SOBRE ALGORITMOS RLS QUE UTILIZAM TRIANGULARIZAÇÃO ORTOGONAL

Maria D. Miranda*
Max Gerken**

Resumo

Neste trabalho são apresentadas interpretações para variáveis intrínsecas do método de ortogonalização QR, baseado em rotações de Givens, quando aplicado ao problema dos mínimos quadrados. Em particular, é apresentada uma interpretação original dos elementos da matriz de sistema que descreve as operações de adaptação e filtragem dos algoritmos QR-RLS baseados em rotações de Givens. Essas interpretações são obtidas estabelecendo-se uma conexão entre o método QR baseado em rotações de Givens e o método de Gram-Schmidt. Elas não só permitem uma melhor compreensão dos vários algoritmos recursivos dos mínimos quadrados que utilizam a triangularização ortogonal da matriz de dados como também fornecem subsídios para o desenvolvimento de novos algoritmos.

Abstract

In the context of the least squares adaptive filtering method the paper presents some interpretations for intrinsic variables of the QR method, based on Givens rotations. Particularly, an interpretation is presented for the elements of the system matrix that describes the filtering and parameter update operations of QR-RLS algorithms using Givens rotations. These interpretations are based on the connection between the QR method using Givens rotations and the Gram-Schmidt orthogonalization method. They give an insight of different least squares recursive algorithms that are based on the orthogonal triangularization of the data matrix and are useful for the development of new algorithms.

* Departamento de Pós-Graduação em Engenharia Elétrica da Universidade Presbiteriana Mackenzie. E-mail: mdm@mackenzie.com.br.

** Departamento de Engenharia de Telecomunicações e Controle da Escola Politécnica da Universidade de São Paulo. E-mail: mgk@lcs.poli.usp.br.

1 INTRODUÇÃO

Os algoritmos recursivos dos mínimos quadrados (RLS – *Recursive Least Squares*) têm propriedades de convergência mais robustas quando comparados aos algoritmos do gradiente estocástico (LMS – *Least Mean Squares*).[1] No caso dos primeiros algoritmos RLS propostos, esta vantagem era obtida à custa de uma complexidade computacional significativamente superior. Porém, no âmbito da filtragem adaptativa, o desenvolvimento dos algoritmos RLS rápidos[1,2] reduziu de forma significativa a diferença da complexidade computacional dos dois tipos de algoritmos. Infelizmente, como é reportado na literatura (e.g.),[1] muitos destes algoritmos apresentam problemas de instabilidade numérica.

Do ponto de vista da robustez numérica, a decomposição QR baseada em rotações de Givens é reconhecida como um dos melhores métodos para resolver o problema dos mínimos quadrados.[1,4,9] Aplicada ao problema dos mínimos quadrados, ela resulta em um procedimento em que somente dados transformados são utilizados, sendo que os coeficientes ótimos podem ser obtidos pelo método de substituições sucessivas. Um resultado fundamental para o desenvolvimento de algoritmos QR-RLS rápidos foi introduzido por McWhirter,[8] que observou que a operação de substituições sucessivas não é necessária, caso somente o erro de estimação *a posteriori* seja de interesse. Neste caso, os coeficientes do filtro transversal não aparecem explicitamente. Os erros de estimação para todas as ordens inferiores ou iguais à ordem do problema de estimação são calculados por um único filtro ortogonal variante no tempo, que recebe a denominação de estrutura de McWhirter. Nos últimos anos foram propostos diversos algoritmos QR-RLS rápidos utilizando essa estrutura.[5,7,9]

À primeira vista, as variáveis internas dos algoritmos QR-RLS rápidos são grandezas distintas daquelas dos algoritmos RLS rápidos. Entretanto, Regalia e Bellanger[5] mostraram, em seu artigo, que os coeficientes de reflexão e de regressão de um filtro em treliça normalizado aparecem como variáveis intermediárias quando a decomposição QR é aplicada ao problema dos mínimos quadrados. Conseqüentemente, os algoritmos QR-RLS rápidos podem ser utilizados em aplicações em que os coeficientes ótimos são de interesse, sem que seja necessário efetuar o procedimento de substituições sucessivas.

Complementando as interpretações de Regalia e Bellanger,[5] o presente trabalho relaciona variáveis internas dos algoritmos QR-RLS rápidos, baseados em rotações de Givens, com variáveis dos algoritmos recursivos em ordem que utilizam o procedimento de ortogonalização de Gram-Schmidt.[6,3] Para estabelecer essas relações, parte-se da descrição das operações de filtragem e atualização dos coeficientes no algoritmo RLS convencional, que podem ser descritas por um sistema linear variante no tempo. A matriz de sistema variante no tempo, correspondente aos algoritmos RLS, obtida a partir da ortogonalização QR utilizando rotações de Givens, é interpretada como resultante da aplicação de uma transformação similar à matriz de sistema que descreve o algoritmo RLS convencional. A comparação entre os dois métodos de triangularização ortogonal permite uma interpretação original dos elementos da matriz de sistema que descrevem os algoritmos QR-RLS baseados em rotações de Givens. Essas interpretações facilitam também a dedução de uma série de relações matemáticas envolvendo os vetores de erro de predição regressivo *a priori* e *a posteriori* e o fator de conversão. Baseado nesse conhecimento, é possível deduzir outros algoritmos, como, por exemplo, o apresentado no texto de Miranda & Gerken.[12]

No texto a seguir, são introduzidas inicialmente a notação utilizada e algumas relações matemáticas referentes ao algoritmo RLS convencional. Nos itens 3 e 4 é feita uma curta revisão dos métodos de triangularização ortogonal QR e de Gram-Schmidt, quando utilizados para resolver o problema dos mínimos quadrados. Segue então a comparação entre os dois métodos e a conseqüente interpretação dos elementos da matriz de sistema ortogonal, que descreve a atualização dos coeficientes e a operação de filtragem dos algoritmos QR-RLS.

