Minimum Variance Distortionless Response Estimators
For Linear Discrete State-Space Models

Eric Chaumette, Benoit Priot, François Vincent, Gael Pages, Arnaud Dion

To cite this version:

Eric Chaumette, Benoit Priot, François Vincent, Gael Pages, Arnaud Dion. Minimum Variance Distortionless Response Estimators For Linear Discrete State-Space Models. IEEE Transactions on Automatic Control, Institute of Electrical and Electronics Engineers, 2016, vol. 11 (n° 4), pp.0. 10.1109/TAC.2016.2594384. hal-01414738

HAL Id: hal-01414738
https://hal.archives-ouvertes.fr/hal-01414738
Submitted on 12 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: [http://oatao.univ-toulouse.fr/](http://oatao.univ-toulouse.fr/)
Eprints ID: 16658

To link this article: [http://dx.doi.org/10.1109/TAC.2016.2594384](http://dx.doi.org/10.1109/TAC.2016.2594384)

To cite this version: Chaumette, Eric and Priot, Benoit and Vincent, François and Pages, Gael and Dion, Arnaud Minimum Variance Distortionless Response Estimators For Linear Discrete State-Space Models. (2016) IEEE Transactions on Automatic Control, vol. 11 (n° 4). ISSN 0018-9286

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr
Minimum Variance Distortionless Response Estimators For Linear Discrete State-Space Models

Eric Chaumette, Benoit Priot, François Vincent, Gael Pages and Arnaud Dion

Abstract—For linear discrete state-space models, under certain conditions, the linear least-mean-squares filter estimate has a convenient recursive predictor/corrector format, aka the Kalman filter. The purpose of this paper is to show that the linear minimum variance distortionless response (MVDR) filter shares exactly the same recursion, except for the initialization which is based on a weighted least-squares estimator. If the MVDR filter is sub-optimal in mean-squared error sense, it is an infinite impulse response distortionless filter (a deconvolver) which does not depend on the prior knowledge (first and second order statistics) on the initial state. In other words, the MVDR filter can be pre-computed and its behaviour can be assessed in advance independently of the prior knowledge on the initial state.

Index Terms—State estimation, unbiased filter, filtering, smoothing, minimum mean-squared error upper bound.

I. INTRODUCTION

We consider the general class of linear discrete state-space (LDSS) models represented with the state and measurement equations, respectively,

\[ x_k = F_{k-1}x_{k-1} + w_{k-1} \quad (1a) \]
\[ y_k = H_kx_k + v_k \quad (1b) \]

where the time index \( k \geq 1 \), \( x_k \) is the \( P \)-dimensional state vector, \( y_k \) is the \( N \)-dimensional measurement vector and the model matrices \( F_k \) and \( H_k \) are known. Unless otherwise stated, the process noise sequence \( \{w_k\} \) and the measurement noise sequence \( \{v_k\} \), as well as the initial state \( x_0 \) are random vectors with arbitrary distributions but at least known covariance and cross-covariance matrices containing elements with finite modulus. The process and the measurement noise sequences have zero-mean values and the initial state has a finite known mean value. The objective is to estimate \( x_k \) based on the measurements and our knowledge of the model dynamics. If the estimate of \( x_k \) is based on measurements up to and including time \( l \), we denote the estimator as \( \hat{x}_{k|l} \) and we use the term estimator to refer to the class of algorithms that includes filtering, prediction, and smoothing. A filter estimates \( x_k \) based on measurements up to and including time \( k \). A predictor estimates \( x_k \) based on measurements prior to time \( k \). A smoother estimates \( x_k \) based on measurements prior to time \( k \), at time \( k \), and later than time \( k \).

Since the seminal paper of Kalman [1], it is known that, if the process noise sequence, the measurement noise sequence and the initial state \( x_0 \) verify certain uncorrelation conditions [2, [9.1],[3, [7.1][4, [8.2] and are Gaussian, the minimum variance or minimum mean-squared error (MSE) filter estimate for LDSS models has a convenient recursive predictor/corrector format, so-called the Kalman filter (KF). Even if the noise is non-Gaussian, the KF is the linear least-mean-squares (LLMS) filter (LLMSE) estimate.

Interestingly enough, the Kalman (stochastic) filtering problem has been related to several variants of deterministic least-squares estimation, i.e. the estimation of \( x_0 \) for a LDSS model (1a-1b) where \( x_k = x_{k-1} = \ldots = x_0 \) and \( x_0 \) is deterministic and unknown [2, [2.6][3, [3.1][4, [4.1][5, [5.12][6]. In the deterministic framework, under Gaussian measurement noise, the weighted least-squares estimator (WLSE) is identical to the maximum likelihood estimator (MLE). Even if the measurement noise is non-Gaussian, the WLSE coincides with the best linear unbiased estimator (BLUE) [4, [4.6] ak the linear minimum variance distortionless response estimator (LMVDRF) [7, [6][8, [5.6], which is a particular case of constrained linear estimation [2, p98]. The LMVDRF is a well known sub-optimal technique in the deterministic framework used to circumvent the problem of the large computational cost of the MLE [7] and its sensitivity to imperfect, incomplete, or erroneous knowledge about the observation model [9]. For instance, in the fields of radar, sonar, and wireless communication, it is common place to design a LMVDRF for the most studied estimation problem: that of separating the components of data formed from a linear superposition of individual signals to noisy data [7][8][9]. This is the reason why, sometimes, the LMVDRF is also called a deconvolution filter [7, [6][8, [5.6].

