Constrained Adaptive Filters

Marcello L. R. de Campos, Stefan Werner and José A. Apolinário Jr.

Federal University of Rio de Janeiro (UFRJ), Brazil, *campos@lps.ufrj.br*
Helsinki University of Technology (HUT), Finland, *stefan.werner@hut.fi*
Military Institute of Engineering (IME), Brazil, *apolin@ieee.org*

Abstract. This chapter provides a survey of known and state-of-the-art linearly constrained adaptive filters focusing on updating algorithms and implementation structures. Presentation of the material is general to fit various applications where linear constraints can be incorporated in the problem specification in order to enhance performance or simplify the solution. This includes, for example, narrowband and broadband arrays. In the next pages the reader will find an introduction to optimal constrained filters, some of the most widely used adaptation algorithms, alternatives to the direct-form structure for implementation, such as the generalized sidelobe canceller, a newly proposed structure based on Householder transformations, reduced-complexity algorithms based on set-membership filtering, and links to key references.

1 Introduction

Linearly-constrainted adaptive filters (LCAF) have found application in numerous areas, such as spectrum analysis, spatial-temporal processing, antenna arrays, and interference suppression in multiple access communications, among others. LCAF algorithms incorporate into the solution application-specific requirements translated into a set of linear equations to be satisfied by the coefficients. By imposing linear constraints on the coefficients one can, for example, improve robustness of the solution or relax the necessity of a training signal. This is equivalent to saying that the coefficient vector is required to belong to a particular hyperplane specified by the constraints. In general, the constraints are deterministic in nature and are derived from prior knowledge of the particular system at hand. For example, if direction of arrival of the signal of interest is known, jammer suppression can take place through spatial filtering without the need of training signal [8, 13]; or in systems with constant-envelope modulation (e.g., M-PSK), a constant-modulus constraint can mitigate multipath propagation effects [12, 20].

When compared to conventional adaptive filtering algorithms, LCAF algorithms require a mechanism to guarantee the imposed constraints are satisfied in every iteration. Several structures have been proposed in the past decades that setup the framework for LCAF algorithms. The two main approaches for derivation of LCAF algorithms rely on the *direct-form* structure and on the *generalized sidelobe canceller* (GSC) structure. The former uses the method of Lagrange multipliers during algorithm derivation such that the constraints are incorporated into the algorithm itself, explicitly solving a constrained optimization problem. The latter transforms the constrained optimization problem into a problem free from constraints to be solved in a subspace of reduced dimension, which is orthogonal to the subspace defined by the constraint equations. One advantage of the direct-form structure is potential lower computational complexity as compared to the GSC structure. On the other hand, the GSC structure offers, as advantage, the possibility to use conventional training-based adaptation algorithms. In the following sections these concepts will be discussed in more detail.

This chapter contains an overview of classical and state-of-the-art material related to direct-form and GSC structures. In particular, for the direct-form structure we present low-complexity algorithms and a computationally efficient implementation scheme based on the Householder transformation of the input signal. We consider equivalence of transient behavior of different algorithms in the direct-form and GSC structures, as well as the set-membership filtering approach to constrained adaptive filtering.

2 Optimal Constrained FIR Filter

In this section we present the basic concepts of optimal linearly-constrained filters with respect to the minimum mean squared error (MSE) criterion and to the (deterministic) least squares (LS) criterion. These criteria are suitable for adaptive implementation, which will be discussed in the next sections.

The basic setup of a multiple-input-single-output filter is depicted in Fig. 1, where the signal of each of the M channels is fed to an FIR filter with N taps. The output of the filter is expressed as $y(k) = \mathbf{w}^{\mathrm{H}}(k)\mathbf{x}(k)$, where

$$\mathbf{w}(k) = \begin{bmatrix}\mathbf{w}_1^{\mathrm{T}}(k) & \mathbf{w}_2^{\mathrm{T}}(k) & \cdots & \mathbf{w}_M^{\mathrm{T}}(k)\end{bmatrix}^{\mathrm{T}} \tag{1}$$

$$\mathbf{x}(k) = \begin{bmatrix}\mathbf{x}_1^{\mathrm{T}}(k) & \mathbf{x}_2^{\mathrm{T}}(k) & \cdots & \mathbf{x}_M^{\mathrm{T}}(k)\end{bmatrix}^{\mathrm{T}} \tag{2}$$

$$\mathbf{x}_i(k) = \begin{bmatrix}x_i(k) & x_i(k-1) & \cdots & x_i(k-N+1)\end{bmatrix}^{\mathrm{T}} \tag{3}$$

The structure can be applied to several applications of interest. In the particular application of a narrowband beamformer, N is chosen equal to 1 or, equivalently, one tap per antenna, and $\mathbf{x}(k)$ contains the signals at the sensor outputs. A broadband beamformer represents the more general case of having $N > 1$, which allows both spatial and temporal filtering.

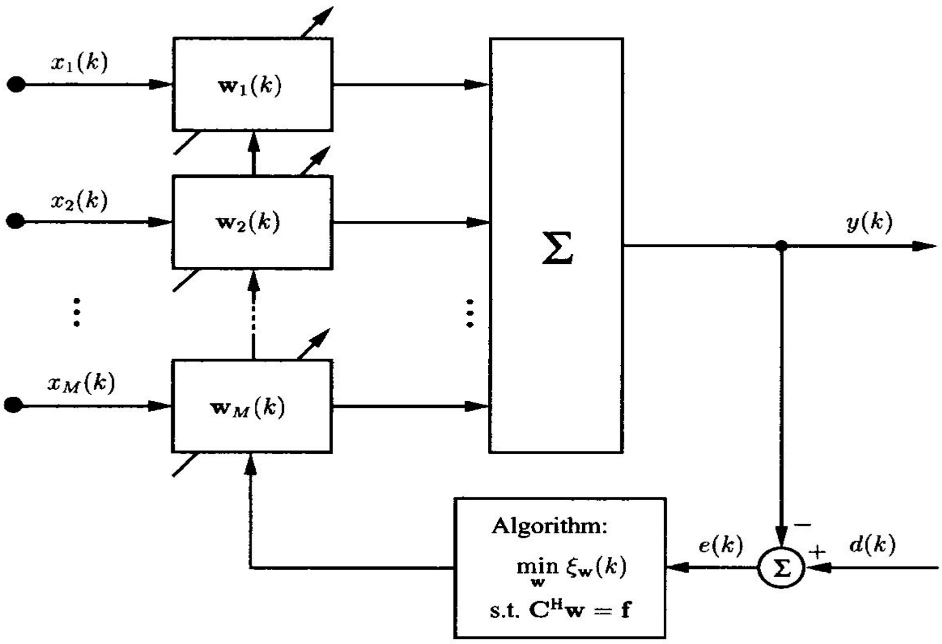

Fig. 1. The basic setup of the constrained filter

The coefficient vector in Eq. (1) is chosen to minimize a function $\xi_{\mathbf{w}}(k)$ subjected to a set of linear constraints, $\mathbf{C}^H\mathbf{w} = \mathbf{f}$, i.e.,

$$\mathbf{w}_{opt} = \arg\min_{\mathbf{w}} \xi_{\mathbf{w}}(k) \quad \text{subject to} \quad \mathbf{C}^H\mathbf{w} = \mathbf{f} \tag{4}$$

where $\mathbf{C}$ is the $MN \times p$ constraint matrix and $\mathbf{f}$ is the $p \times 1$ gain vector, p being the number of constraints. In the following subsections we discuss two particular choices of $\xi_{\mathbf{w}}(k)$.

