This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3103905, IEEE Transactions on Automatic Control

State Conditional Filtering

Algo Carè, *Member, IEEE* Marco Claudio Campi, *Fellow, IEEE,* and Erik Weyer, *Member, IEEE*

***Abstract*—In many dynamical state estimation problems, not all the values that the state can take have the same importance; hence, missing to deliver an appropriate estimate has more severe consequences for certain state values than for others. In many applications, such important state values correspond to events that have low a priori probability to happen (e.g., unsafe situations or conditions that one tries to avoid by design). Provably, Kalman filtering techniques are inadequate to correctly estimate such rare events. In this paper, a new state estimation paradigm is introduced to build confidence regions that contain the true state value, whatever this value is, with a user-chosen probability. Among regions having this property, an algorithm is introduced that generates in a Gaussian setup the region that satisfies a minimum-volume condition.**

***Index Terms*—Optimal Filtering, Kalman Filtering, State Conditional Property, Estimation Theory, Gaussian Process**

NOMENCLATURE

KF	Kalman Filter.
SCF	State Conditional Filter.

Dynamical system variables and dimensions

x_t	$(n \times 1)$-dimensional system state.
y_t	$(q \times 1)$-dimensional measurement.
v_t	$(n \times 1)$-dimensional state noise.
w_t	$(q \times 1)$-dimensional measurement noise.
F	$n \times n$ state transition matrix.
H	$q \times n$ output transform matrix.
Γ, V, W	Covariance matrices of x_t, v_t and w_t.
F_b	Backward state transition matrix.
v_t^b, V_t^b	Backward state noise and its covariance matrix.
$[y]_1^t$	Short form for vector $[y_1^\top, \dots, y_t^\top]^\top$.
ζ_t	$(s \times 1)$-dimensional sub-vector of x_t (components of x_t to be estimated).

Dynamical point estimates and set estimates

α	probability that the set estimate includes x_t (user-chosen value).
$\hat{x}_t^{\rm KF}$, $\mathcal{X}_t^{\rm KF}$	KF point estimate and KF set estimate (of x_t given the measurements $y_1, \dots, y_t$).
$\hat{x}_t$, $\mathcal{X}_t^{\rm SCF}$	SCF point estimate and SCF set estimate (of x_t given the measurements $y_1, \dots, y_t$).
$\hat{\zeta}_t$, $\mathcal{Z}_t^{\rm SCF}$	SCF point estimate and SCF set estimate (of ζ_t given the measurements $y_1, \dots, y_t$).

Static estimation variables and dimensions

x	$(n \times 1)$-dimensional vector to be estimated.
y	$(p \times 1)$-dimensional output vector.
ν	$(p \times 1)$-dimensional noise vector.
A	$p \times n$ regressor matrix ($Ax + \nu = y$).
ζ	$(s \times 1)$-dimensional sub-vector of x (components of x to be estimated).

Static estimation tools and operators

$\mathbb{X}$, $\mathbb{Y}$, $\mathbb{Z}$	The linear span of the random variables in vector x, y and ζ, respectively.
$\mathcal{P}_{\mathbb{X}}(\cdot), \mathcal{P}_{\mathbb{Y}}(\cdot), \mathcal{P}_{\mathbb{Z}}(\cdot)$	Component-wise projection of random vector onto $\mathbb{X}$, $\mathbb{Y}$ and $\mathbb{Z}$, respectively.
$\mathscr{P}_{\mathcal{U}}$	The *unconstrained* minimum variance linear estimation problem.
$\hat{x}^{\mathcal{U}}$	Solution to problem $\mathscr{P}_{\mathcal{U}}$ (Bayesian estimator).
$Z^{\mathcal{U}}$	Matrix such that $\mathcal{P}_{\mathbb{X}}(\hat{x}^{\mathcal{U}}) = Z^{\mathcal{U}}x$.
$\mathscr{P}_{\mathcal{C}}(Z)$	The *constrained* minimum variance linear estimation problem instantiated with a matrix Z.
$\hat{x}_Z^{\mathcal{C}}$	Solution to problem $\mathscr{P}_{\mathcal{C}}(Z)$.
$\mathscr{P}'_{\mathcal{C}}(Z)$	Reformulation of $\mathscr{P}_{\mathcal{C}}(Z)$ in matrix terms.
T_Z	Matrix solution to problem $\mathscr{P}'_{\mathcal{C}}(Z)$ (relation $T_Z y = \hat{x}_Z^{\mathcal{C}}$ holds).
U	Shorthand for $T_{Z^{\mathcal{U}}}$.
$\mathcal{X}$	Set estimate of x, with conditional property, minimum-volume as per Theorem 1.
C	$n \times p$ matrix representing a generic linear compression.
$\mathcal{S}(\cdot)$	Generic map from a vector in $\mathbb{R}^n$ to a subset of $\mathbb{R}^n$.

