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

Aleksej Chinaev, Reinhold Haeb-Umbach

Department of Communications Engineering, University of Paderborn, 33098, Paderborn, Germany
e-mail: {chinaev,haeb}@nt.upb.de, web: nt.upb.de

ABSTRACT

In this contribution we derive the Maximum A-Posteriori (MAP) estimates of the parameters of a Gaussian Mixture Model (GMM) in the presence of noisy observations. We assume the distortion to be white Gaussian noise of known mean and variance. An approximate conjugate prior of the GMM parameters is derived allowing for a computationally efficient implementation in a sequential estimation framework. Simulations on artificially generated data demonstrate the superiority of the proposed method compared to the Maximum Likelihood technique and to the ordinary MAP approach, whose estimates are corrected by the known statistics of the distortion in a straightforward manner.

***Index Terms*—** Gaussian mixture model, Maximum likelihood estimation, Maximum a posteriori estimation

1. INTRODUCTION

Gaussian Mixture Models, i.e., weighted linear combinations of Gaussians whose weights sum up to unity, are often a much more realistic model of real-world data than a single Gaussian. They have found widespread use in many different fields such as image processing (clustering-based image segmentation) and biometric systems (speaker-independent automatic speech recognition).

The Maximum Likelihood (ML) estimates of the parameters of the GMM, i.e., of weights, means and variances of the component densities, can be derived by the Expectation Maximization (EM) algorithm [1, 2]. This has been extended to the MAP estimation by employing conjugate a-priori distributions, see, e.g., [3] for a textbook treatment. The MAP estimation of GMM parameters has also important applications, such as the adaptation of an acoustic model of a speech recognizer to the statistics of a target speaker [4].

In many applications, however, the GMM process is not directly observable, but is superposed by additive noise. The problem occurs, for example, in speaker recognition or MAP-based speaker adaptation if the input data are corrupted by noise. In these applications noise estimation algorithms, such as the minimum statistics method [5], are able to provide estimates of the statistics of the corrupting noise process. But even if the noise statistics are known, it is unclear how to obtain the MAP estimates of the GMM parameters.

In an earlier work we have shown how MAP estimates of a single Gaussian can be obtained from noisy observations: the posterior was approximated by a probability density function (PDF) from the same family as the prior to allow for an efficient realization in a sequential estimation framework [6]. The mode of the approximate posterior was chosen to match the mode of the exact posterior.

In this work we adopt this idea and extend it to develop a computationally efficient MAP estimation of the GMM parameters in the presence of noisy observations. We assume a block sequential setup, where the GMM parameters and the noise statistics are taken to be constant for the duration of N observations. From block to block both the target process and the noise statistics are allowed to change. Thus the derived method is able to track the parameters of a non-stationary GMM process in the presence of time-variant additive noise. The simulations show that improved mean and variance estimates are obtained compared to standard methods.

The paper is organized as follows. In Section 2 the MAP estimator is derived. Due to lack of space we concentrate on those parts which are different from the ordinary MAP estimation in the absence of noise on the observations. In Section 3 we present simulation results followed by conclusions drawn in Section 4.

2. MAP-BASED GMM PARAMETER ESTIMATION FROM NOISY OBSERVATIONS

We consider a real-valued stationary random process $\{X_n\}$, whose realization x_n of the n-th random variable X_n is drawn from an univariate GMM $x_n \sim \sum_{k=1}^{K} \omega_k \cdot \mathcal{N}(\mu_k, \sigma_k^2)$ with given model order K for $n = 1, \dots, N$. Our goal is to estimate the weights ω_k, means μ_k and variances σ_k^2 of the GMM from noisy observations $\tilde{x}_1^N = \{\tilde{x}_1, \dots, \tilde{x}_n, \dots, \tilde{x}_N\}$, where $\tilde{x}_n = x_n + e_n$, given an a-priori model for the involved parameters. The distortion is the realization of a white Gaussian process with the known mean and variance, i.e., $e_n \sim \mathcal{N}(\mu_E, \sigma_E^2)$. Denoting the parameters to be estimated by

The work was in part supported by Deutsche Forschungsgemeinschaft under contract no. Ha3455/8-1.

