

Univerzita Hradec Králové

Pedagogická fakulta

Katedra Fyziky

Spektrální vlastnosti rodin planetek podle přehledky Sloan Digital Sky Survey

Bakalářská práce

Autor: Lenka Trojanová

Studijní program: M 7530 Fyzikálně—technická měření a výpočetní technika

Vedoucí práce: Mgr. Miroslav Brož, Ph.D.

Hradec Králové:

2010

Poděkování

Za neuvěřitelnou ochotu a pomoc při psaní práce děkuji mému vedoucímu Miroslavu Brožovi.

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracovala (pod vedením vedoucího bakalářské práce) samostatně a uvedla jsem všechny použité prameny a literaturu.

V Hradci Králové dne 15. 5. 2010

Anotace

TROJANOVÁ, Lenka. Spektrální vlastnosti rodin planetek podle přehlídky Sloan Digital Sky Survey. Hradec Králové: Pedagogická fakulta Univerzity Hradec Králové, 2010. 54 s. Bakalářská práce.

V práci se zabýváme aplikací metody hlavních komponent na fotometrická data planetek. Vstupem práce jsou obecná matematická metoda hlavních komponent a katalog fotometrických dat pro 471 569 planetek programu Sloan Digital Sky Survey. Popsali jsme postup výpočtu hlavních komponent a vytvořili algoritmy, které celý výpočetní proces zprostředkovávají. Na základě našich výsledků doporučujeme pro studium planetek používat tři hlavní komponenty, pro něž jsme odvodili určitou fyzikální interpretaci. Také jsme zobrazili hlavní komponenty v prostoru dráhových elementů planetek (především velké poloosy, excentricitě a sklonu), a ověřili tím souvislost dráhových a fotometrických dat.

Klíčová slova: metoda hlavních komponent, fotometrická data, planetky

Annotation

TROJANOVÁ, Lenka. Spectral properties of asteroid families based on the Sloan Digital Sky Survey. Hradec Králové: Faculty of Education, University of Hradec Králové, 2010. 54 pp. Bachelor Degree Thesis.

In this work we apply a principal component analysis (PCA) to photometric data of asteroids. The basic input for this work is a general mathematical PCA method and a catalogue of photometric data for 471,569 asteroids, which were observed by the Sloan Digital Sky Survey. We describe the calculation of principal components and we prepare algorithms, which enable practical computations. According to our results, we recommend to use three principal components, for which we inferred a specific physical interpretation, in studies of asteroids. We finally display principal components in the space of orbital elements of asteroids (most importantly semimajor axis, eccentricity and inclination) and verified a close relation between orbital and photometric data.

Keywords: principal component method, photometric data, asteroids

Obsah

1	Úvod	2
2	Statistické pojmy — Základy algebry	3
2.1	Střední hodnota	3
2.2	Medián	3
2.3	Směrodatná odchylka	4
2.4	Rozptyl	4
2.5	Kovariance	4
2.6	Kovarianční matice	5
2.7	Vlastní vektor a vlastní číslo	5
3	Metoda hlavních komponent	7
3.1	Stručný popis metody	7
3.2	Podrobnější výklad metody PCA	9
3.2.1	Transformace souřadnic	9
3.2.2	Vlastní čísla a vlastní vektory	9
3.2.3	Operace s maticemi	11
3.2.4	Matematická formulace metody PCA	12
3.3	Dokumentace užitých programů	13
4	Užité fyzikální principy	17
4.1	Radiometrie	17
4.1.1	Fotometrické systémy	19
4.1.2	Hvězdná velikost (magnituda)	20
4.2	Barevný index	22
5	Aplikace metody PCA na fotometrická data planetek	23
5.1	Přehledka Sloan Digital Sky Survey	23
5.2	Katalog SDSS MOC	24
5.2.1	Uspořádání katalogu MOC	26
5.3	Výpočet hlavních komponent	31
6	Fyzikální interpretace výsledků	33
6.1	Hlavní komponenty a refleční spektra	33
6.2	Hlavní komponenty a dráhové elementy	38
7	Závěr	41
A	Kódy programů	42
B	Použití algoritmu z Numerical Recipes	52
	Literatura	54
	Rejstřík	56

1 Úvod

Současná doba přinesla automatizaci nejrůznějších procesů a podstatně změnila i způsob získávání astronomických dat. Příkladem může být projekt Sloan Digital Sky Survey, který poměrně rychle získal terabyty digitalizovaných dat. V této práci jsme si vzali za úkol použít obecnou metodu hlavních komponent (PCA) na fotometrická data, získaná právě z katalogu „sloanské“ přehlídky, a diskutovat jejich interpretaci. Při optimistickém pohledu do budoucnosti můžeme očekávat opakování nebo rozšiřování stávajících přehlídek oblohy či nástup nových, ještě rozsáhlejších projektů. Ve světle těchto skutečností je vhodné automatizovat zpracování dat tak, aby při budoucích aktualizacích katalogu bylo možné rychle získat nové výsledky.

Nejprve v části 2 „Statistické pojmy — Základy algebry“ připomeneme základní používání statistiky a popíšeme matematiku nezbytnou k dalšímu kroku, tedy představení metody hlavních komponent. V kapitole 3 „Metoda hlavních komponent“ se postupně dopracujeme až k úplnému popisu jejích vlastností. Protože pracujeme s fotometrickými daty, uvádíme v kapitole 4 „Užité fyzikální principy“ definice několika běžně užívaných fyzikálních veličin a pojmů. V kapitole 5 „Aplikace metody PCA na fotometrická data planetek“ nalezneme popis sloanského katalogu a především podrobný výpočetní postup redukce původních čtyřrozměrných dat na tři hlavní komponenty. V následující kapitole 6 nazvané „Fyzikální interpretace výsledků“, zkonstruujeme určité grafy a na základě jejich porovnání budeme diskutovat fyzikální význam tří hlavních komponent.

2 Statistické pojmy — Základy algebry

V této práci se jedná především o analýze velkého souboru dat. Obecně lze data charakterizovat pomocí mnoha statistických pojmů jako střední hodnota, směrodatná odchylka, rozptyl, kovariance, kovarianční matice, vlastní vektor, vlastní čísla a další. Pro pochopení toho, co nám jednotlivé výsledky těchto operací o množině dat říkají, je nezbytné, abychom mohli rozhodnout, která charakteristika je vzhledem k způsobu získání našich dat a vzhledem k jejich interpretaci nejpříhodnější. Uvědomíme si také nedostatky rozličných metod, a shrneme poznatky k porozumění metodě hlavních komponent. Při zpracování kapitoly 2 jsem použila práce [3], [10], [18].

2.1 Střední hodnota

Množina dat má mnoho vlastností, které můžeme vypočítat. Například střední hodnotu (aritmetický průměr) měřených hodnot X_i :

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}. \quad (1)$$

Vztah (1) platí pokud jsou hodnoty v množině rozloženy zcela náhodně v okolí předpokládané správné hodnoty X . Pro případ, že tomu tak není, existuje obecnější definice střední hodnoty například *vážený průměr*:

$$\bar{X} = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n w_i}, \quad (2)$$

kde w_i je *statistická váha* (důležitost) příslušející hodnotě X_i .

Střední hodnota dobře charakterizuje velikost hodnot množiny dat. Pokud si představíme jednotlivá čísla množiny X jako hmotné body umístěné na ose, přičemž jejich hodnoty udávají vzdálenost od počátku, snadno nahlédneme, že aritmetický průměr je shodný s těžištěm (hmotným středem) takové soustavy. Přes své výhody průměr neposkytuje dostatečný popis vlastností množiny dat. Uvažme například tyto dvě množiny:

$$[0 \ 8 \ 12 \ 20], \quad [8 \ 9 \ 11 \ 12], \quad (3)$$

mají totožný průměr, ale očividně jsou velmi rozdílné.

2.2 Medián

Statistika zařazuje medián obdobně jako střední hodnotu, mezi míry centrální tendence. Pokud seřadíme hodnoty dat posuzované množiny podle velikosti, bude mediánem ta hodnota, pro niž nejméně 50 % všech hodnot množiny je menších nebo ji rovných a nejméně 50 % všech hodnot množiny je větších nebo rovných. Hlavní výhodou mediánu jako statistického ukazatele polohy dat je, že na rozdíl od průměru není citlivý na ovlivňování způsobené extrémními (většinou nežádoucími) hodnotami.

2.3 Směrodatná odchylka

Tím, co od sebe odlišuje zmiňované dvě množiny (3), je rozsah hodnot, jakých nabývají jednotlivá data. Takovou vlastnost vyjadřuje *směrodatná odchylka*. Směrodatná odchylka souboru dat je mírou rozptýlení dat kolem předpokládaného středu. Obvykle ji značíme s a platí pro ni¹:

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)}}. \quad (4)$$

Ve jmenovateli je výraz $(n-1)$ na místo očekávaného n . To proto, že pokud sada našich dat je nějakým vzorkem, čili je jistou podmnožinou ze skutečného množství dat (například výběrový vzorek), pak je třeba v jmenovateli použít $n-1$, abychom získali výsledek bližší směrodatné odchylce příslušející celé populaci, a nikoli jen našemu vzorku. Pokud však pracujeme s celou populací (nikoliv jen s částí) měli bychom v jmenovateli užívat n . Podrobněji je vše popsáno v [3].

2.4 Rozptyl

Kvadrát směrodatné odchylky nazýváme rozptylem². Je další charakteristikou vyjadřující míru rozprostření dat v množině. Platí pro něj :

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)}, \quad (5)$$

2.5 Kovariance

Výše zmíněné charakteristiky jsou jednorozměrné. Jako je například jednorozměrná množina dat obsahující výšky jednotlivých lidí v místnosti. Existují i vícerozměrné množiny dat. Například soubor dat obsahující výšku jednotlivých lidí v místnosti a velikost jejich platu. Cílem statistické analýzy vícerozměrných dat, je nalezení nějakého vztahu mezi rozměry (tj. veličinami, které k popisu používáme). Samozřejmě, pokud takový vztah vůbec existuje. Přidržíme-li se předchozího příkladu, šlo by o určení závislosti mezi výškou člověka a jeho platem.

Směrodatnou odchylku a varianci sice můžeme stanovit i pro vícerozměrná data, ale pouze pro každou množinu dat zvlášť, nezávisle.

Kovariance je takovou charakteristikou množin dvourozměrných dat, jež popisuje, jak se liší průměr každé z množin v závislosti na sobě. Kovarianci lze stanovit pouze pro dvourozměrná, říkáme dvoudimenzionální data. Počítáme-li kovarianci mezi jednou a tou samou dimenzí, obdržíme rozptyl. Vzorec pro kovarianci odvodíme právě prostřednictvím rozptylu. Jistě lze napsat:

$$s^2 = \text{var}(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})}{(n-1)}. \quad (6)$$

¹V případě dokonalých dat (Gaussovo rozložení) je ideální mírou.

²V anglicky psaných učebnicích je standardně rozptyl značen $\text{var}(X)$ (od slova variance), v česky psaných učebnicích se spíše setkáme s označením s^2 .

Pokud v druhé závorce zavedeme Y místo X dostaneme vzorec pro kovarianci:

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}. \quad (7)$$

Použijme nyní pro konkrétnost náš příklad s výškou osob a platem. V prvním rozměru H jsou uloženy hodnoty výšky osob, v druhém rozměru Z je plat. Pokud bychom konkrétní data opravdu měli, lze je dosadit do vzorce pro kovarianci a získat určitý výsledek. Přesná hodnota výsledku není teď důležitá. Důležité je znaménko, tedy jestli je výsledek kladný nebo záporný. Pokud je znaménko kladné, dochází se zvětšujícími se hodnotami dat prvního rozměru k zvětšování hodnoty dat druhého rozměru. Takovou situaci nazýváme v ideálním případě přímá úměrnost. V případě záporného znaménka, hodnoty dat v jednom rozměru rostou a v druhém klesají. Je-li hodnota kovariance rovna nule, oba rozměry jsou na sobě nezávislé. „Přepych“ v podobě možnosti vidět data v grafu, si můžeme dovolit u jedno-, dvou- nebo třírozměrných datových množin, což v praxi zdaleka nestačí³.

2.6 Kovarianční matice

V předchozí kapitole jsme upozornili, že kovarianci lze vypočítat pouze pro dvourozměrnou množinu dat. Máme-li tři (nebo vícerozměrnou) množinu dat (x, y, z) můžeme vypočítat $\text{cov}(x, y)$, $\text{cov}(x, z)$ a $\text{cov}(y, z)$. Snadno nahlédneme, že pro n rozměrnou množinu dat můžeme vypočítat $\frac{n!}{(n-2)! \cdot 2}$ různých kovariancí.

Pokud určíme kovarianci pro všechny možné kombinace rozměrů n rozměrné množiny dat a hodnoty uspořádáme do matice podle následujícího pořadí:

$$C^{n \times n} \equiv C_{i,j} = \text{cov}(\text{Dim}_i, \text{Dim}_j), \quad (8)$$

kde $C^{n \times n}$ je matice s n řádky a n sloupci a Dim_i je i -tý rozměr, obdržíme *kovarianční matici* pro třírozměrnou množinu dat:

$$\begin{pmatrix} \text{cov}(x, x) & \text{cov}(x, y) & \text{cov}(x, z) \\ \text{cov}(y, x) & \text{cov}(y, y) & \text{cov}(y, z) \\ \text{cov}(z, x) & \text{cov}(z, y) & \text{cov}(z, z) \end{pmatrix}. \quad (9)$$

Na diagonále matice jsme získali rozptyly jednotlivých rozměrů a matice je podle diagonály symetrická.

2.7 Vlastní vektor a vlastní číslo

Na vlastní vektory lze nahlížet jako na zvláštní případy násobení dvou matic. Podívejme se na obr. 1. V prvním případě výsledná matice $\begin{pmatrix} 11 \\ 5 \end{pmatrix}$ není celočíselným násobkem původní matice $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ zatímco v druhém příkladu je výsledná matice $\begin{pmatrix} 12 \\ 8 \end{pmatrix}$ čtyřikrát větší než původní.

³Samozřejmě n -rozměrná data zobrazovat lze — jako „řezy“, barvami, a jinými postupy.

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

Obr. 1: Příklad vektoru, který není vlastní, a vektoru, který je vlastní.

Matice $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$ je v podstatě vektorem s počátkem v bodě $[0, 0]$ směřujícím do bodu $[3, 2]$. Čtvercovou matici z příkladu můžeme chápat jako transformační. Pokud ji násobíme vektorem, výsledkem je vektor vzniklý transformací (přeměnou) původního vektoru. Dodejme, vlastní vektory se definují právě při dané transformaci.

Uvažujme například transformaci, která původní obecný vektor umístěný na ploše, tedy o souřadnicích $[x, y]$, převede na vektor umístěný na přímce $y = x$. Vektory již ležící na přímce pak jistě zůstanou po transformaci nezměněné. Takové vektory (a jejich libovolné násobky) označujeme jako vlastní vektory příslušné transformační matice.

Vlastní vektory lze nalézt jen pro čtvercové matice. Zároveň ne každá čtvercová matice vlastní vektor má. Maximální počet vlastních vektorů matice $n \times n$ je n . Další vlastností vlastních vektorů je, že pokud je vynásobíme libovolným číslem, tedy je libovolně zvětšíme nebo zmenšíme, dostáváme ve výsledku stále stejný násobek původního vektoru.

Poznamenejme ještě, že všechny vlastní vektory matice, nezávisle na tom kolikarozměrné, jsou na sebe vzájemně kolmé, říkáme *ortogonální*. Právě tato vlastnost umožňuje popsat soubory dat pomocí vlastních vektorů, namísto obvyklého vyjadřování pomocí os souřadnic $x, y, z \dots$. Později, v části zabývající se popisem metody PCA, budeme této vlastnosti hojně využívat.

Velikost vektorů nerozhoduje o tom, jestli jsou nebo nejsou vlastní. V zájmu usnadnění práce je proto vhodné velikost vlastních vektorů standardizovat. Obvykle každý nalezený vlastní vektor dělíme jeho velikostí, a získáme tak vektor jednotkový. Například vektor:

$$\begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

má velikost:

$$\sqrt{(3^2 + 2^2)} = \sqrt{13}$$

a jednotkový vektor nabývá tvaru:

$$\begin{pmatrix} 3 \\ 2 \end{pmatrix} \div \sqrt{13} = \begin{pmatrix} 3/\sqrt{13} \\ 2/\sqrt{13} \end{pmatrix}$$

Vlastní vektory se pro malé matice (ne větší jak 3×3) hledají jednoduše. Vlastní vektory větších matic se hledají pomocí komplikovanějších iteračních metod, které tento úvod nepojme.

3 Metoda hlavních komponent

V anglickém jazyce je metoda hlavních komponent nazývána Principal Components Analysis. V dalším textu jen (PCA). Následující kapitola byla zpracována podle [1], [9], [10].