The main novel contribution of the present paper is the proof that in the stochastic framework, for the fairly general subset of LDSS models for which \( F_k, k \geq 1 \), are invertible and \( H_1 \) is full rank, the LMVDR filter (LMVDRF) shares the same recursion as the KF, except at time \( k = 1 \). This surprising result is worth knowing for users of the KF since, each time one initializes the KF recursion at time \( k = 1 \) with the WLSE associated to the measurement model (1b), i.e. \( \bar{x}_{1|1} = P_{1|1} H_1^\text{T} C_1^{-1} y_1 \) and \( P_{1|1} = (H_1^\text{T} C_1^{-1} H_1)^{-1} \), because the first and second order statistics of the initial state \( x_0 \) are not available, one does not implement any longer the KF but the LMVDRF. Although the LMVDRF is sub-optimal in terms of MSE, it has a number of merits: a) it does not depend on the prior knowledge (first and second order statistics) on the initial state, b) it is an upper bound on the performance of the KF whatever the prior knowledge on the initial state, c) if the LDSS is time-invariant, it shares the same discrete algebraic
Riccati equation (DARE) as the KF. These features are quite interesting for filtering performance analysis and design of a LDSS system since they allow to synthesize an infinite impulse response (IIR) distortionless filter (a deconvolver) which performance is robust to an unknown initial state. However, it seems likely that, since the LMVDRF shares the same recursion as the KF, except at time \( k = 1 \); this result allows to show in Section V, that the two filters share as well, for the so-called “standard LDSS model”, the same possible extensions (prediction, smoothing, filtering with fading memory, constrained filtering,...) and alternate formulations (information filtering, innovations approach,...). Last, the LMVDRF’s properties and performance (in comparison with the KF) are illustrated with an example in Section VI.

II. NOTATIONS AND SIGNAL MODEL

The notational convention adopted is as follows: we shall use italic, small boldface and capital boldface letters to denote respectively scalars, column vectors and matrices. \( \mathcal{M}_C(N, P) \) denotes the vector space of complex matrices with \( N \) rows and \( P \) columns. The scalar/matrix/vector transpose, conjugate and transpose conjugate are respectively indicated by the superscripts \( \top, * \) and \( H \). \( I \) is the identity matrix. \( [A \, B] \) denotes the matrix resulting from the horizontal concatenation of \( A \) and \( B \). \( S = \text{span} \{A\} \) denotes the linear span of the set of the column vectors of \( A \). \( A \geq B \) means that \( A - B \) is positive semi-definite. If \( C^H = C \) and \( C > 0 \), \( \Pi_k^C = (A^H C A)^{-1} A^H C \) denotes the orthogonal projection matrix on \( \text{span} \{A\} \) and \( \Pi_k \perp B \) means that \( A \) and \( B \) are orthogonal, both for the Hermitian inner product defined by \( C \). \( S \perp \) denotes the orthogonal complement of the subspace \( S \) for the canonical Hermitian inner product. \( E[\cdot] \) denotes the expectation operator. If \( x \) is a random vector, \( \mathbf{m}_x \triangleq E[ x ] \) is the expectation of \( x \). If \( x \) and \( y \) are two complex random vectors: a) \( \mathbf{R}_x = E[ x x^H ] \), \( \mathbf{R}_y = E[ y y^H ] \) and \( \mathbf{R}_{x,y} = E[ x y^H ] \) are respectively the correlation matrices of \( x \), of \( y \) and the cross-correlation matrix of \( x \) and \( y \); b) \( \mathbf{C}_x = \mathbf{R}_x - \mathbf{m}_x \mathbf{m}_x^H \), \( \mathbf{C}_y = \mathbf{R}_y - \mathbf{m}_y \mathbf{m}_y^H \) and \( \mathbf{C}_{x,y} = \mathbf{R}_{x,y} - \mathbf{m}_x \mathbf{m}_y^H \) are respectively the covariance matrices of \( x \), of \( y \) and the cross-covariance matrix of \( x \) and \( y \); c) if \( \mathbf{R}_y \) is invertible, then \( \mathbf{R}_{x,y} \triangleq \mathbf{R}_x - \mathbf{R}_{x,y} \mathbf{R}_y^{-1} \mathbf{R}_{x,y}; \) d) if \( \mathbf{C}_y \) is invertible, then \( \mathbf{C}_{x,y} \triangleq \mathbf{C}_x - \mathbf{C}_{x,y} \mathbf{C}_y^{-1} \mathbf{C}_{x,y}^H \).

A. Joint proper complex signals

As in [2, §3] and [8, §5.1], we adopt a joint proper (proper and cross-proper) complex signals assumption for the set of vector \( \{ x_0, \{ w_k \}, \{ v_k \} \} \) which allows to resort to standard estimation in the mean square 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 [8], and a zero mean proper complex random vector is said to be second-order circular [2, §3.2.5]. Indeed, in diverse fields such as communications, control, and signal processing, many practical problems of interest modelize the observation vector as complex and consisting of a bandpass signal which is the output of an Hilbert filtering leading to an “in-phase” real part associated to a “quadrature” imaginary part. These observation vectors belongs to the general class of analytic signals for which a fundamental result states that wide-sense stationary analytic signals, and also complex baseband representations of wide-sense stationary real bandpass signals, must be proper [8, §2.6.1]. Besides its capability to characterize the statistical behavior of the underlying physics of these problems of interest, the proper complex signals assumption also simplify computations and, in many aspects, make complex random signals look and behave like real random signals. Indeed, 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\(^2\).

