International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:4, No:2, 2010

A Family of Minimal Residual Based Algorithm for Adaptive Filtering

Noor Atinah Ahmad

***Abstract*—**The Minimal Residual (MR) is modified for adaptive filtering application. Three forms of MR based algorithm are presented: i) the low complexity SPCG, ii) MREDSI, and iii) MREDSII. The low complexity is a reduced complexity version of a previously proposed SPCG algorithm. Approximations introduced reduce the algorithm to an LMS type algorithm, but, maintain the superior convergence of the SPCG algorithm. Both MREDSI and MREDSII are MR based methods with Euclidean direction of search. The choice of Euclidean directions is shown via simulation to give better misadjustment compared to their gradient search counterparts.

***Keywords*—**Adaptive filtering, Adaptive least square, Minimal residual method.

I. Introduction

ALGORITHMS for finding the optimum weight vector of an adaptive filter can be classified into two: direct method and iterative method. Most of these methods are derived from standard method for solving least squares problem, but modified to suit the stochastically recursive cost function. The main difference between direct and iterative method is that iterative method offers much lower computational complexity for every system update. For example, for an adaptive filter with order N, the Least Mean Square (LMS) algorithm which is derived from the steepest descent method has computational complexity of $O(N)$, the lowest thus far, whereas algorithm based on direct method such as the Recursive Least Squares (RLS) and the QRD-RLS, both have computational complexity of $O(N^2)$, which is considered to be too high for many practical applications.

Several conjugate gradient (CG) based adaptive filtering algorithm have also been derived [1, 2]. The main issue in these methods is to try and find the best estimate for gradient and search directions in the absence of the full knowledge of the autocorrelation matrix, and, in the presence of noise in the system. Noisy estimates of gradient can easily lead to loss of conjugacy among search directions which may cause instability of the algorithm.

Recently, the Stochastic Pairwise Conjugate Gradient (SPCG) algorithm is introduced [3]. The method is a form of stochastic gradient descent method where the stepsize is chosen to insist pairwise conjugation of successive gradient.

Noor Atinah Ahmad is with the Universiti Sains Malaysia , Malaysia
e-mail: atinah@cs.usm.my

The method was shown to have comparable performance with other CG-based method while having lower computational complexity because it avoids explicit computation of conjugate
search directions. It was also noted in [3] that the SPCG algorithm is a form of recursive Minimal Residual (MR) method where the direction of search is the current residual vector.

The MR method can also be performed with direction of search other than the residual vector [4]. In this correspondence, we explore this idea further in search of a better alternative to the SPCG algorithm. The objective is two fold: i) to improve computational complexity, and, ii) to avoid gradient search that usually leads to poor convergence when the eigenvalue spread of the autocorrelation matrix is large. Three different forms of MR-like methods are presented here, namely the low complexity SPCG and MR methods with Euclidean direction of search (MREDSI and MREDSII).

A. Adaptive Filtering Problem

The linear adaptive filtering problem is the following minimization problem,

$$\min_{\mathbf{x}\in R^N} J_n(\mathbf{x}) = \sum_{i=m}^{n} \lambda^{n-i} \left(\mathbf{a}_i^T \mathbf{x} - s(i)\right)^2 , \qquad (1)$$

where $s(i) \in R$ is the desired signal, and, $y(i) = \mathbf{a}_i^T \mathbf{x}$ is the filter output at the i th sample instant.
For a transversal finite impulse response (FIR) adaptive filter, vectors $\mathbf{a}_i \in R^N$ are formed by the input $u(i)$, such that $\mathbf{a}_i = \left[u(i) \quad u(i-1) \quad \cdots \quad u(i-N+1)\right]^T$, and vector $\mathbf{x} \in R^N$ is an estimate of the filter coefficient vector. The quantity $e_i = \mathbf{a}_i^T \mathbf{x} - s(i)$ is the error signal and $\mathbf{x}$ is updated by minimizing the sum of squared error cost function $J_n(\mathbf{x})$. The constant $\lambda \in [0,1]$ is known as the forgetting factor.

