak napovídala předchozí motivační sekce, cílem této práce je prozkoumat možnosti registrace pomocí jiných radiálních bázových funkcí než TPS, konkrétně Gaussových funkcí, které používají Arad a spol. [2]. S tím také souvisí určit nejvhodnější metodu jak automaticky volit vyhlazovací parametr.

Práce zahrnuje rešerši existujících metod registrace pomocí RBF s přihlédnutím k jiným funkcím než jsou TPS. Je zde navrhována vlastní metoda pro automatickou volbu parametrů, jedná se o více kritériální optimalizaci, při které se hledá optimální hodnota všech volných parametrů v modelu. Tyto „optimální“ hodnoty se určují pomocí kritéria, které odhaduje vhodnost daného parametru.

Dále jsou popsána nastudovaná kritéria, která se ve statistice používají pro volbu vyhlazovacího parametru, kromě běžně používané GCV jsou to: Cross Validation (CV), Robustified Cross Validation (RCV), Akaike Information Criterion (AIC) a Corrected Akaike Information Criterion (cAIC). Na základě provedených experimentů bylo jako nejvhodnější kritérium pro Gaussovy funkce nakonec zvoleno kritérium GCV, neboť funguje dobře i pro malý počet vstupů, má nejméně problému s existencí minima a produkuje nejlepší výsledky (měřeno pomocí střední kvadratické odchylky (MSE)). To odpovídá dřívějším pokusům s kritérii pro TPS, které nevedly, jak uvádí Berman [3], k nalezení lepšího kritéria než obvykle používaného GCV.

Provedené pokusy ukázaly, že Gaussovy RBF s navrhovanou metodou jsou vhodné pro registraci snímku, při kterých dochází k lokální deformaci. Na testovaném obrázku, který byl lokálně zdeformován, Gaussovy RBF s použitím navrhované metody dokonce překonaly TPS.

Kapitola 2

Konstrukce mapovací funkce

Jak již bylo uvedeno v části 1.2, úloha o konstrukci mapovací funkce zní následovně. Máme zadáno N korespondujících dvojic bodů (x_i, y_i) a (u_i, v_i) . Hledáme zobrazení $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ splňující interpolační podmínku:

$$T(x_i, y_i) = (u_i, v_i) \quad \forall i = 1, 2, \dots, N \quad (2.1)$$

Toto zobrazení by mělo co nejlépe vystihovat skutečnou transformaci. Někdy můžeme podmínku (2.1) oslabit, potom mluvíme o aproximaci.

V případě radiálních bazových funkcí (RBF) lze konstrukci mapovací funkce separovat a řešit mapování zvlášť pro každou ze souřadnic x, y . Hledá se potom $T_U, T_V : \mathbb{R}^2 \rightarrow \mathbb{R}$ splňující:

$$T_U(x_i, y_i) = u_i \quad \& \quad T_V(x_i, y_i) = v_i \quad \forall i = 1, 2, \dots, N \quad (2.2)$$

Nejprve se budeme zabývat teorií schovanou za interpolací pomocí RBF a potom ukážeme jak najít řešení. Oboje uvádím pro funkci T_U , pro T_V je situace analogická.

2.1 Interpolační Thin Plate Spline

Úlohou interpolace je nalezení funkce $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, která prochází zadanými body:

$$f(x_i, y_i) = u_i \quad \forall i = 1, 2, \dots, N \quad (2.3)$$

Těchto funkcí je však nekonečně mnoho, proto je potřeba najít způsob, jak určit, která z funkcí je nejlepší. V našem případě to znamená najít funkci, která prochází danými body a přitom se nejméně mění neboli povrch $[x, y, f(x, y)]$ má nejmenší míru křivosti; při fyzikálním pohledu je to povrch s nejmenší potenciální energií.

Najdeme funkcionál, který bude označovat míru vhodnosti funkce, optimální funkce potom bude odpovídat minimální hodnotě funkcionálu. Klíčový

matematický problém spočívá v nalezení takového funkcionálu, který by měl nejen výše uvedené vlastnosti, ale zároveň by zaručoval existenci a jednoznačnost řešení. Postup hledání takových funkcionálů ukazuje Grimson [10], stejně jako následující tvrzení:

Tvrzení 2.1.1 *Nechť \mathcal{F} je semi-Hilbertův prostor všech možných povrchů a $\Theta(f) = \mu(f, f)^{\frac{1}{2}}$ je seminorma, přičemž $\mu(f, f)^{\frac{1}{2}}$ je semiskalární součin prostoru \mathcal{F} , potom existuje jednoznačně určený povrch v \mathcal{F} (případně všechny prvky vektorového podprostoru, kde je norma Θ nulová), který minimalizuje seminormu Θ přes všechny povrchy.*

Funkcionál tedy musí splňovat výše uvedenou podmínku a zároveň by měl měřit velikost změn funkce. Dva velmi přirozené funkcionály, které mají požadované vlastnosti, jsou: první založený na Laplaceovu operátoru

$$\Theta_1(f) = \left[\iint (\nabla^2 f)^2 dx dy \right]^{\frac{1}{2}} \quad (2.4)$$

a druhý obsahující kvadratickou varianci

$$\Theta_2(f) = \left[\iint \left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 dx dy \right]^{\frac{1}{2}} \quad (2.5)$$

Dá se dokonce říci že, ostatních funkcionálů, které splňují výše uvedené podmínky, neexistuje mnoho druhů. Jak opět ukazuje Grimson [10] všechny takové funkcionály jsou lineární kombinací výše uvedených $\Theta_1(f)$ a $\Theta_2(f)$.

V čem se tedy funkcionály liší a který je nejlepší? Jedním z důležitých kritérií je velikost nulového prostoru funkcionálu (vektorový podprostor, kde je funkcionál $\Theta(f)$ nulový). Prvky nulového prostoru nelze pomocí funkcionálu rozlišit, přitom rozdíly mezi funkcemi mohou být velké. Nulový prostor je nejmenší u funkcionálu $\Theta_2(f)$. Pro ostatní funkcionály (lineární kombinace $\Theta_1(f)$ a $\Theta_2(f)$) je jejich nulový prostor vždy větší a obsahuje nulový prostor příslušející k $\Theta_2(f)$. Další výhodou je, že pro $\Theta_2(f)$ je nulový prostor tvořen právě množinou všech lineárních funkcí a hodnota normy $\Theta_2(f)$ je invariantní vzhledem k otočení funkce. Důkazy je opět možné nalézt v článku [10].

Nejvhodnější je tedy funkce minimalizující funkcionál $\Theta_2(f)$, nazývá se *thin plate spline* (TPS) nebo *interpolační thin plate spline*. Pokud jsou součástí zadaných bodů alespoň tři body neležící v přímce, je řešení určeno jednoznačně. Explicitní tvar řešení a postup jeho výpočtu je popsán v sekci (2.3).

Pojmenování *thin plate spline*, známé z mechaniky, je odvozeno od toho, že výraz $\Theta_2(f)$ aproximuje potenciální energii zohýbaného tenkého ocelového plátu. Minimalizace $\Theta_2(f)$ potom odpovídá snaze ocelového plátu zaujmout

tvar s nejmenší potenciální energií (použití TPS v mechanice popisuje Harder [12]).

První, kdo představili interpolační TPS ve zpracování obrazu, byli Grimson [10] a Bookstein [4]. Ačkoliv Grimson tehdy použil TPS pro rekonstrukci povrchu ve stereo vidění, využívá se nyní této metody pro širokou škálu problémů. Například vytváření lebečních implantátů, které popisují Carr a spol. v článku [5], odstranění deformace na snímcích pořízených z letadla představené skupinou Wiemker a spol. v [21] a samozřejmě registrace. Kvůli existenci šumu v získaných datech se však v aplikacích často používá aproximační Thin Plate Spline.

2.2 Aproximační Thin Plate Spline

Zpracovávané údaje bývají často zkresleny šumem. Na získané hodnoty u_i (analogicky v_i) se pak díváme jako na součet hladké funkce $f(x, y)$ a šumu:

$$u_i = f(x_i, y_i) + \epsilon_i \quad \forall i = 1, 2, \dots, N \quad (2.6)$$

Přičemž $\epsilon_i \sim N(0, \sigma^2)$ jsou nezávislé náhodné veličiny s normálním rozdělením, které má nulovou střední hodnotu a neznámý rozptyl (ale stejný pro všechna ϵ_i). Neznámá funkce $f(x, y)$ je právě hledaná funkce $T_U(x, y)$ (viz. výraz (2.2), analogicky to platí pro $T_V(x, y)$).

Abyste vliv šumu omezil, uvolňuje se interpolační podmínka (2.3) a přidává se regularizační člen. Hledá se potom funkce minimalizující výraz:

$$\begin{aligned} \tilde{J}_\lambda(f) = & \sum_{i=1}^N [f(x_i, y_i) - u_i]^2 \\ & + \lambda \iint \left(\frac{\partial^2 f}{\partial x \partial x} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y \partial y} \right)^2 dx dy \quad (2.7) \end{aligned}$$

Řešení se nazývá *aproximační thin plate spline* nebo *smoothing spline*. První část výrazu vyjadřuje velikost chyby funkce $f(x, y)$ na zadaných bodech, měřeno kvadrátem euklidovské vzdálenosti. Druhá část je převzaná z interpolační TPS a měří hladkost hledané funkce.

Nový parametr λ určuje, zda klademe důraz na přesnost nebo na hladkost řešení. Pokud je parametr λ příliš velký, řešení je zkresleno šumem. Naopak je-li λ příliš malé, ztratí se část informace a řešení je hrubé. Existují dva hraniční případy: první extrém je $\lambda \rightarrow 0$, kdy se jedná o interpolaci, zatímco druhý extrém $\lambda \rightarrow \infty$ znamená maximální vyhlazení. Tím, jak určit hodnotu λ , aby řešení nejlépe vystihovalo skutečnou funkci, se zabývá kapitola 3 „Volba vyhlazovacího parametru“.

Kromě rovnoměrně rozložených chyb, jak je obvykle předpokládáno (viz Rohr [18]) a vysvětleno zde, je možné uvažovat také nerovnoměrné rozložení

chyb, kterému se věnují Rohr a spol. [17]. Matematickým pohledem na aproximační TPS se potom zabývá Wahba [22]. Obecnou podobu výrazu $\tilde{J}_\lambda(f)$ pro d dimenzí, případně pro spojitost vyšších derivací, lze nalézt v člancích autorů Berman [3] nebo Wood [26]. Omezíme-li se pouze na jednu dimenzi, je řešením dobře známý kubický spline, řešení ve dvou dimenzích ukazuje následující sekce.

Aproximační TPS lze opět využít pro řešení různých problémů, příklady jsou odstranění šumu na snímcích zemského povrchu, které popisuje Berman [3], a registrace medicínských snímků, viz Rohr v knize [19]. Porovnáním TPS a mapování pomocí polynomů a multikvadrík se zabývá Wiemker [23] a Fogel [9].

2.3 Řešení mapování pomocí TPS

Obvykle se řešení problému interpolace i aproximace hledá v Sobolevově prostoru W_2^2 (prostor funkcí, které mají první derivaci absolutně spojitou a druhé parciální derivace jsou v kvadrátu lebesgueovsky integrovatelné). Řešení má tedy spojitě první derivace, jeho přesný výpočet popisuje zbytek této sekce.

Označme \mathbf{x} vektor (x, y) , \mathbf{x}_i vektor (x_i, y_i) . Podle Arad a spol. [2] lze použitím variančního počtu dokázat, že řešení minimalizace $\Theta_2(f)$ ((2.5)), případně $\tilde{J}_\lambda(f)$ ((2.7)) má tvar

$$s(\mathbf{x}) = \sum_{i=1}^N a_i g_i(\|\mathbf{x} - \mathbf{x}_i\|), \quad (2.8)$$

kde $\|\cdot\|$ je euklidovská norma na \mathbb{R}^2 a $g_i(t)$ jsou bazické funkce definované takto:

$$g_i(t) = t^2 \log t, \quad g_i(0) = 0$$

Jak je vidět, těchto bazických funkcí je stejně jako zadaných bodů a jejich středy ($t = 0$) jsou umístěny na stejných souřadnicích, jaké mají zadané body \mathbf{x}_i . Výše uvedené bazické funkce se někdy díky své rotační symetrii nazývají *radiální bazové funkce*. O použití jiných bazických funkcí než $t^2 \log t$ pojednává sekce 2.4.

2.3.1 interpolační TPS

Podívejme se nejprve na řešení interpolace pomocí TPS, uvedené v Arad a spol. [2] nebo Carr a spol. [5]. V rovnici (2.8) máme N neznámých koeficientů a_i . Potřebných N podmínek pro nalezení řešení poskytuje přímo zadání úlohy interpolace, viz rovnice (2.3). Výsledná soustava lineárních rovnic má potom tvar

$$\mathbf{G}\mathbf{a} = \mathbf{u}, \quad (2.9)$$

kde $\mathbf{G} = g_{ij} = g_i(\|\mathbf{x}_i - \mathbf{x}_j\|)$, $\mathbf{a} = (a_1, a_2, \dots, a_N)^T$ je vektor neznámých proměnných a $\mathbf{u} = (u_1, u_2, \dots, u_N)^T$ je vektor známých hodnot.

Nevýhoda tohoto mapování spočívá v tom, že nedokáže reprodukovat polynomy. Proto tento model nedokáže zachytit jednoduché afinní mapování (viz [2]) a dochází k špatné aproximaci v místech, která jsou daleko od zadaných bodů \mathbf{x}_i . To je důvod, proč se obvykle do formulace funkce $s(\mathbf{x})$ přidává polynom p_1 , který afinní mapování reprodukovat umožní. Hledáme tedy funkci

$$s(\mathbf{x}) = p_1(\mathbf{x}) + \sum_{i=1}^N a_i g_i(\|\mathbf{x} - \mathbf{x}_i\|), \quad p_1(\mathbf{x}) \in \Pi_1, \quad (2.10)$$

přičemž Π_1 označuje prostor všech algebraických polynomů na \mathbb{R}^2 stupně nejvýše 1. Hodnota minimalizovaného výrazu $\Theta_2(f)$ se přidáním polynomu p_1 nezmění, protože Π_1 je obsažen v nulovém prostoru funkcionálu $\Theta_2(f)$. Poněvadž p_1

$$p_1(\mathbf{x}) = p_1(x, y) = \alpha_1 + \alpha_2 x + \alpha_3 y$$

přidává tři nové proměnné, potřebujeme další tři rovnice, abychom mohli vypočítat vektor neznámých $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^T$. Uvádí se následující okrajové podmínky (viz. Arad a spol. [2]):

$$\mathbf{C}\boldsymbol{\alpha} = \mathbf{0}, \quad \mathbf{C} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_N \\ y_1 & y_2 & \dots & y_N \end{bmatrix} \quad (2.11)$$

Okrajové podmínky zabezpečují konzistenci v nekonečnu. Zajišťují, aby v nekonečnu neměli RBF žádný vliv a řešení konvergovalo k rovině, která je určena metodou nejmenších čtverců. Obecněji okrajové podmínky rozebírá Wood a Augustin v článku [25].

Vyřešit úlohu interpolace pomocí TPS potom znamená nalézt řešení následující soustavy lineárních rovnic:

$$\begin{bmatrix} \mathbf{G} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{u} \\ \mathbf{0} \end{bmatrix} \quad (2.12)$$

Podle článku [5] má soustava jednoznačné řešení, právě když jsou zadány alespoň tři body a zároveň všechny body $[x, y]$ neleží v jedné přímce. Pokud jsou zadány právě tři body splňující tuto podmínku, řešením je rovina určená body $[x, y, u_i]$ (analogicky to platí pro body $[x, y, v_i]$), celá mapovací funkce $T(x, y)$ (viz (2.14)) potom odpovídá afinnímu mapování.

2.3.2 aproximační TPS

Jak najít řešení aproximační TPS ukazuje opět Arad a spol. [2]. Potřebné vektory neznámých koeficientů $\mathbf{a}, \boldsymbol{\alpha}$ (z rovnice (2.10)) získáme řešením následující soustavy lineárních rovnic, která je zobecněním soustavy

interpolčních rovnic s okrajovými podmínkami.

$$\begin{bmatrix} (\mathbf{G} + \lambda \mathbf{I}) & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{u} \\ \mathbf{0} \end{bmatrix} \quad (2.13)$$

Kde \mathbf{I} označuje jednotkovou matici $N \times N$ a λ je vyhlazovací parametr zavedený ve výrazu $\tilde{J}_\lambda(f)$ ((2.7)).

2.3.3 shrnutí

Shrneme-li dosavadní poznatky dostaneme, že výsledná mapovací funkce je vyjádřena formulí

$$T(x, y) = \left(\begin{aligned} &\alpha_1 + \alpha_2 x + \alpha_3 y + \sum_{i=1}^N a_i g_i(\|x - x_i, y - y_i\|), \\ &\beta_1 + \beta_2 x + \beta_3 y + \sum_{i=1}^N b_i g_i(\|x - x_i, y - y_i\|) \end{aligned} \right) \quad (2.14)$$

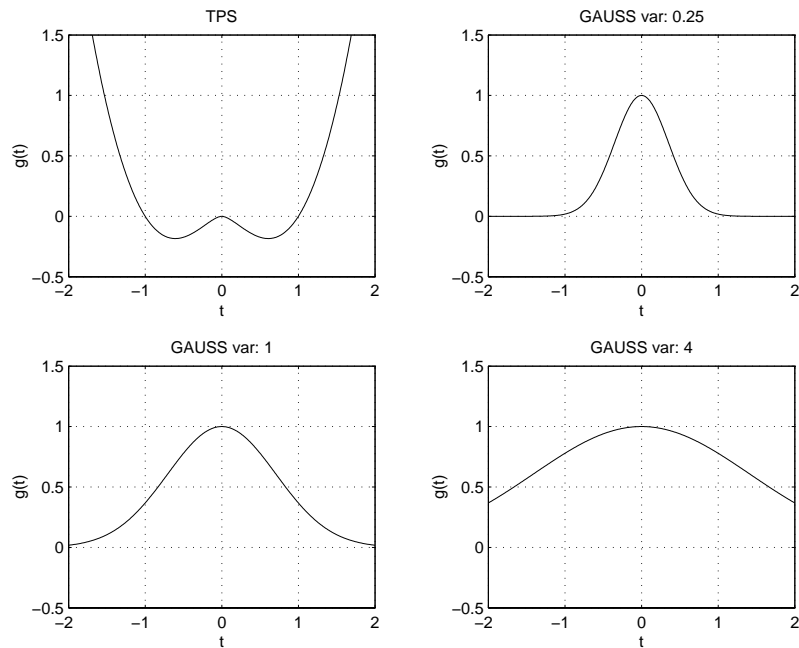
s $g_i(t) = t^2 \log t$. Neznámé proměnné $\alpha_1, \alpha_2, \alpha_3$ a a_1, a_2, \dots, a_N získáme řešením soustavy (2.12) pro interpolaci, případně soustavy (2.13) pro aproximaci. Neznámé $\beta_1, \beta_2, \beta_3$ a b_1, b_2, \dots, b_N vypočítáme řešením analogických soustav rovnic. Pouze nahradíme α_i za β_i , a_i za b_i a u_i za v_i .