B. Equivalent linear observation model

Here, \( F_{k,l} \in \mathcal{M}_C(P, P) \) and \( H_k \in \mathcal{M}_C(N, P) \). First, as (1a) can be rewritten as:

\[
x_k = B_k x_0 + \sum_{l=0}^{k-1} B_{k,l+1} w_l, B_{k,l} = \begin{cases} F_{k-l}, & k > l \\ I, & k = l \\ 0, & k < l \end{cases}
\]

an equivalent form of (1b) is:

\[
y_k = A_k x_0 + n_k, A_k = H_k B_{k,0}, n_k = \sum_{l=0}^{k-1} H_k B_{k,l+1} w_l + v_k.
\]

Second, (2) can be extended on a horizon of \( k \) points from the first observation as:

\[
\begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} A_1 \\ \vdots \\ A_k \end{bmatrix} x_0 + \begin{bmatrix} n_1 \\ \vdots \\ n_k \end{bmatrix} = \overline{A}_k x_0 + \overline{n}_k,
\]

\[\text{Note that if } x \text{ and } y \text{ are (proper) normal complex random vectors, then } C_{y|x} \text{ is exactly the covariance matrix of } x \text{ given } y.\]

\[\text{The case where } \{ x_0, \{ w_k \}, \{ v_k \} \} \text{ are improper complex random vector or hybrid vector (real and complex components) is addressed in Section V-D.}\]
where \( \overline{y}_k, \overline{\eta}_k \in \mathcal{M}_C(Nk, 1) \) and \( \overline{A}_k \in \mathcal{M}_C(Nk, P) \).

III. BACKGROUND ON LINEAR/AFFINE ESTIMATORS

Let us recall that the hypothesis of zero-mean process and measurement noise sequences is equivalent to the assumption of non-zero but known mean values\(^3\); it suffices to center the random variables before processing. Since this transformation is reversible, there is no loss of information in making such a preprocessing [2, p84].

A. Linear least-mean-squares estimator (LLMSE)

Let us consider two complex random vectors \( x \) and \( y \). The error between the signal \( x \) and the affine estimator \( \hat{x} = \hat{x}(y) = Ky + a, K \in \mathcal{M}_C(\dim(x), \dim(y)) \), \( a \in \mathcal{M}_C(\dim(x), 1) \), is \( e = e(y, x) = \hat{x}(y) - x \) and the error covariance matrix is:

\[
P(K, a) = E[ee^H] = E [\hat{x}(y) - x] (\hat{x}(y) - x)^H
\] (4)

As:

\[
e = K(y - m_y) - (x - m_x) + (a - (m_x - Km_y)) \quad (5)
\]

then, if \( C_y \) is invertible, (4) can be rewritten as [8, p121]:

\[
P(K, a) = C_{xy} + (K - C_{xy}C_{y}^{-1}) C_{y} (K - C_{xy}C_{y}^{-1})^H + (a - (m_x - Km_y)) (a - (m_x - Km_y))^H. \] (6)

This quadratic form in \( a \) is positive semidefinite, so \( P(K, a) \geq C_{xy} + (K - C_{xy}C_{y}^{-1}) C_{y} (K - C_{xy}C_{y}^{-1})^H \) with equality for:

\[
a^b = \arg\min_a \{ P(K, a) \} = m_x - Km_y, \] (7)

yielding:

\[
P(K, a^b) = C_{xy} + (K - C_{xy}C_{y}^{-1}) C_{y} (K - C_{xy}C_{y}^{-1})^H, \]

where the superscript \( b \) is used in the following to remind the reader that the value under consideration (scalar, vector, matrix) is the “best” one according to an unambiguous criterion previously defined (here the MSE). Then (5) reduces to:

\[
e = K(y - m_y) - (x - m_x) \] (leading to):

\[
K^b = \arg\min \{ P(K, a^b) \} = C_{xy}C_{y}^{-1}, \quad a^b = m_x - K^b m_y, \] (8)

and:

\[
\hat{x}^b = m_x - \hat{m}_x = C_{xy}C_{y}^{-1}(y - m_y) \] (9)

where:

\[
P(K^b, a^b) = C_{xy}. \] (10)

As \( K^b (8) \) is also the solution of the linear LMS after the centering of \( x \) and \( y \) (9), we follow [2] hereinafter and only use the term linear to designate both linear and affine LMS.

B. LMVDR estimator (LMVDR)

We adopt the notation used in the deterministic framework for the LMVDR [7, §6][8, §5.6] to stress the fact that the LMVDR is different from the LLMSE. Let us consider three complex random vectors \( x, y \) and \( v \) verifying:

\[
y = Hx + v, \quad H \in \mathcal{M}_C(\dim(y), \dim(x)) \]

is full rank. The error between the signal \( x \) and the affine estimator \( \hat{x} = \hat{x}(y) = Wy + a, W \in \mathcal{M}_C(\dim(y), \dim(x)) \), \( a \in \mathcal{M}_C(\dim(x), 1) \), is \( e = e(y, x) = \hat{x}(y) - x \) and the error covariance matrix is given by (4). In this setting, \( W \) can be seen as a “state-former” in the same way as a beamformer in array processing or a frequency-bin former in spectral analysis [7, §6][8, §5.6]. We look for a distortionless “state-former”, that is \( W^b \) and \( a^b \) such that:

\[
\{ W^b, a^b \} = \arg\min \{ P(W^H, a) \} \text{ s.t. } W^H H = I. \] (11)

Since (7) is valid whatever \( K = W^H \), then (5) reduces to:

\[
e = W^H(y - m_y) - (x - m_x) = \left( W^H H - I \right)(x - m_x) + W^H(v - m_v) = W^H(v - m_v), \]

leading to the following well known solution of (11) [3, p83][8, §5.6]:

\[
W^b = C_{v}^{-1}H \left( H^H C_{v}^{-1}H \right)^{-1}, \quad a^b = -\left( W^b \right)^H m_v, \] (12)

and:

\[
P(W^b, a^b) = \left( H^H C_{v}^{-1}H \right)^{-1}. \] (13)

Additionally, since we have restricted \( W \) to the subset satisfying \( W^H H = I \), we can assert that:

\[
P(W^b, a^b) \geq P(K^b, a^b). \] (14)

Then again, as \( W^b (12) \) is also the solution of the linear MVDRE after the centering of \( x, y \) and \( v \), we only use the term linear to designate both linear and affine MVDRE.

