OPTIMAL CEPSTRUM ESTIMATION USING MULTIPLE WINDOWS

Maria Hansson-Sandsten, Johan Sandberg

Lund University
Centre for Mathematical Sciences
Box 118, SE-221 00 Lund, Sweden
{sandsten, sandberg}@maths.lth.se

ABSTRACT

The aim of this paper is to find a multiple window estimator that is mean square error optimal for cepstrum estimation. The estimator is compared with some known multiple window methods as well as with the parametric AR-estimator. The results show that the new estimator has high performance, especially for data with large spectral dynamics, and that it is also robust against parameter choices. Simulated speech data is used for the evaluation. It is also shown that the windows of the estimator can be approximated with the sinusoidal multiple windows and that the weighting factors of the different periodograms can be analytically computed.

Index Terms— cepstrum analysis, multiple windows, multitaper, speech analysis

1. INTRODUCTION

Speech analysis is important in coding, classification as well as in other applications. To estimate relevant parameters of speech, one approach is the cepstrum related methods where mostly the LPCC (linear prediction cepstral coefficients) and MFCC (mel-frequency cepstral coefficients) are used to extract features for further analysis and classification.

Cepstrum methods based on robust spectrum analysis techniques, e.g., multiple windows could be applied to achieve higher performance. The Thomson multiple window method, [1], outperforms the Welch method in terms of leakage, resolution and variance. It has also been applied to speech analysis, [2]. For highly varying spectra, however, which often is the case for speech, the performance of the Thomson method degrades due to cross-correlation between subspectra, [3]. The sinusoid windows in [4] and the Peak Matched multiple windows, [5], have better bias properties for such spectra. Here we suggest a similar approach as the one in [5] for a mean square error optimal cepstrum estimate.

Thanks to the Swedish Research Council for funding.

2. CEPSTRUM MODEL

A discrete-time symmetrical cepstrum can be defined as

$$r_c(n) = \int_{-0.5}^{0.5} \log S_x(f) e^{i2\pi f n} df,\ n = -N+1 \ldots N-1, \tag{1}$$

where $r_c(-n) = r_c(n)$ and $S_x(f)$ is the even, real-valued spectral density function for the real-valued stationary process $\{x(n)\}_0^{N-1}$. The mean square error (MSE) of the cepstrum is obtained as

$$\sum_{n=-N+1}^{N-1} E[(\hat{r}_c(n) - r_c(n))^2] = \int_{-0.5}^{0.5} E[(\hat{S}_c(f) - S_c(f))^2] df, \tag{2}$$

where

$$S_c(f) = \sum_{n=-N+1}^{N-1} r_c(n) e^{-i2\pi f n},\ -0.5 < f \leq 0.5.$$

It has been shown, e.g. in [1, 5], that a frequency local estimator is important in spectrum estimation. To estimate the cepstrum, we need a frequency local estimator of the logarithmic spectrum which is given by,

$$\max \int_{-B/2}^{B/2} |H(f)|^2 S_c(f) df, \tag{3}$$

subject to $\int_{-0.5}^{0.5} S_z(f)|H(f)|^2 df = 1$ where $S_z(f)$ could be defined as the penalty spectrum in [5].

We define a symmetrical cepstrum where the M:th coefficient (and the mirrored coefficient at $-M$) are differing from zero,

$$r_c^M(n) = C_1\delta(n) + C_2\delta(n-M) + C_2\delta(n+M), \tag{4}$$

for $n = -N+1 \ldots N-1$. The zeroth coefficient is also included although this coefficient is usually omitted in applications using the cepstrum. The Fourier transform is given as,

$$S_c^M(f) = C_1 + 2C_2 \cos(2\pi f M), \quad -0.5 < f \leq 0.5. \tag{5}$$

The model cepstrum and corresponding logarithmic spectrum of Eqs. (4,5) has two parameters C_1 and C_2 to be chosen. The logarithmic spectrum shape which is a comb-spectrum with the frequency-distance $1/M$ between the peaks should be estimated. We use the shape of one such peak, (located at $f = 0$), as the bandlimited model. The width of the peak is $1/M$ which naturally is chosen as the bandwidth B in the estimation procedure. The spectrum of Eq. (5), $S_c^M(f)$ should be positive and close to zero at the frequency value $1/2M$, i.e. $S_c^M(1/2M) = C_1 - 2C_2 = 0$ giving $C_1 = 2C_2$. The resulting model spectrum $S_B^M(f)$ with $C_1 = 1$ and $C_2 = 0.5$ is seen in Figure 1.

