CONJUGATE GRADIENT ALGORITHM WITH DATA SELECTIVE UPDATING

José A. Apolinário Jr.,[1] *Stefan Werner,*[2] *and Paulo S. R. Diniz*[3]

[1]Instituto Militar de Engenharia
Depto. de Engenharia Elétrica
Praça Gal. Tibúrcio, 80
22.290-270 Rio de Janeiro, RJ, Brazil
apolin@ieee.org

[2]Helsinki University of Technology
Signal Processing Laboratory
P.O. Box 3000
FIN-02015 HUT, Finland
Stefan.Werner@hut.fi

[3]COPPE/Univ. Federal do Rio de Janeiro
Programa de Engenharia Elétrica
P.O. Box 68504
21.945-970 Rio de Janeiro, RJ, Brazil
diniz@lps.ufrj.br

ABSTRACT

This paper applies data selective updating to the Modified Conjugate Gradient algorithm. In search for a new conjugate-gradient-like filtering algorithm, two different approaches are developed: the first one results in the recently proposed set-membership affine projection (SM-AP) algorithm and the second one reduces the computational requirements of the modified congujate gradient algorithm while keeping approximately the same good results in terms of convergence speed and misadjustment. Simulation results for a system identification experiment show the claimed performance with a considerable reduced number of updates.

1. INTRODUCTION

Fast convergence and low complexity as well as small misadjustment are desired characteristics of adaptive filters always aimed by design engineers specially when the input signal is highly correlated. The RLS family of adaptive filter algorithms is known to have faster convergence than the LMS-like algorithms but the cost in terms of computational complexity are sometimes too high [1]. The Conjugate Gradient (CG) method applied to adaptive filtering as in [2] can attain similar fast convergence and small misadjustment compared to the RLS algorithm without the need of performing matrix inversion. Inspired by recent developments in set-membership filtering (SMF), this paper investigates the possibility of lowering the computation requirements of the modified conjugate gradient algorithm [2] by restricting the number of updates or, in other words, using a data selective scheme.

In set-membership filtering (SMF) [3] an upper bound of the output estimation error is specified. The resulting adaptation algorithms are data-selective which in turn can reduce the computational complexity of the algorithms considerably. Furthermore, the sparse updating also results in a low misadjustment because the algorithms does not utilize the input data if it does not imply innovation. The set-membership NLMS (SM-NLMS) algorithm proposed in [3] was shown to achieve fast convergence and low misadjustment, and its sparse updating and low computational complexity per update makes it attractive in various applications. The set-membership affine projection (SM-AP) algorithm [4] generalizes the ideas of the SM-NLMS algorithm to improve the performance for correlated inputs.

In this paper we recast the ideas of [4] in the framework of the CG algorithm. This paper is organized as follows. Section 2 reviews the basic concepts of set-membership filtering. In Section 3, the contrained CG (CCG) [5] applied to certain contraints is shown to yield the SM-AP algorithm. Section 4 combines data selectivity with the modified CG (MCG) [2] algorithm in an attempt to reduce the computational complexity. Section 5, contains the simulations and Section 6, the concluding remarks.

2. SET-MEMBERSHIP FILTERING

In set-membership filtering, the filter is designed to upper bound the output estimation error $d_k - \mathbf{w}_k^T\mathbf{x}$ with a predefined threshold γ. Therefore, all vectors satifying the bound contraint are considered feasible. Define the *feasibility set* as the set of all filter vectors satifying the error contraint for all possible input-desired data pairs $(d,\ \mathbf{x})$, i.e.,

$$\Theta = \bigcap_{(\mathbf{x},d)\in\mathcal{S}} \{\mathbf{w} \in \mathcal{R}^{N+1} : |d - \mathbf{w}^T\mathbf{x}| \leq \gamma\} \qquad (1)$$

Adaptive solutions try to find estimates belonging to this feasibility set. In many application it is impossible to predict all possible data pairs and, therefore, adaptive methods work with the membership set constructed from the observed data pairs,

$$\psi_k = \bigcap_{i=1}^{k} \mathcal{H}_i \qquad (2)$$

where $\mathcal{H}_k = \{\mathbf{w} \in \mathcal{R}^N : |d_k - \mathbf{w}^T\mathbf{x}_k| \leq \gamma\}$ is the *constraint set* formed by the input data pair at time instant k.

