Woodbury's Identity (Matrix Inversion Lemma)

Let **A** and **B** be two positive definite $M \times M$ matrices related by

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{X}\boldsymbol{\Delta}^{-1}\mathbf{X}^{H}$$

where **D** is another positive definite $N \times N$ matrix, and **C** is an $M \times N$ matrix. We may express the inverse of **A** as

$$\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B}\mathbf{C}\left(\mathbf{D} + \mathbf{C}^{H}\mathbf{B}\mathbf{C}\right)^{-1}\mathbf{C}^{H}\mathbf{B}.$$

We now apply Woodbury's identity to the recursive computation of $\Phi^{-1}(n)$ as follows. Substitute the following into Woodbury

$$\begin{aligned} \mathbf{A} &= \Phi(\nu) \\ \mathbf{B}^{-1} &= \lambda\Phi(\nu-1) \\ \mathbf{X} &= \boldsymbol{\upsilon}(\nu) \\ \boldsymbol{\Delta} &= 1 \end{aligned}$$

and we have our desired result

$$\begin{aligned} \Phi^{-1}(n) &= \lambda^{-1}\Phi^{-1}(n-1) - \frac{\lambda^{-2}\Phi^{-1}(n-1)\mathbf{u}(n)\mathbf{u}^{H}(n)\Phi^{-1}(n-1)}{1+\lambda^{-1}\mathbf{u}^{H}(n)\Phi^{-1}(n-1)\mathbf{u}(n)} \\ &= \lambda^{-1}\Phi^{-1}(n-1) - \lambda^{-1}\frac{\lambda^{-1}\Phi^{-1}(n-1)\mathbf{u}(n)}{1+\lambda^{-1}\mathbf{u}^{H}(n)\Phi^{-1}(n-1)\mathbf{u}(n)}\mathbf{u}^{H}(n)\Phi^{-1}(n-1) \end{aligned}.$$

We note that we've replaced matrix inversion with a simple scalar division in the recursion. For convenience let

$$\mathbf{P}(n) = \Phi^{-1}(\nu) \quad \text{(inverse correlation matrix)}$$

and

$$\mathbf{k}(n) = \frac{\lambda^{-1}\boldsymbol{\Pi}(\nu-1)\boldsymbol{\upsilon}(\nu)}{1+\lambda^{-1}\boldsymbol{\upsilon}^{H}(\nu)\boldsymbol{\Pi}(\nu-1)\boldsymbol{\upsilon}(\nu)} \quad \text{(gain vector).}$$

We thus rewrite our recursion as

$$\begin{aligned} \mathbf{P}(n) &= \lambda^{-1}\mathbf{P}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{u}^{H}(n)\mathbf{P}(n-1) \\ &= \lambda^{-1}[\mathbf{I} - \mathbf{k}(n)\mathbf{u}^{H}(n)]\mathbf{P}(n-1) \end{aligned} \quad \text{(Riccati equation).}$$

Note that

$$\begin{aligned} \mathbf{k}(n) &= \lambda^{-1}\boldsymbol{\Pi}(\nu-1)\boldsymbol{\upsilon}(\nu) - \lambda^{-1}\boldsymbol{\kappa}(\nu)\boldsymbol{\upsilon}^{H}(\nu)\boldsymbol{\Pi}(\nu-1)\boldsymbol{\upsilon}(\nu) \\ &= [\lambda^{-1}\boldsymbol{\Pi}(\nu-1) - \lambda^{-1}\boldsymbol{\kappa}(\nu)\boldsymbol{\upsilon}^{H}(\nu)\boldsymbol{\Pi}(\nu-1)]\boldsymbol{\upsilon}(\nu) \\ &= \boldsymbol{\Pi}(\nu)\boldsymbol{\upsilon}(\nu) \end{aligned}$$

which is to say that the gain vector is $\mathbf{u}(n)$ transformed by $\mathbf{P}(n)$ or $\Phi^{-1}(n)$, i.e. projected onto the subspace defined by $\mathbf{P}(n)$.