Zběžně lze říci, že PCA slouží k zjištění některých zákonitostí v daném souboru dat. Je to velmi užitečná metoda pro zpracovávání dat uspořádaných do vícerozměrného matematického prostoru. Ta obvykle nelze snadno graficky znázornit a podle grafu interpretovat.

Velkou výhodou metody PCA je, že jakmile najdeme významné vztahy v datech, můžeme snížit jejich dimenzi (množiny prostoru dat) a bez velké ztráty informace tak data redukovat. Takové vlastnosti se využívá například při kompresi obrázků.

Nejvýznamnější veličiny jsou metodou rozeznávány prostřednictvím rozptylu. Čím větší má určitá veličina rozptyl, tím většího rozsahu hodnot nabývá, a má pro nás větší důležitost. S klesajícím rozptylem klesá důležitost veličiny. Jde tedy o to, najít v množině původních dat nové směry (osy). A to takové směry, že jakmile je „povýšíme“ na osy veličinových souřadnic a spočítáme rozptyl dat v těchto nových souřadnicích, bude tento rozptyl maximální.

V kapitole 3.1 jsou popsány základní myšlenky PCA. Prohloubení a důkazy uvedených tvrzení budou podány následně.

3.1 Stručný popis metody

Na počátku máme množinu obecně n — rozměrných dat. V této kapitole pro názornost pracujeme s množinou třírozměrných dat. Řekněme, že veličiny jsou zdánlivé hvězdné velikosti ve třech vzájemně různých vlnových délkách, přičemž máme data pro m planetek. Začneme vypočtením kovariance pro každou dvojici veličin (hvězdných velikostí v určitých vlnových délkách) a sestrojením kovarianční matice (viz (9)):

$$\mathbf{C} = \begin{pmatrix} \text{cov}(x, x) & \text{cov}(x, y) & \cdots & \text{cov}(x, z) \\ \text{cov}(y, x) & \text{cov}(y, y) & \cdots & \text{cov}(y, z) \\ \vdots & & \ddots & \\ \text{cov}(z, x) & \text{cov}(z, y) & \cdots & \text{cov}(z, z) \end{pmatrix}.$$

Na diagonále jsme získali rozptyly jednotlivých veličin. Matice \mathbf{C} je symetrická, proto diagonalizovatelná, a má reálná vlastní čísla. Vlastní čísla λ a vlastní vektory \mathbf{v} matice \mathbf{C} určíme relativně jednoduchým výpočtem. Podívejme se nejprv na jejich význam.

Vlastní vektory kovarianční matice určují ty směry, ve kterých mají data největší rozptyl (nejvíce se mění jejich hodnoty vzhledem k předpokládanému mediánu). Vlastní vektory tedy určují směry os, vzhledem k nimž (jako významným směrům) budeme hledat vhodný výklad, např. zákonitosti mezi hodnotami hvězdných velikostí v daném souboru m planetek. Naším přáním je „otočit“ osy souřadnic původní množiny třírozměrných dat (na těchto původních osách máme vyneseny hvězdné velikosti planetek v různých vlnových délkách) tak, aby splývaly se směry vlastních vektorů. Vlastní čísla pak odpovídají rozptylům v nových souřadnicích. Aby se nám vše povedlo tak, jak

jsme popsali, je třeba vyjít z definice vlastního vektoru a splnit podmínku:

$$\mathbf{C} \vec{v} = \lambda \vec{v} , \quad (10)$$

neboli:

$$(\mathbf{C} - \lambda \mathbf{I}) \vec{v} = 0 , \quad (11)$$

kde \mathbf{I} označuje jednotkovou matici. Ekvivalentem předchozího zápisu je:

$$\det |\mathbf{C} - \lambda \mathbf{I}| = 0 , \quad (12)$$

rozepsáno:

$$\begin{aligned} \det |\mathbf{C} - \lambda \mathbf{I}| &= \begin{vmatrix} c_{11} - \lambda & c_{12} & c_{13} \\ c_{21} & c_{22} - \lambda & c_{23} \\ c_{31} - \lambda & c_{32} & c_{33} - \lambda \end{vmatrix} = \\ &= (c_{11} - \lambda)(c_{22} - \lambda)(c_{33} - \lambda) + c_{21}c_{32}c_{13} + c_{31}c_{12}c_{23} + c_{21}c_{32}c_{13} \\ &\quad - c_{13}(c_{22} - \lambda)c_{31} - c_{23}c_{32}(c_{11} - \lambda) - (c_{33} - \lambda)c_{12}c_{21} = \\ &= -\lambda^3 + (c_{11} + c_{22} + c_{33})\lambda^2 + \text{„mnoho } c\text{“} \cdot \lambda + \text{„zbylá } c\text{“} = 0 . \end{aligned} \quad (13)$$

Pro λ tedy získáváme polynom 3. stupně, nazýváme ho *charakteristickou rovnicí*. Jejím řešením získáme tři hodnoty (kořeny) $\lambda_1, \lambda_2, \lambda_3$. Pro každou z nich musíme ještě zjistit souřadnice příslušného vlastního vektoru, a to vyřešením soustavy tří lineárních rovnic:

$$\begin{aligned} (c_{11} - \lambda_i)v_1 + c_{12}v_2 + c_{13}v_3 &= 0 \\ c_{21}v_1 + (c_{22} - \lambda_i)v_2 + c_{23}v_3 &= 0 \\ c_{31}v_1 + c_{32}v_2 + (c_{33} - \lambda_i)v_3 &= 0 \end{aligned} \quad , \quad i = 1, 2, 3 .$$

Zbývá určit polohu původních dat v nových souřadnicích. Na osách nových souřadnic leží nové veličiny, nazýváme je hlavní komponenty:

$$PC_{1,j} = v_1(x_j - \bar{x}) + v_2(y_j - \bar{y}) + v_3(z_j - \bar{z}) , \quad (14)$$

kde m je počet planetek a $j = 1 \dots m$

Vlastní číslo příslušující PC_4 bude pravděpodobně malé. V takovém případě samotná hodnota PC_4 pozbývá na významu a nemá smysl ji počítat.

Pokud by původní množina dat byla vícerozměrná (obecně n rozměrná), příslušelo by každé planetce n veličin. Matice \mathbf{C} by měla rozměr $n \times n$, stupeň charakteristického polynomu by byl n a maximální počet hlavních komponent n . V takovém případě se vlastní čísla a vlastní vektory hledají pomocí numerických iteračních metod [15].

Výpočet je po matematické stránce ukončen, ale fyzikální význam hlavních komponent musíme ještě nalézt. Nejlepším způsobem je vynesení dat do grafu s ortogonálními souřadnicemi PC_1, PC_2 a podrobné porovnání původních dat, jež mají malé nebo velké nebo jinak „zvláštní“ hodnoty PC_1, PC_2 . O tom, jak se podařilo interpretovat hlavní komponenty pro planetky zaznamenané v rámci přehlídky Sloan Digital Sky Survey, informuje kapitola 5.3.

3.2 Podrobnější výklad metody PCA

V předchozí části (kap. 3.1) jsou mnohá tvrzení podána bez důkazu a bez zavedení matematických definic a vět, na kterých postup spočívá. Výhodou takového textu je jeho přehlednost a přímočarost. V této kapitole se pokusíme odpovědět na otázku, jak to, že právě ve směrech vlastních vektorů kovarianční matice dat nabývají data největších rozptylů. K tomu bude nutné připomenout několik matematických definic a vět z lineární algebry.

Shrnout celý použitý matematický aparát v této práci není možné.

3.2.1 Transformace souřadnic

Definice 1 *Nechť je dán vektorový prostor U s uspořádanou bází $\Upsilon = (\vec{u}_1, \vec{u}_2 \dots \vec{u}_n)$. Zvolme v prostoru U další uspořádanou bází $\mathbf{V} = (\vec{v}_1, \vec{v}_2 \dots \vec{v}_n)$. Nechť $P : U \rightarrow U$ je lineární zobrazení definované předpisem :*

$$P(u_i) = \vec{v}_i \quad \text{pro každý vektor } \vec{u}_i \text{ z báze } \Upsilon.$$

Vyjádříme-li vektory nové báze \mathbf{V} jako lineární kombinaci vektorů staré báze Υ , dostaneme rovnosti:

$$\begin{aligned}\vec{v}_1 &= p_{11}\vec{u}_1 + p_{21}\vec{u}_2 + \dots + p_{n1}\vec{u}_n, \\ \vec{v}_2 &= p_{12}\vec{u}_1 + p_{22}\vec{u}_2 + \dots + p_{n2}\vec{u}_n, \\ &\dots \\ \vec{v}_n &= p_{1n}\vec{u}_1 + p_{2n}\vec{u}_2 + \dots + p_{nn}\vec{u}_n.\end{aligned}$$

Matice \mathbf{P} lineárního zobrazení P v bázích \mathbf{V} a Υ :

$$\mathbf{P} = \begin{pmatrix} p_{11}, p_{21}, \dots, p_{n1} \\ p_{12}, p_{22}, \dots, p_{n2} \\ \dots \\ p_{1n}, p_{2n}, \dots, p_{nn} \end{pmatrix}$$

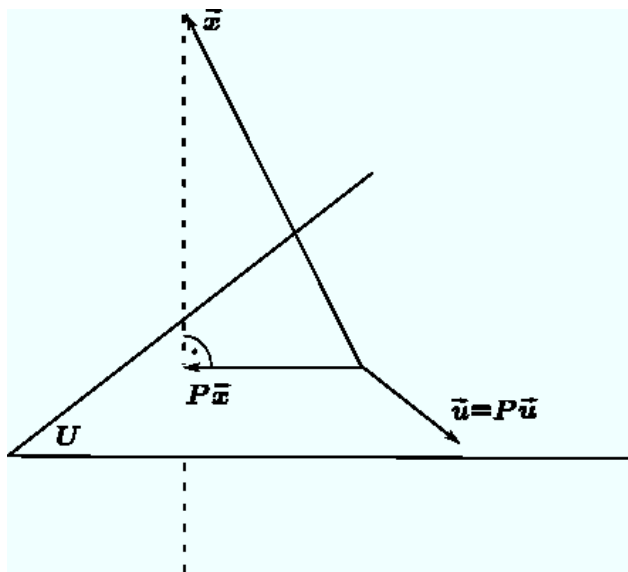
se nazývá *matice přechodu* od báze \mathbf{V} k bázi Υ . Nechť nyní \vec{u} je libovolný vektor z prostoru U . Označme $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$ resp. $\vec{\eta} = (\eta_1, \eta_2, \dots, \eta_n)$ vyjádření vektoru prostřednictvím souřadnic ξ_i vzhledem k bázi Υ resp. η vzhledem k bázi \mathbf{V} . Pak platí :

$$\vec{\xi} = \mathbf{P}\vec{\eta}. \tag{15}$$

Vztah ukazuje, jak lze pomocí souřadnic vektoru v jedné uspořádané bázi spočítat jeho souřadnice v jiné uspořádané bázi, když známe příslušnou matici přechodu. Podobnost předchozí rovnosti a definice vlastních čísel $\lambda\vec{v} = \mathbf{A}\vec{v}$ by neměla zůstat bez povšimnutí.

3.2.2 Vlastní čísla a vlastní vektory

Definice 2 *Mějme matici $\mathbf{A} \in \mathbf{C}^{n,n}$. Jestliže platí $\mathbf{A}\vec{v} = \lambda\vec{v}$, kde λ je jisté komplexní číslo pro jistý vektor $\vec{v} \neq \vec{0}$, potom číslo λ se nazývá *vlastním číslem* a vektor \vec{v} *vlastním vektorem* matice \mathbf{A} .*



Obr. 2: Ortogonální projekce v prostoru R^3 , příklad vlastního vektoru.

Nejvýznačnější vlastností vlastních vektorů je, že jsou vzhledem k transformaci, tedy násobení maticí, invariantní (neměnné). Matematicky můžeme takovou vlastnost vyjádřit:

$$\mathbf{A}\vec{v} = \lambda\vec{v}, \quad (16)$$

číslo λ nazýváme vlastním číslem a \vec{v} vlastním vektorem. Rovnici (16) můžeme upravit:

$$(\mathbf{A} - \lambda\mathbf{I})\vec{v} = 0, \quad (17)$$

kde \mathbf{I} je jednotková matice. Po rozepsání dostaneme:

$$\begin{pmatrix} a_{(1,1)} - \lambda & a_{(1,2)} & \dots & a_{(1,n)} \\ a_{(2,1)} & a_{(2,2)} - \lambda & \dots & a_{(2,n)} \\ \vdots & \ddots & & \\ a_{(n,1)} & a_{(n,2)} & \dots & a_{(n,n)} - \lambda \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = 0.$$

Pro názornost jeden příklad: Je-li \mathbf{A} matice ortogonální projekce (jde o lineární zobrazení P) v prostoru R^3 na nějaký podprostor $U \in R^3$, dejme tomu na rovinu dle obr. 2. Je vidět, že pro $\forall \vec{u} \in U$ platí $\mathbf{P}\vec{u} = \vec{u}$. Všechny vektory z U jsou vlastními vektory matice \mathbf{P} , příslušné vlastnímu číslu $\lambda = 1$.

$\det |\mathbf{A} - \lambda\mathbf{I}| = 0$, je polynomem stupně n v proměnné λ , který nazýváme *charakteristický polynom* matice a jeho kořeny jsou vlastní čísla matice.

Způsobů zavedení vlastních vektorů a vlastních čísel je více. Definice v odstavci 3 je zcela ekvivalentní výše popsanému příkladu. Následující definice snad poslouží k úplnému pochopení toho, čím vlastní vektory a čísla jsou.

Buď V vektorový prostor nad číselnou množinou M . Lineární zobrazení $f : V \rightarrow V$ nazveme *lineární transformací* vektorového prostoru V nebo též *lineární operátor* na

vektorovém prostoru V . Každému vektoru $\vec{v} \in V$ je pak přiřazen obraz $f(\vec{v})$ z téhož prostoru V .

Definice 3 Vektor $\vec{v} \in V \setminus \{0\}$ se nazývá *vlastní vektor lineární transformace* $f : V \rightarrow V$, právě když

$$f(\vec{v}) = \lambda \vec{v}$$

pro některý skalár $\lambda \in M$. Skalár λ se nazývá *vlastní hodnota*, příslušná vlastnímu vektoru \vec{v} . Pak se λ nazývá *vlastní číslo*.

3.2.3 Operace s maticemi

V dalším textu používáme většinou základní maticové operace, které se studují obvykle v prvním ročníku vysoké školy. Následující věta je pro porozumění dalšímu textu důležitá a proto ji uvádíme.

Definice 4 *Ortogonální maticí rozumíme matici, pro niž $\mathbf{A}\mathbf{A}^T = \mathbf{E}$, kde \mathbf{A}^T označuje matici transponovanou k matici \mathbf{A} .*

Věta 1 *Matice \mathbf{A} je ortogonální právě tehdy, když*

$$\mathbf{A}^{-1} = \mathbf{A}^T. \tag{18}$$

Důkaz 1

Vyjdeme ze známé definice inverzní matice \mathbf{A}^{-1} k matici \mathbf{A} , pro kterou $\det \mathbf{A} \neq 0$

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{E} = \mathbf{A}^{-1}\mathbf{A}.$$

z definice 4 :

$$\mathbf{A}\mathbf{A}^T = \mathbf{E},$$

násobením zleva maticí \mathbf{A}^{-1} :

$$\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^T = \mathbf{A}^{-1}\mathbf{E},$$

$$\mathbf{E}\mathbf{A}^T = \mathbf{A}^{-1}\mathbf{E} = \mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1} = \mathbf{E}\mathbf{A}^{-1},$$

proto:

$$\mathbf{A}^T = \mathbf{A}^{-1}$$

■

3.2.4 Matematická formulace metody PCA

Z odstavce 3.1 víme, k čemu lze metodu PCA použít. Víme, jaký je vztah mezi původními proměnnými (veličinami) a hlavními komponentami. Zbývá se přesvědčit, že navržený matematický postup, opravdu nevyhnutelně spěje k nalezení takových hlavních komponent, které mají zmiňované vlastnosti.

