ON-LINE MODEL SELECTION OF NONSTATIONARY TIME SERIES USING GERSCHGORIN DISKS

*Patrice Michel†, Jean-Yves Tourneret† and Petar M. Djurić**

† ENSEEIHT/TéSA, 2 rue Charles Camichel, BP 7122, 31071 Toulouse Cedex 7, France
* Department of Electrical and Computer Engineering, SUNY at Stony Brook, NY 11794, USA

ABSTRACT

The paper proposes a method for on-line model selection of nonstationary time series. The method is based on computation of the covariance matrix of the data, transformation of the matrix by Housholder's tridiagonalization, and application of a clustering algorithm that can separate the Gerschgorin disks of the transformed covariance matrix into disks that correspond to the signals and noise, respectively. The method is applied to on-line estimation of the the number of harmonic signals in noise. Simulation results are presented that show the performance of the proposed method.

1. INTRODUCTION

In many applications [3] [7] [8], observed data contain harmonic signals that appear for a short period of time, and then disappear. They may appear again at some random instant of time, or will remain absent for the rest of the observation interval. The amplitude of these harmonics may not be necessarily constant, and in fact, frequently it varies with time in an unpredictable way. An important signal processing problem in such cases is the determination of the number of harmonics in the data at every instant of time.

We addresses this problem by using the theory of Gerschgorin disks. Initially, Gerschgorin disks were used for estimation of number of sources corrupted by additive white Gaussian noise [11]. The problem with the method employed in [11] is that it cannot be used in on-line applications because it is computationally very demanding. In this paper we propose a scheme that exploits Gerschgorin disks but it does not require operations that cannot be implemented on standard DSP chips. Since the information about the number of harmonic signals is in the covariance matrix of the most recent data, and in particular, its eigenvalues, the idea is to determine the sets of values they may take, and based on these sets, estimate the number of harmonic signals currently in the data. From linear algebra, it is known that crude estimates of these sets are provided by the Gerschgorin disks. If these disks are determined from the original covariance matrix, they are not of much use because most of them have significant overlaps. However, if we apply unitary transformation to the covariance matrix, the disks may shrink and separate and become much more informative. One such transformation is Housholder's tridiagonalization. Once this transformation is applied to the original covariance matrix, we can use a standard classification algorithm to separate the disks that correspond to signal eigenvalues from those that correspond to noise eigenvalues.

Section 2 provides a review of the theory of Gerschgorin disks. The following section proposes a transformation of the covariance matrix of the data that facilitates the estimation of the number of harmonic signals in the data. The proposed method is outlined in Section 4. In Section 5 simulation results are provided, and in Section 6 some final remarks are made.

The work of P. M. Djurić was supported by the National Science Foundation under Award No. CCR-9903120.

2. GERSCHGORIN DISKS AND MODEL SELECTION

Gerschgorin theorem ([6], p. 341), [11], [10] states that the eigenvalues of a $q \times q$ squared matrix $A = (a_{ij})_{i,j=1,\dots,q}$ belong to the union of q disks called Gershgorin disks whose centers and radii are $c_i = a_{ii}$ and $r_i = \sum_{j \neq i} |a_{ij}|$, respectively. An additional property of interest here states that if a collection of k Gerschgorin disks are isolated from the remaining Gerschgorin disks, there exist exactly k eigenvalues of A contained in this collection ([10], pp. 71-72), [11]. This last property is appealing and can be used for development of techniques for model order determination. In particular, these techniques would be based on the location of the Gerschgorin disks associated with the covariance matrix of the noisy observations. Indeed, if the Gerschgorin disks can be split into two disjoints sets corresponding to the signal and noise eigenvalues, the number of signal eigenvalues can be estimated easily from these disks. Unfortunately, the Gerschgorin radii may be large and the Gerschgorin centers may be clustered together with the Gerschgorin disks being overlapped in most practical applications. Several authors have recently proposed to solve the problem of disks overlapping by transforming the estimated covariance matrix of the observed time series using unitary transformations [1], [2], [11], [12].

