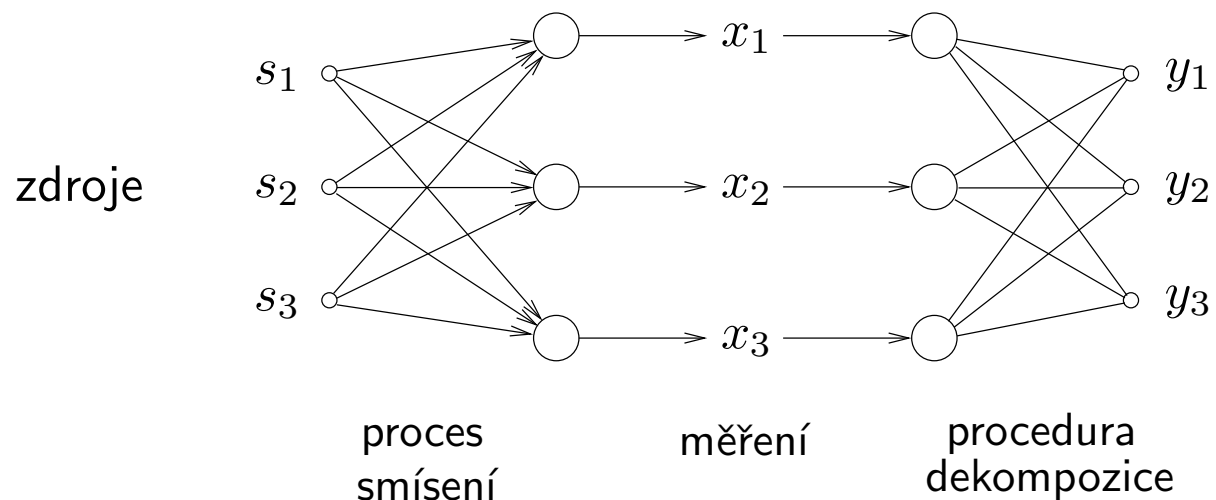


PCA nerozkládá na nezávislé složky



Pokud je měření x_i neznámou směsí signálu z malého počtu zdrojů $x_i = \mathbf{a}_i^\top \mathbf{s}$

- tak můžeme požadovat separaci na komponenty $y_i = \mathbf{w}_i^\top \mathbf{X}$ tak, aby měly vlastnosti původních složek s_i
- volba: hlavní vlastností je nezávislost
- nebude možné určit rozptyl s_i , protože $x_i = \mathbf{a}_i^\top \mathbf{s} = \mathbf{a}_i^\top \sigma \frac{1}{\sigma} \mathbf{s}$
- komponenty nebudou mít ‚pořadí‘ jako v PCA
- zdroje musí být negausovské náhodné proměnné, protože ortogonální transformace \mathbf{R} gausovské náhodné proměnné je gausovská a u normálního rozdělení je nekorelovanost a nezávislost totéž $x_i = \mathbf{a}_i^\top \mathbf{s} = \mathbf{a}_i^\top \mathbf{R}^\top \mathbf{R} \mathbf{s}$
- pomocí PCA dostaneme nekorelované složky, ale ne nezávislé, protože když místo $x_i = \sum_j a_{ij} s_j$ máme $x_i = \sum_j a_{ij} f(s_j)$, tak dostaneme jiné řešení

ICA: míra nezávislosti

Hledáme $y_i = \mathbf{w}_i^T \mathbf{X}$ tak, aby y_i byly navzájem nezávislé, tj, aby platilo

$$\begin{array}{l} 1 \times k \\ (1 \times d)^T \end{array} \quad \begin{array}{l} d \times k \\ \end{array}$$

$$p(y_1, \dots, y_s) = \prod_{i=1}^s p(y_i)$$

Pokud jsou y_i nezávislé, platí

tento vzorec pro $h_i(p) = \log(p)$ známe z vlastností entropie

$$E\{h_1(y_1) \cdot h_2(y_2) \cdots h_s(y_s)\} = \prod_{i=1}^s E\{h_i(y_i)\}$$

Vhodnou mírou nezávislosti je vzájemná informace

$$I(y_1, \dots, y_s) = \sum_{i=1}^s H(y_i) - H(y_1, \dots, y_s)$$

pokračování

Problém $y_i = \mathbf{w}_i^\top \mathbf{X}$, $i = 1, 2, \dots, s$ lze zapsat v maticovém tvaru $\mathbf{y} = \mathbf{W}^\top \mathbf{X}$, kde \mathbf{W} je matice $s \times d$

