Australian Journal of Basic and Applied Sciences, 5(3): 69-74, 2011
ISSN 1991-8178

Least Squares Problem for Adaptive Filtering

[1]Noor Atinah Ahmad, [2]Friday Zinzendoff Okwonu

[1]School of Mathematical Sciences Universiti Sains Malaysia Pulau-Pinang, Malaysia
[2]School of Mathematical Sciences Universiti Sains Malaysia Pulau-PinangCity, Malaysia

Abstract: We discus and demonstrate ways of deriving solutions to least squares problem in adaptive filtering without forming the conventional triangular system of equation, thus, generalized inverse techniques is used to obtain the tap weight coefficient vector when the autocorrelation matrix is singular, this approach is further extended using matrix inversion lemma to derive normal equation incorporating autocorrelation matrix and cross correlation vector.

Key words: Least squares problem;Adaptive filtering; Autocorrelation matrix;Crosscorrelation matrix.

INTRODUCTION

The methods of least squares has been the standard procedure for analyzing data from the beginning of the 1800s [12]. One important such example is when Gauss successfully predicted the orbit of the asteroid Ceres in 1801.Many problems in signal processing can be formulated as a least squares problem

$$\min_{\mathbf{w}} \left\| \mathbf{Xw} - \mathbf{s} \right\|_2 \tag{1}$$

$\mathbf{X} \in \Re^{m\times n}$, $\mathbf{s} \in \Re^{m\times 1}$ and $\| \ \|_2$ is the Euclidean vector norm and w is the least squares solution of the form

$$Xw=s \tag{2}$$

Then the residual error vectors denoted and define as

$$r=s-Xw \tag{3}$$

Least squares solution minimizes the sum of squared residual. If rank of the data matrix X is less than n ,then the solution to (1) is not unique Jian-Shu, &W.Xue-gang, (2006), R(X) is the rang space of data matrix. The basic computational tool to solve (1) introduced by Gauss was to form and solve normal equation

$$\mathbf{X}^T\mathbf{Xw} = \mathbf{X}^T\mathbf{s}$$

The method of least squares is used to approximately solve an over determined system, i.e., systems of equation in which there are more equations than unknown. The method was first described by Carl Friedrich Gauss around 1794[11]. Least squares corresponds to the maximum likelihood criterion if the experimental errors have a normal distribution and can be derived as a method of moment estimator. The least squares method grew out of the field of Astronomy and Geodesy as scientist and Mathematician sought to provide solutions to the challenges of navigating the earth's ocean during the[11] age of exploration. The accurate description of the behavior of the celestial bodies was a key to enabling ships to sail in open seas where before sailors had relied on lands to determine the position of their ship. The method was the culmination of several advances that took place during the eighteen century.

Carl Friedrich Gauss is credited with developing the fundamentals of least squares analysis in 1795. Legendre was the first to publish the method however an early demonstration of the strength of Gauss method came when it was used to predict the future location of the newly discovered asteroid Ceres. Another astronomer Franz Xvaer Von Zach also[10] used the method of least squares to locate Ceres. Gauss did not publish his idea not until 1809 in his work on Celestial mechanics titled "Theoria Motus corporum coelestium in secionbus corricis solem ambientium".

Corresponding Author: Noor Atinah Ahmad, School of Mathematical Sciences Universiti Sains Malaysia Pulau-Pinang, Malaysia
Email: atinah@cs.usm.my

In 1829 he was able to establish and relate the least squares approach to regression analysis which is optimal in the sense that in a linear model where the errors have a mean of zero which are uncorrelated and have equal variances, the best linear unbiased estimator of the coefficient is the least square estimator.

This paper is organized as follow. Formulation of least squares problem in adaptive filtering is discussed in section II.In section III, recursive formulation of least squares problem is described. Conclusion follows in section IV.

Formulation of Least Squares Problem in Adaptive Filtering:

Suppose at any time instant i>0 the adaptive filter parameters are computed so that the quantity

$$\varepsilon(k) = \sum_{k=0}^{i} \beta(k)\mathbf{e}_i^2(k) \tag{4}$$

is minimized. When k=1, this means that the iteration has just started, e(k) for k=1,2..i are the sample error estimate that would be obtained if the filter run from time k to i. Using the set of parameters computed at time i, $\beta(k)$ is the forgetting factor. In this approach the filter parameter are optimized by using all the observations from the time the filter starts till the end and minimizing the sum of squared values of the error samples of the filter output. Consider the following notations and definitions of the parameters.

