SKEW EFFECTS DUE TO OPTICAL PATH LENGTH VARIATION IN ITERATIVE NEURAL PROCESSORS

R. J. Marks II, S. Oh, D. C. Park and L. E. Atlas

Interactive System Design Laboratory
University of Washington, FT-10
Seattle, WA 98195

Abstract:

Little attention is normally given to the consequences of different optical path lengths required within an iterative processor on the processor performance. The resulting clock skew can have significant degrading effects on the predicted accuracy, stability and speed of the processor. A similar problem occurs in iterative asynchronous artificial neural networks when, for example, the time delay between two neurons is proportional to their physical separation. We have shown that, in the absence of temporal dispersion, certain iterative algorithms have stable steady state solutions that are independent of clock skew [1]+. In this paper, we analyze stability criterion for skewed systems and analyze the skew obtained in some commonly used optical processors.

Introduction:

Shamir [2] has noted that in many optical processors, the propagation time from input to output can vary significantly due to the variation of optical path lengths [3]. When the processor is used iteratively, disregarding this *clock skew* can lead to either unstable or drastically different implementation results. A similar problem occurs in iterative asynchronous artificial neural networks where the communication time delay between two neurons is proportional to their physical separation.

Our analysis will be restricted to temporally nondispersive systems. For such systems, a temporal impulse stimulus at any input coordinate can appear later only as a single temporal impulse at any specified output coordinate. Thus, for each input-output coordinate pair, there exists a single temporal delay. If this delay varies from coordinate pair to coordinate pair, the system is skewed.

We have previously shown, that, in certain feedback algorithms, temporally nondispersive clock skew does not affect the stability or the steady state solution of the processor. When an iterative algorithm uses a (possibly nonlinear) contractive operation in the feedback path, the resulting steady state solution is shown to be unaffected by clock skew. Clock skew is shown, however, to have an effect on systems such as Hopfield artificial neural networks [4-5] when hard nonlinearities are used in the feedback path [1].

In this paper, we analyze sufficient conditions for stability of skewed iteration and show that a contractive iteration is stable when the skew is additively separable. The Stanford matrix-vector multiplier is shown to have such skew.

Preliminaries:

In this section, we develop a general description for temporally nondispersive clock skew in a feedback processor. Let a field of N states, $\{S_n \mid 1 \le n \le N\}$, be altered by feedback in a temporally nondispersive skewed processor. Let ϑ_n denote the instantaneous operator that maps the previous states into the current nth state at time t. We can then write:

$$S_n(t) = \vartheta_n[\ \{S_m(t-\tau_{nm}) \mid 1 \le m \le N\}\ ]\ ;\ 1 \le n \le N \qquad (1)$$

where τ_{nm} is the clock skew corresponding to the time required for the state S_m to make a contribution to the state S_n. If we let $t \to \infty$ and assume a stable steady state, then (1) becomes:

$$S_n(\infty) = \vartheta_n[\ \{S_m(\infty) \mid 1 \le m \le N\}\ ]\ ;\ 1 \le n \le N \qquad (2)$$

Although not explicitly noted, this steady state may depend on the clock skew. For example, consider the linear iteration *

$$S_n(t) = \sum_m a_{nm} S_m(t-\tau_{nm}) + f_n(t);\ \ 1 \le n \le N \qquad (3)$$

where f_n is the n-th element of f and $f_n(t) \to f_n$ allows explicitly for input rise time. Letting $t \to \infty$ and assuming a stable result gives

$$S_n(\infty) = \sum_m a_{nm} S_m(\infty) + f_n;\ \ 1 \le n \le N$$

or, equivalently, in matrix- vector form

$$S(\infty) = \underline{A}\, S(\infty) + f$$

If $(\underline{I} - \underline{A})$ is not singular, then the solution to this equation is unique and given by

$$S(\infty) = (\underline{I} - \underline{A})^{-1} f \qquad (4)$$

Clock skew therefore does not affect the solution. The *alternating projection neural network* when interpreted either homogeneously [6-7] or, in layered form [8-9] from the hidden to output layer, is a special case of this example.

