Teorie kognitivních systémů

X Slepá separace zdrojů

- Motivace k separaci zdrojů
- Definice problému separace
- Přehled metod slepé separace
- Independent Component Analysis
- Princip činnosti ICA
- Algoritmus ICA
- Vlastnosti ICA
- Použití ICA

Copyright © 2012 Kamil Ekštein

Slepá separace zdrojů informace (signálů)
Učení bez učitele

Slepá separace zdrojů (**BSS**, *Blind Signal/Source Separation*) je problém učení bez učitele (*Unsupervised Learning*). Je to:

- rodina výpočetních postupů, užívaných zejm. v DSP, statistice, klasifikaci a rozpoznávání, atp.
- cílem je výpočet **odhadu** původních signálů, jejichž směsi jsou vstupem algoritmu (tj. nejedná se o kouzelné postupy, které ze zašuměného signálu udělají samy čistý)
- BSS separuje signály ze směsí bez (nebo s minimálním) užití informace o původu či charakteru signálů nebo způsobu, jakým byly smíchány
- separované signály musí splňovat určité omezující požadavky
- existuje řada (více či méně účinných) metod slepé separace zdrojů, některé jsou vhodné jen pro určité konkrétní situace

Slepá separace zdrojů informace (signálů)
Základní vlastnosti

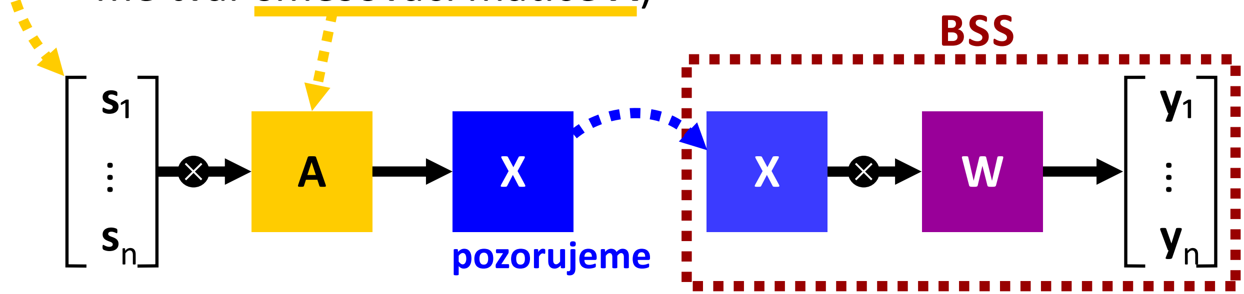
BSS lze úspěšně provádět pouze tehdy, jsou-li zdroje (tj. pův. signály, ze kterých vznikly zpracovávané směsi):

- **nekorelované** (nejdůležitější požadavek)
- statisticky nezávislé

BSS využívá pouze předpoklad vzájemné nezávislosti signálů

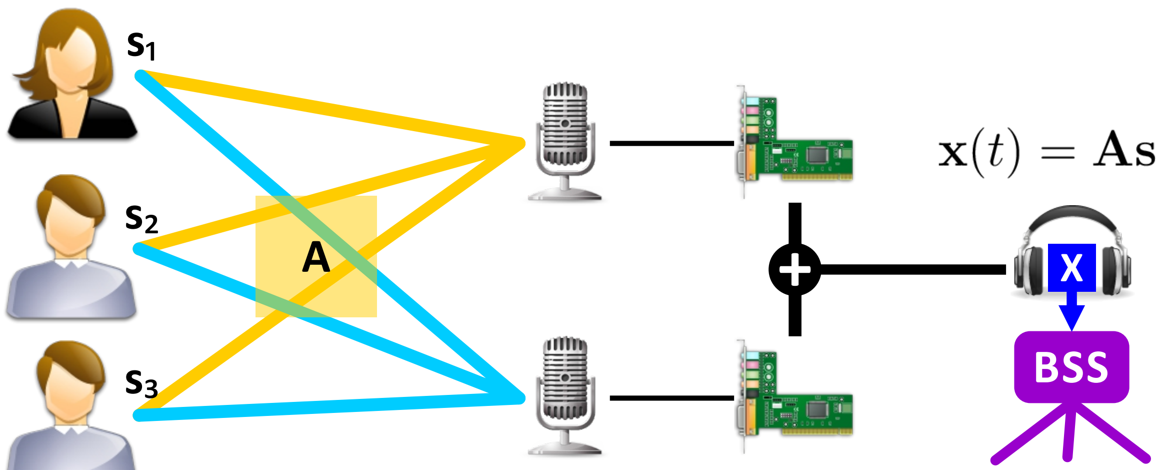
Proč **slepá**?

- zdrojové (pův.) signály **nepozorujeme** (nejsou k dispozici)
- není k dispozici informace o původu jejich směsi (tj. neznáme tvar směšovací matice **A**)



Slepá separace zdrojů informace (signálů)
Aplikace

Problém koktejlového večírku (*Cocktail Party Problem*) – velmi jednoduchá klasická ukázka aplikace BSS, poprvé po-popsaná Edwardem Colinem Cherrym roku 1957:



$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$$

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$$

Slepá separace zdrojů informace (signálů)
Praktické aplikace – DSP

- **magnetoencefalografie (MEG)** – separace signálů za účelem odstranění *artefaktů*, tj. následků rušení vnějšími zdroji elmg polí (např. kovové hodinky pacienta)
- **echo cancellation & dereverbation** – odečítání zpožděných (odrazem či přenosem) signálů, používá se u mobilních telefonů, Skype, apod.
- **potlačení šumu** (*Denoising*) – akustických signálů (1D) i obr. snímků (2D), také u multispektrálních (nD) snímků při dálkovém průzkumu Země
- **analýza komplexních systémů** – např. chemických provozů, jaderných elektráren, apod. na základě množství signálů z fyzikálně různých senzorů (mechanických, chemických, elmg), mezi senzory existují tzv. přeslechy, smyslem BSS je získat čisté signály bez přeslechů

