

# Orthogonal Eigensubspace Estimation Using Neural Networks

George Mathew and V. U. Reddy, *Senior Member, IEEE*

**Abstract**—In this paper, we present a neural network (NN) approach for simultaneously estimating all or some of the orthogonal eigenvectors of a symmetric nonindefinite matrix corresponding to its repeated minimum (in magnitude) eigenvalue. This problem has its origin in the constrained minimization framework and has extensive applications in signal processing. We recast this problem into the NN framework by constructing an appropriate energy function which the NN minimizes. The NN is of feedback type with the neurons having sigmoidal activation function. The proposed approach is analyzed to characterize the nature of the minimizers of the energy function. The main result is that “the matrix  $W^*$  is a minimizer of the energy function if and only if the columns of  $W^*$  are the orthogonal eigenvectors with a given norm corresponding to the smallest eigenvalue of the given matrix.” Further, all minimizers are global minimizers. Bounds on the integration time-step that is required to numerically solve the system of differential equations (which define the dynamics of the NN) have also been derived. Results of computer simulations are presented to support our analysis.

## I. INTRODUCTION

**E**STIMATION of orthogonal eigenvectors corresponding to the repeated smallest eigenvalue of a symmetric positive definite matrix is a problem of much importance in the area of signal processing. In this paper, we propose and analyze a neural network- (NN) based solution to this problem. Even though the development and analysis of this approach are given for the case of symmetric positive definite matrices, we show in Section III-C that it can be generalized to the case of a symmetric nonindefinite matrix.

Let  $\mathbf{R}$  denote an  $N \times N$  symmetric positive definite matrix. Suppose

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_P > \lambda_{P+1} = \lambda_{P+2} = \dots = \lambda_N \quad (1.1)$$

represent the  $N$  eigenvalues of  $\mathbf{R}$  in the decreasing order of magnitude and  $\mathbf{q}_i$ ,  $i = 1, \dots, N$ , the corresponding orthonormal eigenvectors. The eigenvectors corresponding to the minimum eigenvalue  $\lambda_{\min}$  (i.e.,  $\lambda_i$ ,  $i = P + 1, \dots, N$ ) are called minimum eigenvectors. Let  $\mathbf{S}$  be the space of minimum eigenvectors of  $\mathbf{R}$ . Then, the problem addressed in this paper is the estimation of an orthogonal basis for the subspace  $\mathbf{S}$ . For the case of narrowband signals in additive white noise, matrix  $\mathbf{R}$  represents the asymptotic covariance matrix and

Manuscript received August 28, 1992; revised September 17, 1993. The associate editor coordinating the review of this paper and approving it for publication was Prof. J. N. Hwang.

The authors are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore-560 012, India.

IEEE Log Number 9401291.

$\lambda_{\min}$  is the noise variance. As a result, the minimum eigenvalues and eigenvectors are known as noise eigenvalues and noise eigenvectors, respectively, and  $\mathbf{S}$  as the noise subspace. The remaining eigenvalues and eigenvectors are called signal eigenvalues and signal eigenvectors, respectively. Estimation of the noise/signal subspace is a primary requirement in many of the “superresolution” spectral estimation methods [22].

An adaptive approach for estimating the signal eigenvectors was first developed by Owsley [1]. Sharman [2] developed an adaptive algorithm, based on the QR-recursions, to estimate the complete eigenstructure of the sample covariance matrix. Yang and Kaveh [3] reported an adaptive approach for the estimation of complete noise or signal subspace. They also proposed an inflation technique for the estimation of noise subspace. Iterative algorithms based on steepest descent and conjugate gradient techniques for estimating the first few large/small eigenvalues and the corresponding eigenvectors of a Hermitian matrix were reported by Sarkar and Yang [4]. Recently, Mathew *et al.* [5] proposed a technique using a Newton type algorithm and an inflation approach for adaptively seeking the eigensubspace.

Under the topic “modified eigenvalue problem” many researchers have reported algorithms for estimating the eigenstructure of a modified covariance matrix given the prior knowledge of the eigenstructure of the unmodified covariance matrix. A fundamental work in this area is that of Bunch *et al.* [9]. Some of the very recent contributions are those of DeGroat and Roberts [10], Yu [11], Bischof and Shroff [12], and DeGroat [13]. DeGroat and Roberts [10] developed a parallel and numerically stable algorithm for rank-1 recursive updating of the eigenstructure. A parallel algorithm for rank-k recursive updating of the eigenstructure was reported by Yu [11]. Bischof and Shroff [12] reported an algorithm for updating the noise subspace using a rank-revealing QR-factorization. A noniterative and computationally inexpensive subspace updating algorithm was proposed by DeGroat [13].

Many NN-based algorithms have also been reported for estimation of multiple eigenvectors. All these algorithms are developed for feedforward NN structure with most of them using linear activation function. They estimate the principal eigenvectors (i.e., eigenvectors corresponding to the larger eigenvalues) of the covariance matrix of the sequence of input vectors. The problem of learning in a two (or more) layered NN by minimizing a quadratic error function was considered by Baldi and Hornik [14]. They proved that the error function has a unique minimum corresponding to the projection onto the subspace generated by the principal eigenvectors and

all other stationary points are saddle points. Sanger [15] proposed the generalized Hebbian algorithm for a single layer NN and showed its convergence to the principal eigenvectors. Kung and Diamantaras [16] proposed an approach (called APEX) for the recursive computation of principal eigenvectors. Their technique combines the Hebbian [18] and orthogonal learning rules. Kung [17] extended this approach to the case of constrained principal component analysis. In contrast to the algorithms in [14]–[18] which estimate principal eigenvectors, Xu *et al.* [19] proposed a method for estimating the eigenvector corresponding to the minimum eigenvalue using a modified anti-Hebbian learning rule. Very recently, Wang and Mendel [20] proposed a three-layered linear NN with certain constraints on the weights of neurons (called structured network) for solving the symmetric eigenvalue problem.

In this paper, unlike the above mentioned approaches, we present a feedback NN scheme for estimating orthogonal minimum eigenvectors of  $\mathbf{R}$  (i.e., an orthogonal basis for  $\mathbf{S}$ ). The proposed approach is developed in an unconstrained minimization framework by constructing an appropriate cost function which the NN minimizes. This approach is useful in applications where it is required to seek the (minimum) eigenvectors of a given symmetric positive definite matrix.

The paper is organized as follows. The NN formulation of the problem is presented in Section II. In Section III, analysis of the proposed approach is presented. Simulation results are presented in Section IV and finally, Section V concludes the paper.

## II. NEURAL NETWORK FORMULATION OF THE PROBLEM

Research in the area of artificial neural networks (ANN) is a result of the quest to build intelligent machines, drawing inspiration from the working mechanism of human brain. An ANN typically consists of many richly interconnected, simple and similar processing elements (called artificial neurons) operating in parallel. High computational rates and robustness (fault tolerance) are two important features which result from such an architecture of the ANN's [6]. In a feedback type ANN, the network dynamics are such that the stable stationary states of the network correspond to minima of its energy function. Hence, one of the most important applications of the feedback type ANN's is in solving optimization problems. By virtue of their architecture, they can provide fast and collectively computed solutions to difficult optimization problems [7].

In order to arrive at a feedback-type neural network (FBNN) scheme to solve the problem of estimating orthogonal minimum eigenvectors, we first need to construct an appropriate cost function whose minimizer corresponds to the required eigenvectors. To motivate this, we shall first take a brief look at the constrained minimization formulation of the same problem.

Let  $D$  be the dimension of the subspace  $\mathbf{S}$  where  $D = N - P$  (cf. (1.1)). Define an  $N \times M$  ( $M \leq D$ ) matrix  $\mathbf{W}$  as

$$\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_M] \quad (2.1)$$

where  $\mathbf{w}_k = [w_{k1}, \dots, w_{kN}]^T$ ,  $k = 1, \dots, M$ , are  $N \times 1$  vectors. Then, solution of the following constrained minimization problem

$$\min \sum_{i=1}^M \mathbf{w}_i^T \mathbf{R} \mathbf{w}_i$$

subject to

$$\mathbf{w}_i^T \mathbf{w}_j = \delta_{ij} \quad \forall i, j \in \{1, \dots, M\} \quad (2.2)$$

where  $\delta_{ij}$  is the Kronecker delta function, is the set of  $M$  orthonormal minimum eigenvectors of  $\mathbf{R}$  [3].