STABILITY: The above analysis is conditioned on the stability of the skewed iterations. By letting $n \to \infty$, for example, we might predict that the iteration $x(n) = 2\,x(n-1) + 1$ would converge to $x(\infty) = -1$. The difference equation, however, is clearly unstable and $x(\infty) = \pm\infty$ if $x(0) \ne -1$. From the viewpoint of z-transform analysis, the pole of this difference equation lies outside of the unit circle.

A sufficient condition for stability of linear skewed iteration is given by the following Lemma:

(Lemma 1)
Let $\underline{A} = (a_{ij})$ *denote a square matrix of complex numbers. Define* $\underline{A}(s) = (a_{ij}\exp[-s\tau_{ij}])$ *where* $s = \sigma + j\omega$. *If* $||\underline{A}(s)|| < 1$ *for* $Re(s) \ge 0$, *then (3) converges to (4).*

A proof is given in Appendix A. Note that, as a special case, we conclude that a iteration without skew converges if $||\underline{A}|| < 1$, since

$$||\underline{A}(s)|| = |\exp(-s\tau)|\ ||\underline{A}|| \le ||\underline{A}|| < 1$$

where $\tau_{nm} = \tau$ for all (n,m).

Two important results built on this lemma follow:

(Lemma 2)
Let $\underline{B} = (|a_{ij}|)$. *If* $||\underline{B}|| < 1$, *then (3) converges stably to (4) for any* $\tau_{ij} \ge 0$.

Therefore, varying the phase terms in the matrix does not affect the convergence stability if the zero phase iteration is stable.

(Lemma 3)
If $\tau_{ij} = u_i + v_j$, *and* $||\underline{A}|| < 1$, *then (3) converges stably to (4) for any* $u_i \ge 0$ *and* $v_j \ge 0$.

+ The fusion of [1] and this paper are tentatively scheduled to be published as S. Oh, D. C. Park, R. J. Marks II and L. E. Atlas, "Nondispersive propagation skew in iterative neural network and optical feedback processors." ***Optical Engineering*** (May, 1989)

* All summations (Σ) in this paper are from 1 to N.

ISCAS '89

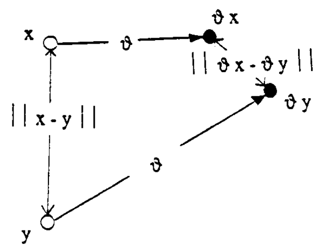

Figure 1: A geometrical illustration of a contractive operator. After the contractive operation, the signals are closer together (solid dots) than originally (hollow dots).

Proofs of lemmas 2 and 3 are in Appendices B and C, respectively. The last lemma will be applied to optical feedback systems later in the paper.

More generally, we can write (2) in vector form as

$$S(\infty) = \underline{\vartheta}\, S(\infty) \tag{5}$$

If the vector operator $\underline{\vartheta}$ is a (possibly nonlinear) *contractive* operator, then

$$\| \underline{\vartheta}\, x - \underline{\vartheta}\, y \| \leq r \, \| x - y \| \tag{6}$$

where the norm for a vector a, is defined by $\| a \|^2 = a^T a$ and $0 \leq r < 1$. If $0 \leq r \leq 1$, then $\underline{\vartheta}$ is said to be *non-expansive*. The reason for the terminology is evident from the geometry in Figure 1. Operating on two signals, x and y, by the operator $\underline{\vartheta}$ results in two signals closer together (contractive) or, at least not as far apart (non-expansive).

If $\underline{\vartheta}$ is contractive, then (5) has a unique solution [1,7-10,11] and there is no contribution of clock skew to the steady state result. When $\underline{\vartheta}$ is non-expansive, (5) can have a number of solutions.

