Applications of matrix completion in spectral analysis & synthesis

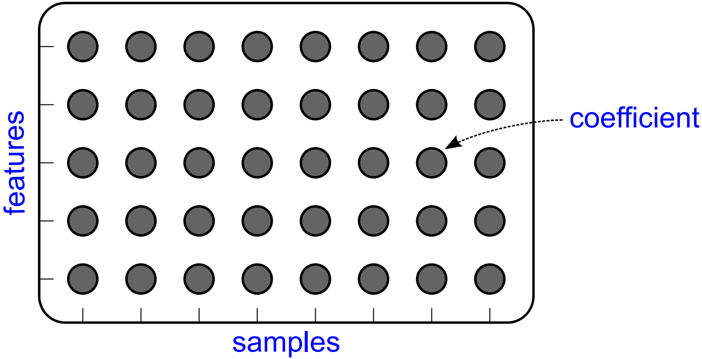
Cédric Févotte
Laboratoire Lagrange, Nice

Missing data in physics
Nice, May 2015

Outline

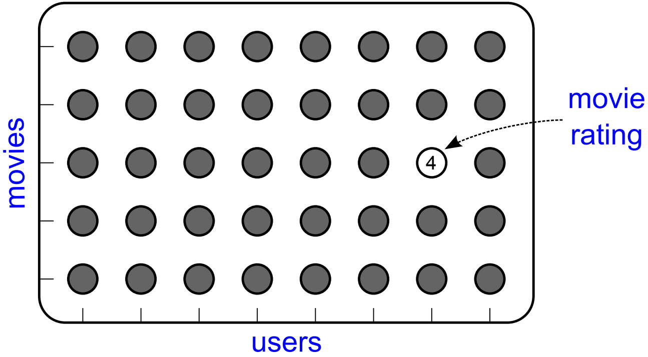
Matrix factorisation models

Data often available in matrix form.



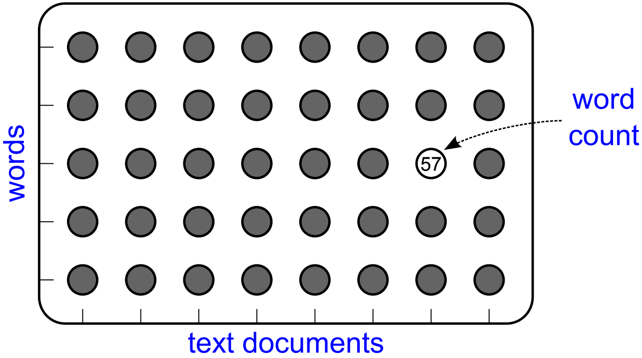
Matrix factorisation models

Data often available in matrix form.



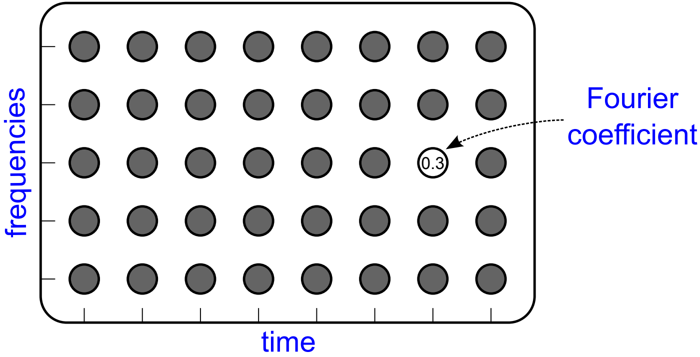
Matrix factorisation models

Data often available in matrix form.