IV. LINEAR FILTERS FOR LDSS MODELS

A. LLMFS for LDSS models

It has been known for ages [1][19] that when \( x_0 \) is also zero-mean, the LLMSE of \( x_k \) based on measurements up to and including time \( k \) is simply (8):

\[
\hat{x}_k|k = [G_{k-1}^b K_k^b] y_k \quad \text{where} \quad G_{k-1}^b \in \mathcal{M}_C(P, N(k - 1)) \quad \text{and} \quad K_k^b \in \mathcal{M}_C(P, N). \]

What is less known, to the best of our knowledge (and not disclosed in recent monographs [2-5] or papers [14-17]), is the following general form of the LLMFS [18, Section IIJ]

\[
\hat{x}_k|k = F_{k-1} \hat{x}_{k-1}|k-1 + K_k^b (y_k - H_k F_{k-1} \hat{x}_{k-1}|k-1) + (I - K_k^b H_k) \left( C_{x_{k-1}|k-1}^{-1} \right) y_k - K_k^b \left( C_{\hat{x}_{k-1}|k-1}^{-1} \right) y_k - K_k^b \left( C_{\hat{x}_{k-1}|k-1}^{-1} \right) \] (16)

for \( k \geq 2 \), where \( K_k^b \) is given by [18, (13)] (\( K_k^b = W_k^b \)). One can notice that (16) has two additional terms in comparison with the recursive predictor/corrector form:

\[
\hat{x}_k|k = F_{k-1} \hat{x}_{k-1}|k-1 + K_k^b (y_k - H_k F_{k-1} \hat{x}_{k-1}|k-1) \] (17)

\(^3\)The case of unknown or unavailable mean values is addressed Section V-C.

\(^4\)Note that if \( x \) and \( y \) are (proper) normal complex random vectors, then \( \overline{x}^b \) is the mean value of \( \overline{x} \) given \( y \), i.e. the minimum MSE estimator.
introduced by Kalman [1]. Therefore, the general assumptions required to obtain the Kalman form (17) of the LLMSF for LDSS are:

\[ C_{\omega_{k-1}|y_{k-1}} = 0, \quad C_{v_k|y_{k-1}} = 0, \quad k \geq 2. \] (18)

For any LDSS model satisfying (18), then the associated LLMSF has a recursive predictor/corrector format (17), aka the KF, where \( K_k^d \) must be computed according to the following recursion for \( k \geq 1 \) [18, Section III:]

\[
\begin{align*}
P_{k|k-1} & = F_{k-1}P_{k-1|k-1}F_{k-1}^H + C_{w_{k-1}} \quad \text{(19a)} \nonumber \\
S_{k|k} & = H_kP_{k|k-1}H_k^H + C_{v_k} + H_kC_{v_{k-1}}H_k^H \quad \text{(19b)} \\
P_{k|k} & = (I - K_k^dH_k)P_{k|k-1} - K_k^dC_{v_k} \quad (19c)
\end{align*}
\]

provided that \( P_{0|0} = C_{x_0} \) and \( \hat{x}_{0|0} = 0^T \). Let us remind that:

\[
P_{k|k-1} = E \left[ (\hat{x}_{k|k-1} - x_k)(\hat{x}_{k|k-1} - x_k)^H \right] = C_{x_k|y_{k-1}},
\]

\( \hat{x}_{k|k-1} = F_k \hat{x}_{k-1} \) is the a priori estimate of \( x_k \), and

\[
S_{k|k} = E \left[ e_k e_k^H \right] = C_{v_k|y_{k-1}},
\]

\( e_k = y_k - H_k \hat{x}_{k|k-1} \) is the innovations vector. Note that the so-called "standard LDSS model" mentioned in monographs [2, §9.1][3, §7.1][4, §8.2], which satisfies:

\[ C_{x_0, w_0} = 0, \quad C_{x_0, v_0} = 0, \quad C_{w_{i}, w_{j}} = C_{w} \delta_{i,j}, \]

\[ C_{v_{i}, v_{j}} = C_{v} \delta_{i,j}, \quad C_{w_{i}, v_{j}} = C_{w_{i,j}} \delta_{i,j+1}, \]

and which has been regarded so far as leading to the general form of the KF (without extension of the state and measurement equations) including correlated process and measurement noise, is in fact a special case of (18) yielding the following simplified expressions of (19a-19c):

\[
\begin{align*}
P_{k|k-1} & = F_kP_{k-1|k-1}F_k^H + C_{w_{k-1}} \quad \text{(20a)} \\
S_{k|k} & = H_kP_{k|k-1}H_k^H + C_{v_k} + H_kC_{v_{k-1}}H_k^H \quad \text{(20b)} \\
P_{k|k} & = (I - K_k^dH_k)P_{k|k-1} - K_k^dC_{v_k} \quad \text{(20c)}
\end{align*}
\]

However, a thorough characterization of the subset of LDSS models compliant with (18) is out of the scope of the paper and is left for future research.

\[ Y\]
Then, according to (26a-26b):
\[ C_{n_{k-1}, n_k} = C_{\overline{n}_{k-1}, n_k} - x_{k-1} \]