For example [1], let

$$S_n(t) = \eta_n \left[\sum_m a_{nm} S_m(t-\tau_{nm}) + g_n(t) \right] + h_n(t); \; 1 \leq n \leq N \tag{7}$$

where $f_n(t) \rightarrow f_n$ and $g_n(t) \rightarrow g_n$ are forcing functions. Assuming stability, the steady state solution in vector form is

$$S(\infty) = \underline{\eta}\, [\underline{A}\, S(\infty) + g] + h$$

where $\underline{\eta}$ is a pointwise nonlinear vector operator, i.e, if $w = \underline{\eta} z$, then the n^{th} element of w is equal to $\eta_n(z_n)$ where η_n is a given function. In the parlance of neural networks, η_n could be referred to as a *sigmoid* operator [12-13]. Using (6), the corresponding operator is contractive if

$$\| \underline{\eta}(\underline{A}\, x + f) - \underline{\eta}(\underline{A}\, y + f) \| \leq r \, \| x - y \|$$

We have shown [1] that the operator is contractive if the spectral radius of $\underline{A}$ does not exceed one and $\underline{\eta}$ contains *soft nonlinearities*. That is

$$| \, d\eta_n(z) / dz \, | \leq 1 \, ; \; 1 \leq n \leq N \tag{8}$$

for all z.

STABILITY: The following lemma establishes a sufficient condition for stability of the skewed operation in (7).

(Lemma 4)

For a given matrix $\underline{A}$ and time-delays $\{\tau_{nm}\}$, if (3) converges for every $f_n(t)$ and $\underline{\eta}$ is non-expansive, then (7) is stable.

A proof is given in the Appendix D.

Hard Nonlinearities:

Clock skew can be a factor when implementing an iterative algorithm with hard nonlinearities. Consider the following example of Hopfield's content addressable memory neural network [4,12,14-16].

(Example 1)

From the three library vectors

v_1 = [0010 0101 0110]^T,
v_2 = [1011 0001 0001]^T,
v_3 = [1101 0110 1000]^T,

we form in accordance to Hopfield's recipe the interconnection matrix $\underline{A}$

$$\underline{A} = (\underline{B}\underline{B}^T - N\underline{I})$$

where $N=3$, $\underline{V} = [\, v_1 : v_2 : v_3 \,]$, $\underline{B} = 2\underline{V} - \underline{1}$ and $\underline{1}$ is a matrix of ones. We form the iteration $S(n+1) = \underline{\eta}\, [\, \underline{A}\, S(n) \,]$, where, for $1 \leq n \leq N$, $\eta_n(\bullet)$ is the unit step function ($\eta_n(x) = 1$ if $x \geq 0$ and is 0 otherwise). If we initialize with $S(0)$ = [1011 0001 1010]^T and iterate synchronously the solution of the operation converges to v_2. However, if this operation has a skew of 3 clocks delays for $\tau_{4,3}$ and $\tau_{4,11}$ and 2 clocks delays for those remaining, then the iteration oscillates. The iteration converges to [0010 1001 0111]^T in the case that the skew is 3 clocks delays for $\tau_{4,3}$, $\tau_{4,11}$, $\tau_{6,9}$ and $\tau_{6,12}$ and 2 clocks delays for those remaining. These three examples are respectively illustrated in Fig.2a,b and c.

The energy of the neural network at the nth iteration is defined as

$$E(n) = - S(n)^T \underline{A}\, S(n)/2$$

A plot of the energy for these three examples is in Fig. (2d).