An example of such unitary transformation consists of partitioning the covariance matrix C as

$$C = \begin{pmatrix} C_1 & c \\ c^H & c_{qq} \end{pmatrix} \tag{1}$$

where the subscript ^{H} denotes Hermitian transpose of a matrix. If the eigendecomposition of the so-called reduced covariance matrix C_1 is denoted by $C_1 = U_1 D_1 U_1^H$, and the unitary matrix

$$U = \begin{pmatrix} U_1 & 0 \\ 0^H & 1 \end{pmatrix} \tag{2}$$

is applied to C, one obtains the transformed covariance matrix $S = U^H C U$ (whose expression is detailed in [11]). Other unitary transformations such as the steering vector transformation [2] or a transformation based on the concept of sample correlation coefficient [12] have also been studied. The effect of these unitary transformations is to provide two separate collections of Gerschgorin disks. The first collection consists of Gerschgorin disks with large radii associated with the signal (or source), whereas the second collection is composed of Gerschgorin disks with small radii associated with the noise. Moreover, these unitary transformations require only $(q-1)$-dimensional eigendecomposition instead of q-dimensional eigendecomposition, needed in the standard approach. Unfortunately, the computational complexity of all these eigendecompositions prevent them from real time applications (in particular, the eigenvalues cannot be computed easily with digital signal processors).

For the estimation of number of sources, several likelihood and heuristic approaches based on Gerschgorin disks have been proposed in the literature [11][2]. Most of these approaches consist of comparing an appropriate criterion that depends on Gerschgorin radii with a suitable threshold. For instance, some authors have proposed to evaluate the Gerschgorin Disk Estimator (GDE) criterion

$$GDE(k) = r_k - \frac{D(N)}{q-1}\sum_{i=1}^{q-1} r_i \qquad (3)$$

for $k = 1, ..., q-1$, where $D(N)$ is a non-increasing function of the number of samples N (between 0 and 1) and to compare this criterion to 0. The number of sources is then determined as $k-1$, where k is the first value such that $GDE(k) < 0$. The GDE criterion was motivated by noting that Gerschgorin radii associated with noise are usually smaller than $\frac{D(N)}{q-1}\sum_{i=1}^{q-1} r_i$, which is contrary to the radii associated to the sources. The performance of the GDE criterion has been recently improved by considering a normalized distance based on the radii and centers of the Gerschgorin disks [2]. However, the threshold $D(N)$ may be difficult to adjust in practical applications. In addition, the Gerschgorin disk radii are not always interesting, since a Gerschgorin disk center remote from zero may have a small radius and a disk center near zero may also have a large radius (as noted in [2]).

3. GERSHGORIN DISKS OF COVARIANCE MATRICES TRANSFORMED BY HOUSEHOLDER TRIDIAGONALIZATION

The objective in this paper is to develop a scheme for model selection that is based in Gershgorin disks. To that end, before computing the Gerschgorin radii of the covariance matrix of the observations, we propose that first it is transformed by Householder tridiagonalization ([6], p. 420). Since the Householder tridiagonalization consists of successive unitary transformations, it is important to note that the eigenvalues of the covariance matrix are not altered by such transformation. Moreover, the preprocessing of the covariance matrix by the Householder tridiagonalization separates the noise Gerschgorin disks from the signal Gerschgorin disks, without requiring eigendecompositions. This property is illustrated in Figures 1 and 2, which show the Gerschgorin disks corresponding to a $q \times q$ covariance matrix of two sinusoids embedded in additive white Gaussian noise:

$$x_t = \sum_{i=1}^{2} A_i \sin(2\pi f_i t) + n_t, \qquad t = 1, ..., N \qquad (4)$$

with $N = 32, q = 8, A_1 = A_2 = 4.47, f_1 = 0.25, f_2 = 0.27$ (normalized frequencies) and $E\left(n_t^2\right) = 1$ (same parameters as in [1]). Figure 1 shows that the Gerschgorin disks of the original covariance matrix are overlapped. The Gerschgorin disks in Figure 2 correspond to the covariance matrix after it was tranformed by Householder tridiagonalization. The disks are clustered into two groups, one corresponding to the sinusoids, and the other to the noise.

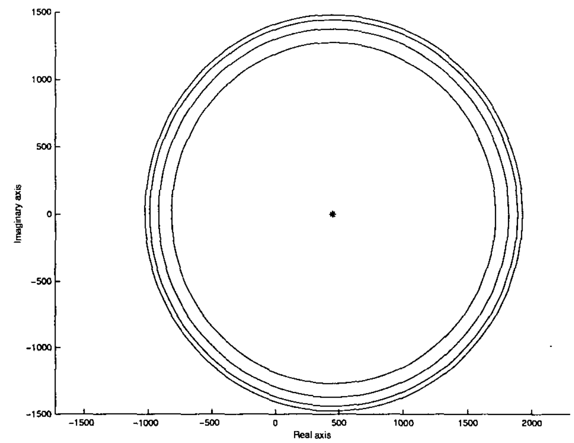

Fig. 1. Gerschgorin disks of the original covariance matrix of x_t.

