GAUSS-NEWTON BASED ADAPTIVE SUBSPACE ESTIMATION

G. Mathew*, V.U. Reddy* and S. Dasgupta**

ABSTRACT

In this paper, we present an adaptive approach for estimating all (or some) the orthogonal eigenvectors of the data covariance matrix (of a time series consisting of real narrowband in additive white noise). We use inflation approach to estimate each of these vectors as minimum eigenvectors (eigenvectors corresponding to the minimum eigenvalue) of appropriately constructed symmetric positive definite matrices. This reformulation of the problem is made possible by the fact that the problem of estimating the minimum eigenvector of a symmetric positive definite matrix can be restated as the unconstrained minimization of an appropriately constructed non-linear non-convex cost function. The modular nature of the algorithm, that results from this reformulation, makes the proposed approach highly parallel, resulting in a high-speed adaptive approach for subspace estimation.

1. INTRODUCTION

The problem of estimating the parameters of narrowband signals in additive white noise has been a subject of active research recently. The various methods which can be applied to this kind of problem can be broadly classified into two categories; i) eigenstructure based methods which exploit the eigenstructure of the underlying covariance matrix and ii) non-eigenstructure based methods. The eigenstructure based methods are preferred to the other since they yield high resolution and asymptotically exact results. Implementation of these methods calls for the estimation of all or some of the eigenvectors of the covariance matrix. Even though these eigenvectors can be estimated using any of the known block approaches (if we are given the covariance matrix) for real time applications we need methods to adaptively estimate them. The problem that we address in this paper is the adaptive estimation of these eigenvectors formulated in a non-linear least-squares framework.

Most of the researchers used the fact that eigenvectors can be estimated by minimizing a specific cost function subject to certain non-linear constraints [2], [3], [8], for developing algorithms for adaptive estimation.

While most of the reported work concentrates on estimating a single eigenvector, only few have addressed the problem of estimating more then one eigenvector adaptively. An adaptive approach for estimating the orthogonal eigenvectors corresponding to the signal subspace of the covariance matrix was first developed by Owsley [2]. Thompson [2] exploited the constrained minimization formulation to develop a constrained stochastic gradient algorithm for seeking the eigenvector corresponding to the minimum eigenvalue of the covariance matrix. Later, Larimore [3] studied the covergence characteristics of Thompson's [2] approach. Reddy et al. [4] restated the constrained minimization problem into an unconstrained framework and developed an approximate Gauss-Newton recursive algorithm for seeking the minimum eigenvector. The development of adaptive algorithms for the single eigenvector case is further explored by Durrani and Sharman [5] and Fuhrman and Liu [6]. Sharman [7] has developed an adaptive algorithm, based on h QR-recursions, to estimate the complete eigenstructure of the covariance matrix. Recently, Yang and Kaveh [8] proposed an adaptive approach for the estimation of the complete noise subspace or the signal subspace of the covariance matrix.

In this paper, we present an adaptive approach, which combines a Gauss-Newton algorithm and an inflation method, for estimating the eigensubspace of the covariance matrix. We develop this algorithm in an unconstrained minimization framework instead of the commonly used constrained framework. The motivation for this work has come from an earlier work of the authors [9]. The basic principle of this approacch can be used to compute the complete eigenstructure (set of all orthogonal eigenvectors and the corresponding eigenvectors and the corresponding eigenvalues) of any symmetric positive definite matrix. Applications of the proposed adaptive approach can be found in super resolution methods which make use of either the complete noise subspace or the signal subspace.

