

# Bayesian linear regression onto data-driven dictionaries with application to music transcription

Cédric Févotte

CNRS-TELECOM ParisTech



3 July 2008  
Acoustics'08, Paris

Acknowledgements to N. Bertin & J.-L. Durrieu

# Regression

$$\mathbf{v} \approx \mathbf{W} \mathbf{h}$$

data vector

“explanatory variables”  
“basis”, “dictionary”  
“patterns”

“regressors”  
“expansion coefficients”  
“activation coefficients”

# Regression

$$[\mathbf{v}_1 \dots \mathbf{v}_N] \approx \mathbf{W} [\mathbf{h}_1 \dots \mathbf{h}_N]$$

data vectors                      “explanatory variables”                      “regressors”  
   “basis”, “dictionary”                      “expansion coefficients”  
   “patterns”                      “activation coefficients”

Usually,  $\mathbf{W}$  is learnt from a set of data vectors  $[\mathbf{v}_1 \dots \mathbf{v}_N]$ .

$$[\mathbf{v}_1 \dots \mathbf{v}_N] \approx \mathbf{W} [\mathbf{h}_1 \dots \mathbf{h}_N]$$

data vectors                      “explanatory variables”                      “regressors”  
   “basis”, “dictionary”                      “expansion coefficients”  
   “patterns”                      “activation coefficients”

Usually,  $\mathbf{W}$  is learnt from a set of data vectors  $[\mathbf{v}_1 \dots \mathbf{v}_N]$ .

E.g : Principal Component Analysis (PCA), Independent Component Analysis (ICA), Vector Quantization (VQ), and more recently Nonnegative Matrix Factorization (NMF).

# Nonnegative matrix factorization (NMF)

Non-subtractive, part-based decomposition of nonnegative data

[Lee and Seung, 99]

$$\min_{\mathbf{W}, \mathbf{H} \geq 0} D(\mathbf{V} | \mathbf{WH}) = \sum_{f=1}^F \sum_{n=1}^N d([\mathbf{V}]_{fn} | [\mathbf{WH}]_{fn})$$

E.g,  $d_{EUC}(x|y) = \frac{1}{2}(x - y)^2$ ,  $d_{KL}(x|y) = x \log \frac{x}{y} - x + y$

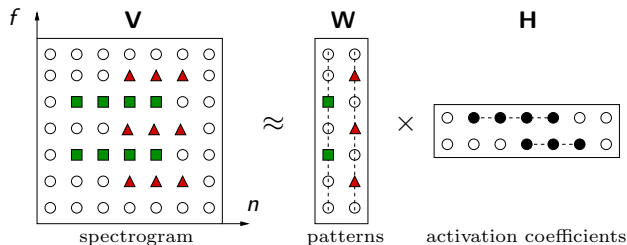
# Nonnegative matrix factorization (NMF)

Non-subtractive, part-based decomposition of nonnegative data

[Lee and Seung, 99]

$$\min_{\mathbf{W}, \mathbf{H} \geq 0} D(\mathbf{V} | \mathbf{WH}) = \sum_{f=1}^F \sum_{n=1}^N d([\mathbf{V}]_{fn} | [\mathbf{WH}]_{fn})$$

E.g.,  $d_{EUC}(x|y) = \frac{1}{2}(x - y)^2$ ,  $d_{KL}(x|y) = x \log \frac{x}{y} - x + y$



[Smaragdis and Brown, 03]

We address NMF with the **Itakura-Saito** (IS) divergence

$$d_{IS}(x|y) = \frac{x}{y} - \log \frac{x}{y} - 1$$

Our point is to show that

- 1) IS-NMF is underlain by a statistical model of superimposed Gaussian components,
- 2) this model is relevant to the representation of audio signals,
- 3) this model can accommodate regularization constraints through Bayesian approaches.

Builds on works by [Benaroya et al, 03] in single-sensor source separation and by [Abdallah and Plumbley, 04] in sparse coding.

# NMF with the Itakura-Saito divergence

Theorem 1: IS-NMF is ML estimation in sum of Gaussian components

Consider the generative model defined by,  $\forall n = 1, \dots, N$

$$\mathbf{x}_n = \sum_{k=1}^K \mathbf{c}_{k,n} \in \mathbb{C}^{F \times 1} \quad \text{with} \quad \mathbf{c}_{k,n} \sim \mathcal{N}_c(0, h_{kn} \text{diag}(\mathbf{w}_k))$$

Define  $[\mathbf{V}]_{fn} = |x_{fn}|^2$ .