Now, if  $\bar{\mathbf{w}} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \dots, \mathbf{w}_M^T]^T$ , then the above minimization problem can be restated in terms of the  $NM \times 1$  vector  $\bar{\mathbf{w}}$  as

$$\min \bar{\mathbf{w}}^T \bar{\mathbf{R}} \bar{\mathbf{w}}$$

subject to

$$\bar{\mathbf{w}}^T \mathbf{E}_{ij} \bar{\mathbf{w}} = \delta_{ij} \quad \forall i, j \in \{1, \dots, M\} \quad (2.3)$$

where  $\bar{\mathbf{R}}$  is a block diagonal matrix of size  $MN \times MN$  defined as

$$\bar{\mathbf{R}} = \text{bdiag}[\mathbf{R}, \mathbf{R}, \dots, \mathbf{R}]. \quad (2.4)$$

The matrices  $\mathbf{E}_{i,j}$  are given by

$$\mathbf{E}_{i,j} = \mathbf{E}_i^T \mathbf{E}_j \quad \forall i, j \in \{1, \dots, M\} \quad (2.5)$$

where  $\mathbf{E}_i$  is an  $MN \times MN$  matrix with one in  $(k, (i-1)N + k)$ th location for  $k = 1, \dots, N$ , and zeros elsewhere. Note that  $\bar{\mathbf{w}}^T \mathbf{E}_{ij} \bar{\mathbf{w}} = \mathbf{w}_i^T \mathbf{w}_j \quad \forall i, j \in \{1, \dots, M\}$ .

Using the penalty function method [8], it is sufficient to solve the following unconstrained problem of the form

$$\min (\bar{\mathbf{w}}^T \bar{\mathbf{R}} \bar{\mathbf{w}} + \mu P(\bar{\mathbf{w}})) \quad (2.6)$$

where

$$P(\bar{\mathbf{w}}) = \sum_{i=1}^M (\bar{\mathbf{w}}^T \mathbf{E}_{ii} \bar{\mathbf{w}} - 1)^2 + \alpha' \sum_{i=1}^{M-1} \sum_{j=i+1}^M (\bar{\mathbf{w}}^T \mathbf{E}_{ij} \bar{\mathbf{w}})^2. \quad (2.7)$$

Here,  $\mu$  and  $\alpha'$  are positive constants and  $P$  is a penalty function such that 1)  $P$  is continuous, 2)  $P(\bar{\mathbf{w}}) \geq 0 \quad \forall \bar{\mathbf{w}}$ , and 3)  $P(\bar{\mathbf{w}}) = 0$  if and only if  $\bar{\mathbf{w}}$  belongs to the constraint space.

Let  $\{\mu_k\}$ ,  $k = 1, 2, 3, \dots$ , be a sequence tending to infinity such that  $\mu_k \geq 0$  and  $\mu_{k+1} > \mu_k$ . Further, let  $\bar{\mathbf{w}}_k$  be the minimizer of  $J(\bar{\mathbf{w}}, \mu_k)$  where  $J(\bar{\mathbf{w}}, \mu_k) = \bar{\mathbf{w}}^T \bar{\mathbf{R}} \bar{\mathbf{w}} + \mu_k P(\bar{\mathbf{w}})$ . Then, we have the following theorem from [8]:

*Theorem 2.1:* Limit point of the sequence  $\{\bar{\mathbf{w}}_k\}$  is a solution of (2.3).

That is, if  $\bar{\mathbf{w}}^* = [\mathbf{w}_1^{*T}, \dots, \mathbf{w}_M^{*T}]^T$  is the limit point of  $\{\bar{\mathbf{w}}_k\}$ , then the vectors  $\mathbf{w}_i^*$ ,  $i = 1, \dots, M$ , are orthonormal minimum eigenvectors of  $\mathbf{R}$ . However, because of special structure in the cost function (2.6), "a minimizer of  $J(\bar{\mathbf{w}}, \mu)$  is the set of orthogonal minimum eigenvectors of  $\mathbf{R}$  for any  $\mu > \frac{\lambda_{\min}}{2}$ " (proved in Section III-A). Further, the problem under consideration (i.e., estimation of  $\mathbf{S}$ ) does not demand the norm of the estimated eigenvectors to be unity.

We can rewrite the cost function in (2.6) as

$$J(\bar{\mathbf{w}}, \mu) = \sum_{i=1}^M \mathbf{w}_i^T \mathbf{R} \mathbf{w}_i + \mu \sum_{i=1}^M (\mathbf{w}_i^T \mathbf{w}_i - 1)^2 + \alpha \sum_{i=1}^{M-1} \sum_{j=i+1}^M (\mathbf{w}_i^T \mathbf{w}_j)^2 \quad (2.8)$$

where  $\alpha = \mu\alpha'$ . Our aim is to design a feedback NN for which  $J$  could be the energy function. It must have  $MN$  neurons with the output of each neuron,  $w_{ij}$ , representing one element of the matrix  $\mathbf{W}$ . The activation function of each neuron is assumed to be the sigmoidal nonlinearity, i.e.,

$$w_{ij}(t) = f(u_{ij}(t)) = \frac{2}{1 + \exp(-u_{ij}(t))} - 1 \quad (2.9)$$

$i = 1, \dots, M; \quad j = 1, \dots, N$

where  $u_{ij}(t)$  and  $w_{ij}(t)$  are the net input and output, respectively, of the  $(i, j)$ th neuron at time  $t$ . The network dynamics should be such that the time derivative of  $J$  is negative. For the cost function (2.8), the time derivative is given by

$$\frac{dJ}{dt} = \sum_{i=1}^M \sum_{j=1}^N \frac{\partial J}{\partial w_{ij}(t)} \cdot f'(u_{ij}(t)) \cdot \frac{du_{ij}(t)}{dt} \quad (2.10)$$

where  $f'(u)$  is the derivative of  $f(u)$  with respect to  $u$ .

Now, let us define the dynamics (i.e., the updating rule) of the  $(i, j)$ th neuron as

$$\begin{aligned} \frac{du_{ij}(t)}{dt} &= -\frac{\partial J}{\partial w_{ij}(t)} \\ &= -2 \sum_{k=1}^N R_{jk} w_{ik}(t) \\ &\quad - 4\mu (\mathbf{w}_i^T(t) \mathbf{w}_i(t) - 1) w_{ij}(t) \\ &\quad - 2\alpha \sum_{\substack{p=1 \\ p \neq i}}^M (\mathbf{w}_i^T(t) \mathbf{w}_p(t)) w_{pj}(t) \end{aligned} \quad (2.11)$$

with the input-output transformation given by (2.9). Fig. 1 shows the model of  $(i, j)$ th neuron.