2.4 Jiné radiální bázevé funkce

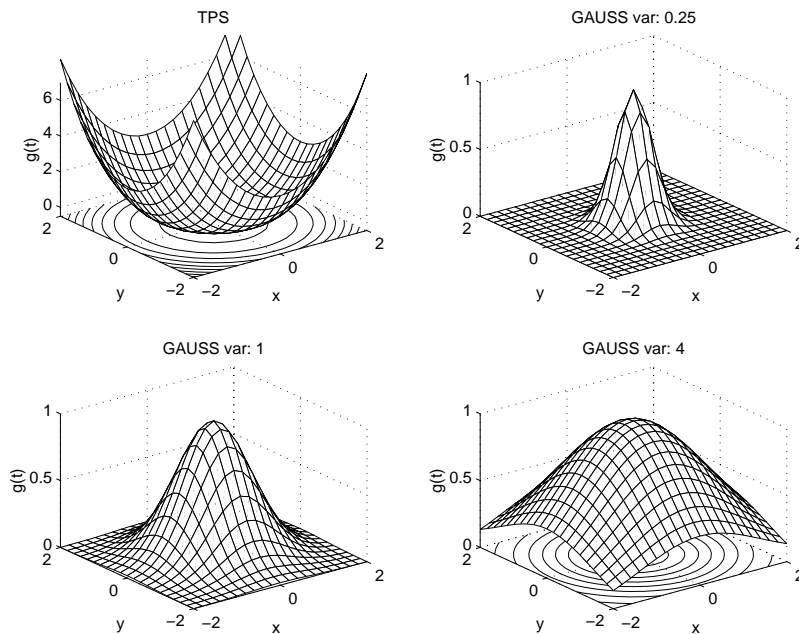
Ačkoliv se TPS používají v mnoha aplikacích, někdy by mohlo být výhodné použít jiné radiální bázevé funkce $g_i(t)$, které dosadíme do formule (2.10), případně (2.14). Velmi známé jiné možnosti RBF, které zaručují jednoznačné řešení při zadání N různých bodů a dosazení do rovnic bez okrajových podmínek, uvádí například Arad a spol. [2]:

1. $g_i(t) = (t^2 + c^2)^\gamma$, $0 < \gamma < 1$ (multikvadriky)
2. $g_i(t) = \log(t^2 + c^2)^{\frac{1}{2}}$, $c^2 \geq 1$ (posunutý logaritmus)
3. $g_i(t) = \exp\left(\frac{-t^2}{\sigma^2}\right)$, $\sigma > 0$ (Gaussova funkce)

Použití multikvadrík při zpracování leteckých snímků a jejich porovnání s metodami využívajícími polynomů třetího stupně popisují Ehlersnd a Fogel [6]. Porovnání multikvadrík a Akimovy metody uvádí Wiemker v [23].



Obrázek 2.1: Příklady radiálních bázových funkcí v 1D – TPS a Gassovy RBF s rozptly: 0.25, 1 a 4.



Obrázek 2.2: Příklady radiálních bázových funkcí ve 2D – TPS a Gassovy RBF s rozptly: 0.25, 1 a 4.

2.4.1 Gaussovy RBF

Nevýhoda použití TPS může spočívat v globální povaze funkcionálu Θ_2 , protože malá odchylka jednoho bodu ovlivní celou funkci $s(x)$. Někdy je výhodné, aby zadané body měly pouze lokální platnost. Potom je užitečné použít RBF s parametrem, který řídí míru lokality příslušných bazových funkcí. Například Arad a spol. [2] používají Gaussovu funkci, ve které úlohu lokálního parametru hraje parametr σ^2 . Se vzrůstajícím σ^2 se rozptyl Gaussovy funkce zvětšuje a tím příslušný bod ovlivňuje větší oblast mapovací funkce. Určení parametru σ závisí na povaze řešeného problému, teoreticky může mít každá funkce $g_i(t)$ jiný parametr σ . Tvar bazových funkcí pro TPS a pro Gaussovy funkce s různým rozptylem ukazují obrázky 2.1 a 2.2.

Pokud si jako radiální bazové funkce z výrazu (2.10) zvolíme Gaussovy funkce, neminimalizuje již tento výraz funkcionál Θ_2 . Pro ilustraci, jak je uvedeno v [2], RBF s Gaussovými funkcemi, bez polynomiálního členu a s parametrem $\sigma = 1$ minimalizují funkcionál (f, f) , kde (f, h) je

$$(f, h) = \iint \hat{f}(x) \overline{\hat{h}(x)} \exp(x^2) dx$$

a \hat{f} označuje Fourierovu transformaci f .

Úspěšné použití RBF s Gaussovými funkcemi ukazují Arad a spol. v článku [2]. Využívají zde toho modelu mapovací funkce pro warping obrázků, konkrétně pro změnu výrazu tváře. Nejprve na obrázku zadají řídicí body, potom určí novou polohu těchto řídicích bodů a parametr λ , σ^2 spočítají podle rozměrů obrázku. Tím je mapování souřadnic, kterým obrázek deformují, určeno. Ukazují, že použitím tohoto mapování se produkují velmi přirozené změny tváře a to díky lokálnímu charakteru Gaussových funkcí. Narozdíl od TPS, které generují na pohled velmi nepřirozené deformace, kdy změna řídicích bodů v jedné části znatelně ovlivní i vzdálené části obrázku. Obrázky, které dokládají tyto vlastnosti, jsou umístěny v úvodu – v části popisující motivaci.

Je známé, že při použití TPS je potřeba zadat poměrně hodně řídicích bodů, právě proto aby nedocházelo k výše popsané deformaci vzdálených oblastí. Některé body potom jen definují, že „tady už k deformaci nedochází“. Zadávání těchto bodů je nejen pracné, ale jejich počet zbytečně zvětšuje výpočetní nároky. Tak vznikla motivace této práce, jak bylo již vysvětleno v úvodu.

2.5 Jiné modely

Jinou možnost jak snížit výpočetní nároky, nabízejí *low rank thin plate spline*, které představil S. N. Wood v článku [26]. Tento model je velice podobný TPS, báze vychází z báze TPS (výraz (2.10)), zmenší se však počet prvků v bázi. Z původních N se jejich počet zmenší na k , přičemž k je

zvolený parametr, který je však větší než dimenze Π_1 (výraz (2.10)). Tím se dosahuje snížení výpočetní náročnosti, neboť místo původních N neznámých koeficientů jich stačí určit k .

Středů RBF se potom už neumísťují na stejné souřadnice, jaké mají zadané body (\mathbf{x}_i) (viz. sekce (2.3)). Nejsou umístěné ani v pravidelné mřížce, ale optimální umístění se určuje na základě souřadnic zadaných bodů (x_i) .

Jak uvádí Wood [26], výraz minimalizovaný TPS ((2.7)) lze vyjádřit také jako:

$$\|\mathbf{u} - \mathbf{G}\mathbf{a} - \mathbf{C}^T\boldsymbol{\alpha}\|^2 + \lambda\mathbf{a}^T\mathbf{G}\mathbf{a}, \quad \text{při podmínce } \mathbf{C}\mathbf{a} = 0 \quad (2.15)$$

Nechť $\mathbf{\Gamma}^k$ je matice, jejíž sloupce tvoří ortonormální bázi s dimenzí k pro prostor parametrů \mathbf{a} . Vektor \mathbf{a} , skládající se z N neznámých koeficientů, lze přibližně spočítat pomocí vektoru \mathbf{a}^k , který obsahuje pouze k koeficientů:

$$\mathbf{a} \approx \mathbf{\Gamma}^k \mathbf{a}^k,$$

Jak ale zvolit vhodnou bázi $\mathbf{\Gamma}^k$, aby se po dosazení $\mathbf{a} \approx \mathbf{\Gamma}^k \mathbf{a}^k$ do výrazu (2.15), změnil tento výraz minimálně? V oslabeném smyslu „optimální“ bázi tvoří prvních k vlastních vektorů matice \mathbf{G} , viz [26]. Přesný popis algoritmu, včetně efektivního získání vlastních vektorů je uveden také v článku [26].

Ohledně vyhlazovacích parametrů navrhuje Wood volit $\lambda = 0$, protože pro tento model je podstatná volba parametru k . Ten, podobně jako λ , určuje složitost modelu, ale pro tento model je důležitější. K jeho volbě lze použít některou z metod uvedených v kapitole 3.

Složitost se využitím představeného modelu a efektivních algoritmů snižuje z původních $O(N^3)$ na $O(kN^2)$. Tento model je také více numericky stabilní. Wood na příkladech ukazuje, že při redukované bázi je výše uvedené rozmístění lepší než rozmístění v pravidelné mřížce; někdy překoná dokonce neredukované TPS. Jako aplikaci potom ukazuje použití modelu při modelování produkce jiker.

Další možné modely jako *tensor product splines* nebo *partial splines* popisuje Gu v článku [11].

Kapitola 3

Volba vyhlazovacího parametru

V úvodu této kapitoly je nejdříve velmi krátce shrnuta konstrukce mapovací funkce pomocí RBF. Následuje zadání pro úlohu volby vyhlazovacího parametru (*smoothing parameter selection*) spolu s možnými způsoby jejího řešení. Tato kapitola navazuje na sekci 2.2 „Aproximační Thin Plate Spline“ a 2.3 „Řešení mapování pomocí TPS“. Zároveň však obsahuje obecné principy, které se používají pro volbu vyhlazovacího parametru; nejsou tedy omezeny na modely založené TPS, RBF nebo úlohu konstrukce mapovací funkce.

Vstupem pro konstrukci mapovací funkce je N korespondujících dvojic bodů (x_i, y_i) a (u_i, v_i) , které určují zobrazení. Toto zobrazení hledáme po částech pomocí dvou funkcí $T_U, T_V : \mathbb{R}^2 \rightarrow \mathbb{R}$, které dohromady dávají celkové zobrazení $T(x_i, y_i) = (T_U(x_i, y_i), T_V(x_i, y_i))$. Dále se práce zabývá funkcí T_U , pro T_V je situace analogická. Uvažujme, jak je obvyklé, zkreslení dat aditivním šumem s normálním rozdělení $\epsilon_i \sim N(0, \sigma^2)$:

$$u_i = T_U(\mathbf{x}_i) + \epsilon_i \quad \forall i = 1, 2, \dots, N, \quad (3.1)$$

Aproximační TPS řeší úlohu hledáním funkce, která minimalizuje výraz $\tilde{J}_\lambda(f)$ ((2.7)). Vyhlazovací parametr λ přitom určuje, zda klademe důraz na přesnost aproximace či na hladkost.

Problém je, jak určit vyhlazovací parametr, aby se funkce minimalizující výraz $\tilde{J}_\lambda(f)$ nejvíce blížila skutečné deformaci snímku. Pokud bude parametr λ příliš velký, bude sice výsledná funkce dobře aproximovat zadaná data, ale bude do tohoto modelu zahrnut i šum v datech. Naopak, pokud bude parametr λ příliš malý, bude generalizace příliš velká. Ztratíme část informace z dat a model bude hrubý, bude postrádat detaily.

Existuje několik možností volby vyhlazovacího parametru, hlavní jsou následující:

- První možností je manuální volba, měníme parametr a sledujeme, kdy je výsledek z hlediska nějakého subjektivního kritéria nejlepší.
- Druhou možností je spočítat optimální hodnotu parametru ze znalosti distribuce nebo alespoň rozptylu šumu. Taková informace sice většinou není známa, obvykle se však předpokládá normální rozdělení a rozptyl je možné odhadnout na základě metody, která byla použita při získávání vstupních dat, pokud taková studie pro metodu existuje.
- Poslední je odhad optimální hodnoty pouze na základě vstupních dat, v našem případě jsou to souřadnice korespondujících bodů.

V této práci se zabývám poslední z výše uvedených možností, neboť navržená metoda by měla být automatická a analýzy chyb pro metody používané při získávání dat nejsou často k dispozici. Úlohou tedy je určit nejvhodnější hodnotu vyhlazovacího parametru λ , známe-li pouze souřadnice korespondujících bodů. Pro funkci T_U to znamená určit vhodnou hodnotu λ ze znalosti souřadnic (x_i, y_i) a funkčních hodnot u_i . Kritéria, která se k tomu používají, popisují následující sekce.

Problém volby vyhlazovacího parametru není vázaný pouze na TPS, vyskytuje se v nejrůznějších modelech, obzvláště v těch, které se používají ve statistice, kde se tyto oblasti nazývají *spline regression*, případně *non-linear regression*. Podrobněji volbu vyhlazovacího parametru rozebírají ve svých článcích Eubank [7], van der Linde [15] a Gu [11].

3.1 Mean Square Error

Mean square error (MSE), česky nazývaná *střední kvadratická odchylka*, je nejlepší kritérium, v praxi ale bohužel nepoužitelné, neboť vyžaduje znalost neznámé funkce $T_U(\mathbf{x})$. Své použití ovšem nalézá při testování kritérií, jako referenční kritérium.

Uvažujme, že známe skutečnou funkci $T_U(\mathbf{x})$ z výrazu (3.1) a máme v bodech \mathbf{x}_i naměřená data u_i , která jsou zkreslena šumem. Dále nechť $s_\lambda(\mathbf{x})$ je odhad funkce $T_U(\mathbf{x})$ pomocí aproximační TPS, neboli je to funkce minimalizující výraz $\tilde{J}_\lambda(f)$. Definujeme potom

$$MSE(\lambda) = \sum_{i=1}^N \left(s_\lambda(\mathbf{x}_i) - T_U(\mathbf{x}_i) \right)^2 \quad (3.2)$$

Ideální λ je následně definováno jako λ , při kterém je výraz $MSE(\lambda)$ minimální, značme ho λ^* .

Jak jsem již řekl, MSE je prakticky nepoužitelné kritérium; důvodem je potřeba znát funkci $T_U(\mathbf{x})$, ale tu se právě snažíme nalézt. Toto kritérium lze však úspěšně použít pro testování, jak dobrý odhad funkce $T_U(\mathbf{x})$ daná

metoda dává. Na MSE se potom nedíváme jako na funkci lambda, ale jako na kritérium, které pro daný odhad určuje velikost chyby tohoto odhadu.

Postup porovnání úspěšnosti metod je následující. Jednoduše si zvolíme $T_U(\mathbf{x})$, přidáme vygenerovaný šum a nějakým způsobem najdeme optimální λ_{M1}^* a λ_{M2}^* postupně příslušející metodám $M1$ a $M2$. Spočítáme odhady $s_{\lambda_{M1}^*}(\mathbf{x}_i)$ a $s_{\lambda_{M2}^*}(\mathbf{x}_i)$, dosadíme postupně tyto odhady do MSE a získáme velikost chyby pro každý z odhadů.

3.2 Cross Validation

Cross validation (CV), někdy také psaná *krosvalidace*, je všestranně použitelná a ve statistice velmi používána metoda. Kritérium krosvalidace je založené na měření průměrné chyby predikce.

Označme $s_{\lambda}^{[k]}(\mathbf{x})$ funkci zkonstruovanou pro pevné lambda a se znalostí všech bodů kromě k -tého. Hodnotu kritéria pro pevné lambda spočítáme následovně. Vynecháme k -tý vstupní bod a spočítáme model funkce bez tohoto bodu $s_{\lambda}^{[k]}(\mathbf{x})$. Pomocí této funkce vyčíslíme odhad v bodě \mathbf{x}_k a spočítáme kvadrát chyby oproti naměřené hodnotě u_k . Tento postup opakuje pro všechny body a zprůměrujeme chyby predikce:

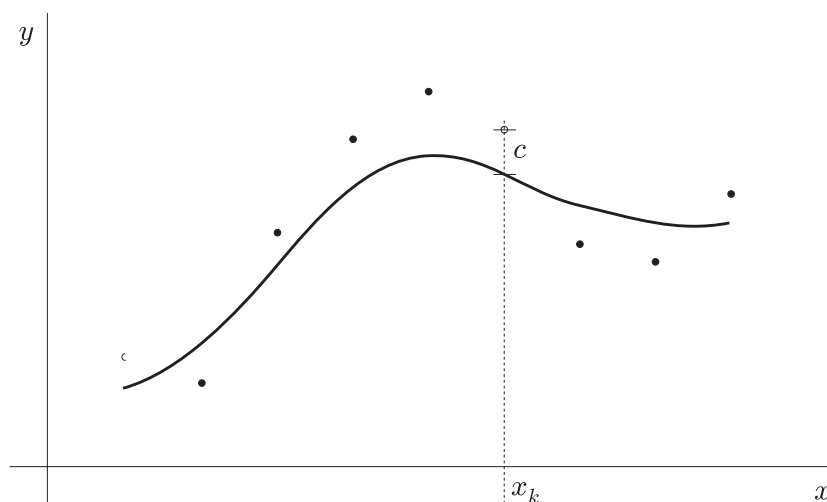
$$CV(\lambda) = \frac{1}{N} \sum_{k=1}^N \left(s_{\lambda}^{[k]}(\mathbf{x}_k) - u_k \right)^2 \quad (3.3)$$

V případě TPS se $s_{\lambda}^{[k]}(\mathbf{x})$ vypočítá podle vztahu (2.13), samozřejmě bez k -tého řádku a sloupce. Predikovaná hodnota $s_{\lambda}^{[k]}(\mathbf{x}_k)$ se potom spočítá ze vztahu (2.10), opět bez k -tého členu.

Optimální λ vzhledem ke kritériu krosvalidace je opět λ , při kterém je výraz $CV(\lambda)$ minimální. Takové lambda odpovídá reprezentaci modelu, který dává v průměru nejlepší predikci (měřeno pomocí MSE). Intuitivně, pokud model funkce dobře predikuje, znamená to, že dobře vystihuje neznámou funkci. Tedy není ani příliš hrubý, ani nezahrnuje zbytečně šum do modelu, nebo-li λ je zvoleno správně.

Obecně se při krosvalidaci nemusí vynechávat pouze jeden prvek. Lze vynechat třeba $\frac{1}{10}$ prvků, spočítat model bez nich a potom vypočítat chybu predikce na těchto prvcích. Aby se nemusely zkoušet všechny kombinace, je možné volit vynechané prvky náhodně.

Největší nevýhodou krosvalidace je výpočetní náročnost, obzvláště pak ve více dimenzích, což je případ obrazových dat. Pro každý bod je totiž potřeba spočítat model, který tento bod neobsahuje, a vyčíslit predikci v tomto bodě. Výpočetní náročnost odstraňuje metoda Generalized Cross Validation popsaná dále, nejprve je však potřeba model TPS trochu přeformulovat, o čemž pojednává následující sekce.



Obrázek 3.1: Krosvalidace v jedné dimenzi – vynecháme bod x_k , spočítáme model bez tohoto bodu a vyčíslíme chybu predikce c , nakonec chyby přes všechny body zprůměrujeme.