However, (18) applied to (24) leads to:
\[ C_{y_{k-1}, y_k} = \overline{A}_{k-1} C_{x_{k-1}, y_k} + C_{\overline{n}_{k-1}, y_k} = 0. \]

Thus, under (18):
\[ \overline{A}_{k-1} F_{k-1} C_{x_{k-1}, y_k} + C_{\overline{n}_{k-1}, y_k} = 0. \]

Therefore, under (18):
\[ C_{n_{k-1}, n_k} = -\overline{A}_{k-1} F_{k-1} C_{x_{k-1}, n_k} - \overline{A}_{k-1} C_{w_{k-1}, n_k}, \]

that is:
\[ C_{n_{k-1}, n_k} = -\overline{A}_{k-1} \lambda_{k-1}, \quad \lambda_{k-1} = C_{x_{k-1}, n_k}, \]

which is a sufficient condition (22) to obtain a predictor/corrector format for the LMVDRF of the current state (21):
\[ \hat{x}_{k|k} = \hat{x}_{n|0} + (W_{b|k})^H (y_k - A_{k} \hat{x}_{n|k-1}) \quad (28) \]

where \( A_{k} = H_k (26b). \) Moreover some additional calculus detailed in [18, Section III.B] allow to show that under (18)
\[ \hat{x}_{n|k-1} = F_{k-1} \hat{x}_{n|k-1}, \quad (28) \]

finally becoming for \( k \geq 2: \)
\[ \hat{x}_{k|k} = F_{k-1} \hat{x}_{n|k-1} + (W_{b|k})^H (y_k - H_k F_{k-1} \hat{x}_{n|k-1}) \quad (29) \]

where:
\[ P_{d|k-1} = F_{k-1} P_{d|k-1} F_{k-1}^H + C_{w_k - 1} \]

\[ S_{d|k} = H_k P_{d|k-1} F_{k-1}^H + C_{v_k} + H_k C_{x_{k-1}, v_k} + C_{x_{k-1}, v_k} H_k^H \]

\[ W_{b|k} = (S_{d|k})^{-1} (H_k P_{d|k-1} + C_{x_{k-1}, v_k}) \quad (31a) \]

\[ P_{d|k} = (I - (W_{b|k})^H H_k) P_{d|k-1} - (W_{b|k})^H C_{x_{k-1}, v_k} \quad (32) \]

For \( k = 1, \) provided that \( H_1 \) is full rank, then (23):
\[ \hat{x}_{1|1} = P_{d|1}^{-1} H_{1}^{-1} C_{v_1} y_1 \quad (33a) \]
\[ P_{l|1} = (H_{1}^{-1} C_{v_1}^{-1} H_{1})^{-1} \quad (33b) \]

It is a bit of a surprise to obtain exactly the same recursion for the KF and the LMVDRF under the same conditions of existence (18), since the two solutions (8) and (12) do not look that similar in the general case. Anyway, this result highlights the key role of the initial condition at \( k = 1 \) for a recursive predictor/corrector filter. Indeed, despite they share the same recursion, the two filters have different properties:

- the LMVDRF is suboptimal in the MSE sense (14):
  \[ \forall \{m_{x_0}, C_{x_0}\} : P_{d|k} \leq P_{l|k}, \]

- since \( P_{d|k} \) (32)(33b) does not depend on \( \{m_{x_0}, C_{x_0}\}, \) therefore \( P_{d|k} \) is an upper bound on the KF performance,

- as the LMVDRF does not depend on the prior knowledge on the initial state \( x_0 \) (first and second order statistics), its behaviour can be assessed in advance and it can be pre-computed.

Thus, it is worth having in mind that anytime a KF is initialized with a WLSE as in (33a)(33b) because \( C_{x_0} \) and \( m_{x_0} \) are not known, it becomes a LMVDRF. However, the awareness of this implicit filter change is less important if the LDSS model is time-invariant, since \( P_{d|k|k} \) and \( P_{d|k-1} \) shares the same discrete algebraic Riccati equation (DARE). Therefore for the standard LDSS model (20), if \((F, H)\) is detectable and \((F - C_{w_k} C_{v_k}^{-1} H)\) is stabilizable \([3, \S 7.3]\), the LMVDRF and the KF share the same steady-state gain and performance.

Last, for completeness, let us recall that for the standard LDSS model (20), (33a)(33b) are also the asymptotic expressions of \( x_{1|1} \) and \( P_{d|1} \) in the KF when \( C_{x_0} = \alpha C \) and \( \alpha \) tends to infinity \([2, \S 3.4.2]\), i.e. a KF with “no a priori knowledge” is actually a LMVDRF.

V. LMVDR PREDICTOR, SMOOTHER AND FURTHER CONSIDERATIONS

A. LMVDR Predictor and Smoother

If the KF and the LMVDRF can share the same convenient recursive predictor/corrector format (17), it is first because this format propagates both (global) unbiasedness and distortionless properties. Indeed, it is known that rewriting (17) as [3]:
\[ \hat{x}_{k|k} - x_k = (I - K_{k|k} H_k) F_{k-1} (\hat{x}_{k-1|k-1} - x_{k-1}) - (I - K_{k|k} H_k) w_{k-1} + K_{k|k} v_k \]

allows to check the unbiasedness property propagation:
\[ E[\hat{x}_{k|k} - x_k] = 0 \quad E[\hat{x}_{k|k} - x_k] = 0. \quad (34) \]

