Neurocomputing 3 (1991) 69–95
Elsevier

Homogeneous alternating projection neural networks

Seho Oh, Robert J. Marks II and Dennis P. Sarr
Department of Electrical Engineering, FT-10, University of Washington, Seattle, WA 98195, USA

Abstract

Oh, S., R.J. Marks II and D.P. Sarr, Homogeneous alternating projection neural networks, Neurocomputing 3 (1991) 69–95

The homogeneous form of the alternating projection neural network (APNN) performs as a content-addressable memory. We analyze and illustrate the characteristics and performance of the homogeneous APNN. Convergence of the iterative reconstruction becomes faster when the percentage of the clamped neurons, corresponding to known states, increases and the number of stored library vectors decreases. For a bipolar (± 1) library, we demonstrate one-step convergence when the number of output neurons is sufficiently small. A new per-step minimization method for relaxation is introduced and is favorably contrasted in performance to other relaxation methods. We also propose a modified training procedure that requires neither a global norm operation nor division. Lastly, the noise characteristics of the APNN are examined and illustrated.

Keywords. Content-addressable memories; associative memories; image processing; noise.

1. Introduction

The homogeneous form of the *alternating projection neural network* (APNN) is a content-addressable memory. It is structured as a maximally connected array of L neurons wherein a set of library vectors, $\{\vec{f}_n | 1 \leq n \leq N\}$, are stored in the interconnects. The interconnection matrix projects onto the space spanned by library. The neurons perform clamping and thresholding operations. In synchronous form, the APNN can be viewed as alternatively projecting between two or more convex sets [1–3]. The APNN has also been shown to perform successfully in an asynchronously and skewed mode [4].

Neurons in an APNN can be clamped to preassigned value and provide the network stimulus. Alternately, a neuron's state can float in accordance to the stimulus of other neurons. The status of a neuron as clamped or floating may change from application to application. Under certain conditions to be stated, the APNN can reconstruct any one library vector by clamping an otherwise arbitrary subset of the neurons to the values equal to the elements of that vector. The states of the remaining floating neurons will then converge to the unknown vector elements. The capacity of the APNN is proportional to the number of clamped neurons.

An example of an APNN's operation is shown in *Fig. 1*. A total of 40 images were stored in the net. One of the images is the girl's face in the upper left hand corner. As is shown in the upper middle picture, a portion of the image is lost. The neurons corresponding to the known portion of the figure are clamped to their known values.

0925-2312/91/$03.50

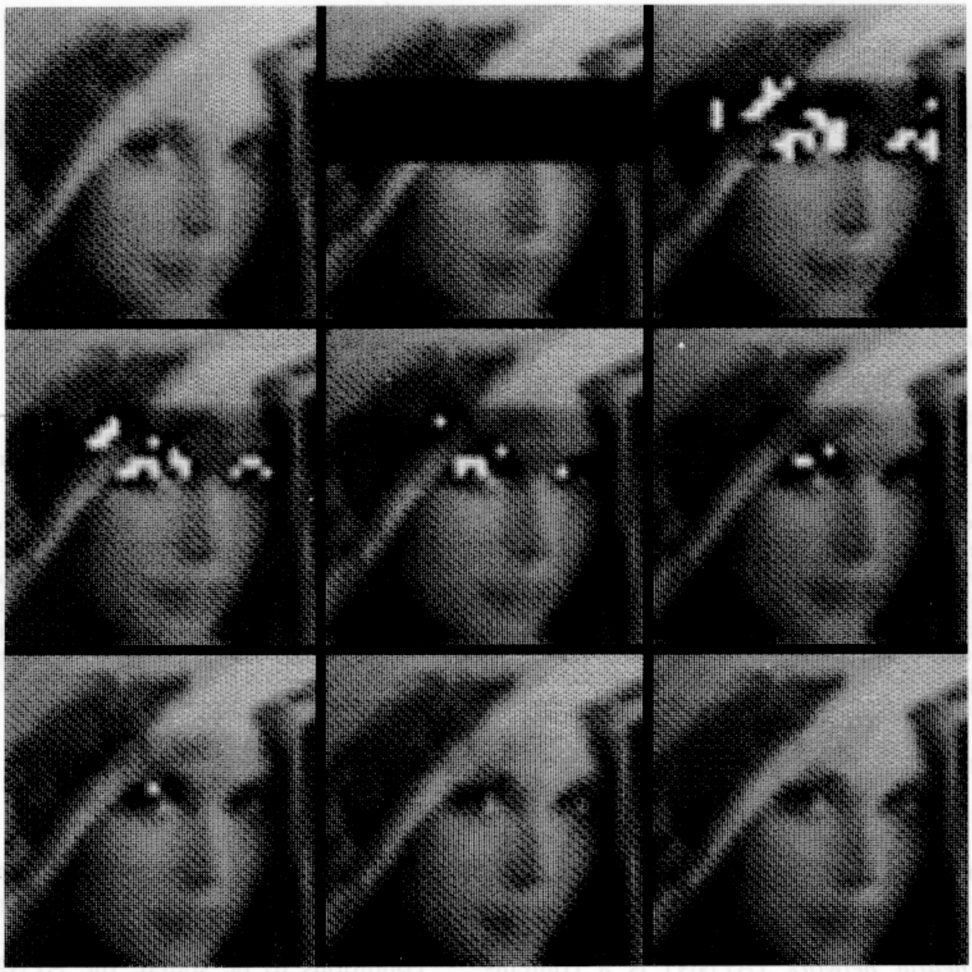


Fig. 1. Image reconstruction sequence. Top left to lower right sequence: original image, clamped image, iteration 0, iteration 1, iteration 2, iteration 3, iteration 5, iteration 10, iteration 19. A total of 40 images were stored on the APNN. Each image is 32×32.

Neurons corresponding to the unknown values, in this case in the proximity of the girl's eyes, float. The net begins to update the neural values until, as shown, convergence to the desired image is achieved.

The APNN also has interpolatory associative memory properties as is illustrated in *Fig. 2*. As shown in the top row of images, a composite initialization is made of the hair of one girl and the nose and mouth of a second. Both images have been stored in the neural network. The APNN interpolated eyes as shown in the lower right figure. Specifically, we have found the closest image in the library vector subspace to the linear variety formed by all images equal to the clamped values in the initialization. The same neural network used in the example in *Fig. 1.* was used here.

In this paper, we present significant APNN properties beyond those reported previously

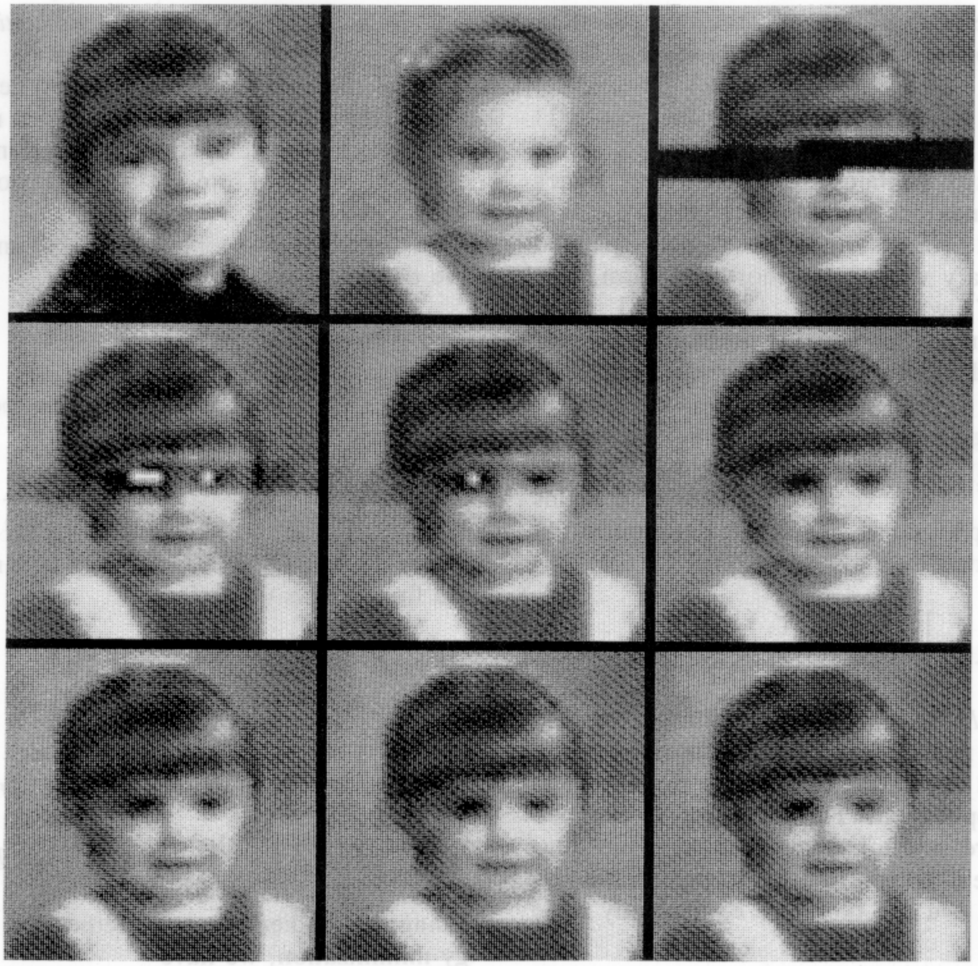


Fig. 2. Interpolatory associative memory property illustration. In the top row of images, initialization is made of the hair of a girl and the nose and mouth of a second. Top left to lower right sequence: original image, clamped image, iteration 0, iteration 1, iteration 2, iteration 5, iteration 8, iteration 10.

[5, 4]. We show, for example, that the APNN can be viewed as a gradient descent algorithm. The convergence rate of the network is shown to worsen as the percentage of floating neurons and/or number of library vectors increase. For a bipolar library, convergence is shown to always occur after finite number of iterations. A bound for this iteration number is established. We also propose improved techniques for both training and iteration. Finally, the performance of the APNN is examined in the presence of various types of inexactitudes. We show that, in general, the noise sensitivity decreases as the percentage of clamped neurons increases.

2. Preliminaries

2.1. Synchronous operation

Let $\vec{s}(M)$ denote the vector of the neuron states at time M. For synchronous operation without use of thresholds [4], the iterative state equation for the APNN can be written as:

$$\vec{s}(M+1) = \eta \underline{T} \vec{s}(M) . \tag{1}$$

The matrix $\underline{T}$ contains the interconnect values among the neurons and is equal to the projection matrix that projects onto the subspace spanned by the library vectors. In shorthand form, $\underline{T} = \underline{F}(\underline{F}^{\mathrm{T}}\underline{F})^{-1}\underline{F}^{\mathrm{T}}$ where $\underline{F} = [\vec{f}_1 | \vec{f}_2 | \cdots | \vec{f}_N]$ is the library matrix. The clamping operator, η, resets all clamped neurons to their preassigned clamped states. For a specified partition of clamped and floating neurons we can assume, with no loss of generality, that neurons 1 through P are clamped. Let $\vec{f}^P$ be a vector of length P contains these clamped values. Then the iteration in (1) can be written as

$$\begin{bmatrix} \vec{f}^P \\ \vec{s}^Q(M+1) \end{bmatrix} = \eta \begin{bmatrix} \underline{T}_2 & \underline{T}_1 \\ \underline{T}_3 & \underline{T}_4 \end{bmatrix} \begin{bmatrix} \vec{f}^P \\ \vec{s}^Q(M) \end{bmatrix} , \tag{2}$$

where $\vec{s}^Q(M)$ denotes the vector of the last $Q = L - P$ neuron states. The status of a neuron as clamped or floating can change from application to application. The states of P clamped neurons are not affected by their input sum. Thus, there is no contribution to the iteration by $\underline{T}_1$ and $\underline{T}_2$. We can therefore equivalently write (2) as

$$\vec{s}^Q(M+1) = \underline{T}_3 \vec{f}^P + \underline{T}_4 \vec{s}^Q(M) . \tag{3}$$

If $\underline{F}_P$ is full rank, then the norm of $\underline{T}_4$ is strictly less than one [4, 5]. It follows that the steady state solution of (3) is

$$\vec{s}^Q(\infty) = (\underline{I} - \underline{T}_A)^{-1} \underline{T}_3 \vec{f}^P = \vec{f}^Q , \tag{4}$$

i.e. the steady state solution is the extrapolation of the library vector.

$$\vec{f} = \begin{bmatrix} \vec{f}^P \\ \vec{f}^Q \end{bmatrix} .$$

2.2. Relation to the energy minimization ANN

As is the case with many other neural networks, the APNN can be viewed as a reduction of an energy metric operation. This is in contrast to the geometrical interpretation previously presented [4].

