Recursive Linearly Constrained Wiener Filter for Robust Multi-Channel Signal Processing

Jordi Vilà-Valls[a], Damien Vivet[a], Eric Chaumette[a], François Vincent[a], Pau Closas[b]

[a] University of Toulouse/ISAE-SUPAERO, DEOS Dept., 31055 Toulouse, France.
[b] Electrical and Computer Engineering Dept., Northeastern University, Boston, MA, USA.

Abstract

This article introduces a new class of recursive linearly constrained minimum variance estimators (LCMVEs) that provides additional robustness to modeling errors. To achieve that robustness, a set of non-stationary linear constraints are added to the standard LCMVE that allow for a closed form solution that becomes appealing in sequential implementations of the estimator. Indeed, a key point of such recursive LCMVE is to be fully adaptive in the context of sequential estimation as it allows optional constraints addition that can be triggered by a preprocessing of each new observation or external information on the environment. This methodology has significance in the popular problem of linear regression among others. Particularly, this article considers the general class of partially coherent signal (PCS) sources, which encompasses the case of fully coherent signal (FCS) sources. The article derivates the recursive LCMVE for this type of problems and investigates, analytically and through simulations, its robustness against mismatches on linear discrete state-space models. Both errors on system matrices and noise statistics uncertainty are considered. An illustrative multi-channel array processing example is treated to support the discussion, where results in different model mismatched scenarios are provided with respect to the standard case with only FCS sources.

Key words: Parameter estimation, linearly constrained minimum variance estimator, model mismatch, robust adaptive beamforming.

1. Introduction

In the literature of signal processing parameter estimation, one of the most studied estimation problems is that of identifying the components of an N-dimensional complex observation vector ($\mathbf{y}$) formed as the linear superposition

Email addresses: jordi.vila-valls@isae.fr (Jordi Vilà-Valls), damien.vivet@isae.fr (Damien Vivet), eric.chaumette@isae.fr (Eric Chaumette), francois.vincent@isae.fr (François Vincent), closas@northeastern.edu (Pau Closas)

of P individual complex signals ($\mathbf{x}$) and complex noisy data ($\mathbf{v}$), also known as (*a.k.a.*) linear regression problem[1],

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v},\ \mathbf{H} \in \mathbb{C}^{N\times P},\ \mathbf{v} \in \mathbb{C}^{N}. \tag{1}$$

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]. In addition, in many practical problems: *a*) $\mathbf{v}$ is zero mean; *b*) $\mathbf{x}$ is uncorrelated with $\mathbf{v}$; and *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, recall that the weighted least squares estimator of $\mathbf{x}$ [3],

$$\hat{\mathbf{x}}^b = \arg\min_{\mathbf{x}} \left\{ (\mathbf{y} - \mathbf{H}\mathbf{x})^H \mathbf{C}_{\mathbf{v}}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}) \right\} = \left(\mathbf{H}^H \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{y}, \tag{2a}$$

coincides with the maximum-likelihood estimator [4] if $\mathbf{x}$ is deterministic and $\mathbf{v}$ is Gaussian, and is known to minimize the mean-squared error (MSE) matrix (with respect to the Löwner ordering for positive Hermitian matrices [5]) among all linear unbiased estimators of $\mathbf{x}$, that is, $\hat{\mathbf{x}}^b = (\mathbf{L}^b)^H \mathbf{y}$, where [6]

$$\begin{aligned} \mathbf{L}^b &= \arg\min_{\mathbf{L}} \left\{ E\left[\left(\mathbf{L}^H\mathbf{y} - \mathbf{x}\right)\left(\mathbf{L}^H\mathbf{y} - \mathbf{x}\right)^H \right] \right\} \text{ s.t. } \mathbf{L}^H\mathbf{H} = \mathbf{I} \\ &= \mathbf{C}_{\mathbf{v}}^{-1}\mathbf{H}\left(\mathbf{H}^H\mathbf{C}_{\mathbf{v}}^{-1}\mathbf{H}\right)^{-1}, \end{aligned} \tag{2b}$$

regardless of $\mathbf{x}$ being a deterministic or random quantity. Furthermore, since the matrix $\mathbf{L}^b$ (2b) is as well the solution of [2, 6],

$$\mathbf{L}^b = \arg\min_{\mathbf{L}} \left\{ \mathbf{L}^H \mathbf{C_v} \mathbf{L} \right\} \text{ s.t. } \mathbf{L}^H\mathbf{H} = \mathbf{I}, \tag{2c}$$

$\hat{\mathbf{x}}^b$ is also known as the minimum variance distortionless response estimator/filter (MVDRE/MVDRF) [1, 2, 6]. However, it is well known that the achievable performance of the MVDRE strongly depends on the accurate knowledge of the observation parametric model - that is, on $\mathbf{H}$ and $\mathbf{C_v}$ [1, § 6.7], - and are particularly sensitive to different types of mismatches between the model and the actual signal [1, § 6.6][7, §1][8]. In order to robustify the MVDRE, the linearly constrained minimum variance estimator/filter (LCMVE/LCMVF) [6, 9]

[1] The notational convention adopted is as follows: scalars, vectors and matrices are represented, respectively, by italic, bold lowercase and bold uppercase characters. $\mathbf{O}$ and $\mathbf{0}$ stand for the all-zeros matrix and all-zeros column vector. The scalar/matrix/vector conjugate transpose is indicated by the superscript $(\cdot)^H$. $\mathbf{1}_N$ denotes an N-dimensional vector with components equal to 1. $\mathbf{I}$ is the identity matrix. $[\mathbf{A}\ \mathbf{B}]$ and $\left[\begin{smallmatrix}\mathbf{A}\\ \mathbf{B}\end{smallmatrix}\right]$ denote the matrix resulting from the horizontal and the vertical concatenation of matrices $\mathbf{A}$ and $\mathbf{B}$, respectively. The vector resulting from the vertical concatenation of k vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ is denoted as $\overline{\mathbf{a}}_k$. The matrix resulting from the vertical concatenation of k matrices $\mathbf{A}_1, \ldots, \mathbf{A}_k$ of same column dimension is denoted as $\overline{\mathbf{A}}_k$. $E[\cdot]$ denotes the expectation operator. If $\mathbf{x}$ and $\mathbf{y}$ are two complex random vectors: a) $\mathbf{m_x} \triangleq E[\mathbf{x}]$ and $\mathbf{m_y} \triangleq E[\mathbf{y}]$, b) $\mathbf{C_x}$, $\mathbf{C_y}$ and $\mathbf{C_{x,y}}$ are respectively the covariance matrices of $\mathbf{x}$, of $\mathbf{y}$, and the cross-covariance matrix of $\mathbf{x}$ and $\mathbf{y}$; c) if $\mathbf{C_y}$ is invertible, then $\mathbf{C_{x|y}} \triangleq \mathbf{C_x} - \mathbf{C_{x,y}}\mathbf{C_y}^{-1}\mathbf{C}_{\mathbf{x},\mathbf{y}}^H$. The superscript $(\cdot)^b$ denotes that the considered value is the "best" one according to a given criterion.

is leveraged in this article, in which additional linear constraints $\mathbf{L}^H \mathbf{\Omega} = \mathbf{\Upsilon}$ are imposed (where $\mathbf{\Omega}$ and $\mathbf{\Upsilon}$ are known matrices of the appropriate dimensions) [1, § 6.7][7, §1][8], that is,

$$\begin{aligned} \mathbf{L}^b &= \arg\min_{\mathbf{L}} \left\{ \mathbf{L}^H \mathbf{C}_{\mathbf{v}} \mathbf{L} \right\} \text{ s.t. } \mathbf{L}^H \left[\mathbf{H} \ \mathbf{\Omega} \right] = \left[\mathbf{I} \ \mathbf{\Upsilon} \right] \\ &= \mathbf{C}_{\mathbf{v}}^{-1} \left[\mathbf{H} \ \mathbf{\Omega} \right] \left(\left[\mathbf{H} \ \mathbf{\Omega} \right]^H \mathbf{C}_{\mathbf{v}}^{-1} \left[\mathbf{H} \ \mathbf{\Omega} \right] \right)^{-1} \left[\mathbf{I} \ \mathbf{\Upsilon} \right]^H . \end{aligned} \tag{3}$$

Robustness is understood as the ability to achieve *close-to-optimal* performance in situations with imperfect, incomplete, or erroneous knowledge about the system under consideration and its environment, while minimal impact on performance under nominal conditions is caused. This comes at the expense of an increase in the achieved MSE, since additional degrees of freedom are used by LCMVEs (3) in order to satisfy the additional constraints $\mathbf{L}^H \mathbf{\Omega} = \mathbf{\Upsilon}$. In this article we further explore the use of linear constraints to robustify the MVDRE in the context of recursive filtering, and for realistic multi-channel signal processing applications, with system model and noise statistics mismatch. Notice that $\mathbf{C}_{\mathbf{v}}$ may be unknown and must be learned by an adaptive technique. Remarkably, if $\mathbf{x}$ and $\mathbf{v}$ are uncorrelated, $\mathbf{C}_{\mathbf{v}}$ can be replaced by $\mathbf{C}_{\mathbf{y}}$ in (2a)(2b)(3), 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.

When several observations are available, $\mathbf{y}_l \in \mathbb{C}^{N_l}$, $1 \leq l \leq k$, with l the discrete-time index, recursive adaptive implementations of the LCMVE have been developed resorting to constrained stochastic gradient [6], constrained recursive least squares [10, 11] and constrained Kalman-type [12] algorithms. The equivalence between the LCMVE and the generalized side lobe canceller [9, 13] allows to resort as well to standard stochastic gradient or recursive least squares [2] solutions. However, the above recursive algorithms can only sequentially update the LCMVE (3) in non-stationary environments for a given set of linear constraints [2, 6, 10, 11, 12]. More explicitly, when the observation model changes over time

$$\mathbf{y}_l = \mathbf{H}_l \mathbf{x}_l + \mathbf{v}_l, \tag{4}$$

these recursive LCMVEs provide a solution for stationary constraints of the form $\mathbf{L}_l^H \left[\mathbf{H}_l \ \mathbf{\Omega} \right] = \left[\mathbf{I} \ \mathbf{\Upsilon} \right]$, which may not be the case of interest in practice.

On another note, in presence of fully coherent signal (FCS) sources, i.e. $\mathbf{x}_l = \mathbf{x}$, one can concatenate the available observations ($\mathbf{y}_l = \mathbf{H}_l \mathbf{x} + \mathbf{v}_l$, $1 \leq l \leq k$) to obtain an augmented observation model,

$$\overline{\mathbf{y}}_k = \overline{\mathbf{H}}_k \mathbf{x} + \overline{\mathbf{v}}_k, \quad \overline{\mathbf{y}}_k, \overline{\mathbf{v}}_k \in \mathbb{C}^{\mathcal{N}_k}, \quad \overline{\mathbf{H}}_k \in \mathbb{C}^{\mathcal{N}_k \times P}, \quad \mathcal{N}_k = \textstyle\sum_{l=1}^k N_l, \tag{5a}$$

and considering non-stationary constraints we have that

$$\hat{\mathbf{x}}_k^b = (\overline{\mathbf{L}}_k^b)^H \overline{\mathbf{y}}_k \tag{5b}$$

$$\overline{\mathbf{L}}_k^b = \arg\min_{\overline{\mathbf{L}}_k} \left\{ \overline{\mathbf{L}}_k^H \mathbf{C}_{\overline{\mathbf{v}}_k} \overline{\mathbf{L}}_k \right\} \text{ s.t. } \overline{\mathbf{L}}_k^H \left[\overline{\mathbf{H}}_k \ \overline{\mathbf{\Omega}}_k \right] = \left[\mathbf{I} \ \mathbf{\Upsilon}_k \right]. \tag{5c}$$

Provided that the noise sequence $\{\mathbf{v}_l\}_{l=1}^{k}$ is temporally uncorrelated, authors in [14] have lately introduced the family of recursive LCMVEs with non-stationary constraints associated to (5a)-(5c), which can equivalently be computed recursively according to a Kalman-like recursion [15, §1] as

$$\hat{\mathbf{x}}_k^b = \hat{\mathbf{x}}_{k-1}^b + (\mathbf{L}_k^b)^H \left(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k-1}^b\right). \tag{5d}$$

However, in real-life experiments some experimental factors, such as source motion, environmental non-stationarities, medium non-homogeneity, direction errors, or local scattering, may prevent from observing FCS sources. On the contrary, it is likely that one rather observes PCS sources because their amplitudes undergo a partial random walk between observations,

$$\mathbf{x}_1 = \mathbf{x},\ l \geq 2 : \mathbf{x}_l = \mathbf{F}_{l-1}\mathbf{x}_{l-1} + \mathbf{w}_{l-1},\ \mathbf{x}_1, \mathbf{x}_l, \mathbf{w}_{l-1} \in \mathbb{C}^P,\ \mathbf{F}_{l-1} \in \mathbb{C}^{P \times P}, \tag{6a}$$

where the fluctuation noise sequence $\{\mathbf{w}_l\}_{l=1}^{k-1}$ is a priori uncorrelated with $\mathbf{x}_1$ and the measurement noise sequence $\{\mathbf{v}_l\}_{l=1}^{k}$. The amplitude fluctuation model (6a) has a number of merits, including its simplicity and its capability to model most cases of PCS amplitudes, including the situation where $\mathbf{C}_{\mathbf{x}_l}$ is invariant, i.e. $\mathbf{C}_{\mathbf{x}_l} = \mathbf{C}_{\mathbf{x}_1}$, with an adjustable correlation matrix $\mathbf{C}_{\mathbf{x}_l,\mathbf{x}_{l-1}}$ between observations. Unfortunately, even in the case of perfect knowledge of the parametric model of the observations, a slight loss of coherence of the signal source introduces a severe breakdown on the MVDRE performance in the large sample regime if the loss of coherence is not considered (see Section 2.4 for an example). Thus, the most noteworthy merit of (6a) is to introduce an observation model belonging to the general class of linear discrete state-space (LDSS) models [15] represented with the state and measurement equations

$$k \geq 2 : \mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \qquad k \geq 1 : \mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k. \tag{6b}$$

Indeed, in this setting, the extension of the results derived in [14] to the case of PCS sources, under milder (w.r.t. [14]) regularity conditions on the noise covariance matrices, follows from [16] which very recently identified the family of linear constraints for which the linearly constrained Wiener filter (LCWF) associated to LDSS models can be computed recursively according to a Kalman-like recursion of the form

$$\hat{\mathbf{x}}_{k|k}^b = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1}^b + (\mathbf{L}_k^b)^H \left(\mathbf{y}_k - \mathbf{H}_k\mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1}^b\right). \tag{6c}$$

The main contribution of this article is the derivation of robust recursive LCMVEs for PCS sources, generalizing the results in [14, 16], in order to improve the performance of existing methodologies in realistic multi-channel signal processing applications. The main significance of the proposed methodology is herein summarized:

- If the parameters $\left(\mathbf{F}_{k-1}, \mathbf{C}_{\mathbf{w}_{k-1}}\right)$ of the fluctuation model (6a) are known, it allows for a recursive computation of the optimal estimate $\hat{\mathbf{x}}_k^b$ of the amplitude $\mathbf{x}_k$ of the sources (6c) and of its covariance error matrix.

