Consistent Nonparametric Spectrum Estimation Via Cepstrum Thresholding

M. Venkatanarayana narayanamoram@yahoo.co.in
Associate Professor/ECE/
K.S.R.M.College of Engg.
Kadapa-516003, India

Dr. T.Jayachandra Prasad jp.talaril@gmail.com
Principal, Professor in ECE
RGMCET
Nandyal-518 501, India

Abstract

For stationary signals, there are number of power spectral density estimation techniques. The main problem of power spectral density (PSD) estimation methods is high variance. Consistent estimates may be obtained by suitable processing of the empirical spectrum estimates (periodogram). This may be done using window functions. These methods all require the choice of a certain resolution parameters called bandwidth. Various techniques produce estimates that have a good overall bias Vs variance tradeoff. In contrast, smooth components of this spectral required a wide bandwidth in order to achieve a significant noise reduction. In this paper, we explore the concept of cepstrum for non parametric spectral estimation. The method developed here is based on cepstrum thresholding for smoothed non parametric spectral estimation. The algorithm for Consistent Minimum Variance Unbiased Spectral estimator is developed and implemented, which produces good results for Broadband and Narrowband signals.

Keywords: Cepstrum, Consistency, Cramer Rao Lower Bound, Unbiasedness.

1. INTRODUCTION

The main objective of spectrum estimation is the determination of the Power Spectral density (PSD) of a random process. The estimated PSD provides information about the structure of the random process, which can be used for modeling, prediction, or filtering of the deserved process. Digital Signal Processing (DSP) Techniques have been widely used in estimation of power spectrum. Many of the phenomena that occur in nature are best characterized statistically in terms of averages [20].

Power spectrum estimation methods are classified as parametric and non-parametric. Former one a model for the signal generation may be constructed with a number of parameters that can be estimated from the observed data. From the model and the estimated parameters, we can compute the power density spectrum implied by the model. On the other hand, do not assume any specific parametric model of the PSD. They are based on the estimate of autocorrelation sequence of random process from the observed data. The PSD estimation is based on the assumption that the observed samples are wide sense stationary with zero mean. Traditionally four techniques are used to estimate non parametric spectrum such as Periodogram, Bartlett method (Averaging periodogram), Welch method (Averaging modified periodogram) and Blackman-Tukey method (smoothing periodogram) [18] and [19].

2. CEPSTRUM ANALYSIS

The cepstrum of a signal is defined as the Inverse Fourier Transform of the logarithm of the Periodogram. The cepstrum of $\{y(t)_{t=0}^{t=N-1}\}$ can be defined as [7],[8] and [13]

$$c_k = \frac{1}{N}\sum_{p=0}^{N-1}\ln(\phi_p)e^{j\omega_k p}\ ;k = 0,......N-1 \tag{1}$$

Consider a stationary, discrete-time, real valued signal $\{y(t)_{t=0}^{t=N-1}\}$, the Periodogram estimate is given by

$$\hat{\phi}_p = \frac{1}{N}\left|\sum_{t=0}^{N-1} y(t)e^{-j2\pi ft}\right|^2 \tag{2}$$

A commonly used cepstrum estimate is obtained by replacing ϕ_p with the periodogram $\hat{\phi}_p$.

$$\hat{c}_k = \frac{1}{N}\sum_{p=0}^{N-1}\ln(\hat{\phi}_p)e^{j\omega_k p}\ ; \quad k = 0,......, N-1 \tag{3}$$

to make unbiased estimate the cepstrum coefficients only at origin is modified, remaining are unchanged.

$$\begin{cases} \bar{c}_0 = \hat{c}_0 + 0.577126 \\ \bar{c}_k = \hat{c}_k \quad k = 1,......N/2 \end{cases} \tag{4}$$

In this approach, we smooth $\left\{\ln \hat{\phi}_p\right\}$ by thresholding the estimated cepstrum $\{\bar{c}_k\}$, not by direct averaging of the values of $\left\{\ln \hat{\phi}_p\right\}$. The following test can be used to infer whether c_k is likely to be equal or close to zero and, there fore, whether $\overline{c_k}$ should be truncated to zero [9]-[12].

$$\tilde{c}_k = \begin{cases} 0 & if\left|\bar{c}_k\right| \le \dfrac{\mu\pi}{(d_k N)^{1/2}} \\ \bar{c}_k & else \end{cases} \tag{5}$$

The spectral estimate corresponding to $\{\tilde{c}_k\}$ is given by

$$\tilde{\phi}_p = \exp\left[\sum_{k=0}^{N-1}\tilde{c}_k e^{-j\omega_p k}\right]; \quad p = 0,........N-1 \tag{6}$$

The proposed non parametric spectral estimate is obtained from $\tilde{\phi}_p$ by a simple scaling

$$\hat{\hat{\phi}}_p = \hat{\alpha}\tilde{\phi}_p, \ p = 0,.....N-1 \tag{7}$$

where $\hat{\alpha} = \frac{\sum_{p=0}^{N-1} \hat{\phi}_p \tilde{\phi}_p}{\sum_{p=0}^{N-1} \tilde{\phi}_p^{\,2}}$; $\hat{\alpha}$ *isa scalingfactor*