Pokud $\mathbf{y} = \mathbf{W}^\top \mathbf{X}$, tak platí

$$f(y_1, \dots, y_s) = \frac{1}{\det \mathbf{W}} f(x_1, x_2, \dots, x_s) \quad s \leq d$$

$y_i = \mathbf{w}_i^\top \mathbf{X}$

$$I(y_1, \dots, y_s) = \sum_i H(y_i) - \underbrace{H(x_1, \dots, x_k)}_{\log \det \mathbf{W}}$$

$$H(y_1, \dots, y_s) = \int f(y_1, \dots, y_s) \log f(y_1, \dots, y_s) dy_1 \dots dy_s = \int f(\mathbf{x}) \log \frac{f(\mathbf{x})}{\det \mathbf{W}} d\mathbf{x}$$

Když y_i budou mít jednotkovou kovarianční matici, tak

$$1 = \det \mathbf{I} = \det \mathbf{y}\mathbf{y}^\top = \det (\mathbf{W}^\top \mathbf{X}\mathbf{X}^\top \mathbf{W}) = (\det \mathbf{W})^2 \det \mathbf{X}\mathbf{X}^\top$$

takže $\det \mathbf{W}$ je konstanta. Z toho plyne, že

$$I(y_1, \dots, y_s) = C - \sum_{i=1}^s H(y_i)$$

$\swarrow \mathbf{w}_i^\top \mathbf{X}$

ICA: aproximace kritériální funkce

Hledáme tedy \mathbf{w}_i tak, že

$$\mathbf{W}^* = \arg \max_{\mathbf{W}} \sum_{i=1}^s H(\mathbf{w}_i^\top \mathbf{X}), \quad \text{kde } \mathbf{w}_i \text{ jsou řádky matice } \mathbf{W}$$

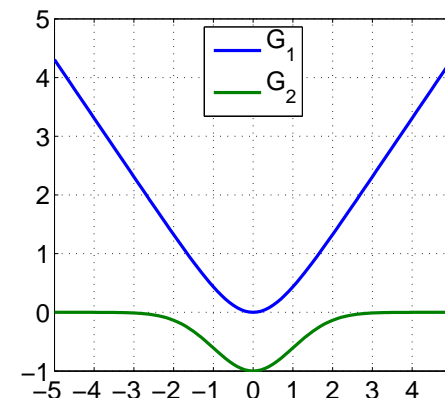
Z výpočetních důvodů se entropie $H(u) = E\{-\log(u)\}$ aproximuje funkcí

$$J(y) = (E\{G(y)\} - E\{G(\nu)\})^2,$$

$$G(y) = y^2 \text{ kulha}$$

kde $\nu \sim N(0, \mathbf{E})$ a typické volby jsou

$$G_1(u) = \frac{1}{a_1} \log \cosh a_1 u \quad \text{nebo} \quad G_2(u) = -e^{-a_2 \frac{u^2}{2}}$$



$$a_1 = a_2 = 1$$

Procedura ICA

Postup pro $s = 1$

1. normalizuj data $\bar{\mathbf{x}}_i = \mathbf{S}^{-\frac{1}{2}}(\mathbf{x}_i - \mathbf{x}_0)$ pro $i = 1, 2, \dots, k$,
kde $\mathbf{S} = \sum_{i=1}^k (\mathbf{x}_i - \mathbf{x}_0)(\mathbf{x}_i - \mathbf{x}_0)^\top$ je kovarianční matice a $\mathbf{S} = \mathbf{S}^{\frac{1}{2}} \mathbf{S}^{\frac{1}{2}}$

Choleski

2. zvol náhodný vektor \mathbf{w} tak, aby $\|\mathbf{w}\| = 1$

3. necht $\mathbf{w}^+ = \sum_{i=1}^k \bar{\mathbf{x}}_i G'(\mathbf{w}^\top \bar{\mathbf{x}}_i) - G''(\mathbf{w}^\top \bar{\mathbf{x}}_i) \mathbf{w}$