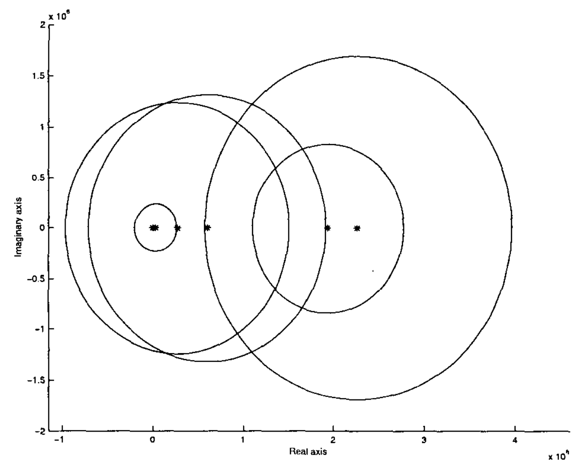

Fig. 2. Gerschgorin disks of the transformed covariance matrix of x_t.

Here we also propose to use clustering techniques [5] [9] to separate the Gerschgorin disks that correspond to noise from those that correspond to harmonic signals, that is, to estimate the number of harmonic signals in the observed data without computing the eigenvalues of the covariance matrix. More precisely, the K-means/Isodata algorithm [5] [9] is used to form a partition of the Gerschgorin centers into two classes corresponding to signal and noise. The Isodata algorithm is an iterative 3 step procedure which consists of 1) choosing initial centroids, 2) classifying the samples by assigning them to the class of the closest centroid (this step requires definition of an appropriate distance, which is here the Euclidean distance) and, 3) recomputing the centroid as the average of the samples in each class. The steps 1), 2) and 3) are repeated as soon as the centroids change. The Isodata algorithm forms a partition which minimizes the trace of the within scatter matrix. An advantage of the Isodata algorithm is its computational simplicity ([9], p. 483), since it converges in a finite number of iterations. In our application, the two initial centroids are chosen as the smallest and largest Gerschgorin centers and the algorithm converges in few iterations. After convergence of the Isodata algorithm, the Gerschgorin disk classes can be either isolated or not. In the former case, the number of signal and noise eigenvalues are the numbers of Gerschgorin centers associated with the signal and noise classes. In the latter case, the eigenvalues which are larger (resp. smaller) than the centroid of the two classes are attributed to the signal (resp. to noise). The number of these eigenvalues is evaluated by using the Sturm sequence property ([6], p. 438). This property states that the eigenvalues of the leading r-by-r principal submatrix of a tridiagonal matrix A strictly separate the eigenvalues of the $(r+1)$-by-$(r+1)$ principal submatrix. Such property can be used to evaluate the number of eigenvalues less than a given threshold (for more details, see [6], p. 438).

4. A NEW METHOD FOR ESTIMATING THE NUMBER OF HARMONIC SIGNALS

In a non stationary context, the covariance matrix of the signal that is embedded in noise at time t is classicaly estimated from the data samples as follows (see [7])

$$R(t) = \sum_{k=1}^{t} \lambda^{t-k} z_L(k) z_L^T(k) \qquad (5)$$

where λ is a forgetting factor, $z_L(k) = [y_{k-1}, ..., y_{k-L}]^T$, and L is the sliding window length. This paper proposes to estimate the number of harmonic signals at time t by using the following 4-step procedure:

- Computation of the covariance matrix $R(t)$,
- Householder tridiagonalization of the covariance matrix $R(t)$,
- Determination of the Gerschgorin centers of the tridiagonalized $R(t)$,
- Assigning the Gerschgorin centers to signal or noise using the K-means/Isodata algorithm,
- Estimation of the number of signal eigenvalues using the Sturm sequence property.

In the proposed implementation, the 4-step procedure is repeated sequentially for each time t.

5. SIMULATION RESULTS

Many simulations have been performed to illustrate the performance of the proposed algorithm.

Example 1 (stationary case): the observed signal is the sum of two sinusoids corrupted by additive white Gaussian noise:

$$x_t = \sum_{i=1}^{2} A_i \sin(2\pi f_i t + \phi_i) + n_t, \qquad t = 1, ..., N \qquad (6)$$

where $N = 64$, $A_1 = A_2 = \sqrt{20}, f_1 = 0.2, f_2 = 0.2 + \frac{1}{N}$ (normalized frequencies), $\phi_1 = 0, \phi_2 = \frac{\pi}{4}$. The noise variance is chosen in order to obtain an appropriate signal to noise ratio $SNR = 10\log_{10}\frac{a_i^2}{2\sigma^2} = 5dB$ (same parameters as in [4]). Table I shows the estimated number of sinusoids obtained with the conventional minimum description length (MDL) criterion, the maximum *a posteriori* (MAP) algorithm and the Gerschgorin algorithm (GA) for 100 Monte Carlo runs.

