Signal Processing 149 (2018) 229–235

Contents lists available at ScienceDirect

Signal Processing

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

Short communication

Recursive linearly constrained minimum variance estimator in linear models with non-stationary constraints

François Vincent, Eric Chaumette*

Department of Electronics, Optronics and Signal, ISAE-SUPAERO, University of Toulouse, 10 Avenue Edouard Belin, Toulouse 31055, France

ARTICLE INFO

Article history:
Received 11 August 2017
Revised 18 March 2018
Accepted 21 March 2018
Available online 22 March 2018

Keywords:
Parameters estimation
Linearly constrained minimum variance estimator
Robust adaptive beamforming

ABSTRACT

In parameter estimation, it is common place to design a linearly constrained minimum variance estimator (LCMVE) to tackle the problem of estimating an unknown parameter vector in a linear regression model. So far, the LCMVE has been mainly studied in the context of stationary constraints in stationary or non-stationary environments, giving rise to well-established recursive adaptive implementations when multiple observations are available. In this communication, provided that the additive noise sequence is temporally uncorrelated, we determine the family of non-stationary constraints leading to LCMVEs which can be computed according to a predictor/corrector recursion similar to the Kalman Filter. A particularly noteworthy feature of the recursive formulation introduced is to be fully adaptive in the context of sequential estimation as it allows at each new observation to incorporate or not new constraints.

1. Introduction

In the signal processing literature dealing with parameter estimation, one of the most studied estimation problem is that of identifying the components of a N-dimensional observation vector ($\mathbf{y}$) formed from a linear superposition of P individual signals ($\mathbf{x}$) to noisy data ($\mathbf{v}$): $\mathbf{y} = \mathbf{Hx} + \mathbf{v}$[1], a.k.a. the linear regression problem, where $\mathbf{H}$ is a N-by-P matrix and $\mathbf{v}$ is a N-dimensional vector. The importance of this problem stems from the fact that a wide range of problems in communications, array processing, and many other areas can be cast in this form [1,2]. As in [3, Section 5.1], we adopt a joint proper complex signals assumption for $\mathbf{x}$ and $\mathbf{v}$, which allows to resort to standard estimation in the mean squared error (MSE) sense defined on the Hilbert space of complex random variables with finite second-order moment. A proper complex random variable is uncorrelated with its complex conjugate. Any result derived with joint proper complex random vectors are valid for real random vectors provided that one substitutes the matrix/vector transpose conjugate for the matrix/vector transpose. Additionally, it is assumed that: (a) $\mathbf{v}$ is zero mean, (b) $\mathbf{x}$ is uncorrelated with $\mathbf{v}$, (c) the model matrix $\mathbf{H}$ and the noise covariance matrix $\mathbf{C_v}$ are either known or specified according to known parametric models. In this setting, the weighted least squares estimator of $\mathbf{x}$ [4]:[2]

$$\widehat{\mathbf{x}}^b = \arg\min_{\mathbf{x}} \left\{ (\mathbf{y} - \mathbf{Hx})^H \mathbf{C}_\mathbf{v}^{-1} (\mathbf{y} - \mathbf{Hx}) \right\} \tag{1a}$$

$$= \left(\mathbf{H}^H \mathbf{C}_\mathbf{v}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{C}_\mathbf{v}^{-1} \mathbf{y}, \tag{1b}$$

coincides with the maximum-likelihood estimator [5], if $\mathbf{x}$ is deterministic and $\mathbf{v}$ is Gaussian, and is known to minimize the MSE matrix among all linear unbiased estimators of $\mathbf{x}$, that is $\widehat{\mathbf{x}}^b = \mathbf{W}^{bH}\mathbf{y}$ where [6]:

$$\mathbf{W}^b = \arg\min_{\mathbf{W}} \left\{ E\left[\left(\mathbf{W}^H\mathbf{y} - \mathbf{x}\right)\left(\mathbf{W}^H\mathbf{y} - \mathbf{x}\right)^H \right] \right\} \text{ s.t. } \mathbf{W}^H\mathbf{H} = \mathbf{I} \tag{2a}$$

$$= \mathbf{C}_\mathbf{v}^{-1}\mathbf{H}\left(\mathbf{H}^H \mathbf{C}_\mathbf{v}^{-1} \mathbf{H}\right)^{-1}, \tag{2b}$$

whatever $\mathbf{x}$ is deterministic or random. Furthermore, since the matrix $\mathbf{W}^b$ is as well the solution of [2,6]:

$$\mathbf{W}^b = \arg\min_{\mathbf{W}} \left\{ \mathbf{W}^H \mathbf{C}_\mathbf{v} \mathbf{W} \right\} \text{ s.t. } \mathbf{W}^H\mathbf{H} = \mathbf{I}, \tag{2c}$$

$\widehat{\mathbf{x}}^b$ is also known as the minimum variance distortion less response estimator (MVDRE) [1,2,6]. However, it is well known that the performance achievable by the MVDRE strongly depends on the accurate knowledge on the parametric model of the observations, that is on $\mathbf{H}$ and $\mathbf{C_v}$, and are not particularly robust in

* Corresponding author.
E-mail addresses: francois.vincent@isae.fr (F. Vincent), eric.chaumette@isae.fr (E. Chaumette).