# NMF with the Itakura-Saito divergence

Theorem 1: IS-NMF is ML estimation in sum of Gaussian components

Consider the generative model defined by,  $\forall n = 1, \dots, N$

$$\mathbf{x}_n = \sum_{k=1}^K \mathbf{c}_{k,n} \in \mathbb{C}^{F \times 1} \quad \text{with} \quad \mathbf{c}_{k,n} \sim \mathcal{N}_c(0, h_{kn} \text{diag}(\mathbf{w}_k))$$

Define  $[\mathbf{V}]_{fn} = |x_{fn}|^2$ . Then, ML estimation of  $\mathbf{W}$  and  $\mathbf{H}$  from  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$  is equivalent to NMF of  $\mathbf{V}$  into  $\mathbf{V} \approx \mathbf{WH}$ , where the IS divergence is used.

# NMF with the Itakura-Saito divergence

Theorem 1: IS-NMF is ML estimation in sum of Gaussian components

Consider the generative model defined by,  $\forall n = 1, \dots, N$

$$\mathbf{x}_n = \sum_{k=1}^K \mathbf{c}_{k,n} \in \mathbb{C}^{F \times 1} \quad \text{with} \quad \mathbf{c}_{k,n} \sim \mathcal{N}_c(0, h_{kn} \text{diag}(\mathbf{w}_k))$$

Define  $[\mathbf{V}]_{fn} = |x_{fn}|^2$ . Then, ML estimation of  $\mathbf{W}$  and  $\mathbf{H}$  from  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$  is equivalent to NMF of  $\mathbf{V}$  into  $\mathbf{V} \approx \mathbf{WH}$ , where the IS divergence is used.

- ▶ Nonnegativity arises naturally (model on  $\mathbf{x}_n$ , not  $|\mathbf{x}_n|$ ),
- ▶ Conservative Wiener reconstruction of the components

$$\hat{c}_{k,fn} = \frac{w_{fk} h_{kn}}{\sum_{l=1}^K w_{fl} h_{ln}} x_{fn} \quad \Rightarrow \quad \sum_{k=1}^K \hat{\mathbf{c}}_k = \mathbf{x}_k$$

# NMF with the Itakura-Saito divergence

Theorem 2: IS-NMF is ML estimation in multiplicative Gamma noise

Consider the generative model defined by

$$\mathbf{V} = (\mathbf{WH}) \cdot \mathbf{E}$$

where  $\mathbf{E}$  is multiplicative i.i.d Gamma noise with mean 1.

# NMF with the Itakura-Saito divergence

Theorem 2: IS-NMF is ML estimation in multiplicative Gamma noise

Consider the generative model defined by

$$\mathbf{V} = (\mathbf{WH}) \cdot \mathbf{E}$$

where  $\mathbf{E}$  is multiplicative i.i.d Gamma noise with mean 1. Then, ML estimation of  $\mathbf{W}$  and  $\mathbf{H}$  is equivalent to NMF of  $\mathbf{V}$  into  $\mathbf{V} \approx \mathbf{WH}$ , where the IS divergence is used.

# NMF with the Itakura-Saito divergence

Theorem 2: IS-NMF is ML estimation in multiplicative Gamma noise

Consider the generative model defined by

$$\mathbf{V} = (\mathbf{WH}) \cdot \mathbf{E}$$

where  $\mathbf{E}$  is multiplicative i.i.d Gamma noise with mean 1. Then, ML estimation of  $\mathbf{W}$  and  $\mathbf{H}$  is equivalent to NMF of  $\mathbf{V}$  into  $\mathbf{V} \approx \mathbf{WH}$ , where the IS divergence is used.