In the LMVDRF case, let us remind that (25a):
\[ \hat{x}_{k|k} - x_k = W_{k-1} y_{k-1} = x_{k-1} + v'_{k-1} \]

where (24)
\[ \bar{y}_{k-1} = \overline{A}_{k-1} x_{k-1} + \overline{W}_{k-1}, v'_{k-1} = W_{k-1} \bar{y}_{k-1} \]

and \( W_{k-1} \overline{A}_{k-1} = I. \) Therefore (17) can be rewritten as:
\[ \hat{x}_{k|k} = (W_{b|k})^H H_k x_k + (W_{b|k}^H) H_k F_{k-1} (x_{k-1} + v'_{k-1}) \]

\[ = x_k + v'_k \]

where:
\[ v'_k = (W_{b|k})^H v_k + (I - (W_{b|k})^H H_k) (F_{k-1} v'_{k-1} - w_{k-1}), \]

demonstrating the distortionless property propagation:
\[ \hat{x}_{k-1|k-1} = x_{k-1} + v'_{k-1} \Rightarrow \hat{x}_{k|k} = x_k + v'_k. \quad (35) \]

By noticing that the KF and the LMVDRF both satisfy:
\[ \hat{x}_{k|k} - x_k = K_{k|k} v_k + (I - K_{k|k} H_k) (F_{k-1} (\hat{x}_{k-1|k-1} - x_{k-1}) - w_{k-1}) \]

they both can be derived as:
\[ (W_{b|k})^H K_{k|k} = \arg \min \{P_{d|k} (K_k)\}, \]
where:
\[ P_{k|k} (K_k) = (I - K_kH_k) P_{k|k-1} (I - K_kH_k)^H \]
\[ - (I - K_kH_k) C_{x_k,v_k} K_k^H - K_k C_{x_k,v_k} (I - K_kH_k)^H + K_k C_{x_k,v_k} K_k^H , \]
and \( P_{k|k-1} \) is given by (19a), leading to general form of the Joseph stabilized version of the covariance measurement update equation.

1) Standard LDSS model: As the LMVDRF only differs from the KF in the initial values of \( \hat{x}_{1|1} \) and \( P_{1|1} \), the LMVDRF can always be regarded as a particular instantiation of the KF for the following initial conditions: \( m_{x_0} = 0, \)
\[ C_{v_1,w_0} = -H_1 (H_1^H C_{v_1}^{-1} H_1) \]
\[ F_0 C_{x_0} F_0^H + C_{w_0} = 2 (H_1^H C_{v_1}^{-1} H_1)^{-1} , \]
since, then:
\[ S_{1|1} = C_{v_1}, K_1 = (H_1^H C_{v_1}^{-1} H_1)^{-1} H_1^H C_{v_1}^{-1}, \]
\[ \hat{x}_{1|1} = K_1 y_1. \]
Therefore, for the standard LDSS model, the two filters share:
a) the same possible extensions: colored process noise, colored measurement noise, prediction, smoothing, filtering with fading memory, constrained filtering,..., b) the same alternate formulations: information filtering, innovations approach,... [2][3]. It seems likely that the two filters also share the same drawbacks as sensitivity to modelling error [3, §10][10] or uncertainties in the system matrices [12][13]. However, a thorough comparison of the robustness [3, §10.4][11] of the two filters is far beyond the scope of this paper and is left for future research.

2) Non-standard LDSS model: As mentioned in Section IV-A, (18) defines the general assumptions required to obtain the filter recursive predictor/corrector format (without extension of the state and measurement equations) for LDSS models and encompass the standard LDSS model (20). To the best of our knowledge, the innovations process properties has been derived only for the standard LDSS model [2, §9] and for the standard linear continuous state space model [2, §16][20]:
\[ y(t) = z(t) + v(t), \quad z(t) = H(t) x(t) \]
\[ x(t) = F(t) x(t) + w(t) \]
where \( t \geq 0, \), \( z(t), v(t), w(t) \) are zero mean process verifying:
\[ C_{x,w} (0,t) = 0, \quad C_{w} (t,s) = C_{w} (t) \delta (t-s) , \]
\[ C_{v} (t,s) = C_{v} (t) \delta (t-s) , \quad C_{w,x} (t,s) = C_{w,x} (t) \delta (t-s) , \]
\[ \forall s < t : C_{w,x} (t,s) = 0, \quad \forall s < t : C_{w,y} (t,s) = 0, \]
and \( C_{x,y} (t,s) = E[x(t)y(s)^H] \). Therefore generalization of the innovations process properties to (18) can not easily be drawn from existing results and is not addressed in the present paper. However some straightforward extensions are available, at least the ones that are based on an explicit use of the predictor/corrector format (17)(19a)(19b)(19c)(36) as augmented system techniques, forward-backward recursion,..., [3] allowing to take into account colored process noise, colored measurement noise, prediction, smoothing, filtering with fading memory, constrained filtering, reduced order-filtering,...