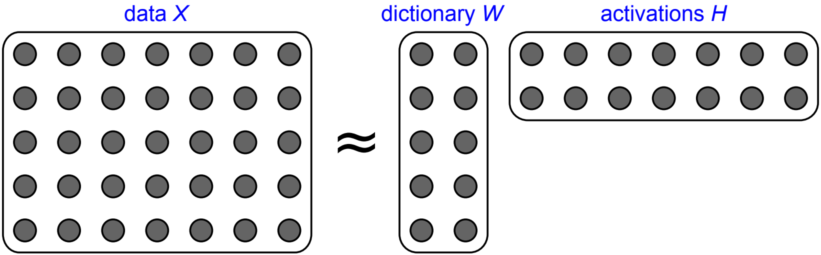
Matrix factorisation models

Data often available in matrix form.



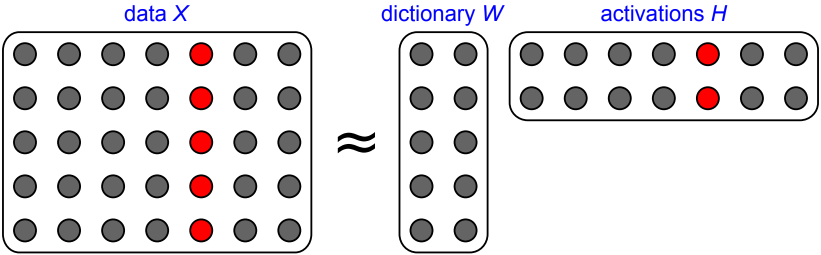
Matrix factorisation models

$\approx$ **dictionary learning**
low-rank approximation
factor analysis
latent semantic analysis



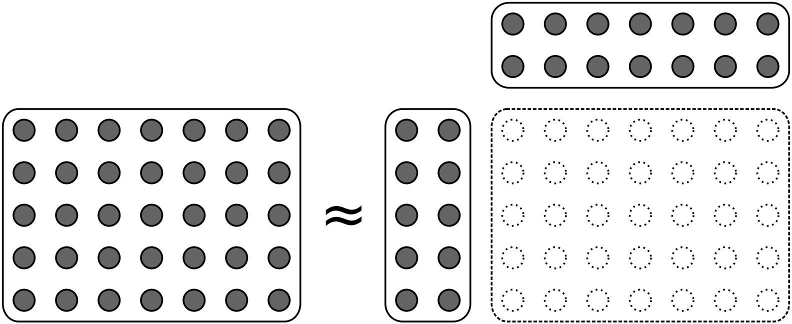
Matrix factorisation models

$\approx$ **dictionary learning**
low-rank approximation
factor analysis
latent semantic analysis



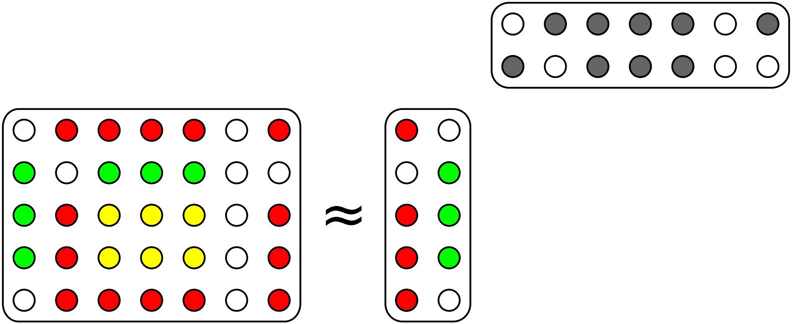
Matrix factorisation models

for dimensionality reduction (coding, low-dimensional embedding)



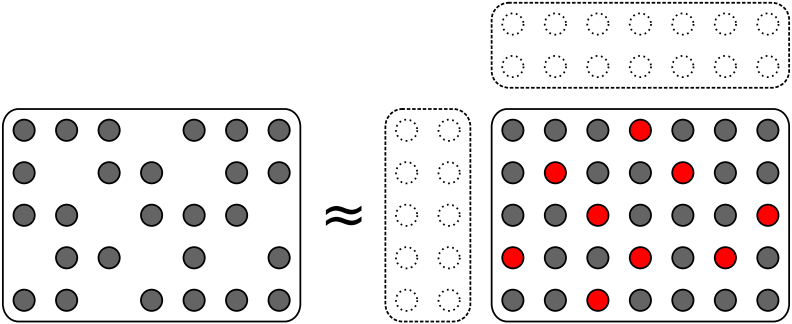
Matrix factorisation models

for unmixing (source separation, latent topic discovery)

