ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning. Bruges (Belgium), 22-24 April 2009, d-side publi., ISBN 2-930307-09-9.

BAYESIAN PERIODOGRAM SMOOTHING FOR SPEECH ENHANCEMENT

Xueru Zhang[1*], Alexander Ypma[2] and Bert de Vries[1,2]

1- Signal Processing Systems, Dept. of Electrical Engineering - TU Eindhoven
Eindhoven, - The Netherlands.
Xueru.Zhang@esat.kuleuven.be

2 - GN Research Centre - GN ReSound A/S -Eindhoven, - The Netherlands.
AYpma@gnresound.com, bdevries@gnresound.com

Abstract. Periodogram smoothing of the received noisy signal is a challenging problem in speech enhancement. We present a Bayesian approach, where the instantaneous periodogram is smoothed through an adaptive smoothing parameter. By updating sufficient statistics using new samples of the noisy signal, the smoothing parameter is adjusted on-line. The performance of the novel smoothing algorithm is studied in a speech enhancement context. It is demonstrated that with respect to Mean Square Error, the proposed Bayesian smoothing algorithm performs better than the other non-Bayesian smoothing algorithms in higher signal-to-noise ratio environments.

1 INTRODUCTION

Periodogram smoothing of the received noisy signal is an important component of a speech enhancement system. This is relevant for e.g. hearing aids and automatic speech recognition systems. The noise power spectrum is then estimated using the smoothed periodogram of the noisy signal. The variance of the instantaneous periodogram of the noisy signal affects the variance of the estimated noise power spectrum. Through smoothing, the variance of the instantaneous periodogram is reduced. In [1], a first order recursive model is used to smooth the periodogram, with a constant smoothing parameter. One disadvantage of this method is that it is difficult to keep a good balance between tracking relevant variations of the noisy signal, and at the same time generating a low bias. Therefore, an adaptive smoothing parameter was proposed by Rainer Martin and co-workers in [2]. In [2], the smoothing parameter is determined by extracting information about the current dynamic regime from the previous noise power spectrum estimate. This requires additional assumptions on the signal and induces a time delay to the tracking method.

In this paper, we present a Bayesian algorithm for instantaneous periodogram smoothing in speech enhancement systems. The paper is organized as follows. In Section 2, we explain the Bayesian periodogram smoothing model, with a

*The first author is currently with Katholieke Universiteit Leuven, Belgium. The authors would like to thank Václav Smídl, Tjeerd Dijkstra and Hugo Van hamme for useful discussions. Authors would thank IBBT for funding.

time-frequency dependent smoothing parameter. A forgetting factor and an alternative probability are introduced to express the prior probability over model parameters. An updating algorithm for the alternative prior probability is explained. In Section 3, we list the results by applying the proposed algorithm and a reference algorithm to synthetic data. The periodogram smoothing experiment results are shown in Section 4 and Section 5 presents our conclusions.

2 BAYESIAN PERIODOGRAM SMOOTHING

We represent a sampled noisy signal as $x(m)$, where m is the sample index. $X(k,l)$ is the short-time Fourier transform (STFT) of $x(m)$, with $k \in 1,2,...,K$ the frequency subband index, and $l \in 1,2,...,L$ the time frame index. The instantaneous periodogram P'_X [dB] of the noisy signal is given as:

$$P'_X(k,l) = 20\log_{10}|X(k,l)| \tag{1}$$

2.1 Periodogram smoothing model

The periodogram smoothing of P'_X is given by a first order recursion, using a time-frequency dependent smoothing parameter.

$$P_X(k,l) = [1-\alpha_X(k,l+1)]P_X(k,l-1) + \alpha_X(k,l+1)P'_X(k,l) \tag{2}$$

with P_X [dB] the smoothed output and $\alpha_X(k,l+1)$ the time-varying frequency-dependent smoothing parameter, with range [0,1]. For lucidity, $P_X(k,l)$ will be represented by P_l. In this way, eqn. (2) can be rewritten as:

$$P_l = P_{l-1} + \alpha_{l+1}(P'_l - P_{l-1}) \tag{3}$$

In eqn. (3), when α_{l+1} is set to be 0, P_l will be the same as the previous smoothed output P_{l-1}. The variance of the instantaneous periodogram is reduced, but the variations of the instantaneous periodogram will not be tracked. When $\alpha_{l+1} = 1$, P_l will be the same as the instantaneous periodogram P'_l. The variations of the instantaneous periodogram are tracked reliably. However, the instantaneous periodogram is not smoothed. Therefore, by adjusting the smoothing parameter, we change the degree of smoothing and tracking variations of the smoothed periodogram.