$$\begin{aligned}\mathbf{X}(k) &= [\mathbf{x}_0(k)\mathbf{x}_1(k)\mathbf{x}_2(k)...\mathbf{x}_{k-1}(k)]^T \\ \mathbf{w}(k) &= [w_0(k)w_1(k)w_2(k)...w_{k-1}(k)]^T\end{aligned} \tag{5}$$

are the input vector signal and adaptive tap weight coefficient vectors respectively. The adaptive filter output is obtained as an inner product of the adaptive tap weight coefficient vector and the input signal vector.

$$y(k) = \sum_{i=0}^{k-1} \mathbf{w}_i(k)\mathbf{x}_i(k) \tag{6}$$

Our objective here is to minimize (6).We can rewrite (6) in vector form as

$$y_i(k) = \mathbf{w}^T(k)\mathbf{x}(k) \tag{7}$$

k=1,2..I (7) is the adaptive filter output generated using adaptive tap weight coefficient vector w(k), therefore, the error signal can be written as

$$\mathbf{e}_i(k) = \mathbf{d}(k) - y_i(k) \tag{8}$$

From (4) and (7) the subscript i added to the adaptive filter output and error is to indicate that these quantities are obtain using the solution w(k) at time up to i. In this formulation, we transform the following into matrix/vector form

$$\begin{aligned}\mathbf{d}(k) &= [\mathbf{d}(1)\mathbf{d}(2)...\mathbf{d}(k)]^T \\ y(k) &= [y_i(1)y_i(2)...y_i(k)]^T \\ \mathbf{e}(k) &= [\mathbf{e}_i(1)\mathbf{e}_i(2)...\mathbf{e}_i(k)]^T\end{aligned} \tag{9}$$

We define information matrix (data matrix)

$$\mathbf{X}(k) = [\mathbf{x}(1)\mathbf{x}(2)...\mathbf{x}(k)] \tag{10}$$

Substituting (7) and (8) into (9, 10) we obtain

$$\begin{aligned}y(k) &= \mathbf{X}^T(k)\mathbf{w}(k) \\ \mathbf{e}(k) &= \mathbf{d}(k) - y)(k)\end{aligned} \tag{11}$$

Recursive Least Squares Algorithm:

Before we begin this section, we discuss the advantages and disadvantages associated to this approach. This technique is known to pursue fast convergence even when the eigenvalue spread of the input signal autocorrelation matrix is large. This approach have excellent performance when working in time varying environment Ahmad and Okwonu, (2010). The acceptance of the recursive least squares filter has been impeded by sometimes unacceptable numerical performance in limited precision environments. This degradation of performance is explicitly noticeable for the family of fast recursive least squares filters and it has been proved to be numerically unstable Athanasios, & Sergios (1996). Our objective here is to select coefficient of the adaptive filter in a way that the adaptive filter output signal $y(k)$during the period of observation will correspond with the desired signal. Since the minimization process requires the information of the input signal to be available. From (4), let the forgetting factor be

$\beta(k) = \lambda^{i-k}$, k=1,2..i, the forgetting factor should be chosen in the range $0 \le \lambda \le 1$,since the information of

the distance past has an increasing negligible effect on the coefficient update Ahmad and Okwonu (2010). Suppose that the forgetting factor is less than one, then the forgetting factor define above will give more weight to the recent samples of the error estimates compare with the former. If the forgetting factor is one this implies that the process lay more emphasis on the recent sample of the observed data and tends to forget the past (Bjock, 1996; Alexander and Ghirnikar, 1993; Bhouri *et al.,* 2005; Bhouri *et al.,* 1998; Gaye & Wood, 2008; Jian-Shu & Xue-gang, 2006; Raymond *et al.,* 2007). The ratio 1/(1λ) is used to measure the memory of the algorithm otherwise if the forgetting factor is one this means that the algorithm has infinite memory. From (4) we have

$$\varepsilon(k) = \sum_{k=0}^{i} \lambda^{i-k}[\mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{w}(k)]^2 \tag{12}$$

From (4) we observed that the error signal is the difference between the desired signal and the adaptive filter output vector using the most recent adaptive tap weight coefficient vector w(k). Differentiating (12) with respect to w(k) gives

$$\frac{\partial \varepsilon(k)}{\partial \mathbf{w}(k)} = -2\sum_{k=0}^{i} \lambda^{i-k}\mathbf{X}(k)[\mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{w}(k)] \tag{13}$$

Equating (13) to zero, we find that the adaptive tap weight coefficient vector w(k) that minimizes the least squares error is given by the following relation

$$-\sum_{k=0}^{i} \lambda^{i-k}\mathbf{X}(k)\mathbf{X}^T(k)\mathbf{w}(k) + \sum \lambda^{i-k}\mathbf{X}(k)\mathbf{d}(k) = \begin{bmatrix} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{14}$$