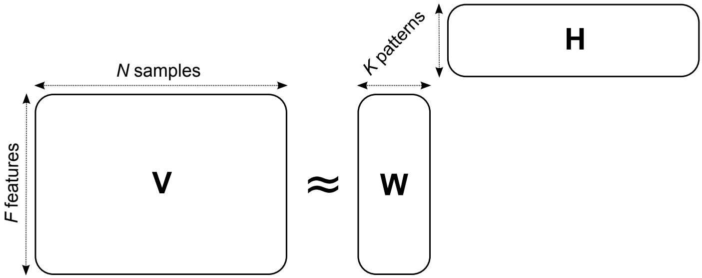


Matrix factorisation models

for interpolation (collaborative filtering, image inpainting)



Nonnegative matrix factorisation

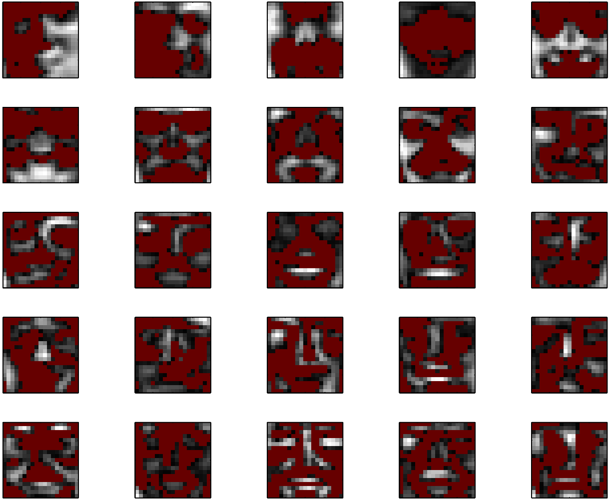


- data $\mathbf{V}$ and factors $\mathbf{W}$, $\mathbf{H}$ have nonnegative entries.
- nonnegativity of $\mathbf{W}$ ensures interpretability of the dictionary, because patterns $\mathbf{w}_k$ and samples $\mathbf{v}_n$ belong to the same space.
- nonnegativity of $\mathbf{H}$ tends to produce part-based representations, because subtractive combinations are forbidden.

Early work by Paatero and Tapper (1994), landmark *Nature* paper by Lee and Seung (1999)

49 images among 2429 from MIT's CBCL face dataset

PCA dictionary with $K = 25$



red pixels indicate negative values

NMF dictionary with $K = 25$

experiment reproduced from (Lee and Seung, 1999)

NMF as a constrained minimisation problem

Minimise a measure of fit between $\mathbf{V}$ and $\mathbf{WH}$, subject to nonnegativity:

$$\min_{\mathbf{W},\mathbf{H}\geq \mathbf{0}} D(\mathbf{V}|\mathbf{WH}) = \sum_{fn} d([\mathbf{V}]_{fn}|[\mathbf{WH}]_{fn}),$$

