The Constrained Conjugate Gradient Algorithm

J. A. Apolinário, Jr., *Member, IEEE*, M. L. R. de Campos, *Member, IEEE*, and C. P. Bernal O.

***Abstract*—Based on the condition for equivalence between linearly constrained minimum-variance (LCMV) filters and their generalized sidelobe canceler (GSC) implementations, we derive the new constrained conjugate gradient (CCG) algorithm. We discuss the use of orthogonal and nonorthogonal blocking matrices for the GSC structure and how the choice of this matrix may affect the relationship with the LCMV counterpart. The newly derived algorithm was tested in a computer experiment for adaptive multiuser detection and showed excellent results.**

***Index Terms*—Conjugate gradient algorithms, constrained adaptive filtering.**

I. Introduction

LINEARLY constrained adaptive filters have been used in many applications including adaptive beamforming with sensor arrays and blind adaptive interference cancellation in multiuser mobile communication systems. The constrained version of the least mean square (LMS) algorithm (CLMS) was proposed in [1] for the minimization of the output-error energy of a finite impulse response (FIR) filter subject to a set of known linear constraints, i.e., $\min_{\mathbf{w}} E[e^2]$ subject to $\mathbf{C}^T\mathbf{w} = \mathbf{f}$, where $\mathbf{w}$ is the length M coefficient vector, e is the filter output error, $\mathbf{C}$ is the $M \times p$ constraint matrix, and $\mathbf{f}$ is the length p gain vector. In [2], an alternative structure was presented whereby only a smaller set of coefficients are updated, which are confined to the subspace orthogonal to the space spanned by the constraint matrix $\mathbf{C}$. This structure, known as the generalized sidelobe canceler (GSC), is able to transform the linearly constrained minimization problem into an unconstrained minimization problem, and therefore can accommodate virtually any adaptation algorithm. Although the constrained algorithm and its GSC implementation are assumed to present identical steady-state performance [2] in a stationary environment, different choices of the blocking matrix $\mathbf{B}$ such that $\mathbf{B}^T\mathbf{C} = \mathbf{0}$ leads to different results. Moreover, this matrix determines the computational complexity of the adaptation algorithm implemented in the GSC structure. This paper revisits the condition of equivalence between a constrained adaptive filter and its GSC counterpart and uses this condition to introduce a new constrained algorithm, the constrained conjugate gradient (CCG) algorithm.

Manuscript received June 20, 2000. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. G. Ramponi.

J. A. Apolinário, Jr. and C. P. Bernal O. are with the Facultad de Ingeniería Electrónica, Escuela Politécnica del Ejército, Sangolquí, Ecuador (e-mail: apolin@ieee.org).

M. L. R. de Campos is with the Programa de Engenharia Elétrica, COPPE/Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21.945-970 Brazil (campos@lps.ufrj.br).

Publisher Item Identifier S 1070-9908(00)10183-X.

II. Preliminaries

The CLMS solution to the linearly constrained minimum-variance (LCMV) problem is given by [1]

$$\mathbf{w}(k+1) = \mathbf{P}\mathbf{w}(k) - \mu y(k)\mathbf{P}\mathbf{x}(k) + \mathbf{F} \tag{1}$$

where

$\mathbf{F}$ $= \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{f}$;

$\mathbf{P} = \mathbf{I} - \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T$ projection matrix onto the subspace orthogonal to the subspace spanned by the constraint matrix, and the output signal;

$y(k)$ $= \mathbf{w}^T(k)\mathbf{x}(k)$, output signal.

$\mathbf{x}(k)$ is the input-signal vector containing present and past input-signal samples $[x(k)\, x(k-1)\, \cdots\, x(k-M+1)]^T$. We recall the fact that although $\mathbf{P}\mathbf{w}(k) + \mathbf{F}$ corresponds to $\mathbf{w}(k)$ in infinite precision, the computation as in (1) is necessary in a limited-precision-arithmetic machine in order to avoid any drift from the constraint plane [1].