We will assume hereafter that all signals are zero-mean wide-sense stationary processes.

2.1 The Optimum Constrained MSE filter

Cost functions are usually related to the output error signal, $e(k) = d(k) - y(k)$, and the most widely used is the MSE, defined as

$$\xi_{\mathbf{w}}(k) = \mathrm{E}[|e(k)|^2] \tag{5}$$

which gives as optimal solution [6, 19]

$$\mathbf{w}_{opt} = \mathbf{R}^{-1}\mathbf{p} + \mathbf{R}^{-1}\mathbf{C}(\mathbf{C}^H\mathbf{R}^{-1}\mathbf{C})^{-1}(\mathbf{f} - \mathbf{C}^H\mathbf{R}^{-1}\mathbf{p}) \tag{6}$$

where $\mathbf{R}$ is the input-signal autocorrelation matrix defined as

$$\mathbf{R} = \mathrm{E}[\mathbf{x}(k)\mathbf{x}^H(k)] \tag{7}$$

and $\mathbf{p}$ is the cross-correlation vector between the training signal and the input signal vector, defined as

$$\mathbf{p} = \mathrm{E}[d^*(k)\mathbf{x}(k)] \tag{8}$$

In the absence of a training signal, $d(k) = 0$ and consequently $\mathbf{p} = \mathbf{0}$. The cost function, in this case, is the minimum output energy (MOE), $\xi_{\mathbf{w}}(k) = \mathbf{w}^{\mathrm{H}}\mathbf{R}\mathbf{w}$. The solution, known as the linearly-constrained minimum variance (LCMV) filter, can be obtained from Eq. (6) with $\mathbf{p} = \mathbf{0}$. The rest of the chapter treats the more general case when a training signal is present; applications without training signals can be modeled by setting $d(k) = 0$.

An application where the training signal is present was discussed in [19] where constraints were used to ensure linear phase filters. In many wireless systems a training signal is periodically retransmitted to aid time-varying parameter estimation, for example, multipath channel coefficients. In such applications we can think of a solution that switches in time between the two cases, i.e., with and without training signal. In beamforming the constraints can be used to relax the need of training signal by setting spatial or temporal (or both) constraints on the filter coefficients. For example, in a narrowband beamformer, minimization of MOE with unit gain on a specific direction of arrival can cancel jammer signals impinging the array from other directions [24]. In a broadband beamformer, constraints can also be imposed in time [13].

2.2 The Constrained Least Squares Filter

Another possible (deterministic) cost function is the weighted least-squares (LS) function defined as

$$\xi_{\mathbf{w}}(k) = \sum_{i=0}^{k} \lambda^{(k-i)} \left| d(i) - \mathbf{w}^{\mathrm{H}}\mathbf{x}(i) \right|^2 \tag{9}$$

where λ is a forgetting factor usually chosen close to one. Equation (9) substituted in Eq. (4) gives the following constrained LS solution at time instant k

$$\mathbf{w}_{\mathrm{opt}}(k) = \mathbf{R}^{-1}(k)\mathbf{p}(k) + \mathbf{R}^{-1}(k)\mathbf{C}(\mathbf{C}^{\mathrm{H}}\mathbf{R}^{-1}(k)\mathbf{C})^{-1}[\mathbf{f} - \mathbf{C}^{\mathrm{H}}\mathbf{R}^{-1}(k)\mathbf{p}(k)] \tag{10}$$

where the *deterministic* autocorrelation matrix $\mathbf{R}(k)$ and the *deterministic* cross-correlation vector $\mathbf{p}(k)$ are given by

$$\mathbf{R}(k) = \sum_{i=0}^{k} \lambda^{(k-i)} \mathbf{x}(i)\mathbf{x}^{\mathrm{H}}(i) \tag{11}$$

$$\mathbf{p}(k) = \sum_{i=0}^{k} \lambda^{(k-i)} \mathbf{x}(i) d^*(i) \tag{12}$$

respectively. For $d(k) = 0$, then $\mathbf{p}(k) = \mathbf{0}$, and Eq. (10) yields the least squares version of the LCMV filter.

3
Direct-Form Constrained FIR Adaptive Filters

Adaptive filters are useful in cases where the statistics of the signals are unknown or time-varying and, therefore, optimal filters cannot be directly obtained. When signal statistics are not available but can be estimated recursively or in batch, the formulas presented in the previous section can be used. However, in these cases, implementation may be prohibitively expensive in terms of computational complexity, required memory, and introduced delay. Adaptation algorithms are usually more efficient as alternative to estimate recursively the optimal solution.

Adaptation algorithms differ in terms of speed of convergence towards the optimal solution, steady-state error, and computational complexity. Computational complexity is usually tied to their sensitivity to input-signal correlation: complexity and speed of convergence are usually conflicting performance measures in correlated scenarios. Therefore, choice of algorithm should be application dependent, such that acceptable performance is obtained with minimum complexity. In this section, we discuss three FIR LCAF algorithms implemented using the direct-form structure: the constrained least mean square (CLMS) algorithm, the constrained affine-projection (CAP) algorithm, and the constrained recursive least squares (CRLS) algorithm. These algorithms cover a wide range of applications with different performance requirements.

3.1
The Constrained LMS Algorithm

The CLMS algorithm [8] has been widely used due to its simplicity, proven stability, and low computational complexity.