Další, někdy uváděnou nevýhodou, je nedostatečné vyhlazení a tedy zbytečně zahrnutí šumu do modelu, v angličtině popisované jako *under-smoothing*. Krosvalidace totiž v některých případech, jak uvádí Robinson a Moyeed [20], volí λ menší než by bylo optimální, někdy je zvolené λ tak drasticky malé, že se jedná téměř o interpolaci. Celkově se volba malého λ může projevit spíše při menším počtu zadaných dat.

Odstraněním nedostatečného vyhlazení a zvětšením stability krosvalidace, např. proti některým špatně změřeným hodnotám, se zabývá sekce 3.5 Robustified Cross Validation, obsahující kritérium se stejným názvem.

3.3 Formulace TPS pomocí hat matrix

Tato sekce nepopisuje žádné kritérium, ukazuje však, jak upravit výpočet TPS v určitých bodech do tvaru, který vyžadují některá dále uvedená kritéria.

Označme $\mathbf{u} = (u_1, u_2, \dots, u_N)^T$ vektor složený ze známých hodnot, které přísluší bodům \mathbf{x}_i , a dále necht' \mathbf{s}_λ je vektor hodnot TPS v těchto bodech. Cílem je vyjádřit každou z hodnot \mathbf{s}_λ jako lineární kombinaci známých hodnot, to znamená přeformulovat výpočet \mathbf{s}_λ do tvaru:

$$\mathbf{s}_\lambda = \mathbf{A}(\lambda)\mathbf{u} \quad (3.4)$$

Čtvercová matice $\mathbf{A}(\lambda)$ s rozměry $N \times N$ je závislá na hodnotě parametru λ a v anglicky psané literatuře nazývá *hat matrix* nebo *influence matrix*.

Prvně je ukázáno přeformulování pro jednoduchý model TPS bez lineárního členu, později pak pro model TPS s lineárním členem, který se opravdu používá.

Jednoduchý model

Uvažujme vyjádření TPS funkce $s_\lambda(\mathbf{x})$, které neobsahuje lineární člen (viz sekce 2.3). Počítáme hodnoty funkce $s_\lambda(\mathbf{x})$ pouze v bodech \mathbf{x}_j (body, ve kterých jsou známé hodnoty u_i):

$$s_\lambda(\mathbf{x}_j) = \sum_{i=1}^N a_i g_i(\|\mathbf{x}_j - \mathbf{x}_i\|) \quad \forall i = 1, 2, \dots, N \quad (3.5)$$

Vektor koeficientů \mathbf{a} se při daném λ vypočítá jako $\mathbf{a} = (\mathbf{G} + \lambda \mathbf{I})^{-1} \mathbf{u}$, kde \mathbf{G} je matice definovaná v sekci 2.3.1. Dosazením vektoru \mathbf{a} do vztahu (3.5) získáme požadovaný tvar (3.4), kde

$$\mathbf{A}(\lambda) = \mathbf{G} (\mathbf{G} + \lambda \mathbf{I})^{-1} \quad (3.6)$$

Model s okrajovými podmínkami

Nyní přejdeme k vyjádření TPS funkce $s_\lambda(\mathbf{x})$ se zahrnutým lineárním členem (jak je uvedené v sekci 2.3). Opět počítáme pouze hodnoty funkce $s_\lambda(\mathbf{x})$ v bodech \mathbf{x}_j :

$$s_\lambda(\mathbf{x}_j) = p_1(\mathbf{x}_j) + \sum_{i=1}^N a_i g_i(\|\mathbf{x}_j - \mathbf{x}_i\|) \quad \forall i = 1, 2, \dots, N, \quad (3.7)$$

kde $p_1(\mathbf{x}) = p_1(x, y) = \alpha_1 + \alpha_2 x + \alpha_3 y$.

Protože přibyly tři nové neznámé, přidávají se tři hraniční podmínky, viz soustava (2.11). Vektory koeficientů \mathbf{a} , $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^T$ se potom získají řešením soustavy (2.13), jak uvedeno v sekci 2.3 popisující řešení mapování pomocí TPS. Dosazením do (3.7) získáme:

$$[\mathbf{s}_\lambda] = [\mathbf{G} \quad \mathbf{C}^T] \begin{bmatrix} (\mathbf{G} + \lambda \mathbf{I}) & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{u} \\ \mathbf{0} \end{bmatrix}$$

Přeformulování do požadovaného tvaru se provede triviálním rozložením poslední matice pomocí \mathbf{I}_N :

$$\mathbf{A}(\lambda) = [\mathbf{G} \quad \mathbf{C}^T] \begin{bmatrix} (\mathbf{G} + \lambda \mathbf{I}) & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{I}_N \\ \mathbf{0}_{3,N} \end{bmatrix} \quad (3.8)$$

Matice \mathbf{I}_N označuje jednotkovou matici $N \times N$ a $\mathbf{0}_{3,N}$ je matice $3 \times N$ obsahující pouze nuly.

Jak odstranit okrajové podmínky v obecném případě (nejen pro TPS nebo RBF) a dosáhnout požadovaného tvaru (3.4), ukazuje Wood [25]. Předpokladem jsou hraniční podmínky ve tvaru $\mathbf{C}\boldsymbol{\alpha} = \mathbf{0}$.

3.4 Generalized Cross Validation

Kritérium *Generalized Cross Validation* (GCV) je zřejmě nejpoužívanější kritérium, platí to nejen pro registraci a jiné oblasti zpracování obrazu. Princip GCV vychází z krosvalidace, odstraňuje však její výpočetní náročnost. K tomuto urychlení se používá průměrování, a proto nejsou kritéria ekvivalentní.

Metoda GCV byla původně odvozena ve statistice pro regresní analýzu dat pomocí spline funkcí, jak však uvádí Robinson a Moyeed [20], GCV je možné použít obecně pro lineární modely. Lineární modely jsou modely, kde se nové hodnoty v zadaných bodech počítají jako lineární zobrazení hodnot zadaných. To znamená výpočet nových hodnot lze přeformulovat do tvaru (3.4), jak to lze provést pro TPS a jiné RBF ukazovala předchozí sekce.

Odvození GCV z krosvalidace shrnují Robinson a Moyeed [20] nebo Fang a Nure [8]. Hlavní myšlenkou je vyjádření $s_\lambda^{[k]}(\mathbf{x}_k)$ pouze pomocí $s_\lambda(\mathbf{x}_k)$. Model potom stačí spočítat pouze jednou a ne n -krát. Klíčovým vztahem je „deletion theorem“:

$$s_\lambda^{[k]}(\mathbf{x}_k) - u_k = \frac{s_\lambda(\mathbf{x}_k) - u_k}{1 - a_{kk}}, \quad (3.9)$$

kde a_{ii} je i -tý diagonální prvek matice $\mathbf{A}(\lambda)$. Dosazením do vztahu pro krosvalidaci (3.3) získáme:

$$CV_2(\lambda) = \frac{1}{N} \sum_{k=1}^N \frac{(s_\lambda(\mathbf{x}_k) - u_k)^2}{(1 - a_{kk})^2} \quad (3.10)$$

Jak uvádí Robinson a Moyeed [20], hodnoty a_{ii} se nahrazují jejich průměrem $\mu_1(\lambda) = N^{-1} \text{tr} \mathbf{A}(\lambda)$. Dosazením získáváme kritérium nazývané Generalized Cross Validation:

$$GCV(\lambda) = \frac{N^{-1} \|(\mathbf{I} - \mathbf{A}(\lambda)) \mathbf{u}\|^2}{(N^{-1} \text{tr}(\mathbf{I} - \mathbf{A}(\lambda)))^2} \quad (3.11)$$

Čitatel výrazu vyjadřuje velikost chyby v zadaných bodech (měřeno kvadrátem euklidovské vzdálenosti), zatímco jmenovatel penalizuje složitost modelu. Výraz $\text{tr} \mathbf{A}(\lambda) = \sum_{i=1}^N a_{ii}$ je stopu matice \mathbf{A} , označuje *efektivní stupně volnosti* a odhaduje složitost modelu. V angličtině se tento pojem označuje *effective degrees of freedom* nebo *equivalent degrees of freedom* (EDF), více podrobností uvádí Linde [15].

Optimální hodnota λ odpovídá opět minimální hodnotě GCV kritéria, narozdíl od krosvalidace však stačí spočítat model pouze jednou. Jak uvádí Linde [15] nebo Robinson a Moyeed [20], GCV kritérium je asymptoticky ekvivalentní krosvalidaci, tedy asymptoticky minimalizuje střední kvadratickou odchylku predikce. Více teoretických vlastností ukazuje opět Robinson a Moyeed [20] nebo Wood [25].

Podobně jako u krosvalidace dochází i u GCV někdy k nedostatečnému vyhlazení (undesmoothering), zvláště při menším počtu zadaných dat. Řešením tohoto problému se zabývá, jak jsem už zmínil, následující sekce a v ní uvedená metoda Robustified Cross Validation.

Minimalizaci GCV kritéria je možné dále urychlit pomocí aproximací. Značné urychlení výpočtu GCV kritéria pro TPS, založené na Fourierově reprezentaci aproximační TPS, popisuje Berman [3]. Aproximaci pro kubický spline nazvanou *asymptotic cross validation* ukazují Robinson a Moyeed [20].

Možnosti použití GCV kritéria jsou obrovské, dokazují to aplikace z nej-různějších oborů. Příklady jsou odstranění šumu na snímcích zemského povrchu pomocí aproximační TPS, které popisuje Berman [3], a vyhlazování snímku hlavy v medicíně pomocí diskretních Gaussových filtrů od autorů Fang a Nure [8]. Dále odstranění šumu použitím waveletové transformace, kterou představili Jansen a spol. [14], GCV se zde používá pro určení prahu v této metodě (anglický název je *wavelet-thresholding*). Slepá rekonstrukce obrazu a zvětšení rozlišení od skupiny autorů Ngueyen a spol. [16], zde je také uvedena aproximace pro výpočet GCV kritéria.

3.5 Robustified Cross Validation

Kritérium *Robustified Cross Validation* (RCV) vzniklo, aby odstranilo problém nedostatečného vyhlazení (undersmoothering), který se někdy objevuje u kritérií krosvalidace a GCV, a aby se zvětšila robustnost těchto kritérií. Jedná se opět o kritérium založené na krosvalidaci a stejně jako GCV je RCV možné použít obecně pro lineární modely. Kritérium RCV navrhli Robinson a Moyeed, úplné odvození spolu s jinými podrobnostmi, je možné nalézt v jejich článku [20].

Odvození RCV navazuje na odvození GCV kritéria, výraz (3.10), který je ekvivalentním vyjádřením kritéria krosvalidace, lze upravit dále na:

$$\frac{1}{N} \sum_{k=1}^N \left(\sum_{j=1}^N s_{\lambda}(\mathbf{x}_j) - s_{\lambda}^{[k]}(\mathbf{x}_j) \right)^2 \quad (3.12)$$

Kritérium GCV je potom pouze váženou variantou tohoto výrazu. Jak je zde vidět, krosvalidace minimalizuje sumu chyb, bez ohledu na znaménka. V případě mnoha zadaných dat není pravděpodobné, že by to vadilo, ale pokud je dat málo může tento nedostatek robustnosti způsobit problémy.

Proto Robinson a Moyeed [20] navrhuji pro robustní volbu výraz:

$$\frac{1}{N} \sum_{k=1}^N \sum_{j=1}^N \left(s_{\lambda}(\mathbf{x}_j) - s_{\lambda}^{[k]}(\mathbf{x}_j) \right)^2 \quad (3.13)$$

Toto kritérium klade větší důraz na hladkost a je velmi odolné proti odchylkám dat, v praxi ale příliš silně vyhlazuje. Abychom získali kritérium s

vhodnou mírou robustnosti, navrhují Robinson a Moyeed [20] zkombinovat se stejnou váhou výrazy (3.12) a (3.13). Podobnými algebraickými úpravami, jakými vznikl výraz (3.12), ale aplikovanými opačným směrem, a podobným průměrováním jako odvození GCV vznikne výraz pro RCV kritérium:

$$RCV(\lambda) = N^{-1} \frac{1 + N^{-1} \operatorname{tr} \mathbf{A}(\lambda)^2}{(1 - N^{-1} \operatorname{tr} \mathbf{A}(\lambda))^2} \|(\mathbf{I} - \mathbf{A}(\lambda)) \mathbf{u}\|^2 \quad (3.14)$$

Tímto jsme získali kritérium se stejnou výpočetní náročností jako GCV, ale s větší robustností.

Optimální hodnota λ odpovídá opět minimální hodnotě RCV kritéria. Jak ukazují Robinson a Moyeed [20], RCV kritérium je stejně jako GCV asymptoticky ekvivalentní krosvalidaci.

Narozdíl od krosvalidace a GCV, opět podle [20] má RCV kritérium někdy tendenci volit λ příliš velké a tím zbytečně odstraňovat jemnosti modelu. V anglicky psané literatuře toto přílišné vyhlazování nazývá *over-smoothing*. Robinson a Moyeed uvádí, že je možné do výrazu RCV kritéria zavést parametr, který kontroluje míru robustnosti kritéria, tímto bychom ale získali další neznámý parametr.

Aproximace RCV kritéria pro kubický spline, podobně jako u GCV, ukazují opět Robinson a Moyeed citerobinson.

3.6 Akaike Information Criterion

Akaike Information Criterion (AIC), česky nazývané *Akaikovo informační kritérium*, byl odvozeno jako přibližný nevychýlený odhad Kullback-Leiberovy informace (viz Hurvich [13]). Kritérium odhaduje přirozenou míru nesrovnalosti nebo, jinak řečeno, vzdálenost od „skutečného“ modelu. Předpokladem použití AIC ve tvaru uvedeném zde je opět lineární model.

V původní verzi AIC se používá parametr, který reguluje složitost modelu. V případě TPS tato vlastnost není parametrem λ explicitně vyjádřena, proto se jako odhad složitosti modelu používá, již uvedený, odhad stupňů volnosti $EDF = \operatorname{tr} \mathbf{A}(\lambda)$. Akaikovo informační kritérium (viz Robinson a Moyeed [20]) má potom tvar:

$$AIC(\lambda) = \log \frac{\|(\mathbf{I} - \mathbf{A}(\lambda)) \mathbf{u}\|^2}{N} + \frac{2 \operatorname{tr} \mathbf{A}(\lambda)}{N} \quad (3.15)$$

První část výrazu tvoří logaritmus reziduálního rozptylu, který samozřejmě chceme minimalizovat, druhá část je penalizace za složitost modelu. Podle jiného pohledu odhaduje AIC ztrátu informace způsobenou tím, že místo „skutečných“ parametrů modelu použijeme parametry spočítané metodou maximální věrohodnosti. Druhý člen zajišťuje korekci, aby odhad nebyl vychýlený.

Optimální hodnota λ odpovídá opět minimální hodnotě AIC. Podle autorů Hurvich a spol. [13], je AIC asymptoticky ekvivalentní GCV kritériu.

AIC bylo původně odvozeno pro větší množství dat, proto, podobně jako GCV, někdy nedostatečně vyhlazuje, obzvláště při menších vzorcích dat, více uvádí Hurvich a spol. [13]. Odstraněním tohoto problému se zabývá následující sekce a v ní obsažené zlepšení AIC.

3.7 Corrected Akaike Information Criterion

Corrected Akaike Information Criterion (cAIC), česky nazývané *opravené Akaikeovo kritérium*, je verzí AIC snažící se zlepšit chování AIC pro malý počet vzorků. Jak již bylo řečeno, AIC v těchto případech volí zbytečně složité modely. Předpokladem použití cAIC je opět lineární model.

Celé odvození je poměrně dlouhé, v plném rozsahu je popsáno autory Hurvich a spol. [13]. Cílem bylo opravit AIC, aby bylo méně vychýlené pro malé vzorky dat. Výsledný tvar je aproximací vyvinutých kritérií a už je v praxi dobře použitelný:

$$cAIC(\lambda) = \log \frac{\|(\mathbf{A}(\lambda) - \mathbf{I})\mathbf{u}\|^2}{N} + 1 + \frac{2(\text{tr } \mathbf{A}(\lambda) + 1)}{N - \text{tr } \mathbf{A}(\lambda) - 2} \quad (3.16)$$

První výraz je shodný s AIC, druhý představuje opravené vychýlení kritéria. Autoři poznamenávají, že ačkoliv bylo kritérium odvozeno pro šum s normálním rozdělením (což je splněno, viz výraz (3.1)), pracuje dobře i když tento předpoklad splněn není.

Optimální hodnota λ odpovídá opět minimální hodnotě cAIC. Jak uvádí Hurvich a spol. [13], $cAIC - 1$ je asymptoticky ekvivalentní AIC i GCV. Ve zmíněném článku je také ukázáno, že v praxi se kritéria opravdu liší chováním na menších datech. Pokud je vzorků dat mnoho, rozdíly mezi kritérii se stírají.

V předchozích sekcích bylo uvedeno, že kritéria CV a RCV jsou asymptoticky ekvivalentní kritériu GCV. Zkombinujeme-li to s tvrzením v předchozím odstavci, dostáváme, že všechna kritéria pro odhad vyhlazovacího parametru, které jsou zde uvedena (kromě MSE to se pro odhad nedá použít), jsou asymptoticky ekvivalentní. Čím se kritéria liší, je jejich chování pro malý počet vzorků dat.

3.8 Jiné metody

Velmi rychlou metodu odhadu parametrů, obzvláště pokud je parametrů více, navrhl Wood v článku [24]. Jedná se o metodu založenou na GCV

kritériu a aproximacích. Předpokladem je model, který minimalizují výraz

$$\left\| \mathbf{W}^{\frac{1}{2}} (\mathbf{A}\mathbf{a} - \mathbf{u}) \right\|^2 \lambda + \sum_{i=1}^m \theta_i \mathbf{a}^T \mathbf{S}_i \mathbf{a} \quad (3.17)$$

s omezujícími podmínkami

$$\mathbf{C}\mathbf{a} = \mathbf{0},$$

kde \mathbf{W} je diagonální matice obsahující váhy a θ_i jsou další vyhlazovací parametry. Jak do tohoto tvaru přeformulovat TPS, ukazují Wood [26], Wood a Augustin [25].

Při vlastním algoritmu se optimální hodnoty parametrů hledají iterativně, přičemž GCV kriterium se v každém kroku nepočítá znovu, ale pouze koriguje. Přínosem tohoto efektivního algoritmu je možnost značně zvýšit počet parametrů modelu. Jak ukazuje Wood [24], bylo by možné rozšířit model TPS o neisotropické vyhlazování. V každé ose modelu by bylo vyhlazování řízeno jiným parametrem, který by se samozřejmě určoval automaticky z dat, právě pomocí jím navržené metody.

Myslím, že by bylo zajímavé vyzkoušet tuto metodu na TPS, s tím že se zavede různý vyhlazovací parametr λ_i pro každý zadaný bod. To by znamenalo, že některé body jsou volnější než jiné nebo-li by to modelovalo různou přesnost zadaných bodů. To by mohlo vést k lepšímu modelování neznámé mapovací funkce. Optimální hodnoty parametrů λ_i by se určovali z dat, pomocí metody, kterou navrhl Wood [24].