We define $\sigma_{l+1}e_{l+1}$ as the prediction error, which is the difference between the instantaneous periodogram P'_{l+1} and the smoothed output P_l, and assume $e_{l+1} \sim \mathcal{N}(0,1)$. $\mathcal{N}(0,1)$ is a Gaussian distribution, with mean 0 and variance 1. The instantaneous spectrum P'_{l+1} can then be expressed by the smoothed power spectrum P_l as follows:

$$P'_{l+1} = P_l + \sigma_{l+1}e_{l+1} \tag{4}$$

From eqn. (3) and eqn. (4), P'_{l+1} can be given as:

$$P'_{l+1} = P_{l-1} + \alpha_{l+1}(P'_l - P_{l-1}) + \sigma_{l+1}e_{l+1} \tag{5}$$

For simplicity, precision ω_{l+1} is rewritten as $(\sigma^2_{l+1})^{-1}$.

ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning. Bruges (Belgium), 22-24 April 2009, d-side publi., ISBN 2-930307-09-9.

2.2 Bayesian modelling

By rewriting eqn. (3) and eqn. (4) to eqn. (5), we can write down the (normal) likelihood over P'_{l+1}, given model parameters α_{l+1} and ω_{l+1}. Let $A = [-1, 1, a_{l+1}]^T$ and $B = [P'_{l+1}, P_{l-1}, P'_l - P_{l-1}]^T$. The likelihood is given as:

$$\begin{aligned} p(P'_{l+1}|\alpha_{l+1}, \omega_{l+1}, P'_l) \equiv& \mathcal{N}(P_{l-1} + \alpha_{l+1}(P'_l - P_{l-1}), \omega_{l+1}^{-1}) \\ =& (2\pi)^{-0.5}\omega_{l+1}^{0.5} \exp\{-\frac{1}{2}\omega_{l+1}A^T BB^T A\} \end{aligned} \tag{6}$$

The observations of the model, expressed by eqn. (4), are shown by the shaded circles in figure 1. Given the estimated model parameters, the smoothed output (which is shown in the second layer from the bottom in figure 1) is given by eqn. (3). The previous estimate of the posterior mean of P_{l-1} is now considered as a pseudo-observation [3].

Regarding the likelihood eqn. (6), we choose a Normal-Gamma (NG) distribution as the (conjugate) prior distribution for the joint probability of the model parameters α_{l+1} and ω_{l+1} [4], with sufficient statistics V_{l+1}, ν_{l+1}. We rewrite the unknown model parameters α_{l+1} and ω_{l+1} as θ_{l+1} with $\theta_{l+1} = [\alpha_{l+1}, \omega_{l+1}]^T$. This is shown by the third layer from below in figure 1.

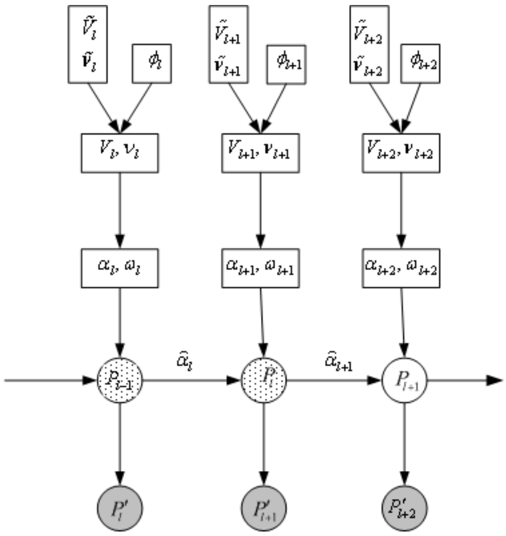

Fig. 1: Graphical model for Bayesian periodogram smoothing. Shaded circles represent observations. Dotted circles represent the pseudo-observations.