Other symbols and shorthands

$\mathbb{P}r\{E\}$	Probability of the event E.
$\mathbb{P}r\{E_1 \vert E_2\}$	Probability of the event E_1 conditional on E_2.
$\mathbb{E}[\,\cdot\,]$	Expectation operator.
$\mathcal{G}(\mu, \Sigma)$	Gaussian distribution with mean μ and covariance matrix Σ.
$\chi^2(\alpha, n)$	Quantile at probability α of the Chi-square distribution with n degrees of freedom.
$\mathcal{E}_\Sigma^\alpha(\mu)$	Minimum volume ellipsoid covering a probability α of the $\mathcal{G}(\mu, \Sigma)$ distribution.
$\mathcal{B}_r$	Closed ball centered in 0 with radius r.
${\rm Vol}(S)$	Lebesgue measure of the set S.
$\mathbb{1}(\cdot)$	The indicator function (1 when statement is true; 0 otherwise).

Matrices

$M^\top$	Transpose of matrix M.
M^{-1}	Inverse matrix of M.
$M_1 \preceq M_2$	Matrix $M_2 - M_1$ is positive semidefinite.

A. Carè and M.C. Campi are with the Department of Electrical and Computer Engineering, University of Brescia, Brescia, Italy (e-mail: [algo.care,marco.campi]@unibs.it). E. Weyer is with the Department of Electrical and Electronic Engineering, The University of Melbourne, Melbourne, VIC, Australia (email: ewey@unimelb.edu.au).

I. INTRODUCTION

We consider a process $\{x_t, y_t\}$, $x_t \in \mathbb{R}^n$ and $y_t \in \mathbb{R}^q$ generated by a linear time-invariant system

$$\left\{ \begin{array}{rcl} x_{t+1} & = & Fx_t + v_t \\ y_t & = & Hx_t + w_t, \end{array} \right. \tag{1}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3103905, IEEE Transactions on Automatic Control

where $\{v_t\}$ and $\{w_t\}$ are Gaussian stationary white processes, independent of each other. The state x_t is not directly accessible and the goal is to estimate x_t based on the measurements $y_1, \dots, y_t$. One way to address this problem is to use the Kalman Filter (KF), [1], [2], which has found ubiquitous applications in myriad fields, see e.g. [3-11]. In this paper, we argue that KF lacks to satisfy a property – which we call "state-conditional property" – that plays a significant role in various applications (see e.g. Example 1 below and the comments thereafter). Therefore, we introduce a new state estimation method, called State Conditional Filter (SCF), which has this property.

To better understand the new estimation method, we first revisit the KF approach. In the Gaussian case, KF recursively computes the conditional density of x_t given the measurements $y_1, \dots, y_t$. Once this density has been found, the region of minimum volume that contains a given percentage α, say $\alpha = 99\%$, of the conditional distribution can be constructed. This region, which we name $\mathcal{X}_t^{\rm KF}$, is the ellipsoid centered at the conditional mean $\hat{x}_t^{\rm KF}$ of x_t given $y_1, \dots, y_t$ while its shape is determined by the conditional variance of x_t given $y_1, \dots, y_t$, which does not depend on the measurements and is obtained by solving the Riccati equation. As a result, $\mathcal{X}_t^{\rm KF}$ has the same shape and volume irrespective of the measurements while its location in the state space is measurement-dependent. Moreover, since for any given value of the measurements the conditional probability with which $x_t \in \mathcal{X}_t^{\rm KF}$ equals α, α is also the total probability with which $x_t \in \mathcal{X}_t^{\rm KF}$:

$$\mathbb{P}r\{x_t \in \mathcal{X}_t^{\rm KF}\} = \alpha. \tag{2}$$

In this paper we are interested in constructing a region $\mathcal{X}_t$ that, regardless of the actual value of the system state x_t, has the additional property to contain x_t with a given probability α. We call this property the "state-conditional property", which is formalized in the following.