Slepá separace zdrojů informace (signálů)
Přehled metod

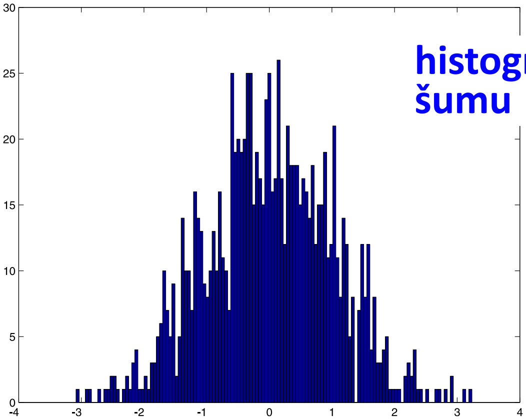
- **Independent Component Analysis** (ICA)
- Principal Component Analysis (PCA)
- Singular Value Decomposition (SVD)
- Dependent Component Analysis
- Non-negative Matrix Factorization (NMF)
- Low-complexity Coding & Decoding
- Stationary Subspace Analysis (SSA)
- Common Spatial Pattern (CSP)
- a další...

Řada z nich vychází z relativně starých metod a postupů aplikovaných v ekonometrii, statistice, chemii, apod.

Independent Component Analysis (ICA)
Definice

Analýza nezávislých komponent (ICA) – výpočetní postup pro separaci vícerozměrného signálu, resp. směsi signálů, na aditivní komponenty za předpokladu statistické nezávislosti **ne-gaussovských** zdrojových signálů...

➔ **POZOR:** Přímo z definice vyplývá, že ICA **nedokáže** ze směsi separovat např. **bílý šum** (je gaussovský)



histogram bílého šumu

Independent Component Analysis (ICA)
Definice z matematického pohledu

ICA je postup nalezení směšovací matice **A**, resp. jejího odhadu, tj. „rozmíchávací" matice **W**, $\mathbf{W} = \mathbf{A}^{-1}$, v daných vztazích

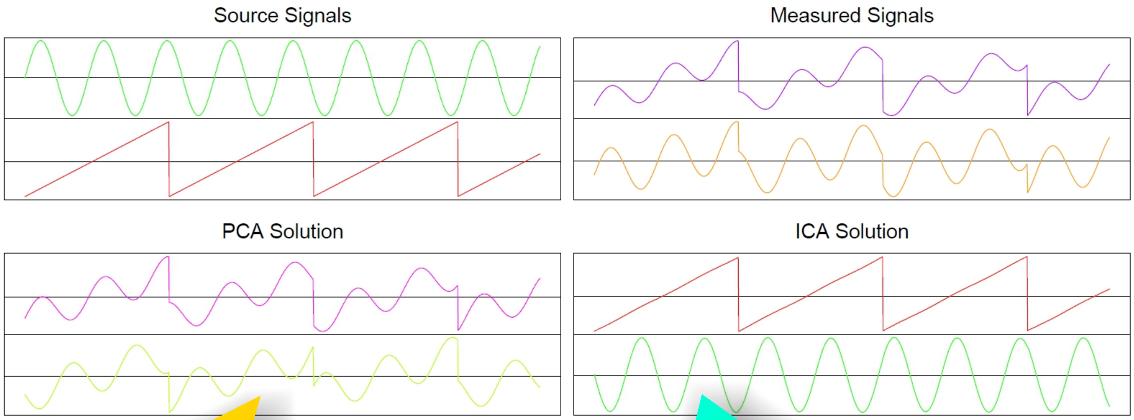
$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$$
$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$$

tak, aby statistická **nezávislost** separovaných komponent byla **nejvyšší možná**.

Míru statistické nezávislosti komponent je možné stanovit na základě řady různých metrik, např.:

- entropie, příp. negentropie
- vzájemné informace (*Mutual Information*)
- koeficientu špičatosti (*Kurtosis*)
- apod.