Newton

4. renormalizuj $\mathbf{w} = \frac{\mathbf{w}^+}{\|\mathbf{w}^+\|}$

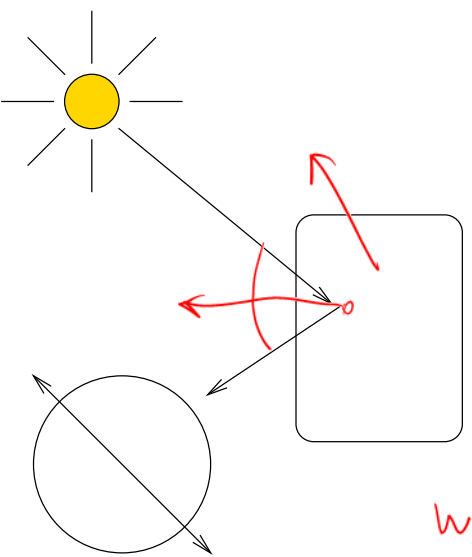
5. opakuj kroky 3–4 do konvergence

6. nalezená složka je $y = \mathbf{w}^\top \bar{\mathbf{X}}$

y je vektor délky k

- $G'(u)$ je derivace $\frac{dG(u)}{du}$
- dobrá implementace je podstatně složitější

Aplikace ICA: separace odlesků



analyzátor polarizace

$s = 3$

w_1
 w_2
 w_3

$x_1 = k \times 1$

vstupní obrazy x_2

x_3

komponenta 1

komponenta 2

komponenta 3

The diagram shows a yellow sun icon emitting light rays towards a rectangular box representing a polarizing filter. A red arrow indicates the direction of polarization. Below the box is a circular polarizer analyzer with a diagonal line and two arrows. To the right of the diagram are three columns of images. The top row shows the original input images, and the bottom row shows the separated components. Handwritten red text includes $x_1 = k \times 1$ above the first image, x_2 above the second, and x_3 above the third. The labels 'komponenta 1', 'komponenta 2', and 'komponenta 3' are placed below each column. The weights w_1 , w_2 , and w_3 are written in red next to the $s = 3$ label.

- rovina polarizace závisí na úhlu normály \Rightarrow pohyb odlesku
- 3. komponenta částečně obsahuje tento pohyb

Fisherova lineární diskriminační analýza

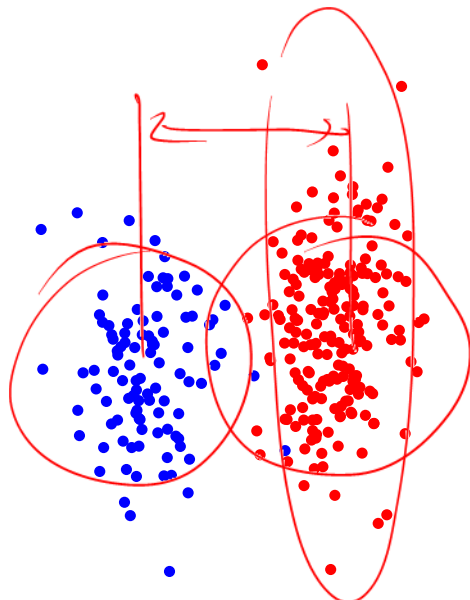
Dána množina vektorů $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ v prostoru dimenze d , u každého identifiátor třídy ω_i , $i = 1, 2, \dots, k$, $\omega_i \in \Omega$ (Ω je množina velikosti o), po sloupcích sestavíme \mathbf{X}^ω pro třídu $\omega \in \Omega$.

Nechť \mathbf{W} je matice $s \times d$, $s < d$, která redukuje dimenzi vektorů lineární transformací

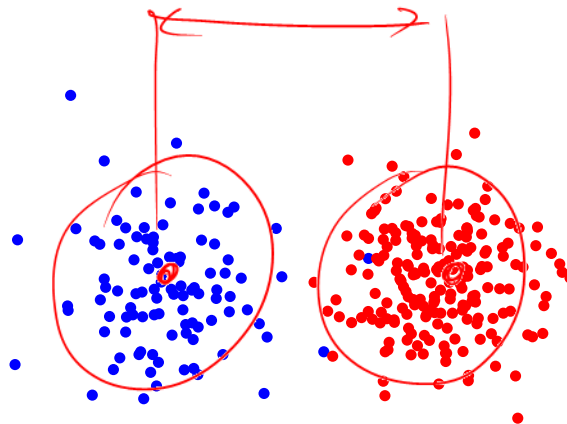
$$\mathbf{Y} = \mathbf{W}\mathbf{X}.$$

Budeme hledat takovou projekci, ve které jsou třídy co nejdále od sebe, měřeno vzhledem k rozptylu uvnitř tříd.

