Low-rank time-frequency synthesis

Cédric Févotte\textsuperscript{1} & Matthieu Kowalski\textsuperscript{2}

\textsuperscript{1}Laboratoire Lagrange, Nice
(CNRS, Observatoire de la Côte d’Azur & Université Nice Sophia Antipolis)

\textsuperscript{2}Laboratoire des Signaux et Systèmes (LSS), Gif-sur-Yvette
(CNRS, CentraleSupelec & Université Paris-Sud)

SPARS 2015
Cambridge, England
Spectral unmixing by NMF

▶ $y_{fn} = \sum_t x(t) \phi_{fn}^\ast(t)$: short-time Fourier transform (STFT) of temporal signal $x(t)$.

▶ $s_{fn} = |y_{fn}|^2$: power spectrogram.

▶ NMF extracts recurring spectral patterns from the data by solving

$$\min_{W,H \geq 0} D(S | WH).$$

▶ Successful applications in audio source separation and music transcription.
Itakura-Saito NMF & the Gaussian composite model
(Févotte, Bertin, and Durrieu, 2009)

- Low-rank variance model of analysis coefficients (STFT):
  \[ y_{fn} \sim N_c(0, [WH]_{fn}). \]

- Log-likelihood equivalent to Itakura-Saito (IS) divergence:
  \[ -\log p(Y|WH) = D_{IS}(|Y|^2|WH) + \text{cst.} \]

- Underlies a Gaussian composite model (GCM):
  \[ y_{fn} = \sum_k y_{kfn}, \]
  \[ y_{kfn} \sim N_c(0, w_{fk}h_{kn}). \]

- Given estimates of \( W \) and \( H \), latent STFT components can be estimated by Wiener filter:
  \[ \hat{y}_{kfn} = \frac{w_{fk}h_{kn}}{[WH]_{fn}} y_{fn}. \]

- Inverse-STFT of \( \{\hat{y}_{kfn}\}_{fn} \) produces temporal components such that
  \[ x(t) = \sum_k \hat{c}_k(t). \]
Low-rank time frequency synthesis (LRTFS)  
(Févotte and Kowalski, 2014)

- Low-rank variance model of synthesis coefficients:
  \[ x(t) = \sum_{f_n} \alpha_{fn} \phi_{fn}(t) + e(t), \]
  \[ \alpha_{fn} \sim N_c(0, [WH]_{fn}), \]
  \[ e(t) \sim N_c(0, \lambda). \]

- \( \phi_{fn}(t) \): time-frequency atom (e.g., from a Gabor frame),
- \( \alpha_{fn} \): synthesis coefficient,
- \( e(t) \): residual term.

- LRTFS is a generative model of raw data \( x(t) \).
- Like in the GCM, the synthesis coefficients have a latent composite structure:
  \[ \alpha_{fn} = \sum_k \alpha_{kfn}, \]
  \[ \alpha_{kfn} \sim N_c(0, w_{fk} h_{kn}). \]

- Given estimates of \( W \) of \( H \), latent coefficients \( \alpha_{kfn} \) can be estimated from their posterior mean and temporal components can be reconstructed as 
  \[ \hat{c}_k(t) = \sum_{f_n} \hat{\alpha}_{kfn} \phi_{fn}(t). \]
Relation to sparse Bayesian learning (SBL)

- Generative signal model in vector/matrix form:

\[ x = \Phi \alpha + e. \]

- **x, e**: vectors of signal and residual time samples (size \( T \)),
- **\( \alpha \)**: vector of synthesis coefficients \( \alpha_{fn} \) (size \( FN \)),
- **\( \Phi \)**: time-frequency dictionary (size \( T \times FN \)).

- Synthesis coefficients model in vector/matrix form:

\[ p(\alpha|v) = N_c(\alpha|0, \text{diag}(v)). \]

- **v**: vector of variance coefficients \( v_{fn} = [WH]_{fn} \) (size \( FN \)).

