

# Automatic relevance determination in nonnegative matrix factorization with the $\beta$ -divergence

Vincent Y. F. Tan<sup>1</sup> and Cédric Févotte<sup>2</sup>

<sup>1</sup>University of Wisconsin-Madison



<sup>2</sup>CNRS LTCI; Télécom ParisTech  
Paris, France



NIPS “Sparse Low Rank” workshop  
December 2011

# Nonnegative matrix factorization (NMF)

Given a *nonnegative* matrix  $\mathbf{V}$  of dimensions  $F \times N$ , NMF is the problem of finding a factorization

$$\mathbf{V} \approx \mathbf{WH}$$

where  $\mathbf{W}$  and  $\mathbf{H}$  are *nonnegative* matrices of dimensions  $F \times K$  and  $K \times N$ , respectively.

# Nonnegative matrix factorization (NMF)

Given a *nonnegative* matrix  $\mathbf{V}$  of dimensions  $F \times N$ , NMF is the problem of finding a factorization

$$\mathbf{V} \approx \mathbf{WH}$$

where  $\mathbf{W}$  and  $\mathbf{H}$  are *nonnegative* matrices of dimensions  $F \times K$  and  $K \times N$ , respectively.

Constrained optimization problem:

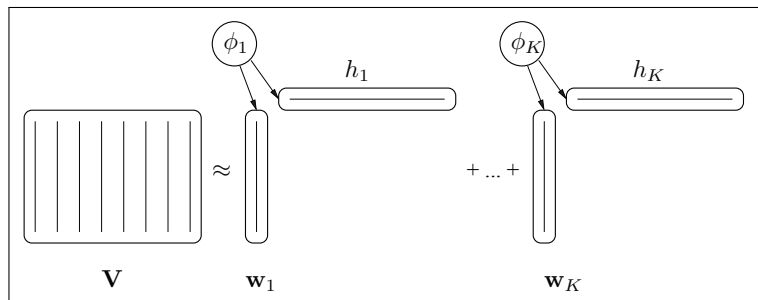
$$\min_{\mathbf{W}, \mathbf{H} \geq 0} D(\mathbf{V} | \mathbf{WH}) = \sum_{fn} d([\mathbf{V}]_{fn} | [\mathbf{WH}]_{fn})$$

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

Objective of this work is to identify the “right” value of  $K$ .

## Automatic relevance determination in NMF

Inspired by Bayesian PCA (Bishop, 1999): each “component”  $k$  is assigned a relevance (= variance) parameter  $\phi_k$ .



Half-Gaussian or exponential priors on  $w_k$  and  $h_k$ .

$$\text{E.g., } p(w_k | \phi_k) = \prod_f \phi_k^{-1} \exp -\phi_k^{-1} w_{fk}, \quad p(h_k | \phi_k) = \prod_n \phi_k^{-1} \exp -\phi_k^{-1} h_{kn}$$

# Automatic relevance determination in NMF

After a few manipulations, we are essentially left with the minimization of

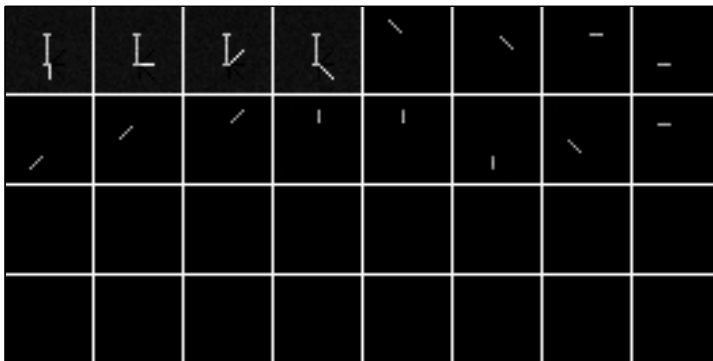
$$C(\mathbf{W}, \mathbf{H}) = D_{\beta}(\mathbf{V}|\mathbf{WH}) + \rho \sum_{k=1}^K \log(\|\mathbf{w}_k\| + \|\mathbf{h}_k\| + \varepsilon)$$

where

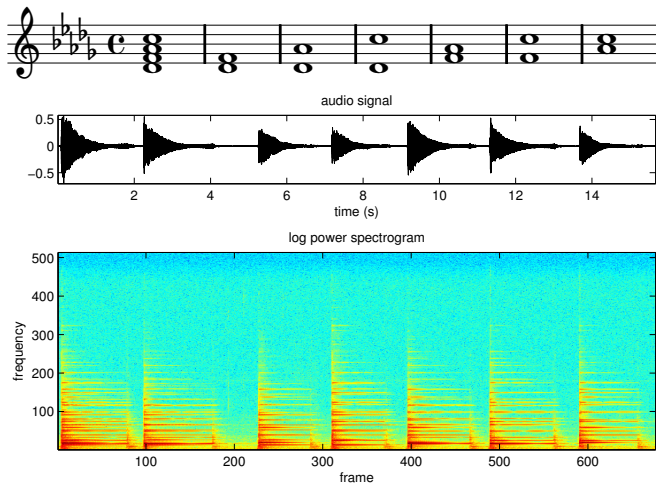
- ▶  $D_{\beta}(\mathbf{V}|\mathbf{WH})$  is the measure of fit (in this work,  $\beta$ -divergence)
- ▶  $\|\mathbf{x}\| = \frac{1}{2}\|\mathbf{x}\|_2^2$  (half-Gaussian priors) or  $\|\mathbf{x}\| = \|\mathbf{x}\|_1$  (exponential priors).

## Swimmer decomposition results

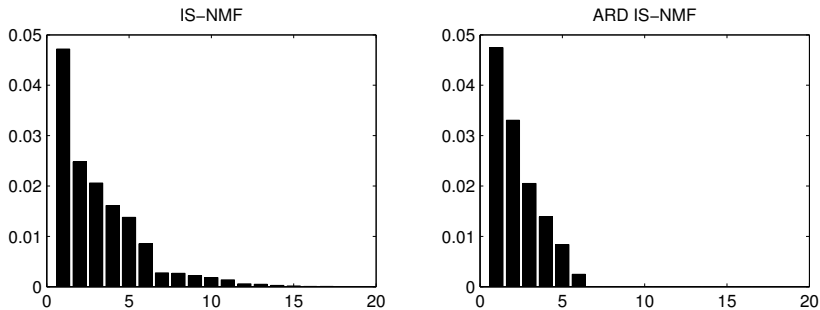
8 data samples (among 256)

Estimated  $\mathbf{W}$  using with exponential priors /  $\ell_1$  penalization

# Audio decomposition results



# Audio decomposition results



**Figure:** Histograms of standard deviation values of all  $K = 18$  components produced by Itakura-Saito NMF and ARD Itakura-Saito NMF (with  $\ell_2$  penalization). ARD IS-NMF only retains the 6 “right” components.

*Check our full-length technical report available on arxiv.*