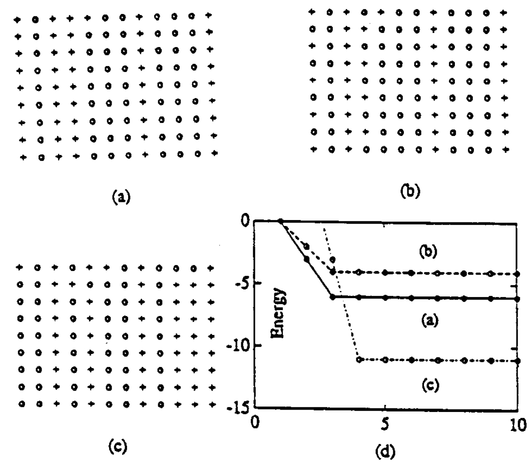

Figure 2: Examples of stability and convergence in Hopfield' s model. Iteration time in (a), (b) and (c) are in rows top to bottom. (+ denotes +1 in (a),(b) and (c)). The first row is the result of the first iteration. We see that S(0) (a) converges to v_2 without skew, (b) oscillates with skew, (c) converges to a vector which is not in library with a different skew. (d) Energy transitions with iteration for each case.

Skew in Optical Processors

The major source of skew in optical processors is the time delay resulting from the differing optical lengths (OL) [17] between input and output. For nondispersive clock skew, the time delay from the input point (ξ,η) to the output point (x,y) can be written as

$$\tau(x,y;\xi,\eta) = OL/c$$

where c is the speed of light in free space. In this section, we show some examples of optical processor operations which are not affected by skew.

(Example 2)

A commonly used processor for performing matrix-vector multiplication is shown at the top of Fig.3. The top view of this processor, shown in Fig.(3c), resembles a point source collimator. Since we are interested only with a single point at the output, there is no clock skew due to OL differences from this perspective. The side view of the processor shown in Fig.(3b), is

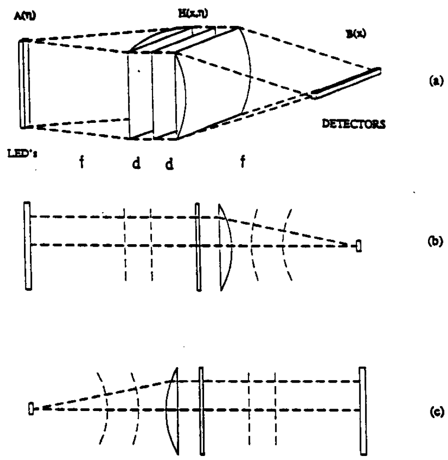

Figure 3: (a) A diagram of a matrix-vector multiplier (b) side view (c) top view

equivalent to that in Fig.(3c) except that the input and output are reversed. Since there is no skew from this view, the composite processor has no clock skew due to OL differences. The total OL from $(0,\eta)$ at the input plane to $(x,0)$ at the output plane is given by

$$OL_t = OL_1 + 2d + OL_2 = 2f + 2\Delta_0 + 2d$$

where OL_1 is the OL from $(0,\eta)$ to (x,η) and OL_2 is from (x,η) to $(x,0)$ and $OL_1 = OL_2 = f + \Delta_0$ and Δ_0 denotes the OL through the center of lens. This equation states that OL_t is constant for all (x,η) pairs.

(Example 3)

Fig.(4a) illustrates the Stanford matrix-vector multiplier [18] that is more light-efficient than the one in Fig.(3a). Since the performance is similar to that in Fig.(3b), there is no skew apparent in the side view in Fig.(4b). From the perspective of the top view in Fig.(4c), however, the apparent point source input is incident on the detector as a cylindrical wave. Under a Fresnel approximation, the skew is therefore quadratic. The total OL for this processor is given by

$$OL_t = 5f + \Delta + \eta^2/f = K + x^2/f$$

where Δ denotes sum of the OL's through the center of lenses and K is a constant. The time delay for this processor is thus

$$\tau(x,\eta) = OL_t/c = K/c + x^2/(f\,c)$$

Therefore, this processor is temporally skewed, but the skew is separable. By Lemmas 3 and 4, if (8) is true, any iterative processors which employ this Stanford matrix-vector multiplier using a pointwise soft nonlinearity in the feedback path that satisfies (8) will converge independent of this skew. An example of such a processor is the *alternating projection neural network* (APNN) which uses linear feedback [19,20].

