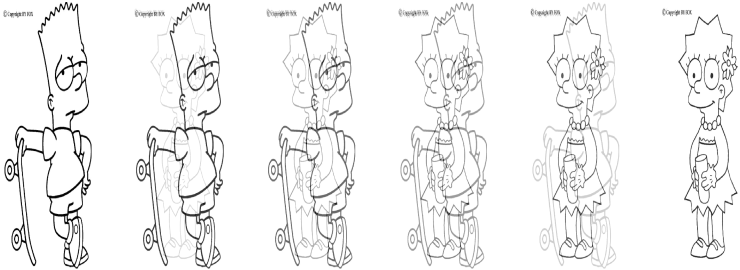
Blind Source Separation

Václav Šmídl

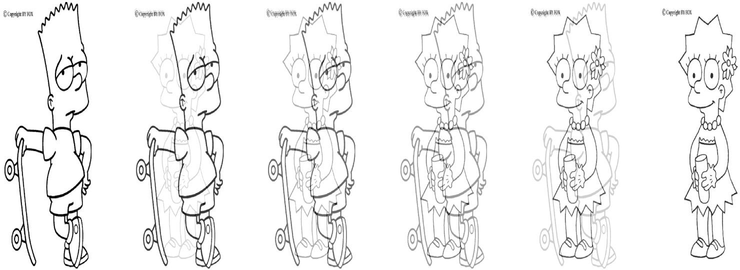
April 4, 2018

Blind source separation – image sequence



- 2 sources

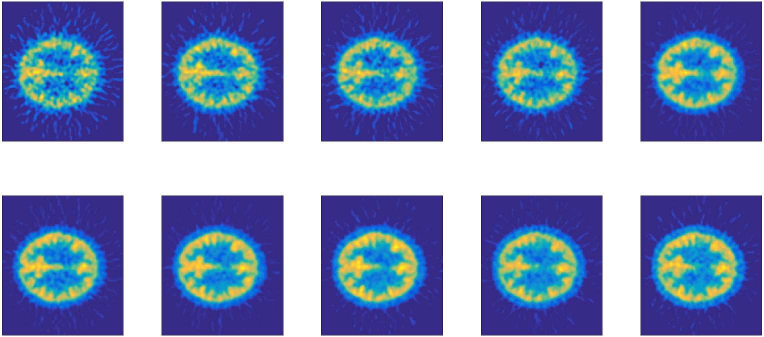
Blind source separation – image sequence



- 2 sources
- weights:

1	0.8	0.6	0.4	0.2	0
0	0.2	0.4	0.6	0.8	1

Medical imaging



- number of sources?
- source images
- time activity of the source

Mathematical description

Linear model of $p \times n$ matrix

$$d_i = Ax_i + e_i, i = 1..n \qquad\qquad D = AX^T + E,$$

where

A is the $p \times r$ matrix of source images, $r < \min(n, p)$

X is the $n \times r$ matrix of time activities,

E is the $p \times n$ noise matrix (Gaussian)

Both images and time activities are unknown!

Issue: ambiguity

$$AX^T = ATT^{-1}X^T$$
$$(AT)(T^{-1}X^T) = \overline{AX}^T$$

Principal component analysis

Consider n, p-dimensional vectors $d_i, i = 1, \dots n$, and their covariance matrix

$$S = \frac{1}{n} \sum (d_i - \overline{d})(d_i - \overline{d})^T.$$

Then r-dimensional vectors x_i,

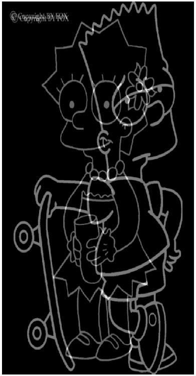
$$x_i = U(:, 1:r)^T d_i, \qquad\qquad S = U \Lambda U^T,$$

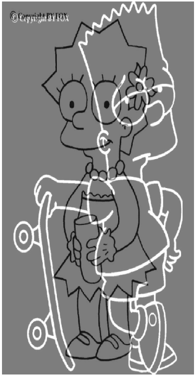
has maximum variance from all possible projections to r dimensions. U are eigenvectors of S sorted with decreasing eigenvalue.

