Infinite non-negative matrix factorization

Mikkel N. Schmidt and Morten Mørup
Technical University of Denmark

Infinite non-negative matrix factorization

Non-negative matrix factorization (NMF)

- Matrix factorization
  - Algorithms in multivariate analysis / linear algebra
  - Matrix is factored into the product of two matrices
  - Assumptions about the factors lead to different algorithms
    - PCA, FA, ICA, NMF, VQ, etc.

$$\underset{I\times J}{\boldsymbol{V}} = \underset{I\times D}{\boldsymbol{W}}\ \underset{D\times J}{\boldsymbol{H}} + \underset{I\times J}{\boldsymbol{E}}$$

- NMF: Factors constrained to be non-negative

$$\text{s.t. } \boldsymbol{W}, \boldsymbol{H} \geq \boldsymbol{0}$$

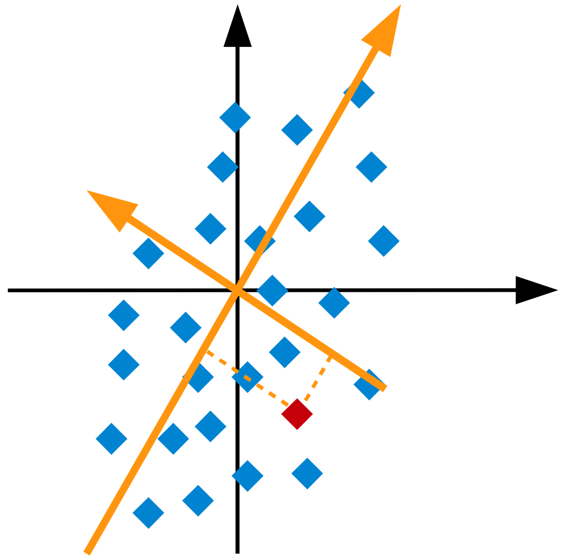
- Problem: Efficient model order selection

Infinite non-negative matrix factorization

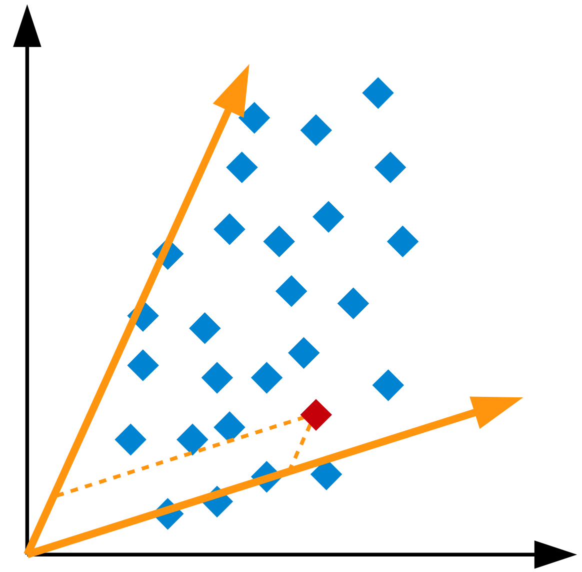
Non-negative matrix factorization (NMF)

$$\underset{I\times J}{\boldsymbol{V}} = \underset{I\times D}{\boldsymbol{W}}\ \underset{D\times J}{\boldsymbol{H}} + \underset{I\times J}{\boldsymbol{E}}$$

Principal component analysis



Non-negative matrix factorization



DTU Informatics
Department of Informatics and Mathematical Modeling

Infinite non-negative matrix factorization

Non-negative matrix factorization (NMF)

$$\underset{I\times J}{\boldsymbol{V}} = \underset{I\times D}{\boldsymbol{W}}\ \underset{D\times J}{\boldsymbol{H}} + \underset{I\times J}{\boldsymbol{E}} \quad \text{s.t. } \boldsymbol{W}, \boldsymbol{H} \geq \boldsymbol{0}$$

Applications of NMF

- Document clustering and topic discovery
  - Word occurence counts
- Learning image basis functions
  - Pixel intensities
- Unmixing spectral data (hyperspectral imaging, chemometrics, etc.)
  - Spectral amplitudes

Infinite non-negative matrix factorization

Bayesian NMF

$$\underset{I\times J}{\boldsymbol{V}} = \underset{I\times D}{\boldsymbol{W}}\ \underset{D\times J}{\boldsymbol{H}} + \underset{I\times J}{\boldsymbol{E}} \qquad \text{s.t. } \boldsymbol{W}, \boldsymbol{H} \geq \mathbf{0}$$

- Principled framework
  - Data and parameters modeled as stochastic variables
  - Models uncertainty about the factors
  - All assumptions are made explicit
- Many extensions are special cases
- Gives predictions with error-bars (credible intervals)
- Model selection is integrated in the framework

Infinite non-negative matrix factorization

Choosing the number of factors

Resampling

- Crossvalidation, Bootstrapping, etc.