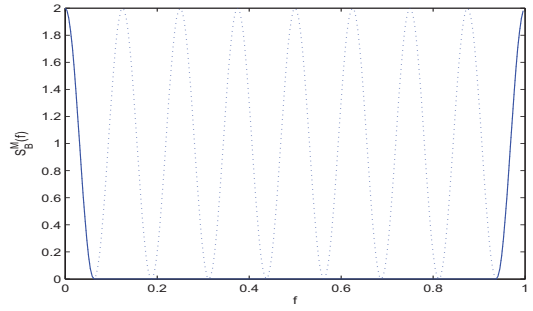

Fig. 1. Example of the spectrum shape $S_c^M(f)$ (dotted line) and the model spectrum $S_B^M(f)$ (solid line), $M = 8$.

The aim is to find the mean square error optimal estimator, based on a multiple window approach, to this spectrum and evaluate the performance for cepstrum estimation. The approach is similar to [5].

3. MULTIPLE WINDOW SPECTRUM ESTIMATION

The multiple window spectral estimator is defined as

$$\hat{S}_x(f) = \sum_{k=0}^{K-1} \alpha_k \left| \sum_{n=0}^{N-1} x(n) h_k(n) e^{-i2\pi f} \right|^2, \; -0.5 < f \leq 0.5, \tag{6}$$

where the set $\mathbf{h}_k = [h_k(0) \dots h_k(N-1)]^T$ is found as the eigenvectors of the (generalized) eigenvalue problem

$$\mathbf{R}_B^M \mathbf{q}_k = \lambda_k \mathbf{R}_Z \mathbf{q}_k, \quad k = 0 \dots N-1, \tag{7}$$

which is the solution of Eq. (3). The $(N \times N)$ Toeplitz covariance matrix $\mathbf{R}_B^M$ corresponds to $S_B^M(f)$ and $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{N-1}$. The eigenvectors corresponding to the K largest eigenvalues are used as windows, $\mathbf{h}_k = \mathbf{q}_k$, $k = 0 \; \dots \; K-1$. The covariance matrix $\mathbf{R}_Z$ could be chosen to grant low sidelobe level to the estimate $\hat{S}_x(f)$, [5]. Without sidelobe suppression $\mathbf{R}_Z = \mathbf{I}$. The weighting factors should be chosen for a mean square error optimal cepstrum and therefore also for the logarithmic spectrum. The mean square error for each frequency value f in Eq. (2),

$$E[(\hat{S}_c(f) - S_c(f))^2] = E[(\log \hat{S}_x(f) - \log S_x(f))^2] \tag{8}$$

$$\begin{aligned} &= \left(E[\log \hat{S}_x(f)] - \log S_x(f)\right)^2 + V\left[\log \hat{S}_x(f)\right] \\ &\approx \left(\log \frac{E[\hat{S}_x(f)]}{S_x(f)}\right)^2 + \frac{V[\hat{S}_x(f)]}{E^2[\hat{S}_x(f)]} \\ &\approx \left(\frac{E[\hat{S}_x(f)] - S_x(f)}{E[\hat{S}_x(f)]}\right)^2 + \frac{V[\hat{S}_x(f)]}{E^2[\hat{S}_x(f)]} \\ &= \frac{MSE\left[\hat{S}_x(f)\right]}{E^2[\hat{S}_x(f)]}, \end{aligned} \tag{9}$$

assuming that $S_x(f) \approx E[\hat{S}_x(f)]$. This approximation shows that the normalized mean square error of the spectral estimator $\hat{S}_x(f)$ is a reasonable estimate for the mean square error of the estimator $\hat{S}_c(f)$. In [6], choosing weighting factors as

$$\alpha_k = \frac{\lambda_k}{\sum_{k=0}^{K-1} \lambda_k}, \quad k = 0 \dots K-1, \tag{10}$$

showed to give approximately a minimum normalized mean square error estimate for the spectrum estimate. Therefore we can choose the weighting factors according to Eq. (10). The number of eigenvalues K will depend on the parameter M as $K \approx N/M$.

