MAP-BASED ESTIMATION OF THE PARAMETERS OF A GAUSSIAN MIXTURE MODEL IN THE PRESENCE OF NOISY OBSERVATIONS

UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Computer Science, Electrical Engineering and Mathematics
Communications Engineering
Prof. Dr.-Ing. Reinhold Häb-Umbach

Aleksej Chinaev and Reinhold Haeb-Umbach, University of Paderborn, 33098 Paderborn, Germany, {chinaev, haeb}@nt.upb.de

Introduction

- Maximum A-Posteriori (MAP) estimate of the parameters of an univariate Gaussian Mixture Model (GMM)
- GMM process superposed by an additive noise with known statistics
- Derivation of an approximate conjugate prior of the GMM parameters

MAP-based GMM Parameter Estimation

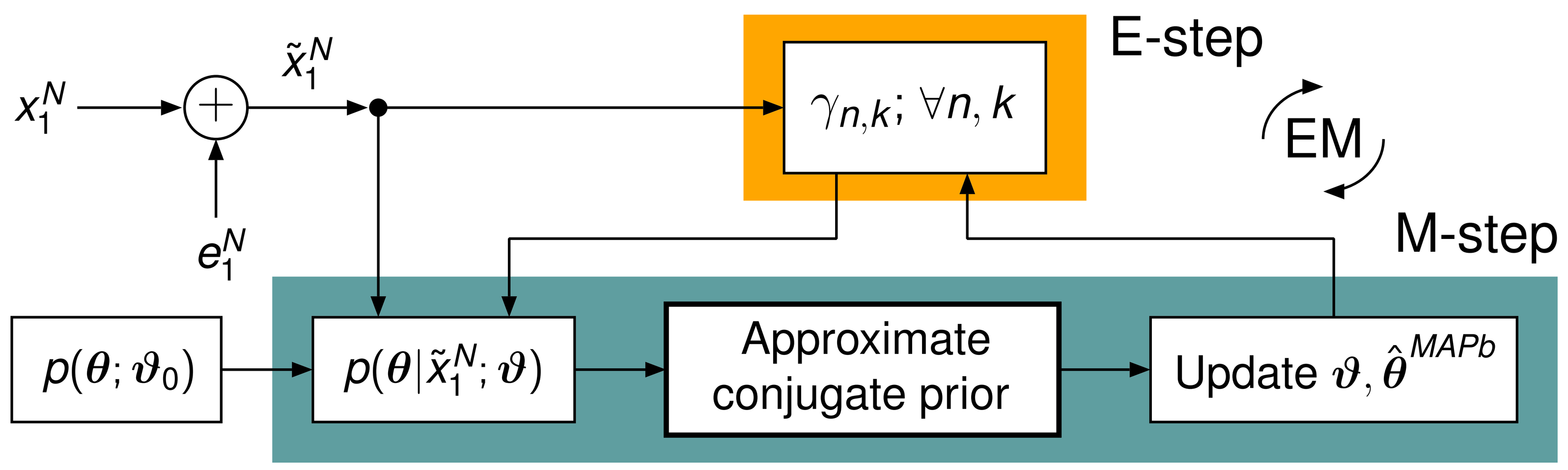
- Given noisy observation data block $\tilde{x}_1^N = \{\tilde{x}_n\}_{n\in[1,N]}$ with $\tilde{x}_n = x_n + e_n$
  - $x_n \sim \sum_{k=1}^K \omega_k \cdot \mathcal{N}(\mu_k, \sigma_k^2)$ with model order K and $e_n \sim \mathcal{N}(\mu_E, \sigma_E^2)$
- MAP-based estimate of a GMM parameter vector $\boldsymbol{\theta} = \{\omega_k, \mu_k, \sigma_k^2; \forall k\}$ by using the EM approach with the auxiliary function $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$

$$\hat{\boldsymbol{\theta}}^{\mathrm{MAPb}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ p(\boldsymbol{\theta}|\tilde{x}_1^N; \boldsymbol{\vartheta}) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ e^{Q(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})} \cdot p(\boldsymbol{\theta}; \boldsymbol{\vartheta}_0) \quad (1)$$

- Error free observations
  - A conjugate prior $p(\boldsymbol{\theta}; \boldsymbol{\vartheta}_0) = p(\boldsymbol{\omega}) \cdot \prod_{k=1}^K p(\mu_k|\sigma_k^2) \cdot p(\sigma_k^2)$ with a hyperparameter vector $\boldsymbol{\vartheta}_0 = \{\xi_{k0}, \mathrm{m}_{k0}, \kappa_{k0}, \nu_{k0}, \lambda_{k0}^2; \forall k\}$

$$p(\boldsymbol{\omega}) = \mathrm{Dir}(\boldsymbol{\omega}; \xi_{k0}) \qquad p(\mu_k|\sigma_k^2) = \mathcal{N}(\mu_k; \mathrm{m}_{k0}, \sigma_k^2/\kappa_{k0}) \qquad p(\sigma_k^2) = \mathrm{Si}\text{-}\chi^2(\sigma_k^2; \nu_{k0}, \lambda_{k0}^2)$$