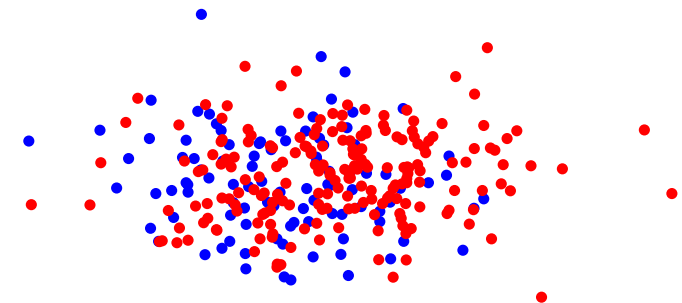
Př: body v dimenzi $d = 3$



projekce $x - y$



projekce $x - z$



projekce $y - z$

Úloha pro $o = 2$

Promítáme $\mathbf{y} = \mathbf{w}^\top \mathbf{X}$, kde \mathbf{y} je ‚matice‘ dimenze $1 \times k$.

Před projekcí jsou vektory středních hodnot

$$\mathbf{m}^\omega = \frac{1}{k^\omega} \sum_{i=1}^{k^\omega} \mathbf{x}_i^\omega, \quad \text{kde } k^\omega \text{ je počet prvků ve třídě } \omega$$

Vzdálenost mezi třídami po projekci měříme rozdílem středních hodnot

and good!

$$|n^1 - n^2|^2 = |\mathbf{w}^\top \mathbf{m}^1 - \mathbf{w}^\top \mathbf{m}^2|^2 = \dots = \mathbf{w}^\top (\mathbf{m}^1 - \mathbf{m}^2)(\mathbf{m}^1 - \mathbf{m}^2)^\top \mathbf{w} \stackrel{\text{def}}{=} \mathbf{w}^\top \mathbf{S}_b \mathbf{w}$$

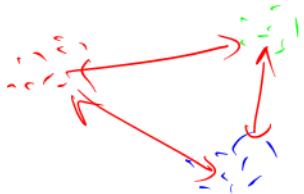
Rozptyl uvnitř tříd po projekci měříme jako

$$(s^1)^2 + (s^2)^2 = \sum_{i=1}^{k^1} |\mathbf{w}^\top (\mathbf{x}_i^1 - \mathbf{m}^1)|^2 + \sum_{j=1}^{k^2} |\mathbf{w}^\top (\mathbf{x}_j^2 - \mathbf{m}^2)|^2 = \dots = \mathbf{w}^\top (\mathbf{S}^1 + \mathbf{S}^2) \mathbf{w} \stackrel{\text{def}}{=} \mathbf{w}^\top \mathbf{S}_w \mathbf{w}$$

pokračování

Fisherův diskriminant je

známo jako Rayleigho poměr



$$F = \frac{|n^1 - n^2|^2}{(s^1)^2 + (s^2)^2} = \frac{\mathbf{w}^\top \mathbf{S}_b \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_w \mathbf{w}}.$$

Věta: Vektor \mathbf{w} maximalizující F musí pro nějaké $\lambda \in \mathbb{R}$ vyhovovat podmínce

$$\mathbf{S}_b \mathbf{w} = \lambda \mathbf{S}_w \mathbf{w}.$$

$$\hat{\mathbf{S}}_w^{-1} \mathbf{S}_b \mathbf{w} = \lambda \mathbf{w} \quad (1)$$

