10 Seguimiento de sistemas variantes en el tiempo

- Se han discutido las familias de algoritmos LMS y RLS considerando el comportamiento **promedio** operando en un ambiente ESA. En ese ambiente la superfície de desempeño es fija y el requerimiento esencial es buscar, paso a paso, el punto de mínimo de esa superfície y en consecuencia asegurar desempeño óptimo o casi óptimo.
- Ahora estudiaremos la operación de esos algoritmos en un **contexto no estacionario**, para el cual la solución óptima de Wiener tiene una forma variante en el tiempo. El resultado es que el punto de mínimo de la superfície de desempeño ya no es fijo.
- En consecuencia, el algoritmo de filtrado adaptivo tiene que adicionar la tarea de **seguimiento** (tracking) del punto de mínimo de la superfície de desempeño. En otras palabras, se requiere que el algoritmo siga continuamente las variaciones estadísticas de la entrada, las que se suponen "suficientemente" lentas para que el seguimiento sea posible.
- El seguimiento es un **fenómeno estacionario**. Esto está en contraste con convergencia que es un fenómeno transitorio. De esta manera, para que un filtro adaptivo ejecute el seguimiento debe pasar del estado de operación transitorio al estacionario y debe poder ajustar los coeficientes del filtro.
- En general estas dos, convergencia y seguimiento, son características bien diferentes de un algoritmo. En particular, un algoritmo de filtrado adaptivo con buenas propiedades de convergencia no tiene necesariamente características de seguimiento rápida y visceversa.

10.1 Modelo de Markov para identificación de sistemas

- Un ambiente no estacionario puede encontrarse en la práctica como:

  1. **El marco de referencia que provee la respuesta deseada puede ser variante en el tiempo.** Estas situaciones aparecen, por ejemplo, en identificación de sistemas cuando se usa un filtro adaptivo para modelar un sistema variante en el tiempo. En este caso, la matriz correlación se mantiene fija (como en un ambiente ESA), mientras que el vector de correlación cruzada es variante en el tiempo.
  2. **El proceso estocástico que provee la entrada del filtro adaptivo es no estacionario.** Estas situaciones aparecen, por ejemplo, cuando un filtro adaptivo se utiliza para ecualizar un canal variante en el tiempo. En este caso tanto la matriz correlación como el vector de correlación cruzada son variantes en el tiempo.

- De esta forma, las características de seguimiento de un sistema variante en el tiempo no son solo dependientes del tipo de filtro adaptivo empleado si no también del **problema específico**.

- Discutiremos un modelo de sistema variante en el tiempo muy popular para la aplicación de identificación de sistemas. Este modelo está gobernado por tres ecuaciones básicas

  1. **Proceso de Markov de primer orden.** El sistema de dinámica desconocida se modela con un filtro de coeficientes $\boldsymbol{w}_o(n)$, cuya variación está definida por un proceso de Markov de primer orden,

$$\boldsymbol{w}_o(n+1) \; = \; a\boldsymbol{w}_o(n) + \boldsymbol{\omega}(n)$$

donde a es un parámetro constante y $\boldsymbol{\omega}(n)$ es un proceso de ruido, de media cero y matriz correlación $\boldsymbol{Q}$. Esto puede interpretarse como $\boldsymbol{\omega}(n)$ como entrada de un filtro con función transferencia $1/(1-a z^{-1})$. Se supone que a es muy próximo a 1. De esta manera el ancho de banda del filtro pasabajos es mucho menor que el de los datos de entrada. En forma equivalente, con este modelo se requerirán varias iteraciones para producir un cambio significativo en $\boldsymbol{w}(n)$.

2. **Respuesta deseada**. Está definida por

$$d(n) \;=\; \boldsymbol{w}_o^H(n)\boldsymbol{u}(n) + \nu(n)$$

donde $\nu(n)$ es el **ruido de medición**, supuesto blanco, de media cero y varianza σ^2. A pesar que $\boldsymbol{u}(n)$ y $\nu(n)$ podrían ser procesos ESA, la respuesta deseada $d(n)$ es un proceso **no estacionario**, dado que $\boldsymbol{w}_o(n)$ varia en el tiempo.

3. **Señal de error**. La señal de error $e(n)$ está definida por

$$e(n) \;=\; d(n) - y(n) = \boldsymbol{w}_o^H(n)\boldsymbol{u}(n) + \nu(n) - \hat{\boldsymbol{w}}^H(n)\boldsymbol{u}(n)$$

donde $\hat{\boldsymbol{w}}(n)$ es el vector de coeficientes del filtro adaptivo supuesto de estructura transversal. Se supone además que el filtro adaptivo tiene el mismo número de coeficientes que el sistema desconocido.