	$p=0$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
MDL	0	0	74	17	3	6
MAP	0	0	99	1	0	0
GA	0	14	86	0	0	0

Table I : Performance Comparison for MDL, MAP and GA criteria.

The MAP strategy outperforms the Gerschgorin algorithm and the MDL criterion. The performances of the MDL criterion and the Gerschgorin algorithm are similar (the main difference is that the MDL criterion tends to overestimate the order p whereas the Gerschgorin algorithm tends to underestimate it). However, it is important to note that the Gerschgorin algorithm can be readily implemented in real-time processing, whereas the MAP algorithm described in [4] cannot.

Example 2 (non-stationary case): The observed signal is constructed as follows

$$x_t = \begin{cases} A_1 \sin(2\pi f_1 t + \phi_1) + n_t & t = 1, ..., 300 \\ \sum_{i=1}^{2} A_i \sin(2\pi f_i t + \phi_i) + n_t & t = 301, ..., 700 \\ A_1 \sin(2\pi f_1 t + \phi_1) + n_t & t = 701, ..., 1000 \end{cases}$$

where $A_1 = A_2 = \sqrt{20}, f_1 = 0.2, f_2 = 0.2 + \frac{1}{32}, \phi_1 = 0, \phi_2 = \frac{\pi}{4}$. The results obtained with the proposed approach are compared with an on-line order selection algorithm described in [7]. However, since the Akaike information criterion (AIC) frequently overestimates the model order (as noticed in [4]), the AIC has been replaced by the MDL criterion. Figures 3 to 6 compare the estimated probabilities of selecting the order p, obtained with the on-line MDL criterion [7] and the Gerschgorin algorithm, for different values of p and different SNR's (note that there were 100 trials for each SNR). The Gerschgorin algorithm clearly outperforms the MDL criterion, specially for $SNR = 15dB$, where the MDL criterion estimates only one sinusoid for $t = 301, ..., 700$. Many experiments have shown that the Gerschgorin algorithm continues to work well provided that $SNR \geq 5dB$.

6. CONCLUSIONS

The paper addressed the problem of on-line determination of the number of harmonic signals in nonstationary data. A method was proposed that is based on the locations of the Gerschgorin disks of the transformed covariance matrix of the data. Once the disks are located, the K-means/Isodata clustering algorithm is applied to separate the disks correponding to the signals and noise. The method was tested on stationary and nonstationary data, and it showed very promising results.

7. REFERENCES

[1] O. Caspary, P. Nus and T. Cecchin, "The source number estimation based on Gerschgorin radii disks," *in Proc. ICASSP*, pp. 1993-1997, 1998.

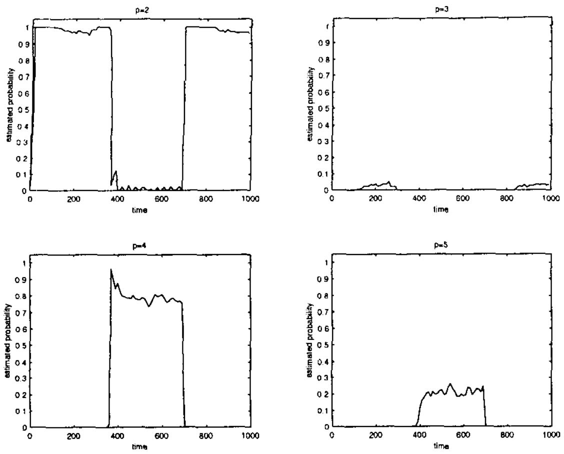

Fig. 3. Estimated probabilities (MDL criterion, SNR=20dB).

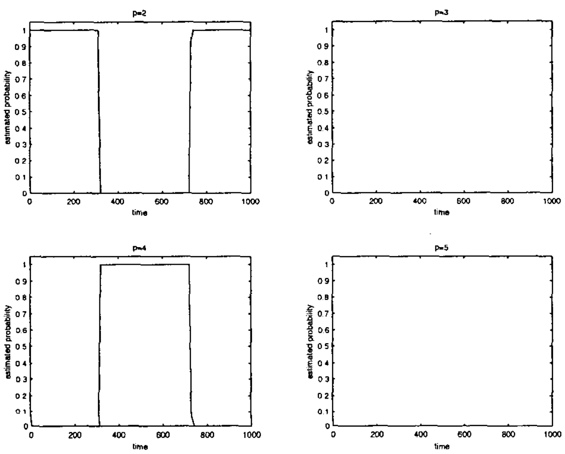

Fig. 4. Estimated probabilities (GA, SNR=20dB).