(14) is the normal equation, that is

$$\sum_{k=0}^{i} \lambda^{i-k}\mathbf{X}(k)\mathbf{X}^T(k)\mathbf{w}(k) = \sum \lambda^{i-k}\mathbf{X}(k)\mathbf{d}(k) \tag{15}$$

From (15) we have

$$\mathbf{w}(k)=\left[\sum_{k=0}^{i}\lambda^{i-k}\mathbf{X}(k)\mathbf{X}^{T}(k)\right]^{-1}\times$$

$$\left[\sum_{k=0}^{i}\lambda^{i-k}\mathbf{X}(k)\mathbf{d}(k)\right] \tag{16}$$

$$=\varphi^{-1}(k)\wp(k)$$

Where

$$\varphi(k)=\left[\sum_{k=0}^{i}\lambda^{i-k}\mathbf{X}(k)\mathbf{X}^{T}(k)\right]$$

$$\wp(k)=\left[\sum_{k=0}^{i}\lambda^{i-k}\mathbf{X}(k)\mathbf{d}(k)\right] \tag{17}$$

Where $\varphi(k)$ in (17) is the autocorrelation matrix and $\wp(k)$ is the cross-correlation matrix between the input signal and the desired signal vectors. The autocorrelation matrix $\varphi(k)$ is assumed to be nonsingular, suppose also that the autocorrelation matrix $\varphi(k)$ is singular the only technique to obtain the adaptive tap weight coefficient vector that minimize the error signal vector i(k) is by using the generalize inverse (Athanasios, &T. Sergios, 1996; Bhouri *et al.,* 1998; Ahmad and Okwonu 2010). Suppose we assumed that the input signal is zero for $i<0$ then the autocorrelation matrix will always be singular for $1<k<i$ that is during the initialization period which utilizes the back substitution (Alexander and Ghirnikar 1993; Ahmad and Okwonu, 2010), the inverse of theautocorrelation matrix $\varphi^{-1}(k)$ requires order of N cube that is ($O(N^{3})$) computational complexity. The recursion from (17) and result from matrix algebra form the basis of the derivation of the recursive least squares algorithm. We apply the matrix inversion lemma as follow

$$(\mathbf{A}+\alpha\mathbf{a}\mathbf{a}^{T})^{-1}=\mathbf{A}^{-1}-\frac{\alpha\mathbf{A}^{-1}\mathbf{a}\mathbf{a}^{T}\mathbf{A}^{-1}}{1+\alpha\mathbf{a}^{T}\mathbf{A}^{-1}\mathbf{a}} \tag{18}$$

Where A is $N\times N$ matrix, a is an $N\times 1$ vector and α is a scalar. By (18) the inverse autocorrelation matrix can be computed as

$$\varphi^{-1}(k)=$$

$$\frac{1}{\lambda}\left[\varphi^{-1}(k-1)-\frac{\varphi^{-1}(k-1)\mathbf{X}(k)\mathbf{X}^{T}(k)\varphi^{-1}(k-1)}{\lambda+\mathbf{X}^{T}(k)\varphi^{-1}(k-1)\mathbf{X}(k)}\right] \tag{19}$$

$$\wp(k)=\lambda\wp(k-1)+\mathbf{d}(k)\mathbf{X}(k)$$

$$\mathbf{w}(k)=\varphi^{-1}(k)\wp(k)$$

(16) Can be computed in a more simplified form as follow

$$\left[\sum_{k=0}^{i}\lambda^{i-k}\mathbf{X}(k)\mathbf{X}^{T}(k)\right]\mathbf{w}(k)=\lambda\left[\sum_{k=0}^{i}\mathbf{X}(k)\mathbf{d}(k)\right]+\mathbf{X}(k)\mathbf{d}(k) \tag{20}$$

Assume that $\varphi(k-1)\mathbf{w}(k-1)=\wp(k-1)$, then the following relations follows

$$\begin{aligned}&\left[\sum_{k=0}^{i}\lambda^{i-k}\mathbf{X}(k)\mathbf{X}^T(k)\right]\mathbf{w}(k)\\&=\lambda\wp(k-1)+\mathbf{X}(k)\mathbf{d}(k)\\&=\lambda\varphi(k-1)\mathbf{w}(k-1)+\mathbf{X}(k)\mathbf{d}(k)\\&=\left[\sum_{k=0}^{i}\lambda^{i-k}\mathbf{X}(k)\mathbf{X}^T(k)-\mathbf{X}(k)\mathbf{X}^T(k)\right]\times\mathbf{w}(k-1)+\mathbf{X}(k)\mathbf{d}(k)\end{aligned}\tag{21}$$

