

15.4 Recursive Identification

Exercise 4.1 (9.1): Derivation of the real-time RLS algorithm.

Show that the weighted RLS algorithm

$$\begin{cases} \hat{\theta}_t = \hat{\theta}_{t-1} + K_t \epsilon_t \\ \epsilon_t = y_t - \varphi_t^T \hat{\theta}_{t-1} \\ K_t = \mathbf{P}_t \varphi_t \\ \mathbf{P}_t = \frac{1}{\lambda_t} \left[\mathbf{P}_{t-1} - \frac{\mathbf{P}_{t-1} \varphi_t \varphi_t^T \mathbf{P}_{t-1}}{\lambda_t + \varphi_t^T \mathbf{P}_{t-1} \varphi_t} \right] \end{cases}$$

solves in each step the problem

$$\theta_t = \underset{\theta}{\operatorname{argmin}} \sum_{s=1}^t \lambda^{t-s} \epsilon_s^2(\theta)$$

where $\epsilon_s^2(\theta) = y_s - \varphi_s^T \hat{\theta}$ for all $s = 1, \dots, t$.

Solution.

The solution to the least squares problem is given as

$$\hat{\theta}_t = \mathbf{R}_t^{-1} \sum_{s=1}^t \lambda^{t-s} \varphi_s y_s$$

where

$$\mathbf{R}_t = \sum_{s=1}^t \lambda^{t-s} \varphi_s \varphi_s^T.$$

Then we have that

$$\begin{aligned} \hat{\theta}_t &= \mathbf{R}_t^{-1} \left(\sum_{s=1}^t \lambda^{t-s} \varphi_s (y_s - \varphi_s^T \hat{\theta}_{t-1}) \right) + \hat{\theta}_{t-1} \\ &= \hat{\theta}_{t-1} + \mathbf{R}_t^{-1} \left(\lambda \sum_{s=1}^{t-1} \lambda^{t-s} \varphi_s (y_s - \varphi_s^T \hat{\theta}_{t-1}) + \varphi_t (y_t - \varphi_t^T \hat{\theta}_{t-1}) \right) \end{aligned}$$

The first of the two terms within the large brackets equals zero, Hence

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \mathbf{R}_t^{-1} (y_t - \varphi_t^T \hat{\theta}_{t-1}) \varphi_t \quad (i)$$

. Turn now to computing the terms \mathbf{R}_t^{-1} in a recursive manner, since

$$\mathbf{R}_t = \lambda \mathbf{R}_{t-1} + \varphi_t \varphi_t^T$$

and by the matrix inversion lemma, we have

$$\mathbf{R}_t^{-1} = \frac{1}{\lambda} \mathbf{R}_{t-1}^{-1} - \frac{1}{\lambda^2} \frac{\mathbf{R}_{t-1}^{-1} \varphi_t \varphi_t^T \mathbf{R}_{t-1}^{-1}}{1 + \frac{1}{\lambda} \varphi_t^T \mathbf{R}_{t-1}^{-1} \varphi_t}.$$

Using the above expression, we get a computationally more attractive expression for the gain vector in (i) as

$$\mathbf{K}_t = \mathbf{R}_t^{-1} \varphi_t = \frac{\mathbf{R}_{t-1}^{-1} \varphi_t}{\lambda + \varphi_t^T \mathbf{R}_{t-1}^{-1} \varphi_t}$$

which concludes the proof.

Exercise 4.2 (9.2): Influence of forgetting factor on consistency properties of parameter estimates.

Consider the static-gain system

$$y_t = bu_t + e_t, \quad \forall t = 1, 2, \dots$$

where

$$\mathbb{E}[e_t] = 0, \quad E[e_s e_t] = \delta_{t,s}$$

and u_t is a persistently exciting nonrandom signal. The unknown parameter b is estimated as

$$\hat{b} = \operatorname{argmin}_b \sum_{t=1}^n \lambda^{n-t} (y_t - bu_t)^2$$

where n denotes the number of datapoints, and the forgetting factor λ satisfies $0 < \lambda \leq 1$. Determine $\operatorname{var}(\hat{b})$. Show that for $n \rightarrow \infty$ one has $\operatorname{var}(\hat{b}) = 0$. Also, show that for $\lambda < 1$ there are signals u_t for which consistency is not obtained.