$\boldsymbol{\theta}=\{\omega_k, \mu_k, \sigma_k^2; k=1,\ldots,K\}$, the MAP estimate is given by

$$\hat{\boldsymbol{\theta}}_{\text{MAP}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ p(\boldsymbol{\theta}|\tilde{x}_1^N) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ p(\tilde{x}_1^N|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}). \quad (1)$$

The MAP estimate is determined with the help of the EM algorithm using the auxiliary function

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = E\left[\log p(\tilde{z}_1^N|\boldsymbol{\theta})|\tilde{x}_1^N, \hat{\boldsymbol{\theta}}\right], \quad (2)$$

where $\hat{\boldsymbol{\theta}}$ are the parameter estimates of the last iteration. Here, $\tilde{z}_1^N = [\tilde{x}_1^N, l_1^N]$ denote the complete data comprising the noisy observations $\tilde{x}_1^N$ and the unknown labels $l_1^N = l_1, \ldots, l_N$ with $l_n \in \{1, \ldots, K\}$. After convergence of the EM algorithm the MAP estimate is obtained via the maximization of

$$L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) := e^{Q(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})} \cdot p(\boldsymbol{\theta}). \quad (3)$$

For the GMM viewed here the auxiliary function is given by

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \quad (4)$$
$$\sum_{k=1}^{K}\sum_{n=1}^{N} P(l_n = k|\tilde{x}_n, \hat{\boldsymbol{\theta}}) \cdot \ln\left[P(l_n = k|\boldsymbol{\theta}) \cdot p(\tilde{x}_n|l_n = k, \boldsymbol{\theta})\right],$$

which gives, after some manipulations [7],

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \sum_{k=1}^{K}\sum_{n=1}^{N} \frac{\alpha_{n,k}}{\alpha_n} \cdot \ln\left[\omega_k \cdot p(\tilde{x}_n|l_n = k, \boldsymbol{\theta})\right] \quad (5)$$

with $\alpha_{n,k} = \hat{\omega}_k \cdot p(\tilde{x}_n|l_n = k, \hat{\boldsymbol{\theta}})$ and $\alpha_n = \sum_{k=1}^{K}\alpha_{n,k}$, where $\hat{\omega}_k = P(l_n = k|\hat{\boldsymbol{\theta}})$ and $\omega_k = P(l_n = k|\boldsymbol{\theta})$.

Using $p(\tilde{x}_n|l_n = k, \boldsymbol{\theta}) = \mathcal{N}(\tilde{x}_n; \mu_k + \mu_E, \sigma_k^2 + \sigma_E^2)$ we arrive at

$$e^{Q(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})} \propto \prod_{k=1}^{K} \frac{\omega_k^{\gamma_k}}{(\sigma_k^2 + \sigma_E^2)^{\frac{\gamma_k}{2}}} \cdot e^{-\frac{\gamma_k(\bar{x}_k - (\mu_k + \mu_E))^2 + S_k}{2(\sigma_k^2 + \sigma_E^2)}}, \quad (6)$$

with $\gamma_{n,k} = \frac{\alpha_{n,k}}{\alpha_n}$, $\gamma_k = \sum_{n=1}^{N}\gamma_{n,k}$, $\bar{x}_k = \frac{1}{\gamma_k}\sum_{n=1}^{N}\gamma_{n,k} \cdot \tilde{x}_n$ and $S_k = \sum_{n=1}^{N}\gamma_{n,k} \cdot (\tilde{x}_n - \bar{x}_k)^2$.