2. Computation of the Complete Eigenstructure of a Symmetric Positive Definite Matrix.

Let x(n) denote the time series consisting of the sum of P real narrowband signals corrupted with additive white noise of variance σ^2. Let $\mathbf{R}$ be the asymptotic covariance matrix of size N x N, $(N \geq 2P + 1)$ of x(n). Since $\mathbf{R}$ is symmetric, we can express it as

$$\mathbf{R} = \sum_{i=1}^{n} \lambda_i e_i e_i^T \qquad (2.1)$$

where $\sigma^2 = \lambda_1 = \lambda_2 = \ldots = \lambda_{N-2P} < \lambda_{N-2P+1} \leq \lambda_{N-2P+2} \leq \ldots \leq \lambda_N$ are the eigenvalues of $\mathbf{R}$ in the ascending order and e_i is the orthonormalized eigenvector of $\mathbf{R}$ corresponding to the eigenvalue λ_i (i = 1,...,N). The subspace spanned by the eigenvectors $e_{N-2P+1},\ldots,e_N$ is called the signal subspace (S_R) and that spanned by $e_1,\ldots,e_{N-2P}$ is called the noise subspace

* Indian Institute of Science, Bangalore, India
**Dept. of Elec. & Comp. Eng., Univ. of Iowa, Iowa City, IA 52242, U.S.A.
Supported by NSF Grant MIP-9001170

0-7803-0532-9/92 $3.00

(N_R) of **R**. Further, any linear combination of $e_1,...,e_{N-2P}$ is also an eigenvector of **R** with the eigenvalue σ^2 and in what follows we refer to this as a minimum eigenvector.

2.1 Computation of a Minimum Eigenvector of R

It is well known that a minimum eigenvector of **R** can be obtained as the solution of the following constrained minimization problem [2]:

$$\min_{w} w^T \mathbf{R} w \text{ subject to } w^T w = 1 \qquad (2.2)$$

where $w = [w_1, w_2, ..., w_N]^T$ is an N-dimensional vector.

In an earlier paper [9], we have recast the above constrained minimization problem (2.2) into an unconstrained minimization framework different from that used in [4]. In order to facilitate an easy recollection, we recall the results from [9] which are relevant to the present problem.

Define a function J as

$$J(w,\mu) = w^T \mathbf{R} w + \mu(w^T w - 1)^2; \mu > 0 \qquad (2.3)$$

Since **R** is a positive definite matrix and μ is positive, the function J is always positive. The main results are: i) If λ_{min} represents the minimum eigenvalue of **R**, then μ is bounded as $\mu > \lambda_{min}/2$. ii) All the stationary points of J lie within the unit hypersphere in the N-dimensional space. iii) w^* is a stationary point of J if and only if w^* is an eigenvector of **R** corresponding to the eigenvalue λ, with $\|w^*\|^2 = \beta^2 = 1 - \frac{\lambda}{2\mu}$ where $\|\cdot\|$ denotes the 2-norm. iv) w^* is a global minimizer of J if and only if w^* is a minimum eigenvector of **R** corresponding to the eigenvalue λ_{min}, with $\|w^*\| = \beta^2 = 1 - \frac{\lambda_{min}}{2\mu}$. v) For a given μ, every local minimizer of J is also a global minimizer. vi) The minimizer of J is unique only when N = 2P+1. vii) The signal eigenvectors of **R** correspond to the saddle points of J.

2.2 Multiple Eigenvectors Case

We now present an inflation approach to extend the results of Section 2.1 to compute multiple eigenvectors.

Let w_i^* be the eigenvector (with norm β_i) of **R** corresponding to the eigenvalue λ_i and assume that $w_i^*, ..., w_{k-1}^*$ ($2 \le k \le N$) are orthogonal. That is,

$$\mathbf{R} w_i^* = \lambda_i w_i^*, \quad i = 1,...,k-1$$

$$w_i^{*T} w_j^* = \qquad \forall\, i \ne j, \quad i,j \in \{1,...,k-1\}$$

Our objective is to obtain the next orthogonal eigenvector w_k^*, assuming that $w_i^*, i = 1,...,k-1$ are available.