Asymptotic theory

- Bayesian information criterion (BIC), etc.

Bayesian model comparison

- Chib's method, Themodynamic integration, etc.

Formulating a super-model

- A comprehensive model that comprises factorizations of any order
- Model selection is integrated in the inference procedure

Infinite non-negative matrix factorization

Related infinite matrix factorization methods

- Based on the Indian buffet process
  (Griffiths and Ghahramani, 2006)
  - Infinite binary matrix factorization
    (Meeds et al. 2007)
  - Infinite sparse coding, Infinite independent component analysis
    (Knowles and Ghahramani, 2007)

$$\boldsymbol{V} = \boldsymbol{A}(\boldsymbol{S} \odot \boldsymbol{Z}) + \boldsymbol{E}$$

Unbounded binary matrix

- Our approach
  - Not Indian buffet process
  - Reversible jump Markov chain Monte Carlo (RJMCMC)
    (Green, 1995)

Infinite non-negative matrix factorization
Infinite NMF

$$\underbrace{p(\boldsymbol{W}, \boldsymbol{H}, D|\boldsymbol{V})}_{\text{Posterior}} \propto \underbrace{p(\boldsymbol{V}|\boldsymbol{W}, \boldsymbol{H}, D)}_{\text{Likelihood}}\underbrace{p(\boldsymbol{W}|D)p(\boldsymbol{H}|D)p(D)}_{\text{Priors}}$$

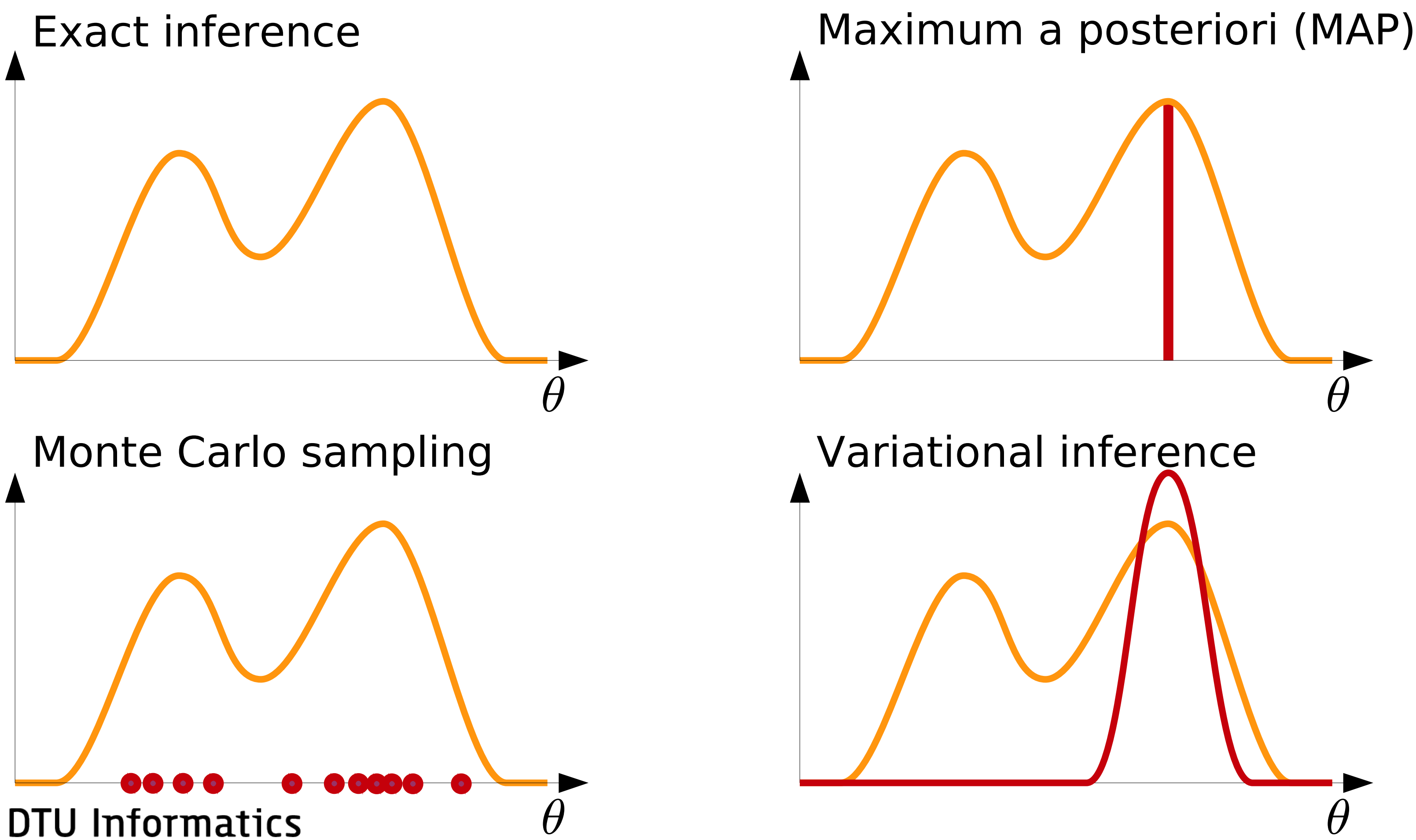
Model of data
Distribution of noise etc.