Independent Component Analysis (ICA)
ICA není PCA

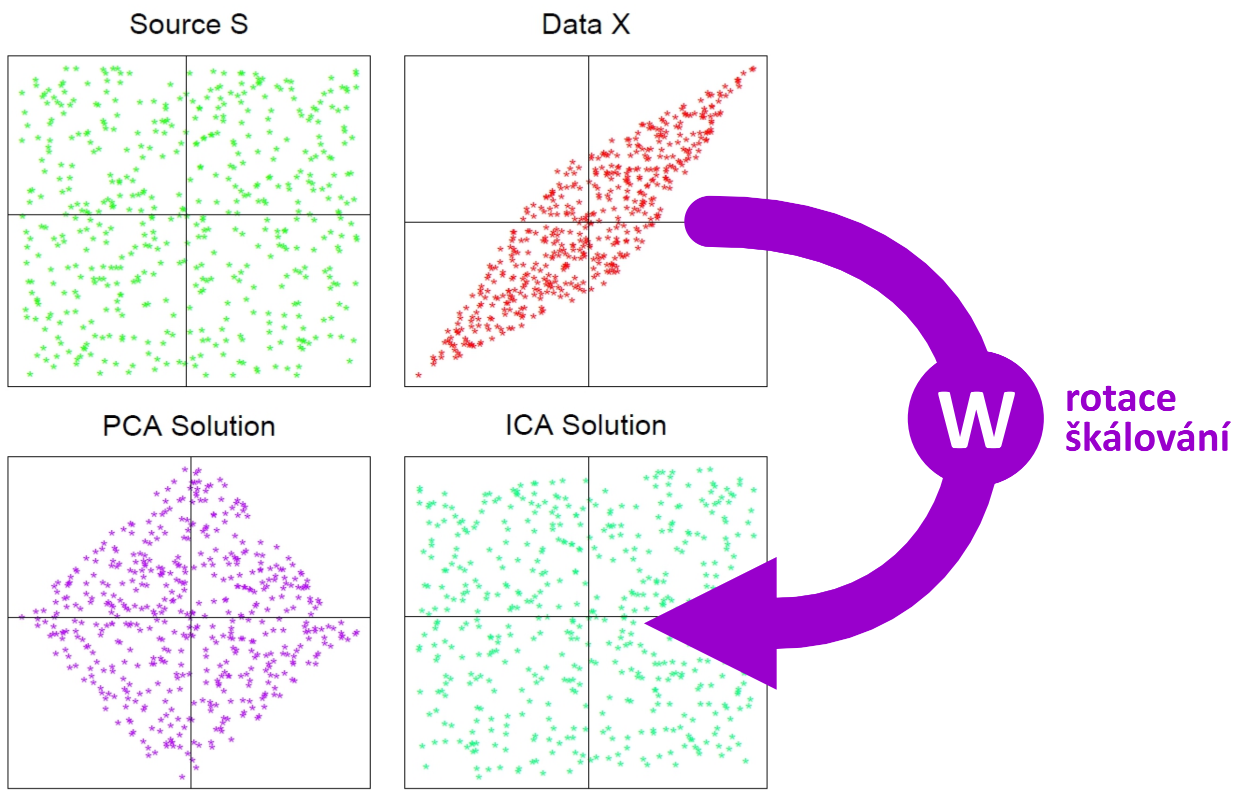


PCA maximalizuje **rozptyl dat**, tj. zajišťuje mezi kom-ponentami v cílovém pro-storu maximální vzdálenost

ICA maximalizuje **nezávis-lost komponent**, tj. hledá takovou transformaci, aby se od sebe komponenty co nejvíce lišily

(obrázek z knihy Hastie, Tibshirani, Friedman: The Elements of Statistical Learning)

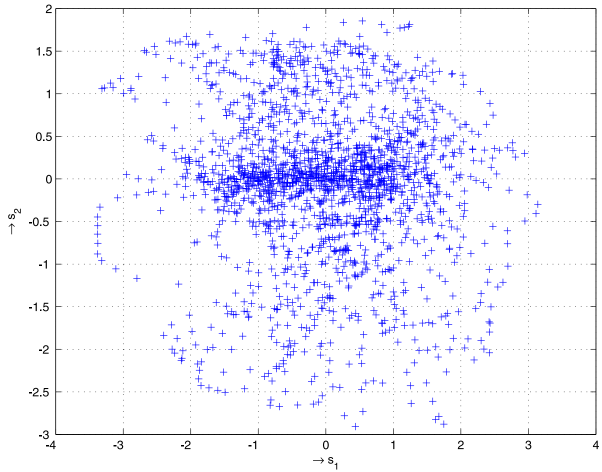
Independent Component Analysis (ICA)
ICA není PCA



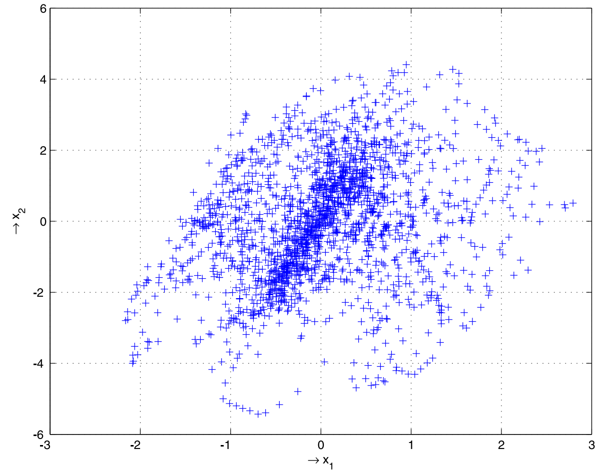
(obrázek z knihy Hastie, Tibshirani, Friedman: The Elements of Statistical Learning)

Independent Component Analysis (ICA)
Odvození algoritmu

Pův. signály $\mathbf{s}(t) = [s_1(t), s_2(t)]$



Směs $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$



Matematický model ICA:

pozorujeme

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$$

chceme znát

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$$

vypočítá ICA

JAK?

dostaneme odhad

Independent Component Analysis (ICA)
Nejednoznačnosti při ICA

- **nelze určit rozptyl (energii) nezávislých komponent** – z mat. modelu ICA je zřejmé, že neznáme-li ani **s**, ani **A**, pak libovolný skalár násobící některý ze zdrojů s_i může být kompenzován vydělením sloupce $\mathbf{a}_i$ matice **A** tímtéž skalárem → tj. abs. hodnota energie signálu je volitelná...

- **nelze určit pořadí nezávislých komponent** – jelikož neznáme ani **s**, ani **A**, můžeme pořadí sčítanců při směšování signálů maticí **A** chápat jako libovolné → tj. nelze říci, která z nezávislých komponent je první, druhá, atd.

Např. při odstraňování šumu apriori nevíme, který z výsledných odhadů signálů je šum (a jaká je jeho celková energie) a který je užitečný signál...

Independent Component Analysis (ICA)
Co je nezávislost?

Má-li ICA vypočítat odhad pův. signálů takový, aby jejich **vzájemná nezávislost byla co nejvyšší**, je třeba definovat nezávislost. To lze např.:

- pomocí **hustoty pravděpodobnosti**:
  $p(y_1, y_2)$ je sdružená hustota pravděpodobnosti (*Probability Density Function, PDF*) náhodných proměnných y_1 a y_2, $p_1(y_1)$ je pak marginální hustota p-sti daná takto:
  $$p_1(y_1) = \int p(y_1, y_2)\, dy_2$$
  a analogicky pro y_2. Potom y_1 a y_2 jsou nezávislé tehdy a jen tehdy, je-li sdružená hustota p-sti rozložitelná takto:
  $$p(y_1, y_2) = p_1(y_1)\, p_2(y_2)$$
  tj. **známá hodnota** y_1 **naprosto nic neříká o hodnotě** y_2.

Independent Component Analysis (ICA)
Co je nezávislost?

Tento výraz lze zobecnit pro n náhodných proměnných, pak sdružená hustota musí být součinem n marginálních hustot.

Z uvedené definice lze odvodit jednu z **nejdůležitějších vlastností nezávislých náhodných proměnných**:
Mějme dvě funkce h_1 a h_2, potom platí

$$E\{h_1(y_1)h_2(y_2)\} = E\{h_1(y_1)\}E\{h_2(y_2)\}$$

očekávaná/střední hodnota náh. prom.