Matrix SVD approach:

$$\begin{aligned} D &\overset{\text{svd}}{=} USV, \\ &= AX^T \end{aligned}$$

Application to sequences of images





- Popular few years ago for speed of computation,
- Requires to find the rotation matrix T,
- Can we do better?
  - Independent Component Analysis (higher order moments)?
  - Structural priors

Probabilistic PCA [Tipping, Bishop, 1999]

Consider model:

$$\begin{aligned} p(d_i|A, x_i, \sigma) &= \mathcal{N}(Ax_i, \sigma I_p), \\ p(x_i) &= \mathcal{N}(0, I_r), \end{aligned}$$

Marginalization over x_i yields

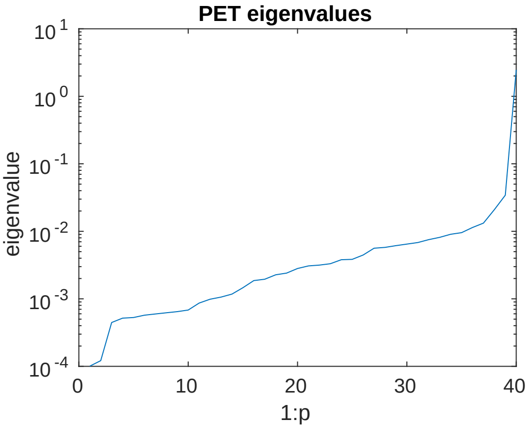
$$\begin{aligned} p(d_i|W, \sigma) &= \mathcal{N}(0, C), \\ C &= W^T W + \sigma I_p, \end{aligned}$$

with maximum likelihood

$$\begin{aligned} \hat{W} &= U_{1:r}(\Lambda_{1:r} - \sigma I)^{\frac{1}{2}} T, \\ \hat{\sigma} &= \frac{1}{d-r} \sum_{i=r+1}^{N} \lambda_i, \end{aligned}$$

where $S = U \Lambda U^T$ is eigen-decomposition of S.

Image Sequence (PET)



- no plateau
- Bayesian solution?

Toy matrix decomposition

Consider 1×1 matrix d, decomposed

$$p(d|a,x) = \mathcal{N}(ax, r_e),$$

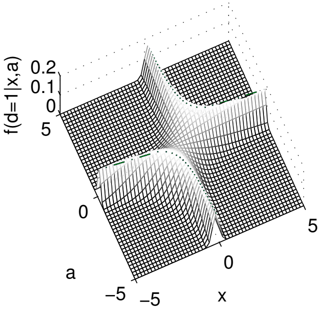
Find a, x.

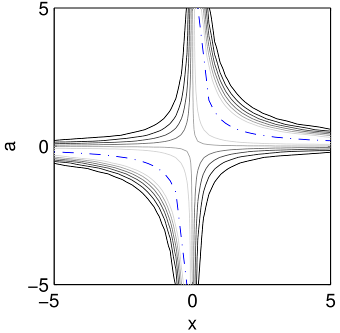
Toy matrix decomposition

Consider 1×1 matrix d, decomposed

$$p(d|a,x) = \mathcal{N}(ax, r_e),$$

Find a, x.

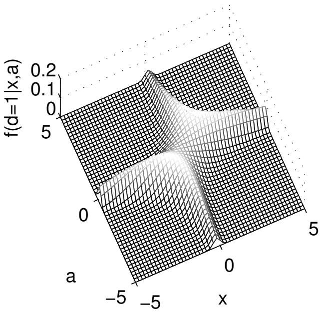


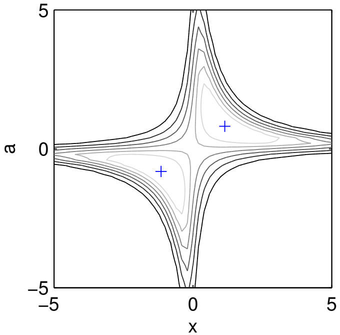


Toy matrix decomposition

Consider 1×1 matrix d, decomposed

$$
\begin{aligned}
p(d|a,x) &= \mathcal{N}(ax, r_e),\\
p(x) &= \mathcal{N}(0, r_x),\\
p(a) &= \mathcal{N}(0, r_a)
\end{aligned}
$$