- En base a la definición, y dado que el sistema es variante en el tiempo, $d(n)$ aplicada al filtro adaptivo es no estacionaria. En consecuencia, a pesar que $\boldsymbol{R}$ es constante, o sea, el problema tendrá una superfície de desempeño con la forma de un paraboloide, el valor del mínimo estará continuamente variando.

Suposiciones: Típicamente, las variaciones representadas por $\boldsymbol{\omega}(n)$ en el modelo de Markov son lentas (o sea acotadas). De esta forma algoritmos adecuados de filtrado adaptivo (LMS o RLS) puede **seguir** las variaciones estadísticas del comportamiento dinámico del sistema desconocido.

Para proceder con el análisis de seguimiento de los algoritmos LMS y RLS en el contexto anterior se realizarán las siguientes hipótesis en general

1. El proceso de ruido $\boldsymbol{\omega}(n)$ es independiente de $\boldsymbol{u}(n)$ y el ruido de medición $\nu(n)$.
2. $\boldsymbol{u}(n)$ y el ruido de medición $\nu(n)$ son idependientes.
3. $\nu(n)$ es blanco, de media cero y varianza $\sigma_\nu^2 < \infty$.

Las dos últimas supocisiones fueron realizadas también cuando se analizó los algoritmos LMS y RLS. Dado que las tres cantidades aleatórias asociadas al fenómeno físico considerado no se relacionan en forma simple, la primera suposición permite además un tratamiento matemático accesible. Colectivamente se las denomina **suposición de independencia**.

10.2 Grado de no estacionaridad

- Para discutir una definición clara de la noción, algo ambigua, del significado de variaciones estadísticas lentas o rápidas del modelo, introduciremos el **grado de no estacionaridad**.

- En el contexto del modelo de Markov descripto, el grado de no estacionaridad, α, se define formalmente como

$$\alpha = \left(\frac{E\left[|\boldsymbol{\omega}^H(n)\boldsymbol{u}(n)|^2\right]}{E\left[|\nu(n)|^2\right]} \right)^{1/2} \tag{146}$$

- El grado de no estacionaridad es entonces una característica asociada solo al sistema variante en el tiempo y no tiene nada que ver con el filtro adaptivo.

- El numerador de esta ecuación puede reescribirse como sigue

$$E\left[|\boldsymbol{\omega}^H(n)\boldsymbol{u}(n)|^2\right] = E\left[\boldsymbol{\omega}^H(n)\boldsymbol{u}(n)\boldsymbol{u}^H(n)\boldsymbol{\omega}(n)\right] = tr\left[\boldsymbol{Q}\,\boldsymbol{R}\right]$$

  donde $\boldsymbol{Q}$ es la matriz correlación de $\boldsymbol{\omega}(n)$.

- El denominador de (146) es simplemente la varianza σ_ν^2 del ruido de medición $\nu(n)$. De esta forma

$$\alpha = \frac{1}{\sigma_\nu}(tr[\boldsymbol{R}\,\boldsymbol{Q}])^{1/2} = \frac{1}{\sigma_\nu^2}(tr[\boldsymbol{Q}\,\boldsymbol{R}])^{1/2}$$

- El grado de no estacionaridad α se puede relacionar con el desajuste del filtro adaptivo$\mathcal{M}$.

- Primero, es posible notar que el error medio cuadrático mínimo J_{min} que el filtro puede alcanzar es σ_ν^2.

- Luego, lo mejor que puede realizar un filtro adaptivo en un contexto variante en el tiempo como el definido es producir una vector de error en los coeficientes $\boldsymbol{\epsilon}(n)$ igual al proceso de ruido $\boldsymbol{\omega}(n)$.

- Entonces, en base a la terminología introducida en la discusión del algoritmo LMS, es posible hacer que su matriz correlación asociada $\boldsymbol{K}(n)$ sea igual a $\boldsymbol{Q}$.

- Teniendo en cuenta que el exceso de error medio cuadrático $J_{ex}(n) = tr[\boldsymbol{R}\,\boldsymbol{K}(n)]$, es posible establecer que el exceso de error medio cuadrático a ser alcanzado por el filtro nunca puede ser menor que $tr[\boldsymbol{R}\,\boldsymbol{Q}]$. Teniendo en cuenta la definición de desajuste se tiene entonces que

$$\mathcal{M} = \frac{J_{ex}}{J_{min}} \geq \frac{tr[\boldsymbol{R}\,\boldsymbol{Q}]}{\sigma_{\nu}^{2}} = \alpha^{2}$$

  o en otras palabras, la raíz cuadrada del desajuste $\mathcal{M}$ impone un límite superior al grado de no estacionaridad. En base a esta ecuación es posible concluir que

  1. Para variaciones estadísticas pequeñas, α es pequeño. Esto a su vez significa que debería ser posible construir un filtro adaptivo que pueda seguir al sistema variante en el tiempo.
  2. Cuando las variaciones estadísticas del contexto son muy ráidas α debe ser mayor que uno. En tal caso, el desajuste producido por el filtro adaptivo excede 100 %, lo que significa que no existe ventaja alguna en construir un filtro adaptivo para resolver el problema de seguimiento.