Statistics of log periodogram

The mean and variance of the k th component of the log periodogram of the signal, $\log\left|\overline{Y}_k\right|^2$, assuming that the spectral component $\overline{Y}_k$ is Gaussian, are, respectively, given by [1]-[6],

$$E\{\log\left|\overline{Y}_k\right|^2\} = \begin{cases} \log(\lambda_{Y_k}) - \gamma - \log 2 & k = 0, K/2 \\ \log(\lambda_{Y_k}) - \gamma & k = 1,.........K/2-1 \end{cases} \quad (8)$$

where $\gamma = 0.57721566490$ is the Euler constant, and

$$\operatorname{var}(\log\left|\overline{Y}_k\right|^2) = \begin{cases} \sum_{n=1}^{\infty} \frac{n!}{(0.5)_n} \frac{1}{n^2} & k = 0, K/2 \\ \sum_{n=1}^{\infty} \frac{1}{n^2} & k = 1,....K/2-1 \end{cases} \quad (9)$$

where $(a)_n \cong 1.a.(a+1).(a+2)........(a+n-1)$. Furthermore,

$$\sum_{n=1}^{\infty} \frac{n!}{(0.5)_n} \frac{1}{n^2} = \frac{\pi^2}{2}; \ \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6};$$

Note from (8) that the expected value of the k th component of the log-periodogram equals the logarithm of the expected value of the periodogram plus some constant. This surprising linear property of the expected value operator is of course a result of the Gaussian model assumed here. From (9) the variance of the k th log-periodogram component of the signal is given by the constant.

Statistics of Cepstrum

The mean of the cepstral component of the signal is obtained from (8) and is given by [1], [2] and [7]

$$E\{c_y(n)\} = \frac{1}{K} \sum_{k=0}^{K-1} \log(\lambda_{Y\,k}) \exp\{ j \frac{2\pi}{K} kn \} - \frac{1}{K} \xi_n \quad (10)$$

where $\xi_n = \begin{cases} 2\log 2, & if\ n = 0\ or\ n\ even \\ 0, & if\ n\ odd \end{cases}$

the variance of the cepstral components is obtained from (9) and given by for $n = 0,..........K/2$

$$\begin{aligned} \operatorname{var}(c_y(n)) &= \operatorname{cov}(c_y(n), c_y(n)) \\ &= \begin{cases} \frac{2}{K} k_1 + \frac{2}{K^2}(k_0 - 2k_1), & if\ n = 0, \frac{K}{2} \\ \frac{1}{K} k_1 + \frac{2}{K^2}(k_0 - 2k_1), & if\ 0 < n < \frac{K}{2} \end{cases} \end{aligned} \quad (11)$$

and for $n, m = 0,1,............, K/2, n \neq m$

$$\operatorname{cov}(c_y(n), c_y(m)) = \begin{cases} \frac{2}{K^2}(k_0 - 2k_1) & \text{if } n-m = \pm 2, \pm 4, \ldots\ldots\ldots \pm \frac{K}{2} \\ 0, & otherwise \end{cases} \tag{12}$$

where $k_0 = \frac{\pi^2}{2}$; $k_1 = \frac{\pi^2}{6}$

The covariance matrix of cepstral components of the signal, assuming the spectral components of the signal are statistically independent complex Gaussian random variables. The covariance matrix of cepstral components given by (11) and (12) is independent of the underlying power spectral density which characterizes the signal under the Gaussian assumption. The covariance of cepstral components under the Gaussian assumption is a fixed signal independent matrix that approaches, for large K a diagonal matrix given by

$$\operatorname{cov}(c_y(n), c_y(m)) = \begin{cases} \frac{1}{K}\frac{\pi^2}{3}, & \text{if } n = m = 0, \frac{K}{2} \\ \frac{1}{K}\frac{\pi^2}{6}, & \text{if } 0 < n = m < \frac{K}{2} \\ 0, & otherwise \end{cases} \tag{13}$$

Cepstrum algorithm

1. Let a stationary, discrete-time, real valued signal $\{ y(t)_{t=0}^{t=N-1} \}$
2. Compute the periodogram estimate of ϕ_p using FFT.

$$\hat{\phi}_p(\omega) = \frac{1}{N} \left| \sum_{t=0}^{N-1} y(t) e^{-j\omega t} \right|^2$$

3. First apply natural logarithm and take IFFT to compute the cepstrum estimate.

$$\hat{c}_k = \frac{1}{N} \sum_{p=0}^{N-1} \ln(\hat{\phi}_p) e^{j\omega_k p};$$

$$k = 0, \ldots\ldots, N-1$$

4. Compute the threshold by choosing the appropriate value of μ depending on the type of signal and determine the cepstral coefficients

$$\tilde{c}_k = \begin{cases} 0 & \text{if} \left|\bar{c}_k\right| \le \frac{\mu\pi}{(d_k N)^{1/2}} \\ \bar{c}_k & else \end{cases}$$