Let $\vec{i}$ be in the column space of $\underline{F}$ and $\vec{i} = \underline{T}\vec{s}$. We define the energy function as

$$E = \| \eta \vec{i} - \vec{i} \|^2 = \| \vec{f}^P - \vec{i}^P \|^2 , \tag{5}$$

where $\vec{i}^P$ is a vector of the first P components of $\vec{i}$. Because $\vec{i} = \underline{T}\vec{i}$,

$$\vec{i}^P = \underline{T}_P \vec{i} , \tag{6}$$

where $\underline{T}_P = \underline{F}_P(\underline{F}^{\mathrm{T}}\underline{F})^{-1}\underline{F}^{\mathrm{T}}$ and $\underline{F}_P$ is a matrix of the first P rows of $\underline{F}$. Thus

$$E = \| \vec{f}^P - \underline{T}_P \vec{i} \|^2 .$$

From (6), it follows that

$$\begin{aligned} \vec{\nabla} E &= 2\underline{T}_P^{\mathrm{T}} \underline{T}_P \vec{i} - 2\underline{T}_P^{\mathrm{T}} \vec{f}^P = 2\underline{T}_P^{\mathrm{T}}(\vec{i}^P - \vec{f}^P) \\ &= 2\underline{F}(\underline{F}^{\mathrm{T}}\underline{F})^{-1}\underline{F}^{\mathrm{T}}(\vec{i} - \eta\vec{i}) = 2(\vec{i} - \underline{T}\eta\vec{i}) . \end{aligned}$$

Using the gradient descent method for the energy minimization, we have

$$\begin{aligned} \vec{i}(M+1) &= \vec{i}(M) - \beta\, \vec{\nabla} E \\ &= (1 - 2\beta)\vec{i}(M) + 2\beta \underline{T}\eta\vec{i}(M) . \end{aligned}$$

Premultiplying η both sides and using $\vec{s}(M) = \eta\vec{i}(M)$, we obtain

$$\vec{s}(M+1) = \theta\eta\underline{T}\vec{s}(M) + (1-\theta)\vec{s}(M) ,$$

where $\theta = 2\beta$. If $\theta = 1$, then the gradient descent method is exactly the same as the APNN operation. For $\theta \neq 1$, we have a relaxed version of the APNN [4].

3. APNN properties

3.1. Convergence

3.1.1. The effects of the percentage of clamped neuron and library size on the convergence rate

The convergence rate of the APNN is linear with a time constant on the order of the norm (i.e. maximum eigenvalue or spectral radius) of $\underline{T}_4$ [5]. In this section, we discuss the effect on convergence of the number of library vectors, N, and the number of clamping neurons, P.

As the percentage of clamped neurons decreases, the norm of $\underline{T}_4$ increases and convergence slows. This is proven in Appendix A and illustrated in *Fig. 3.* The same neural network is used as in *Fig. 1.* In *Fig. 3*, however, fewer neurons are clamped in the initialization. A total

Fig. 3. Image reconstruction using the same net used in *Fig. 1.* Here, the percentage of clamped neurons is reduced to 75%. Top left to lower right sequence: original image, clamped image, iteration 0, iteration 2, iteration 20, iteration 40, iteration 60, iteration 80, iteration 101. Clearly, more iterations are required.

of 750 synchronous iterations were needed to generate the result at the bottom right of *Fig. 3*. Only 19 iterations were required for the bottom right image in *Fig. 1*.

Similarly, increasing the number of library vectors increases the norm of $\underline{T}_4$ and therefore slows convergence. A proof is in Appendix A. To illustrate, the neural network used for the example in *Fig. 1* was trained with additional images for the total of 160 stored images. As is shown in *Fig. 4*, a total of 208 iterations were required to generate the image in the bottom right hand corner as opposed to 19 iterations in *Fig. 1*.

3.1.2. Convergence for bipolar library

Assume that our library is bipolar, i.e. all library elements are ± 1. In this case, we can perfoım a sign $(\cdot)$ operation of the neural state after a finite number of iterations. Clearly, if the

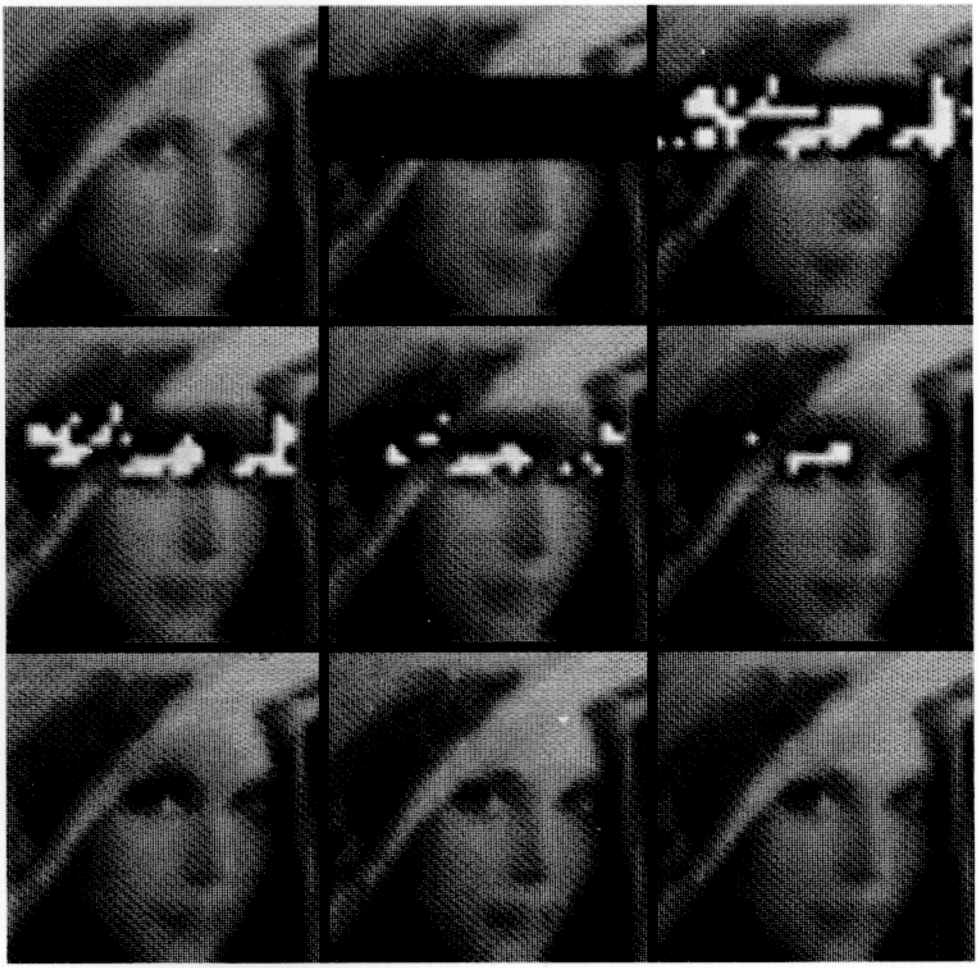


Fig. 4. Image reconstruction sequence when the number of the library images is 160. Top left to lower right sequence: original image, clamped image, iteration 1, iteration 5, iteration 21, iteration 20, iteration 50, iteration 150, iteration 208. Comparing to *Fig. 1*, convergence is slowed as more images are stored in the net.

sign of each vector in the series

$$s_i(M) \cdots s_i(M+j) \cdots s_i(\infty) \cdots$$

is the same, then

$$s_i(\infty) = \mathrm{sign}[s_i(M)] .$$

This means that convergence can be achieved after a finite number of iterations. Consider, then, the following lemma.

Lemma 1. *If* $\|\vec{s}^Q(M) - \vec{s}^Q(M-1)\| < 1 - \|\underline{T}_4\|$, *then* $\vec{f}^Q = \mathrm{sign}[\vec{s}^Q(M-1)]$.

We can also establish a bound for the required number of iterations.

Lemma 2. *If*

$$M > -\frac{\ln(Q)}{2\ln(\|\underline{T}_4\|)} ,$$

then $\vec{f}^Q = \mathrm{sign}[\vec{s}^Q(M-1)]$.

Proofs of Lemma 1 and 2 are in Appendix A.

From Lemma 2, we have a condition for one-step convergence in the sense that

$$\vec{f}^Q = \mathrm{sign}[\vec{s}^Q(1)] \tag{7}$$

when

$$\|\underline{T}_4\| Q^{1/4} < 1 .$$

If there is a single output neuron ($Q = 1$) then the sufficient condition is $\|\underline{T}_4\| < 1$. This condition is satisfied if $\underline{F}_P$ is of full column rank.

3.2 Relaxation methods

Both the projection and clamping operations can be relaxed to alter the network without affecting its steady state solution [6]. For the interconnects, we choose an appropriate value of a nonstationary relaxation parameter $\theta(M)$ at time M and redefine the interconnect matrix as

$$\underline{T}^{\theta}(M) = \theta(M)\underline{T} + [1 - \theta(M)]\underline{I} .$$

If this operation is stable, then

$$\vec{s}^Q(\infty) = (\underline{I} - \underline{T}_4)^{-1}\underline{T}_3\vec{f}^P .$$

This steady state solution is our desired result.

We now consider and contrast a number of other relaxation methods.

3.2.1. Constant relaxation parameter

Consider the case where $\theta(M)$ is constant in time (i.e. stationary). Let an eigenvalue of $\underline{T}_4^{\theta}$ be λ^{θ}:

$$\lambda^{\theta} = \theta\lambda + (1 - \theta) .$$

We have shown [4, 5] that good stationary relaxation choices are

$$\theta = \frac{\mathrm{tr}(\underline{I} - \underline{T}_4)}{\mathrm{tr}[(\underline{I} - \underline{T}_4)^2]} , \tag{8}$$

where $\mathrm{tr}(\cdot)$ denotes the matrix trace, and

$$\theta = \frac{1}{1 - (\lambda_{\max} + \lambda_{\min})/2} , \tag{9}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ denote the maximum and minimum eigenvalues of $\underline{T}_4$. These relaxation parameters correspond to the use of ℓ_2 and ℓ_∞ norms respectively. The performance of (9) is better than (8), but the operation of (9) needs the calculation of the maximum and minimum eigenvalue of $\underline{T}_4$.

3.2.2. Stark's relaxation method

We define the error for the nonstationary relaxation parameter as

$$\epsilon(M) = \|\underline{\eta}\vec{s}(M) + \theta(M)\underline{\eta}(\underline{T} - \underline{I})\vec{s}(M) - \vec{s}(\infty)\|^2 .$$

Ideal relaxation minimizes $\epsilon(M)$ at each step, but requires knowledge of the steady state solution. Stark [7] proposes the suboptimal solution

$$\theta(M) = 1 + \frac{\|\underline{T}\vec{s}(M) - \eta\underline{T}\vec{s}(M)\|^2}{\|\eta\underline{T}\vec{s}(M) - \vec{s}(M)\|^2} .$$

3.2.3. Relaxation by per-step minimized method

We propose a different relaxation parameter at time M that minimizes the error

$$\epsilon(M) = \|\eta\underline{T}\vec{s}(M) - \vec{s}(M)\|^2 . \tag{10}$$

Define $\vec{u}(M) = \eta\underline{T}\vec{s}(M-1)$ and $\vec{w}(M) = \eta\underline{T}\vec{u}(M)$. The recurrence relation of $\vec{s}(M)$ is

$$\begin{aligned}\vec{s}(M) &= \theta(M)\eta\underline{T}\vec{s}(M-1) \\ &\quad + [1-\theta(M)]\vec{s}(M-1) \\ &= \theta(M)\vec{u}(M) \\ &\quad + [1-\theta(M)]\vec{s}(M-1) .\end{aligned} \tag{11}$$

Using (11), Eq. (10) is

$$\begin{aligned}\epsilon(M) = \|&\theta(M)[2\vec{u}(M) - \vec{w}(M) - \vec{s}(M-1)] \\ &- [\vec{u}(M) - \vec{s}(M-1)]\|^2 .\end{aligned} \tag{12}$$

From the minimization of (12), the optimal value of $\theta(M)$ is

$$\theta(M) = 1 + \frac{\vec{\delta}^{\mathrm{T}}\vec{\epsilon}}{\|\vec{\epsilon}\|^2} ,$$

where $\vec{\epsilon} = 2\vec{u}(M) - \vec{w}(M) - \vec{s}(M-1)$ and $\vec{\delta} = \vec{w}(M) - \vec{u}(M)$. Using (11), the recurrent relationship of $\vec{u}(M)$ is

$$\begin{aligned}\vec{u}(M+1) &= \eta\underline{T}\vec{s}(M-1) \\ &= \theta(M)\vec{w}(M) + [1-\theta(M)]\vec{u}(M) .\end{aligned}$$

We can now state the following relaxation procedure:

1. Initialize ($M = 0$)
   (a) Choose $\vec{s}^{\,Q}(0)$
   (b) Calculate $\vec{u}(1) = \eta\underline{T}\vec{s}(0)$.
2. Set $M = M + 1$
   (a) Calculate $\theta(M)$

$$\begin{aligned}\vec{w}(M) &= \eta\underline{T}\vec{u}(M) \\ \vec{\epsilon} &= 2\vec{u}(M) - \vec{w}(M) \\ &\quad - \vec{s}^{\,Q}(M-1) \\ \vec{\delta} &= \vec{w}(M) - \vec{u}(M) \\ \theta(M) &= 1 + \frac{\vec{\delta}^{\mathrm{T}}\vec{\epsilon}}{\|\vec{\epsilon}\|^2}\end{aligned}$$

   (b) Update $\vec{u}$ and $\vec{s}$ using the relaxation operator

$$\begin{aligned}\vec{s}(M) &= \theta(M)\vec{u}(M) \\ &\quad + [1-\theta(M)]\vec{s}(M-1) \\ \vec{u}(M+1) &= \theta(M)\vec{w}(M) \\ &\quad + [1-\theta(M)]\vec{u}(M) .\end{aligned}$$

   (c) Go to step 2.