Note that the feasibility set is included in the membership set and if all possible data pairs are traversed up to time instant k, the membership becomes equal to the feasibility set. The simplest adaptive approaches computes a point estimate provided a subset of the mebershipset, e.g., the information provided by $\mathcal{H}_k$ like in the SM-NLMS algorithm or by utilizing L past contraint sets like in the SM-AP algorithm. Other approaches try to outer bound the membership set with ellipsiods leading to the family of optimal bounding ellipsoid (OBE) algorithms [6].

3. CONSTRAINED CONJUGATE GRADIENT ALGORITHM

For this first formulation, we partition the *membership set* ψ_k as $\psi_k = \psi_k^{k-L} \bigcap \psi_k^L$ where ψ_k^L corresponds to the intersection of the L last constraint sets, i.e.,

$$\psi_k^L = \bigcap_{i=k-L+1}^{k} \mathcal{H}_i \tag{3}$$

Our goal is the derivation of an algorithm with Conjugate Gradient update such that the updated coefficient vector $\mathbf{w}_k$ belongs to the last L constraint-sets, i.e., $\mathbf{w}_k \in \psi_k^L$. For this, we define $\mathcal{S}_{k-i+1}$ as the hyperplane which contains all vectors $\mathbf{w}$ such that $d_{k-i+1} - \mathbf{w}^T\mathbf{x}_{k-i+1} = g_{k-i+1}$ for $i = 1, \cdots, L$. The parameters g_{k-i+1} should be any satisfying the bound constraint or $|g_{k-i+1}| < \gamma$ such that $\mathcal{S}_{k-i+1} \in \mathcal{H}_{k-i+1}$. For the sake of simplicity, we shall use these parameters equal to zero, see [4] for a discussion on particular choices.

When the Conjugate Gradient algorithm is used to solve $\mathbf{Rw} = \mathbf{p}$ ($\mathbf{R}$ being the input signal autocorrelation matrix and $\mathbf{p}$ the cross-correlation between the input vector and the reference signal), it is also minimizing $\frac{1}{2}\mathbf{w}^T\mathbf{Rw} - \mathbf{p}^T\mathbf{w}$. In our case, we can impose an aditional restriction whenever $\mathbf{w}_{k-1} \notin \psi_k^L$ and state the following optimization criterion

$$\begin{aligned} &\min \frac{1}{2}\mathbf{w}_k^T\mathbf{R}\mathbf{w}_k - \mathbf{p}^T\mathbf{w}_k \\ &\text{subject to: } \mathbf{X}_k^T\mathbf{w}_k = \mathbf{d}_k \end{aligned} \tag{4}$$

where

$$\begin{aligned} \mathbf{d}_k &= [d_k \ \ d_{k-1} \ \ \cdots \ \ d_{k-L+1}]^T \\ \mathbf{X}_k &= [\mathbf{x}_k \ \ \mathbf{x}_{k-1} \ \ \cdots \ \ \mathbf{x}_{k-L+1}]^T \end{aligned} \tag{5}$$

and

$$\mathbf{x}_k = [x_k \ \ x_{k-1} \ \ \cdots \ \ x_{k-N+1}]^T \tag{6}$$

N being the filter order.

The minimization of this constrained cost function using the Conugate Gradient metfod may correspond to the Constrained Conjugate Gradient (CCG) algorithm [5] if we make the constraint $\mathbf{C}^T\mathbf{w}_k = \mathbf{f}$ such that the constraint matrix $\mathbf{C}$ is time variant and corresponding to $\mathbf{X}_k$. Moreover, the gain vector $\mathbf{f}$ should also be time variant and corresponding to $\mathbf{d}_k$. Table 1 shows the CCG algorithm imposing this condition.

Table 1: The Constrained Conjugate Gradient Algorithm.