With definitions

$$\mathbf{A}_n = \left[\sqrt{\lambda^{n-m}}\mathbf{a}_m \quad \mathbf{a}_{m+1} \quad \cdots \quad \mathbf{a}_n\right]^T , \; 0 \le m < n$$

and, $$\mathbf{p}_n = \mathbf{A}_n^T \left[\sqrt{\lambda^{n-m}}\, s(m), s(m+1), \ldots, s(n)\right]^T ,$$

the minimization problem in (1) can be shown to be equivalent to solving the normal equation

$$\Phi_n \mathbf{x} = \mathbf{p} , \qquad (2)$$

where $\Phi = \mathbf{A}_n^T \mathbf{A}_n$. Recursive updates of Φ_n and $\mathbf{p}_n$ are possible [5] through the formulas,

$$\Phi_n = \lambda \Phi_{n-1} + \mathbf{a}_n \mathbf{a}_n^T,$$
$$\mathbf{p}_n = \lambda \mathbf{p}_{n-1} + \mathbf{a}_n s(k).$$

B. Minimal Residual (MR) Method for Adaptive Filtering

The standard MR method for solving linear system of equation of the form $\mathbf{Ax} = \mathbf{b}$ updates the approximation by the formula

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)} \mathbf{d}^{(k)},$$

where the stepsize $\alpha^{(k)}$ is chosen so that the residual two norm squared $\left\| \mathbf{b} - \mathbf{A}\mathbf{x}^{(k+1)} \right\|_2^2$ is minimized along the direction of search $\mathbf{d}^{(k)}$. This procedure is also equivalent to an orthogonal projection technique where the stepsize $\alpha^{(k)}$ is chosen so that the residual vector $\mathbf{r}^{(k+1)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(k+1)}$ is orthogonal to the subspace $span\left\{ \mathbf{A}\mathbf{d}^{(k)} \right\}$ (see Figure 1).

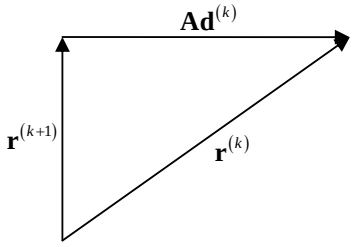

Fig. 1 Geometrical interpretation of the MR iteration

When applying this method to the adaptive filtering problem, we seek to update the adaptive filter coefficient vector using a similar idea, that is to choose stepsize $\alpha^{(k)}$ that minimizes the residual 2-norm squared $\left\| \mathbf{p}_k - \Phi_k \mathbf{x}^{(k+1)} \right\|_2^2$ along the stochastic direction of search $\mathbf{d}^{(k)}$. Since matrices $\mathbf{p}_k$ and Φ_k are updated recursively every time the system is updated, the method becomes a stochastic minimization algorithm. The long term behaviour of the method can be understood by considering the recursive form for the residual vector $\mathbf{r}^{(k)}$. From [2],

$$\begin{aligned} \mathbf{r}^{(k+1)} &= \mathbf{p}_{k+1} - \Phi_{k+1} \mathbf{x}^{(k+1)} \\ &= \lambda \mathbf{r}^{(k)} - \alpha^{(k)} \lambda \Phi_k \mathbf{d}^{(k)} + \mathbf{a}_{k+1} \left[s_{k+1} - \mathbf{a}_{k+1}^T \mathbf{x}^{(k+1)} \right] \\ &= \lambda \mathbf{r}^{(k)} - \alpha^{(k)} \lambda \Phi_k \mathbf{d}^{(k)} + \mathbf{a}_{k+1} e_{k+1} \end{aligned}$$

Let

$$\hat{\mathbf{r}}^{(k+1)} = \mathbf{p}_k - \Phi_k \mathbf{x}^{(k+1)} = \mathbf{r}^{(k)} - \alpha^{(k)} \Phi_k \mathbf{d}^{(k)}.$$

By construction of the MR method, α_k is chosen so that $\left(\hat{\mathbf{r}}^{(k+1)}, \Phi_k \mathbf{d}^{(k)} \right) = 0$. Therefore, we have,

$$\begin{aligned} \left(\mathbf{r}^{(k+1)}, \Phi_k \mathbf{d}^{(k)} \right) &= \lambda \left(\hat{\mathbf{r}}^{(k+1)}, \Phi_k \mathbf{d}^{(k)} \right) + e_{k+1} \left(\mathbf{a}_{k+1}, \Phi_k \mathbf{d}^{(k)} \right) \\ &= e_{k+1} \left(\mathbf{a}_{k+1}, \Phi_k \mathbf{d}^{(k)} \right) \end{aligned} \quad (3)$$

Assuming, $\mathbf{d}_k$ is a descent direction, (3) implies that, as $e_{k+1} \to 0$, the behaviour of the stochastic MR based adaptive filtering algorithm approaches that of its deterministic counterpart.