The GSC decomposes the coefficient vector by using a transformation matrix given by $\mathbf{T} = [\mathbf{C} \vdots \mathbf{B}]$ where $\mathbf{B}$ is called blocking matrix, and it spans the null space of the constraint matrix $\mathbf{C}$. The GSC-transformed coefficient vector in $\mathbf{w}(k) = \mathbf{T}\overline{\mathbf{w}}(k)$ is partitioned as $\overline{\mathbf{w}}(k) = [\overline{\mathbf{w}}_U^T \vdots -\overline{\mathbf{w}}_L^T]^T$, where the upper part is constant and chosen such that $\mathbf{C}\overline{\mathbf{w}}_U$ corresponds to $\mathbf{F} = \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{f}$, and $-\overline{\mathbf{w}}_L = \mathbf{w}_{\text{GSC}}(k)$ is updated according to an unconstrained adaptive filter such that the overall coefficient vector corresponds to $\mathbf{w}(k) = \mathbf{F} - \mathbf{B}\mathbf{w}_{\text{GSC}}(k)$.

The inverse of the GSC transformation matrix (guaranteed by linearly independent columns of $\mathbf{B}$ and $\mathbf{C}$, and by $\mathbf{B}^T\mathbf{C} = \mathbf{0}$ [3]) can be partitioned as $\mathbf{T}^{-1} = [\mathbf{A}_1^T \vdots \mathbf{A}_2^T]^T$ where $\mathbf{A}_1 = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T$ and $\mathbf{A}_2 = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$.

By replacing $\mathbf{A}_1$ and $\mathbf{A}_2$ in $\mathbf{T}^{-1}$ and then in $\mathbf{T}\mathbf{T}^{-1} = \mathbf{I}$, we find another expression for the projection matrix $\mathbf{P}$, as obtained in [4]

$$\mathbf{P} = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T. \tag{2}$$

III. Equivalence Condition Revisited

In this section, we obtain the CLMS algorithm from its GSC implementation in order to find under which circumstances they are equivalent in infinite precision. The GSC coefficient-vector update equation using the LMS algorithm relates to the coefficient-vector update equation for the constrained LMS algorithm according to

$$\begin{aligned}\mathbf{w}(k+1) &= \mathbf{F} - \mathbf{B}[\mathbf{w}_{\text{GSC}}(k+1)] \\ &= \mathbf{F} - \mathbf{B}[\mathbf{w}_{\text{GSC}}(k) + \mu e_{\text{GSC}}(k)\mathbf{x}_{\text{GSC}}(k)]\end{aligned} \tag{3}$$

TABLE I
CONSTRAINED CONJUGATE GRADIENT ALGORITHM

Initialization:

λ, η with $(\lambda - 0.5) \leq \eta \leq \lambda$

δ small number

$\mathbf{w}(0) = \mathbf{F} = \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{f}$

$\bar{\mathbf{R}}(0) = \mathbf{P} = \mathbf{I} - \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T$

$\mathbf{g}(0) = \mathbf{c}(0) = zeros(M, 1)$

Running the algorithm:

for each k

{

$\bar{\mathbf{x}}(k) = \mathbf{P}\mathbf{x}(k)$

$\bar{\mathbf{R}}(k) = \lambda\bar{\mathbf{R}}(k-1) + \bar{\mathbf{x}}(k)\bar{\mathbf{x}}^T(k)$

$y(k) = \mathbf{w}^T(k-1)\mathbf{x}(k)$

$e(k) = d(k) - y(k)$

$\alpha(k) = \eta\frac{\mathbf{c}^T(k)\mathbf{g}(k-1)}{[\mathbf{c}^T(k)\bar{\mathbf{R}}(k)\mathbf{c}(k)+\delta]}$ (8a)

$\mathbf{g}(k) = \lambda\mathbf{g}(k-1) - \alpha(k)\bar{\mathbf{R}}(k)\mathbf{c}(k) - \bar{\mathbf{x}}(k)e(k)$ (11)

