

Bayesian non-negative matrix factorization

Mikkel N. Schmidt
University of Cambridge
mns@imm.dtu.dk
www.mikkelschmidt.dk

Ole Winther and Lars Kai
Hansen
Technical University of Denmark

Non-negative matrix factorization

Bi-linear model

$$X = AB + E$$

$\mathbb{R}^{I \times J}$ Data $\mathbb{R}^{I \times N}$ Non-negative factors $\mathbb{R}^{N \times J}$ Non-negative factors $\mathbb{R}^{I \times J}$ Noise

Maximum likelihood

$$\arg \min_{A, B \geq 0} D(X, AB)$$

where D is a divergence derived from the noise model

Bayesian NMF model

- Normal likelihood

$$p(\mathbf{X} | \mathbf{A}, \mathbf{B}, \sigma^2) = \prod_{i,j} \mathcal{N}(X_{i,j} | [\mathbf{AB}]_{i,j}, \sigma^2)$$

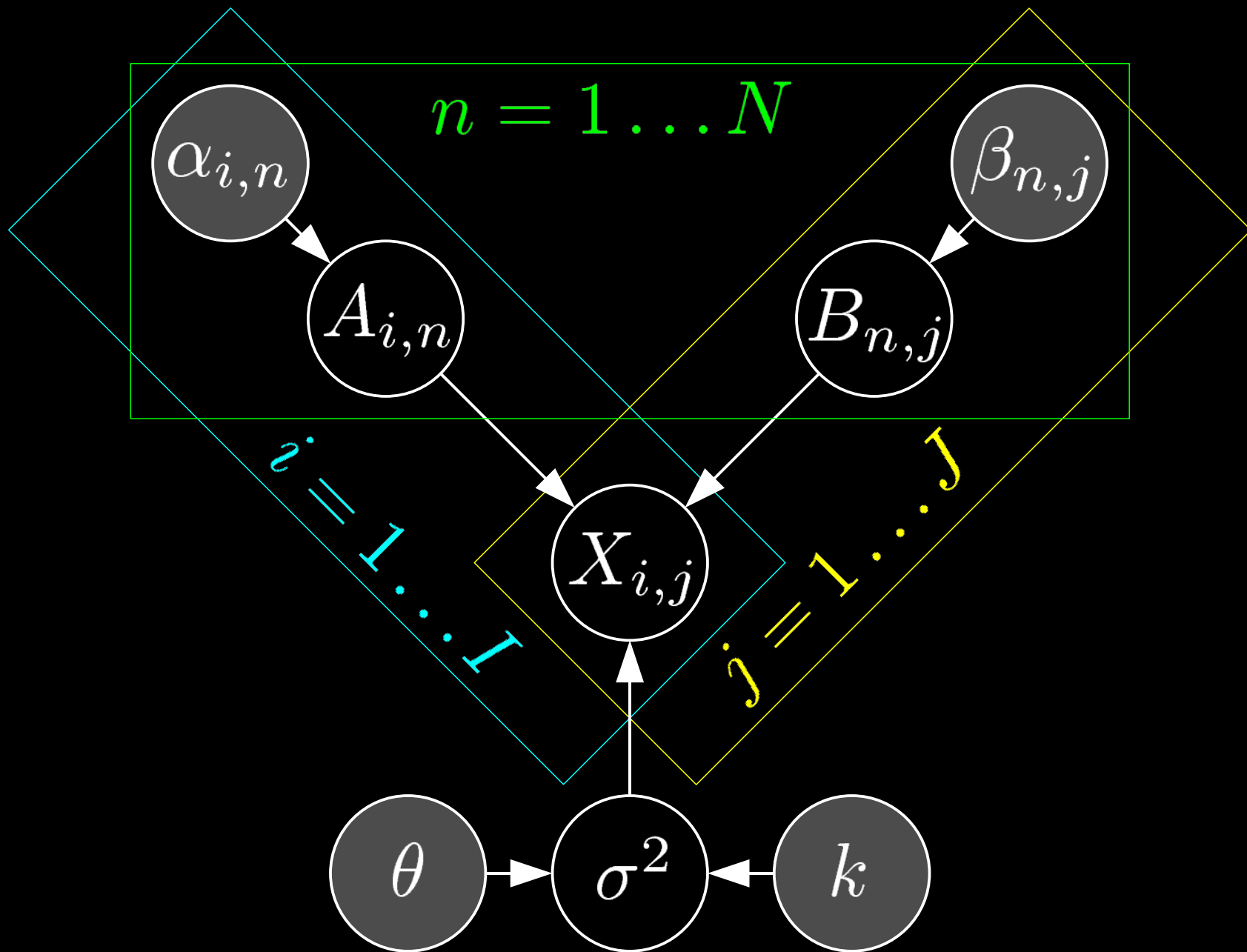
- Exponential priors over A and B

$$p(\mathbf{A} | \boldsymbol{\alpha}) = \prod_{i,n} \mathcal{E}(A_{i,n} | \alpha_{i,n}) \quad p(\mathbf{B} | \boldsymbol{\beta}) = \prod_{n,j} \mathcal{E}(B_{i,n} | \beta_{i,n})$$

- Inverse Gamma prior over noise variance

$$p(\sigma^2 | k, \theta) = \mathcal{G}^{-1}(\sigma^2 | k, \theta)$$

Graphical model



Gibbs sampling algorithm

Initialize A and B

Repeat

For each column in A

 | Sample from rectified Gaussian

For each row in B

 | Sample from rectified Gaussian

Noise variance: Sample from inverse Gamma

Iterated conditional modes algorithm for MAP estimation

Initialize A and B

Repeat

For each column in A

 Set to mode of rectified Gaussian

For each row in B