B. The deterministic least-squares problem
If \( x_k = x_{k-1} = \ldots = x_0 \) where \( x_0 \) is deterministic and unknown, then the LMDRE and the WSE are identical (duality) [2, §3.4][4, §4], that is:
\[ \hat{x}_{0|0} = \arg \min \left\{ (y_k - \hat{A}_k x_0)^H C_{v_k}^{-1} (y_k - \hat{A}_k x_0) \right\} \]
(37)
Then one can easily notice that usual assumptions on the recursive WSE [2, §2.2.2][5, §4.12][6]:
\[ w_k = 0, \quad C_{v_1,v_k} = C_{v_k} \delta_k, \]
(38)
leads to (2):
\[ \hat{m}_k = \hat{v}_k, \quad \hat{C}_{m_k-1,n_k} = C_{v_k-1,v_k} = 0, \]
which is a particular case of (22) and can be regarded as an alternative proof of the recursive predictor/corrector format of the WSE in that particular case, i.e.:
\[ \hat{x}_{0|k}^d = \arg \min \left\{ \sum_{i=1}^{k} (y_i - A_i x_0)^H C_{v_i}^{-1} (y_i - A_i x_0) \right\} \]
\[ = \hat{x}_{0|k-1} + \left( W_{0|k}^b \right)^H \left( y_k - A_k \hat{x}_{0|k-1} \right) \]
where:
\[ W_{0|k}^b = \left( A_k \hat{A}_{0|k-1}^d + C_{v_k} \right)^{-1} A_k \hat{A}_{0|k-1}^d, \]
\[ \hat{A}_{0|k}^d = \left( I - \left( W_{0|k}^d \right)^H A_k \right) P_{0|k-1}. \]
The extension of this result to the regularized WSE (RWLSE) [2, §2.4][6]:
\[ \hat{x}_{0|k}^d = \arg \min \left\{ (x_0 - c)^H \Lambda_0^{-1} (x_0 - c) \right\} \]
\[ + \sum_{i=1}^{k} (y_i - A_i x_0)^H C_{v_i}^{-1} (y_i - A_i x_0) \]
where \( \Lambda_0 \) is an Hermitian invertible matrix, is simply obtained by adding a fictitious observation at time \( k = 0 \):
\[ y_0 = H_0 x_0 + v_0, \quad C_{v_0} = \Lambda_0, \quad y_0 = c, \quad H_0 = I, \]
and by starting the recursion at time \( k = 0 \):
\[ \hat{x}_{0|0} = P_{0|0} H_0^H C_{v_0}^{-1} y_0 = c, \quad P_{0|0} = (H_0^H C_{v_0}^{-1} H_0)^{-1} = \Lambda_0, \]
instead of time \( k = 1 \) (23). In the light of the above, the RWLSE is primarily a special case of the LMDRE, and its relation to the KF highlighted in [6] is actually purely formal.

C. Unknown mean values
So far, we have followed the didactic approach usually adopted in monographs dealing with linear filtering [2, §3.2.4][3][4][5, §4] which assumes that the mean values of \( (x_0, \{ w_k \}, \{ v_k \}) \) are available (theoretically known or estimated). However an analysis of the benefits of the knowledge of the mean values in term of estimation performance is of some interest for LDSS systems design. Indeed, on the one hand, the knowledge of the mean values allows to move from linear LMS/MVDR estimators to affine LMS/MVDR estimators in order to reduce the minimum attainable MSE (10)(13). On the other hand, in some systems, the centering of signals...
has a cost in terms of mean values estimation (DC offsets for instances) and of hardware or software implementation. First, when the mean values of \( \{x_0, \{w_k\}, \{v_k\}\} \) are not known, the following modification of the usual definition of proper complex random variables is required:

- a proper complex random variable \( x \) is orthogonal with its complex conjugate \( x^* \), i.e. \( E[x(x^*)^H] = E[xx^T] = 0 \)
- \( x \) and \( y \) are joint proper if: \( E[xx^T] = 0 \), \( E[yy^T] = 0 \) and \( E[xy^T] = 0 \).

Second, all the results introduced in Section IV hold provided one replaces covariance matrices for correlation matrices and set \( \tilde{x}(y) = Ky \) or \( \tilde{x}(y) = W^2y \). Then, a measure of the loss in MSE sense that occurs, if one does not take the mean values into account for the synthesis of the LLMS, is [18, Section IV]:

\[
\Delta P = R_{x|y} - C_{x|y} = (1 + m_2^H C_2^{-1} m_2)^{-1} a^b (a^b)^H \tag{39}
\]

As intuitively expected \( \Delta P \geq 0 \) with equality if and only if \( m_2 = K^m_2 m_2 \), what is unlikely to be true in the general case. Therefore in the general case, the diagonal element of \( \Delta P \) are strictly positive which is an evidence of the soundness of the centering if we are interested in the estimation of \( x \) only. Indeed, it is worth noting that \( \Delta P \) is a rank one positive matrix and that the improvement only occurs along \( span \{a^b\} \), which means that for any linear combination of \( x : \lambda^T x \mid \lambda \in span \{a^b\} \), there is no performance gain obtained by centering the random vectors. As \( \dim(\{a^b\})^\perp = P - 1 \), if one is primarily interested only in some linear combinations of \( \{\lambda^T x\} \), it is worth checking if \( \{\lambda\} \in span \{a^b\} \), especially if the random vectors centering is costly.

Last, the common recursive predictor/corrector format for the KF and the LMVDRF (\( \tilde{x}_{k|k} \equiv \tilde{x}_{k|k}, W_{k|k} \equiv (K_k^H) \)) is still:

\[
\tilde{x}_{k|k} = F_{k-1} \tilde{x}_{k-1|k-1} + K_k^b (y_k - H_k F_{k-1} \tilde{x}_{k-1|k-1})
\]

and is obtained for LDSS satisfying:

\[
R_{w_{k-1},y_{k-1}} = 0, \quad R_{v_{k},y_{k-1}} = 0, \quad k \geq 2, \tag{40}
\]

where:

\[
P_{k|k-1} = F_{k-1} P_{k-1|k-1} F_{k-1}^H + R_{w_{k-1}} \tag{41a}
\]

\[
S_{k|k} = H_k P_{k|k-1} H_k^H + R_{v_{k}}, \quad K_k^b = (P_{k|k-1} H_k^H + R_{v_{k},x_k}^H) S_{k|k}^{-1}, \tag{41b}
\]