5. Compute the spectral estimate corresponding to $\{\tilde{c}_k\}$ is given by

$$\tilde{\phi}_p = \exp\left[\sum_{k=0}^{N-1} \tilde{c}_k e^{-j\omega_p k} \right]; \quad p = 0, \ldots\ldots\ldots N-1$$

6. Obtain the proposed non parametric spectral estimate by a simple scaling

$$\hat{\hat{\phi}}_p = \hat{\alpha}\tilde{\phi}_p, \; p = 0, \ldots\ldots N-1$$

Simulation Results

In this section, we present experimental results on the proposed algorithm for simulated data to estimate the power spectrum. The performance of proposed method is verified for simulated data, generated by applying Gaussian random input to a system, which is either broad band or narrow band.The MA broad band signal is generated by using the difference equation [18]

$$\begin{aligned} & y(t)-1.3817y(t-1)+1.5632y(t-2)-0.8843y(t-3)+ \\ & 0.4096y(t-4)=e(t)+0.3544e(t-1)+0.3508e(t-2)+ \\ & 0.1736e(t-3)+0.2401e(t-4), \\ & \quad t=0,1,\ldots N-1 \end{aligned} \tag{14}$$

where $e(t)$ is a normal white noise with mean zero and unit variance. The ARMA narrow band signal is generated by using the difference equation

$$\begin{aligned} & y(t)-0.2y(t-1)+1.61y(t-2)-0.19y(t-3)+ \\ & 0.8556y(t-4)=e(t)-0.21e(t-1)+0.25e(t-2), \\ & \quad t=0,1,\ldots N-1 \end{aligned} \tag{15}$$

The number of samples in each realization is assumes as N=256.

After performing 1000 Monte Carlo Simulations, the comparison of the mean Power Spectrum, Variance and Mean Square Error for the broad band signal and narrow band signals, obtained using periodogram and cepstrum approach along with the true power spectrum are shown in Figure 1 (a) , (b) and (c) and Figure 2 (a), (b) and (c) respectively.