2 A ESTIMAÇÃO DOS MÍNIMOS QUADRADOS

O problema de filtragem dos mínimos quadrados de ordem M consiste em determinar no instante n o vetor de coeficientes $\mathbf{w}_M(n)$, que minimiza a função

$$\xi_M(n) = \left\|\mathbf{e}_M(n)\right\|^2,$$

onde $\mathbf{e}_M(n) = \mathbf{\Lambda}^{\frac{1}{2}}(n)[e_M(0) \cdots e_M(n)]^T$ é o vetor dos erros de estimação *a posteriori*. Este vetor é definido como a diferença entre o vetor de respostas desejadas $\mathbf{d}(n) = \mathbf{\Lambda}^{\frac{1}{2}}(n)[d(0) \cdots d(n)]^T$ e o vetor estimado destas respostas, $\hat{\mathbf{d}}(n) = \mathbf{A}_M(n)\mathbf{w}_M(n)$. A matriz de dados de entrada

$$\mathbf{A}_M(n) = \left[A_M(j,i)\right] = \left[\lambda^{\frac{(n-j+1)}{2}} u(j-i)\right]$$

é uma matriz de dimensão $(n+1) \times M$ $(n+1)$ com $\{u(j-i)\}$ representando a seqüência de dados pré-janelada ($u(k) = 0$ para $k < 0$). A matriz

$$\mathbf{\Lambda}(n) = diag\left\{\lambda^{\frac{n}{2}}, \lambda^{\frac{(n-1)}{2}}, \cdots, 1\right\}$$

é a matriz de ponderação, sendo λ o fator de esquecimento ($0 << \lambda < 1$). A função (1) é minimizada quando o vetor de coeficientes satisfaz

$$\mathbf{w}_M(n) = \mathbf{\Phi}_M^{-1}(n)\mathbf{z}_M(n), \tag{2}$$

onde

$$\mathbf{\Phi}_M(n) = \mathbf{A}_M^T(n)\mathbf{A}_M(n) \tag{3}$$

é a matriz de autocorrelação dos dados de entrada e

$$\mathbf{z}_M(n) = \mathbf{A}_M^T(n)\mathbf{d}(n) \tag{4}$$

o vetor de correlação cruzada de dimensão $M \times 1$. A matriz de autocorrelação $\Phi_M(n)$ de dimensão $M \times M$ é admitida como sendo não singular. Os correspondentes vetores de estimação e de erro de estimação ótimos satisfazem

$$\hat{\mathbf{d}}(n) = \mathbf{P}_M(n)\mathbf{d}(n) \tag{5}$$

e

$$\mathbf{e}_M(n) = \left(\mathbf{I} - \mathbf{P}_M(n)\right)\mathbf{d}(n), \tag{6}$$

sendo $\mathbf{P}_M(n) = \mathbf{A}_M(n)\mathbf{\Phi}_M^{-1}(n)\mathbf{A}_M^T(n)$ o operador de projeção responsável pela projeção do vetor de respostas desejadas que é ortogonal ao erro de estimação ótimo.

2.1 O algoritmo RLS convencional

No algoritmo RLS convencional, o vetor de coeficientes $\mathbf{w}_M(n)$ pode ser obtido através da seguinte operação adaptativa:[1]

$$\mathbf{w}_M(n) = \mathbf{w}_M(n-1) + \mathbf{g}_M(n)\alpha(n), \tag{7}$$

sendo $\alpha(n)$ o erro de estimação *a priori* obtido através da operação de filtragem

$$\alpha(n) = d(n) - \mathbf{u}_M^T(n)\mathbf{w}_M(n-1), \tag{8}$$

onde $\mathbf{u}_M(n) = [u(n) \;\; u(n-1) \; \dots \; u(n-M+1]^T$ é o vetor de dados de entrada no instante n para a ordem M. O vetor $\mathbf{g}_M(n)$, conforme,[2,1] pode ser obtido como

$$\mathbf{g}_M(n) = \lambda^{-1}\gamma_M(n)\mathbf{\Phi}_M^{-1}(n-1)\mathbf{u}_M(n)$$

ou

$$\mathbf{g}_M(n) = \mathbf{\Phi}_M^{-1}(n)\mathbf{u}_M(n). \tag{9}$$

Por sua vez, o fator de conversão $\gamma_M(n)$, definido como o quociente entre os erros de estimação *a posteriori* e *a priori*,[1,10] satisfaz

$$\boxed{\gamma_M(n) = 1 - \mathbf{u}_M^T(n)\mathbf{\Phi}_M^{-1}(n)\mathbf{u}_M(n)} \tag{10}$$

Introduzindo-se o erro normalizado em ângulo[1,8]

$$\boxed{\varepsilon_M(n) = \gamma_M^{\frac{1}{2}}(n)\alpha_M(n) = \gamma_M^{-\frac{1}{2}}(n)e_M(n)} \tag{11}$$

é possível reunir (7), (8) e (11) no seguinte conjunto de equações recursivas no tempo que caracterizam as operações de adaptação e de filtragem do algoritmo RLS convencional:

$$\boxed{\begin{bmatrix} \mathbf{w}_M(n) \\ \varepsilon_M(n) \end{bmatrix} = \mathbf{S}^w(n) \begin{bmatrix} \mathbf{w}_M(n-1) \\ d(n) \end{bmatrix}} \tag{12}$$

onde $\mathbf{S}^w(n)$ é definida como

$$\mathbf{S}^w(n) = \begin{bmatrix} \mathbf{I} - \mathbf{g}_M(n)\mathbf{u}(n) & \mathbf{g}_M(n) \\ -\gamma_M^{\frac{1}{2}}(n)\mathbf{u}_M^T(n) & \gamma_M^{\frac{1}{2}}(n) \end{bmatrix}.$$

O algoritmo RLS convencional é suscetível ao ruído numérico em aritmética de precisão finita.[1] Uma forma de amenizar este problema consiste em efetuar uma triangularização ortogonal conveniente na matriz de dados, o que significa efetuar uma troca de coordenadas em (12). Na verdade, o sistema (12) pode ser transformado em um número ilimitado de sistemas com o mesmo comportamento entrada-saída e resolvendo o mesmo problema dos mínimos quadrados.[11] Apesar de os sistemas resultantes serem teoricamente equivalentes, o comportamento numérico pode variar conforme o sistema de coordenadas usado. A equivalência entre eles é exata somente no caso de aritmética de precisão infinita.