To enable an efficient estimation procedure, the a-priori PDF $p(\boldsymbol{\theta}) = P(\boldsymbol{\omega}) \cdot \prod_{k=1}^{K} p(\mu_k|\sigma_k^2) \cdot p(\sigma_k^2)$ is chosen to be a conjugate prior to the likelihood-function for error free observations $\tilde{x}_n = x_n$. The weights ω_k are modeled by a Dirichlet distribution $P(\boldsymbol{\omega}) = \prod_{k=1}^{K}\omega_k^{\xi_{k,0}-1}$, the means μ_k by a Gaussian distribution $p(\mu_k|\sigma_k^2) \propto \frac{1}{\sigma_k} \cdot \exp\left(-\frac{(\mu_k - m_{k,0})^2}{2\sigma_k^2/\kappa_{k,0}}\right)$ and the variances σ_k^2 by a scaled inverse chi-squared (SICS) distribution $p(\sigma_k^2) \propto \left(\frac{1}{\sigma_k}\right)^{\nu_{k,0}+2} \cdot \exp\left(-\frac{\nu_{k,0}\lambda_{k,0}^2}{2\sigma_k^2}\right)$ with the following hyperparameters: concentration parameters $\xi_{k,0}$, means $m_{k,0}$, compactness degrees $\kappa_{k,0}$, degrees of freedom $\nu_{k,0}$ and scale factors $\lambda_{k,0}^2$. Hence, the a-priori PDF is given by:

$$p(\boldsymbol{\theta}) \propto \prod_{k=1}^{K} \frac{\omega_k^{\xi_{k,0}-1}}{\sigma_k^{\nu_{k,0}+3}} \cdot e^{-\frac{\kappa_{k,0}(\mu_k - m_{k,0})^2 + \nu_{k,0}\lambda_{k,0}^2}{2\sigma_k^2}}. \quad (7)$$

However, in the case of noisy observations, the term $L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ in (3) no longer has the same algebraic form as the prior $p(\boldsymbol{\theta})$ in (7). While an update formula for the concentration parameters can be found according to

$$\xi_k = \xi_{k,0} + \gamma_k, \quad (8)$$

trying to find an update formula for the hyperparameters of the Gaussian PDF for μ_k leads to

$$\kappa_k(\sigma_k^2) = \kappa_{k,0} + \frac{\sigma_k^2}{\sigma_k^2 + \sigma_E^2} \cdot \gamma_k, \quad (9)$$

$$m_k(\sigma_k^2) = m_{k,0} + \frac{\gamma_k \cdot \sigma_k^2 \cdot (\bar{x}_k - \mu_E - m_{k,0})}{\kappa_{k,0} \cdot (\sigma_k^2 + \sigma_E^2) + \gamma_k \cdot \sigma_k^2}, \quad (10)$$

which depend on the unknown variance σ_k^2. Similar to [6] we suggest to eliminate this dependence by replacing σ_k^2 by its estimate $\hat{\sigma}_k^2$, obtained from the last EM iteration:

$$\kappa_k := \kappa_k(\hat{\sigma}_k^2); \quad m_k := m_k(\hat{\sigma}_k^2). \quad (11)$$

Estimates of ω_k and μ_k are calculated from the updated hyperparameters of (8) and (10) by

$$\hat{\omega}_k = \frac{1 - \xi_k}{K - \sum_{k=1}^{K}\xi_k} \quad \text{and} \quad \hat{\mu}_k = m_k. \quad (12)$$

Using the approximation (11) and concentrating on the terms in $L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ that depend on the variance σ_k^2 leaves

$$f(\sigma_k^2) = \frac{e^{-\frac{\kappa_{k,0}\gamma_k(\bar{x}_k - \mu_E - m_{k,0})^2}{2(\sigma_k^2(\kappa_{k,0}+\gamma_k) + \sigma_E^2\kappa_{k,0})} - \frac{S_k}{2(\sigma_k^2 + \sigma_E^2)} - \frac{\nu_{k,0}\lambda_{k,0}^2}{2\sigma_k^2}}}{(\sigma_k^2 + \sigma_E^2)^{\frac{\gamma_k}{2}} \cdot \sigma_k^{\nu_{k,0}+3}}, \quad (13)$$

which no longer has the algebraic form of a SICS distribution.

In order to obtain a computationally efficient MAP estimation we need to establish a conjugate prior. This is achieved by approximating $f(\sigma_k^2)$ by a SICS distribution having the maximum at the same position as $f(\sigma_k^2)$.