$\mathbf{w}(k) = \mathbf{P}\mathbf{w}(k-1) + \mathbf{F} - \alpha(k)\mathbf{c}(k)$ (10)

$\beta(k) = \frac{[\mathbf{g}(k)-\mathbf{g}(k-1)]^T\mathbf{g}(k)}{[\mathbf{g}^T(k-1)\mathbf{g}(k-1)+\delta]}$ (8b)

$\mathbf{c}(k+1) = \mathbf{g}(k) + \beta(k)\mathbf{c}(k)$ (9)

}

where $e_{\text{GSC}}(k) = y(k)$ and $\mathbf{x}_{\text{GSC}}(k) = \mathbf{B}^T\mathbf{x}(k)$. Moreover, using $\mathbf{w}(k) = \mathbf{F} - \mathbf{B}\mathbf{w}_{\text{GSC}}(k)$ to express $\mathbf{w}_{\text{GSC}}(k)$ as $(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{F} - (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{w}(k)$ and (2), straightforward algebraic manipulations lead to

$$\mathbf{w}(k+1) = \mathbf{F} + \mathbf{P}\mathbf{w}(k) - \mu y(k)\mathbf{B}\mathbf{B}^T\mathbf{x}(k). \tag{4}$$

From (4), (2), and (1), we see that if $\mathbf{B}^T\mathbf{B} = \mathbf{I}$, then $\mathbf{P} = \mathbf{B}\mathbf{B}^T$, and the GSC implementation is equivalent to the constrained algorithm. This equivalence condition, mentioned in [2], is necessary and sufficient for the CLMS algorithm, and may be extended to the conjugate gradient (CG) algorithm to be presented in the next section.

IV. THE CONSTRAINED CONJUGATE ALGORITHM

Based on the modified CG algorithm detailed in [5] and following the same approach described in the previous section for the CLMS algorithm, we propose next a constrained version of the CG algorithm. Condition $\mathbf{B}^T\mathbf{B} = \mathbf{I}$ will be used in the derivation.

It is worth mentioning that the reference algorithm [5] uses a degenerated scheme of the CG algorithm in order to have only one iteration per coefficient-vector update. If we use the exponentially weighted data strategy to estimate the input signal autocorrelation matrix for the unconstrained CG algorithm of the GSC structure, then

$$\begin{aligned}\mathbf{R}_{\text{GSC}}(k) &= \lambda\mathbf{R}_{\text{GSC}}(k-1) + \mathbf{x}_{\text{GSC}}(k)\mathbf{x}_{\text{GSC}}^T(k)\\ &= \mathbf{B}^T\mathbf{R}(k)\mathbf{B}\end{aligned} \tag{5}$$

where $\mathbf{R}(k) = \sum_{i=1}^{k}\lambda^{k-i}\mathbf{x}(i)\mathbf{x}^T(i) + \mathbf{R}(0)$ with $\mathbf{R}(0)$ initialized as $\mathbf{P}$ in order to force identical results when compared to the GSC implementation initialized with the identity matrix.

The derivation starts with the updating equation of the GSC CG algorithm [5]

$$\begin{aligned}\mathbf{w}(k) &= \mathbf{F} - \mathbf{B}\mathbf{w}_{\text{GSC}}(k)\\ &= \mathbf{F} - \mathbf{B}[\mathbf{w}_{\text{GSC}}(k-1) + \alpha_{\text{GSC}}(k)\mathbf{c}_{\text{GSC}}(k)]\\ &= \mathbf{P}\mathbf{w}(k-1) + \mathbf{F} - \alpha_{\text{GSC}}(k)\mathbf{B}\mathbf{c}_{\text{GSC}}(k)\end{aligned} \tag{6}$$

