ANALYSIS OF THE EUCLIDEAN DIRECTION SET ADAPTIVE ALGORITHM

Guo-Fang Xu and Tamal Bose

Department of Electrical Engineering
University of Colorado
Denver, CO 80217-3364

ABSTRACT

A mathematical analysis is performed on a recently reported gradient-based adaptive algorithm named the Euclidean Direction Set (EDS) method. It has been shown that the EDS algorithm has a computational complexity of $O(N)$ for each system update, and a rate of convergence (based on computer simulations) comparable to the RLS algorithm. In this paper, the stability of the EDS method is studied and it is shown that the algorithm converges to the true solution. It is also proved that the convergence rate of the EDS method is superior to that of the steepest descent method.

1. INTRODUCTION

Adaptive filtering algorithms have been very successful in a wide variety of applications. The LMS algorithm is still very popular due to its computational simplicity. The slow convergence rate of the LMS algorithm is well known. On the other hand, the RLS algorithm has a fast convergence rate but its computational complexity of $O(N^2)$ makes its use prohibitive in many applications. A variety of other adaptive filtering algorithms have been recently reported including the Conjugate Gradient (CG) based methods and the Fast Transversal Filters (FTF) which have their own sets of pros and cons. Very recently, a direction set (DS) based algorithm was introduced in [1] for adaptive filtering applications. The algorithm keeps the minimum search either on a set of near conjugate directions with respective to the Hessian matrix or simply on a set of Euclidean coordinate directions. The proposed algorithm was shown to have a computational complexity of $O(N)$ for each coefficient update and a rate of convergence, based on computer simulations, comparable to the RLS algorithm [1], [2]. The application of this algorithm was also investigated and it performed very well in system identification and spectral estimation.

In this paper, we perform a mathematical analysis of the Euclidean direction set (EDS) method and prove that it converges to the true parameters and that its convergence rate is superior to the steepest descent method. The paper is organized as follows: In Section 2, we show that the EDS method is a new gradient-based adaptive process and it can be implemented by a Euclidean direction search procedure. A mathematical analysis of stability and convergence rate are presented in Section 3. Concluding remarks are given in Section 4.

This work was supported in part by a grant from the Colorado Advanced Software Institute.

2. EUCLIDEAN DIRECTION SET METHOD

Consider the following quadratic optimization problem,

$$J(W) = \min_{W \in R^N} \{W^T R W - 2W^T P + d\}, \tag{1}$$

where R is an $N \times N$ real symmetric positive definite matrix (Hessian); P and variable W are vectors of order N, and d is a scaler. The notation $(.)^T$ means the transpose of $(.)$.

The minimum of $J(W)$ can be obtained by setting the gradient $\nabla = 0$, where $\nabla \triangleq \frac{\partial J(W)}{\partial W} = 2RW - 2P$. This yields the normal equation : $RW = P$. Since R is nonsingular, the optimal solution W_* equals $R^{-1}P$. Rewriting this in terms of the initial value $W(0)$ and the gradient $\nabla(0)$, where $\nabla(0) \triangleq \nabla|_{W=W(0)}$, gives the so called one-step Newton's method $W_* = W(0) - \frac{1}{2}R^{-1}\nabla(0)$.

By introducing a constant μ to regulate the convergence rate, Newton's method can be expressed as:

$$W(k+1) = W(k) - \frac{\mu}{2}R^{-1}\nabla(k), \tag{2}$$

where $W(k)$ denotes the variable vector at step k, $\nabla(k) \triangleq \nabla|_{W=W(k)} = 2RW(k) - 2P$, and $0 < \mu < 2$.

Since matrix R is unknown in most practical applications, R^{-1} in Newton's method must be estimated based on a statistical sample, which in general, is computationally expensive.

The steepest descent method avoids dealing with matrix R^{-1} and updates the variable vector W on the negative gradient direction as follows:

$$W(k+1) = W(k) - \frac{\mu}{2}\nabla(k) , \tag{3}$$

where $0 < \mu < \frac{2}{\lambda_{\max}}$ and $\lambda_{\max}$ is the largest eigenvalue of R. This simplification makes the steepest descent method very popular in practical applications. However, the fact that the convergence rate is bounded by the condition number of matrix R is well known [3].