where $d(x|y)$ is a scalar cost function, e.g.,

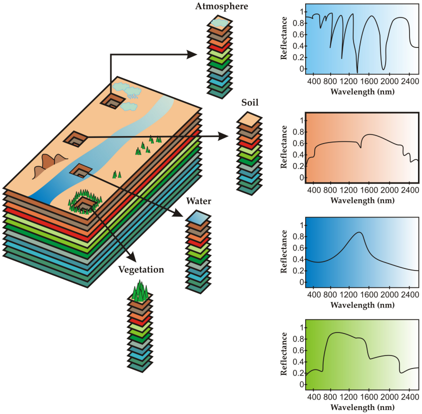
- Euclidean distance (Paatero and Tapper, 1994; Lee and Seung, 2001)
- Kullback-Leibler divergence (Lee and Seung, 1999; Finesso and Spreij, 2006)
- Itakura-Saito divergence (Févotte, Bertin, and Durrieu, 2009)
- α-divergence (Cichocki et al., 2008)
- β-divergence (Cichocki et al., 2006; Févotte and Idier, 2011)
- Bregman divergences (Dhillon and Sra, 2005)
- and more in (Yang and Oja, 2011)

Regularisation terms often added to $D(\mathbf{V}|\mathbf{WH})$ for sparsity, smoothness, dynamics, etc.

Common algorithmic design: alternative updates of $\mathbf{W}$ and $\mathbf{H}$ with majorisation-minimisation.

NMF for hyperspectral unmixing

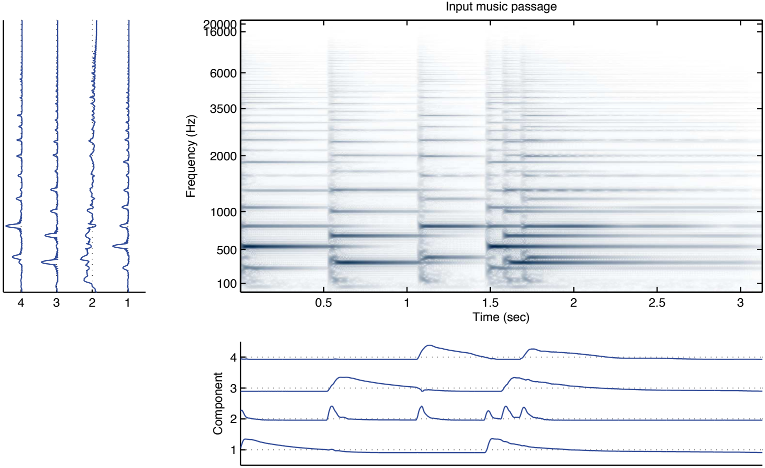
(Berry, Browne, Langville, Pauca, and Plemmons, 2007)



reproduced from (Bioucas-Dias et al., 2012)

NMF for audio spectral unmixing

(Smaragdis and Brown, 2003)



reproduced from (Smaragdis, 2013)

Outline

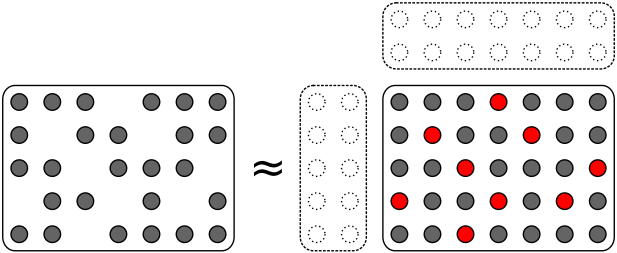
Model selection

- NMF based on the minimisation of a measure of fit:

$$\min_{\mathbf{W},\mathbf{H}\geq\mathbf{0}} D(\mathbf{V}|\mathbf{WH}) = \sum_{fn} d([\mathbf{V}]_{fn}|[\mathbf{WH}]_{fn})$$

- what is the right measure of fit ?
- can sometimes be derived from a probabilistic model, but not always.
- squared Euclidean distance often a default choice, but not always optimal.