Mějme libovolný vektor \vec{x}_i reprezentující data. Představuje jednu položku, tedy v našem případě jednu planetku. Ve složkách vektoru mohou být uloženy například hodnoty hvězdné velikosti planetky v různých filtrech. Tento vektor má své pevné místo v soustavě souřadnic. Nyní určíme nějakou novou, zatím zcela obecnou osu souřadnic. Bude jednotková a označíme ji \vec{U} . Náš vektor \vec{x}_i tímto získá Nové souřadnice na nové ose. Osy Souřadnic obvykle značíme písmeny x, y, z nebo písmenem iks s indexem i . Nově vytvořenou osu označíme PC_1 a pro její hodnotu jistě bude platit:

$$PC_1 = \vec{x}_i \bullet \vec{U}. \quad (19)$$

Skalární součin je skalár, PC_1 je také skalár a vše je tak v pořádku. Dosud popsané provedeme pro všechny položky (planetky). Získáme n nových souřadnic PC_1 příslušející n planetkám. Následně spočteme jejich aritmetický průměr: (n označuje počet planetek)

$$\overline{PC_1} = \frac{1}{n} \sum_{i=1}^n \vec{x}_i \bullet \vec{U},$$

dále vypočteme rozptyl nové souřadnice (tedy rozptyl dat po transformaci souřadnic):

$$\text{var}(PC_1) = \frac{1}{n-1} \sum_{i=1}^n (PC_{1,i} - \overline{PC_1})^2, \quad (20)$$

což lze rozepsat a upravit [9] (použitím rovností $(\mathbf{AB})^T = \mathbf{A}^T \mathbf{B}^T$):

$$\text{var}(PC_1) = \frac{1}{n-1} \sum_{i=1}^n (\vec{x}_i \bullet \vec{U} - \frac{1}{n} \sum \vec{x}_i \bullet \vec{U})^2, \quad (21)$$

$$\text{var}(PC_1) = \frac{1}{n-1} \sum_{i=1}^n (\vec{U} \bullet (\vec{x}_i - \frac{1}{n} \sum \vec{x}_i))^2, \quad (22)$$

$$\text{var}(PC_1) = \frac{1}{n-1} \sum_{i=1}^n (\vec{U} \bullet (\vec{x}_i - \frac{1}{n} \sum \vec{x}_i) \cdot (\vec{U} \bullet (\vec{x}_i - \frac{1}{n} \sum \vec{x}_i))^T), \quad (23)$$

$$\text{var}(PC_1) = \frac{1}{n-1} \vec{U} \cdot \left[\sum_{i=1}^n ((\vec{x}_i - \frac{1}{n} \sum \vec{x}_i) \cdot ((\vec{x}_i - \frac{1}{n} \sum \vec{x}_i)^T)) \right] \cdot \vec{U}^T, \quad (24)$$

výraz v hranatých závorkách odpovídá definici kovarianční matice. V tom případě:

$$\text{var}(PC_1) = \frac{1}{n-1} \vec{U} \mathbf{C} \vec{U}^T, \quad (25)$$

na levé straně rovnosti (25) je skalár. Jednotlivé členy rovnice (25) vhodně přeskupíme a získáme:

$$\lambda \vec{U} = \mathbf{C} \vec{U}, \quad (26)$$

což není nic jiného než definice vlastního vektoru, tedy \vec{U} evidentně musí být vlastním vektorem. Na počátku jsme \vec{U} volili libovolně. Při použití metody hlavních komponent vybíráme takový vektor \vec{U} , jež přísluší maximálnímu vlastnímu číslu λ . Porovnáme-li rovnice (25) a (26) určíme:

$$\lambda = (n - 1) \text{var}(PC_1), \quad (27)$$

to znamená, že pro největší vlastní číslo bude maximální možný i rozptyl $\text{var}(PC_1)$. Obržet vektor, v jehož směru budou data nabývat největšího rozptylu, je přesně tím, čeho jsme chtěli dosáhnout, a důkaz lze považovat za dokončený.

3.3 Dokumentace užítých programů

Nezbytnou součástí práce jsou programy vykonávající výpočty. Stojí na výše popsaných matematických základech. V této kapitole přehledně představíme, jakým způsobem programy použít a jakou mají funkci. Celý výpočet je prováděn prostřednictvím tří samostatných programů. Výstupem jsou hlavní komponenty příslušné datovému souboru na vstupu.

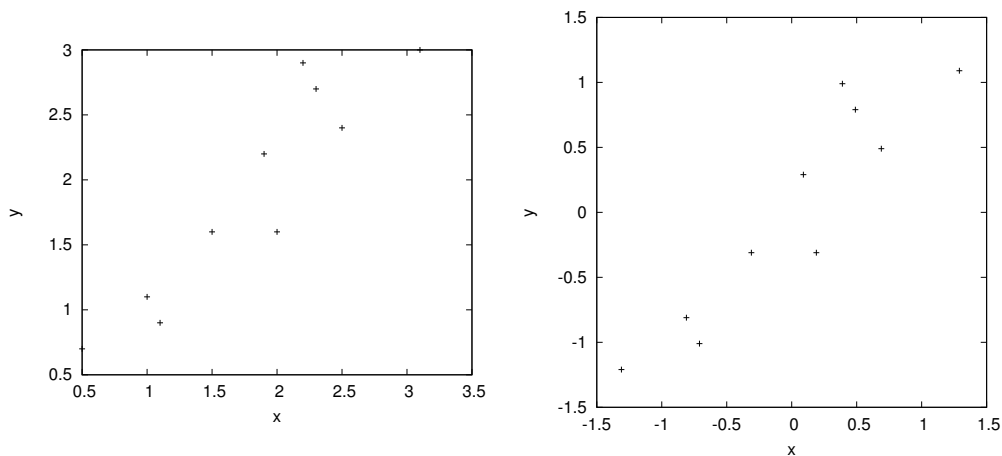
Jednotlivé podprogramy a programy jsou vytvořeny v jazyce Pascal. Abychom zachovali přehlednost při popisu výpočetních kroků a aby bylo jasné, co programy dělají, opustíme na chvíli reálná rozsáhlá data. Místo nich uvedeme cvičnou, testovací dvou-rozměrnou množinu dat. Množina je vypsána v tabulce 1. Její hodnoty vyneseny do grafu jsou v obr. 3.

Tab. 1: Vstupní data ukázkového příkladu

x	y
2,5	2,4
0,5	0,7
2,2	2,9
1,9	2,2
3,1	3,0
2,3	2,7
2,0	1,6
1,0	1,1
1,5	1,6
1,1	0,9

Prvním použitým programem, je `kovariancni_matice.pas`. Spustit jej můžeme například z terminálu (příkazového řádku) příkazem:

```
./kovariancni_matice < kovariancni_matice.in
```



Obr. 3: Grafické znázornění datového souboru ukázkového příkladu. Vlevo původní data z tabulky 1, vpravo původní data po redukcí.

Vstupní soubor `kovariancni_matice.in` obsahuje název souboru s daty, pro která počítáme hlavní komponenty, a hodnotu přepínače zobrazení mezivýsledků na standardním výstupu (0 pro nezobrazování, 1 pro zobrazení). Soubor `kovariancni_matice.in` může vypadat například takto:

```
SDC.dat
0
```

Po spuštění programu se nejprve stanoví průměr každého sloupečku. Hodnota průměru se od každého členu sloupečku odečte. V našem jednoduchém případě budou všechny hodnoty x zmenšené o \bar{x} a všechny hodnoty y zmenšené o \bar{y} . Průměr upravených dat je nula.

Podle definice (7) vypočteme postupně hodnoty kovariance mezi proměnnými. Po uložení jednotlivých hodnot do matice (podle (9)) získáme matici kovariance, která se vypíše na standardní výstup. Standardní výstup je uložen do souboru pojmenovaného `kovariancni_matice.out`:

```
# Zadej jmeno vstupniho souboru:
# Pro vypis matice dat zadej 1, jinak 0:
# filename = SDC.dat
# n = 10
# m = 2
# CovarianceMatrix =
0.6165555556 0.6154444444
0.6154444444 0.6165555556
```

Nediagonální prvky kovarianční matice vyšly kladné, lze tedy očekávat, že závislost mezi x a y se bude blížit přímé úměře, nikoli nepřímé.

V dalším kroku určíme vlastní čísla a jim příslušné vlastní vektory. Pro dvourozměrná data provádí tento úkol program `vlastni_vektory.pas`. Spustit ho lze příkazem:

```
./vlastni_vektory < vlastni_vektory.in
```

Vstupní soubor programu `vlastni_vektory.pas` obsahuje kovarianční matici vypočtenou v předchozím kroku.

Program nejprve určí koeficienty charakteristické rovnice podle vztahu (13). Ta je pro dvourozměrná data polynomem druhého stupně a je dále řešena jako kvadratická rovnice. Kořeny, neboli vlastní čísla, se vypíší do výstupního souboru. V dalším se postupně pro obě vlastní čísla určí příslušné vlastní vektory, a to řešením obecně n rozměrné (v našem případě dvourozměrné) soustavy rovnic. Užíváme k tomu Gaussovu eliminační metodu. Získané vlastní vektory se taktéž objeví ve výstupním souboru `vlastni_vektory.out`:

```
# Zadej kovariancni matici 2 x 2:
# lambda1 = 1.284027712172784e+00
# lambda2 = 4.908339893832728e-02
# vlastní vektor prislusny lambda1 =
 6.778733985280118e-01  7.351786555444081e-01
# vlastní vektor prislusny lambda2 =
-7.351786555444080e-01  6.778733985280118e-01
```

Výpočet hlavních komponent provádí program `hlavni_komponenta.pas`, a to dle vztahu (14). Vstupní soubor se nazývá `hlavni_komponenta.in`. Program dokáže pracovat s libovolným menším počtem rozměrů než čtyři, a proto je třeba do prvního řádku vstupního souboru `hlavni_komponenta.in` zapsat rozměr dat. Podobně i počet hlavních komponent, které budou spočteny, je volitelný se stejným omezením, tedy maximálně do čtyř. Do následujících řádků zkopírujeme vlastní vektory získané programem `vlastni_vektory` a v posledním řádku definujeme jméno souboru s daty, pro která hlavní komponenty počítáme. Příklad vstupního souboru následuje:

```
2
2
 6.778733985280118e-01  7.351786555444081e-01
-7.351786555444080e-01  6.778733985280118e-01
SDC.dat
```

Vlastní vektory jsme normalizovali, a jsou tedy jednotkové.

Pokud ponecháme oba vlastní vektory, obdržíme dvě hlavní komponenty. Jejich hodnoty jsou vypsány v tabulce 3.3 a zobrazeny v obr. 4. Vidíme, že původní data jsou pouze otočená tak, aby osy x a y přešly ve vlastní vektory. Je patrné, že takovým přeuspořádáním žádné informace o vnitřní struktuře v datech neztratíme.

Ponecháme-li pouze jeden vlastní vektor, získáme jednu hlavní komponentu a na výstupu očekáváme data jednorozměrná. Vypsána jsou v tabulce 3.3 a zobrazeny v grafu 5. V tomto případě jsme jistou, ale v mnoha případech zanedbatelnou informaci o datech ztratili. Zřetelně to uvidíme, pokud provedeme inverzní transformaci,

Tab. 2: Vlastní komponenty příslušející původním datům

PC_1	PC_2
-0,82797018	-0,175115307
1,77758033	0,142857227
-0,99219749	0,384374989
-0,27421041	0,130417207
1,67580142	-0,209498461
-0,91294910	0,175282444
0,09910943	-0,349824698
1,14457216	0,046417258
0,43804613	0,017764629
1,22382056	-0,162675287

tzn. z hlavních komponent PC_1 a nulové PC_2 spočteme x a y . V posledním kroku, prostřednictvím kterého jsme obdrželi matici nových souřadnic (hlavních komponent), jsme vlastní vektory násobili maticí redukovaných původních dat (viz (14)). Použitím maticového zápisu získáme:

$$\mathbf{PC} = \mathbf{Vlv} \overline{\mathbf{XY}}, \quad (28)$$

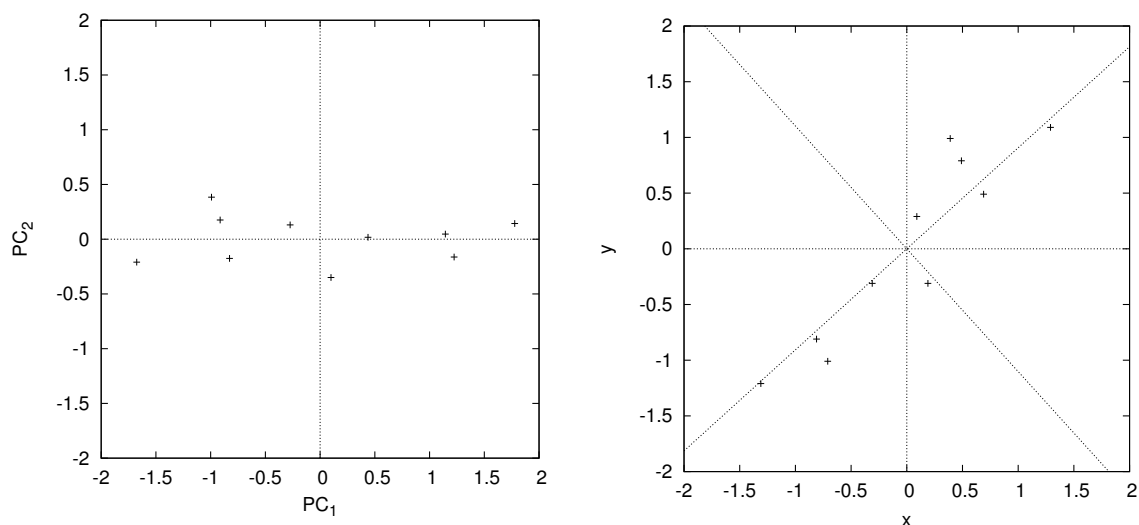
označení \mathbf{PC} je maticí hlavních komponent, \mathbf{Vlv} je matice vlastních vektorů a $\overline{\mathbf{XY}}$ je matice redukovaných dat. Zápis inverzní transformace:

$$\mathbf{Vlv}^{-1} \mathbf{PC} = \overline{\mathbf{XY}}, \quad (29)$$

ekvivalentně podle (4) dostaneme:

$$\mathbf{Vlv}^T \mathbf{PC} = \overline{\mathbf{XY}}, \quad (30)$$

Na obrázku 5 vidíme příslušný graf.



Obr. 4: Vlevo grafické znázornění dat po transformování metodou hlavních komponent. Vpravo původní, normalizovaná data se znázorněním směru vlastních vektorů.

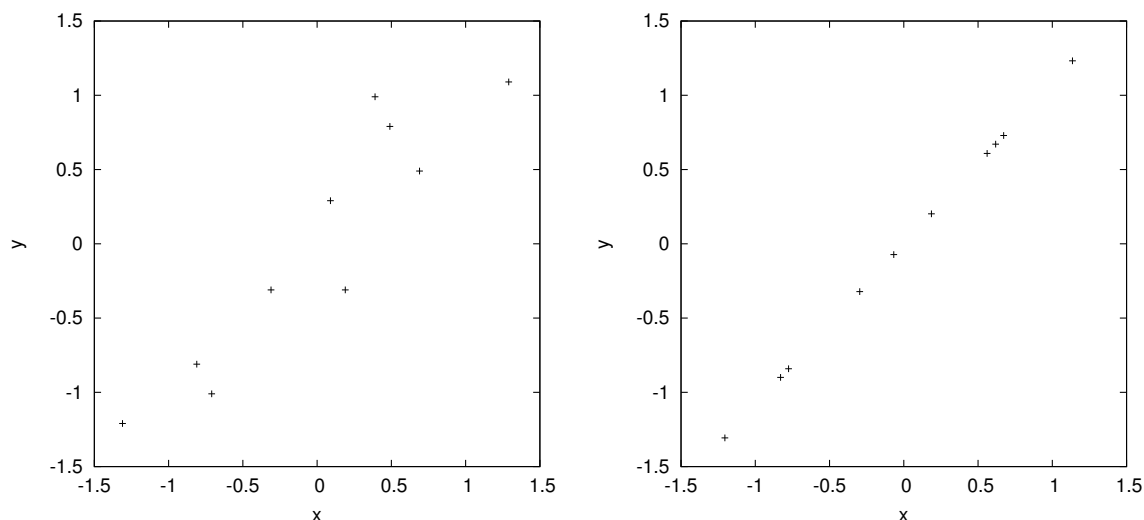
Tab. 3: Vlastní komponenta PC_1 příslušející původním datům

PC_1
-0,82797018
1,77758033
-0,99219749
-0,27421041
1,67580142
-0,91294910
0,09910943
1,14457216
0,43804613
1,22382056

4 Užité fyzikální principy

4.1 Radiometrie

Radiometrické veličiny [7] popisují množství světla. Jedná se nám především o měření světla dopadajícího na čidlo přístroje. Protože potřeba určovat například množství světla přicházejícího od hvězd byla přítomna i v době neexistence objektivních čidel záření, používalo se k tomu účelu lidské oko. Oko ovšem pracuje nepřesně a nelineárně a hodnoty okem naměřených veličin se obecně liší od těch přístrojových. V současném názvosloví tedy paralelně s radiometrií zavádíme i fotometrii. Fotometrie popisuje



Obr. 5: Data zpětně převedená do původních proměnných. Vlevo za použití obou komponent. Vpravo při vynechání méně důležité komponenty PC_2 .

množství světla v závislosti na tom, jakým účinkem působí na lidské oko⁴.

Než světlo od zdroje (například planetky) dopadne na detektor, prochází atmosférou Země. Musíme proto počítat s tím, že část záření bude absorbována nebo rozptýlena, přičemž absorpce velmi závisí na vlnové délce záření. Navíc, pro planetku nízko nad obzorem je absorpce zcela jiná než pro planetku v zenitu. S takovými efekty si lze porovnat měření standardních hvězd a vhodným přepočtem. Podrobněji se o problematice můžete dočíst v [14], [4].