II. Variation of MR Adaptive Filtering Algorithm

The SPCG algorithm proposed in [3] can be viewed as a (stochastically) recursive form of the standard Minimal Residual (MR) method, where the direction of search is set to be the current residual. It can also be treated as a form of stochastic gradient descent algorithm. Although the SPCG algorithm shows higher convergence rate compared to the LMS algorithm, the computational complexity of SPCG algorithm is still rather high (refer Table 1). The SPCG algorithm also has the tendency to produce high misadjustment especially when the autocorrelation matrix Φ has large eigenvalue spread.

In order to overcome the disadvantages of the SPCG, we introduce several variations of MR –like algorithm,

i) Low complexity SPCG;
ii) MREDSI;
iii) MREDSII.

The low complexity SPCG is a modification of the SPCG algorithm, where the residual vector is approximated by the stochastic gradient approximation used in the LMS algorithm. Algorithms MREDSI and MREDSII explores two forms of updating the coefficient vector by MR-like iterations along Euclidean directions.

A. Low Complexity SPCG

The low complexity SPCG has $O(N)$ complexity, and this is achieved by making approximations to steps (7) and (9) in Table 1, i.e.,

$$\mathbf{r}^{(k+1)} \approx \hat{\mathbf{r}}^{(k+1)} = -\mathbf{a}_{k+1} e_{k+1};$$
$$\theta_k = \mathbf{a}_k^T \mathbf{a}_k;$$
$$\mathbf{b}^{(k+1)} = \lambda e_k \mathbf{R}_k \mathbf{a}_k + \mathbf{a}_k \theta_k e_k = \lambda \mathbf{b}^{(k)} + \mathbf{a}_k \theta_k e_k.$$

The steps above has $4N+1$ multiplications and $N-1$ additions, and, explicit formation of Φ_k and $\mathbf{p}_k$ is avoided. These approximations reduce the algorithm to an LMS type algorithm with a stepsize that minimizes the residual. The low complexity SPCG needs to be periodically restarted with the exact values for $\mathbf{r}^{(k)}$ and $\mathbf{b}^{(k)}$ to avoid stagnation [6]. Restarting after every N sample update, maintains the complexity at $O(N)$ per sample update.

TABLE 1
THE SPCG ALGORITHM

	$\times/\div$	$+/-$
Initialization:	(if $m=1$)	
(1a) $\mathbf{X}_m = \mathbf{a}_m\mathbf{a}_m^T$	1	
(1b) $\Phi_m, \mathbf{p}_m, \mathbf{x}^{(0)} = \mathbf{0} \in R^N$	-	
(1c) $\mathbf{r}^{(m)} = \mathbf{p}_m - \Phi_m\mathbf{x}^{(0)}$,	-	
(1d) $\mathbf{b}^{(m)} = \Phi_m\mathbf{r}^{(m)}$	1	
For $k = m+1,\ldots,N$		
(2) $\alpha^{(k)} = \dfrac{\mathbf{r}^{(k)T}\mathbf{b}^{(k)}}{\mathbf{b}^{(k)T}\mathbf{b}^{(k)}}$	$2N+1$	$2(N-1)$
(3) $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{r}^{(k)}$	N	N
New sample update: $\mathbf{a}_k \to \mathbf{a}_{k+1}$,		
(4a) $\mathbf{X}_{k+1}(2:N, 2:N) = \mathbf{X}_k(1:N-1, 1:N-1)$	-	
(4b) $\mathbf{X}_{k+1}(1,:) = u(n)\mathbf{a}_{k+1}^T$	N	
(4c) $\mathbf{X}_{k+1}(:,1) = \mathbf{X}_{k+1}(1,:)^T$	-	
(5) $\Phi_{k+1} = \lambda\Phi_k + \mathbf{X}_{k+1}$	$\frac{1}{2}N(N+1)$	$\frac{1}{2}N(N+1)$ (considering symmetry)
(6) $e_{k+1} = \mathbf{a}_{k+1}^T\mathbf{x}^{(k+1)} - s(k+1)$	N	$2N-1$
(7) $\mathbf{r}^{(k+1)} = \lambda\mathbf{r}^{(k)} - \alpha\lambda\mathbf{b}^{(k)} + \mathbf{a}_{k+1}e_{k+1}$	$3N+1$	$2N$
(8) $\mathbf{b}^{(k+1)} = \Phi_{k+1}\mathbf{r}^{(k+1)}$	N^2	$N(N-1)$
EndFor		
Total	$\frac{3}{2}N^2 + \frac{17}{2}N + 4$	$\frac{1}{2}(3N^2+13N-3)$
(if $\lambda=1$)	N^2+7N+3	

