



Úvod do geometrické morfometrie

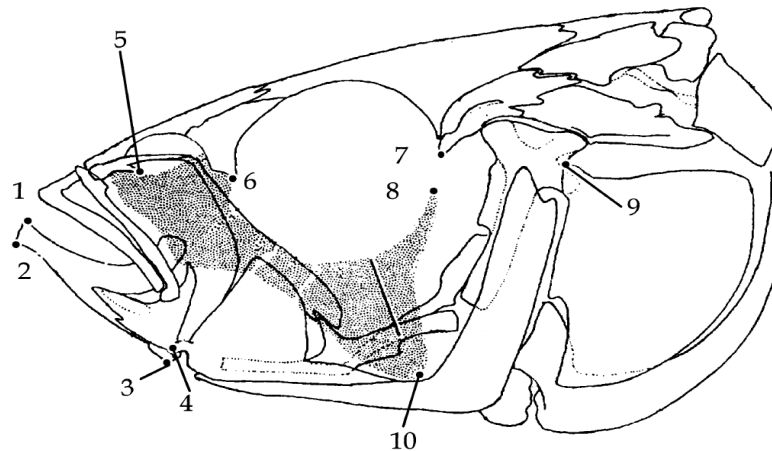
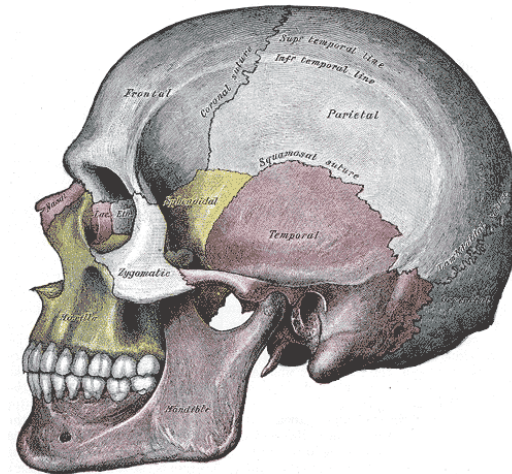
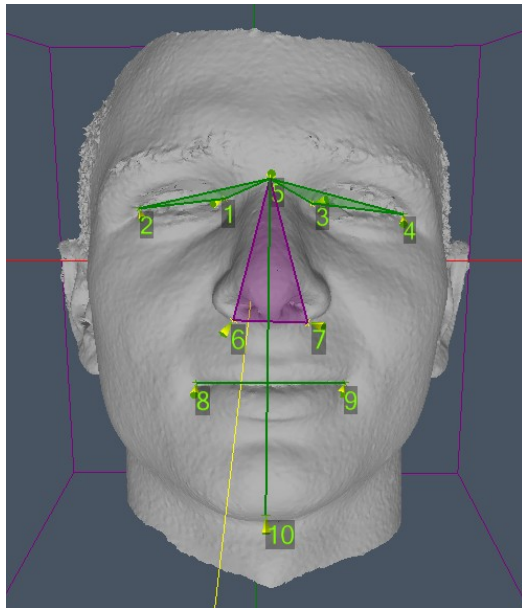
© 2011 **Josef Pelikán, Václav Krajíček**
KSVI MFF UK Praha

<http://cgg.mff.cuni.cz/~pepca/>

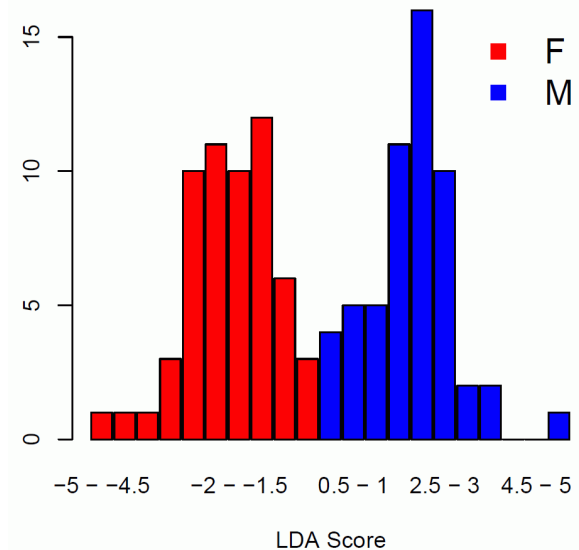
<http://cgg.mff.cuni.cz/~vajicek/>

Seminář KDM: Aplikace matematiky pro učitele,
18. 10. 2011, K6

Geometrická morfometrie



Semilandmarks by arc





Obsah přednášky

- ◆ co je geometrická morfometrie?
 - ◆ aplikace, cíle a základní postupy
- ◆ postupy v klasické GMM
- ◆ tvar, velikost, registrace tvarů
 - ◆ Bookstein, Prokrustes,
- ◆ tvarové prostory
 - ◆ Kendallův prostor trojúhelníků
 - ◆ analýza 2D křivek
- ◆ statistické modely tvarů (množiny jedinců)
 - ◆ PCA, shluková analýza, diskriminační analýza

Geometrická morfometrie



- ◆ **morfometrie:** disciplína zabývající se zkoumáním změn tvarů objektů, jejich variabilitou a klasifikací
 - ◆ biologie, archeologie, forenzní vědy
- ◆ **geometrická morfometrie (GMM):** geometrický popis tvaru vychází z tzv. landmarků
 - ◆ 2D nebo 3D souřadnice biologicky významných bodů
- ◆ **moderní GMM:** další metody, jak popsat tvar
 - ◆ trojúhelníkové sítě, analytické reprezentace křivek a ploch, objemová data

Cíle GMM



- **odhalit skrytá fakta a souvislosti**
- **než začneme** se vztahy → popis exempláře
 - landmarkové metody, křivky, trojúhelníkové sítě
- **nic nevíme** → získávání nových poznatků studiem tvaru u skupiny exemplářů
 - PCA, shluková analýza, vizualizace
- **něco tušíme** → ověření nebo vyvrácení hypotézy
 - statistické testy, korelace, regresní analýza
- **víme co, nevíme kolik** → stanovení míry odlišnosti
 - metriky

Úlohy GMM



- ◆ porovnávání dvou (mnoha) tvarů
- ◆ hledání odlišností mezi dvěma či více skupinami
- ◆ hledání průměrného jedince v daném souboru
- ◆ studium symetrií
- ◆ hledání trendu vývoje jedince (ontogeneze) nebo druhu (fylogeneze)
- ◆ klasifikace tvarů
- ◆ interpolace a extrapolace (predikce)



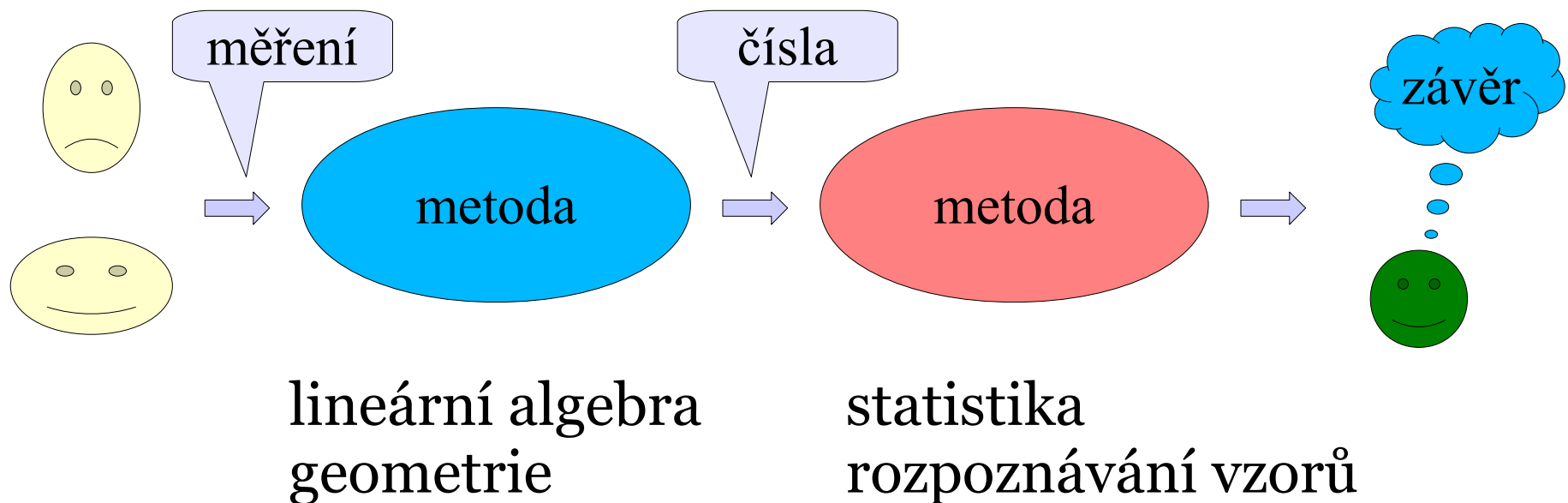
Morfometrické postupy

- ◆ co je geometrická morfometrie?
 - ◆ aplikace, cíle a základní postupy
- ◆ **postupy v klasické GMM**
- ◆ tvar, velikost, registrace tvarů
 - ◆ Bookstein, Prokrustes,
- ◆ tvarové prostory
 - ◆ Kendallův prostor trojúhelníků
 - ◆ analýza 2D křivek
- ◆ statistické modely tvarů (množiny jedinců)
 - ◆ PCA, shluková analýza, diskriminační analýza

Postup GMM



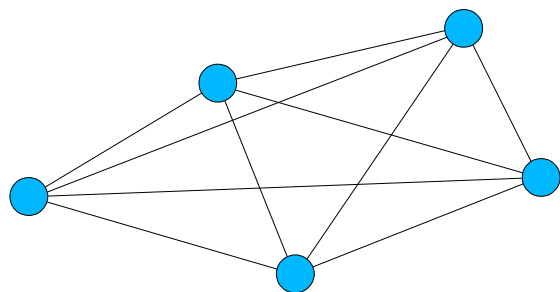
1. popsat tvar čísly
2. čísla statisticky vyhodnotit



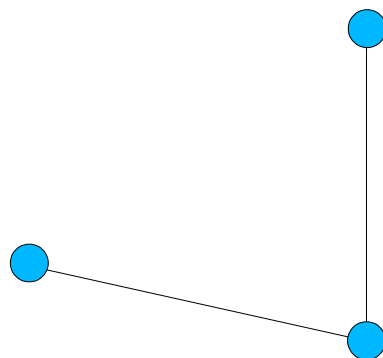


Tradiční (stará) morfometrie

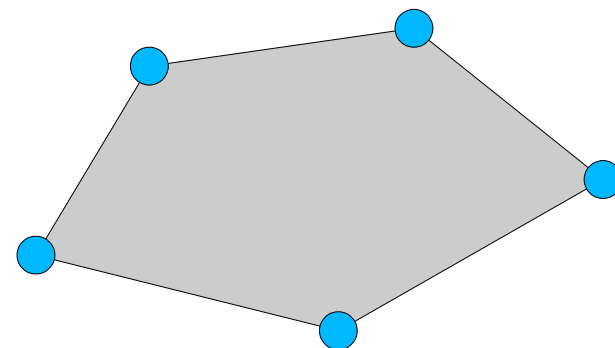
- poloha „bodů“ nehraje roli (ani se nedala dobře měřit)
- pracovalo se s
 - ♦ vzdálenostmi mezi dvěma „body“, poměry vzdáleností
 - ♦ úhly mezi třemi „body“
 - ♦ plochou oblasti (polygonu)



$$\binom{p}{2}$$



$$p \binom{p-1}{2}$$

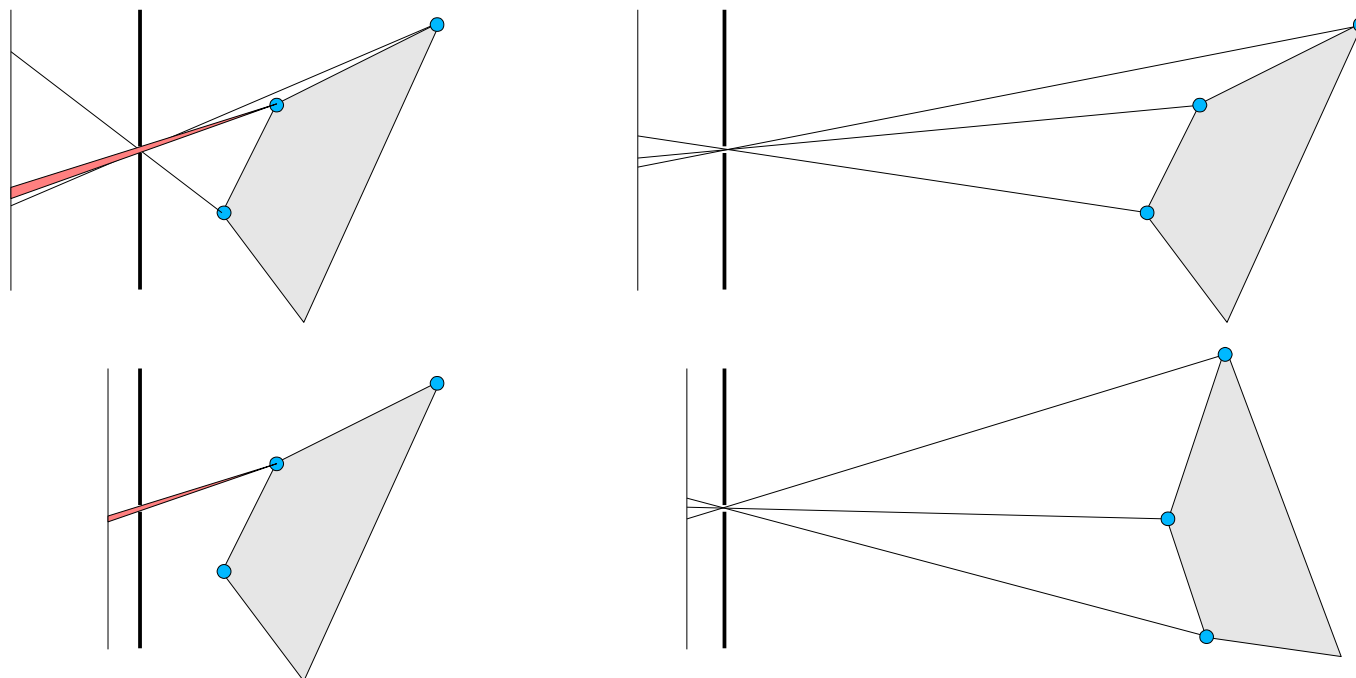


Plocha složená z ploch nepřekrývajících se trojúhelníků

Pořizování dat



- ◆ **2D**: fotografie, video
 - ◆ perspektivní projekce – potenciální zdroj chyb měření
- ◆ **3D**: objemová data, povrchové skenery, polohovací zařízení

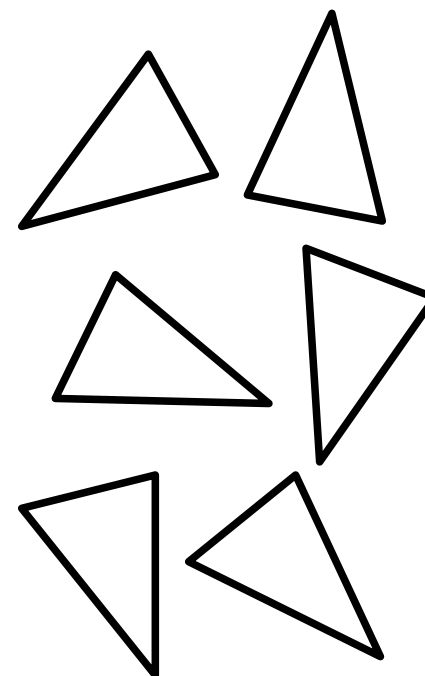
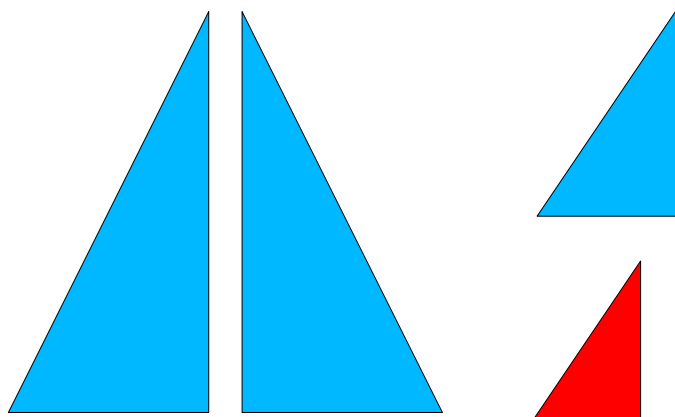
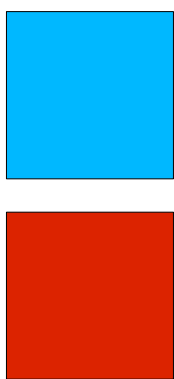




Moderní metody GMM

- ♦ metody měření, zpracování, analýzy a zobrazení s cílem studia tvaru a jeho změn
- ♦ **tvarové proměnné** – popisují tvar, nemění se s velikostí

forma = velikost + tvar





Co je vlastně tvar?