Initialization:
- λ, η with $(\lambda - 0.5) \le \eta \le \lambda$
- L is the number of constraint sets
- δ small number
- $\mathbf{w}_0 = zeros(N+1, 1)$
- $\bar{\mathbf{R}}_0 = \mathbf{I}$
- $\mathbf{g}_0 = \mathbf{c}_0 = zeros(N+1, 1)$

Running the algorithm:
for each k
{
- $\mathbf{C} = \mathbf{X}_k$
- $\mathbf{f} = \mathbf{d}_k$
- $\mathbf{P} = \mathbf{I} - \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T$
- $\mathbf{F} = \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{f}$
- $\bar{\mathbf{x}}_k = \mathbf{P}\mathbf{x}_k$
- $\bar{\mathbf{R}}_k = \lambda\bar{\mathbf{R}}_{k-1} + \bar{\mathbf{x}}_k\bar{\mathbf{x}}_k^T$
- $y_k = \mathbf{w}_{k-1}^T\mathbf{x}_k$
- $e_k = d_k - y_k$
- $\alpha_k = \eta\frac{\mathbf{c}_k^T\mathbf{g}_{k-1}}{[\mathbf{c}_k^T\bar{\mathbf{R}}_k\mathbf{c}_k + \delta]}$
- $\mathbf{g}_k = \lambda\mathbf{g}_{k-1} - \alpha_k\bar{\mathbf{R}}_k\mathbf{c}_k - \bar{\mathbf{x}}_k e_k$
- $\mathbf{w}_k = \mathbf{P}\mathbf{w}_{k-1} + \mathbf{F} - \alpha_k\mathbf{c}_k$
- $\beta_k = \frac{[\mathbf{g}_k - \mathbf{g}_{k-1}]^T\mathbf{g}_k}{[\mathbf{g}_{k-1}^T\mathbf{g}_{k-1} + \delta]}$
- $\mathbf{c}_{k+1} = \mathbf{g}_k + \beta_k\mathbf{c}_k$

}

From the above table, we see that the CCG algorithm is updated according to $\mathbf{w}_k = \mathbf{P}\mathbf{w}_{k-1} + \mathbf{F} - \alpha_k\mathbf{c}_k$ where for this particular application we have $\mathbf{P} = \mathbf{I} - \mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^{-1}\mathbf{X}_k^T$ and $\mathbf{F} = \mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^{-1}\mathbf{d}_k$.

Nevertheless, $\mathbf{P}$ is actually the projection matrix onto to the subspace spanned by the constraint matrix which in our case corresponds to $\mathbf{X}_k$ defined in (5). It means that $\bar{\mathbf{x}}_k = \mathbf{P}\mathbf{x}_k$ is a null vector because $\mathbf{x}_k$ is the first column of $\mathbf{X}_k$. Since the computation of the last right-sided term of the coefficient vector update ($-\alpha_k\mathbf{c}_k$) is based on $\bar{\mathbf{x}}_k$, its value will be zero (or close to zero due to δ).

By considering $\alpha_k \mathbf{c}_k = 0$ and replacing $\mathbf{P}$ and $\mathbf{F}$, the coefficient vector update becomes

$$\begin{aligned}\mathbf{w}_k =& \mathbf{P}\mathbf{w}_{k-1} + \mathbf{F} \\ =& [\mathbf{I} - \mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^{-1}\mathbf{X}_k^T]\mathbf{w}_{k-1} \\ &+ \mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^{-1}\mathbf{d}_k \\ =& \mathbf{w}_{k-1} + \mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^{-1}\mathbf{e}_k \end{aligned} \quad (7)$$

where $\mathbf{e}_k = \mathbf{d}_k - \mathbf{X}_k^T\mathbf{w}_{k-1}$.

The last equation corresponds to the affine projection algorithm update and the corresponding set-membership algorithm is actually the SM-AP of [4].