- PCS sources introduce a lower limit in the achievable estimation performance in the large sample regime even with a perfect knowledge of the LDSS model (6b), as highlighted in Section 2.4. This allows for the computation of the estimator performance.
- If the parameters $\left(\mathbf{F}_{k-1}, \mathbf{C}_{\mathbf{w}_{k-1}}\right)$ are unknown, it nevertheless allows to perform a parametric study of the robustness of the LCMVE for FCS sources (5a) against partial coherency, by comparing its performance – assessed via Monte-Carlo simulations – and the best performance achievable for each likely (or possible) value of the parameters $\left(\mathbf{F}_{k-1}, \mathbf{C}_{\mathbf{w}_{k-1}}\right)$.
- Capability to mitigate both modelling errors in system matrices $(\mathbf{F}_{k-1}, \mathbf{H}_k)$ and system noise statistics uncertainty, in presence of FCS or PCS sources.

In that perspective, in complement of the results introduced in [14, 16], we provide a detailed analysis (illustrated by representative examples) of the robust recursive LCWF capabilities. Regarding the possible lack of knowledge on the statistics of the measurement noise $(\mathbf{v}_k)$ and/or the amplitude fluctuation noise $(\mathbf{w}_{l-1})$, we consider the case resulting from the addition of nuisances whose parametric models are partially known. This analysis relies on the particularly noteworthy feature of the recursive LCWF to be fully adaptive in the context of sequential estimation, as it allows optional addition of constraints that can be triggered by a preprocessing of each new observation or external information regarding the environment. This is a key feature since in numerous real-world applications [15] the observations become available sequentially and, immediately upon reception of new observations, it may be desirable to determine new estimates based upon all previous observations (including the current ones). It is also 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. Finally, this can be computationally beneficial in cases in which the number of observations is much larger than the number of signals [15].

As shown in Section 2.3, recursive LCMVEs matched to FCS or PCS are specific instantiations of the recursive LCWF where at time $k = 1$ a distortionless constraint is introduced and (implicitly) propagated from observation to observation in order to enforce the MVDR property of the estimator of signal source amplitudes. As a consequence, the MVDRE and LCMVEs are sub-optimal in terms of MSE but they do not depend on the prior knowledge (first and second order statistics) on the initial state $\mathbf{x}_1$. Hence, the MVDRE and LCMVEs can be pre-computed and their behaviour can be assessed in advance independently of the prior knowledge on $\mathbf{x}_1$. If this prior knowledge on the initial state $\mathbf{x}_1$ is available, then one should incorporate it into the standard form of the LCWF in order to attain the minimum achievable MSE for a given set of linear constraints.

The reminder of the paper is organized as follows. Section 2.1, along with the material introduced in the present section, sets the basics of the notation used in the paper as well as the mathematical modeling of the problem at hand. The main results on the proposed LCWF are provided in Sections 2.2 and 2.3.

Section 2.4 discusses a motivating example in the context of a multi-channel signal processing application, which is used throughout the article. Section 3 discusses the sensitivity to modeling mismatches on the system matrices, as well as mitigation strategies. Conversely, Section 4 investigates the case of mismatch on the statistical characterization of the noise. The paper concludes with some final remarks in Section 5.

2. On the LCWF, MVDRE and LCMVE for LDSS Models

2.1. Signal Model

As in [14][17, §5.1], we adopt a joint proper complex signals assumption for the set of vectors $(\mathbf{x}_1, \{\mathbf{w}_k\}, \{\mathbf{v}_k\})$ which allows to resort to standard estimation in the 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 conjugate transpose for the matrix/vector transpose. Moreover, we adopt the standard notation used in array processing for linear estimators, a.k.a. filter estimates or simply filters [1, §6][2, §2.5][17, §5.6], as in (2b)(2c)(3)(5c), in the case of LDSS models (6b) since one can define a "state-former" in the same way as a beamformer in array processing or a frequency-bin former in spectral analysis. Last, the estimate of $\mathbf{x}_l$ (6b) based on measurements up to and including time k is denoted $\hat{\mathbf{x}}_{l|k}$. Since the partial random walk (6a) of the individual signals $\mathbf{x}_1$ can be recast as, $k \geq 2$,

$$\mathbf{x}_k = \mathbf{B}_{k,1}\mathbf{x}_1 + \mathbf{G}_k\overline{\mathbf{w}}_{k-1}, \; \mathbf{G}_k\overline{\mathbf{w}}_{k-1} = \sum_{i=1}^{k-1} \mathbf{B}_{k,i+1}\mathbf{w}_i,$$
$$\mathbf{B}_{k,i} = \left| \begin{array}{cl} \mathbf{F}_{k-1}\mathbf{F}_{k-2}...\mathbf{F}_i & , k > i \\ \mathbf{I} & , k = i \\ \mathbf{0} & , k < i \end{array} \right. ,$$

where $\mathbf{G}_k \in \mathbb{C}^{P\times(k-1)P}$, $\overline{\mathbf{w}}_{k-1} \in \mathbb{C}^{(k-1)P}$ and $\mathbf{B}_{k,i} \in \mathbb{C}^{P\times P}$, the observation model becomes

$$\mathbf{y}_k = \mathbf{A}_k\mathbf{x}_1 + \mathbf{n}_k, \; \mathbf{A}_k = \mathbf{H}_k\mathbf{B}_{k,1}, \; \mathbf{n}_1 = \mathbf{v}_1, \; \mathbf{n}_{k\geq 2} = \mathbf{v}_k + \mathbf{H}_k\mathbf{G}_k\overline{\mathbf{w}}_{k-1} \tag{7a}$$

leading to the updated augmented observation model

$$\overline{\mathbf{y}}_k = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_k \end{pmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_k \end{bmatrix} \mathbf{x}_1 + \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_k \end{bmatrix} = \overline{\mathbf{A}}_k\mathbf{x}_1 + \overline{\mathbf{n}}_k. \tag{7b}$$

2.2. Formulation of the LCWF for LDSS Models

It is known that, if $\mathbf{x}$ and $\mathbf{y}$ are two zero mean complex random vectors, then the linear estimator of $\mathbf{x}$, i.e. $\hat{\mathbf{x}} = \mathbf{W}^H\mathbf{y}$, which minimizes the MSE matrix $\mathbf{P}(\mathbf{W}) = E\left[\left(\mathbf{W}^H\mathbf{y}-\mathbf{x}\right)\left(\mathbf{W}^H\mathbf{y}-\mathbf{x}\right)^H\right]$ with respect to the Löwner ordering [5], is the Wiener filter (WF) estimate [2, §2.4]. If $\mathbf{C_y}$ is invertible, the WF estimate is given by

$$\hat{\mathbf{x}}^b = (\mathbf{W}^b)^H\mathbf{y}, \quad \mathbf{W}^b = \arg\min_{\mathbf{W}}\{\mathbf{P}(\mathbf{W})\} = \mathbf{C}_{\mathbf{y}}^{-1}\mathbf{C}_{\mathbf{x},\mathbf{y}}^H, \quad \mathbf{P}\left(\mathbf{W}^b\right) = \mathbf{C}_{\mathbf{x}|\mathbf{y}}. \tag{8a}$$

If it is required to impose some linear constraints (LCs) on the filter coefficients $\mathbf{W}$, the WF (8a) becomes the linearly constrained WF (LCWF) defined as

$$\hat{\mathbf{x}}^b = (\mathbf{L}^b)^H\mathbf{y}, \ \mathbf{L}^b = \arg\min_{\mathbf{L}}\{\mathbf{P}(\mathbf{L})\} \ \text{s.t.} \ \mathbf{L}^H\boldsymbol{\Lambda} = \mathbf{T}, \tag{9a}$$

where $\boldsymbol{\Lambda}$ and $\mathbf{T}$ are known matrices of the appropriate dimensions. If $\boldsymbol{\Lambda}$ has full rank, then [2, (2.113)]

$$\mathbf{L}^b = \mathbf{W}^b + \mathbf{C}_{\mathbf{y}}^{-1}\boldsymbol{\Lambda}\left(\boldsymbol{\Lambda}^H\mathbf{C}_{\mathbf{y}}^{-1}\boldsymbol{\Lambda}\right)^{-1}\left(\mathbf{T}^H - \boldsymbol{\Lambda}^H\mathbf{W}^b\right), \tag{9b}$$

$$\mathbf{P}\left(\mathbf{L}^b\right) = \mathbf{P}\left(\mathbf{W}^b\right) + \left(\mathbf{T}^H - \boldsymbol{\Lambda}^H\mathbf{W}^b\right)^H\left(\boldsymbol{\Lambda}^H\mathbf{C}_{\mathbf{y}}^{-1}\boldsymbol{\Lambda}\right)^{-1}\left(\mathbf{T}^H - \boldsymbol{\Lambda}^H\mathbf{W}^b\right), \tag{9c}$$

where $\mathbf{W}^b$ is the unconstrained WF (8a). Among many successful applications of the WF [2], a widely studied application is the estimation of the state vector $\mathbf{x}_k \in \mathbb{C}^P$ of LDSS models (6b). In this context, the model matrices $\mathbf{F}_k$ and $\mathbf{H}_k$ are known, and the state noise sequence $\{\mathbf{w}_k\}$ and the measurement noise sequence $\{\mathbf{v}_k\}$, as well as the initial state $\mathbf{x}_1$, are random vectors with zero-mean values and known covariance and cross-covariance matrices.

The WF estimate of $\mathbf{x}_k$ based on measurements up to and including time k is then given by (8a) if $\mathbf{C}_{\overline{\mathbf{y}}_k}$ is invertible:

$$\hat{\mathbf{x}}_{k|k}^b = (\mathbb{W}_k^b)^H\overline{\mathbf{y}}_k, \quad \mathbb{W}_k^b = \arg\min_{\mathbb{W}_k}\left\{\mathbf{P}_{k|k}(\mathbb{W}_k)\right\} = \mathbf{C}_{\overline{\mathbf{y}}_k}^{-1}\mathbf{C}_{\mathbf{x}_k,\overline{\mathbf{y}}_k}^H, \quad \mathbf{P}_{k|k}^b = \mathbf{C}_{\mathbf{x}_k|\overline{\mathbf{y}}_k}, \tag{10a}$$

where $\mathbb{W}_k \in \mathbb{C}^{N_k\times P}$, $\mathbf{P}_{l|k}(\mathbb{W}_k) = E\left[\left(\mathbb{W}_k^H\overline{\mathbf{y}}_k - \mathbf{x}_l\right)\left(\mathbb{W}_k^H\overline{\mathbf{y}}_k - \mathbf{x}_l\right)^H\right]$ and $\mathbf{P}_{k|k}^b = \mathbf{P}_{k|k}\left(\mathbb{W}_k^b\right)$. Since the seminal paper of Kalman [15][19], it is known that, if $\{\mathbf{w}_k, \mathbf{v}_k, \mathbf{x}_1\}$ verify certain uncorrelation conditions, lately extended in [20],

$$\mathbf{C}_{\mathbf{w}_{k-1},\overline{\mathbf{y}}_{k-1}} = \mathbf{0}, \ \mathbf{C}_{\mathbf{v}_k,\overline{\mathbf{y}}_{k-1}} = \mathbf{0}, \ \forall k \geq 2, \tag{10b}$$

then $\hat{\mathbf{x}}_{k|k}^b$ (10a) admits the following convenient recursive form

$$\hat{\mathbf{x}}_{k|k}^b = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1}^b + (\mathbf{W}_k^b)^H\left(\mathbf{y}_k - \mathbf{H}_k\mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1}^b\right), \ k \geq 2, \tag{10c}$$

so-called the Kalman filter (KF) estimate of $\mathbf{x}_k$ [15][19], where $\mathbf{W}_k^b \in \mathbb{C}^{N_k\times P}$ is the KF gain matrix at time k. The KF gain matrix $\mathbf{W}_k^b$ minimizes the MSE matrix among all linear filters $\hat{\mathbf{x}}_{k|k}$ of the form

$$\hat{\mathbf{x}}_{k|k}(\mathbf{W}_k) = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1}^b + \mathbf{W}_k^H\left(\mathbf{y}_k - \mathbf{H}_k\mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1}^b\right), \tag{11a}$$

that is $\mathbf{W}_k^b = \arg\min_{\mathbf{W}_k} \left\{ \mathbf{P}_{k|k}^J (\mathbf{W}_k) \right\}$ where

$$\mathbf{P}_{k|k}^J (\mathbf{W}_k) = E\left[\left(\hat{\mathbf{x}}_{k|k}(\mathbf{W}_k) - \mathbf{x}_k \right) \left(\hat{\mathbf{x}}_{k|k}(\mathbf{W}_k) - \mathbf{x}_k \right)^H \right], \tag{11b}$$

and can be computed according to [20]

$$\mathbf{P}_{k|k-1} = \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}^b\mathbf{F}_{k-1}^H + \mathbf{C}_{\mathbf{w}_{k-1}} + \mathbf{F}_{k-1}\mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}}^H + \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}}\mathbf{F}_{k-1}^H \tag{11c}$$

$$\begin{aligned} \mathbf{S}_{k|k-1} &= \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^H + \mathbf{C}_{\mathbf{v}_k} + \mathbf{H}_k\mathbf{C}_{\mathbf{v}_k,\mathbf{x}_k}^H + \mathbf{C}_{\mathbf{v}_k,\mathbf{x}_k}\mathbf{H}_k^H \\ \mathbf{W}_k^b &= \mathbf{S}_{k|k-1}^{-1}\left(\mathbf{H}_k\mathbf{P}_{k|k-1} + \mathbf{C}_{\mathbf{v}_k,\mathbf{x}_k}\right) \qquad (11\text{d}) \\ \mathbf{P}_{k|k}^b &= \left(\mathbf{I} - (\mathbf{W}_k^b)^H\mathbf{H}_k\right)\mathbf{P}_{k|k-1} - (\mathbf{W}_k^b)^H\mathbf{C}_{\mathbf{v}_k,\mathbf{x}_k}. \qquad (11\text{e}) \end{aligned}$$

The above recursion (11c-11e) is also valid for $k = 1$ provided that $\mathbf{P}_{1|0} = \mathbf{C}_{\mathbf{x}_1}$ and $\hat{\mathbf{x}}_{1|1}^b = (\mathbf{W}_1^b)^H\mathbf{y}_1$. Introducing a set of LCs, i.e. $\mathbb{W}_k^H\overline{\mathbf{\Lambda}}_k = \mathbf{T}_k$, into model (10a) yields the following LCWF

$$\hat{\mathbf{x}}_{k|k}^b = (\mathbb{L}_k^b)^H\overline{\mathbf{y}}_k,\ \mathbb{L}_k^b = \arg\min_{\mathbb{L}_k}\left\{\mathbf{P}_{k|k}(\mathbb{L}_k)\right\} \text{ s.t. } \mathbb{L}_k^H\overline{\mathbf{\Lambda}}_k = \mathbf{T}_k, \tag{12a}$$

whose "batch form" solution is given by (9b)

$$\mathbb{L}_k^b = \mathbb{W}_k^b + \mathbf{C}_{\overline{\mathbf{y}}_k}^{-1}\overline{\mathbf{\Lambda}}_k\left(\overline{\mathbf{\Lambda}}_k^H\mathbf{C}_{\overline{\mathbf{y}}_k}^{-1}\overline{\mathbf{\Lambda}}_k\right)^{-1}\left(\mathbf{T}_k^H - \overline{\mathbf{\Lambda}}_k^H\mathbb{W}_k^b\right). \tag{12b}$$