- ◆ co je geometrická morfometrie?
 - ◆ aplikace, cíle a základní postupy
- ◆ postupy v klasické GMM
- ◆ **tvar, velikost, registrace tvarů**
 - ◆ Bookstein, Prokrustes,
- ◆ tvarové prostory
 - ◆ Kendallův prostor trojúhelníků
 - ◆ analýza 2D křivek
- ◆ statistické modely tvarů (množiny jedinců)
 - ◆ PCA, shluková analýza, diskriminační analýza

Tvar



- ◆ definice (Kendall):

Tvar je veškerá geometrická informace, která objektu zůstane po odstranění vlastností polohy, měřítka a otočení

- ◆ „tradiční“ morfometrie

- ◆ pracuje přímo se vzdálenostmi (vylepšení – poměry vzdáleností)
- ◆ samotné vzdálenosti stěží odrážejí tvar
- ◆ dva exempláře jsou tak tvarem neporovnatelné

- ◆ **oddělení velikosti a tvaru** v datech

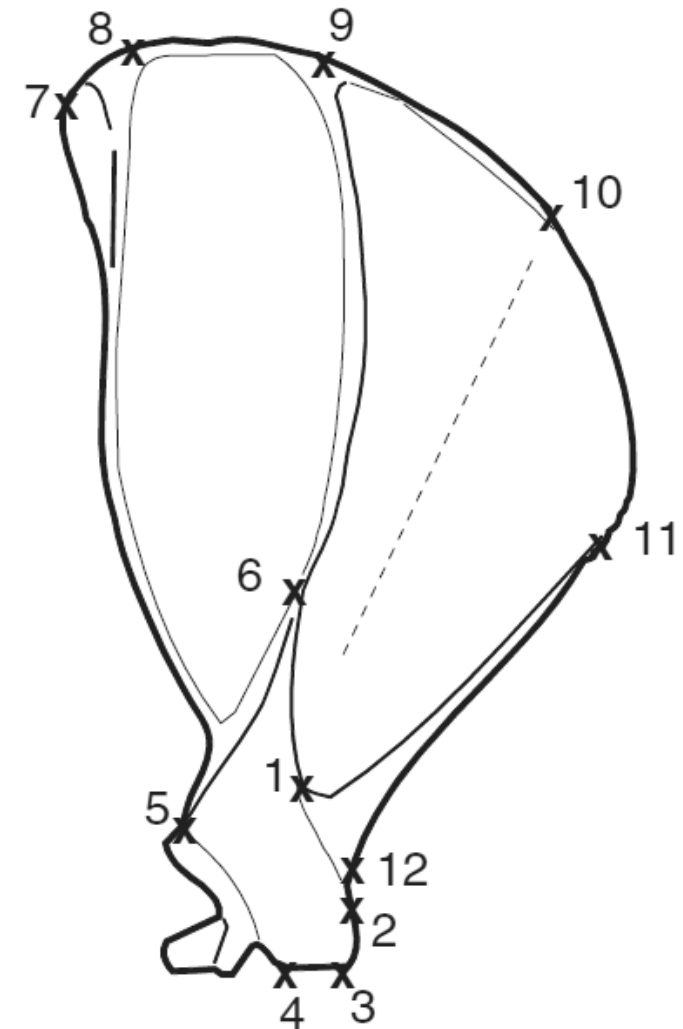
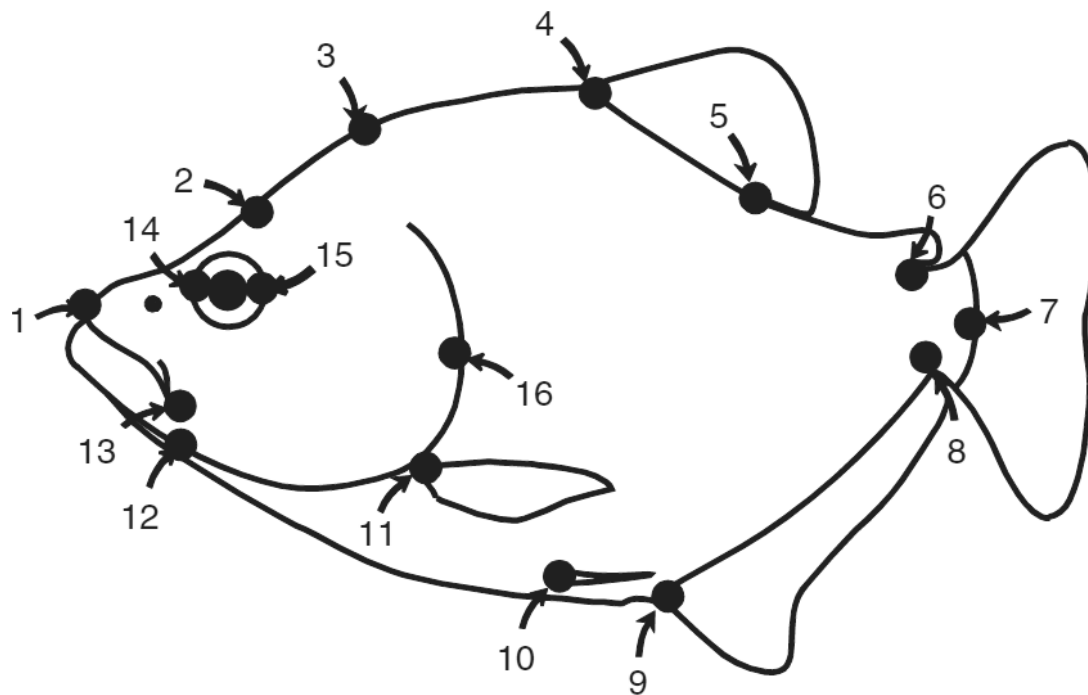
- ◆ jedna z hlavních myšlenek geometrické morfometrie



Landmarky

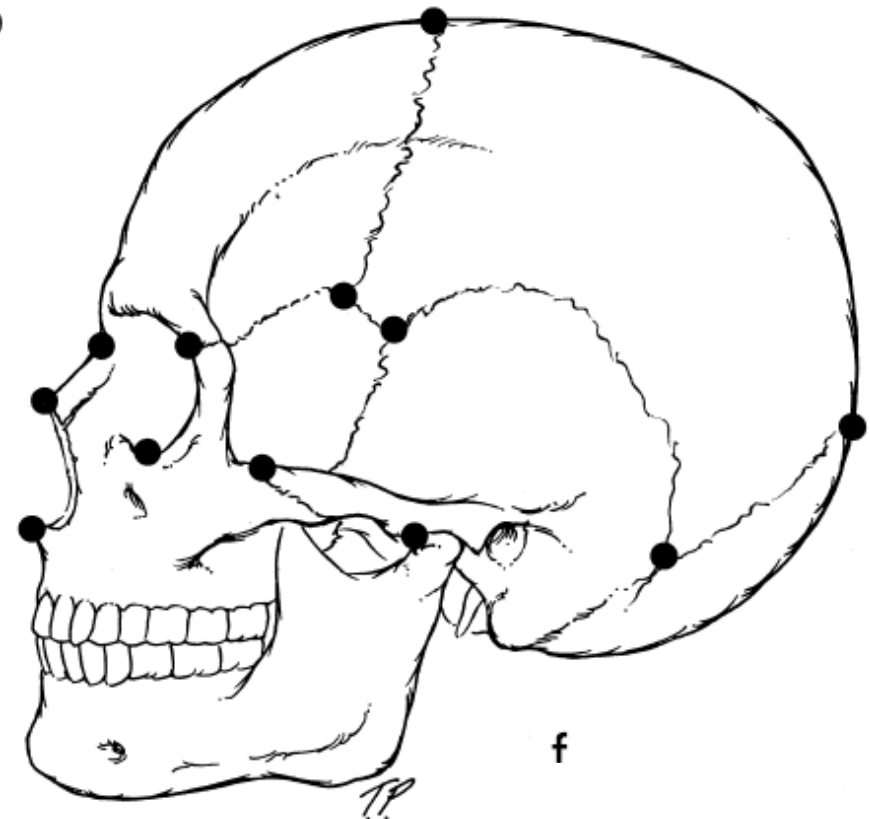
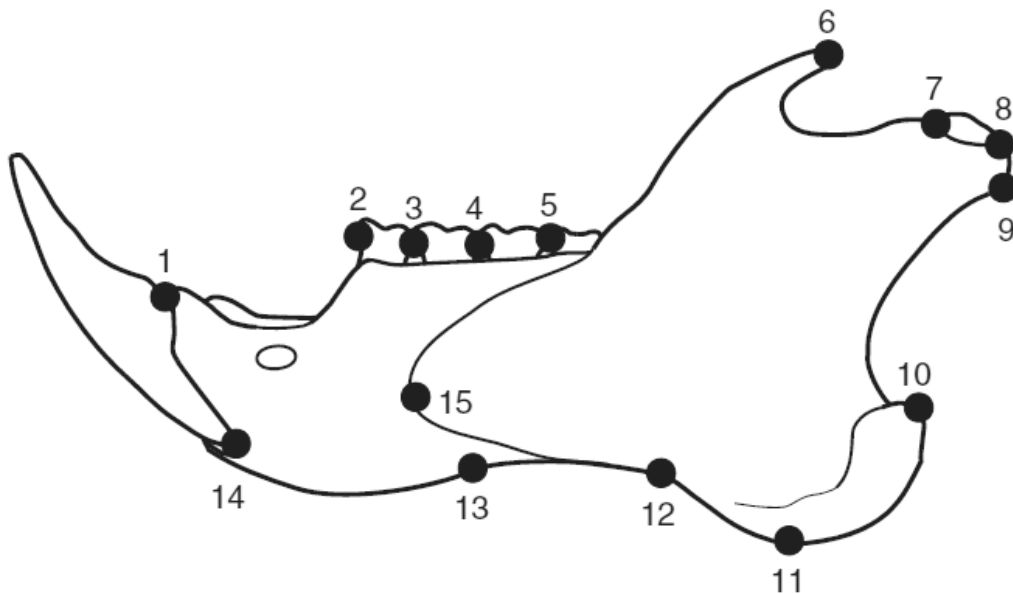
- ▶ **významné body** umístěné na objektu
 - ♦ biologický význam (pojmenování)
 - ♦ souřadnice v 2D nebo 3D prostoru
- ▶ **typické umístění**, snadno reprodukovatelné:
 - ♦ extrém tvaru (roh)
 - ♦ průsečík dvou linií
 - ♦ extrém křivosti, ohyb, příp. inflexní bod
- ▶ **semi-landmarky**
 - ♦ pomocné body dopočítané nějakým matematickým postupem

Landmarky – příklady



(převzato ze Zelditch et al: Geometric morphometrics for biologists: the primer)

Landmarky – příklady



(převzato ze Zeldich et al: Geometric morphometrics for biologists: the primer)

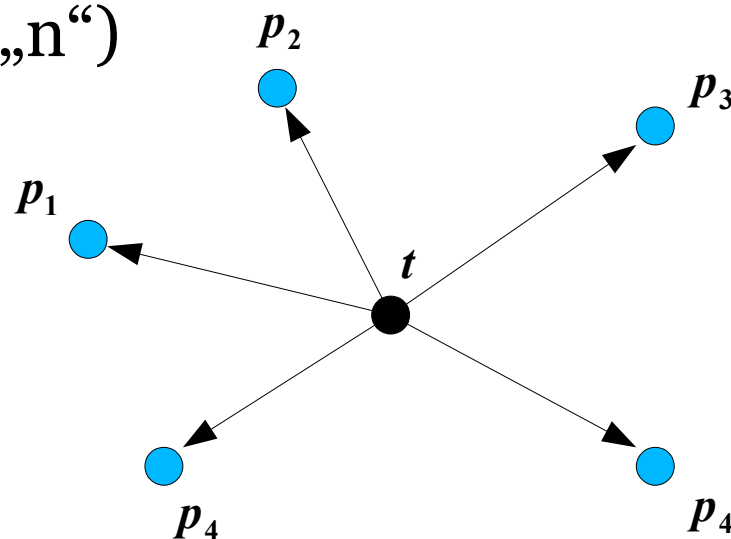


Velikost

- ◆ nezáporný atribut, linearita
- ◆ **středová velikost** (Centroid Size)
 - ◆ matematicky je „nezávislá na tvaru“
 - ◆ výpočet z landmarků (pozor na „n“)

$$\mathbf{t} = \frac{1}{n} \sum_{i=1}^n \mathbf{p}_i$$

$$CS = \sqrt{\sum_{i=1}^n (\mathbf{p}_i - \mathbf{t})^T (\mathbf{p}_i - \mathbf{t})}$$



