ECE 417 Identification & Adaptive Control Fall 2002

Handout: ***Convergence of Least Squares***

In this handout we consider the noisy linear system

$$y_{k+1} = \varphi_k^T \theta^\circ + v_{k+1}. \tag{1}$$

The least squares algorithm for estimating θ° is defined by the equation

$$\hat{\theta}_{k+1} = \left(\sum_{i=0}^{k} \varphi_i \varphi_i^T \right)^{-1} \left(\sum_{i=0}^{k} \varphi_i y_{i+1} \right) \tag{2}$$

Substituting the expression for y, the parameter estimate is thus given by

$$\hat{\theta}_{k+1} = \theta^\circ + \left(\sum_{i=0}^{k} \varphi_i \varphi_i^T \right)^{-1} \left(\sum_{i=0}^{k} \varphi_i v_{i+1} \right)$$

This equation makes the parameter estimation error explicit. The main question is whether or not the estimates converge, so that $\hat{\theta}_n \to \theta^\circ$ as $n \to \infty$. To ensure convergence, we must impose some conditions on the disturbance sequence $\mathbf{v}$, as the following example shows.

Example of bias Take the simple scalar model

$$y_{t+1} = \theta^\circ y_t + c_1 w_t + w_{t+1}$$

with $\theta^\circ = 0$, and $\mathbf{w}$ a wide sense white process. Then

$$\hat{\theta}_{k+1} = \left(\frac{1}{k} \sum_{i=0}^{k} y_i^2 \right)^{-1} \left(\frac{1}{k} \sum_{i=0}^{k} y_i y_{i+1} \right)$$

The process $\mathbf{y}$ is wide sense stationary with

$$\begin{aligned} R_y(0) &= \lim \frac{1}{N} \sum_{i=0}^{N} y_i^2 = E[(c_1 w_0 + w_1)^2] = \sigma_w^2 (c_1^2 + 1) \\ R_y(1) &= \lim \frac{1}{N} \sum_{i=0}^{N} y_i y_{i+1} = E[(c_1 w_0 + w_1)(c_1 w_1 + w_2)] = \sigma_w^2 c_1 \end{aligned}$$

It follows that

$$\lim_{n \to \infty} \hat{\theta}_n = \frac{c_1}{1 + c_1^2},$$

and hence, the parameter estimates do not converge unless $c_1 = 0$. This is precisely the case

$$y_{t+1} = \theta^\circ y_t + w_{t+1},$$

so the noise is wide sense white.

White noise The sequence $\mathbf{v}$ is called *white* if

(i) $\{v(0), v(1), \ldots\}$ are independent;

(ii) $\{v(0), v(1), \ldots\}$ have identical distributions;

(iii) The mean is zero: $\mathsf{E}[v(t)] = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} v(k) = 0$;

(iv) The variance is finite: $\sigma_v^2 = \mathsf{E}[v(t)^2] = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} v(k)^2 < \infty$;

(v) The fourth moment is also finite: $\mathsf{E}[v(t)^4] = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} v(k)^4 < \infty$.

White noise sequences satisfy the following important properties. Let $z_k = f(v(1), \ldots, v(k), k)$ be a non-anticipative function of the noise. We then have

(P1) If $\sum z_k^2 = \infty$ then

$$\lim_{N\to\infty} \left(\sum_{k=0}^{N} z_k^2 \right)^{-1} \sum_{k=0}^{N} z_k v_{k+1} = 0;$$

while if $\sum z_k^2 < \infty$ then

$$\lim_{N\to\infty} \sum_{k=0}^{N} z_k v_{k+1} \quad \text{exists as a finite number.}$$

(P2) If $|z_k| \le \bar{z}$ for some constant $\bar{z}$, and all $k \ge 0$, then

$$\sum_{k=0}^{N} |z_k| v_{k+1}^2 = O(\sum_{k=0}^{N} |z_k|)$$

The "big O" notation in (P2) means that $\sum_{k=0}^{N} |z_k| v_{k+1}^2 \le K(\sum_{k=0}^{N} |z_k|)$ for some $K < \infty$. A "little o" notation is also commonly used: property (P1) is often written

$$\sum_{k=0}^{N} z_k v_{k+1} = o(\sum_{k=0}^{N} z_k^2)$$

The proof of these results is beyond the scope of this course, though the proofs are not difficult after ECE 434. For those of you with some probability background: the properties (P1) and (P2) follow from the Martingale Convergence Theorem and Kronecker's Lemma.

Convergence of Least Squares in White Noise

One of the most elegant bounds in system identification is due to Lai, 1982. The Recursive Least Squares (RLS) algorithm satisfies the recursive equations:

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \frac{P_{n-1}\varphi_n e_{n+1}}{1+\varphi_n^T P_{n-1}\varphi_n}, \qquad P_n = P_{n-1} - \frac{P_{n-1}\varphi_n\varphi_n^T P_{n-1}}{1+\varphi_n^T P_{n-1}\varphi_n}$$

where $e_{n+1} = y_{n+1} - \hat{\theta}_n^T\varphi_n$ is the *prediction error*. The *parameter estimation error* is defined as $\tilde{\theta}_n = \theta^\circ - \hat{\theta}_n$.