The Euclidean direction set (EDS) method introduced in [1] in adaptive filtering is a gradient-based method, which can be written as:

$$W(k+1) = W(k) - \frac{\mu}{2}A^{-1}\nabla(k) \ , \tag{4}$$

where $0 < \mu \leq 1$, and A is an $N \times N$ matrix with elements $\{a_{i,j}\}$. Matrix A is related to matrix R as below.

In the Direction Set method, the algorithm performs line search along the Euclidean directions systematically in a cyclic manner. When one direction is chosen at a time, the method is called EDS1. In this case, A is a lower triangular matrix with

$$a_{i,j} = \left\{ \begin{array}{l} r_{i,j}, \text{ for } i = 1, 2, \cdots N; \ 1 \leq j \leq i; \\ 0, \text{ otherwise} \end{array} \right\}. \tag{5}$$

When two directions are used for line searches at a time, the algorithm is called EDS2. In this case, A becomes a lower block triangular matrix with the row size in each block being 2, i.e.

$$a_{i,j} = \left\{ \begin{array}{l} r_{i,j}, \text{ for } i = 1, 2, \cdots N; \ 1 \leq j \leq i; \\ r_{i,j}, \text{ for } i = 1, 3, \cdots M; \ j = i+1; \\ 0, \text{ otherwise} \end{array} \right\} \tag{6}$$

where $r_{i,j}$ is the element of R and $M = N-1$ for N is even, otherwise, $M = N-2$.

Define $\alpha(k) = W(k+1) - W(k)$. Substituting $\alpha(k)$, $\nabla(k)$ into Eq. (4) and multiplying both sides by A, gives:

$$\left\{ \begin{array}{l} A\alpha(k) = -\mu(RW(k) - P); \\ W(k+1) = W(k) + \alpha(k). \end{array} \right\}$$

Note that if A is a triangular or block triangular matrix, $\alpha(k)$ can be solved explicitly.

Assume $R = \{r_{i,j}\}$, $P = [p_1, p_2, \cdots, p_N]^T$, $\mu = 1$, and at time k, $W(k) = [w_1, w_2, \cdots, w_N]^T$. $W(k+1)$ can be updated by the following EDS1 or EDS2 search cycle with $O(N)$ multiplications in each loop.

EDS1: (one Euclidean direction search at each loop.)

{ for $i = 1 : N$

(1): $\varepsilon_i = \sum_{j=1}^{N} r_{i,j}w_j - p_i;$

(2): $\alpha_i = -\frac{\varepsilon_i}{r_{i,i}};$

(3): $w_i = w_i + \alpha_i.$ }

EDS2: (double Euclidean direction search at each loop.)

{ for $i = 1 : N/2$

(1): $\varepsilon_{2i-1} = \sum_{j=1}^{N} r_{2i-1,j}w_j - p_{2i-1},$

$\varepsilon_{2i} = \sum_{j=1}^{N} r_{2i,j}w_j - p_{2i};$

(2): $r = r_{2i-1,2i-1}r_{2i,2i} - r_{2i-1,2i}r_{2i,2i-1};$

(3): $\alpha_{2i-1} = -\frac{\varepsilon_{2i-1}r_{2i-1,2i-1} - \varepsilon_{2i}r_{2i-1,2i}}{r},$

$\alpha_{2i} = -\frac{\varepsilon_{2i}r_{2i,2i} - \varepsilon_{2i}r_{2i,2i-1}}{r};$

(4): $w_{2i-1} = w_{2i-1} + \alpha_{2i-1},$

$w_{2i} = w_{2i} + \alpha_{2i}.$

if N is odd

(1): $\varepsilon_N = \sum_{j=1}^{N} r_{N,j}w_j - p_N;$

(2): $\alpha_N = -\frac{\varepsilon_N}{r_{N,N}};$

(3): $w_N = w_N + \alpha_N.$ }

The above search procedures are named Euclidean direction set (EDS) algorithm because it is equivalent to performing an orderly minimum search on a set of Euclidean directions, which was originally developed in [1], [2]. Such a one-by-one search scheme ensures that the sample-based EDS algorithm has $O(N)$ computational complexity in each system update.