Zářivý tok (nazývaný též výkon) odpovídá celkové energii vyzářené zdrojem ve všech vlnových délkách za jednotku času. Zářivý tok značíme L a jeho jednotkou je W .

Fotonová zářivost I v daném směru je definovaná zářivým tokem vyslaným do jednotkového prostorového úhlu v daném směru. Jednotkou je $W \text{ m}^{-2} \text{ sr}^{-1}$

Plošná hustota F zářivého toku odpovídá toku záření, který za sekundu projde 1 m^2 plochy postavené kolmo ke směru přicházejících paprsků. Jednotka je W/m^2 .

Signál, který jsme schopni zaznamenat⁵, má k ideálnímu, izotropnímu daleko. V praxi popisované objekty září v různých vlnových délkách s různou spektrální hustotou zářivosti. Proto zavádíme radiometrické veličiny jako *spektrální hustotu zářivého toku*, které jsou *funkcí vakuové vlnové délky* anebo *frekvence* $\nu = \frac{\lambda}{c}$. Pro spektrální hustotu zářivého toku platí $H = f(\lambda)$. Určuje, jaká část celkové energie vyzářené zdrojem připadá na záření o vlnové délce mezi λ a $\lambda + \Delta\lambda$. Jednotkou je $W \cdot \text{m}^{-2} \cdot \text{m}^{-1} = W \cdot \text{m}^{-3}$. Celková hustota F zářivého toku je:

⁴V astronomii se často hovoří o fotometrii, i když se měří objektivními přístroji v nevizuálním oboru

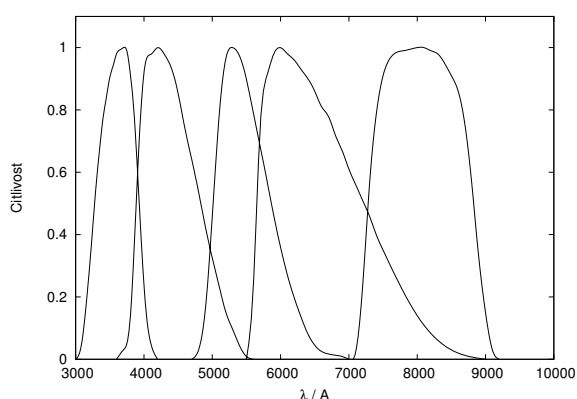
⁵Mluvíme nyní o příjmu signálu v programu SDSS. Obecně lze i tak nepatrné toky, které přicházejí od planetek, změřit v celé rozhodně širší spektra. Jedná se o téměřbolometrické hodnoty. Takové experimenty jsou prováděny mimo atmosféru Země, na oběžné dráze.

$$F = \int_0^{\infty} H(\lambda)d\lambda, \quad (31)$$

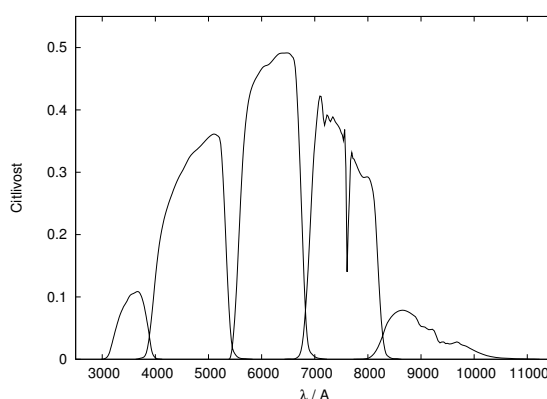
Jak vidíme, měření radiometrických veličin je komplikováno řadou faktorů. Zemská atmosféra snižuje možnost měření tím, že je pro řadu oborů elektromagnetického záření zcela nepropustná a my zachycujeme jen část intervalů vlnových délek.

Další obtíží je nestejná spektrální citlivost detektorů záření (konkrétně CCD kamer).

Z těchto a dalších důvodů se v řadě astrofyzikálních aplikací, jež v sobě přirozeně zahrnují celý rozsah elektromagnetického spektra, používají veličiny, vztažené na určitou část spektra elektromagnetického záření. Jednotlivé části se nejčastěji vymezují filtry s přesně danou propustností. Nejinak tomu je v případě projektu SDSS.



Obr. 6: Propustnost Johnsonovy soustavy filtrů.



Obr. 7: Propustnost filtrů použitých v SDSS.

4.1.1 Fotometrické systémy

Hned z počátku poznamenejme, že správně bychom měli mluvit o radiometrických systémech. Protože s popisovanými systémy pracují především astronomové, budeme užívat v astronomii vžitého označení „fotometrické“.

Fotometrické systémy jsou definovány soustavami barevných filtrů přesně daných propustností. Standardizace jednotlivých fotometrických systémů byla podřízena historicko-technickým vlivům. Jednotlivé filtry jsou charakterizovány pološířkou spektrální propustnosti, maximální propustností a odrazivostí. Každý systém je navíc doplněn o seznam *standardních hvězd*, které slouží ke vzájemné kalibraci jednotlivých fotometrických soustav. V astronomické praxi se nejčastěji setkáváme s pásmovými filtry. Ty členíme podle pološířky spektrální propustnosti na širokopásmové (propustnost v stovkách nm), středněpásmové (desítky nm) a úzkopásmové (méně než 10 nm).

Johnsonův fotometrický UBV systém je nejpoužívanějším systémem. Udává jasnosti v barvě U (365 nm), B (440 nm), V (550 nm); graf spektrální propustnosti filtrů viz obr. 7.

UBVRI fotometrický systémfotometrický systém UBVRI je rozšířením Johnsonova do červené a infračervené oblasti elektromagnetického záření. Jednotlivé filtry mají

maximum propustnosti v R (700 nm), I (900 nm) případně J (1250 nm), K (2200 nm), L (3400 nm), M (5000 nm), N (10200 nm).

Strömgrerův fotometrický systém je dalším z často používaných. Maximální propustnosti jednotlivých filtrů jsou u (350 nm), v (410 nm), b (470 nm), y (550 nm). Pološířka spektrální propustnosti je oproti předchozímu systému menší.

Do dnešní doby vzniklo přes 200 prakticky používaných fotometrických systémů. Například systém speciálně vytvořený pro kosmickou družici HIPPARCOS.

Měřením jasností planetek ve vícero oborech si lze učinit uspokojivou představu o rozložení energie ve spektru. Ta je funkcí povrchových vlastností planetek. Fotometrický systém používaný v průběhu „sloanské“ přehlídky je odvozen ze systému AB_{ν} (Oke & Gunn 1983). Tvůrci sloanské přehlídky původní AB_{ν} systém upravili, aby převod mezi magnitudami a zářivými toky byl ještě spolehlivější, a nazvali ho AB_{95} . Pět filtrů pokrývá oblast vlnových vakuových délek od ultrafialové, kterou již zemská atmosféra propouští, až po limit citlivosti používaného křemíkového CCD snímače. Maxima propustnosti jednotlivých filtrů jsou následující: u (350 nm), g (480 nm), r (625 nm), i (770 nm), z (910 nm); graf spektrální propustnosti je uveden na obr. 7. K překryvu jednotlivých filtrů prakticky nedochází.

4.1.2 Hvězdná velikost (magnituda)

Hvězdná velikost je veličina popisující fotonovou zářivost objektů. Jedná se o jednu z nejzákladnějších astronomických veličin, používanou již od počátku astronomie samotné. Její geneze je proto rozsáhlá a jejímu vymezení je nutné věnovat určitou pozornost.

Původně se hvězdná velikost určovala vizuálně. Proto byla později zavedena jako fotometrická veličina, jež nepřímou vyjadřuje hustotu světelného toku z hvězd (či jiných nebeských objektů). V tomto případě mluvíme o *vizuální* hvězdné velikosti. V současnosti získáváme hvězdnou velikost v různých oborech elektromagnetického záření. Postupný vývoj techniky umožnil určovat hvězdnou velikost pomocí různých detektorů záření. Získali jsme tak ryze objektivní veličinu, nezávislou na zraku pozorovatele.

Objev Weberova — Fechtnerova psychofyzického zákona určil další vývoj definice hvězdné velikosti. Zákon říká, že *mění-li se fyzikální podmínky působící na lidské smysly řadou geometrickou, vnímá člověk jejich změnu v řadě aritmetické*. Přesto, že zákon není univerzálně platný, dal podnět k logaritmickému měření změny energie. Matematicky vyjádřeno, subjektivní vjem veličiny není úměrný veličině samotné, ale jejímu logaritmu. Pokud tedy m_1 a m_2 jsou hvězdné velikosti dvou nebeských objektů a F_1 a F_2 příslušné hustoty zářivých toků, platí:

$$m_1 - m_2 = -k(\log F_1 - \log F_2)$$

Záporné znaménko u konstanty k signalizuje nepřímou úměrnost mezi hvězdnou velikostí a hustotou zářivého toku anebo veličinou která je hustotě zářivého toku úměrná. Konstanta samotná je určena podle měřeného výsledku, že pro objekty s podílem hustoty světelného toku 100 (tedy od jednoho objektu přijímáme stokrát více energie než od druhého) vyhodnotí lidské oko jeden z objektů jasnější pouze šestkrát. Pro konstantu

pak platí $k = -2, 5$, což je spočteno z rovnice:

$$6 - 1 = -k(\log 100 - \log 1).$$

Nyní můžeme rovnici napsat tak, jak ji roku 1850 definoval anglický astronom Norman Robert Pogson. Dodnes ji nazýváme Pogsonova rovnice:

$$m_1 - m_2 = -2,5 \log \frac{F_1}{F_2}.$$

Hvězdné velikosti, o které jsme mluvili doposud, přísluší ještě další vymežující adjektivum: zdánlivá. *Zdánlivá hvězdná velikost* neodpovídá skutečné zářivosti těchto těles, neboť je významně zkreslena jejich vzdáleností. Její použití je výhodné především v observační astronomii. Například zdravé lidské oko je schopno uvidět hvězdy do 6. hvězdné velikosti⁶.

Zbývá uvést definici *absolutní hvězdné velikosti* M . Jedná se o hvězdnou velikost, jakou by objekt nabýval ve vzdálenosti 10 parseků (32,6 světelných roků). Hustota F zářivého toku určitého zdroje o zářivosti I je nepřímo úměrná kvadrátu vzdálenosti r , v níž jasnost měříme. Pokud porovnáme jasnosti F_1 a F_2 jednoho zdroje v obecně libovolných vzdálenostech r_1 a r_2 , obdržíme pro jejich poměr:

$$\frac{F_2}{F_1} = \frac{I}{r_2^2} \frac{r_1^2}{I} = \left(\frac{r_1}{r_2} \right)^2. \quad (32)$$

Pogsonovu rovnici tudíž můžeme upravit:

$$m_2 - m_1 = -2,5 \log \frac{F_2}{F_1} = -5 \log \frac{r_1}{r_2}. \quad (33)$$

Změříme-li osvětlení od zdroje ve vzdálenosti deseti parseků, jasnost kterou získáme, bude jistě splňovat podmínky definice absolutní hvězdné velikosti. Můžeme tedy psát:

$$m - M = -5 \log \frac{10}{r} = -5 \log 10 + 5 \log r. \quad (34)$$

Vztah pro převod mezi absolutní a zdánlivou hvězdnou velikostí píšeme ve tvaru:

$$M = m + 5 - 5 \log r,$$

kde r označuje vzdálenost v parsecích. Kdyby se všechny hvězdy v naší Galaxii přemístily do vzdálenosti 10 pc od Země, odpovídala by jejich zdánlivá hvězdná velikost energii, kterou hvězdy vysílají. Našemu Slunci přísluší absolutní hvězdná velikost +34,71 mag, Polárce −3,61 mag.

Pro málo zářící tělesa, jako planetky, komety nebo planety zavádíme absolutní hvězdnou velikost poněkud odlišným způsobem. Jednoduše proto, že podle výše uvedené definice by absolutní hvězdná velikost vycházela nezvykle velká. Tedy *absolutní hvězdná velikost pro planetku* je zdánlivá hvězdná velikost v případě, že vzdálenost planetky od Slunce (které ji osvětluje) je 1 AU, vzdálenost planetky od Země (na níž sedí pozorovatel) je 1 AU a fázový úhel (Slunce, planetka, Země) $\alpha = 0^\circ$ ⁷.

⁶Někdy se nesprávně říká 6. magnitudy. Magnituda je ale jednotkou veličiny — hvězdné velikosti.

⁷Definice absolutní hvězdné velikosti pro planetky je zavádějící, neboť nás jako pozorovatele umísťuje do středu Slunce.

4.2 Barevný index

Barva hvězd úzce souvisí s jejich fotosférickou teplotou. Načervenalé hvězdě odpovídá teplota asi 3000 K, oranžové asi 4500 K, nažloutlé 6000 K, bílé 10 000 K a teplotě nad 10 000 K odpovídá modrá barva. Pokud je celý systém barev hvězd a příslušných teplot správně kalibrován, lze prostřednictvím informace o barvě rozhodnout, kterému spektrálnímu typu hvězda přísluší.

Barva planetek, vypovídá především o jejich povrchových vlastnostech. Ve vizuální a blízké infračervené oblasti, kde planetky především „odrážejí“ sluneční světlo, lze pak usuzovat na mineralogické složení povrchu a jeho strukturu. V dalekém IR oboru lze zachytit tepelné záření planetek samotných, pak lze usuzovat na povrchovou teplotu (řádově 100 K) nebo na tepelné vlastnosti (vodivost, kapacita, hustota povrchových vrstev) materiálu planetky.

V astronomii se barva nebeského objektu obvykle vyjadřuje číslem. Toto číslo nazýváme barevným indexem a jedná se o rozdíl hvězdných velikostí měřeného objektu ve dvou různých spektrálních oborech. Barevný index zavedl roku 1900 Karl Schwarzschild. Pro určení barevného indexu hvězd používal fotografickou a vizuální hvězdnou velikost. První odpovídá vlnové délce 425 nm druhá 590 nm. Takto zavedený barevný index nazýváme *mezinárodní*. V současnosti se nejčastěji používá tříbarevný *UBV* systém, zvaný též *standardní systém*. V něm jsou indexy $U - B$ a $B - V$ definovány jako rozdíly mezi ultrafialovou 365 nm a modrou 440 nm, respektive mezi modrou a vizuální 550 nm hvězdnou velikostí.

Barevný index zavádíme především z toho důvodu, že samotná hvězdná velikost má pro nás jen malou výpovědní hodnotu. Například i tělesa se zcela různými spektrálními vlastnostmi mohou nabývat pro určité vlnové délky stejných hodnot hvězdných velikostí. Použití barevného indexu v případě planetek zajistí určitou „normalizaci“ jasností jednotlivých planetek, těch které by se lišily vlivem jejich různé vzdálenosti od Slunce a od Země. Je totiž zřejmé, že blízkozemní planetka dané velikosti má obecně ve všech barevných filtrech menší hvězdnou velikost (je jasnější) než například planetka transneptunická téže velikosti (jednoduše pro to, že transneptunická planetka je mnohem dále od Země i Slunce než blízkozemní), přesto, že by jejich spektrum a tedy složení mohlo být totožné.

5 Aplikace metody PCA na fotometrická data planetek

5.1 Přehledka Sloan Digital Sky Survey

Sloan Digital Sky Survey (SDSS) je moderní elektronická přehledka oblohy [17]. Snaha o zaznamenání stavu oblohy však lidstvo provází nejméně od roku 350 př.n.l, kdy řecký básník Arastos v básni Phenomena popsal tehdy známá souhvězdí. V té době se vědě dařilo a mohutně se rozvíjela. V roce 150 př.n.l. se na světě objevil první katalog hvězd, sestavený řeckým učencem Hipparchem. Ve své první verzi obsahoval asi 800 stálic a později byl dokonce rozšířen na 1200. To zhruba odpovídá počtu hvězd, viditelných na naší světlem znečištěné obloze. Čas běžel.

Starověká souhvězdí se udržela díky arabské tradici. Ve středověku byla znovu objevena podobně jako jiná antická díla. Roku 1603 německý právník Johannes Bayer vydává první velký hvězdný atlas, ve kterém uvádí všechna starověká souhvězdí. O 90 let později, roku 1690, Johan Hevelius doplňuje „prázdná místa“ na severní obloze a tiskem mapy vydává.

V devatenáctém století stojí za připomenutí na svou dobu velký projekt, jehož cílem bylo fotografické mapování celého nebe z osmnácti observatoří celého světa. Potřeba většího množství dat vedla v padesátých letech dvacátého století k vytvoření ještě rozsáhlejší přehledky — slavné Palomar Observatory Sky Survey (POSS). Zaznamenávalo se na fotografické desky o velikosti 35×35 cm. Na každé desce bylo zachyceno pole o průměru 5° vždy v červené a modré barvě. Známý vesmír se tenkrát rozšířil dvacetpětkrát. Každá deska byla verzně vytištěna na papír a ve formě katalogu distribuována do rukou astronomů. Přestože přehledka POSS byla později opakována a rozšířena, vyvstala potřeba (jako už mnohokrát) podrobnějších a přesnějších informací.