Nutnou podmínkou pro použití této metody je, aby model minimalizoval výraz ve tvaru (3.17). Nepodařilo se mi však nalézt, zda lze tento převod provést pro RBF s Gaussovými funkcemi.

Kapitola 4

Navrhovaná metoda

Tato kapitola obsahuje popis navržené metody volby vyhlazovacích parametrů pro Gaussovy RBF. Nejprve jsou vysvětleny základní myšlenky metody a důvody které k nim vedly, následuje algoritmus a kapitola uzavírá sekce s popisem jak se přesně mapovací funkce konstruuje.

V rešerši známé literatury se píše, že vyhlazovací parametr λ u TPS se buď volí ručně nebo pro automatickou volbu se obvykle používá GCV. Volbu parametrů u Gaussových RBF řešili Arad a spol. [2] v jejich článku, kde se Gaussovy RBF používají pro „warping“ obličejů. Parametry λ i σ^2 se zde určují ručně, a považují se za parametry pro „warping“, případně se při animaci skokově mění, provádí se však jejich normalizace k rozměrům deformovaného obrázku.

Proč však nehledat optimální hodnotu parametru σ^2 také pomocí nějakého kritéria? Tento pohled na parametr σ^2 jako na jiný vyhlazovací parametr podporují pokusy v sekci 5.3, kde se ukazuje, že optimální hodnota parametru se při zafixovaném λ mění v závislosti na rozptylu přidávaného šumu. To přesně bychom očekávali od vyhlazovacího parametru a také se tak λ při zafixovaném σ^2 chová. Dále, s rostoucím σ^2 se zvětšuje oblast, kterou bod ovlivňuje, změny se stávají pozvolnější a výsledná funkce je proto hladší, to však přesně funkce vyhlazovacího parametru.

Proto vznikla myšlenka více kriteriální minimalizace, kdy se budou oba parametry λ i σ^2 automaticky volit tak, aby bylo dosaženo nejlepšího výsledku. Algoritmus metody je popsán v následující sekci, ještě zbývá vybrat z používaných kritérií pro volbu vyhlazovacích parametrů to nejvhodnější, pomocí kterého se budou hodnoty parametrů λ a σ^2 volit, to je diskutováno v sekci 4.2.

4.1 Minimalizace

Vstupem algoritmu jsou korespondující dvojice řídicích bodů a kritérium, které se má použít, výstupem jsou optimální hodnoty λ a σ^2 vzhledem

k zadanému kritériu, nebo-li hodnoty parametrů, kde je kritérium minimální.

Počáteční hodnoty jsou standardně nastaveny na $\log(\lambda) = -3$ a σ^2 je průměr vzdáleností dvou nejbližších a dvou nejvzdálenějších řídicích bodů. Algoritmus dále postupuje iterativně, s krokem standardně 10^{-3} pro $\log(\lambda)$ a 10^{-2} pro σ^2 , jak je vidět algoritmus pracuje s logaritmickou stupnicí u parametru λ . V každé iteraci se algoritmus podívá do čtyř směrů podél os (to je lehce možné změnit za jiný vzor, ale tento se nejvíce osvědčil) do vzdálenosti kroku a pokračuje směrem s nejmenší hodnotou kritéria.

Podmínky ukončení iterace jsou následující, pokud je v sousedních bodech hodnota větší než v aktuálním bodě, bodem se myslí dvojice $[\log(\lambda), \sigma^2]$. Pokud se λ ocitne mimo interval (e^{-10}, e^{10}) , tento interval doporučuje Ber-
man [3] jako dobrou aproximaci pro $(0, \infty)$. Samozřejmě podmínkou ukončení je také dosažení maximálního počtu iterací.

Při pokusech popsaných v sekci 5.3 se ukázalo, že pro daný vstup existuje nějaký vhodný poměr λ a σ^2 , při jehož zachování a rostoucím σ^2 hodnota kritéria stále klesá. Protože však hodnota kritéria později klesá pouze nepatrně, řekl bych zanedbatelně, iterace se ukončí, když desetkrát po sobě klesne hodnota kritéria méně než je stanovená mez ($5 \cdot 10^{-6}$). Přidání této ukončovací podmínky se ukázalo prospěšné, neboť vhodný poměr se u kritérií a MSE malinko liší. Pokud by σ^2 při zachování poměru stále vrůstalo, tak ačkoliv by hodnota kritéria klesala, MSE by stoupalo.

Při použití TPS se pro hledání optimální hodnoty parametru λ používá standardní funkce `FMINBND` z programu MATLAB. Minimalizace se provádí na logaritmické stupnici v intervalu (e^{-10}, e^{10}) .

4.2 Mapování

Vstupem pro algoritmus konstrukce mapovací funkce jsou korespondující dvojice řídicích bodů a obrázek, který se má transformovat.

Konstrukce mapovací funkce se provádí standardně, pouze pro volbu parametrů λ , případně σ^2 , se používá výše uvedená metoda minimalizace. Přesný postup konstrukce mapovací funkce je popsán v sekcích 2.3 a 2.4. Používá se však zpětné mapování, to znamená, že se nehledá mapovací funkce $T : zdroj \mapsto cil$, ale hledá se rovnou funkce inverzní $T^{-1} : cil \mapsto zdroj$. V implementaci to znamená pouze zaměnit zdrojové a cílové řídicí body, postup se nemění. Při převzorkování se používá lineární interpolace.

Jako nejvhodnější kritérium pro minimalizaci a volbu parametrů bylo nakonec na základě pokusů, které jsou popsány v následující kapitole, vybráno GCV kritérium. Ukázalo se že toto kritérium produkuje stabilně většinou nejlepší výsledky (podle MSE) a nemá problémy s existencí minima nebo malým počtem vstupních bodů.

Kapitola 5

Experimenty

Tato kapitola obsahuje experimenty s konstrukcí mapovací funkce pomocí TPS a Gaussových RBF. První dvě sekce ukazují, jak se mění výsledná mapovací funkce při změně parametrů λ a σ^2 . Další sekce se zabývá studiem různých kritérií pro automatickou volbu vyhlazovacích parametrů a zkoumá, jak tyto kritéria reagují na různou úroveň šumu ve vstupních datech. Kapitola potom uzavírá výsledky navrhované metody, spolu s předvedením možností použití.

Všechny experimenty byly implementovány v prostředí MATLAB a jsou dostupné na přiloženém CD.

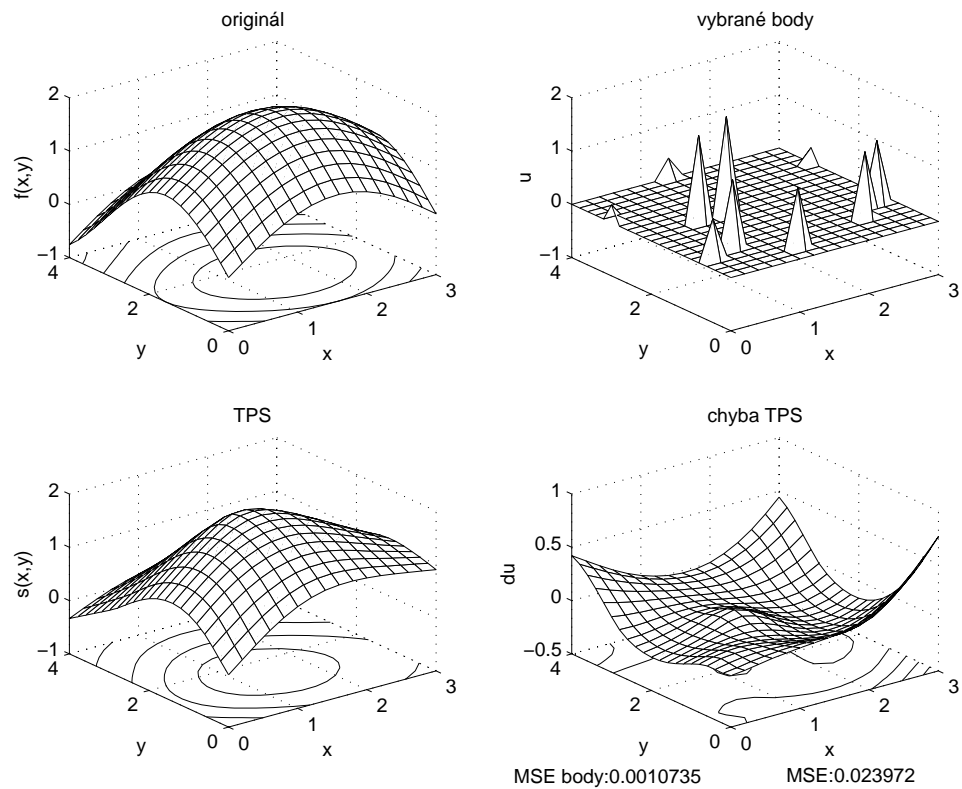
5.1 Interpolace

Tato sekce ukazuje, jak se liší výsledná mapovací funkce při použití různých bází, konkrétně se jedná o TPS a Gaussovy RBF s parametrem $\sigma^2 = 1, 4, 8, 32$.

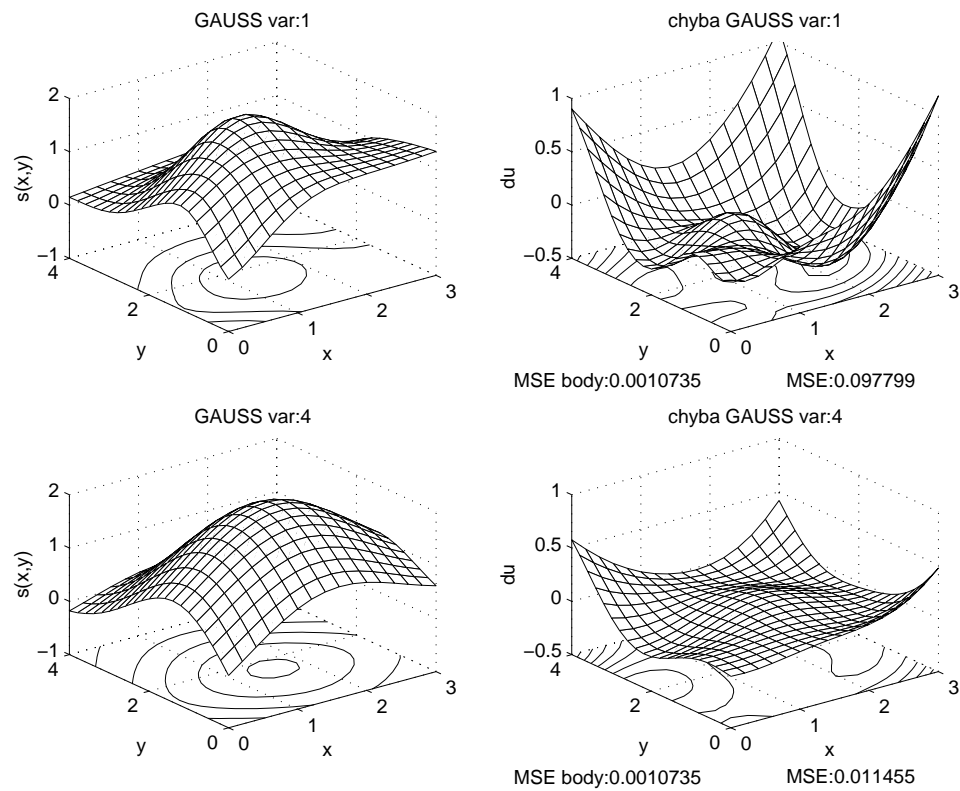
Pro interpolaci byla zvolena funkce $\mathbb{R}^2 \rightarrow \mathbb{R} : \sin(x) + \sin(y)$. Na kompaktní oblasti $[0, 3] \times [0, 4]$ bylo vybráno náhodně 10 bodů, k jejichž hodnotám byl přičten vygenerovaný šum $\sim N(0, \frac{1}{625})$. Tyto body a hodnoty byly potom vstupem pro interpolaci (obr. 5.1). Na dalších obrázcích je vždy vlevo interpolace pomocí zvolené báze, postupně jsou to TPS (obr. 5.1), Gaussovy RBF s rozptyly 1, 4 (obr. 5.2) a Gaussovy RBF s rozptyly 8, 32 (obr. 5.3). Vpravo od grafu interpolace je zobrazen rozdíl mezi touto interpolací a originální funkcí, pod grafy rozdílů jsou spočítány hodnoty MSE na bodech, které tvořily zadání, a MSE vyčíslená v pravidelné mřížce po 0.2 na oblasti $[0, 3] \times [0, 4]$.

Jak je vidět na obrázku 5.1, už pouhých 10 bodů stačí TPS, aby docela dobře vystihlo původní tvar funkce. Výsledky interpolace pomocí Gaussových RBF hodně závisí na volbě parametru σ^2 . Pokud je rozptyl zvolen příliš malý, chyba vzniklá při odhadu původní funkce je velká (obr. 5.2). Se zvyšujícím σ^2 se však vlastnosti Gaussových RBF zlepšují až dokonce překonají

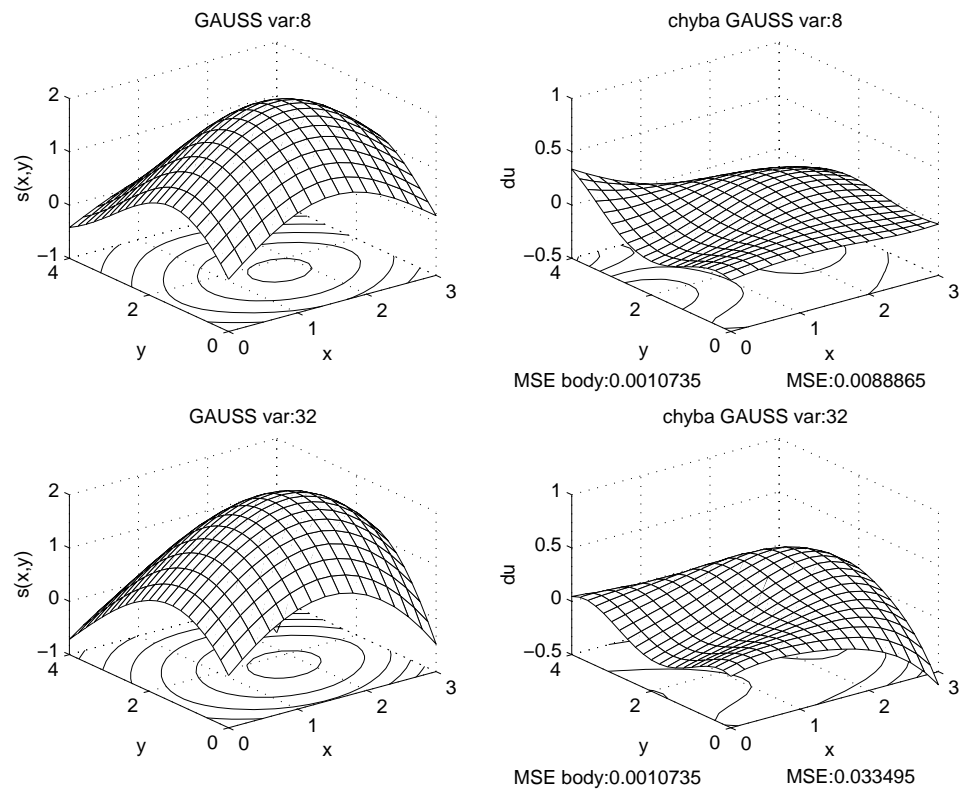
TPS (obr. 5.1, $\sigma^2 = 8$), měřeno pomocí MSE na pravidelné mřížce. Pokud parametr σ^2 dále zvyšujeme, chyba interpolace začne opět narůstat.



Obrázek 5.1: Interpolace pomocí TPS – nahoře je graf originální funkce, vpravo jsou z ní náhodně vybrané body, k jejichž hodnotám byl přičten šum $\sim N(0, \frac{1}{625})$, dole je TPS zkonstruovaná se znalostí těchto bodů a hodnot. Vpravo je rozdíl interpolace a originálu, pod grafem rozdílu jsou hodnoty MSE na zadaných bodech a pravidelné mřížce.



Obrázek 5.2: Interpolace pomocí Gaussových RBF – nahoře pro $\sigma^2 = 1$ a dole pro $\sigma^2 = 4$; vlevo je interpolace zkonstruovaná se znalostí bodů a hodnot z obrázku 5.1. Vpravo je rozdíl interpolace a originálu, pod grafy rozdílů jsou hodnoty MSE na zadaných bodech a pravidelné mřížce.



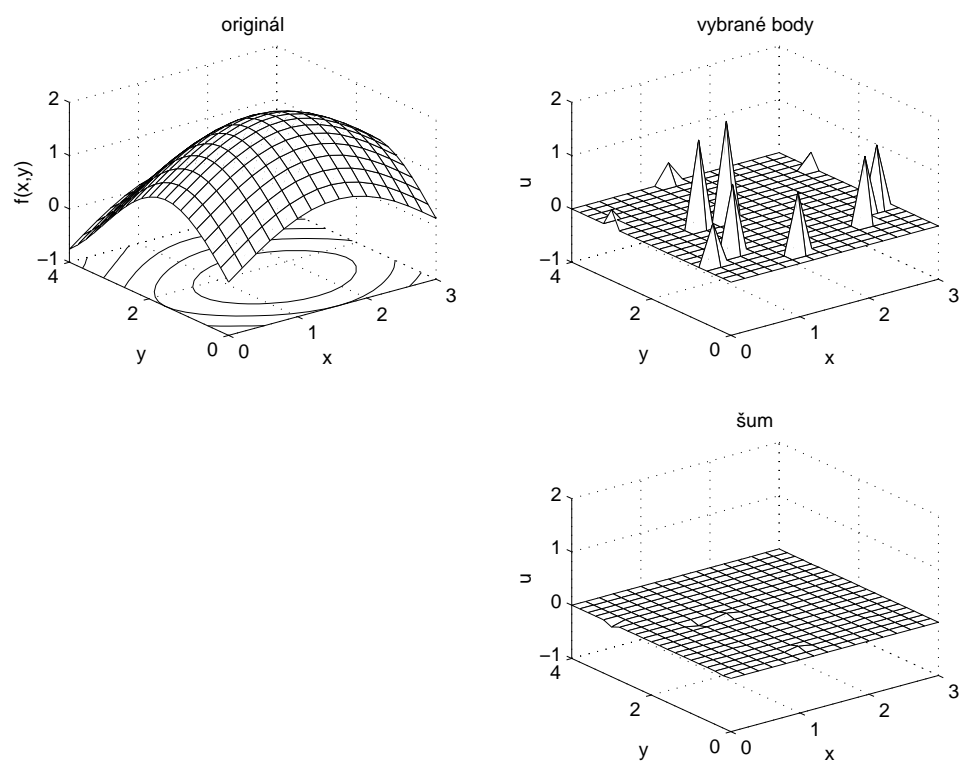
Obrázek 5.3: Interpolace pomocí Gaussových RBF – nahoře pro $\sigma^2 = 8$ a dole pro $\sigma^2 = 32$; vlevo je interpolace zkonstruovaná se znalostí bodů a hodnot z obrázku 5.1. Vpravo je rozdíl interpolace a originálu, pod grafy rozdílů jsou hodnoty MSE na zadaných bodech a pravidelné mřížce.