From the above equation the autocorrelation matrix $\mathbf{X}(k)\mathbf{X}^T(k)$ was subtracted and added to the right hand side of (21), from the information available we define the a priori error as

$$\mathbf{e}(k)=\mathbf{d}(k)-\mathbf{X}^T(k)\mathbf{w}(k-1)\tag{22}$$

The desired signal defined in (22) is expressed as the function of the a priori error and further mathematical manipulation in (21) gives

$$\mathbf{w}(k)=\mathbf{w}(k-1)+\mathbf{e}(k)\wp(k)\mathbf{X}(k)\tag{23}$$

Suppose further that the information matrix and the desired signal vector is given by

$$\begin{aligned}\mathbf{X}(k)&=\left[\mathbf{x}(k)\quad\lambda^{1/2}\mathbf{x}(k-1)...\quad\lambda^{k/2}\mathbf{x}(0)\right]\\\mathbf{d}(k)&=\left[\mathbf{d}(k)\quad\lambda^{1/2}\mathbf{d}(k-1)...\quad\lambda^{k/2}\mathbf{d}(k)\right]^T\end{aligned}\tag{24}$$

Where X(k) is (N+1)×(N+1 and the desired signal is (k+1)×1. At this point, we want to use the orthogonality principle to verify that the weighted error vector is the null space of the information matrix. We rewrite (16) as

$$\mathbf{X}(k)\mathbf{X}^T(k)\mathbf{w}(k)=\mathbf{X}(k)\mathbf{d}(k)\tag{25}$$

since $\mathbf{X}^T(k)\mathbf{w}(k)$ constitute vectors that includes all the adaptive filter output when the coefficients are given by the adaptive tap weight coefficient vector then

$$y(k)=\left[y(k)\quad\lambda^{1/2}y(k-1)...\quad\lambda^{k/2}y(0)\right]^T=\mathbf{X}^T(k)\mathbf{w}(k)$$

From above equations we have

$$\mathbf{X}(k)\mathbf{X}^T(k)\mathbf{w}(k)-\mathbf{X}(k)\mathbf{d}(k)=\mathbf{X}(k)[y(k)-\mathbf{d}(k)]=0$$

This shows that the weighted error vector is the null space of the information matrix which means that the weighted error vector is orthogonal to all row vectors of the information matrix; this shows that equation (25) is the normal equation.

ACKNOWLEDGMENT

The authours acknowledge the support from the Higher Education Department of Malaysia and Universiti Sains Malaysia for financial support through the Fundamental Research Grant Scheme (FRGS).

Conclusion:

We have shown that the constituent component of adaptive tap weight coefficient vector (solution to the problem) is the error signal minimization.Thus, the method presented also discuss on how to measure the memory of the algorithm, the ensuing discussion also shield light on the formation of normal equation from the conventional method of multiplying the triangular system with the transpose of the information matrix to more classical method based on matrix inversion lemma.

REFERENCES

Ahmad, N.A., F.Z. Okwonu, 2010. Techniques and analysis of adaptive least squares problem. Global journal of pure and applied Mathematics, 6: 53-62.

Alexander, S.T. and A.L. Ghirnikar, 1993. A method for recursive least squares filtering based upon an inverse QR decomposition. IEEE transaction on signal processing, 41: 20-30.

Athanasios, A.R. & T. Sergios, 1996. An adaptive least squares algorithm based on orthogonal Householder transformation. Electronic,circuits and systems,ICECS'96, proceedings of third IEEE international conference., 2: 800-806.

Bhouri, M., M. Bonne, M.A. Mboup, 2005. A new QRD based block adaptive algorithm,. IEEE, 1497-1500.

Bhouri, M.B., M.A. Mboup, 1998. A new QRD based block adaptive algorithm. IEEE, Proceedings of the 1998 international Conference on acoustics and singal processing (ICASSP) 3: 1497-1500.

Bjock, A., 1996. Numerical methods for least squares problems. SIAM, Philadelphia.

Gaye, L., & R. Wood, 2008. QR recursive least squares IP core example 15th annual IEEE international conference and workshop on the engineering of computer based systems, pp: 369-374.

Jian-Shu, C. & W. Xue-gang, 2006. LSMI algorithm based on inverse QR decomposition. IEEE, pp: 262-265.

Raymond, C.G.H.C., & D.P. O'Leary, 2007. (Ed.), Milestones in matrix computation:Selected works of Gene H.Golub,with commentaries, Oxford University press.