Model of factors
Non-negativity, sparsity, etc.

Model of number of components
E.g. flat (improper) distribution

Infinite non-negative matrix factorization

Approximate inference procedures



Infinite non-negative matrix factorization

Gibbs / Metropolis-Hastings sampling

- Sampling factorizing matrices etc.
  - Standard Gibbs sampling

$$\boldsymbol{W} \sim p(\boldsymbol{W}|\boldsymbol{H}, D, \boldsymbol{V})$$

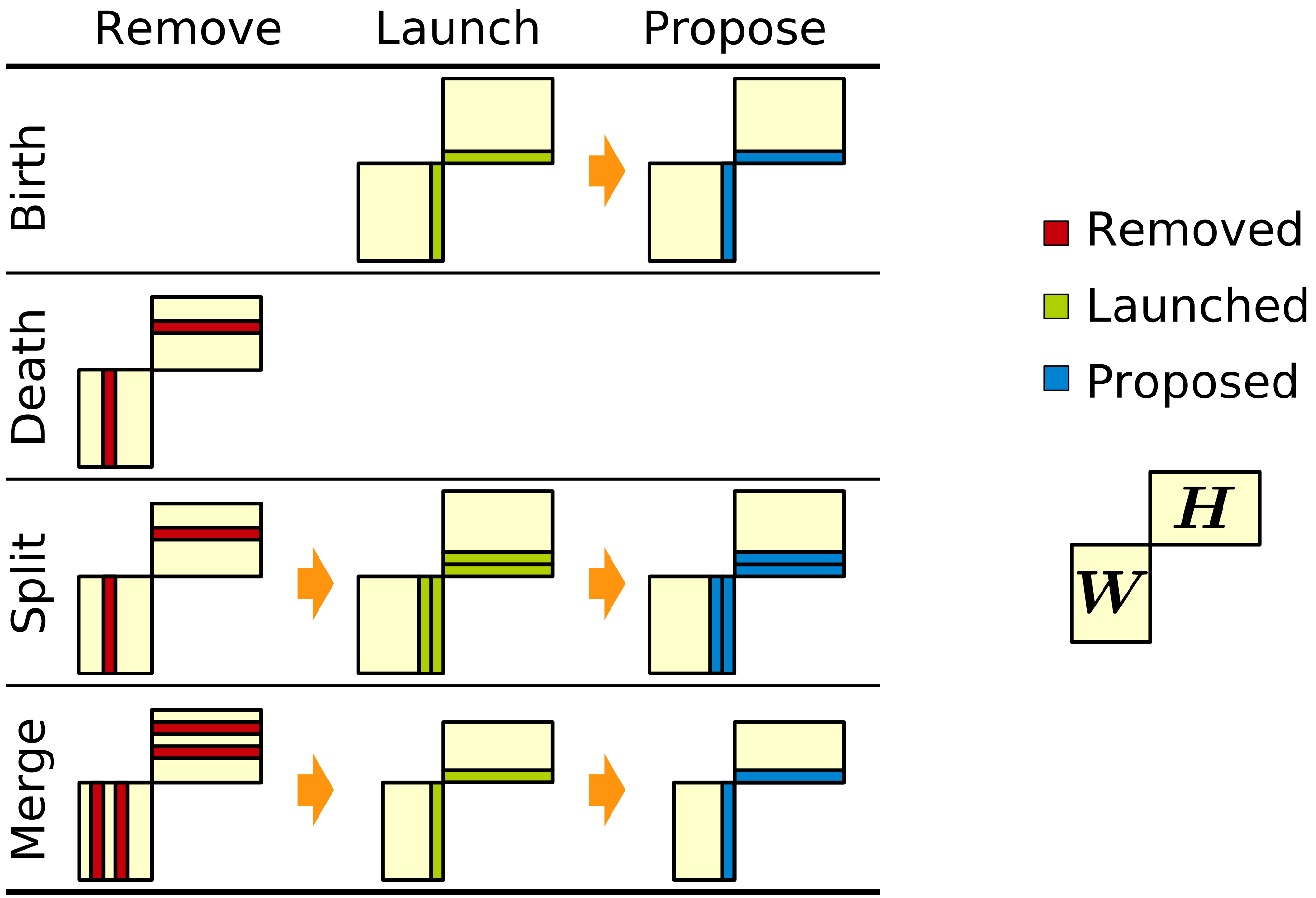
$$\boldsymbol{H} \sim p(\boldsymbol{H}|\boldsymbol{W}, D, \boldsymbol{V})$$

