Neural Information Processing Systems Conference

ADAPTIVE TEMPLATE MATCHING WITH SHIFT-INVARIANT SEMI-NMF

Jonathan Le Roux[1,2], Alain de Cheveigné[2], Lucas C. Parra[3]

[1]The University of Tokyo, Tokyo, Japan; [2]Ecole Normale Supérieure, Paris, France; [3]City University of New York, New York, USA

Abstract

The signal to be analyzed is decomposed into a weighted sum of templates. Templates are learned from the data in contrast to matching pursuit; Weights are positive in contrast to ICA, PCA, etc; Template sign is unconstrained in contrast to NMF; Templates are allowed to shift in time in contrast to semi-NMF. Sparsity is enforced. The method is applied to single-electrode extracellular recordings in the cochlear nucleus, and to audio signals. The algorithm achieves good performance if the SNR is above 6dB and templates are sufficiently distinct.

Motivation

- Task: Spike sorting in extracellular recordings
  - Challenge: overlapping spikes
- Task: Template matching
  - Challenge: unknown templates
- Task: Signal decomposition
  - Challenge: "good" decomposition, e.g. sparse, non-negative amplitudes.
- Exploit waveform regularity/reproducibility
  - Spike trains in extracellular recordings
  - Electronic music, instruments like piano, drums

Goal

Explain a time series with a decomposition that

- Detects timing of "events"
- Events characterized and identified by their time course
- Time course of events discovered from data de novo
- Events have variable amplitudes (but positive)
- Overlapping events are additive

Signal model

We model data sequence with a convolutive factorization

$$\hat{x}_i(t) = \sum_k \sum_l b_k(t) a_{ik}(t-l)$$

- Each k stands for a different type of event with time course $b_k(t)$
- Each i stands for a different example of a time sequence
- We think of $a_{ik}(t)$ as amplitudes, therefore demand

$$a_{ik}(t) \geq 0$$

- Further, if we think of $a_{ik}(t)$ as a spike train (of variable amplitude), we should demand that $a_{ik}(t)$ is sparse, i.e., follows the prior distribution of the form:

$$p(a) \propto e^{-|\frac{a}{\sigma}|^\alpha}$$

Relation to other criteria:

NMF (Lee&Seung) $\hat{X} = AB,\ A \geq 0,\ B \geq 0$	**Shifted-NMF** (Morup et al.) $\hat{x}(t) = A(t)*b(t),\ A(t) \geq 0,\ b(t) \geq 0$
Semi-NMF (Ding et al.) $\hat{X} = AB,\ A \geq 0$	**Shift-invariant semi-NMF** (prop.) $\hat{x}(t) = A(t)*b(t),\ A(t) \geq 0$

Optimization of the model

- In short, including noise, the data $x(t)$ is modeled as

$$x(t) = A(t) * b(t) + n(t)$$

- With Gaussian independent noise and independent amplitudes, the MAP estimate is then

$$\underset{A(t)\geq 0,\ \sum_t |b_k(t)|^2=1}{\operatorname{argmax}} \frac{1}{2}\sum_t \|x(t) - A(t)*b(t)\|_2^2 + \beta \sum_t \|A(t)\|_\alpha^\alpha$$

  - Templates are normalized to avoid scaling ambiguity
  - $\|\ \|_p$ stands for the L_p norm (quasi-norm for $0<p<1$)
  - Strength of L_α penalty depends on SNR:

$$\beta = \frac{\sigma_N^2}{\sigma_A^\alpha}\left(\Gamma(3/\alpha)/\Gamma(1/\alpha)\right)^{\alpha/2}$$

  - Sparseness determined by α depends on firing rate
- Can also be written based on Einstein notation ($a^k b_k = \sum_k a_k b_k$):

$$\hat{x}_t = \tilde{A}_t^{kl} B_{kl} = A_{kl}\tilde{B}_t^{kl}$$

with the shifted amplitudes and templates $\begin{cases}\tilde{A}_{klt} = A_k(t-l)\\ \tilde{B}_{klt} = b_k(t-l)\end{cases}$

- A update with sparseness term

$$A_{nk} \leftarrow A_{nk}\sqrt{\frac{\left(\tilde{X}_n^l B_{kl}\right)^+ + A^{n'k'}\left(\tilde{B}^t_{n'k'}\tilde{B}_{nkt}\right)^-}{\left(\tilde{X}_n^l B_{kl}\right)^- + A^{n'k'}\left(\tilde{B}^t_{n'k'}\tilde{B}_{nkt}\right)^+ + \alpha\beta A_{nk}^{\alpha-1}}}$$

  - Convergence proof including sparseness term can be obtained based on inequality derived by Kameoka:

$$A_{nk}^{\alpha} \leq \frac{\alpha A_{nk}'^{\alpha-2}}{2}A_{nk}^{2} + \left(1-\frac{\alpha}{2}\right)A_{nk}'^{\alpha},\ \forall\alpha, 0<\alpha<2$$

- B update with normalization: constrained least squares solution using Lagrange multipliers

$$B_{kl} = \left(\tilde{A}^t_{kl}\tilde{A}_{tk'l'} + \Lambda_{kl,k'l'}\right)^{-1}\tilde{A}^{k'l'}_t X^t$$

Diagonal matrix with K different Lagrange multipliers adjusted so that $B^t_{kl}B_{kl} = 1$

Discussion

Related techniques

- Update equation for A contains correlations, i.e. the algorithm does a form of **template matching**.
- For independent amplitudes A the updated equation for B results in **reverse correlation**.
- Extracts templates directly from the data without supervision, i.e. it is a form of **unsupervised clustering**.