- Noisy observations
  - Only $p(\boldsymbol{\omega})$ holds as a conjugate prior with an update $\xi_k = \xi_{k0} + \sum_{n=1}^N \gamma_{n,k}$
    - $\gamma_{n,k}$ a responsibility that component k takes for 'explaining' the observation $\tilde{x}_n$
  - Improper dependence of the update equations $\mathrm{m}_k(\sigma_k^2)$ and $\kappa_k(\sigma_k^2)$
  - $\mathrm{Si}\text{-}\chi^2(\sigma_k^2; \nu_{k0}, \lambda_{k0}^2)$ is no longer a conjugate prior
- EM framework with an approximate conjugate prior



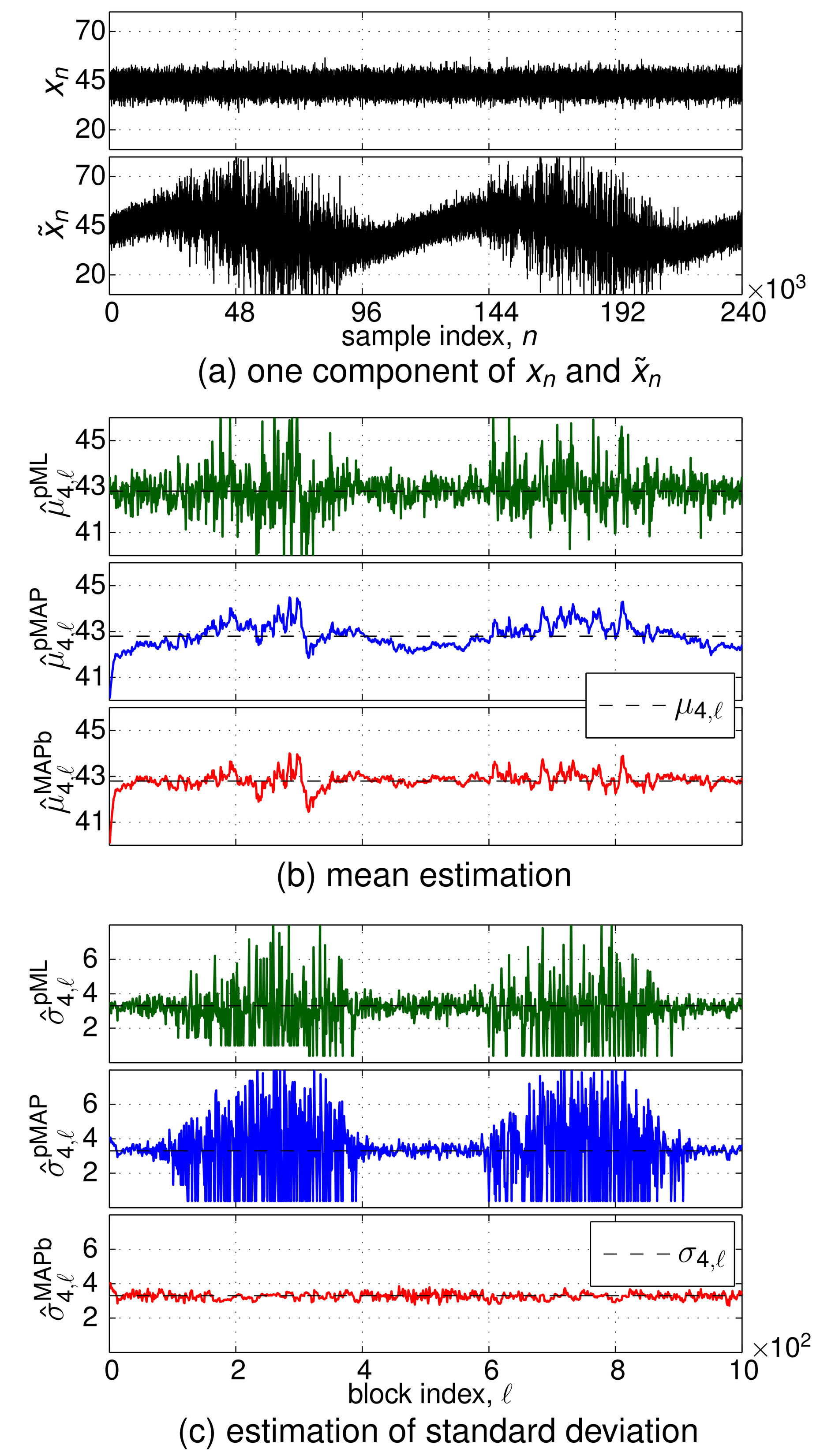


Fig. 1: Samples (a); Estimated trajectories (b) and (c) of 4th GMM component ($K = 8$, $C = 16$).