10.3 Criterios de seguimiento

- Cuando el sistema dinámico desconocido se modela con un filtro transversal $\boldsymbol{w}_o(n)$ y se utiliza un filtro adaptivo $\hat{\boldsymbol{w}}(n)$, es posible definir el **vector de error en los coeficientes** como

$$\boldsymbol{\epsilon}(n) = \hat{\boldsymbol{w}}(n)\boldsymbol{w}_o(n)$$

- Teniendo en cuenta $\boldsymbol{\epsilon}(n)$, es posible definir dos figuras de mérito para establecer la capacidad de seguimiento de un filtro adaptivo:

  1. **Desviación media cuadrática**. Definida por

$$\begin{aligned}\mathcal{D}(n) &= E[\|\hat{\boldsymbol{w}}(n) - \boldsymbol{w}_o(n)\|^2] \\ &= E[\|\boldsymbol{\epsilon}(n)\|^2] \\ &= tr[\boldsymbol{K}(n)]\end{aligned}$$

     donde n se supone suficientemente grande como para que el filtro adaptivo se encuentre en estado estacionario. Claramente $\mathcal{D}(n)$ debería ser pequeña para un buen desempeño.
     El vector de error en los coeficientes $\boldsymbol{\epsilon}(n)$ puede expresarse como la suma de dos componentes

$$\begin{aligned}\boldsymbol{\epsilon}(n) &= \boldsymbol{\epsilon}_1(n) + \boldsymbol{\epsilon}_2(n) \\ \boldsymbol{\epsilon}_1(n) &= \hat{\boldsymbol{w}}(n) - E[\hat{\boldsymbol{w}}(n)] \\ \boldsymbol{\epsilon}_2(n) &= E[\hat{\boldsymbol{w}}(n)] - \boldsymbol{w}_o(n)\end{aligned}$$

     donde $\boldsymbol{\epsilon}_1(n)$ y $\boldsymbol{\epsilon}_2(n)$ son las componentes debidas al ruido y retardo, respectivamente.
     En base a las suposiciones anteriores $E[\boldsymbol{\epsilon}_1^H(n)\boldsymbol{\epsilon}_2(n)] = 0$, de forma que la desviación cuadrática media también se puede expresar como la suma de dos componentes

$$\mathcal{D}(n) = \mathcal{D}_1(n) + \mathcal{D}_2(n) = E\left[\|\boldsymbol{\epsilon}_1(n)\|^2\right] + E\left[\|\boldsymbol{\epsilon}_2(n)\|^2\right]$$

2. **Desajuste**. Otra figura de mérito utilizada comunmente para analizar la capacidad de tracking de un filtro adaptivo es el **desajuste**, definido por

$$\mathcal{M}(n) = \frac{J_{ex}(n)}{\sigma_\nu^2}$$

donde de nuevo se supone que la cantidad de iteraciones es suficientemente grande como para considerar el período transitorio finalizado. Teniendo en cuenta las suposiciones de independencia, $J_{ex}(n) = tr[\boldsymbol{R}\,\boldsymbol{K}(n)]$, tal que

$$\mathcal{M}(n) = \frac{tr[\boldsymbol{R}\,\boldsymbol{K}(n)]}{\boldsymbol{\nu}^2}$$

Para un buen desempeño de seguimiento es posible concluir que el desajuste $\mathcal{M}(n)$ debería ser pequeño comparado con la unidad.

De la misma forma que para $\mathcal{D}(n)$, es posible expresar $J_{ex}(n)$ en términos de dos compoenentes, una asociada al ruido de estimación y la otra asociada al retardo de estimación. De esta forma es posible expresar el desajuste como

$$\mathcal{M}(n) = \mathcal{M}_1(n) + \mathcal{M}_2(n)$$

correspondientes al desajuste por ruido y al desajuste por retardo respectivamente.