Decomposition criteria

- Spike/events can be **overlapping**.
- Templates **do not need to be non-negative** as in NMF.
- Templates **do not need to be orthogonal** as in PCA or ICA. But works best if templates are different ("orthogonal").
- **Positivity and sparseness** of amplitudes is the key criterion.

Properties

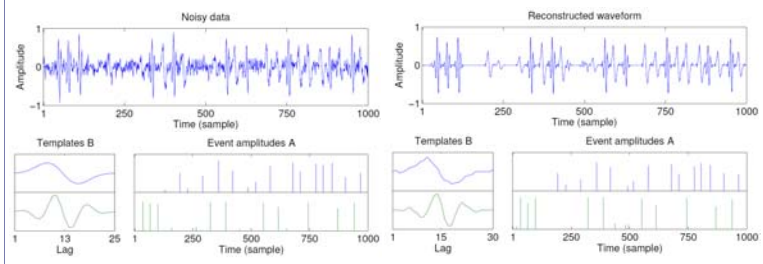
- Algorithm is guaranteed to **converge**.
- Events may occur at **any point in time** (in contrast to PCA, ICA, NMF), i.e. decomposition is shift invariant.
- Algorithm decomposes signals in **spike trains convolved with basis functions** (as Smith & Lewicki).
- Model has the form of a **convolutive mixture**. However, non-negative constraint, sparseness and the length of A and B make the interpretation very different than that of blind source separation.

Quantitative evaluation on synthetic data

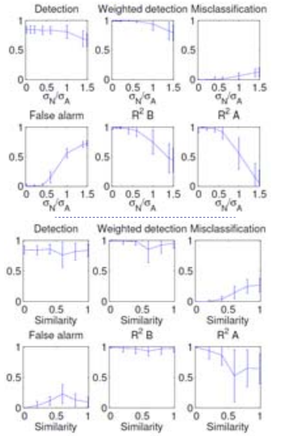
Conditions:

- Synthetic spike trains with additive Gaussian noise
- Two types of "spikes", with a cosine of 74° → similarity 0, while similarity 1=identical templates
- 30 events, amplitude uniformly distributed in [0,1], 100 retries.

Example of result for SNR=2



Performance as a function of **SNR** and **similarity** between templates



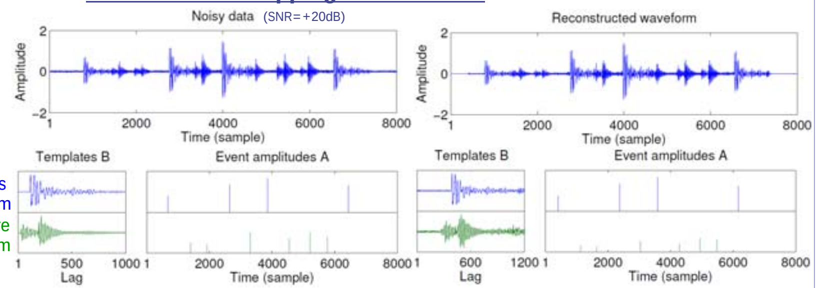
Results on overlapping drum sounds

Conditions:

- Single channel
- Unknown sounds
- Unknown timings
- No training data

Goal: Estimate the sound patterns, their timings & amplitudes

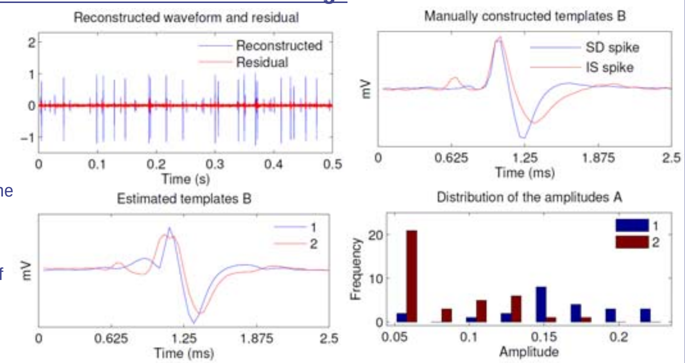
Data: 2 s drum loop, 4 kHz sampling rate



Results on extracellular recordings

Extracellular recordings from primary-like cells within anteroventral cochlear nucleus with a single electrode typically show a succession of events made up of three sub-events:

- a small pre-synaptic spike from the large auditory nerve fiber terminal,
- a medium-sized post-synaptic spike from the initial segment of the axon where it is triggered (the IS spike)
- a large-sized spike produced by back-propagation into the soma and dendrites of the cell (the soma-dendritic or SD spike)



Future work

- Alternative Priors
- Extraction of fundamental patterns from databases of speech, music and environmental sounds
- Enforcing orthogonality between templates
- Application to multichannel data

Bibliography

1) C. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorization for clustering and low-dimension representation," Tech. Rep. LBNL-60428, 2006.
2) M. Mørup, M. N. Schmidt, and L. K. Hansen, "Shift invariant sparse coding of image and music data," *submitted to Journal of Machine Learning Research, 2007.*
3) H. Kameoka, N. Ono, K. Kashino, and S. Sagayama, "Complex NMF: A new sparse representation for acoustic signals," *submitted to ICASSP 2009.*
4) J. Le Roux, "Exploiting regularities in natural acoustic scenes", Ph.D. dissertation, The University of Tokyo and University Paris VI, to be defended in Jan. 2009
5) S. Arkadiusz, M. Sayles, and I. M. Winter, "Spike waveforms in the anteroventral cochlear nucleus revisited," in *ARO midwinter meeting, no. Abstract #678, 2008.*