[1] Throughout the present communication, scalars, vectors and matrices are represented, respectively, by italic, bold lowercase and bold uppercase characters. The scalar/matrix/vector transpose conjugate is indicated by the superscript ^{H}. $[\mathbf{A}\,\mathbf{B}]$ and $\begin{bmatrix}\mathbf{A}\\\mathbf{B}\end{bmatrix}$ denotes respectively the matrix resulting from the horizontal and the vertical concatenation of $\mathbf{A}$ and $\mathbf{B}$. $E[\cdot]$ denotes the expectation operator.

[2] The superscript ^{b} is used to remind the reader that the value under consideration is the "best" one according to a given criterion.

https://doi.org/10.1016/j.sigpro.2018.03.016
0165-1684/

the presence of various types of differences between the model and the actual environment [1, Section 6.6], [7, Section 1], [8]. Thus linearly constrained minimum variance estimators (LCMVEs) [6,9,10] have been developed in which additional linear constraints are imposed to make the MVDRE more robust [1, Section 6.7], [7, Section 1], [8]:

$$\mathbf{W}^b = \arg\min_{\mathbf{W}} \left\{ \mathbf{W}^H \mathbf{C}_\mathbf{v} \mathbf{W} \right\} \text{ s.t. } \mathbf{W}^H \mathbf{\Lambda} = \mathbf{\Gamma}, \ \mathbf{\Lambda} = [\mathbf{H} \ \mathbf{\Omega}], \ \mathbf{\Gamma} = [\mathbf{I} \ \mathbf{\Upsilon}], \tag{3a}$$

$$= \mathbf{C}_\mathbf{v}^{-1} \mathbf{\Lambda} \left(\mathbf{\Lambda}^H \mathbf{C}_\mathbf{v}^{-1} \mathbf{\Lambda} \right)^{-1} \mathbf{\Gamma}^H, \tag{3b}$$

where $\mathbf{\Omega}$ and $\mathbf{\Upsilon}$ are known matrices of the appropriate dimensions, at the expense of an increase of the minimum MSE achieved, since additional degrees of freedom are used by the LCMVE in order to satisfy these constraints. However, firstly, the closed-form solution of the LCMVE (3b) requires the inversion of $\mathbf{C}_\mathbf{v}$, which can be too computationally complex for numerous real-world applications. Secondly, $\mathbf{C}_\mathbf{v}$ may be unknown and must be learned by an adaptive technique. Interestingly enough, if $\mathbf{x}$ and $\mathbf{v}$ are uncorrelated, $\mathbf{C}_\mathbf{v}$ can be replaced by $\mathbf{C}_\mathbf{y}$ in (1b), (2b) and (3b), which means that either $\mathbf{C}_\mathbf{v}$ can be learned from auxiliary data containing noise only, if available, or $\mathbf{C}_\mathbf{y}$ can be used instead and learned from the observations. Therefore, when several observations $\{\mathbf{y}_1, \ldots, \mathbf{y}_k\}$ are available, adaptive implementations of the LCMVE have been developed resorting to constrained stochastic gradient [6,11], constrained recursive least squares [12,13] and constrained Kalman-type [14,15] algorithms. The known equivalence between the LCMVE and the generalized side lobe canceller processor [9,10,16] allows to resort as well to standard (unconstrained) stochastic gradient or recursive least squares [2]. These recursive algorithms belongs to the set of sequential estimation algorithms compatible with applications where the observations become available sequentially and, immediately upon receipt of new observations, it is desirable to determine new estimates based upon all previous observations (including the current ones). It is an attractive formulation for embedded systems in which computational time and memory are at a premium, since it does not require that all observations are available for simultaneous ("batch") processing. Last, this can be computationally beneficial in cases in which the number of observations is much larger than the number of signals [17].