In vector notation, (2.11) can be expressed as

$$\begin{aligned} \frac{d\mathbf{u}_i(t)}{dt} &= -2\mathbf{R}\mathbf{w}_i(t) - 4\mu(\mathbf{w}_i^T(t)\mathbf{w}_i(t) - 1)\mathbf{w}_i(t) \\ &\quad - 2\alpha \sum_{\substack{p=1 \\ p \neq i}}^M (\mathbf{w}_i^T(t)\mathbf{w}_p(t))\mathbf{w}_p(t) \end{aligned} \quad (2.12)$$

$i = 1, \dots, M$

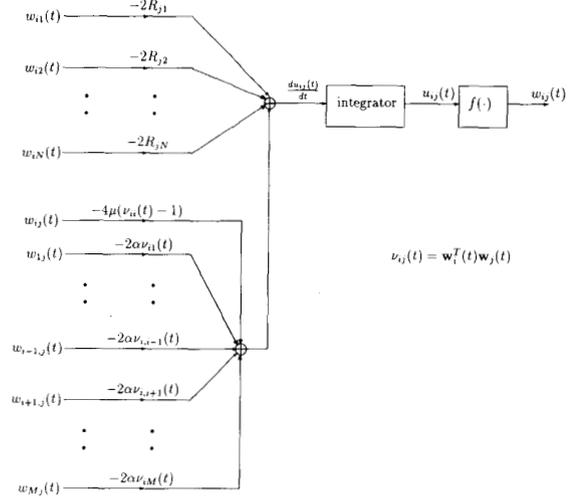


Fig. 1. Model of  $(i, j)$ th neuron.

where  $\mathbf{u}_i(t) = [u_{i1}(t), \dots, u_{iN}(t)]^T$ . Substituting (2.11) in (2.10) and using the fact that  $f(u)$  is a monotonically increasing function, we get

$$\frac{dJ}{dt} < 0 \quad \text{if} \quad \frac{du_{ij}(t)}{dt} \neq 0$$

for at least one  $i$  and  $j$

and

$$\frac{dJ}{dt} = 0 \quad \text{if and only if} \quad \frac{du_{ij}(t)}{dt} = 0$$

for all  $i$  and  $j$ . (2.13)

Equation (2.13) implies that the FBNN with dynamics defined by (2.11) and (2.9) has stable stationary points at the local minima of  $J$ . The analysis of Section III shows that every minimizer of  $J$  corresponds to a subset of an orthogonal basis for the subspace  $\mathbf{S}$  and all minimizers are global minimizers.

We would now like to comment on the similarity between the updating equation (2.11) and some of the existing algorithms for eigenvector estimation. The first two terms in (2.11) are similar to the Hebbian rule of Oja [18]. In the generalized Hebbian algorithm of Sanger [15], if the lower triangular operation is removed and the terms are grouped together appropriately, the resulting algorithm resembles (2.11). The work of Brockett [23] is similar to our approach in the sense that he showed how the symmetric eigenvalue problem can be solved using a system of differential equations that describe certain gradient flow on the space of orthogonal matrices. The system is initialized using the given symmetric matrix and it converges to a diagonal matrix with the eigenvalues as the diagonal elements.

### III. ANALYSIS OF THE PROPOSED APPROACH

In this section, we present an analysis of the proposed FBNN approach. This is primarily aimed at establishing the correspondence between the minimizers of  $J$  and orthogonal

bases of  $\mathbf{S}$ . We also derive bounds on the integration time-step ( $h$ ) that is used in the numerical solution of the system dynamics (cf. (2.11), (2.9)).

Combining (2.6) and (2.7), we get

$$J(\bar{\mathbf{w}}, \mu) = \bar{\mathbf{w}}^T \bar{\mathbf{R}} \bar{\mathbf{w}} + \mu \sum_{i=1}^M (\bar{\mathbf{w}}^T \mathbf{E}_{ii} \bar{\mathbf{w}} - 1)^2 + \alpha \sum_{i=1}^{M-1} \sum_{j=i+1}^M (\bar{\mathbf{w}}^T \mathbf{E}_{ij} \bar{\mathbf{w}})^2. \quad (3.1)$$

In the following analysis, we assume that the parameters  $\mu$  and  $\alpha$  are fixed at some appropriately chosen values. Section III-B contains guidelines for choosing the values for  $\mu$  and  $\alpha$ . The gradient vector and Hessian matrix, respectively, of  $J$  with respect to  $\mathbf{w}_k$  are given by

$$\mathbf{g}_k(\bar{\mathbf{w}}) = 2\mathbf{R}\mathbf{w}_k + 4\mu(\mathbf{w}_k^T \mathbf{w}_k - 1)\mathbf{w}_k + 2\alpha \sum_{\substack{i=1 \\ i \neq k}}^M \mathbf{w}_i \mathbf{w}_i^T \mathbf{w}_k \quad (3.2)$$

$$\mathbf{H}_k(\bar{\mathbf{w}}) = 2\mathbf{R} + 8\mu\mathbf{w}_k \mathbf{w}_k^T + 4\mu(\mathbf{w}_k^T \mathbf{w}_k - 1)\mathbf{I}_N + 2\alpha \sum_{\substack{i=1 \\ i \neq k}}^M \mathbf{w}_i \mathbf{w}_i^T. \quad (3.3)$$

If  $\mathbf{g}(\bar{\mathbf{w}})$  and  $\mathbf{H}(\bar{\mathbf{w}})$  denote the gradient and Hessian, respectively, of  $J$  with respect to  $\bar{\mathbf{w}}$ , then  $\mathbf{g}_k(\bar{\mathbf{w}})$  is the  $k$ th  $N$ -vector segment of  $\mathbf{g}(\bar{\mathbf{w}})$  and  $\mathbf{H}_k(\bar{\mathbf{w}})$  is the  $N \times N$  principal submatrix of  $\mathbf{H}(\bar{\mathbf{w}})$  beginning at  $((k-1)N+1, (k-1)N+1)$ th location.

#### A. Correspondence Between the Minimizers of $J$ and Orthogonal Bases of $\mathbf{S}$

In this section, we prove the main result of the paper, i.e., correspondence between the minimizers of  $J$  and orthogonal bases of the subspace  $\mathbf{S}$ . Because of the use of sigmoidal nonlinearity, the space in which  $\bar{\mathbf{w}}$  lies is the  $NM$ -dimensional unit open hypercube which we denote by  $\mathbf{T}$ . Further, it is evident that  $\mathbf{g}(\bar{\mathbf{w}}) = 0$  if and only if  $\mathbf{g}_k(\bar{\mathbf{w}}) = 0 \forall k = 1, \dots, M$ . We have the following results.

*Theorem 3.1:*  $\bar{\mathbf{w}}^* = [\mathbf{w}_1^{*T}, \dots, \mathbf{w}_M^{*T}]^T$  is a stationary point of  $J$  if and only if  $\mathbf{w}_k^*$  is an eigenvector, with Euclidean norm  $\beta_k$ , of the matrix  $\mathbf{A} = \mathbf{R} + \alpha \sum_{i=1}^M \mathbf{w}_i^* \mathbf{w}_i^{*T}$  corresponding to the eigenvalue  $a_k$  where  $a_k = 2\mu(1 - \beta_k^2) + \alpha\beta_k^2$  for  $k = 1, \dots, M$ .

This result immediately follows from (3.2).

*Theorem 3.2:* Let  $\bar{\mathbf{w}}^*$  be a stationary point of  $J$ . Then, the vectors  $\mathbf{w}_1^*, \dots, \mathbf{w}_M^*$  are orthogonal if and only if  $\mathbf{w}_k^*$  is an eigenvector, with norm  $\beta_k$ , of  $\mathbf{R}$  corresponding to the eigenvalue  $2\mu(1 - \beta_k^2) \in \{\lambda_1, \dots, \lambda_N\}$  for  $k = 1, \dots, M$ .

This result also can be easily arrived at using (3.2).

*Remark:* Theorems 3.1 and 3.2 are valid even if  $D < M \leq N$ .

*Theorem 3.3:*  $\bar{\mathbf{w}}^*$  is a global minimizer of  $J$  if and only if  $\mathbf{w}_1^*, \dots, \mathbf{w}_M^*$  form the subset of an orthogonal basis for  $\mathbf{S}$  with  $\beta_k^2 = 1 - \frac{\lambda_{\min}}{2\mu}$  for  $k = 1, \dots, M$ .