10.4 Desempeño de seguimiento del algoritmo LMS

- En el contexto no estacionario descripto es posible considerar que el filtro adaptivo se implementa usando un algoritmo LMS. En ese caso

$$\hat{\boldsymbol{w}}(n+1) = \hat{\boldsymbol{w}}(n) + \mu \boldsymbol{u}(n) e^*(n)$$

donde teniendo en cuenta la señal de error con el modelo de regresión lineal se obtiene

$$\hat{\boldsymbol{w}}(n+1) = [\boldsymbol{I} - \mu \boldsymbol{u}(n)\boldsymbol{u}^H(n)]\hat{\boldsymbol{w}}(n) + \boldsymbol{u}(n)\boldsymbol{u}^H(n)\boldsymbol{w}_o(n) + \mu \boldsymbol{u}(n)\nu^*(n)$$

o equivalentemente

$$\begin{aligned} \overline{\boldsymbol{\epsilon}}(n+1) &= \hat{\boldsymbol{w}}(n+1) - \boldsymbol{w}_o(n+1) \\ &= [\boldsymbol{I} - \mu \boldsymbol{u}(n)\boldsymbol{u}^H(n)]\overline{\boldsymbol{\epsilon}}(n) + (1-a)\boldsymbol{w}_o(n) + \mu \boldsymbol{u}(n)\nu^*(n) - \boldsymbol{\omega}(n) \end{aligned}$$

- Con la suposición de a es muy próximo a 1 es posible simplificarla de la siguiente forma

$$\overline{\boldsymbol{\epsilon}}(n+1) = [\boldsymbol{I} - \mu \boldsymbol{u}(n)\boldsymbol{u}^H(n)]\overline{\boldsymbol{\epsilon}}(n) + \mu \boldsymbol{u}(n)\nu^*(n) - \boldsymbol{\omega}(n)$$

- Con μ suficientemente pequeño es posible resolver utilizando el **método de promediación directa**,

$$\boldsymbol{\epsilon}(n+1) = [\boldsymbol{I} - \mu \boldsymbol{R}]\boldsymbol{\epsilon}(n) + \mu \boldsymbol{u}(n)\nu^*(n) - \boldsymbol{\omega}(n) \quad (147)$$

- Para evaluar la matriz correlación de $\boldsymbol{\epsilon}(n+1)$ es posible invocar las suposiciones de independencia descriptas anteriormente. En ese caso

$$\begin{aligned} \boldsymbol{K}(n+1) &= E[\boldsymbol{\epsilon}(n+1)\boldsymbol{\epsilon}^H(n+1)] \\ &= [\boldsymbol{I} - \mu \boldsymbol{R}]\boldsymbol{K}(n)[\boldsymbol{I} - \mu \boldsymbol{R}] + \mu \sigma_\nu^2 \boldsymbol{R} - \boldsymbol{Q} \end{aligned}$$

- Para obtener la **solución de estado estacionario**, para n suficientemente grande, consideramos $\boldsymbol{K}(n+1) = \boldsymbol{K}(n)$. Además, suponiendo que μ es pequeño como para eliminar el término $\mu^2 \boldsymbol{R}\boldsymbol{K}(n)\boldsymbol{R}$ en comparación con $\boldsymbol{I}$, es posible obtener la aproximación

$$\boldsymbol{R}\boldsymbol{K}(n) + \boldsymbol{K}(n)\boldsymbol{R} \cong \mu\sigma_\nu^2\boldsymbol{R} + \frac{1}{\mu}\boldsymbol{Q} \tag{148}$$

Desviación media cuadrática del algoritmo LMS

- Premultiplicando (148) por $\boldsymbol{R}^{-1}$ y aplicamos la traza del resultado,

$$tr[\boldsymbol{K}(n)] + tr[\boldsymbol{R}^{-1}\boldsymbol{K}(n)\boldsymbol{R}] \cong \mu M\sigma_\nu^2 + \frac{1}{\mu}tr[\boldsymbol{R}^{-1}\boldsymbol{Q}]$$

- Como $\boldsymbol{K}(n)$ y $\boldsymbol{R}$ tienen las mismas dimensiones: $tr[\boldsymbol{R}^{-1}\boldsymbol{K}(n)\boldsymbol{R}] = tr[\boldsymbol{K}(n)\boldsymbol{R}\boldsymbol{R}^{-1}] = tr[\boldsymbol{K}(n)]$. Entonces, usando la definición,

$$\mathcal{D}(n) \cong \frac{\mu}{2}M\sigma_\nu^2 + \frac{1}{2\mu}tr[\boldsymbol{R}^{-1}\boldsymbol{Q}]$$

donde el primer término es la varianza de estimación debido al ruido de medición $\nu(n)$ (varia linealmente con μ). El segundo término es la varianza de retardo debido al proceso de ruido $\boldsymbol{\omega}(n)$ (varia en forma inversa con μ).