- Sampling the number of components
  - Requires joint update of factorizing matrices
  - Reversible jump MCMC (Metropolis-Hastings) with suitable proposal

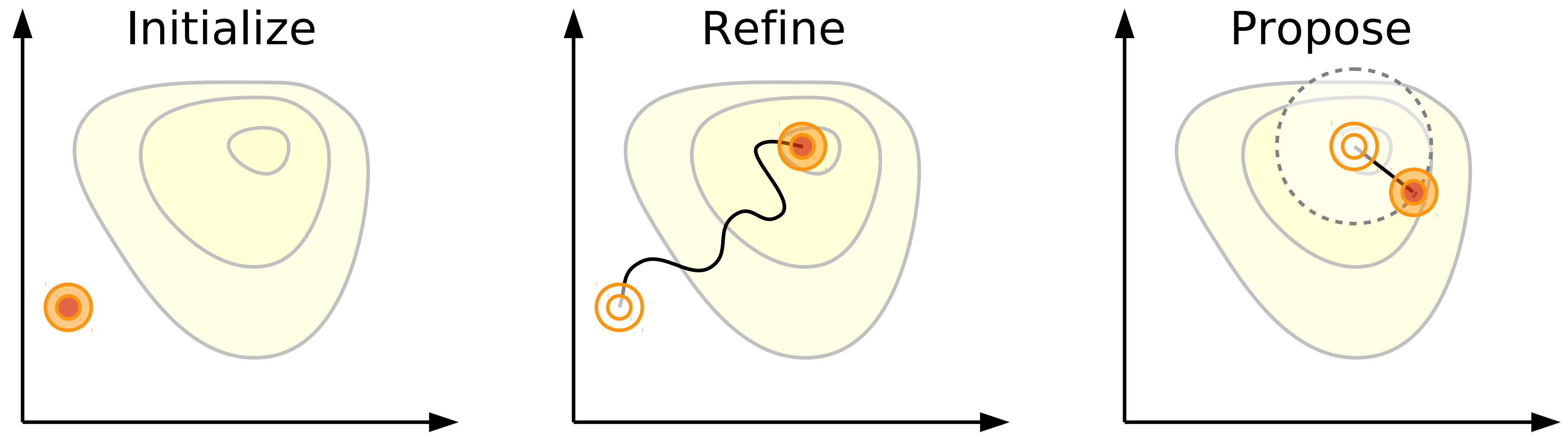
$$\boldsymbol{W}^*, \boldsymbol{H}^*, D^* \sim q(\boldsymbol{W}^*, \boldsymbol{H}^*, D^*|\boldsymbol{W}, \boldsymbol{H}, D)$$

Infinite non-negative matrix factorization

Reversible jump moves



Infinite non-negative matrix factorization
Proposal based on launch state



- Initialization: Draw from prior
- Refinement: Restricted Gibbs sampling
- Proposal: One final restricted Gibbs sweep

Details and expression for acceptance rate given in the paper

Infinite non-negative matrix factorization
Summary and discussion

- Bayesian NMF with an a priori unbounded number of factors
  - Learn the number of factors from data
  - Model order selection integrated in inference
- Efficient sampling scheme for cross-dimensional jumps
  - Based on reversible jump MCMC
  - Efficient proposals through high probability launch states
- Demonstrated on real and synthetic data
  - Reliably extract the correct model order
  - Lower comptational complexity than competing approaches