Let us assume the following block matrix decomposition: $\mathbb{L}_k = \begin{bmatrix}\mathbb{J}_{k-1}\\ \mathbf{L}_k\end{bmatrix}$, $\mathbb{J}_{k-1} \in \mathbb{C}^{\mathcal{N}_{k-1}\times P}$ and $\mathbf{L}_k \in \mathbb{C}^{N_k\times P}$ leading to $\mathbb{L}_k^H\overline{\mathbf{y}}_k = \mathbb{J}_{k-1}^H\overline{\mathbf{y}}_{k-1} + \mathbf{L}_k^H\mathbf{y}_k$. It appears [16] that the subset of LCs $\mathbb{L}_k^H\overline{\mathbf{\Lambda}}_k = \mathbf{T}_k$ allowing to compute the "batch form" of the LCWF (12b) with a Kalman-like recursion, that is,

$$\hat{\mathbf{x}}_{k|k}^b = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1}^b + (\mathbf{L}_k^b)^H\left(\mathbf{y}_k - \mathbf{H}_k\mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1}^b\right),\ k \geq 2, \tag{13a}$$

consists of the following three types of LCs:

$$\mathcal{C}_k^1 \quad : \quad \left[\mathbb{J}_{k-1}^H\ \mathbf{L}_k^H\right]\begin{bmatrix}\mathbf{0}\\ \mathbf{\Delta}_k\end{bmatrix} = \mathbf{T}_k \tag{13b}$$

$$\mathcal{C}_k^2 \quad : \quad \left[\mathbb{J}_{k-1}^H\ \mathbf{L}_k^H\right]\begin{bmatrix}\overline{\mathbf{\Lambda}}_{k-1}\\ \mathbf{H}_k\mathbf{F}_{k-1}\mathbf{T}_{k-1}\end{bmatrix} = \mathbf{F}_{k-1}\mathbf{T}_{k-1} \tag{13c}$$

$$\mathcal{C}_k^3 \quad : \quad \left[\mathbb{J}_{k-1}^H\ \mathbf{L}_k^H\right]\begin{bmatrix}\overline{\mathbf{\Lambda}}_{k-1} & \mathbf{0}\\ \mathbf{H}_k\mathbf{F}_{k-1}\mathbf{T}_{k-1} & \mathbf{\Delta}_k\end{bmatrix} = \left[\mathbf{F}_{k-1}\mathbf{T}_{k-1}\ \mathbf{T}_k\right] \tag{13d}$$

where

- $\mathcal{C}_k^1$ is dedicated to introduce the first subset of LCs at time k,
- $\mathcal{C}_k^2$ corresponds to the *implicit* propagation at time k via recursion (13a) of the LCs already set from time 1 up to time $k-1$,

- $\mathcal{C}_k^3$ combines $\mathcal{C}_k^2$ and $\mathcal{C}_k^1$, that is, propagation of previously set LCs before time k and addition of a new subset of LCs at time k.

Under $\mathcal{C}_k^2$, $\mathbf{L}_k^b$ is given by $\mathbf{L}_k^b = \arg\min_{\mathbf{L}_k} \left\{ \mathbf{P}_{k|k}^J(\mathbf{L}_k) \right\}$ and can be computed from the "unconstrained" KF recursion (11c-11e). Under $\mathcal{C}_k^1$ and $\mathcal{C}_k^3$, $\mathbf{L}_k^b$ is given by $\mathbf{L}_k^b = \arg\min_{\mathbf{L}_k} \left\{ \mathbf{P}_{k|k}^J(\mathbf{L}_k) \right\}$ s.t. $\mathbf{L}_k^H \mathbf{\Delta}_k = \mathbf{T}_k$, and can be computed from the following "constrained" KF recursion,

$$\mathbf{P}_{k|k-1} = \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}^b\mathbf{F}_{k-1}^H + \mathbf{C}_{\mathbf{w}_{k-1}} + \mathbf{F}_{k-1}\mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}}^H + \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}}\mathbf{F}_{k-1}^H \tag{13e}$$

$$\begin{aligned}
\mathbf{S}_{k|k-1} &= \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^H + \mathbf{C}_{\mathbf{v}_k} + \mathbf{H}_k\mathbf{C}_{\mathbf{x}_k,\mathbf{v}_k} + \mathbf{C}_{\mathbf{v}_k,\mathbf{x}_k}\mathbf{H}_k^H \\
\mathbf{W}_k &= \mathbf{S}_{k|k-1}^{-1}\left(\mathbf{H}_k\mathbf{P}_{k|k-1} + \mathbf{C}_{\mathbf{v}_k,\mathbf{x}_k}\right) \\
\mathbf{\Gamma}_k &= \mathbf{T}_k^H - \mathbf{\Delta}_k^H\mathbf{W}_k, \quad \mathbf{\Psi}_k = \mathbf{\Delta}_k^H\mathbf{S}_{k|k-1}^{-1}\mathbf{\Delta}_k \\
\mathbf{L}_k^b &= \mathbf{W}_k + \mathbf{S}_{k|k-1}^{-1}\mathbf{\Delta}_k\mathbf{\Psi}_k^{-1}\mathbf{\Gamma}_k
\end{aligned} \tag{13f}$$

$$\mathbf{P}_{k|k}^b = \left(\mathbf{I} - \mathbf{W}_k^H\mathbf{H}_k\right)\mathbf{P}_{k|k-1} - \mathbf{W}_k^H\mathbf{C}_{\mathbf{v}_k,\mathbf{x}_k} + \mathbf{\Gamma}_k^H\mathbf{\Psi}_k^{-1}\mathbf{\Gamma}_k \tag{13g}$$

The above recursion (13e-13g) is also valid for $k = 1$ provided that $\mathbf{P}_{1|0} = \mathbf{C}_{\mathbf{x}_1}$ and $\hat{\mathbf{x}}_{1|1}^b = (\mathbf{L}_1^b)^H\mathbf{y}_1$. The case of a non-zero mean initial state $\mathbf{x}_1$, with mean $\mathbf{m}_{\mathbf{x}_1}$, is addressed by simply setting

$$\hat{\mathbf{x}}_{1|1}^b = \mathbf{m}_{\mathbf{x}_1} + (\mathbf{W}_1^b)^H\left(\mathbf{y}_1 - \mathbf{H}_1\mathbf{m}_{\mathbf{x}_1}\right) \ ; \ \hat{\mathbf{x}}_{1|1}^b = \mathbf{m}_{\mathbf{x}_1} + (\mathbf{L}_1^b)^H\left(\mathbf{y}_1 - \mathbf{H}_1\mathbf{m}_{\mathbf{x}_1}\right).$$

Interestingly, the recursive formulation of the LCWF for LDSS models introduced is fully adaptive in the context of sequential estimation as it allows at each new observation to incorporate or not new LCs. Lastly, since the KF is the recursive form of the WF obtained for LDSS models, it makes sense to refer to linearly constrained KF (LCKF) to denote the recursive form of the LCWF obtained for LDSS models.

2.3. MVDRE and LCMVE for LDSS Models

As the computation of both KF and LCKF depends on prior information on the mean ($\mathbf{m}_{\mathbf{x}_1}$) and covariance matrix ($\mathbf{C}_{\mathbf{x}_1}$) of $\mathbf{x}_1$, they can be looked upon as "initial state first and second order statistics" matched filters. However in numerous applications $\mathbf{m}_{\mathbf{x}_1}$ and/or $\mathbf{C}_{\mathbf{x}_1}$ are unknown. A commonly used solution to circumvent this lack of prior information is the Fisher initialization [21][22, §II]. The Fisher initialization consists in initializing the KF recursion at time $k = 1$ with the weighted least squares estimator of $\mathbf{x}_1$ (2a) associated to the measurement model (6b), which coincides with the MVDRE of $\mathbf{x}_1$

$$\hat{\mathbf{x}}_{1|1}^b = (\mathbf{L}_1^b)^H\mathbf{y}_1, \ \mathbf{L}_1^b = \mathbf{C}_{\mathbf{v}_1}^{-1}\mathbf{H}_1\left(\mathbf{H}_1^H\mathbf{C}_{\mathbf{v}_1}^{-1}\mathbf{H}_1\right)^{-1}, \ \mathbf{P}_{1|1}^b = \left(\mathbf{H}_1^H\mathbf{C}_{\mathbf{v}_1}^{-1}\mathbf{H}_1\right)^{-1}, \tag{14a}$$

that is the LCKF defined by

$$\hat{\mathbf{x}}_{1|1}^b = (\mathbf{L}_1^b)^H\mathbf{y}_1, \ \mathbf{L}_1^b = \arg\min_{\mathbf{L}_1}\left\{\mathbf{P}_{1|1}(\mathbf{L}_1)\right\} \text{ s.t. } \mathbf{L}_1^H\mathbf{H}_1 = \mathbf{I}. \tag{14b}$$

A closer examination of LCs $\mathcal{C}_k^2$ and $\mathcal{C}_k^3$ shows that, if a time $k = 1$ the LCs $\mathbf{L}_1^H \mathbf{H}_1 = \mathbf{I}$ are set (as in (14b)), then the LCs $\mathbb{L}_k^H \overline{\mathbf{A}}_k = \mathbf{B}_{k,1}$ are set at time k. In this case, according to (7b),

$$\begin{aligned}\hat{\mathbf{x}}_{k|k} &= \mathbb{L}_k^H \overline{\mathbf{y}}_k = \mathbb{L}_k^H \overline{\mathbf{A}}_k \mathbf{x}_1 + \mathbb{L}_k^H \overline{\mathbf{n}}_k = (\mathbf{B}_{k,1}\mathbf{x}_1 + \mathbf{G}_k \overline{\mathbf{w}}_{k-1}) + \mathbb{L}_k^H \overline{\mathbf{n}}_k - \mathbf{G}_k \overline{\mathbf{w}}_{k-1} \\ &= \mathbf{x}_k + \mathbb{L}_k^H \overline{\mathbf{n}}_k - \mathbf{G}_k \overline{\mathbf{w}}_{k-1}, \end{aligned} \tag{14c}$$

which means that $\mathbb{L}_k$ is a distortionless response filter. Thus the use at time $k = 1$ of LCs of the form

$$\mathbf{L}_1^H \boldsymbol{\Delta}_1 = \mathbf{T}_1, \ \{\boldsymbol{\Delta}_1 = \mathbf{H}_1, \mathbf{T}_1 = \mathbf{I}\} \ \text{or} \ \{\boldsymbol{\Delta}_1 = [\mathbf{H}_1 \ \boldsymbol{\Omega}_1], \mathbf{T}_1 = [\mathbf{I} \ \boldsymbol{\Upsilon}_1]\}, \tag{15a}$$

leading to

$$\hat{\mathbf{x}}_{1|1}^b = (\mathbf{L}_1^b)^H \mathbf{y}_1, \mathbf{L}_1^b = \mathbf{C}_{\mathbf{v}_1}^{-1} \boldsymbol{\Delta}_1 \left(\boldsymbol{\Delta}_1^H \mathbf{C}_{\mathbf{v}_1}^{-1} \boldsymbol{\Delta}_1\right)^{-1} \mathbf{T}_1^H, \tag{15b}$$

$$\mathbf{P}_{1|1}^b = \mathbf{T}_1 \left(\boldsymbol{\Delta}_1^H \mathbf{C}_{\mathbf{v}_1}^{-1} \boldsymbol{\Delta}_1\right)^{-1} \mathbf{T}_1^H, \tag{15c}$$

combined with any combination of LCs $\mathcal{C}_l^2$ and $\mathcal{C}_l^3$, $2 \leq l \leq k$, transforms the LCKF into either a MVDRE or a LCMVE. Although the MVDRE and LCMVEs are sub-optimal in terms of MSE (due to the LCs (15a) introduced at time $k = 1$), they have a number of merits: a) according to (15b) they do not depend on the prior knowledge (first and second order statistics) on the initial state $\mathbf{x}_1$, b) they may outperform the LCKF in case of misspecification of the prior knowledge on $\mathbf{x}_1$ [20, 23]. In other words, the MVDRE and LCMVEs can be pre-computed and their behaviour can be assessed in advance independently of the prior knowledge on $\mathbf{x}_1$. As the MVDRE is a special case of LCMVEs, in the following we only mention LCMVEs.

2.4. On the Impact of PCS Sources on Recursive MVDRE Performance

To further motivate the need of robust recursive MVDRE in real-life multi-channel signal processing applications, we first show an example for the performance degradation on the achievable performance due to PCS sources (w.r.t. the FCS source hypothesis in [14]). This example is considered throughout the paper to support the discussion on the proposed methodologies.

Let us consider a uniform linear array with $N = 21$ sensors equally spaced at $\hat{d} = \lambda/2$ (half-wavelength) and an impinging signal source x_1 with broadside angle $\alpha = 10^{\circ}$, embedded in a spatially and temporally white noise,

$$\mathbf{y}_k = \mathbf{h}_k\left(\hat{d}, \alpha\right) x_1 + \mathbf{v}_k, \mathbf{h}_k^T(d, \alpha) = \left(1, \ldots, e^{j2\pi \frac{(N-1)d\sin(\alpha)}{\lambda}}\right), \mathbf{C}_{\mathbf{v}_l, \mathbf{v}_k} = \mathbf{I}\delta_k^l. \tag{16a}$$

The signal source x_1 is random, Gaussian complex circular with unit variance ($C_{x_1} = 1$), and is assumed to be fully coherent ($x_k = x_1$). However, fluctuation of the propagation medium are sometime unavoidable during the whole observation time interval, which prevents from observing a perfectly coherent signal source. Indeed, the random fluctuation of the propagation medium induces a random fluctuation of the signal amplitude. If the propagation medium fluctuations are small, then the mean power received from the signal source remains unchanged [18], which can be modeled via (6a) as:

$$x_k = f_{k-1} x_{k-1} + w_{k-1},\ C_{x_k} = C_{x_1},\ \rho_{x_{k-1},x_k} = C_{x_k,x_{k-1}}/C_{x_{k-1}} = f_{k-1},\tag{16b}$$

where f_{k-1} is the correlation coefficient between x_{k-1} and x_k which fully characterizes the loss of coherence between observation $k-1$ and k ($|f_{k-1}|^2 \leq 1$). Firstly, we investigate the impact of a slight loss of coherence of the signal source on the performance of the recursive MVDRE [14, (17a)-(17c)] computed under the hypothesis of a FCS source, that is:

$$\hat{\mathbf{x}}_k^b = (\overline{\mathbf{L}}_k^b)^H \overline{\mathbf{y}}_k,\ \overline{\mathbf{L}}_k^b = \arg\min_{\overline{\mathbf{L}}_k} \left\{ \overline{\mathbf{L}}_k^H \mathbf{C}_{\overline{\mathbf{v}}_k} \overline{\mathbf{L}}_k \right\} \text{ s.t. } \overline{\mathbf{L}}_k^H \overline{\mathbf{H}}_k = \mathbf{I}.$$

To this end, we compute the MSE in the estimation of x_k, both for a FCS source (reference case) denoted "MVDRE (FCS)", and for a PCS source, denoted by "MVDRE Mismatched to PCS". Secondly, we highlight the benefit of the formulation of a recursive MVDRE taking into account (16b), with f_{k-1} and $C_{w_{k-1}}$ known, and denoted as "MVDRE Matched to PCS". The results are summarized in Fig. 1, where the empirical MSE (denoted "...(Sim)...") is assessed with 10^4 Monte-Carlo trials. Three cases of very small loss of coherence are considered $\left(\sigma_{w_l}^2 = \sigma_w^2 \in \left\{10^{-6}, 10^{-5}, 10^{-4}\right\}\right)$. Fig. 1 clearly exemplifies the impact of a slight loss of coherence of the signal source on the MVDRE performance in the large sample regime, which introduces a severe performance breakdown when the loss of coherence is not taken into account. Thanks to the results introduced in Section 2.2, we can also evaluate which is the minimum achievable MSE when the amplitude fluctuation model is known (16b). Fig. 1 also shows that, when the signal source amplitude becomes partially coherent, there exists a lower limit in the achievable MSE, and an optimal number of observations that can be combined to estimate the amplitude with a nearly minimum achievable MSE. Hence, the significance of the derivation of a recursive LCMVE for PCS sources introduced in this section.