[2] O. Caspary and T. Cecchin, Detection of the source number by the Gerschgorin disks," *in Proc E SIPC* , 2000.

[3] Y. S. Cho, S. B. Kim and E. J. Powers, Time-varying spectral estimation using AR models with variable forgetting factors," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. 39, no 6, pp. 1422-1n426, 1991.

[4] P. Djuric, A model selection rule for sinusoids in white Gaussian noise," *IEEE Trans. Sig. Proc.*, vol. 44, no 7, pp. 1744-1751, 1996.

[5] R. Duda and P. Hart, *Pattern Classi cation and Scene Anal- sis* , New York: Wiley, 1973.

[6] G. H. Golub and C. F. Van Loan, *Matri computations* , Johns Hopkins University Press, Baltimore, 1983.

[7] S. Goto, M. Nakamura and K. Uosaki, On-line spectral estimation of nonstationary time series based on AR model parameter estimation and order selection with a forgetting factor," *IEEE Trans. Sig. Proc.*, vol. 43, no 6, pp. 1519-1522, 1995.

[8] M. Niedzwiecki, *Identi cation of Time- ar ing Processes* , New York: Wiley, 2000.

[9] S. Theodoridis and K. Koutroumbas, *Pattern Recognition.* Academic Press, 1999.

[10] J. H. Wilkinson, *The Algebraic Eigen alue Problem.* Oxford, UK: Clarendon, 1965.

[11] H. T. Wu, J. F. Yang and F. K. Chen, Source number estimators using transformes Gerschgorin radii," *IEEE Trans. Sig. Proc.*, vol. 43, no 6, pp. 1325-1333, June 1995.

[12] H. T. Wu, A new Gerschgorin radii based method for source number detection," *in Proc SP or shop Stat. Signal Arra Proc.*, pp. 104-107, 2000.

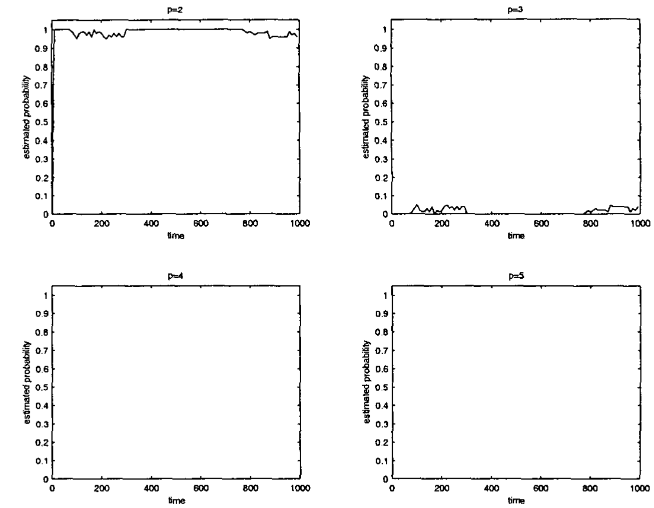

Fig. 5. Estimated probabilities (MDL, SNR=15dB).

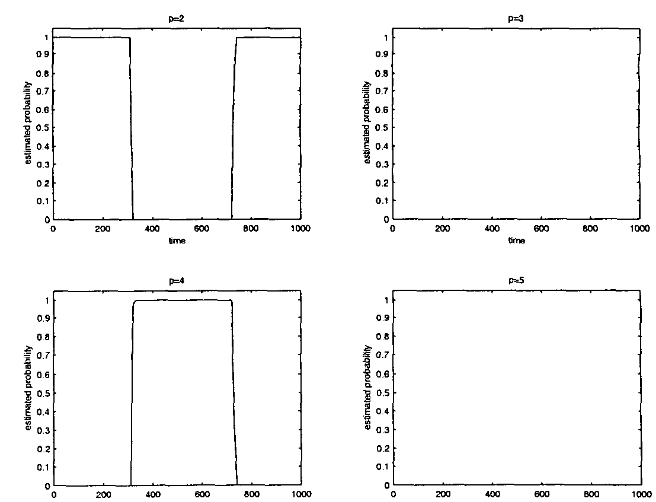

Fig. . Estimated probabilities (GA, SNR=15dB).