4. THE POINT-WISE APPROACH

In this second formulation, we are interested in testing if the previous coefficient vector belongs to the membership subset ψ_k^L and make a decision of update. This procedure shall decrease the number of updates and consequently the computational burden of the Conjugate Gradient algorithm. To verify if $\mathbf{w}_{k-1} \in \psi_k^L$ we need to test if all elements of vector $\mathbf{e}_k$ defined above have absolute values lower than γ. Tabel 2 shows this Selective Updating Conjugate Gradient algorithm. As mentioned before, this version is based on the Modified Conjugate Gradient (MCG) algorithm presented in [2] which uses a degenerated scheme in order to have only one iteration per coefficient-vector update. Moreover, it uses an exponentially decaying data window to estimate the input signal autocorrelation matrix. Although, the MCG does not in itself guarantee that the updates will end up in the subset ψ_k^L like the CCG in the previous section, it will only perform updates whenever $\mathbf{w}_{k-1} \notin \psi_k^L$ resulting in a lower computational complexity than the conventional MCG.

Table 2: The MCG Algorithm with selective updating.

Initialization:
  λ, η with $(\lambda - 0.5) \leq \eta \leq \lambda$
  L is the number of constraint sets
  γ upper bound of the estimation error
  δ small number
  $\mathbf{g}_0 = \mathbf{c}_0 = \delta\cdot\text{ones}(N+1, 1)$
  $\mathbf{w}_0 = \text{zeros}(N+1, 1)$
  $\mathbf{R}_0 = \mathbf{I}$
Running the algorithm:
for k=1:k_{MAX}
{
  $\mathbf{e}_k = \mathbf{d}_k - \mathbf{X}_k^T\mathbf{w}_{k-1}$
  if $|e_k(i)| < \gamma$ for all $i = 1 : L$
  {
  $\mathbf{w}_k = \mathbf{w}_{k-1}$
  }
  else
  {
    $\mathbf{R}_k = \lambda\mathbf{R}_{k-1} + \mathbf{x}_k\mathbf{x}_k^T$
    $\alpha_k = \eta\frac{\mathbf{c}_k^T\mathbf{g}_{k-1}}{[\mathbf{c}_k^T\mathbf{R}_k\mathbf{c}_k + \delta]}$
    $\mathbf{g}_k = \lambda\mathbf{g}_{k-1} - \alpha_k\mathbf{R}_k\mathbf{c}_k + \mathbf{x}_k e_k(1)$
    $\mathbf{w}_k = \mathbf{w}_{k-1} + \alpha_k\mathbf{c}_k$
    $\beta_k = \frac{[\mathbf{g}_k - \mathbf{g}_{k-1}]^T\mathbf{g}_k}{[\mathbf{g}_{k-1}^T\mathbf{g}_{k-1} + \delta]}$
    $\mathbf{c}_{k+1} = \mathbf{g}_k + \beta_k\mathbf{c}_k$
  }
}

5. SIMULATION RESULTS

In this section we consider a system identification setup where the input signal consists of white Gaussian noise filtered through an IIR filter defined by $y_k = x_k - 1.1y_{k-1} - 0.99y_{k-2}$. The additional noise was such that its variance corresponds to $\sigma_n^2 = 10^{-6}$ and the unknown plant has 20 coefficients ($N = 19$). We also used $\lambda = 0.99$, $\eta = 0.8$, and $\delta = 10^{-5}$. The bound on the estimation error was chosen to $\gamma = \sqrt{5\sigma_n^2}$.

For this experiment we have used the number of constraint sets equal to $L = 3$. Fig. 1 depicts the learning curve obtained from an average of 1000 independent runs. We observe that the performance of the proposed algorithm is quite similar to the original Conjugate Gradient algorithm. Moreover, for comparison reasons, we present the results for the NLMS and RLS algorithms.

Fig. 2 shows the norm of the coefficient-error vector $(\mathbf{w}_k - \mathbf{w}_{\text{Opt}})$ for the same experiment. It is worth mentioning that the iterations without updating of the coefficient vector corresponds, for this experiment, to 39.34% of the total number of iterations. This value could increase by increasing the γ at the expense of a higher misadjustment or with more practical situations where we have lower SNR and, therefore, can allow larger γ. As an example, we have simulated the same experiment for $\sigma_n^2 = 10^{-3}$ and obtained near 80% of non-updating iterations with a slight decrease in performance.