Define the function J_k as

$$J_k(w_k,\mu,\alpha) = w_k^T \mathbf{R} w_k + \mu(w_k^T w_k - 1)^2 + \alpha \sum_{i=1}^{k-1} (w_k^T w_i^*)^2$$

$$= w_k^T \mathbf{R}_k w_k + \mu(w_k^T w_k - 1)^2 \qquad (2.4)$$

where $w_k \in R^N$ and $\alpha > 0$ and

$$\mathbf{R}_k = \mathbf{R} + \alpha \sum_{i=1}^{k-1} w_i^* w_i^{*T} \qquad (2.5)$$

Defining $e_i = \frac{w_i^*}{\beta_i}$, i = 1,...,k-1, we get from (2.1) and (2.5)

$$\mathbf{R}_k = \sum_{i=1}^{N} \lambda_i e_i e_i^T + \alpha \sum_{i=1}^{k-1} \beta_i^2 e_i e_i^T$$

$$= \sum_{i=1}^{N} \left[\lambda_i + \alpha \sum_{j=1}^{k-1} \beta_j^2 \delta_{i,j}\right] e_i e_i^T \qquad (2.6)$$

where, $\delta_{i,j}$ is the Kronecker delta function. Now, if α is chosen such that

$$\lambda_k < \lambda_i + \alpha \sum_{j=1}^{k-1} \beta_j^2 \delta_{i,j}, \quad i=1,...,k-1 \qquad (2.7)$$

then the minimum eigenvector of $\mathbf{R}_k$ is e_k and the corresponding eigenvalue is λ_k. So, if w_k^* is a minimizer of J_k, then from the results i) through vii) of Section 2.1 we get that w_k^* is a minimum eigenvector of $\mathbf{R}_k$ with eigenvalue λ_k. It can be easily verified from (2.6) that this minimum eigenvalue-eigenvector pair of $\mathbf{R}_k$ corresponds to the kth eigenvalue-eigenvector pair of **R**. That is,

$$\mathbf{R} w_k^* = \lambda_j w_k^* \text{ with } \|w_k^*\|^2 = 1 - \frac{\lambda k}{2\mu} \qquad (2.8)$$

Further, the strict inequality in (2.7) guarantees that

$$w_k^{*T} w_i^* = 0 \quad \forall \quad i=1,...,k-1. \qquad (2.9)$$

Thus, constructing N cost functions with

$$J_1(w_1,\mu,\alpha) = w_1^T \mathbf{R}_1 w_1 + \mu(w_1^T w1 - 1)^2, \quad \mathbf{R}_1 = \mathbf{R}$$

and J_k, k=2,...,N as defined in (2.4), and finding the minimizers of each of these, we get the N orthogonal eigenvectors of **R** and the corresponding eigenvalues. The procedure detailed above can be considered as the k^{th} step of an algorithm to compute the eigen-subspace of the covariance matrix. The norm of each eigenvector thus obtained depends upon the corresponding eigenvalue and the parameter μ. In order for the computation of all the orthogonal eigenvectors to become feasible, the parameters μ and α must satisfy the following conditions

$$\mu > \frac{\lambda_{max}}{2} \qquad (2.10)$$

which follows from the result i) of Section 2.1 where $\lambda_{max} = \lambda_N$, the maximum eigenvalue of **R**, and

$$\lambda_{max} < \lambda_i + \alpha\beta_i^2 \quad \text{for } i=1,...,N-1 \tag{2.11}$$

Substituting for β_i^2 in (2.11) and after some manipulations, we get

$$\alpha > \frac{\lambda_{max} - \lambda_{min}}{1 - \frac{\lambda_{min}}{2\mu}} \tag{2.12}$$

Since the eigenvalues of **R** are not known a priori, it is necessary to derive practically computable lower bounds for μ and α. Since

$$\lambda_{max} < \text{Trace}(\mathbf{R}) \text{ and } \lambda_{min} \leq \frac{\text{Trace}(\mathbf{R})}{N}$$

we get the practical values of the lower bounds for μ and α as

$$\mu > \frac{\text{Trace}(\mathbf{R})}{2} \tag{2.13}$$

$$\alpha > \frac{\text{Trace}(\mathbf{R})}{(1 - \frac{\text{Trace}(\mathbf{R})}{2\mu N})} \tag{2.14}$$

Thus, the selection of μ and α satisfying (2.13) and (2.14) guarantees the computation of all the orthogonal eigenvectors.