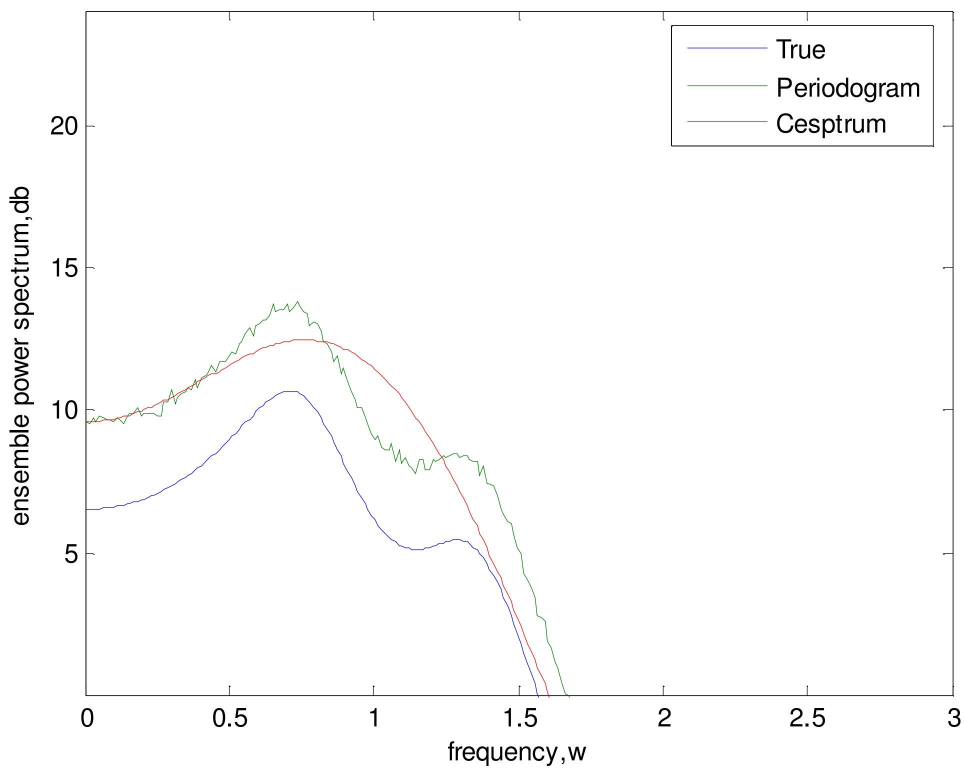

FIGURE 1: (a) PSD vs frequency for broadband signal

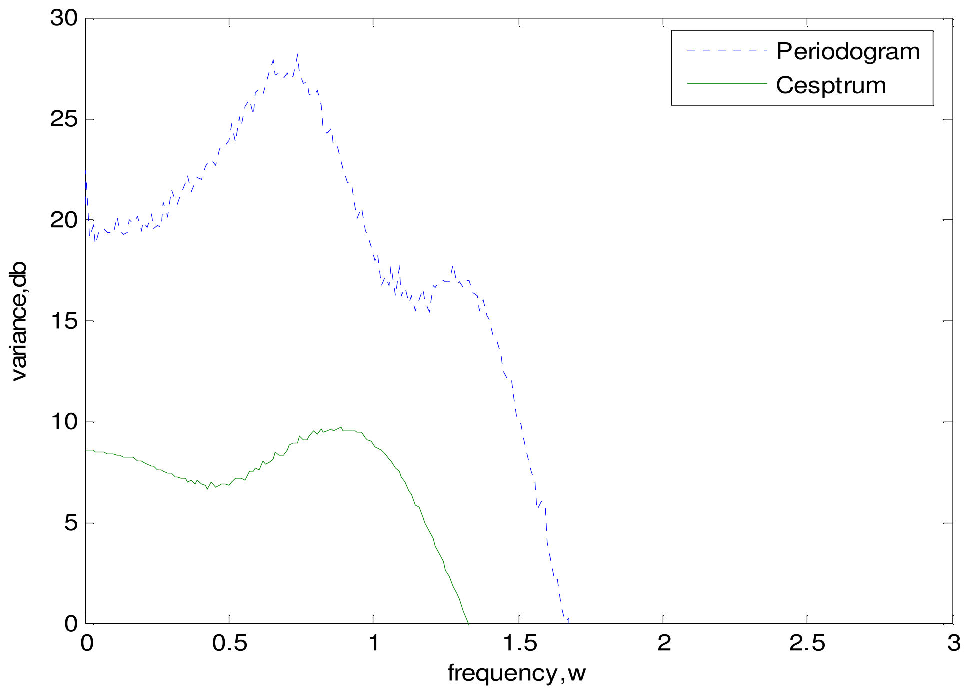

FIGURE 1: (b) Variance vs frequency for broadband signal

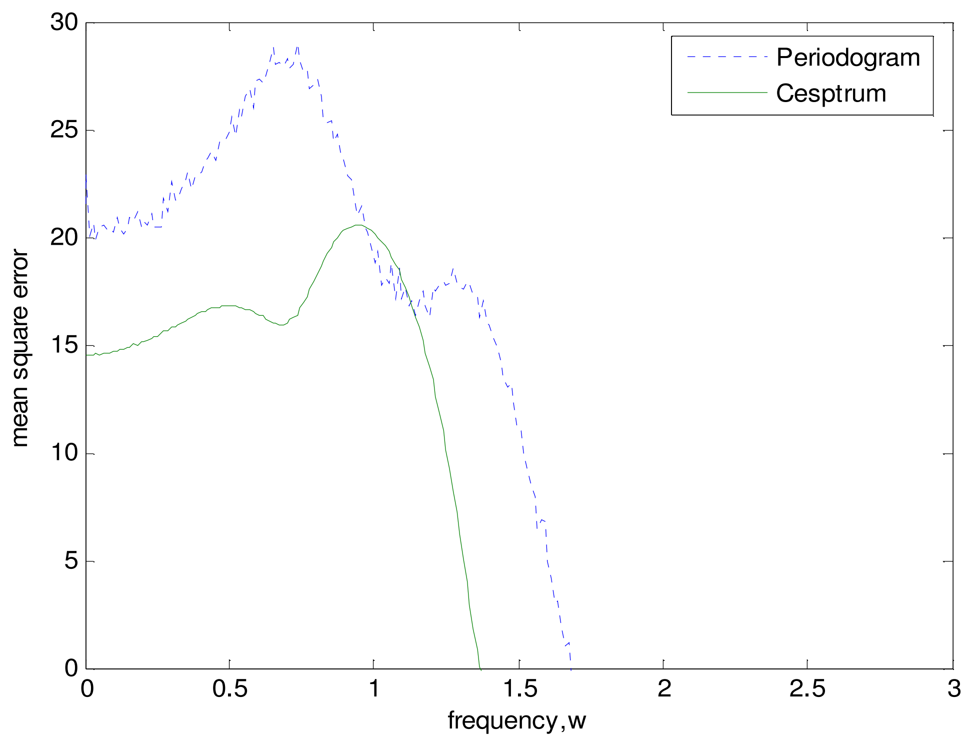

FIGURE 1: (c) Mean Square Error vs frequency for broadband signal

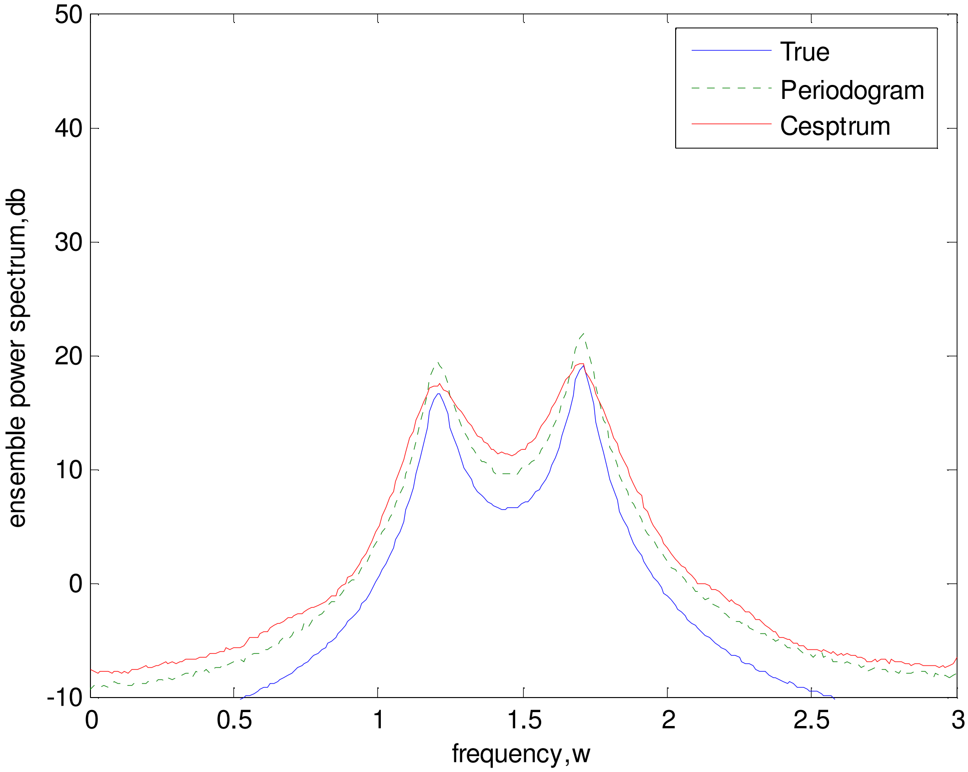

FIGURE 2: (a) PSD vs frequency for narrowband signal

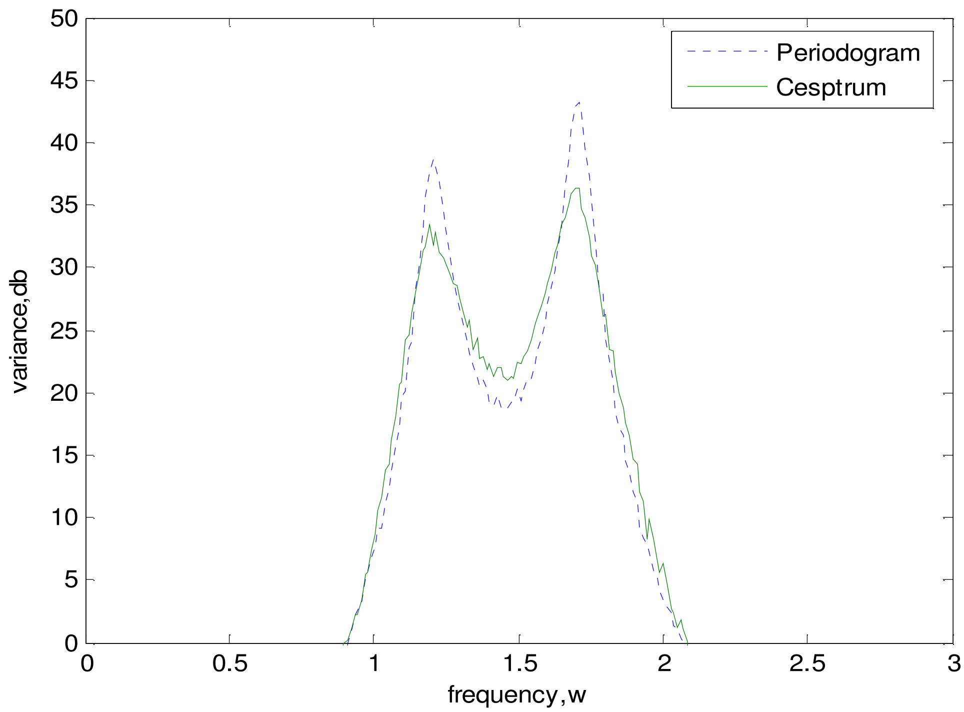

FIGURE 2: (b) Variance vs frequency for narrowband signal

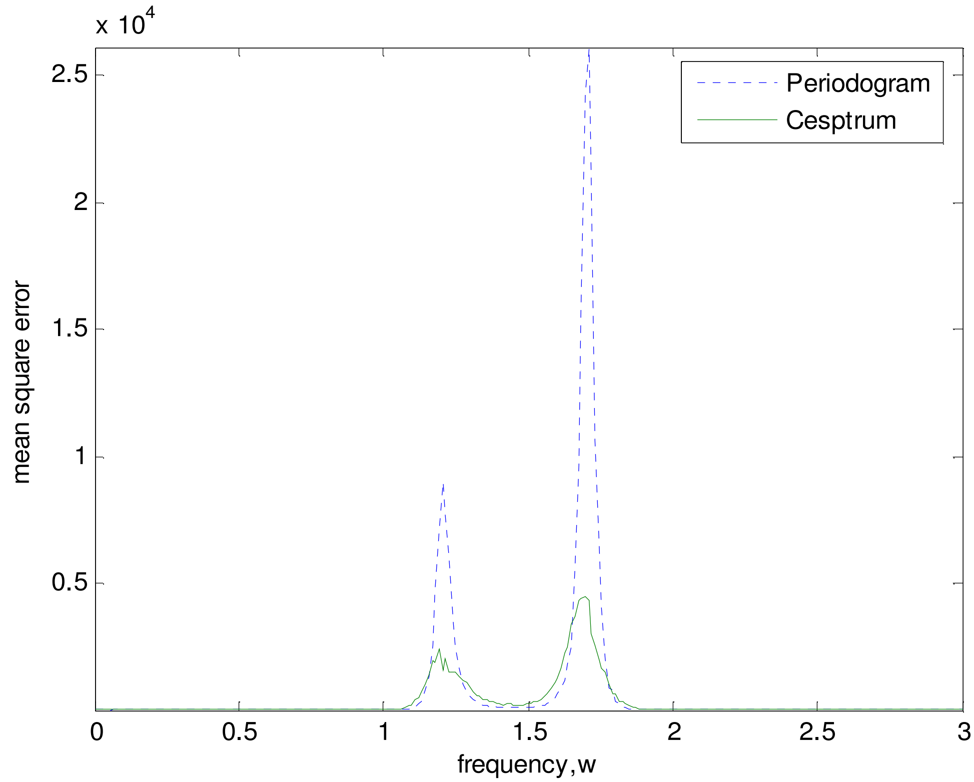

FIGURE 2: (c) Mean Square Error vs frequency for narrowband signal

From the above results we can say that

1. In the case of broad band signal the spectral estimates through cepstrum approach has very smooth response compared to the periodogram approach. However it can be observed that the mean square error is more in the case of periodogram and least with cepstrum thresholding approach.
2. In the case of broad band signals, variance obtained through cepstrum thresholding approach is very small as compared to the periodogram approach.