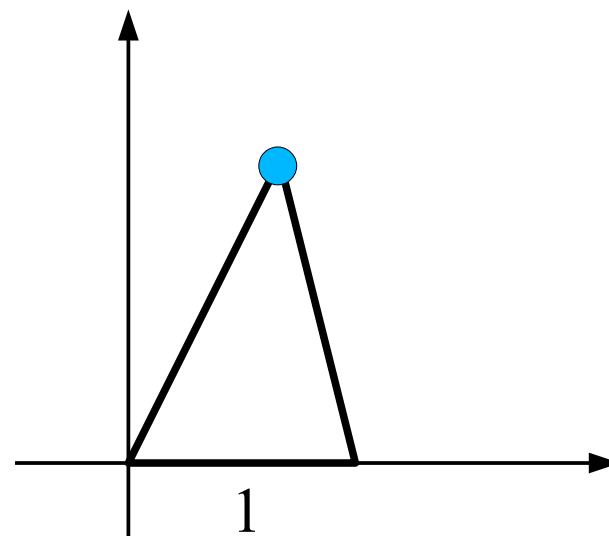
- ◆ často se hledá konkrétní souvislost středové velikosti s tvarem



Příklad – trojúhelník

- nejjednodušší geometrický objekt, u kterého se dá hovořit o tvaru
- úplný popis tvaru trojúhelníku v rovině: dvojice čísel
- **stupeň volnosti tvaru**: počet proměnných po odstranění velikosti, otočení a posunutí

- 2D: $2p - 4$
- 3D: $3p - 7$
- k-dim: $pk - k - k(k - 1)/2 - 1$
p - počet bodů

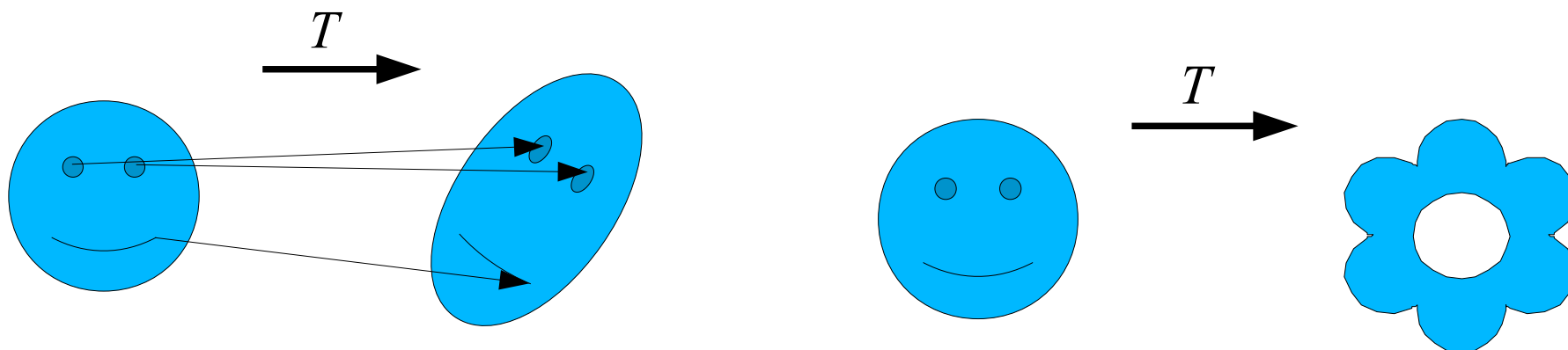




Registrace

- ◆ definice pojmu:

Hledání transformace dvou datových množin do společného systému souřadnic



- ◆ transformací vzorků do společného „jednotkového“ souřadného systému odstraníme vliv velikosti na tvarové proměnné



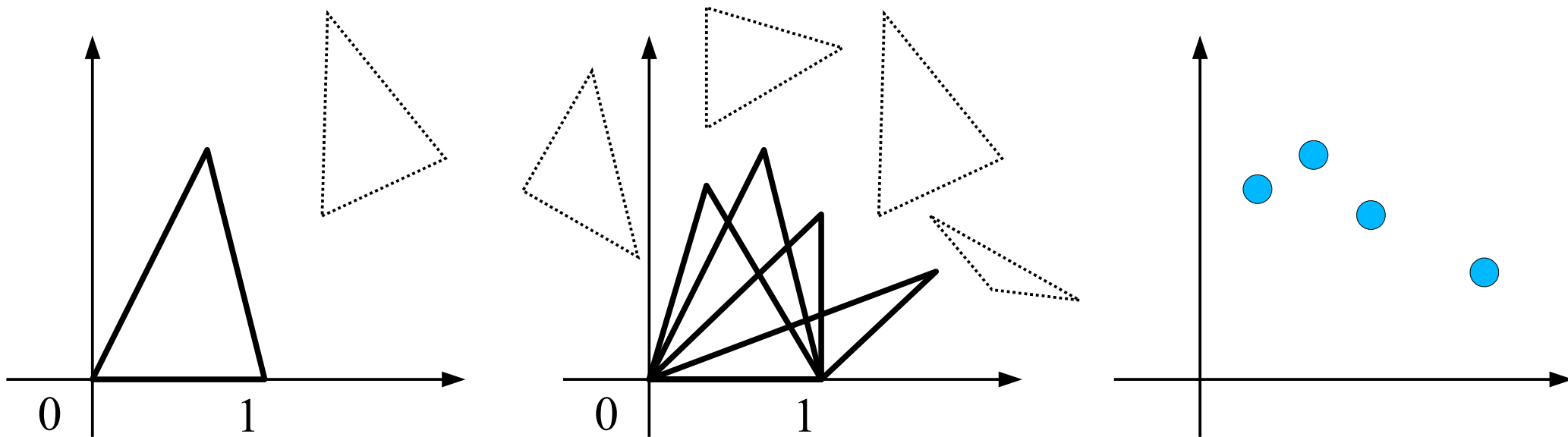
Typy registrace

- různé typy registrace podle:
 - ♦ dat se kterými pracují: 2D, 3D, objem, obrázek, body, trojúhelníkové sítě, ..
 - ♦ typu transformace které připouštějí: rigidní, elastická
 - ♦ způsobu hledání: uzavřené formy, gradientní metody
 - ♦ způsobu porovnání dvou datových množin
- pro účely získání tvarových proměnných připustíme pouze transformaci **polohy, orientace a měřítka**
 - ♦ lineární algebra

Dvoubodová registrace (Bookstein)

„Booksteinova registrace“:

- volba dvojice landmarků: tzv. **základna** (base-line)
- transformace celého vzorku tak, aby základna ležela na ose x mezi **0** a **1**.

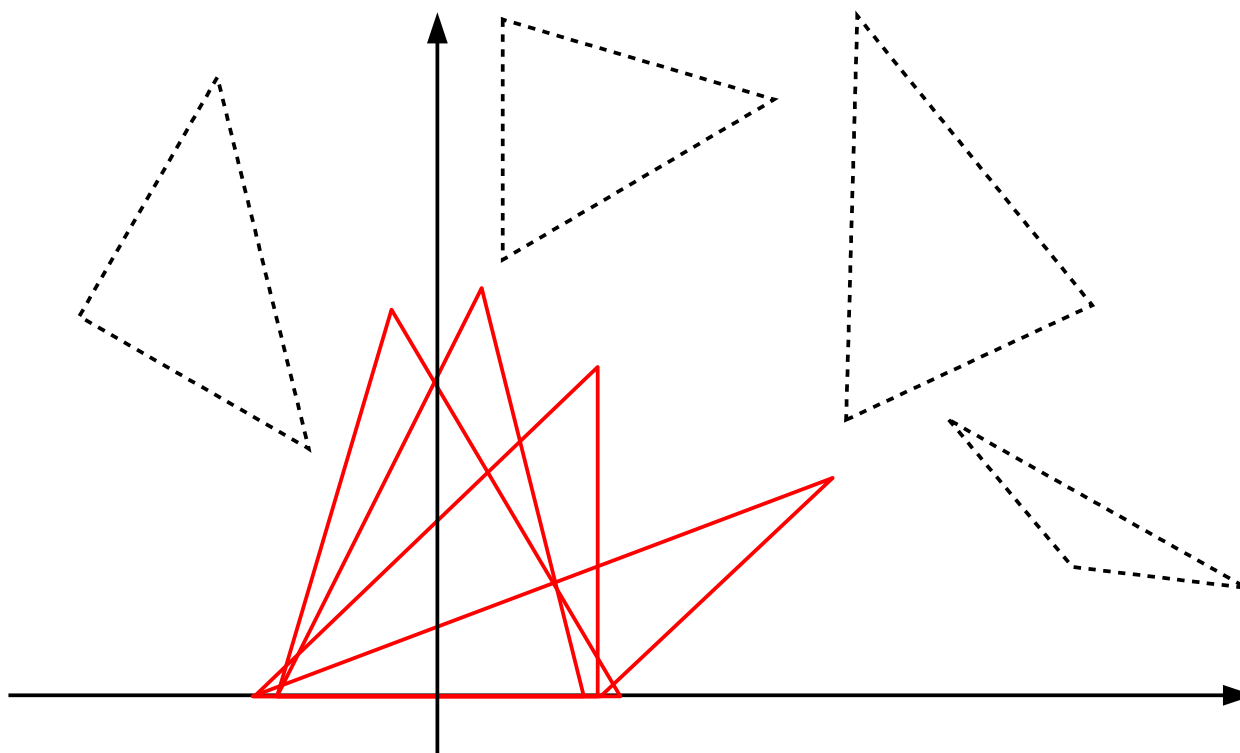




Klouzající základna

Podobný předchozímu postupu:

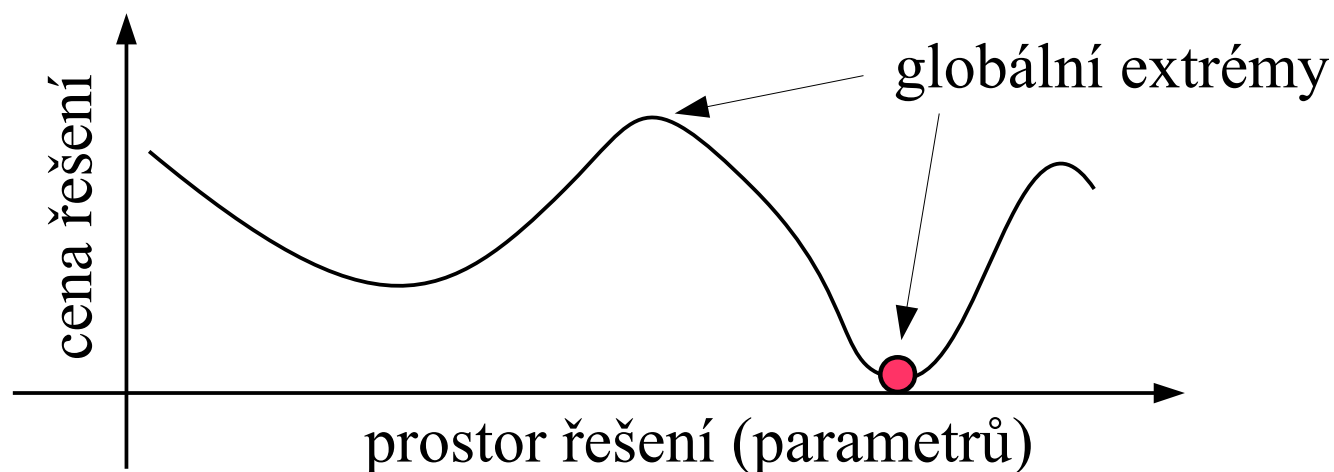
- ◆ škálování celého objektu na jednotkovou velikost
- ◆ do počátku se umístí **střed základny**





Prokrustovská analýza (GPA)

- ◆ „lepší“ alternativa:
 - ◆ funguje i v 3D, není ovlivněna výběrem žádné základny
- ◆ hledání nejlepších parametrů transformace (posunutí, škálování, rotace) ve smyslu **nejmenší vzdálenosti** odpovídajících si landmarků (nejmenší čtverce)
- ➔ optimalizační problém



Prokrustovská analýza II



- ♦ jak ohodnotit naši transformaci?

$$X' = \frac{X - \mathbf{1}^T \mathbf{t}}{CS} H = YH \quad \|YH - \bar{X}\| \rightarrow \min$$

- ♦ \bar{X} .. referenční tvar, kterému se chci nejvíce přiblížit
- ♦ \mathbf{t} .. poloha těžiště, posun těžiště do počátku
- ♦ **CS** .. středová velikost, škálování na jednotkovou velikost
- ♦ **H** .. rotační matice, úhel, který ji definuje, se spočítá minimalizací vzdálenosti k referenčnímu tvaru
- ♦ **norma** $\| \|$ odpovídá součtu čtverců (vzdáleností odpovídajících si landmaků)

Prokrustovská analýza III



- norma a stopa matice $\|A\| = \text{trace}(A^T A)$
- dosazení $\|YH - \bar{X}\| = \text{trace}(Y^T Y + \bar{X}^T \bar{X}) - 2 \text{trace}(\bar{X}^T YH)$
- maximalizace posledního členu a SVD
 $\text{trace}(\bar{X}^T YH) = \text{trace}(USV^T H) = \text{trace}(SV^T HU) = \text{trace}(SQ)$
 $\text{trace}(SQ) \rightarrow \max$
- Q je jednotková, protože tak maximalizuje stopu
 $Q = V^T HU = I$

proč to funguje

$$\bar{X}^T Y = USV^T \quad H = VU^T$$

**co musíme
spočítat**



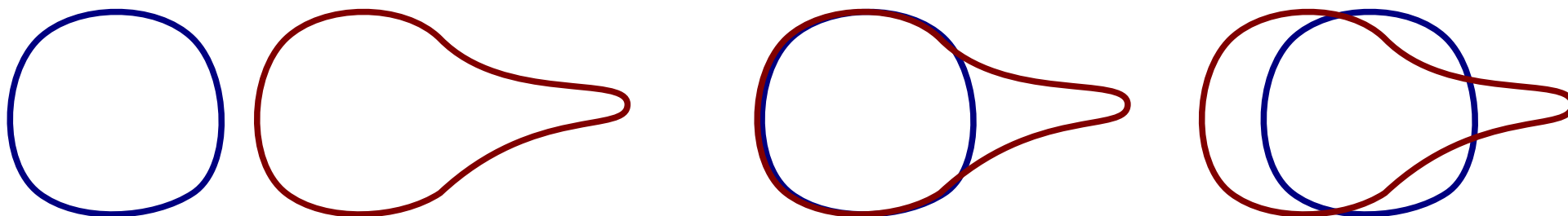
GPA – shrnutí

♦ výhody:

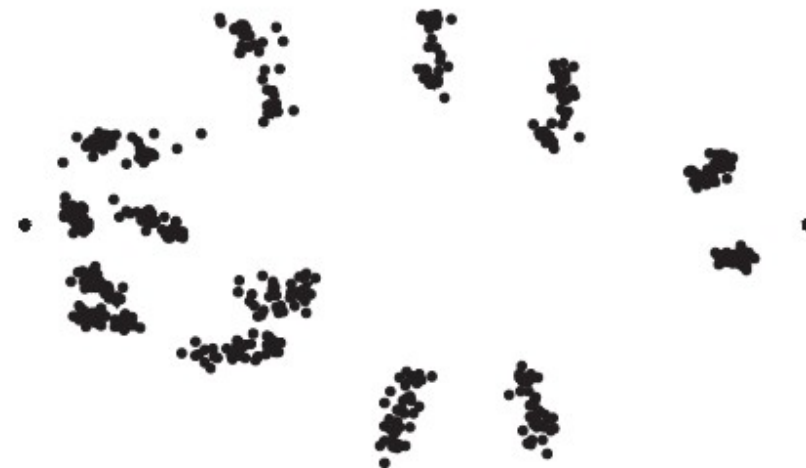
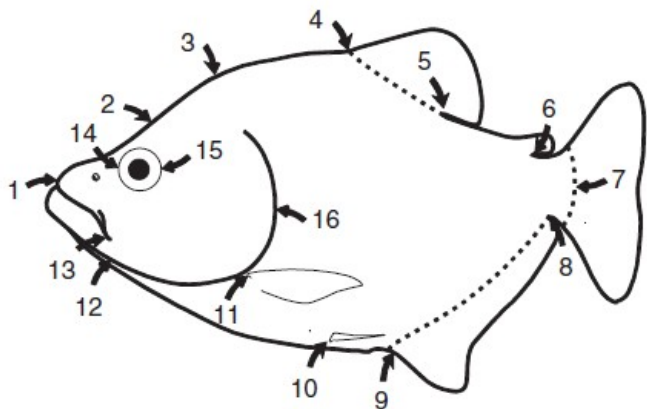
- ♦ rozšiřitelnost do 3D, stabilní, rychlá

♦ nevýhody:

- ♦ v základní variantě neumí vážené landmarky
- ♦ jeden jedinec („outlier“) může rozhodit zarovnání celé skupiny („Pinocchio effect“)
 - řešení: použití mediánu, předvýběr



Srovnání registrací



Bookstein



klouzavá základna



Prokrustes



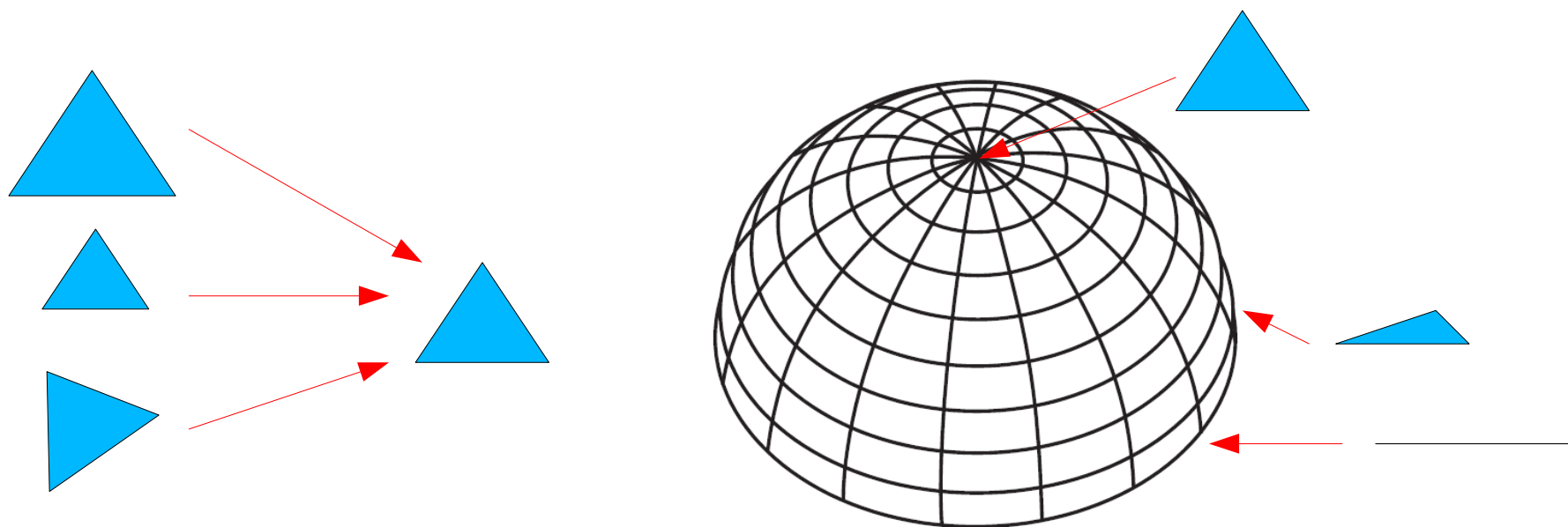
Tvarové prostory

- ◆ co je geometrická morfometrie?
 - ◆ aplikace, cíle a základní postupy
- ◆ postupy v klasické GMM
- ◆ tvar, velikost, registrace tvarů
 - ◆ Bookstein, Prokrustes,
- ◆ **tvarové prostory**
 - ◆ Kendallův prostor trojúhelníků
 - ◆ analýza 2D křivek
- ◆ statistické modely tvarů (množiny jedinců)
 - ◆ PCA, shluková analýza, diskriminační analýza



Prostor trojúhelníků

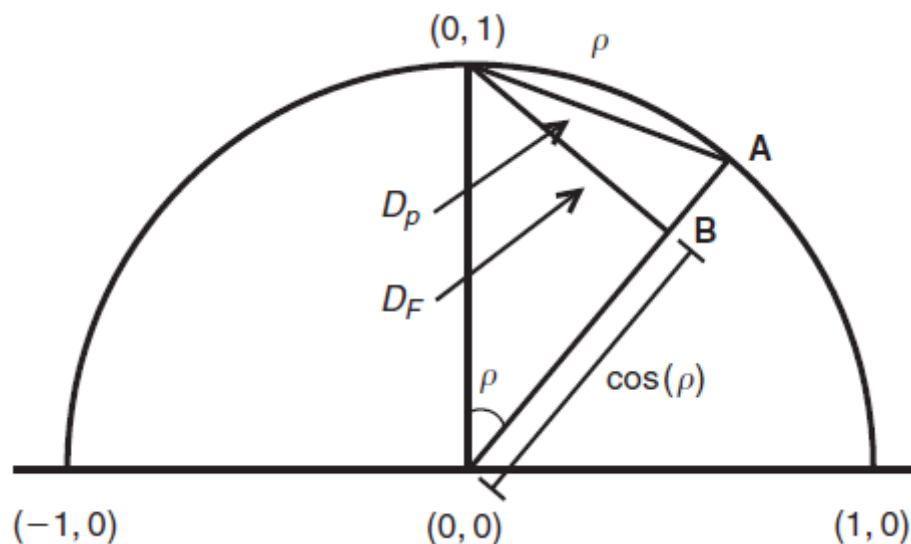
1. souřadnice vrcholů trojúhelníků
2. zarovnání a normalizace → omezený prostor (2D)
3. omezenost prostoru → povrch polokoule
 - zemská „šířka“ a „výška“





Prostor trojúhelníků II

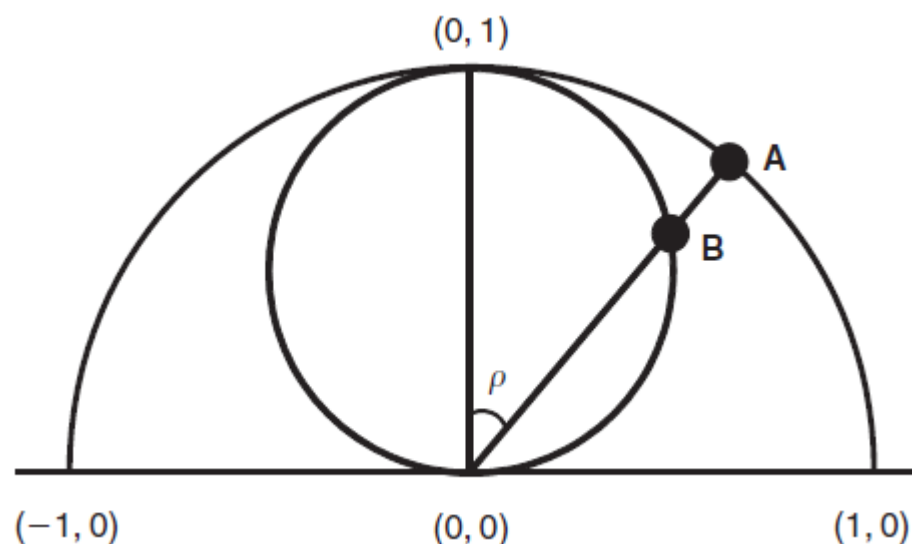
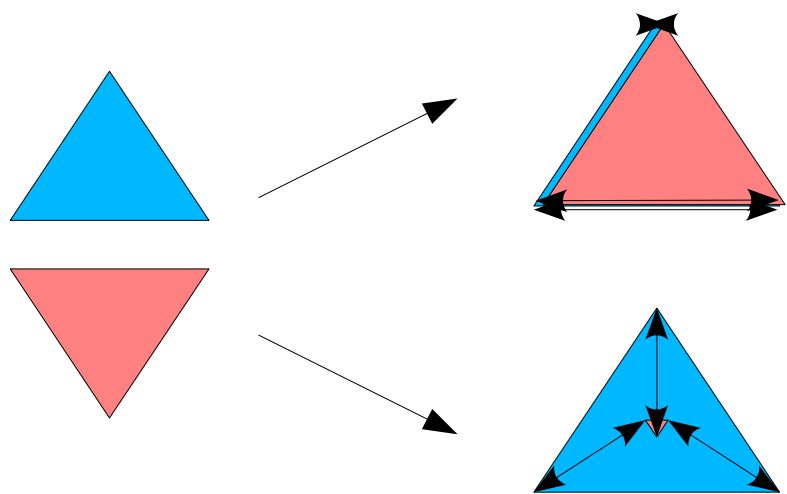
- částečná prokrustovská vzdálenost D_p
 - přímá vzdálenost mezi dvěma body na povrchu
 - přechodový tvar nemá jednotkovou velikost
 - odpovídá $2 \sin(\rho/2)$
- prokrustovská vzdálenost
 - vzdálenost po povrchu
 - oblouk ρ
- úplná prokrustovská vzdálenost D_F
 - bez ohledu na normalizaci, vzdálenost je $\cos(\rho)$
 - nejkratší vzdálenost, odpovídá $\sin(\rho)$



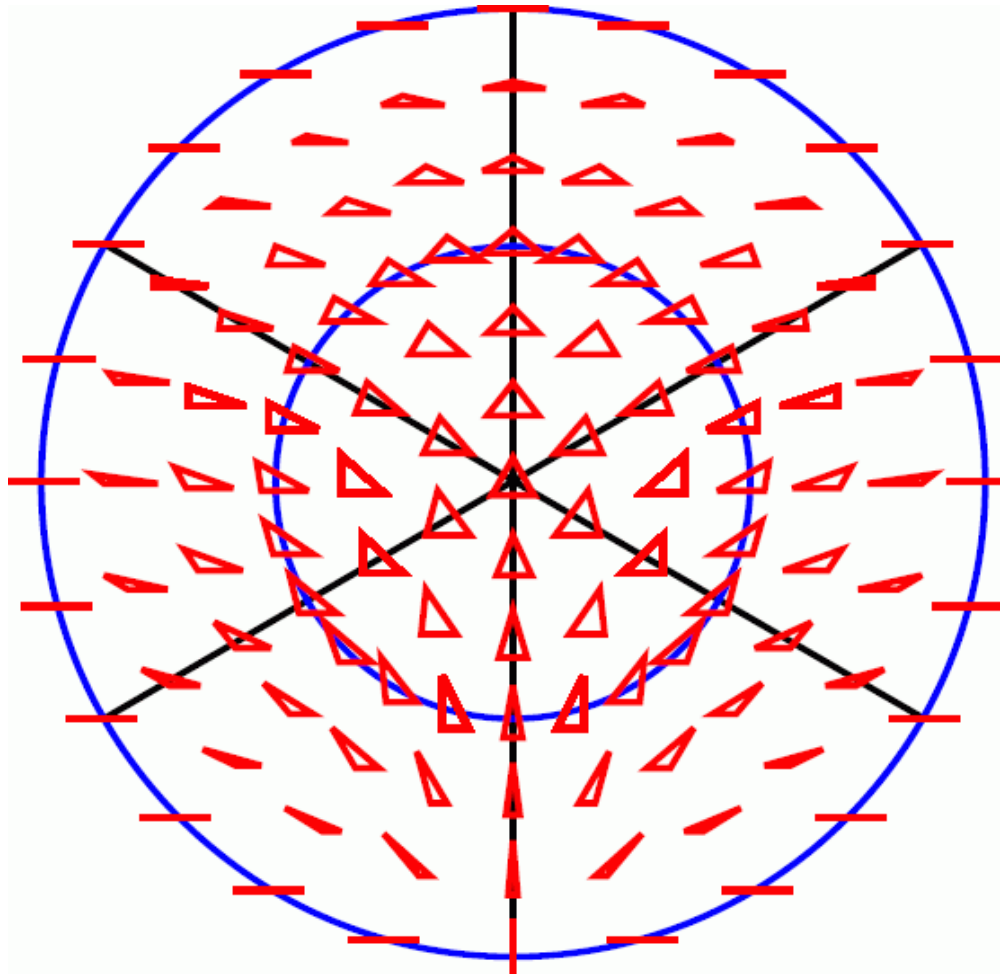


Kendallův prostor

- indukován úplnou prokrustovskou vzdáleností
- vzájemné vzdálenosti libovolných bodů odpovídají úplné prokrustovské vzdálenosti
 - to neplatilo u prostoru trojúhelníků!
- odpovídá povrchu koule o poloměru $1/2$



Kendallův prostor



Jedna hemisféra pozorovaná z pólu



Důsledky zakřivení prostoru

- ▶ statistické metody pracují **na ploše** a ne na sféře
 - ♦ některé mohou pracovat v zakřiveném prostoru..
- ▶ chceme používat jednoduchou **eukleidovskou vzdálenost** mezi vektory tvarových proměnných

$$E_d(\mathbf{a}, \mathbf{b}) = \sqrt{\sum_i^n (a_i - b_i)^2}$$

- ▶ dá se použít pro porovnávání podobných tvarů při projekci do tečné roviny okolo referenčního tvaru
 - ♦ teoreticky můžu srovnávat lebky odlišných živočichů pokud mají stejný počet landmarků
 - ♦ vzdálenosti nebudou extrémně odlišné



Analýza křivky (kontury)

- ♦ hladká hranice, kontura
 - ♦ neobsahuje přímo landmarky (jen koncové body)
- ♦ segmentace kontury

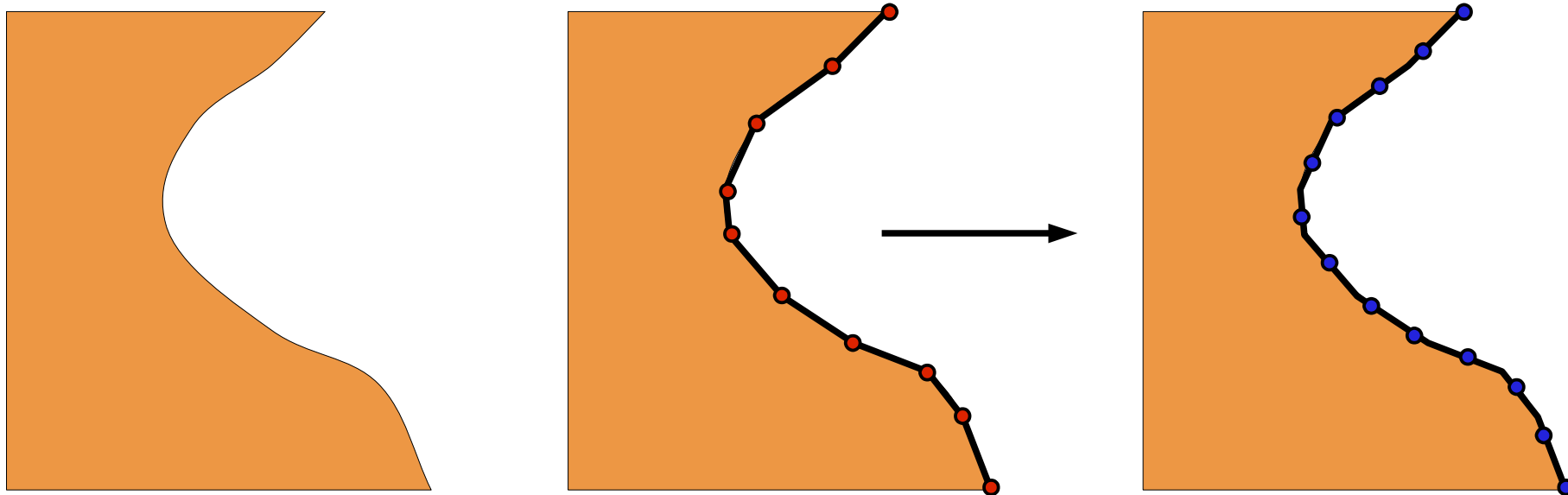
Jak konturu reprezentovat čísla?

