Non-negative TD regularization with an augmented Lagrangian

Miguel Arjona Ramírez
miguel@lps.usp.br
University of São Paulo, Brazil

1 Motivation

- Nonuniform sampling and interpolation of spectral envelopes for speech signals by temporal decomposition.
- Non-negative matrix factorization (NMF) applies naturally to difference event vectors and functions in TD.
- TD constraints: line spectral frequency (LSF) ordering in event vectors and monotonic event functions.

2 Goals

- Third constraint: Event function complementarity.
- Inclusion of third constraint in NMF starting from penalty condition.

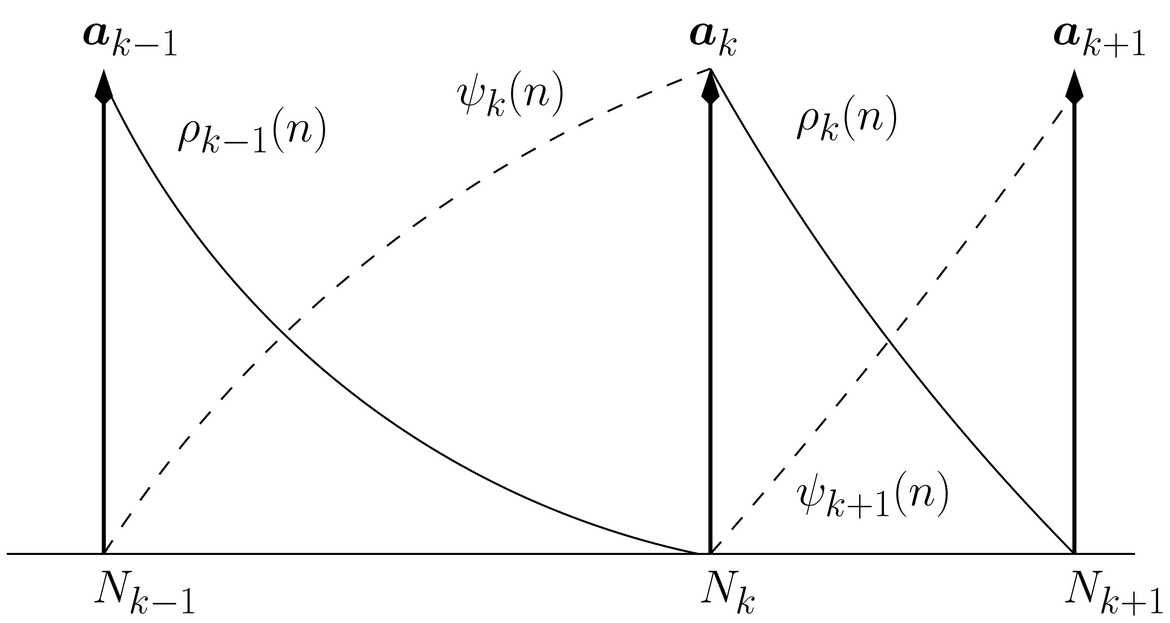
3 Temporal decomposition (TD)

$$\hat{\mathbf{Y}} = \mathbf{A}\mathbf{\Phi} \tag{1}$$

$$\hat{y}_i(n) = a_{i,k-1}\phi_{k-1}(n) + a_{i,k}\phi_k(n) \tag{3}$$

$$\hat{y}_i(n) = a_{i,k-1}\rho_{k-1}(n) + a_{i,k}\psi_k(n) \tag{4}$$

4 Temporal decomposition in adjacent superframes



5 Model fitting stages

1. Event locations N_k and the number of events K are determined. Triangular interpolation spectral measure (TRISM) is used.
2. Half event functions $\rho_{k-1}(n)$ and $\psi_k(n)$ and right event vector $\boldsymbol{a}_k$ are determined minimizing cost function.

6 Cost functions and augmented Lagrangian

$$J_0 = \sum_{n=N_{k-1}}^{N_k} \sum_{i=1}^{p} \left[y_i(n) - a_{i,k-1}\rho_{k-1}(n) - a_{i,k}\psi_k(n)\right]^2 \tag{6}$$

$$J_1 = J_0 + \sum_{n=N_k+1}^{N_{k+1}} \sum_{i=1}^{p} \left[y_i(n) - a_{i,k}\rho_k(n) - a_{i,k+1}\psi_{k+1}(n)\right]^2$$

$$J = J_1 + \frac{\mu}{2} \sum_{n=N_{k-1}}^{N_k} \left[\rho_{k-1}(n) + \psi_k(n) - 1\right]^2 + \sum_{n=N_{k-1}}^{N_k} \lambda_n \left[\rho_{k-1}(n) + \psi_k(n) - 1\right] \tag{17}$$

7 Non-negative difference LSF event vector

$$d_{i,k} = \begin{cases} a_{i,k} & \text{if } i = 1 \\ a_{i,k} - a_{i-1,k} & \text{if } 1 < i \leq p. \end{cases} \tag{12}$$

8 Non-negative difference half event functions

$$\delta_{k-1}(n) = \begin{cases} \rho_{k-1}(n) & \text{if } n = N_k \\ \rho_{k-1}(n) - \rho_{k-1}(n+1) & \text{if } N_{k-1} \leq n < N_k \end{cases}$$

$$\gamma_k(n) = \begin{cases} \psi_k(n) & \text{if } n = N_{k-1} \\ \psi_k(n) - \psi_k(n-1) & \text{if } N_{k-1} < n \leq N_k. \end{cases} \tag{13}$$