- Si μ_{opt} es el valor óptimo del factor de convergencia para el cual la desviación media cuadrática alcanza el mínimo $\mathcal{D}_{min}$, es simple verificar que

$$\mu_{opt} \cong \frac{1}{\sigma_\nu\sqrt{M}}(tr[\boldsymbol{R}^{-1}\boldsymbol{Q}])^{1/2} \text{ y } \mathcal{D}_{min} \cong \sigma_\nu\sqrt{M}(tr[\boldsymbol{R}^{-1}\boldsymbol{Q}])^{1/2}$$

Desajuste del algoritmo LMS

- En base a la traza de (148) es posible escribir

$$tr[\boldsymbol{R}\boldsymbol{K}(n)] + tr[\boldsymbol{K}(n)\boldsymbol{R}] \cong \mu\sigma_\nu^2 tr[\boldsymbol{R}] + \frac{1}{\mu} tr[\boldsymbol{Q}]$$

- Teniendo en cuenta que la traza de $\boldsymbol{R}\boldsymbol{K}(n)$ y $\boldsymbol{K}(n)\boldsymbol{R}$ es igual, entonces usando la definición,

$$\mathcal{M}(n) \cong \frac{\mu}{2} tr[\boldsymbol{R}] + \frac{1}{2\mu\sigma_\nu^2} tr[\boldsymbol{Q}]$$

  donde, en este caso, el primer término es el desajuste de ruido $\nu(n)$, y tiene una forma similar que para un ambiente estacionario. El segundo término es el desajuste de retardo causado por $\boldsymbol{\omega}(n)$, el cual es representativo del contexto no estacionario.

- El desajuste de ruido varia linealmente con μ mientras que el desajuste de retardo varía inversamente con μ. De esta forma, optimizando μ para un desajuste mínimo se obtiene

$$\begin{aligned}
\mu_{opt} &\cong \frac{1}{\sigma_\nu}\left(\frac{tr[\boldsymbol{Q}]}{tr[\boldsymbol{R}]}\right)^{1/2} \\
\mathcal{M}_{min} &\cong \frac{1}{\sigma_\nu}\left(tr[\boldsymbol{Q}]\, tr[\boldsymbol{R}]\right)^{1/2}
\end{aligned}$$

10.5 Desempeño de seguimiento del algoritmo RLS

Teniendo en cuenta el mismo modelo de no estacionaridad que en el caso anterior, es posible tener en cuenta ahora el algoritmo RLS aplicado a un filtro adaptivo transversal, dado por

$$\hat{\boldsymbol{w}}(n) = \hat{\boldsymbol{w}}(n-1) + \boldsymbol{\Phi}^{-1}(n)\boldsymbol{u}(n)\xi^*(n)$$

donde $\boldsymbol{\Phi}(n) = \sum_{i=1}^{n} \lambda^{n-i}\boldsymbol{u}(i)\boldsymbol{u}^H(i)$ y $\xi(n)$ es el error de estimación a priori dado por

$$\xi(n) = d(n) = \hat{\boldsymbol{w}}(n-1)\boldsymbol{u}(n)$$

Entonces, considerando que

$$\boldsymbol{w}_o(n) = a\boldsymbol{w}_o(n-1) + \boldsymbol{\omega}(n) = d(n) - \boldsymbol{w}_o^H(n-1)\boldsymbol{u}(n) + \nu(n)$$

es posible escribir la ecuación de adaptación del error en el vector de coeficientes para el algoritmo RLS como

$$\overline{\boldsymbol{\epsilon}}(n) = [\boldsymbol{I} - \boldsymbol{\Phi}^{-1}(n)\boldsymbol{u}(n)\boldsymbol{u}^H(n)]\overline{\boldsymbol{\epsilon}}(n-1) + \boldsymbol{\Phi}^{-1}(n)\boldsymbol{u}(n)\nu^*(n) + (1-a)\boldsymbol{w}_o(n-1) - \boldsymbol{\omega}(n)$$

Realizando una simplificación similar que en el caso del algoritmo LMS, suponiendo a próximo a 1, esta ecuación se reduce a

$$\overline{\boldsymbol{\epsilon}}(n) = [\boldsymbol{I} - \boldsymbol{\Phi}^{-1}(n)\boldsymbol{u}(n)\boldsymbol{u}^H(n)]\overline{\boldsymbol{\epsilon}}(n-1) + \boldsymbol{\Phi}^{-1}(n)\boldsymbol{u}(n)\nu^*(n) - \boldsymbol{\omega}(n) \quad (149)$$

En este punto, es útil obtener una aproximación de la matriz $\boldsymbol{\Phi}^{-1}(n)$ para permitir un manejo tratable del estudio en cuestión.

$$E[\mathbf{\Phi}(n)] = \sum_{i=1}^{n} \lambda^{n-i} E[\boldsymbol{u}(i)\boldsymbol{u}^H(i)] = \sum_{i=1}^{n} \lambda^{n-i} \boldsymbol{R}$$
$$= \boldsymbol{R}(1+\lambda+\lambda^2+\cdots+\lambda^{n-1}) = \frac{\boldsymbol{R}}{1-\lambda} \qquad \text{para } n \text{ grande}$$

lo que permite (en una aproximación no trivial)

$$\mathbf{\Phi}(n) \cong \frac{\boldsymbol{R}}{1-\lambda} \text{ ó } \mathbf{\Phi}^{-1}(n) \cong (1-\lambda)\boldsymbol{R}^{-1}, \qquad \text{para } n \text{ grande}$$