Přehledka SDSS začala v devadesátých letech dvacátého století. Jedná se o ještě rozsáhlejší projekt s jednou zásadní změnou — účastí výpočetní techniky. Celý proces od pozorování přes snímání po vyhodnocování dat je automatizován. SDSS je první přehledkou snímající na CCD čip. Dvou a půl metrový dalekohled na observatoři Apache Point v Novém Mexiku byl projektován a postaven výhradně za účelem vytvoření katalogu SDSS. Světlo ze 3° velkého zorného pole je buď vedeno na fotometrickou kameru, nebo pomocí optických vláken až k dvěma spektrografům. Kamera je sestavena z třiceti čipů, každý čip má 2048×2048 pixelů. Čipy jsou seřazeny do 6 sloupců po pěti tak, aby na sebe jednotlivé sloupečky pixelů jednotlivých čipů ($6 \times 5 = 30$) navazovaly. Před každým sloupcem je identická sada pěti barevných filtrů, od fialové po infračervenou (r, i, u, z, g) s propustností dle obr. 7. Dalšími snímacími prvky, kterými dalekohled disponuje, je 24 menších čipů na okraji zorného pole. Slouží k určení polohy snímaných objektů a ke kontrole zaostření.

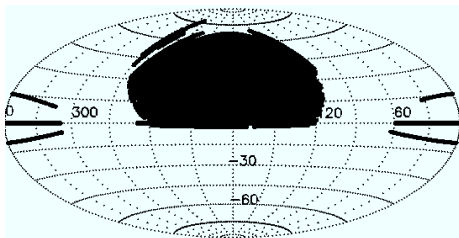
Snímání probíhá tak, že obraz hvězdy vstoupí nejprve na jeden z pomocných, astro-metrických čipů. Následně se obraz přesune na první řádek fotometrických čipů hlavní části kamery, poté na druhý řádek, a tak dále. Nakonec obraz hvězdy projde přes čip pátého řádku. Tímto postupem je možno za 55 sekund zaznamenat informace o poloze hvězdy i o její hvězdné velikosti v pěti barvách. Aby hvězda přejela přes čipy kamery v žádaném směru a žádané rychlosti, je nutné dalekohledem aktivně (automaticky)

posouvat.

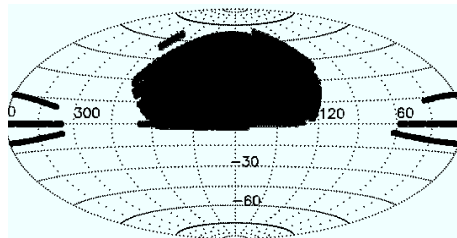
Spektra je možné pořizovat, pokud je do ohniska dalekohledu zasunuta deska s optickými vlákny. Vlákna uchycených v provrtaných dírkách desky je až 640. Obraz objektu, jehož spektrum zamýšlíme pořídít, musí „dosednout“ přesně na nějaký otvor desky s optickým vláknem. Proto jsou tyto desky vrtány na míru, podle fotografie získané fotometrickou kamerou.

5.2 Katalog SDSS MOC

Data získaná během třech hlavních⁸ pozorovacích programů plní katalog nesoucí označení SDSS. Prohlédnuto bylo pět pásů oblohy v oblasti severního galaktického pólu o celkové rozloze 7646 čtverečních stupňů tj (23,5 % oblohy), viz obr. 8, obr. 9. Také pak 750 čtverečních stupňů (2,3 % oblohy) v okolí jižního galaktického pólu. Započítáme-li i oblasti podrobené spektroskopickým měření (galaxie a kvasary), získáme celkově 8423 (26 %) čtverečních stupňů prohlédnuté, zaznamenané, komukoliv přístupné části oblohy. Rozšíření sloanské přehlídky o 3240 čtverečních stupňů (10% oblohy), pokrývající nižší galaktické šířky, má sloužit ke studiu struktury Mléčné dráhy. Tento rozšiřující program nese označení SEGUE (Sloan Extension for Galactic Understanding and Exploration), obr. 10, obr. 11. Nakonec zmiňme pozorovací část s výstižným názvem Supernova. Zahrnuje 80 opakovaných snímkování v pásu širokém 2,5°. Situován je do oblasti rovníku, s počátkem na hranicích souhvězdí Býka a Eridanus a koncem u hranice Vodnáře a Orla. Tento program probíhal převážně v období zhoršených pozorovacích podmínek, při oblačných nocích, noci s úplňkem aj.



Obr. 8: Pokrytí oblohy hlavním pozorovacím programem přehlídky SDSS — spektroskopická měření

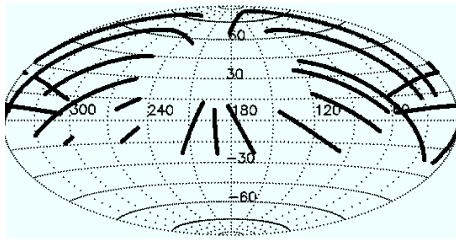


Obr. 9: Pokrytí oblohy hlavním pozorovacím programem přehlídky SDSS — pořizování fotografických snímků

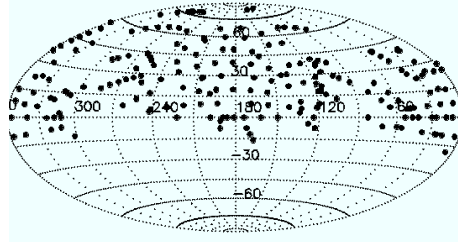
Vyjádřeno na kusy, přehlídka pořídila obrazy více jak 350 milionů nebeských těles, spektra 930 tisíců galaxií, 120 tisíců kvasarů a 460 tisíc hvězd. Získané informace jsou plně kalibrovány, redukovány a efektivně seřazeny do databáze, která byla každým rokem rozšiřována právě o pozorování za rok přibývší. Pro úplnost - posledního rozšíření se dostalo označení DR7 (Data Release).

Naše pozornost by se měla soustředit na základní pozorovací program, zvaný Legacy. Určitá část oblohy byla v rámci tohoto programu fotografována vícekrát. Tímto opatřením se předně, prodloužila expoziční doba a byly zaznamenány slaběji zářící

⁸V rámci SDSS proběhly desítky speciálních pozorovacích „prográmečků“ (slovo prográmeček chápeme v kontextu mamutího programu SDSS).



Obr. 10: Pokrytí oblohy pozorovacím programem SEGUE přehlídky SDSS — pořizování fotografických snímků.



Obr. 11: Rozsah pokrytí oblohy pozorovacím programem SEGUE přehlídky SDSS — spektrická měření



Obr. 12: Duchové, odlesky a identifikované létající objekty na snímcích SDSS. První obrázek ukazuje fotografii zorného pole s jasnou hvězdou v blízkém okolí. Její světlo se odrazí od tubusu dalekohledu a dopadne na čip. Konstrukce dalekohledu byla sice navržena tak, aby takové odrazy eliminovala, ale jak vidíme, velmi silný zdroj světla dokáže své. Druhý obrázek zachycuje satelit na nízké orbitě. Jasnost satelitu kolísá v důsledku jeho rotace. Na třetím obrázku máme štěstí a vidíme právě prolétávající jasný meteor. Čtvrtý obrázek předvádí difrakční kříže hvězd, vzniklé na konstrukci dalekohledu. Šestý obrázek by měl být nejnápadnější světelným halem okolo jasné hvězdy. Pokud totiž jeden pixel CCD kamery značně osvětlíme, přeplní se záporným nábojem. Náboje se mohou přemístit i do okolních pixelů.

objekty. Za druhé byly rozpoznány objekty, pohybující se vůči hvězdnému pozadí. Výsledkem je katalog SDSS-II o informačním rozsahu 10 TB, v jehož rámci se podařilo identifikovat 471 569 pohybujících se těles — planetek.

Ve snaze maximálně usnadnit uživatelům použití dat, jsou data od všech planetek zpracována zvláště do textového souboru s označením MOC (Moving Object Catalog). Zájemcům o jiný způsob zpracování, než je použit pro MOC, jsou k dispozici zmiňované původní obrázky v katalogu SDSS-II.

Data z pozorovací oblasti v rozmezí $\pm 15^\circ$ podél galaktické roviny jsou však zrádná. Hodnoty astrometrických a fotometrických veličin nejsou zcela přesné. Příčinou je ohromná koncentrace objektů v této části oblohy. Většina obrazů hvězd splývá a vzniká jen obtížně redukovatelný snímek. Situace je obzvláště nepříjemná, hledáme-li tělesa se zvláštními barvami. Podrobněji o zrádnosti dat pojednává [13].

V katalogu SDSS jsou k nalezení astrometrické veličiny — pozice a rychlost a fotometrické veličiny — hvězdné velikosti v pěti barevných filtrech. U všech veličin jsou samozřejmě uvedeny i chyby měření. Katalog MOC obsahuje i parametry drah katalogizovaných (známých) planetek.

Katalog MOC je tvořen daty, která byla původně zahrnuta v katalogu SDSS-II. Vyčlenění proběhlo užitím databázových dotazů (query). Nejprve byly vyjmuty všechny

příliš jasné a saturované objekty. U takových si totiž můžeme být jisti, že mají reálný obraz na zpracovaných snímcích a že se tedy nejedná o nějakou šmouhu, šum či cokoliv jiného nepoctivého. Do dalšího kola „soutěže“ o zahrnutí do MOC postoupily jen ty objekty, které prokázaly, že se oproti hvězdnému pozadí výrazně pohybují. Z kandidátů byly dále vyřazena tělesa, u kterých se nepodařilo určit parametry drah nebo⁹ taková, jejichž relativní hvězdná velikost klesla pod 21,5 mag. Posledním kritériem výběru je velikost vektoru rychlosti, která musí ležet v intervalu $\langle 0,05; 0,5 \rangle^\circ \cdot \text{den}^{-1}$. Tímto postupem se podařilo najít 95% ze 100% zachycených planetek. Objektů, které nejsou planetkami, a přesto „sítém“ výběru do katalogu MOC propadly, je asi 6%. Převědno v cifry, získáváme 471 569 planetek a 220 101 „pravděpodobných“ planetek, ale s neznámými parametry drah.

5.2.1 Uspořádání katalogu MOC

Katalog lze stáhnout jako ASCII soubor (o velikosti 219 MB). Data jsou vyrovnána do řádků. Každý řádek náleží jednomu pozorovanému objektu. Každý řádek obsahuje parametry vypsané v následných tabulkách tab.13 až 19.

Prázdné hodnoty jakéhokoliv pole byly nahrazeny podtržítkem. Taková úprava usnadňuje programům čtení záznamu. Pokud hodnota nějaké veličiny resp. chyby, byla změřena nepřesně, nabývá v katalogu taková hodnota velikosti 99,99 resp. 9,99 podle definice.

⁹jedná se o matematické nebo

Obr. 13: Všeobecné identifikátory.

sloupec	počet znaků	označení	popis
1	8	moID	Katalogové číslo objektu.
2	4	Run	Číslo observačního běhu.
3	2	Col	Sloupec oblohy, který byl pozorován
4	5	Field	Ukazatel políčka na obloze, ve kterém byl objekt pozorován.
5	5	Object	Identifikátor objektu. 0 – neznámý, 1 – kosmické záření, 2 – defekt, 3 – galaxie, 4 – odlesk, 5 – známý, 6 – hvězda, 7 – stopa hvězdy, 8 – type sky.
6	9	rowc	Řádková souřadnice středu objektu na fotografii (pixely).
7	9	colc	Sloupečková souřadnice středu objektu na fotografii (pixely).

Obr. 14: Astrometrické veličiny.

sloupec	počet znaků	označení	popis
8	12	Time (MJD)	Čas (modifikované juliánské datum) pozorování (střední hodnota z celkové expozice).
9	11	R.A.	Rektascenze objektu, v části čipu s červeným filtrem.
10	11	Dec	Deklinace objektu, v části čipu s červeným filtrem.
11	11	λ	Ekliptikální délka v okamžiku pozorování MJD.
12	11	β	Ekliptikální šířka v okamžiku pozorování MJD.
13	12	Φ	Vzdálenost od místa opozice.
14 + 15	8+7	$v_{\mu}, v_{\mu\text{error}}$	Velikost složky rychlosti rovnoběžné se směrem pohybu, vykonávaném dalekohledem během pozorování, a její chyba.
16 + 17	8+7	$v_{\nu}, v_{\nu\text{error}}$	Velikost kolmé složky rychlosti na směr pohybu dalekohledu během pozorování, a její chyba.
18	8	v_{λ}	Složka rychlosti rovnoběžná s ekliptikou ($^{\circ} \cdot \text{den}^{-1}$)
19	8	v_{β}	Složka rychlosti kolmá na ekliptiku ($^{\circ} \cdot \text{den}^{-1}$)

Obr. 15: Fotometrické veličin

sloupec	počet znaků	označení	popis
20 + 21	6	u, u_{error}	Hvězdná velikost ve filtru u a její chyba.
22 + 23	6	g, g_{error}	Hvězdná velikost ve filtru g a její chyba.
24 + 25	6	r, r_{error}	Hvězdná velikost ve filtru r a její chyba.
26 + 27	6	i, i_{error}	Hvězdná velikost ve filtru i a její chyba.
28 + 29	6	z, z_{error}	Hvězdná velikost ve filtru z a její chyba.
30 + 31	6	a, a_{error}	$a^{*10} = 0,85(g-r) + 0,45(r-i) + 0,57$. a^* je zavedeno tak, že planetky s $a^* > 0$ jsou červené a planetky s $a^* < 0$ jsou modré.
32	6	V	Absolutní velikost planety v oboru V — Johnsonova systému vypočtená podle $V = r + 0,44(g - r) - 0,02 \text{ mag}$
33	6	B	Absolutní velikost planety v pásu B — Johnsonova systému vypočtená podle $B = V + 1,44(g - r) - 0,19 \text{ mag}$

Obr. 16: Identifikátory objektu

sloupec	počet znaků	označení	popis
34	2	identifikation flag	pokud byl objekt rozpoznán jako známá planetka, nabývá příslušnost hodnoty 1, v opačném případě 0.
35	8	Numeration	Katalogové číslo planetky. Pokud objekt nebyl rozpoznán jako planetka, obdržel v tomto sloupci nulu.
36	11	Designation	Objekty dosud nepojmenované nebo neidentifikované jako planetky se označují znakem "-".
37	3	Designation counter	Každou další planetkou, objevenou v rámci SDSS, vzroste hodnota čítače o jednu.
38	3	Total Detection Count	Celkový počet pozorování planetky v rámci SDSS.
39	9	Flags	Interní označení objektu pro správce projektu SDSSMOC.

Obr. 17: Porovnávací identifikátory pro planetky

sloupec	počet znaků	označení	popis
40	11	R.A. (calc.)	Předpovězená rektascenze objektu během pozorování.
41	11	Dec (calc.)	Předpovězená deklinace objektu během pozorování.
42	6	Mag (calc.)	Vypočtená hvězdná velikost.
43	8	R_{\odot}	Vzdálenost od Slunce v okamžiku pozorování.
44	8	R_{\oplus}	Vzdálenost od Země v okamžiku pozorování.
45	6	Phase	Fáze.

Obr. 18: Elementy drah

sloupec	počet znaků	označení	popis
46	11		Jméno katalogu, ze kterého byly dráhové elementy okopírovány.
47	6	H	Absolutní hvězdná velikost.
48	5	G	Popisuje jak se mění magnituda v závislosti na změně osvětlení planety.
49	6	Arc	Oblouk, rozsah dní pokrývajících pozorování planety, za účelem stanovení dráhových elementů.
50	14	Epoch	
51	13	a	
52	11	e	
53	11	i	
54	11	Ω	
55	11	ω	
56	11	M	

Pozorovatel sleduje soustavu ze záporné části osy z . Základní rovina je kolmá na z

Obr. 19: Vlastní elementy drah planetek

sloupec	počet znaků	označení	popis
57	11	Cat ID	Pojmenování katalogu, ze kterého byly vlastní elementy okopírovány
58	13	a'	Vlastní velká poloosa
59	11	e'	Vlastní excentricita
60	11	$\sin i'$	Sklon
61	128		Binary processing flags

5.3 Výpočet hlavních komponent

Doposud jsme vysvětlili, co a jakým mechanismem lze pomocí metody hlavních komponent spočítat. Zbývá uvést metodu v život a aplikovat ji na konkrétní data.¹¹ Ovšem pozor! Není to tak, že bychom se nejprve rozhodli pohovořit o PCA a poté vyhledali vhodnou datovou množinu, na které bychom PCA prakticky předvedli. Realita věci je přesně opačná. Na počátku dalekohled „sloanské“ přehlídky zaznamenal hvězdnou velikost 691670 planetek v pěti filtrech. V takto ohromném souboru dat je jistě skryta spousta vztahů a závislostí mezi hodnotami jednotlivých hvězdných velikostí planetek i jejich kombinacemi. Volba matematické metody, která by vnitřní strukturu v množině dat mohla objevit, byla provedena následně.