In nonstationary environments, the model parameters will change, but their evolution model is not known. Therefore, we adopt the approach in [5][6][7], where a time-varying forgetting factor ϕ_{l+1} and an 'alternative' fall-back distribution $\widetilde{p}$ [5] are introduced to model the prior probability over the model parameters from $p(\theta_l)$ to $p(\theta_{l+1})$. $\widetilde{V}$ and $\widetilde{\nu}$ are sufficient statistics of the alternative distribution, see the top layer in figure 1. The forgetting factor expresses the degree to which the model parameters for the next time frame are similar to the previous model parameters. If the environment is stationary, the forgetting factor ϕ_{l+1} is almost 1. Then next time model parameters are similar to the previous model parameters. If the environment rapidly changes, ϕ_{l+1} will be approximating 0. The alternative prior will then determine the next model parameters.

We use variational Bayes Expectation Maximization (VEM) to estimate the model parameters, in the same spirit as [5]. Details of our on-line Bayesian algorithm can be found in a technical report [8].

2.3 Updating the Alternative Prior Probability

The alternative prior distribution $\widetilde{p}(\theta_{l+1}|\mathbf{P}_l)$ is chosen to have the same form as $p(\theta_l|\mathbf{P}'_l)$ [8]. $\widetilde{V}_{l+1}$ is chosen as a combination of V_l and V_0 with coefficient β, see eqn. (7). Therefore, the alternative distribution takes into account both the model parameter estimated from the previous observations and the new initialization. In contrast, in [5] the alternative prior is a fixed matrix V_0.

$$\widetilde{V}_{l+1} = \beta V_l + (1 - \beta)V_0 \tag{7}$$

If $\beta = 1$, V_l from the previous time frame acts as the alternative prior. Therefore, model parameter α is assumed to be stationary.

If $\beta = 0$, the alternative prior is the same as initialization value V_0, and we obtain the Smidl algorithm. In this case, new exploration is included in the estimation of the (nonstationary) model parameter α.

3 SYNTHETIC DATA EXPERIMENTS

We apply the proposed alternative prior probability algorithm to a univariate second-order AR model, similar to the example in section 5.4 in [5]. The noisy speech signal is the sum of a clean speech signal and an uncorrelated additive white noise. Figure 2, left subfigure, shows the average square root of the mean square error (sqrt(MSE)) over 100 realizations of the AR process, with different updating coefficient β. The average square root of the MSE of our alternative prior updating algorithm is smaller than that of the original Smidl algorithm for $\beta \neq 0$. An analysis based on a 99% confidence interval on the mean difference between the methods corroborated that the performance improvement of prior updating over Smidl's original method is significant for $\beta \neq 0$ [8].

4 PERIODOGRAM SMOOTHING EXPERIMENTS

We implemented the previously discussed smoothing algorithms in a speech enhancement system. In this section, we compare performance of Bayesian periodogram smoothing algorithm with the adaptive step size Least Mean Square (ASLMS) algorithm [8], and Martin's algorithm.

ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning. Bruges (Belgium), 22-24 April 2009, d-side publi., ISBN 2-930307-09-9.

The speech utterances are taken from the database TIMIT [9]. Noise signals are taken from database Noisex92 [10]. The input SNR at which speech and noise are mixed is varying from -6 dB to 26 dB with step size 2 dB. For each time step of the variational Bayes EM algorithm, the forgetting factor is initialized to $\phi = 0.7$. We use 3 VEM iterations per time step, to ensure real-time performance.

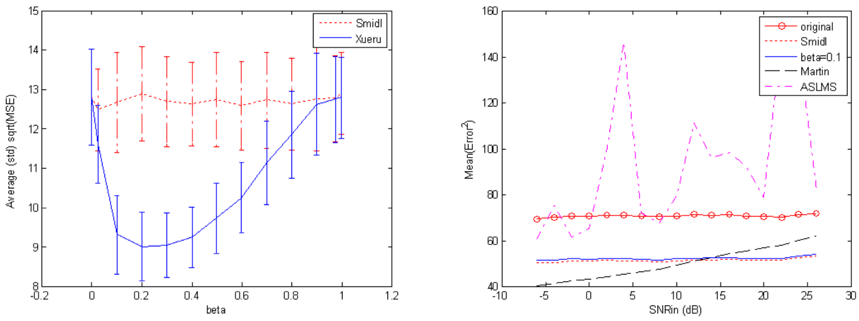

Fig. 2: *Left subfigure:* Assessing the effect of updating the alternative prior based on the Average Square Root of the MSE. The red dotted line ('Smidl') is obtained without prior updating. The blue solid line ('Xueru') is obtained with our alternative prior updating algorithm. Vertical bars represent standard deviations of the MSE. The horizontal axis displays the updating coefficient β. *Right subfigure:* Assessing our Bayesian periodogram smoother based on MSE. The horizontal axis displays input SNRs. The magenta dash dotted line represents the MSE based on ASLMS smoothing. The red solid line with circle marker is from the spectrum estimator of a reference ('original') speech enhancement system, with a fixed smoothing parameter. The blue solid line represents our Bayesian periodogram smoothing algorithm with $\beta \neq 0$. The red dotted line is our proposed Bayesian algorithm with $\beta = 0$. The black dashed line is Martin's smoothing algorithm.