- Nalezení takového \mathbf{w} a λ se nazývá **zobecněný problém vlastních čísel**. Pokud je \mathbf{S}_w regulární, lze ho vyřešit jako obyčejný problém $\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{w} = \lambda \mathbf{w}$.
- λ je hodnota v optimu \Rightarrow vybereme \mathbf{w} odpovídající největšímu λ
- Pro $o = 2$ není třeba řešit (1), protože $\mathbf{S}_b \mathbf{w}$ je vždy paralelní s vektorem $\mathbf{m}^1 - \mathbf{m}^2$, takže

$$\mathbf{w} = \mathbf{S}_w^{-1} (\mathbf{m}^1 - \mathbf{m}^2)$$

normalizovaný na jednotkovou délku

důkaz: *1

- Pokud je málo měření a dimenze \mathbf{S} je velká, tak je \mathbf{S}_w singulární a problém (1) musíme řešit obecně.

V Matlabu pak použijeme $[\mathbf{W}, \mathbf{D}] = \text{eig}(\mathbf{S}_b, \mathbf{S}_w)$, \mathbf{D} obsahuje λ_i na diagonále a \mathbf{W} obsahuje zobecněné vlastní vektory po sloupcích

Úloha pro $o > 2$

Matice vnitrotřídního rozptylu zůstává $\mathbf{S}_w = \sum_{i=1}^o \mathbf{S}^i$.

Zavedeme vážené střední hodnoty

vážený průměr pro \mathbf{S}_b lze použít i pro $o = 2$

$$\mathbf{m} = \frac{1}{k} \sum_{i=1}^o k^i \mathbf{m}^i \quad \mathbf{S}_b = \sum_{i=1}^o k^i (\mathbf{m}^i - \mathbf{m})(\mathbf{m}^i - \mathbf{m})^\top \quad \text{Platí } \sum_{i=1}^k (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^\top = \mathbf{S}_w + \mathbf{S}_b$$

Všechny projekce sdružíme maticovým zápisem

$$\mathbf{Y} = \mathbf{W}\mathbf{X}, \quad \text{kde } \mathbf{Y} \text{ je } s \times k, \mathbf{W} \text{ je } s \times d \text{ a } \mathbf{X} \text{ je } d \times k \text{ matice}$$

Pak je Fisherův diskriminant

det – měří objem ‚elipsoidu kovariance‘ po projekci

$$F = \frac{\det(\mathbf{W}^\top \mathbf{S}_b \mathbf{W})}{\det(\mathbf{W}^\top \mathbf{S}_w \mathbf{W})}$$

Věta: Sloupce \mathbf{W} jsou tvořeny zobecněnými vlastními vektory odpovídajícími s největším zobecněným vlastním číslům $\mathbf{S}_b \mathbf{w}_i = \lambda_i \mathbf{S}_w \mathbf{w}_i, \quad i = 1, 2, \dots, s$