Proof of convergence is in Appendix B.

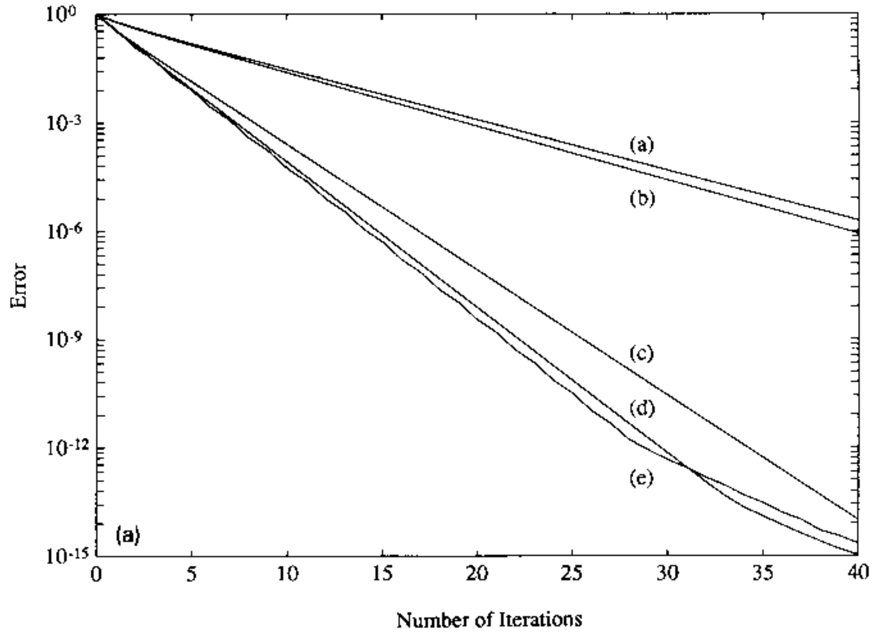
Figure 5 shows the convergence rates for these relaxation methods for two different cases. Stark's method and the per-step minimization outperform the cases of ℓ_2, ℓ_∞ and no relaxation.

3.3. Training

The equation for the interconnect matrix in (1) is unacceptable because of the required prior computation of the inverse of a matrix, which, due to the library matrix structure, may be singular or ill-conditioned. Furthermore, we desire a technique whereby training data can be incrementally learned in a neural network structure one library vector at a time.

3.3.1. Gram–Schmidt with norm operation

Assume we have an interconnect matrix, $\underline{T}$, and wish to update the interconnects correspond-



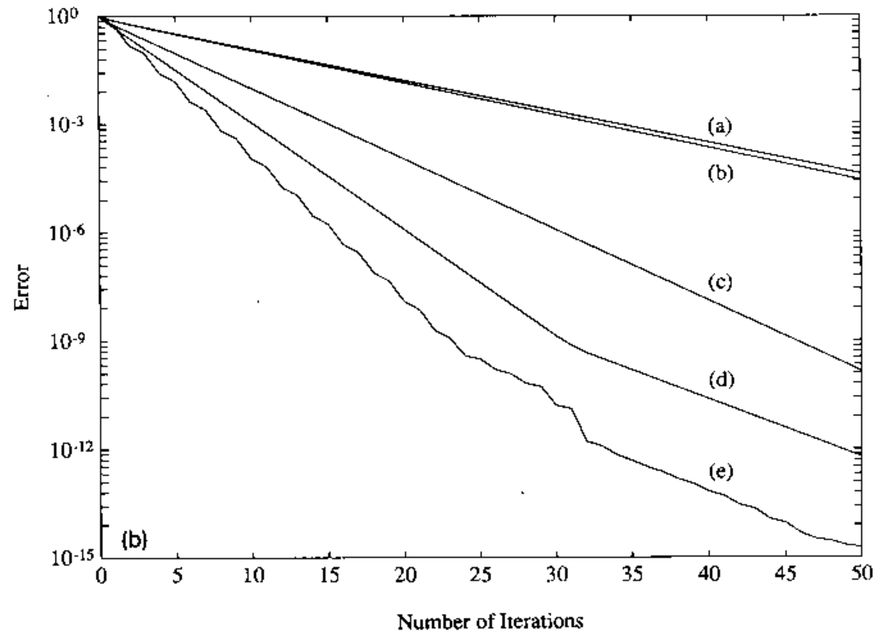


Fig. 5. Convergence for various relaxation methods. The eigenvalues are 0.7245, 0.6036, 0.5538 and 0.3586 for the top figure and 0.8182, 0.7751, 0.6984 and 0.4244 for the lower. (a) no relaxation; (b) ℓ_2 norm; (c) ℓ_∞ norm; (d) Stark's method; (e) per-step minimization method.

ing to a new library vector, $\vec{f}$. The updated interconnect matrix is [4, 5]

$$\underline{T}^{+} = \underline{T} + \frac{\vec{\epsilon}\vec{\epsilon}^{\mathrm{T}}}{\vec{\epsilon}^{\mathrm{T}}\vec{\epsilon}}$$

where $\vec{\epsilon} = (\underline{I} - \underline{T})\vec{f}$. This is similar to Gram–Schmidt orthogonalization.

3.3.2. Gram–Schmidt without norm operation

Algorithm and basic properties. The normal Gram–Schmidt process is mathematically straightforward, but requires evaluation of the error norm, $\|\vec{\epsilon}\|^2 = \vec{\epsilon}^{\mathrm{T}}\vec{\epsilon}$. Also, since the process involves division, sensitivity to this operation can be high when the error norm is small.

In this section, we propose the iterative learning algorithm using the error outer product rather than the norm. We assume that the norms of all library vectors are bounded by B. Let $\underline{T}_n(i)$ be the interconnect matrix after the ith for the nth library vector. The training algorithm is as follows:

Algorithm

1. Choose a convergence factor β in the interval $(1, 4]$ and the iteration number, I (see (13)).
2. Choose a library vector $\vec{f}_n$ and set $\underline{T}_n(0) = \underline{T}_{n-1}(I)$, $(\underline{T}_1(0) = \underline{0})$.
3. Calculate
$$\vec{\epsilon}_i = \vec{f}_n - \underline{T}_{n-1}(i-1)\vec{f}_n$$
$$\underline{T}_n(i) = \underline{T}_n(i-1) + \frac{\beta^{i-1}}{B^2}\,\vec{\epsilon}_i\vec{\epsilon}_i^{\mathrm{T}}\,.$$
4. If $i < I$, then go to step 3. Otherwise go to step 2.

Let the procedure end after N iterations. The resulting interconnect matrix is symmetric and positive semidefinite. Also, $\|\underline{T}_N(I)\| \leq 1$. The proofs of these characteristics are in Appendix C.

Let $\underline{T}_w = \underline{T}_N(I)$. We evaluate $\|\vec{f} - \underline{T}_w\vec{f}\|^2$ when $\vec{f}$ is in the column space of $\underline{F}$. If

$$\vec{f} = \sum_{n=1}^{N} c_n \vec{f}_n$$

then

$$\|\vec{f} - \underline{T}_w\vec{f}\|^2 \leq \frac{B}{\beta^{I/2}} \sum_{n=1}^{N} |c_n|\,.$$

If $\vec{f} = \vec{f}_n$, then

$$\|\vec{f}_n - \underline{T}_w\vec{f}_n\| \leq \frac{B}{\beta^{I/2}}\,.$$

A sufficient condition for $\|\vec{f}_n - \underline{T}_w\vec{f}_n\| \leq \delta$ is

$$I > 2\,\frac{\ln(B/\delta)}{\ln(\beta)}\,. \tag{13}$$

The proof is in Appendix C.

Operation characteristics of the algorithm. We will here evaluate an upper bound for the nonzero eigenvalues of $\underline{T}_w$. As shown in Appendix C,

$$\|\underline{T}_w - \underline{T}_w^2\| \leq \xi\,,$$

where

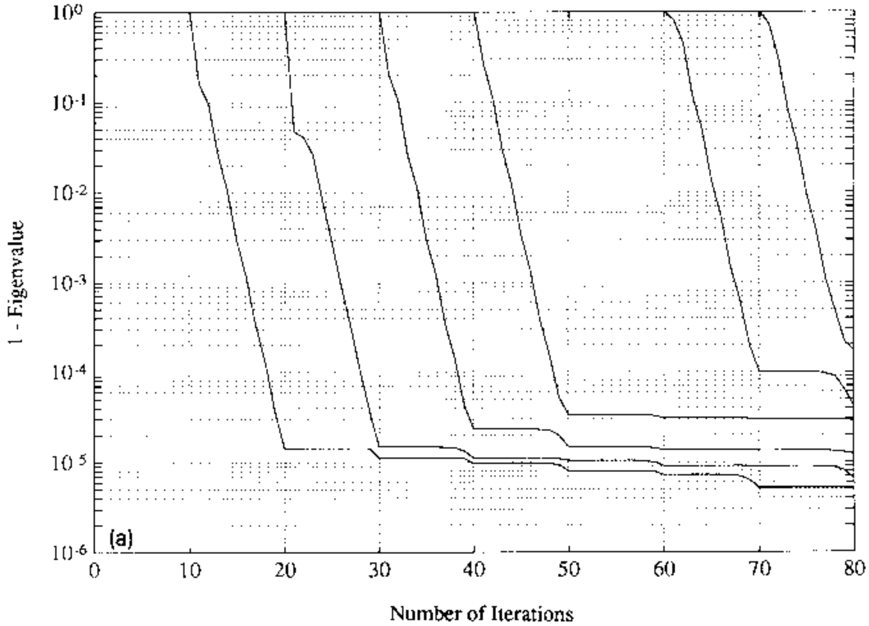
$$\xi = \frac{B}{\beta^{I/2}}\sqrt{\frac{N}{\lambda_F}}$$

and λ_F is the smallest nonzero eigenvalue of $\underline{F}^{\mathrm{T}}\underline{F}$. If $\xi < 0.25$, then the nonzero eigenvalues of $\underline{T}_w$ satisfy the following condition:

$$0 \leq 1 - \lambda \leq \frac{1 - \sqrt{1 - 4\xi}}{2} = \xi_1 < \xi + \xi^2\,.$$

Also from (13), ξ will be

$$\xi = \delta\sqrt{\frac{N}{\lambda_F}}\,.$$



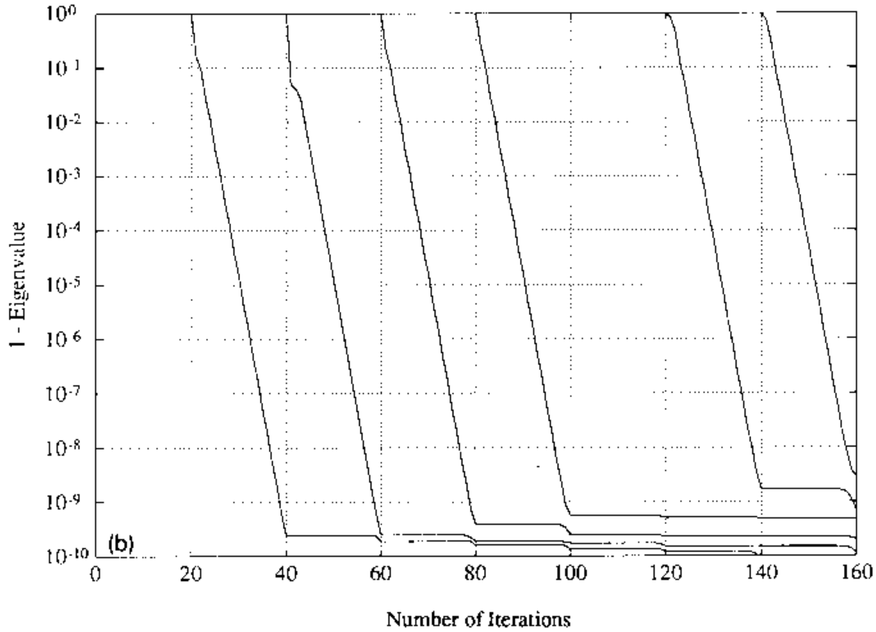


Fig. 6. Eigenvalues versus the number of iterations for the proposed training algorithm. $\beta = 3$ and $\lambda_F = 0.5617$. $\vec{f}_6 = c[\vec{f}_1 + \cdots + \vec{f}_5]$, where c is a constant for $\|\vec{f}_6\|^2 = 10$. (a) $I = 10$ and (b) $I = 20$.

Figure 6 illustrates convergence as a function of the number of iterations, I.

4. Noise characteristics

In this section, we examine the sensitivity of the APNN to computational inexactness and data noise.