5.2 Aproximace

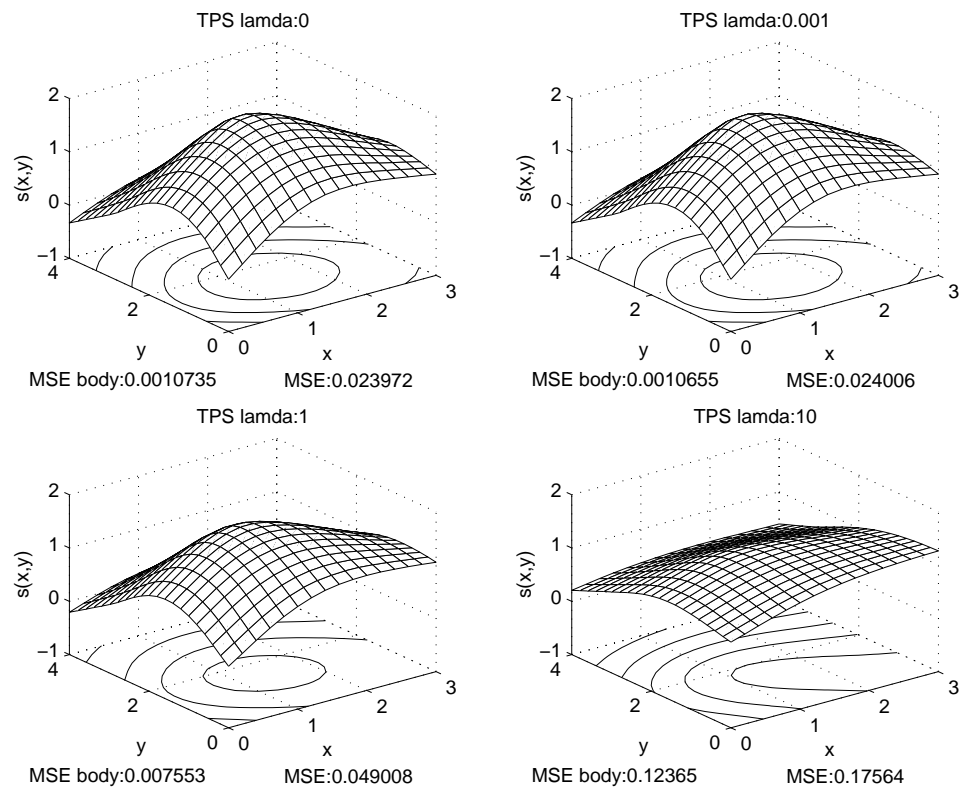
Tato sekce ukazuje, jak se mění výsledná mapovací funkce při změně vyhlazovacího parametru λ , zadání je stejné jako pro interpolaci. Porovnávají se opět TPS a Gaussovými RBF, které mají zafixovaný parametr $\sigma^2 = 8$, mění se parametr λ .

Zadání opět tvoří funkce $\sin(x) + \sin(y)$ na oblasti $[0, 3] \times [0, 4]$, ze které bylo vybráno 10 bodů a k jejich hodnotám byl přičten šum $\sim N(0, \frac{1}{625})$ (body i hodnoty šumu jsou stejné jako u interpolace). Na obrázku 5.5 je pro aproximaci použit TPS, postupně s parametrem $\lambda = 0, 0.001, 0.1, 10$. Na následujícím obrázku 5.5 je použita aproximace pomocí Gaussových RBF, opět postupně s $\lambda = 0, 0.001, 0.1, 10$.

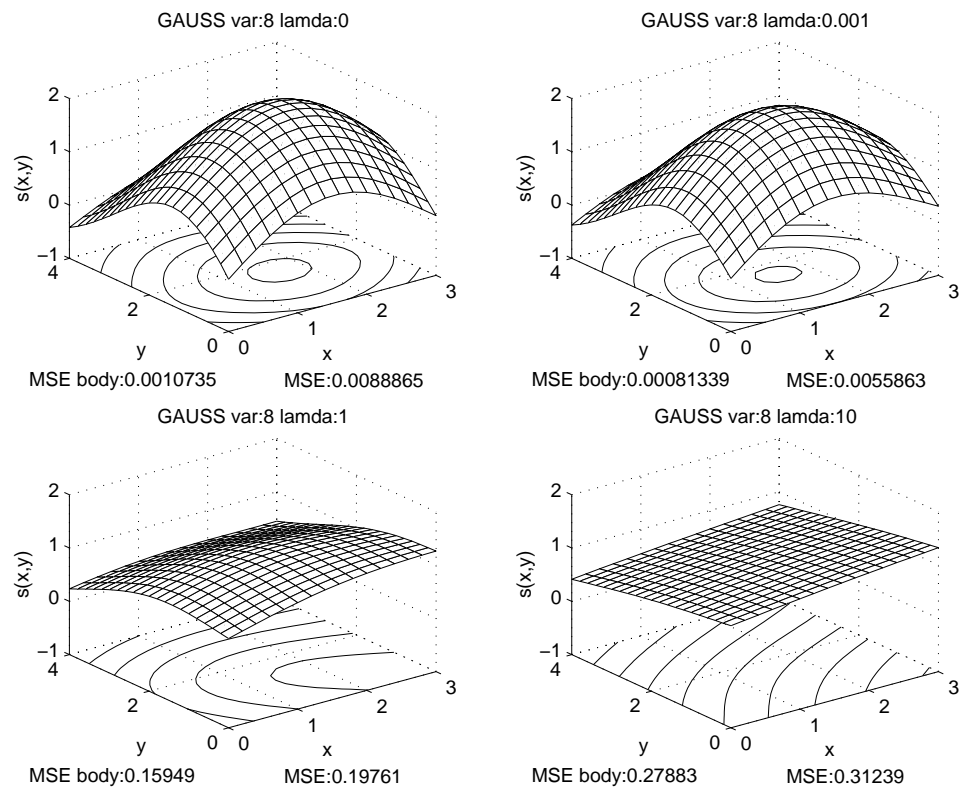
Z obrázků můžeme jasně vidět, že se zvyšující hodnotou parametru λ se zvětšuje míra vyhlazení a funkce se blíží rovině, která je proložená zadanými body metodou nejmenších čtverců. Je také vidět, že při zvolené hodnotě $\sigma = 8$ dochází u Gaussových RBF k vyhlazení dříve než u TPS. Obrázek 5.6 ukazuje, že při $\lambda = 0.001$ a $\sigma = 8$ překonávají Gaussovy RBF výsledky, kterých se dosáhlo aproximací pomocí TPS, alespoň podle MSE. Stále jsou to však spíše demonstrace, opravdové porovnání lze nalézt v sekci 5.4.



Obrázek 5.4: Zadání pro aproximaci – stejné se zadáním interpolace, nahoře je originální funkce a vpravo z ní vybrané body; dole je šum $\sim N(0, \frac{1}{625})$, který se přičte k hodnotám v bodech (body i hodnoty šumu jsou stejné jako u interpolace).



Obrázek 5.5: Aproximace pomocí TPS – zkonstruovaná se znalostí bodů a hodnot z obrázku 5.4, vyhlazovací parametr nabývá postupně hodnot $\lambda = 0, 0.001, 0.1, 10$, pod grafy jsou hodnoty MSE na zadaných bodech a pravidelné mřížce.



Obrázek 5.6: Aproximace pomocí Gassových RBF – zkonstruovaná se znalostí bodů a hodnot z obrázku 5.4, vyhlazovací parametr nabývá postupně hodnot $\lambda = 0, 0.001, 0.1, 10$, rozptyl je zvolený pevně $\sigma^2 = 8$, pod grafy jsou hodnoty MSE na zadaných bodech a pravidelné mřížce.

5.3 Kritéria

Sekce kritéria zkoumá kriteria pro volbu vyhlazovacích parametrů, která byla popsána v kapitole 3. Je zde ukázáno, jak se mění průběhy kritérií při změně rozptylu šumu v zadání, případně při zvětšení počtu zadaných bodů.

Pro první tři příklady bylo použito stejné zadání jako pro příklady interpolace a aproximace (včetně bodů a hodnot šumu), pouze byly hodnoty šumu vynásobeny konstantou, aby bylo možné sledovat reakci kritérií na změnu rozptylu šumu. Poslední příklad je výjimkou, pracuje se 100 vstupními body, v tomto příkladě zůstalo stejné zadání aproximované funkce, pouze vybrané body a šum byly znovu vygenerovány (body s rovnoměrným rozdělením, šum s rozdělením normálním).

Jedná se tedy opět o funkci $\sin(x) + \sin(y)$ na oblasti $[0, 3] \times [0, 4]$, na které byl vybrán příslušný počet bodů a k jejich hodnotám byl přičten šum (viz tabulka 5.1). Z těchto bodů a hodnot se pro pevné λ a σ^2 vypočítá "hat matrix" (viz sekce 3.3) a následně kritérium, celý postup se opakuje pro jiné hodnoty parametrů.

Číslo příkladu	Vstupní funkce	Počet bodů	Rozptyl šumu
1	$\sin(x) + \sin(y)$	10	1
2	$\sin(x) + \sin(y)$	10	$\frac{1}{9}$
3	$\sin(x) + \sin(y)$	10	$\frac{1}{900}$
4	$\sin(x) + \sin(y)$	100	$\frac{1}{9}$

Tabulka 5.1: Tabulka zadání příkladu pro studium průběhu kritérií

Porovnávána byla následující kritéria:

1. krosvalidace (CV2) se vztahem (3.10), tento vztah umožňuje rychlejší výpočet a výsledky jsou stejné jako při použití (3.3)
2. Generalized Cross Validation (GCV) se vztahem (3.11)
3. Robustified Cross Validation (RCV) s vzorcem (3.14)
4. Akaike Information Criterion (AIC) počítané pomocí vztahu (3.15)
5. Corrected Akaike Information Criterion (cAIC) (3.16)
6. pro porovnání i graf střední kvadratické odchylky (MSE) viz (3.2)

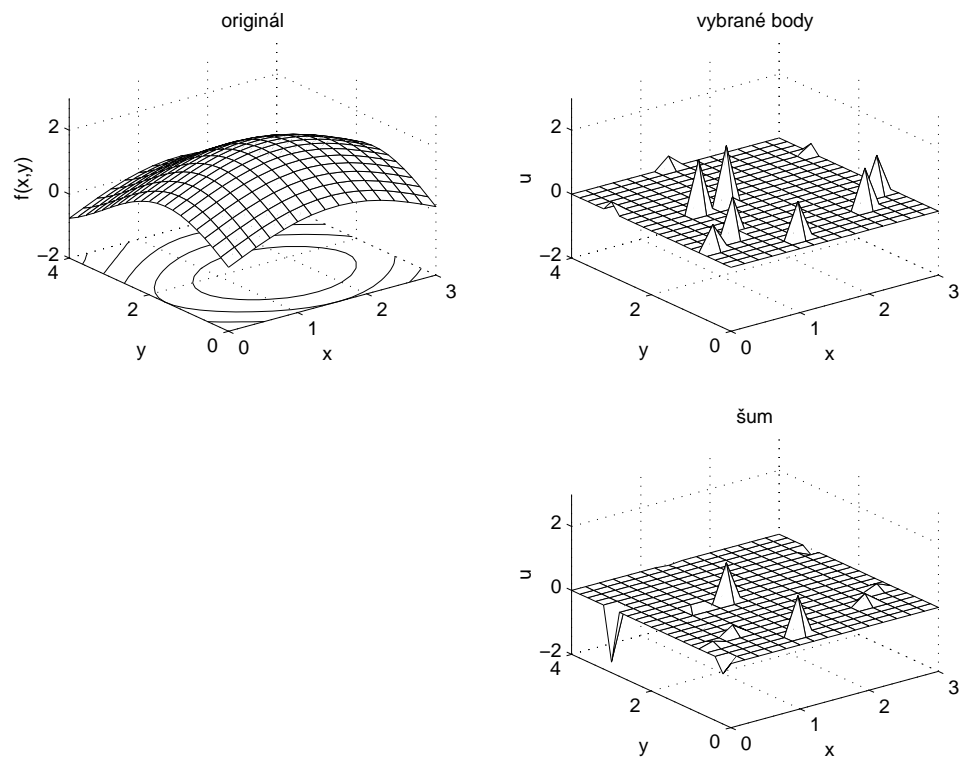
Každý příklad obsahuje grafy průběhu kritérií pro TPS v závislosti na λ a pro Gaussovy RBF v závislosti na σ^2 při pevném λ a opačně v závislosti na λ při pevném σ^2 . Příklad vždy uzavírá graf vývoje kritérií pro Gaussovy RBF v závislosti na λ i σ^2 .

Na grafech, které tvoří zbytek této sekce, můžeme vidět následující skutečnosti. Grafy jsou v souladu s tvrzením uvedeném v kapitole 3, že cAIC a obzvláště AIC špatně fungují pro malý počet vstupních hodnot. Jak je vidět na příkladech 1 až 3, deset vstupních hodnot je opravdu málo a kritéria nedávají žádné rozumné výsledky. Pokud v příkladu 4 zadáme 100 vstupních bod, situace se alespoň pro Gaussovy RBF zlepšuje (viz obr. 5.22 a 5.23).

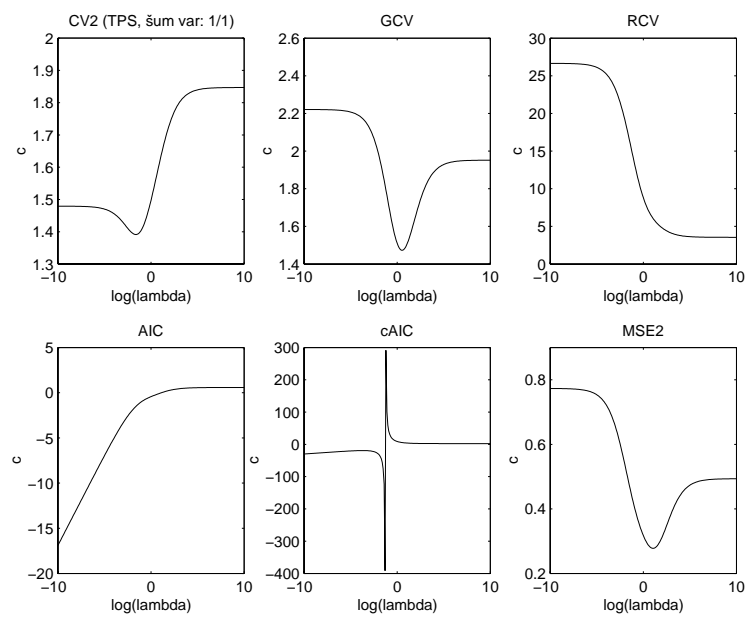
Na dobře fungujících kriteriích jako CV2, GCV a MSE je zřetelně vidět, jak se posouvá hodnota λ , kterou považují kritéria za optimální, podle toho jak se mění rozptyl přidávaného šumu. Při větším rozptylu šumu volí kritéria větší λ a tedy větší vyhlazení, zatímco při menším rozptylu označují kritéria jako optimální menší hodnoty λ . Kritéria se tedy v tomto chovají tak, jak bychom od nich očekávali. RCV kritérium, v souladu s teorií, upřednostňuje větší vyhlazení než GCV, někdy však má problémy s konvergencí a volí extrémní vyhlazení.

Při pevném λ se při změně rozptylu šumu mění volené σ^2 , to opravňuje pohled na σ^2 jako na vyhlazovací parametr. Dále se zdá, že pro konkrétní vstup existuje nějaký výhodný poměr mezi λ a σ^2 , při jehož zachování, s klesajícím λ a stoupajícím σ^2 hodnota kriterií (CV2, GCV, MSE) stále klesá, od jisté hranice však zanedbatelně. Toto jsou pouze náhledy, přesná čísla je možné nalézt v následující sekci.

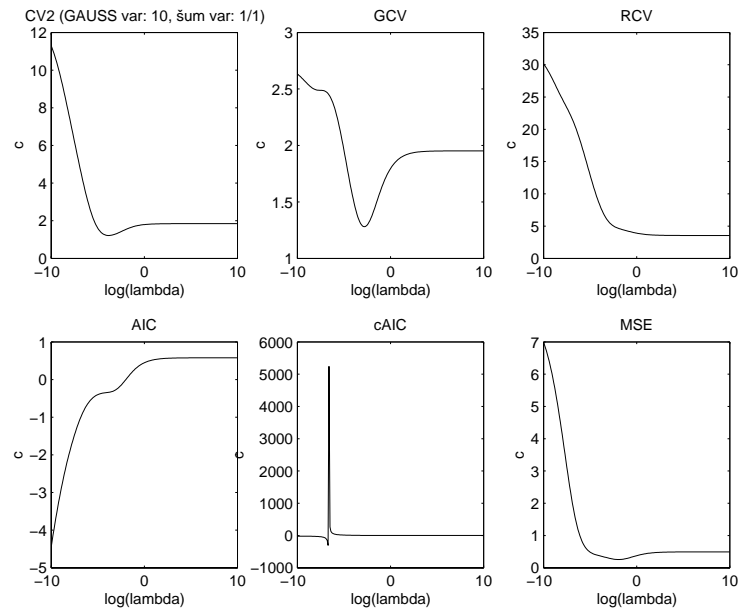
5.3.1 příklad 1



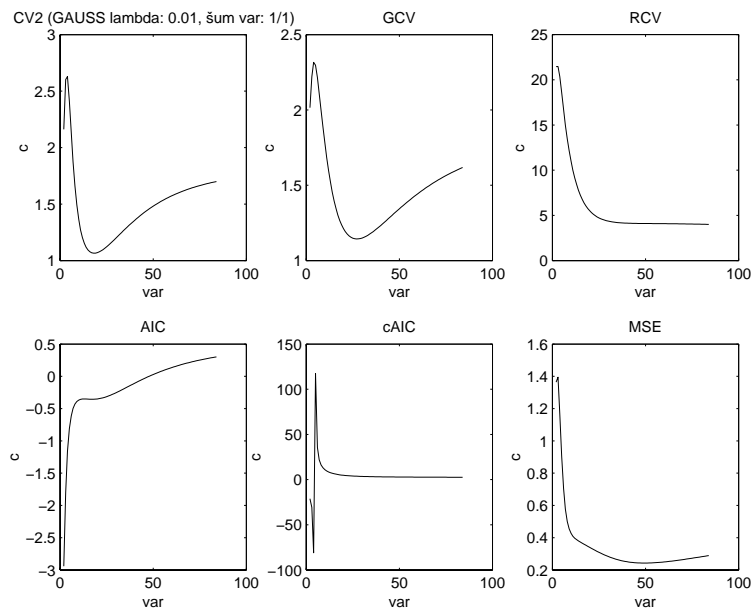
Obrázek 5.7: Zadání prvního příkladu – podobné zadání pro interpolaci, nahoře je originální funkce a vpravo z ní vybrané body, dole je šum $\sim N(0, 1)$, který se přičte k hodnotám v bodech (body jsou stejné jako u interpolace, hodnoty šumu jsou pouze vynásobené konstantou podle tabulky 5.1).