We define MSE[1] as the mean square error between P_l and P'_{l+1}. In figure 2, right subfigure, we show the MSE for different periodogram smoothing algorithms. We see that smoothing based on Martin's method and our Bayesian method gives better performance than ASLMS and the periodogram estimator of the reference ('original') system. The performance of the ASLMS algorithm depends on a meta-parameter ρ for the step size. In our opinion, the slow convergence speed of ASLMS results in bigger MSE at several points. The system has not converged to the optimum values at the high MSE points around SNR = 4, 12, 14, 16, 22 and 24 dB. However, when it did converge (e.g. at SNR -6, -2, 0, 6, 8, 20, 26), we get a performance slightly better than 'original' for

[1] We use MSE to evaluate the performance of the approaches, because we address the task of smoothing in a statistically optimal way, instead of measuring the performance in terms of an index relevant for speech enhancement. However, the results indicate that better spectrum estimation maybe achieved, which could be exploited in speech enhancement applications.

SNRs below 5 dB, and slightly worse for higher SNRs. Martin's algorithm gives better performance than our Bayesian algorithm at lower input SNR. However, if the input SNR is higher, Bayesian smoothing algorithm gives lower MSE than Martin's algorithm. We think that for higher SNR, speech plays an important role in the periodogram smoothing and the Bayesian smoothing algorithm keeps better track of the nonstationary variations in the original speech signal. The tracking ability of our Bayesian periodogram smoother does not seem to benefit from updating the alternative prior, since the performance for $\beta = 0$ (referred to as 'Smidl') is almost indistinguishable from the performance for $\beta = 0.1$.

5 CONCLUSION

In this paper, we proposed a novel Bayesian algorithm for tracking of the smoothing parameter to reliably estimate a signal spectrum from the periodogram. The smoothing parameter is on-line updated for each time frame and frequency subband. Our Bayesian periodogram smoothing method yields an improved tracking ability at higher input SNRs. Our novel prior updating scheme allows for improved tracking performance with data from an AR(2) model.

Possible future improvements could be adjusting the coefficient for the alternative prior updating algorithm rather than using a fixed value. One way may be to adjust the β coefficient based on the tracking error, along the line of the Martin smoother. Through adjusting the coefficient, our Bayesian smoothing algorithm could be improved in stationary environments.

References

[1] R. Martin. Spectral subtraction based on minimum statistics. *in: Proceedings of the European Signal Processing Conference (EUSIPCO)*, pages 1182–1185, 1994.

[2] R. Martin, D. Malah, R.V. Cox, and A.J. Accardi. A noise reduction preprocessor for mobile voice communication. *EURASIP Journal on Applied Signal Processing*, (8):1046–1058, 2004.

[3] G. Sibley, G.S. Sukhatme, and L. Matthies. The iterated sigma point kalman filter with applications to long range stereo. *Robotics Science and Systems*, 2006.

[4] C.M. Bishop. Pattern recognition and machine learning. *Springer*, 2006.

[5] V. Šmídl. The variational bayes approach in signal processing. PhD thesis, Trinity College Dublin, 2004.

[6] V. Peterka. Bayesian approach to system identification. *in Trends and Progress in System Identification* P. Eykhoff, Ed, pages 239–304, 1981.

[7] R. Kulhavý and M.B. Zarrop. On a general concept of forgetting. *International Journal of Control*, 58(4):905–924, 1993.

[8] X. Zhang. Bayesian periodogram smoothing for speech enhancement. Technical report, Eindhoven University of Technology, 2008.

[9] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, and N. Dahlgren. The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM. National Institute of Standards and Technology (NIST), October 1990.

[10] A. Varga and H. J. M. Steeneken. Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems. *Speech Communication*, 12(3):247–251, July 1993.