4.1. Noise modeling

Figure 7 is a block diagram of the APNN iteration in (3) corrupted with additive noise [8]. The vectors $\vec{n}_i$ and $\vec{n}_d$ denote the input (data source) and output (detector) noise respectively, and $\vec{n}_f$ is the feedback noise vector. These vectors are added appropriately to the vector components (neuron states). The matrices $\underline{N}_3$ and $\underline{N}_4$ denote the system noise which is associated with the interconnects. They may represent the inexactness of analog multiplication or, for digital implementation, round-off error [9]. We assume that each neural noise process consists of elements with identical and independent distributions (iid) in the spatial domain, and, temporally, are either white or static. For system noise, we assume that the noise is spatially iid with temporally white noise. All noise vectors and matrices are assumed to be zero mean and statistically independent.

Let

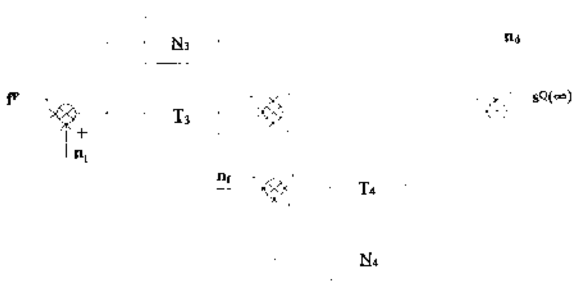


Fig. 7. The block diagram of the APNN operation with additive noise. The matrices $\underline{N}_3$ and $\underline{N}_4$ indicate the interconnect noise.

$$\underline{\tilde{T}}_3(M) = \underline{T}_3 + \underline{\tilde{N}}_3(M)$$

and

$$\underline{\tilde{T}}_4(M) = \underline{T}_4 + \underline{\tilde{N}}_4(M)$$

Then the relationship between the noisy $\vec{s}^Q(M)$ and $\vec{s}^Q(M-1)$ will be

$$\begin{aligned}\vec{s}^Q(M) = {} & \underline{\tilde{T}}_4(M-1)\vec{s}^Q(M-1) \\ & + \underline{\tilde{T}}_3(M-1)[\vec{f}^P + \tilde{\vec{n}}_i(M-1)] \\ & + \tilde{\vec{n}}_f(M-1)\,.\end{aligned}$$

The solution of the above equations is

$$\begin{aligned}\vec{s}^Q(M) = {} & \sum_{k=0}^{M-1} \underline{A}_M(k)\underline{\tilde{T}}_3(M-1-k)\vec{f}^P \\ & + \sum_{k=0}^{M-1} \underline{A}_M(k)\tilde{\vec{n}}_f(M-1-k) \\ & + \sum_{k=0}^{M-1} \underline{A}_M(k)\underline{\tilde{T}}_3(M-1-k) \\ & \times \tilde{\vec{n}}_i(M-1-k)\,, \end{aligned} \tag{14}$$

where

$$\underline{A}_M(k) = \begin{cases} \underline{I} & ; \quad k=0 \\ \underline{\tilde{T}}_4(M-1)\underline{\tilde{T}}_4(M-2) \cdots \underline{\tilde{T}}_4(M-k) & ; \quad k>0 \end{cases}$$

For notational convenience, we define

$$\underline{T}_3(k) = \underline{\tilde{T}}_3(M-k)\,,$$

$$\underline{T}_4(k) = \underline{\tilde{T}}_4(M-k)\,,$$

$$\vec{n}_i(k) = \tilde{\vec{n}}_i(M-k-1)\,,$$

and

$$\vec{n}_f(k) = \tilde{\vec{n}}_f(M-k-1)\,.$$

Equation (14) becomes

$$\vec{s}^Q(M) = \sum_{k=0}^{M-1} \underline{A}(k)\underline{T}_3(k+1)\vec{f}^P + \sum_{k=0}^{M-1} \underline{A}(k)\underline{T}_3(k+1)\vec{n}_i(k) + \sum_{k=0}^{M-1} \underline{A}(k)\vec{n}_f(k)\,,$$

where

$$\underline{A}(k) = \begin{cases} \underline{I} & ; \quad k=0 \\ \underline{T}_4(1)\underline{T}_4(2)\cdots\underline{T}_4(k) & ; \quad k>0 \end{cases}.$$

Therefore the noisy steady state result is

$$\vec{s}^Q(\infty) = \vec{r}_1 + \vec{r}_2 + \vec{r}_3 + \vec{n}_d\,,$$

where

$$\vec{r}_1 = \sum_{k=0}^{\infty} \underline{A}(k)\underline{T}_3(k+1)\vec{f}^P$$

$$\vec{r}_2 = \sum_{k=0}^{\infty} \underline{A}(k)\underline{T}_3(k+1)\vec{n}_i(k)$$

$$\vec{r}_3 = \sum_{k=0}^{\infty} \underline{A}(k)\vec{n}_f(k)\,.$$

The expectation of $\vec{s}^Q(\infty)$ is

$$E[\vec{s}^Q(\infty)] = E[\vec{r}_1] + E[\vec{r}_2] + E[\vec{r}_3] + E[\vec{n}_d] = \sum_{k=0}^{\infty} \underline{T}_4^k\underline{T}_3\vec{f}^P = (\underline{I}-\underline{T}_4)^{-1}\vec{f}^P = \vec{f}^Q\,.$$

Thus, $\vec{s}^Q(\infty)$ is an unbiased estimate of our desired steady state result.

4.2. Second order analysis

The second order statistics of the noise are an indicator of the uncertainty of the final result. The covariance of $\vec{s}^Q(\infty)$ is

$$\begin{aligned}\mathrm{Cov}[\vec{s}^Q(\infty)] &= E[(\vec{r}_1+\vec{r}_2+\vec{r}_3+\vec{n}_d) \\ &\quad\times(\vec{r}_1+\vec{r}_2+\vec{r}_3+\vec{n}_d)^{\mathrm{T}}] - \vec{f}^Q\vec{f}^{Q\mathrm{T}} \\ &= \underline{C}_s + \underline{C}_i + \underline{C}_f + \underline{C}_d\,, \end{aligned} \tag{15}$$

where

$$\underline{C}_s = E[\vec{r}_1\vec{r}_1^{\mathrm{T}}] - \vec{f}^Q\vec{f}^{Q\mathrm{T}}\,,$$

$$\underline{C}_i = E[\vec{r}_2\vec{r}_2^{\mathrm{T}}]\,,$$

$$\underline{C}_f = E[\vec{r}_3\vec{r}_3^{\mathrm{T}}]$$

and

$$\underline{C}_d = E[\vec{n}_d\vec{n}_d^{\mathrm{T}}]\,.$$

The subscripts refer to respectively, system, input, feedback and detector noise.

Let the variance of each of the elements of $\underline{N}_3$ and $\underline{N}_4$ be σ_3^2 and σ_4^2 respectively. We can show that, if

$$\sigma_4^2 < \frac{1 - \|\underline{T}_4\|^2}{Q}\,,$$

then we guarantee the convergence of $\vec{s}^Q(\infty)$. The covariance of the system noise, $\underline{C}_s$, can be shown to be

$$\underline{C}_s = [\sigma_4^2\|\vec{f}^Q\|^2 + \sigma_3^2\|\vec{f}^P\|^2]\,\gamma(\underline{I}-\underline{T}_4^2)^{-1}\,, \tag{16}$$

where

$$1/\gamma = 1 - \sigma_4^2\,\mathrm{tr}[(\underline{I}-\underline{T}_4^2)^{-1}]\,.$$

By assumption, $\underline{C}_d = \sigma_d^2\underline{I}$ for the both static and time varying case. The other covariances of the static and time varying cases, however, are different. We will consider each case separately.

1. Static

Effects of static input noise are illustrated in *Fig. 8* for various noise levels. The floating neurons, in each case, have better resolution than

Fig. 8. Image reconstruction with static gaussian noise for 25% floating neurons. Here, and in the figure to follow noise is measured as the ratio's of the standard deviation to the full dynamic range of the gray level. The top row has 1% noise (iterations 0, 4, 10). The middle row has 5% noise (iterations 0, 5, 11). The bottom row has 10% noise (iterations 0, 5, 10).

the clamped neurons. Improvement of the clamped neuron values can be achieved by a projection onto the column space of $\underline{F}$.

We can show that, if the feedback and input noise are static, i.e. $E[\vec{n}_i\vec{n}_i^{\mathrm{T}}] = \sigma_i^2\underline{I}$ and $E[\vec{n}_f\vec{n}_f^{\mathrm{T}}] = \sigma_f^2\underline{I}$, then

$$\underline{C}_f = \sigma_f^2(\underline{I} - \underline{T}_4)^{-2} + \sigma_f^2\sigma_4^2 \operatorname{tr}[(\underline{I} - \underline{T}_4)^{-2}]\,\gamma(\underline{I} - \underline{T}_4^2)^{-1} \quad (17)$$

and

$$\underline{C}_i = \sigma_i^2(\underline{I} - \underline{T}_4)^{-1}\underline{T}_4 + \sigma_i^2\{\sigma_4^2 \operatorname{tr}[(\underline{I} - \underline{T}_4)^{-1}\underline{T}_4] + P\sigma_3^2\}\,\gamma(\underline{I} - \underline{T}_4^2)^{-1}\,. \quad (18)$$

We will use this result later in establishing reconstruction probability bounds.

2. Time varying

The effects of time varying input noise is illustrated in *Fig. 9.* In the time varying case, from the assumption of a white noise process, $E[\vec{n}_i(k)\vec{n}_i(l)] = \sigma_i^2 \underline{I}\delta_{k-l}$ and $E[\vec{n}_f(k)\vec{n}_f(l)] = \sigma_f^2 \underline{I}\delta_{k-l}$. Expressions for $\underline{C}_f$ and $\underline{C}_i$ follow as

$$\underline{C}_f = \sigma_f^2 \gamma (\underline{I} - \underline{T}_4^2)^{-1}$$

$$\begin{aligned}\underline{C}_i = {} & \sigma_i^2 (\underline{I} + \underline{T}_4)^{-1} \underline{T}_4 \\ & + \sigma_i^2 \{\sigma_4^2 \operatorname{tr}[(\underline{I} + \underline{T}_4)^{-1} \underline{T}_4] \\ & + P\sigma_3^2\} \gamma (\underline{I} - \underline{T}_4^2)^{-1} . \end{aligned} \tag{20}$$

We now have all of the information required to evaluate (15) for both the static and white noise cases.

Fig. 9. Image reconstruction with white noise for 25% unclamped neurons. The top row has 1% noise (iterations 0, 5, 12). The middle row has 5% noise (iterations 0, 3, 7). The bottom row has 10% noise (iterations 0, 2, 5).

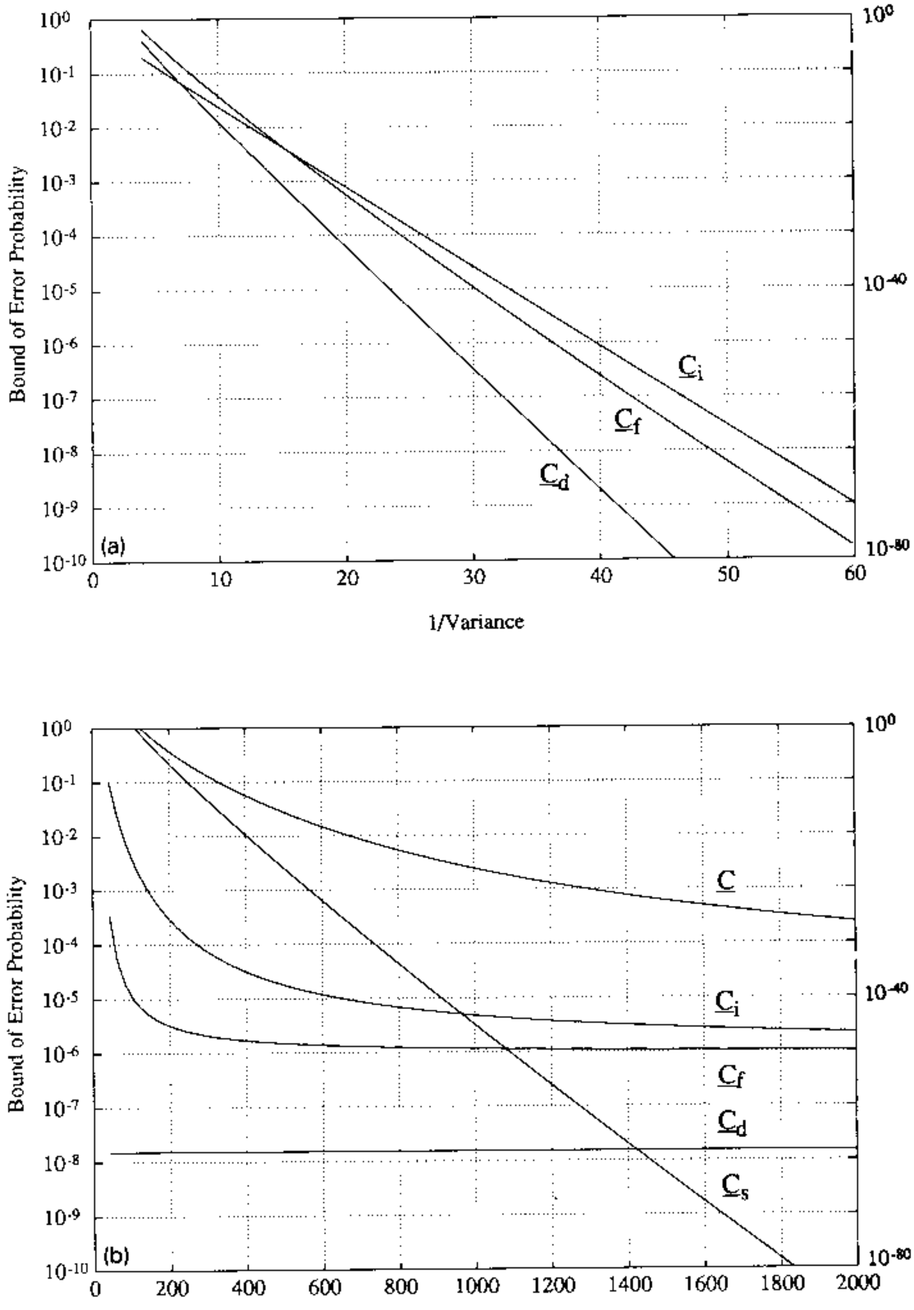


Fig. 10. The error probabilities for the APNN with white noise. The scale on right side is for $\underline{C}i$. (a) The error probabilities decrease with the decrease of σ_d, σ_i and σ_f. The value of σ_3 and σ_4 are $\frac{1}{36}$. (b) The error probabilities from $\underline{C}_i$, $\underline{C}_d$ and $\underline{C}_f$ saturate as σ_3 and σ_4 decrease, but that from $\underline{C}_s$ decreases. The value of σ_d, σ_i and σ_f are $\frac{1}{6}$.