Intuitively, we expect the EDS algorithm to converge faster than the steepest descent method and to be computationally less complex than Newton's method, because the EDS method utilizes matrix R partially and does not have to compute the inverse of matrix A.

3. ANALYSIS OF THE EDS METHOD

3.1. Stability of The EDS Algorithm

We begin the stability analysis by defining an error vector as $C(k) \triangleq W(k) - W_*$, where $W_* = R^{-1}P$ is the optimal solution of (1). Subtracting W_* from both sides of (4), and using $\nabla(k) = 2(RW(k) - RW_*)$, the error recursive equation becomes

$$C(k+1) = (I - \mu A^{-1}R)C(k),$$

where I is the identity matrix. With an initial value of $C(0)$, for the error vector, the solution of (??) is

$$C(k) = (I - \mu A^{-1}R)^k C(0). \tag{7}$$

It is well known that $C(k) \to 0$ as $k \to \infty$, if and only if $\rho(I - \mu A^{-1}R) < 1$, where $\rho(.)$ denotes the maximum absolute eigenvalue of (.).

Based on the fact that matrix A is a lower triangular submatrix of R, the following theorem provides a sufficient condition for the convergence of the EDS algorithm. It is worth noting that this sufficient condition is independent of the largest eigenvalue of matrix R.

Theorem 1 Let R be an $N \times N$ symmetric positive definite matrix, and matrix A be as defined as (5) or (6), with the step-size parameter $0 < \mu \leq 1$. The Euclidean direction set (EDS) algorithm converges to the optimal solution.

Proof: Let $B \triangleq A + A^T - R$. It is easy to verify that matrix B is a block diagonal submatrix of R, with each block size being 1×1 or 2×2. As proven in the next subsection, matrix B is symmetric and positive definite. Now, assume that λ and x are the eigenvalue and associated eigenvector of matrix $A^{-1}R$ where

$$(I - A^{-1}R)x = (1 - \lambda)x.$$

Multiplying by x^*A, taking the absolute value of both sides, and rearranging gives

$$|1 - \lambda| = \frac{|x^*Ax - x^*Rx|}{|x^*Ax|}, \tag{8}$$

where $(.)^*$ denotes the conjugate transpose of (.).

Recall that for $x \in C^N; x \neq 0$, $x^*Rx > 0$, $x^*Bx > 0$, and $\mathrm{Re}\{x^*Ax\} = x^*(\frac{A+A^T}{2})x = x^*(\frac{R+B}{2})x$. If $x^*Rx = 2a$, $x^*Bx = 2c$, then, $x^*Ax = a + c + jb$, where b is real. So, substituting a, b, c into (8), we have

$$|1 - \lambda| = \left(\frac{(a-c)^2 + b^2}{(a+c)^2 + b^2} \right)^{1/2} < 1,$$

which implies that $\rho(I - A^{-1}R) < 1$.

Since for $0 < \mu \leq 1$, $\rho(I - \mu A^{-1}R) < \rho(I - A^{-1}R)$ [6], we have $C(k) \to 0$ as $k \to \infty$ in (7), and the conclusion follows. ♯

3.2. Convergence Rate of The EDS Method

In the previous section, we have shown that $|1-\lambda| < 1$, which implies that $\lambda \neq 0$. Therefore, using similarity transformation, we may express (7) as

$$C(k) = Q(I - \mu\Lambda)^k Q^{-1} C(0),$$

where matrix Λ is a diagonal matrix containing the eigenvalues of matrix $(A^{-1}R)$, and $A^{-1}R = Q\Lambda Q^{-1}$. For convenience, define a new vector $V(k) = Q^{-1}\ C(k)$, so that

$$V(k) = (I - \mu\Lambda)^k V(0).$$

Clearly, the convergence rate of each element $v_i(k)$ in vector $V(k)$ is dependent on the associated eigenvalue λ_i of matrix $(A^{-1}R)$.

The quantity $r_i = 1 - \mu\lambda_i$ is known as the "geometric ratio". Note that when the absolute value of r_i is less than 1, the rate of convergence increases as r_i decreases.

As mentioned in [4], the overall convergence rate r cannot be expressed in a simple closed form. But fortunately, the absolute value of geometric ratio r is lower bounded. So, we indicate that the convergence performance of the EDS method is superior to the steepest descent method by showing that the lower bound $|r|_{bound}$ in the EDS method is lower than that of the steepest descent method.