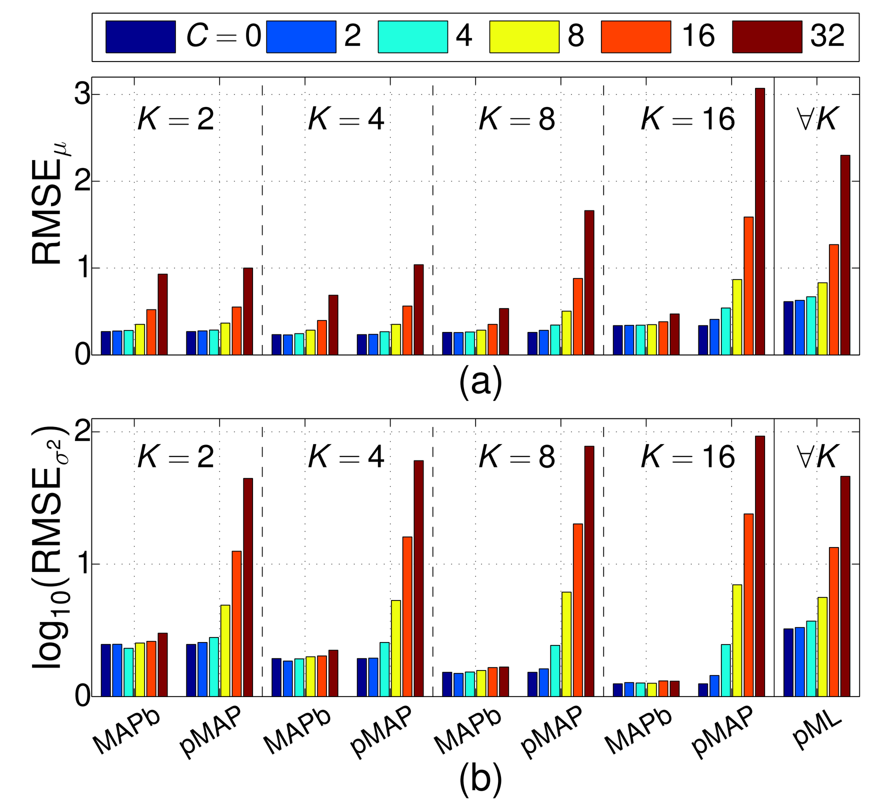


Fig. 2: Root-Mean-Squared Error (RMSE) values averaged over 100 experiments: (a) RMSE_μ and (b) $\log_{10}(\mathrm{RMSE}_{\sigma^2})$.

Approximations to Obtain Conjugate Prior

(1) Replace σ_k^2 in $\mathrm{m}_k(\sigma_k^2)$ and $\kappa_k(\sigma_k^2)$ by $\hat{\sigma}_k^2$

(2) Approximate remaining terms in $p(\boldsymbol{\theta}|\tilde{x}_1^N; \boldsymbol{\vartheta})$, which depend on σ_k^2, by $\mathrm{Si}\text{-}\chi^2$ distribution with the same mode $\hat{\sigma}_k^2$

(3) Determine $\hat{\sigma}_k^2$ by the interval bisection and the Newton method

Simulation Results

- $N = 30 \cdot K$ observations per data block for $L = 1000$ blocks
- Nonstationary noise $e_n \sim \mathcal{N}(\mu_{E,\ell}, \sigma_{E,\ell}^2)$ with $\mu_{E,\ell} = C \cdot \sin(4\pi\ell/L)/2$ and $\sigma_{E,\ell} \sim U\left(0, C \cdot \sin^2(2\pi\ell/L)\right)$, Fig. 1(a)
- Reference approaches: plain ML (pML) and plain MAP (pMAP)

with $\Omega \in \{\text{'ML'}, \text{'MAP'}\}$

$$\hat{\mu}_{k,\ell}^{\mathrm{p}\Omega} = \hat{\mu}_{k,\ell}^{\Omega} - \mu_{E,\ell} \qquad \hat{\sigma}_{k,\ell}^{\mathrm{p}\Omega} = \begin{cases} \sqrt{(\hat{\sigma}_{k,\ell}^{\Omega})^2 - \sigma_{E,\ell}^2} & \text{for } \hat{\sigma}_{k,\ell}^{\Omega} > \sigma_{E,\ell} \\ \min(\hat{\sigma}_{k,1:\ell}^{\mathrm{p}\Omega}) & \text{otherwise} \end{cases}$$

- Sequential estimation setup for both MAP approaches
  - Posterior estimated on the previous block as a-priori for the next block
- Proposed MAPb method delivers the best trajectories, Fig.1(b)-(c)
  - Estimates $\hat{\mu}_{4,\ell}^{\mathrm{pML}}$, $\hat{\sigma}_{4,\ell}^{\mathrm{pML}}$ and $\hat{\sigma}_{4,\ell}^{\mathrm{pMAP}}$ result in a too large variance
- For all tested conditions MAPb performs better than pML, Fig.2
  - Superiority of the MAPb over pML and pMAP becomes larger with growing GMM order K and for large noise power

Conclusions

- Proposed method delivers GMM parameter estimates with significantly lower error variance compared to plain ML and MAP methods
- Superiority of MAPb approach holds also in tracking time-variant $\boldsymbol{\theta}_\ell$
- An efficient estimation of GMM parameters in a sequential framework

International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2013, Vancouver, Canada

Work supported by DFG under contract no. HA 3455/8-1