4.3. Probability of error for bipolar library

In this section, we will discuss the probability of error for correct reconstruction of a bipolar library vector. The bipolar response is obtained from noisy output by a sign(·) operation.

$$\vec{s}_o^Q = \text{sign}[\vec{s}^Q(\infty)] .$$

The probability of error is

$$p_e = \sum_{n=1}^{N} Q_n[1 - p_{c_n}] ,$$

where Q_n is the priori probability of the library vector n and p_{c_n} is the probability that our classification is correct. We cannot, however, evaluate p_{c_n} easily. We can, though, compute the *union bound* [10] for p_{e_n}. For Gaussian noise, a probability of error bound for the nth library vector can be written as

$$p_{e_n} = 1 - p_{c_n} \leq \sum_{l=1}^{Q} \int_0^{\infty} \frac{\exp[-(x-1)^2/(2c_l)]}{\sqrt{2\pi c_l}} \, dx ,$$

where c_l is the lth diagonal element of covariance matrix $\underline{C}$.

4.3.1 Computer simulation

We demonstrate the probability of error bounds for the 5 randomly generated bipolar library of dimension $L = 30$. The average value of the diagonal elements of $\underline{T}$ is $\frac{1}{6}$. *Figure 10a* shows the probability of errors corresponding to $\underline{C}_d$, $\underline{C}_i$ and $\underline{C}_f$ vs. σ_d^2, σ_i^2 and σ_f^2 respectively when σ_3 and σ_4 are $\frac{1}{36}$. From *Fig. 10a*, the error bound decreases with the decrease of σ_d, σ_i and σ_f. *Figure 10b* shows the probability of errors

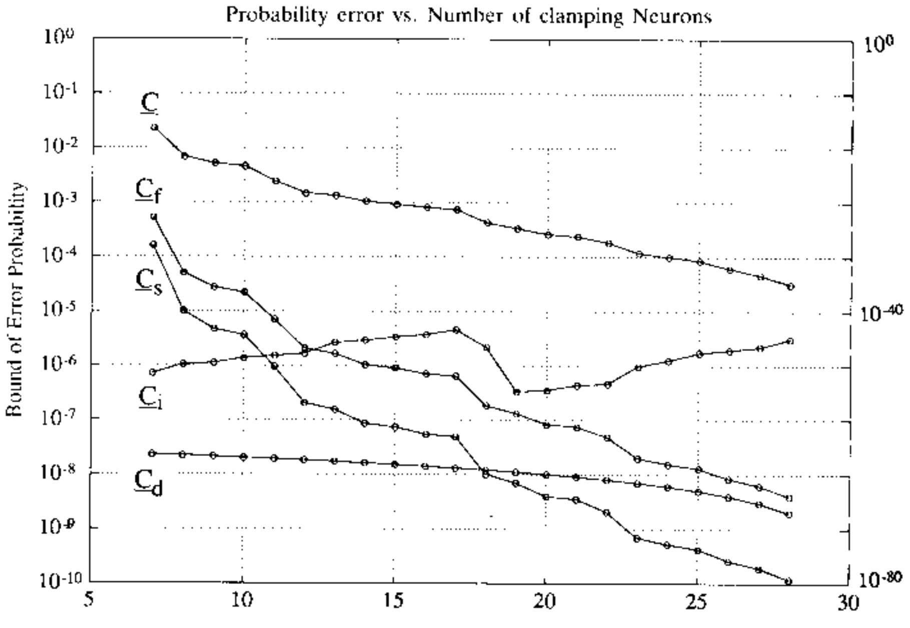


Fig. 11. The error probability as a function of the clamped neurons for the APNN. The error probabilities from $\underline{C}_f$, $\underline{C}_i$ and $\underline{C}_s$ decrease with an increase in the percentage of clamped neurons. The values of σ_d, σ_i and σ_f are $\frac{1}{6}$, σ_3 and σ_4 $\frac{1}{36}$. The scale on right side is for $\underline{C}_i$.

corresponding to $\underline{C}_d$, $\underline{C}_i$, $\underline{C}_f$ and $\underline{C}_s$ vs. σ_3^2 and σ_4^2 when σ_d, σ_i and σ_f are $\frac{1}{6}$. The probabilities of error from $\underline{C}_d$, $\underline{C}_i$ and $\underline{C}_f$ saturate with the decreases of σ_3 and σ_4, but that from $\underline{C}_s$ decreases as shown in *Fig. 10b*. *Figure 11* shows that the probability of error from $\underline{C}_d$, $\underline{C}_f$ and $\underline{C}_s$ decrease as the percentage of clamped neurons increases.

5. Conclusion

We have analyzed a number of characteristics of the homogeneous alternating projection neural network. The net's convergence becomes faster when the percentage of the clamping neurons increases and the number of the library vectors is decreased. For the bipolar library APNN, convergence can be achieved after a finite number of iterations. We proposed a per-step minimization method for the relaxation and showed its superiority to ℓ_2 and ℓ_∞ relaxation. We modified the training Gram–Schmidt method, so that no division or norm operation were required. The noise sensitivity of the APNN was analyzed. The noise performance improves as the percent of the clamped neurons increases.

Appendices

A. Proofs of theorems in convergence

A.1. The convergence rate and N and P

Lemma A.1. *Let $\underline{T}_4$ be the interconnect matrix between the last Q neurons and let $\underline{T}_4^*$ be the interconnect matrix between the last $Q-1$ neurons. Then $\|\underline{T}_4^*\| \leqslant \|\underline{T}_4\|$.*

Proof. Let

$$\underline{F}_Q = \begin{bmatrix} \vec{g}_{P+1}^{\mathrm{T}} \\ \underline{F}_{Q-1} \end{bmatrix},$$

where $\underline{F}^{\mathrm{T}} = [\vec{g}_1, \vec{g}_2, \ldots, \vec{g}_L]$ and let

$$(\underline{F}^{\mathrm{T}}\underline{F})^{-1} = \underline{G}^{\mathrm{T}}\underline{G}\,.$$

Then

$$\underline{F}_Q^{\mathrm{T}}\underline{F}_Q = \underline{F}_{Q-1}^{\mathrm{T}}\underline{F}_{Q-1} + \vec{g}_{P+1}\vec{g}_{P+1}^{\mathrm{T}}\,.$$

For every $\vec{x}$,

$$\begin{aligned}\vec{x}^{\mathrm{T}}\underline{G}\underline{F}_Q^{\mathrm{T}}\underline{F}_Q\underline{G}^{\mathrm{T}}\vec{x} &= \vec{x}^{\mathrm{T}}\underline{G}\underline{F}_{Q-1}^{\mathrm{T}}\underline{F}_{Q-1}\underline{G}^{\mathrm{T}}\vec{x} + [\vec{g}_{P+1}^{\mathrm{T}}\underline{G}^{\mathrm{T}}\vec{x}]^2 \\ &\geqslant \vec{x}^{\mathrm{T}}\underline{G}\underline{F}_{Q-1}^{\mathrm{T}}\underline{F}_{Q-1}\underline{G}^{\mathrm{T}}\vec{x}\,,\end{aligned}$$

so

$$\|\underline{G}\underline{F}_{Q-1}^{\mathrm{T}}\underline{F}_{Q-1}\underline{G}^{\mathrm{T}}\| \leqslant \|\underline{G}\underline{F}_Q^{\mathrm{T}}\underline{F}_Q\underline{G}^{\mathrm{T}}\|\,.$$

Therefore

$$\begin{aligned}\|\underline{T}_4^*\| &= \|\underline{F}_{Q-1}\underline{G}^{\mathrm{T}}\underline{G}\underline{F}_{Q-1}^{\mathrm{T}}\| \leqslant \|\underline{F}_Q\underline{G}^{\mathrm{T}}\underline{G}\underline{F}_Q^{\mathrm{T}}\| \\ &= \|\underline{T}_4\|\,. \quad \square\end{aligned}$$

Lemma A.2. *Let $\underline{T}_4^+$ be the interconnect matrix between the floating neurons of the library $\{\vec{f}_1, \vec{f}_2, \ldots, \vec{f}_{N+1}\}$. Then $\|\underline{T}_4\| \leqslant \|\underline{T}_4^+\|$.*

Proof. By the learning algorithm

$$\underline{T}^+ = \underline{T} + \frac{\vec{e}\vec{e}^{\mathrm{T}}}{\|\vec{e}\|^2}\,,$$

so

$$\underline{T}_4^+ = \underline{T}_4 + \frac{\vec{e}^Q\vec{e}^{Q\mathrm{T}}}{\|\vec{e}\|^2}\,,$$

where $\vec{e}^Q$ is the last Q values of $\vec{e}$. Also

$$\vec{x}^{\mathrm{T}}\underline{T}_4^+\vec{x} = \vec{x}^{\mathrm{T}}\underline{T}_4\vec{x} + \frac{[\vec{e}^{Q\mathrm{T}}\vec{x}]^2}{\|\vec{e}\|^2} \geqslant \vec{x}^{\mathrm{T}}\underline{T}_4\vec{x}\,.$$

Therefore

$$\|\underline{T}_4\| \leq \|\underline{T}_4'\| \quad \square.$$

A.2. Proof of Lemma 1

From (3), $\vec{s}^{Q}(M)$ will be

$$\vec{s}^{Q}(M) = \sum_{k=0}^{M} \underline{T}_4^k \underline{T}_3 \vec{f}^{P} \tag{21}$$

and

$$\vec{f}^{Q} = \sum_{k=0}^{\infty} \underline{T}_4^k \underline{T}_3 \vec{f}^{P}. \tag{22}$$

Thus

$$\begin{aligned}
&\vec{f}^{Q} - \vec{s}^{Q}(M-1) \\
&\quad = \sum_{k=M}^{\infty} \underline{T}_4^k \underline{T}_3 \vec{f}^{P} = \underline{T}_4^M \sum_{k=0}^{\infty} \underline{T}_4^k \underline{T}_3 \vec{f}^{P} \\
&\quad = \underline{T}_4^M (\underline{I} - \underline{T}_4)^{-1} \underline{T}_3 \vec{f}^{P} = (\underline{I} - \underline{T}_4)^{-1} \underline{T}_4^M \underline{T}_3 \vec{f}^{P}
\end{aligned}$$

because

$$\vec{s}^{Q}(M) - \vec{s}^{Q}(M-1) = \underline{T}_4^M \underline{T}_3 \vec{f}^{P}.$$

Therefore

$$\begin{aligned}
\vec{f}^{Q} - \vec{s}^{Q}(M-1) = (\underline{I} - \underline{T}_4)^{-1} [\vec{s}^{Q}(M) \\
- \vec{s}^{Q}(M-1)].
\end{aligned}$$

It follows that

$$\begin{aligned}
&\|\vec{f}^{Q} - \vec{s}^{Q}(M-1)\| \\
&\quad = \|(\underline{I} - \underline{T}_4)^{-1} [\vec{s}^{Q}(M) - \vec{s}^{Q}(M-1)]\| \\
&\quad \leq \|(\underline{I} - \underline{T}_4)^{-1}\| \cdot \|\vec{s}^{Q}(M) - \vec{s}^{Q}(M-1)\| \\
&\quad \leq \frac{\|\vec{s}^{Q}(M) - \vec{s}^{Q}(M-1)\|}{1 - \|\underline{T}_4\|} < 1.
\end{aligned}$$