Figure 2 shows an example of the resulting multiple windows and weighting factors when $\mathbf{R}_Z = \mathbf{I}$. This estimator is named Multiple Window Cepstral Estimator ($MWCE$). In the analysis we also use the possibility to further suppress sidelobes with G dB of the multiple windows windows using the penalty matrix $\mathbf{R}_Z$ defined in [5]. The estimator is then called $MWCE_G$.

3.1. Approximated estimator

As this estimator is mean square error optimal only for a process with cepstrum defined by Eq. (4), a number of different eigenvalue problems need to be solved in an actual application to find the best estimator. This is not a practical scenario. The multiple window spectrum estimator can be computationally effective only if the set of multiple windows are the same for all values of M and we propose an approximation using the sinusoidal windows, [4].

If we assume that the K eigenvalues λ_k, $k = 0 \dots K-1$, approximately are equal to the spectrum values at frequency $f = k/2N$, an analytic approximation of the weighting factors is proposed as

$$\alpha_k^a = \frac{\cos(2\pi k M/2N) + 1}{\sum_{k=0}^{K-1} \cos(2\pi k M/2N) + 1}, \quad k = 0 \dots K-1, \tag{11}$$

which also is plotted as the dashed line in Figure 2b). The approximation is shown to be sufficiently close for all values of M.

The weighted square error is

$$e_{tot} = \sum_{k=0}^{K-1} \alpha_k^a \sum_{n=0}^{N-1} |h_k(n) - h_k^s(n)|^2, \qquad (12)$$

where $h_k^s(n)$ is the k:th sinusoidal window. An evaluation is performed for $N = 128$ and for values of $M = 1 \ldots 32$. The error is plotted in Figure 2c) where it can be seen that the approximation error is very small for low values of M but significantly larger for higher M. This is mainly due to the mismatch between the windows as the analytic approximation of the weighting factors are fairly good for all values of M.

However, using the sinusoidal windows, for a specific process realization of length N, the windowed periodograms only need to be computed once. Then, using the analytic formula of Eq. (11), the weighting of the periodograms for the mean square error optimal estimate valued for each M, is done. In this case no solution of eigenvalue problems are needed. We name this estimator the Sinusoidal Window Cepstral Estimator, ($SWCE$).

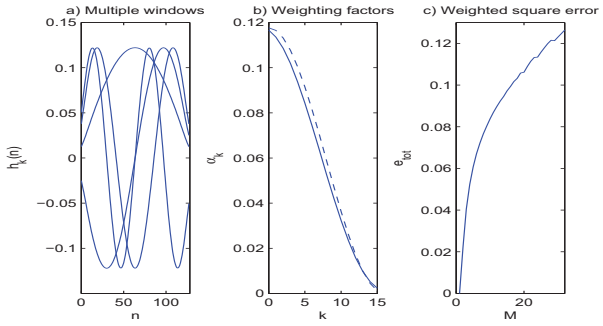

Fig. 2. Example of the eigenvectors and eigenvalues that are used as multiple windows and weighting factors for the spectrum estimation: a) The four first eigenvectors for $M = 8$; b) The eigenvalues, $k = 0 \ldots K-1$, (solid line) and the analytic approximation (dashed line), $M = 8$; c) The weighted square error of the approximation for different M.

4. EVALUATION

The proposed estimator is evaluated and compared with other well-known algorithms. We compare with AR-estimation (AR_M), equally weighted sinusoidal multiple windows, ($SIN\,MW$) and the Thomson multiple windows ($TH\,MW$). In all simulations $N = 128$ and the number of realizations are 500.