Model selection by completion

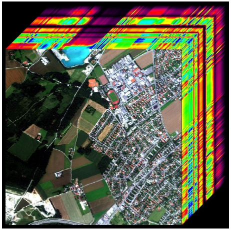


- randomly remove coefficients from $\mathbf{V}$ (with indices in $\mathcal{M}$)
- for a set of candidate measures $d(\cdot|\cdot)$, solve

$$\min_{\mathbf{W},\mathbf{H}\geq 0} D(\mathbf{V}|\mathbf{WH}) = \sum_{(f,n)\in\mathcal{O}} d([\mathbf{V}]_{fn}|[\mathbf{WH}]_{fn})$$

- reconstruct missing entries with $[\mathbf{WH}]_{fn}$, compare with original data $[\mathbf{V}]_{fn}$ (for indices in $\mathcal{M}$)
- choose the measure $d(\cdot|\cdot)$ that provides best reconstruction (according to a task-specific performance measure)

Hyperspectral data completion



- two datasets of dimensions $F \sim 150$ and $N = 50 \times 50$, from
  - the Aviris hyperspectral cube over Moffett Field (CA)
  - the Madonna hyperspectral cube over Villelongue (FR)
- candidate measures of fit from the β-divergence family
- evaluation using the average spectral angle mapper (aSAM)

$$\mathrm{aSAM}(\mathbf{V}, \hat{\mathbf{V}}) = \frac{1}{N} \sum_{n=1}^{N} \mathrm{acos}\left(\frac{\langle \mathbf{v}_n, \hat{\mathbf{v}}_n \rangle}{\|\mathbf{v}_n\| \|\hat{\mathbf{v}}_n\|} \right)$$

β-divergence

Popular cost function in NMF (Basu et al., 1998; Cichocki and Amari, 2010):

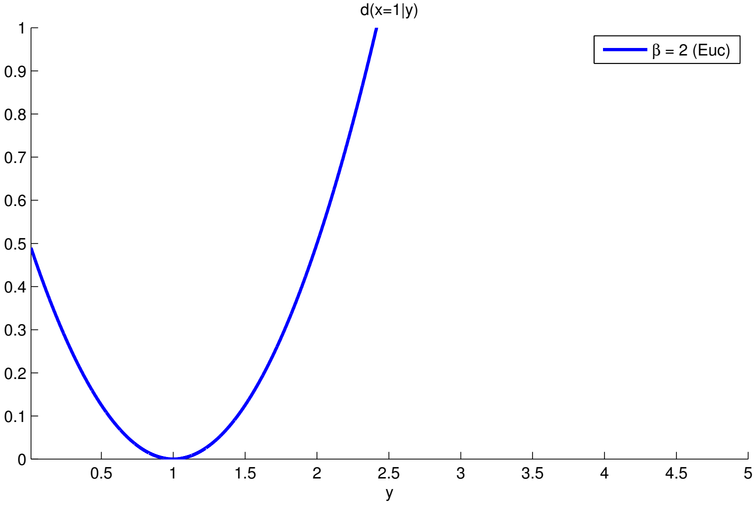
$$d_\beta(x|y) \stackrel{\text{def}}{=} \begin{cases} \frac{1}{\beta(\beta-1)} \left(x^\beta + (\beta-1)\, y^\beta - \beta\, x\, y^{\beta-1}\right) & \beta \in \mathbb{R}\backslash\{0,1\} \\ x \log \frac{x}{y} + (y - x) & \beta = 1 \\ \frac{x}{y} - \log \frac{x}{y} - 1 & \beta = 0 \end{cases}$$

Special cases:

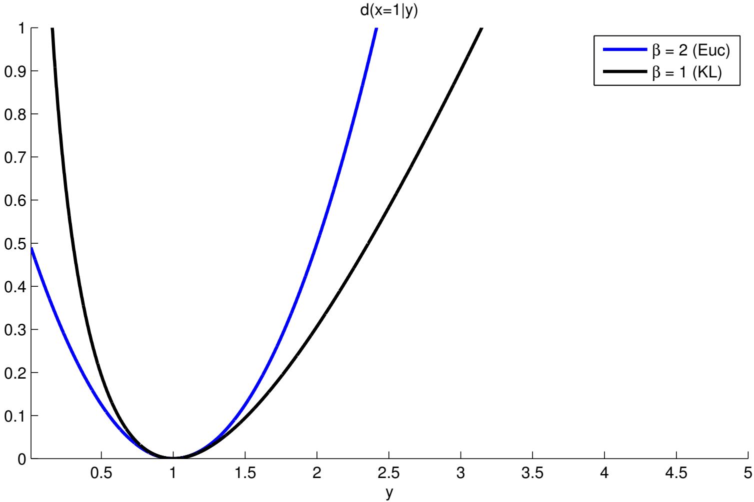
- squared Euclidean distance ($\beta = 2$)
- Kullback-Leibler (KL) divergence ($\beta = 1$)
- Itakura-Saito (IS) divergence ($\beta = 0$)

Behaviour with respect to scale: $d_\beta(\lambda x|\lambda y) = \lambda^\beta d_\beta(x|y)$.