Náznak důkazu:

$$\begin{aligned}
E\{h_1(y_1)h_2(y_2)\} &= \int\int h_1(y_1)h_2(y_2)p(y_1,y_2)dy_1dy_2 \\
&= \int\int h_1(y_1)p_1(y_1)h_2(y_2)p_2(y_2)dy_1dy_2 = \\
&= \int h_1(y_1)p_1(y_1)dy_1 \int h_2(y_2)p_2(y_2)dy_2 = \\
&= E\{h_1(y_1)\}E\{h_2(y_2)\}
\end{aligned}$$

Nezávislost náhodných proměnných
Jsou nekorelované proměnné nezávislé?

Nekorelovanost je slabší forma nezávislosti: Dvě náhodné proměnné jsou nekorelované tehdy, je-li jejich **kovariance** nulová:

$$E\{y_1 y_2\} - E\{y_1\}E\{y_2\} = 0$$

Jsou-li proměnné nezávislé, pak jsou nekorelované (to vyplývá z výrazu na předchozím snímku, dosadíme-li za $h_1(y_1) = y_1$ a $h_2(y_2) = y_2$). **Obrácená implikace však neplatí**!

Důkaz: Mějme diskrétní proměnné y_1, y_2 takové, že pár (y_1, y_2) nabývá s p-stí 0.25 kteroukoliv z hodnot (0, 1), (0, -1), (1, 0) a (-1, 0). Pak lze snadno spočítat, že y_1 a y_2 jsou nekorelované, ale nejsou nezávislé, protože

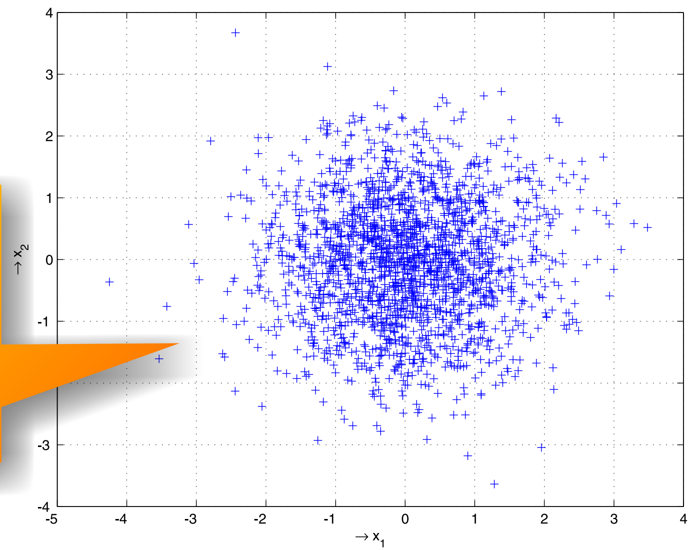
$$E\{y_1^2 y_2^2\} = 0 \neq \frac{1}{4} = E\{y_1^2\}E\{y_2^2\}$$

Omezení ICA
Proč nelze separovat gaussovské proměnné?

Jsou-li směšované signály s_i gaussovské a směšovací matice **A** ortogonální, pak směsi x_i jsou gaussovské, nekorelované, s jednotkovým rozptylem. Jejich sdružená hustota p-sti je

$$p(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right)$$

což vypadá takto:



Distribuje je zcela symetrická, a tudíž neobsahuje žádnou informaci o směru vektorů/sloupců matice **A**. Proto nelze **A** odhadnout!

Princip ICA
Negaussovské je nezávislé...

Centrální limitní věta (CLV): Rozdělení p-sti součtu nezávislých náhodných proměnných (za určitých podmínek) konverguje k normálnímu, tj. gaussovskému rozdělení.

Odtud plyne, že součet dvou nezávislých náhodných proměnných má rozdělení blíže normálnímu, než obě sčítané náhodné proměnné.

Hledáme vektor $\mathbf{w}$ – řádek matice $\mathbf{W}$ inverzní k $\mathbf{A}$:
označíme $\mathbf{z} = \mathbf{A}^T\mathbf{w}$, pak $\mathbf{y} = \mathbf{w}^T\mathbf{x} = \mathbf{w}^T\mathbf{A}\mathbf{s} = \mathbf{z}^T\mathbf{s}$, tj. $\mathbf{y}$ je lineární kombinací zdrojových signálů s_i váhovaných hodnotou z_i.

Jelikož platí CLV, pak lineární kombinace signálů $\mathbf{z}^T\mathbf{s}$ je více gaussovská, než kterýkoliv ze zdrojových signálů samotných.

Princip ICA
Negaussovské je nezávislé...

Nejméně gaussovská je pak směs $\mathbf{z}^T\mathbf{s}$ tehdy, rovná-li se právě jednomu ze zdrojových signálů – tehdy je zřejmě pouze jedna složka z_i vektoru **z** nenulová...