9 Update rule for Lagrange multipliers

$$\mathbf{0} = \boldsymbol{\nabla} J_1 + \sum_{n=N_{k-1}}^{N_k} \underbrace{\left[\lambda_n + \mu\left[\rho_{k-1}(n) + \psi_k(n) - 1\right]\right]} \boldsymbol{\nabla}\left[\rho_{k-1}(n) + \psi_k(n) - 1\right] \tag{18}$$

$$\lambda_n \leftarrow \frac{\mu}{\lambda_n + \mu\left[\rho_{k-1}(n) + \psi_k(n)\right]} \lambda_n. \tag{20}$$

10 Update rules for difference event functions and event vector

$$
\begin{aligned}
\delta_{k-1}(n) &\leftarrow \frac{\Sigma_{m=N_{k-1}}^{n}\left[2\,\Sigma_{i=1}^{p} a_{i,k-1}y_i(m)+\mu\right]}{\Sigma_{m=N_{k-1}}^{n}\left\{2\,\Sigma_{i=1}^{p}\left[a_{i,k-1}a_{i,k}\psi_k(m)+a_{i,k-1}^2\rho_{k-1}(m)\right]+\mu\left[\rho_{k-1}(m)+\psi_k(m)\right]+\lambda_m\right\}}\delta_{k-1}(n)\\
\gamma_k(n) &\leftarrow \frac{\Sigma_{m=n}^{N_k}\left[2\,\Sigma_{i=1}^{p} a_{i,k}y_i(m)+\mu\right]}{\Sigma_{m=n}^{N_k}\left\{2\,\Sigma_{i=1}^{p}\left[a_{i,k-1}a_{i,k}\rho_{k-1}(m)+a_{i,k}^2\psi_k(m)\right]+\mu\left[\rho_{k-1}(m)+\psi_k(m)\right]+\lambda_m\right\}}\gamma_k(n) \qquad (21)\\
d_{i,k} &\leftarrow \frac{2\left\{\Sigma_{n=N_{k-1}}^{N_k}\Sigma_{j=i}^{p}\psi_k(n)y_j(n)+\Sigma_{n=N_k+1}^{N_{k+1}}\Sigma_{j=i}^{p}\rho_k(n)y_j(n)\right\}}{2\left\{\Sigma_{n=N_{k-1}}^{N_k}\Sigma_{j=i}^{p}\left[a_{j,k-1}\rho_{k-1}(n)\psi_k(n)+a_{j,k}\psi_k^2(n)\right]+\Sigma_{n=N_k+1}^{N_{k+1}}\Sigma_{j=i}^{p}\left[a_{j,k+1}\rho_k(n)\psi_{k+1}(n)+a_{j,k}\rho_k^2(n)\right]\right\}}d_{i,k}
\end{aligned}
$$

(15)

11 Experimental setup

- 210 speech signals from TIMIT test partition
- Spectrogram: TANDEM-STRAIGHT at 200 frame/s
- LSF vector generation: Wiener-Khinchin theorem and Linear Predictive analysis of order $p = 16$
- Event rate: 12.32 ev/s

12 Performance of non-negative TD (NTD) with several regularizations.

Method	SD (dB)	μ_c ($\times 10^{-3}$)	$c_{\min}$ ($\times 10^{-3}$)	$c_{\max}$ ($\times 10^{-3}$)	e_m (%)	e_o (%)
$(\boldsymbol{\rho}, \boldsymbol{\psi}, \boldsymbol{a})$	1.15	0.225	-170.8	185.1	0.27	0.012
$(\boldsymbol{\rho}, \boldsymbol{\psi}, \boldsymbol{d})$	1.28	0.219	-173.3	184.1	0.26	0
$(\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{a})$	1.22	0.222	-118.8	107.5	0	0.012
$(\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{d})$	1.35	0.231	-120.7	107.8	0	0
$(\boldsymbol{\rho}, \boldsymbol{\psi}, \boldsymbol{a}, \boldsymbol{\lambda})$	1.34	-2.77	-6.69	0	0	0.012
$(\boldsymbol{\rho}, \boldsymbol{\psi}, \boldsymbol{d}, \boldsymbol{\lambda})$	1.49	-2.76	-6.69	0	0	0
$(\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{d}, \boldsymbol{\lambda})$	1.49	-0.502	-6.42	4.22	0	0

- SD: mean log spectral distortion
- ordering error at k if, for any i, $a_{i,k} - a_{i-1,k} \leq 0$
- e_o : percentage of event vectors with ordering errors
- monotonicity error at k if $\boldsymbol{\rho}_{k-1}$ or $\boldsymbol{\psi}_k$ nonmonotonic
- e_m : percentage of frames with monotonicity error
- complementarity deviation: $c(n) = \rho(n) + \psi(n) - 1$
- μ_c : mean value of complementarity deviation
- $c_{\min}$: minimum complementarity deviation
- $c_{\max}$: maximum complementarity deviation

13 Conclusion

Baseline
- Difference LSF event vectors provide stable synthesis filters.
- Non-negative difference event functions make event functions monotonic.
- Regularization of event function complementarity includes a quadratic penalty term as previous NMF TD.

Proposed
- Smoother TANDEM-STRAIGHT spectral envelopes improve NMF for TD of speech.
- Adaptive regularization of event function complementarity by means of a linear term;
  - linear term adjustable by Laplace multipliers (LMs);
  - multiplicative rule updates LMs.

Result
- This approach has eliminated ordering and monotonicity errors and greatly reduced the amplitude of complementarity errors.

14 Thanks