3. It is also observed that the mean square error estimated through cepstrum approach for narrowband signals is less compared to broadband signals.

Comparison among the traditional methods and the cepstrum method

In order to evaluate the performance of the cepstrum technique, which is compared with the traditional methods such basic Peridogram, Bartlett method, Welch method and Blackman and Tukey [21] for simulated ARMA narrow band signal, which is generated by using equation (15).

The various PSD techniques	Mean	Variance
Cepstrum	0.0090	2.4023e-004
Periodogram	0.0092	4.8587e-004
Black-man and Tukey	0.0521	0.0047
Welch	0.0138	8.9491e-004
Bartlett	0.2474	0.0637

TABLE 1: Comparison table for the parameters mean and variance (Record length N=128).

From the comparison table 1, for short record length, with respect to mean and variance, the cepstrum technique produces better results in comparison with the traditional methods. For longer record length, with reduced computational complexity, the cepstrum method produces the values of mean and variance as same as that of the Welch method, but these methods are better than the remaining techniques. For 1000 Monte carlo simulations, the ensemble power spectrum for various techniques is shown in figure 3.

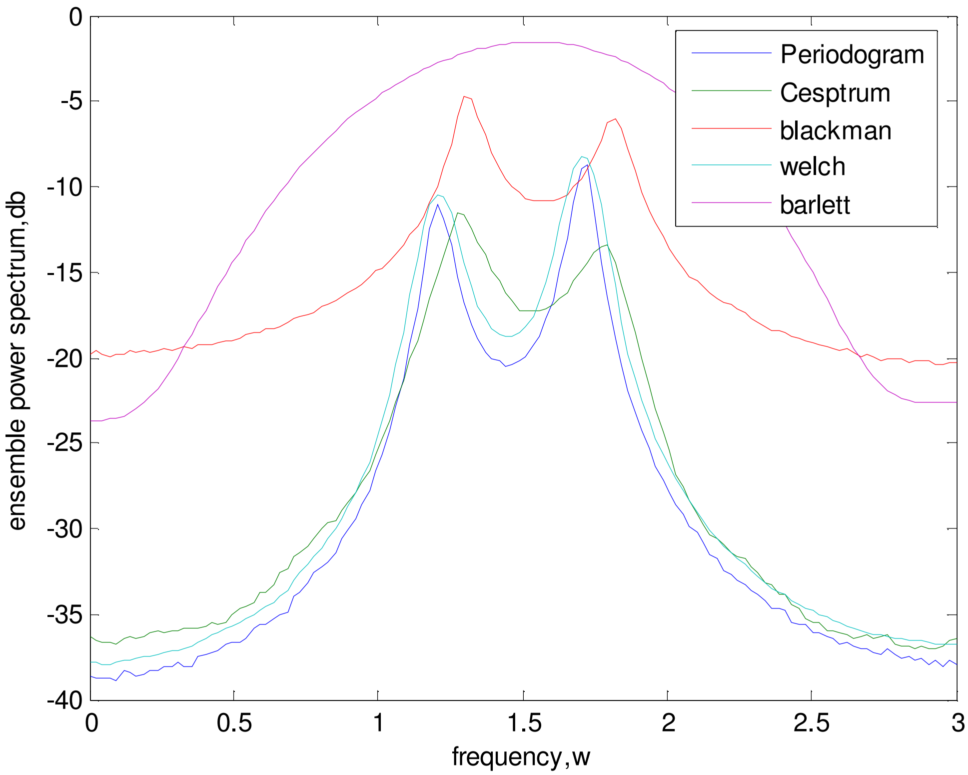

FIGURE 3: an ensemble power spectrum of an ARMA narrowband signal by using the traditional methods and the cepstrum method

Results for MST Radar data

The concept of cepstrum is applied to atmospheric data collected from the MST Radar on 10th August 2008 at Gadhanki, Tirupati, India. 150 sample functions, each having 256 samples are used to know the performance of cepstrum in comparison with the standard periodogram. The better results are obtained through the cepstrum than the periodogram. The comparison of the mean Power Spectrum, Variance for Radar data, obtained using periodogram and cepstrum approach are shown in Figure 4 (a) and (b) respectively. It is observed that the smooth power spectra and less variance in cepstrum than that of the periodogram.

