Convergence and bias in the LSG algorithm for adaptive lattice filters

ROHIT NEGI¹ and P G POONACHA²

Department of Electrical Communication Engineering, Indian Institute of Technology, Powai, Bombay 400 076, India

Present address: ¹Information Systems Laboratory, Durand 112, Stanford University, Stanford, CA 94305-9510, USA

²Silicon Automation Systems, # 3008, 12th B Main, 8th Cross, HAL II Stage, Bangalore 560 008, India

e-mail: negi@isl.stanford.edu; poonacha@sasi.com

Abstract. The lattice filter has several desirable characteristics that make it attractive in adaptive applications. The Lattice Stochastic Gradient (also called Gradient Adaptive Lattice algorithm) is popularly used to adapt the lattice filter. However, a theoretical study of the bias in the reflection coefficient and the convergence of the LSG algorithm have not been studied extensively yet. This paper presents some theoretical results on these issues. It is hoped that the results will also present some insights into the factors affecting the convergence of the filter.

Keywords. Adaptive filters; lattice filters; convergence analysis; adaptive algorithms.

1. Introduction

The lattice filter is useful in adaptive applications because it possesses several desirable characteristics such as orthogonalization of the input, faster convergence, simple stability checking criteria etc. The Lattice Stochastic Gradient (LSG), also called Gradient Adaptive Lattice (GAL) algorithm, is popularly used to adapt the (FIR) lattice filter as it is fast and has a reasonably low computation.

In adaptive filtering literature, there is a lack of rigorous theoretical analysis on the issues of bias in filter coefficient and the convergence of the algorithms. The analysis of the convergence of the LMS algorithm for the transversal filter is handled by making the ‘independence assumption’, which though handy, seems somewhat artificial. An analysis which does not use this assumption is detailed by Douglas & Meng (1992) for a restricted kind of input.

Such rigorous analysis is absent in the case of the lattice filter. Honig & Messerschmitt have proposed convergence models by Honig & Messerschmitt (1981) for analysis of speed of convergence that are partly analytic.

This paper attempts to prove absence of bias in reflection coefficient and the convergence of the LSG algorithm using more rigorous arguments. The notation used is that followed by Honig & Messerschmitt (1984).

The order update equations of the lattice filter are:

$$\begin{aligned} e_f(T/n) &= e_f(T/n - 1) - k_n^b \cdot e_b(T - 1/n - 1), \\ e_b(T/n) &= e_b(T - 1/n - 1) - k_n^f \cdot e_f(T/n - 1), \end{aligned} \quad (1)$$

where e_f and e_b represent the forward and backward errors respectively, T is the time index, k_n^f and k_n^b are forward and backward reflection coefficients of the n th stage. In the LSG algorithm, $k_n^f = k_n^b$.

The lattice stochastic gradient algorithms is:

$$\begin{aligned} k_n(T + 1) &= k_n(T) + \frac{1}{E(T/n - 1)} \\ &\quad [e_f(T/n - 1) \cdot e_b(T/n) + e_b(T - 1/n - 1) \cdot e_f(T/n)], \\ E(T/n - 1) &= \lambda \cdot E(T - 1/n - 1) + e_f^2(T/n - 1) + e_b^2(T - 1/n - 1), \end{aligned} \quad (2)$$

with $e_f(T/0) = e_b(T/0) = u(T)$ and $E(0/n - 1) = +ve \text{ constant } \forall n = 1, \dots, N$.

2. Theoretical results on the LSG algorithm

Both the results presented below require assumption of ergodic input $u(T)$ to the lattice filter Middleton (1960) for the definition of ergodic process used). The proof on convergence of the LSG algorithm requires, in addition, the assumption of ‘independent enough’ input (defined later).

