E4.4

HIGH PERFORMANCE SUD-LIKE PROCEDURE FOR SPECTRAL ESTIMATION USING RAYLEIGH FUNCTION ESTIMATES

Miguel A. Lagunas

Signal Processing on Communications.
E.T.S.I. Telecomunicación. Apdo. 30002
08080 Barcelona, SPAIN

ABSTRACT

This work describes how Rayleigh estimates can be viewed as a method which performs as a SVD procedure without doing it. As a short cut to get principal component reduction, Rayleigh quotients allows the resolution of frequency detectors yet preserving the asymptotic behavior to the actual power spectral density. In a filtering framework the estimate is extended to adaptive schemes and 2-D spectral estimation.

The resulting estimate provides the way out for adaptive processing with low computational complexity which, in general, are the two drawbacks associated to the principal component analysis. It avoids also the crucial decision between signal subspace and noise subspace which promotes undesired distortion and false peaks in spectral estimation applications.

1. INTRODUCTION

Principal component analysis (PCA) is becoming a major tool in signal analysis. There are many reported works of great interest in the topic which reveals that principal component analysis is a high performance procedure in filtering problems as well as a high resolution one in frequency detection problems or angle of arrival detection. In this sense, the works of D. Tufts and R. Kumaresan /1/ /2/ and J. Cadzow /3/ should be mentioned here.

Nevertheless, as gradient algorithms in adaptive filtering or Levinson method in spectral estimation, PCA needs of procedures of low complexity at the expense of the resulting performance in order to achieve a higher degree of incidence in signal processing applications. Let us summarize which are the main drawbacks of principal component analysis. First one is the computational load; the eigenvalues and eigenvectors have to be computed from the given data matrix, and, after selection of signal subspace in the filtering problem or the noise subspace in the frequency detection one, a low rank matrix is computed again. Second problem lies with the difficult decision about subspaces; regardless the decision, it is worthwhile to remain that in a spectral estimation problem both noise and signal have to be keeped along the procedure. In other words, resolution uses to be associated with a distorted version of the actual spectral estimate. Third one is related with the previous one, which is the use of PCA only in the case of sinusoides in white noise; probably this is motivated with the association of singular value decomposition with frequency detectors, without any attemp to use it in real spectral estimation. Fourth problem is related with the absence of adaptive schemes in PCA; due to the computation load associated and the necessary computation of eigenvalues and eigenvectors at every update in the adaptation process.

The purpose of this work is to show up to what degree, and for spectral estimation problems, the hereafter defined Rayleigh spectral estimates can cope with the above mentioned problems.

The structure of the paper is as follows: In Section II the Rayleigh quotients are proposed as a family of spectral estimates, starting from the concept of power function estimates as reported by Pisarenko some years ago. The concept of asymptotic convergence as a necessary condition for any candidate in spectral estimation is also introduced.

In section III the properties of the Rayleigh estimates are shown as well as its relationship, under some parameter selection, with currently reported methods for power spectral density estimates. What is really important on this section is up to what degree Rayleigh estimates is a short cut to enhance the SVD or PCA methods per-

---

This work is supported by CAICYT research grant number 2106/84 (Spain).

formance. Because the filter bank framework supports the reported methods included in the family of Rayleigh estimates, this framework is used in section IV to provide an adaptive scheme for the family.

2. Rayleigh spectral estimates

Any spectral estimate steems from the data autocorrelation matrix with a finite order. One way to understand the concept of asymptotic convergence is that when the order of the matrix Q tends to infinity the eigenvalues tend to be the samples of the actual power spectral density.

$$\lambda_q \rightarrow S(2\pi q/Q) \quad \text{as} \quad Q\rightarrow\infty \qquad (1)$$

Thus when Q is high the autocorrelation values could be represented by the eigenvalues S(w) and the eigenfunctions exp(jwm) as shown in (2).

$$r(m) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S(w)\, e^{jmw}\, dw \qquad (2)$$

This is the reason to consider that a good candidate to be an spectral estimate have to converge in some sense to the eigenvalues of the data autocorrelation matrix, being at the same time homogeneous (i.e. if λ_q changes to $\alpha.\lambda_q$, then S(w) changes to α.S(w)).