The coefficient updating equation for the CLMS algorithm solves

$$\min_{\mathbf{w}} \left|e(k)\right|^2 \quad \text{subject to} \quad \mathbf{C}^H\mathbf{w} = \mathbf{f} \tag{13}$$

using a gradient approach and the method of Lagrange multipliers, and is given by [8]

$$\mathbf{w}(k+1) = \mathbf{P}\left[\mathbf{w}(k) + \mu e^*(k)\mathbf{x}(k)\right] + \mathbf{F} \tag{14}$$

where μ is a step size controlling stability and speed of convergence,

$$\mathbf{P} = \mathbf{I} - \mathbf{C}\left(\mathbf{C}^H\mathbf{C}\right)^{-1}\mathbf{C}^H \tag{15}$$

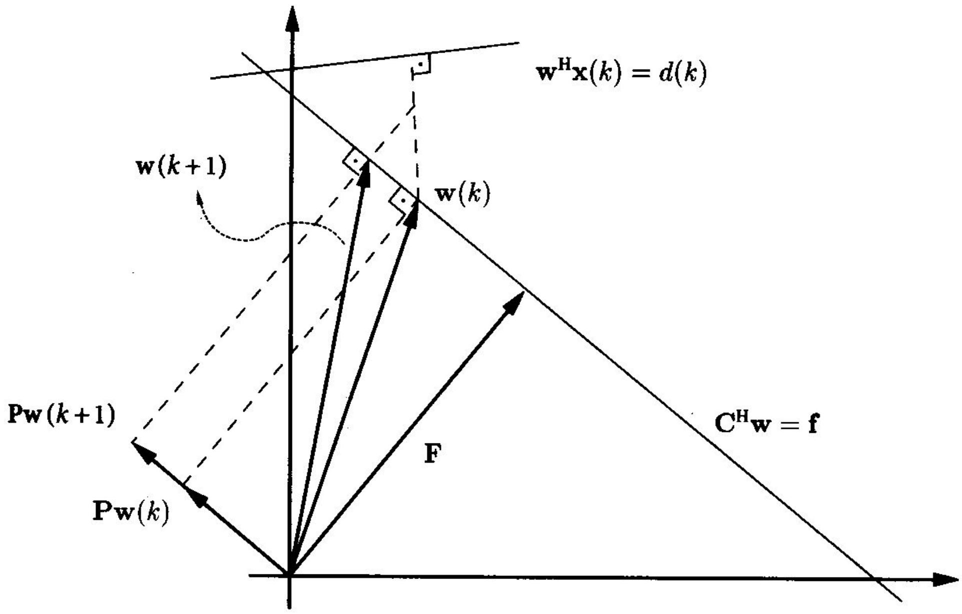

Fig. 2. Illustration of CLMS algorithm

performs a projection onto the hyperplane defined by $\mathbf{C}^H\mathbf{w} = \mathbf{0}$, and

$$\mathbf{F} = \mathbf{C}\left(\mathbf{C}^H\mathbf{C}\right)^{-1}\mathbf{f} \tag{16}$$

is a vector used to move the projected solution back to the constraint hyperplane, as illustrated in Fig. 2. Notice in this figure that the CLMS solution corresponds to the LMS solution projected onto the hyperplane defined by $\mathbf{C}^H\mathbf{w} = \mathbf{f}$.

The simplification $\mathbf{Pw}(k) + \mathbf{F} = \mathbf{w}(k)$ should be avoided in a finite-precision environment, for accumulation of round-off errors may cause the solution to drift away from the constraint hyperplane [8].

In order to guarantee mean-squared stability, the step size should be kept within the range [8]

$$0 < \mu < \frac{2}{3\,\mathrm{tr}\left(\mathbf{R}\right)} \tag{17}$$

The equations of the CLMS algorithm are presented in Table 1.

Table 1. The CLMS Algorithm

CLMS Algorithm
$\mathbf{w}(0) = \mathbf{F}$
for each k
{
$e(k) = d(k) - \mathbf{w}^H(k)\mathbf{x}(k)$
$\mathbf{w}(k+1) = \mathbf{P}[\mathbf{w}(k) + \mu e^*(k)\mathbf{x}(k)] + \mathbf{F}$
}

3.2 The Constrained Affine-Projection Algorithm

The main drawback of the CLMS algorithm is the slow convergence speed for highly correlated signals. The CAP algorithm balances convergence speed and computational complexity by using a suitable number of past data pairs in the coefficient update.

The CAP algorithm can also be viewed as a generalization of the normalized constrained LMS (NCLMS) [1] and binormalized data-reusing constrained LMS (BNDRCLMS) [1] algorithms to include an arbitrary number L of previous data pairs.

The coefficient updating equation for the CAP algorithm solves

$$\min_{\mathbf{w}} \left\|\mathbf{w}(k)-\mathbf{w}\right\|^2 \text{ subject to } \begin{cases}\mathbf{X}^{\mathrm{T}}(k)\mathbf{w}^* = \mathbf{d}(k)\\ \mathbf{C}^{\mathrm{H}}\mathbf{w} = \mathbf{f}\end{cases} \tag{18}$$

using the method of Lagrange multipliers, and the updating equation is given

$$\mathbf{w}(k+1) = \mathbf{P}\left[\mathbf{w}(k) + \mu\mathbf{X}(k)\left(\mathbf{X}^{\mathrm{H}}(k)\mathbf{P}\mathbf{X}(k)+\delta\mathbf{I}\right)^{-1}\mathbf{e}^*(k)\right] + \mathbf{F} \tag{19}$$

where $\mathbf{P}$ and $\mathbf{F}$ are defined as in Eqs. (15) and (16), δ is a small positive number used to regularize the inverse, $\mathbf{I}$ is the $L\times L$ identity matrix,

$$\mathbf{e}(k) = \mathbf{d}(k) - \mathbf{X}^{\mathrm{T}}(k)\mathbf{w}^*(k) \tag{20}$$

$$\mathbf{d}(k) = \begin{bmatrix} d(k) & d(k-1) & \cdots & d(k-L+1)\end{bmatrix}^{\mathrm{T}} \tag{21}$$

$$\mathbf{X}(k) = \begin{bmatrix} \mathbf{x}(k) & \mathbf{x}(k-1) & \cdots & \mathbf{x}(k-L+1)\end{bmatrix} \tag{22}$$

The equations of the CAP algorithm are presented in Table 2.

Table 2. The CAP Algorithm

CAP Algorithm

$\mathbf{w}(0) = \mathbf{F}$
for each k
{
$\quad\mathbf{e}(k) = \mathbf{d}(k) - \mathbf{X}^{\mathrm{T}}(k)\mathbf{w}^*(k)$
$\quad\mathbf{t}(k) = [\mathbf{X}^{\mathrm{H}}(k)\mathbf{P}\mathbf{X}(k) + \mathbf{I}]^{-1}\mathbf{e}^*(k)$
$\quad\mathbf{w}(k+1) = \mathbf{P}[\mathbf{w}(k) + \mu\mathbf{X}(k)\mathbf{t}(k)] + \mathbf{F}$
}

3.3 The Constrained Recursive Least-Squares Algorithm

Adaptive implementations of the direct-form CRLS algorithm have been considered in, e.g., [3, 18]. The main advantage of the CRLS algorithm is fast convergence, which is independent of the eigenvalue spread of the input-signal autocorrelation matrix. The CRLS algorithm in [18] uses recursive updates of $\mathbf{R}^{-1}(k)$ and other internal variables which are based on the matrix inversion lemma [6]. This algorithm is more prone to diverge due to numerical errors than its conventional training-based counterpart even for well-behaved input signals, and its computational complexity is of the order of $(MN)^2$ multiplications per iteration. However, as it will be shown in the next section, transient and steady-state behavior of the CRLS algorithm in direct-form and in GSC structures are identical. Therefore, one should consider carefully alternatives to the CRLS in a direct-form structure, such as stable and efficient RLS algorithm implementations in the GSC structure.