Therefore $|f_i - s_i(M-1)| < 1$ for $P+1 \leq i \leq L$. If $f_i = 1$, then $0 < s_i(M-1) < 2$. Therefore

$$\mathrm{sign}[s_i(M-1)] = 1.$$

If $f_i = -1$, then $-2 < s_i(M-1) < 0$. Therefore

$$\mathrm{sign}[s_i(M-1)] = -1$$

and our proof is complete. $\square$

A.3. Proof of Lemma 2

From (21) and (22), $\vec{f}^{Q} - \vec{s}^{Q}(M-1)$ will be

$$\begin{aligned}
\vec{f}^{Q} - \vec{s}^{Q}(M-1) &= \underline{T}_4^M (\underline{I} - \underline{T}_4)^{-1} \underline{T}_3 \vec{f}^{P} \\
&= \underline{T}_4^M \vec{f}^{Q},
\end{aligned} \tag{23}$$

so the norm of the error will be

$$\begin{aligned}
\|\vec{f}^{Q} - \vec{s}^{Q}(M-1)\| &\leq \|\underline{T}_4^M\| \cdot \|\vec{f}^{Q}\| \\
&= \|\underline{T}_4\|^M \|\vec{f}^{Q}\|.
\end{aligned}$$

Here, $\|\vec{f}^{Q}\| = \sqrt{Q}$. Assuming

$$\|\underline{T}_4\|^M < Q^{-1/2}$$

we have

$$\|\vec{f}^{Q} - \vec{s}^{Q}(M-1)\| < Q^{-1/2}\sqrt{Q} = 1$$

and our proof is complete. $\square$

B. Proof of algorithm convergence

We take

$$\begin{aligned}
&\vec{u}(M+1) - \vec{s}(M) \\
&\quad = [\theta(M)\vec{w}(M+1) + (1 - \theta(M))\vec{u}(M)] \\
&\quad\quad - [\theta(M)\vec{u}(M) \\
&\quad\quad + (1 - \theta(M))\vec{s}(M-1)] \\
&\quad = \theta(M)[(\vec{w}(M+1) - \vec{u}(M)) \\
&\quad\quad - (\vec{u}(M) - \vec{s}(M-1)] + [\vec{u}(M) - \vec{s}(M-1)],
\end{aligned}$$

where

$$\vec{w}(M+1) - \vec{u}(M) = \eta \underline{T}\vec{u}(M) - \eta \underline{T}\vec{s}(M-1) = \eta \underline{T}[\vec{u}(M) - \vec{s}(M-1)]$$

$$\vec{u}(M+1) - \vec{s}(M) = [\theta(M)(\eta \underline{T} - \underline{I}) + \underline{I}] \cdot [\vec{u}(M) - \vec{s}^{Q}(M-1)] .$$

Let $\vec{x}(M) = \vec{u}^{Q}(M) - \vec{s}^{Q}(M-1)$. Because of $\|\vec{u}(M) - \vec{s}(M-1)\| = \|\vec{x}\|$, then

$$\|\vec{x}(M+1)\|^2 = \|\vec{x}(M)\|^2 - 2\theta(M)\vec{x}(M)^{\mathrm{T}}(\underline{I} - \underline{T}_4)\vec{x}(M) + \theta(M)^2\|(\underline{I} - \underline{T}_4)\vec{x}(M)\|^2$$

and

$$\theta(M) = \frac{\vec{x}(M)^{\mathrm{T}}(\underline{I} - \underline{T}_4)\vec{x}(M)}{\|(\underline{I} - \underline{T}_4)\vec{x}(M)\|^2} .$$

From the above two equations,

$$\frac{\|x(M+1)\|^2}{\|\vec{x}(M)\|^2} = 1 - \frac{[\vec{x}(M)^{\mathrm{T}}(\underline{I} - \underline{T}_4)\vec{x}(M)]^2}{\|\vec{x}(M)\|^2\|(\underline{I} - \underline{T}_4)\vec{x}(M)\|^2} . \tag{24}$$

Consider the spectral decomposition matrix $\underline{E}_i$. Let

$$k_i = \frac{\|\underline{E}_i\vec{x}(M)\|^2}{\|\vec{x}(M)\|^2}$$

then clearly,

$$\sum_{i=1}^{t} k_i = 1 .$$

Also

$$\vec{x}(M)^{\mathrm{T}}(\underline{I} - \underline{T}_4)\vec{x}(M) = \left[\sum_{i=1}^{t} \underline{E}_i\vec{x}(M)\right]^{\mathrm{T}}(\underline{I} - \underline{T}_4)\left[\sum_{i=1}^{t} \underline{E}_i\vec{x}(M)\right]$$

$$= \sum_{i=1}^{t} (1 - \lambda_i)\vec{x}(M)^{\mathrm{T}}\underline{E}_i\vec{x}(M) = \sum_{i=1}^{t} (1 - \lambda_i)\|\underline{E}_i\vec{x}(M)\|^2 = \sum_{i=1}^{t} (1 - \lambda_i k_i)\|\vec{x}(M)\|^2$$

and

$$\|(\underline{I} - \underline{T}_4)\vec{x}(M)\|^2 = \|(\underline{I} - \underline{T}_4)\sum_{i=1}^{t} \underline{E}_i\vec{x}(M)\|^2 = \|\sum_{i=1}^{t} (1 - \lambda_i)\underline{E}_i\vec{x}(M)\|^2 = \sum_{i=1}^{t} (1 - \lambda)^2 k_i\|\vec{x}(M)\|^2 .$$

Therefore (24) will be

$$\frac{\|\vec{x}(M+1)\|^2}{\|\vec{x}(M)\|^2} = 1 - \frac{\left[\sum_{i=1}^{t} (1 - \lambda_i)k_i\right]^2}{\sum_{i=1}^{t} (1 - \lambda_i)^2 k_i} .$$

From Cauchy's inequality,

$$\frac{\sum_{i=1}^{t} (1 - \lambda_i)^2 k_i}{\left[\sum_{i=1}^{t} (1 - \lambda_i)k_i\right]^2} \leqslant \max_{i,j} \frac{[(1 - \lambda_i) + (1 - \lambda_j)]^2}{4(1 - \lambda_i)(1 - \lambda_j)} ,$$

Therefore (24) is bounded

$$\frac{\|\vec{x}(M+1)\|^2}{\|\vec{x}(M)\|^2} \leqslant \max_{i,j}\left[\frac{(1 - \lambda_i) - (1 - \lambda_j)}{(1 - \lambda_i)(1 - \lambda_j)}\right]^2 = \frac{(\lambda_{\max} - \lambda_{\min})/2}{1 - (\lambda_{\max} + \lambda_{\min})/2} < 1 .$$

We thus conclude that

$$\lim_{M\to\infty} \|\vec{x}(M)\| = 0 .$$

Because $\vec{x}(M)$ lies in a compact set, $\vec{x}(M)\to\vec{0}$. Therefore $\vec{s}^{\,Q}(M)$ will be

$$\vec{s}^{\,Q}(\infty) = \underline{T}_4\vec{s}^{\,Q}(\infty) + \underline{T}_3\vec{f}^{\,P}$$

or

$$\vec{s}^{\,Q}(\infty) = (\underline{I} - \underline{T}_4)^{-1}\underline{T}_3\vec{f}^{\,P} ,$$

and our proof is complete. □

C. Proof of training procedure

C.1. Basic lemmas

Lemma C.1. *We assume that $\underline{T}_n(0)$ is symmetric, positive semidefinite and $\|\underline{T}_n(0)\| < 1$.*

Let

$$y_i = \frac{\beta^i\vec{\epsilon}_i^{\,\mathrm{T}}\vec{f}_n}{B^2} ,$$

then

1. $0 \leqslant y_i \leqslant 1$.
2. $\|\vec{\epsilon}_i\| \leqslant \|\vec{\epsilon}_{i-1}\|$.

Proof.

$$\vec{\epsilon}_i = \vec{f}_n - \underline{T}_n(i)\vec{f}_n$$
$$= \vec{f}_n - \underline{T}_n(i-1)\vec{f}_n - \frac{\beta^{i-1}\vec{\epsilon}_{i-1}^{\,\mathrm{T}}\vec{f}_n\vec{\epsilon}_{i-1}}{B^2}$$
$$= \left[1 - \frac{\beta^{i-1}\vec{\epsilon}_{i-1}^{\,\mathrm{T}}\vec{f}_n}{B^2}\right]\vec{\epsilon}_{i-1} .$$

Therefore,

$$y_i = \beta[1 - y_{i-1}]y_{i-1} .$$

Now, we will establish the bounds for y_i. Define

$$\vec{\epsilon}_0 = \vec{f}_n - \underline{T}_n(0)\vec{f}_n = [\underline{I} - \underline{T}_n(0)]\vec{f}_n .$$

Thus

$$y_0 = \frac{\vec{f}_n^{\,\mathrm{T}}[\underline{I} - \underline{T}_n(0)]\vec{f}_n}{B^2}$$

so

$$0 \leqslant y_0 \leqslant \frac{\|\vec{f}_n\|^2}{B^2} \leqslant 1 .$$

If $0 \leqslant y_{i-1} \leqslant 1$, then

$$0 \leqslant y_i \leqslant \frac{\beta}{4} \leqslant 1 .$$

Therefore, $0 \leqslant y_i \leqslant 1$. Since

$$\vec{\epsilon}_i = (1 - y_{i-1})\vec{\epsilon}_{i-1}$$

then

$$\|\vec{\epsilon}_i\| = (1 - y_{i-1})\|\vec{\epsilon}_{i-1}\| \leqslant \|\vec{\epsilon}_{i-1}\| \quad \square$$

C.2. Proof that $\|\underline{T}_N(I)\| < 1$

Lemma C.2. *If $\|\underline{T}_n(i-1)\| \leqslant 1$, then $\|\underline{T}_n(i)\| \leqslant 1$.*

Proof. For every $\vec{x}$,

$$e(i) = \vec{x}^{\,\mathrm{T}}\vec{x} - \vec{x}^{\,\mathrm{T}}\underline{T}_n(i)\vec{x} = \vec{x}^{\,\mathrm{T}}[\underline{I} - \underline{T}_n(i)]\vec{x}$$
$$= \vec{x}^{\,\mathrm{T}}\left[\underline{I} - \underline{T}_n(i-1) - \frac{\beta^{i-1}\vec{\epsilon}_{i-1}\vec{\epsilon}_{i-1}^{\,\mathrm{T}}}{B^2}\right]\vec{x} .$$

Let $\underline{S}(i) = \underline{I} - \underline{T}_n(i-1)$. Then

$$e(i) = \vec{x}^{\,\mathrm{T}}\underline{S}(i)\vec{x} - \frac{[\vec{f}_n^{\,\mathrm{T}}\underline{S}(i)\vec{x}]^2}{\vec{f}_n^{\,\mathrm{T}}\underline{S}(i)\vec{f}_n}\, y_i .$$

Let $\langle\vec{x}|\vec{y}\rangle = \vec{x}^{\,\mathrm{T}}\underline{S}(i)\vec{y}$. Then $\langle\vec{x}|\vec{y}\rangle$ satisfies the inner product condition because $\|\underline{T}_n(i-1)\| \leqslant 1$.

$$e(i) = \langle\vec{x}|\vec{x}\rangle - \frac{[\langle\vec{f}_n|\vec{x}\rangle]^2}{\langle\vec{f}_n|\vec{f}_n\rangle} + (1 - y_i)\frac{[\langle\vec{f}_n|\vec{x}\rangle]^2}{\langle\vec{f}_n|\vec{f}_n\rangle} .$$

From Lemma (C.1) and Schwarz's inequality, $e(i) \geq 0$. Thus

$$\vec{x}^{\mathrm{T}}\vec{x} \geq \vec{x}^{\mathrm{T}}\underline{T}_n(i)\vec{x}$$

and our proof is complete □

Lemma C.3. *$\|\underline{T}_n(i)\| \leq 1$ for all n and i.*

Proof. Since $\underline{T}_1(0) = \underline{0}$, it follows that $\|\underline{T}_1(0)\| \leq 1$.

From Lemma (C.2), clearly $\|\underline{T}_1(i)\| \leq 1$ for all i.

