Signal Processing 89 (2009) 894–900

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

A generalized data windowing scheme for adaptive conjugate gradient algorithms

Shengkui Zhao [a,*], Zhihong Man [b], Suiyang Khoo [b]

[a] School of Computer Engineering, Nanyang Technological University (NTU), Nanyang Avenue, Singapore 639798, Singapore
[b] School of Engineering, Monash University (Malaysia), Jalan Lagoon Selatan, Bandar Sunway, Petaling Jaya 46150, Malaysia

ARTICLE INFO

Article history:
Received 20 February 2008
Received in revised form
12 November 2008
Accepted 13 November 2008
Available online 20 November 2008

Keywords:
Conjugate gradient method
Adaptive filtering
LMS algorithm
Recursive least squares

ABSTRACT

The performance of the modified adaptive conjugate gradient (CG) algorithms based on the iterative CG method for adaptive filtering is highly related to the ways of estimating the correlation matrix and the cross-correlation vector. The existing approaches of implementing the CG algorithms using the data windows of exponential form or sliding form result in either loss of convergence or increase in misadjustment. This paper presents and analyzes a new approach to the implementation of the CG algorithms for adaptive filtering by using a generalized data windowing scheme. For the new modified CG algorithms, we show that the convergence speed is accelerated, the misadjustment and tracking capability comparable to those of the recursive least squares (RLS) algorithm are achieved. Computer simulations demonstrated in the framework of linear system modeling problem show the improvements of the new modifications.

1. Introduction

Adaptive conjugate gradient (CG) algorithms based on the CG method have been successfully introduced to solve the adaptive filtering problems. The advantages are that they have convergence properties superior to those of least mean square (LMS) algorithms and computational cost less than that of the classic recursive least squares (RLS) algorithm. Moreover, the instability problems existing in the RLS algorithm are not likely to occur in the CG algorithm [1,2]. In the reported modified adaptive CG algorithms [1–6], two data windowing schemes, i.e., the finite sliding data windowing scheme and the exponentially decaying data windowing scheme, of estimating the correlation matrix and the cross-correlation vector have been applied. However, the analysis and simulations presented in [1,2] show that the modified CG algorithms, which are implemented using the finite sliding data windowing scheme and run several iterations per data update, have a convergence behavior and misadjustment which depend on the length of data window. A small window size produces slow convergence, whereas a large window size introduces high computational cost and high misadjustment. On the other hand, the modified CG algorithms, which are implemented using the exponentially decaying data windowing scheme and run one iteration per coefficient and data update, have the convergence dependent of the input eigenvalue spread. When the input with large eigenvalue spread is applied, the convergence is considerably slow.

To improve the convergence performance as well as misadjustment of the modified CG algorithms, this paper presents and analyzes a new approach of implementing the CG algorithm using a generalized data windowing scheme which combines the features of the finite sliding data windowing scheme and the exponentially decaying data windowing scheme. In Section 2, we first introduce the generalized data windowing scheme, and analyze the computational complexity and its advantages for CG

* Corresponding author.
E-mail addresses: zhao0024@ntu.edu.sg (S. Zhao), man.zhihong@eng.monash.edu.my (Z. Man), khooyang@yahoo.com (S. Khoo).

0165-1684/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2008.11.007

algorithms. The numerical problems and the convergence behavior of the modified CG algorithms are then studied. Section 3 presents computer simulations to compare the convergence speed, misadjustment and tracking capabilities of the new modified CG algorithms with those of the existing approaches as well as the RLS algorithm. Finally, Section 4 concludes this paper.

2. Generalized data windowing scheme and accelerated CG algorithms

Consider a minimization problem for the following quadratic performance function:

$$J(\mathbf{w}) = \tfrac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{R}\mathbf{w} - \mathbf{b}^{\mathrm{T}}\mathbf{w}, \tag{1}$$

where $\mathbf{R}$ is $N \times N$ square matrix and positive definite, $\mathbf{b}$ and $\mathbf{w}$ are vectors of dimension N. Solving for the vector $\mathbf{w}_o$ that minimizes the quadratic performance function (1) is equivalent to solving the linear equation $\mathbf{R}\mathbf{w}_o = \mathbf{b}$. It has been shown that the CG algorithm can be used to solve this linear equation iteratively and efficiently in [7,8]. The algorithm uses the following weight update equation:

$$\mathbf{w}(k) = \mathbf{w}(k-1) + \alpha(k)\mathbf{p}(k), \tag{2}$$

where $\mathbf{w}(k)$ is the kth estimate of $\mathbf{w}_o$, and the step size $\alpha(k)$ is chosen to minimize the performance function $J(\mathbf{w}(k-1) + \alpha(k)\mathbf{p}(k))$ along the update direction, and the current direction vector $\mathbf{p}(k)$ is to be linearly independent and $\mathbf{R}$-conjugate (i.e., $\mathbf{p}^{\mathrm{T}}(i))\mathbf{R}\mathbf{p}^{\mathrm{T}}(j) = 0$, $i \neq j$) to all the previous direction vectors $\mathbf{p}(1), \mathbf{p}(2), \ldots, \mathbf{p}(k-1)$. In this algorithm, the estimate $\mathbf{w}(k)$ finally converges to the optimal solution $\mathbf{w}_o$ in finite number of iterations.