B. MR Iterations along Euclidean Directions

The SPCG and its low complexity version produce large misadjustment especially when the autocorrelation matrix has large eigenvalue spread. The eigenvalueapread (or the spectral condition number) is the ratio of the largest and the smallest eigenvalue of Φ. It has a value ≈ 1 when the input signal to the filter is white (i.e. uncorrelated) and >1 for coloured (i.e. correlated) signal. It is well known that gradient based methods are sensitive to eigenvalue spread.

In an effort to reduce sensitivity to eigenvalue spread, we introduce MR like method where solutions are updated along Euclidean directions. This choice of search direction is not only orthogonal but also easy to implement. We describe two different methods, MR-EDSI and MR-EDSII, both are of $O(N^2)$ complexity.

1) MR-EDSI

We assume the approximate solution is updated as

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_1^{(k)}\mathbf{e}_1 + \alpha_2^{(k)}\mathbf{e}_2 + \ldots + \alpha_N^{(k)}\mathbf{e}_N = \mathbf{x}_k + \Sigma_k,$$

where $\mathbf{e}_i = [0 \;\; 0 \;\; \cdots \;\; 1 \;\; \cdots \;\; 0]^T$, with 1 appearing in the i th place. The stepsizes $\alpha_i, i = 1,\ldots,N$, are computed as follows:

$$\text{Set } \alpha_1^{(k)} = \frac{\mathbf{r}_k^T\mathbf{R}_1}{\|\mathbf{R}_1\|_2^2},$$

and, for $i = 2,\ldots,N$,

$$\alpha_i^{(N)} = \left\{\mathbf{R}_i^T\left(\mathbf{r} - 2\sum_{j=1}^{i-1}\alpha_j^{(k)}\mathbf{R}_j\right)\right\} \Big/ \|\mathbf{R}_i\|_2^2 .$$

The choice of α_i's above guarentees $\tilde{\mathbf{r}}_{k+1}^T(\mathbf{R}_k\Sigma_k) = 0$, where $\tilde{\mathbf{r}}_{k+1} = \mathbf{r}_k - \mathbf{R}_k\Sigma_k$.

2) MR-EDSII

Another version of the MR-EDS algorithm takes after the EDS algorithm [7] where the MR projections are performed along the Euclidean direction cyclically. In both algorithm the current weight vector is updated as follows,

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \sum_{i=1}^{N}\alpha_i^{(k)}\mathbf{e}_i .$$

The major difference between MR-EDS II and the EDS algorithm is in the calculation of the stepsize $\alpha_i^{(k)}$. In EDS algorithm, $\alpha_i^{(k)}$ is the minimizer of $\nabla_\alpha J(\mathbf{x}_i^{(k)} + \alpha\mathbf{e}_i)$, whereby in the MR-EDS II, $\alpha_i^{(k)}$ is the minimizer of $\nabla_\alpha \left\|\mathbf{p}_k - \Phi_k(\mathbf{x}_i^{(k)} + \alpha\mathbf{e}_i)\right\|$, i.e., for MR-EDSII,

$$\alpha_i^{(k)} = \frac{\mathbf{r}_k^{(i)T}\Phi_k^{(i)}}{(\Phi_k^{(i)})^T(\Phi_k^{(i)})}, \qquad i = 1,\ldots,N$$

with

$$\mathbf{r}_i^{(k)} = \mathbf{r}_{i-1}^{(k)} - \alpha_i^{(k)}\Phi_k^{(i)},$$

$\Phi_k^{(i)}$ - the i th row of Φ_k.