*Proof:* If part: From the hypothesis, we have for all  $k, j \in \{1, \dots, M\}$

$$\mathbf{R}\mathbf{w}_k^* = \lambda_{\min} \mathbf{w}_k^*, \quad \lambda_{\min} = 2\mu(1 - \beta_k^2)$$

and

$$\mathbf{w}_k^{*T} \mathbf{w}_j^* = \beta_k^2 \delta_{kj}. \quad (3.4)$$

Substituting (3.4) in (3.2) gives  $\mathbf{g}_k(\bar{\mathbf{w}}^*) = 0$ ,  $k = 1, \dots, M$ , implying that  $\bar{\mathbf{w}}^*$  is a stationary point of  $J$ . To prove that  $\bar{\mathbf{w}}^*$  is a global minimizer of  $J$ , we need to show that  $J(\mathbf{x}, \mu) \geq J(\bar{\mathbf{w}}^*, \mu) \forall \mathbf{x} \in \mathbf{T}$ . Let  $\mathbf{x} = \bar{\mathbf{w}}^* + \mathbf{p}$  where  $\mathbf{p} \in \mathbf{T}$ . From (3.4), we obtain

$$\mathbf{R}\bar{\mathbf{w}}^* = \lambda_{\min} \bar{\mathbf{w}}^*. \quad (3.5)$$

Evaluating  $J$  at  $\mathbf{x}$  and making use of (3.5), we get

$$\begin{aligned} J(\mathbf{x}, \mu) &= J(\bar{\mathbf{w}}^*, \mu) + 2\lambda_{\min} \mathbf{p}^T \bar{\mathbf{w}}^* \\ &\quad - \lambda_{\min} \sum_{i=1}^M [2\mathbf{p}^T \mathbf{E}_{ii} \bar{\mathbf{w}}^* + \mathbf{p}^T \mathbf{E}_{ii} \mathbf{p}] \\ &\quad + \alpha \sum_{i=1}^{M-1} \sum_{j=i+1}^M [\bar{\mathbf{w}}^{*T} \mathbf{E}_{ij} \mathbf{p} + \mathbf{p}^T \mathbf{E}_{ij} \bar{\mathbf{w}}^* \\ &\quad + \mathbf{p}^T \mathbf{E}_{ij} \mathbf{p}]^2 + \mathbf{p}^T \bar{\mathbf{R}} \mathbf{p} \\ &\quad + \mu \sum_{i=1}^M [2\mathbf{p}^T \mathbf{E}_{ii} \bar{\mathbf{w}}^* + \mathbf{p}^T \mathbf{E}_{ii} \mathbf{p}]^2. \end{aligned} \quad (3.6)$$

Note that

$$\sum_{i=1}^M \mathbf{p}^T \mathbf{E}_{ii} \bar{\mathbf{w}}^* = \mathbf{p}^T \bar{\mathbf{w}}^*, \quad \sum_{i=1}^M \mathbf{p}^T \mathbf{E}_{ii} \mathbf{p} = \mathbf{p}^T \mathbf{p}$$

and

$$\mathbf{p}^T \bar{\mathbf{R}} \mathbf{p} \geq \lambda_{\min} \mathbf{p}^T \mathbf{p}. \quad (3.7)$$

Substituting (3.7) in (3.6) and simplifying, we get

$$\begin{aligned} J(\mathbf{x}, \mu) &\geq J(\bar{\mathbf{w}}^*, \mu) + \mu \sum_{i=1}^M [2\mathbf{p}^T \mathbf{E}_{ii} \bar{\mathbf{w}}^* + \mathbf{p}^T \mathbf{E}_{ii} \mathbf{p}]^2 \\ &\quad + \alpha \sum_{i=1}^{M-1} \sum_{j=i+1}^M [\bar{\mathbf{w}}^{*T} \mathbf{E}_{ij} \mathbf{p} + \mathbf{p}^T \mathbf{E}_{ij} \bar{\mathbf{w}}^* \\ &\quad + \mathbf{p}^T \mathbf{E}_{ij} \mathbf{p}]^2 \\ &\geq J(\bar{\mathbf{w}}^*, \mu) \forall \mathbf{x} \in \mathbf{T}. \end{aligned}$$

Thus,  $\bar{\mathbf{w}}^*$  is a global minimizer of  $J$ .

Only if part: From the hypothesis,  $\bar{\mathbf{w}}^*$  is a stationary point. Hence, we get from (3.2) (after some manipulation)

$$\left( \mathbf{R} + \alpha \sum_{i=1}^M \mathbf{w}_i^* \mathbf{w}_i^{*T} \right) \mathbf{w}_k^* = [2\mu(1 - \beta_k^2) + \alpha\beta_k^2] \mathbf{w}_k^* \quad \forall k = 1, \dots, M \quad (3.8)$$

$$2\mu(1 - \beta_k^2) \geq \lambda_{\min} + \alpha \sum_{\substack{i=1 \\ i \neq k}}^M \frac{(\mathbf{w}_k^{*T} \mathbf{w}_i^*)^2}{\beta_k^2} \quad \forall k = 1, \dots, M. \quad (3.9)$$

We denote

$$\mathbf{A} = \mathbf{R} + \mathbf{V} \text{ with } \mathbf{V} = \alpha \sum_{i=1}^M \mathbf{w}_i^* \mathbf{w}_i^{*T}. \quad (3.10)$$

It is easy to see from (3.8) that  $\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{A}$ . Further,  $\mathbf{A}$  and  $\mathbf{V}$  are symmetric matrices. Hence, there exists an  $N \times N$  orthonormal matrix [21]  $\tilde{\mathbf{Q}} = [\tilde{\mathbf{q}}_1, \dots, \tilde{\mathbf{q}}_N]$  such that  $\tilde{\mathbf{Q}}$  simultaneously diagonalizes  $\mathbf{A}$  and  $\mathbf{V}$ . Since  $\mathbf{R} = \mathbf{A} - \mathbf{V}$ , it follows that  $\mathbf{R}$  is also diagonalizable by  $\tilde{\mathbf{Q}}$ . Hence, we get

$$\mathbf{A} = \sum_{i=1}^N a_i \tilde{\mathbf{q}}_i \tilde{\mathbf{q}}_i^T, \quad \mathbf{R} = \sum_{i=1}^N r_i \tilde{\mathbf{q}}_i \tilde{\mathbf{q}}_i^T$$

and

$$\mathbf{V} = \sum_{i \in S_v} v_i \tilde{\mathbf{q}}_i \tilde{\mathbf{q}}_i^T \quad (3.11)$$

where  $a_i, r_i$  and  $v_i$  are the eigenvalues of  $\mathbf{A}$ ,  $\mathbf{R}$  and  $\mathbf{V}$ , respectively, and  $S_v$  is defined as  $S_v = \{i : v_i > 0\}$ . Substituting (3.11) in (3.3) and rearranging, we obtain

$$\begin{aligned} \mathbf{H}_k(\bar{\mathbf{w}}^*) &= \sum_{i \notin S_v} 2[r_i - 2\mu(1 - \beta_k^2)] \tilde{\mathbf{q}}_i \tilde{\mathbf{q}}_i^T \\ &+ \sum_{i \in S_v} 2[r_i + v_i - 2\mu(1 - \beta_k^2)] \tilde{\mathbf{q}}_i \tilde{\mathbf{q}}_i^T \\ &+ 2(4\mu - \alpha) \mathbf{w}_k^* \mathbf{w}_k^{*T}. \end{aligned} \quad (3.12)$$

Now, define the following subspaces

$$\mathbf{V}_2 = \text{Span}\{\tilde{\mathbf{q}}_i : i \in S_v\}$$

and

$$\mathbf{V}_1 = \mathbf{V}_2^\perp. \quad (3.13)$$

Clearly, range of  $\mathbf{V}$  is same as  $\mathbf{V}_2$ . Hence, we can rewrite (3.12) as

$$\begin{aligned} \mathbf{H}_k(\bar{\mathbf{w}}^*) &= 2 \sum_{i \notin S_v} [r_i - 2\mu(1 - \beta_k^2)] \tilde{\mathbf{q}}_i \tilde{\mathbf{q}}_i^T \\ &+ 2 \sum_{i=1}^{r(\mathbf{V})} \mathbf{g}_i \tilde{\mathbf{q}}_i \tilde{\mathbf{q}}_i^T \end{aligned} \quad (3.14)$$

where the second term in (3.14) is an eigenvalue decomposition of the last two terms in (3.12) and  $r(\mathbf{V})$  is the rank of  $\mathbf{V}$ . The vectors  $\tilde{\mathbf{q}}_i, i = 1, \dots, r(\mathbf{V})$ , constitute an orthonormal basis for  $\mathbf{V}_2$ . Thus, R.H.S of (3.14) is an eigenvalue decomposition of  $\mathbf{H}_k(\bar{\mathbf{w}}^*)$ . Since  $\bar{\mathbf{w}}^*$  is a global minimizer of  $J$ , the Hessian  $\mathbf{H}(\bar{\mathbf{w}}^*)$  should be at least positive semidefinite. Hence,  $\mathbf{H}_k(\bar{\mathbf{w}}^*)$  also should be positive semidefinite since it is a principal submatrix of  $\mathbf{H}(\bar{\mathbf{w}}^*)$ . Thus, we get

$$r_i \geq 2\mu(1 - \beta_k^2) \quad \forall i \notin S_v, k = 1, \dots, M \quad (3.15)$$

$$g_i \geq 0 \quad \forall i = 1, \dots, r(\mathbf{V}). \quad (3.16)$$

Our goal now is to prove the following points:

- i)  $\mathbf{V}_2 \subseteq \mathbf{S}$  and
- ii)  $\mathbf{w}_i^*, i = 1, \dots, M$  are orthogonal.