However, the aforementioned recursive algorithms can only update sequentially the LCMVE (3b) in non-stationary environments, i.e. when the observation model changes over time ($\mathbf{y}_l = \mathbf{H}_l \mathbf{x} + \mathbf{v}_l$, $1 \le l \le k$), for a given set of linear constraints $\mathbf{W}^H \mathbf{\Lambda} = \mathbf{\Gamma}$ [2,6,11–15], which defines the set of recursive LCMVEs for stationary constraints. An example of a recursive LCMVE for non-stationary constraints in non-stationary environments is given by the MVDRE $\widehat{\mathbf{x}}_k^b$ of $\mathbf{x}$, based on observations up to and including time k. Indeed, provided that the additive noise sequence $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is temporally uncorrelated, $\widehat{\mathbf{x}}_k^b$ follows a predictor/corrector recursion similar to the Kalman Filter [17, Section 1] [18]:

$$\widehat{\mathbf{x}}_k^b = \widehat{\mathbf{x}}_{k-1}^b + \mathbf{W}_k^{bH} \left(\mathbf{y}_k - \mathbf{H}_k \widehat{\mathbf{x}}_{k-1}^b \right), \ \widehat{\mathbf{x}}_1^b = \left(\mathbf{H}_1^H \mathbf{C}_{\mathbf{v}_1}^{-1} \mathbf{H}_1 \right)^{-1} \mathbf{H}_1^H \mathbf{C}_{\mathbf{v}_1}^{-1} \mathbf{y}_1, \tag{4}$$

where $\mathbf{W}_k^b$ is analogous to a Kalman gain at time k. In this case, the set of constraints (2c) is non-stationary since it is defined as $\overline{\mathbf{W}}^H \overline{\mathbf{H}}_k = \mathbf{I}$, where $\overline{\mathbf{H}}_k$ is the matrix resulting from the vertical concatenation of k matrices $\mathbf{H}_1$, ... ,$\mathbf{H}_k$, and $\overline{\mathbf{W}}$ is an unknown matrix of the appropriate dimensions. Off course, from a theoretical point of view, if all the observations $\{\mathbf{y}_1, \ldots, \mathbf{y}_k\}$ are stacked into a single vector $\overline{\mathbf{y}}_k^T = \left(\mathbf{y}_1^T, \ldots, \mathbf{y}_k^T \right)$, the "batch form" (3b) obtained from $\overline{\mathbf{y}}_k$ allows to implement LCMVEs with non-stationnary constraints, which are, unfortunately, hardly likely to be computable as the size of $\overline{\mathbf{y}}_k$ increases. Therefore the novel contribution of the present communication is to introduce, provided that the additive noise sequence $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is temporally uncorrelated, the family of linear constraints yielding a LCMVE which can be computed recursively in the form of (4) in place of the "batch form" (3b). It appears that this family only contains non-stationary constraints, including the aforementioned MVDRE. A particularly noteworthy feature of the recursive formulation introduced is to be fully adaptive in the context of sequential estimation as it allows at each new observation to incorporate or not new constraints. The relevance of the proposed recursive formulation of the LCMVE is exemplified in Section 3 in the context of array processing.

2. Recursive linearly constrained minimum variance estimators

In the following: a) the vector space of complex matrices with N rows and P columns is denoted $\mathcal{M}_\mathbb{C}(N, P)$, b) the matrix resulting from the vertical concatenation of k matrices $\mathbf{A}_1$, ... , $\mathbf{A}_k$ of same column number is denoted $\overline{\mathbf{A}}_k$. We consider the linear measurement/observation model:

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{v}_k, \quad k \ge 1, \tag{5a}$$

where $\mathbf{x}$ is a P-dimensional complex unknown vector, $\mathbf{y}_k$ is a N_k-dimensional complex measurement/observation vector, $\mathbf{H}_k \in \mathcal{M}_\mathbb{C}(N_k, P)$ and the complex noise sequence $\{\mathbf{v}_k\}_{k \ge 1}$ is zero-mean and temporally uncorrelated. Then (5a) can be extended on a horizon of k points from the first observation as:

$$\overline{\mathbf{y}}_k = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_k \end{pmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_k \end{bmatrix} \mathbf{x} + \begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_k \end{pmatrix}$$
$$= \overline{\mathbf{H}}_k \mathbf{x} + \overline{\mathbf{v}}_k, \ \left| \begin{array}{c} \overline{\mathbf{y}}_k, \overline{\mathbf{v}}_k \in \mathcal{M}_\mathbb{C}(\mathcal{N}_k, 1) \\ \overline{\mathbf{H}}_k \in \mathcal{M}_\mathbb{C}(\mathcal{N}_k, P) \\ \mathcal{N}_k = \sum_{l=1}^k N_l \end{array} \right. . \tag{5b}$$