3 A TRIANGULARIZAÇÃO ORTOGONAL DA MATRIZ DE DADOS

A triangularização da matriz de dados pode ser convenientemente descrita por

$$\boxed{\mathbf{Q}_M(n)\mathbf{A}_M(n) = \begin{bmatrix} \mathbf{Q1}_M(n) \\ \mathbf{Q2}_M(n) \end{bmatrix} \mathbf{A}_M(n) = \begin{bmatrix} \mathbf{R}_M(n) \\ \mathbf{0} \end{bmatrix}} \tag{13}$$

onde $\mathbf{Q}_M(n)$ é uma matriz unitária de dimensão $(n + 1) \times (n + 1)$ e $\mathbf{R}_M(n)$ é uma matriz triangular superior de dimensão $M \times M$. Como $A_M(n)$ é uma matriz de dimensão $(n + 1) \times M$, é interessante considerar a partição da matriz $\mathbf{Q}_M(n)$ em duas matrizes, uma de dimensão $M \times (n + 1)$ representada por $\mathbf{Q1}_M(n)$ e outra de dimensão $(n + 1 - M) \times (n + 1)$ representada por $\mathbf{Q2}_M(n)$.

A função custo do problema dos mínimos quadrados (1) fica inalterada se incluirmos nos cálculos a matriz unitária $\mathbf{Q}_M(n)$, isto é ,

$$\|\mathbf{e}_M(n)\|^2 = \|\mathbf{Q}_M(n)\mathbf{e}_M(n)\|^2 . \tag{14}$$

Considerando o vetor do erro de estimação *a posteriori*

$$\mathbf{e}_M(n) = \mathbf{d}(n) - \mathbf{A}_M(n)\mathbf{w}_M(n)$$

e introduzindo-se a notação

$$\mathbf{Q1}_M(n)\mathbf{d}(n) = \mathbf{dq1}(n) \tag{15}$$

e

$$\mathbf{Q2}_M(n)\mathbf{d}(n) = \mathbf{dq2}(n) \tag{16}$$

obtém-se com (13) a seguinte expressão:

$$\mathbf{Q}_M(n)\mathbf{e}_M(n) = \begin{bmatrix} \mathbf{dq1}_M(n) \\ \mathbf{dq2}_M(n) \end{bmatrix} - \begin{bmatrix} \mathbf{R}_M(n) \\ \mathbf{0} \end{bmatrix} \mathbf{w}_M(n) . \tag{17}$$

De (17) conclui-se que a norma euclidiana do vetor dos erros de estimação *a posteriori* é minimizada quando a seguinte equação é satisfeita:

$$\mathbf{dq1}(n) = \mathbf{R}_M(n)\mathbf{w}_M(n) , \tag{18}$$

condição em que

$$\xi_M(n) = \|\mathbf{e}_M(n)\|^2 = \|\mathbf{dq2}(n)\|^2 . \tag{19}$$

Como $\mathbf{R}_M(n)$ é uma matriz triangular superior, o vetor de coeficientes ótimos pode ser obtido de (18) pelo método de substituições sucessivas.[1]

Observando que

$$\mathbf{Q}_M^T(n)\mathbf{Q}_M(n) = \mathbf{Q}_M(n)\mathbf{Q}_M^T(n) = \mathbf{I},$$

obtém-se da equação (13) a igualdade

$$\mathbf{A}_M(n) = \mathbf{Q1}_M^T(n)\mathbf{R}_M(n) . \tag{20}$$

Com (20), (18) e (15), conclui-se que o vetor estimado ótimo satisfaz

$$\boxed{\hat{\mathbf{d}}(n) = \mathbf{A}_M(n)\mathbf{w}_M(n) = \mathbf{Q1}_M^T\mathbf{Q1}_M(n)\mathbf{d}(n)} \quad (21).$$

De (18), (17) e (16) resulta que o erro de estimação mínimo satisfaz

$$\boxed{\mathbf{e}_M(n) = \mathbf{Q2}_M^T\mathbf{Q2}_M(n)\mathbf{d}(n)} \quad (22).$$

Desta forma, o operador de projeção fica definido pela matriz $\mathbf{Q1}_M(n)$, isto é ,

$$\boxed{\mathbf{P}_M(n) = \mathbf{Q1}_M^T(n)\mathbf{Q1}_M(n)} \quad (23)$$

e o complementar do operador de projeção pela matriz, $\mathbf{Q^2}_{\mathrm{M}}(n)$

$$\boxed{\mathbf{I} - \mathbf{P}_M(n) = \mathbf{Q2}_M^T(n)\mathbf{Q2}_M(n)} \quad (24).$$

A partir de (20) e (15) resulta que o vetor de correlação cruzada (4) da equação normal é dado por

$$\boxed{\mathbf{z}_M(n) = \mathbf{R}_M^T(n)\mathbf{dq1}(n)} \quad (25).$$

Finalmente, a matriz de autocorrelação satisfaz

$$\boxed{\mathbf{\Phi}_M(n) = \mathbf{R}_M^T(n)\mathbf{R}_M(n)} \quad (26).$$

o que decorre das equações (20) e (3).