2.1. A generalized data windowing scheme

When implementing the adaptive CG algorithm for solving adaptive filtering problems where the input vector of dimension N denoted by $\mathbf{x}(k)$ and the desired output denoted by $d(k)$ are assumed to be stationary processes, the correlation matrix $\mathbf{R} \triangleq E[\mathbf{x}(k)\mathbf{x}^{\mathrm{T}}(k)]$ and the cross-correlation vector $\mathbf{b} \triangleq E[d(k)\mathbf{x}(k)]$ are usually not available. The two data windowing schemes that have been used to estimate $\mathbf{R}$ and $\mathbf{b}$ are summarized as follows [2].

(a) *Finite sliding data windowing scheme*:

$$\mathbf{R}(k) = \frac{1}{M}\sum_{j=k-M+1}^{k} \mathbf{x}(j)\mathbf{x}^{\mathrm{T}}(j), \tag{3}$$

$$\mathbf{b}(k) = \frac{1}{M}\sum_{j=k-M+1}^{k} d(j)\mathbf{x}(j). \tag{4}$$

(b) *Exponentially decaying data windowing scheme*:

$$\mathbf{R}(k) = \lambda\mathbf{R}(k-1) + \mathbf{x}(k)\mathbf{x}^{\mathrm{T}}(k), \tag{5}$$

$$\mathbf{b}(k) = \lambda\mathbf{b}(k-1) + d(k)\mathbf{x}(k), \tag{6}$$

where λ is the forgetting factor.

Notice that in the first approach the time ensemble average over a window of finite length M is evaluated. The accuracy of the estimated $\mathbf{R}(k)$ and $\mathbf{b}(k)$ is highly related to the data window size M. When small window size of $M < N$ is used, the estimates of $\mathbf{R}$ and $\mathbf{b}$ have a high variance and nonfull rank. Since the modified CG algorithms using this approach tend to converge to the form $\mathbf{R}(k)^{-1}\mathbf{b}(k)$ for each data update [2], the misadjustment is usually very high. Moreover, since the estimate of the correlation matrix is not full rank, the direction vector that is used to update the filter coefficients cannot be projected to all N directions, thus the modified CG algorithms show slow convergence. On the other hand, though large window size could be used to improve the overall performance, the computational cost will increase dramatically. In the second approach, the exponential form provides a long memory of the estimates, therefore the variances of the estimates, $\mathbf{R}(k)$ and $\mathbf{b}(k)$, eventually converge to zero. As a result, a misadjustment comparable to that of the RLS algorithm is achieved in the modified CG algorithms as seen in [2,5]. The issue that should be pointed out is that at the initial stages $(k < N)$, the estimated correlation matrix is not full rank since the updating matrix $\mathbf{x}(k)\mathbf{x}^{\mathrm{T}}(k)$ is always with rank one. In this case, the exponential decay due to the forgetting factor causes slow initial convergence for the estimation of the correlation matrix, which leads to a slow initial convergence in the modified CG algorithms implemented by this approach.

The generalized data windowing scheme which was previously reported in [9] combines the features of the above two approaches and is shown below:

(c) *Generalized data windowing scheme*:

$$\mathbf{R}(k) = \lambda\mathbf{R}(k-1) + \frac{1}{M}\sum_{j=0}^{M-1} \mathbf{x}(k-j)\mathbf{x}^{\mathrm{T}}(k-j), \tag{7}$$

$$\mathbf{b}(k) = \lambda\mathbf{b}(k-1) + \frac{1}{M}\sum_{j=0}^{M-1} d(k-j)\mathbf{x}(k-j), \tag{8}$$

where λ is again the forgetting factor and M is the window size. As can be seen that the generalized data windowing scheme in (7) and (8) combines the exponentially decaying form and finite sliding form. In general, the exponential form accumulates all the past data for the estimation, thereby decreasing variance of the estimates. Compared to the finite sliding data windowing scheme, this new scheme improves the misadjustment. On the other hand, the finite sliding form is used for the update, resulting in the rank of the updating matrix equal to N when $M \geqslant N$ and $k \geqslant N$. As a result, this scheme changes the weighting of the updating form so that the decay due to the forgetting factor is not exponentially decreasing any more. Though at the initial stages, the estimated correlation matrix is also not full rank, the initial convergence of the estimation increases due to $\mathbf{x}(k-j)\mathbf{x}^{\mathrm{T}}(k-j)$, $j = 1, 2, \ldots, M-1$ are present in $\lambda\mathbf{R}(k-1)$. This improves the initial convergence of the modified CG algorithms.

Compared to the existing data window schemes, the new scheme increases the computational cost. However, the computational cost can be reduced to a reasonable level