2.1 Absence of bias in $k_n(T)$ in the LSG algorithm

Theorem 1. *The lattice filter coefficients $k_n(T)$ are asymptotically unbiased when the LSG adaptive algorithm is used, assuming an ergodic input. ie.*

$$\lim_{T \rightarrow \infty} E[k_{n+1}(T + 1)] = k_{n+1}^{opt}.$$

Proof. The LSG algorithm implements the following expression recursively

$$k_{n+1}(T + 1) = \frac{\sum_{i=0}^T 2 \cdot e_f(i/n) \cdot e_b(i - 1/n)}{\sum_{i=0}^T \{e_f^2(i/n) + e_b^2(i - 1/n)\}}. \quad (3)$$

Denote the numerator and denominator of the above fraction as NUM and DEN respectively, for convenience.

Here, we have considered the case $\lambda = 1$ only, which is usually taken for stationary input. Now define

$$N_0 = E[2 \cdot e_f(T/n) \cdot e_b(T - 1/n)],$$

$$D_0 = E[e_f^2(T/n) + e_b^2(T - 1/n)]. \quad (4)$$

We can write $\frac{1}{T} \text{NUM} = N_0 + n(T)$ and $\frac{1}{T} \text{DEN} = D_0 + d(T)$.

If $u(T)$ is assumed to be ergodic, and we assume that the coefficients k_j , $j = 1, \dots, n$ have converged (to constant values), then $e_f(T/n) \cdot e_b(T - 1/n)$ and $e_f^2(T/n) + e_b^2(T - 1/n)$ can be shown to be ergodic processes. Then $\lim_{T \rightarrow \infty} \frac{1}{T} \text{NUM} = N_0$ with probability 1, and $\lim_{T \rightarrow \infty} \frac{1}{T} \text{DEN} = D_0$ with probability 1, which means

$$\lim_{T \rightarrow \infty} n(T) = 0, \text{ with probability 1, and}$$

$$\lim_{T \rightarrow \infty} d(T) = 0, \text{ with probability 1,}$$

where ‘probability 1’ means that the probability of the event that $\lim_{T \rightarrow \infty} n(\Omega_j, T) \neq 0$ is zero ($n(\Omega_j, T)$ denotes the j th sample function of the process $n(T)$). Now

$$k_{n+}(T + 1) = \frac{\text{NUM}/T}{\text{DEN}/T} = \frac{N_0}{D_0} \cdot \left[\frac{1 + \frac{n(T)}{N_0}}{1 + \frac{d(T)}{D_0}} \right], \quad (5)$$

$$E[k_{n+1}(T + 1)] = k_{n+1}^{\text{opt}} \cdot E \left[\frac{1 + \frac{n(T)}{N_0}}{1 + \frac{d(T)}{D_0}} \right], \quad (6)$$

where k_{n+1}^{opt} is the optimum Wiener solution in the MMSE sense.

Now, we prove that for the expectation on the RHS, denoted by $E[(\cdot)]$ for convenience, $\lim_{T \rightarrow \infty} E[(\cdot)] = 1$ under the ergodicity assumptions made above.

$$\begin{aligned} E[(\cdot)] &= E \left[\frac{1 + \frac{n(T)}{N_0}}{1 + \frac{d(T)}{D_0}} \right] = \sum_{j \in \mathcal{S}} \left[\frac{1 + \frac{n(\Omega_j, T)}{N_0}}{1 + \frac{d(\Omega_j, T)}{D_0}} \right] \cdot P(\Omega_j), \quad (7) \\ |E[(\cdot)] - 1| &= \left| \sum_{j \in \mathcal{S}_c} \left[\frac{\frac{n(\Omega_j, T)}{N_0} - \frac{d(\Omega_j, T)}{D_0}}{1 + \frac{d(\Omega_j, T)}{D_0}} \right] \cdot P(\Omega_j) \right. \\ &\quad \left. + \sum_{j \in \bar{\mathcal{S}}_c} \left[\frac{\frac{n(\Omega_j, T)}{N_0} - \frac{d(\Omega_j, T)}{D_0}}{1 + \frac{d(\Omega_j, T)}{D_0}} \right] \cdot P(\Omega_j) \right| \\ &\leq \sum_{j \in \mathcal{S}_c} \left| \frac{\frac{n(\Omega_j, T)}{N_0} - \frac{d(\Omega_j, T)}{D_0}}{1 + \frac{d(\Omega_j, T)}{D_0}} \right| \cdot P(\Omega_j) \\ &\quad + \sum_{j \in \bar{\mathcal{S}}_c} \left| \frac{\frac{n(\Omega_j, T)}{N_0} - \frac{d(\Omega_j, T)}{D_0}}{1 + \frac{d(\Omega_j, T)}{D_0}} \right| \cdot P(\Omega_j) \end{aligned}$$