A matriz $\mathbf{Q}_M(n)$ pode ser interpretada como uma rotação do subespaço gerado pelas colunas da matriz $\mathbf{A}_M(n)$ e pelo erro de estimação ótimo $\mathbf{e}_M(n)$, o qual é ortogonal às colunas da matriz $\mathbf{A}_M(n)$. No caso, a matriz $\mathbf{Q}_M(n)$ faz coincidir o subespaço gerado pelas colunas da matriz $\mathbf{A}_M(n)$, de dimensão M, com o subespaço correspondente aos M primeiros vetores da base do espaço euclidiano, de dimensão $(n + 1) \times (n + 1)$, utilizado na representação. O subespaço de dimensão $n + 1 - M$, que contém o erro de estimação ótimo $\mathbf{e}_M(n)$ é ortogonal ao espaço das colunas de $\mathbf{A}_M(n)$, coincide então com o subespaço gerado pelos demais $n + 1 - M$ vetores da base euclidiana. Esta transformação afeta o sinal de entrada, o sinal desejado e o operador de projeção, sem alterar a matriz de autocorrelação, o vetor de correlação cruzada e o vetor de coeficientes ótimos.

3.1 Atualização no tempo

As atualizações no tempo da matriz unitária $\mathbf{Q}_M(n)$, da matriz de dados triangularizada $\mathbf{R}_M(n)$ e do vetor de respostas desejadas rotacionado $\mathbf{dq}(n)$ podem ser convenientemente descritas (ver Haykin)[1] por

$$\underbrace{\mathbf{T}(n)\begin{bmatrix} \mathbf{Q}_M(n-1) & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{Q}_M(n)}\begin{bmatrix} \lambda^{\frac{1}{2}}\mathbf{A}_M(n-1) & \lambda^{\frac{1}{2}}\mathbf{d}(n-1) \\ \mathbf{u}_M^T(n) & d(n) \end{bmatrix} =$$

$$= \mathbf{T}(n)\begin{bmatrix} \lambda^{\frac{1}{2}}\mathbf{R}_M(n-1) & \lambda^{\frac{1}{2}}\mathbf{dq1}(n-1) \\ \mathbf{0} & \lambda^{\frac{1}{2}}\mathbf{dq2}(n-1) \\ \mathbf{u}_M^T(n) & d(n) \end{bmatrix} = \begin{bmatrix} \mathbf{R}_M(n) & \mathbf{dq1}(n) \\ \mathbf{0} & \lambda^{\frac{1}{2}}\mathbf{dq2}(n-1) \\ \mathbf{0}_M^T & \varepsilon_M(n) \end{bmatrix}. \tag{27}$$

A matriz $\mathbf{T}(n)$ é o produto de M rotações de Givens utilizadas para anular os M elementos do vetor $\mathbf{u}_M^T(n)$. Destas operações decorre que o elemento não nulo da última linha da matriz resultante em (27) é o erro de estimação *a posteriori*, normalizado em ângulo $\varepsilon_M(n)$ definido em (11) (ver McWhirter).[8] Eliminando-se da matriz $\mathbf{T}(n)$ as linhas e colunas de índices $M + 1$ até n, obtém-se uma matriz unitária de dimensão $(M + 1) \times (M + 1)$, aqui denominada de $\mathbf{S}^q(n)$. Com o auxílio desta matriz pode-se reescrever parte da equação (27) como sendo

$$\mathbf{S}^q(n)\begin{bmatrix} \lambda^{\frac{1}{2}}\mathbf{R}_M(n-1) & \lambda^{\frac{1}{2}}\mathbf{dq1}(n-1) \\ \mathbf{u}_M^T(n) & d(n) \end{bmatrix} = \begin{bmatrix} \mathbf{R}_M(n) & \mathbf{dq1}(n) \\ \mathbf{0}_M^T & \varepsilon_M(n) \end{bmatrix}. \tag{28}$$

Esta igualdade mostra que a matriz ortogonal $\mathbf{S}^q(n)$, além de atualizar a matriz $\mathbf{R}_M(n)$ no tempo, também faz o papel de matriz de sistema que descreve as operações de adaptação e filtragem do algoritmo QR-RLS. Cabe observar que em (28) o vetor $\mathbf{dq1}(n)$ tem o mesmo papel que $\mathbf{w}_M(n)$ em (12).

4 O PROCEDIMENTO DE ORTOGONALIZAÇÃO DE GRAM-SCHMIDT

No procedimento de ortogonalização de Gram-Schmidt, é determinada uma base de M vetores mutuamente ortogonais, a partir de um conjunto de M vetores linearmente independentes. No caso, esse conjunto é formado pelas colunas

$$\mathbf{a}(n-i+1) = \mathbf{\Lambda}^{\frac{1}{2}}(n)\begin{bmatrix} u(1-i) \\ \vdots \\ u(n-i) \\ u(n-i+1) \end{bmatrix},$$

$1 \leqslant i \leqslant M$, da matriz de dados $\mathbf{A}_M(n)$. O procedimento básico é formar os vetores $\mathbf{e}^b_{i-1}(n)$, de dimensão $1 \times (n+1)$, a partir das colunas $\mathbf{a}(n-i+1)$, subtraindo de cada coluna $\mathbf{a}(n-i+1)$ as suas componentes no espaço gerado pelos vetores $\mathbf{e}^b_{j-1}(n)$, $1 \leqslant j \leqslant i-1$, isto é,

$$\boxed{\begin{aligned} &\mathbf{e}^b_0(n) = \mathbf{a}(n) \\ &\mathbf{e}^b_{i-1}(n) = \mathbf{a}(n-i+1) - \hat{\mathbf{a}}(n-i+1) \end{aligned}} . \tag{29}$$

O vetor $\hat{\mathbf{a}}(n-i+1)$ representa a projeção da coluna $\mathbf{a}(n-i+1)$ no subespaço descrito pelos vetores $\mathbf{e}^b_{j-1}(n)$:

$$\boxed{\hat{\mathbf{a}}(n-i+1) = \sum_{j=1}^{i-1} k_{ji}(n)\mathbf{e}} \tag{30}$$

para, $i = 2,3, \ldots, M$, com

$$\boxed{k_{ji}(n) = \frac{\mathbf{e}^{bT}_{j-1}(n)\mathbf{a}(n-i+1)}{\left\|\mathbf{e}^b_{j-1}(n)\right\|^2}, j < i} . \tag{31}$$