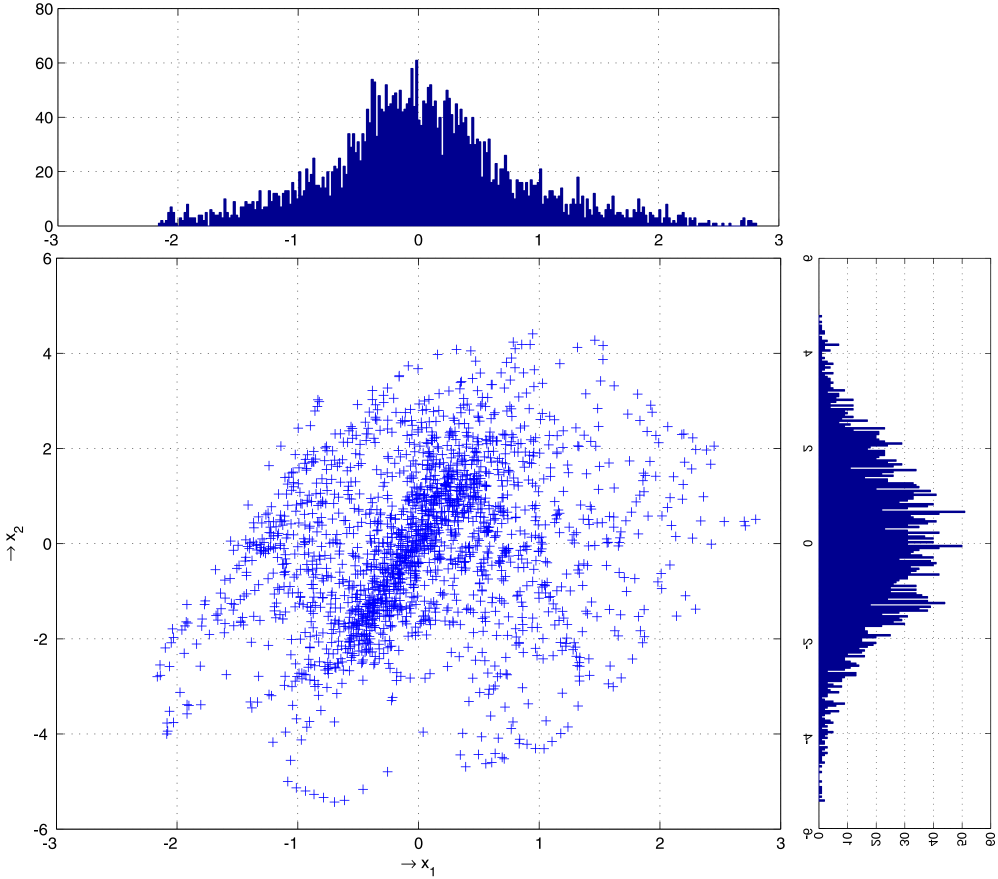
Hledáme tedy vektor **w** takový, aby součin $\mathbf{w}^T\mathbf{x}$ byl co nejméně gaussovský.

Takový vektor **w** nutně odpovídá (v transformovaném souřadnicovém systému) vektoru **z** s jedinou nenulovou složkou.

To znamená, že $\mathbf{w}^T\mathbf{x} = \mathbf{z}^T\mathbf{s}$, což je pak právě jedna z nezávislých komponent směsi... stačí najít

→ potřebujeme míru „negaussovskosti” a alg. optimalizace

Princip ICA
Negaussovské je nezávislé...



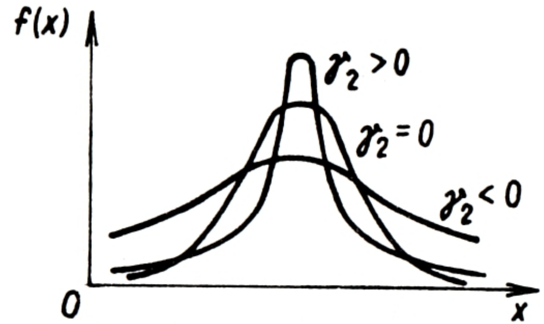
Míry „negaussovskosti”
Koeficient špičatosti čili exces (*Kurtosis*)

Klasická míra „negaussovskosti” je tzv. **koeficient špičatosti** neboli **exces** (*kurtosis*):

$$\mathrm{kurt}(y) = E\{y^4\} - 3(E\{y^2\})^2$$

Protože jsme přepokládali, že **y** má jednotkový rozptyl, pravá strana se redukuje na $E(y^4) - 3$, tj. exces je vlastně normalizovaný centrální moment 4. řádu $E(y^4)$.

Gaussovská náhodná proměnná má exces nulový.



Obr. 430. Druhy excesu γ_2 při symetrickém rozdělení s týmž průměrem (pro $\gamma_2 = 0$ jde o normální rozložení)

(obrázek z knihy H.-J. Bartsch: Matematické vzorce)

Míry „negaussovskosti”
Koeficient špičatosti čili exces (*Kurtosis*)

Exces teoreticky může být použit jako optimalizační kritérium při výpočtu ICA, ale **v praxi** (když exces počítáme z naměřených hodnot či navzorkovaného signálu) **to není rozumné**:

- je bohužel velmi citlivý na marginální hodnoty – jediný např. chybně zaznamenaný vzorek může hodnotu excesu značně pozměnit
- nejedná se o dostatečně robustní míru „negaussovskosti”
- existují jiné míry, lépe přizpůsobené reálnému nasazení

Míry „negaussovskosti”
Negentropie

Entropie náhodné proměnné je množství informace, kterou tato náh. proměnná nese. Čím je „více náhodná”, tj. méně predikovatelná a uspořádaná, tím je vyšší její entropie.

Přesněji (za určitých zjednodušujících předpokladů) je entropie náhodné proměnné délka kódování takové proměnné.

Entropie H náhodné proměnné Y je definovaná takto:

$$H(Y) = -\sum_i P(Y = a_i) \log P(Y = a_i)$$

kde a_i jsou hodnoty, jichž může Y nabývat.

Zásadní poznatek z teorie informace: Gaussovská náhodná proměnná má nejvyšší entropii ze všech náh. proměnných se stejným rozptylem.

Míry „negaussovskosti” Negentropie

➔ entropii lze využít jako míru „negaussovskosti”, **gaussovské rozdělení p-sti je nejnáhodnější ze všech**, zatímco u rozdělení, u nichž hustota p-sti roste kolem nějakých specifických hodnot, je entropie malá...

Jelikož chceme míru „negaussovskosti”, která je nulová pro gaussovskou proměnnou a vždy nezáporná, definujeme tzv. **negentropii** J (vlastně přizpůsobenou diferenciální entropii):

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y})$$

$\mathbf{y}_{\text{gauss}}$ je gaussovská náhodná proměnná se stejnou kovarianční maticí jako **y**.

✔ Negentropie je ze statistického hlediska optimálním odhadem „negaussovskosti” náhodné proměnné.