Let $\overline{\mathbf{W}}_k = \begin{bmatrix} \overline{\mathbf{D}}_{k-1} \\ \mathbf{W}_k \end{bmatrix}$ where $\overline{\mathbf{D}}_{k-1} \in \mathcal{M}_\mathbb{C}(\mathcal{N}_{k-1}, P)$ and $\mathbf{W}_k \in \mathcal{M}_\mathbb{C}(N_k, P)$. The aim is to look for the family of linear constraints:

$$\overline{\mathbf{W}}_k^H \overline{\mathbf{\Lambda}}_k = \mathbf{\Gamma}_k, \ \overline{\mathbf{\Lambda}}_k = \left[\overline{\mathbf{H}}_k \ \overline{\mathbf{\Omega}}_k \right], \ \mathbf{\Gamma}_k = [\mathbf{I} \ \mathbf{\Upsilon}_k], \tag{6}$$

where $\overline{\mathbf{\Omega}}_k$ and $\mathbf{\Upsilon}_k$ are known matrices of the appropriate dimensions, yielding a LCMVE $\widehat{\mathbf{x}}_k^b = \overline{\mathbf{W}}_k^{bH} \overline{\mathbf{y}}_k$ where (3a) and (3b):

$$\overline{\mathbf{W}}_k^b = \arg\min_{\overline{\mathbf{W}}_k} \left\{ \overline{\mathbf{W}}_k^H \mathbf{C}_{\overline{\mathbf{v}}_k} \overline{\mathbf{W}}_k \right\} \text{ s.t. } \overline{\mathbf{W}}_k^H \overline{\mathbf{\Lambda}}_k = \mathbf{\Gamma}_k \tag{7a}$$

$$= \mathbf{C}_{\overline{\mathbf{v}}_k}^{-1} \overline{\mathbf{\Lambda}}_k \left(\overline{\mathbf{\Lambda}}_k^H \mathbf{C}_{\overline{\mathbf{v}}_k}^{-1} \overline{\mathbf{\Lambda}}_k \right)^{-1} \mathbf{\Gamma}_k^H, \tag{7b}$$

which can be computed according to a predictor/corrector recursion of the form, $\forall k \ge 2$:

$$\widehat{\mathbf{x}}_k^b = \widehat{\mathbf{x}}_{k-1}^b + \mathbf{W}_k^{bH} \left(\mathbf{y}_k - \mathbf{H}_k \widehat{\mathbf{x}}_{k-1}^b \right) = \left(\mathbf{I} - \mathbf{W}_k^{bH} \mathbf{H}_k \right) \widehat{\mathbf{x}}_{k-1}^b + \mathbf{W}_k^{bH} \mathbf{y}_k. \tag{8}$$

A key point to solve the problem at hand is to notice that, since $\mathbf{C}_{\mathbf{v}_l, \mathbf{v}_k} = \mathbf{C}_{\mathbf{v}_k} \delta_k^l$, then for any $\overline{\mathbf{W}}_k$ satisfying (6):

$$\mathbf{P}_k\left(\overline{\mathbf{W}}_k\right) = \overline{\mathbf{W}}_k^H \mathbf{C}_{\overline{\mathbf{v}}_k} \overline{\mathbf{W}}_k = \overline{\mathbf{D}}_{k-1}^H \mathbf{C}_{\overline{\mathbf{v}}_{k-1}} \overline{\mathbf{D}}_{k-1} + \mathbf{W}_k^H \mathbf{C}_{\mathbf{v}_k} \mathbf{W}_k = \mathbf{P}_k\left(\overline{\mathbf{D}}_{k-1}, \ \mathbf{W}_k\right), \tag{9}$$

which suggests that some ad hoc linear constraints (6) could yield separable solutions for $\overline{\mathbf{D}}_{k-1}$ and $\mathbf{W}_k$, which is investigated in a first step.

- **First step**

If we recast $\overline{\mathbf{\Lambda}}_k = \left[\overline{\mathbf{H}}_k \ \overline{\mathbf{\Omega}}_k \right]$ as $\overline{\mathbf{\Lambda}}_k = \begin{bmatrix} \overline{\mathbf{\Phi}}_{k-1} \\ \mathbf{\Phi}_k \end{bmatrix}$ where $\overline{\mathbf{\Phi}}_{k-1} = \left[\overline{\mathbf{H}}_{k-1} \ \overline{\mathbf{\Omega}}_{k-1} \right]$ and $\mathbf{\Phi}_k = [\mathbf{H}_k \ \mathbf{\Omega}_k]$, then an equivalent form of (6) is:

$$\overline{\mathbf{W}}_k^H \overline{\mathbf{\Lambda}}_k = \mathbf{\Gamma}_k \ \Leftrightarrow \ \overline{\mathbf{D}}_{k-1}^H \overline{\mathbf{\Phi}}_{k-1} = \mathbf{\Gamma}_k - \mathbf{W}_k^H \mathbf{\Phi}_k. \tag{10}$$