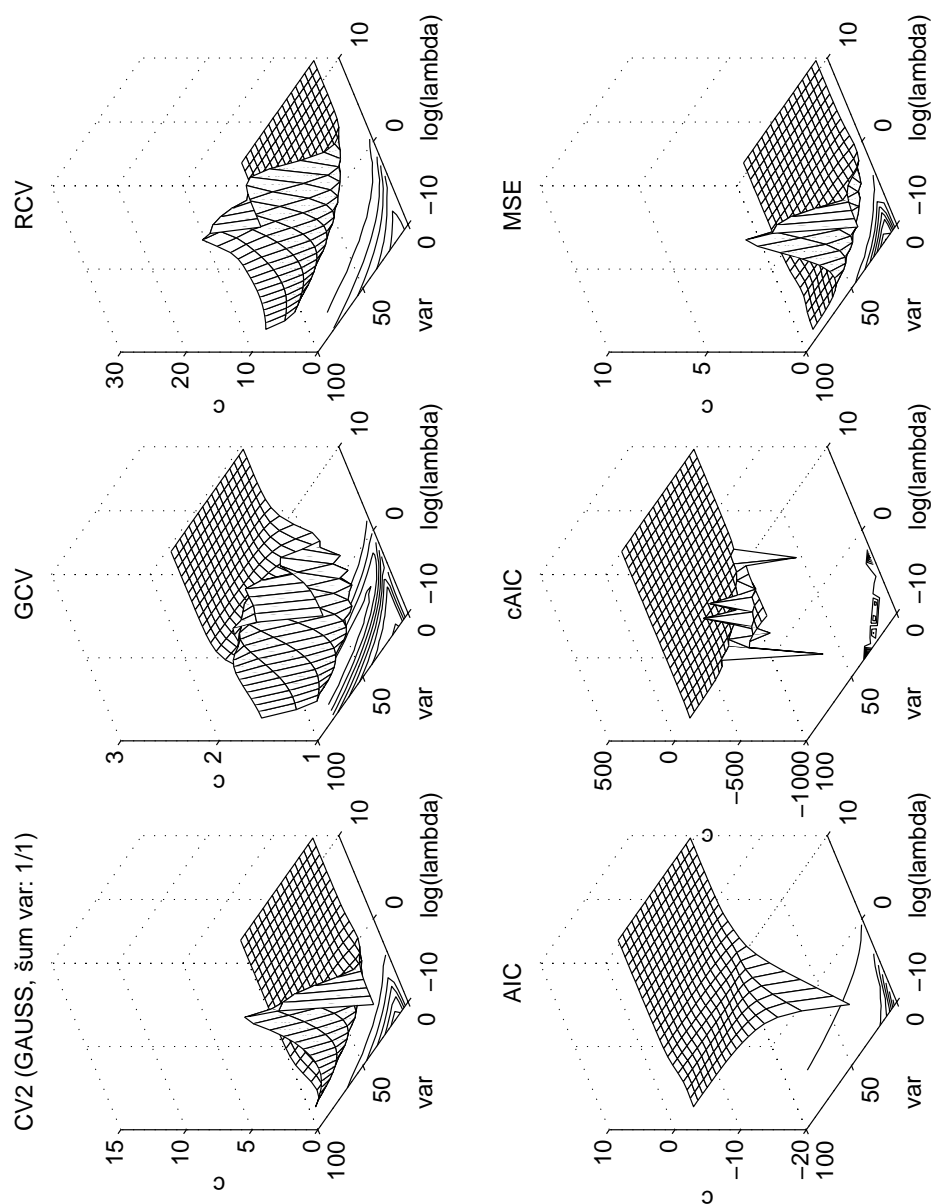
Obrázek 5.8: př.1: Průběh kritérií pro TPS v závislosti na λ – vstupem jsou hodnoty a body z obrázku 5.7 (šum $\sim N(0, 1)$), kritéria měří vhodnost vyhlazovacího parametru λ , menší hodnota je lepší.



Obrázek 5.9: př.1: Průběh kritérií pro Gaussovy RBF v závislosti na λ – vstupem jsou hodnoty a body z obrázku 5.7 (šum $\sim N(0,1)$), kritéria měří vhodnost vyhlazovacího parametru λ , menší hodnota je lepší.

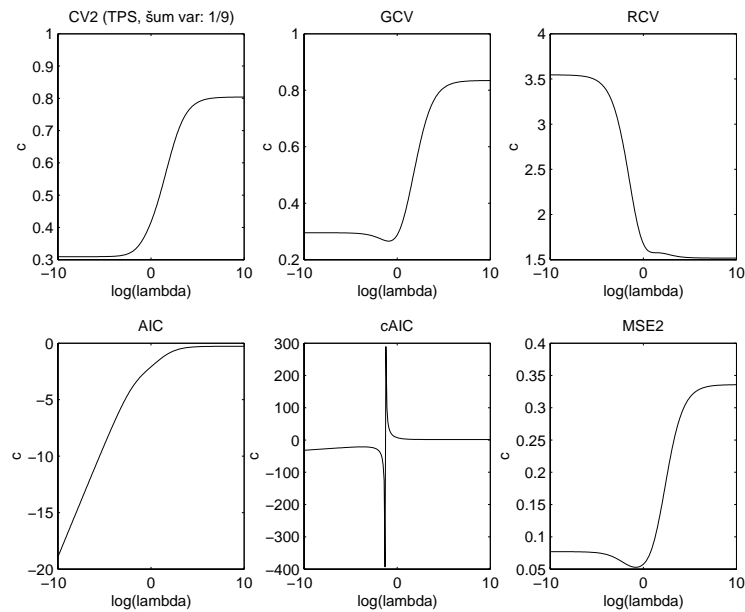


Obrázek 5.10: př.1: Průběh kritérií pro Gaussovy RBF v závislosti na σ^2 – vstupem jsou hodnoty a body z obrázku 5.7 (šum $\sim N(0,1)$), kritéria měří vhodnost parametru σ^2 , menší hodnota je lepší.

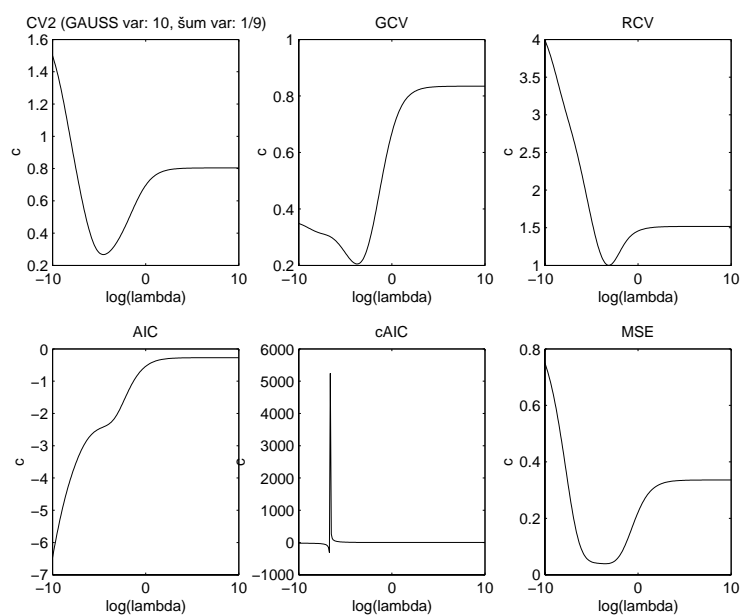


Obrázek 5.11: př.1: Průběh kritérií pro Gaussovy RBF v závislosti na σ^2 a λ – vstupem jsou hodnoty a body z obrázku 5.7 (šum $\sim N(0, 1)$), kritéria měří vhodnost parametrů σ^2 a λ , menší hodnota je lepší.

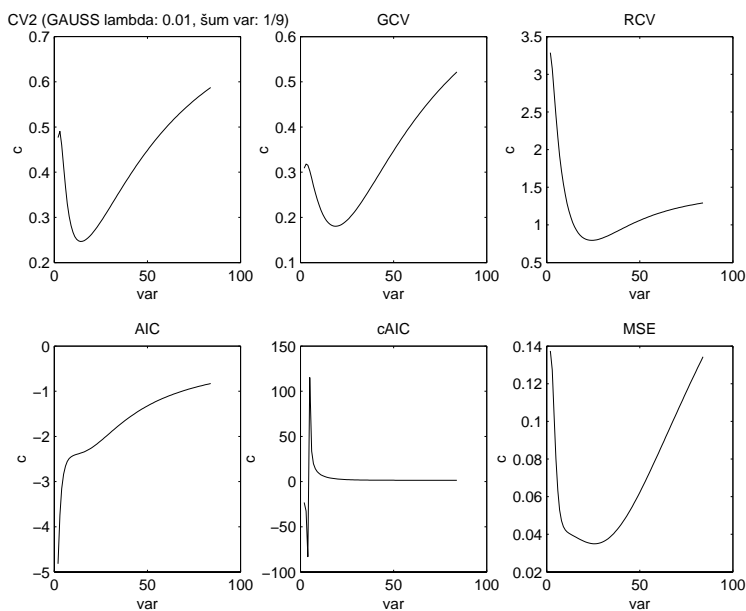
5.3.2 příklad 2



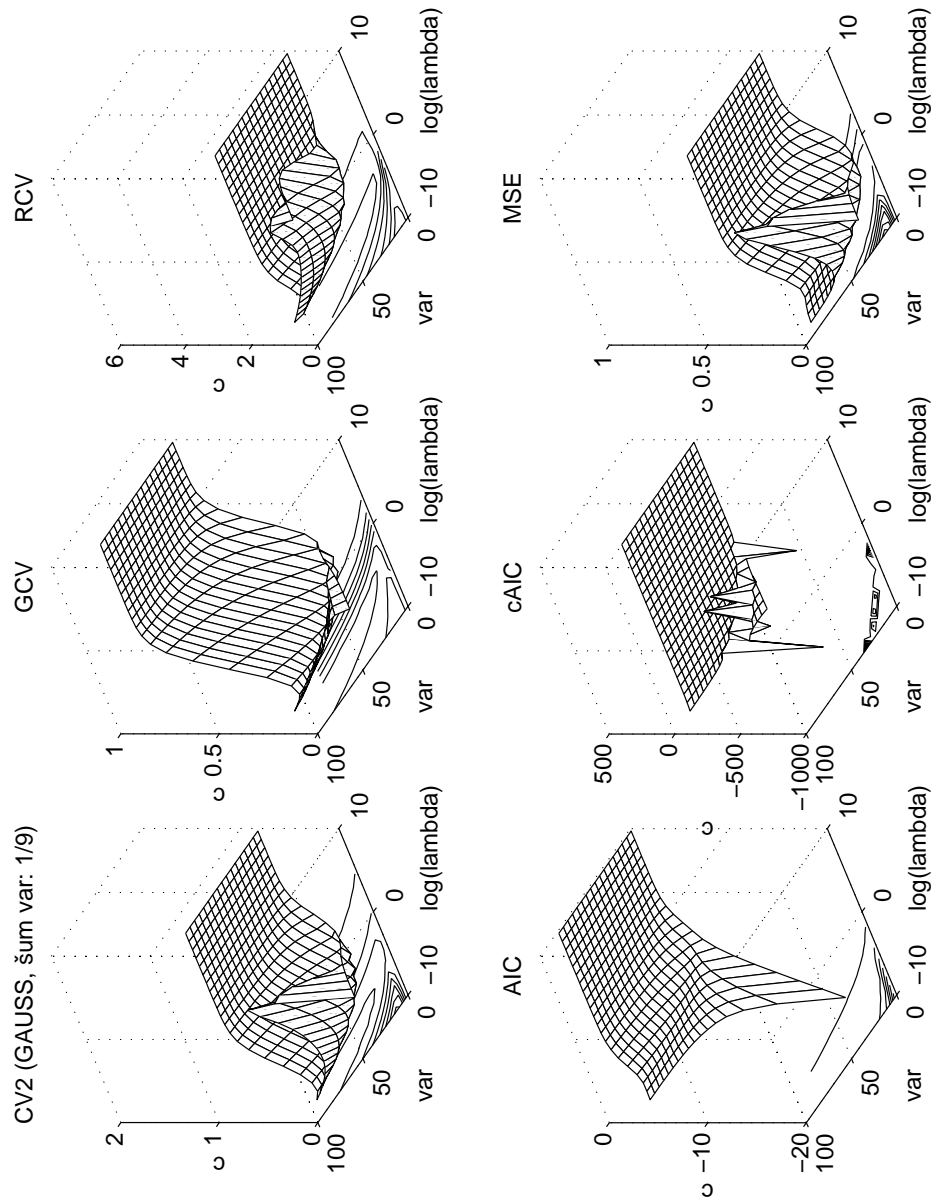
Obrázek 5.12: př.2: Průběh kritérií pro TPS v závislosti na λ – vstupem jsou hodnoty a body z obrázku 5.7 (šum $\sim N(0, \frac{1}{9})$), kritéria měří vhodnost vyhlazovacího parametru λ , menší hodnota je lepší.



Obrázek 5.13: př.2: Průběh kritérií pro Gaussovy RBF v závislosti na λ – vstupem jsou hodnoty a body z obrázku 5.7 (šum $\sim N(0, \frac{1}{9})$), kritéria měří vhodnost vyhlazovacího parametru λ , menší hodnota je lepší.

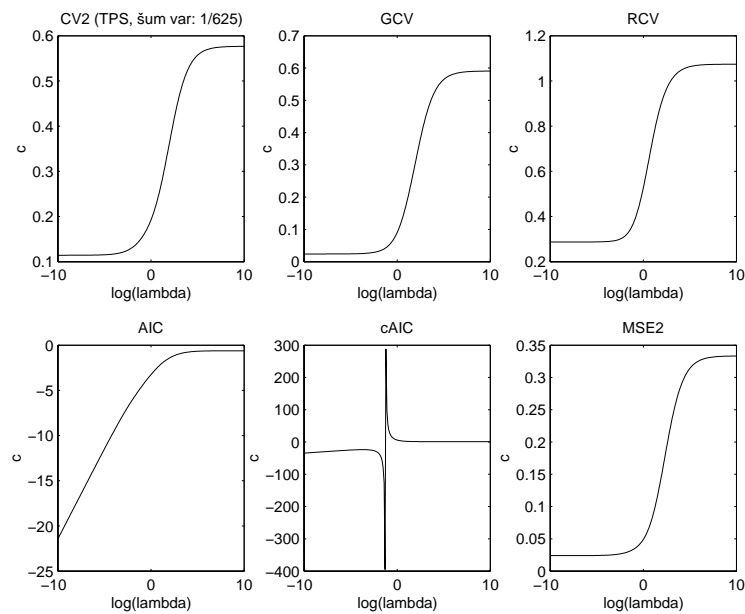


Obrázek 5.14: př.2: Průběh kritérií pro Gaussovy RBF v závislosti na σ^2 – vstupem jsou hodnoty a body z obrázku 5.7 (šum $\sim N(0, \frac{1}{9})$), kritéria měří vhodnost parametru σ^2 , menší hodnota je lepší.

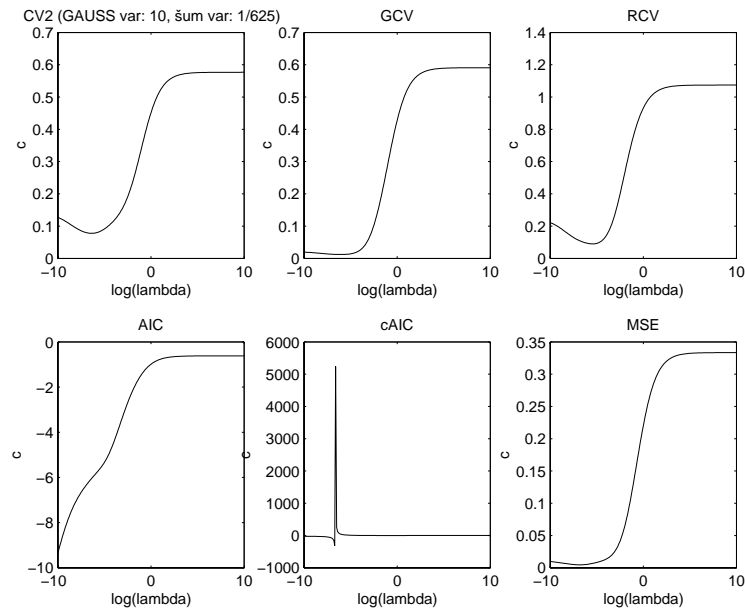


Obrázek 5.15: př.2: Průběh kritérií pro Gaussovy RBF v závislosti na σ^2 a λ – vstupem jsou hodnoty a body z obrázku 5.7 (šum $\sim N(0, \frac{1}{9})$), kritéria měří vhodnost parametrů σ^2 a λ , menší hodnota je lepší.