Míry „negaussovskosti”
Aproximace negentropie

Negentropii je velmi obtížné získat výpočetně – je třeba využít pro výpočet ICA její jednodušší aproximace:

$$J(y) \approx \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} \operatorname{kurt}(y)^2$$

$$J(y) \approx \sum_{i=1}^{p} k_i [E\{G_i(y)\} - E\{G_i(\nu)\}]^2$$

kde k_i jsou kladné konstanty, ν je gaussovská náh. proměnná se stř. hodnotou 0 a jednotkovým rozptylem; proměnná y by měla mít $\mu = 0$ a $\sigma^2 = 1$ (předpoklad). Funkce Gi jsou nějaké nekvadratické funkce, např. se osvědčila volba

$$G_1(u) = \frac{1}{a_1} \log\cosh a_1 u, \quad G_2(u) = -\exp(-u^2/2)$$

kde $1 \leq a_1 \leq 2$. (Hyvärinen, Oja, 2000)

Další možnosti ICA
Minimalizace vzájemné informace

Vzájemná informace (*Mutual Information*) – přirozená míra vzájemné závislosti dvou (nebo více) náhodných (skalárních) proměnných, nejběžnějším vyjádřením míry je **bit**.

Definice vzájemné informace:
Mějme diskrétní náhodné proměnné X a Y. Pak

$$I(X,Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

kde p(x, y) je sdružená hustota p-sti X a Y, a p(x) a p(y) jsou marginální hustoty X a Y.

I(X, Y) = 0 tehdy a jen tehdy, jsou-li X a Y nezávislé.

D: Jsou-li nezávislé, pak p(x, y) = p(x) p(y), atd.

Další možnosti ICA
Vztah vzájemné informace a (neg)entropie

$$\begin{aligned} I(X,Y) &= H(X) - H(X|Y) \quad \leftarrow \text{podmíněná entropie} \\ \text{marginální entropie} \quad &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y) \quad \leftarrow \text{sdružená entropie} \\ &= H(X,Y) - H(X|Y) - H(Y|X) \end{aligned}$$

(Papoulis, 1991) a (Cover a Thomas, 1991) ukázali, že pro vratnou lineární transformaci **y** = **Wx** platí:

$$I(y_1, y_2, \dots, y_n) = \sum_i H(y_i) - H(\mathbf{x}) - \log|\det \mathbf{W}|$$

z čehož lze (poměrně složitě) odvodit základní vztah mezi vzájemnou informací a negentropií:

$$I(y_1, y_2, \dots, y_n) = C - \sum_i J(y_i)$$

kde C je konstanta nezávislá na **W**.

Další možnosti ICA
Některé další odhady vzájemné závislosti

- **Maximum Likelihood Estimation (MLE)** –
  v podstatě ekvivalent minimalizace vzájemné informace...

- **Princip Infomax** (*The Infomax Principle*) –
  pův. použit v neuronových sítích za účelem maximalizace entropie výstupní vrstvy, pro potřeby ICA rozpracován v roce 1995 Bellem a Sejnowskim...

- zřejmě je možné vyjádřit nezávislost náhodných proměnných/signálů ještě řadou dalších postupů, mnohé jsou adaptované na konkrétní charakter signálů, atp.

Předzpracování dat pro ICA
Ustředění (*Centering*)

Signál nesmí mít tzv. ***offset***, tzn. jeho **střední hodnota musí být nulová**.

Toho dosáhneme tím, že od pozorované směsi signálů **x**(t) odečteme střední hodnotu E(**x**(t)):

$$x_i(t) = x_i(t) - E(x_i(t))$$

MATLAB

```
X(:, 1) = X(:, 1) - mean(X(:, 1));
X(:, 2) = X(:, 1) - mean(X(:, 2));
```

Smyslem ustředění je zjednodušit algoritmus ICA – neznamená to, že by nebylo možné offset výpočtem ICA odhadnout...

Pokud je v datech offset významný (což např. u audiosignálů **není**), pak je třeba k výslednému odhadu „rozmíchaných” signálů jejich offsety získané při předzpracování opět přičíst...

Předzpracování dat pro ICA
Bělení (*Whitening*)

Lineární transformace pozorované směsi signálů **x**(t) taková, že složky transformované směsi jsou nekorelované a jejich rozptyl je jednotkový, tzn. kovarianční matice $\bar{\mathbf{x}}(t)$ se rovná jednotkové, tj. $E\{\bar{\mathbf{x}}(t)\,\bar{\mathbf{x}}(t)^T\} = \mathbf{I}$:

$$x_i(t) = x_i(t) \,/\, \sqrt{D(x_i(t))}$$

MATLAB

```
X(:, 1) = X(:, 1) / std(X(:, 1));
X(:, 2) = X(:, 1) / std(X(:, 2));
```

Provádí se **po** ustředění signálu, **před** výpočtem ICA. Nejsnazší postup bělení je vydělení složek jejich směrodatnými odchylkami...

Jednoduchý algoritmus ICA
Gradientní metoda odhadu pomocí excesu

Algoritmus:

(1) inicializace váhového vektoru **w** (náhodně)

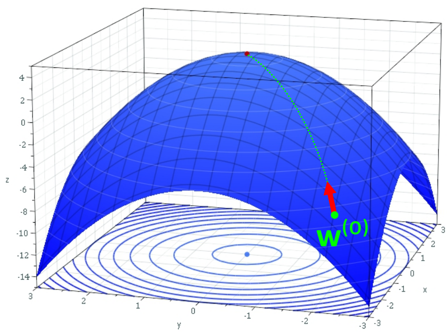
(2) *sphering*, tj. SVD matice směsi $\mathbf{X} = \mathbf{USV}^T$

(3) **for** i ← 1 .. maximální zvolený počet iterací

- (a) odhad aproximace „rozmíchaného” signálu $\mathbf{y} \leftarrow \mathbf{Uw}^T$
- (b) výpočet gradientu **g** z odhadu excesu **y**
- (c) oprava váhového vektoru **w**: $\mathbf{w} \leftarrow \mathbf{w} + \delta\mathbf{g}$
- (d) normalizace váhového vektoru **w**: $\mathbf{w} \leftarrow \mathbf{w} \,/\, |\mathbf{w}|$