Throughout the rest of the paper, $\lambda_i(.)$, $\lambda_{\max}(.)$ and $\lambda_{\min}(.)$ will denote the *ith*, the largest and the smallest eigenvalues of matrix (.), respectively.

The *ith* geometric ratio, for the steepest descent method, is $r_i = 1 - \mu\lambda_i(R)$, where $0 < \mu < \frac{2}{\lambda_{\max}(R)}$. The overall convergence rate is lower bounded as

$$|r| \geq \max\{|1 - \mu\lambda_{\max}(R)|, |1 - \mu\lambda_{\min}(R)|\}. \qquad (9)$$

The best step size μ for the convergence occurs at $\mu\lambda_{\max}(R) - 1 = 1 - \mu\lambda_{\min}(R)$, which gives $\mu = \frac{2}{\lambda_{\max}(R)+\lambda_{\min}(R)}$. Substituting it into (9), yields

$$|r| \geq \frac{\lambda_{\max}(R) - \lambda_{\min}(R)}{\lambda_{\max}(R) + \lambda_{\min}(R)}. \qquad (10)$$

So, in the steepest descent method,

$$\begin{aligned} |r|_{bound} &= \frac{\lambda_{\max}(R) - \lambda_{\min}(R)}{\lambda_{\max}(R) + \lambda_{\min}(R)} \\ &= 1 - \frac{2}{\frac{\lambda_{\max}(R)}{\lambda_{\min}(R)} + 1} \end{aligned} \qquad (11)$$

where $\frac{\lambda_{\max}(R)}{\lambda_{\min}(R)}$ equals the condition number of R.

Note that $|r|_{bound}$ decreases as the ratio $\frac{\lambda_{\max}(R)}{\lambda_{\min}(R)}$ decreases. Hence, the convergence rate increases as the ratio $\frac{\lambda_{\max}(R)}{\lambda_{\min}(R)}$ decreases.

In order to compare the bounds between two methods, let's prove the following lemma first.

<u>Lemma 1</u> Assume R is an $N \times N$ symmetric positive definite matrix, and matrix B is a block diagonal submatrix of R, with each block being square, then, matrix B is symmetric, positive definite, and

$$\lambda_{\min}(B) \geq \lambda_{\min}(R);\ \lambda_{\max}(B) \leq \lambda_{\max}(R).$$

Proof: For $i = 1, 2, \cdots, N$, the *ith* principle submatrix R_i of R is the $i \times i$ submatrix consisting of the intersection of the first i rows and columns of R. Let B_j denote the *jth* block submatrix in B, i.e. $B = \{B_1 \oplus B_2 \cdots \oplus B_L\}$. Since any principle submatrix of a symmetric positive definite matrix is symmetric positive definite and $\lambda_k(R) \leq \lambda_k(R_i) \leq \lambda_{k+N-i}(R)$ for each integer k such that $1 \leq k \leq i$ [5], it therefore, follows that the first block in B is symmetric, positive definite, and

$$\lambda_{\min}(B_1) \geq \lambda_{\min}(R);\ \lambda_{\max}(B_1) \leq \lambda_{\max}(R).$$

Recall that there exists the permutation matrices which can bring the other blocks being the first block without effect of the symmetric and positive definite properties [6], and

$$\begin{aligned} \lambda_{\min}(B) &= \min_{j=1,\cdots L} \{\lambda_{\min}(B_j)\}; \\ \lambda_{\max}(B) &= \max_{j=1,\cdots L} \{\lambda_{\max}(B_j)\}. \end{aligned}$$

We thus proved that matrix B is symmetric, positive definite, and

$$\lambda_{\min}(B) \geq \lambda_{\min}(R);\ \lambda_{\max}(B) \leq \lambda_{\max}(R). \quad ♯ \qquad (12)$$

In the EDS method, for $\mu = 1$, the overall convergence rate

$$|r| \geq \max_{i=1,2,\cdots,N} \{|1 - \lambda_i(A^{-1}R)|\},$$

and

$$|r|_{bound} = \sup_{A^{-1}Rx=\lambda x; x\neq 0} \left\{\left|1 - \frac{x^* Rx}{x^* Ax}\right|\right\}. \qquad (13)$$