The performance is evaluated for processes with logarithmic spectra $S_c^M(f)$ for values of $M = 1 \ldots 32$. The realizations of the processes are simulated from the spectrum $S_x^M(f) = e^{S_c^M(f)}$. For each value of M, optimal choices for all algorithms in the evaluation are made. For the AR-estimator, we choose the order to be equal to M, which is a natural choice for estimation of a process that has a spectrum with M equally spaced peaks. The choice for the Thomson multiple windows are based on the assumption that the bandwidth of the peak of the combspectrum is $1/M$ and therefore $B = 1/M$ and the number of windows is chosen as an integer close to $BN - 2 = N/M - 2$, [1]. The number of sinusoidal multiple windows is also chosen according to the same rule. The mean square error of the cepstrum estimate is computed as

$$MSE_c = \sum_{n=1}^{N} E[(\hat{r}_c^M(n) - r_c^M(n))^2], \qquad (13)$$

for the different algorithms. Note that the cepstrum coefficient at $n = 0$ is excluded in the analysis. The reason is that the zeroth coefficient is usually omitted in cepstrum applications, e.g., speech analysis.

Figure 3 shows the results. The solid line is the $MWCE$ and the dash-dotted line the $SWCE$ which as expected give a slightly higher MSE than the optimal method. The dotted line and the dashed line are the results from the $SIN\,MW$ and the $TH\,MW$ respectively. The best result for all M is, as expected, given from the AR_M, (crossed line).

However, a comb-spectrum with all frequency peaks at the same level is not very close to a real world application. We instead simulate processes from a symmetrical spectrum which decreases according to,

$$S_x^M(f) = e^{S_c^M(f)} \cdot e^{-fd}, \quad ,0 \le f \le 0.5, \qquad (14)$$

where $d = 0.2$. The results from the evaluation for different values of M are depicted in Figure 4. The $MWCE$ (solid line) gives about the same performance as AR_M (crossed line). A smaller MSE is given from $SWCE$ (dash-dotted line), the $SIN\,MW$ (dotted line) and the $TH\,MW$ (dashed line). The smallest MSE, at least for processes of lower M, is given from the $MWCE_{30}$ (stars) where the sidelobe suppression of 30 dB combined with the optimal weighting of the periodograms give a better result than the other methods.

5. SPEECH ANALYSIS

To evaluate the performance for speech data, different AR-models are estimated from sounds of the syllables 'A' and 'L' of sampled data, $F_s = 11$ kHz. The choices of the orders of the AR-models are made from the final prediction error. Simulated data are extracted from the different models. The number of realizations are 500 and the data length is $N = 128$. We evaluate using the same set of algorithms and parameter choices as in Section 4. The results show that the AR_M performs well for the lower order models AR_8 for 'A', Figure 5b) (crossed line, minimum for $M = 8$) and

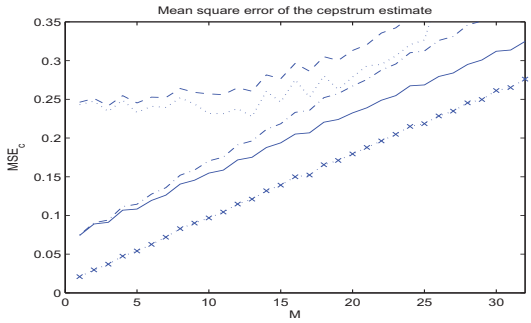

Fig. 3. MSE_c of the model cepstrum; $MWCE$ (solid line), $SWCE$ (dash-dotted line), $TH\,MW$ (dashed line), $SIN\,MW$ (dotted line), AR_M (crossed line)