where $\mathcal{S}$ denotes the joint sample space of the random processes $n(T)$ and $d(T)$, $\mathcal{S}_c$ denotes the subset comprising all sample functions $n(\Omega_j, T)$ and $d(\Omega_j, T)$ which converge to zero as $T \rightarrow \infty$, and $\bar{\mathcal{S}}_c$ denotes the complement of $\mathcal{S}_c$.

Considering the second summation in above inequality, we see that since $[1 + d(\Omega_j, T)/D_0] = \text{DEN}/D_0T$, and DEN is always positive non-zero, hence we can bound the absolute value of the fraction by an upper bound M . Then,

$$\left| \sum_{j \in \bar{\mathbf{S}}_c} [(\cdot)] \cdot P(\Omega_j) \right| \leq M \cdot \sum_{j \in \bar{\mathbf{S}}_c} P(\Omega_j),$$

which is zero under the ergodicity assumptions and above.

Considering the first summation in same inequality, we see that since all $n(\Omega_j, T)$ and all $d(\Omega_j, T) \in \mathbf{S}_c$ converge to zero as $T \rightarrow \infty$, hence, given a $\delta > 0$, we can find a T_0 such that $|n(\Omega_j, T)|$ and $|d(\Omega_j, T)|$ are both $\leq \delta$ for $T \geq T_0$.

Therefore, for $T \geq T_0$, the absolute value of the fraction in the summation $\leq k \cdot \delta$, where k has a finite value. Hence,

$$\left| \sum_{j \in \mathbf{S}_c} [(\cdot)] \cdot P(\Omega_j) \right| \leq k \cdot \delta \cdot \underbrace{\sum_{j \in \mathbf{S}_c} P(\Omega_j)}_{1 \text{ by ergodicity}},$$

which implies that

$$|E[(\cdot)] - 1| \leq k \cdot \delta, \quad \text{for } T \geq T_0. \quad (8)$$

Hence, $\lim_{T \rightarrow \infty} E[(\cdot)] = 1$ which implies that

$$\lim_{T \rightarrow \infty} E[k_{n+1}(T + 1)] = k_{n+1}^{opt}. \quad (9)$$

By following a method similar to the one used above, if we consider

$$\begin{aligned} & E[\{k_{n+1}(T + 1) - k_{n+1}^{opt}\}^l] \\ &= (k_{n+1}^{opt})^l \cdot E\left[\left\{\left(\frac{n(T)}{N_0} - \frac{d(T)}{D_0}\right) \middle/ \left(1 + \frac{d(T)}{D_0}\right)\right\}^l\right] \quad l = 2, 3, \dots, \end{aligned}$$

then, it can be shown that the term $E[(\cdot)]$ on the RHS converges to zero as $T \rightarrow \infty$.

Thus, all central moments of $k_n(T) \rightarrow 0$ as $T \rightarrow \infty$, which means that $k_n(T)$ converges to a constant value.

Thus, under the assumption of ergodic $u(T)$, it is proved that $k_1(T)$ converges to a fixed value which is also the optimum value i.e. k_1 obtained has no bias.