Hint. Consider the signal where u_t is constant.

Solution. Simple calculation gives that

$$\hat{b} = \frac{\sum_{t=1}^n \lambda^{n-t} y_t u_t}{\sum_{t=1}^n \lambda^{n-t} u_t^2} = b + \frac{\sum_{t=1}^n \lambda^{n-t} e_t u_t}{\sum_{t=1}^n \lambda^{n-t} u_t^2}.$$

Thus

$$\operatorname{var}(\hat{b}) = \frac{(\sum_{t=1}^n \sum_{s=1}^n \lambda^{2n-s-t} u_s u_t \mathbb{E}[e_s e_t])}{(\sum_{t=1}^n \lambda^{n-t} u_t^2)^2}$$

or

$$\operatorname{var}(\hat{b}) = \frac{(\sum_{t=1}^n \lambda^{2(n-t)} u_t^2)}{(\sum_{t=1}^n \lambda^{n-t} u_t^2)^2} \quad (i)$$

Let $\lambda = 1$, then (i) gives

$$\operatorname{var}(\hat{b}) = \frac{1}{\sum_{t=1}^n u_t^2}.$$

Since u_t is PE, it follows that $\sum_{t=1}^n u_t^2 \rightarrow \infty$. Hence part 3 of the question follows.

If $\lambda < 1$, then consistency of \hat{b} might be lost. To exemplify this, let $u_t = 1$ for all $t = 1, \dots, n$. This is PE of order 1. Then from (i) above we have that

$$\operatorname{var}(\hat{b}) = \frac{(1 - \lambda^{2n}) (1 - \lambda)^2}{(1 - \lambda^2) (1 - \lambda^n)^2} = \frac{(1 + \lambda^n) (1 - \lambda)}{(1 - \lambda^n) (1 + \lambda)} \rightarrow \frac{1 - \lambda}{1 + \lambda} > 0$$

if $n \rightarrow \infty$. The lack of consistency in such case might be explained as follows For $\lambda < 1$, 'old' measurements are weighted out from the criterion, so that the effective number of samples used in estimating \hat{b} does not increase with growing n .

Exercise 4.3 (9.4): Convergence properties and dependence on initial conditions of the RLS estimate.

Consider the model

$$y_t = \varphi_t^T \theta_0 + \epsilon_t$$

Let the *offline* weighted LS estimate of θ_0 up to instant t be

$$\hat{\theta}_t = \left(\sum_{s=1}^t \lambda^{t-s} \varphi_s \varphi_s^T \right)^{-1} \left(\sum_{s=1}^t \lambda^{t-s} \varphi_s y_s \right)$$

Consider also the online weighted RLS estimates $\{\bar{\theta}_s\}_s$

- (i) Derive the difference equations for \mathbf{P}_t^{-1} and $\mathbf{P}_t^{-1} \hat{\theta}_t$. Solve these equations to find how $\hat{\theta}_t$ depends on the initial values $\hat{\theta}_0$ and \mathbf{P}_0 and on the forgetting factor λ .
- (ii) Let $\mathbf{P}_0 = \rho I_n$, then prove that for every t where $\bar{\theta}_t$ exists

$$\lim_{\rho \rightarrow \infty} \hat{\theta}_t = \bar{\theta}_t$$

- (iii) Suppose that $\bar{\theta}_t$ is bounded, and suppose that $\lambda^t \mathbf{P}_t \rightarrow 0$ as $t \rightarrow \infty$. Prove that

$$\lim_{t \rightarrow \infty} (\hat{\theta}_t - \bar{\theta}_t) = 0$$

Solution.