Consider the following two cases.

*Case 1*  $M < D$ : Since  $M < D$ , we have  $(\mathbf{V}_1 \cap \mathbf{S}) \setminus \{\mathbf{0}\} \neq \emptyset$ . This implies that  $\exists l \notin S_v$  such that

$$\tilde{\mathbf{q}}_l \in \mathbf{V}_1 \cap \mathbf{S} \text{ and } \mathbf{R}\tilde{\mathbf{q}}_l = \lambda_{\min} \tilde{\mathbf{q}}_l$$

which in view of the second term of (3.11) gives  $r_l = \lambda_{\min}$ . Substituting this in (3.15) with  $i = l$ , we get

$$r_l = \lambda_{\min} \geq 2\mu(1 - \beta_k^2) \quad \forall k = 1, \dots, M. \quad (3.17)$$

For the inequalities of (3.9) and (3.17) to be true simultaneously, we must have

$$\begin{aligned} \mathbf{w}_k^{*T} \mathbf{w}_i^* &= 0 \text{ and } \lambda_{\min} = 2\mu(1 - \beta_k^2) \\ &\forall k, i = 1, \dots, M; k \neq i. \end{aligned} \quad (3.18)$$

Combining (3.18) with (3.8), we get  $\mathbf{w}_i^*, i = 1, \dots, M$ , as orthogonal minimum eigenvectors of  $\mathbf{R}$ , thus proving (i) and (ii).

*Case 2*  $M = D$ : Since  $M = D$ , we need to establish  $\mathbf{V}_2 = \mathbf{S}$  (or  $\mathbf{V}_1 = \mathbf{S}^\perp$ ) in order to prove (i). We show this by contradiction. Suppose

$$(\mathbf{V}_2 \cap \mathbf{S}^\perp) \setminus \{\mathbf{0}\} \neq \emptyset \quad (3.19)$$

implying that  $(\mathbf{V}_1 \cap \mathbf{S}) \setminus \{\mathbf{0}\} \neq \emptyset$ . Then, following the steps as in Case 1, we get

$$\begin{aligned} \exists l \notin S_v \text{ such that } r_l &= \lambda_{\min} \geq 2\mu(1 - \beta_k^2) \\ &\forall k = 1, \dots, D. \end{aligned} \quad (3.20)$$

It is then possible to choose  $k = k_1 \in S_v$  such that  $\mathbf{w}_{k_1}^* \cap \mathbf{S}^\perp \neq \emptyset$ . But, for the inequalities (3.9) and (3.20) to be true simultaneously for  $k = k_1$ , we must have

$$\begin{aligned} \mathbf{w}_{k_1}^{*T} \mathbf{w}_i^* &= 0 \text{ and } \lambda_{\min} = 2\mu(1 - \beta_{k_1}^2) \\ &\forall i = 1, \dots, D; i \neq k_1. \end{aligned} \quad (3.21)$$

Combining (3.21) with (3.8), we get

$$\mathbf{R}\mathbf{w}_{k_1}^* = 2\mu(1 - \beta_{k_1}^2) \mathbf{w}_{k_1}^* = \lambda_{\min} \mathbf{w}_{k_1}^* \quad (3.22)$$

which is a contradiction since any eigenvector of  $\mathbf{R}$  in  $\mathbf{S}^\perp$  must correspond to a nonminimum eigenvalue. Hence,  $\mathbf{V}_1 \cap \mathbf{S} = \{\mathbf{0}\}$  and  $\mathbf{V}_2 \cap \mathbf{S}^\perp = \{\mathbf{0}\}$ . This means that  $\mathbf{V}_1 \subset \mathbf{S}^\perp$  and  $\mathbf{V}_2 \subset \mathbf{S}$ . This, combined with the facts that  $\mathbf{V}_1 = \mathbf{V}_2^\perp, \mathbf{V}_1 \oplus \mathbf{V}_2 = \mathbb{R}^N$  and  $\mathbf{S} \oplus \mathbf{S}^\perp = \mathbb{R}^N$ , where ‘ $\oplus$ ’ stands for ‘direct sum’ of subspaces, gives  $\mathbf{V}_2 = \mathbf{S}$ . Hence,  $\{\mathbf{w}_1^*, \dots, \mathbf{w}_D^*\}$  is a basis for  $\mathbf{S}$ . Using this result in (3.8) and rearranging, we obtain

$$[\lambda_{\min} - 2\mu(1 - \beta_k^2)] \mathbf{w}_k^* + \alpha \sum_{\substack{i=1 \\ i \neq k}}^D (\mathbf{w}_k^{*T} \mathbf{w}_i^*) \mathbf{w}_i^* = \mathbf{0}. \quad (3.23)$$

Since  $\mathbf{w}_1^*, \dots, \mathbf{w}_D^*$  are linearly independent, it follows from (3.23) that

$$\begin{aligned} \lambda_{\min} - 2\mu(1 - \beta_k^2) &= 0 \text{ and } \mathbf{w}_k^{*T} \mathbf{w}_i^* = 0 \\ &\forall k, i \in \{1, \dots, D\}, \quad k \neq i \end{aligned} \quad (3.24)$$

thus establishing (i) and (ii).

The following corollaries bring out some significant aspects of Theorem 3.3.

*Corollary 3.1:* The value of  $\mu$  should be greater than  $\frac{\lambda_{\min}}{2}$ .

*Corollary 3.2:* For a given  $\mu$  and  $\alpha$ , every local minimizer of  $J$  is also a global minimizer.

This result follows from Theorem 3.3 by recognizing that  $\mathbf{H}(\bar{\mathbf{w}}^*)$  is at least positive semidefinite when  $\bar{\mathbf{w}}^*$  is a local minimizer of  $J$ .

*Corollary 3.3:* Minimizer of  $J$  is unique (except for the sign) only when  $D = 1$ .

### B. Discussion

Theorems 3.1 and 3.2 characterize all stationary points of  $J$ . While Theorem 3.1 gives a general characterization, Theorem 3.2 establishes the condition of orthogonality for the set of vectors under consideration. However, the main result is that of Theorem 3.3 which along with Corollary 3.2 shows that it is enough to find a minimizer of  $J$  in order to get an orthogonal basis for  $\mathbf{S}$ . Since all minimizers are global minimizers, any simple search technique (e.g., gradient descent) will suffice to reach the correct answer. This is very significant since nonlinear optimization problems usually encounter difficulties due to local minima.

Note from Theorem 3.3 that the norms of all the estimated eigenvectors are same and it is decided by the values of  $\mu$  and  $\lambda_{\min}$ . The higher the value of  $\mu$ , the closer will be the norm to unity. According to Corollary 3.1, *a priori* knowledge of the minimum eigenvalue of  $\mathbf{R}$  is required for the selection of  $\mu$ . But in practice,  $\lambda_{\min}$  will not be known *a priori* and hence we suggest the following practical lower bound

$$\mu > \frac{\text{Trace}(\mathbf{R})}{2N}. \quad (3.25)$$

Note that the second parameter  $\alpha$  can be any positive scalar.