Final Remarks

The primary source of clock skew in optical system is differing OL's. We have investigated the effects of clock skew on the performance of iterative processors and have shown that clock skew does not affect the convergence and stability of the solution when the feedback is contractive. We also have shown some examples of optical systems which have no skew or are not affected by skew when used iteratively.

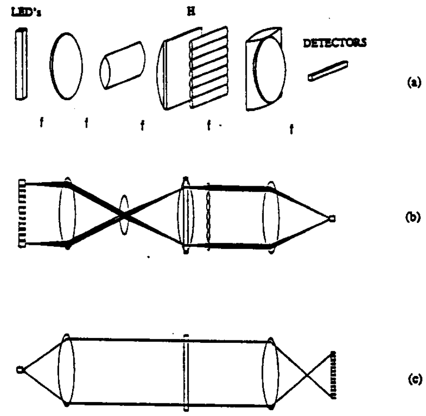

Figure 4: (a) Illustration of the Stanford matrix-vector multiplier (b) side view (c) top view

Acknowledgements:

This work was supported by the SDIO/IST administered by ONR through the Optical Systems Lab at Texas Tech University in Lubbock and by the Washington Technology Center at the University of Washington. In addition, L.E. Atlas was supported by an NSF Presidential Young Investigator's Award.

References:

[1] S. Oh, L.E. Atlas, R.J. Marks II and D.C. Park "Effects of clock skew in iterative neural network and optical feedback processors", **Proceedings of the IEEE International Conference on Neural Networks**, San Diego, July 24-27, 1988, vol.II, pp.429-436.

[2] Joseph Shamir, "Fundamental speed limitations on parallel processing," ***Applied Optics*** 26,pp. 1567 (1987).

[3] B. Clymer and J. W. Goodman," Timing uncertainty for receivers in optical clock distribution for VLSI," ***Optical Engineering***, vol. 27, pp.944-954 (1988)

[4] J.J. Hopfield "Neural networks and physical systems with emergent collective computational abilities", ***Proc. of the National Academy of Sciences***, USA, 81, 1982, pp.3088-3092.

[5] K.F. Cheung, L.E. Atlas and R.J. Marks II "Synchronous versus asynchronous behavior of Hopfield's content addressable memory", ***Applied Optics***, vol. 26, pp.4808-4813 (1987).

[6] R.J. Marks II "A class of continuous level associative memory neural nets", ***Applied Optics***, vol. 26, pp.2005-2009 (1987).

[7] K.F. Cheung, S. Oh, R.J. Marks II and L.E. Atlas "Neural net associative memories based on convex set projections", ***Proceedings of the IEEE First International Conference on Neural Networks***, San Diego, June 1987.

[8] R.J. Marks II, S. Oh, L.E. Atlas, and J.A. Ritcey, "Alternating projection neural networks", ***IEEE Transaction on Circuits and Systems*** (in press).

[9] R.J. Marks II, L.E. Atlas, S. Oh and J.A. Ritcey "The performance of convex set projection based neural networks", ***Proceedings of the IEEE Conference on Neural Information Processing Systems - Natural and Synthetic***, Boulder Colorado, November 1987.

[10] R. J. Marks, II, and D. K. Smith, "Gerchberg - type linear deconvolution and extrapolation algorithms" in ***Transformations in Optical Signal Processing***, edited by W. T. Rhodes, J. R. Fienup and B. E. A. Saleh, SPIE vol. 373, pp. 161-178 (1984).

[11] Arch W. Naylor and George R. Sell, ***Linear Operator Theory in Engineering and Science***, Springer-Verlag, New York (1982).

[12] R. P. Lippman, "An introduction to computing with neural nets," ***IEEE ASSP Mag.***, pp. 7 (April 1987).