Volviendo a (149) y usando la aproximación anterior

$$\overline{\boldsymbol{\epsilon}}(n) \cong [\boldsymbol{I} - (1-\lambda)\boldsymbol{R}\boldsymbol{u}(n)\boldsymbol{u}^H(n)]\overline{\boldsymbol{\epsilon}}(n-1) + (1-\lambda)\boldsymbol{R}^{-1}\boldsymbol{u}(n)\nu^*(n) - \boldsymbol{\omega}(n)$$

Típicamente λ es próximo a la unidad, de forma que $1-\lambda$ tiene un valor pequeño. De esta forma, invocando el método de promediación directa

$$\boldsymbol{\epsilon}(n) \cong \lambda\boldsymbol{\epsilon}(n-1) + (1-\lambda)\boldsymbol{R}^{-1}\boldsymbol{u}(n)\nu^*(n) - \boldsymbol{\omega}(n) \tag{150}$$

Evaluando ahora la matriz correlación de $\boldsymbol{\epsilon}(n)$

$$\boldsymbol{K}(n) \cong \lambda^2\boldsymbol{K}(n-1) + (1-\lambda)^2\sigma_\nu^2\boldsymbol{R}^{-1} + \boldsymbol{Q}$$

Para obtener una **solución de estado estacionario**, para n suficientemente grande, hacemos $\boldsymbol{K}(n=1) = \boldsymbol{K}(n)$, tal que

$$(1-\lambda^2)\boldsymbol{K}(n) \cong (1-\lambda)^2\sigma_\nu^2\boldsymbol{R}^{-1} + \boldsymbol{Q}$$

y usando para $\lambda \cong 1$ que $1-\lambda^2 = (1+\lambda)(1-\lambda) \cong 2(1-\lambda)$, es posible llegar finalmente a

$$\boldsymbol{K}(n) \cong \frac{(1-\lambda)}{2}\sigma_\nu^2\boldsymbol{R}^{-1} + \frac{1}{2(1-\lambda)}\boldsymbol{Q} \tag{151}$$

Desviación media cuadrática del algoritmo RLS

Aplicando la ecuación anterior a definición de desviación media cuadrática se obtiene

$$\mathcal{D}(n) \cong \frac{(1-\lambda)}{2}\sigma_\nu^2\, tr[\boldsymbol{R}^{-1}] + \frac{1}{2(1-\lambda)}\, tr[\boldsymbol{Q}]$$

tal que, el primer término esla varianza de estimación debida al ruido de medición $\nu(n)$ y el segundo término es la varianza de retardo debida al proceso de ruido $\boldsymbol{\omega}(n)$. Estas dos contribuciones varian en proporción a $(1-\lambda)$ y $(1-\lambda)^{-1}$, respectivamente. El valor óptimo del factor de ponderación y la desviación media cuadrática mínima para el algoritmo RLS están dadas por

$$\lambda_{opt} \cong 1 - \frac{1}{\sigma_\nu}\left(\frac{tr[\boldsymbol{Q}]}{tr[\boldsymbol{R}^{-1}]}\right)^{1/2} \qquad \mathcal{D}_{min} \cong \sigma_\nu (tr[\boldsymbol{R}^{-1}]tr[\boldsymbol{Q}])^{1/2}$$

Desajuste del algoritmo RLS

Multiplicando ambos lados de (151) por la matriz correlación $\boldsymbol{R}$, se tiene que

$$\boldsymbol{R}\boldsymbol{K}(n) \cong \frac{(1-\lambda)}{2}\sigma_\nu^2\boldsymbol{I} + \frac{1}{2(1-\lambda)}\boldsymbol{R}\boldsymbol{Q}$$

y en base a la traza de esta ecuación

$$tr[\boldsymbol{R}\boldsymbol{K}(n)] \cong \frac{(1-\lambda)}{2}\sigma_\nu^2 M + \frac{1}{2(1-\lambda)}\, tr[\boldsymbol{R}\boldsymbol{Q}]$$

tal que aplicando la definición, el desajuste asociado al algoritmo RLS será

$$\mathcal{M}(n) \cong \frac{1-\lambda}{2}M + \frac{1}{2(1-\lambda)\sigma_\nu^2}\, tr[\boldsymbol{R}\boldsymbol{Q}]$$

El primer término representa el desajuste del algoritmo RLS debido a $\nu(n)$, que varía linealmente con $1-\lambda$ y depende también de M.

El segundo término representa el desajuste de retardo del algoritmo RLS debido al proceso de ruido $\boldsymbol{\omega}(n)$. Este término varía inversamente con $1-\lambda$.