- Similar to sparse Bayesian learning (Tipping, 2001; Wipf and Rao, 2004) except that the variance parameters are tied together by the low-rank structure \( WH \).
Estimation in LRTFS
Maximum joint likelihood (JL)

▶ Optimise

\[ C_{\text{JL}}(\alpha, W, H, \lambda) \overset{\text{def}}{=} -\log p(x, \alpha|W, H, \lambda) \]
\[ = \frac{1}{\lambda} \|x - \Phi \alpha\|_2^2 + D_{\text{IS}}(|\alpha|^2|v) + \log(|\alpha|^2) + \text{cst}. \]

▶ Possible EM algorithm using the procedure of (Figueiredo and Nowak, 2003) based on the hidden variable \( z \) such that

\[
\begin{align*}
    x &= \Phi z + e_1, \\
    z &= \alpha + \sqrt{\beta} e_2,
\end{align*}
\]

with \( e_1 \sim N_c(0, \lambda I - \beta \Phi \Phi^*) \) and \( e_2 \sim N_c(0, I) \) (condition applies on \( \beta \)).

▶ Leads to a form of iterative shrinkage algorithm that scales well with real-world signal dimension. Shrinkage operator involves Itakura-Saito NMF of power posterior expectation of \( \alpha \) at each iteration.
Estimation in LRTFS
Maximum marginal likelihood (ML)

- Integration of $\alpha$ from the joint likelihood (like in SBL)

$$C_{ML}(W, H, \lambda) \overset{\text{def}}{=} - \log p(x|W, H, \lambda)$$
$$= - \log \int_{\alpha} p(x|\alpha, \lambda)p(\alpha|WH)d\alpha$$

- Possible EM algorithm treating $\alpha$ as the hidden variable. Tractable algorithm but does not scale well with dimension.
Figure: Three representations of data.
Multi-resolution LRTFS

LRTFS allows for multi-resolution hybrid representations:

\[ x = \Phi_a \alpha_a + \Phi_b \alpha_b + e. \]

- \( \Phi_a \) and \( \Phi_b \) are time-frequency dictionaries with different resolutions,
- \( \alpha_a \) and \( \alpha_b \) have their own latent low-rank structure \( W_a H_a \) and \( W_b H_b \).

Not possible with standard NMF!

Previous optimisation strategies apply by concatenation:

\[ \Phi = [\Phi_a \ \Phi_b] \]
\[ \alpha = \begin{bmatrix} \alpha_a \\ \alpha_b \end{bmatrix} \]

Other hybrid decompositions are possible. E.g., low-rank layer + sparse layer (forthcoming EUSIPCO paper)
Hybrid decomposition of jazz music

\[ x = \Phi_{\text{tonal}} \alpha_{\text{tonal}} + \Phi_{\text{transient}} \alpha_{\text{transient}} + e \]

- spectrogram of \( x \)
- \( \log |\hat{\alpha}_{\text{transient}}| \)
- \( \log |\hat{\alpha}_{\text{tonal}}| \)

Latent rank-1 components from transient layer

Latent rank-1 components from tonal layer
Speech enhancement in applause noise

\[ \mathbf{x} = \Phi_{\text{tonal}} \left( \alpha_{\text{speech,tonal}} + \alpha_{\text{noise,tonal}} \right) + \Phi_{\text{transient}} \left( \alpha_{\text{speech,transient}} + \alpha_{\text{noise,transient}} \right) + \mathbf{e} \]

Spectrograms of \( \mathbf{x} \) with short and large resolutions

Noisy signal: long window STFT analysis

Noisy signal: short window STFT analysis

\[ \log | \hat{\alpha}_{\text{speech,tonal}} | \]

Denoised signal: Tonal Layer

\[ \log | \hat{\alpha}_{\text{speech,transient}} | \]

Denoised signal: Transient Layer
A dedication to Bill Fitzgerald

Prof. William J. Fitzgerald (1948-2014)

An Introduction to Bayesian Inference Applied to Signal and Data Processing

W J Fitzgerald

Signal Processing Laboratory, Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK
wjf@eng.cam.ac.uk
16/5/2001

(my first tutorial on Bayesian inference)