Combinando as equações (29) e (30) e considerando as M colunas da matriz $\mathbf{A}_M(n)$, chega-se à seguinte expressão matricial:

$$\boxed{\underbrace{\begin{bmatrix} \mathbf{a}^T(n) \\ \mathbf{a}^T(n-1) \\ \vdots \\ \mathbf{a}^T(n-M+1) \end{bmatrix}^T}_{\mathbf{A}_M(n)} = \underbrace{\begin{bmatrix} \mathbf{e}_0^{bT}(n) \\ \mathbf{e}_1^{bT}(n) \\ \vdots \\ \mathbf{e}_{M-1}^{bT}(n) \end{bmatrix}^T}_{\mathbf{G}_M(n)} \mathbf{K}_M(n)}, \tag{32}$$

onde $\mathbf{K}_M(n)$ é uma matriz triangular superior, de dimensão $M \times M$, com os elementos da diagonal principal unitários. Os elementos da coluna i de $\mathbf{K}_M(n)$ para $i > j$ são os coeficientes $k_{ji}(n)$ obtidos das projeções sucessivas da coluna $\mathbf{a}(n - i + 1)$ sobre os vetores $\mathbf{e}_{j-1}^b(n), 1 \leqslant j \leqslant M$. As M colunas ortogonais da matriz $\mathbf{G}_M(n)$ não são normalizadas. Desta forma, o produto

$$\mathbf{G}_M^T(n)\mathbf{G}_M(n) = \mathbf{D}_M(n)$$

é uma matriz diagonal, cujos elementos representam a norma ao quadrado dos vetores ortogonais $\mathbf{e}_{j-1}^b(n), 1 \leqslant i \leqslant M$, obtidos no procedimento de Gram-Schmidt.

4.1 Interpretações

Como a estimativa de $\mathbf{a}(n - i + 1)$, obtida em (30), é ortogonal a $\mathbf{e}_{i-1}^b(n)$ e as colunas da matriz $\mathbf{A}_M(n)$ são formadas por dados seqüenciais deslocados, então, utilizando-se o princípio da ortogonalidade, conclui-se que $\mathbf{e}_{i-1}^b(n)$ representa o vetor dos erros *a posteriori* de predição regressiva de ordem $i - 1$. Assim, a matriz $\mathbf{G}_M(n)$ pode ser reescrita como:

$$\boxed{\mathbf{G}_M(n) = \left[\mathbf{e}_0^b(n) \cdots \mathbf{e}_{M-1}^b(n)\right] = \left[\mathbf{b}(0) \cdots \mathbf{b}(n)\right]^T}. \tag{33}$$

Os elementos $b_{i-1}(n)$, $1 \leqslant \mathrm{i} \leqslant M$, do vetor coluna $\mathbf{b}(n)$ são os erros de predição regressiva

$$b_{i-1}(n) = u(n-i+1) - \hat{u}\left(n-i+1 \middle| U_n\right),$$

onde U_n representa o conjunto das $i - 1$ amostras, $\{u(n) \;\; u(n-1) \; \ldots \; u(n-i+2)\}$. A partir da equação (32), obtém-se

34
$$\boxed{\left[\mathbf{b}(0)\cdots\mathbf{b}(n)\right] = \mathbf{K}_M^{-T}(n)\left[\mathbf{u}_M(0)\cdots\mathbf{u}_M(n)\right]\Lambda^{\frac{1}{2}}(n)},$$

e, portanto, $\mathbf{b}(n) = \mathbf{K}_M^{-T}(n)\mathbf{u}_M(n)$, o que permite fazer as seguintes interpretações (ver também, Haykin[1] e Proakis et al):[3]

- os elementos não nulos da linha i da matriz $\mathbf{K}_M^{-T}(n)$ são os coeficientes do filtro de erro de predição regressiva de ordem $i - 1$;
- a coluna i de $\mathbf{G}_M(n)$, $1 \leqslant i \leqslant M$ é a seqüência de erros de predição regressiva, correspondente ao filtro da i-ésima linha de $\mathbf{K}_M^{-T}(n)$;
- os elementos da matriz diagonal $\mathbf{D}_M(n)$ correspondem à energia de erro de predição regressiva,

$$\mathbf{D}_M(n) = diag\left\{\xi_0^b(n), \xi_1^b(n), \ldots, \xi_{M-1}^b(n)\right\},$$

sendo que a energia de erro da predição regressiva $\xi_{i-1}^b(n)$ está associada ao filtro de erro de predição da i-ésima linha de $\mathbf{K}_M^{-T}(n)$. Desta forma, $\xi_{i-1}^b(n) = \left\|\mathbf{e}_{i-1}^b(n)\right\|^2$.

Aplicando-se o procedimento de ortogonalização de Gram-Schmidt à matriz $[\mathbf{A}_M(n) \;\; \mathbf{d}(n)]$, resulta a matriz ortogonal $[\mathbf{G}_M(n) \;\; \mathbf{e}_M(n)]$, onde $\mathbf{e}_M(n) = \mathbf{d}(\mathbf{n}) - \hat{\mathbf{d}}(n)$. O vetor de respostas desejadas estimado $\hat{\mathbf{d}}(n)$ pode ser obtido a partir dos vetores ortogonais $\mathbf{e}_{j-1}^b(n)$, $1 \leqslant j \leqslant M$:

35
$$\boxed{\hat{\mathbf{d}}(n) = \sum_{j=1}^{M} k_{j,M+1}^d(n)\mathbf{e}_{j-1}^b(n) = \mathbf{G}_M(n)\mathbf{K}_M^d(n)}$$

onde

36
$$\boxed{k_{j,M+1}^d(n) = \frac{\mathbf{e}_{j-1}^{bT}(n)\mathbf{d}(n)}{\xi_{j-1}^b(n)}}$$

e

37
$$\boxed{\mathbf{K}_M^d(n) = \left[k_{j,M+1}^d \;\cdots\; k_{j,M+1}^d\right]^T = \mathbf{D}_M^{-1}(n)\mathbf{G}_M^T(n)\mathbf{d}(n)}.$$

Neste caso, $\mathbf{e}_M(n)$, que é ortogonal a $\hat{\mathbf{d}}(n)$, é o vetor dos erros de estimação *a posteriori* que minimiza a função custo definida em (1). O vetor de estimação ótimo é dado por

$$\boxed{\hat{\mathbf{d}}(n) = \mathbf{G}_M(n)\mathbf{D}_M^{-1}(n)\mathbf{G}_M^T(n)\mathbf{d}(n)} \tag{38}$$

Conseqüentemente, para o operador de projeção, resulta a seguinte expressão:

$$\mathbf{P}_M(n) = (\mathbf{G}_M(n)\mathbf{D}_M^{-\frac{1}{2}}(n)\mathbf{.}\ (\mathbf{D}_M^{-\frac{1}{2}}(n)\mathbf{G}_M^T(n)) = \sum_{i=1}^{M} \bar{\mathbf{e}}_{i-1}^{b}(n)\bar{\mathbf{e}}_{i-1}^{bT}(n),$$

sendo

$$\boxed{\bar{\mathbf{e}}_{i-1}^{b}(n) = \frac{\mathbf{e}_{i-1}^{b}(n)}{\xi_{i-1}^{\frac{b}{2}}(n)}} \tag{39}$$

o vetor dos erros de predição regressivos *a posteriori* normalizados.

A estimativa de $\mathbf{d}$(n) obtida utilizando-se os erros de predição regressiva $\mathbf{e}_{j-1}^{b}(n)$, $1 \leqslant j \leqslant M$ é conhecida como estimativa do processo conjunto, e o vetor $\mathbf{K}_M^d(n)$ representa o vetor dos coeficientes da estimação do processo conjunto, também chamados de coeficientes de regressão. Como

$$\boxed{\hat{\mathbf{d}}(n) = \mathbf{A}_M(n)\mathbf{w}_M(n) = \mathbf{G}_M(n)\mathbf{K}_M^d(n)} \tag{40}$$

o vetor de coeficientes ótimos $\mathbf{w}_M(n)$ satisfaz

$$\boxed{\mathbf{K}_M(n)\mathbf{w}_M(n) = \mathbf{K}_M^d(n)} \tag{41}$$

As matrizes $\mathbf{R}_M(n)$ e $\mathbf{D}_M^{\frac{1}{2}}(n)\mathbf{K}_M(n)$ são iguais, pois ambas são fator de Cholesky da matriz de autocorrelação $\mathbf{A}_M^T(n)\mathbf{A}_M(n)$. Isto resulta das equações (20) e (32) e das propriedades das matrizes envolvidas. Assim, comparando (18) e (41), conclui-se que os vetores $\mathbf{K}_M^d(n)$ e $\mathbf{dq1}(n)$, resultantes dos procedimentos de ortogonalização de Gram-Schmidt e de Givens da matriz $\mathbf{A}_M(n)$, estão relacionados pela matriz diagonal $\mathbf{D}_M^{\frac{1}{2}}(n)$

$$\boxed{\mathbf{dq1}(n) = \mathbf{D}_M^{\frac{1}{2}}(n)\mathbf{K}^d(n)} \tag{42}$$

Conseqüentemente, $\mathbf{dq1}(n)$ tem uma interpretação equivalente ao coeficiente de regressão do estimador conjunto. Substituindo (37) em (42), conclui-se que $\mathbf{Q1}^T(n) = \mathbf{G}_M(n)\mathbf{D}_M^{-\frac{1}{2}}(n)$. Logo, as colunas de $\mathbf{Q1}^T(n)$ são os erros da predição regressiva *a posteriori* normalizados $\overline{\mathbf{e}}_{i-1}^{b}(n)$, $1 \leq i \leq M$, introduzidos em (39). Desta forma, como $\mathbf{dq1}(n) = \mathbf{Q1}(n)\mathbf{d}(n)$, os elementos de $\mathbf{dq1}(n)$ são os coeficientes das projeções do vetor de respostas desejadas sobre os erros de predição regressiva *a posteriori* normalizados.

5 CONSIDERAÇÕES SOBRE A MATRIZ DE SISTEMAS $\mathbf{S}^q(n)$

A equação (28) mostra que as operações de adaptação e de filtragem dos algoritmos QR-RLS podem ser descritas pelo seguinte sistema variante no tempo:

$$\begin{bmatrix} \mathbf{dq1}(n) \\ \varepsilon_M(n) \end{bmatrix} = \mathbf{S}^q(n) \begin{bmatrix} \lambda^{\frac{1}{2}}\mathbf{dq1}(n-1) \\ d(n) \end{bmatrix}, \tag{43}$$

onde a matriz unitária $\mathbf{S}^q(n)$ é a matriz de sistema. A equação (18), $\mathbf{dq1}(n) = \mathbf{R}_M(n)\mathbf{w}_M(n)$, sugere que a matriz $\mathbf{R}_M(n)$ pode ser interpretada como uma transformação similar que modifica o sistema (12) no sistema (43). Realmente, aplicando-se a transformação

$$\begin{bmatrix} \mathbf{R}_M(n) & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}$$

em (12), resulta, após algumas manipulações algébricas, o sistema equivalente (43) sendo

$$\mathbf{S}^q(n) = \gamma_M^{\frac{1}{2}}(n) \begin{bmatrix} \lambda^{\frac{1}{2}}\gamma_M^{-\frac{1}{2}}(n)\mathbf{\Gamma}(n) & \gamma_M^{-\frac{1}{2}}(n)\overline{\mathbf{b}}(n) \\ -\overline{\boldsymbol{\psi}}^T(n) & 1 \end{bmatrix}, \tag{44}$$

onde

$$\boxed{\mathbf{\Gamma}(n)=\mathbf{R}_M^{-T}(n)\mathbf{R}_M^{T}(n-1)} \tag{45}$$,

$$\boxed{\overline{\mathbf{b}}(n)=\mathbf{R}_M^{-T}(n)\mathbf{u}_M(n)} \tag{46}$$,

$$\boxed{\overline{\boldsymbol{\psi}}(n)=\lambda^{-\frac{1}{2}}\mathbf{R}_M^{-T}(n-1)\mathbf{u}_M(n)} \tag{47}$$.