In the following we show how the maximum of $f(\sigma_k^2)$ can be found efficiently. First we note that there is only a single local maximum in $\mathbb{R}^+$. To see this, consider the function:

$$g(\psi) = -2 \cdot \ln(f(\psi)) = \quad (14)$$
$$\frac{K_2}{K_3\psi + K_4} + \frac{K_6}{\psi + K_0} + \frac{K_1}{\psi} + K_5 \cdot \ln\psi + K_7 \cdot \ln(\psi + K_0)$$

with constants $K_1 = \nu_{k,0} \cdot \lambda_{k,0}^2$, $K_2 = \kappa_{k,0} \cdot (\bar{x}_k - m_{k,0} - \mu_E)^2$, $K_3 = 1 + \kappa_{k,0}/\gamma_k$, $K_4 = \kappa_{k,0} \cdot \sigma_E^2/\gamma_k$, $K_5 = \nu_{k,0} + 3$, $K_6 = S_k$, $K_7 = \gamma_k$, $K_0 = \sigma_E^2$ and $\psi = \sigma_k^2 > 0$. Since $g(\psi)$ is continuous and a sum of strictly monotonically increasing and strictly monotonically decreasing functions with $\lim_{\psi\to 0} g(\psi) = \lim_{\psi\to\infty} g(\psi) = \infty$, $g(\psi)$ has exactly one minimum ψ_k, which is the maximum of $f(\psi)$.

At the same time ψ_k is the single positive root of the derivative $f'(\psi) = h(\psi)/(\psi^2(K_3\psi + K_4)^2(\psi + K_0)^2)$ with $h(\psi) = (K_3\psi + K_4)^2(\psi + K_0) \cdot h_a(\psi) + \psi^2 \cdot h_b(\psi)$, where

$$h_a(\psi) = (K_5 + K_7)\psi^2 + (K_0K_5 - K_1)\psi - K_0K_1,$$
$$h_b(\psi) = -K_6(K_3\psi + K_4)^2 - K_2K_3(\psi + K_0)^2.$$

Since $h(\psi)$ is a fifth order polynomial, the analytical determination of ψ_k is very complicated. We suggest to find it numerically by a combination of the interval bisection and the Newton method. To determine a start interval $[b_L, b_U]$ for the bisection approach with $b_L, b_U > 0$, $h(b_L) < 0$ and $h(b_U) > 0$, the functions $h_a(\psi)$ and $h_b(\psi)$ are analyzed. It can be verified that $h_a(b_L) < 0$ for

$$b_L = \frac{K_1 - K_0 K_5 + \sqrt{(K_0 K_5 - K_1)^2 + 4 K_0 K_1 (K_5 + K_7)}}{2(K_5 + K_7)}, \tag{15}$$

which can be used as the lower limit b_L of the start interval, since $h_b(\psi) < 0$ for $\forall\psi$. As an upper limit b_U we suggest to use the global sample variance. After executing a number of bisection loops, a rough estimate of ψ_k is calculated, which is used as a starting value of the Newton approach. The resulting estimate of the Newton method $\hat{\psi}_k$ is taken as the mode of posterior PDF of the variance: $\hat{\sigma}_k^2 = \hat{\psi}_k$.

In a sequential estimation setup, where the posterior estimated on the preceding data block l is taken as the prior for the next block $l+1$, the prior PDF of the variance is chosen to be a SICS with mode $\hat{\sigma}_{k,l}^2$ found by the procedure described before. The scale factors of the established conjugate prior for the variances can then be calculated by $\lambda_{k,l+1}^2 = \frac{\nu_{k,0}+2}{\nu_{k,0}} \cdot \hat{\sigma}_{k,l}^2$. The mean hyperparameters are set to $m_{k,l+1} = \hat{\mu}_{k,l}$.

If the target process is stationary, all hyperparameters should be updated after each block of data, resulting in more reliable estimates as more data come in. If, however, the target process x_n is non-stationary, an update of all hyperparameters would result in an estimator, which is unable to track the time-variant statistics, due to its increasingly narrow bandwidth. The importance of the prior relative to the current data block, and thus the bandwidth of estimator, is controlled by the values of the hyperparameters $\xi_{k,l}$, $\kappa_{k,l}$ and $\nu_{k,l}$. Thus, they should be kept to constant values chosen according to the degree of time-variance of the GMM parameters.

