

Gregori Vázquez, Antoni Gasull, Miguel A. Lagunas.  
E.T.S.I. Telecomunicación de Barcelona.  
C/ Jordi Girona Salgado, s/n.08034 Barcelona - España

ABSTRACT

This work deals with the use of previous or colateral information to improve the behaviour of adaptive algorithms. The study is made on gradient-based methods due to the relatively simple and good performances that they use to exhibit.

This paper shows that the complete knowledge of the data at the input of the adaptive filter (and in consequence of its autocorrelation matrix and its inverse) can be used to modify the classic L.M.S. algorithm leading to better expressions for the gradient and a new adaptive 'step size'.

Finally, the description is completed with the comparison between the variation ranges and a natural generalization of 'step size' parameter is obtained.

1. INTRODUCTION

Consideremos el problema clásico del filtrado de Wiener. Podrían adoptarse dos alternativas para la resolución del problema en el supuesto de un conocimiento exacto de la matriz de correlación de los datos, esto es, hacer uso directamente de la ecuación óptima de Wiener, que garantiza un comportamiento óptimo en un bloque de datos, o considerar que una aproximación adaptativa puede ser mejor en situaciones reales en las que se suelen imponer aritméticas finitas y en que las condiciones pueden ser no estacionarias.

Si se adopta la segunda posibilidad, la pregunta que se plantea de forma inmediata es como reflejar o hacer uso de toda la información previa o colateral disponible en la mejora del comportamiento del algoritmo adaptativo.

Desde nuestro punto de vista, habrían dos posibles maneras de reflejar estas informaciones adicionales en un esquema adaptativo con un objetivo cuadrático:

a) Utilizarla para una mejor estimación de los parámetros o funciones asociadas involucradas en el algoritmo adaptativo.

b) Incluir la información colateral como restricciones o bien en el mismo proceso de minimización.

Este trabajo se dirige en ambos sentidos. El primer punto es el más inmediato y será utilizado para la mejora de la estimación del gradiente. Por otro lado, el punto dos no es tan claro como el anterior, y, en general, tratará de modificar la estructura del esquema adaptativo para que se satisfagan las restricciones impuestas o la minimización del criterio de error. Por lo tanto, aunque parece una posibilidad atractiva, el diseñador deberá prestar atención al hecho de que muy frecuentemente la nueva estructura obtenida requerirá de un elevado costo computacional. En nuestro caso, únicamente se propone una adaptación

del 'step-size', pero manteniendo en todo momento el esquema adaptativo usual.

2. FILTRADO LINEAL OPTIMO (m.m.s.e)  
[1], [2], [3], [4].

Se pretende minimizar el error cuadrático medio entre una señal de referencia dada  $y(n)$  y la salida  $\hat{y}(n)$  de un filtro F.I.R. definido por el vector de coeficientes  $\underline{W}$ . Así pues, para una señal de entrada  $x(n)$  se dispone de un vector de datos  $\underline{X}_n$ , y de la estimación  $y(n) = \underline{X}_n^T \underline{W}$ .

$$\underline{X}_n^T = (x(n), x(n-1), \dots, x(n-Q+1))$$

$$\underline{W}^T = (w(0), w(1), \dots, w(Q-1)) ;$$

Q: ORDEN DEL FILTRO F.I.R.

El error cuadrático medio vendrá dado por:

$$\begin{aligned} \epsilon^2 &= E((y(n) - \hat{y}(n))^2) = \\ &E((y(n) - \underline{X}_n^T \underline{W})^2) \end{aligned} \quad (2.1)$$

La minimización de esta expresión lleva a la conocida solución óptima de Wiener:

$$\underline{W}_{opt} = R_{xx}^{-1} \underline{P} \quad (2.2)$$

donde:  $R_{xx} = E(\underline{X}_n \underline{X}_n^T)$  es la matriz  $Q \times Q$  de autocorrelación de la secuencia de datos  $x(n)$  a la entrada y  $\underline{P} = E(y(n) \underline{X}_n)$  es un vector de correlaciones cruzadas entre las muestras  $\underline{X}_n$  y la referencia  $y(n)$ .