Now, ergodic $u(T)$ and convergence of $k_1(T)$ to optimum value is sufficient to prove the convergence of $k_2(T)$ to its optimum value, and so on.

Thus, all $k_n(T)$ converge to their respective optimum values. $\square$

2.2 Proof of convergence of $k_n(T)$ to a stable stationary point when the LSG algorithm is used

Theorem 2. Assuming an ergodic and ‘independent enough’ input (defined below), the lattice filter coefficients $k_n(T)$ converge to stable stationary points, when the LSG adaptive algorithm is used.

Proof. This proof uses the method of conversion of a recursive difference equation to an ordinary differential equation (see Goodwin & Payne 1977, pp 192–196).

Consider the LSG algorithm (2). It can be rewritten as

$$k_n(T+1) = k_n(T) + \frac{1}{E(T/n-1)} \cdot [2 \cdot e_f(T/n-1) \cdot e_b(T-1/n-1) - k_n(T) \cdot \{e_f^2(T/n-1) + e_b^2(T-1/n-1)\}]. \quad (10)$$

$$E(T/n-1) = \sum_{j=0}^T \lambda^{T-j} \cdot \{e_f^2(j/n-1) + e_b^2(j-1/n-1)\}. \quad (11)$$

Assume ergodic input $u(T)$. Also, assume that the stages $1, \dots, n-1$ have ‘adapted’; by which we mean that $\{k_i(T)\}, i = 1, \dots, n-1$ have each converged to a stable stationary point.

We prove that in such a case $k_n(T)$ will also converge to a stationary point

For convenience, denote expectation operator $E[(\cdot)]$ as $(\cdot)$.

From (10), we can write

$$k_n(T+M) = k_n(T) + \sum_{j=0}^{M-1} \left[\frac{1}{E(T+j/n-1)} \right] \cdot [2 \cdot e_f(T+j/n-1) \cdot e_b(T+j-1/n-1) - k_n(T+j) \cdot \{e_f^2(T+j/n-1) + e_b^2(T+j-1/n-1)\}]. \quad (12)$$

Let $T = 0$ denote the time when all the first $(n-1)$ stages have ‘adapted’. Then from (11), we get

$$\bar{E}(T/n-1) = \sum_{j=0}^T \lambda^{T-j} \cdot \{2 \cdot P_{n-1}\},$$

where

$$P_{n-1} = \bar{e}_f^2(j/n-1) = \bar{e}_b^2(j-1/n-1), \text{ and } P_{n-1}$$

is independent of time by the assumption that $k_i(j), i = 0, \dots, n-1$ have converged to stable stationary solution.

$$\begin{aligned} \Rightarrow \bar{E}(T/n-1) &= \left[\frac{1 - \lambda^{T+1}}{1 - \lambda} \right] 2 \cdot P_{n-1} \\ \Rightarrow \bar{E}(T/n-1) &\approx \underbrace{\left[\frac{\ln(1/\lambda)}{1 - \lambda} \cdot 2 \cdot P_{n-1} \right]}_{\text{constant}=\beta} T. \end{aligned} \quad (13)$$

The approximation holds when $T \ll |1/\ln(\lambda)|$.

Now, since $\bar{E}(T/n-1) = \beta \cdot T$, we use the heuristic that for large T , $E(T/n-1) = \beta \cdot T$ also. Then, since increment in $k_n(T)$ is inversely proportional to $(T+j)$, for $T \gg M$, we

can replace $k_n(T + j)$ in (12) by $k_n(T)$, which gives

$$\begin{aligned}
 k_n(T + M) &= k_n(T) + \sum_{j=0}^{M-1} \left[\frac{1}{E(T + j/n - 1)} \right] \\
 &\quad \cdot \overbrace{[2 \cdot e_f(T + j/n - 1) \cdot e_b(T + j - 1/n - 1) - k_n(T) \cdot \{e_f^2(T + j/n - 1) + e_b^2(T + j - 1/n - 1)\}]}^{\text{denote by PRO}(T+j)} \\
 &\quad \underbrace{\hspace{10em}}_{\text{denote by SUM}(T+j)} \\
 \bar{k}_n(T + M) &= \bar{k}_n(T) + \sum_{j=0}^{M-1} \left[\frac{1}{E(T + j/n - 1)} \right] \\
 &\quad \cdot [\text{PRO}(T + j) - k_n(T) \cdot \text{SUM}(T + j)]. \tag{14}
 \end{aligned}$$