### C. Generalization

The development and analysis of the proposed approach given in the previous sections assume the given matrix to be symmetric and positive definite. However, it can be extended to the case of any symmetric matrix (as long as it is not indefinite) as shown below. Let  $\mathbf{X}$  be the given  $N \times N$  matrix. Then, construct the matrix  $\mathbf{R}$  as

- 1)  $\mathbf{R} = -\mathbf{X}$  if  $\mathbf{X}$  is negative definite
- 2)  $\mathbf{R} = -\mathbf{X} + \eta\mathbf{I}_N$  if  $\mathbf{X}$  is negative semidefinite
- 3)  $\mathbf{R} = \mathbf{X} + \eta\mathbf{I}_N$  if  $\mathbf{X}$  is positive semidefinite

where  $\eta > 0$ . The proposed approach can now be applied to  $\mathbf{R}$  since it is symmetric and positive definite, and it yields orthogonal eigenvectors corresponding to the repeated smallest magnitude eigenvalue of  $\mathbf{X}$ .

### D. Bounds on the Integration Time-Step

The system of  $NM$  ordinary (nonlinear) differential equations, which describe the dynamics of the proposed FBNN, is given by (cf. (2.11), (2.9))

$$\frac{d\bar{\mathbf{u}}(t)}{dt} = -\nabla_{\bar{\mathbf{w}}} J(\bar{\mathbf{w}}(t), \mu) = -2\mathbf{B}(t)\bar{\mathbf{w}}(t) \quad (3.26)$$

where

$$\begin{aligned} \mathbf{B}(t) = & \bar{\mathbf{R}} + 2\mu \sum_{i=1}^M (\bar{\mathbf{w}}^T(t)\mathbf{E}_{ii}\bar{\mathbf{w}}(t) - 1)\mathbf{E}_{ii} \\ & + \alpha \sum_{i=1}^{M-1} \sum_{j=i+1}^M (\bar{\mathbf{w}}^T(t)\mathbf{E}_{ij}\bar{\mathbf{w}}(t)) \cdot (\mathbf{E}_{ij} + \mathbf{E}_{ji}) \end{aligned} \quad (3.27)$$

$$\bar{w}_k(t) = f(\bar{u}_k(t)) = \frac{2}{1 + \exp(-\bar{u}_k(t))} - 1 \quad \forall k = 1, \dots, NM \quad (3.28)$$

with  $\bar{u}_k(t) = u_{ij}(t)$  and  $\bar{w}_k(t) = w_{ij}(t)$  denoting the  $k$ th elements of  $\bar{\mathbf{u}}(t)$  and  $\bar{\mathbf{w}}(t)$ , respectively, for  $k = (i-1)N + j$ , and  $\bar{\mathbf{u}}(t) = [\mathbf{u}_1^T(t), \dots, \mathbf{u}_M^T(t)]^T$ . To solve this system, we resort to some numerical technique and this calls for the selection of an appropriate integration time-step, say  $h$ , in order to guarantee convergence to the correct solution. Below, we present an approximate analysis of the discretized system of equations to obtain the bounds for  $h$ .

For sufficiently small  $h$ , we have the following approximation

$$\left. \frac{d\bar{\mathbf{u}}(t)}{dt} \right|_{t=nh} \approx \frac{\bar{\mathbf{u}}(n+1) - \bar{\mathbf{u}}(n)}{h}$$

where  $n$  is the discrete time index. Substituting this in (3.26), we get

$$\bar{\mathbf{u}}(n+1) \approx \bar{\mathbf{u}}(n) - 2h\mathbf{B}(n)\bar{\mathbf{w}}(n). \quad (3.29)$$

Observe that (3.29) and (3.28) implement the discrete-time gradient descent with step-size  $h$ . Since  $J$  has only global minimizers, this search technique will reach a minimizer provided  $h$  is sufficiently small. Hence, we assume that the trial solution  $\bar{\mathbf{w}}(n)$  is very close to the desired solution, say

$$\bar{\mathbf{w}}^* = [\mathbf{w}_1^{*T}, \dots, \mathbf{w}_M^{*T}]^T$$

for  $n > K$  where  $K$  is a large enough positive integer. Since  $\mathbf{w}_i^{*T} \mathbf{w}_j^* = \beta_i^2 \delta_{ij}$  for all  $i, j \in \{1, \dots, M\}$  (cf. Theorem 3.3), we also assume that the vectors  $\mathbf{w}_i(n), i = 1, \dots, M$ , which constitute the trial solution  $\bar{\mathbf{w}}(n)$ , are approximately orthogonal and that the norm of each of them remains approximately constant at  $\beta_i$  for  $n > K$ . Using these approximations in (3.29), we get

$$\bar{\mathbf{u}}(n+1) \approx \bar{\mathbf{u}}(n) - 2h\mathbf{B}\bar{\mathbf{w}}(n) \quad \forall n > K \quad (3.30)$$

where  $\mathbf{B} = \bar{\mathbf{R}} + 2\mu \sum_{i=1}^M (\beta_i^2 - 1)\mathbf{E}_{ii}$ .

Now, writing the Taylor's series expansion for  $f(u)$  around  $u = 0$  and neglecting terms of order 5 and above, we obtain

$$w = f(u) \approx \frac{u}{2} - \frac{u^3}{24}. \quad (3.31)$$

Substituting (3.31) in (3.30) gives

$$\bar{\mathbf{u}}(n+1) \approx \mathbf{C}\bar{\mathbf{u}}(n) + \frac{h}{12}\mathbf{B}\bar{\mathbf{u}}(n) \quad (3.32)$$

where  $\mathbf{C} = \mathbf{I}_{NM} - h\mathbf{B}$  and  $\tilde{\mathbf{u}}(n)$  is a vector with  $k$ th element  $\tilde{u}_k(n) = u_k^3(n)$ . Iterating (3.32) from  $n$  to  $K$  (downward), we get

$$\bar{\mathbf{u}}(n) \approx \mathbf{C}^{n-K} \bar{\mathbf{u}}(K) + \sum_{i=0}^{n-1-K} \frac{h}{12} \mathbf{C}^i \mathbf{B} \tilde{\mathbf{u}}(n-1-i). \quad (3.33)$$

Define two block-diagonal matrices  $\bar{\mathbf{Q}}$  and  $\bar{\mathbf{\Lambda}}$  of size  $NM \times NM$  as

$$\bar{\mathbf{Q}} = \text{bdiag} [\mathbf{Q}, \mathbf{Q}, \dots, \mathbf{Q}]$$

and

$$\bar{\mathbf{\Lambda}} = \text{bdiag} [\mathbf{\Lambda}, \mathbf{\Lambda}, \dots, \mathbf{\Lambda}] \quad (3.34)$$

where  $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N]$  and  $\mathbf{\Lambda} = \text{diag} [\lambda_1, \lambda_2, \dots, \lambda_N]$  (cf. Section I). Then, from (2.4) and (3.34) the eigenvalue decomposition of  $\bar{\mathbf{R}}$  is obtained as

$$\mathbf{R} = \bar{\mathbf{Q}} \bar{\mathbf{\Lambda}} \bar{\mathbf{Q}}^T = \sum_{k=1}^M \sum_{j=1}^N \gamma_{kj} \mathbf{a}_{kj} \mathbf{a}_{kj}^T \quad (3.35)$$

where  $\gamma_{kj} = \lambda_j$  and  $\mathbf{a}_{kj} = [\mathbf{0}^T, \dots, \mathbf{0}^T, \mathbf{q}_j^T, \mathbf{0}^T, \dots, \mathbf{0}^T]^T$  with  $\mathbf{q}_j$  occupying the  $k$ th  $N$ -vector segment in the  $NM \times 1$  vector  $\mathbf{a}_{kj}$ . Substituting (3.35) in (3.33), we get

$$\begin{aligned} \bar{\mathbf{u}}(n) \approx & \sum_{k=1}^M \sum_{j=1}^N \{1 - h[\gamma_{kj} - 2\mu(1 - \beta_k^2)]\}^{n-K} \\ & \cdot \mathbf{a}_{kj} \mathbf{a}_{kj}^T \bar{\mathbf{u}}(K) + \sum_{i=0}^{n-1-K} \sum_{k=1}^M \sum_{j=1}^N \frac{h}{12} \{1 - h[\gamma_{kj} \\ & - 2\mu(1 - \beta_k^2)]\}^i [\gamma_{kj} - 2\mu(1 - \beta_k^2)] \\ & \cdot \mathbf{a}_{kj} \mathbf{a}_{kj}^T \tilde{\mathbf{u}}(n-1-i). \end{aligned} \quad (3.36)$$