3. Mitigation of Modelling Errors in System Matrices

Since the LCKF and LCMVEs of $\mathbf{x}_k$ are based on the measurements and our knowledge of the model dynamics, any mismatch between the true model dynamics and the the assumed model dynamics leads to a suboptimal filter, and possibly to a filter with bad performance, as the discrepancy between the two models increases. Thus we consider the situation where we do not know perfectly the system matrices $(\mathbf{F}_{k-1}, \mathbf{H}_k)$, i.e. there is the true LDSS model and the one we assume:

$$\text{True}: \begin{cases} \mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1} \\ \mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \end{cases} \quad \text{Assumed}: \begin{cases} \widetilde{\mathbf{x}}_k = \widehat{\mathbf{F}}_{k-1}\widetilde{\mathbf{x}}_{k-1} + \mathbf{w}_{k-1} \\ \mathbf{y}_k = \widehat{\mathbf{H}}_k \widetilde{\mathbf{x}}_k + \mathbf{v}_k \end{cases}.\tag{18}$$

Let us denote $d\mathbf{F}_{k-1} = \mathbf{F}_{k-1} - \widehat{\mathbf{F}}_{k-1}$ and $d\mathbf{H}_k = \mathbf{H}_k - \widehat{\mathbf{H}}_k$.

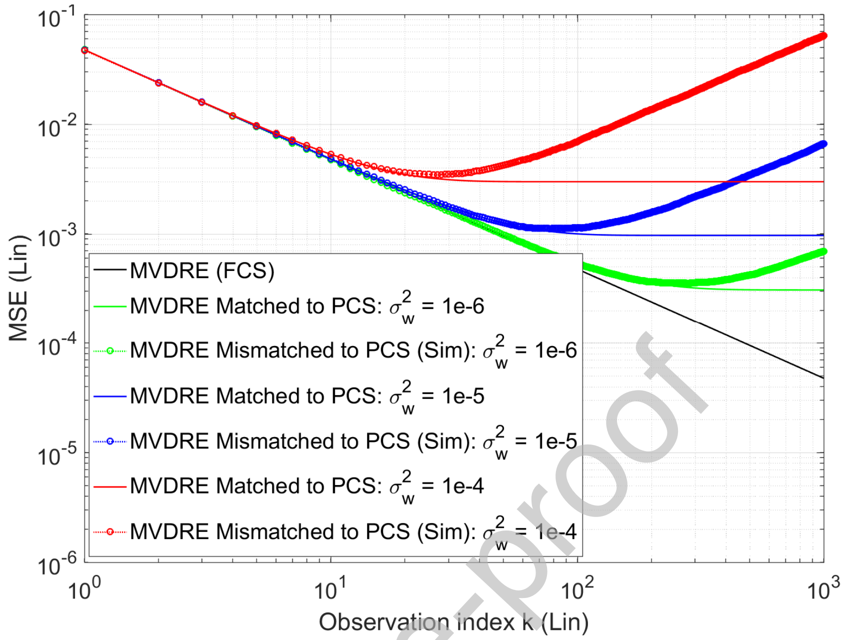

Figure 1: MSE of the recursive MVDRE of x_k (16a)-(16b) over time.

3.1. Parametric Modelling Errors

The existence of uncertainty on system matrices $(\mathbf{F}_{k-1}, \mathbf{H}_k)$ can be illustrated by the case where a parametric modelling of system matrices $(\mathbf{F}_{k-1}, \mathbf{H}_k)$ is known: $\mathbf{F}_{k-1} \triangleq \mathbf{F}_{k-1}(\boldsymbol{\omega}) = \left[\mathbf{f}_{k-1}^1(\boldsymbol{\omega}) \ \ldots \ \mathbf{f}_{k-1}^P(\boldsymbol{\omega})\right]$ and $\mathbf{H}_k \triangleq \mathbf{H}_k(\boldsymbol{\theta}) = \left[\mathbf{h}_k^1(\boldsymbol{\theta}) \ \ldots \ \mathbf{h}_k^P(\boldsymbol{\theta})\right]$, where $\boldsymbol{\omega}$ and $\boldsymbol{\theta}$ are supposed to be deterministic vector values determined via an ad hoc calibration process. In many cases, such calibration process provides estimates $\hat{\boldsymbol{\omega}} = \boldsymbol{\omega} + d\hat{\boldsymbol{\omega}}$ and $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + d\hat{\boldsymbol{\theta}}$ of the true values $\boldsymbol{\omega}$ and $\boldsymbol{\theta}$. If the calibration process is accurate enough, i.e. $d\hat{\boldsymbol{\omega}}$ and $d\hat{\boldsymbol{\theta}}$ are small, then the true state and measurement matrices, i.e. $\mathbf{F}_{k-1} \triangleq \mathbf{F}_{k-1}(\boldsymbol{\omega})$ and $\mathbf{H}_k \triangleq \mathbf{H}_k(\boldsymbol{\theta})$, differ from the assumed ones, i.e. $\widehat{\mathbf{F}}_{k-1} = \mathbf{F}_{k-1}(\hat{\boldsymbol{\omega}})$ and $\widehat{\mathbf{H}}_k = \mathbf{H}_k\left(\hat{\boldsymbol{\theta}}\right)$, via first order Taylor series

$$\mathbf{H}_k \simeq \widehat{\mathbf{H}}_k + d\mathbf{H}_k, \qquad d\mathbf{H}_k = -\left[\frac{\partial \mathbf{h}_k^1\left(\hat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}^T} d\hat{\boldsymbol{\theta}} \ \ldots \ \frac{\partial \mathbf{h}_k^P\left(\hat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}^T} d\hat{\boldsymbol{\theta}}\right] \tag{19a}$$

$$\mathbf{F}_{k-1} \simeq \widehat{\mathbf{F}}_{k-1} + d\mathbf{F}_{k-1}, \quad d\mathbf{F}_{k-1} = -\left[\frac{\partial \mathbf{f}_{k-1}^1(\hat{\boldsymbol{\omega}})}{\partial \boldsymbol{\omega}^T} d\hat{\boldsymbol{\omega}} \ \ldots \ \frac{\partial \mathbf{f}_{k-1}^P(\hat{\boldsymbol{\omega}})}{\partial \boldsymbol{\omega}^T} d\hat{\boldsymbol{\omega}}\right] \tag{19b}$$

3.2. Impact of Modelling Errors in System Matrices

At time $k = 1$, any LCKF or LCMVE of $\mathbf{x}_1$ is of the form $\hat{\mathbf{x}}_{1|1}(\mathbf{L}_1) = \mathbf{L}_1^H \mathbf{y}_1$ where $\mathbf{L}_1 \triangleq \mathbf{L}_1^b$ is the solution of $\mathbf{L}_1^b = \arg\min_{\mathbf{L}_1} \left\{\mathbf{P}_{1|1}(\mathbf{L}_1)\right\}$ or the solution of $\mathbf{L}_1^b = \arg\min_{\mathbf{L}_1} \left\{\mathbf{P}_{1|1}(\mathbf{L}_1)\right\}$ s.t. $\mathbf{L}_1^H \boldsymbol{\Delta}_1 = \mathbf{T}_1$, computed with the assumed LDSS model. Since $\mathbf{y}_1 = \widehat{\mathbf{H}}_1 \mathbf{x}_1 + \mathbf{v}_1 + d\mathbf{H}_1 \mathbf{x}_1$, the error made by the assumed filter in estimating the true $\mathbf{x}_1$ is given by

$$\hat{\mathbf{x}}_{1|1} - \mathbf{x}_1 = -\left(\mathbf{I} - \mathbf{L}_1^H \widehat{\mathbf{H}}_1\right)\mathbf{x}_1 + \mathbf{L}_1^H \mathbf{v}_1 + \boldsymbol{\varepsilon}_1(\mathbf{L}_1), \tag{20a}$$

$$\boldsymbol{\varepsilon}_1(\mathbf{L}_1) = \mathbf{L}_1^H d\mathbf{H}_1 \mathbf{x}_1. \tag{20b}$$

At time $k \geq 2$, provided that LCs (13b)-(13d) are considered, any LCKF or LCMVE of $\mathbf{x}_k$ is obtained from the Kalman-like recursion

$$\hat{\mathbf{x}}_{k|k}(\mathbf{L}_k) = \widehat{\mathbf{F}}_{k-1}\hat{\mathbf{x}}_{k-1|k-1}^b + \mathbf{L}_k^H\left(\mathbf{y}_k - \widehat{\mathbf{H}}_k \widehat{\mathbf{F}}_{k-1}\hat{\mathbf{x}}_{k-1|k-1}^b\right) \tag{21}$$

where $\mathbf{L}_k \triangleq \mathbf{L}_k^b$ is the solution of $\mathbf{L}_k^b = \arg\min_{\mathbf{L}_k}\left\{\mathbf{P}_{k|k}^J(\mathbf{L}_k)\right\}$ (11c-11e), or the solution of $\mathbf{L}_k^b = \arg\min_{\mathbf{L}_k}\left\{\mathbf{P}_{k|k}^J(\mathbf{L}_k)\right\}$ s.t. $\mathbf{L}_k^H \boldsymbol{\Delta}_k = \mathbf{T}_k$ (13e-13g), computed with the assumed LDSS model. Since

$$\begin{aligned}
\mathbf{F}_{k-1}\mathbf{x}_{k-1} &= \widehat{\mathbf{F}}_{k-1}\mathbf{x}_{k-1} + d\mathbf{F}_{k-1}\mathbf{x}_{k-1}, \\
\mathbf{y}_k &= \widehat{\mathbf{H}}_k\left(\widehat{\mathbf{F}}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}\right) + \mathbf{v}_k + d\mathbf{H}_k\mathbf{x}_k + \widehat{\mathbf{H}}_k d\mathbf{F}_{k-1}\mathbf{x}_{k-1},
\end{aligned}$$

the error made by the assumed filter (21) in estimating the true $\mathbf{x}_k$ is given by

$$\begin{aligned}
\hat{\mathbf{x}}_{k|k}(\mathbf{L}_k) - \mathbf{x}_k = &\left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right)\left(\widehat{\mathbf{F}}_{k-1}\left(\hat{\mathbf{x}}_{k-1|k-1}^b - \mathbf{x}_{k-1}\right) - \mathbf{w}_{k-1}\right) + \mathbf{L}_k^H \mathbf{v}_k \\
&+ \boldsymbol{\varepsilon}_k(\mathbf{L}_k),
\end{aligned} \tag{22a}$$

$$\begin{aligned}
\boldsymbol{\varepsilon}_k(\mathbf{L}_k) = &\left(\mathbf{L}_k^H d\mathbf{H}_k\left(\widehat{\mathbf{F}}_{k-1} + d\mathbf{F}_{k-1}\right) - \left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right) d\mathbf{F}_{k-1}\right)\mathbf{x}_{k-1} \\
&+ \left(\mathbf{L}_k^H d\mathbf{H}_k\right)\mathbf{w}_{k-1}.
\end{aligned} \tag{22b}$$

3.3. Mitigation of Modelling Errors in System Matrices

At time $k = 1$, if the subset of gain matrices $\mathcal{L}_1 = \left\{\mathbf{L} \in \mathbb{C}^{N_1 \times P} \mid \boldsymbol{\varepsilon}_1(\mathbf{L}) = \mathbf{0}\right\}$ is non empty, then $\forall \mathbf{L}_1 \in \mathcal{L}_1$, (20a) and $\mathbf{P}_{1|1}(\mathbf{L}_1)$ reduce to

$$\hat{\mathbf{x}}_{1|1} - \mathbf{x}_1 = -\left(\mathbf{I} - \mathbf{L}_1^H \widehat{\mathbf{H}}_1\right)\mathbf{x}_1 + \mathbf{L}_1^H \mathbf{v}_1, \tag{23a}$$

$$\begin{aligned}
\mathbf{P}_{1|1}(\mathbf{L}_1) = &\left(\mathbf{I} - \mathbf{L}_1^H \widehat{\mathbf{H}}_1\right)\mathbf{C}_{\mathbf{x}_1}\left(\mathbf{I} - \mathbf{L}_1^H \widehat{\mathbf{H}}_1\right)^H + \mathbf{L}_1^H \mathbf{C}_{\mathbf{v}_1}\mathbf{L}_1 \\
&- \left(\mathbf{I} - \mathbf{L}_1^H \widehat{\mathbf{H}}_1\right)\mathbf{C}_{\mathbf{x}_1,\mathbf{v}_1} - \mathbf{C}_{\mathbf{x}_1,\mathbf{v}_1}^H\left(\mathbf{I} - \mathbf{L}_1^H \widehat{\mathbf{H}}_1\right)^H
\end{aligned} \tag{23b}$$

and the best $\mathbf{L}_1 \in \mathcal{L}_1$ in the MSE sense,

$$\mathbf{L}_1^b = \arg\min_{\mathbf{L}_1}\left\{\mathbf{P}_{1|1}(\mathbf{L}_1)\right\} \text{ s.t. } \mathbf{L}_1 \in \mathcal{L}_1, \tag{23c}$$

computed from the assumed LDSS model minimizes the MSE matrix associated with the true state $\mathbf{x}_1$. In that sense, $\mathcal{L}_1$ defines the set of gain matrices which match the true observations $\mathbf{y}_1$ with the assumed LDSS model. We then obtain the performance of LCKF and LCMVEs for the assumed LDSS model with an increase of the achievable MSE due to the introduction of additional LCs ($\mathbf{L}_1 \in \mathcal{L}_1$). At time $k \geq 2$, if the subset of gain matrices

$$\mathcal{L}_k = \left\{ \mathbf{L} \in \mathbb{C}^{N_k \times P} \mid \boldsymbol{\varepsilon}_k(\mathbf{L}) = \mathbf{0} \right\}, \tag{23d}$$

is non empty, then for any $\mathbf{L}_k \in \mathcal{L}_k$ (22a) reduces to

$$\hat{\mathbf{x}}_{k|k}(\mathbf{L}_k) - \mathbf{x}_k = \left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right)\left(\widehat{\mathbf{F}}_{k-1}\left(\hat{\mathbf{x}}_{k-1|k-1}^b - \mathbf{x}_{k-1}\right) - \mathbf{w}_{k-1}\right) + \mathbf{L}_k^H \mathbf{v}_k. \tag{23e}$$