3. SIMULATION RESULTS

To verify the proposed estimation method we generated data samples $x_1^M = x_1, \ldots, x_M$ by drawing them from a GMM of given model order K with constant parameters ω_k, μ_k and σ_k, randomly drawn for each simulation under the constraint of multi-modality of the GMM. To generate the noisy observations $\tilde{x}_n$, we divided the sample sequence x_1^M in $L = 1000$ blocks of length $N = M/L$. In the l-th block for $l = 1, \ldots, L$, realizations e_n of an additive noise were generated by draws from a Gaussian PDF with time-variant mean $\mu_{E,l} = C \cdot \sin(4\pi l/L)/2$ and standard deviation $\sigma_{E,l}$ randomly drawn from a uniform distribution on $[0; C \cdot \sin^2(2\pi l/L)]$. Thus, the statistics of the distortion is constant within each block and changes from block to block. The constant C controls the maximum values of $\mu_{E,l}$ and $\sigma_{E,l}$ and thus the signal-to-noise ratio. In order to have a sufficient number of observations for a pure ML estimation of the parameters in each block, which we consider for comparison purposes, the number of observations per block was set to $N = 3K \cdot 10$. Note that the total number of observations M thus depends on the GMM order K.

The proposed MAP-based approach, denoted by MAPb, was then applied to obtain estimates $\hat{\omega}_{k,l}^{\text{MAPb}}$, $\hat{\mu}_{k,l}^{\text{MAPb}}$ and $\hat{\sigma}_{k,l}^{\text{MAPb}}$ for each block l. Despite of the stationary target process we kept the hyperparameters $\xi_{k,l} = \xi_{k,0} = 5$, $\kappa_{k,l} = \kappa_{k,0} = 5N$ and $\nu_{k,l} = \nu_{k,0} = N$ constant for all data blocks. The start values of the hyperparameters $m_{k,0}$ and $\lambda_{k,0}^2$ were set such that the corresponding modes of $p(\boldsymbol{\theta})$ are located close to the true values μ_k and σ_k^2 respectively. It should be noted that the assumption of a good a-priori knowledge was used for all tested approaches. While the hyperparameter $m_{k,l}$ and $\lambda_{k,l}^2$ are updated from block to block, as mentioned before, they remain unchanged within iterations of the EM-algorithm on the same data block. The maximum number of EM iterations was set to 5 for all tested estimators. The root $\hat{\psi}_k$ was calculated for each EM iteration by executing 10 bisections and a maximum of 10 iterations of the Newton method.

Furthermore, the ML method and the ordinary MAP approach from [4] were applied on the same data $\tilde{x}_1^M$ (both without consideration of the presence of the distortion) with resulting estimates $\hat{\omega}_{k,l}^{\text{ML}}$, $\hat{\mu}_{k,l}^{\text{ML}}$, $\hat{\sigma}_{k,l}^{\text{ML}}$ and $\hat{\omega}_{k,l}^{\text{MAP}}$, $\hat{\mu}_{k,l}^{\text{MAP}}$, $\hat{\sigma}_{k,l}^{\text{MAP}}$ respectively. Since knowledge of the statistics $\mu_{E,l}$ and $\sigma_{E,l}$ is assumed, these estimates can be corrected in a straightforward manner to obtain estimates, which are denoted as plain ML (pML) and plain MAP (pMAP) respectively:

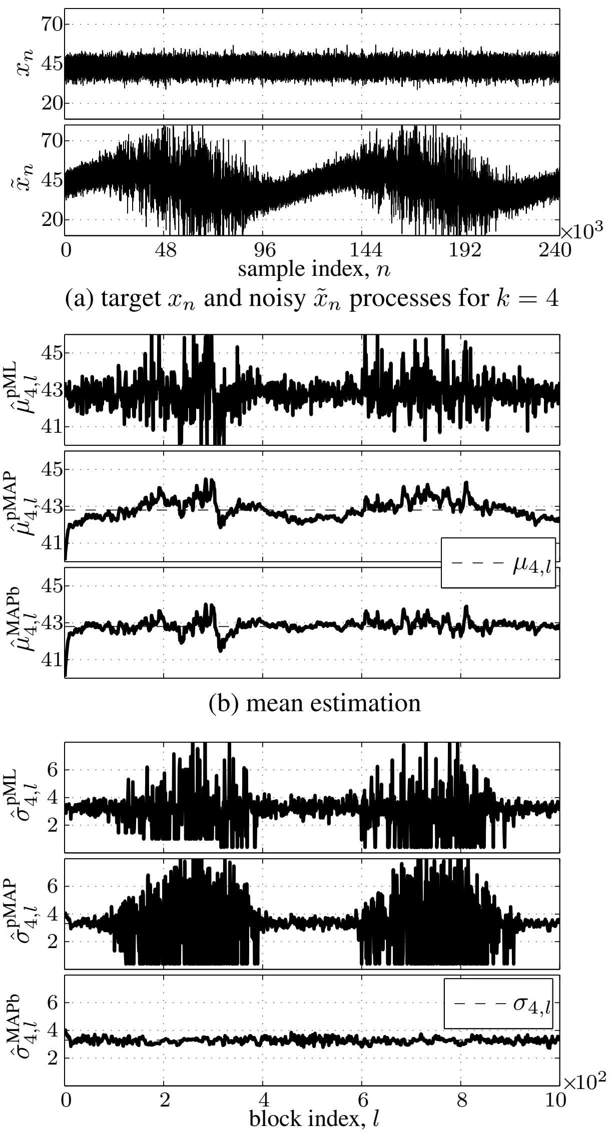
$$\hat{\omega}_{k,l}^{\text{p}\Omega} = \hat{\omega}_{k,l}^{\Omega}, \qquad \hat{\mu}_{k,l}^{\text{p}\Omega} = \hat{\mu}_{k,l}^{\Omega} - \mu_{E,l}, \tag{16}$$

$$\hat{\sigma}_{k,l}^{\text{p}\Omega} = \begin{cases} \sqrt{(\hat{\sigma}_{k,l}^{\Omega})^2 - \sigma_{E,l}^2} & \text{for } \hat{\sigma}_{k,l}^{\Omega} > \sigma_{E,l} \\ \min(\hat{\sigma}_{k,l'}^{\text{p}\Omega}) & \text{otherwise} \end{cases} \tag{17}$$

for $l' = 1, \ldots l-1$ and $\Omega = \{\text{'ML'}, \text{'MAP'}\}$. The pMAP approach used the same set of hyperparameters as the proposed MAPb method.

Examples of the performance of the pML, pMAP and MAPb methods are illustrated in Fig. 1. The trajectory of the 4-th component of the target process, $x_{n,4} \sim \omega_4 \cdot \mathcal{N}(\mu_4, \sigma_4^2)$, where $\omega_4 = 0.23$, $\mu_4 = 42.8$ and $\sigma_4 = 3.3$, and of the associated noisy process $\tilde{x}_{n,4} = x_{n,4} + e_n$ are depicted in Fig. 1(a). The estimates $\hat{\mu}_{4,l}^{\text{pML}}$ and $\hat{\sigma}_{4,l}^{\text{pML}}$ given in Fig. 1(b) and in Fig. 1(c) result in a too large variance. Furthermore, the larger the noise variance $\sigma_{E,l}^2$ is, the stronger the ML estimates scatter. This behavior can also be seen in the pMAP estimates. While $\hat{\mu}_{4,l}^{\text{pMAP}}$ varies only slightly with $\sigma_{E,l}^2$, the estimates $\hat{\sigma}_{4,l}^{\text{pMAP}}$ exhibit large variations. The proposed MAPb method delivers the best trajectories.

To quantitatively evaluate the performance of the estimators we calculated the root-mean-squared error (RMSE) of the weight, mean and variance estimates denoted by RMSE_ω, RMSE_μ and RMSE_{σ^2} respectively, averaged over all K component densities. The RMSE_μ and RMSE_{σ^2} values of the



(a) target x_n and noisy $\tilde{x}_n$ processes for $k = 4$

(b) mean estimation

(c) estimation of standard deviation

Fig. 1. Sample trajectories of pML, pMAP and MAPb methods for one GMM component (k=4, K=8, C=16).