4 Decomposed-Form FIR Adaptive Filters

This section deals with alternative implementations of LCAFs that allow training-based adaptation algorithms to be applied to linearly-constrained problems. Firstly, the classical generalized sidelobe canceller (GSC) structure [13] is reviewed, followed by a discussion on the conditions for transient-equivalence of the adaptive implementations of the GSC and the direct-form counterparts. Finally, the Householder-Transform (HT) structure [5] for solving constrained problems is presented as an alternative to the GSC structure. In terms of computational complexity, the HT structure compares to the most efficient implementations of the direct-form and GSC structures.

4.1 Generalized Sidelobe Canceller

The GSC structure in Fig. 3 solves the linearly-constrained problem given in Eq. (4) by dividing the filter vector, $\mathbf{w}$, into two orthogonal components

$$\mathbf{w} = \mathbf{F} - \mathbf{B}\mathbf{w}_{\text{GSC}}, \tag{23}$$

where: the component in the upper branch, $\mathbf{F}$, as given in Eq. (16), alone satisfies the constraints, i.e., $\mathbf{C}^H\mathbf{F} = \mathbf{f}$; the $MN \times (MN - p)$ matrix $\mathbf{B}$, referred to as the *blocking matrix*, is in the left nullspace of $\mathbf{C}$, i.e., $\mathbf{B}^H\mathbf{C} = \mathbf{0}$; and $\mathbf{w}_{\text{GSC}}$ is the solution to the unconstrained optimization problem

$$\mathbf{w}_{\text{GSC,opt}} = \arg \min_{\mathbf{w}_{\text{GSC}}} \xi_{\text{w}}(k) \tag{24}$$

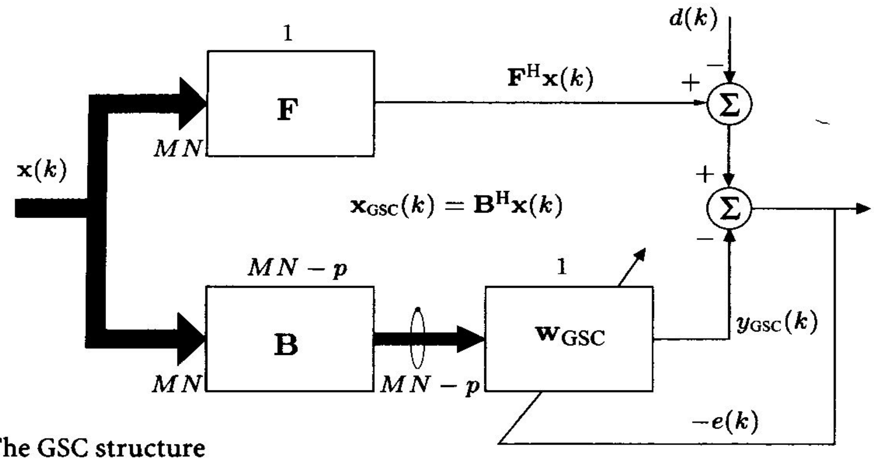

Fig. 3. The GSC structure

with w given by Eq. (23). Figure 3 shows the schematic of the GSC structure where $\mathbf{w}_{\mathrm{GSC}}$ is updated using a training-based adaptation algorithm. The desired signal, as defined in the figure, incorporates an external reference signal $d(k)$ such that the resulting desired signal fed back to the adaptation algorithm in the GSC structure becomes $d_{\mathrm{GSC}}(k) = \mathbf{F}^{\mathrm{H}}\mathbf{x}(k) - d(k)$. This more general case includes common applications where $d(k) = 0$, e.g., blind beamforming [8] and blind multiuser detection [15]. If we choose to minimize the MSE, i. e., $\xi_{\mathrm{w}}(k) = \mathrm{E}[|d(k) - \mathbf{w}^{\mathrm{H}}\mathbf{x}(k)|^2]$, the optimal solution is given by

$$\mathbf{w}_{\mathrm{GSC,opt}} = \mathbf{R}_{\mathrm{GSC}}^{-1}\mathbf{p}_{\mathrm{GSC}} \tag{25}$$

where $\mathbf{R}_{\mathrm{GSC}}$ and $\mathbf{p}_{\mathrm{GSC}}$ are the autocorrelation matrix and the cross-correlation vector, given by

$$\mathbf{R}_{\mathrm{GSC}} = \mathbf{B}^{\mathrm{H}}\mathbf{R}\mathbf{B} \tag{26}$$

$$\mathbf{p}_{\mathrm{GSC}} = -\mathbf{B}^{\mathrm{H}}\mathbf{p} + \mathbf{B}^{\mathrm{H}}\mathbf{R}\mathbf{F} \tag{27}$$

respectively. Choosing the LS objective function instead of the MSE, implies that $\mathbf{R}_{\mathrm{GSC}}$ and $\mathbf{p}_{\mathrm{GSC}}$ are substituted by their deterministic versions $\mathbf{R}_{\mathrm{GSC}}(k)$ and $\mathbf{p}_{\mathrm{GSC}}(k)$, respectively. Since $\mathbf{w}_{\mathrm{GSC}}$ is the solution to an unconstrained optimization problem, as indicated in Eq. (24), several training-based algorithms available in literature can be applied for the recursive update of the filter coefficients.

4.1.1
Choice of blocking matrix B

The only requirement put on the blocking matrix **B** is that it spans the left nullspace of **C**, i. e., $\mathbf{C}^{\mathrm{H}}\mathbf{B} = \mathbf{0}$. As a consequence, implementation allows several possible choices of **B**. However, matrix **B** affects the overall computational complexity and the stability of the implementation. If a unitary blocking matrix is to be chosen, e.g., to minimize problems in a finite precision environment [15], singular-value decomposition or any other unitary decomposition of **C** could be used. In such cases, **B** may not have any special structure that can reduce the computational complexity of the matrix-vector multiplication $\mathbf{B}\mathbf{x}(k)$ performed at each iteration. As a consequence, computational complexity of the filtering operation may be up to an order of magnitude higher than that of the adaptation algorithm. This observation is particularly important if low-complexity algorithms, such as the LMS or affine-projection algorithms [6] are used. In case the RLS algorithm is used, the computational complexity of the coefficient update is at least of the same order of magnitude as that of the product $\mathbf{B}\mathbf{x}(k)$.

A non-unitary blocking matrix is suggested in [21] and is implemented as a sequence of sparse blocking matrices $\mathbf{B} = \mathbf{B}_1 \cdots \mathbf{B}_{p-1}\mathbf{B}_p$ where $\mathbf{B}_i$ is a $(MN-i+1)\times(MN-i)$ matrix of full rank. In beamforming with presteering, the requirement for spatial blocking of the look direction is that the rows of **B** sum up to zero [13]. A commonly used blocking matrix fulfilling this requirement contains +1 in the main diagonal and −1 in the upper diagonal. If the number of antennas is such that $M = 2^n$, $n > 0$, an orthogonal blocking matrix can be constructed using Walsh functions [13].