Původní katalog (ADR4.dat) je popsán v kapitole 5.1. Byl stažen ze serveru astro.washington.edu kompletní link viz [16]. Dalším logickým krokem bylo vybrání pouze pěti sloupečků ze souboru ADR4.dat, jenž v sobě obsahovaly jednotlivé hvězdné velikosti.

Zároveň jsme z pěti hvězdných velikostí pro každou planetku, spočítali čtyři barevné indexy, čímž se odstranila rozmanitost ve velikosti magnitud způsobená různou vzdáleností planetek od Země a od Slunce.

Při dalším zpracování jsme si museli uvědomit, že planetky nezáří, nýbrž odrážejí světlo přicházející od Slunce. Pokud by se táž planetka znenadání ocitla v blízkosti jiného zdroje světla než Slunce, tak i přesto, že její odrazivé vlastnosti (složení povrchu) by zůstaly nezměněné, hodnoty hvězdných velikostí v jednotlivých filtrech by se změnily. Proto je nezbytné „oprostit“ planetku od zkreslující informace nesené světlem Slunce, a od jednotlivých barevných indexů odečíst hodnotu příslušného barevného indexu Slunce. Všechny tři popsané operace jsme provedli pomocí skriptu AWK, spuštěného příkazem:

```
./sloan-read.awk < ADR4.dat > sloan-read.out
```

Dalším nezbytným krokem bylo vymazání záznamů s velkou chybou měření. Pokud uvážíme, že typické rozdíly mezi reflekcními spektry různých planetek stejného typu (obr. 21) se nacházejí v intervalu jedné desetiny, lze označit hvězdné velikosti měřené s chybou větší než 0,2 mag za málo průkazné a ze souboru je nutné odstranit. Uvedené chyby byly určeny při redukci snímků a na jejich velikost má vliv Poissonův šum zdroje, šum od oblohy, šum temného snímku, vyčítací a šum od CCD kamery. Data s velkou chybou se chovají jako náhodná veličina a nemá smysl provádět s nimi výpočty.

Použili jsme následující skript:

```
./err-range.awk < sloan-read.out > err-range.out
```

Doposud popsané kroky vedou k redukci původního souboru čítajícího 471 569 planetek na soubor o 70 887 členech.

Protože data, která analyzujeme, jsou čtyřrozměrná, jsme v průběhu výpočtu nuceni řešit polynom čtvrtého stupně a to je úkol značně obtížný. Abychom si práci

¹¹Veškeré výpočty jsou realizovány v Linuxu. Programy i skripty jsou však přenositelné a je možné je použít i ve Windows

ulehčili, sáhli jsme po subrutinách publikovaných v Numerical Recipes [15] a vytvořili jednoduchý program v jazyku Fortran, který je postupně volá. Program se nazývá `eigen.f`. Vstupní soubor `eigen.in` obsahuje rozměr dat se kterými operujeme a kovarianční matici. Obsah výstupního souboru `eigen.out` je při pohledu na jeho výpis zřejmý:

```
# pocet prvku n =
      4
# symetricka matice a(n,n) =
  5.95948622E-02  9.38026514E-03  1.71851907E-02  1.93552729E-02
  9.38026514E-03  4.45557572E-02  4.47760150E-02  3.77184525E-02
  1.71851907E-02  4.47760150E-02  7.05474243E-02  5.98190278E-02
  1.93552729E-02  3.77184525E-02  5.98190278E-02  8.55809450E-02
# vlastni cisla matice a():
  0.17270036
  5.31925820E-02
  2.49198601E-02
  9.46618244E-03

# vlastni vektory matice a() v~radcich:
  0.23546046      0.41565195      0.59843224      0.64317226
  0.96814215     -0.17290485     -0.14676926     -0.10612940
  0.07798612      0.60094780      0.32988197     -0.72384918
 -0.03431175     -0.66045374      0.71520305     -0.22607116
```

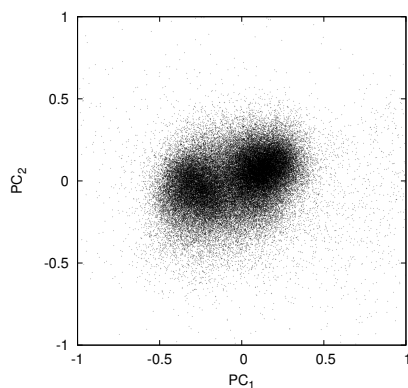
Nakonec postupným spuštěním programů:

```
./kovariancni_matice < kovariancni_matice.in
./eigen < eigen.in
./hlavni_komponenta < hlavni_komponenta.in
```

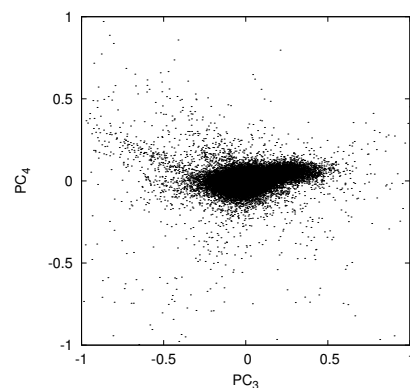
obdržíme soubor se čtyřmi sloupci, přičemž v prvním jsou hlavní komponenty PC_1 příslušející největšímu vlastnímu číslu v druhém jsou hlavní komponenty PC_2 příslušející druhému největšímu vlastnímu číslu atd. Zdrojový kód použitých programů je uveden v příloze A.

závisejí na velikostech a vzdálenostech planetek. Zobrazení normalizovaných spekter se tak výrazně zpřehlední.

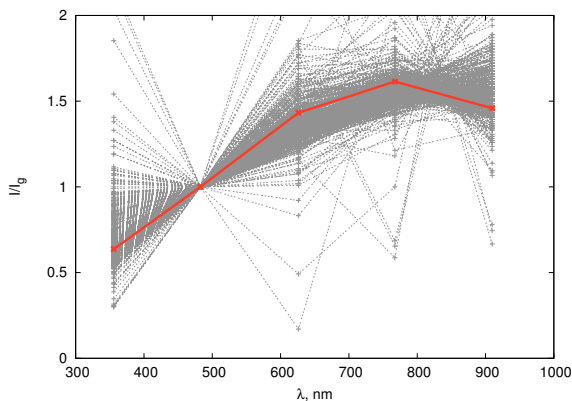
Hlavní komponenta (PC_1) je určitá vlastnost planetky. Planetky s velkou hodnotou hlavní komponenty by měly být odlišné od planetek s malou hodnotou hlavní komponenty. Tuto odlišnost kvantitativně popisuje právě hlavní komponenta.¹² Použitelnost hlavních komponent, tedy jejich schopnost kvalitativně rozlišit jednotlivé planetky ověříme například sestrojením grafu s PC_1 a PC_2 na osách pro planetky známých typů (viz [12], obr. 20, 22). Jak vidíme planetky určitých typů vytvářejí skupiny v určitých oblastech grafu. Kritérium použitelnosti hlavních komponent je tedy naplněno. Grafy spektrálního průběhu sestrojujeme pro planetky s velkou a malou hodnotou PC_1 nebo PC_2 (obr. 24 a 25).



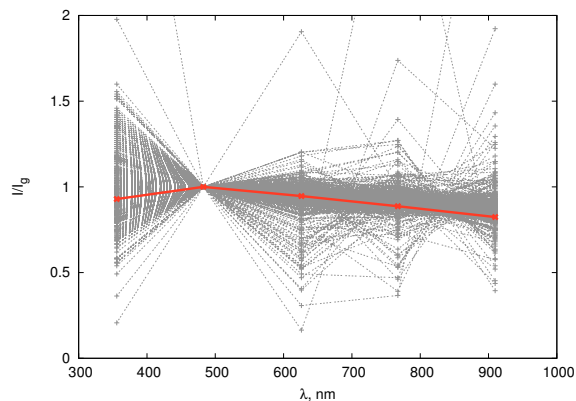
Obr. 22: Graf závislosti mezi první a druhou komponentou. Vidíme větší rozptyl komponenty PC_1 .



Obr. 23: Graf závislosti mezi třetí a čtvrtou komponentou. Rozptyl čtvrté hlavní komponenty je oproti třetí komponentě výrazně menší.

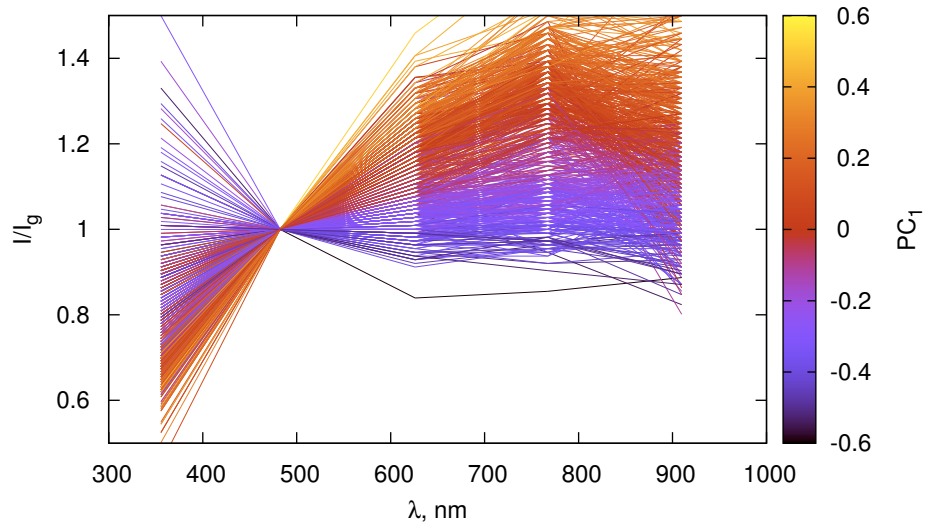


Obr. 24: Spektrální průběhy planetek s velkou hodnotou PC_1 ($PC_1 \in (0,4; 0,5)$). Typická hodnota průběhu je zvýrazněna tlustou čarou.



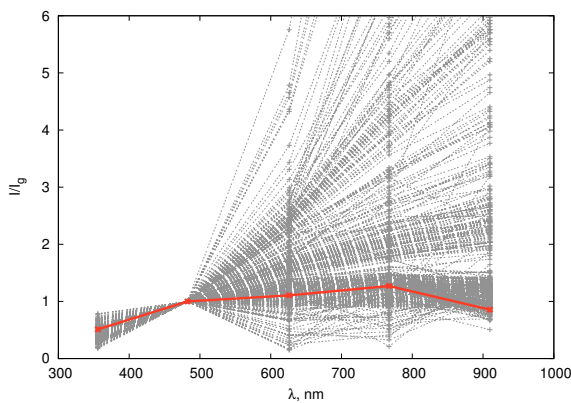
Obr. 25: Spektrální průběhy planetek s malou hodnotou PC_1 ($PC_1 \in (-0,7; -0,54)$). Typická hodnota průběhu je zvýrazněna tlustou čarou.

¹²Obdobně jako odlišnost mezi tenisovým míčkem a medicinbalem popisuje například hmotnost.

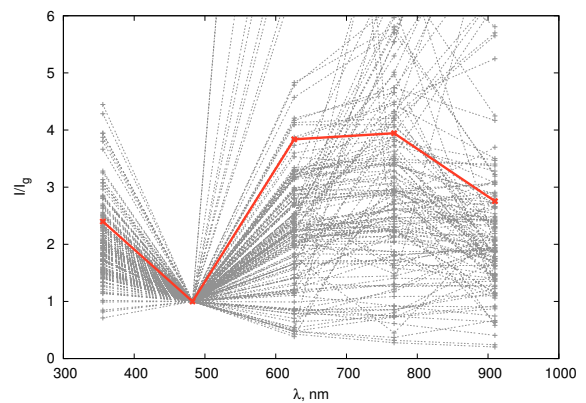


Obr. 26: Spektrální průběhy náhodného výběru ze všech planetek. Hlavní komponenty PC_1 jsou zobrazeny pomocí barevné škály.

Při pohledu na grafy je zřejmé, že planetky s malou hodnotou PC_1 mají spektrum spíše ploché. Naopak planetky s velkou hodnotou PC_1 značně skloněné (říkáme též zčervenalé). Ploché spektrum, tak jak ho vidíme na obr. 24, je charakteristické pro planetky taxonomického typu C. Skloněná spektra (obr. 25) odpovídají základnímu taxonomickému typu S. Známé tvary spekter jednotlivých taxonomických typů planetek jsou na obr. 21. Sklon spektra je závislý na povrchových vlastnostech planetek a odpovídá odrazivosti minerálů.



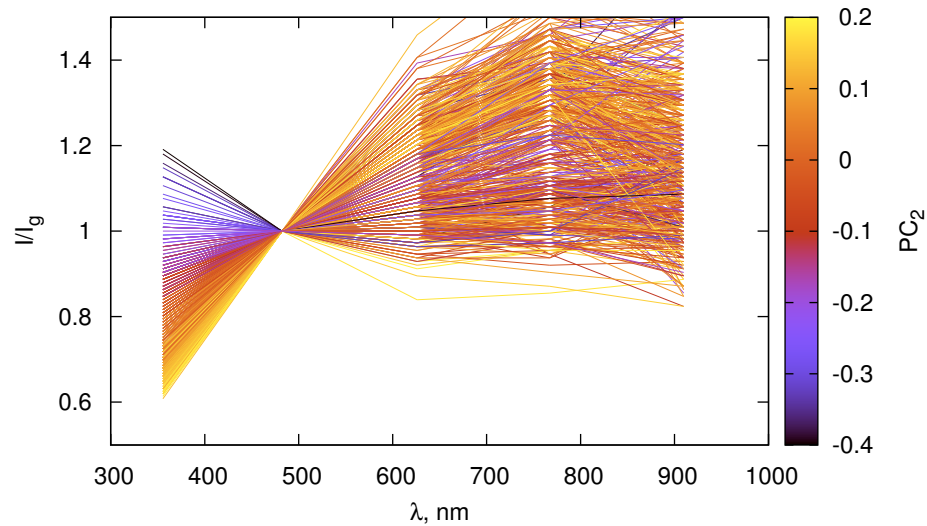
Obr. 27: Spektrální průběhy planetek s malou hodnotou PC_2 ($PC_2 \in (0,4; 0,5)$). Typická hodnota průběhu je zvýrazněna tlustou čarou.



Obr. 28: Spektrální průběhy planetek s velkou hodnotou PC_2 ($PC_2 \in (-1,2; -0,9)$). Typická hodnota průběhu je zvýrazněna tlustou čarou.

PC_2 , druhá hlavní komponenta, je veličinou, pro jejíž hodnoty data nabývají také

značného rozptylu, ovšem již menšího než pro PC_1 . Veličina, kterou PC_2 zastupuje, je pro rozlišení jednotlivých planetek a skupin planetek méně průkazná. Podobně jako výška člověka je méně významnějším rozlišovacím kritériem než například barva jeho pleti.



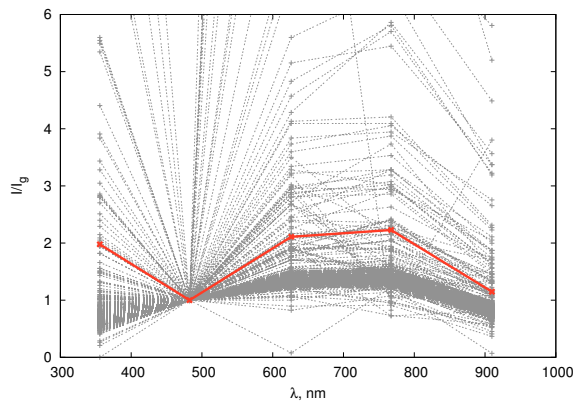
Obr. 29: Spektrální průběhy náhodného výběru ze všech planetek. Hlavní komponenty PC_2 jsou zobrazeny pomocí barevné škály.

Grafy na obr. 27, 28 a 29 ukazují tvar spektra pro velkou a malou hodnotu PC_2 . I v tomto případě nacházíme jistou odlišnost. Spektrum planetek, jimž přísluší nízké hodnoty PC_2 zachovává sklon a ve svém průběhu se nijak zásadně nelomí. Naproti tomu pro planetky s velkou hodnotou PC_2 lze identifikovat pokles spektrálního průběhu v blízkosti ultrafialové oblasti nebo pokles v oblasti 770 nm. Jednotlivé filtry integrují intenzity vlnových délek v příslušných šířkách jejich propustnosti. Určit přesnou hodnotu vlnové délky na které k poklesu dochází, je vzhledem k tomu nemožné. V práci [12] je tento pokles v průběhu spektra identifikován jako přítomnost absorpčního pásu na 1 000 nm. Absorpci způsobují minerály pyroxen a olivín, které se nacházejí na povrchu kamenných planetek.¹³

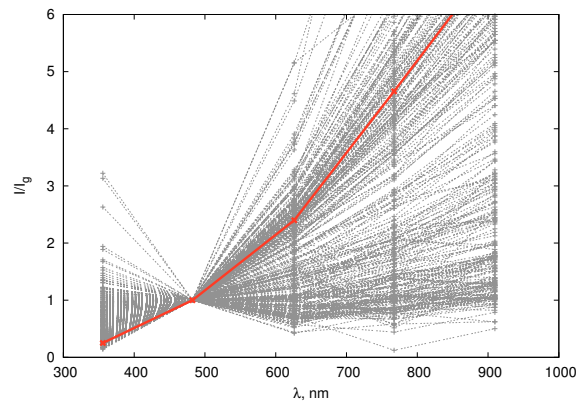
Ještě menšího rozptylu nabývají hodnoty třetí hlavní komponenty. Podle velikosti vlastních čísel, příslušejících druhé a třetí komponentě (0,05319 pro PC_2 a 0,02492 pro PC_3) můžeme rozhodnout, že významnost obou komponent je podobná. Sestrojíme proto stejné grafy i pro třetí hlavní komponentu PC_3 (obr. 30, 31, 32).