Denoting  $\hat{v}_{fn} = [\mathbf{WH}]_{fn}$

Cost	$p(v_{fn}   \hat{v}_{fn})$	Scale
$d_{EUC}(x y)$	$\mathcal{N}(v_{fn}   \hat{v}_{fn}, \sigma^2)$	$d_{EUC}(\lambda x   \lambda y) = \lambda^2 d_{EUC}(x y)$
$d_{KL}(x y)$	$\mathcal{P}(v_{fn}   \hat{v}_{fn})$	$d_{KL}(\lambda x   \lambda y) = \lambda d_{KL}(x y)$
$d_{IS}(x y)$	$\frac{1}{\hat{v}_{fn}} \mathcal{G}\left(\frac{v_{fn}}{\hat{v}_{fn}}   \alpha, \frac{1}{\alpha}\right)$	$d_{IS}(\lambda x   \lambda y) = d_{IS}(x y)$

# Multiplicative gradient descent algorithms

**Input** : nonneg. matrix  $\mathbf{V}$

**Output** : nonneg. matrices  $\mathbf{W}$  and  $\mathbf{H}$  such that  $\mathbf{V} \approx \mathbf{WH}$

Initialize  $\mathbf{W}$  and  $\mathbf{H}$  with positive values

**for**  $i = 1 : n_{iter}$  **do**

$$\mathbf{H} \leftarrow \mathbf{H} \cdot \frac{[\nabla_{\mathbf{H}} D(\mathbf{V}|\mathbf{WH})]_-}{[\nabla_{\mathbf{H}} D(\mathbf{V}|\mathbf{WH})]_+}$$

$$\mathbf{W} \leftarrow \mathbf{W} \cdot \frac{[\nabla_{\mathbf{W}} D(\mathbf{V}|\mathbf{WH})]_-}{[\nabla_{\mathbf{W}} D(\mathbf{V}|\mathbf{WH})]_+}$$

**end for**

[Lee and Seung, 01], [Cichocki et al, 06]

- ▶ Alternative structure implies local convergence
- ▶ Continuous descent of the cost is proven for  $d_{EUC}$  and  $d_{KL}$  (relies on convexity of these costs),
- ▶ Continuous descent of the cost is only observed for  $d_{IS}$  but is not proven (the proof does not generalize because  $d_{IS}$  is not convex).

# Small-scale example

Decomposition of a short piano sequence



(MIDI numbers : 61, 65, 68, 72)

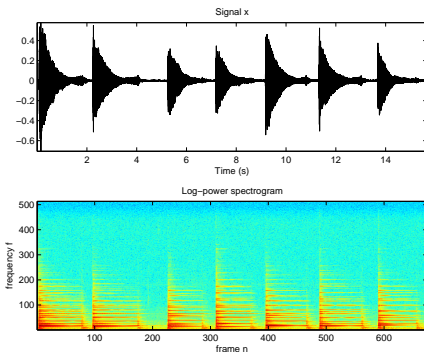
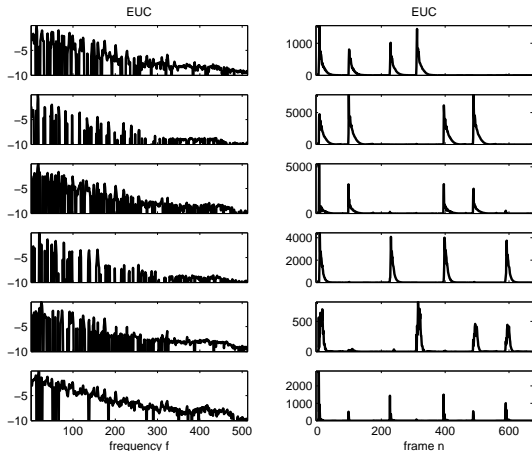


Figure : Three representations of data.

# Factorization with $K = 6$

## Euclidean NMF

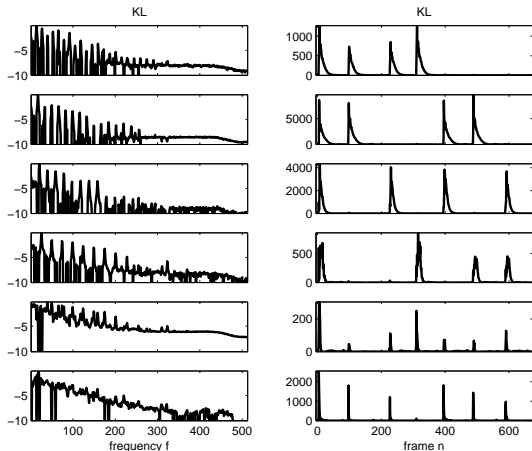


Pitch estimates: 61.0 65.0 65.4 68.0 72.0 0

(True values: 61, 65, 68, 72)