Toy maximum likelihood

Joint distribution:

$$\log p(d, a, x) \propto -\frac{1}{2r_e}(d - ax)^2 - \frac{1}{2r_a}a^2 - \frac{1}{2r_x}x^2$$

Find

$$\hat{x}, \hat{a} = \arg\max_{a,x}(\log p(d, a, x))$$

For $d < \frac{r_e}{\sqrt{r_a r_x}}$,

$$\hat{x} = 0, \hat{a} = 0,$$

For $d \geq \frac{r_e}{\sqrt{r_a r_x}}$,

$$\hat{x} = \pm\left(d\sqrt{\frac{r_x}{r_a}} - \frac{r_e}{r_a}\right)^{\frac{1}{2}}, \qquad \hat{a} = \pm\left(d\sqrt{\frac{r_a}{r_x}} - \frac{r_e}{r_x}\right)^{\frac{1}{2}}.$$

Note that the product of the maxima is

$$\hat{a}\hat{x} = d - \frac{r_e}{\sqrt{r_a r_x}}.$$

Marginal likelihood (PPCA)

Joint distribution:

$$\log p(d, a, x) \propto -\frac{1}{2r_e}(d - ax)^2 - \frac{1}{2r_a}a^2 - \frac{1}{2r_x}x^2$$

Marginal

$$\begin{aligned} p(a|d) &\propto \int p(d, a, x)dx \\ &\propto \exp(-\frac{1}{2}d^2(a^2 r_x + r_e)^{-1})\sqrt{r_e r_x}(a^2 r_x + r_e)^{-\frac{1}{2}}, \end{aligned}$$

with maximum:

$$\hat{a} = \begin{cases} \frac{\sqrt{d^2 - r_e}}{\sqrt{r_x}} & \text{if } d^2 > r_e, \\ 0 & \text{otherwise}, \end{cases}$$

Variational Bayes

Joint distribution:

$$\log p(d, a, x) \propto -\frac{1}{2r_e}(d - ax)^2 - \frac{1}{2r_a}a^2 - \frac{1}{2r_x}x^2$$

Factor $q(a|d)$

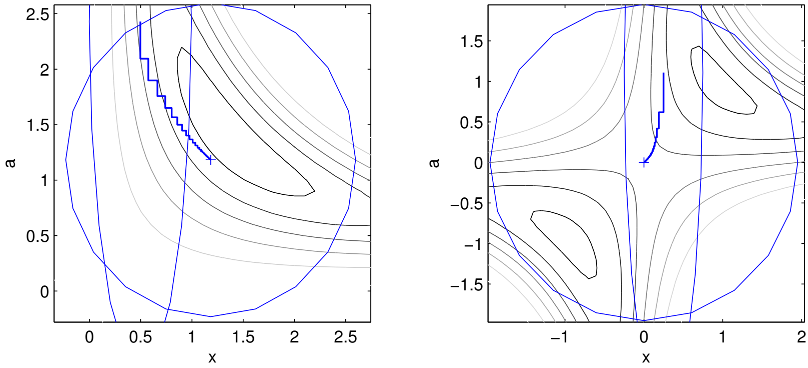
$$\begin{aligned} p(a|d) &\propto \exp\left(\mathrm{E}_x\left[-\frac{1}{2r_e}(d - ax)^2 - \frac{1}{2r_a}a^2\right]\right) \\ &= \mathcal{N}(\hat{a}, \sigma_a), \\ \hat{a} &= \sigma_a d \langle x \rangle r_e^{-1}, \quad \sigma_a^{-1} = \left\langle a^2 \right\rangle r_e^{-1} + r_a^{-1} \end{aligned}$$