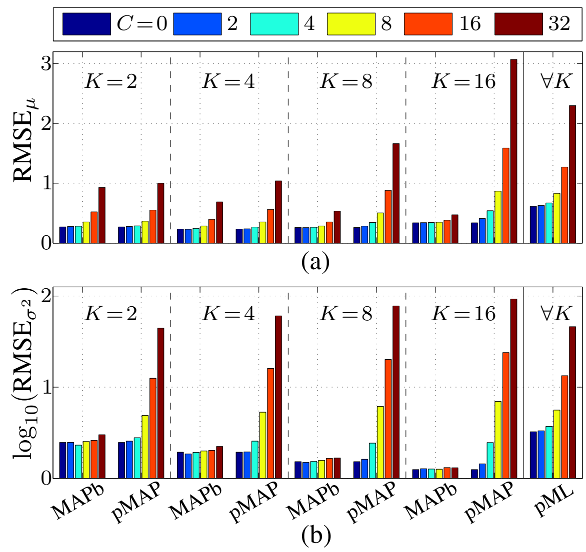


Fig. 2. (a) RMSE$_\mu$ and (b) $\log_{10}(\text{RMSE}_{\sigma^2})$ of the proposed (MAPb), plain MAP (pMAP) and plain ML (pML) estimators, averaged over 100 experiments, for model order $K \in [2, 4, 8, 16]$ and for values of $C \in [0, 2, 4, 8, 16, 32]$.

tested estimators, averaged over 100 experiments, for model orders $K \in [2, 4, 8, 16]$ and values of $C \in [0, 2, 4, 8, 16, 32]$ are depicted in Fig. 2. The RMSE values of the pML method appeared to be independent of K and are thus given only once. For all tested conditions the MAPb approach performs better than the pML method. As expected the pMAP approach reaches comparable performance to that of the proposed MAPb method only for small values of C.

Fig. 2(a) furthermore shows that the RMSE$_\mu$ values of all approaches increase with growing variance of the distortion, as controlled by the parameter C. Superiority of the MAPb approach over pML and pMAP methods becomes larger with growing GMM order K and especially for large values of C. Please note the logarithmic scaling of the ordinate in Fig. 2(b)! While the RMSE$_{\sigma^2}$ values of the pML and pMAP approaches increase with growing C, they seem to be independent of C for the MAPb approach. Further, it should be mentioned that the $RMSE_\omega$ values of the MAPb approach are slightly smaller than those of pML and pMAP. This is based on the better performance of MAPb in estimating the means $\mu_{k,l}$ and variances $\sigma_{k,l}^2$ compared to the other methods.

In experiments not reported here we verified the superiority of the proposed MAPb approach over pML and pMAP methods in tracking time-variant GMM parameters.

4. CONCLUSIONS AND RELATION TO PRIOR WORK

We have shown how MAP estimation of GMM parameters can be carried out in the presence of additive white Gaussian noise of known mean and variance. We have proposed a method to find an approximate posterior from the same family as the prior, which allows for an efficient estimation in a sequential estimation framework. Estimates with significantly lower error variance are obtained compared to other methods.

While the MAP estimation of the GMM parameters is already well-known [3], we present for the first time an approximate MAP estimator for the case that the target process is superposed by additive noise, a situation often occurring in practice. This is an extension of our earlier work [6], where we introduced an approximate MAP estimator in the presence of noisy observations, where the target process was a plain Gaussian random process.

5. REFERENCES

[1] Victor Hasselblad, "Estimation of parameters for a mixture of normal distributions," *Technometrics*, vol. 8, no. 3, pp. 431–444, 1966.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[3] Christopher M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[4] J. Gauvain and C. Lee, "Bayesian learning for hidden markov model with gaussian mixture state observation densities," 1991.

[5] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, 2001.

[6] A. Krueger and R. Haeb-Umbach, "MAP-based estimation of the parameters of non-stationary gaussian processes from noisy observations," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3596–3599, 2011.

[7] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol. 26, no. 2, pp. 195–239, 1984.