where $\alpha_{\text{GSC}}(k) = \eta((\mathbf{c}_{\text{GSC}}^T(k)\mathbf{g}_{\text{GSC}}(k-1))/(\mathbf{c}_{\text{GSC}}^T(k)\mathbf{R}_{\text{GSC}}(k)\mathbf{c}_{\text{GSC}}(k)))$ and $\mathbf{c}_{\text{GSC}}(k+1) = \mathbf{g}_{\text{GSC}}(k) + \beta_{\text{GSC}}(k)\mathbf{c}_{\text{GSC}}(k)$ with $\lambda - 0.5 \leq \eta \leq \lambda$ [5]. Moreover, completing the equations of the GSC-CG, we have

$$\begin{aligned}\mathbf{g}_{\text{GSC}}(k) &= \lambda\mathbf{g}_{\text{GSC}}(k-1) - \alpha_{\text{GSC}}(k)\mathbf{R}_{\text{GSC}}(k)\mathbf{c}_{\text{GSC}}(k)\\ &\quad + e_{\text{GSC}}(k)\mathbf{x}_{\text{GSC}}(k)\end{aligned} \tag{7a}$$

$$\beta_{\text{GSC}}(k) = \frac{[\mathbf{g}_{\text{GSC}}(k) - \mathbf{g}_{\text{GSC}}(k-1)]^T\mathbf{g}_{\text{GSC}}(k)}{\mathbf{g}_{\text{GSC}}^T(k-1)\mathbf{g}_{\text{GSC}}(k-1)} \tag{7b}$$

where $e_{\text{GSC}}(k) = y(k)$.

For the new CCG algorithm, if we make $\mathbf{c}(k) = \mathbf{B}\mathbf{c}_{\text{GSC}}(k)$ and $\mathbf{g}(k) = \mathbf{B}\mathbf{g}_{\text{GSC}}(k)$, the new $\alpha(k)$ and $\beta(k)$ will be equal to $\alpha_{\text{GSC}}(k)$ and $\beta_{\text{GSC}}(k)$, respectively, or

$$\alpha(k) = \alpha_{\text{GSC}}(k) = \eta\frac{\mathbf{c}^T(k)\mathbf{g}(k-1)}{[\mathbf{c}^T(k)\overline{\mathbf{R}}(k)\mathbf{c}(k) + \delta]} \tag{8a}$$

$$\beta(k) = \beta_{\text{GSC}}(k) = \frac{[\mathbf{g}(k) - \mathbf{g}(k-1)]^T\mathbf{g}(k)}{[\mathbf{g}^T(k-1)\mathbf{g}(k-1) + \delta]} \tag{8b}$$

where $\overline{\mathbf{R}}(k) = \mathbf{B}\mathbf{R}_{\text{GSC}}(k)\mathbf{B}^T = \mathbf{P}\mathbf{R}(k)\mathbf{P}$ and a small number δ was introduced to avoid division by zero.

The new search direction is easily obtained, and the result is

$$\mathbf{c}(k+1) = \mathbf{B}\mathbf{c}_{\text{GSC}}(k+1) = \mathbf{g}(k) + \beta(k)\mathbf{c}(k). \tag{9}$$

Note that the above definitions of $\alpha(k)$ and $\mathbf{c}(k)$ in (6) result in the updating equation of the new algorithm given by

$$\mathbf{w}(k) = \mathbf{P}\mathbf{w}(k-1) + \mathbf{F} - \alpha(k)\mathbf{c}(k). \tag{10}$$

Furthermore, it can be easily verified that the search direction and the residual are such that $\mathbf{c}(k) = \mathbf{P}\mathbf{c}(k)$ and $\mathbf{g}(k) = \mathbf{P}\mathbf{g}(k)$ for every k. From (7a) and the observations above, it follows that the updating equation of the residual is carried out as

$$\begin{aligned}\mathbf{g}(k) &= \mathbf{B}\mathbf{g}_{\text{GSC}}(k)\\ &= \mathbf{B}[\lambda\mathbf{g}_{\text{GSC}}(k-1) - \alpha_{\text{GSC}}(k)\mathbf{R}_{\text{GSC}}(k)\mathbf{c}_{\text{GSC}}(k)\\ &\qquad + y(k)\mathbf{B}^T\mathbf{x}(k)]\\ &= \lambda\mathbf{g}(k-1) - \alpha(k)\overline{\mathbf{R}}(k)\mathbf{c}(k) + y(k)\overline{\mathbf{x}}(k)\end{aligned} \tag{11}$$