El valor óptimo del factor de ponderación exponencial y el valor mínimo del desajuste están dados por

$$\begin{aligned} \lambda_{opt} &\cong 1 - \frac{1}{\sigma_\nu}\left(\frac{1}{M}\, tr[\boldsymbol{R}\,\boldsymbol{Q}]\right)^{1/2} \\ \mathcal{M}_{min} &\cong \frac{1}{\sigma_\nu}(M\;\; tr[\boldsymbol{R}\,\boldsymbol{Q}])^{1/2} \end{aligned}$$

10.6 Comparación del desempeño de seguimiento del LMS y RLS

- Teniendo en cuenta que los algoritmos LMS y RLS se formulan de forma totalmente diferente, no es sorprendente que exiban propiedades de convergencia diferentes y también propiedades de seguimiento diferentes.
- En el algoritmo RLS el vector de entrada está premultiplicado por $\boldsymbol{R}^{-1}$, lo que indica la diferencia fundamental entre los algoritmos RLS y LMS.
- $(1-\lambda)$ en el algoritmo RLS juega el papel de μ en el algoritmo LMS. Teniendo en cuenta esta analogía, pero además que λ es adimensional en tanto que μ tiene dimensiones de potencia se harán de aqui en más las siguientes consideraciones
  - Para el algoritmo LMS definimos el **factor de convergencia normalizado**, $\nu = \mu\sigma_u^2$, donde σ_u^2 es la varianza de la señal de entrada $u(n)$.

– Para el algoritmo RLS definimos la **velocidad de convergencia**, $\beta = 1 - \lambda$.

- Con el objetivo de realizar la comparación consideraremos la relación entre la desviación media cuadrática obtenida para ambos métodos y también la relación entre sus desajustes. Específicamente es posible escribir

$$\frac{\mathcal{D}_{min}^{LMS}}{\mathcal{D}_{min}^{RLS}} \cong \left(\frac{M \;\; tr[\boldsymbol{R}^{-1}\boldsymbol{Q}]}{tr[\boldsymbol{R}^{-1}]\; tr[\boldsymbol{Q}]} \right)^{1/2}$$

$$\frac{\mathcal{M}_{min}^{LMS}}{\mathcal{M}_{min}^{RLS}} \cong \left(\frac{tr[\boldsymbol{R}]\; tr[\boldsymbol{Q}]}{M \;\; tr[\boldsymbol{R}\,\boldsymbol{Q}]} \right)^{1/2}$$

- Claramente, la comparación entre los algoritmos LMS y RLS sobre esta base es función de $\boldsymbol{R}$ y $\boldsymbol{Q}$, de forma que será importante considerar algunos casos particulares.

Ejemplo 1: $\boldsymbol{Q} = \sigma_\omega^2 \boldsymbol{I}$.

Consideremos el caso de un proceso de ruido $\boldsymbol{\omega}(n)$ del modelo de primer orden de Markov originado de una fuente de ruido blanco de media cero y varianza σ_ω^2. Su matriz correlación será

$$\boldsymbol{Q} = \sigma_\omega^2 \boldsymbol{I}$$

Luego usando las relaciones anteriores es posible concluir que

$$\begin{aligned} \mathcal{D}_{min}^{LMS} &\cong \mathcal{D}_{min}^{RLS} \\ \mathcal{M}_{min}^{LMS} &\cong \mathcal{M}_{min}^{RLS} \end{aligned}$$

tal que el comportamiento para ese tipo de no estacionaridad es similar en ambos algoritmos.

Ejemplo 2: $\boldsymbol{Q} = c_1 \boldsymbol{R}$.

En este ejemplo consideraremos que la matriz $\boldsymbol{Q}$ es proporcional, a través de una constante c_1, a la matriz correlación de la entrada. Se introduce la constante c_1 con dos objetivos:

1. Para tener en cuenta que el proceso de ruido $\boldsymbol{\omega}(n)$ y el vector de entradas $\boldsymbol{u}(n)$ se miden en unidades diferentes.
2. Para asegurar que el μ óptimo para el algoritmo LMS y el λ óptimo para el algoritmo RLS tengan valores equivalentes.

Los resultados de la comparación, así como los del ejemplo siguiente se muestran en la tabla a continuación

	$\boldsymbol{Q} = c_1 \boldsymbol{R}$	$\boldsymbol{Q} = c_2 \boldsymbol{R}^{-1}$
$\frac{\mathcal{D}_{min}^{LMS}}{\mathcal{D}_{min}^{RLS}}$	$\frac{M}{(tr[\boldsymbol{R}^{-1}]\, tr[\boldsymbol{R}])^{1/2}}$	$\frac{(M\, tr[\boldsymbol{R}^{-2}])^{1/2}}{tr[\boldsymbol{R}^{-1}]}$
$\frac{\mathcal{M}_{min}^{LMS}}{\mathcal{M}_{min}^{RLS}}$	$\frac{tr[\boldsymbol{R}]}{(M\, tr[\boldsymbol{R}^{2}])^{1/2}}$	$\frac{1}{M}\left(tr\left[\boldsymbol{R}\right]\, tr\left[\boldsymbol{R}^{-1}\right]\right)^{1/2}$

Ejemplo 3: $\boldsymbol{Q} = c_2\boldsymbol{R}^{-1}$.