En general, sin información adicional sobre  $R_{xx}$  y  $\underline{P}$ , deberá recurrirse a estimaciones tanto para  $R_{xx}$  como para  $\underline{P}$ . El uso de estas estimaciones suele hacerse en expresiones obtenidas a partir de una aproximación adaptativa. La más frecuente de ellas consiste en el uso del gradiente del error cuadrático medio.

Si se dispone de toda la información el esquema adaptativo es de la forma:

$$\underline{W}_{n+1} = \underline{W}_n - \mu_n \underline{V}_n \quad (2.3)$$

con:

$$\underline{V}_n = R_{xx} \underline{W}_n - \underline{P} \quad (\text{estimación exacta del gradiente del error}). \quad (2.4)$$

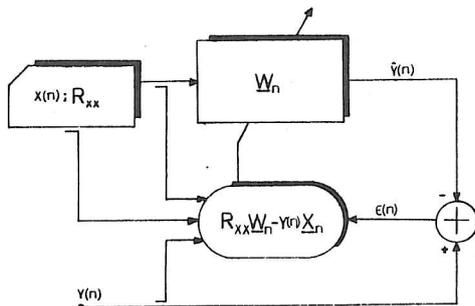
$\mu_n$  es el "step size".

Pero, como se ha dicho anteriormente, en general,  $R_{xx}$  y  $\underline{P}$  se conocen sólo parcialmente. Así pues, el gradiente exacto se sustituye por una estimación ( $\hat{\underline{V}}_n$ ) de este:

$$\underline{W}_{n+1} = \underline{W}_n - \mu_n \hat{\underline{V}}_n \quad (2.5)$$

En primer lugar, Widrow propuso como gradiente a  $\hat{\underline{V}}_n = (y(n) - \hat{y}(n)) \cdot \underline{X}_n$  (denominado gradiente instantáneo). Más tarde L.J. Griffiths introdujo como gradiente a  $\hat{\underline{V}}_n = y(n) \underline{X}_n - \underline{P}$  en el contexto de los arrays adaptativos.

Como se observa, la aproximación de Widrow es una estimación completamente instantánea lo cual hace que sea computacionalmente la más eficiente.



### 3. DESCRIPCIÓN DEL NUEVO ALGORITMO

En algunos casos, la secuencia de datos  $x(n)$  es perfectamente conocida (sonar activo, medida de tiempos de retardo, ...) y en consecuencia su matriz de autocorrelación puede considerarse también como dato.

En estas condiciones puede parecer que la mejor alternativa sea directamente la ecuación de Wiener para las estimaciones de los coeficientes. Es decir:

$$\underline{W}_n = R_{xx}^{-1} \hat{\underline{P}}_n \quad (3.1)$$

Siendo  $\hat{\underline{P}}_n$  una estimación instantánea del vector de correlaciones cruzadas, por ejemplo:

$$\hat{\underline{P}}_n = y(n) \cdot \underline{X}_n \quad (3.2)$$

Pero es claro que este método hace poco uso de la información pasada. De entre muchas alternativas, una posible variante puede ser la versión alisada siguiente:

$$\underline{W}_n = \alpha \cdot \underline{W}_{n-1} + (1-\alpha) R_{xx}^{-1} \hat{\underline{P}}_n \quad (3.3)$$

donde  $\alpha$  es una constante tal que  $0 < \alpha < 1$ . Sin embargo, en la práctica, el cálculo de las expresiones suele hacerse en aritméticas finitas y entre ellas, fundamentalmente, en coma fija. Por otro lado, la elección

del parámetro  $\alpha$  es una cuestión difícil ya que afectará decisivamente en el tiempo de convergencia, y no es clara.

Así pues, aún en este caso parece más que razonable que se haga uso de las expresiones basadas en el gradiente del error.