4.1.2 Equivalence of direct-form and GSC structures

Analysis of the CLMS algorithm in the direct-form structure and the LMS algorithm in the GSC structure reveals that the transients of both algorithms become equal only if **B** is unitary, i.e., $\mathbf{B}^H\mathbf{B} = \mathbf{I}$ [13]. As discussed above, a requirement of a unitary blocking matrix can lead to a computationally complex implementation of the GSC structure, making it hard to motivate the use of simple algorithms from the computational complexity point-of-view. Non-unitary matrices can render computational complexity of the order of magnitude comparable to that of simple algorithms. However, transient, or equivalently, convergence speed, may depend on the step size and the particular blocking matrix chosen. In other words, if the blocking matrix changes, the step size changes, including the limits for stability. With respect to the RLS algorithm, the transients of the RLS algorithm in the GSC structure and that of the CRLS algorithm are identical regardless of the blocking matrix chosen in the GSC structure [28]. This result is formalized in the following lemma:

Lemma [28]: For $\mathbf{B}^H\mathbf{C} = \mathbf{0}$, if $\mathbf{R}^{-1}(k)$ exists and is symmetric, if rank $(\mathbf{B}) = MN-p$, and if rank $(\mathbf{C}) = p$, the GSC-RLS and the CRLS algorithms have identical solutions for all k.

For a proof of the lemma, the reader is referred to [28]. This is a looser requirement than the transient equivalence of the CLMS algorithm and the GSC structure using the LMS algorithm, which in addition to $\mathbf{B}^H\mathbf{C} = \mathbf{0}$, requires $\mathbf{B}$ to be unitary.

For reasons of computational complexity and robustness, the result just presented serves as an indication that implementing the unconstrained form of the RLS algorithm may be preferable, either using the Householder transform structure to be discussed in the following section, or using the GSC structure. Based on the previous section, one may argue that the choice of a non-unitary blocking matrix is of less importance with respect to computational complexity due to the complexity of the coefficient updating of the RLS algorithm itself. However, at least for the particular case of a broadband beamformer where only spatial constraints are imposed and $M-p$ FIR filters are placed after the blocking operation [13], multichannel implementations of the fast RLS algorithm may be used for each filter [16]. In this case, computational complexity can be reduced by using non-unitary matrices, like the one proposed in [21] and mentioned in the previous section.

4.2 Householder-Transform Constrained Filters

In this subsection we consider an alternative implementation of the GSC structure, which applies a unitary transformation to the input-signal vector. The basic idea to be presented below can be seen as a particularly efficient implementation of a GSC structure with a unitary blocking matrix. This is illustrated in Fig. 4.

Let the unitary matrix $\mathbf{Q}$ transform the adaptive filter-coefficient vector in order to generate a modified vector $\bar{\mathbf{w}}(k) = \mathbf{Q}\mathbf{w}(k)$. If the same transformation is applied to the input-signal vector $\mathbf{x}(k)$, i.e., $\bar{\mathbf{x}}(k) = \mathbf{Q}\mathbf{x}(k)$ the output signal from the transformed filter will be the same as that of the original filter, i.e.,

$$\bar{y}(k) = \bar{\mathbf{w}}^H(k)\bar{\mathbf{x}}(k) = \mathbf{w}^H(k)\mathbf{Q}^H\mathbf{Q}\mathbf{x}(k) = \mathbf{w}^H(k)\mathbf{x}(k) \tag{28}$$

If matrix $\mathbf{Q}$ is chosen such that it triangularizes matrix $\mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1/2}$ through a sequence of Householder transformations, as proposed in [5], then vector $\bar{\mathbf{w}}_U(0)$, which constitutes the $p \times 1$ upper part of the transformed vector $\bar{\mathbf{w}}(k)$, is constant, whereas vector $\bar{\mathbf{w}}_L(0)$, which constitutes its $(MN-p) \times 1$ lower part, can be updated using one of several conventional training-based adaptation algorithms. The advantage of such transformation approach, as compared to the GSC structure, lies in the efficient implementation of $\mathbf{Q}\mathbf{x}(k)$ that can be carried out through the following product of p matrices

$$\bar{\mathbf{x}}(k) = \mathbf{Q}_p \cdots \mathbf{Q}_2\mathbf{Q}_1\mathbf{x}(k) \tag{29}$$

where

$$\mathbf{Q}_i = \begin{bmatrix} \mathbf{I}_{i-1\times i-1} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{Q}}_i \end{bmatrix} \tag{30}$$

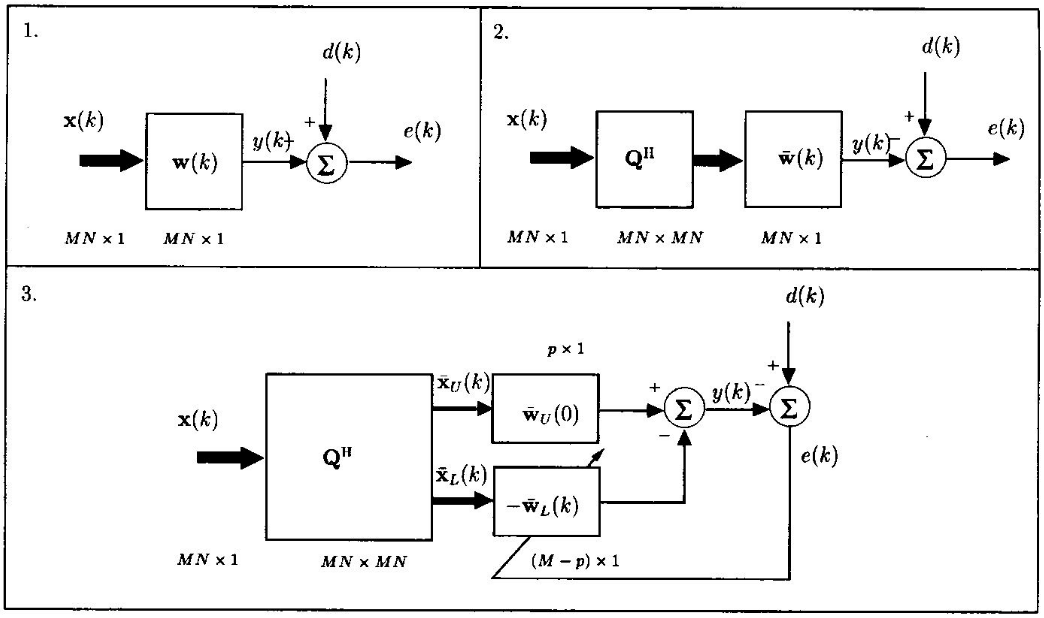

Fig. 4. (1) Untransformed adaptive filter. (2) Applying the transformation as in Eq. (29). (3). Splitting the transformed vector $\bar{\mathbf{w}}(k)$ into a constant $p \times 1$ vector $\bar{\mathbf{w}}_U(0)$ and an $MN-p \times 1$ vector $\bar{\mathbf{w}}_L(k)$ to be updated adaptively

and matrix $\bar{\mathbf{Q}}_i = \mathbf{I} - 2\bar{\mathbf{v}}\bar{\mathbf{v}}^H$ is an ordinary $(MN - i+1) \times (MN - i+1)$ Householder transformation.