1. semilandmarky - body
 - ♦ rozdělení kontury na úseky podle délky, nebo úhlu
2. transformace na soubor koeficientů
 - ♦ aproximace nějakou křivkou a práce s koeficienty
 - ♦ Waveletová (vlnková) / Fourierova transformace navzorkované křivky



Semilandmarky

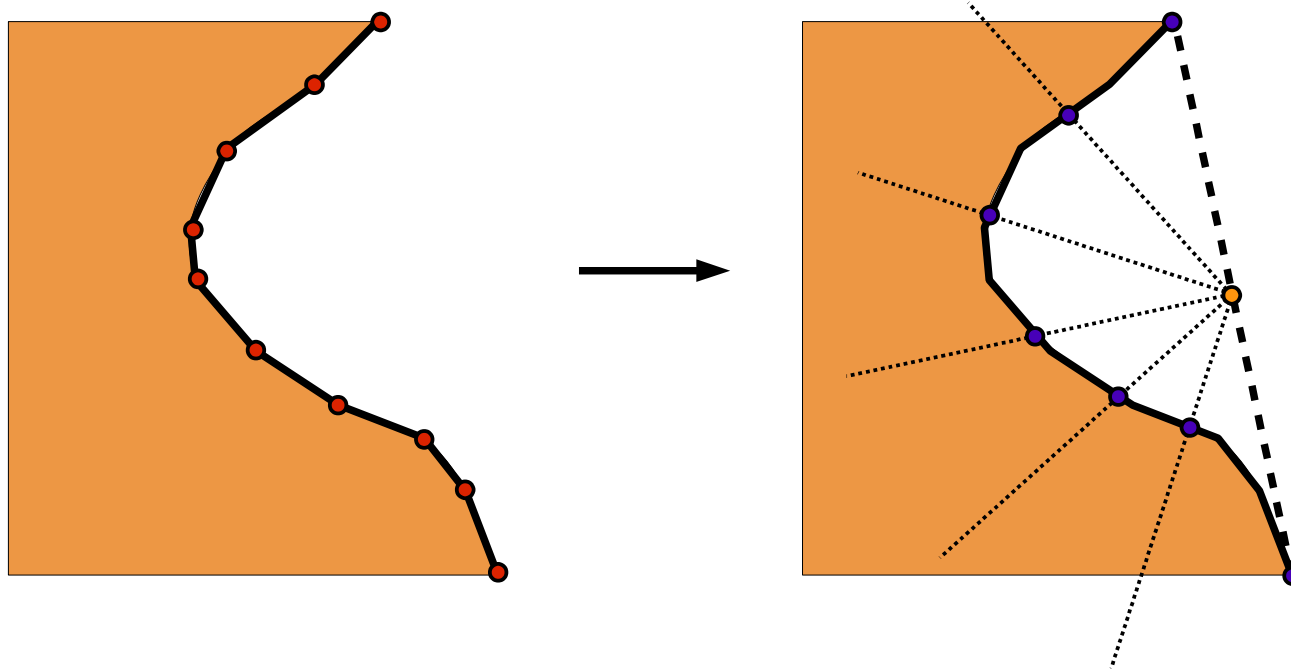
- dělení podle délky křivky (arc) na potřebný počet n úseků



Semilandmarky II



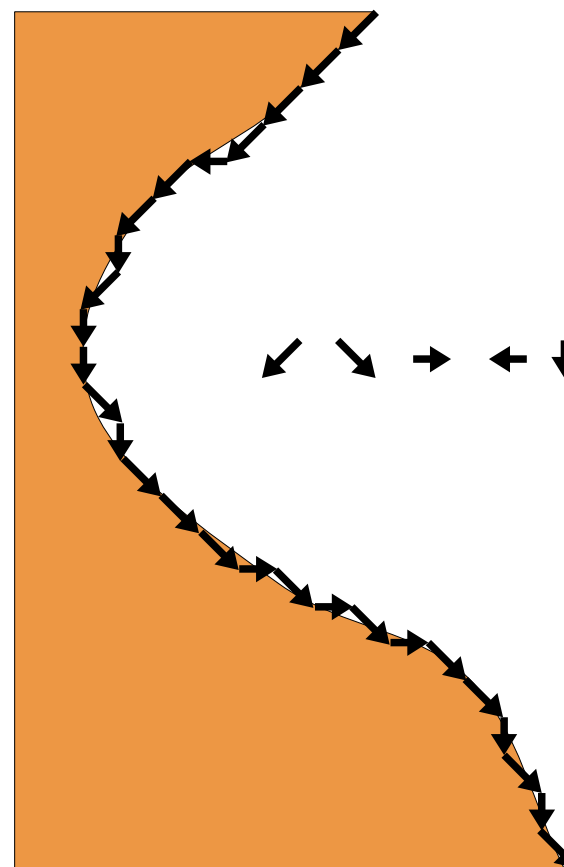
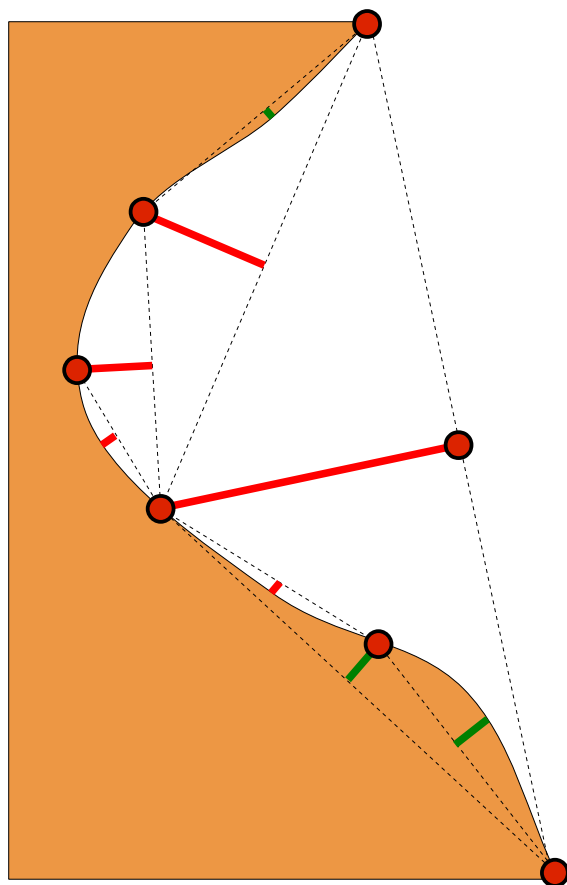
♦ dělení podle úhlu





Jiné reprezentace křivek

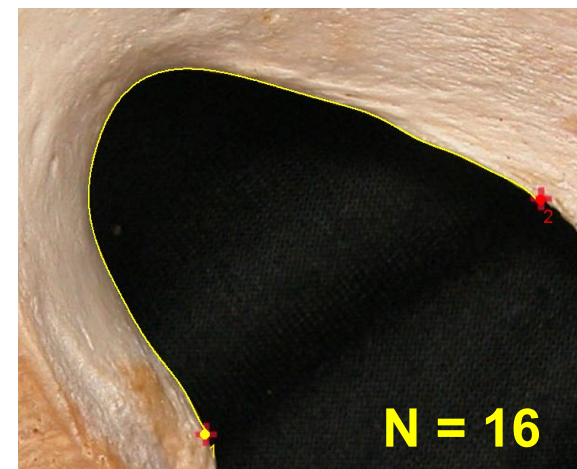
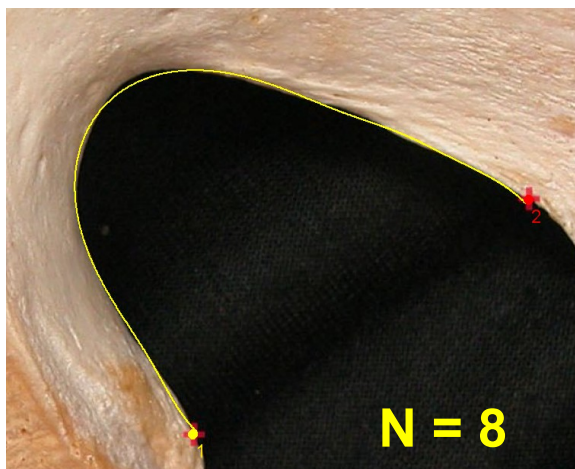
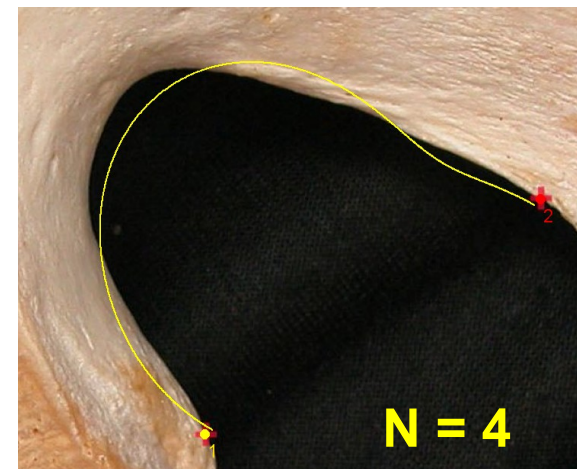
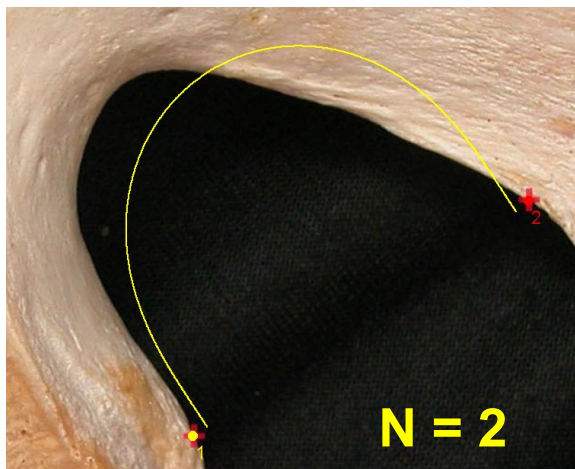
- mnoho různých způsobu jak zachytit tvar kontury



Cirkulární harmoniky



- ◆ Legendrové polynomy nebo Fourierovy koeficienty



Statistické modely



- ◆ co je geometrická morfometrie?
 - ◆ aplikace, cíle a základní postupy
- ◆ postupy v klasické GMM
- ◆ tvar, velikost, registrace tvarů
 - ◆ Bookstein, Prokrustes,
- ◆ tvarové prostory
 - ◆ Kendallův prostor trojúhelníků
 - ◆ analýza 2D křivek
- ◆ **statistické modely tvarů (množiny jedinců)**
 - ◆ PCA, shluková analýza, diskriminační analýza



Statistický model tvaru

- **model:** matematický popis jevu z reálného světa
- příklady:
 - ♦ model tvaru lebky může být „vhodným“ reprezentantem
 - splňuje dokonale pouze on sám, extrémny jsou daleko
 - ♦ průměrný reprezentant ze zkoumané populace
 - nesplňuje ho nikdo, ale všichni jsou mu dostatečně blízko
 - ♦ průměr + odchylka
 - může popisovat i nemožné exempláře
- parametrický model
 - ♦ každý konkrétní exemplář odpovídá číselné konfiguraci



Statistický model tvaru II

- zohledňuje tendence vyskytující se v pozorované **množině exemplářů (jedinců)**, vzorku

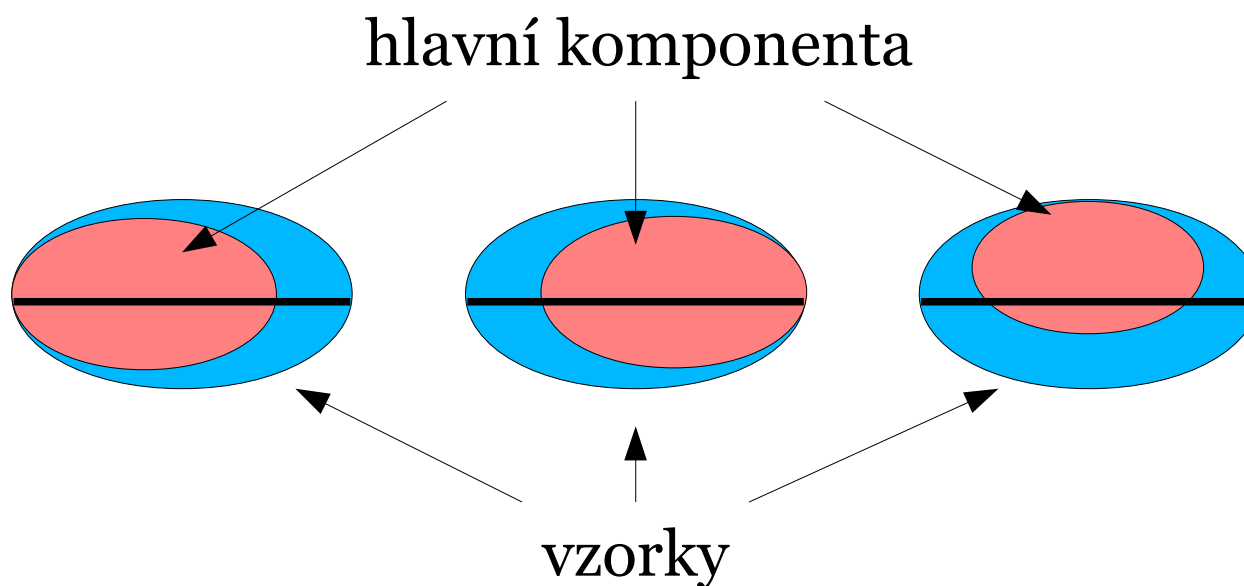
$$\{P_1, P_2, \dots, P_n\} \rightarrow F(k_1, \dots, k_m)$$