3. ADAPTIVE ESTIMATION OF THE EIGENSUBSPACE OF THE DATA COVARIANCE MATRIX

Let $a = [a_1, a_2, ..., a_N]^T$ represent the coefficient vector of a N-tap tapped delay line filter whose input is the sequence $\{x(n)\}$. Then the filter output is given by

$$y(n) = a^T x_n \tag{3.1}$$

Following the theory developed in Section 2, we define a least-squares criterion function for a block of L data samples as

$$J(a,\mu,L) = \frac{1}{L}\sum_{n=1}^{L} y^2(n) + \mu(a^T a - 1)^2$$
$$= a^T \mathbf{R}(L) a + \mu(a^T a - 1)^2 \tag{3.2}$$

where

$$x(n) = [x(n), x(n-1), ..., x(n-N+1)]^T$$
$$\mathbf{R}(L) = \frac{1}{L}\sum_{n=1}^{L} x_n x_n^T .$$

In the following sub-section, we derive the adaptive algorithm for seeking the minimum eigenvector of the data covariance matrix.

3.1 Single Eigenvector Case

Minimization of J has been chosen as the adaptation criterion for the filter. The Newton-type adaptive algorithm for minimizing J is of the form

$$a(n) = a(n-1) - \{[\mathbf{H}]^{-1} g\}\,|_{a=a(n-1)}$$
$$= a(n-1) - [\mathbf{H}(n-1)]^{-1} g(n-1) \tag{3.3}$$

where a(n) is the filter coefficient vector at the nth adaptation instant, g and H are the first derivative vector (gradient) and the second derivative matrix (Hessian), respectively, of J with respect to a. Differentiating (3.2) with respect to a, we get

$$g(n) = 2\mathbf{R}(n+1)a(n) + 4\mu(a^T(n)a(n) - 1)a(n) \tag{3.4}$$

$$\mathbf{H}(n) = 2\mathbf{R}(n+1) + \delta\mu(a(n)a^T(n)) + 4\mu(a^T(n)a(n) - 1)\mathbf{I}_N \tag{3.5}$$

Equation (3.3) implicitely assumes that the inverse of **H** (n) exists for all n. Further, in order to keep the computational requirements to a minimum it is preferable to have a recursive updating rule for computing the inverse of H(n). These two requirements are met if we approximate the Hessian by dropping the last term in (3.5). Using this approximation, we get the Gauss-Newton adaptive algorithm in place of (3.3) as

$$a(n) = a(n-1) - [\tilde{\mathbf{H}}(n-1)]^{-1} g(n-1) \tag{3.6}$$

where

$$[\tilde{\mathbf{H}}(n) = 2\mathbf{R}(n+1) + 8\mu a(n)a^T(n) \tag{3.7}$$

Applying the matrix inversion lemma to the R H S of (3.7), we get

$$[[\tilde{\mathbf{H}}(n)]^{-1} = \frac{1}{2}\,[\mathbf{R}^{-1}(n+1) - \frac{\mathbf{R}^{-1}(n+1)a(n)aT(n)\mathbf{R}^{-1}(n+1)}{\frac{1}{4\mu} + aT(n)\,\mathbf{R}^{-1}(n+1)a(n)}] \tag{3.8}$$

where $\mathbf{R}^{-1}(n)$ is given by

$$\mathbf{R}^{-1}(n) = \frac{n}{n-1}\,[\mathbf{R}^{-1}(n-1) - \frac{\mathbf{R}^{-1}(n-1)x_n x_n^T \mathbf{R}^{-1}(n-1)}{n-1 + x_n^T \mathbf{R}^{-1}(n-1)x_n}] \tag{3.9}$$