As an example of the utilization of the scheme presented above, the equations of the Householder-based CLMS (HCLMS) algorithm [5] are presented in Table 3.

In order to relate the HT structure to the GSC structure, it was shown in [5] that if a Householder transformation matrix Q is constructed to triangularize

Table 3. The HCLMS Algorithm

HCLMS Algorithm
$\bar{\mathbf{w}}_U(0) =$ first p elements of **QF**
for each k
{
$\bar{\mathbf{x}}(k) = \mathbf{Q}\mathbf{x}(k) = \left[\bar{\mathbf{x}}_U^T(k) \quad \bar{\mathbf{x}}_L^T(k)\right]^T$
$\bar{\mathbf{w}}(k) = \left[\bar{\mathbf{w}}_U^T(0) \quad \bar{\mathbf{w}}_L^T(k)\right]^T$
$e(k) = d(k) - \bar{\mathbf{w}}^H(k)\bar{\mathbf{x}}(k)$
$\bar{\mathbf{w}}_L(k) = \bar{\mathbf{w}}_L(k) + \mu e^*(k)\bar{\mathbf{x}}_L(k)$
}

matrix $C(C^H C)^{-1/2}$, the following is true

$$QC\left(C^H C\right)^{-1} C^H Q^H = \bar{C}\left(\bar{C}^H \bar{C}\right)^{-1} \bar{C}^H = \begin{bmatrix} I_{p\times p} & 0 \\ 0 & 0 \end{bmatrix} \tag{31}$$

where $\bar{C} = QC$. Let the transformation matrix Q and the adaptive filter vector $w(k)$ be partitioned as follows

$$Q = \begin{bmatrix} Q_U \\ Q_L \end{bmatrix}, \quad w(k) = \begin{bmatrix} w_U(0) \\ w_L(k) \end{bmatrix} \tag{32}$$

where Q_U has dimension $p \times MN$ and Q_L has dimension $(MN - p) \times MN$. It will be shown that using this partitioning of Q, the upper part Q_U can be related to the upper part of the GSC structure, which consists of the filter F, and Q_L can be regarded as a valid blocking matrix. We first show that $Q_L^H \bar{w}_U(0) = F$ by pre-multiplying Eq. (31) with Q^H and post-multiplying with $\bar{w}(k)$

$$\begin{aligned} Q^H \bar{C}\left(\bar{C}^H \bar{C}\right)^{-1} \bar{C}^H \bar{w}(k) &= F \\ &= \begin{bmatrix} Q_U^H & 0 \end{bmatrix} \bar{w}(k) = \begin{bmatrix} Q_U^H & 0 \end{bmatrix} \begin{bmatrix} \bar{w}_U(0) \\ \hline \bar{w}_L(k) \end{bmatrix} = Q_U^H \bar{w}_U(0) \end{aligned} \tag{33}$$

which is the upper part in the GSC structure. In other words, given F and Q, the constant part $\bar{w}_U(0)$ is given by the first p elements of the vector QF. In order to show that Q_L constitutes a valid blocking matrix, notice that

$$QC = \begin{bmatrix} Q_U C \\ \hline Q_L C \end{bmatrix} = \bar{C}\left(\bar{C}^H \bar{C}\right)^{-1} \bar{C}^H \bar{C} = \begin{bmatrix} I_{p\times p} & 0 \\ 0 & 0 \end{bmatrix} QC = \begin{bmatrix} Q_U C \\ \hline 0 \end{bmatrix} \tag{34}$$

and, therefore, $Q_L C = 0$. Together with the fact that Q_L has full rank, we can conclude that Q_L fulfills the requirement for a valid blocking matrix.

For a large number of constraints ($p > MN/2$) it may be more economical to implement the upper and lower parts of the GSC structure separately. However, for small to moderate (or practical) values of p, implementation as discussed above will be more efficient with the additional advantage that the transformation is guaranteed to be unitary.

5
Reduced-Complexity Constrained Filters

Demanding applications of LCAF, e.g., wireless communications or smart antennas, may require long filters for an acceptable performance on one hand, and cheap or low-power hardware on the other hand. An acceptable compromise solution to these conflicting requirements must also take into account that constrained versions of adaptive filters are usually more complex than their unconstrained counterparts. This is the case even when low-complexity implementations are favored, such as the Householder-based implementation or the GSC stucture, with trivial or sparse blocking matrices, discussed in the previous sections. In such cases, reducing computational complexity of the LCAF may mean cheaper hardware, simpler software, or may even render the application viable.

In this section we discuss two reduced-complexity strategies for LCAF implementation: *Set-membership filtering*, which acts on the algorithm limiting the frequency of updates, and *partial adaptation*, which acts on the LCAF structure limiting the degrees of freedom (or number of parameters) to adapt.

5.1
Set-Membership Affine-Projection Algorithm

Set-membership filtering (SMF) [11, 26] is a recent approach to adaptive filtering, where a specification on the filter parameters is achieved by constraining the output estimation error to be smaller than a deterministic threshold. For a properly chosen bound γ on the estimation error $e(k)$, there are infinitely many valid estimates for $\mathbf{w}$. As a result of the bounded error specification, adaptive filters derived within the SMF framework will not perform update for all incoming samples, in other words they are *data selective*, and can reduce the overall complexity considerably as compared to their conventional counterparts. Set-membership CAF has been considered in [17, 27].

Adaptive SMF algorithms work with the so-called *exact membership set*[1] $\psi(k)$ constructed from the observed data pairs,

$$\psi(k) = \bigcap_{i=1}^{k} \mathcal{H}(i) \tag{35}$$

[1] The optimal set-membership filter seek solutions that belong to the *feasibility set* $\Theta = \cap_{(d,\mathbf{x})\in\mathcal{S}} \{\mathbf{w} \in \Re^{MN} : |d - \mathbf{w}^H\mathbf{x}| \leq \gamma\}$ with $\mathcal{S}$ being the set of all possible data-pairs $(\mathbf{x}, d)$. In practice, it may be difficult to predict all possible data pairs making $\psi(k)$ more suitable for adaptive methods. Notice that Θ is a subset of $\psi(k)$ and also the limiting set of $\psi(k)$, i.e., the two sets will be equal if the training signal traverses all pairs belonging to $\mathcal{S}$.

where $\mathcal{H}(k)$ is referred to as the *constraint set* containing all the vectors **w** for which the associated output error at time instant k is upper bounded by $\gamma(k)$, i.e.,

$$\mathcal{H}(k)=\left\{\mathrm{w}\in\Re^{MN}:\left|d(k)-\mathrm{w}^{\mathrm{H}}\mathrm{x}(k)\right|\le\gamma(k)\right\} \tag{36}$$

Simple training-based adaptive SMF algorithms compute a point estimate provided part of the information in $\psi(k)$, e.g., the information provided by $\mathcal{H}(k)$ [11] or several past constraint sets [7,29].