A raiz quadrada do fator de conversão $\gamma_M^{\frac{1}{2}}(n)$, definido em (10), neste caso pode ser calculada através da seguinte expressão:[1]

$$\boxed{\gamma_M^{\frac{1}{2}}(n)=\prod_{i=1}^{M}\cos\theta_i(n)} \tag{48}$$.

Os ângulos $\{\theta i(n)\}$ são calculados de forma a eliminar os elementos do vetor $\mathbf{u}_M(n)$ em (28), sendo que os algoritmos QR-RLS rápidos utilizam a predição progressiva e regressiva para fazer este cálculo de forma eficiente.

Os erros de estimação *a posteriori* podem ser obtidos diretamente, a partir de (43), sem efetuar a operação de substituições sucessivas necessária no cálculo do vetor de coeficientes do filtro transversal $\mathbf{w}_M(n)$.[8] A equação (43) caracteriza um sistema ortogonal variante no tempo, devido às características da própria matriz $\mathbf{S}^q(n)$, a qual resulta do produto M de rotações de Givens. Esse sistema pode ser representado como uma cascata de sistemas ortogonais de primeira ordem, conhecida como estrutura de McWhirter.[5] Nesta estrutura, cada bloco da cascata é caracterizado por uma rotação de Givens definida por $\theta_i(n)$.

Substituindo em $\mathbf{R}_M(n)=\mathbf{D}_M^{\frac{1}{2}}(n)\mathbf{K}_M(n)$, em (46) e (47), obtém-se

$$\boxed{\overline{\mathbf{b}}(n)=\mathbf{D}_M^{-\frac{1}{2}}(n)\underbrace{\mathbf{K}_M^{-T}(n)\mathbf{u}_M(n)}_{\mathbf{b}(n)}} \tag{49}$$

e

$$\boxed{\overline{\boldsymbol{\psi}}(n)=\lambda^{-\frac{1}{2}}\mathbf{D}_M^{-\frac{1}{2}}(n-1)\underbrace{\mathbf{K}_M^{-T}(n-1)\mathbf{u}_M(n)}_{\boldsymbol{\psi}(n)}} \tag{50}$$.

A partir das observações sobre as matrizes $\mathbf{K}_M^{-T}(n), \mathbf{D}_M(n)$ e $\mathbf{G}_M(n)$, efetuadas na seção anterior, é possível interpretar adequadamente as equações (49) e (50) da seguinte forma:

- Na equação (49) o vetor $\mathbf{b}(n)$ representa as saídas, no tempo n, dos filtros definidos por $\mathbf{K}_M^{-T}(n)$. Isto é, o vetor dos erros de predição regressivos *a posteriori* no tempo n. Conseqüentemente, os elementos do vetor

$$\overline{\mathbf{b}}(n) = \left[\overline{b}_0(n), \overline{b}_0(n), \cdots, \overline{b}_0(n)\right]^T$$

são os erros de predição regressivos *a posteriori* normalizados

51
$$\boxed{\overline{b}_k(n) = \frac{b_k(n)}{\xi_k^{\frac{b}{2}}(n)}},$$

para $0 \leqslant k \leqslant M-1$.

- Da equação (50), conclui-se que os elementos do vetor $\overline{\boldsymbol{\psi}}(n)$ são saídas, no tempo n, dos filtros de erro de predição regressiva definidos por $\mathbf{K}_M^{-T}(n-1)$. Isto é, dos filtros de erro de predição regressiva com os coeficientes otimizados para o instante de tempo $n-1$. Portanto, $\overline{\boldsymbol{\psi}}(n)$ é o vetor dos erros de predição regressiva *a priori* no tempo n. Os elementos de

$$\overline{\boldsymbol{\psi}}(n) = \left[\overline{\psi}_0(n), \overline{\psi}_1(n), \cdots, \overline{\psi}_{M-1}(n)\right]^T$$

são os erros de predição regressiva *a priori*, normalizados pela energia do erro de predição regressiva *a posteriori* no tempo $n-1$ e ponderados por $\lambda^{-\frac{1}{2}}$, ou seja:

52
$$\boxed{\overline{\psi}_k(n) = \frac{\psi_k(n)}{\lambda^{\frac{1}{2}}\boldsymbol{\xi}_k^{\frac{b}{2}}(n-1)}},$$

para $0 \leqslant k \leqslant M-1$.