Without any additional assumptions, the best $\mathbf{L}_k \in \mathcal{L}_k$ in the MSE sense,

$$\mathbf{L}_k^b = \arg\min_{\mathbf{L}_k} \left\{ \mathbf{P}_{k|k}^J(\mathbf{L}_k) \right\} \text{ s.t. } \mathbf{L}_k \in \mathcal{L}_k \tag{24}$$

is computed according to (13e-13g) relying in part on the knowledge of

$$\left\{ \begin{array}{rl} \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}} = \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{F}_{k-2}\mathbf{x}_{k-2}+\mathbf{w}_{k-2}} & = \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-2}}\mathbf{F}_{k-2}^H + \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{w}_{k-2}} \\ \mathbf{C}_{\mathbf{x}_k,\mathbf{v}_k} = \mathbf{C}_{\mathbf{F}_{k-1}\mathbf{x}_{k-1}+\mathbf{w}_{k-1},\mathbf{v}_k} & = \mathbf{F}_{k-1}\mathbf{C}_{\mathbf{x}_{k-1},\mathbf{v}_k} + \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{v}_k} \end{array} \right. ,$$

however we only have access to the knowledge of

$$\left\{ \begin{array}{rl} \widehat{\mathbf{C}}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}} & = \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-2}}\widehat{\mathbf{F}}_{k-2}^H + \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{w}_{k-2}} \\ \widehat{\mathbf{C}}_{\mathbf{x}_k,\mathbf{v}_k} & = \widehat{\mathbf{F}}_{k-1}\mathbf{C}_{\mathbf{x}_{k-1},\mathbf{v}_k} + \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{v}_k} \end{array} \right. .$$

Thus, if we restrict to the "standard" LDSS model mentioned in monographs [24, §9.1][25, §7.1], which satisfies

$$\begin{aligned} &\mathbf{C}_{\mathbf{x}_1,\mathbf{w}_k} = \mathbf{0}, \quad \mathbf{C}_{\mathbf{x}_1,\mathbf{v}_k} = \mathbf{0}, \quad \mathbf{C}_{\mathbf{w}_l,\mathbf{w}_k} = \mathbf{C}_{\mathbf{w}_k}\delta_k^l, \\ &\mathbf{C}_{\mathbf{v}_l,\mathbf{v}_k} = \mathbf{C}_{\mathbf{v}_k}\delta_k^l, \mathbf{C}_{\mathbf{w}_l,\mathbf{v}_k} = \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{v}_k}\delta_k^{l+1}, \end{aligned} \tag{25}$$

then $\mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}} = \mathbf{0}$ and $\mathbf{C}_{\mathbf{x}_k,\mathbf{v}_k} = \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{v}_k}$, leading to

$$\widehat{\mathbf{C}}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}} = \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}} = \mathbf{0} \text{ and } \widehat{\mathbf{C}}_{\mathbf{x}_k,\mathbf{v}_k} = \mathbf{C}_{\mathbf{x}_k,\mathbf{v}_k} = \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{v}_k}.$$

In this case, $\mathbf{P}_{k|k}^J(\mathbf{L}_k)$ (11b) reduces to

$$\begin{aligned} \mathbf{P}_{k|k}^J(\mathbf{L}_k) &= \left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right)\mathbf{P}_{k|k-1}\left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right)^H + \mathbf{L}_k^H \mathbf{C}_{\mathbf{v}_k}\mathbf{L}_k \\ &\quad - \left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right)\mathbf{C}_{\mathbf{w}_{k-1},\mathbf{v}_k}\mathbf{L}_k - \mathbf{L}_k^H \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{v}_k}^H\left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right)^H, \end{aligned} \tag{26a}$$

$$\mathbf{P}_{k|k-1} = \widehat{\mathbf{F}}_{k-1}\mathbf{P}_{k-1|k-1}^b \widehat{\mathbf{F}}_{k-1}^H + \mathbf{C}_{\mathbf{w}_{k-1}}, \tag{26b}$$

$$\mathbf{P}_{k-1|k-1}^b = E\left[\left(\hat{\mathbf{x}}_{k-1|k-1}^b - \mathbf{x}_{k-1}\right)\left(\hat{\mathbf{x}}_{k-1|k-1}^b - \mathbf{x}_{k-1}\right)^H\right], \tag{26c}$$

and the solution of (24) only depends on $\widehat{\mathbf{F}}_{k-1}$, $\widehat{\mathbf{H}}_k$, $\mathbf{C}_{\mathbf{x}_1}$ and $\mathbf{m}_{\mathbf{x}_1}$.
Finally, if $\mathcal{L}_k$ is non empty at each time k, then for the "standard" LDSS model (25), LCKF and LCMVEs computed from the assumed LDSS model are matched to the true observations $\mathbf{y}_k$ and the recursion (13e-13g) minimizes the MSE associated with the true state $\mathbf{x}_k$. We then obtain the performance of LCKF and LCMVEs for the assumed LDSS model with an increase of the achievable MSE due to the introduction of additional LCs ($\mathbf{L}_k \in \mathcal{L}_k$). However, under (25) $\mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}} = \mathbf{0}$, therefore

$$\boldsymbol{\varepsilon}_k(\mathbf{L}_k) = \mathbf{0} \;\Leftrightarrow\; \left\{ \begin{array}{r} \mathbf{L}_k^H d\mathbf{H}_k = \mathbf{0} \\ \mathbf{L}_k^H d\mathbf{H}_k\left(\widehat{\mathbf{F}}_{k-1} + d\mathbf{F}_{k-1}\right) - \left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right) d\mathbf{F}_{k-1} = \mathbf{0} \end{array} \right. ,$$

where $\mathbf{L}_k, \widehat{\mathbf{H}}_k, d\mathbf{H}_k \in \mathbb{C}^{N_k \times P}$, $d\mathbf{F}_{k-1} \in \mathbb{C}^{P \times P}$, that is,

$$\boldsymbol{\varepsilon}_k(\mathbf{L}_k) = \mathbf{0} \;\Leftrightarrow\; \left\{ \mathbf{L}_k^H d\mathbf{H}_k = \mathbf{0}, \quad \left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right) d\mathbf{F}_{k-1} = \mathbf{0} \right\}, \tag{28a}$$

which leads to two possible cases: $rank\,(d\mathbf{F}_{k-1}) = P$ and $rank\,(d\mathbf{F}_{k-1}) < P$,

- Case 1) $rank\,(d\mathbf{F}_{k-1}) = P$

$$\left\{ \left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right) d\mathbf{F}_{k-1} = \mathbf{0}, \quad rank\,(d\mathbf{F}_{k-1}) = P_k \right\} \;\Leftrightarrow\; \mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k = \mathbf{0},$$

which leads to a degenerated form of the Kalman-like recursion (21)

$$\hat{\mathbf{x}}_{k|k}(\mathbf{L}_k) = \left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right) \widehat{\mathbf{F}}_{k-1} \hat{\mathbf{x}}^b_{k-1|k-1} + \mathbf{L}_k^H \mathbf{y}_k = \mathbf{L}_k^H \mathbf{y}_k$$

and a degenerated form of LCKF and LCMVEs

$$\hat{\mathbf{x}}^b_{k|k} = (\mathbf{L}_k^b)^H \mathbf{y}_k, \mathbf{L}_k^b = \arg\min_{\mathbf{L}_k} \left\{ \mathbf{L}_k^H \mathbf{C}_{\mathbf{v}_k} \mathbf{L}_k \right\} \text{ s.t. } \left\{ \mathbf{L}_k^H d\mathbf{H}_k = \mathbf{0}, \mathbf{L}_k^H \widehat{\mathbf{H}}_k = \mathbf{I} \right\} \tag{28b}$$

Thus if $rank\,(d\mathbf{F}_{k-1}) = P$, the introduction of LCs to mitigate modelling errors in state matrices $\mathbf{F}_{k-1}$ removes the KF main merit, that is the ability to combine previous observations to improve the estimation of the current state.

- Case 2) $rank\,(d\mathbf{F}_{k-1}) < P$

  In this case, (28a) can be recast as

$$\left\{ \mathbf{L}_k^H d\mathbf{H}_k = \mathbf{0}, \quad \mathbf{L}_k^H \left(\widehat{\mathbf{H}}_k d\mathbf{F}_{k-1}\right) = d\mathbf{F}_{k-1} \right\}$$

while the Kalman-like recursion (21) does not degenerate as above. More specifically, let $d\mathbf{F}_{k-1} = \mathbf{U}_{k-1} d\boldsymbol{\Phi}_{k-1}$ be the singular value decomposition (SVD) of $d\mathbf{F}_{k-1}$ where $\mathbf{U}_{k-1} \in \mathbb{C}^{P \times R_{k-1}}$ has full rank $R_{k-1} < P$ and $d\boldsymbol{\Phi}_{k-1} \in \mathbb{C}^{R_{k-1} \times P}$ [5]. Then

$$\left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right) d\mathbf{F}_{k-1} = \mathbf{0} \;\Leftrightarrow\; \left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right) \mathbf{U}_{k-1} d\boldsymbol{\Phi}_{k-1} = \mathbf{0}, \; \forall d\boldsymbol{\Phi}_{k-1}.$$

Since $\mathbf{U}_{k-1}$ has full rank, the above LCs are equivalent to

$$\left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right) \mathbf{U}_{k-1} = \mathbf{0} \;\Leftrightarrow\; \mathbf{L}_k^H \left(\widehat{\mathbf{H}}_k \mathbf{U}_{k-1}\right) = \mathbf{U}_{k-1}$$

and (28a) becomes

$$\boldsymbol{\varepsilon}_k\left(\mathbf{L}_k\right) = \mathbf{0} \;\Leftrightarrow\; \left\{\mathbf{L}_k^H d\mathbf{H}_k = \mathbf{0},\;\; \mathbf{L}_k^H \left(\widehat{\mathbf{H}}_k \mathbf{U}_{k-1}\right) = \mathbf{U}_{k-1}\right\}. \tag{28c}$$

3.4. Special Case: Mitigation of Modelling Error in State Matrices where $d\mathbf{F}_{k-1}$ has not full rank

If the measurement matrices $\mathbf{H}_k$ are perfectly known, then $d\mathbf{H}_k = \mathbf{0}$. At time $k = 1$ there is no longer any mismatch to mitigate, whereas at time $k \geq 2$ (22a) reduces to

$$\begin{aligned}\hat{\mathbf{x}}_{k|k}\left(\mathbf{L}_k\right) - \mathbf{x}_k = \left(\mathbf{I} - \mathbf{L}_k^H \mathbf{H}_k\right)\left(\widehat{\mathbf{F}}_{k-1}\left(\hat{\mathbf{x}}^b_{k-1|k-1} - \mathbf{x}_{k-1}\right) - \mathbf{w}_{k-1}\right) + \mathbf{L}_k^H \mathbf{v}_k \\ + \boldsymbol{\varepsilon}_k\left(\mathbf{L}_k\right),\end{aligned}$$

where $\boldsymbol{\varepsilon}_k\left(\mathbf{L}_k\right) = \left(\mathbf{L}_k^H \mathbf{H}_k - \mathbf{I}\right) d\mathbf{F}_{k-1}\mathbf{x}_{k-1}$. Since $rank\left(d\mathbf{F}_{k-1}\right) < P$, (28a) becomes

$$\boldsymbol{\varepsilon}_k\left(\mathbf{L}_k\right) = \mathbf{0} \;\Leftrightarrow\; \left\{\mathbf{L}_k^H \left(\mathbf{H}_k \mathbf{U}_{k-1}\right) = \mathbf{U}_{k-1}\right\} \tag{29a}$$

where $d\mathbf{F}_{k-1} = \mathbf{U}_{k-1} d\boldsymbol{\Phi}_{k-1}$ is the SVD of $d\mathbf{F}_{k-1}$, and, $\forall \mathbf{L}_k \in \mathcal{L}_k$,

$$\hat{\mathbf{x}}_{k|k}\left(\mathbf{L}_k\right) - \mathbf{x}_k = \left(\mathbf{I} - \mathbf{L}_k^H \mathbf{H}_k\right)\left(\widehat{\mathbf{F}}_{k-1}\left(\hat{\mathbf{x}}^b_{k-1|k-1} - \mathbf{x}_{k-1}\right) - \mathbf{w}_{k-1}\right) + \mathbf{L}_k^H \mathbf{v}_k, \tag{29b}$$

leading to, under (25),

$$\begin{aligned}\mathbf{P}^J_{k|k}\left(\mathbf{L}_k\right) &= \left(\mathbf{I} - \mathbf{L}_k^H \mathbf{H}_k\right) \mathbf{P}_{k|k-1} \left(\mathbf{I} - \mathbf{L}_k^H \mathbf{H}_k\right)^H + \mathbf{L}_k^H \mathbf{C}_{\mathbf{v}_k} \mathbf{L}_k \\ &\quad - \left(\mathbf{I} - \mathbf{L}_k^H \mathbf{H}_k\right) \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{v}_k} \mathbf{L}_k - \mathbf{L}_k^H \mathbf{C}^H_{\mathbf{w}_{k-1},\mathbf{v}_k} \left(\mathbf{I} - \mathbf{L}_k^H \mathbf{H}_k\right)^H, \\ \mathbf{P}_{k|k-1} &= \widehat{\mathbf{F}}_{k-1} \mathbf{P}^b_{k-1|k-1} \widehat{\mathbf{F}}^H_{k-1} + \mathbf{C}_{\mathbf{w}_{k-1}}, \\ \mathbf{P}^b_{k-1|k-1} &= E\left[\left(\hat{\mathbf{x}}^b_{k-1|k-1} - \mathbf{x}_{k-1}\right)\left(\hat{\mathbf{x}}^b_{k-1|k-1} - \mathbf{x}_{k-1}\right)^H\right],\end{aligned}$$

where all terms are known. Finally, if $\mathcal{L}_k$ is non empty at each time k, then under (25), LCKF and LCMVEs computed from the assumed LDSS model are matched to the true observations $\mathbf{y}_k$ and the recursion (13e-13g) minimizes the MSE associated with the true state $\mathbf{x}_k$. We then obtain the performance of LCKF and LCMVEs for the assumed LDSS model with an increase of the achievable MSE due to the introduction of additional LCs (29a).

3.4.1. Example on State Matrix Error $d\mathbf{F}_{k-1}$ Mitigation

As an example of model mismatch on the state transition matrix we consider the same scenario as above (16a), but a second FCS source x_2 is now located at a broadside angle α_2 in the angular vicinity of the signal of interest, i.e., $\alpha_2 = \alpha + \alpha_{3dB}/8$, where α_{3dB} denotes the beamwidth. The second source to noise power is 40 dB, which is high enough to induce a crosstalk affecting the first FCS source x_1 leading to the following true and mismatched LDSS models,

$$\text{True}: \left\{ \begin{array}{l} \mathbf{x}_k = \begin{bmatrix} 1 & d\phi \\ 0 & 1 \end{bmatrix} \mathbf{x}_{k-1} \\ \mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \end{array} \right. \text{Assumed}: \left\{ \begin{array}{l} \widetilde{\mathbf{x}}_k = \widetilde{\mathbf{x}}_{k-1} \\ \mathbf{y}_k = \mathbf{H}_k \widetilde{\mathbf{x}}_k + \mathbf{v}_k \end{array} \right. , \tag{30a}$$

where $d\phi$ is the unknown crosstalk coefficient which depends on various features, including the distance between the two sources. Note that due to crosstalk the first source x_1 turns into a PCS source.