 Set to mode of rectified Gaussian

Noise variance: Set to mode of inverse Gamma

Estimating the marginal likelihood

Chib's method

$$p(\mathbf{X}) = \frac{p(\mathbf{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathbf{X})}$$

Can easily be evaluated
Difficult to compute

Segment $\boldsymbol{\theta}$ into K blocks amenable to Gibbs sampling

$$p(\boldsymbol{\theta}|\mathbf{X}) = p(\boldsymbol{\theta}_1|\mathbf{X}) p(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1, \mathbf{X}) \cdots p(\boldsymbol{\theta}_K|\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{K-1}, \mathbf{X})$$

$p(\mathbf{X})$

K

can be evaluated by K runs of the Gibbs sampler

Experimental evaluation: Chemical shift imaging of human skull

- Data: 369 point chemical shift spectra measured at 256 positions in human skull [1]
- Previously analyzed
 - in [1] using Bayesian spectral decomposition: Bilinear model with atomic point-mass prior estimated using MCMC. **Computationally expensive**
 - in [2] using a constrained NMF method **Fast, but computes only maximum likelihood**

[1] Ochs, M.F., Stoyanova, R.S., Arias-Mendoza, F., Brown, T.R.: A new method for spectral decomposition using a bilinear bayesian approach. Journal of Magnetic Resonance 137 (1999) 161-176

[2] Sajda, P., Du, S., Brown, T.R., Stoyanova, R., Shungu, D.C., Mao, X., Parra, L.C.: Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain. Medical Imaging, IEEE Transactions on 23(12) (Dec 2004) 1453-1465