with, for the case of orthogonal blocking matrix where $\mathbf{P} = \mathbf{B}\mathbf{B}^T$, $\overline{\mathbf{x}}(k) = \mathbf{P}\mathbf{x}(k)$ such that $\overline{\mathbf{R}}(k) = \lambda\overline{\mathbf{R}}(k-1) + \overline{\mathbf{x}}(k)\overline{\mathbf{x}}^T(k)$.

Table I shows the resulting CCG algorithm. Note that in this table, the more general case of reference signal not equal to zero was addressed. The same result can alternatively be obtained using the fact that the constrained algorithm can be viewed as a special case of the GSC structure with the projection matrix $\mathbf{P}$

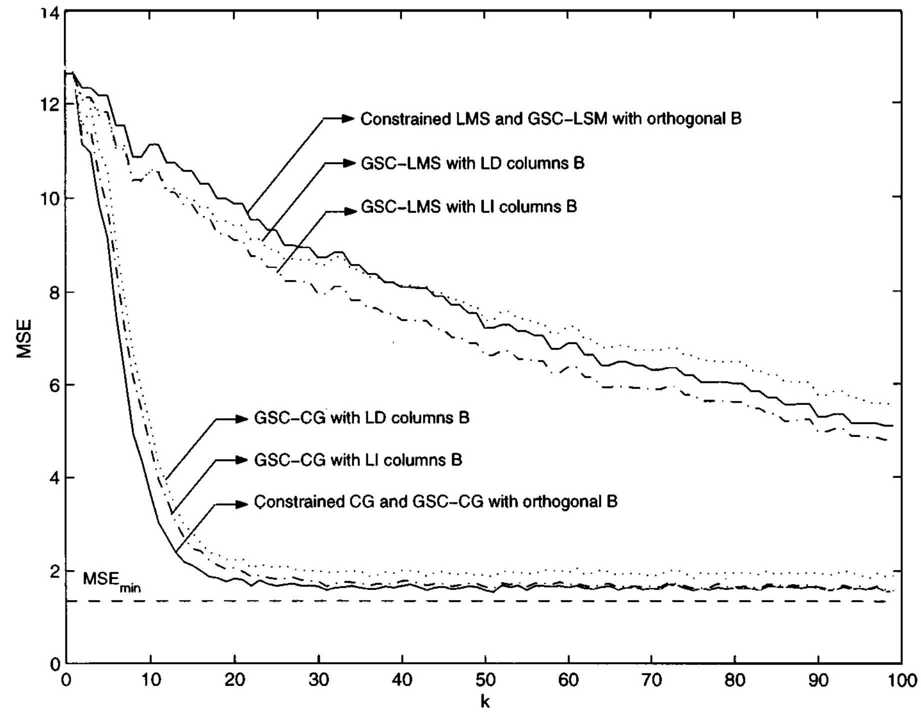

Fig. 1. Mean squared error (MSE).

used instead of the blocking matrix $\mathbf{B}$, as pointed out in [4] for the constrained LMS algorithm.

V. Simulation Results

In order to test the new algorithm, we applied the CCG algorithm to a single-user detection in a DS-CDMA mobile communication system. For this experiment, we assumed a simple model for a downlink synchronous transmission of K users.