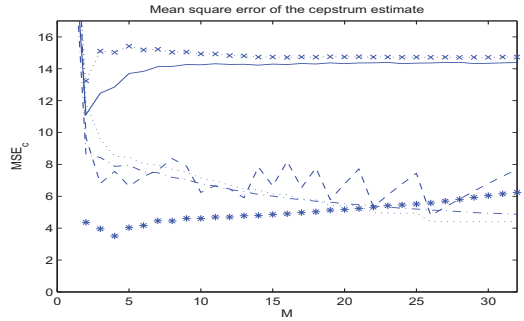

Fig. 4. MSE_c of exponentially decreasing spectrum; $MWCE$ (solid line), $SWCE$ (dash-dotted line), $MWCE_{30}$ (stars), $TH\,MW$ (dashed line), $SIN\,MW$ (dotted line), AR_M (crossed line)

AR_{11} for 'L', Figure 5d) (minimum for $M = 11$). However, the AR-model does not give the smallest MSE in any of the tested cases. The best results are given from the $SWCE$ (dash-dotted line) which seems to give the smallest error for $15 \leq M \leq 20$ for these data. It is closely followed by the $MWCE_{30}$ (stars) with minimum for $10 \leq M \leq 15$. The $SIN\,MW$ and the $TH\,MW$ are more sensitive to the choice of M but with a proper choice these methods also give small MSE. The $MWCE$ gives a significantly larger error which certainly is explained by the bad sidelobe suppression of the original windows, see the large jumps at the start and end of the windows, Figure 2a).

6. CONCLUSIONS

Multiple windows for mean square error optimal cepstrum estimation are proposed where windows and weighting factors are estimated from a comb-spectrum model. An approximation using the sinusoidal multiple windows and an analytical expression for the weighting factors are also suggested and evaluated. The new estimator is shown to give a smaller mean

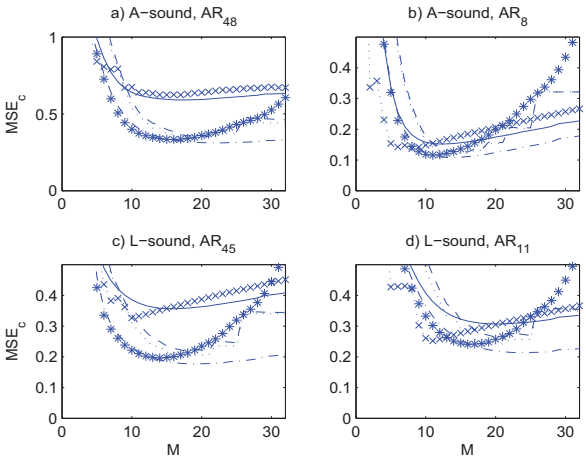

Fig. 5. MSE_c of different AR-models of the sounds 'A' and 'L'; $MWCE$ (solid line), $SWCE$ (dash-dotted line), $MWCE_{30}$ (stars), $TH\,MW$ (dashed line), $SIN\,MW$ (dotted line), AR_M (crossed line)

square error than AR-estimators and other multiple window estimators as the Thomson multiple windows and the sinusoidal multiple windows.

7. REFERENCES

[1] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. of the IEEE*, vol. 70, no. 9, pp. 1055–1096, Sept 1982.

[2] L. P. Ricotti, "Multitapering and a wavelet variant of MFCC in speech recognition," *IEE Proc.-Vis. Image Signal Processing*, vol. 152, no. 1, pp. 29–35, 2005.

[3] A. T. Walden, E. McCoy, and D. B. Percival, "The variance of multitaper spectrum estimates for real gaussian processes," *IEEE Trans. on Signal Processing*, vol. 42, no. 2, pp. 479–482, Feb 1994.

[4] K. S. Riedel, "Minimum bias multiple taper spectral estimation," *IEEE Trans. on Signal Processing*, vol. 43, no. 1, pp. 188–195, January 1995.

[5] M. Hansson and G. Salomonsson, "A multiple window method for estimation of peaked spectra," *IEEE Trans. on Signal Processing*, vol. 45, no. 3, pp. 778–781, March 1997.

[6] M. Hansson, "Optimized weighted averaging of peak matched multiple window spectrum estimates," *IEEE Trans. on Signal Processing*, vol. 47, no. 4, pp. 1141–1146, April 1999.