Analysis of chemical shift imaging data

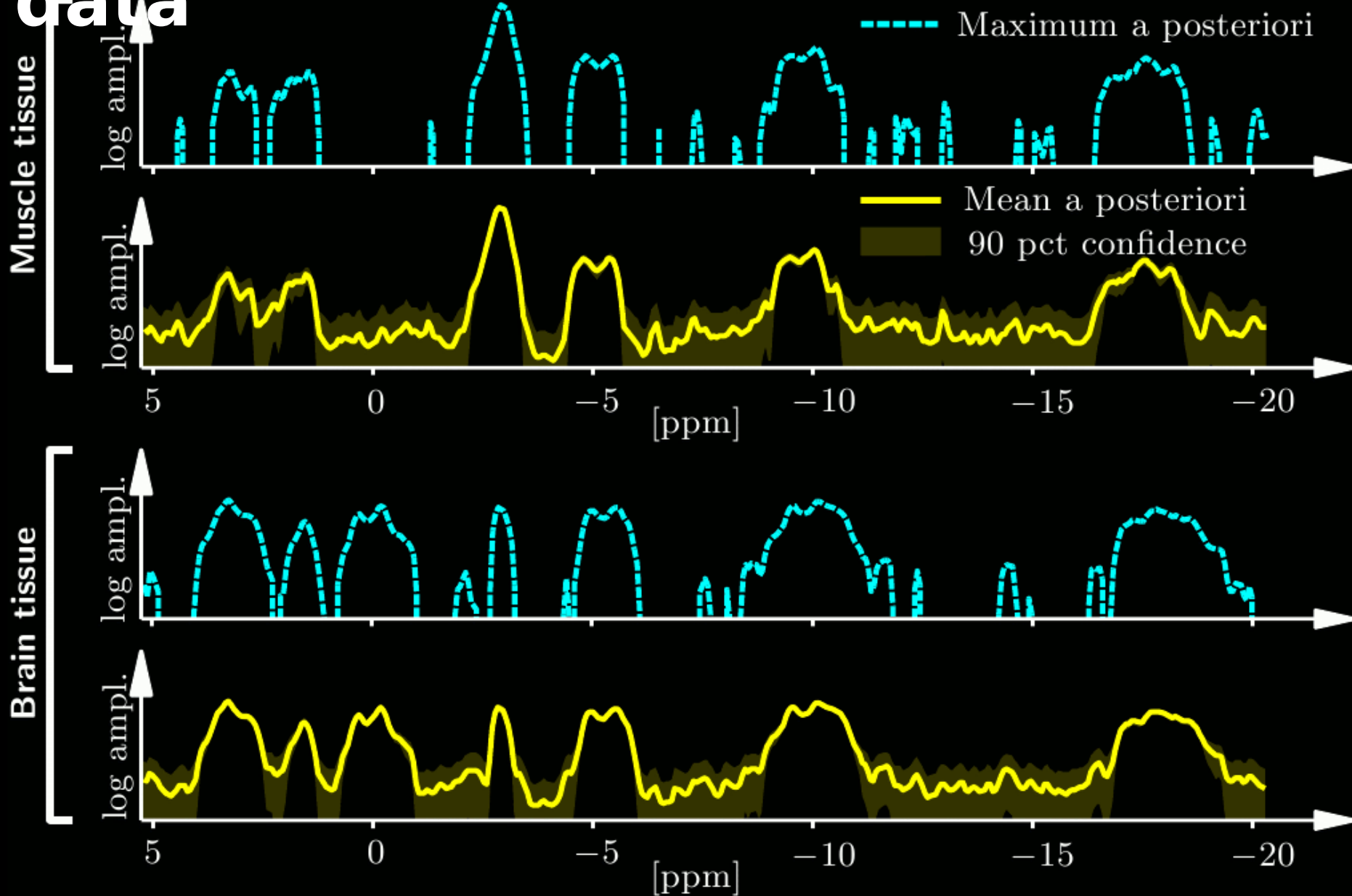
In our experiments, we

- used $N=2$ factors (brain and muscle tissue)
- generated 40,000 samples (which took 80 seconds) and discarded the first 20,000 to allow burn-in
- used a point where factors are expected to differ to resolve permutation ambiguity
- computed mean and 5th and 95th percentile
- computed MAP estimate for comparison

[3] Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature 401(6755) (1999) 788-791