Bajo las condiciones del problema, nuestra propuesta para la estimación del gradiente será:

$$\hat{\underline{V}}_n = R_{xx} \underline{X}_n - y(n) \cdot \underline{X}_n \quad (3.4)$$

frente a las ya enumeradas, dadas por Widrow y por Griffiths.

Otro aspecto importante es la actualización del "step size" ( $\mu$ ) en cada iteración. Powell argumenta que una elección adecuada es la que minimiza el e.c.m. en cada una de las iteraciones. La expresión que se obtiene es:

$$\mu_n = - \frac{\hat{\underline{V}}_n^T \hat{\underline{V}}_n}{\hat{\underline{V}}_n^T R_{xx} \hat{\underline{V}}_n} \quad (3.5)$$

Nuestra propuesta difiere de la anterior en el objetivo a minimizar. Dado que se conoce  $R_{xx}$  de forma exacta, se propone que en cada etapa se minimice el error asociado a la envolvente visual de los coeficientes dado por la expresión:

$$\epsilon_n^2 = E((\underline{W}_n - \underline{W}_{opt})^T (\underline{W}_n - \underline{W}_{opt})) \quad (3.6)$$

Se comprueba que en este caso viene dada por:

$$\mu'_n = - \frac{\hat{\underline{V}}_n^T R_{xx}^{-1} \hat{\underline{V}}_n}{\hat{\underline{V}}_n^T \hat{\underline{V}}_n} \quad (3.7)$$

A partir de (3.5) y (3.7) se sigue la expresión general dada por:

$$\mu_n^k = \frac{\hat{\underline{V}}_n^T R_{xx}^{-k} \hat{\underline{V}}_n}{\hat{\underline{V}}_n^T \hat{\underline{V}}_n} \quad (3.8)$$

Este conjunto completo de posibles valores lleva a (3.5) y (3.7) para  $k=0$  y  $k=-1$ , respectivamente, y presenta para cualquier valor de  $k$  el mismo margen de variación dado por:

$$1/\lambda_{m\acute{a}x} < \mu_n^k < 1/\lambda_{m\acute{i}n}$$

con  $\lambda_{m\acute{a}x}$  y  $\lambda_{m\acute{i}n}$  los autovalores máximo y mínimo de la matriz de autocorrelación  $R_{xx}$ .

### 4. CONCLUSIONES

El resultado final es un algoritmo con un comportamiento excelente respecto a los otros métodos basados en el uso del gradiente, así como con una estructura fácilmente implementable en V.L.S.I. mediante cualquier técnica sistólica.

REFERENCIAS

- | 1 | R.A. Monzingo and T.W. Miller. "Gradient-Based Algorithms". John Wiley & Sons, Inc. 1980.
- | 2 | L.J. Griffiths. "A simple Adaptive Algorithm for Real-Time Processing in Antenna Arrays". Proc. of the IEEE, Vol. 57, No. 10, Oct. 1969.
- | 3 | B. Widrow et al. "Adaptive Antenna Systems". Proc. IEEE, Vol. 55, pp.2143-2159, Dec. 1967.
- | 4 | B. Widrow et al. "A Comparison of Adaptive Algorithms Based on the Methods of Steepest Descent and Random Search". IEEE Trans. Antennas and Prop., Vol. AP-24, No. 5, Sep. 1976.
- | 5 | M. Bertran & C. Nadeu. "On the Inclusion of Prior Information in the Window Method of Spectral Estimation". Proc. of MELECON'85 Vol. II, pp.59-61. Oct. 1985. Madrid.
- | 6 | R.A. Monzingo and T.W. Miller. "Recursive Methods for Adaptive Array Processing". John Wiley & Sons, Inc. 1980.
- | 7 | J.S. Walther. "A Unified Algorithm for Elementary Functions". 1971, Spring JCC, pp.379-385.
- | 8 | G. Vazquez et al. "Known Input Power Spectrum in Adaptive L.M.S. and A.G. Algorithms". EUSIPCO 86.