The Exponentially Weighted Recursive Least-Squares Algorithm

We begin with our solution to the normal equations and our recursion for $\mathbf{z}(n)$

$$\begin{aligned} \hat{\mathbf{w}}(n) &= \mathbf{\Pi}(\nu)\boldsymbol{\zeta}(\nu) \\ &= \mathbf{\Pi}(\nu)[\lambda\boldsymbol{\zeta}(\nu-1) + \boldsymbol{\upsilon}(\nu)\delta^{*}(\nu)] \\ &= \lambda\mathbf{\Pi}(\nu)\boldsymbol{\zeta}(\nu-1) + \mathbf{\Pi}(\nu)\boldsymbol{\upsilon}(\nu)\delta^{*}(\nu) \end{aligned} .$$

We next substitute our recursion for **P**(n) into the first term on the RHS

$$\begin{aligned} \tilde{w}(n) &= \mathbf{\Pi}(\nu-1)\boldsymbol{\zeta}(\nu-1) - \kappa(\nu)\upsilon^{H}(\nu)\mathbf{\Pi}(\nu-1)\boldsymbol{\zeta}(\nu \\ &= \omega(\nu-1) - \kappa(\nu)\upsilon^{H}(\nu)\omega(\nu-1) + \mathbf{\Pi}(\nu)\upsilon(\nu \\ &= \omega(\nu-1) + \kappa(\nu)[\delta^{*}(\nu) - \upsilon^{H}(\nu)\omega(\nu-1)] \\ &= \omega(\nu-1) + \kappa(\nu)\xi^{*}(\nu) \end{aligned}$$

where

$$\xi(n) = d(n) - \hat{\mathbf{w}}^{H}(n-1)\mathbf{u}(n) .$$

We'll call ξ(n) the *a priori estimation error* as opposed to the *a posteriori estimation error*

$$e(n) = \delta(\nu) - \hat{\boldsymbol{\omega}}^{H}(\nu)\boldsymbol{\upsilon}(\nu)$$

which involves the current filter coefficient vector.

Table 13.1 Summary of the RLS Algorithm.

Initialization of the RLS Algorithm (Soft-Constrained Initialization)

In order to get RLS started we must initialize the Riccati equation

$$\mathbf{P}(n) = \lambda^{-1}\boldsymbol{\Pi}(\nu-1) - \lambda^{-1}\boldsymbol{\kappa}(\nu)\boldsymbol{\upsilon}^{H}(\nu)\boldsymbol{\Pi}(\nu-1)$$

by choosing $\mathbf{P}(0) = \boldsymbol{\Phi}^{-1}(0)$ appropriately. In order to do this we'll modify the deterministic correlation matrix as follows

$$\Phi(n) = \sum_{i=1}^{n} \lambda^{n-i}\mathbf{u}(i)\mathbf{u}^{H}(i) + \delta\lambda^{n}\mathbf{I}$$

where $\delta < 0.01\sigma_u^2$ where σ_u^2 is the variance of the input signal. Clearly as *n* increases the influence of the modification decreases. We thus have

$$\begin{aligned}\boldsymbol{\Phi}^{-1}(0) &= \mathbf{P}(0)\\ &= \delta^{-1}\mathbf{I}\end{aligned}.$$

Finally we choose

$$\hat{\mathbf{w}}(0) = \mathbf{0}\ .$$

We note that the modification means we are no longer finding the LS solution to our original cost function. Instead our LS solution minimizes a modified cost function

$$\varepsilon(n) = \delta\lambda^{n}\left\|\hat{\mathbf{w}}(n)\right\|^{2} + \sum_{i=1}^{n}\lambda^{n-i}\left|e(i)\right|^{2}$$

but again we note that as *n* increases this modified cost function is asymptotic to our original.

Properties

Orthogonality Principle

In this section we show that when $\hat{\mathbf{w}}(n)$ has minimized the exponentially-weighted LS cost function, the weighted error vector is orthogonal to the weighted input signal.

We begin with the normal equations (matrix form)

$$\Phi(n)\hat{\mathbf{w}}(n) = \mathbf{z}(n)$$

Let

$$\mathbf{U}(n) = \begin{bmatrix} \upsilon(\nu) & \lambda^{1/2}\upsilon(\nu-1) & \Lambda & \lambda^{(\nu-2)/2}\upsilon(2) & \lambda^{(\nu-1)/2}\upsilon(1) \\ \upsilon(\nu-1) & \lambda^{1/2}\upsilon(\nu-2) & \Lambda & \lambda^{(\nu-1)/2}\upsilon(1) & 0 \\ \mathrm{M} & \mathrm{M} & & \mathrm{M} & \mathrm{M} \\ \upsilon(\nu-M+1) & \lambda^{1/2}\upsilon(\nu-M) & \Lambda & 0 & 0 \end{bmatrix}$$

$$= \left[\boldsymbol{\upsilon}(\nu)\ \middle|\ \lambda^{1/2}\boldsymbol{\upsilon}(\nu-1)\ \middle|\ \Lambda\ \middle|\ \lambda^{(\nu-1)/2}\boldsymbol{\upsilon}(1)\right]$$