Gradientní metoda –
iterační algoritmus, další kroky
se konají ve směru gradientu
(detaily viz MA či NM)

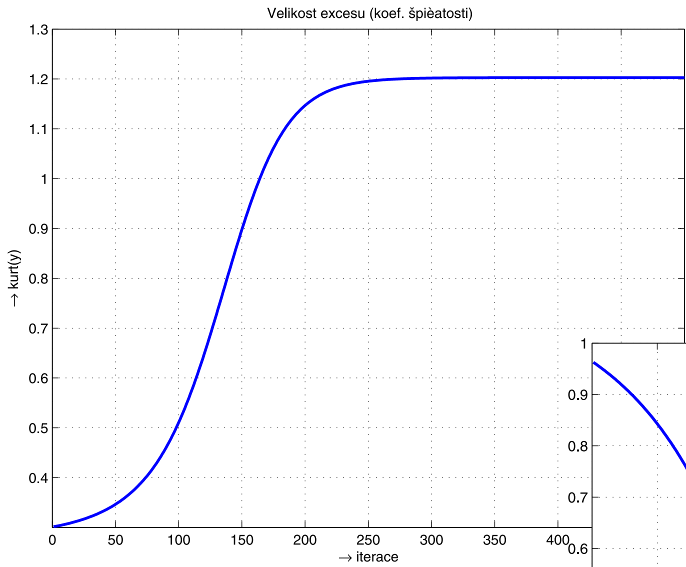
(obrázek z Wikimedia Commons)



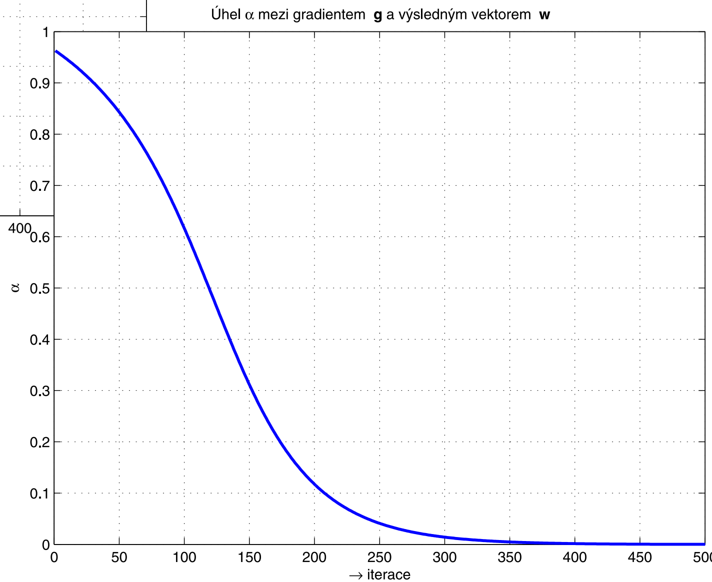
Jednoduchý algoritmus ICA
Zápis algoritmu v MATLABu

```
% sphering dat
[U D V] = svd(X, 0);     % "ekonomická" dekompozice (viz ML>> help svd)
Z = U;
Z = Z ./ repmat(std(Z, 1), n, 1);

% ______________________________________________________________________
%
% ICA
% ______________________________________________________________________
%
w = randn(1, m);         % "rozmíchávací" vektor inicializovaný náhodně
w = w / norm(w);         % normalizovaný
y = Z * w';              % počáteční odhad aproximace zdrojového signálu

max_iters = 500;         % maximální počet iterací
delta = 2e-2;            % velikost kroku gradientní metody

for i = 1:max_iters
    y = Z * w';
    k = mean(y .^ 4) - 3;    % odhad koeficientu špičatosti (kurtosis)

    % výpočet gradientu
    ycub = y .^ 3;
    ycubr = repmat(ycub, 1, 2);
    g = mean(ycubr .* Z);

    % oprava "rozmíchávací" matice W tak, aby se zvýšil koef. špičatosti
    w = w + delta * g;
    w = w / norm(w);
end
```

Jednoduchý algoritmus ICA
Volba počtu iterací



Počet iterací volíme podle vývoje velikosti excesu a úhlu mezi gradientem a výsledným váhovým vek-torem...



Velikost excesu by měla konvergovat k nějakému lokálnímu maximu...

Sofistikované algoritmy ICA
Co lze využít při reálném nasazení?