In the SMF formulation applicable to linearly-constrained problems considered here, the membership set is expressed as $\psi(k)=\psi_{k-L}(k)\cap\psi_L(k)$, where $\psi_L(k)$ corresponds to the intersection of the L past constraint sets, i.e.,

$$\psi_L(k)=\bigcap_{i=k-L+1}^{k}\mathcal{H}(i) \tag{37}$$

Next we consider an adaptive SMF algorithm whose coefficient vector after updating belongs to the hyperplane defined by $\mathbf{C}^{\mathrm{H}}\mathbf{w}=\mathbf{f}$ and also to the L past constraint sets $\psi_L(k)$. The set-membership constrained affine projection (SM-CAP) algorithm [27] solves the following optimization criterion whenever $\mathbf{w}(k)\notin\psi_L(k)$

$$\mathbf{w}(k+1)=\arg\min_{\mathbf{w}}\left\|\mathbf{w}-\mathbf{w}_k\right\|^2 \text{ s.t. } \mathbf{C}^{\mathrm{H}}\mathbf{w}=\mathbf{f} \text{ and } \mathbf{d}(k)-\mathbf{X}^{\mathrm{T}}(k)\mathbf{w}^*(k)=\mathbf{g}(k) \tag{38}$$

where $\mathbf{g}(k)=[g_1(k)\ g_2(k)\ \dots\ g_L(k)]^{\mathrm{T}}$ with $g_i(k)$ chosen such that $\mathbf{w}(k+1)\in\psi_L(k)$ or, equivalently, $|g_i(k)|\le\gamma(k-i)$ for $i=1\ \dots\ L$, and the most general version has the recursions given by

$$\mathbf{w}(k+1)=\begin{cases}\mathbf{P}\left[\mathbf{w}(k)+\mathbf{X}(k)\left(\mathbf{X}^{\mathrm{H}}(k)\mathbf{P}\mathbf{X}(k)\right)^{-1}\left(\mathbf{e}^*(k)-\mathbf{g}^*(k)\right)\right]+\mathbf{F}, & \text{if } \left|e(k)\right|>\gamma(k)\\ \mathbf{w}(k), & \text{otherwise}\end{cases} \tag{39}$$

Choice of $g_i(k)$ can vary for different problems. A particularly simple SM-CAP version shown in Table 4 is obtained for fixed thresholds, $\gamma(k)=\gamma$, if $g_i(k)$ for $i\neq 1$ is chosen equal to the a posteriori error $g_i(k)=d(k-i+1)-\mathbf{x}^{\mathrm{T}}(k-i+1)\mathbf{w}^*(k)$, and $g_1(k)=e(k)/|e(k)|$ [27].

5.2 Partially-Adaptive Filters

Partial adaptation will be presented here in the context of the GSC structure, whereby only a subset of the $MN-p$ degrees of freedom is used. A modification in the structure of Fig. 3 is necessary to represent the reduction to $MN-p-L$ in the number of degrees of freedom (or parameters to be adapted),

Table 4. The SM-CAP Algorithm

SM-CAP with constant a posteriori errors

$\mathbf{w}(0) = \mathbf{F}$
 $\mathbf{u}_1 = [1\ 0\ \ldots\ 0]^T$
 for each k
 {
 $e(k) = d(k) - \mathbf{w}^H(k)\mathbf{x}(k)$
 if $\lvert e(k)\rvert > \gamma$
 $\alpha(k) = 1 - \gamma/\lvert e(k)\rvert$
 $\mathbf{t}(k) = [\mathbf{X}^H(k)\mathbf{P}\mathbf{X}(k) + \mathbf{I}]^{-1}\alpha(k)e(k)\mathbf{u}_1$
 $\mathbf{w}(k+1) = \mathbf{P}[\mathbf{w}(k) + \mathbf{X}(k)\mathbf{t}(k)] + \mathbf{F}$
 else
 $\mathbf{w}(k+1) = \mathbf{w}(k)$
 }

originally equal to $MN-p$. This reduction in dimension is carried out via an $L\times(MN-p)$ matrix **T**, to be placed after the blocking matrix **B**. The operation performed by matrix **T** can be incorporated in Fig. 3 as a modification in the blocking matrix, to be represented as $\mathbf{B}' = \mathbf{BT}$. This operation can also be alternatively viewed as a modification introduced in the set of constraints [22–25], represented by an auxiliary constraint matrix $\mathbf{C}_A$, such that:

- $\mathbf{C}_A$ is in the null space of **C**, i.e., $\mathbf{C}^H\mathbf{C}_A = \mathbf{0}$;
- the new composite blocking matrix $\mathbf{B}' = \mathbf{BT}$ spans the null space of the new modified constraint matrix $\mathbf{C}' = [\mathbf{C}\ \mathbf{C}_A]$; and
- the rank of matrix $[\mathbf{C}\ \mathbf{C}_A\ \mathbf{BT}]$ is equal to MN.

This new formulation of the minimization problem can be stated as follows:

$$\mathbf{w}_{opt} = \arg\min_{\mathbf{w}} \xi_w(k) \quad \text{subject to } \mathbf{C}'^H\mathbf{w} = \mathbf{f}' \tag{40}$$

where $\mathbf{f}' = [\mathbf{f}^T\ \mathbf{f}_A^T]^T$. The optimal solution to the new modified minimization problem with reduced dimension is

$$\mathbf{w}'_{opt} = \left(\mathbf{T}^H\mathbf{B}^H\mathbf{R}\mathbf{B}\mathbf{T}\right)^{-1}\mathbf{T}^H\mathbf{B}^H\mathbf{R}\mathbf{F} + \left(\mathbf{T}^H\mathbf{B}^H\mathbf{R}\mathbf{B}\mathbf{T}\right)^{-1}\mathbf{T}^H\mathbf{B}^H\mathbf{p} \tag{41}$$

This presentation of the partial adaptation of LCAF suggests that an efficient implementation is possible via the Householder method presented in Sect. 4.2.

A judicious choice of matrix **T** may influence positively performance of the system. The additional information may improve performance at the same time that it reduces the number of degrees of freedom, and consequently the computational complexity. In a beamforming application where some knowledge of interference is available, the extra L constraints can be used to minimize the output interference power [22–25].