For convergence of  $\bar{\mathbf{u}}(n)$ , we must have

$$\begin{aligned} |1 - h[\gamma_{kj} - 2\mu(1 - \beta_k^2)]| & \leq 1 \\ \forall k = 1, \dots, M; j = 1, \dots, N. \end{aligned} \quad (3.37)$$

Now, we know from Section III-A that the minimizer of  $J$ , to which the above algorithm converges, is the set of minimum eigenvectors of  $\mathbf{R}$ . Further, we do not have any *a priori* knowledge about the eigenvalues of  $\mathbf{R}$  so as to make any precise choice of the value of  $h$ . Taking these into account as well as the ordering of the eigenvalues, we get the conditions on  $h$ , for each  $k = 1, \dots, M$ , as

$$1 - h[\gamma_{kj} - 2\mu(1 - \beta_k^2)] = 1 \quad \forall j = P+1, \dots, N \quad (3.38)$$

$$|1 - h[\gamma_{kj} - 2\mu(1 - \beta_k^2)]| < 1 \quad \forall j = 1, \dots, P \quad (3.39)$$

which on simplification results in

$$\lambda_{\min} = 2\mu(1 - \beta_k^2) \quad \forall k = 1, \dots, M \quad (3.40)$$

$$0 < h < \frac{2}{\gamma_{kj} - \lambda_{\min}} \quad \forall k = 1, \dots, M; j = 1, \dots, P. \quad (3.41)$$

TABLE I  
QUALITY ( $E$ ) AND  $Orth_{\max}$  OF THE ESTIMATED  
EIGENVECTORS ( $N = 8$ , SNR = 10 dB)

	$D$	$M$	$E$	$Orth_{\max}$
Single sinusoid case	6	6	$5.138 \times 10^{-5}$	$1.786 \times 10^{-6}$
	6	3	$5.129 \times 10^{-5}$	$9.550 \times 10^{-7}$
Two sinusoid case	4	4	$5.737 \times 10^{-4}$	$5.057 \times 10^{-7}$
	4	2	$5.480 \times 10^{-4}$	$4.317 \times 10^{-8}$

In order to obtain a single upper-bound which will be satisfied for all  $k$  and  $j$  in (3.41), we replace  $\gamma_{kj}$  by  $\lambda_{\max}(= \lambda_1)$ , giving

$$0 < h < \frac{2}{\lambda_{\max} - \lambda_{\min}}. \quad (3.42)$$

Since the eigenvalues of  $\mathbf{R}$  are not known *a priori*, practical bounds for  $h$  can be given as

$$0 < h < \frac{2}{\text{Trace}(\mathbf{R})}. \quad (3.43)$$

Recall that in arriving at (3.31), the Taylor's series was truncated at fourth order term. Inclusion of higher order terms would cause extra terms on the R.H.S of (3.32). But, all these extra terms will have the same form as the second term. As it is evident from the above analysis, presence of these higher order terms will not affect the final answer in (3.42) in any way.

#### IV. SIMULATION RESULTS

We now present some simulation results to support the theoretical assertions made in the previous section.

In the simulation studies, the matrix  $\mathbf{R}$  was chosen as the asymptotic covariance matrix of the signal consisting of  $P/2$  real sinusoids in additive white noise of unit variance (i.e.,  $\lambda_{\min} = 1$ ). Size of  $\mathbf{R}$  was fixed at 8 (i.e.,  $N = 8$ ). Amplitudes of the sinusoids were chosen to give the desired signal-to-noise ratio (SNR). If  $\zeta$  is the amplitude of the sinusoid, then the SNR(dB) is given by  $10 \log(\zeta^2/2)$ . We considered two cases: 1) single sinusoid of 0.2 Hz (normalized) in noise, i.e.,  $P = 2$ , and 2) two equipower sinusoids of 0.2 and 0.24 Hz. (normalized) in noise, i.e.,  $P = 4$ . The dimension of  $\mathbf{S}$  (i.e.,  $D$ ) for these two cases is 6 and 4, respectively.

The performance measures used for evaluating the quality of the estimated eigenvectors are as follows. Let  $\hat{\mathbf{w}}_i^*$ ,  $i = 1, \dots, M$ , represent the estimated eigenvectors where  $M \leq D$ . To see the quality of the estimated eigenvectors, i.e., how close they are to the true subspace, we use the projection error measure

$$E = \sqrt{\sum_{i=1}^M \left\| \frac{\hat{\mathbf{w}}_i^*}{\|\hat{\mathbf{w}}_i^*\|} - \sum_{j=P+1}^N \mathbf{q}_j \mathbf{q}_j^T \frac{\hat{\mathbf{w}}_i^*}{\|\hat{\mathbf{w}}_i^*\|} \right\|^2}. \quad (4.1)$$

The true eigenvectors ( $\mathbf{q}_j$ ) are obtained using the MATLAB eigenvalue decomposition routine.

Observe from (4.1) that  $E$  does not reveal the extent of orthogonality among the estimated eigenvectors. It only conveys how close the estimated eigenvectors are to  $\mathbf{S}$ . The smaller the value of  $E$ , the closer the estimated subspace is

TABLE II  
CONVERGENCE PERFORMANCE AS A FUNCTION OF SNR

	SNR (dB)	$D$	$M$	$E$	$Orth_{\max}$	$N_{\text{con}}$	$P(\lambda)$
Single sinusoid case	5	6	4	$2.288 \times 10^{-4}$	$7.785 \times 10^{-7}$	201	12.6722
	15	6	4	$1.425 \times 10^{-5}$	$7.955 \times 10^{-7}$	30	117.7189
Two sinusoid case	5	4	2	$1.821 \times 10^{-3}$	$1.741 \times 10^{-7}$	1054	2.2608
	15	4	2	$1.658 \times 10^{-4}$	$1.494 \times 10^{-7}$	148	13.6078

to  $\mathbf{S}$ , and vice versa. Since the proposed approach results in orthogonal eigenvectors, we define an orthogonality measure,  $Orth_{\max}$ , as

$$Orth_{\max} = \max_{i,j \in \{1, \dots, M\}} \left\{ \left| \frac{\hat{\mathbf{w}}_i^* \hat{\mathbf{w}}_j^*}{\|\hat{\mathbf{w}}_i^*\| \cdot \|\hat{\mathbf{w}}_j^*\|} \right| \right\}_{i \neq j} \quad (4.2)$$

Observe that  $Orth_{\max}$  is the magnitude cosine of the angle between those two vectors which have the smallest angle between them. Thus, the larger the value of  $Orth_{\max}$ , the less orthogonal the estimated eigenvectors are and vice versa.

The system of differential equations (cf. (3.26)–(3.28)) was solved numerically choosing the integration time-step  $h$  according to (3.43). The vector  $\bar{\mathbf{u}}(n)$  was initialized to an  $NM \times 1$  random vector and the iterations were stopped when the norm of the difference between the consecutive solution vectors was less than a predetermined threshold  $\delta$  (i.e.,  $\|\bar{\mathbf{w}}(n+1) - \bar{\mathbf{w}}(n)\|_2 < \delta$ ). The parameters  $\delta$ ,  $\mu$  and  $\alpha$  were chosen as  $10^{-5}$ , 50 and 100, respectively.

Table I gives the results for  $M = D$  and  $M < D$ , for the cases of single and two sinusoids. It may be observed that the quality of estimated subspace ( $E$ ) is quite good and for a given  $D$  it does not change appreciably with the value of  $M$ . The entries under  $Orth_{\max}$  show that the estimated eigenvectors are almost orthogonal.