Consideraremos que existe también una relación directa entre $\boldsymbol{Q}$ y $\boldsymbol{R}^{-1}$ a través de una constante c_2 de proporcionalidad, utilizada por las mismas razones que en el ejemplo anterior. Los resultados de la comparación se incluyeron en la tabla anterior.

Los resultados mostrados en esa tabla muestran **simetría recíproca**, cuyo significado específico se discutirá más adelante.

Los resultados de los términos en la antidiagonal de esa tabla conducen a la aplicación de la **desigualdad de Schartz**, lo cual implica las siguientes cotas

$$
\begin{array}{rcll}
\mathcal{M}_{min}^{LMS} & \leq & \mathcal{M}_{min}^{RLS} & \text{para } \boldsymbol{Q} = c_1\boldsymbol{R} \\
\mathcal{D}_{min}^{RLS} & \leq & \mathcal{D}_{min}^{LMS} & \text{para } \boldsymbol{Q} = c_2\boldsymbol{R}^{-1}
\end{array}
$$

lo que conduce a las siguientes proposiciones

1. Para $\boldsymbol{Q} = c_1\boldsymbol{R}$, el algoritmo LMS tiene mejor desempeño que el RLS, ya que produce el menor desajuste mínimo.

2. Para $\boldsymbol{Q} = c_2\boldsymbol{R}^{-1}$, el algoritmo RLS tiene mejor desempeño que el LMS ya que produce la menor desviación media cuadrática.

No es posible ser tan conclusivos en relación a los términos de la diagonal de la tabla, aunque en función de la simetría recíproca comentada es posible establecer que si, para $\boldsymbol{Q} = c_1\boldsymbol{R}$, $\mathcal{D}_{min}^{LMS} \leq \mathcal{D}_{min}^{RLS}$, entonces es cierto que, para $\boldsymbol{Q} = c_2\boldsymbol{R}^{-1}$, $\mathcal{M}_{min}^{RLS} \leq \mathcal{M}_{min}^{LMS}$.

Para ilustrar la validez de esta proposición consideraremos el caso especial de un filtro adaptivo de orden $M = 2$, tal que

$$
\boldsymbol{R} = \begin{bmatrix} r_{11} & r_{21} \\ r_{21} & r_{22} \end{bmatrix}
$$

Los valores de comparación obtenidos en la tabla anterior se incluyen en la tabla siguiente para este caso particular

	$\boldsymbol{Q} = c_1\boldsymbol{R}$	$\boldsymbol{Q} = c_2\boldsymbol{R}^{-1}$
$\frac{\mathcal{D}_{min}^{LMS}}{\mathcal{D}_{min}^{RLS}}$	$\frac{2\sqrt{r_{11}r_{22}-r_{12}^2}}{r_{11}+r_{22}}$	$\frac{\sqrt{2(r_{11}^2+2r_{21}^2+r_{22}^2}}{r_{11}+r_{22}}$
$\frac{\mathcal{M}_{min}^{LMS}}{\mathcal{M}_{min}^{RLS}}$	$\frac{r_{11}+r_{22}}{\sqrt{2(r_{11}^2+2r_{21}^2+r_{22}^2}}$	$\frac{r_{11}+r_{22}}{2\sqrt{r_{11}r_{22}-r_{12}^2}}$

Teniendo en cuenta que $(r_{11} - r_{22})^2 + (2r_{21})^2 \geq 0$, esta tabla permite establecer las siguientes conclusiones

1. Para $\boldsymbol{Q} = c_1\boldsymbol{R}$, el algoritmo LMS se desempeña mejor que el algoritmo RLS de forma que produce los menores valores de $\mathcal{D}_{min}$ y $\mathcal{M}_{min}$.

2. Para $\boldsymbol{Q} = c_2\boldsymbol{R}^{-1}$, el algoritmo RLS se desempeña mejor que el algoritmo LMS de forma que produce los menores valores de $\mathcal{D}_{min}$ y $\mathcal{M}_{min}$.

Estos dos ultimos ejemplos muestran que los algoritmos LMS y RLS no tienen un completo domínio de buen desempeño en seguimiento, si no que más bien uno u otro algoritmo será preferido para el seguimiento en un ambiente no estacionario dependiendo del contexto prevaleciente.