Even in situations where no prior knowledge is available and extra constraints cannot be used, partial adaptation will still reduce computational complexity. An optimal approach to the selection of matrix **T**, in a sense that it provides the lowest MSE out of all possible matrices formed as a combination of L eigenvectors of matrix $\mathbf{B}^H\mathbf{R}\mathbf{B}$, is the one that selects the L eigenvectors that maximize the cross-spectral metric (CSM), defined as [9, 14]

$$\frac{\left|\mathbf{v}_i^H\mathbf{p}\right|^2}{\lambda_i} \tag{42}$$

where $(\lambda_i, \mathbf{v}_i)$ is the ith eigenvalue-eigenvector pair of $\mathbf{B}^H\mathbf{R}\mathbf{B}$, and $\mathbf{p}$ is the cross-correlation vector between $\mathbf{B}^H\mathbf{x}$ and $d_0 = \mathbf{F}^H\mathbf{x} - d$.

6 Discussions

Research on LCAF has targeted different optimization criteria yielding algorithms tailored to specific applications. Objective functions like the MOE may suffer from signal cancellation if interreference is correlated with the signal of interest. We can overcome this by changing the objective function, or by carefully designing the constraint equations, or by spatial smoothing. For example, the linearly-constrained constant-modulus algorithm [12, 20] has been used for blind equalization in multipath environments. Algorithms with fast convergence even for highly correlated input signals that compromise computational complexity with better robustness in finite-precision arithmetic have also been developed recently [2–4].

The multistage Wiener filter (MWF) [9] is an alternative implementation of the Wiener filter where partial adaptation via reduction of the number of degrees of freedom can be easily done [10]. In linearly-constrained problems the MWF can be used in the lower branch of the GSC structure. Contrary to other decompositions of the Wiener filter that aim at the best approximation to the covariance matrix (e.g., principal components analysis), the MWF acts on affine directions defined by cross-correlations and precise knowledge of signal-subspace dimension is not required for good performance.

Performance of LCAF is directly tied to a proper design of the constraints, which are application dependent and based on system knowledge. This chapter presented algorithms and structures that can be used directly in general problems, as well as tools for the development of more refined solutions.

7 References

1. Apolinário Jr. JA, Werner S, Laakso T, Diniz PSR (1998) Constrained normalized adaptive filtering for CDMA mobile communications. In: European Signal Processing Conference, pp 2053–2056
2. Apolinário Jr. JA, de Campos MLR, Bernal CP (2000) The constrained conjugate gradient algorithm. IEEE Signal Processing Letters 12: 351–354
3. Chern S-J, Chang C-Y (2002) Adaptive linearly constrained inverse QRD-RLS beamforming algorithm for moving jammers suppression. IEEE Transactions on Antennas and Propagation 8: 1138–1150
4. de Campos MLR, Werner S, Apolinário Jr. JA, Laakso TI (1998) Constrained quasi-Newton algorithm for CDMA mobile communications. In: Proceedings of the International Telecommunications Symposium, pp 371–376
5. de Campos MLR, Werner S, Apolinário Jr. JA (2002) Constrained adaptation algorithms employing Householder transformation. IEEE Transactions on Signal Processing 9: 2187–2195
6. Diniz PSR (2002) Adaptive filtering, algorithms and practical implementations, Second edition. Kluwer Academic Publishers, Boston, London, Dordrecht
7. Diniz PSR, Werner S (2003) Set-membership binormalized data reusing algorithms. IEEE Transactions on Signal Processing 1: 124–134
8. Frost III OL (1972) An algorithm for linearly constrained adaptive array processing. Proceedings of IEEE 8: 926–935
9. Goldstein JS, Reed IS, Scharf LL (1998) A multistage representation of the Wiener filter based on orthogonal projections. IEEE Transactions on Information Theory 7: 2943–2959
10. Goldstein JS, Reed IS, Zulch PA (1999) Multistage partially adaptive STAP CFAR detection algorithm. IEEE Transactions on Aerospace and Electronic Systems 2: 645–661
11. Gollamudi S, Nagaraj S, Kapoor S, Huang Y-F (1998) Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size. IEEE Signal Processing Letters 5: 111–114
12. Gray SD (1992) Multipath reduction using constant modulus conjugate gradient techniques. IEEE Journal on Selected Areas in Communications 8: 1300–1305
13. Griffiths LJ, Jim CW (1982) An alternative approach to linearly constrained adaptive beamforming. IEEE Transactions on Antennas and Propagation 1: 27–34
14. Guerci JR, Goldstein JS, Reed IS (2000) Optimal and adaptive reduced-rank STAP. IEEE Transactions on Aerospace and Electronic Systems 2: 647–663
15. Honig M, Tsatsanis MK (2000) Adaptive techniques for multiuser CDMA receivers. IEEE Signal Processing Magazine 3: 49–61
16. Medina CA, Apolinário Jr. JA, Siqueira MG (2002) A unified framework for multichannel fast QRD-LS adaptive filters based on backward prediction errors. In: Proceedings of the 45^{th} Midwest Symposium on Circuits and Systems, pp 668–671
17. Nagaraj S, Gollamudi S, Kapoor S, Huang Y-F, Deller Jr. JR (2000) Linear and adaptive interference suppression for CDMA systems with a worst-case estimation error criterion. IEEE Transactions on Signal Processing 1: 284–289
18. Resende LS, Romano JMT, Bellanger MG (1996) A fast least-squares algorithm for linearly constrained adaptive filtering. IEEE Transactions on Signal Processing 5: 1168–1174
19. Resende LS, Romano JMT, Bellanger MG (1998) Simplified FLS algorithm for linear phase adaptive filtering. In: European Signal Processing Conference, pp 1237–1240

20. Rude MJ, Griffiths LJ (1989) Incorporation of linear constraints into the constant modulus algorithm. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, pp 968–971
21. Tseng C-Y, Griffiths LJ (1988) A systematic procedure for implementing the blocking matrix in decomposed form. In: Proceedings of the 22nd Conference on Signals, Systems, and Computers, pp 808–812
22. Van Veen BD (1988) Eigenstructure based partially adaptive array design. IEEE Transactions on Antennas and Propagation 3: 357–362
23. Van Veen BD (1989) An analysis of several partially adaptive beamformer designs. IEEE Transactions on Acoustics, Speech, and Signal Processing 2: 192–203
24. Van Veen BD, Buckley KM (1988) Beamforming: A versatile approach to spatial filtering. IEEE ASSP Magazine, pp 4–24
25. Van Veen BD, Roberts RA (1987) Partially adaptive beamformer design via output power minimization. IEEE Transactions on Acoustics, Speech, and Signal Processing 11: 1524–1532
26. Werner S (2002) Reduced complexity adaptive filtering algorithms with applications to communications systems. D.Sc. thesis, Helsinki University of Technology
27. Werner S, Apolinário Jr. JA, de Campos MLR (2001) The data-selective constrained affine-projection algorithm. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, pp 3745–3748
28. Werner S, Apolinário Jr. JA, de Campos MLR (2003) On the equivalence of RLS Implementations of LCMV and GSC Processors. IEEE Signal Processing Letters 12: 356–359
29. Werner S, Diniz PSR (2001) Set-membership affine projection algorithm. IEEE Signal Processing Letters 8: 231–235