[13] J. G. McDonnel, R. J. Marks, II, and L. E. Atlas "An introduction to neural networks for solving combinatorial search problems," ***Northcon/88 Conference Record***, vol.II, pp.868-876, Seattle WA. October 1988 (Western Periodicals Co., North Hollywood, CA)

[14] D. Psaltis and N. Farhat, "Optical information processing based on an associative memory model of neural nets with thresholding and feedback," ***Optics Letters***, vol. 10, pp.98 (1985)

[15] N. Farhat, D. Psaltis, A. Prata and E. Paek, "Optical implementation of the Hopfield model," ***Applied Optics*** vol. 24, pp.1469 (1985)

[16] R. J. Marks and L. E. Atlas, "Geometrical interpretation of Hopfield's content addressable memory neural network," ***Northcon/88 Conference Record***, vol.II, pp.964-977, Seattle WA. October 1988 (Western Periodicals Co., North Hollywood, CA)

[17] M. Born and E. Wolf, ***Principles of Optics***, Pergamon Press, New York (1980)

[18] J.W. Goodman, A.R. Dias, and L.M. Woody," Fully parallel, high speed incoherent optical method for performing discrete Fourier transforms," ***Optics Letters***, vol. 2, pp. 1-3(1978)

[19] R. J. Marks, II, L. E. Atlas, S. Oh and K. F. Cheung, "Architectures for a continuous level neural network based on alternating orthogonal projections," ***Proceedings of O-E/LASE '88 Conference on Neural Network Models for Optical Computing***, Los Angeles, January 1988.

[20] R.J. Marks II, L.E. Atlas, S. Oh and K.F. Cheung "Optical processor architectures for alternating projection neural networks", ***Optics Letters***, vol.13, pp.533-525 (1988).

Appendices

Appendix A. Proof of Lemma 1

We take the Laplace transform of (3)

$$S_n(s) = \sum_m a_{nm} S_m(s) \exp(-s\tau_{nm}) + F_n(s)$$

or, equivalently, in matrix form,

$$S(s) = \underline{A}(s)\, S(s) + F(s)$$

Since $||\underline{A}(s)|| < 1$, $\det[\underline{I} - \underline{A}(s)] \neq 0$ and $S(s)$ becomes

$$S(s) = [\underline{I} - \underline{A}(s)]^{-1} F(s)$$

By applying the final value theorem, we obtain

$$S(\infty) = \lim_{s\to 0} s\,S(s) = \lim_{s\to 0} [\underline{I} - \underline{A}(s)]^{-1}[s\,F(s)]$$
$$= [\underline{I} - \underline{A}(s)]^{-1} f$$

which is our desired result.

Appendix B. Proof of Lemma 2

Let y be an N dimensional vector and $\underline{A}(s)=(a_{ij}\exp[-s\tau_{ij}])$ an NxN matrix. Then

$$||\underline{A}(s)\, y||^2 = y^* \underline{A}(s)^* \underline{A}(s)\, y$$
$$= \sum_i\sum_j\sum_k y_i^* a_{ki}^* \exp(-s^*\tau_{ki})\, a_{kj}\exp(-s\tau_{kj})\, y_j$$
$$\leq \sum_i\sum_j\sum_k |y_i|\,|a_{ki}|\,|a_{kj}|\,|y_j|$$

where $*$ denotes the complex conjugate for scalars and, for matrices, denotes the complex conjugate transpose. Let $z_i = |y_i|$, $b_{ki} = |a_{ki}|$ and $z = (z_i)$. Then

$$||\underline{A}(s)\, y||^2 \leq \sum_i\sum_j\sum_k z_i b_{ki} b_{kj} z_j = ||\underline{B}\, z||^2$$

Since $||y|| = ||z||$, $||\underline{A}(s)|| \leq ||\underline{B}|| < 1$. Since convergence is assured for $\underline{B}$ by assumption, we conclude from that (3) converges to (4) and our proof is complete.