Substituting (3.8) and (3.4) in (3.6), and simplifying, we get

$$a(n) = 1(n-1)\mathbf{R}^{-1}(n)a(n-1) \tag{3.10}$$

where

$$1(n-1) = \frac{1 + a^T(n-1)a(n-1)}{\frac{1}{2\mu} + 2a^T(n-1)\mathbf{R}^{-1}(n)\,a(n-1)} \tag{3.11}$$

3.2 Extension to the Multiple Eigenvectors Case

Extension of the adaptive algorithm developed in Section 3.1 to the multiple eigenvectors case directly follows from the principle described in Section 2.2 Consider the unit k. Here, $a_k(n)$ is the estimate of the k^{th} eigenvector of **R**(n). The adaptation criterion for the k^{th} unit is the minimization of the least-squares cost function J_k defined as

$$J_k(a_k,\mu,L) = a_k^T \mathbf{R}_k(L) a_k + \mu(a_k^T a_k - 1)^2 \quad k=1,...,N \tag{3.12}$$

where

$$\mathbf{R}_k(l) = \mathbf{R}_k\text{-}1(L) + \alpha a_{k-1} a_{k-1}^T, \quad k=2,...,N \tag{3.13}$$

with

$$\mathbf{R}_1(L) = \mathbf{R}(L).$$

Then, the Gauss-Newton algorithm for updating $a_k(n-1)$ to $a_k(n)$ follows from (3.9) to (3.11) and can be given as

$$a_k(n) = 1_k(n-1)\mathbf{R}_k^{-1}(n)a_k(n-1), \text{ where} \tag{3.14}$$

$$1_k(n-1) = \frac{1 + a_k^T(n-1)a_k(n-1)}{\frac{1}{2\mu} + 2a_k^T(n-1)\mathbf{R}_\kappa^{-1}(n)a_k(n-1)}), \; k=1,...,N \tag{3.15}$$

$$\mathbf{R}_k^{-1}(n) = \mathbf{R}_{k-1}^{-1}(n) - \frac{\mathbf{R}_{k-1}^{-1}(n)\, a_{k-1}(n)\, a_{k-1}^T(n)\mathbf{R}_{k-1}^{-1}(n)}{\frac{1}{\alpha} + a_{k-1}^T(n)\mathbf{R}_{k-1}^{-1}(n)\, a_{k-1}(n)} \tag{3.16}$$

$$\mathbf{R}_1^{-1}(n) = \mathbf{R}^{-1}(n)$$

$$= \frac{n}{n-1}\left[R_1^{-1}(n-1) - \frac{R^{-1}(n-1)\,x_n x_n^T R^{-1}(n-1)}{n-1+x_n^T R^{-1}(n-1)x_n}\right] \quad (3.17)$$

As we did in Section 3.1, we comment that the recursive algorithm described by equations (3.14) to (3.17) converge to the k^{th} orthogonal eigenvector of the data covariance matrix.

The above described algorithm is sequential in nature since the k^{th} unit needs the eigenvector estimates from all the previous k-1 units, to update its estimate (see (3.12) and (3.13)). However, it can be made parallel by making all the units to start computations simultaneously as described below. At any instant n, each unit goes through a three step procedure. This is given below for the k^{th} unit

Step 1 Pass on the current weight vector $a_k(n-1)$ and the matrix $R_{k-1}^{-1}(n-1)$ to the next unit k + 1.

Step 2 Accept the weight vector $a_{k-1}(n-1)$ and the matrix $R_{k-1}^{-1}(n-1)$ from the just previous unit k - 1.