We have that

$$\hat{\theta}_t = \left(\sum_{s=1}^t \lambda^{t-s} \varphi_s \varphi_s^T \right)^{-1} \left(\sum_{s=1}^t \lambda^{t-s} \varphi_s y_s \right)$$

or

$$\begin{cases} \hat{\theta}_t = \hat{\theta}_{t-1} + \mathbf{P}_t \varphi_t (y_t - \varphi_t^T \hat{\theta}_{t-1}) \\ \mathbf{P}_t = \frac{1}{\lambda} \left(\mathbf{P}_{t-1} - \frac{\mathbf{P}_{t-1} \varphi_t \varphi_t^T \mathbf{P}_{t-1}}{\lambda + \varphi_t^T \mathbf{P}_{t-1} \varphi_t} \right) \end{cases}$$

- (a) We can also write that

$$\mathbf{P}_t^{-1} = \lambda (\lambda \mathbf{P}_{t-1}^{-1} + \varphi_t \varphi_t^T).$$

This is a linear difference equation in \mathbf{P}_t^{-1} or

$$\mathbf{P}_t = \mathbf{P}_0^{-1} + \sum_{s=1}^t \varphi_s \varphi_s^T$$

Define $\mathbf{x}_t = \mathbf{P}_t^{-1} \hat{\theta}_t$, then

$$\mathbf{x}_t = \mathbf{P}_t^{-1} \left(\hat{\theta}_{t-1} + \mathbf{P}_t \varphi_t (y_t - \varphi_t^T \hat{\theta}_t) \right) = (\mathbf{P}_t^{-1} - \varphi_t \varphi_t^T) \hat{\theta}_{t-1} + \varphi_t^T y_t = \lambda \mathbf{x}_{t-1} + \varphi_t y_t.$$

Solving this linear difference equation in \mathbf{x}_t gives

$$\mathbf{x}_t = \lambda^t \mathbf{x}_0 + \sum_{s=1}^t \lambda^{t-s} \varphi_s y_s.$$

Thus we find that

$$\hat{\theta}_t = \mathbf{P}_t \mathbf{x}_0 = \left(\lambda^t \mathbf{P}_0^{-1} + \sum_{s=1}^n \lambda^{t-s} \varphi_s \varphi_s^T \right)^{-1} \left(\lambda^t \mathbf{P}_0^{-1} \hat{\theta}_0 + \sum_{s=1}^n \lambda^{t-s} \varphi_s y_s \right)$$

(b). Let $\mathbf{P}_0 = \rho I_d$. Assume θ_t exists, then $\sum_{s=1}^t \lambda^{t-s} \varphi_s \varphi_s^T$ is invertible. We get

$$\hat{\theta}_t = \left(\frac{\lambda^t}{\rho} I_d + \sum_{s=1}^t \lambda^{t-s} \varphi_s \varphi_s^T \right)^{-1} \left(\frac{\lambda^t}{\rho} \theta_0 + \sum_{s=1}^t \lambda^{t-s} \varphi_s y_s \right)$$

which tends to $\bar{\theta}_t$ when $t \rightarrow \infty$.

(c). If $\lambda^t \mathbf{P}_t \rightarrow 0$ when $t \rightarrow \infty$, then

$$\hat{\theta}_t - \bar{\theta}_t = \mathbf{P}_t \left(\lambda^t \mathbf{P}_0^{-1} \hat{\theta}_t + \sum_{s=1}^t \lambda^{t-s} \varphi_s y_s - \mathbf{P}_t^{-1} \bar{\theta}_t \right)$$

and hence

$$= \mathbf{P}_t \left(\lambda^t \mathbf{P}_0^{-1} \hat{\theta}_t - \lambda^t \mathbf{P}_t^{-1} \bar{\theta}_t \right)$$

or

$$= \mathbf{P}_t \lambda^t \mathbf{P}_0^{-1} (\hat{\theta}_t - \bar{\theta}_t)$$

which tends to 0 if $t \rightarrow \infty$.

Exercise 4.4 (9.10): An RLS algorithm with a sliding window

Consider the parameter estimate

$$\hat{\theta}_t = \operatorname{argmin}_{\theta} \sum_{s=t-m+1}^t \epsilon_s^2(\theta)$$

where $\epsilon_s(\theta) = y_s - \varphi_s^T \theta$. The number m is the size of the sliding window. Show that such $\hat{\theta}_t$ can be computed recursively as