The effect of crosstalk on the MVDRE is shown in Figures 2 and 3. In these figures, firstly, we compare the performance of the MVDRE computed under the hypothesis of FCS sources [14, (17a)-(17c)] when $d\phi = 0$ (denoted “MVDRE (FCS)” and “MVDRE (FCS,Sim)”) and when $d\phi = 10^{-4}$ (denoted “MVDRE (CT,Sim)”). Fig. 2 clearly exemplifies the impact of a loss of coherence of the signal source x_1 on the MVDRE performance in the large sample regime, which introduces a severe performance breakdown. On the other side, since the second source x_2 remains a FCS, its MSE remains unchanged. Secondly, in the considered scenario $d\mathbf{F}_{k-1} = \mathbf{u}_{k-1} d\phi$ where $\mathbf{u}_{k-1} = \binom{1}{0}$, and (29a) reduces to

$$\mathbf{L}_k^H \left(\mathbf{H}_k \mathbf{u}_{k-1} \right) = \mathbf{u}_{k-1} \;\Leftrightarrow\; \mathbf{L}_k^H \mathbf{h}_k \left(\hat{d}, \alpha \right) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{30b}$$

Therefore, according to the analysis introduced above, the effect of the crosstalk can be mitigated by resorting to a recursive LCMVE based on (13e-13g) where $\mathbf{T}_k = \binom{1}{0}$ and $\boldsymbol{\Delta}_k = \mathbf{h}_k \left(\hat{d}, \alpha \right)$, which is exemplified in Figures 2 and 3 with curves ”LCMVE (CT)” and ”LCMVE (CT,Sim)”. The empirical MSEs (denoted “...Sim)...”) are assessed with 10^4 Monte-Carlo trials. Interestingly enough, the mitigation of a modelling error on the first source x_1 does not have an impact on the estimation of the second source x_2. Last, but not least, even if the transformation of a FCS source into a PCS relies on a different mechanism (crosstalk instead of random walk), the formulation of a recursive LCMVE (instead of a recursive MVDRE) taking this phenomenon into account yields a similar behaviour in the large sample regime, i.e. a minimum achievable MSE.

3.5. Special Case: Mitigation of Modelling Error in Measurement Matrices

If the state matrices $\mathbf{F}_{k-1}$ are perfectly known, then $d\mathbf{F}_{k-1} = \mathbf{0}$. At time $k = 1$ the above analysis is still valid, whereas at time $k \geq 2$ (22a) reduces to

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} \left(\mathbf{L}_k \right) - \mathbf{x}_k = \left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k \right) \left(\mathbf{F}_{k-1} \left(\hat{\mathbf{x}}_{k-1|k-1}^b - \mathbf{x}_{k-1} \right) - \mathbf{w}_{k-1} \right) + \mathbf{L}_k^H \mathbf{v}_k \\ + \boldsymbol{\varepsilon}_k \left(\mathbf{L}_k \right), \end{aligned}$$

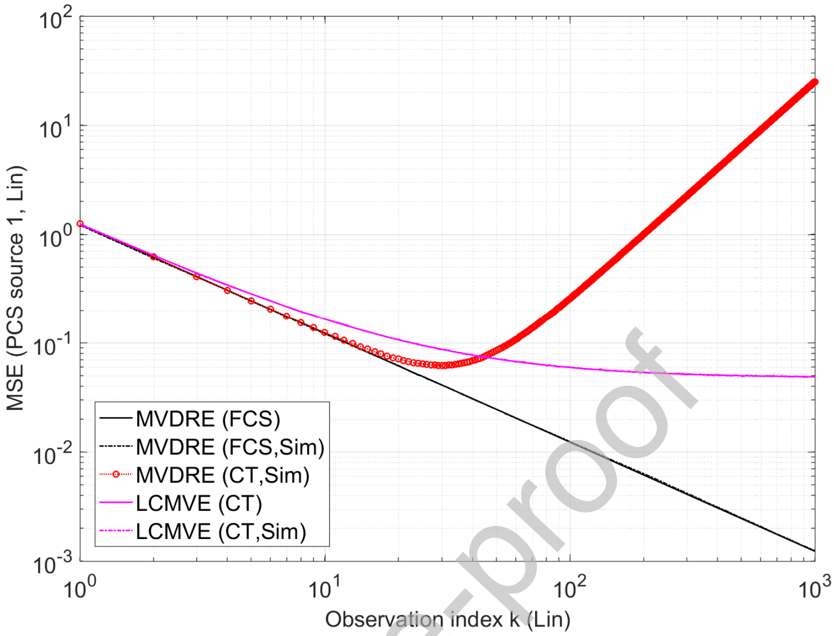

Figure 2: MSE of recursive MVDRE and LCMVE of $(\mathbf{x}_k)_1$ over time k, in presence of crosstalk (30a).

where $\boldsymbol{\varepsilon}_k(\mathbf{L}_k) = \mathbf{L}_k^H d\mathbf{H}_k \mathbf{x}_k$. Thus (28a) becomes

$$\boldsymbol{\varepsilon}_k(\mathbf{L}_k) = \mathbf{0} \;\Leftrightarrow\; \left\{\mathbf{L}_k^H d\mathbf{H}_k = \mathbf{0}\right\}, \tag{31a}$$

and, $\forall \mathbf{L}_k \in \mathcal{L}_k$,

$$\hat{\mathbf{x}}_{k|k}(\mathbf{L}_k) - \mathbf{x}_k = \left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right)\left(\mathbf{F}_{k-1}\left(\hat{\mathbf{x}}_{k-1|k-1}^b - \mathbf{x}_{k-1}\right) - \mathbf{w}_{k-1}\right) + \mathbf{L}_k^H \mathbf{v}_k, \tag{31b}$$

leading to, under (10b),

$$\begin{aligned}
\mathbf{P}_{k|k}^J(\mathbf{L}_k) &= \left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right)\mathbf{P}_{k|k-1}\left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right)^H + \mathbf{L}_k^H \mathbf{C}_{\mathbf{v}_k}\mathbf{L}_k \\
&\quad - \left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right)\mathbf{C}_{\mathbf{x}_k,\mathbf{v}_k}\mathbf{L}_k - \mathbf{L}_k^H \mathbf{C}_{\mathbf{x}_k,\mathbf{v}_k}^H\left(\mathbf{I} - \mathbf{L}_k^H \widehat{\mathbf{H}}_k\right)^H, \\
\mathbf{P}_{k|k-1} &= \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}^b\mathbf{F}_{k-1}^H + \mathbf{C}_{\mathbf{w}_{k-1}} + \mathbf{F}_{k-1}\mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}}^H + \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}}\mathbf{F}_{k-1}^H, \\
\mathbf{P}_{k-1|k-1}^b &= E\left[\left(\hat{\mathbf{x}}_{k-1|k-1}^b - \mathbf{x}_{k-1}\right)\left(\hat{\mathbf{x}}_{k-1|k-1}^b - \mathbf{x}_{k-1}\right)^H\right].
\end{aligned}$$

where all terms, including $\mathbf{C}_{\mathbf{x}_k,\mathbf{v}_k}$ and $\mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}}$, are known. Finally, if $\mathcal{L}_k$ is non empty at each time k, then under (10b), LCKF and LCMVEs computed from the assumed LDSS model are matched to the true observations $\mathbf{y}_k$

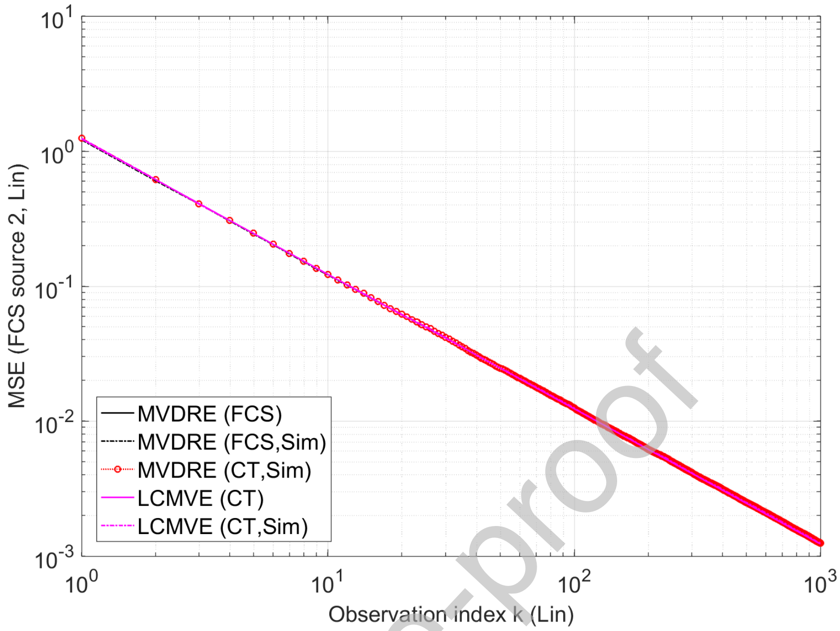

Figure 3: MSE of recursive MVDRE and LCMVE of $(\mathbf{x}_k)_2$ versus k, in presence of crosstalk (30a).

and the recursion (13e-13g) minimizes the MSE associated with the true state $\mathbf{x}_k$. We then obtain the performance of LCKF and LCMVEs for the assumed LDSS model with an increase of the achievable MSE due to the introduction of additional LCs (31a).

3.5.1. Example on Measurement Matrix Error $d\mathbf{H}_k$ Mitigation

Considering the example introduced in Section 2.4, let us assume now that, due to a calibration error, or array deformation (e.g., thermal effects, aging, etc.), the actual inter-sensor distance is $d = 0.98\hat{d}$, i.e. $\hat{d} - d = \lambda/100$. Thus, we are in the presence of a parametric modelling error in measurement vectors $\mathbf{h}_k(d, \alpha)$ which leads to the computation of a recursive MVDRE that does not match the true observations (16a). The effect of such kind of "miscalibration" on the MVDRE is shown in Fig. 4 where we compare the performance of MVDREs based on recursions [14, (17a)-(17c)] computed with the true value d ("Cal MVDRE (FCS)") and with the assumed value $\hat{d}$ ("MisCal MVDRE (FCS,SIM)") for a FCS source. From a more general perspective, in this case (31a) becomes

$$\mathbf{L}_k^H d\mathbf{H}_k = \mathbf{0} \;\Leftrightarrow\; \mathbf{L}_k^H \left[\frac{\partial \mathbf{h}_k^1\left(\hat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}^T} d\hat{\boldsymbol{\theta}} \;\ldots\; \frac{\partial \mathbf{h}_k^P\left(\hat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}^T} d\hat{\boldsymbol{\theta}} \right] = \mathbf{0}, \; \forall d\hat{\boldsymbol{\theta}},$$

which yields the sufficient condition

$$\mathbf{L}_k^H \left[\frac{\partial \mathbf{h}_k^1 \left(\hat{\boldsymbol{\theta}} \right)}{\partial \boldsymbol{\theta}^T} \; \dots \; \frac{\partial \mathbf{h}_k^P \left(\hat{\boldsymbol{\theta}} \right)}{\partial \boldsymbol{\theta}^T} \right] = \mathbf{0}. \tag{33}$$

Under (33), identity (31b) becomes a first order approximation

$$\hat{\mathbf{x}}_{k|k} \left(\mathbf{L}_k \right) - \mathbf{x}_k \simeq \left(\mathbf{I} - \mathbf{L}_k \widehat{\mathbf{H}}_k \right) \left(\mathbf{F}_{k-1} \left(\hat{\mathbf{x}}_{k-1|k-1}^b - \mathbf{x}_{k-1} \right) - \mathbf{w}_{k-1} \right) + \mathbf{L}_k \mathbf{v}_k.$$

Thus, in order to mitigate the effect on $\mathbf{h}_k \left(d, \alpha \right)$ of a small change in the system parameter d, at each iteration the constraint $\mathbf{l}_k^H \partial \mathbf{h}_k \left(\hat{d}, \alpha \right) / \partial d = 0$ is taken into account ("MisCal LCMVE (FCS,SIM)"). We also assess the impact of a PCS source on recursive LCMVE performance by considering the amplitude fluctuation model (16b) where $\sigma_w^2 = 10^{-4}$. For this purpose, we compare the performance of the recursive LCMVE computed under the hypothesis of a FCS source [14, (19a)-(19d)], denoted by "MisCal LCMVE Mismatched to PCS (SIM)" in Fig. 4, and the proposed extension (13e-13g), that is the ability to resort to a recursive LCMVE taking into account (16b), denoted by "MisCal LCMVE Matched to PCS (SIM)" in Fig. 4. Again, even a slight loss of coherence introduces a severe LCMVE performance breakdown when the loss of coherence is ignored, breakdown which can be mitigated when the amplitude fluctuation model (16b) is taken into account thanks to the proposed methodology. Last but not least, in case of a "small" miscalibration effect, the analytic LCMVE recursion (13e-13g) provides a tight prediction of the actual behaviour of the LCMVE, both in presence of a FCS source ("MisCal LCMVE (FCS,Pred)") and of a PCS source ("MisCal LCMVE Matched to PCS (Pred)").

4. Mitigation of System Noise Statistics Uncertainty

Since LCKF and LCMVEs are derived as solutions of the minimization of the MSE matrix of a linear estimator under LCs, only the first and second order moments are taken into account, and nothing is said about the probability distribution of the system noise. In that sense, LCKF and LCMVEs can be looked upon as "system noise first and second order statistics" matched filters. Therefore any mismatch between the true system noise statistics and the assumed system noise statistics leads to a suboptimal filter, and possibly to a filter with bad performance, as the discrepancy between the two models increases. Regarding the possible lack of knowledge on the measurement noise ($\mathbf{v}_k$) and/or the amplitude fluctuation noise ($\mathbf{w}_{l-1}$) mean and covariance, we consider the case where it results from the addition of nuisances whose parametric models are partially known, leading to the problem of estimating the state of a nominal

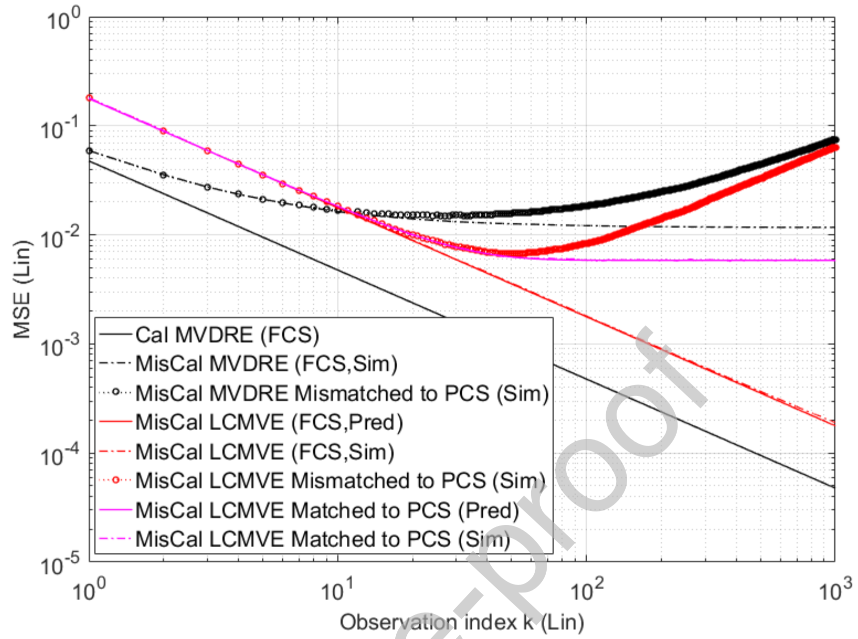

Figure 4: MSE of the recursive LCMVE of x_k (16a-16b) over time k, $\sigma_w^2 = 10^{-4}$

LDSS model from an observed LDSS model corrupted by the nuisances:

$$\text{Observed}: \begin{cases} \widetilde{\mathbf{x}}_k = \mathbf{F}_{k-1}\widetilde{\mathbf{x}}_{k-1} + \mathbf{w}_{k-1} + \boldsymbol{\eta}_{k-1} \\ \mathbf{y}_k = \mathbf{H}_k\widetilde{\mathbf{x}}_k + \mathbf{v}_k + \mathbf{j}_k \end{cases} \tag{34}$$

$$\text{Nominal}: \begin{cases} \mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1} \\ \mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k \end{cases} \tag{35}$$