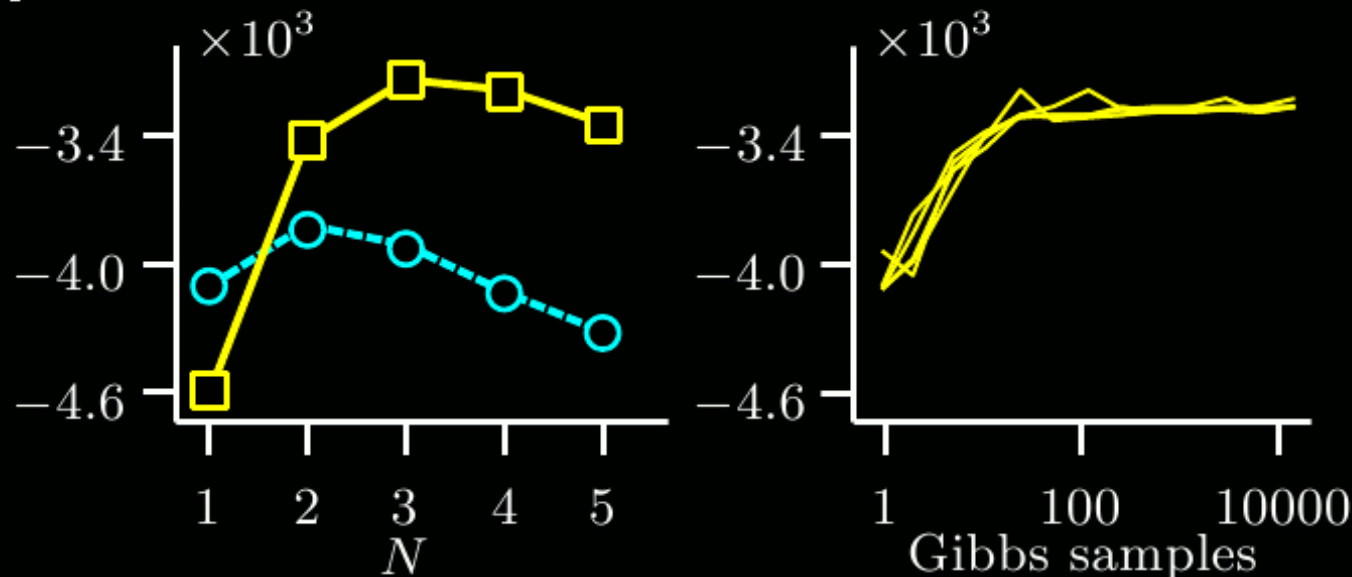
Analysis of chemical shift imaging data

data



Model selection

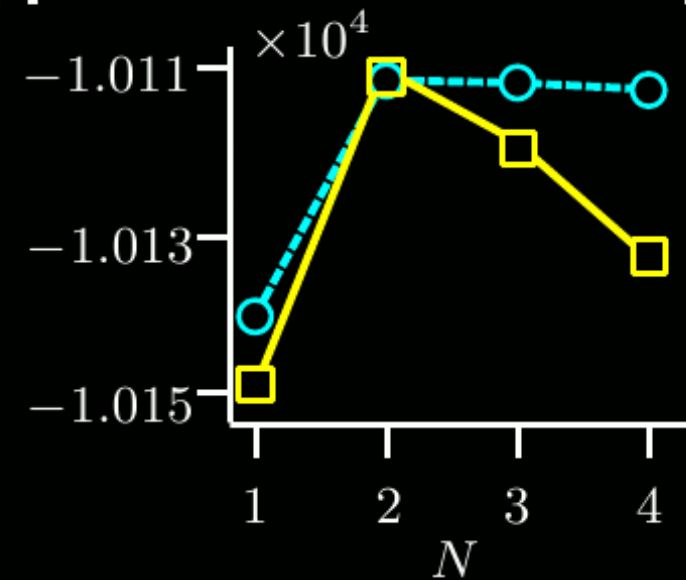
3-component toy example



—○— Bayesian information criterion

—□— Chib's method

CSI data



Summary and conclusions

We have presented a

- Bayesian approach to NMF
- fast Gibbs sampling algorithm
- procedure for computing the marginal likelihood
- iterated conditional modes algorithm for MAP est.

MATLAB software available:
mikkelschmidt.dk/ica2009