# Factorization with $K = 6$

## Kullback-Leibler NMF

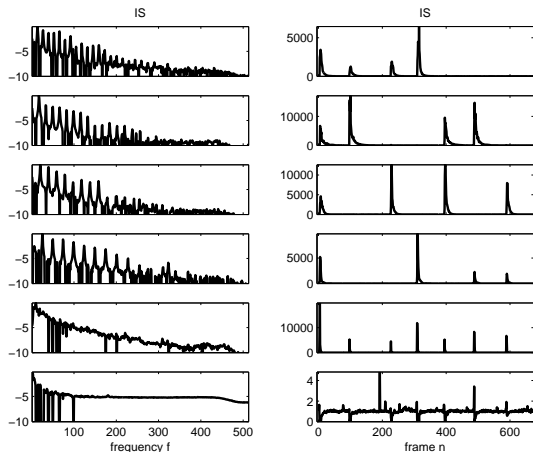


Pitch estimates: 61.0 65.0 68.0 72.0 81.0 0

(True values: 61, 65, 68, 72)

# Factorization with $K = 6$

Itakura-Saito NMF



Pitch estimates: 61.0 65.0 68.0 72.0 0 0

(True values: 61, 65, 68, 72)

# Influence of $K$

- ▶ When  $K$  is decreased :
  - ▶ Some components are merged
  - ▶ IS-NMF always produces a low-energy component (scale-invariance)
  
- ▶ When  $K$  is increased :
  - ▶ Euclidean NMF and KL-NMF tend to split existing components  
Ex: note sustain + decay subcomponents
  - ▶ IS-NMF tends to add and refine low-energy components  
Ex : hammer hits and pedal releases eventually get split

(all results available on my webpage)

**Message:** IS-NMF appears to follow our object-based comprehension of sound.

# Expectation-Maximization algorithm

Towards Bayesian approaches

More precisely a **Space Alternating Generalized EM (SAGE) algorithm** where component  $\mathbf{C}_k$  is a hidden-data space for column  $\mathbf{w}_k$  of  $\mathbf{W}$  and row  $h_k$  of  $\mathbf{H}$ .

# Expectation-Maximization algorithm

Towards Bayesian approaches

More precisely a **Space Alternating Generalized EM (SAGE) algorithm** where component  $\mathbf{C}_k$  is a hidden-data space for column  $\mathbf{w}_k$  of  $\mathbf{W}$  and row  $h_k$  of  $\mathbf{H}$ .

**Input** : nonneg. matrix  $\mathbf{V}$

**Output** : nonneg. matrices  $\mathbf{W}$  and  $\mathbf{H}$  such that  $\mathbf{V} \approx \mathbf{WH}$

Initialize  $\mathbf{W}$  and  $\mathbf{H}$  with positive values

**for**  $i = 1 : n_{iter}$  **do**

**for**  $k = 1 : K$  **do**

**E-step**

$$Q_k(\mathbf{w}_k, h_k | \mathbf{W}', \mathbf{H}') = - \int_{\mathbf{C}_k} \log p(\mathbf{C}_k | \mathbf{w}_k, h_k) p(\mathbf{C}_k | \mathbf{X}, \mathbf{W}', \mathbf{H}') d\mathbf{C}_k$$

**M-step**

    Minimize  $Q_k(\mathbf{w}_k, h_k | \mathbf{W}', \mathbf{H}')$

**end for**

**end for**

# Expectation-Maximization algorithm

Towards Bayesian approaches

More precisely a **Space Alternating Generalized EM (SAGE) algorithm** where component  $\mathbf{C}_k$  is a hidden-data space for column  $\mathbf{w}_k$  and row  $h_k$ .

**Input** : nonneg. matrix  $\mathbf{V}$

**Output** : nonneg. matrices  $\mathbf{W}$  and  $\mathbf{H}$  such that  $\mathbf{V} \approx \mathbf{WH}$

Initialize  $\mathbf{W}$  and  $\mathbf{H}$  with positive values

**for**  $i = 1 : n_{iter}$  **do**

**for**  $k = 1 : K$  **do**

**E-step**

$$Q_k(\mathbf{w}_k, h_k | \mathbf{W}', \mathbf{H}') = - \int_{\mathbf{C}_k} \log p(\mathbf{C}_k | \mathbf{w}_k, h_k) p(\mathbf{C}_k | \mathbf{X}, \mathbf{W}', \mathbf{H}') d\mathbf{C}_k$$