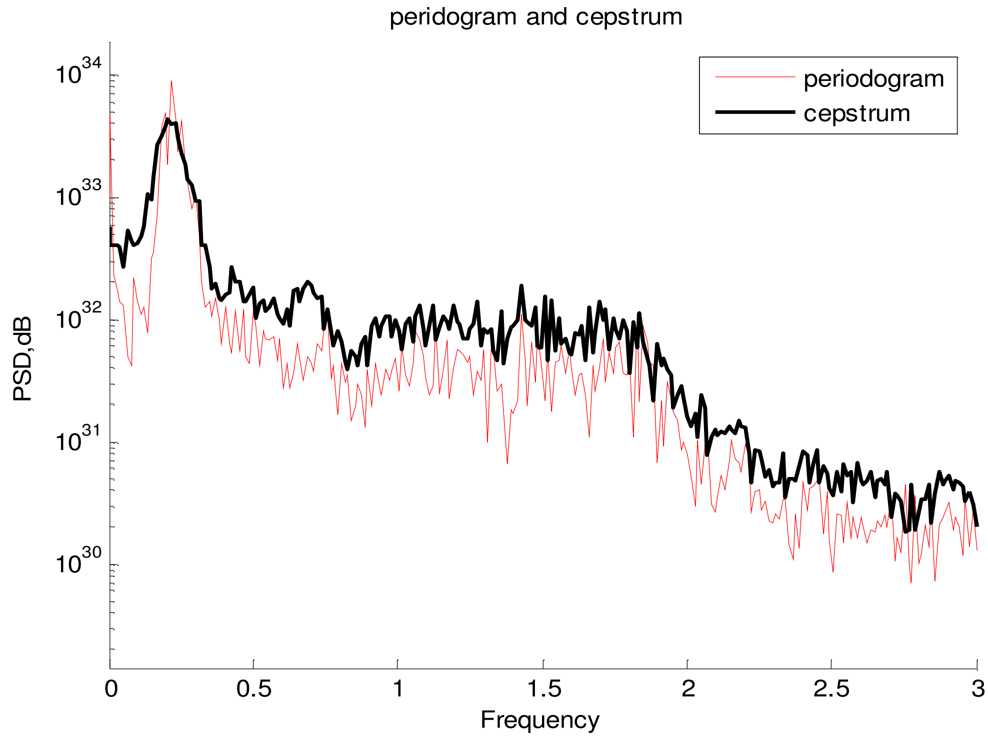

FIGURE 4: (a) Mean Power Spectra Vs Frequency for MST Radar data

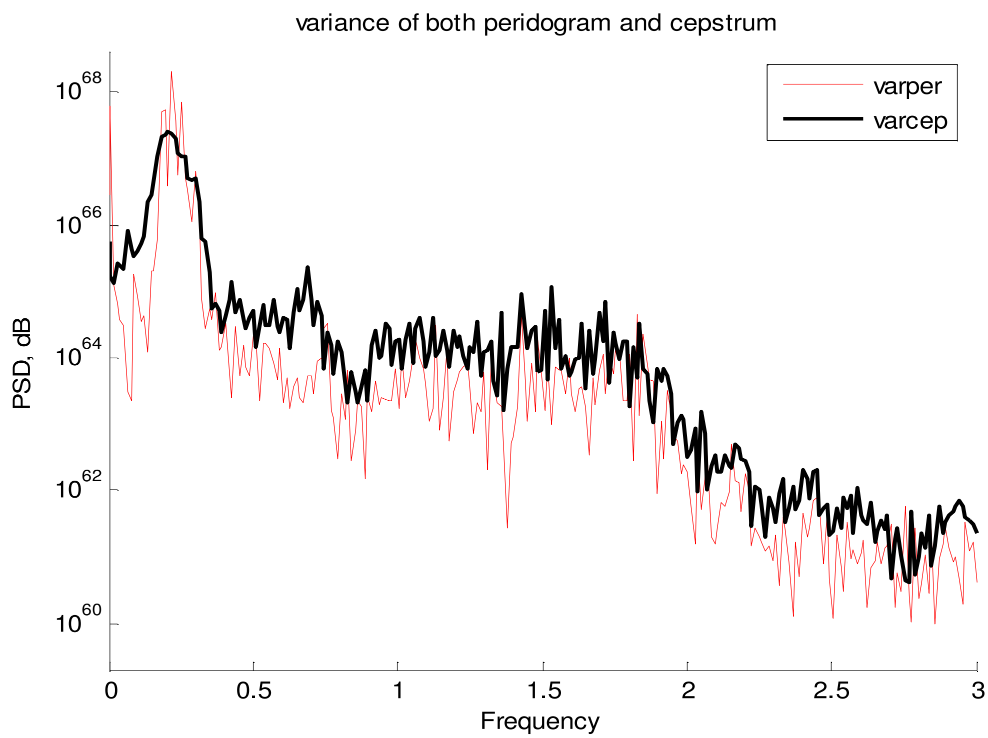

FIGURE 4: (b) Variance Vs Frequency for MST Radar data

3. CONSLUSION & FUTURE WORK

The problem in traditional methods is that the variance becomes proportional to square of power spectrum instead of converging into zero, thus the estimated spectrum is an inconsistent. In this paper the new technique has been proposed, called cepstrum, which gives reduce variance while evaluating the smoothed nonparametric power spectrum estimation. The expression for mean and variance of the cepstrum has been presented. The total variance reduction is more through broadband signals when compared to narrowband signals. All results are verified by using MAT lab 7.0.1. The concept of Cepstrum can be also extended for higher order spectral estimations.

4. REFERENCES

1. Y.Ephraim and M.Rahim, "*On second-order statistics and linear estimation of cepstral Coefficiets*", IEEE Trans.Speech Audio Processing, vol.7, no.2, pp.162-176, 1999.

2 A.H.Gray Jr., "*Log spectra of Gaussian signals*", Journal Acoustical Society of America, Vol.55, No.5, May 1974.

3 Grace Wahba, "*Automatic Smoothing of the Log Periodogram*", JASA, Vol.75, Issue 369, Pages: 122-132.

4. Herbert T. Davis and Richard H. Jones, "*Estimation of the Innovation Variance of a Stationary Time Series*": Journal of the American Statistical Association, Vol. 63, No. 321 (Mar., 1968), pp. 141- 149.

5. Masanobu Taniguchi, "*On selection of the order of the spectral density model for a stationary Process*", Ann.Inst.Statist.Math.32 (1980), part-A, 401-419.

6. Yariv Ephraim and David Malah, "*Speech Enhancement Using a Minimum Mean Square Error Short-Time Spectral Amplitude Estimator*", IEEE trans., on ASSP, vol.32, No.6, December, 1984.

7. P.Stoica and N. Sandgren, "*Smoothed nonparametric spectral estimation via cepstrum Thresholding*" IEEE Signal Processing Magazine, November, 2006, pp. 34-45.