Based on the above, Pisarenko proposed in /4/ the so-called power function estimates with the following formulation:

$$S_m(w) = (\underline{S}^H\, \underline{\underline{R}}^m\, \underline{S})^{1/m} \qquad (3)$$

being $\underline{S}$ the steering vector (1, exp(jw),...,exp(jQw)) and $\underline{\underline{R}}$ the given data autocorrelation matrix of order Q.

It can be seen that the power function estimate includes the cases of the Blackman-Tukey estimate (m=1) and the Capon or maximum likelihood estimate of power level (m=-1), but, does not contain the case of the normalized maximum likelihood or real power density estimate reported by the author in /5/ and denoted in (4).

$$S_{NMLM}(w) = \frac{\underline{S}^H\, \underline{\underline{R}}^{-1}\, \underline{S}}{\underline{S}^H\, \underline{\underline{R}}^{-2}\, \underline{S}} \qquad (4)$$

Based in (4) and (3) a new family of spectral estimates emerges by generalizing the normalized maximum likelihood estimate by a quotient of two quadratic forms with sucessive orders as it is shown in (5). This estimate will be named hereafter as the Rayleigh estimate because its mathematical structure

$$S^m(w) = \frac{\underline{S}^H\, \underline{\underline{R}}^{-m+1}\, \underline{S}}{\underline{S}^H\, \underline{\underline{R}}^{-m}\, \underline{S}} \qquad (5)$$

3. Rayleigh estimates and principal component analysis

Most of the interesting features of the Rayleigh estimates can be cleared by the use of the principal components formulation. To introduce such formulation, let us assume that λ_q denotes the eigenvalues in decreasing order from λ_1 to λ_Q of the data autocorrelation matrix $\underline{\underline{R}}$, and $\underline{u}_q$ the corresponding eigenvectors with Fourier transform $U_q(w)$, where $U_q(w)$ is the squared magnitude of the dot product of the steering vector and the given eigenvector. Thus, using these definitions the numerator or the denominator of a Rayleigh estimate can be writen as:

$$\underline{S}^H\, \underline{\underline{R}}^{-m}\, \underline{S} = \sum_{q=1}^{Q} \lambda_q\, U_q(w) \qquad (6)$$

First at all, note that the estimate shown in (5) contents in the family the Blackman-Tukey estimate (m=0), the Capon estimate (m=1) and the normalized maximum likelihood or true spectral density from the Capon's estimate (m=2).

Going back to the properties and handling the eigenstructure of the estimator,

$$S^m(w) = \sum_{q=1}^{Q} \lambda_q^{-m+1}\, U_q(w) \Big/ \sum_{q=1}^{Q} \lambda_q^{-m}\, U_S(w) \qquad (7)$$

it results clear that the estimate is homogeneous because a scaling factor α in the eigenvalues conveys the same scaling factor in the resulting estimate.

Also, and based in the positive character of $U_q(w)$ for q=1, Q; is easy to prove that

$$S^m(w) \geqslant S^{(m+1)}(w) \qquad (8)$$

In other words, any previous estimate is a bound for the previous one. This allows two very important conclussions. First one concerns with how the very well-known estimates of B-T, ML and NML are bounded each other.

$$S_{BT}(w) \geqslant S_{ML}(w) \geqslant S_{NML}(w) \qquad (9)$$

Second one, formula (8) explains the role of parameter m in the Rayleigh quotient because if increasing m generates a lower bound for the previous estimate that means that resolution increases. In other words, m controls the peaky character of the resulting estimate.

Other property which makes interesting the family of estimates is the asymptotic convergence because when the order Q of the data autocorrelation matrix goes to infinity, the estimate converges, for any value of m, to the actual power spectral density. To see this, just note that when the order tend to infinity, $\lambda_q \rightarrow S(2\ q/Q)$ and only the corresponding $U_q(w)$ will be different from zero because the eigenvectors tend to the successive steering vectors $\underline{S}$. Thus it can be said that:

$$\lim_{Q\to\infty} S^m(w) = S(w)\ ;\ \forall m \qquad (10)$$

This is a very important property that does not hold, or it is not easy to prove, in many reported parametric and non-parametric methods.