If $\|\underline{T}_n(I)\| \leq 1$, then $\|\underline{T}_{n+1}(0)\| \leq 1$ because $\underline{T}_n(I) = \underline{T}_{n+1}(0)$, and our proof is complete. □

C.3. Proof of error bound

Lemma C.4. *Let $\vec{\epsilon}_i = \vec{f}_n - \underline{T}_n(i)\vec{f}_n$, then*

$$\|\vec{\epsilon}_i\|^2 \leq \frac{B^2}{\beta^i} ,$$

Proof.

$$\begin{aligned}\|\vec{\epsilon}_i\|^2 &\leq \vec{f}_n^{\mathrm{T}}[\underline{I} - \underline{T}_n(i)]\vec{f}_n \\ &= \vec{\epsilon}_i^{\mathrm{T}}\vec{f}_n = \frac{y_i B^2}{\beta^i} \leq \frac{B^2}{\beta^i} . \quad \square\end{aligned}$$

Lemma C.5. *For every library vector $\vec{f}_n$,*

$$\|\vec{f}_n - \underline{T}_w\vec{f}_n\| \leq \frac{B}{\beta^{I/2}} .$$

Proof.

$$\begin{aligned}\|\vec{f}_n - \underline{T}_w\vec{f}_n\|^2 &\leq \vec{f}_n^{\mathrm{T}}[\underline{I} - \underline{T}_w]\vec{f}_n \\ &= \vec{f}_n^{\mathrm{T}}[\{\underline{I} - \underline{T}_n(I)\} + \{\underline{T}_n(I) - \underline{T}_w\}]\vec{f}_n \\ &= \vec{f}_n^{\mathrm{T}}[\underline{I} - \underline{T}_n(I)]\vec{f}_n - \vec{f}_n^{\mathrm{T}}[\underline{T}_w - \underline{T}_n(I)]\vec{f}_n \\ &\leq \vec{f}_n^{\mathrm{T}}[\underline{I} - \underline{T}_n(I)]\vec{f}_n \leq \frac{B^2}{\beta^I} ,\end{aligned}$$

and our proof is complete. □

Lemma C.6. *If $\vec{f}$ is the linear combination of the columns of $\underline{F}$,*

$$\vec{f} = \sum_{n=1}^{N} c_n\vec{f}_n ,$$

then

$$\|\vec{f} - \underline{T}_w\vec{f}\| \leq \frac{B}{\beta^{I/2}} \sum_{n=1}^{N} |c_n| .$$

Proof.

$$\begin{aligned}\|\vec{f} - \underline{T}_w\vec{f}\| &= \|(\underline{I} - \underline{T}_w)\vec{f}\| \\ &= \|\sum_{n=1}^{N} c_n(\underline{I} - \underline{T}_w)\vec{f}\| \\ &\leq \sum_{n=1}^{N} |c_n|\|(\underline{I} - \underline{T}_w)\vec{f}\| .\end{aligned}$$

From Lemma (C.5),

$$\|\vec{f} - \underline{T}_w\vec{f}\| \leq \frac{B}{\beta^{I/2}} \sum_{n=1}^{N} |c_n| ,$$

and our proof is complete. □

C.4. Eigenvalues of $\underline{T}_w$

Theorem C.1. *Let λ_w be the nonzero eigenvalue of $\underline{T}_w$, then λ_w satisfies the following inequality:*

$$0 \leq 1 - \lambda_w \leq \frac{1 - \sqrt{1 - 4\xi}}{2} ,$$

where ξ is

$$\xi = \frac{B}{\beta^{I/2}} \sqrt{\frac{N}{\lambda_F}}$$

and λ_F is the smallest nonzero eigenvalues of $\underline{F}^{\mathrm{T}}\underline{F}$.

Proof. For every $\vec{f}$, there is $\vec{c}$ such that

$$\underline{F}\vec{c} = \underline{T}_w\vec{f} = \sum_{n=1}^{N} c_n\vec{f}_n . \tag{25}$$

We now apply singular value decomposition. $\underline{F}$ can be written as

$$\underline{F} = \underline{P}_F \underline{D}_F \underline{Q}_F . \tag{26}$$

From (25) and (26),

$$\vec{c} = \underline{Q}_F^{\mathrm{T}} \underline{D}_F^{-1} \underline{P}_F^{\mathrm{T}} \underline{T}_w \vec{f} .$$

The norm of $\vec{c}$ will be

$$\begin{aligned} \|\vec{c}\| &= \|\underline{Q}_F^{\mathrm{T}} \underline{D}_F^{-1} \underline{P}_F^{\mathrm{T}} \underline{T}_w \vec{f}\| \\ &\leq \|\underline{Q}_F^{\mathrm{T}} \underline{D}_F^{-1} \underline{P}_F^{\mathrm{T}}\| \cdot \|\underline{T}_w\| \cdot \|\vec{f}\| \leq \frac{\|\vec{f}\|}{\sqrt{\lambda_F}} . \end{aligned}$$

Using Schwarz's inequality,

$$\sum_{n=1}^{N} |c_n| \leq \sqrt{N} \|\vec{c}\| .$$

Therefore

$$\sum_{n=1}^{N} |c_n| \leq \sqrt{\frac{N}{\lambda_F}} \cdot \|\vec{f}\| .$$

Now, we evaluate

$$\begin{aligned} \|(\underline{T}_w - \underline{T}_w^2)\vec{f}\| &= \|(\underline{I} - \underline{T}_w)\underline{T}_w \vec{f}\| \\ &= \|(\underline{I} - \underline{T}_w)\underline{F}\vec{c}\| . \end{aligned}$$

From Lemma (C.6),

$$\|(\underline{T}_w - \underline{T}_w^2)\vec{f}\| \leq \frac{B}{\beta^{1/2}} \sum_{n=1}^{N} |c_n| ,$$

so

$$\frac{\|(\underline{T}_w - \underline{T}_w^2)\vec{f}\|}{\|\vec{f}\|} \leq \xi .$$

Therefore the norm of $\underline{T}_w - \underline{T}_w^2$ will be

$$\|\underline{T}_w - \underline{T}_w^2\| \leq \xi .$$

We will now evaluate the bound of the λ_w. Because $\underline{T}_w$ is symmetric, λ_w satisfies

$$\lambda_w - \lambda_w^2 \leq \xi .$$

If $\xi < 0.25$, then, because $\mathrm{rank}(\underline{T}_w) = \mathrm{rank}(\underline{F})$, the nonzero eigenvalue of $\underline{T}_w$ will be

$$\frac{1 + \sqrt{1 - 4\xi}}{2} \leq \lambda_w \leq 1 . \quad \square$$

D. Noise analysis

D.1. General lemmas

We assume that all elements of $\underline{N}$ are identically independent white noise. Also, $\underline{D}$ is symmetric and positive semidefinite. All matrices are elements of $R^{Q \times Q}$.

Some properties are listed below:

1. $E[\underline{N}\underline{A}\underline{N}^{\mathrm{T}}] = \sigma^2 \,\mathrm{tr}(\underline{A})\underline{I}$.
2. $E[(\underline{B} + \underline{N})\underline{A}(\underline{B} + \underline{N})^{\mathrm{T}}]$
   $= \underline{B}\underline{A}\underline{B}^{\mathrm{T}} + \sigma^2 \,\mathrm{tr}(\underline{\mathrm{A}})\underline{I}$.
3. $E[\underline{A}(k)\underline{D}\underline{A}^{\mathrm{T}}(k)]$ is symmetric and positive semidefinite.
4. If $\underline{D}\underline{T}_4 = \underline{T}_4\underline{D}$, then $E[\underline{A}(k)\underline{D}\underline{A}^{\mathrm{T}}(k)]\underline{T}_4 = \underline{T}_4 E[\underline{A}(k)\underline{D}\underline{A}^{\mathrm{T}}(k)]$.
5. If
   $$\sum_{k=0}^{\infty} \mathrm{tr}\{\mathrm{E}[\underline{A}(k)\underline{A}^{\mathrm{T}}(k)]\}$$
   converges, then
   $$\sum_{k=0}^{\infty} E[\underline{A}(k)\underline{D}\underline{A}^{\mathrm{T}}(k)]$$
   converges.

Lemma D.1. *If*

$$\sigma_4^2 < \frac{1 - \|\underline{T}_4\|^2}{Q}$$

and $\underline{D}\underline{T}_4 = \underline{T}_4\underline{D}$, *then*

$$\mathrm{tr}\left\{\sum_{k=0}^{\infty} E[\underline{A}(k)\underline{D}\underline{A}^{\mathrm{T}}(k)]\right\} = \gamma\,\mathrm{tr}[(\underline{I}-\underline{T}_4^2)^{-1}\underline{D}]$$

and

$$1/\gamma = 1 - \sigma_4^2\,\mathrm{tr}[(\underline{I}-\underline{T}_4^2)^{-1}].$$

Proof.

1. We wish to establish the convergence of

$$\mathrm{tr}\left\{\sum_{k=0}^{\infty} E[\underline{A}(k)\underline{D}\underline{A}^{\mathrm{T}}(k)]\right\}:$$

$$\begin{aligned} &E[\underline{A}(k)\underline{A}^{\mathrm{T}}(k)] \\ &\quad = \underline{T}_4^2 E[\underline{A}(k-1)\underline{A}^{\mathrm{T}}(k-1)] \\ &\qquad + \sigma_4^2\,\mathrm{tr}\{E[\underline{A}(k-1)\underline{A}^{\mathrm{T}}(k-1)]\}\underline{I}\,. \end{aligned}$$

Let $c(k) = \mathrm{tr}\{E[\underline{A}(k)\underline{A}^{\mathrm{T}}(k)]\}$, then

$$\begin{aligned} c(k) &= \sigma_4^2 Q c(k-1) \\ &+ \mathrm{tr}\{\underline{T}_4^2 E[\underline{A}(k-1)\underline{A}^{\mathrm{T}}(k-1)]\} \\ &\leqslant \sigma_4^2 Q c(k-1) + \|\underline{T}_4\|^2 c(k-1) \\ &= c(k-1)[\|\underline{T}_4\|^2 + \sigma_4^2 Q] \\ &\cdot \sum_{k=0}^{\infty} \mathrm{tr}\{E[\underline{A}(k)\underline{A}^{\mathrm{T}}(k)]\} \\ &= \sum_{k=0}^{\infty} c(k) \leqslant \frac{Q}{1-\|\underline{T}_4\|^2-\sigma_4^2 Q}\,. \end{aligned}$$

Therefore,

$$\sum_{k=0}^{\infty} \mathrm{tr}\{E[\underline{A}(k)\underline{D}\underline{A}^{\mathrm{T}}(k)]\}$$

converges.

2. Evaluation of

$$\begin{aligned} &\sum_{k=0}^{\infty} E[\underline{A}(k)\underline{D}\underline{A}^{\mathrm{T}}(k)] \\ &\quad = \sum_{k=0}^{\infty} E[\underline{T}_4(1)\underline{A}(k-1)\underline{D}\underline{A}^{\mathrm{T}}(k-1) \\ &\qquad \cdot \underline{T}_4^{\mathrm{T}}(1)] + \underline{D} \\ &\quad = \sum_{k=0}^{\infty} \underline{T}_4^2 E[\underline{A}_0(k)\underline{D}\underline{A}_0^{\mathrm{T}}(k)] \\ &\qquad + \sigma_4^2\,\mathrm{tr}\left\{\sum_{k=0}^{\infty} E[\underline{A}_0(k)\underline{D}\underline{A}_0^{\mathrm{T}}(k)]\right\}\underline{I} + \underline{D}\,, \end{aligned}$$

where

$$\underline{A}_0(k) = \begin{cases} \underline{I} & k=1 \\ \underline{T}_4(2)\underline{T}_4(3)\cdots\underline{T}_4(k) & k>1 \end{cases}$$

gives

$$\begin{aligned} &\sum_{k=0}^{\infty} E[\underline{A}_0(k)\underline{D}\underline{A}_0^{\mathrm{T}}(k)] \\ &\quad = \left[\underline{D} + \sigma_4^2\,\mathrm{tr}\left\{\sum_{k=0}^{\infty} E[\underline{A}_0(k)\underline{D}\underline{A}_0^{\mathrm{T}}(k)]\right\}\underline{I}\right] \\ &\qquad \cdot (\underline{I}-\underline{T}_4^2)^{-1}\,. \end{aligned} \tag{27}$$

Because the noise process is the same all time, we obtain

$$\mathrm{tr}\left\{\sum_{k=0}^{\infty} E[\underline{A}(k)\underline{D}\underline{A}^{\mathrm{T}}(k)]\right\} = \gamma\,\mathrm{tr}[(\underline{I}-\underline{T}_4^2)^{-1}\underline{D}]\,. \quad \square$$

Lemma D.2. *If*

$$\sum_{k=0}^{\infty} \mathrm{tr}\{E[\underline{A}(k)\underline{A}^{\mathrm{T}}(k)]\}$$

converges and $\underline{D}\underline{T}_4 = \underline{T}_4\underline{D}$, *then*

$$\begin{aligned} &\sum_{k=0}^{\infty} E[\underline{A}(k)\underline{D}\underline{A}^{\mathrm{T}}(k)] \\ &= (\underline{I}-\underline{T}_4^2)^{-1}\underline{D} + \sigma_4^2\,\mathrm{tr}[(\underline{I}-\underline{T}_4^2)^{-1}\underline{D}]\gamma(\underline{I}-\underline{T}_4^2)^{-1}\,. \end{aligned}$$

Proof. The proof follows straightforwardly from (27). $\square$

Lemma D.3.

$$\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} E[\underline{A}(k)\underline{D}\underline{A}^{\mathrm{T}}(l)]$$

$$= (\underline{I} - \underline{T}_4)^{-1}\underline{D}(\underline{I} - \underline{T}_4)^{-1} + \sigma_4^2 \operatorname{tr}[(\underline{I} - \underline{T}_4)^{-1}\underline{D}(\underline{I} - \underline{T}_4)^{-1}]\gamma(\underline{I} - \underline{T}_4^2)^{-1} . \tag{28}$$

Proof.