Appendix C. Proof of Lemma 3

From the assumption, we have

$$\underline{A}(s) = (a_{ij}\exp[-s(u_i+v_j)]) = (a_{ij}\exp[-su_i]\exp[-sv_j])$$

By letting

$$\underline{D}_u = diag\{\exp(-su_1), \exp(-su_2), \ldots, \exp(-su_N)\}$$

and $\underline{D}_v = diag\{\exp(-sv_1), \exp(-sv_2), \ldots, \exp(-sv_N)\}$,

$\underline{A}(s)$ becomes

$$\underline{A}(s) = \underline{D}_u \underline{A}\, \underline{D}_v$$

Since $||\underline{D}_u|| \leq 1$ and $||\underline{D}_v|| \leq 1$ for $Re(s) \geq 0$, we have

$$||\underline{A}(s)|| \leq ||\underline{D}_u||\;||\underline{A}||\;||\underline{D}_v|| \leq ||\underline{A}|| < 1$$

From this and Lemma 1, we conclude that (3) converges to (4).

Appendix D. Proof of Lemma 4

We rewrite (7) as

$$S_n(t) = \eta_n[\,y_n(t) + g_n(t)] + h_n(t)$$

where $y_n(t) = \sum_m a_{nm} S_m(t-\tau_{nm})$.

Let $y(t)$ denotes the vector of $y_n(t)$, we take the norm of $S_n(t_1)-S_n(t_2)$.

$$||S(t_1)-S(t_2)||$$
$$\leq ||\eta[y(t_1)+g_n(t_1)] - \eta[y(t_2)+g(t_2)]|| + ||h(t_1)-h(t_2)||$$
$$\leq ||y(t_1) - y(t_2)|| + p(t_1,t_2)$$

or

$$||S(t_1,t_2)|| \leq ||y(t_1,t_2)|| + p(t_1,t_2) \qquad \text{(D.1)}$$

where

$$p(t_1,t_2) = ||g(t_1)-g(t_2)|| + ||h(t_1)-h(t_2)||,$$
$$S(t_1,t_2) = S(t_1) - S(t_2)$$

and $y(t_1,t_2)$ denotes the vector of

$$y_n(t_1,t_2) = \sum_m a_{nm}[S_m(t_1-\tau_{nm}) - S_m(t_2-\tau_{nm})].$$

Substitute Q for S in the linear skewed iteration in (3) and take the difference between (3) at t_1 and t_2 to obtain

$$Q(t_1,t_2) = x(t_1,t_2) + f(t_1,t_2).$$

where

$$f(t_1,t_2) = f(t_1) - f(t_2),$$
$$Q(t_1,t_2) = Q(t_1) - Q(t_2)$$

and $x(t_1,t_2)$ denotes the vector of

$$x_n(t_1,t_2) = \sum_m a_{nm}[Q_m(t_1-\tau_{nm}) - Q_m(t_2-\tau_{nm})].$$

We will construct $f(t_1,t_2)$ to be colinear with $x(t_1,t_2)$. Let

$$f(t_1,t_2) = \begin{cases} \text{any vector} & ;\ \text{for } x(t_1,t_2) = 0 \\ C(t_1,t_2)\, x(t_1,t_2) & ;\ \text{otherwise} \end{cases}$$

where the proportionality constant, $C(t_1,t_2)$, is chosen so that $||f(t_1,t_2)|| = p(t_1,t_2)$. Then, by construction,

$$||Q(t_1,t_2)|| = ||x(t_1,t_2)|| + p(t_1,t_2). \qquad \text{(D.2)}$$

Because $||Q(t_1,t_2)||$ converges to 0 by the assumption, (D.1) also converges:

$$||S(t_1,t_2)|| \to 0 \quad \text{for } t_1 \to \infty,\ t_2 \to \infty.$$

or equivalently, since our vectors are in a finite dimensional Euclidean space,

$$\lim_{t\to\infty} S(t) = S_\infty$$

and our proof is complete.