Finally, it is remarkable what is the behavior of the estimate when parameter m goes to infinity. When m goes to infinity there are two different situations depending on the magnitude of the Fourier transform of the minimum eigenvalue $U_Q(w)$. If $U_Q(w)\neq 0$ we have (11.a)

$$\lim_{m\to\infty} S^m(w)=\lim_{m\to\infty} \frac{\lambda_{min}^{-m+1}}{\lambda_{min}^{-m}}=\lambda_{min} \qquad (11.a)$$

If $U_Q(w)=0$, that means that the steering vector $\underline{S}$ is orthogonal to the minimum eigenvector which belongs to the noise subspace. Thus, in this case, we are just steering the estimate where a source is located and $S^m(w)$ will peak above λ_{min} at this point. Let say that the smallest $U_q(w)$ which is not zero is named as $U_M(w)$, where M<Q, then (11.b) holds

$$\lim_{m\to\infty} S^m(w)=\lim_{m\to\infty} \frac{\lambda_M^{-m+1}\ U_M(w)}{\lambda_M^{-m}\ U_M(w)}=\lambda_M>\lambda_{min} \qquad (11.b)$$

The reader can see that the behavior of the Rayleigh estimates when the parameter m increases has a close relationship with the Pisarenko estimate. Rayleigh peaks at the source location leaving the constant value of λ_{min} in the rest of the frequency range. It is very important to remark that his property is not contradictory with the asymptotic convergence to the actual spectral distribution even when the spectrum under analysis is not a line spectra.

Let see now the relationship of Rayleigh with low rank approximations or principal component analysis.

Low rank reduction is the approximation of R, or its inverse in frequency detectors, by a set of eigenvectors and eigenvalues out of the Q which form the initial data matrix. Thus, assuming the selection for frequency detectors of the first P eigenvalues of the inverse autocorrelation matrix the resulting matrix $\underline{\underline{R}}^D$ is:

$$\underline{\underline{R}}^D = \sum_{m=Q-P}^{Q} \lambda_m^{-1}\ \underline{U}_m\ \underline{U}_m^T \qquad (12)$$

At this moment we can see $\underline{\underline{R}}^D$ as a non-linear operator applied to $\underline{\underline{R}}$, or, in the eigenspace as $\lambda_N=H(\lambda_q)$, i.e. the new eigenvalue is a non linear function of the old one. This situation is summarized in (13).

$$\underset{x=1/\lambda}{H(x)} \begin{cases} 0 & \text{if} \quad x<1/\lambda_{Q-P} \\ x & \text{if} \quad x\geq 1/\lambda_{Q-P} \end{cases} \qquad (13)$$

It is evident that an approximate function for H(x) could be a polynomial in increasing powers of x. The advantage of this are that, to obtain the low rank approximation there is no need of computing the eigenvalues and eigenvectors of $\underline{\underline{R}}^{-1}$ or $\underline{\underline{R}}$, because a polynomial function of the eigenvalues can be done directly over R as shown in (14).

$$\hat{H}(x) = \sum_{s=1}^{m} a_s\ x^s$$
$$\hat{\underline{\underline{R}}}^D = \sum_{s=1}^{m} a_s\ \underline{\underline{R}}^s \qquad (14)$$

In summary, this is the way out to obtain low-rank approximation without performing really singular value decomposition or principal component analysis. In fact, the denominator of a Rayleigh estimates can be viewed as a low rank approximation where H(x) is approximated by $H(x)=x^m$. To justify this there is an ungoing study comparing the denominator of Rayleigh estimates with MUSIC in angle of arrival detection. The preliminary results allow to the author to conclude that we have almost the same performance in resolution with m=5 even m=3 and, of course, with a very low computational cost. Furthermore, when the numerator is used all the properties mentioned previously are enhanced yet preserving the resolution performance.

4. Rayleigh estimates and filter bank analysis

As it is shown in /5/, the normalized maximum likelihood estimate is derived from the power level at the output of a pass band filter normalized by the noise bandwidth of the filter.

The design of the filter is done from a variational problem where the residual of the filter is minimized subject to some constraints of the filter related with its shape around the steering frequency /6/.

$$\underline{A}^T \underline{\underline{R}}\, \underline{A} \,\Big|\, \text{MIN} \qquad \underline{A}\, \underline{\underline{C}} = \underline{f} \tag{15}$$

Where $\underline{A}$ is the FIR filter impulse response in vectorial notation and $\underline{\underline{C}}$ defines the set of linear constraints with vector $\underline{f}$. In general, $(\underline{\underline{C}},\underline{f})$ reduces to the steering vector and zero dB response (i.e. $(\underline{S},1)$).