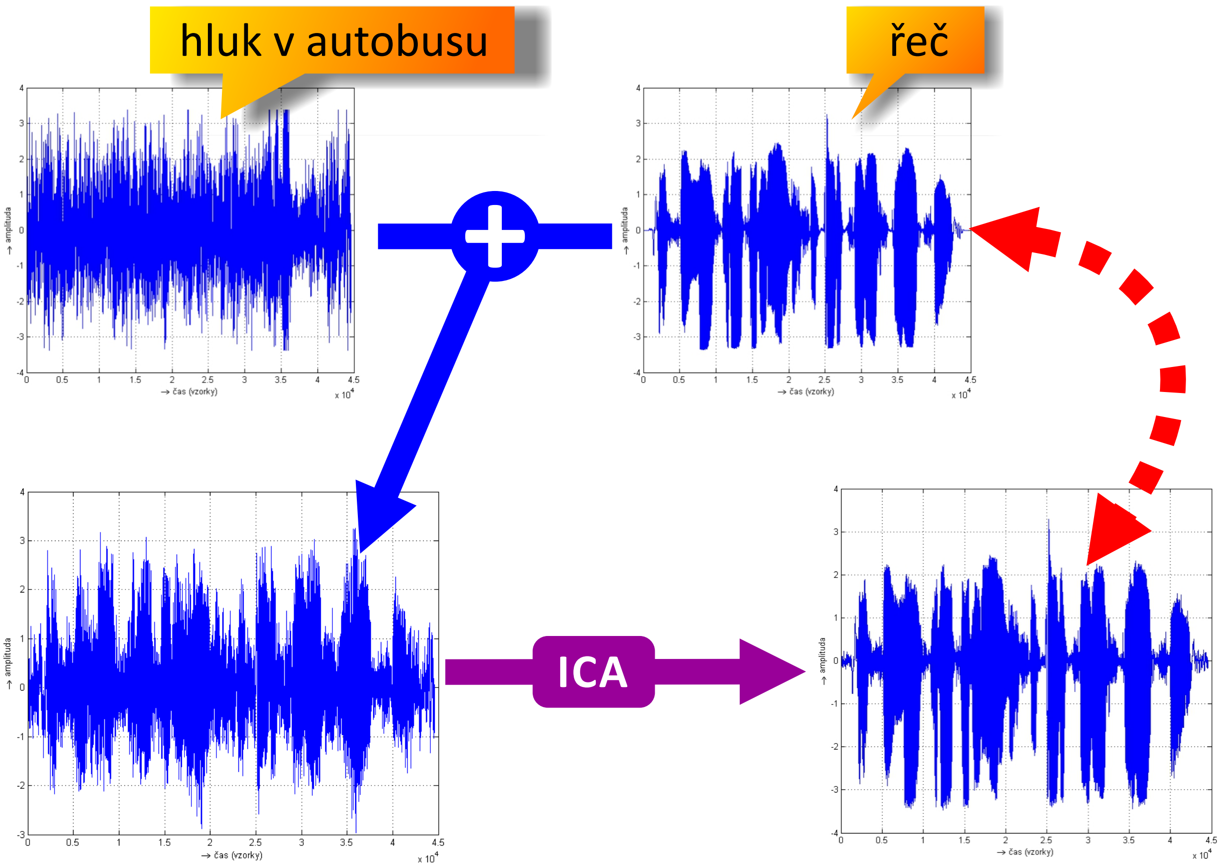
- **FastICA** –
  H. Gävert, J. Hurri, J. Särelä, A. Hyvärinen (Helsinki UT),
  k dispozici pod GPL pro MATLAB, jako součást knihovny **IT++** také pro C++, pro Javu knihovna **FastICA for Java**

  http://research.ics.aalto.fi/ica/fastica/
  http://www.fastica.org/

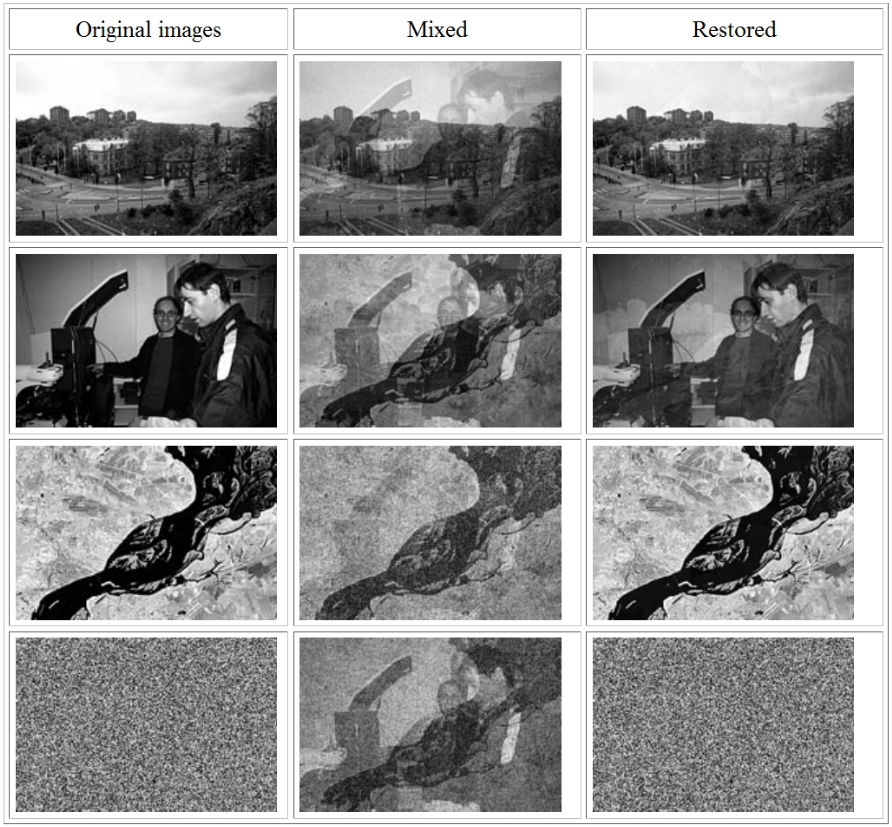
- **MILCA** (Mutual Information Least-dependent Comp. A.) –
  H. Stögbauer, A. Kraskov, S. Astachov a P. Grassberger (Caltech),
  k dispozici pod GPL pro MATLAB a C

  http://www.klab.caltech.edu/~kraskov/MILCA/

Ukázky použití ICA
Odstranění hluku na pozadí audiosignálu

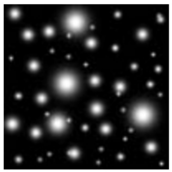
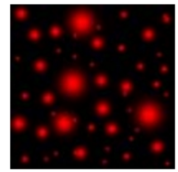
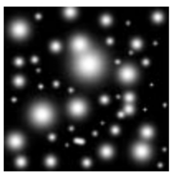
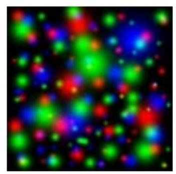
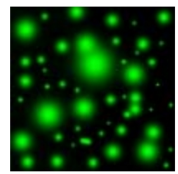
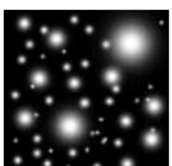
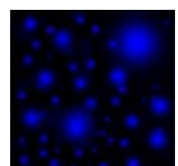


Ukázky použití ICA
Odstranění nežádoucího obsahu ze snímku



(obrázek z http://www.klab.caltech.edu/~kraskov/MILCA/demo/demo.html)

Ukázky použití ICA
Slepá separace spektrálních komponent

Grayscale-coded abundances	Mixture image (RGB)	Decomposed
red 		
green 		
blue 		

(obrázek z http://www.klab.caltech.edu/~kraskov/MILCA/demo/demo.html)