Since x^*Rx is real, it is easy to see that the above supremum occurs when x^*Ax is real. Note that, if x^*Ax is real, then $x^*Ax = x^*\frac{R+B}{2}x$ and $x^*Ax - x^*Rx = x^*\frac{B-R}{2}x$. Therefore,

$$\begin{aligned} &|r|_{bound} \\ \leq\ & \sup_{x^*Ax\in R^N; x\neq 0} \left\{\frac{|x^*Rx - x^*Bx|}{x^*Rx + x^*Bx}\right\} \\ =\ & \left\{ \begin{array}{ll} 1 - \frac{2}{\sup_{x\in R^N; x\neq 0}\left\{\frac{x^*Rx}{x^*Bx}\right\}+1} & \text{if } x^*Rx \geq x^*Bx; \\ 1 - \frac{2}{\sup_{x\in R^N; x\neq 0}\left\{\frac{x^*Bx}{x^*Rx}\right\}+1} & \text{if } x^*Rx < x^*Bx; \end{array} \right\} \\ \leq\ & \left\{ \begin{array}{ll} 1 - \frac{2}{\frac{\lambda_{\max}(R)}{\lambda_{\min}(B)}+1} & \text{if } x^*Rx \geq x^*Bx; \\ 1 - \frac{2}{\frac{\lambda_{\max}(B)}{\lambda_{\min}(R)}+1} & \text{if } x^*Rx < x^*Bx; \end{array} \right\}. \end{aligned} \qquad (14)$$

That is,

$$|r|_{bound} \leq \max\left\{1 - \frac{2}{\frac{\lambda_{\max}(R)}{\lambda_{\min}(B)} + 1}, 1 - \frac{2}{\frac{\lambda_{\max}(B)}{\lambda_{\min}(R)} + 1}\right\}. \qquad (15)$$

Recall Lemma 1, and note that when the equality in (12) occurs, $|r|_{bound} = 0$. Therefore, $\frac{\lambda_{\max(R)}}{\lambda_{\min(R)}} > \frac{\lambda_{\max(R)}}{\lambda_{\min(B)}}$ and $\frac{\lambda_{\max(R)}}{\lambda_{\min(R)}} > \frac{\lambda_{\max(B)}}{\lambda_{\min(R)}}$ are true in general, which asserts that $|r|_{bound}$ in the EDS is lower than that in steepest descent method.

If $\frac{\lambda_{\max(R)}}{\lambda_{\min(R)}}$ is close to 1, with proper step size μ, both EDS and steepest descent method converge fast. However, when $\frac{\lambda_{\max(R)}}{\lambda_{\min(R)}}$ is large, the last " $\leq$ " in (14) is a very conservative step. Since matrix B is a block diagonal submatrix from R and matrix R is central majorized., $\sup_{x \in R^N; x \neq 0} \frac{|x^* Rx - x^* Bx|}{x^* Rx + x^* Bx}$ is most likely close to the ratio $\frac{\lambda_{\min}(B) - \lambda_{\min}(R)}{\lambda_{\min}(B) + \lambda_{\min}(R)}$ instead of $\frac{\lambda_{\max}(R) - \lambda_{\min}(B)}{\lambda_{\max}(R) + \lambda_{\min}(B)}$ or $\frac{\lambda_{\max}(B) - \lambda_{\min}(R)}{\lambda_{\max}(B) + \lambda_{\min}(R)}$, i.e. when $\frac{\lambda_{\max(R)}}{\lambda_{\min(R)}}$ is large,

$$
\begin{aligned}
|r|_{bound} &\leq \sup_{x \in C^N; x \neq 0} \left\{ \frac{|x^* Ax - x^* Rx|}{|x^* Ax|} \right\} \\
&\approx \frac{\lambda_{\min}(B) - \lambda_{\min}(R)}{\lambda_{\min}(B) + \lambda_{\min}(R)}. \qquad (16)
\end{aligned}
$$

This implies that the convergence rate of EDS is limited by the condition number of matrix R somewhat too. In addition, (16) also exploits the fact that the double direction search seems to converge more rapidly than the one direction search because $\lambda_{\min}(B)$ is smaller and more close to $\lambda_{\min}(R)$ in EDS2.