Při prvním pohledu na odkazované grafy si okamžitě všimneme již diskutovaného poklesu spektrálního průběhu okolo hodnoty 1 000 nm. Ba co více, grafy příslušející třetí

¹³Vlivem dlouhodobé interakce mezi povrchem hornin a slunečním nebo kosmickým zářením dochází k přetváření povrchu planetek a postupným změlčováním až vyhlazením absorpce na 1 000 nm. Tomuto procesu se říká kosmické zvětrávání.



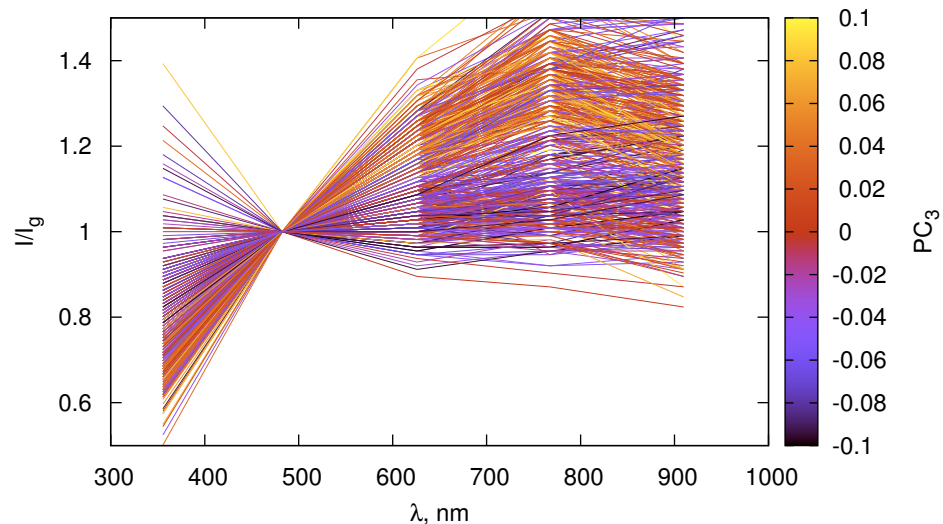
Obr. 30: Spektrální průběhy planetek s malou hodnotou PC_3 . Typická hodnota průběhu je zvýrazněna tlustou čarou.



Obr. 31: Spektrální průběhy planetek s velkou hodnotou PC_3 . Typická hodnota průběhu je zvýrazněna tlustou čarou.

hlavní komponentě zobrazují tento pokles mnohem lépe než grafy zkonstruované pro druhou hlavní komponentu. Výsledkem naší analýzy je tedy identifikace třetí hlavní komponenty PC_3 jako míry absorpce na 1000 nm. Co je velmi překvapivé, lišíme se v tomto od práce [12]. Rozdílnost v interpretaci druhé, respektive třetí hlavní komponenty by mohla vznikat v důsledku různého počtu analyzovaných planetek. Zatímco Nesvorný et al. pracoval s 7 593 planetkami, my jsme jich použili 70 894.

Tento výsledek, tedy fyzikální interpretace tří hlavních komponent namísto dvou, je užitečný pro další studium planetek a jejich rodin.

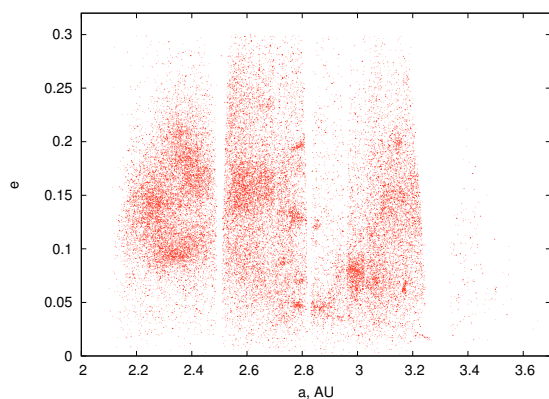


Obr. 32: Spektrální průběhy výběru ze všech planetek. Hlavní komponenty PC_3 jsou zobrazeny pomocí barevné škály.

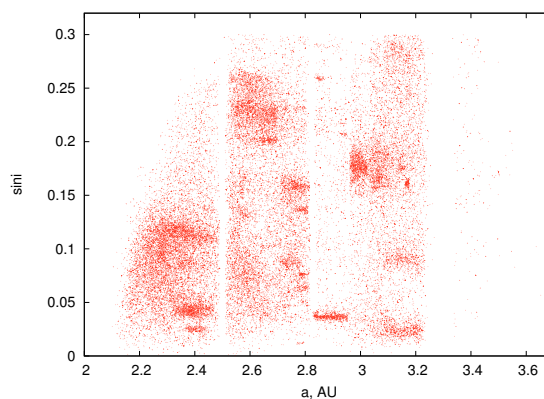
6.2 Hlavní komponenty a dráhové elementy

Planetky s podobnými dráhovými elementy náleží do rodin, tedy skupin kolizního původu, s podobným stářím a složením (viz [6]). Rodiny vznikají rozpadem větších těles. Postupně vzájemnými srážkami dochází k fragmentaci jejich členů. Rodiny planetek lze s úspěchem zobrazit pomocí grafů, na jejichž osách jsou vyneseny vlastní (středované) dráhové elementy. Jednotlivé členy rodiny mají podobné elementy drah a na grafech se „prozradí“ nápadným shlukem, viz obr. 34, obr. 33. I uvnitř jednotlivých shluků, tedy planetek na podobné dráze ve sluneční soustavě, se však mohou skrývat spektroskopicky (barevně) značně odlišné skupiny. Rozlišit detailněji jednotlivé shluky můžeme například právě pomocí hlavních komponent. Grafy (a, e) , $(a, \sin i)$ jsme doplnili o třetí souřadnici obarvením hodnotou hlavních komponent PC_1 a nebo PC_2 , viz obr. 35 až 38.

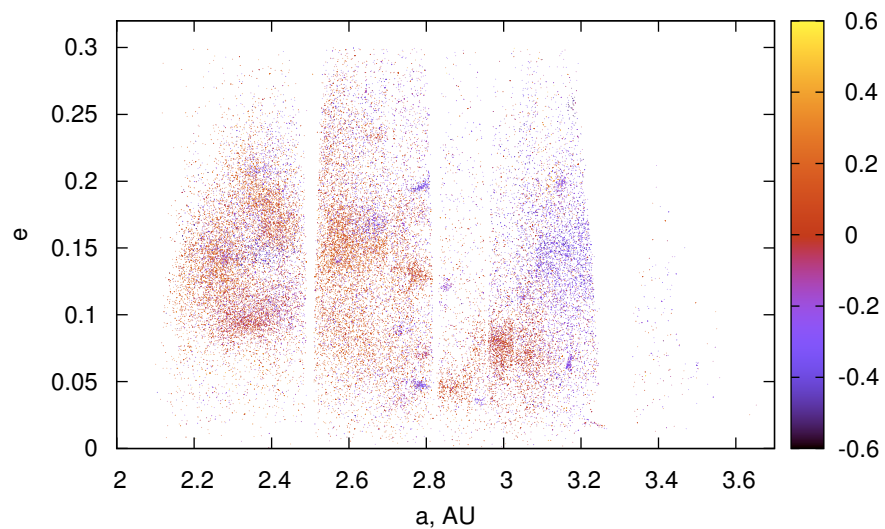
Obdobně jako v práci [6] je zřetelné, že dynamické rodiny planetek mají také podobné hlavní komponenty neboli spektrální vlastnosti.



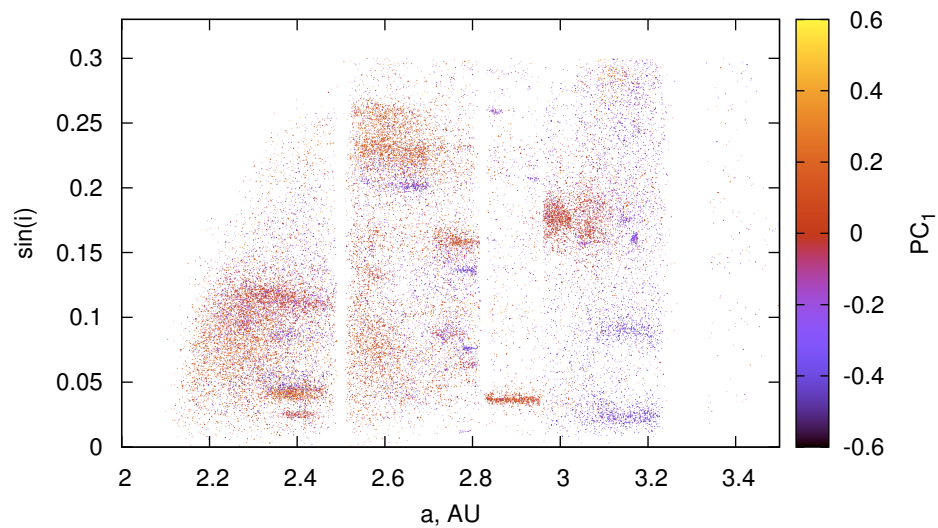
Obr. 33: Rozložení 70 887 planetek rodin hlavního pásu a planetek pozadí. Zobrazeno v prostoru excentricity a hlavní poloosy.



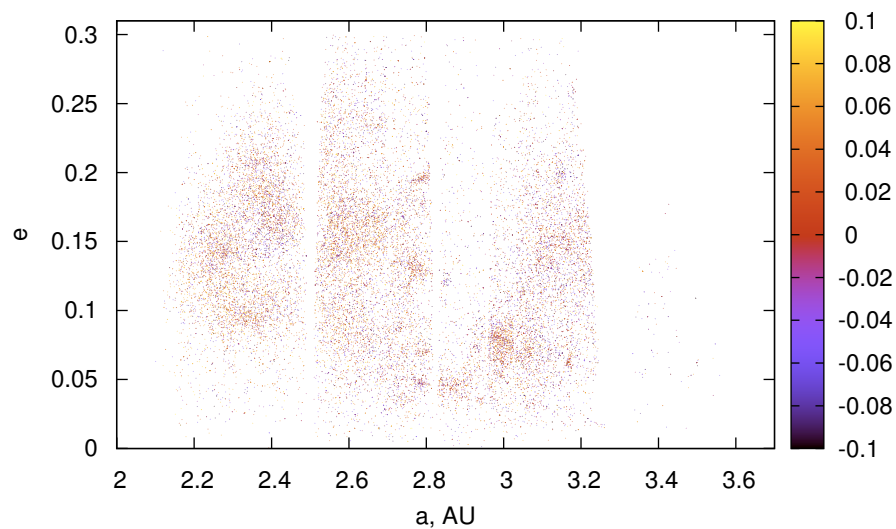
Obr. 34: Rozložení 70 887 planetek rodin hlavního pásu a planetek pozadí. Zobrazeno v prostoru \sin sklonu a hlavní poloosy.



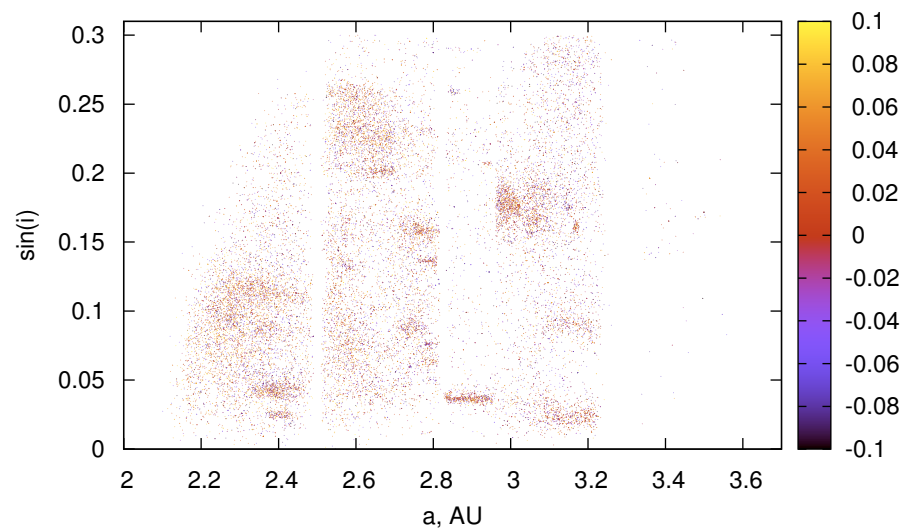
Obr. 35: Rozložení planetek hlavního pásu a pozadí podle velké poloosy a excentricity. Třetím (barevným) rozměrem je PC_1



Obr. 36: Rozložení planetek hlavního pásu a pozadí podle velké poloosy a sklonu. Třetím (barevným) rozměrem je PC_1



Obr. 37: Rozložení planetek hlavního pásu a pozadí podle velké poloosy a excentricity. Třetím (barevným) rozměrem je PC_2



Obr. 38: Rozložení planetek hlavního pásu a pozadí podle velké poloosy a sklonu. Třetím (barevným) rozměrem je PC_2

7 Závěr

Podali jsme přehled základních matematických poznatků o metodě hlavních komponent a předvedli ji na jednoduchém cvičném příkladu. Pro potřebu analýzy objemných datových souborů jsme naprogramovali nezbytné algoritmy v jazyce Pascal, pomocí nichž jsme aplikovali metodu PCA na fotometrická data ze Sloan Digital Sky Survey. Tímto jsme vytvořili dostatečně obecný výpočetní postup pro analýzu v budoucnu aktualizovaných souborů dat. Fyzikálně jsme interpretovali hlavní komponenty jako sklon spektra, pokles v ultrafialové oblasti a pokles v infračervené oblasti. Zobrazili jsme hlavní komponenty společně s dráhovými elementy planetek a potvrdili tak souvislost spektroskopických a orbitálních vlastností planetek. Ukázalo se, že větší množství analyzovaných dat může zásadně změnit pohled na význam hlavních komponent, proto by mohlo být v budoucnu velmi zajímavé, obdobně zpracovat ještě větší množství dat.

A Kódy programů

Výpis programu kovariancni_matice.pas napsaného v jazyce Pascal:

```
{
  kovariancni_matice.pas
  Vypocet kovariancni matice pro data nactena ze souboru.
  Lenka Trojanova, 5. 2. 2010
}

program kovariancni_matice (input,output);

const
  maxn = 100000; { maximalni pocet radku }
  maxm = 4; { maximalni pocet sloupcu }

type
  TwoDimensionalArray = array [1..maxn, 1..maxm] of real;

var
  n, m, debug: integer;           { rozmery matic }
  filename: string(255);         { jmeno vstupniho souboru }
  MD, MDR, CovM: TwoDimensionalArray; { MatrixDate, MatrixDateReduced,
                                       CovarianceMatrix }

{ Nacteni dat ze souboru. }

procedure DataRead(filename: string; var MD: TwoDimensionalArray;
  var n,m: integer);
var
  i,j: integer;
  SDC: text;
begin
  assign(SDC, filename);
  reset(SDC);
  i:=1;
  j:=1;
  n:=0;
  m:=0;
  while not Eof(SDC) do begin
    read(SDC, MD[i,j]);
    j:=j+1;
    if (j>maxm+1) then begin
      writeln("Data v souboru prekracuji rozmer matice!");
      halt;
    end;
    if Eoln(SDC) then begin
      m:=j-1;
      i:=i+1;
      if (i>maxn) then begin
```

```

        writeln("Data v~souboru prekracuji rozmer matice!");
        halt;
    end;
    j:=1;
    readln(SDC);
    end;
end;
n:=i-1;
end;

{ Vypocet aritmetického prumeru; j jsou sloupecky. }

function Mean(var MD: TwoDimensionalArray; n, j: integer): real;
var
    Sum: real;
    i: integer;
begin
    Sum:=0;
    for i:=1 to n do begin
        Sum := Sum + MD[i,j];
    end;
    Mean := Sum/n;
end;

{ Odecteni aritmetickych prumeru. }

procedure MinusMean(MD: TwoDimensionalArray; var MDR: TwoDimensionalArray;
    n,m: integer);
var
    i,j: integer;
    Prum: real;
begin
    for j:=1 to m do begin
        Prum := Mean(MD, n, j);
        for i:=1 to n do begin
            MDR[i,j] := MD[i,j] - Prum;
        end;
    end;
end;

{ Vypocet kovariance. }

function Cov(var MDR: TwoDimensionalArray; n, p, q: integer): real;
var
    i: integer;
    Sum: real;
begin
    Sum:=0;

```

```

    for i:=1 to n do begin
        Sum := Sum + MDR[i,p] * MDR[i,q];
    end;
    Cov := Sum/(n-1);
end;

{ Vypocet kovariancni matice. }

procedure CovMatrix(var MDR, CovM: TwoDimensionalArray; n,m: integer);
var
    i,j: integer;
begin
    for i:=1 to m do begin
        for j:=i to m do begin
            CovM[i,j] := Cov(MDR, n, i, j);
            CovM[j,i] := CovM[i,j];           { matice je symetricka }
        end;
    end;
end;

{ Vypis matice na standardni vystup. }

procedure PrintMatrix(var Matrix: TwoDimensionalArray; n,m: integer);
var
    i,j: integer;
begin
    for i:=1 to n do begin
        for j:=1 to m do begin
            write(Matrix[i,j], ' ');
        end;
        writeln;
    end;
    writeln;
end;

begin
    writeln("# Zadej jmeno vstupniho souboru: ");
    readln(filename);
    writeln("# Pro vypis matice dat zadej 1, jinak 0:");
    readln(debug);
    readln;

    DataRead(filename,MD,n,m);

    writeln("# filename = ", filename);
    writeln("# n = ", n);
    writeln("# m = ", m);

```

```

if debug = 1 then begin
  writeln("# MatrixDate = ");
  PrintMatrix(MD,n,m);
end;

MinusMean(MD,MDR,n,m);

if debug = 1 then begin
  writeln("# MatrixDateReduced = ");
  PrintMatrix(MDR,n,m);
end;

CovMatrix(MDR,CovM,n,m);

writeln("# CovarianceMatrix = ");
PrintMatrix(CovM,m,m);
end.