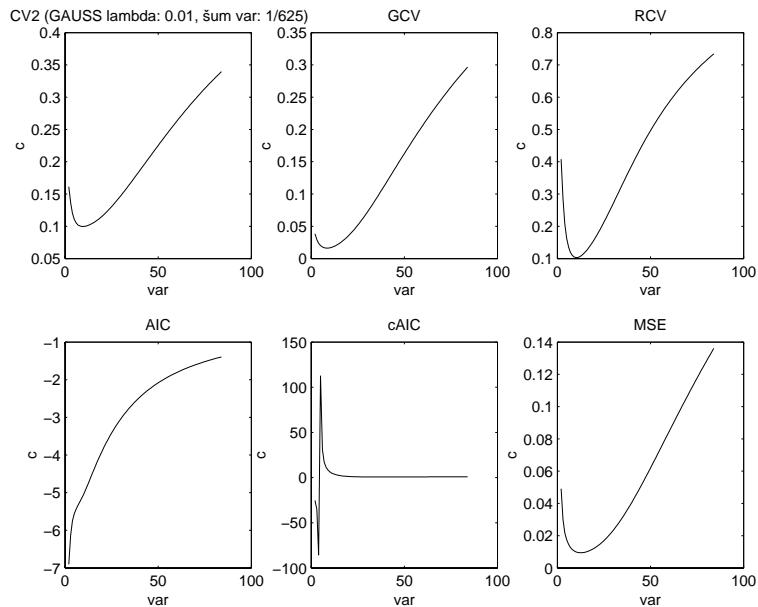
5.3.3 příklad 3



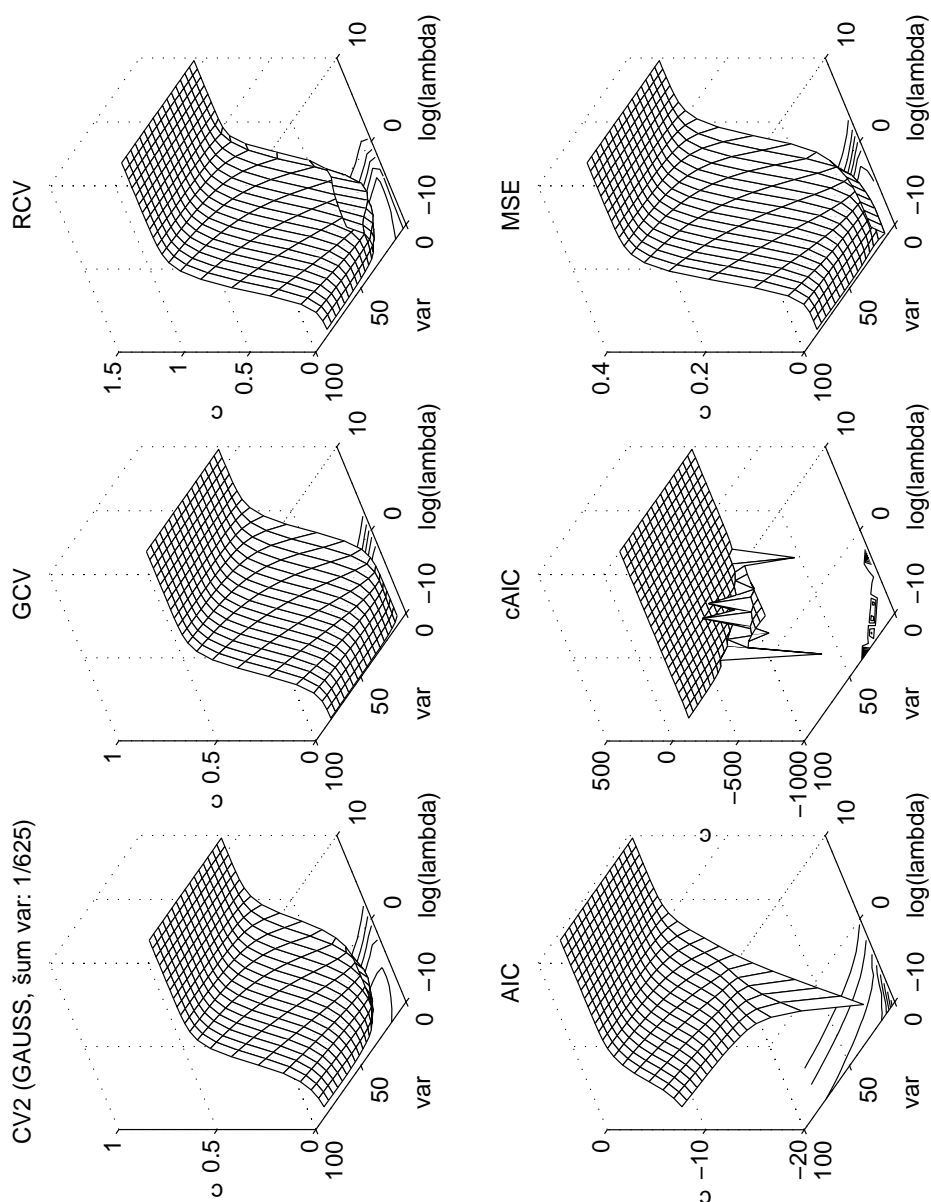
Obrázek 5.16: př.2: Průběh kritérií pro TPS v závislosti na λ – vstupem jsou hodnoty a body z obrázku 5.7 (šum $\sim N(0, \frac{1}{900})$), kritéria měří vhodnost vyhlazovacího parametru λ , menší hodnota je lepší.



Obrázek 5.17: př.2: Průběh kritérií pro Gaussovy RBF v závislosti na λ – vstupem jsou hodnoty a body z obrázku 5.7 (šum $\sim N(0, \frac{1}{900})$), kritéria měří vhodnost vyhlazovacího parametru λ , menší hodnota je lepší.

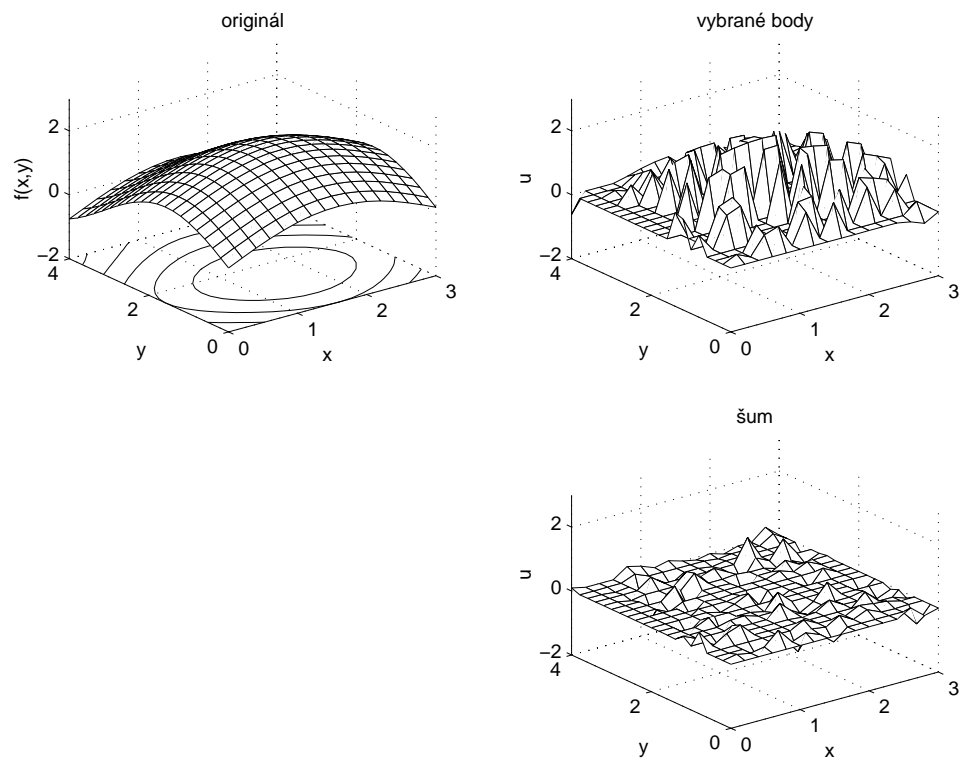


Obrázek 5.18: př.2: Průběh kritérií pro Gaussovy RBF v závislosti na σ^2 – vstupem jsou hodnoty a body z obrázku 5.7 (šum $\sim N(0, \frac{1}{900})$), kritéria měří vhodnost parametru σ^2 , menší hodnota je lepší.

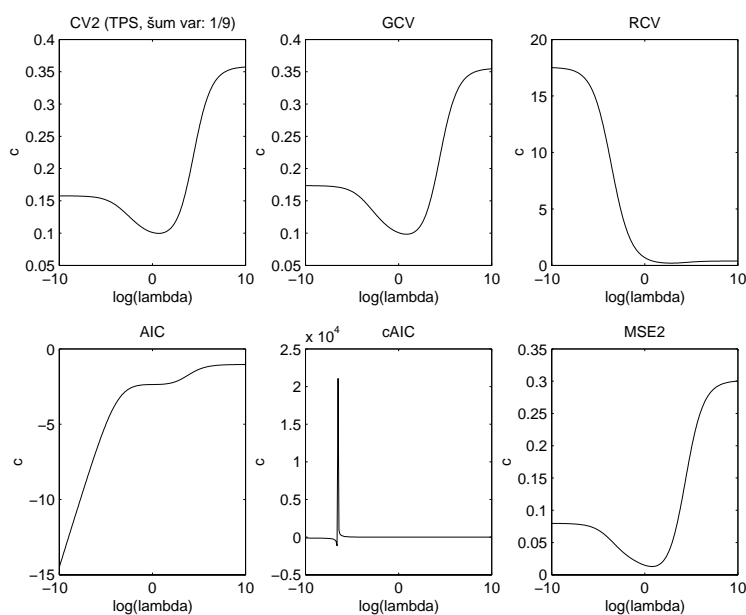


Obrázek 5.19: př.2: Průběh kritérií pro Gaussovy RBF v závislosti na σ^2 a λ – vstupem jsou hodnoty a body z obrázku 5.7 (šum $\sim N(0, \frac{1}{900})$), kritéria měří vhodnost parametrů σ^2 a λ , menší hodnota je lepší.

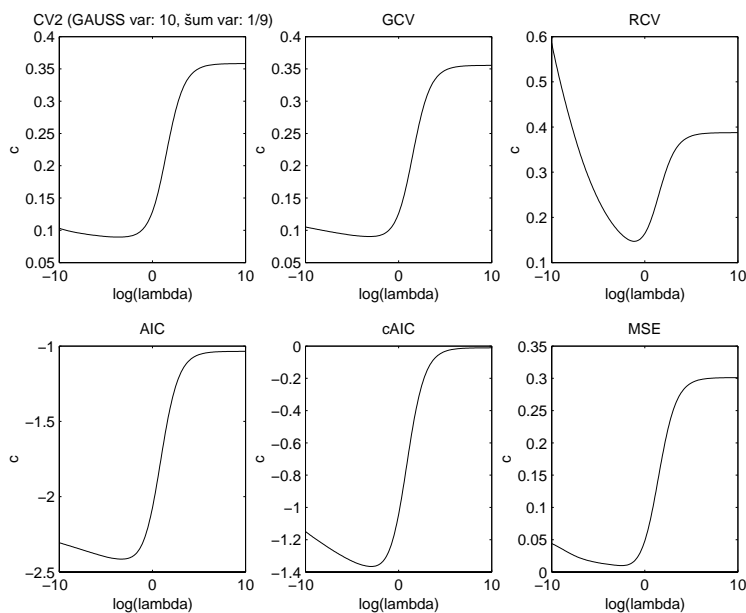
5.3.4 příklad 4



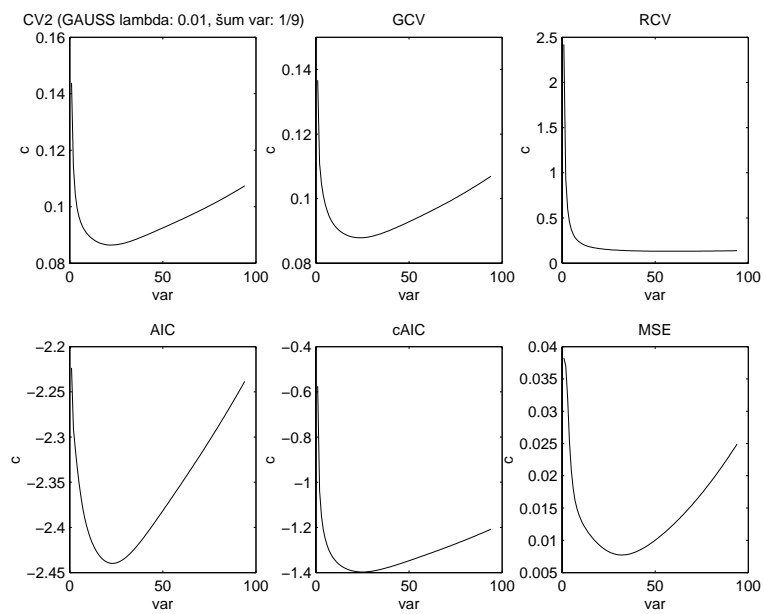
Obrázek 5.20: Zadání čtvrtého příkladu – zadání podobné prvnímu příkladu pouze je více bodů, nahoře je originální funkce a vpravo z ní 100 vybraných bodů, dole je šum $\sim N(0, \frac{1}{9})$, který se přičte k hodnotám v bodech.



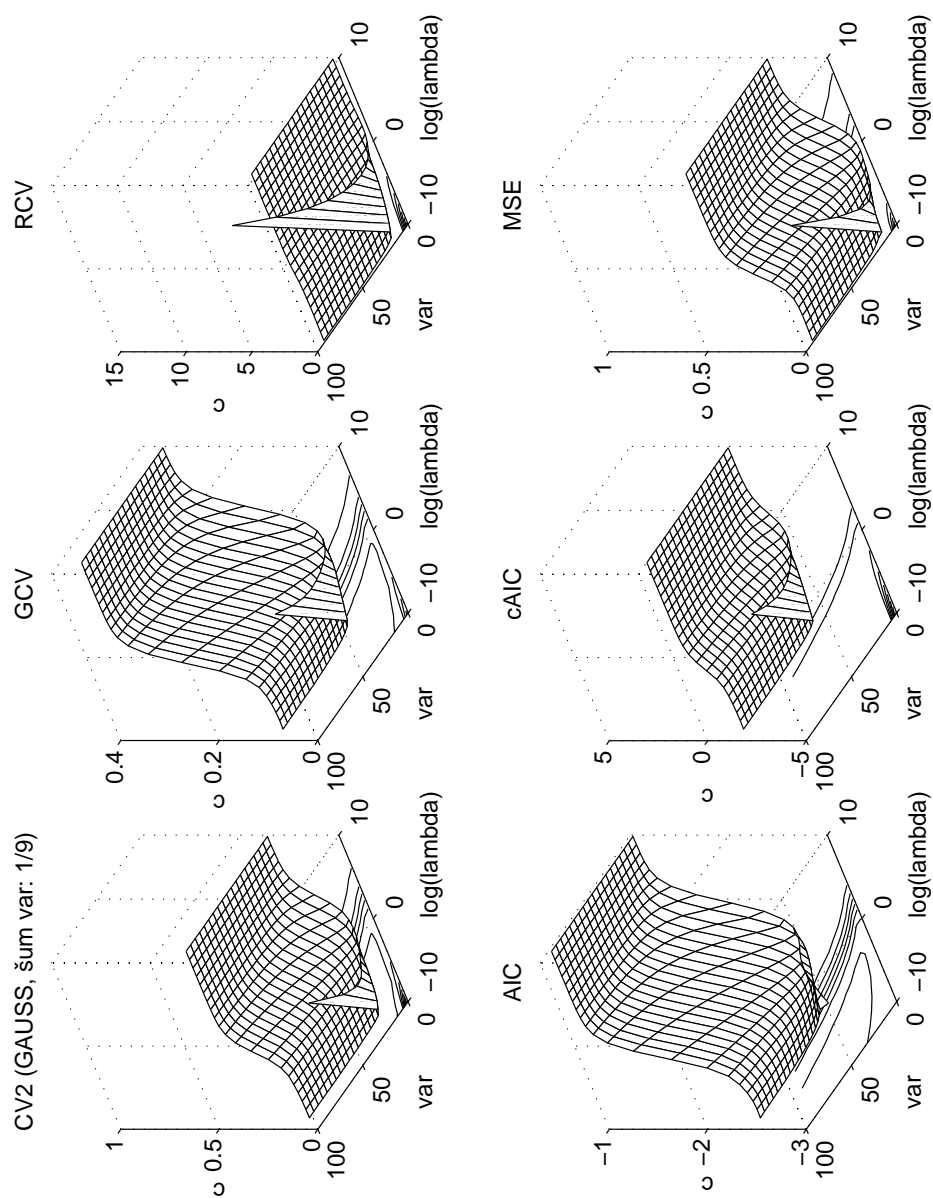
Obrázek 5.21: př.4: Průběh kritérií pro TPS v závislosti na λ – vstupem jsou hodnoty a body z obrázku 5.20, kritéria měří vhodnost vyhlazovacího parametru λ , menší hodnota je lepší.



Obrázek 5.22: př.4: Průběh kritérií pro Gaussovy RBF v závislosti na λ – vstupem jsou hodnoty a body z obrázku 5.20), kritéria měří vhodnost vyhlazovacího parametru λ , menší hodnota je lepší.



Obrázek 5.23: př.4: Průběh kritérií pro Gaussovy RBF v závislosti na σ^2 – vstupem jsou hodnoty a body z obrázku 5.20, kritéria měří vhodnost parametru σ^2 , menší hodnota je lepší.



Obrázek 5.24: př.4: Průběh kritérií pro Gaussovy RBF v závislosti na σ^2 a λ – vstupem jsou hodnoty a body z obrázku 5.20, kritéria měří vhodnost parametrů σ^2 a λ , menší hodnota je lepší.

5.4 Minimalizace

Sekce minimalizace ukazuje jaké výsledky dosahuje navrhovaná metoda a zároveň porovnává, které kritérium pro volbu parametrů je pro tuto metodu nejvhodnější.

Sekce obsahuje dva příklady, zadání příkladů je podobné předchozí sekci. První příklad pracuje s 10 zadanými body zatímco druhý se 100 body, zkoumané rozptyly šumu jsou 1 , $\frac{1}{9}$, $\frac{1}{900}$. Opět máme funkci $\sin(x) + \sin(y)$ na oblasti $[0, 3] \times [0, 4]$, body a hodnoty šumu jsou stejné jako v předchozí sekci, pouze hodnoty šumu se vynásobí příslušnou konstantou, aby se změnil rozptyl.

Průběh výpočtu je podobný jako v předchozí sekci, pouze se nekreslí grafy, ale hledají se hodnoty parametrů, pro které je dané kritérium minimální. Pro TPS se minimalizace provádí pomocí standardní funkce `FMINBND` z programu MATLAB, optimální λ se hledá na intervalu (e^{-10}, e^{10}) . U Gaussových RBF se optimální λ a σ^2 hledá pomocí metody navržené v kapitole 4.

Struktura výsledků je následující: každý příklad obsahuje dvě tabulky, z nichž první popisuje výsledky minimalizace pro TPS a druhá pro Gaussovy RBF. Každá tabulka obsahuje výsledky minimalizace pro všechny zkoumaná kritéria a různé hodnoty šumu.

Minimalizace potvrdila domněnky o vhodnosti kritérií uvedené v předchozí sekci. Jak je vidět z následujících tabulek, kritéria AIC a cAIC fungují špatně pro malý počet případů, RCV mívá problémy s existencí minima a volí příliš velké vyhlazení, ani CV se nepodařilo překonat výsledky GCV. Jako nejlepší kritérium se tedy jeví GCV, neboť funguje dobře i pro malý počet případů, má nejméně problémů s existencí minima a produkuje stabilně dobré a většinou nejlepší výsledky, měřeno pomocí střední kvadratické odchylky (MSE).

Z výsledků by se také dalo soudit, že při použití GCV produkují Gaussovy RBF lepší výsledky než TPS, opět měřeno pomocí střední kvadratické odchylky (MSE).