Factor $q(x|d)$

$$\begin{aligned} p(a|d) &\propto \exp\left(\mathrm{E}_x\left[-\frac{1}{2r_e}(d - ax)^2 - \frac{1}{2r_a}a^2\right]\right) \\ &= \mathcal{N}(\hat{a}, \sigma_a), \\ \hat{a} &= \sigma_a d \langle x \rangle r_e^{-1}, \quad \sigma_a^{-1} = \left\langle a^2 \right\rangle r_e^{-1} + r_a^{-1} \end{aligned}$$

Convergence of VB

1. compute $\hat{a}, \sigma_a$, and $\langle a \rangle = \hat{a}$, $\langle a^2 \rangle = \hat{a}^2 + \sigma_a$,
2. compute $\hat{x}, \sigma_x$, and $\langle x \rangle = \hat{x}$, $\langle x^2 \rangle = \hat{x}^2 + \sigma_x$,



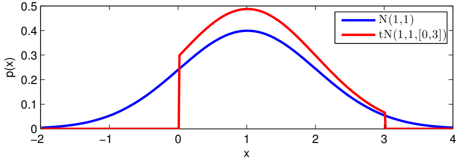
Positive support

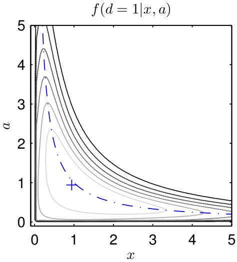
What if we are interested only in the positive solution?

Positive support

What if we are interested only in the positive solution?

$$
\begin{aligned}
p(d|a,x) &= \mathcal{N}(ax, r_e),\\
p(x) &= t\mathcal{N}(0, r_x, \langle 0, \infty\rangle) \propto \mathcal{N}(0, r_x)\chi(x>0),\\
p(a) &= t\mathcal{N}(0, r_a, \langle 0, \infty\rangle) \propto \mathcal{N}(0, r_a)\chi(a>0),
\end{aligned}
$$





Variational Bayes with positive support

Joint distribution:

$$\log p(d,a,x) \propto -\frac{1}{2r_e}(d-ax)^2 - \frac{1}{2r_a}a^2 - \frac{1}{2r_x}x^2 + \log\chi_x + \log\chi_a$$

Factor $q(a|d)$

$$\begin{aligned} p(a|d) &\propto \exp\left(\mathsf{E}_x\left[-\frac{1}{2r_e}(d-ax)^2 - \frac{1}{2r_a}a^2 + \log\chi_x\right]\right) \\ &= t\mathcal{N}(\hat{a},\sigma_a,\langle 0,\infty\rangle), \\ \hat{a} &= \sigma_a d\left\langle x\right\rangle r_e^{-1}, \quad \sigma_a^{-1} = \left\langle x^2\right\rangle r_e^{-1} + r_a^{-1} \end{aligned}$$

Factor $q(x|d)$

$$\begin{aligned} p(x|d) &\propto \exp\left(\mathsf{E}_x\left[-\frac{1}{2r_e}(d-ax)^2 - \frac{1}{2r_a}a^2 + \log\chi_a\right]\right) \\ &= t\mathcal{N}(\hat{a},\sigma_x,\langle 0,\infty\rangle), \\ \hat{x} &= \sigma_x d\left\langle a\right\rangle r_e^{-1}, \quad \sigma_x^{-1} = \left\langle a^2\right\rangle r_e^{-1} + r_x^{-1} \end{aligned}$$

Convergence of VB with positive support

1. compute $\hat{a}, \sigma_a$, and

$$\langle a \rangle = \hat{a} + \frac{\phi(\alpha) - \phi(\beta)}{Z_a},$$

$$\left\langle a^2 \right\rangle = \sigma_a^2 \left\{ 1 + \frac{\alpha\phi(\alpha) - \beta\phi(\beta)}{Z_a} \right\}$$

   where $\alpha = \frac{-\hat{a}}{\sqrt{\sigma_a}}$, $\beta = \infty$, $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/s)$,
   $Z = \Phi(\beta) - \Phi(\alpha)$, $\Phi(x) = \frac{1}{2}(1 + \text{erf}(x))$.