and

$$\mathbf{d}(n) = [\,\delta(\nu) \quad \lambda^{1/2}\,\delta(\nu-1) \quad \Lambda \quad \lambda^{(\nu-1)/2}\,\delta(1)\,]^T$$

where **U**(n) is $M \times n$ and **d**(n) is $n \times 1$. The expanded normal equations can then be factored as

$$\begin{aligned}
&\mathbf{U}(n)\mathbf{Y}^H(\nu)\hat{\boldsymbol{\omega}}(\nu) = \mathbf{Y}(\nu)\boldsymbol{\delta}^*(\nu) \\
&\Leftrightarrow \\
&\mathbf{Y}(\nu)\boldsymbol{\delta}^*(\nu) - \mathbf{Y}(\nu)\mathbf{Y}^H(\nu)\hat{\boldsymbol{\omega}}(\nu) = \mathbf{0}
\end{aligned}.$$

We note that

$$\begin{aligned}
\mathbf{U}^H(n)\hat{\boldsymbol{\omega}}(\nu) &= [\,\psi^*(\nu) \quad \lambda^{1/2}\,\psi^*(\nu-1) \quad \Lambda \quad \lambda^{(\nu-1)/2}\,\psi^*(1)\,]^T \\
&= \boldsymbol{\psi}^*(\nu)
\end{aligned}$$

forms the vector of weighted adaptive filter outputs. Thus we have

$$\begin{aligned}
&\mathbf{U}(n)\boldsymbol{\delta}^*(\nu) - \mathbf{Y}(\nu)\boldsymbol{\psi}^*(\nu) = \mathbf{0} \\
&\Leftrightarrow \\
&\mathbf{Y}(\nu)[\boldsymbol{\delta}^*(\nu) - \boldsymbol{\psi}^*(\nu)] = \mathbf{0} \\
&\Leftrightarrow \\
&\mathbf{Y}(\nu)\boldsymbol{\varepsilon}^*(\nu) = \mathbf{0}
\end{aligned}$$

where

$$\mathbf{e}(n) = [\,\varepsilon(\nu) \quad \lambda^{1/2}\,\varepsilon(\nu-1) \quad \Lambda \quad \lambda^{(\nu-1)/2}\,\varepsilon(1)\,]^T$$

is the weighted error vector.

We thus have our result: *the weighted-error vector is orthogonal to all row vectors of* **U**(n), *the weighted input signal.*

Relation Between Least-Squares and Wiener Solutions

With λ = 1 (infinite memory) we have

$$\begin{aligned}
\Phi(n) &= \sum_{i=1}^{n} \mathbf{u}(i)\mathbf{u}^H(i) \\
\mathbf{z}(n) &= \sum_{i=1}^{n} \lambda^{n-i}\mathbf{u}(i)d^*(i)
\end{aligned}.$$

If the input and desired signals are ergodic and stationary then

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{\nu}\Phi(\nu) &= \lambda\iota\mu_{\nu\to\infty}\frac{1}{\nu}\sum_{\iota=1}^{\nu}\boldsymbol{\upsilon}(\iota)\boldsymbol{\upsilon}^H(\iota) \\
&= E[\boldsymbol{\upsilon}(\iota)\boldsymbol{\upsilon}^H(\iota)] \\
&= \mathbf{P}
\end{aligned}$$

and

$$\begin{aligned} \lim_{n\to\infty}\frac{1}{\nu}\boldsymbol{\zeta}(\nu) &= \underset{\nu\to\infty}{\lambda\iota\mu}\frac{1}{\nu}\sum_{\iota=1}^{\nu}\boldsymbol{\upsilon}(\iota)\delta^{*}(\iota) \\ &= E[\boldsymbol{\upsilon}(\iota)\delta^{*}(\iota)] \\ &= \boldsymbol{\pi} \end{aligned}$$

It can then be shown

$$\begin{aligned} \lim_{n\to\infty}\hat{\boldsymbol{\omega}}(\nu) &= \underset{\nu\to\infty}{\lambda\iota\mu}\Phi^{-1}(\nu)\boldsymbol{\zeta}(\nu) \\ &= \mathbf{P}^{-1}\boldsymbol{\pi} \\ &= \boldsymbol{\omega}_{o} \end{aligned}$$

or in other words that the LS solution (with infinite memory) tends to the Wiener solution if the signals are ergodic and stationary.