The choice of $\alpha_i^{(k)}$ guarentees $\mathbf{r}_{i+1}^{(k)T}(\Phi\mathbf{e}_i) = \mathbf{r}_{i+1}^{(k)T}\Phi_i = 0$.

III. SIMULATION RESULTS

Our simulation is based on an adaptive system modeling configuration given in Figure 2. The input signal is passed through a colouring filter with frequency response $H(z) = \dfrac{\sqrt{1-\alpha^2}}{1-\alpha z^{-1}}$, where $|\alpha| < 1$. The parameter α controls the eigenvalue spread of the input autocorrelation matrix,

where $\alpha = 0$ gives uncorrelated sequence (white) with eigenvalue spread ≈ 1. The aim is to find the parameters of a model $\mathbf{x}$ through an adaptive algorithm so that the difference between the unknown system output, $s(n)$, and the adaptive model output, $y(n)$, is minimized according to the cost function $J_n(\mathbf{x})$. Noise, $\eta(n)$, with a variance of 0.001 is added to the output of the unknown system.

The performance of the algorithm is studied based on the propagation of the ensemble average of the mean error norm. The convergence rate of the algorithm refers to the number of iterations to reach steady state error, and, misadjusment is evaluated based on the value of the steady state error.

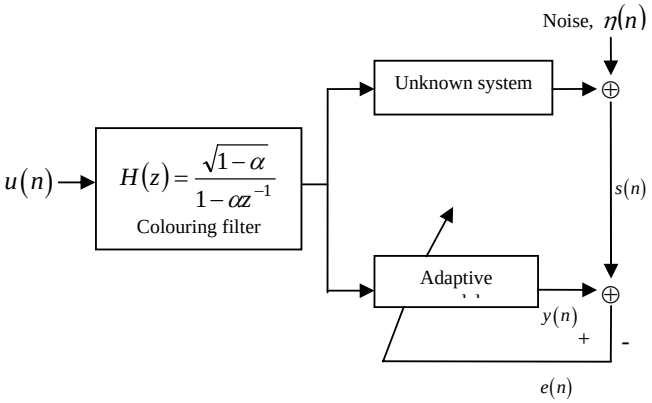

Fig. 2 Adaptive system modeling

A. Convergence Rate

In Figure 3, the performance of MR based algorithms are compared with that of the LMS algorithm. It is clear that the MR based algorithms have superior convergence rate compared to LMS despite having higher computational complexity. Even the low complexity SPCG shares the same rate of convergence with its original version. There is evidence however that the MR based algorithm tend to produce larger misadjustment compared to LMS.