2. compute $\hat{x}, \sigma_x$, and

$$\langle x \rangle = \hat{x} + \frac{\phi(\alpha) - \phi(\beta)}{Z_x},$$

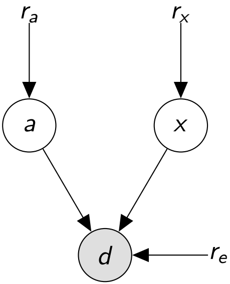
$$\left\langle x^2 \right\rangle = \sigma_x^2 \left\{ 1 + \frac{\alpha\phi(\alpha) - \beta\phi(\beta)}{Z_x} \right\}$$

Zero solution is hard to reach. (ARD)

Decomposition with ARD

Original model:

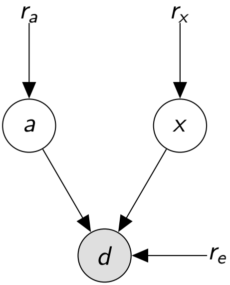
$$
\begin{aligned}
p(d|a,x) &= N(ax, r_e), \\
p(a) &= N(0, r_a), \\
p(x) &= N(0, r_x),
\end{aligned}
$$



Decomposition with ARD

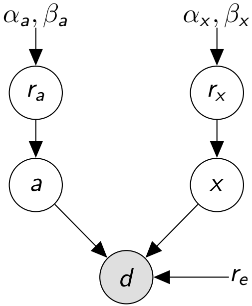
Original model:

$$
\begin{aligned}
p(d|a,x) &= N(ax, r_e),\\
p(a) &= N(0, r_a),\\
p(x) &= N(0, r_x),
\end{aligned}
$$



Unknown hyper-parameters:

$$
\begin{aligned}
p(a|r_a) &= N(0, r_a),\\
p(x|r_e) &= N(0, r_x),\\
p(r_a) &= \Gamma(\alpha_a, \beta_a),\\
p(r_x) &= \Gamma(\alpha_x, \beta_x),
\end{aligned}
$$



Variational PCA [Bishop 200]

Consider model

$$
\begin{aligned}
p(d_i|A, x_i, \sigma) &= \mathcal{N}(Ax_i, \omega I_p), \\
p(x_i) &= \mathcal{N}(0, I_r), \\
p(\underline{a}_j) &= \mathcal{N}(0, I_r),
\end{aligned}
$$

Find $p(A, X, \sigma)$.

Variational PCA [Bishop 200]

Consider model

$$\begin{aligned} p(d_i|A, x_i, \sigma) &= \mathcal{N}(Ax_i, \omega I_p), \\ p(x_i) &= \mathcal{N}(0, I_r), \\ p(\underline{a}_j) &= \mathcal{N}(0, I_r), \end{aligned}$$

Find $p(A, X, \sigma)$.

Joint likelihood:

$$\log p(D, A, X) \propto -\frac{1}{2}\omega \sum_i (d_i - Ax_i)^T (d_i - Ax_i) - \frac{1}{2} \sum_j \underline{a}_j^T \underline{a}_j - \frac{1}{2} \sum_i x_i^T x_i,$$

Variational PCA [Bishop 200]

Consider model

$$p(d_i|A, x_i, \sigma) = \mathcal{N}(Ax_i, \omega I_p),$$
$$p(x_i) = \mathcal{N}(0, I_r),$$
$$p(\underline{a}_j) = \mathcal{N}(0, I_r),$$

Find $p(A, X, \sigma)$.