Now, let us assume that the input $u(T)$ is, in addition to being ergodic, also an ‘independent enough’ process, by which we mean that $\exists$ a finite ‘ l ’ such that $u(T)$ and $u(T - j)$ are independent $\forall j \geq l$. We shall call this the assumption of ‘independent-enoughness’.

Next, we note the input samples that appear in the various terms of the summation in (14)

$E(T + j/n - 1)$ has samples $u(T + j), u(T + j - 1), \dots, u(0)$.

$\text{PRO}(T + j)$ has samples $u(T + j), \dots, u(T + j - n)$ and $k_i, i \leq n - 1$.

$\text{SUM}(T + j)$ has samples $u(T + j), \dots, u(T + j - n)$ and $k_i, i \leq n - 1$.

$k_n(T)$ has samples $u(0), u(1), \dots, u(T - 1)$. However, by the heuristic that $E(T/n - 1) \propto T$, it seems reasonable to believe that the dependence of $k_n(T)$ on $u(\tau)$ will ‘decrease’ as τ increases. At this stage, we cannot rigorously justify this, and hence, this decreasing dependence can be considered an assumption. We call this the assumption of “decreasing dependence”.

Since we are considering large T , we can write $E(T + j/n - 1) \approx \sum_{r=0}^{T+j-n-l} [e_f^2(\tau/n - 1) + e_b^2(\tau - 1/n - 1)]$, so that $E(T + j/n - 1)$ now has samples $u(T + j - n - l), u(T + j - n - l - 1), \dots, u(0)$.

Comparing the samples $u(\tau)$ in $E(T + j/n - 1)$, $\text{PRO}(T + j)$ and $\text{SUM}(T + j)$, and noting that $k_i, i \leq n - 1$ have already ‘adapted’ before $T = 0$, we conclude that by the assumption of ‘independent-enoughness’, $E(T + j/n - 1)$ is independent of $\text{PRO}(T + j)$ and of $\text{SUM}(T + j)$.

Again, comparing the samples $u(\tau)$ in $k_n(T)$ and $\text{PRO}(T + j)$, $\text{SUM}(T + j)$, we can say that for $j \geq (l + n - 1)$, $k_n(T)$ is independent of both $\text{PRO}(T + j)$, and of $\text{SUM}(T + j)$. Choose M large enough (but $\ll T$), such that $\sum_{j=0}^{M-1} (\cdot) \approx \sum_{j=l+n-1}^{M-1} (\cdot)$. Therefore, for all terms in the summation, $k_n(T)$ is independent of PRO and SUM .

Finally, if $E(T + j/n - 1) \approx \sum_{\tau=p}^{T+j-n} [e_f^2(\tau/n - 1) + e_b^2(\tau - 1/n - 1)]$ where p is large enough so that the assumption of “decreasing dependence” of $k_n(T)$ is valid for terms of $E(T + j/n - 1)$, then $k_n(T)$ will also be independent of $E(T + j/n - 1)$. Under these

conditions, we can write (14) as

$$\bar{k}_n(T + M) = \bar{k}_n(T) + \sum_{j=0}^{M-1} \frac{1}{T + j} \cdot \left[\frac{1}{E(T + j/n - 1)/(T + j)} \right] \cdot [\overline{\text{PRO}}(T + j) - \bar{k}_n(T) \cdot \overline{\text{SUM}}(T + j)].$$