4.1. Impact of System Noise Statistics Uncertainty

At time $k = 1$, any LCKF or LCMVE of $\mathbf{x}_1$ is of the form $\hat{\mathbf{x}}_{1|1}(\mathbf{L}_1) = \mathbf{L}_1^H \mathbf{y}_1$ where $\mathbf{L}_1 \triangleq \mathbf{L}_1^b$ is the solution of $\mathbf{L}_1^b = \arg\min_{\mathbf{L}_1} \{\mathbf{P}_{1|1}(\mathbf{L}_1)\}$ or the solution of $\mathbf{L}_1^b = \arg\min_{\mathbf{L}_1} \{\mathbf{P}_{1|1}(\mathbf{L}_1)\}$ s.t. $\mathbf{L}_1^H \boldsymbol{\Delta}_1 = \mathbf{T}_1$, computed with the nominal LDSS model. Since $\mathbf{y}_1 = \mathbf{H}_1\mathbf{x}_1 + \mathbf{v}_1 + \mathbf{j}_1$, the error made by the nominal filter in estimating the nominal $\mathbf{x}_1$ is given by

$$\hat{\mathbf{x}}_{1|1} - \mathbf{x}_1 = -\left(\mathbf{I} - \mathbf{L}_1^H \mathbf{H}_1\right)\mathbf{x}_1 + \mathbf{L}_1^H \mathbf{v}_1 + \boldsymbol{\varepsilon}_1(\mathbf{L}_1), \tag{36a}$$

where $\boldsymbol{\varepsilon}_1(\mathbf{L}_1) = \mathbf{L}_1^H \mathbf{j}_1$. At time $k \geq 2$, provided that LCs (13b)-(13d) are considered, any LCKF or LCMVE of $\mathbf{x}_k$ is obtained from the Kalman-like recursion

$$\hat{\mathbf{x}}_{k|k}(\mathbf{L}_k) = \mathbf{F}_{k-1}\hat{\mathbf{x}}^b_{k-1|k-1} + \mathbf{L}_k^H \left(\mathbf{y}_k - \mathbf{H}_k \mathbf{F}_{k-1}\hat{\mathbf{x}}^b_{k-1|k-1}\right) \tag{37}$$

where $\mathbf{L}_k \triangleq \mathbf{L}_k^b$ is the solution of $\mathbf{L}_k^b = \arg\min_{\mathbf{L}_k} \left\{ \mathbf{P}_{k|k}^J (\mathbf{L}_k) \right\}$ (11c-11e), or the solution of $\mathbf{L}_k^b = \arg\min_{\mathbf{L}_k} \left\{ \mathbf{P}_{k|k}^J (\mathbf{L}_k) \right\}$ s.t. $\mathbf{L}_k^H \mathbf{\Delta}_k = \mathbf{T}_k$ (13e-13g), computed with the nominal LDSS model. Since

$$\mathbf{y}_k = \mathbf{H}_k (\mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}) + \mathbf{v}_k + \mathbf{H}_k \boldsymbol{\eta}_{k-1} + \mathbf{j}_k,$$

the error made by the nominal filter in estimating the nominal $\mathbf{x}_k$ (34) is

$$\hat{\mathbf{x}}_{k|k} (\mathbf{L}_k) - \mathbf{x}_k = \left(\mathbf{I} - \mathbf{L}_k^H \mathbf{H}_k\right) \left(\mathbf{F}_{k-1} \left(\hat{\mathbf{x}}_{k-1|k-1}^b - \mathbf{x}_{k-1}\right) - \mathbf{w}_{k-1}\right) + \mathbf{L}_k^H \mathbf{v}_k + \boldsymbol{\varepsilon}_k (\mathbf{L}_k), \quad (38a)$$

where $\boldsymbol{\varepsilon}_k (\mathbf{L}_k) = \mathbf{L}_k^H \left(\mathbf{H}_k \boldsymbol{\eta}_{k-1} + \mathbf{j}_k\right)$.

4.2. Mitigation of System Noise Statistics Uncertainty

At time $k = 1$, if the subset of gain matrices $\mathcal{L}_1 = \left\{\mathbf{L} \in \mathbb{C}^{N_1 \times P} \mid \boldsymbol{\varepsilon}_1 (\mathbf{L}) = \mathbf{0}\right\}$ is non empty, then $\forall \mathbf{L}_1 \in \mathcal{L}_1$, (20a) and $\mathbf{P}_{1|1} (\mathbf{L}_1)$ reduce to

$$\begin{aligned}
\hat{\mathbf{x}}_{1|1} - \mathbf{x}_1 &= -\left(\mathbf{I} - \mathbf{L}_1^H \mathbf{H}_1\right) \mathbf{x}_1 + \mathbf{L}_1^H \mathbf{v}_1, \\
\mathbf{P}_{1|1} (\mathbf{L}_1) &= \left(\mathbf{I} - \mathbf{L}_1^H \mathbf{H}_1\right) \mathbf{C}_{\mathbf{x}_1} \left(\mathbf{I} - \mathbf{L}_1^H \mathbf{H}_1\right)^H + \mathbf{L}_1^H \mathbf{C}_{\mathbf{v}_1} \mathbf{L}_1 \\
&\quad - \left(\mathbf{I} - \mathbf{L}_1^H \mathbf{H}_1\right) \mathbf{C}_{\mathbf{x}_1,\mathbf{v}_1} - \mathbf{C}_{\mathbf{x}_1,\mathbf{v}_1}^H \left(\mathbf{I} - \mathbf{L}_1^H \mathbf{H}_1\right)^H,
\end{aligned}$$

and the best $\mathbf{L}_1 \in \mathcal{L}_1$ in the MSE sense,

$$\mathbf{L}_1^b = \arg\min_{\mathbf{L}_1} \left\{ \mathbf{P}_{1|1} (\mathbf{L}_1) \right\} \text{ s.t. } \mathbf{L}_1 \in \mathcal{L}_1, \quad (39a)$$

computed from the nominal LDSS model minimizes the MSE matrix associated with the estimation of the nominal state $\mathbf{x}_1$. In that sense, $\mathcal{L}_1$ defines the set of gain matrices which match the true observations $\mathbf{y}_1$ with the nominal LDSS model. We then obtain the performance of LCKF and LCMVEs for the nominal LDSS model with an increase of the achievable MSE due to the introduction of additional LCs ($\mathbf{L}_1 \in \mathcal{L}_1$). At time $k \geq 2$, if the subset of gain matrices $\mathcal{L}_k = \left\{\mathbf{L} \in \mathbb{C}^{N_k \times P} \mid \boldsymbol{\varepsilon}_k (\mathbf{L}) = \mathbf{0}\right\}$ is non empty, then $\forall \mathbf{L}_k \in \mathcal{L}_k$, (38a) reduces to

$$\hat{\mathbf{x}}_{k|k} (\mathbf{L}_k) - \mathbf{x}_k = \left(\mathbf{I} - \mathbf{L}_k^H \mathbf{H}_k\right) \left(\mathbf{F}_{k-1} \left(\hat{\mathbf{x}}_{k-1|k-1}^b - \mathbf{x}_{k-1}\right) - \mathbf{w}_{k-1}\right) + \mathbf{L}_k^H \mathbf{v}_k,$$

leading to, under (10b),

$$\begin{aligned}
&\mathbf{P}_{k|k}^J (\mathbf{L}_k) = \left(\mathbf{I} - \mathbf{L}_k^H \mathbf{H}_k\right) \mathbf{P}_{k|k-1} \left(\mathbf{I} - \mathbf{L}_k^H \mathbf{H}_k\right)^H + \mathbf{L}_k^H \mathbf{C}_{\mathbf{v}_k} \mathbf{L}_k \\
&\qquad - \left(\mathbf{I} - \mathbf{L}_k^H \mathbf{H}_k\right) \mathbf{C}_{\mathbf{x}_k,\mathbf{v}_k} \mathbf{L}_k - \mathbf{L}_k^H \mathbf{C}_{\mathbf{x}_k,\mathbf{v}_k}^H \left(\mathbf{I} - \mathbf{L}_k^H \mathbf{H}_k\right)^H, \\
&\mathbf{P}_{k|k-1} = \mathbf{F}_{k-1} \mathbf{P}_{k-1|k-1}^b \mathbf{F}_{k-1}^H + \mathbf{C}_{\mathbf{w}_{k-1}} + \mathbf{F}_{k-1} \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}}^H + \mathbf{C}_{\mathbf{w}_{k-1},\mathbf{x}_{k-1}} \mathbf{F}_{k-1}^H, \\
&\mathbf{P}_{k-1|k-1}^b = E\left[\left(\hat{\mathbf{x}}_{k-1|k-1}^b - \mathbf{x}_{k-1}\right) \left(\hat{\mathbf{x}}_{k-1|k-1}^b - \mathbf{x}_{k-1}\right)^H\right].
\end{aligned}$$

Finally, if $\mathcal{L}_k$ is non empty at each time k, then under (10b), LCKF and LCMVEs computed from the nominal LDSS model are matched to the true observations $\mathbf{y}_k$ and the recursion (13e-13g) minimizes the MSE matrix associated with the estimation of the nominal state $\mathbf{x}_k$. We then obtain the performance of LCKF and LCMVEs for the nominal LDSS model with an increase in the achievable MSE due to the introduction of additional LCs ($\mathbf{L}_k \in \mathcal{L}_k$). Interestingly, the LCs

$$\boldsymbol{\varepsilon}_k(\mathbf{L}_k) = \mathbf{L}_k^H \left(\mathbf{H}_k \boldsymbol{\eta}_{k-1} + \mathbf{j}_k\right) = \mathbf{0} \tag{40a}$$

can mitigate various cases of noise statistics uncertainty. For instance:

- The presence of interference sources (jammers) in measurement where $\mathbf{j}_k = \boldsymbol{\Psi}_k \mathbf{i}_k$, $\boldsymbol{\Psi}_k$ is known and $\mathbf{i}_k$ is unknown,

$$\mathbf{L}_k^H \boldsymbol{\Psi}_k = \mathbf{0} \;\Rightarrow\; \mathbf{L}_k^H \mathbf{j}_k = \mathbf{0} \tag{40b}$$

- The presence of interference sources or noise statistics uncertainty in the state where $\boldsymbol{\eta}_{k-1} = \boldsymbol{\Phi}_{k-1}\mathbf{g}_{k-1}$, $\boldsymbol{\Phi}_{k-1}$ is known and $\mathbf{g}_{k-1}$ is unknown,

$$\mathbf{L}_k^H \left(\mathbf{H}_k \boldsymbol{\Phi}_{k-1}\right) = \mathbf{0} \;\Rightarrow\; \mathbf{L}_k^H \mathbf{H}_k \boldsymbol{\eta}_{k-1} = \mathbf{0}, \tag{40c}$$

  which occurs, for example, if the mean value of $\mathbf{w}_{k-1}$ is non-zero where $\mathbf{m}_{\mathbf{w}_{k-1}} = \boldsymbol{\Phi}_{k-1}\mathbf{g}_{k-1}$.

- The combination of the two previous cases,

$$\left\{\mathbf{L}_k^H \boldsymbol{\Psi}_k = \mathbf{0}, \;\; \mathbf{L}_k^H \left(\mathbf{H}_k \boldsymbol{\Phi}_{k-1}\right) = \mathbf{0}\right\} \;\Rightarrow\; \mathbf{L}_k^H \left(\mathbf{H}_k \boldsymbol{\eta}_{k-1} + \mathbf{j}_k\right) = \mathbf{0} \tag{40d}$$

4.2.1. Example on System Noise Statistics Uncertainty Mitigation

If calibration uncertainties must be mitigated for each new observations $\mathbf{y}_k$, in some sequential estimation problems it is more optimal to add on-line constraints that are triggered by a preprocessing of $\mathbf{y}_k$ or by external information on the environment. As an example we consider the same scenario as above (16a) when the ULA can be regarded as perfectly calibrated ($\theta = \hat{\theta}$). An intermittent jammer is located at a known broadside angle α_J in the angular vicinity of the signal of interest, i.e., $\alpha = \alpha_J - \alpha_{3dB}/4$, where α_{3dB} denotes the beamwidth. The jammer to noise power is 40 dB and its probability of activation at each observation is denoted by $\mathcal{P}_J$. We assume that the jammer is detected whenever it is activated.

Firstly, for a FCS source (reference case), we compare the standard jammer cancellation procedure [1, §6.7.1] (denoted "Batch LCMVE (FCS)" and "Batch LCMVE (FCS,Sim)" in Figures (5-9)) with the dynamic jammer cancellation (denoted "LCMVE (FCS)" and "LCMVE (FCS,Sim)" in Figures (5-9)) allowed by recursive LCMVEs. The standard procedure [1, §6.7.1] consists in imposing a permanent null constraint $\bar{\mathbf{l}}_k^H \overline{\mathbf{h}}_k(\theta, \alpha_J) = 0$ in the batch form of the LCMVE

(12a-12b), i.e. $\mathbf{\Gamma}_k = [1\ 0]$ and $\overline{\mathbf{\Lambda}}_k = \left[\overline{\mathbf{h}}_k(\theta,\alpha)\ \overline{\mathbf{h}}_k(\theta,\alpha_J)\right]$. Regarding the proposed dynamic jammer cancellation, at each jammer detection, the null constraint (40b) $\mathbf{l}_k^H \mathbf{h}_k(\theta,\alpha_J) = 0$ is added to cancel the jammer signal, and the recursive LCMVE is updated according to [14, (19a)-(19d)] where $T_k = 0$ and $\mathbf{\Delta}_k = \mathbf{h}_k(\theta,\alpha_J)$. In the absence of jammer detection, the recursive LCMVE is updated without additional constraint, that is according to [14, (17a)-(17c)]. The empirical MSEs (denoted "...Sim)...") are assessed with 10^4 Monte-Carlo trials. When the null constraint is set, the jammer signal is cancelled at the expense of an increase of the output noise power in comparison with a jammer free scenario, which increases the minimum MSE achieved. Therefore, to limit the increase of the MSE achieved, the null constraint must be set only when the jammer is activated, which is highlighted by Figures (5-9) displaying the MSE of both solutions obtained for 5 values of $\mathcal{P}_J$: 1, 0.9, 0.5, 0.1, 0. As expected, the superiority of the recursive LCMVE over the batch form LCMVE increases as $\mathcal{P}_J$ decreases.

Secondly, we also assess the impact of a PCS source on recursive LCMVE performance by considering the amplitude fluctuation model (16b) where $\sigma_w^2 = 10^{-4}$. For this purpose, we compare the performance of the dynamic jammer cancellation computed under the hypothesis of a FCS source (as above) denoted by "LCMVE Mismatched to PCS (Sim)" in Figures (5-9), and the proposed extension of [14] to PCS source, denoted by "LCMVE Matched to PCS" and "LCMVE Matched to PCS (Sim)" in Figures (5-9). In the latter case, at each jammer detection, the null constraint $\mathbf{l}_k^H \mathbf{h}_k(\theta,\alpha_J) = 0$ is added via (13e-13g) where $T_k = 0$ and $\mathbf{\Delta}_k = \mathbf{h}_k(\theta,\alpha_J)$. In the absence of jammer detection, the recursive LCMVE is updated according to (11c-11e).

Once again, Figures (5-9) exemplify the impact of a slight loss of coherence of the signal source on the LCMVE performance in the large sample regime, which introduces a severe performance breakdown when the loss of coherence is not taken into account. Thanks to the results introduced in Section 2.2, we can also evaluate which is the minimum achievable MSE when the amplitude fluctuation model is known (16b). Interestingly enough, as illustrated in Figures 5 (permanent jammer) and 9 (no jammer), the minimum achievable MSE does depend on the LCs considered.