Joint likelihood:

$$\log p(D, A, X) \propto -\frac{1}{2}\omega \sum_i (d_i - Ax_i)^T (d_i - Ax_i) - \frac{1}{2} \sum_j \underline{a}_j^T \underline{a}_j - \frac{1}{2} \sum_i x_i^T x_i,$$

Factor $q(x_i)$

$$q(x_i) \propto \exp\left(-\frac{1}{2}\omega(d_i - Ax_i)^T (d_i - Ax_i) - \frac{1}{2} \sum_i x_i^T x_i, \right)$$
$$= \mathcal{N}(\hat{x}_i, \Sigma_x),$$
$$\hat{x}_i = \langle \omega \rangle \Sigma_x \langle A \rangle d_i, \quad \Sigma_x^{-1} = \langle \omega \rangle \left\langle A^T A \right\rangle + I$$

Extensions: Multivariate PCA

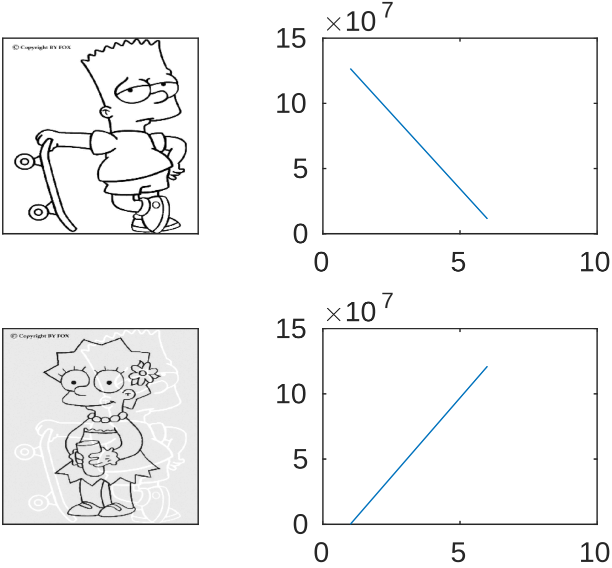
1. Positive support

$$p(x_i) = t\mathcal{N}(0, I_r, \langle 0, \infty\rangle),$$
$$p(\underline{a}_j) = t\mathcal{N}(0, I_r, \langle 0, \infty\rangle),$$

moments

$$[\langle x\rangle,\text{diag}\left\langle x^T x\right\rangle]=\text{momtrun\_low}(\hat{x},\text{sqrt}(\text{diag}(\Sigma_x)))$$

Non-negative Matrix Factorization (NMF)



Extensions: Multivariate PCA

1. Positive support

$$p(x_i) = t\mathcal{N}(0, I_r, \langle 0, \infty\rangle),$$
$$p(\underline{a}_j) = t\mathcal{N}(0, I_r, \langle 0, \infty\rangle),$$

   moments

$$[\langle x\rangle,\text{diag}\left\langle x^T x\right\rangle]\text{=momtrun\_low}(\hat{x}\text{,sqrt(diag}(\Sigma_x)))$$

2. Automatic Relevance Determination (#of factors):

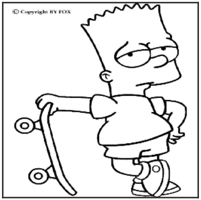
$$p(\underline{a}_j) = \mathcal{N}(0, \text{diag}(\alpha)),$$

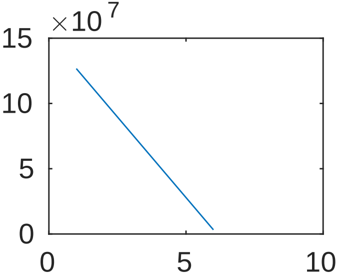
3. Automatic Relevance Determination (#pixels):

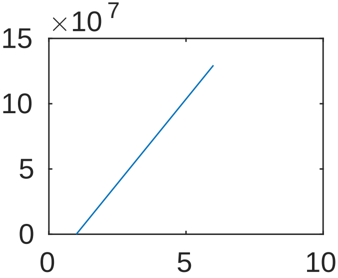
$$p(\underline{a}_j) = \mathcal{N}(0, \text{diag}(\alpha_j)),$$

4. many more

Sparse Non-negative Matrix Factorization (SNMF)