Due to the limitations in the extend of this paper, it will be shown in a few words how Rayleigh estimates could be encompassed in this framework. This is the subject of other publication in preparation by the author. Nevertheless in /5/, /6/ and in this work are the fundamentals to obtain the adaptive and 2-D schemes for Rayleigh estimates.

The key point is to review how matrix $\underline{\underline{R}}$ is formed in (15) from the data signal $\underline{x}^H = (x(n),\ldots,x(n-Q+1))$. If this data are windowed, the new data could be obtained as denoted in (16).

$$\tilde{\underline{x}} = \underline{\underline{W}}\, \underline{x} \quad \text{and} \quad \tilde{R} = \underline{\underline{W}}\, \underline{\underline{R}}\, \underline{\underline{W}}^T \tag{16}$$

Clearly, the use of a classical window reduces $\underline{\underline{W}}$ to a diagonal matrix. But in a generalized window theory $\underline{\underline{W}}$ will be a matrix in its general form. Now, remembering the data dependent window concept first introduced by Capon, an attemp to reduce to the principal components of $\underline{\underline{R}}$ will be to approximate $\underline{\underline{R}}$ by a low rank reduction of it. But, using the approximation mentioned in the previous section, the choice of R for W will give $\tilde{\underline{\underline{R}}}$ as $\underline{\underline{R}}^3$. Doing this, the use of powers of $\underline{\underline{R}}$ as matrix transformation, the spectral estimate obtained as follows will produce Rayleigh estimates.

Design equations:

$$\frac{\underline{A}^T \tilde{\underline{\underline{R}}}\, \underline{A}}{\underline{A}^T \underline{S}} \Big|_1 \text{MIN} \quad \text{with } \tilde{R} = \underline{\underline{W}}\, \underline{\underline{R}}\, \underline{\underline{W}}^T \tag{17.A}$$

Estimate equations:

$$S(w) = \frac{\underline{A}^T \underline{\underline{R}}\, \underline{A}}{\underline{A}^T \underline{A}} \tag{17.b}$$

As an example, a transformation of $\underline{\underline{W}}=\underline{\underline{R}}$ will produce the Rayleigh estimates with parameter m equal to 6.

Note that it is a difference between the design equations and the filtering or estimate equations. It results clear that (17.b) is always the way to obtain the power spectral density from a filtering measurement technique (i.e. power of the output residual, averaged or not, divided by the bandwidth estimate). Thus, the filter designed must be used with the original data to preserve the asymptotic convergence and not over the transformed ones. On the other hand, the filter is designed based on the transformed data in order to enhance the low rank reduction involved in the matrix $\underline{\underline{R}}$.

5. Conclusions

A new spectral estimate which enhances all the properties needed in a valuable spectral estimator is reported. Homogeneity, asymptotic convergence, tradeoff between resolution and statistical stability under control of the user; made it uncomparable with currently reported methods. As a result the procedure can be viewed as the way out to obtain SVD performance without doing it really. The inclussion to adaptive schemes using Frost algorithm or to 2-D problems is also summarized. Both aspects are under study in detail, and will be the subject of further publications.

6. References

/1/ D. Tufts & R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making Linear prediction perform like maximum likelihood", Proc. IEEE, Vol. 70, pp. 975-989, September 1982.

/2/ L.L. Sharf & D. Tufts, "Rank reduction for modelling stationary signals", IEEE Trans. on Acoustics Speech and Signal Processing, ASSP-35, pp. 350-355, March 1987.

/3/ J. Cadzow, "Signal enhancement using canonical projection operators", Proc. ICASSP-87, pp. 673-676, Dallas, April 6-9, 1987.

/4/ V.F. Pisarenko, "On the estimation of spectra by means of non-linear functions of the covariance matrix", Geophys. J.R. Astro. Soc. (1972) 28, pp. 511-531.

/5/ M.A. Lagunas et al., "Maximum likelihood filters in Spectral Estimation problems", Signal Processing 10, Ed. North-Holland, 1986, pp. 19-34.

/6/ M.A. Lagunas, "The variational approach in spectral estimation", Proceedings of the EUSIPCO Conference (Invited), Ed. North-Holland, EUSIPCO-86, The Hague, The Netherlands, Sept. 1986.