8. D. G. *Childers D*. P. SLciAner and R. C. Kernemit, "*The Cepstrum: A Guide to Processing*", Proceedings of the IEEE, Vol. 65, no. 10, October 1977.

9. P. Moulin, "*Wavelet thresholding techniques for power spectrum estimation*" *IEEE Trans. Signal Processing*, vol. 42, no. 11, pp. 3126–3136, 1994.

10. H.-Y. Gao, "*Choice of thresholds for wavelet shrinkage estimate of the spectrum*" *J. Time Series Anal.*, vol. 18, no. 3, pp. 231–251, 1997.

11. A.T. Walden, D.B. Percival, and E.J. McCoy, "*Spectrum estimation by wavelet thresholding of multitaper estimators,*" *IEEE Trans. Signal Processing*, vol. 46, no. 12, pp. 3153–3165, 1998.

12. A.R. Ferreira da Silva, "*Wavelet denoising with evolutionary algorithms*" in *Proc. Digital Sign.*, 2005, vol. 15, pp. 382–399.

13. B.P. Bogert, M.J.R. Healy, and J.W. Tukey, "*The quefrency analysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking,*" in *Time Series Analysis*, M. Rosenblatt, Ed. Ch. 15, 1963, pp. 209–243.

14. E.J. Hannan and D.F. Nicholls, "*The estimation of the prediction error variance,*"*J. Amer.*

Statistic. Assoc., vol. 72, no. 360, pp. 834–840, 1977.

15. M. Taniguchi, "*On estimation of parameters of Gaussian stationary processes*," *J. Appl. Prob.*, vol. 16, pp. 575–591, 1979.

16. PETR SYSEL JIRI MISUREC, "*Estimation of Power Spectral Density using Wavelet Thresholding*", Proceedings of the 7th WSEAS International Conference on circuits, systems, electronics, control and signal processing (CSECS'08).

17. Petre Stoica and Niclas Sandgren, "*Cepstrum Thresholding Scheme for Nonparametric Estimation of Smooth Spectra*", IMTC 2006 - Instrumentation and Measurement Technology Conference Sorrento, Italy 24-27 April 2006.

18. P. Stoica and R.Moses, "*Spectral Analysis of Signals"*, Englewood Cliffs, NJ: Prentice Hall, 2005.

19. John G. Proakis and Dimitris G. Manolakis, "*Digital Signal Processing: Principles and Applications*", PHI publications, 2nd edition, Oct 1987.

20. M.B.Priestley, "*Spectral Analysis and Time series*" Volume-1, Academic Press, 1981.

21. Alexander D.Poularikas and Zayed M.Ramadan, "*Adaptive Filtering Primer with Matlab*", CRC Press, 2006