\[
P_{k|k} = (I - K_k^b H_k) P_{k|k-1} - K_k^b R_{v_{k},x_k} \tag{41c}
\]

with initial conditions:

\[
K_{F}: \tilde{x}_{0|0} = 0, \quad P_{0|0} = R_{x_0} \tag{42a}
\]

\[
LMVDRF: \tilde{x}_{1|1} = P_{1|1} H_k^H R_{x_1, y_1}, \quad P_{1|1} = (H_k^H R_{x_1} H_k)^{-1} \tag{42b}
\]

and a KF MSE loss (39) given by \( (k \geq 1) \):

\[
\Delta P_{k|k} = (1 + \epsilon_k) a_k^b (a_k^b)^H, \tag{43}
\]

\[
a_k^b = u_k - K_k^b \delta_k, \quad \epsilon_k = \epsilon_k - \delta_k^H S_{k|k}^{-1} \delta_k, \tag{43}
\]

\[
u_k = F_{k-1} a_{k-1}^b + m_{w_{k-1}}, \quad \delta_k = H_k u_k + m_{v_k}, \tag{43}
\]

with initial conditions \( a_0^b = m_{x_0} \), and \( \epsilon_0 = 0 \).

### D. On the use of complex signals

It is worth reminding that complex LDSS models (1a)(1b) is simply the compact form of an equivalent real LDSS models:

\[
\begin{align*}
(x_k^c) &= \left[ F_{k-1}^c - F_{k-1}^r \right] (x_{k-1}^c) + \left( w_{k-1}^c \right), \\
(y_k^c) &= \left[ H_k^c - H_k^r \right] (x_{k-1}^c) + \left( v_k^c \right) \tag{41a}
\end{align*}
\]

where \( a^c, a^i, A^c \) and \( A^i \) denotes respectively the real and imaginary part of \( a \) and \( A \). Therefore, handling proper complex random vectors simply offers the capability of a computation burden reduction without loss of estimation performance in the MSE sense [2, [3,2.5][8, [5,4.1]], since the complex LDSS models manipulates vectors with twice less elements and matrices with four time less elements (but twice more additions and multiplications). If the checking of the “proper” property is problematic, one can always use a real LDSS models and all the results released in the present paper provided that one substitutes the matrix/vector transpose conjugate for the matrix/vector transpose.

### VI. AN ILLUSTRATIVE EXAMPLE

Consider the problem of the estimation of a bias from noisy measurements based on the following time invariant LDSS model:

\[
x_k = x_{k-1} + w_{k-1}, \quad y_k = x_k + v_k, \tag{44}
\]

where the noise process, the measurement noise and the initial state \( x_0 \) are Gaussian and uncorrelated (20): \( C_{x_0, w_k} = 0, C_{x_0, v_k} = 0, C_{w_k, w_k} = \sigma_w^2 \), \( C_{v_k, v_k} = \sigma_v^2 \), \( C_{w_k, v_k} = \sigma_{w,v}^2 \), \( C_{w_1, v_k} = 0 \).

Moreover, assume that \( x_0 = 0 \) and \( \sigma_v^2 = 100 \).

Figure (1) displays the root-MSE (RMSE) of the LMVDRF, that is \( \sqrt{P_{k|k}} \) (32), and the RMSEs of the KF, that is \( \sqrt{P_{k|k}} \) (19c), for \( C_{x_0} = \sigma_{x_0}^2 \in \{10^{-2}, 1, 10^2, 10^4\} \), when \( \sigma_{v_0}^2 = \frac{1}{10} \).

Figure (1) illustrates the properties of the LMVDRF mentioned in Section IV-B:

- the LMVDRF is suboptimal in the MSE sense: \( P_{k|k} \leq P_{d|k|k} \),
- \( P_{d|k} \) is an upper bound on the KF performance,
- if the LDSS model is time-invariant, the LMVDRF and the KF share the same steady-state performance,
- when \( \sigma_{x_0}^2 \) tends to infinity the KF tends to the LMVDRF.

As already mentioned, although the LMVDRF is suboptimal in terms of MSE, its performance does not depend on the prior knowledge (first and second order statistics) on the initial state, which can be a valuable feature as illustrated on Figure (2). Indeed in Figure (2) we assume that the true value of \( \sigma_{x_0}^2 \) (\( \sigma_{x_0}^2 = 1 \)) is not available and that one has to initialize the KF recursion with an assumed value
result since the LMVDRF is sub-optimal in MSE sense and is an upper bound on the performance of the KF whatever the first and second order statistics of the initial state. In the light of the above, the KF can be looked upon as a "initial state first and second order statistics" matched filter whereas the LMVDRF is a mismatched filter. Once known, these features are quite interesting for estimation performance analysis and design of a LDSS system since they allow to synthesize an IIR distortionless filter which performance does not depend on the unknown initial state. The LMVDRF may not be the best filters, but its behaviour can be assessed in advance and it can be pre-computed. Additionally we have shown that for the standard LDSS model, the LMVDRF properties (upper bound and pre-computation) can be extended to LMVDRF predictors and LMVDR smoother.

VII. CONCLUSION

For LDSS models, we have identified the general assumptions required to obtain, without extension of the state and measurement equations, a recursive predictor/corrector format of the LLLSF, aka the KF. We have shown that these assumptions, for a fairly general subset of LDSS models, allow the LMVDRF to share exactly the same recursion as the KF, except at time $k = 1$. Interestingly enough, numerous users of the KF, without knowing it, have used a LMVDRF each time they have initialized a KF at time $k = 1$ with a WLSE because the first and second order statistics of the initial state $x_0$ were not available. Indeed, it is worth being aware of this

REFERENCES