```

Výpis programu vlastni_vektory.pas:

```

{
  vlastni_vektory.pas
  Vypocet vlastnich vektoru kovariancni matice rozmeru 2 x 2.
  Lenka Trojanova, 7. 2. 2010
}

program vlastni_vektory (input,output);

const
  maxm = 2; { rozmer kovariancni matice }

type
  TwoDimensionalArray = array [1..maxm, 1..maxm] of real;
  OneDimensionalArray = array [1..maxm] of real;

var
  CovM: TwoDimensionalArray;
  Eigenvector: OneDimensionalArray;
  i, m: integer;
  a, b, c, lambda1, lambda2: real;

{ Reseni kvadraticke rovnice. }

procedure kvadraticka_rce (a,b,c: real; var x1,x2: real);
var
  D: real;
begin
  D := sqr(b)-4*a*c;
  if (D >= 0) then begin

```

```

    x1 := (-b + sqrt(D))/2;
    x2 := (-b - sqrt(D))/2;
end else begin
    writeln("Diskriminant je mensi nez nula!");
    halt;
end
end;

{ Koeficienty charakteristickeho polynomu. }

procedure charakteristica_rce (var CovM: TwoDimensionalArray;
    var a,b,c: real);
begin
    a~:= 1;
    b := -CovM[1,1] - CovM[2,2];
    c := CovM[1,1]*CovM[2,2] - CovM[1,2]*CovM[2,1];
end;

{ Gaussova eliminacni metoda (vcetne prave strany a~overeni singularity). }

procedure Gauss (var a: TwoDimensionalArray; var b: OneDimensionalArray;
    n: integer);
const
    eps = 1.0e-15; { "nula" }
var
    i,j,k: integer;
    cij: real;
begin

    for i:=1 to n-1 do begin { cyklus pres radky }

{ vymena i-teho a~k-teho radku, na nemz se nachazi nejvetsi prvek
v~i-tem sloupci }

        k:=i;
        for j:=i+1 to n do begin
            if (abs(a[i,j]) > abs(a[i,k])) then begin
                k:=j;
            end;
        end;

        if (k<>i) then begin
            for j:=i to n do begin
                cij := a[j,i];
                a[j,i] := a[j,k];
                a[j,k] := cij;
            end;
            cij := b[i]; { s~pravou stranou delame tytez upravu }
        end;
    end;
end;

```

```

        b[i] := b[k];
        b[k] := cij;
    end;

{ diagonalizace matice }

    for j:=i+1 to n do begin
        if (abs(a[i,i]) > eps) then begin
            cij := a[i,j]/a[i,i];
        end else begin
            writeln("Matice je singularni!");
            halt;
        end;
        for k:=i to n do begin
            a[k,j] := a[k,j] - cij*a[k,i];
        end;
        b[j] := b[j] - cij*b[i];
    end;

end; { cyklu pres i~}

{ zpetne dosazeni }

    for i:=n downto 1 do begin
        for j:=i+1 to n do begin
            b[i] := b[i] - a[j,i]*b[j];
        end;
        if (abs(a[i,i]) > eps) then begin
            b[i] := b[i]/a[i,i];
        end else begin
            writeln("Matice je singularni!");
            halt;
        end;
    end;

end; { Gauss }

{ Normalizace vektoru. }

procedure normalizace(var v: OneDimensionalArray; n: integer);
var
    i: integer;
    norm: real;
begin
    norm:=0;
    for i:=1 to n do begin
        norm := norm + sqr(v[i]);
    end;

```

```

if (norm>0) then begin
  norm:=1/sqrt(norm);
  for i:=1 to n do begin
    v[i] := v[i]*norm;
  end;
end else begin
  writeln("Nulovy vektor nelze normalizovat!");
end;
end;

{ Vypocet vlastniho vektoru. }

procedure vlastni_vektor(var CovM: TwoDimensionalArray; m: integer;
  lambda: real; var Eigenvector: OneDimensionalArray);
var
  i,j: integer;
  a: TwoDimensionalArray;
  b: OneDimensionalArray;
begin
  { matice je singularni => jednu slozku vektoru musime zvolit! }
  { Eigenvector[2] := 1; }
  { pro 2 x 2 by reseni soustavy bylo jednoduche... (bez GEM) }
  { Eigenvector[1] := CovM[1,2]*Eigenvector[2]/(CovM[1,1]-lambda); }

  { obecnejsi reseni pro rozmer m x m }
  { => vzhledem k~singularite vezmeme pouze podmatici (m-1) x (m-1) }
  for i:=1 to m-1 do
    for j:=1 to m-1 do
      a[i,j] := CovM[i,j];

  { odedcteni lambda na diagonale }
  for i:=1 to m-1 do
    a[i,i] := a[i,i] - lambda;

  { "zbytek" bude tvorit pravou stranu <= jednu slozku vektoru musime zvolit! }
  Eigenvector[m] := 1;
  for j:=1 to m-1 do
    b[j] := -CovM[m,j]*Eigenvector[m];

  Gauss(a,b,m-1);

  for j:=1 to m-1 do
    Eigenvector[j] := b[j];

  normalizace(Eigenvector, m);
end;

begin

```



```

writeln("# Zadej kovariancni matici 2 x 2: ");

m:=2;
readln(CovM[1,1], CovM[1,2]);
readln(CovM[2,1], CovM[2,2]);

charakteristicka_rce(CovM, a,b,c);

kvadraticka_rce(a,b,c, lambda1,lambda2);

writeln("# lambda1 = ", lambda1);
writeln("# lambda2 = ", lambda2);

vlastni_vektor(CovM, m, lambda1, Eigenvector);

writeln("# vlastni vektor prislusny lambda1 = ");
writeln(Eigenvector[1], " ", Eigenvector[2]);

vlastni_vektor(CovM, m, lambda2, Eigenvector);

writeln("# vlastni vektor prislusny lambda2 = ");
writeln(Eigenvector[1], " ", Eigenvector[2]);
end.

```

Výpis programu hlavni_komponenta.pas:

```

{
  hlavni_komponenty.pas
  Vypocet hlavnich komponent ze zadanych vlastnich vektoru.
  Lenka Trojanova 30. 3. 2010
}

program hlavni_komponenty (input,output);

const
  maxn = 100000; {maximalni pocet radku}
  maxm = 4; {maximalni rozmer dat}

type
  VectorArray = array [1..maxm, 1..maxm] of real;
  TwoDimensionalArray = array [1..maxn, 1..maxm] of real;

var
  Eigenvectors:VectorArray;
  MD, MDR, PC: TwoDimensionalArray;
  i,j,n,m,m_vectors: integer;
  filename: string(255);

{ nasledujici tri podprogramy jsou stejne jako v~kovariancni_matice.pas }

```

```

procedure DataRead
function Mean
procedure MinusMean

procedure PrincipalComponent(var MDR: TwoDimensionalArray; Eigenvectors:
  VectorArray; m, m_vectors,n:integer; var PC: TwoDimensionalArray);
var
  i,j,k: integer;
  PCpom: real;
begin
  for i:=1 to n do begin
    for j:=1 to m_vectors do begin
      PCpom:=0;
      for k:=1 to m do begin
        PC[i,j]:=PCpom + Eigenvectors[j,k] * MDR[i,k];
        PCpom:=PC[i,j];
      end;
    end;
  end;
end;

begin
  writeln("zadej rozmer dat. Maximalne vsak ,'maxm',>");
  readln(m);
  writeln("zadej pocet vlastnich vektoru, pro ktere budu pocitat hlavni
komponenty Maximalne vsak ,'maxm',>");
  readln(m_vectors);

  writeln("zadej vlastni vektory. Vzdy jeden vektor na jeden radek.");
  for i:=1 to m_vectors do begin
    for j:=1 to m do begin
      read(Eigenvectors[i,j]);
    end;
    readln;
  end;

  writeln('zadej jmeno vstupniho souboru'); { souradnice jednoho objektu }
  read(filename);

  DataRead(filename,MD,n,m);

  writeln("# filename = ", filename);
  writeln("# n = ", n);
  writeln("# m = ", m);

  MinusMean(MD,MDR,n,m);

```

```
writeln("souradnice v~hlavnich komponentach:");
PrincipalComponent(MDR, Eigenvectors,m, m_vectors,n,PC );

for i:=1 to n do begin
  for j:=1 to m_vectors do begin
    write(PC[i,j]," ");
  end;
  writeln;
end;
end.
```

B Použití algoritmu z Numerical Recipes

Některé složitější algoritmy lze převzít z literatury. Velmi často používaným zdrojem je kniha Numerical Recipes [15]. Jednoduchý a názorný program `eigen.f` demonstruje použití subrutin pro výpočet vlastních čísel a vlastních vektorů N rozměrné matice:

```
c eigen.f
c Program pro vypocet vlastnich cisel a~vlastnich vektoru.
c Jan 15th 2010

      program eigen

      integer np
      parameter(np = 10) ! maximalni rozmer matice

      integer n           ! skutecny rozmer vstupni matice
      real a(np,np)      ! vstupni matice (symetricka)
      real d(np)         ! diagonalni clen tridiagonalni matice
      real e(np)         ! ostatni clen pod diagonalou
      integer id(np)     ! pole indexu pro trideni vystupu

c nacteni symetricke matice ze standardniho vstupu
      write(*,*) '# pocet prvku n = '
      read(*,*) n
      write(*,*) n

      write(*,*) '# symetricka matice a(n,n) = '
      do i~ = 1, n
        read(*,*) (a(i,j), j = 1, n)
      enddo
      do i~ = 1, n
        write(*,*) (a(i,j), j = 1, n)
      enddo

c volani subrutin z~Numerical Recipes

c redukce symetricke matice na tridiagonalni
      call tred2(a,n,np,d,e)

c a(np,np) ... na vystupu tred2 jde o~transformacni matici,
c           kterou dosazujeme do tqli

c vypocet vlastnich cisel a~vektoru tridiagonalni matice
      call tqli(d,e,n,np,a)

c d(np)    ... na vystupu tqli jsou zde vlastni cisla
c a(np,np) ... byla nahrazena vlastnimi vektory
```

```

c trideni vlastnich cisel podle velikosti
  call srtidx(n,d,id)

c vypis vlastnich cisel (nejvetsi prvni)
  write(*,*) '# vlastni cisla matice a():'
  do i~= n, 1, -1
    write(*,*) d(id(i))
  enddo
  write(*,*)

c vypis odpovidajicich vlastnich vektoru
  write(*,*) '# vlastni vektory matice a() v~radcich:'
  do i~= n, 1, -1
    do j = 1, n
      write(*,20) a(j,id(i))
20      format(f16.8,1x,$)
    enddo
    write(*,*)
  enddo
  write(*,*)

  stop
  end

```

Literatura

- [1] Brož, M. Astronomický kurz (9) — Planetky. *Povětroň*, 2009, č. 2, s. 4–41. ISSN 1213-659X. Dostupné na: <http://www.astrohk.cz/ashk/povetron/>.
- [2] Bus, S. J., Binzel, R. P. Phase II of the small main—belt spectroscopic survey: a feature—based taxonomy. *Icarus*, 158, s. 146–177, 2002.
- [3] Demlová, M., Nagy J. *Algebra*, Praha: SNTL, 1985. 187 s. ISBN 04-010-85.
- [4] Harmanec, P, Brož, M. *Astrofyzika II — Stavba a vývoj hvězd* [online]. Astronomický ústav Univerzity Karlovy, 2009. Dostupné na: <http://sirrah.troja.mff.cuni.cz/~mira/astrofyzika2/>.
- [5] Ivezić, Ž. et al. Solar System objects observed in the Sloan Digital Sky Survey commissioning data. *Astronomical Journal*, 122, s. 2749–2784, 2001.
- [6] Ivezić, Ž. et al. Color confirmation of asteroid families. *Astronomical Journal*, 124, s. 2943–2948, 2002.
- [7] Kleczek, J. *Velká encyklopedie vesmíru*, Praha: Academia, 2002. 582 s. ISBN 80-200-0906-X.
- [8] Koukolík, F., Koubský, P. *Šimpanz a vesmír*, Praha: Vyšehrad, 1998. ISBN 80-7021-204-7.
- [9] Lewis, J. Three derivations of Principal Components [online]. [cit. 20.2.2010]. Dostupné na: <http://scribblethink.org/Work/PCderivations/PCderivations.pdf>.
- [10] Lindsay, I. S. A tutorial on Principal Component Analysis [online]. February 2002 [cit. 11.2.2010]. Dostupné na: http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf.
- [11] Mikulášek, Z., Proměnné hvězdy [online]. [cit. 18.5.2010]. Dostupné na: http://astro.physics.muni.cz/documents/lecture_notes/.
- [12] Nesvorný, D. et al. Evidence for asteroid space weathering from the Sloan Digital Sky Survey. *Icarus*, 173, s. 132–152, 2005.
- [13] Parker, A.H. et al. The size distributions of asteroid families in the SDSS Moving Object Catalog 4. *Icarus*, 198, s. 138–155, 2008.
- [14] Photometric Flux Calibration. [online]. [cit. 18.5.2010]. Dostupné na: <http://www.sdss.org/dr5/algorithms/fluxcal.html>.
- [15] Press, W. H. et al. *Numerical Recipes — The Art of Scientific Computing*, Cambridge: Cambridge University Press, 1997. ISBN 0521880688.
- [16] Sloan Digital Sky Survey, Moving Object Catalogue [online]. Dostupné na: <http://www.astro.washington.edu/users/ivezic/sdssmoc/ADR4.dat>.

- [17] Sloan Digital Sky Survey [online]. [cit. 30.3.2010]. Dostupné na:
⟨<http://www.sdss.org>⟩.
- [18] Wetson, H. A note on standard deviation [online]. [cit. 30.3.2010]. Dostupné na:
⟨<http://mathcentral.uregina.ca/RR/database/RR.09.95/weston2.html>⟩.

Rejstřík

- aritmetický průměr, 3
- barevný index, 22
- báze vektorového prostoru, 9
- bolometrické hodnoty, 18
- fotometrické systémy, 19
- fotometrický systém AB_{95} , 20
- fotometrický systém Johnsonův, 19
- fotometrický systém Strömgerův, 20
- fotometrický systém UBVRI, 19
- fotosférická teplota hvězd, 22
- hlavní komponenta, 8
- hvězdná velikost zdánlivá, 21
- charakteristická rovnice, 8
- charakteristický polynom, 10
- kovariance, 4
- kovarianční matice, 5
- lineární transformace, 10
- lineární zobrazení, 9
- medián, 3
- metoda hlavních komponent, 7
- ortogonalita, 6
- PCA, 7
- plošná hustota zářivého toku, 18
- Principal Components Analysis, 7
- přímá úměrnost, 5
- rozptyl, 4
- směrodatná odchylka, 4
- spektrální hustota zářivého toku, 18
- statistická váha, 3
- střední hodnota, 3
- transformační matice, 6
- variance, 4
- vážený průměr, 3
- vektor jednotkový, 6
- vlastní číslo, 11
- vlastní vektor, 5, 9, 11
- zářivost, 18
- zářivý tok, 18