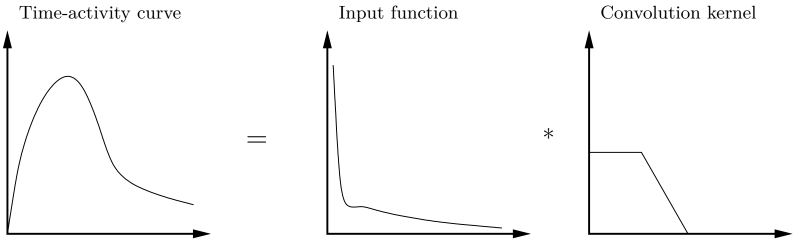
Assignment

Toy Problem		points
Laplace		5
VB		10
Gibbs		10

Matrix factorization		points
Bayesian PCA		20
+ positive support		+10
+ ARD, $x_{i,j} \sim \mathcal{N}(0, \gamma_{i,j})$		+10
ICA (other decomp)		10

BSS with deconvolution

Time-activity curves for brain imaging are results of convolution.



Model of the curve

$$x = b * w = Bw, \qquad B = \begin{pmatrix} b_1 & 0 & 0 & 0 \\ b_2 & b_1 & 0 & 0 \\ \dots & b_2 & b_1 & 0 \\ b_n & \dots & b_2 & b_1 \end{pmatrix}$$

Unknows are b and w. Kernel is sparse.

BSS with deconvolution

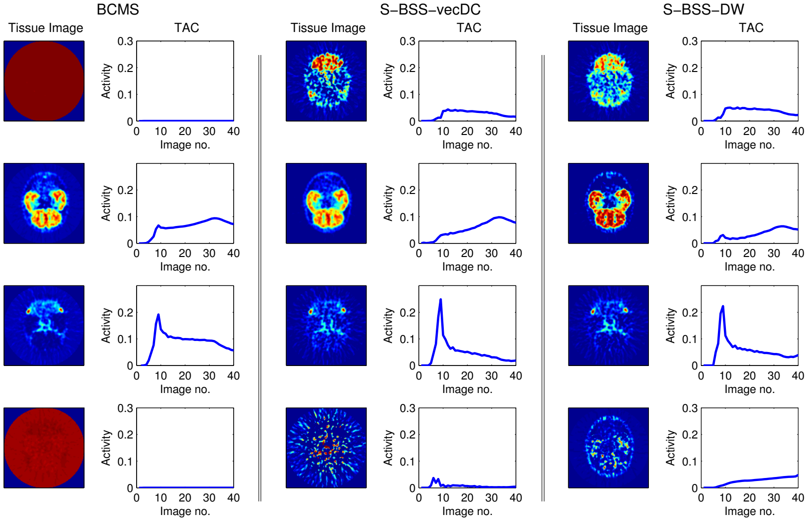


Image deconvolution

(a) out-of-focus blur

Image deconvolution

Mathematical model:

$$d = Ax + e$$

where

d	is the observed (blured) image,	
x	is the original (true) image,	unknown
A	is the convolution matrix,	
e	is the measurement (model) error	

Find: x, A, $var(e)$

Image deconvolution

Mathematical model:

$$d = Ax + e$$

where

d	is the observed (blured) image,	
x	is the original (true) image,	unknown
A	is the convolution matrix,	unknown kernel
e	is the measurement (model) error	

Find: x, A, $var(e)$

Image deconvolution

Mathematical model:

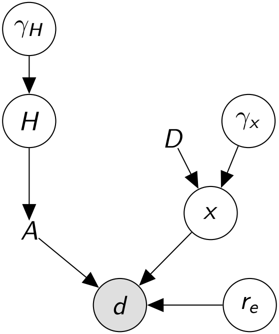
$$d = Ax + e$$

where

d	is the observed (blured) image,	
x	is the original (true) image,	unknown
A	is the convolution matrix,	unknown kernel
e	is the measurement (model) error	unknown variance

Find: x, A, $var(e)$

Model



- number of unknowns $> 3\times$ higher than number of observations
- ARD coefficients γ_H, γ_x, r_e,
- approximations of the covariance matrices by diagonal

Results:

(a) out-of-focus blur

(b) blind deconvolution