Since $u(T)$ is assumed to be ergodic, hence $e_f^2(T/n - 1) + e_b^2(T - 1/n - 1)$ is also ergodic. Therefore, according to lemma 1 (see appendix A), we see that for $\lambda = 1$, $1/[E(\cdot)/(T + j)] \rightarrow 1/[\overline{E}(\cdot)/(T + j)]$ as $T \rightarrow \infty$. We assume that for large T (but $\ll |1/\ln(\lambda)|$), we can replace $1/[E(\cdot)/(T + j)]$ by $1/[\overline{E}(\cdot)/(T + j)]$ even for $\lambda \neq 1$.

Finally, noting that $\overline{\text{PRO}}(T + j) = r_{fb}$, $\overline{\text{SUM}}(T + j) = 2 \cdot P_{n-1}$ and $\overline{E}(\cdot)/T = \beta$, where r_{fb} , P_{n-1} and β are independent of time for the conditions of ergodicity of $u(\tau)$, and ‘adaptation’ of the lower $n - 1$ stages, we can write the above equation as

$$\bar{k}_n(T + M) = \bar{k}_n(T) + \frac{1}{\beta} \cdot [r_{fb} - \bar{k}_n(T) \cdot 2 \cdot P_{n-1}] \cdot \underbrace{\sum_{j=0}^{M-1} \left[\frac{1}{T + j} \right]}_{\text{small}}.$$

Converting this difference equation to an ODE

$$\frac{d}{dt}(\bar{k}_n(t)) = \frac{1}{\beta} \cdot [r_{fb} - \bar{k}_n(t) \cdot 2 \cdot P_{n-1}]. \quad (15)$$

The ODE gives as the stationary point

$$\begin{aligned} \bar{k}_n(t) &= r_{fb}/[2 \cdot P_{n-1}] \\ &= \frac{E[2 \cdot e_f(T/n - 1) \cdot e_b(T - 1/n - 1)]}{E[e_f^2(T/n - 1) + e_b^2(T - 1/n - 1)]}, \end{aligned}$$

which is also the optimum Wiener solution in the MMSE sense.

To check for stability of the stationary point, we write

$$\frac{d}{dt}(\Delta \bar{k}_n(t)) = - \left[\frac{1}{\beta} \cdot 2 \cdot P_{n-1} \right] \Delta \bar{k}_n(t).$$

Hence stability is guaranteed iff $2 \cdot P_{n-1}/\beta > 0$, i.e. $\left[\frac{1-\lambda}{\ln(1/\lambda)} \right] > 0$, which is true.

Hence, we conclude that under the assumptions made, when the lower $n - 1$ stages of the lattice filter have converged to stable stationary points, then the n th stage will also do so. The first stage does not require convergence of any other stage.

Thus, the first stage ‘adapts’ the earliest. Next, the second stage ‘adapts’ (since the first stage has already adapted), and so on for higher stages.

This completes the proof. $\square$

3. Results and discussion

While the two results presented above can both be regarded as proving absence of bias in reflection coefficient as well as convergence of the LSG algorithm, yet they are not quite

identical. The first result holds only for the case $\lambda = 1$ (which of course should be used for truly stationary input).

The second result holds for other values of λ also. The price to be paid is less rigour in the proof as several heuristics have to be introduced. However, we feel that the second result provides insight into the working of the filter in the way that the first result does not. For, example, a look at the ODE obtained shows that the speed of convergence depends on $(1 - \lambda)/\ln(1/\lambda)$, which is independent of the MSE P_{n-1} and is very nearly independent of λ for values close to 1.

4. Conclusion

The paper has attempted to approach the issues of convergence of the LSG algorithm and the absence of bias in the reflection coefficient in a somewhat more rigorous manner than is available in literature at present. The conditions imposed on the input are admittedly very strict, but they would be satisfied by Gaussian processes, which are used in simulation studies. However, it is clear that mere wide-sense stationarity is not enough for convergence. It is to be seen whether the conditions imposed on the input can be weakened.