- dokáže popsat nekonečně velkou populaci ... generování exemplářů
- nástroj: **Analýza hlavních komponent** („Principal Component Analysis“ = PCA)
 - ◆ exemplář se skládá z příspěvků různých komponent
 - ◆ hledáme komponenty, které nejvíc přispívají do každého exempláře ve vzorku (nesou nejvíce informace)

Analýza hlavních komponent



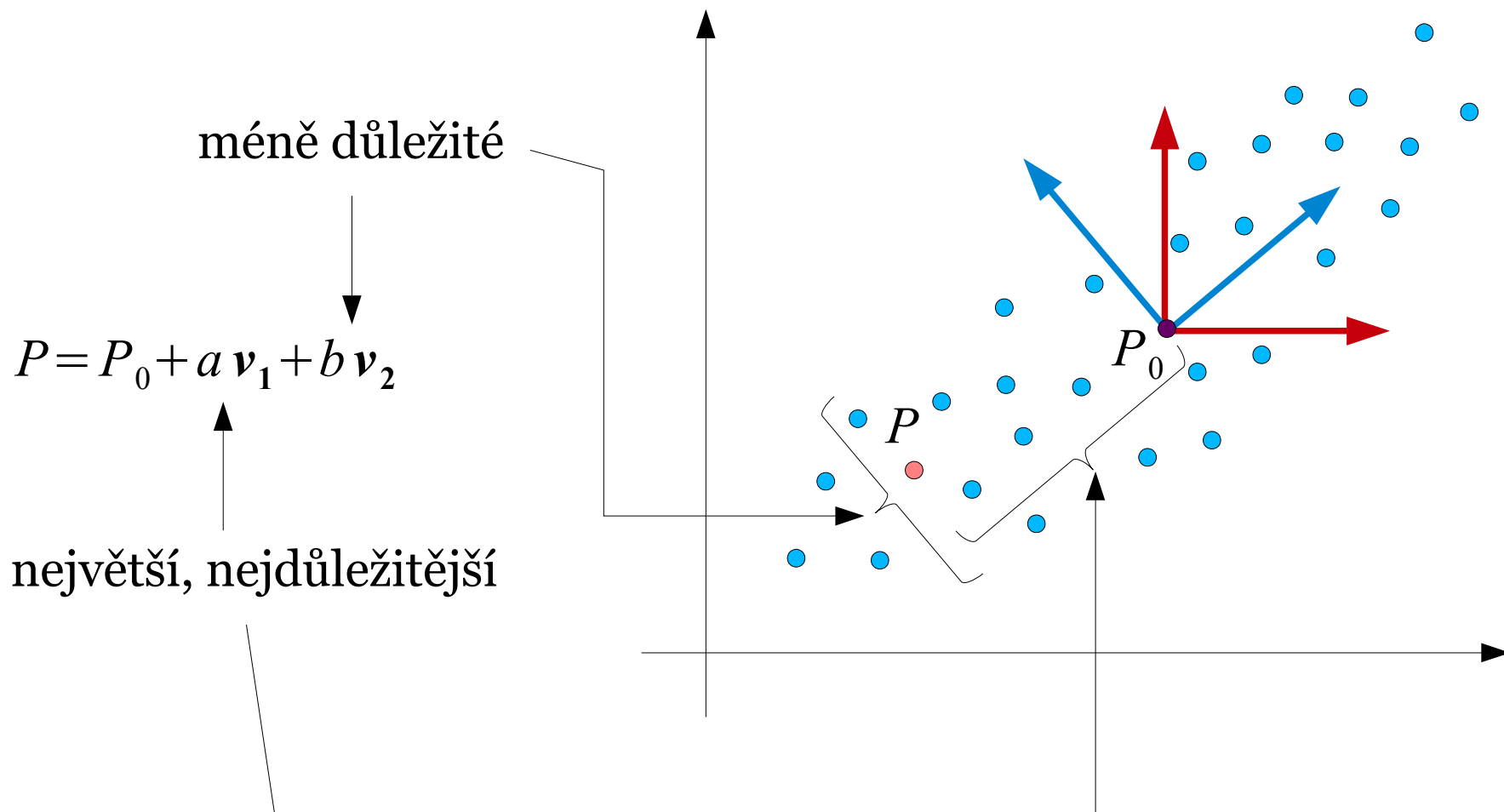
- ◆ máme **N-tici parametrů** popisující exemplář
- ◆ potřebujeme **oddělit významnější složky** od méně významných
 - ◆ např. je potřeba zredukovat dimenzi prostoru kvůli objemu dat, náročnosti měření, apod.
 - ◆ nebo hledáme minimální dimenzi parametrů popisující dobře strukturu nějaké populace (postačující pro separaci druhů..)
 - ◆ pro (interaktivní) vizualizaci a animace je přehledné pohybovat se jen v 2D prostoru parametrů..



◆ matematické vyjádření

- ◆ komponenta $\mathbf{v}_1 = [v_{1,1}, \dots, v_{1,n}]$ $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$
- ◆ vzorek $P = P_0 + \sum \mathbf{v}_i k_i$ $P_j \rightarrow \{k_{j,1}, \dots, k_{j,n}\}$
- ◆ hodnoty $k_i \rightarrow$ *PCA score*
- ◆ komponenty jsou seřazeny podle míry přispění!

PCA II



PCA matematika



Matematically je to jen
lineární transformace souřadnic (změna baze).

1. kovarianční matice S:

$$s_{ij}^2 = \frac{1}{n-1} \sum_{k=1}^n (x_{ki} - \bar{x}_i) \cdot (x_{kj} - \bar{x}_j)$$

2. vlastní vektory matice S tvoří PCA bazi

3. komponenty se setřídí podle velikosti příslušných
vlastních čísel

4. suma velikostí vlastních čísel ... celková informace
(může se počítat podíl informace PCA složek)



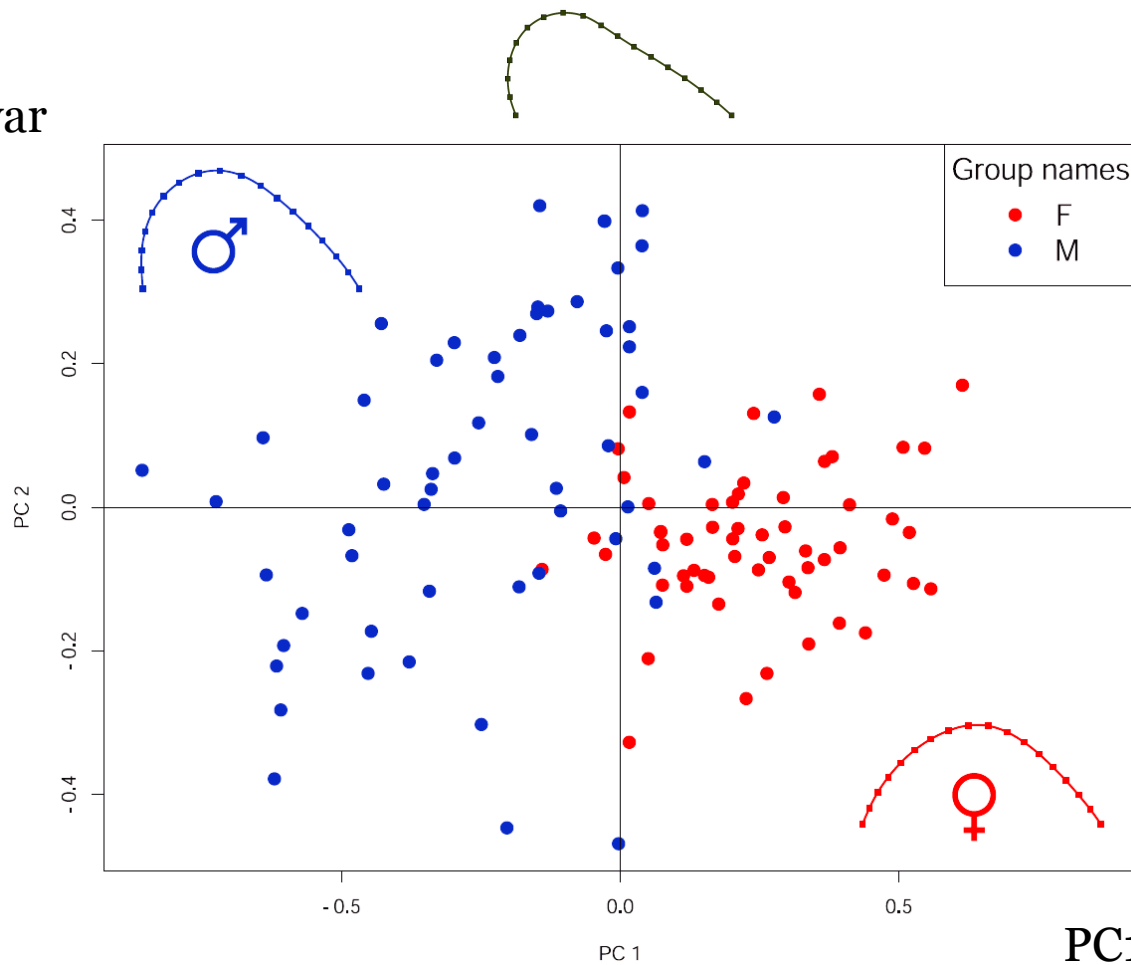
PCA příklady

- ◆ sada 2D, 3D landmarků
 - ◆ klasická úloha GMM
- ◆ 2D křivka
 - ◆ kontura
 - ◆ příklad s incisurami (ženy / muži)
- ◆ trojúhelníková síť
 - ◆ po unifikaci DMC (stejný počet vrcholů všech sítí)
 - ◆ příklad s obličejovými skeny

PCA příklad – incisury



PC2 ... 27% var



PC1 ... 50% var



Statistické metody

- ◆ klasické přístupy známé desetiletí
 - ◆ testování hypotéz (stejná či odlišná distribuce..)
 - ◆ rozhodování na zvolené hladině významnosti
- ◆ N-rozměrná analýza
 - ◆ jedinec .. vektor čísel (souřadnice landmarků, koeficienty, vzdálenosti, úhly, povrchy, objemy..)
- ◆ jsou dvě množiny statisticky významně oddělené?
 - ◆ pro klasifikaci v terénu, forenzní aplikace..
 - ◆ diskriminační analýza (často LDA), další klasifikační postupy (SVM)

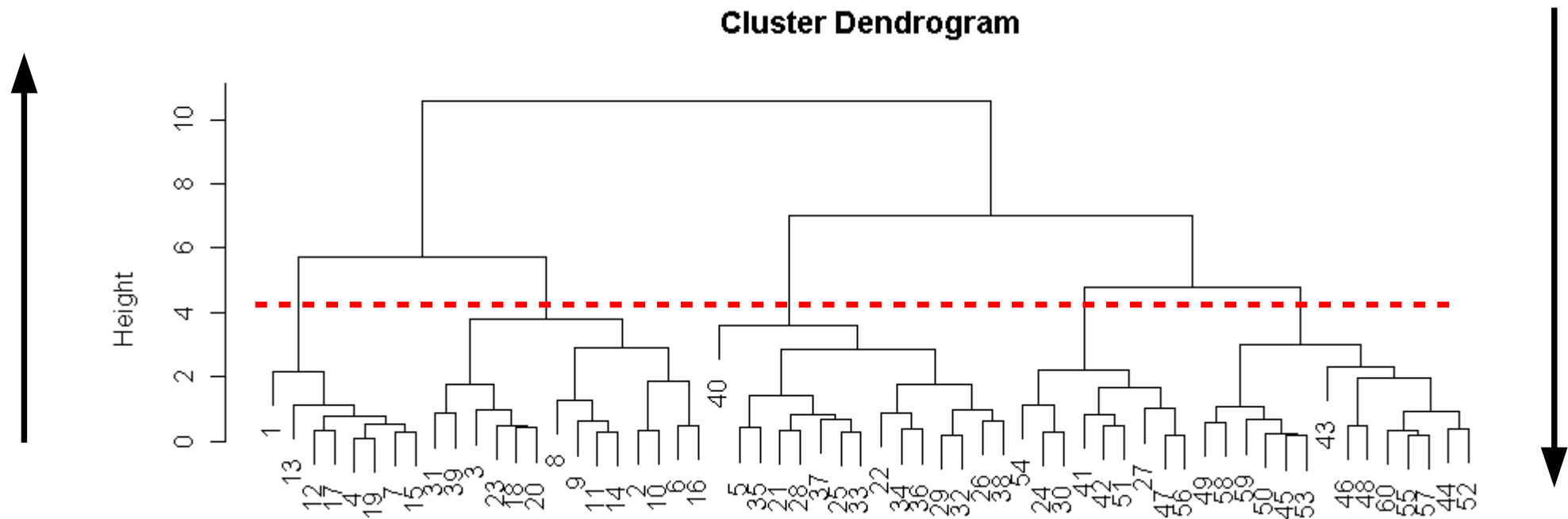
Shluková analýza (Cluster Analysis)

- ◆ **vstup:** soubor k N -rozměrných vektorů popisujících k jedinců
- ◆ **výstup:** dendrogram (hierarchie shlukování)
 - ◆ postupné slévání / oddělování skupin na různých hladinách významnosti
- ◆ problémy:
 - ◆ jakou zvolit metriku (diameter, minimum, mean dist.)
 - ◆ neuniformní distribuce entropie v jednotlivých dimenzích (každá složka N -tice má jinou váhu / důležitost)

Shluková analýza – příklad



- hierarchické shlukování .. výsledný dendrogram:



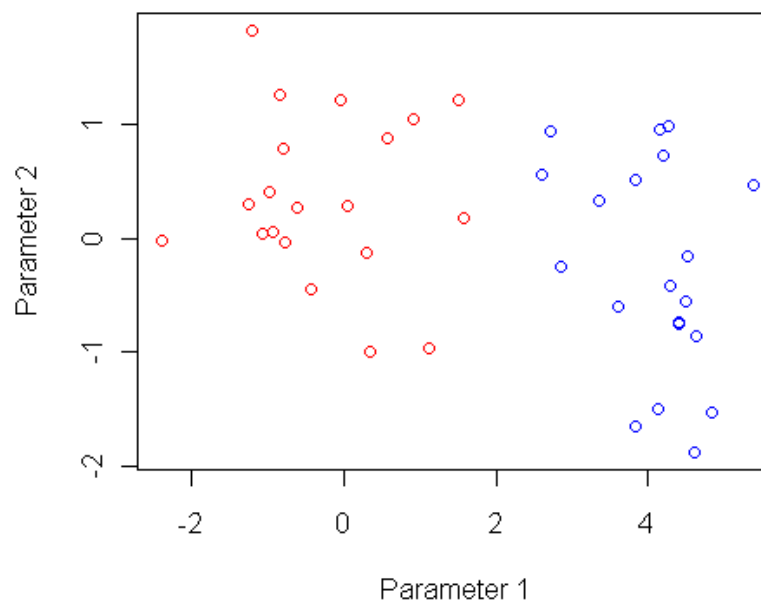


Diskriminační analýza

- ◆ dichotomie – rozdělitelnost souboru na dvě skupiny
- ◆ hledání **diskriminační funkce f** , která jedince \mathbf{x}
 - $f(\mathbf{x}) > 0$ přiřadí do první skupiny
 - $f(\mathbf{x}) < 0$ přiřadí do druhé skupiny
- ◆ lineární diskriminační analýza (**LDA**) → f je lineární

$$f(\mathbf{x}) = \mathbf{w} \mathbf{x} + c$$

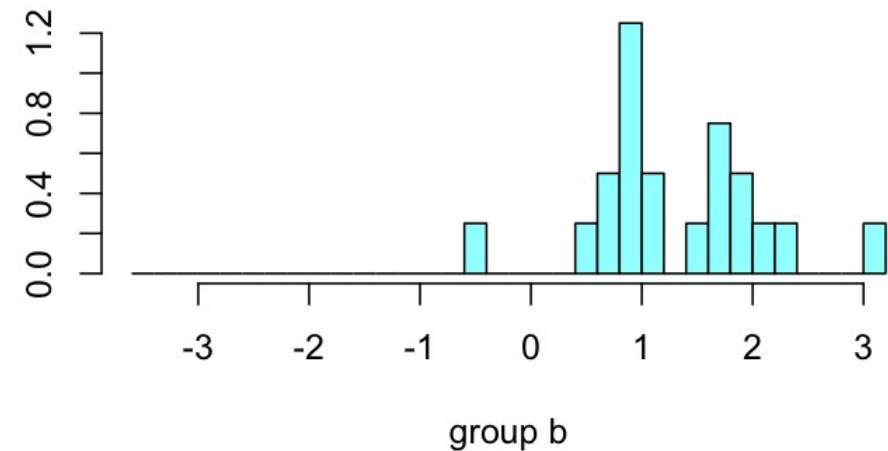
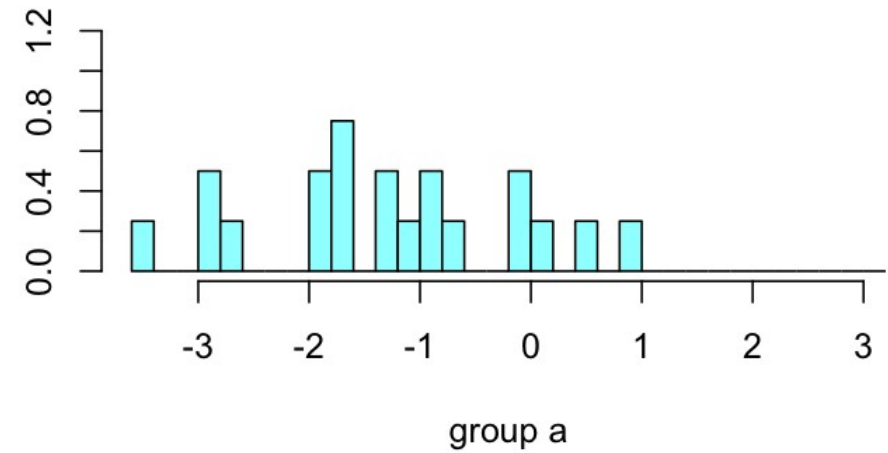
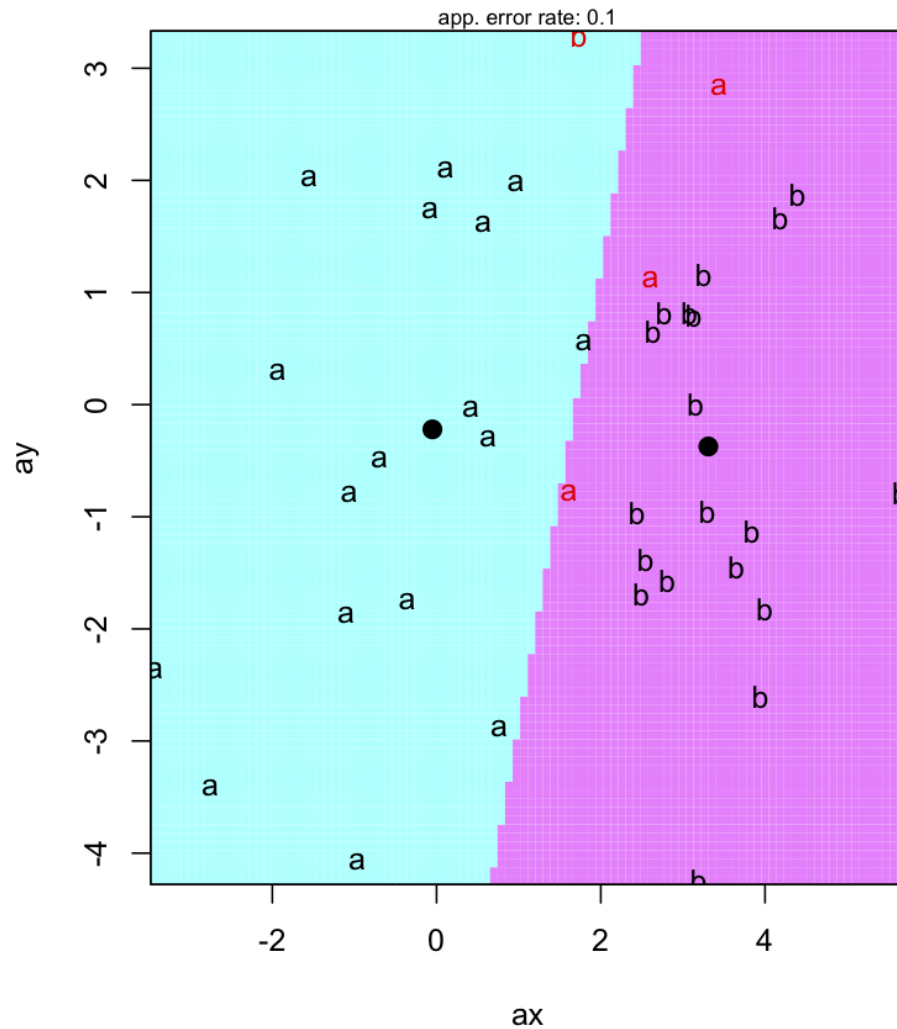
- ◆ hledání vektoru \mathbf{w} a skaláru c
- ◆ dimenze dat je libovolná
- ◆ rozšíření pro více skupin



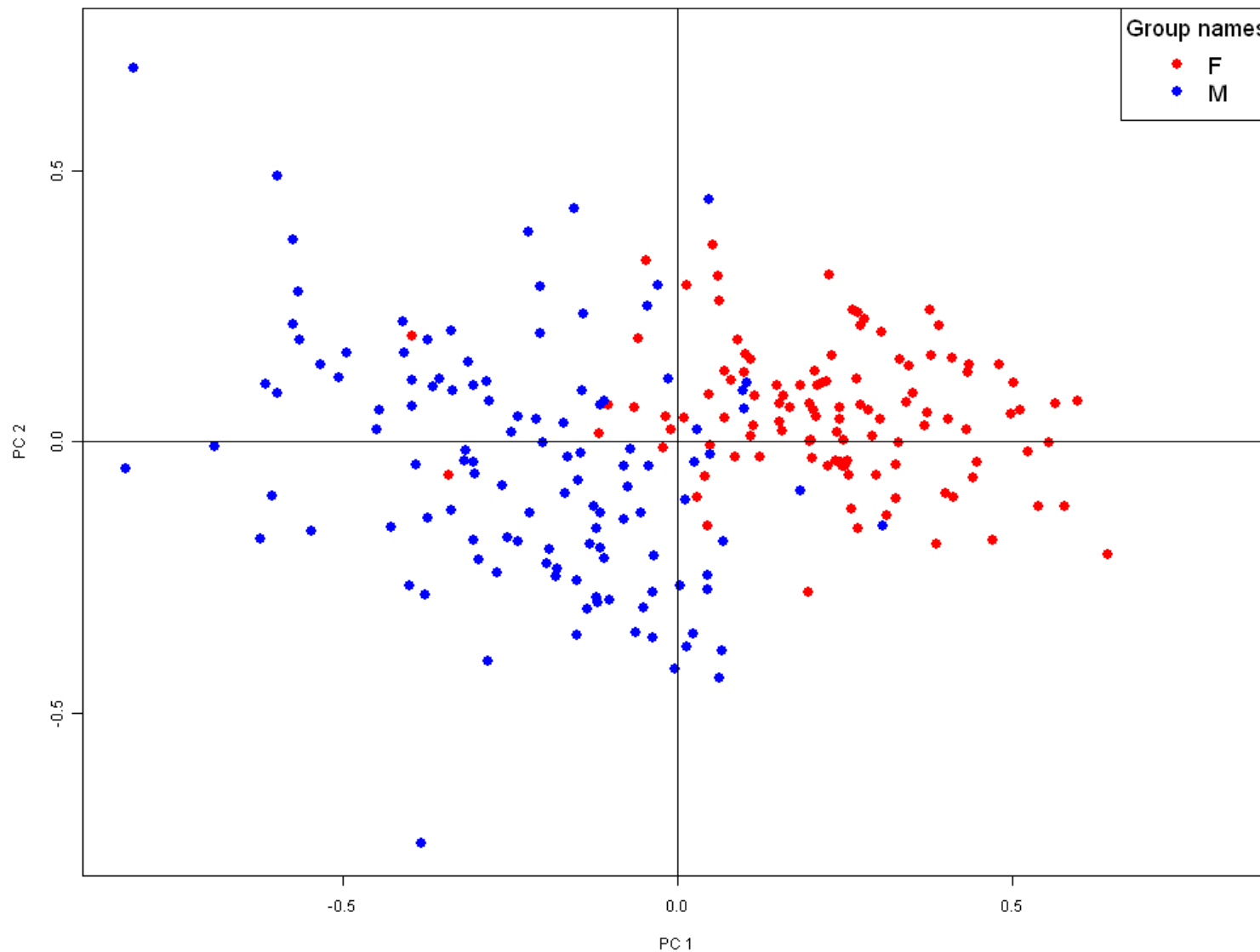
LDA – příklad



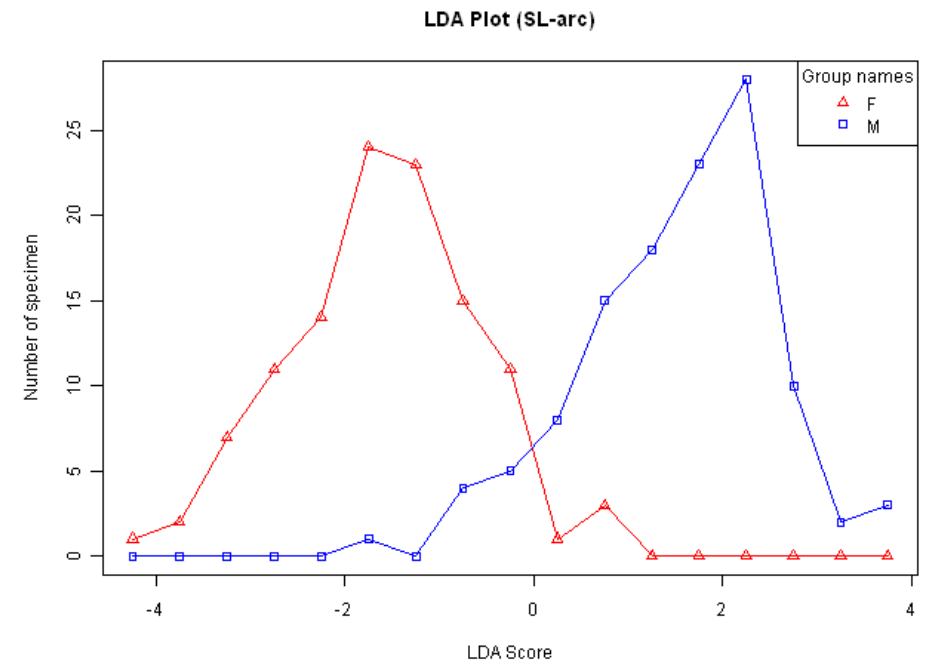
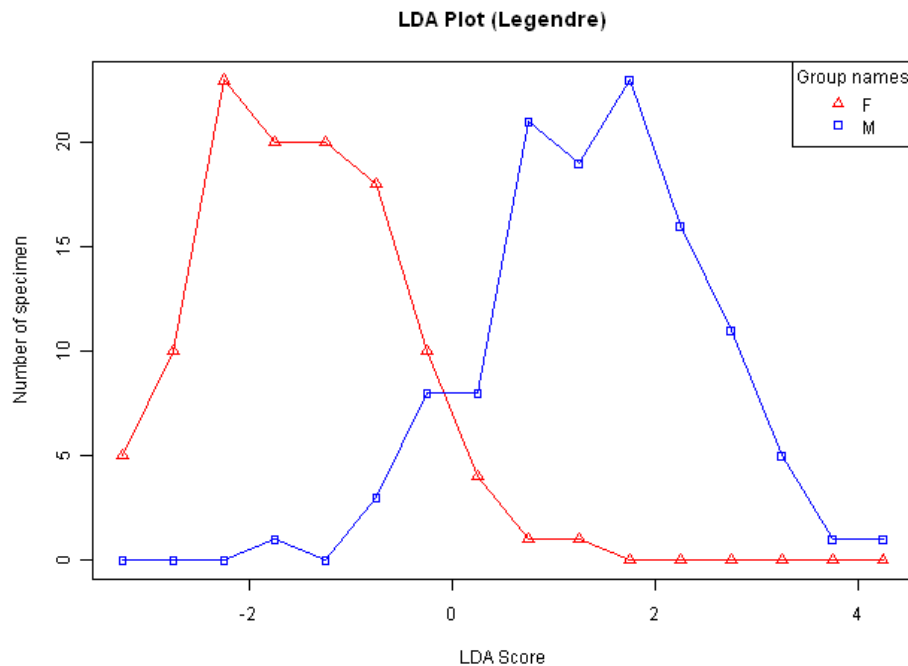
Partition Plot



LDA – incisury (rozdělení podle PCA)



LDA – incisury (dvě LDA)



LDA pro dvě různé reprezentace křivek

(zobrazuje se distribuce souřadnice kolmé na oddělovací nadrovinu ... číselně „t-value”)