Step 3 Compute $R_k^{-1}(n)$ as

$$R_k^{-1}(n) = R_{k-1}^{-1}(n-1) - \frac{R_{k-1}^{-1}(n-1)a_{k-1}(n-1)a_{k-1}^T(n-1)R_{k-1}^{-1}(n-1)}{\frac{1}{\alpha} + a_{k-1}^T(n)R_{k-1}^{-1}(n)a_{k-1}(n)}$$

$k = 2,...,N$ and update $a_k(n-1)$ to $a_k(n)$ using (3.14).

Note from Step 3 that the data x(n) that enters unit 1 reaches unit k after a delay of k time units. Hence at instant n, even though x(1) to x(n) have been received by the first unit, only x(1) to x(n-k) have reached the kth unit. Consequently, there will be a similar delay in getting the final eigenvector estimates from each of these units.

It can be easily shown that adopting the above three-step procedure for parallelizing the Gauss-Newton approach results in the reduction of L . D - L- D time units where L is the data length and D is the number of eigenvectors to be estimated. Also, this algorithm is highly modular in nature since the computations at each unit are exactly identical to one another. However, inherent in the algorithm is the fact unit k will coverage only after all the previous units have converged.

CONCLUSIONS

The problem of computing the orthogonal eigenvectors of a symmetric positive definite matrix has been formulated as an unconstrained minization problem. The methodology evolved from the formulation has been made use of to develop a Gauss-Newton adaptive algorithm for estimating the orthogonal eigenvectors of the data covariance matrix. Unlike the conventional subspace estimation algorithms where the orthogonalization of the estimated eigenvectors is done explicitly, the proposed adaptive algorithm makes use of an implicit orthorthogonalization procedure which is built in through the cost function which the algorithm minimizes.

The algorithm thus resulted is not only suitable for parallel implementation but also highly modular in nature; thus resulting in a high-speed adaptive algorithm for subspace estimation. One can show analytically that the algorithm is convergent under mild assumption. Simulations confirm that its performance is exactly identical to the adaptive block inverse power method.

REFERENCES

[1] N.L. Owsley, "Adaptive Data Orthogonalization," in Proc. ICASSP, pp. 109-112, 1978.

[2] P.A. Thompson, "An Adaptive Spectral Analysis Technique for Unbiased Frequency Estimation in the Presence of White Noise," Proc. 13th Asilomar Conf., pp. 529-533, 1979.

[3] M.G. Larimore, "Adaptive Convergence of Spectral Estimation Based on Pisarenko's Harmonic Retrieval," IEEE ASSP, pp. 955-962, Aug. 1983.

[4] V.U. Reddy, B. Egardt and T. Kailath, "Least Squares Type Algorithm for Adaptive Implementation of Pisarenko's Harmonic Retrieval Method," IEEE Trans., pp. 399-405, June 1982.

[5] T.S. Durrani and K.C. Sharman, "Eigenfilter Approaches to Adaptive Array Processing," IEE Proc., pts. F and H, pp. 22-28, Feb. 1983.

[6] D.R. Fahrman and B. Liu, "Rotational Speech Methods for Adaptive Pisarenko Harmonic Retrieval," IEEE Trans. pp. 1550-1565, Dec. 1986.

[7] K.C. Sharman, "Adaptive Algorithms for Estimating the Complete Covariance Eigenstructure," in Proc. ICASSP, pp. 1401-1403, 1986.

[8] J.F. Yang and M. Kaveh, "Adaptive Eigensubspace Algorithms for Direction or Frequency Estimation and Tracking," IEEE Trans., pp. 241-251, Feb 1988.

[9] G. Mathew and V.U. Reddy, "Development and Analysis of a Neural Network Approach to Pisarenko's Harmonic Retrieval Method," Submitted to IEEE Trans. Signal Processing.

[10] Y.H. Hu, "Adaptive Methods for Real Time Pisarenko's Spectrum Estimate," Proc. ICASSP, pp. 105-108, 1985.