> *Suppose that the disturbance* $\mathbf{v}$ *in (1) is a white noise sequence, and that for each* n*, the noise variable* v_n *is statistically independent of* $\{\phi_i, y_i, v_i, i \le n-1\}$*. Then the RLS algorithm satisfies*
>
> $$\begin{aligned} |\tilde{\theta}_n|^2 &= O\Big(\frac{\log \lambda_{\max}(P_{n-1}^{-1})}{\lambda_{\min}(P_{n-1}^{-1})}\Big) \\ \sum_{i=0}^{n} \frac{|\varphi_i^T\tilde{\theta}_i|^2}{1+\varphi_i^T P_{i-1}\varphi_i} &= O\Big(\log \lambda_{\max}(P_n^{-1})\Big), \qquad n \ge 1. \end{aligned}$$

If the regression sequence is weakly persistently exciting, so that

$$\frac{1}{N}\sum_{1}^{N} \varphi_k\varphi_k^T \to Q > 0, \qquad N \to \infty,$$

it follows that from the first inequality that

$$|\tilde{\theta}_n|^2 \le \text{Const.}\ \frac{\log(n)}{n} \to 0 \qquad \text{as } n \to \infty.$$

The second inequality is useful in bounding the prediction error $e_{i+1} = \varphi_i^T\tilde{\theta}_i + v_{i+1}$.

A Lyapunov Recursion The proof of this result is based upon properties of white noise, and an examination of the "Lyapunov function"

$$Q_n = \tilde{\theta}_n^T P_{n-1}^{-1}\tilde{\theta}_n$$

Finding appropriate bounds on Q_n will give the result, since $|\tilde{\theta}_n|^2 \le Q_n/\lambda_{\min}(P_{n-1}^{-1})$.

From the RLS recursions we have

$$\tilde{\theta}_{n+1} = \left(I - \frac{P_{n-1}\varphi_n\varphi_n^T}{1+\varphi_n^T P_{n-1}\varphi_n}\right)\tilde{\theta}_n - \frac{P_{n-1}\varphi_n v_{n+1}}{1+\phi_n^T P_{n-1}\varphi_n}$$

Substituting the recursion for P_n then gives

$$\tilde{\theta}_{n+1} = P_n P_{n-1}^{-1}\tilde{\theta}_n - \frac{P_{n-1}\varphi_n v_{n+1}}{1+\varphi_n^T P_{n-1}\varphi_n}$$

So, the recursion for Q_n becomes

$$
\begin{aligned}
Q_{n+1} = \tilde{\theta}_{n+1}^T P_n^{-1} \tilde{\theta}_{n+1} &= \tilde{\theta}_{n+1}^T P_{n-1}^{-1} P_n P_{n-1}^{-1} \tilde{\theta}_n - 2\hat{\theta}_n P_{n-1}^{-1} \left(\frac{P_{n-1}\varphi_n v_{n+1}}{1+\varphi_n^T P_{n-1}\theta_n} \right) \\
&\quad + v_{n+1}^2 \frac{\varphi_n^T P_{n-1} P_n^{-1} P_{n-1}\varphi_n}{(1+\varphi_n^T P_{n-1}\varphi_n)^2}
\end{aligned}
$$

The terms involving products of P_i can be calculated as follows:

$$
\begin{aligned}
P_{n-1}^{-1} P_n P_{n-1}^{-1} &= P_{n-1}^{-1} \left(P_{n-1} - \frac{P_{n-1}\varphi_n \varphi_n^T P_{n-1}}{1+\varphi_n^T P_{n-1}\varphi_n} \right) P_{n-1}^{-1} \\
&= P_{n-1}^{-1} - \frac{\varphi_n \varphi_n^T}{1+\varphi_n^T P_{n-1}\varphi_n} \\
P_{n-1} P_n^{-1} P_{n-1} &= P_{n-1}(P_{n-1}^{-1} + \varphi_n \varphi_n^T) P_{n-1} \\
&= P_{n-1} + P_{n-1}\varphi_n \varphi_n^T P_{n-1}
\end{aligned}
$$

We then obtain the following recursive equation:

$$
\begin{aligned}
Q_{n+1} &= Q_n - \frac{\mid \tilde{\theta}_n^T \varphi_n \mid^2}{1+\varphi_n^T P_{n-1}\varphi_n} - 2\frac{\tilde{\theta}_n^T \varphi_n v_{n+1}}{1+\varphi_n^T P_{n-1}\varphi_n} \\
&\quad + v_{n+1}^2 \frac{\varphi_n^T P_{n-1}\varphi_n}{1+\varphi_n^T P_{n-1}\varphi_n}
\end{aligned}
$$