Como a matriz $\mathbf{S}^q(n)$ é o resultado do produto de uma seqüência de rotações de Givens, resulta que $\mathbf{S}^{qT}(n)\mathbf{S}^q(n) = \mathbf{I}$ e $\mathbf{S}^q(n)\mathbf{S}^{qT}(n) = \mathrm{I}$. Desta igualdade resultam as seguintes propriedades:

$$\boxed{\left\|\overline{\boldsymbol{\psi}}(n)\right\|^2 = \gamma_M^{-1}(n) - 1} \tag{53}$$,

$$\boxed{\left\|\overline{\mathbf{b}}(n)\right\|^2 = 1 - \gamma_M(n)} \tag{54}$$,

$$\boxed{\overline{\boldsymbol{\psi}}(n) = \gamma_M^{-1}(n)\lambda^{\frac{1}{2}}\boldsymbol{\Gamma}^T(n)\overline{\mathbf{b}}(n)} \tag{55}$$,

$$\boxed{\overline{\mathbf{b}}(n) = \lambda^{\frac{1}{2}}\boldsymbol{\Gamma}(n)\overline{\boldsymbol{\psi}}(n)} \tag{56}$$,

$$\boxed{\overline{\boldsymbol{\psi}}(n)\overline{\boldsymbol{\psi}}^T(n) = \gamma_M^{-1}(n)\left(\mathbf{I} - \lambda\boldsymbol{\Gamma}^T(n)\boldsymbol{\Gamma}(n)\right)} \tag{57}$$

e

$$\boxed{\overline{\mathbf{b}}(n)\overline{\mathbf{b}}^T(n) = \mathbf{I} - \lambda\boldsymbol{\Gamma}(n)\boldsymbol{\Gamma}^T(n)} \tag{58}$$

Tais propriedades, embora não sejam exploradas neste trabalho, mostram-se úteis tanto para a análise de propriedades de estabilidade como para o desenvolvimento de novos algoritmos QR-RLS.

6 CONSIDERAÇÕES FINAIS

Relacionando-se dois procedimentos de triangularização ortogonal da matriz de dados, um baseado em rotações de Givens e outro no método de Gram-Schmidt, foram apresentadas interpretações para variáveis internas dos algoritmos QR-RLS baseados em rotações de Givens. A análise e as interpretações apresentadas permitem não só uma melhor compreensão dos algoritmos QR-RLS baseados em rotações de Givens e das suas relações com outros algoritmos RLS, mas também o desenvolvimento de novos algoritmos RLS rápidos, numericamente robustos. Por exemplo, na referência[5] o vetor dos erros de predição regressiva normalizado $\overline{\mathbf{b}}(n)$ é utilizado no desenvolvimento de um algoritmo QR-RLS rápido. Neste caso, o vetor $\overline{\mathbf{b}}(n)$ é obtido através de um filtro em treliça normalizado baseado no erro *a posteriori*. A identificação do vetor de erros de predição regressiva *a priori* $\overline{\boldsymbol{\psi}}(n)$, normalizado em relação à energia do erro de predição regressiva *a posteriori*, na matriz $\mathbf{S}^q(n)$ permite o desenvolvimento de algoritmos QR-RLS rápidos baseados nos erros de predição *a priori*. Um algoritmo deste tipo foi proposto por Miranda & Gerken.[12,13]

REFERÊNCIAS BIBLIOGRÁFICAS

1. HAYKIN, S. *Adaptive Filter Theory.* 2 ed. New Jersey: Prentice Hall, 1991.
2. BELLAMGER, M. *Analyse des signaux et filtrage numéerique adapatif.* Paris: Masson et CNET-ENST, 1989.
3. PROAKIS, J.G., RADER, C.M., LING, F., NIKIAS, C.L. *Advanced digital signal processing.* New York: Maxwell Mac-Millan, 1992.
4. REGALIA, P.A. Numerical stability properties of a QR-based fast least squares algorithm. *IEEE Transactions on Signal Processing,* v.41, n.6, p.2096-2109, jun. 1993.
5. REGALIA, P.A., BELLANGER, M.G. On the duality between fast QR methods and lattice methods in least squares adaptive filtering. *IEEE Transactions on Signal Processing,* v.39, n.4, p.879-891, abril 1991.
6. LING, F. Givens rotation based LSL and related algorithms. *IEEE Transactions on Signal Processing,* v.39, n.7, p.1541-1551, jul. 1991.
7. YANG, B., BÖHME, J.F. Rotation-Based RLS algorithms: unified derivations, numerical properties, and parallel implementations. *IEEE Transactions on Signal Processing,* v.40, n.5, p.1151-1167, maio 1992.
8. MCWHIRTER, J. G. Recursive least squares minimization using a systolic array. *Proc. SPIE Int. Soc. Opt. Eng.,* v. 431, p.105-112, 1983.
9. PROUDLER, I.K., MCWHIRTER, J.G., SHEPERD, T.J. Computationally efficient QR decomposition approach to least squares adaptive filtering. *IEE Proceedings – F.,* v.138, n.4, ago. 1991.
10. CARAYANNIS, G., MANOLAKIS, D.G., KALOUPTIDIS, N. A fast sequential algorithm for least-squares filtering an prediction. *IEEE Trans. on Acoustics, Speech, and Signal Processing,* v.31, n.6, p.1394-1402, dez. 1983,
11. LJUNG, S., LJUNG, L. Error propagation properties of recursive least-squares adaptation algorithms. *Automatica,* v.21, n.2, p.157-167, 1985.
12. MIRANDA, M.D., GERKEN, M. *A Hybrid QR-Lattice Least Squares Algorithm Using A Priori Errors.* In: Proceedings of the 38th. Midwest Symposium on Circuits and Systems, 13 a 16 de agosto de 1995, Rio de Janeiro, p.983-986.
13. MIRANDA, M.D., GERKEN, M. Hybrid Least Squares QR-Lattice Algorithm Using a Priori Errors. *IEEE Transactions on Signal Processing,* v.45, n.12, p.2900-2911, dez. 1997.
14. MIRANDA, M.D., AGUAYO, L., GERKEN, M. Performance of the a priori and a posteriori QR-LSL algorithms in a limited precision environment. *Proceedings of the IEEE, ICASSP-97,* Munique, p.2337-2340, abr. 1997.
15. MIRANDA, M.D., GERKEN, M. Sobre algoritmos RLS que utilizam triangularização ortogonal, *Anais* do 13º Simpósio Brasileiro de Telecomunicações, 3 a 6 de setembro de 1995, Águas de Lindóia, p.339-344.