**M-step**

    Minimize  $Q_k(\mathbf{w}_k, h_k | \mathbf{W}', \mathbf{H}') - \log p(\mathbf{w}_k) - \log p(h_k)$  for MAP

**end for**

**end for**

# Favoring smoothness

using Gamma or inverse-Gamma Markov chains

Take  $p(h_k) = \prod_{n=2}^N p(h_{kn}|h_{k(n-1)}) p(h_{k1})$

such that  $\operatorname{argmax} p(h_{kn}|h_{k(n-1)}) = h_{k(n-1)}$

# Favoring smoothness

using Gamma or inverse-Gamma Markov chains

Take  $p(h_k) = \prod_{n=2}^N p(h_{kn}|h_{k(n-1)}) p(h_{k1})$

such that  $\operatorname{argmax} p(h_{kn}|h_{k(n-1)}) = h_{k(n-1)}$

E.g :  $\mathcal{IG}(h_{kn}|\alpha, (\alpha + 1) h_{k(n-1)})$  or  $\mathcal{G}(h_{kn}|\alpha, (\alpha - 1)/h_{k(n-1)})$

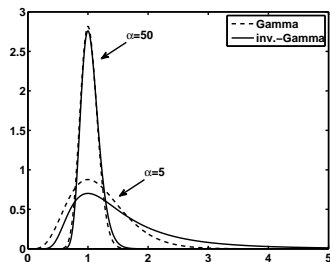
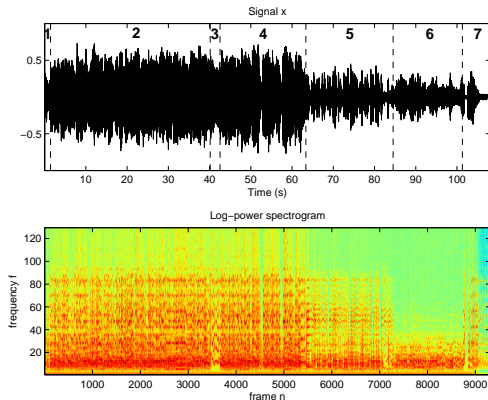


Figure:  $p(h_{kn}|h_{k(n-1)}=1)$

Similar modeling with KL-NMF in [\[Virtanen et al, 08\]](#) (see Taylan Cemgil's talk)

# Large-scale example

Denosing and upmixing and Louis Armstrong and His Hot Five

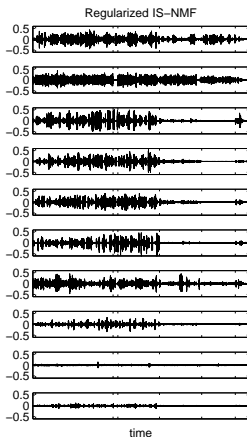
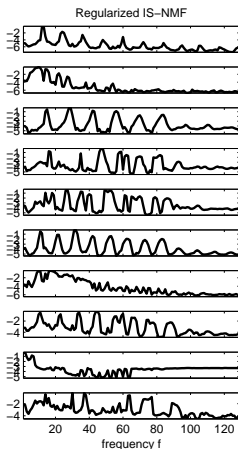


**Figure :** (2,4,7) all instruments : trumpet, clarinet, trombone, piano and double bass  
(1) clarinet only, (3) trumpet solo, (5) clarinet and piano, (6) piano solo.

data

# Factorization with $K = 10$

## Regularized Itakura-Saito NMF



Original

Denoised

Exc. Orig. / Den.

Accompaniment

Comp. 2

Same w/o prior

Trombone

Comp. 7

Trump. + Clar.

Comp. 1, 3-6, 8, 10

Denoised & Upmixed

Comp. 1 2 3 4 5 6 7 8 9 10

# Conclusions

- 1) IS-NMF is underlain by a statistical model of superimposed Gaussian components,  
Nonnegativity arises naturally.  
Wiener reconstruction has sense here.  
Such statistical setting is missing in NMF with most other cost functions.
- 2) this model is relevant to the representation of audio signals,  
Note spectra look like note spectra.  
Great accuracy for low-power component.
- 3) this model can accommodate regularization constraints through Bayesian approaches.  
The Bayesian setting allows complex prior structures...  
... and complex inference tools ! (MCMC, variational Bayes, etc.)

## Technical report and audio samples on my webpage

C. Févotte, N. Bertin and J.-L. Durrieu. “Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis.” May 2008.