Báze	Počet bodů	Rozptyl šumu	Kritérium	$\log(\lambda)$	MSE
TPS	10	$\frac{1}{1}$	CV	-1.621	0.489
TPS	10	$\frac{1}{1}$	CV2	-1.621	0.489
TPS	10	$\frac{1}{1}$	GCV	0.533	0.288
TPS	10	$\frac{1}{1}$	RCV	10.000	0.493
TPS	10	$\frac{1}{1}$	AIC	-10.000	0.773
TPS	10	$\frac{1}{1}$	cAIC	-1.256	0.442
TPS	10	$\frac{1}{9}$	CV	-10.000	0.077
TPS	10	$\frac{1}{9}$	CV2	-10.000	0.077
TPS	10	$\frac{1}{9}$	GCV	-0.903	0.053
TPS	10	$\frac{1}{9}$	RCV	10.000	0.336
TPS	10	$\frac{1}{9}$	AIC	-10.000	0.077
TPS	10	$\frac{1}{9}$	cAIC	-1.256	0.054
TPS	10	$\frac{1}{900}$	CV	-10.000	0.025
TPS	10	$\frac{1}{900}$	CV2	-10.000	0.025
TPS	10	$\frac{1}{900}$	GCV	-10.000	0.025
TPS	10	$\frac{1}{900}$	RCV	-10.000	0.025
TPS	10	$\frac{1}{900}$	AIC	-10.000	0.025
TPS	10	$\frac{1}{900}$	cAIC	-1.256	0.033

Tabulka 5.2: Hledání optimálního λ u TPS – př. 1 s 10 zadanými body

Báze	Počet bodů	Rozptyl šumu	Kritérium	$\log(\lambda)$	σ^2	MSE
Gaussovy RBF	10	$\frac{1}{1}$	CV	-9.127	130.850	0.440
Gaussovy RBF	10	$\frac{1}{1}$	CV2	-9.127	130.850	0.440
Gaussovy RBF	10	$\frac{1}{1}$	GCV	-7.581	99.530	0.356
Gaussovy RBF	10	$\frac{1}{1}$	RCV	2.758	27.370	0.491
Gaussovy RBF	10	$\frac{1}{1}$	AIC	-10.000	0.040	0.475
Gaussovy RBF	10	$\frac{1}{1}$	cAIC	-2.994	2.040	0.677
<hr/>						
Gaussovy RBF	10	$\frac{1}{9}$	CV	-7.934	55.260	0.052
Gaussovy RBF	10	$\frac{1}{9}$	CV2	-7.934	55.260	0.052
Gaussovy RBF	10	$\frac{1}{9}$	GCV	-5.991	34.520	0.040
Gaussovy RBF	10	$\frac{1}{9}$	RCV	-8.426	122.940	0.044
Gaussovy RBF	10	$\frac{1}{9}$	AIC	-10.000	0.130	0.274
Gaussovy RBF	10	$\frac{1}{9}$	cAIC	-2.994	2.040	0.085
<hr/>						
Gaussovy RBF	10	$\frac{1}{900}$	CV	-7.566	22.050	0.004
Gaussovy RBF	10	$\frac{1}{900}$	CV2	-7.566	22.050	0.004
Gaussovy RBF	10	$\frac{1}{900}$	GCV	-4.487	6.930	0.014
Gaussovy RBF	10	$\frac{1}{900}$	RCV	-6.965	23.150	0.005
Gaussovy RBF	10	$\frac{1}{900}$	AIC	-10.000	0.330	0.198
Gaussovy RBF	10	$\frac{1}{900}$	cAIC	-2.994	2.040	0.060

Tabulka 5.3: Hledání optimálního λ a σ^2 u Gaussových RBF – př. 1 s 10 zadanými body

Báze	Počet bodů	Rozptyl šumu	Kritérium	$\log(\lambda)$	MSE
TPS	100	$\frac{1}{1}$	CV	1.894	0.064
TPS	100	$\frac{1}{1}$	CV2	1.894	0.064
TPS	100	$\frac{1}{1}$	GCV	2.103	0.064
TPS	100	$\frac{1}{1}$	RCV	5.329	0.237
TPS	100	$\frac{1}{1}$	AIC	-10.000	0.719
TPS	100	$\frac{1}{1}$	cAIC	2.424	0.068
<hr/>					
TPS	100	$\frac{1}{9}$	CV	0.676	0.013
TPS	100	$\frac{1}{9}$	CV2	0.676	0.013
TPS	100	$\frac{1}{9}$	GCV	0.845	0.013
TPS	100	$\frac{1}{9}$	RCV	2.859	0.046
TPS	100	$\frac{1}{9}$	AIC	-10.000	0.080
TPS	100	$\frac{1}{9}$	cAIC	1.198	0.014
<hr/>					
TPS	100	$\frac{1}{900}$	CV	-4.402	0.001
TPS	100	$\frac{1}{900}$	CV2	-4.402	0.001
TPS	100	$\frac{1}{900}$	GCV	-2.018	0.000
TPS	100	$\frac{1}{900}$	RCV	0.163	0.002
TPS	100	$\frac{1}{900}$	AIC	-10.000	0.001
TPS	100	$\frac{1}{900}$	cAIC	-0.941	0.001

Tabulka 5.4: Hledání optimálního λ u TPS – př. 2 se 100 zadanými body

Báze	Počet bodů	Rozptyl šumu	Kritérium	$\log(\lambda)$	σ^2	MSE
Gaussovy RBF	100	$\frac{1}{1}$	CV	-3.000	14.530	0.056
Gaussovy RBF	100	$\frac{1}{1}$	CV2	-3.000	14.530	0.056
Gaussovy RBF	100	$\frac{1}{1}$	GCV	-3.000	17.410	0.050
Gaussovy RBF	100	$\frac{1}{1}$	RCV	-1.014	31.120	0.120
Gaussovy RBF	100	$\frac{1}{1}$	AIC	-3.000	16.680	0.051
Gaussovy RBF	100	$\frac{1}{1}$	cAIC	-3.000	19.760	0.046
Gaussovy RBF	100	$\frac{1}{9}$	CV	-3.000	9.740	0.010
Gaussovy RBF	100	$\frac{1}{9}$	CV2	-3.000	9.740	0.010
Gaussovy RBF	100	$\frac{1}{9}$	GCV	-3.000	9.470	0.011
Gaussovy RBF	100	$\frac{1}{9}$	RCV	-3.000	24.210	0.011
Gaussovy RBF	100	$\frac{1}{9}$	AIC	-4.474	20.560	0.009
Gaussovy RBF	100	$\frac{1}{9}$	cAIC	-4.356	21.630	0.008
Gaussovy RBF	100	$\frac{1}{900}$	CV	-3.000	3.500	0.001
Gaussovy RBF	100	$\frac{1}{900}$	CV2	-3.000	3.500	0.001
Gaussovy RBF	100	$\frac{1}{900}$	GCV	-3.009	1.990	0.001
Gaussovy RBF	100	$\frac{1}{900}$	RCV	-3.000	4.540	0.001
Gaussovy RBF	100	$\frac{1}{900}$	AIC	-7.878	12.980	0.000
Gaussovy RBF	100	$\frac{1}{900}$	cAIC	-7.564	13.010	0.000

Tabulka 5.5: Hledání optimálního λ a σ^2 u Gaussových RBF – př. 2 se 100 zadanými body

5.5 Mapování

Tato sekce obsahuje výsledky navrhované metody pro konstrukci mapovací funkce pomocí Gaussových RBF tak, jak byla popsána v kapitole 4.

Vstupem pro mapování byl nějakým způsobem deformovaný obrázek a souřadnice korespondujících řídicích bodů, řídicí body byly vybírány ručně. Pro srovnání jsou ukázány výsledky mapování pomocí TPS pro stejný vstupní obrázek a stejné řídicí body. Zdeformované obrázky i zadané řídicí body jsou k dispozici na přiloženém CD.

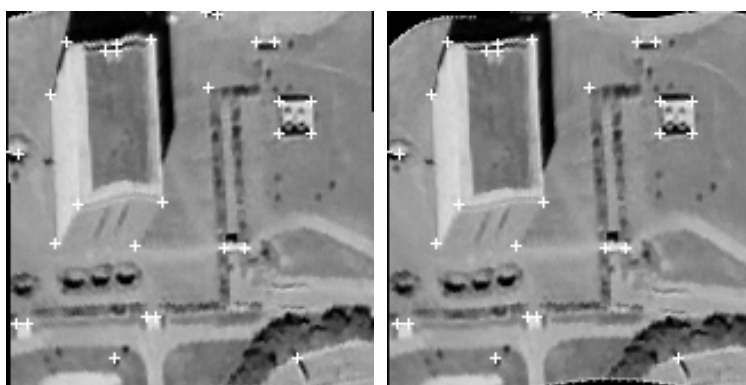
5.5.1 příklad 1

V následujícím příkladu byl zdrojový obrázek zdeformován v programu GIMP filtrem CurveBend, deformace horního i dolního okraje se zadává pomocí polynomiální křivky.

V tomto příkladě vychází lépe mapování pomocí TPS, podle MSE spočítané na obrazových funkcích (narovnaného obrázku proti originálu) jsou výsledky následující: MSE pro TPS = 0.008, MSE pro Gaussovy RBF = 0.021. Využívají se zde totiž globální vlastnosti TPS, kdy pomocí bodů zadaných uprostřed obrázku dokáže TPS pěkně narovnat rohy obrázku. Toho Gaussovy RBF se svým lokálním charakterem schopné nejsou.



Obrázek 5.25: Zadání mapování – vlevo je vstupní obrázek zdeformovaný v programu GIMP, vpravo vidíme obrázek původní, řídicí body jsou označeny křížky.



Obrázek 5.26: Výsledky mapování – vlevo při použití TPS, vprav jsou výsledky Gaussových RBF, řídicí body jsou označeny křížky.

5.5.2 příklad 2

Následující obrázek byl deformován pomocí „Mesh warp“ deformace v programu Corel PHOTO-PAINT. Jedná se o lokální deformaci, při které byla horní část levé věže stlačena ze stran a stlačena směrem dolů. Řídící body byly umístěny i do rohů obrázku, aby se zvýraznil lokální charakter deformace.



Obrázek 5.27: Zadání mapování – vlevo je vstupní obrázek zdeformovaný pomocí „Mesh warp“, vpravo vidíme obrázek původní, řídicí body jsou označeny křížky.



Obrázek 5.28: Výsledky mapování – vlevo při použití TPS, vpravo jsou výsledky Gaussových RBF, řídicí body jsou označeny křížky.

Podle MSE (počítané jako v prvním příkladu) jsou Gaussovy RBF v tomto případě lepší než TPS; přesné hodnoty jsou: MSE pro TPS = 0.007, MSE pro Gaussovy RBF = 0.003. U TPS vidíme zbytečné prohnutí středu obrázku, také při narovnání věže působí Gaussovy RBF napohled lépe.

5.5.3 zjištění

Celkově, při použití Gaussových RBF se obrázek jeví měkčí a snadněji deformovatelný, což plyne z lokalizovanosti Gaussových RBF. Tyto funkce umožňují deformovat část obrázku a jinou část nechat nezměněnou. Při menším rozptylu se však může stát, že když jedna část obrázku zůstane na místě a druhá je posunutá, oblast mezi se výrazně protáhne. Kvůli svému lokálnímu charakteru jsou Gaussovy RBF také citlivější na umístění řídicích bodů.

Naopak, při použití TPS se obrázek jeví velmi tuhý a těžko deformovatelný, což plyne z minimalizace funkcionálu vyjadřující energii zohýbaní. TPS se jeví stabilnější, méně náchylné k výkyvům nebo posunům bodů. Někdy je globální povaha TPS a deformace vzdálených oblastí výhodou (viz obr. 5.26 s pěkným narovnáním rohů) jindy zase nevýhodou (viz obr. 5.28 se zbytečným prohnutím střední části nebo obr. 1.2 s nepřirozeným protáhnutím obličeje v motivačním příkladu).

Výše uvedené vlastnosti plynou z vlastností báze, parametry mají pouze kvantitativní vliv.

Kapitola 6

Závěr

Cílem práce bylo prozkoumat možnosti registrace pomocí jiných radiálních básových funkcí (RBF) než jsou Thin Plate Spline (TPS), konkrétně pomocí Gaussových funkcí. S tím souvisí nalezení vhodné metody pro automatickou volbu vyhlazovacího parametru.

Na začátku byla provedena rešerše existujících metod registrace pomocí RBF s přihlédnutím k jiným básovým funkcím než jsou používané TPS. Dále byla nastudována kritéria, která se ve statistice používají pro volbu vyhlazovacího parametru (například při nelineární regresi). Byla navržena vlastní metoda pro automatickou volbu parametrů, jedná se o více kriteriální optimalizaci, při které se hledá optimální hodnota všech volných parametrů v modelu. Tyto „optimální“ hodnoty se odhadují pomocí vybraného kritéria. V práci byla porovnána následující kritéria: Generalized Cross Validation (GCV), Cross Validation (CV), Robustified Cross Validation (RCV), Akaike Information Criterion (AIC) a Corrected Akaike Information Criterion (cAIC). Jako nejvhodnější kritérium pro Gaussovy funkce bylo nakonec zvoleno kritérium GCV, neboť funguje dobře i pro malý počet vstupů, má nejméně problému s existencí minima a produkuje nejlepší výsledky (měřeno pomocí střední kvadratické odchylky (MSE)).

Zvolení GCV kritéria pro určování parametrů u Gaussových RBF je v souladu s dřívějšími pokusy s kritérii pro TPS, které nevedly k nalezení lepšího kritéria než obvykle používaného GCV, viz Berman [3]. Novým přístupem v navrhované metodě je použití GCV pro volbu všech parametrů modelu, ne pouze pro vyhlazovací parametr. Na testovaném obrázku s lokální deformací se navržené metodě s Gaussovými RBF dokonce podařilo překonat TPS.

Navrhovaná metoda s Gaussovými RBF se ukázala vhodná pro registraci obrazu, kde dochází k lokální deformaci snímku; příkladem by mohlo být lékařství. Dále by se metoda dala využít pro warping nebo pro aproximaci dvourozměrné funkce.

Námětem na vylepšení uvedené metody by mohlo být zavedení různých

rozptylů pro každý prvek báze, tím by se ale velmi zvýšila výpočetní náročnost pro určování těchto parametrů, pokud by nebyla nevyvinuta optimalizovaná metoda.

Literatura

- [1] J. Anděl: *Statistické metody*, Matfyzpress, 2. vydání Praha, 1998
- [2] N. Arad, N. Dyn, D. Reifeld, Y. Yeshurun: „Image Warping by Radial Basis Functions: Application to Facial Expressions“, *CVCIP: Graphical Models and Image Processing*, vol. 56, pp. 161–172, 1994
- [3] M. Berman: „Automated Smoothing of Image and Other Regularly Spaced Data“, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 460–468, 1994
- [4] F. L. Bookstein: „Principal warps: Thin-plate splines and the decomposition of deformations“, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, pp. 567–585, 1989
- [5] J. C. Carr, W. R. Fright, R. K. Beatson: „Surface Interpolation with Radial Basis Functions for Medical Imaging“, *IEEE Transactions on Medical Imaging*, vol. 16, pp. 96–107, 1997
- [6] M. Ehlers, D. N. Fogel: „High-precision geometric correction of airborne remote sensing revisited: The multiquadric interpolation“, *Proc. SPIE: Image and Signal Processing for Remote Sensing*, vol. 2315, pp. 814–824, 1994
- [7] R. L. Eubank: „Spline Regression“, pp. 1–18, in: *Smoothing and Regression Approaches, Computation and Application*, Wiley, New York, 2000
- [8] H. Fang, J. H. Nure: „Smoothing Random Noise from Human Head Scan Data“, *IEEE Transactions on Medical Imaging*, vol. 15, pp. 102–111, 1996
- [9] D. N. Fogel: „Image rectification with radial basis functions: Application to RS/GIS data integration“, *Proc. 3rd Int'l. Conf. on Integrating GIS and Environmental Modelling*, CD-ROM, 19 pp., Santa Fe, New Mexico, 1996.
- [10] B. E. L. Grimson: „A Computational Theory of Visual Surface Interpolation“, *Phil. Trans. R. Lond. B*, vol. 298, pp. 195–427, 1982

-
- [11] C. Gu: „Multivariate Spline Regression“, pp. 329–355, in: *Smoothing and Regression Approaches, Computation and Application*, Wiley, New York, 2000
- [12] R. L. Harder and R. N. Desmarais: „Interpolation using surface splines“, *Journal of Aircraft*, vol. 9, pp. 189–191, 1972
- [13] C. M. Hurvich, J. S. Simonoff, C. Tsai: „Smoothing Parametr Selection in Nonparametric Regression Using an Improved Akaike Information Criterion“, *Journal of the Royal Statistical Society B*, vol. 60, pp. 271–293, 1998
<http://www.stern.nyu.edu/~jsimonof/aicc.ps/aicc.ps>
- [14] M. Jansen, M. Malfait, A. Bulthell: „Generalized Cross Validation for wavelet thresholding“, *Signal Processing*, vol. 56, pp. 33–44, 1997
<http://www.cs.kuleuven.ac.be/~maarten/publications/>
- [15] A. van der Linde: „Variance Estimation and Smothing-Parametr Selection for Spline Regression“, pp. 19–41, in: *Smoothing and Regression Approaches, Computation and Application*, Wiley, New York, 2000
- [16] N. Nguyen, P. Milanfar, G. Golub: „Efficient Generalized Cross-Validation with Applications to Parametric Image Restotation and Resolution Enhancement“, *IEEE Transactions on Image Processing*, vol. 10, pp. 1299–1308, 2001
- [17] K. Rohr, H. S. Stiehl, T. M. Buzug, J. Weese, M. H. Kuhn: „Landmark-based elastic registration using approximating thin-plate splines“, *IEEE Transaction on medical imaging*, vol. 20, pp. 526–534, 2001
- [18] K. Rohr, H. S. Stiehl, R. Sprengel, W. Beil, T. M. Buzug, J. Weese, M. H. Kuhn: „Nonrigid registration of medical images based on anatomical point landmarks and approximating thin-plate splines“, *Proc. Aachen Workshop: Bildverarbeitung für die Medizin*, pp. 41–46, Aachen, Germany, 1996.
- [19] K. Rohr: *Landmark-Based Image Analysis: Using Geometric and Intensity Models*, vol. 21 of *Computational Imaging and Vision Series*, Kluwer Academic Publishers, Dordrecht Boston London, 2001.
- [20] T. Robinson, R. Moyeed: „Making Robust the Cross-validatory Choice of Smoothing Parametr in Spline Smoothing Regression“, *Commun. Statis. Theory Meth*, vol. 18, pp. 523–539, 1989
- [21] R. Wiemker, K. Rohr, L. Binder, R. Sprengel, H. Siegfried Stiehl: „Application of Elastic Registration to Imagery from Airborne Scanners“, *International Archives for Photogrammetry and Remote Sensing*, vol. XXXI, part B, commission IV, 1996

- [22] G. Wahba: *Spline Models for Observational Data*, Philadelphia, SIAM, 1990
- [23] R. Wiemker, K. Rohr, L. Binder, R. Sprengel, H. S. Stiehl: „Application of elastic registration to imagery from airborne scanners“, *Int'l. Archives for Photogrammetry and Remote Sensing*, vol. XXXI-B4, pp. 949–954, 1996
- [24] S. N. Wood: „Modelling and smoothing parameter estimation with multiple quadratic penalties“, *J.R. Statist. Soc. B*, vol. 62, pp. 413–428, 2000
- [25] S. N. Wood, N. H. Augustin: „GAMs with integrated model selection using penalized regression splines and applications to environmental modelling“, *Ecological Modelling*, vol. 157, pp. 157–177, 2002
<http://dolphin.mcs.st-and.ac.uk/simon/papers/wagam.pdf>
- [26] S. N. Wood: Thin-plate regression splines, *J.R. Statist. Soc.*, in press 2003