$$\begin{cases} \hat{\theta}_t = \hat{\theta}_{t-1} + \mathbf{K}_1 \epsilon(t, \hat{\theta}_{t-1}) - \mathbf{K}_2 \epsilon(t-m, \hat{\theta}_{t-1}) \\ \mathbf{K}_1 = \mathbf{P}_{t-1} \varphi_t \left(I + \begin{bmatrix} \varphi_t^T \\ -\varphi_{t-m}^T \end{bmatrix} \mathbf{P}_{t-1} \begin{bmatrix} \varphi_t & \varphi_{t-m} \end{bmatrix} \right)^{-1} \\ \mathbf{K}_2 = \mathbf{P}_{t-1} \varphi_{t-m} \left(I + \begin{bmatrix} \varphi_t^T \\ -\varphi_{t-m}^T \end{bmatrix} \mathbf{P}_{t-1} \begin{bmatrix} \varphi_t & \varphi_{t-m} \end{bmatrix} \right)^{-1} \\ \mathbf{P}_t = \mathbf{P}_{t-1} - \begin{bmatrix} \mathbf{K}_1 & \mathbf{K}_2 \end{bmatrix} \begin{bmatrix} \varphi_t^T \\ -\varphi_{t-m}^T \end{bmatrix} \mathbf{P}_{t-1} \end{cases}$$

Solution.

Set

$$\mathbf{P}_t = \left(\sum_{s=t-m+1}^t \varphi_s \varphi_s^T \right)^{-1}$$

Then

$$\mathbf{P}_t^{-1} = \mathbf{P}_{t-1}^{-1} + \varphi_t \varphi_t^T - \varphi_{t-m} \varphi_{t-m}^T$$

and hence

$$\begin{aligned} \hat{\theta}_t &= \mathbf{P}_t \sum_{s=t-m+1}^t \varphi_s y_s = \mathbf{P}_t \left(\varphi_t y_t - \varphi_{t-m} y_{t-m} + \sum_{s=t-m}^{t-1} \varphi_s y_s \right) \\ &= \mathbf{P}_t \left(\varphi_t y_t - \varphi_{t-m} y_{t-m} + \mathbf{P}_{t-1}^{-1} \hat{\theta}_{t-1} \right). \end{aligned}$$

Set

$$(K_t^1, K_t^2) = \mathbf{P}_t (\varphi_t, \varphi_{t-m})$$

Then

$$\hat{\theta}_t = K_t^1 y_t - K_t^2 y_{t-m} + \mathbf{P}_t (\mathbf{P}_t^{-1} - \varphi_t \varphi_t^T + \varphi_{t-m} \varphi_{t-m}^T)$$

or

$$= \hat{\theta}_0 + K_t^1 e_t(\hat{\theta}_{t-1}) - K_t^2 e_{t-m}(\hat{\theta}_{t-1})$$

Application of the matrix inversion lemma gives that

$$\mathbf{P}_t = \mathbf{P}_{t-1} - \left(\mathbf{P}_{t-1} \begin{bmatrix} \varphi_t & \varphi_{t-m} \end{bmatrix} \left(I_2 + \begin{bmatrix} \varphi_t^T \\ \varphi_{t-m}^T \end{bmatrix} \mathbf{P}_{t-1} \begin{bmatrix} \varphi_t & \varphi_{t-m} \end{bmatrix} \right)^{-1} \begin{bmatrix} \varphi_t^T \\ \varphi_{t-m}^T \end{bmatrix} \mathbf{P}_{t-1} \right)^{-1}.$$

It remains to be verified that K_t^1 and K_t^2 as defined above satisfies the relation stated in the problem. Let

$$\tilde{K}_t = \mathbf{P}_{t-1} (\varphi_t, \varphi_{t-m})$$

Straightforward calculation then gives that

$$(K_t^1, K_t^2) = \mathbf{P}_t (\varphi_t, \varphi_{t-m})$$

or

$$\begin{aligned} &= \tilde{K}_t - \tilde{K}_t \left(I_d + \begin{bmatrix} \varphi_t^T \\ \varphi_{t-m}^T \end{bmatrix} \tilde{K}_t \right)^{-1} \begin{bmatrix} \varphi_t^T \\ \varphi_{t-m}^T \end{bmatrix} \tilde{K}_t \\ &= \tilde{K}_t \left(I_d + \begin{bmatrix} \varphi_t^T \\ \varphi_{t-m}^T \end{bmatrix} \tilde{K}_t \right)^{-1} \end{aligned}$$

as desired.