Table II illustrates the convergence performance of the approach for different SNR's. The quantity  $N_{\text{con}}$  is the number of iterations which the algorithm took to converge and  $P(\lambda)$  is the ratio  $\lambda_P/\lambda_{\min}$ . The latter quantity,  $P(\lambda)$ , plays a significant role in deciding the convergence rate of the algorithm, as explained below.

Note from Theorem 3.3 that for any finite  $\mu$ , the norm of the estimated minimum eigenvectors is less than unity. For moderate values of  $\mu$  (as chosen for the numerical results in Table II), the outputs of the neurons lie in a small neighborhood around origin where the sigmoidal function can be approximated by a straight line. Then, the R.H.S of (3.36) can be approximated by the first term. Since the rate of convergence is controlled by factors of the form  $1 - h(\lambda_j - \lambda_{\min})$ ,  $j = 1, \dots, N$ , it is clear that the convergence speed will be decided by the factor  $1 - h(\lambda_P - \lambda_{\min})$  as this will be closest to unity. Recall that  $P(\lambda) = \lambda_P/\lambda_{\min}$ . A large value for this ratio implies that  $\lambda_P$  is far away from  $\lambda_{\min}$  which results in a smaller value for  $1 - h(\lambda_P - \lambda_{\min})$  thereby yielding faster convergence. Since  $P(\lambda)$  increases with SNR, the convergence speed goes up with SNR.

We may point out here that the values chosen for the parameters  $\mu$  and  $\alpha$  do not affect the quality of the estimated eigenvectors. By choosing a smaller value for the threshold

parameter,  $\delta$ , we can get more accurate estimates. However, for a given  $h$ , the smaller the value of  $\delta$ , the longer will be the convergence time. For  $\mu \leq \frac{\Delta_{\min}}{2}$ , behavior of the system will be erroneous. Also, if we choose  $M > D$ , the algorithm will converge to some arbitrary point which may not have any significance to the problem addressed here.

## V. CONCLUSION

The problem of estimating the orthogonal eigenvectors corresponding to the repeated minimum (in magnitude) eigenvalue of a symmetric nonindefinite matrix is addressed and a feedback neural network solution has been developed. This is done by constructing an appropriate energy function for this NN. An analysis establishing the correspondence between the minimizers of this energy function and the desired set of orthogonal eigenvectors has been presented. Bounds on the integration time-step which is required to numerically solve the system of differential equations have also been derived. Simulation results are presented to corroborate the analysis.

## REFERENCES

- [1] N. L. Owsley, "Adaptive data orthogonalization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1978, pp. 109–112.
- [2] K. C. Sharman, "Adaptive algorithms for estimating the complete covariance eigenstructure," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1986, pp. 1401–1404.
- [3] J. F. Yang and M. Kaveh, "Adaptive eigensubspace algorithms for direction or frequency estimation and tracking," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 2, pp. 241–251, Feb. 1988.
- [4] T. K. Sarkar and X. Yang, "Application of the conjugate gradient and steepest descent for computing the eigenvalues of an operator," *Signal Processing*, vol. 17, pp. 31–38, May 1989.
- [5] G. Mathew, V. U. Reddy, and S. Dasgupta, "Gauss-Newton based adaptive subspace estimation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, Mar. 1992, pp. 205–208, vol. IV.
- [6] R. P. Lippman, "An introduction to computing with neural nets," *IEEE Acoust., Speech, Signal Processing Mag.*, vol. 4, pp. 4–22, Apr. 1987.
- [7] J. J. Hopfield and D. W. Tank, "Neural computations of decisions in optimization problems," *Biological Cybern.*, vol. 52, pp. 141–152, 1985.
- [8] D. G. Luenberger, *Linear and Non-linear Programming*. Reading, MA: Addison-Wesley, pp. 366–369, 1978.
- [9] J. R. Bunch, C. P. Neilsen, and D. C. Sorensen, "Rank-one modification of the symmetric eigenproblem," *Numerische Mathematik*, vol. 31, pp. 31–48, 1978.
- [10] R. D. DeGroat and R. A. Roberts, "Efficient, numerically stabilized rank-one eigenstructure updating," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 2, pp. 301–316, Feb. 1990.
- [11] K. B. Yu, "Recursive updating the eigenvalue decomposition of a covariance matrix," *IEEE Trans. Signal Processing*, vol. 39, no. 5, pp. 1136–1145, May 1991.
- [12] C. H. Bischof and G. M. Shroff, "On updating signal subspaces," *IEEE Trans. Signal Processing*, vol. 40, no. 1, pp. 96–105, Jan. 1992.
- [13] R. D. DeGroat, "Noniterative subspace tracking," *IEEE Trans. Signal Processing*, vol. 40, no. 3, pp. 571–577, Mar. 1992.
- [14] P. Baldi and K. Hornik, "Neural networks and principal component analysis: Learning from examples without local minima," *Neural Networks*, vol. 2, pp. 53–58, 1989.

- [15] T. D. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward neural network," *Neural Networks*, vol. 2, pp. 459-473, 1989.
- [16] S. Y. Kung and K. I. Diamantaras, "A neural network learning algorithm for adaptive principal component extraction (APEX)," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1990, pp. 861-864.
- [17] S. Y. Kung "Constrained principal component analysis via an orthogonal learning network," in *Proc. IEEE Int. Symp. Circuits, Syst.*, pp. 719-722, 1990.
- [18] E. Oja, "A simplified neuron model as a principal component analyzer," *J. Math. Biology*, vol. 15, pp. 267-273, 1982.
- [19] L. Xu, E. Oja and C.Y. Suen, "Modified Hebbian learning for curve and surface fitting," *Neural Networks*, vol. 5, pp. 441-457, 1992.
- [20] L. Wang and J. M. Mendel, "Parallel structured networks for solving a wide variety of matrix algebra problems," *J. Parallel, Distributed Computing*, vol. 14, no. 3, pp. 236-247, Mar. 1992.
- [21] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge University Press, 1989, pp. 103-108,
- [22] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectrum Estimation Workshop*, (New York), 1979.
- [23] R. W. Brockett, "Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems," in *Proc. IEEE Int. Conf. Decision, Contr.*, 1988, pp. 799-803.

**V. U. Reddy** (S'68-M'70-SM'82) received the B.E. and M.Tech. degrees in electronics and communication engineering from Osmania University and the Indian Institute of Technology (IIT), Kharagpur, in 1962 and 1963, respectively, and the Ph.D. degree in electrical engineering from the University of Missouri, Columbia, in 1971.

From 1972 to 1976, he was an Assistant Professor at IIT, Madras, and was a Professor at IIT, Kharagpur, from 1976 to 1979. During 1979-1982 and 1986-1987, he was a Visiting Professor at the Department of Electrical Engineering, Stanford University, Stanford, CA. In April 1982, he joined Osmania University as a Professor and was the Founder-Director of the Research and Training Unit for Navigational Electronics, funded by the Department of Electronics, Government of India. Since April 1988, he has been with the Indian Institute of Science, Bangalore, as a Professor of Electrical Communication Engineering and is presently its Chairman. He has served as a Consultant in signal processing to the Avionics Design Bureau of Hindustan Aeronautics Limited, Hyderabad, and to Central Research Laboratory, Bharat Electronics Limited, Bangalore. His research interests are in sensitivity analysis of high-resolution algorithms, adaptive algorithms, adaptive arrays, and wavelet transform.

Dr. Reddy is a Fellow of the Indian Academy of Sciences, Indian National Academy of Engineering, and Indian National Science Academy, and Fellow of the Institute of Electronics and Telecommunication Engineers (IETE), India. He received the S. K. Mitra Memorial Award (1989) from IETE for the best research paper.



**George Mathew** received the B.E. degree from Karnataka Regional Engineering College, Surathkal, India, in 1987, and the M.Sc.(Engg.) degree from the Indian Institute of Science (IISc), Bangalore, in 1989, both in electrical communication engineering. He is working toward the Ph.D. degree in electrical communication engineering at IISc.

His research interests are in adaptive algorithms and neural networks.