3.3. Example

A simple example of a single-input adaptive linear combiner with two weights is shown in Fig. 1. The input and desired signals are sampled sinusoids at the same frequency, with $M = 6$ samples per cycle. This is the same example as in [3]. The input correlation matrix R and the correlation vector P are calculated as follows:

$$
\begin{aligned}
R &= E\left\{ \begin{bmatrix} x_k^2 & x_k x_{k-1} \\ x_{k-1} x_k & x_{k-1}^2 \end{bmatrix} \right\} = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{bmatrix}; \\
P &= E\{ [\ d_k x_k \quad d_k x_{k-1}\ ] \}^T = \begin{bmatrix} 0 & -\frac{\sqrt{3}}{2} \end{bmatrix}.
\end{aligned}
$$

Solving $\det(\lambda I - R) = 0$, gives that $\lambda_{\min}(R) = 0.25$, $\lambda_{\max}(R) = 0.75$. From (11), $|r|_{bound} = \frac{1}{2}$.

In EDS1, $A = \begin{bmatrix} 0.5 & 0 \\ 0.25 & 0.5 \end{bmatrix}$; $B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$.

$\lambda_{\min}(A^{-1}R) = 0.75$, $\lambda_{\max}(A^{-1}R) = 1$, and $|r|_{boun} = \frac{1}{7}$ based on (13). From (15) and $\lambda_{\min}(B) = \lambda_{\max}(B) = 0.5$, we conclude that $|r| \geq \frac{1}{3}$ minimally.

Fig.2 plots the instantaneous squared error at each adaptive iteration starting with $W(0) = 0$. This shows that for $\mu = 1$ the EDS adaptation needs about 25 iterations to converge and 50 for the steepest descent method with $\mu = \frac{2}{\lambda_{\max}(R) + \lambda_{\min}(R)}$. No doubt, the EDS2 method here is equivalent to Newton's method and can converge in one step.

4. CONCLUDING REMARKS

The recently reported Euclidean Direction Set (EDS) algorithm for adaptive filtering has been considered. It is shown that this algorithm is really a gradient based adaptive process. This formulation makes it suitable for mathematical analysis. It is then proven that the EDS algorithm converges to the true parameters. The convergence rate of this algorithm is also mathematically shown to be superior than the LMS algorithm.

5. REFERENCES

[1] M.Q. Chen, T.Bose and G.F.Xu "A direction set based algorithm for adaptive Filtering", IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Aug. 1997.

[2] T.Bose, M.Q. Chen and G.F.Xu "A Fast Adaptive Algorithm for Spectral Estimation", IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Aug. 1997

[3] Bernard Widerow & Samuel D. Stearns, Adaptive Signal Processing, Prentice Hall, 1985.

[4] S. Haykin, Adaptive Filter Theory, Prentice Hall, Englewood Cliffs, NJ, 1996.

[5] Roger A. Horn & Charles R. Johnson, Matrix Analysis, Combridge University Press 1985.

[6] Rajendra Bhatia, Matrix Analysis, Springer-Verlag New York, 1997.

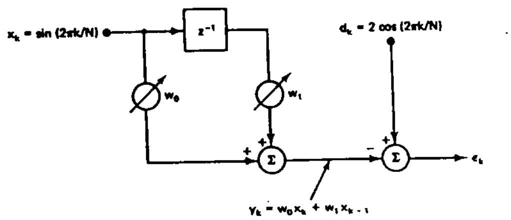

Fig. 1 Example of an adaptive linear combiner with two weights

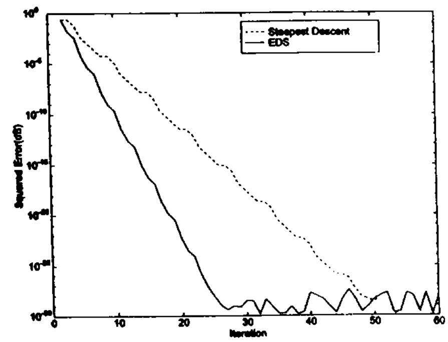

Fig. 2 The instantaneous square error for each iteration