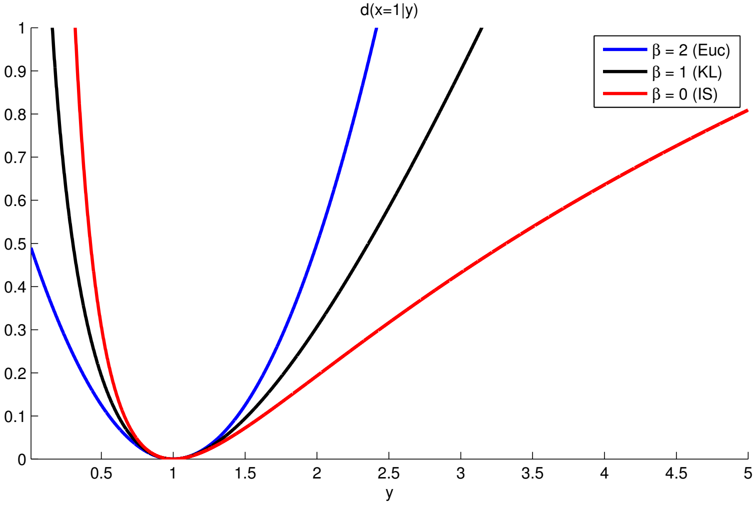
The β-divergence



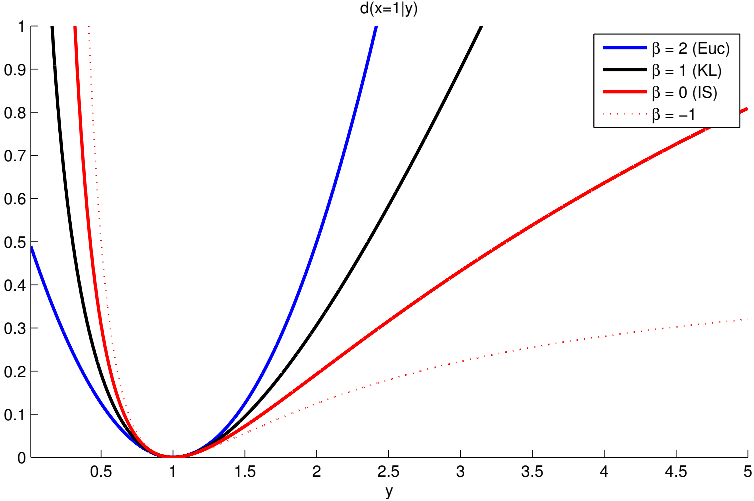
The β-divergence



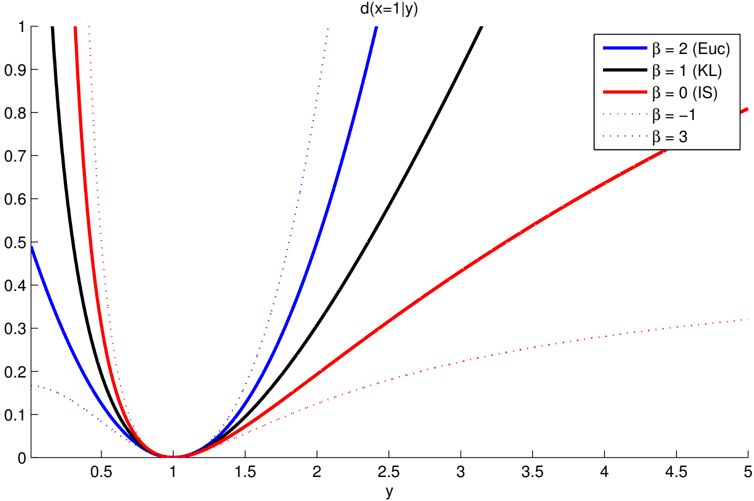
The β-divergence



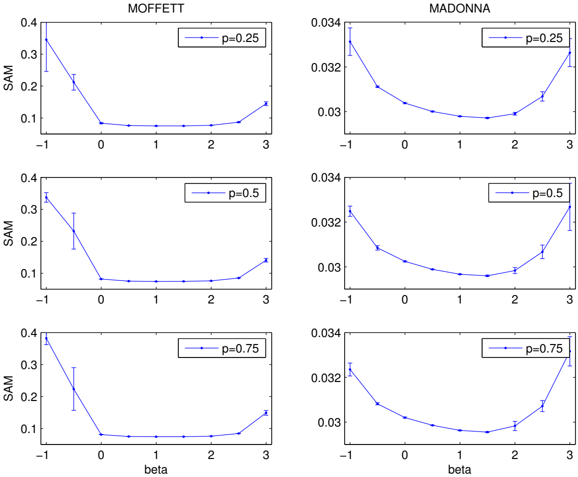
The β-divergence



The β-divergence



Hyperspectral data completion results
(Févotte and Dobigeon, 2014)



Best reconstruction for $\beta \approx 1$ ($\approx$ KL divergence), though values of in [0,2] yield sensibly similar results.

Audio spectral data completion results

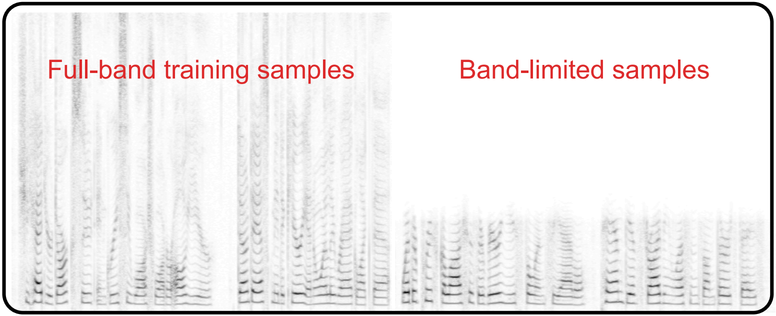
(King, Févotte, and Smaragdis, 2012)

- similar experiment conducted with music data.
- best reconstructions for $\beta \in [0, 1]$, depending on the spectrogram parameters.
- range of divergences more sensitive to small energies – because of $d_\beta(\lambda x|\lambda y) = \lambda^\beta d_\beta(x|y)$.

Outline

Audio bandwidth extension

(Sun and Mazumder, 2013)

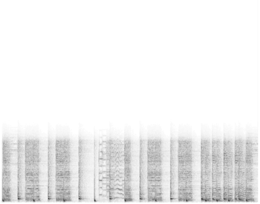
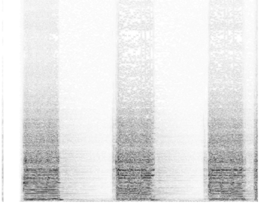
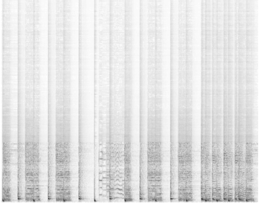
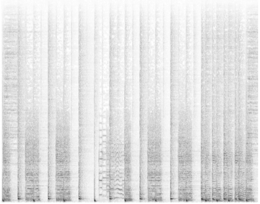
$V =$ 

adapted from (Sun and Mazumder, 2013)