Iterating this bound gives

$$
\begin{aligned}
Q_{n+1} &= Q_0 - \sum_{i=0}^{n} \frac{\mid \tilde{\theta}_i^T \varphi_i \mid^2}{1+\varphi_i^T P_{i-1}\varphi_i} - 2\sum_{i=0}^{n} \frac{\tilde{\theta}_i^T \varphi_i v_{i+1}}{1+\varphi_i^T P_{i-1}\varphi_i} \\
&\quad + \sum_{i=0}^{n} \frac{\varphi_i^T P_{i-1}\varphi_i}{1+\varphi_i^T P_{i-1}\varphi_i} v_{i+1}^2
\end{aligned}
$$

All of the terms on the right hand side are easily bounded using properties of white noise. We have

$$
\begin{aligned}
\sum_{i=0}^{n} \frac{\tilde{\theta}_i^T \varphi_i v_{i+1}}{1+\varphi_i^T P_{i-1}\varphi_i} &= o\left(\sum_{i=0}^{n} \frac{\mid \tilde{\theta}_i^T \varphi_i \mid^2}{(1+\varphi_i^T P_{i-1}\varphi_i)^2} \right) \\
\sum_{i=0}^{n} \frac{\varphi_i^T P_{i-1}\varphi_i}{1+\varphi_i^T P_{i-1}\varphi_i} v_{i+1}^2 &= O\left(\sum_{i=0}^{n} \frac{\varphi_i^T P_{i-1}\varphi_i}{1+\varphi_i^T P_{i-1}\varphi_i} \right)
\end{aligned}
$$

where the first bound follows from (P1), and the second follows from (P2).

Combining these bounds finally gives

$$
Q_{n+1} + (1+o(1)) \sum_{i=0}^{n} \frac{\mid \tilde{\theta}_i^T \varphi_i \mid^2}{1+\varphi_i^T P_{i-1}\varphi_i} = Q_0 + O\left(\sum_{i=0}^{n} \frac{\varphi_i^T P_{i-1}\varphi_i}{1+\varphi_i^T P_{i-1}\varphi_i} \right).
$$

Hence *both* Q_{n+1} and $\sum_{i=0}^{n} \frac{|\tilde{\theta}_i^T \varphi_i|^2}{1+\varphi_i^T P_{i-1}\varphi_i}$ may be bounded by

$$O\left(\sum_{i=0}^{n} \frac{\varphi_i^T P_{i-1}\varphi_i}{1+\varphi_i^T P_{i-1}\varphi_i}\right)$$

To complete the proof, we must obtain

Bounds on $\sum_{i=0}^{n} \frac{\varphi_i^T P_{i-1}\varphi_i}{1+\varphi_i^T P_{i-1}\varphi_i}$ We have by definition,

$$\begin{aligned} P_i^{-1} &= P_{i-1}^{-1} + \varphi_i \varphi_i^T \\ &= P_{i-1}^{-\frac{1}{2}}(I + (P_{i-1}^{\frac{1}{2}}\varphi_i)(P_{i-1}^{\frac{1}{2}}\varphi_i)^T)P_{i-1}^{-\frac{1}{2}} \end{aligned}$$

Let $|\cdot|$ denote the determinant of $(\cdot)$. We have

$$|P_i^{-1}| = |P_{i-1}^{-\frac{1}{2}}|\ |I + (P_{i-1}^{\frac{1}{2}}\varphi_i)(P_{i-1}^{\frac{1}{2}}\varphi_i)^T|\ |P_{i-1}^{-\frac{1}{2}}|$$

One can compute this determinant as follows:

$$|P_i|^{-1} = |P_{i-1}|^{-1}(1 + \varphi_i^T P_{i-1}\varphi_i)$$

Solving this equation gives the following formulae:

$$\begin{aligned} \varphi_i^T P_{i-1}\varphi_i &= \frac{|P_i|^{-1} - |P_{i-1}|^{-1}}{|P_{i-1}|^{-1}} \\ \frac{\varphi_i^T P_{i-1}\varphi_i}{1+\varphi_i^T P_{i-1}\varphi_i} &= \frac{|P_i|^{-1} - |P_{i-1}|^{-1}}{|P_i|^{-1}} \end{aligned}$$

The following calculation almost completes the proof of Lai's result:

$$\sum_{i=0}^{n} \frac{\varphi_i^T P_{i-1}\varphi_i}{1+\varphi_i^T P_{i-1}\varphi_i} = \sum_{i=0}^{n} \frac{|P_i|^{-1} - |P_{i-1}|^{-1}}{|P_i|^{-1}} \le \log\left(\frac{|P_{-1}|}{|P_n|}\right).$$

The last inequality follows from the bound $-x+1 \le -\log(x)$, $x > 0$.

To bound this final term, observe that

$$|P_n|^{-1} = |P_n^{-1}| \le \dim(P_n)\lambda_{\max}(P_n^{-1})$$

□