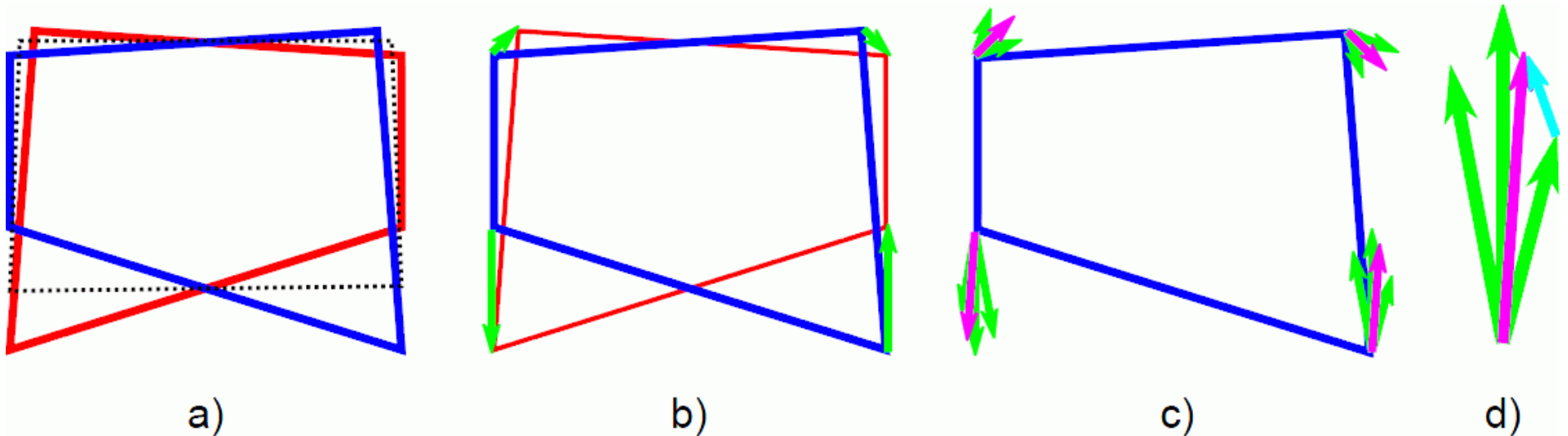


Studie symetrie

- ◆ **symetrie**
- ◆ **asymetrie** (nejčastěji se vyskytuje trend jedním směrem)
- ◆ **anti-symetrie** (nevyskytují se symetrická data)

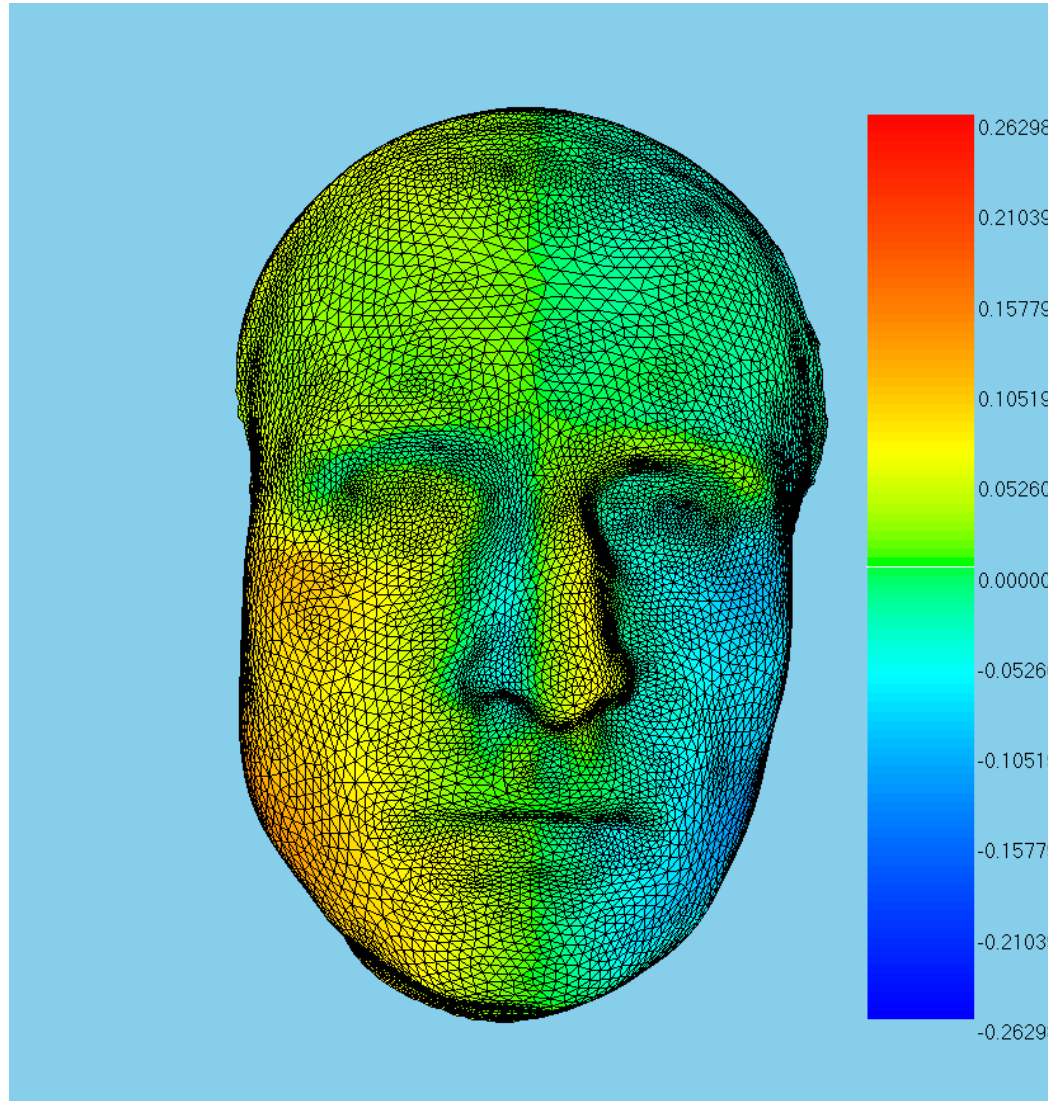
- ◆ **hledání roviny symetrie**
 - ◆ nemusí být jednoznačně definovaná (např. povrch obličeje člověka)
 - ◆ matematický přístup: minimalizace kvadratické vzdálenosti plochy a jejího zrcadlového obrazu

Analýza asymetrie



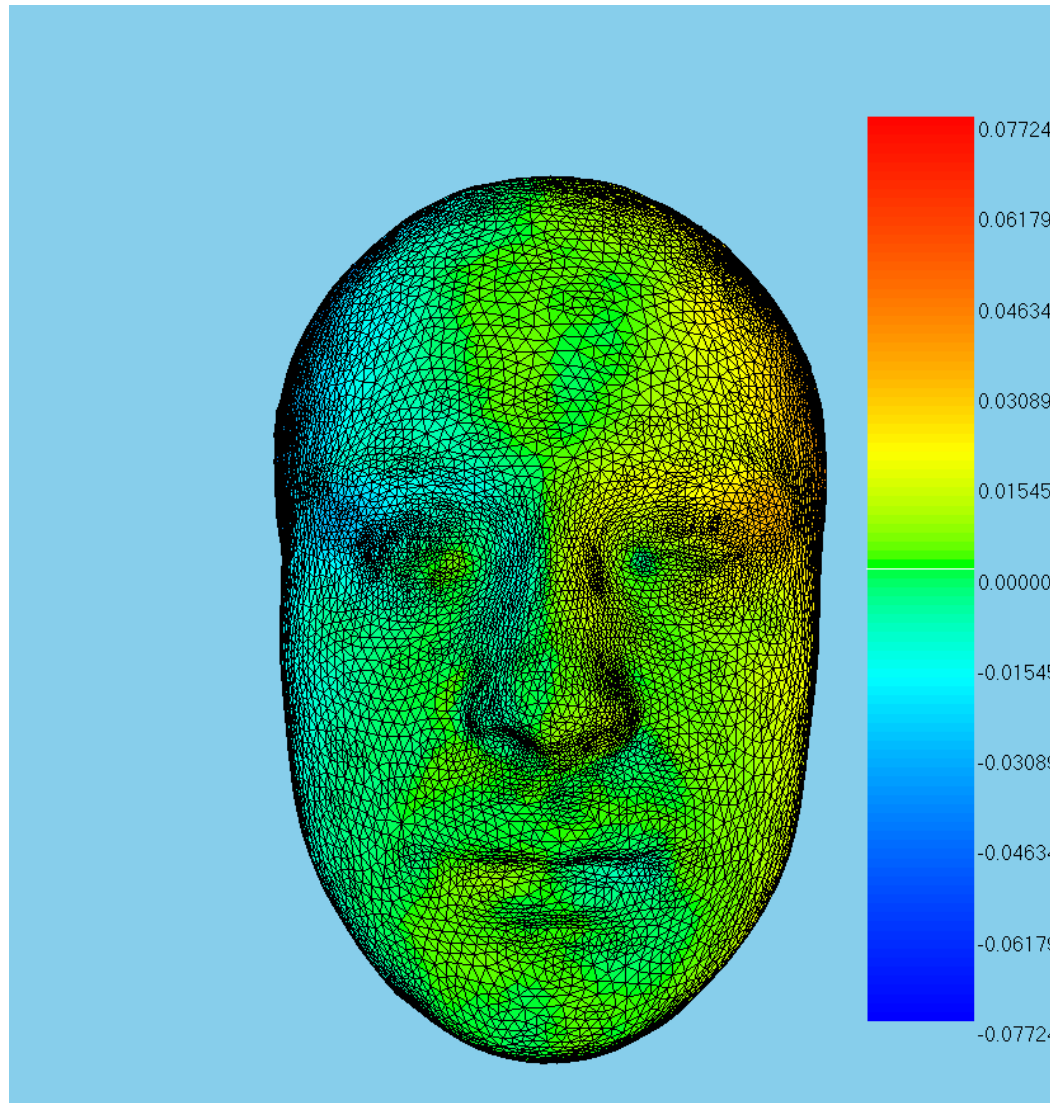
- a) konstrukce zrcadlového a dokonale symetrického jedince
- b) individuální asymetrie (rozdíly mezi landmarky pův. a zrcadl.)
- c) direkcionální asymetrie (zprůměrované individuální asymetrie)
- d) fluktuační asymetrie (rozdíl individuální a direkcionální asymetrie)

Asymetrie – příklady



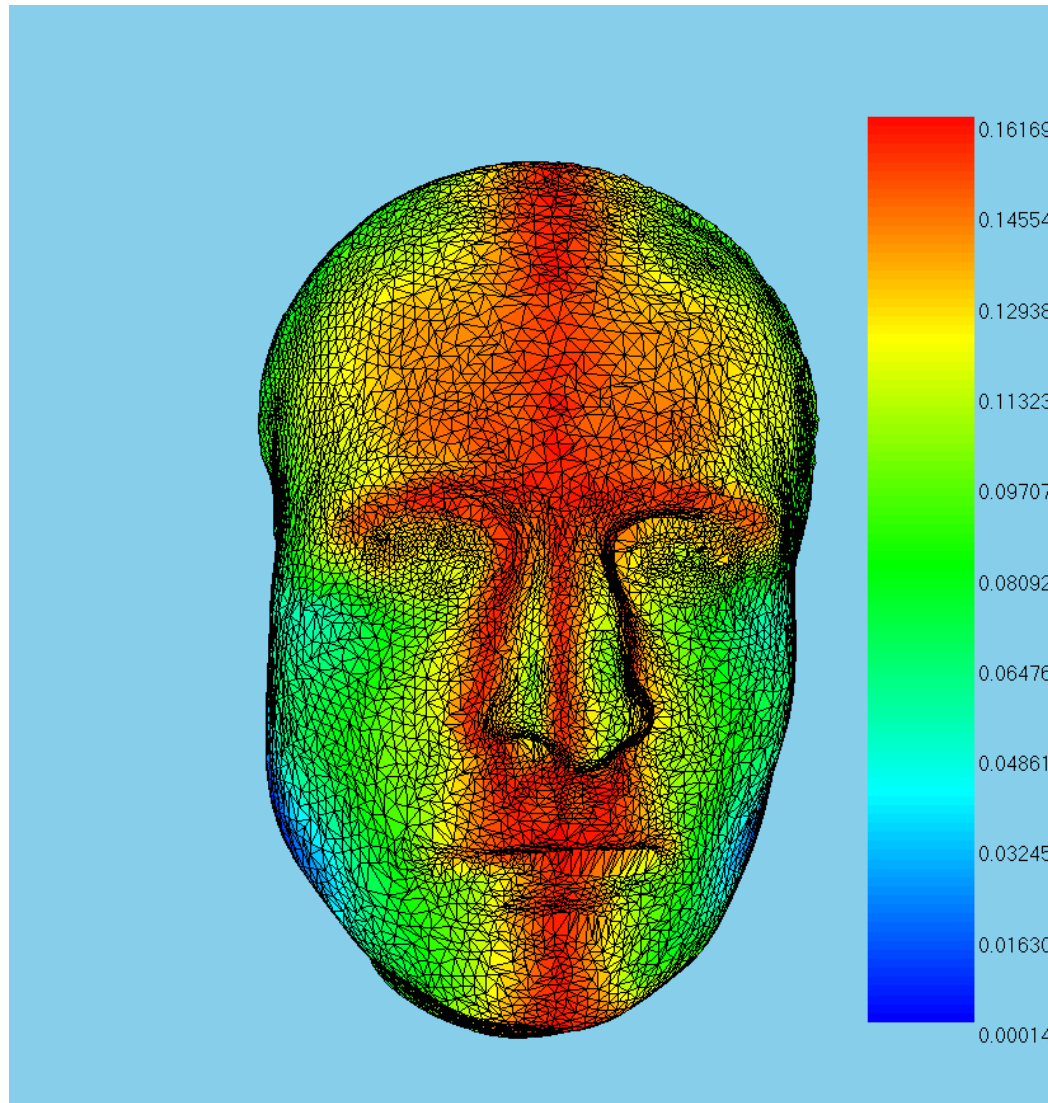
Individuální asymetrie
(orientované rozdíly jsou
znázorněny barvami)

Asymetrie – příklady



Direkcionální asymetrie
(zprůměrované individuální
asymetrie)

Asymetrie – příklady



Fluktuační asymetrie
(rozdíl individuální a
direkcionální asymetrie)



Hledání průměrného jedince

- ◆ průměr je důležitý pro další analýzu
 - měří se proti němu nový jedinec
 - průměr reprezentuje skupinu (ženy, muži)
- ◆ **N-tice** popisující jedince (triviální)
 - jenom musí být model odchylek symetrický
- ◆ plochy v 3D
 - ◆ obtížný úkol
 - ◆ parametrizace ploch (PCA) a potom průměr hodnot parametrů..
 - ◆ použití zavedených (klasických) landmarků

Literatura – knihy



- ◆ Zelditch, Swiderski, Sheets, Fink: *Geometric morphometrics for biologists: the primer*, Elsevier, 2004
- ◆ Dennis E. Slice: *Modern Morphometrics in Physical Anthropology*, Kluwer Academic, 2005
- ◆ Lele, Richtsmeier: *An Invariant Approach to Statistical Analysis of Shapes*, CRC Press, 2001

Literatura



- ◆ Kendall: *A Survey of the Statistical Theory of Shape*, Statistical Science, 1989
- ◆ Bookstein: *Principal Warps: Thin-Plate Splines and the Decomposition of Deformations*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 1989
- ◆ Mitteroecker, Gunz: *Advances in Geometric Morphometrics*, Evol. Biol., 2009
- ◆ Adams, Rohlf, Slice: *Geometric morphometrics: ten years of progress following the 'revolution'*, Ital. J. Zool, 2004