6. CONCLUSIONS

In this paper, data selective conjugate-gradient algorithms were discussed. It was shown that the constrained conjugate gradient algorithm in a set-membership filtering framework resulted in the recently proposed set-membership affine projection algorithm (SM-AP). Furthermore, data selectivity was applied to the modified conjugate gradient algorithm

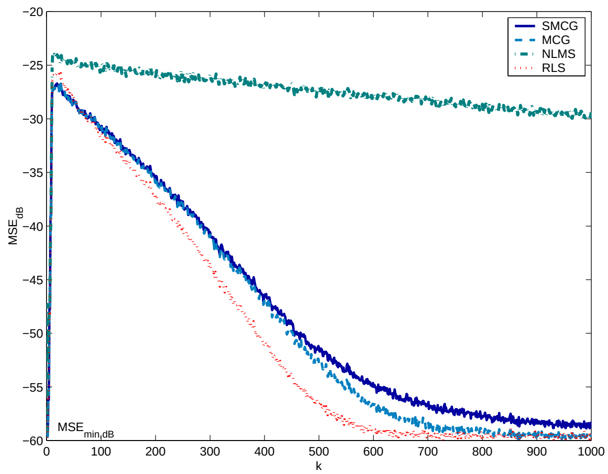

Figure 1: Learning curves for the data selective MCG algorithm with $L = 3$ ($\gamma = \sqrt{5\sigma_n^2}$), the MCG algorithm ($\eta = 0.8$), the RLS algorithm ($\lambda = 0.99$), $\sigma_n^2 = 10^{-6}$, and colored input signal.

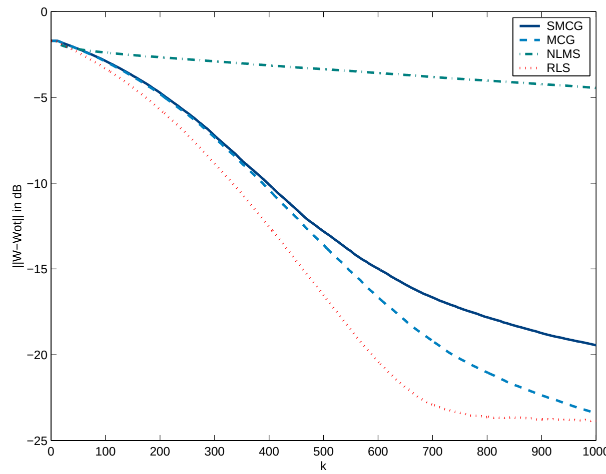

Figure 2: Norm of the coefficient-error vector for the data selective MCG algorithm with $L = 3$ ($\gamma = \sqrt{5\sigma_n^2}$), the MCG algorithm ($\eta = 0.8$), the RLS algorithm ($\lambda = 0.99$), $\sigma_n^2 = 10^{-6}$, and colored input signal.

(MCG) in an attempt to reduce its the computational complexity. Simulations confirmed that the proposed data selective MCG (SMCG) algorithm indeed presents good performance (similar to the original MCG algorithm) with a considerable decrease in the number of updates, consequently, lowering the computational load.

7. REFERENCES

[1] P. S. R. Diniz, *Adaptive Filtering — Algorithms and Practical Implementation*. Boston, MA: Kluwer Academic Press, 1997.

[2] P. S. Chang and A. N. Willson, Jr., “Adaptive filtering using modified conjugate gradient,” Proc. 38th Midwest Symp. Circuits Syst., pp. 243-246, Rio de Janeiro, Brazil, August 1995.

[3] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y.-F. Huang, “Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size,” IEEE Signal Processing Letters, vol. 5, pp. 111-114, May 1998.

[4] S. Werner and P. S. R. Diniz, “Set-membership affine projection algorithm,” sumitted for publication to IEEE Signal Processing Letters, 2000.

[5] J. A. Apolinário Jr., M. L. R. de Campos, and C. P. Bernal, “The constrained conjugate-gradient algorithm,” accepted for publication, IEEE Signal Processing Letters, vol. 7, December 2000.

[6] E. Fogel, and Y. F. Huang, “On the value of information in system identification - bounded noise case,” *Automatica*, vol. 18, pp. 229–238, March 1982.