5. Conclusion

A discussion on the important problem of linear regression was provided – with a focus on a multi-channel signal processing application. Particularly, the case of partially coherent signal sources was addressed, where the source amplitudes undergo a partial random walk between observations, thus extending prior work involving fully coherent signal sources. In that context, the paper derived a new class of recursive linearly constrained minimum variance estimators (LCMVE), which was seen to provide additional robustness to modeling errors through the incorporation of non-stationary linear constraints. Such formulation has the interesting feature of allowing for a closed-form solution. Moreover,

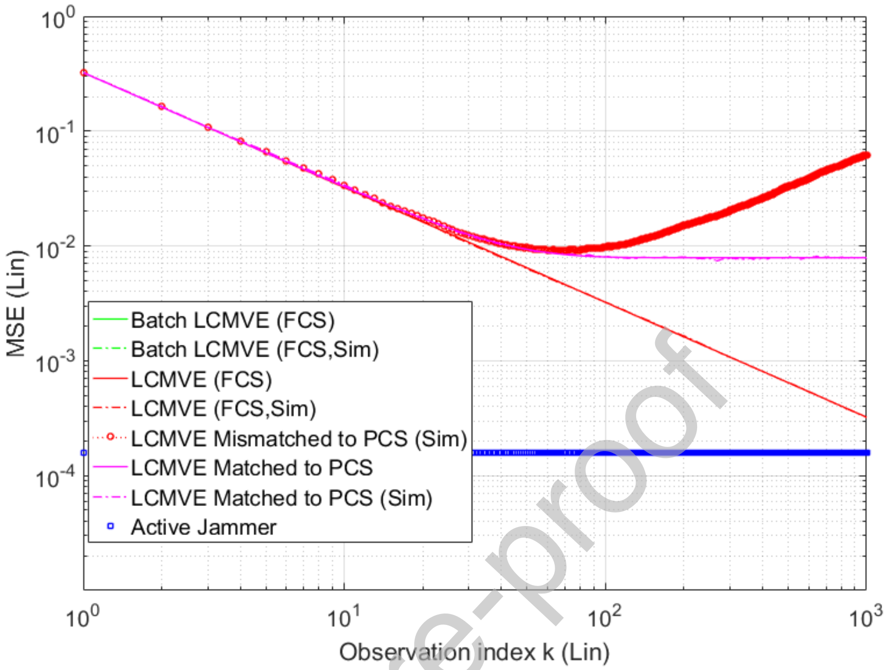

Figure 5: MSE of recursive and batch form LCMVEs of x_k (16a-16b) over time k, $\mathcal{P}_J = 1$.

a noteworthy feature of the recursive LCMVE is to be fully adaptive in the context of sequential estimation as it allows the addition of optional constraints that can be triggered by a preprocessing of each new observation or external information on the environment. The proposed methodology is validated with a running example on array signal processing where a variety of model mismatches are analyzed both analytically and with computer simulations. Those mismatches include erroneous system matrices and noise statistics in linear discrete state-space models. The analyses show the enhanced robustness with respect to standard LCMVE schemes. Notice that the methodology relies only on first and second order moments, then there is no need to impose a statistical model. For slightly nonlinear problems, the proposed solution is still valid if we consider a first-order linearization of the system model, in the vein of the extended Kalman filter. The analysis of the general model taking into account, simultaneously, model errors in system matrices and system noise statistics is not trivial to elaborate and is a topic under study. The main goal of this article is not to cover all possible cases but to show the capabilities of the addition of non-stationary linear constraints in order to robustify LCMVEs.

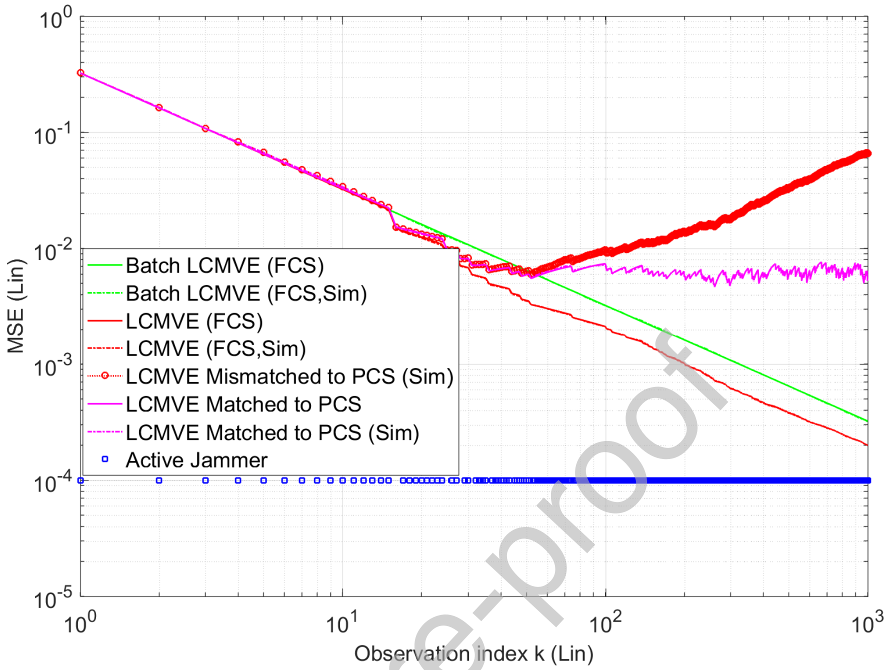

Figure 6: MSE of recursive and batch form LCMVEs of x_k (16a-16b) over time k, $\mathcal{P}_J = 0.9$.

6. Acknowledgments

This work has been partially supported by the DGA/AID (2018.60.0072.00.470.75.01) and NSF under Awards CNS-1815349 and ECCS-1845833.

References

[1] H. L. Van Trees, *Optimum Array Processing*, Wiley-Interscience, 2002.

[2] Paulo S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation (4Ed)*, Springer Science+Business Media, 2013.

[3] R. L. Plackett, "Some Theorems in Least Squares", Biometrika, 37: 149-157,1950.

[4] F. C. Schweppe, "Sensor array data processing for multiple signal sources", IEEE Trans. on IT, 14: 294-305, 1968.

[5] R. A. Horn, C. R. Johnson, *Matrix Analysis (2nd Ed)*. Cambridge University Press, 2013.

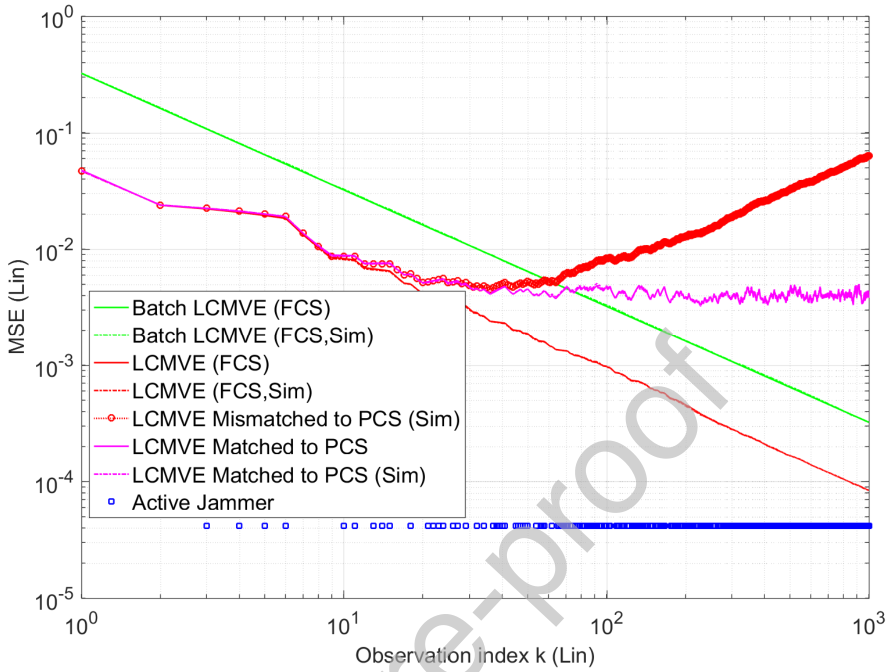

Figure 7: MSE of recursive and batch form LCMVEs of x_k (16a-16b) over time k, $\mathcal{P}_J = 0.5$.

[6] O. L. Frost, “An algorithm for linearly constrained adaptive array processing”, Proceedings of the IEEE, 60(8): 926-935, 1972.

[7] J. Li and P. Stoica, *Robust Adaptive Beamforming*, Wiley-Interscience, 2006.

[8] S. A. Vorobyov, “Principles of minimum variance robust adaptive beamforming design”, Elsevier Signal Processing, 93: 3264-3277, 2013.

[9] L. J. Griffiths and C. W. Jim, “An alternative approach to linearly constrained adaptive beamfonning”, IEEE Trans. on AP, 30(1): 27-34, 1982.

[10] L. S. Resende, J. M. T. Romano, M. G. Bellanger, “A fast least squares algorithm for linearly constrained adaptive filtering”, IEEE Trans. on SP, 44(5): 1168-1174, 1996.

[11] R. C. de Lamare, “Adaptive reduced-rank LCMV beamforming algorithms based on joint iterative optimisation of filters”, Electronics Letters, 44(8): 565-566, 2008.

[12] D. Cherkassky and S. Gannot, “New Insights into the Kalman Filter Beamformer: Applications to Speech and Robustness”, IEEE SP Letters, 23(3): 376-380, 2016.

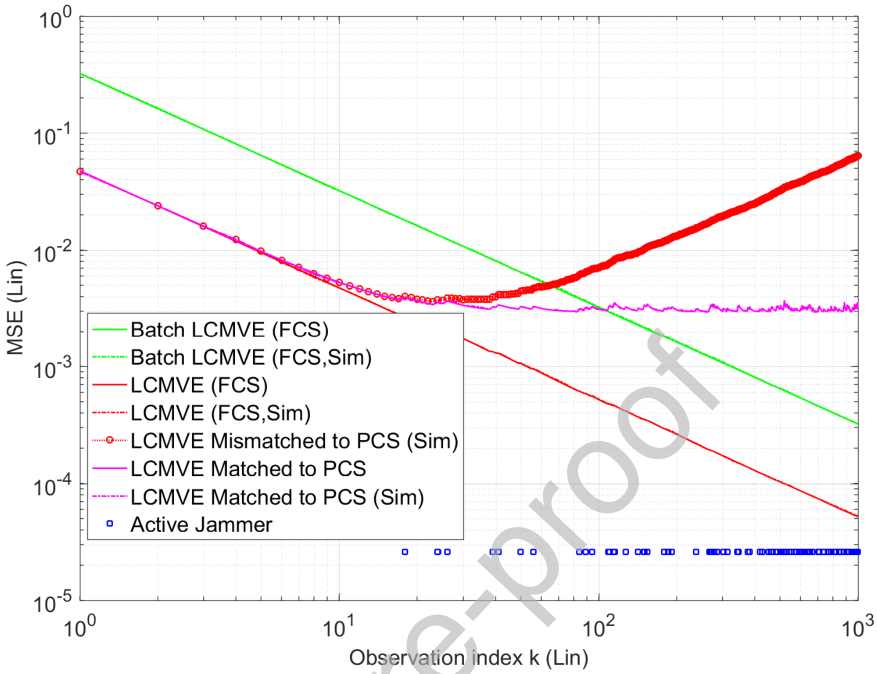

Figure 8: MSE of recursive and batch form LCMVEs of x_k (16a-16b) over time k, $\mathcal{P}_J = 0.1$.

[13] S Werner, J. A. Apolinario Jr, M. L. R. De Campos, "On the equivalence of RLS implementations of LCMV and GSC processors", IEEE SP Letters, 10(12): 356-359, 2003.

[14] F. Vincent and E. Chaumette, "Recursive Linearly Constrained Minimum Variance Estimator in Linear Models with Non-Stationary Constraints", Signal Processing (Elsevier) 149: 229-235, 2018.

[15] J. L. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems (2Ed)*, CRC Press, Taylor & Francis Group, 2012.

[16] E. Chaumette, F. Vincent and J. Vilà-Valls, "Linearly Constrained Wiener Filter Estimates For Linear Discrete State-Space Models", in Proc of Asilomar Conference on Signals, Systems, and Computers, 2018.

[17] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data*, Cambridge University Press 2010.

[18] N. Levanon, *Radar Principles*. Wiley-Interscience, Wiley 1988.

[19] R. Kalman, "A new approach to linear filtering and prediction problems", ASME Journal of Basic Engineering, 82: 35-45, 1960.

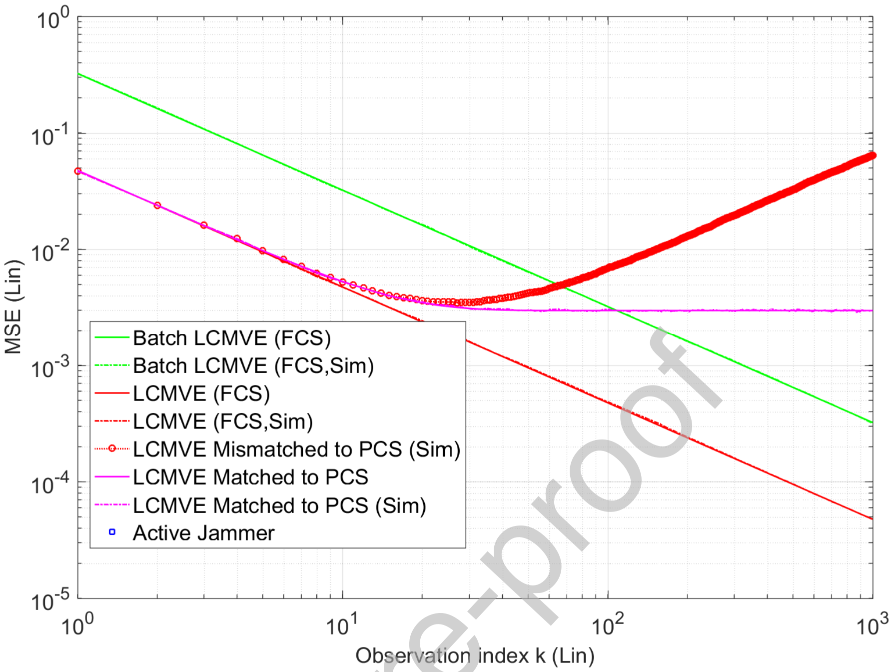

Figure 9: MSE of recursive and batch form LCMVEs of x_k (16a-16b) over time k, $\mathcal{P}_J = 0$.

[20] E. Chaumette et al, "Minimum Variance Distortionless Response Estimators For Linear Discrete State-Space Models", IEEE Trans. on AC, 62(4): 2048-2055, 2017.

[21] D. E. Catlin, "Estimation of random states in general linear models", IEEE Trans. on AC, 36(2): 248-252, 1991.

[22] G. Chen, *Approximate Kalman Filtering*, World Scientific, Singapore, 1993.

[23] E. Chaumette et al., "On LMVDR Estimators For LDSS Models: conditions for existence and further applications", IEEE Trans. on AC, 64(6): 2598-2605, 2019.

[24] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*, Prentice-Hall, Upper Saddle River, New Jersey, 2000.

[25] D. Simon, *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches*, Wiley InterScience, 2006.