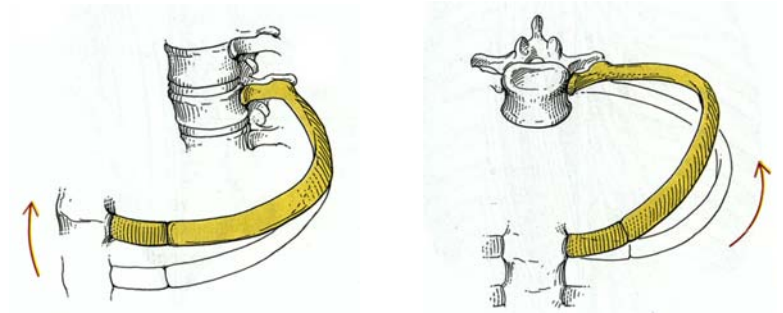
Příklad: Index kvality dechových pohybů

Ve spolupráci s FTVS UK Praha

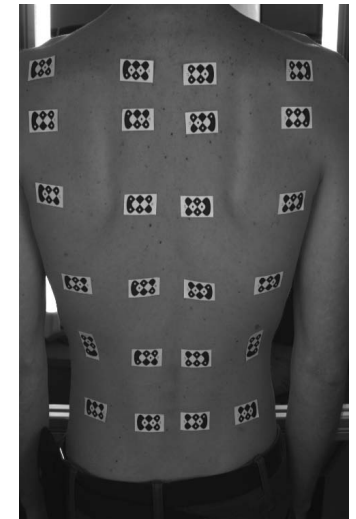
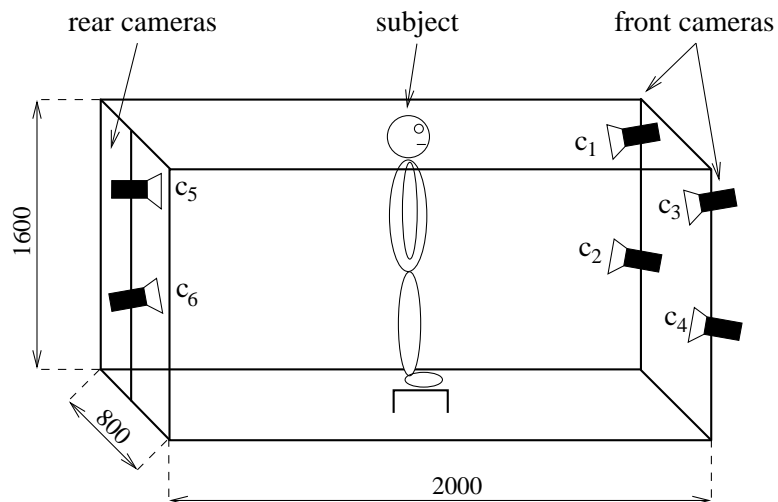
Tři dechové programy:

pump, abdominal, bucket handle

vybráno vzhledem k posturální funkci dechu

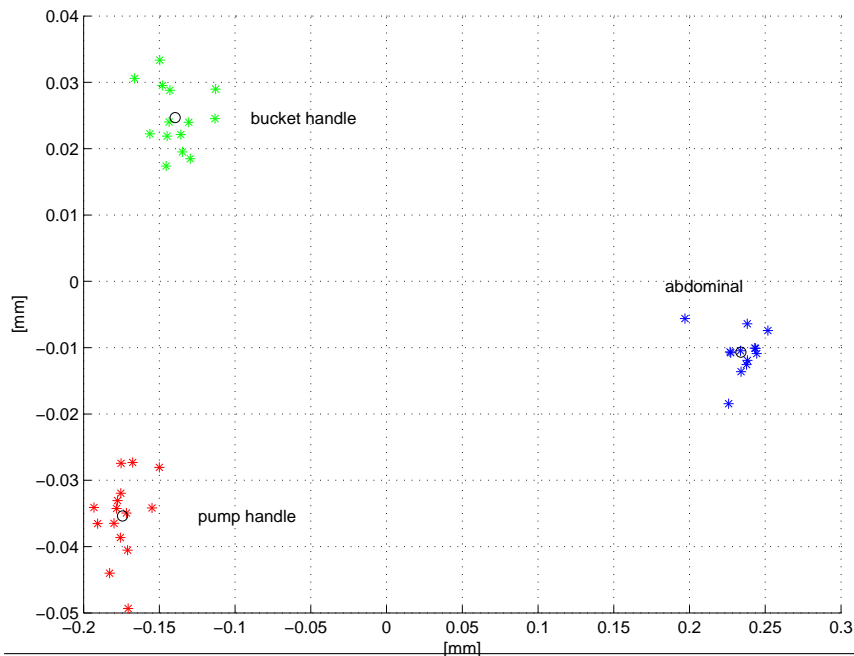


Cíl: Pro měřeného jedince určit podobnost se základními dechovými programy na základě 3D měření.



pokračování

normalizované měření v milimetrech									
linie	pump			abdominal			bucket		
	x	y	z	x	y	z	x	y	z
1	-0.16	0.09	0.18	0.07	0.03	-0.01	-0.22	0.05	0.43
2	-0.25	0.07	0.33	0.08	0.03	-0.02	-0.28	0.02	0.42
3	-0.38	0.03	0.42	0.08	0.04	-0.04	-0.33	0.03	0.22
4	-0.40	0.01	0.33	0.02	0.04	0.04	-0.41	0.04	0.03
5	-0.25	0.02	0.08	-0.02	0.04	0.57	-0.25	0.05	-0.08
6	-0.17	0.01	-0.01	0.15	0.04	0.70	-0.18	0.02	-0.03



- projekce do dimenze $s = 2$
- každý bod je jedno měření, barva je identifikátor typu dechu
- vizualizace
- rozlišení typu dechu je možné

Konec