$$\begin{aligned}
\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} E[\underline{A}(k)\underline{DA}^{\mathrm{T}}(l)]
&= \sum_{k=0}^{\infty}\sum_{s=0}^{\infty} E[\underline{A}(k)\underline{DA}^{\mathrm{T}}(k+s)] + \sum_{t=0}^{\infty}\sum_{l=0}^{\infty} E[\underline{A}(l+t)\underline{DA}^{\mathrm{T}}(l)] - \sum_{k=0}^{\infty} E[\underline{A}(k)\underline{DA}^{\mathrm{T}}(k)] \\
&= \sum_{k=0}^{\infty}\sum_{s=0}^{\infty} E[\underline{A}(k)\underline{DT}_4^{s}\underline{A}^{\mathrm{T}}(k)] + \sum_{l=0}^{\infty}\sum_{t=0}^{\infty} E[\underline{A}(l)\underline{T}_4^{t}\underline{DA}^{\mathrm{T}}(l)] - \sum_{k=0}^{\infty} E[\underline{A}(k)\underline{DA}^{\mathrm{T}}(k)] \\
&= \sum_{k=0}^{\infty} E[\underline{A}(k)\underline{D}(\underline{I} - \underline{T}_4)^{-1}\underline{A}^{\mathrm{T}}(k)] + \sum_{k=0}^{\infty} E[\underline{A}(k)(\underline{I} - \underline{T}_4)^{-1}\underline{DA}^{\mathrm{T}}(k)] - \sum_{k=0}^{\infty} E[\underline{A}(k)\underline{DA}^{\mathrm{T}}(k)] .
\end{aligned}$$

Let $\underline{G} = (\underline{I} - \underline{T}_4)^{-1}\underline{D}(\underline{I} - \underline{T}_4)^{-1}$, then

$$\underline{D}(\underline{I} - \underline{T}_4)^{-1} + (\underline{I} - \underline{T}_4)^{-1} - \underline{D} = \underline{G} - \underline{T}_4\underline{GT}_4 .$$

$$\begin{aligned}
\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} E[\underline{A}(k)\underline{DA}^{\mathrm{T}}(l)] &= \sum_{k=0}^{\infty} E[\underline{A}(k)\underline{GA}^{\mathrm{T}}(k)] - \sum_{k=0}^{\infty} E[\underline{A}(k)\underline{T}_4\underline{GT}_4\underline{A}^{\mathrm{T}}(k)] \\
&= \underline{G} + \sum_{k=0}^{\infty} E[\underline{A}(k-1)\underline{T}_4(k)\underline{GT}_4^{\mathrm{T}}(k)\underline{A}^{\mathrm{T}}(k-1)] - \sum_{k=0}^{\infty} E[\underline{A}(k)\underline{T}_4\underline{GT}_4\underline{A}^{\mathrm{T}}(k)] \\
&= \underline{G} + \sigma_4^2 \operatorname{tr}(\underline{G}) \sum_{k=0}^{\infty} E[\underline{A}(k)\underline{A}^{\mathrm{T}}(k)] .
\end{aligned}$$

The proof follows from Lemma (D2). □

D.2. The second order statistics of the APNN

D.2.1. Calculation of $\underline{C}_s$

$$\begin{aligned}
\underline{C}_s &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \{E[\underline{A}(k)\underline{T}_3(k+1)\vec{f}^{P}\vec{f}^{P\mathrm{T}}\underline{T}_3^{\mathrm{T}}(l+1)\underline{A}^{\mathrm{T}}(l)] - \underline{T}_4^{k}\underline{T}_3\vec{f}^{P}\vec{f}^{P\mathrm{T}}\underline{T}_3^{\mathrm{T}}\underline{T}_4^{l}\} \\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} E[\underline{A}(k)\underline{T}_3\vec{f}^{P}\vec{f}^{P\mathrm{T}}\underline{T}_3^{\mathrm{T}}\underline{A}^{\mathrm{T}}(l)] - \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} [\underline{T}_4^{k}\underline{T}_3\vec{f}^{P}\vec{f}^{P\mathrm{T}}\underline{T}_3^{\mathrm{T}}\underline{T}_4^{l}] + \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} E[\underline{A}(k)\sigma_3^2 \operatorname{tr}[\vec{f}^{P}\vec{f}^{P\mathrm{T}}]\underline{A}^{\mathrm{T}}(l)] .
\end{aligned}$$

Let

$$\underline{G} = (\underline{I} - \underline{T}_4)^{-1}\underline{T}_3\vec{f}^{P}\vec{f}^{P\mathrm{T}}\underline{T}_3^{\mathrm{T}}(\underline{I} - \underline{T}_4)^{-1} .$$

Then from Lemma D3, we have

$$\begin{aligned}
\underline{C}_s &= \left\{\underline{G} - \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} [\underline{T}_4^{k}\underline{T}_3\vec{f}^{P}\vec{f}^{P\mathrm{T}}\underline{T}_3^{\mathrm{T}}\underline{T}_4^{l}]\right\} + [\sigma_4^2 \operatorname{tr}(\underline{G}) + \sigma_3^2 \operatorname{tr}(\vec{f}^{P}\vec{f}^{P\mathrm{T}})]\gamma(\underline{I} - \underline{T}_4^2)^{-1} \\
&= [\sigma_4^2 \operatorname{tr}(\vec{f}^{Q}\vec{f}^{Q\mathrm{T}}) + \sigma_3^2 \operatorname{tr}(\vec{f}^{P}\vec{f}^{P\mathrm{T}})]\gamma(\underline{I} - \underline{T}_4^2)^{-1} .
\end{aligned}$$

From which (16) immediately follows.

D.2.2. Static noise

$$\underline{C}_i = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} E[\underline{A}(k)\underline{T}_3(k+1)\vec{n}_i(k)\vec{n}_i^{\mathrm{T}}(l)\underline{T}_3^{\mathrm{T}}$$

$$(l+1)\underline{T}^{\mathrm{T}}(l)]$$

$$= \sigma_i^2 \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} E[\underline{A}(k)\underline{T}_3(k+1)\underline{A}^{\mathrm{T}}(l)]$$

$$= \sigma_i^2 \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} E[\underline{A}(k)\underline{T}_3\underline{T}_3^{\mathrm{T}}\underline{A}^{\mathrm{T}}(l)]$$

$$+ P\sigma_i^2\sigma_3^2 \sum_{k=0}^{\infty} E[\underline{A}(k)\underline{A}^{\mathrm{T}}(k)]\,.$$

Let

$$\underline{G} = (\underline{I}-\underline{T}_4)^{-1}\underline{T}_3\underline{T}_3^{\mathrm{T}}(\underline{I}-\underline{T}_4)^{-1} = (\underline{I}-\underline{T}_4)^{-1}\underline{T}_4$$

because

$$\underline{T}_3\underline{T}_3^{\mathrm{T}} = \underline{T}_4 - \underline{T}_4^2\,,$$

and by Lemma (D3), we obtain (18)

$$\underline{C}_f = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} E[\underline{A}(k)\vec{n}_f(k)\vec{n}_f^{\mathrm{T}}(l)\underline{A}^{\mathrm{T}}(l)]$$

$$= \sigma_f^2 \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} E[\underline{A}(k)\underline{A}^{\mathrm{T}}(l)]\,.$$

Using Lemma D3 results in (17).

D.2.3. Time varying noise

$$\underline{C}_i = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} [\underline{A}(k)\underline{T}_3(k+1)\vec{n}_i(k)\vec{n}_i^{\mathrm{T}}(l)\underline{T}_3^{\mathrm{T}}$$

$$(l+1)\underline{A}^{\mathrm{T}}(l)]$$

$$= \sigma_i^2 \sum_{k=0}^{\infty} E[\underline{A}(k)\underline{T}_3(k+1)\underline{T}_3^{\mathrm{T}}(k+1)\underline{A}^{\mathrm{T}}(k)]$$

$$= \sigma_i^2 \sum_{k=0}^{\infty} E[\underline{A}(k)\underline{T}_3\underline{T}_3^{\mathrm{T}}\underline{A}^{\mathrm{T}}(k)]$$

$$+ P\sigma_i^2\sigma_3^2 \sum_{k=0}^{\infty} E[\underline{A}(k)\underline{A}^{\mathrm{T}}(k)]\,.$$

$\underline{T}_3\underline{T}_3^{\mathrm{T}} = \underline{T}_4 - \underline{T}_4^2$ and using Lemma D2, we obtain (20)

$$\underline{C}_f = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} E[\underline{A}(k)\vec{n}_f(k)\vec{n}_f^{\mathrm{T}}(l)\underline{A}^{\mathrm{T}}(l)]\,.$$

Using Lemma D2 yields (19).

References

[1] R.J. Marks II, A class of continuous level associative memory neural networks, *Applied Optics* (10) (1987) 2005–2010.

[2] K.F. Cheung, S. Oh, R.J. Marks II and L.E. Atlas, Neural net associative memories based on convex set projection, in: *Proc. IEEE 1st Internat. Conf. on Neural Networks*, San Diego (1987).

[3] R.J. Marks II, L.E. Atlas, S. Oh and K.F. Cheung, Optical-processor architectures for alternating-projection neural networks, *Optics Letters* 13 (6) (1988) 533–535.

[4] R.J. Marks II, S. Oh and L.E. Atlas, Alternating projection neural networks, *IEEE Trans. Circuits Syst.* 36 (6) (1989) 846–855.

[5] R.J. Marks II, S. Oh, L.E. Atlas and J.A. Ritcey, Homogeneous and layered alternating projection neural networks, in: *SPIE Real-Time Signal Processing for Industrial Applications* (1988) 217–232.

[6] D.C. Youla and H. Webb, Image restoration by the method of convex projections: part I – theory, *IEEE Trans. Med. Imaging* MI-1 (1982) 81–94.

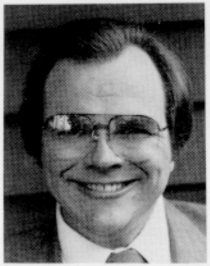
[7] H. Stark, *Image Recovery* Academic Press (New York 1987).

[8] S. Oh and R.J. Marks II, Noise sensitivity of projection neural networks, in *Proc.* 1989 *IEEE Internat. Symp. on Circuits and Systems* (1989) 2201–2204.

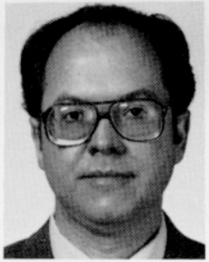
[9] S.G. Batsell, T. Jong, J.F. Walkup and T.F. Krile, Noise limitations in optical linear algebra processor, *Applied Optics* 29 (14) (1990) 2084–2090.

[10] J.M. Wozencraft and I.M. Jacobs, *Principles of Communication Engineering*. (Wiley, New York, 1965).

Seho Oh received the B.S. degree in electronics engineering from Seoul National University and the M.S. Degree in electrical engineering from Korea Advanced Institute of Science and Technology, Seoul. From 1981 through 1986, he was with Goldstar Research Laboratory in Seoul. He received his Ph.D. in Electrical Engineering from the University of Washington, Seattle, in 1989. His research interests are in the areas of signal analysis and processing, artificial neural networks and pattern recognition. Dr. Oh is the co-author of over thirty archival and proceedings papers and has been issued two United States patents.



Robert J. Marks II is a Professor in the Department of Electrical Engineering at the University of Washington, Seattle. Prof. Marks was awarded the Outstanding Branch Councillor award in 1982 by IEEE and, in 1984, was presented wth an IEEE Centennial Medal. He was Chair of IEEE Neural Networks Committee and was the co-founder and first Chair of the IEEE Circuits & Systems Society Technical Committee on Neural Systems & Applications. Prof. Marks was also elected the first President of the IEEE Council on Neural Networks. He is a Fellow of the Optical Society of America and a Senior Member of IEEE. Dr. Marks was also the co-founder and first President of the Puget Sound Section of the Optical Society of America and was recently elected that organization's first honorary member. He is co-founder and current President of Multidimensional Systems Corporation in Bellevue, Washington. Prof. Marks is the topical editor for Optical Signal Processing and Image Science for the *Journal of the Optical Society on America-A*. He is also a member of the Editorial Board for *Neurocomputing – An International Journal*. Prof. Marks is the author of *Introduction to Shannon Sampling and Interpolation Theory* (Springer-Verlag, 1991). He holds two patents in the areas of neural networks. Dr. Marks is a co-founder of the Christian Faculty Fellowship at the University of Washington. He is a member of Eta Kappa Nu and Sigma Xi.



Dennis P. Sarr is a lead engineer in Quality Assurance Research and Development, part of Operations Technology within the Boeing Commercial Airplane Group. He received his MSEE from the University of Washington in 1989, with his thesis research on neural networks and machine vision. His BSEE was at Michigan State University in 1973. He also is a registered Professional Engineer in the State of Washington. Since 1977, he has worked at Boeing with non-destructive testing and dimensional measurement technologies. Current projects involve design of test instrumentation, using microcomputers, machine vision, neural networks, and Statistical Process Control implementations. His contributions to the field of quality inspection has resulted in 10 patents issued as sole or co-author, with several other patents pending.