The authors would like to thank one of our reviewers for bringing to our attention a recent paper by Fan & Liu (1993), which proves the convergence of the LSG algorithm under more general conditions.

Appendix A

In this appendix, we state and prove lemma 1.

Lemma 1. If $w(T)$ is a stationary, ergodic random process with $\overline{w(T)} = w_0$, and $E(T) = \frac{1}{T} \cdot \sum_{j=0}^T w(j)$, then

$$\lim_{T \rightarrow \infty} \left[\frac{1}{E(T)} \right] = \frac{1}{w_0}.$$

Proof. The proof uses arguments similar to those in the proof given in 2.1

Since $w(T)$ is stationary ergodic, hence we can write,

$$E(T) = w_0 + n(T),$$

where $n(T) \rightarrow 0$ with probability 1 as $T \rightarrow \infty$.

Hence, we can write

$$\left[\frac{1}{E(T)} \right] = \frac{1}{w_0} \cdot \left[\frac{1}{1 + n(T)/w_0} \right]. \quad (16)$$

Next, we show that the expectation on the RHS $\rightarrow 1$ as $T \rightarrow \infty$. Write

$$1 - \left[\frac{1}{1 + n(T)/w_0} \right] = \left[\frac{n(T)/w_0}{1 + n(T)/w_0} \right].$$

Then, by writing the expectation as two summations, in a manner similar to that used in § 2.1, we get

$$\left| 1 - \left[\frac{1}{1 + n(T)/w_0} \right] \right| \leq \sum_{k \in \mathbf{S}_c} \left| \frac{n(\Omega_k, T)/w_0}{1 + n(\Omega_k, T)/w_0} \right| \cdot P(\Omega_k) \\ + \sum_{k \in \bar{\mathbf{S}}_c} \left| \frac{n(\Omega_k, T)/w_0}{1 + n(\Omega_k, T)/w_0} \right| \cdot P(\Omega_k),$$

where $\mathbf{S}_c$ consists of all sample functions $n(\Omega_k T)$ of $n(T)$ which converge to zero as $T \rightarrow \infty$, and $\bar{\mathbf{S}}_c$ is the complement of $\mathbf{S}_c$.

Now, the second summation on the RHS is zero since the denominator of the fraction is always positive and $\sum_{k \in \bar{\mathbf{S}}_c} P(\Omega_k) = 0$.

As for the first summation on the RHS, given any δ , we can find a T_0 such that for $T \geq T_0$

$$\sum_{k \in \mathbf{S}_c} \left| \frac{n(\Omega_k, T)/w_0}{1 + n(\Omega_k, T)/w_0} \right| \cdot P(\Omega_k) \leq p \cdot \delta \underbrace{\sum_{k \in \mathbf{S}_c} P(\Omega_k)}_1,$$

where p is a finite number.

Thus, we have shown that as $T \rightarrow \infty$, the expectation on the RHS of (16) $\rightarrow 1$.

This therefore proves the lemma. $\square$

References

- Douglas S, Meng T H Y Exact expectation analysis of the LMS adaptive filter without the independence assumption. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp 61–64
- Fan H, Liu X 1993 GAL and LSL revisited: new convergence results. *IEEE Trans. Signal Process.* 41: 55–66
- Goodwin G C, Payne R L 1977 *Dynamic system identification experimentation, design, and data analysis* (New York: Academic Press)
- Honig M L, Messerschmitt D G 1981 Convergence properties of an adaptive digital lattice filter. *IEEE Trans. Acoust., Speech Signal Processing* 29: 642–653
- Honing K L, Messerschmitt D G 1984 *Adaptive fillters – Structures, algorithms and applications* (Boston: Kluwer Academic)
- Middleton D 1960 *Introduction to statistical communication theory* (New York: McGraw-Hill)