Audio bandwidth extension

(Sun and Mazumder, 2013)

AC/DC example

band-limited data (*Back in Black*)	training data (*Highway to Hell*)
	
bandwidth extended	ground truth
	

Examples from `http://statweb.stanford.edu/~dlsun/bandwidth.html`*, used with permission from the author.*

References I

A. Basu, I. R. Harris, N. L. Hjort, and M. C. Jones. Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3):549–559, Sep. 1998.

M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics & Data Analysis*, 52(1):155–173, Sep. 2007.

J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):354–379, 2012.

A. Cichocki and S. Amari. Families of Alpha- Beta- and Gamma- divergences: Flexible and robust measures of similarities. *Entropy*, 12(6):1532–1568, June 2010.

A. Cichocki, R. Zdunek, and S. Amari. Csiszar's divergences for non-negative matrix factorization: Family of new algorithms. In *Proc. International Conference on Independent Component Analysis and Blind Signal Separation (ICA)*, pages 32–39, Charleston SC, USA, Mar. 2006.

A. Cichocki, H. Lee, Y.-D. Kim, and S. Choi. Non-negative matrix factorization with α-divergence. *Pattern Recognition Letters*, 29(9):1433–1440, July 2008.

I. S. Dhillon and S. Sra. Generalized nonnegative matrix approximations with Bregman divergences. In *Advances in Neural Information Processing Systems (NIPS)*, 2005.

C. Févotte and N. Dobigeon. Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization. Technical report, arxiv, 2014. URL http://arxiv.org/abs/1401.5649.

C. Févotte and J. Idier. Algorithms for nonnegative matrix factorization with the beta-divergence. *Neural Computation*, 23(9):2421–2456, Sep. 2011. doi: 10.1162/NECO_a_00168. URL http://www.unice.fr/cfevotte/publications/journals/neco11.pdf.

References II

C. Févotte, N. Bertin, and J.-L. Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis. *Neural Computation*, 21(3):793–830, Mar. 2009. doi: 10.1162/neco.2008.04-08-771. URL http://www.unice.fr/cfevotte/publications/journals/neco09_is-nmf.pdf.

L. Finesso and P. Spreij. Nonnegative matrix factorization and I-divergence alternating minimization. *Linear Algebra and its Applications*, 416:270–287, 2006.

B. King, C. Févotte, and P. Smaragdis. Optimal cost function and magnitude power for NMF-based speech separation and music interpolation. In *Proc. IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, Santander, Spain, Sep. 2012. URL http://www.unice.fr/cfevotte/publications/proceedings/mlsp12.pdf.

D. D. Lee and H. S. Seung. Learning the parts of objects with nonnegative matrix factorization. *Nature*, 401:788–791, 1999.

D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural and Information Processing Systems 13*, pages 556–562, 2001.

P. Paatero and U. Tapper. Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5:111–126, 1994.

P. Smaragdis. About this non-negative business. WASPAA keynote slides, 2013. URL http://web.engr.illinois.edu/~paris/pubs/smaragdis-waspaa2013keynote.pdf.

P. Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'03)*, Oct. 2003.

D. L. Sun and R. Mazumder. Non-negative matrix completion for bandwidth extension: a convex optimization approach. In *Proc. IEEE Workshop on Machine Learning and Signal Processing (MLSP)*, 2013.

References III

Z. Yang and E. Oja. Unified development of multiplicative algorithms for linear and quadratic nonnegative matrix factorization. *IEEE Transactions on Neural Networks*, 22:1878 – 1891, Dec. 2011. doi: http://dx.doi.org/10.1109/TNN.2011.2170094.