The received continuous-time signal is passed through a chip-matched filter and is sampled at a chip rate such that the received discrete time input-signal vector may be expressed as $\mathbf{x}(k) = \mathbf{SAb}(k) + \mathbf{n}(k)$ where $\mathbf{S} = [\mathbf{s}_1\, \mathbf{s}_2 \cdots \mathbf{s}_K]$ is the spreading matrix containing the sampled spreading sequences of users 1 to K, $\mathbf{A} = \mathrm{diag}[A_1\, A_2\, \cdots\, A_K]$ contains the amplitudes of signals for each user, $\mathbf{b}(k) = [b_1(k)\, b_2(k)\, \cdots\, b_K(k)]^T$ contains the information bits, and $\mathbf{n}(k)$ is the sampled noise sequence. For this example, the constraint is such that $\mathbf{C} = \mathbf{s}_1$, and $\mathbf{f} = 1$. The number of users was set to $K = 5$, with Gold codes of length 7 used for spreading. The SNR for user one was made equal to 8 dB, and the relative power of interfering users was set to 20 dB, i.e., 10 $\log(P_i/P_1) = 20$.

The results are depicted in Fig. 1, where we show the mean squared error (MSE) for the CLMS and the new CCG algorithms as well as their GSC implementations with three different blocking matrices: one orthogonal obtained via SVD and the other two nonorthogonal built with plus and minus ones such that one has linearly independent and the other one linearly dependent columns. The parameters used in the adaptive filters were $\mu = 0.0005$, $\lambda = 0.9$, and $\eta = 0.6$.

From the results of this experiment, we could verify the equivalence between the constrained algorithms and their GSC implementations with an orthogonal blocking matrix. Moreover, the degradation of the GSC results with nonorthogonal blocking matrices are clear, and the fast convergence of the CG algorithm compared to the LMS algorithm is observed. This improvement in performance is typical for the input signal used is highly correlated.

It is worth mentioning that due to the use of an exponentially decaying data window, the proposed algorithm has a tracking capability and a convergence performance comparable to the RLS algorithm. In addition to these features, the misadjustment of the new algorithm is expected to be equal to the misadjustment of the GSC-RLS algorithm, for they minimize equivalent cost functions [5]. The computational complexity of the proposed algorithm may be expressed by $3M^2 + (10 + 4p)M + 1$ multiplications and two divisions, where M is the number of coefficients [size of $\mathbf{w}(k)$], and p is the number of constraints. In rough comparison (number of operations without any refinement or optimization in the computation of each formula), this computational load is higher than the CLMS algorithm ($2pM + 2M + 1$ multiplications), slightly smaller than the GSC-CG algorithm using orthogonal blocking matrix [$4M^2 + (11 - 7p)M + 3p^2 - 10p + 1$ multiplications and two divisions], and considerably smaller than the GSC-RLS algorithm [$5M^2 + (6 - 9p)M + 4p^2 - 5p$ multiplications and one division].

VI. Conclusions

The property that a GSC implementation of an adaptive filter may be equivalent to its constrained version through proper initialization and choice of the blocking matrix was discussed and applied to the derivation of the CCG algorithm. The new algorithm tested favorably against the CLMS algorithm in a computer experiment for multiuser detection. For the particular case where the blocking matrix was chosen to be orthogonal, the

CLMS and the CCG algorithms showed better results in terms of speed of convergence and misadjustment when compared to their GSC counterparts, employing nonorthogonal blocking matrices. This result was somehow expected because the derivation of a constrained adaptive filter imparts a criteria of optimality that is only shared by the GSC counterpart with orthogonal blocking matrices. Finally, we can infer from the equivalence in infinite precision between the CCG and the GSC-CG algorithms and the convergence criterion established in [5] for the CG algorithm that the proposed algorithm imparts the same stability properties.

REFERENCES

[1] O. L. Frost III, "An algorithm for linearly-constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp. 926–935, Aug. 1972.

[2] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propagat.*, vol. AP-30, pp. 27–34, Jan. 1982.

[3] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[4] C.-Y. Tseng and L. J. Griffiths, "A systematic procedure for implementing the blocking matrix in decomposed form," in *Proc. ACSSC*, Pacific Grove, CA, Oct. 1988, pp. 808–812.

[5] P. S. Chang and A. N. Willson Jr., "Analysis of conjugate gradient algorithms for adaptive filtering," *IEEE Trans. Signal Processing*, vol. 48, pp. 409–418, Feb. 2000.