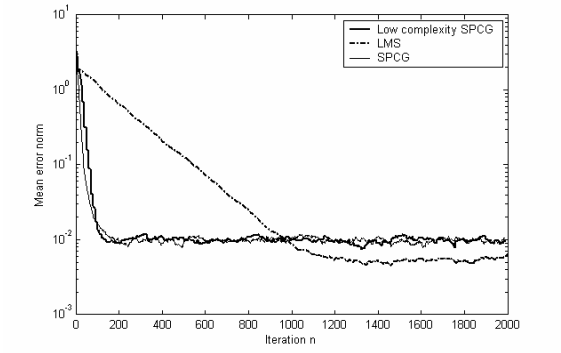

Fig. 3a Convergence rates of low complexity SPCG and SPCG compared with LMS

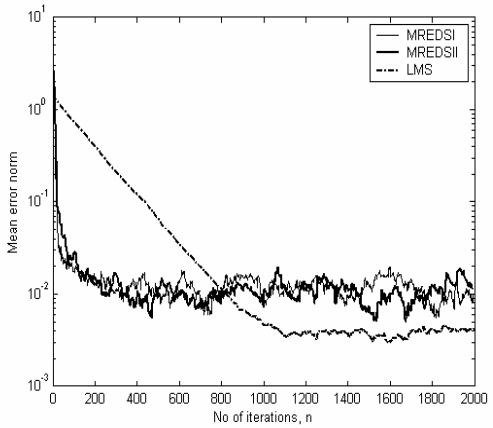

Fig. 3b Convergence rates of MREDSI and MREDSII compared with LMS

B. Stability with respect to Eigenvalue Spread

In Figure 4, the value of α is set to be 0.5 which leads to the value of eigenvalue spread of ≈ 6.6. Unlike SPCG and LMS, the misadjustment of MREDS algorithms seem to be unaffected by the increased eigenvalue spread. Between MREDSI and MREDSII, we see that MREDSII is more stable. MREDSI seems to be showing signs of instability where large errors are produced at the onset of adaptation. The MREDSII algorithm also shows better performance compared to EDS algorithm.

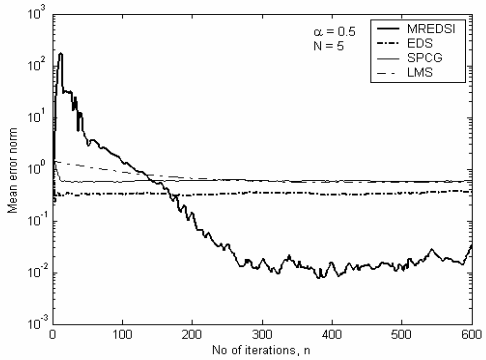

Fig. 4a: MREDSI gives smaller misadjustment compared to SPCG, LMS and EDS. However convergence seems to be slower

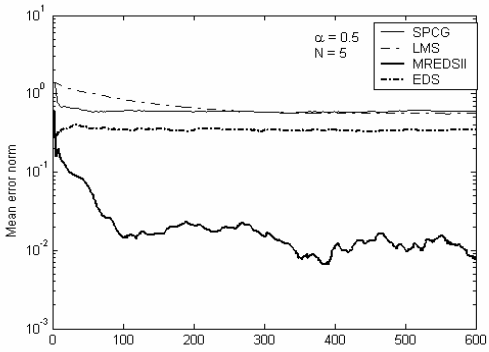

Fig. 4b MREDSII gives small misadjusment as well as high convergence rate

IV. Summary and Discussions

A family of three MR based adaptive filtering algorithms is presented. The low complexity SPCG algorithm provides a lower complexity version of the original SPCG proposed in [Ahmad]. This new version is an LMS-type algorithm with a stepsize calculated via the MR formula. Simulation results verify the superior convergence of low complexity SPCG compared to the LMS algorithm. Although some approximations are made to derive the low complexity version, with restarting, the algorithm maintains the performance of the original version.

Choosing Euclidean vectors as direction of search in MREDSI and MREDSII meets the objective of improving misadjustment which seems to be a disadvantage in their gradient search counterparts. Between the two versions, MREDSII shows better performance in that it gives better rate of convergence compared to MREDSI. It was also noted that although MREDSII and the EDS adopt similar update formulas, the choice of stepsizes in MREDSI allows it to provide better misadjustment than EDS.

Acknowledgment

This work is supported by the Department of Higher Education, Malaysia and Universiti Sains Malaysia under FRGS grant.

References

[1] G. K. Boray and M. D. Srinath, "Conjugate gradient techniques for adaptive filtering", *IEEE Trans. Circuits Syst.*, vol. 39, Jan. 1992, pp. 1-10.

[2] P. S. Chang and A. N. Wilson, Jr., "Analysis of conjugate gradient method for adaptive filtering", *IEEE Trans. Signal Processing*, vol. 48, no. 2, Feb. 2000.

[3] N. A. Ahmad, "A globally convergent stochastic pairwise conjugate gradient based adaptive filtering algorithm", *IEEE Signal Process. Lett.*, vol. 15, Dec. 2008, pp. 914-917.

[4] Y. Saad, *Iterative Methods for Sparse Linear Systems*. Boston, MA: PWS-Kent, 1996.

[5] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1991.

[6] N. A. Ahmad, 'Low Complexity Stochastic Pairwise Conjugate Gradient Based Adaptive Filtering Algorithm', (to be published).

[7] G. F. Xu, T. Bose and J. Schroeder, "The Euclidean direction search algorithm for adaptive filtering", *Proceedings of the 1999 IEEE Int. Symp. on Circuits and Systems*, **3**:146 – 149, 1999.