Property 1 (state conditional property)**.** For a given $\alpha \in (0,1)$, it holds that $\mathbb{P}r\{x_t \in \mathcal{X}_t | x_t = \bar{x}\} = \alpha$, $\bar{x}$-almost everywhere (a.e.). $*$

Although the state value is not a measurable quantity, we shall see that Property 1 can be rigorously enforced in SCF. While Property 1 implies that $\mathbb{P}r\{x_t \in \mathcal{X}_t\} = \alpha$ as in KF, it also bears additional implications that determine a major paradigm shift from KF, and we feel advisable to introduce at this early stage a simple example to highlight the differences between KF and SCF. In Section II, a more concrete example in mobile robotics will also be presented which further motivates the new approach of this paper.

Example 1 (gambling with a game value $v(x)$)**.** Consider the system (1). At a given time t, the player has to deliver a set $\mathcal{X}_t$ and wins the bet if $x_t \in \mathcal{X}_t$. Before the game starts, the dealer chooses a function $v(x)$, called the value of the bet. In case of win, the player is given the amount $v(x_t)$; on the other hand, if $x_t \notin \mathcal{X}_t$ the player loses the bet and has to pay the amount $99 \cdot v(x_t)$. Hence, the average return over N bets is given by

$$r_N = \frac{1}{N} \sum_{i=1}^{N} v(x_t^{(i)}) \left[\mathbb{1}(x_t^{(i)} \in \mathcal{X}_t^{(i)}) - 99 \cdot \mathbb{1}(x_t^{(i)} \notin \mathcal{X}_t^{(i)}) \right],$$

where $\mathbb{1}(\cdot)$ is indicator function and superscript (i) indicates the bet number.

Suppose that a player wishes to construct a set $\mathcal{X}_t$ so that the game is fair, that is, $\mathbb{E}\left[v(x_t)\left[\mathbb{1}(x_t \in \mathcal{X}_t) - 99 \cdot \mathbb{1}(x_t \notin \mathcal{X}_t)\right]\right] = 0$ (under mild assumptions on $v(x)$, in a fair game r_N almost surely tends to zero in the long run). It is a fact that $\mathcal{X}_t^{\rm KF}$ with $\alpha = 99\%$ does not obtain this result. To see this, consider for concreteness the scalar case where $n = q = 1$, assume $F = \sqrt{\frac{5}{6}}$, $H = 1$ and $\{v_t\}$ and $\{w_t\}$ to be zero mean with $\mathbb{E}[v_t^2] = \mathbb{E}[w_t^2] = 1$. Let $\hat{x}_t^{\rm KF} = \mathbb{E}[x_t | y_1, \dots, y_t]$ be the Kalman filter estimate and let $p_t = \mathbb{E}[(\hat{x}_t^{\rm KF} - x_t)^2]$ be its error variance. A simple application of the KF equations gives $\lim_{t\to\infty} p_t = \frac{3}{5} =: \bar{p}$. The conditional probability of x_t given the observations is Gaussian and, if t is large, $\mathcal{X}_t^{\rm KF}$ is the interval $[\hat{x}_t^{\rm KF} - \rho, \hat{x}_t^{\rm KF} + \rho]$, where ρ is such that $\int_{-\rho}^{\rho} \frac{1}{\sqrt{2\pi\bar{p}}} \exp\left(-\frac{x^2}{\bar{p}}\right) \mathrm{d}x = 0.99$, which gives $\rho \approx 2$. Note that the size of the interval ρ does not depend on the observations. Now, the density of $\hat{x}_t^{\rm KF}$ conditional on the state value $x_t = \bar{x}$ is a Gaussian with mean $\left(1 - \frac{p_t}{\mathbb{E}[x_t^2]}\right)\bar{x}$ (which tends to $0.9\bar{x}$ as $t \to \infty$) and variance $p_t\left(1 - \frac{p_t}{\mathbb{E}[x_t^2]}\right)$ (which tends to 0.54 as $t \to \infty$).[1] Hence, asymptotically, the center of the KF interval (which is $\hat{x}_t^{\rm KF}$) has a mean value $0.9\bar{x}$ that takes a gap from $\bar{x}$ that increases with $\bar{x}$; consequently, the probability of $\{x_t \in \mathcal{X}_t^{\rm KF}\}$ conditional on $x_t = \bar{x}$ goes to zero for increasing values of $\bar{x}$, see also Figure 1. As a consequence, if $v(x)$ is large for large values of x and small otherwise, e.g.,

$$v(x) = \begin{cases} 100 & \text{if } |x| > 100 \\ 0 & \text{otherwise,} \end{cases}$$

the game is unfair. This result is perhaps not surprising since KF implements a Bayesian logic that relies on the use of the prior belief that large values of x seldom happen: according to this logic, in this example the optimal exploitation of the information contained in the measurements results in pulling back the value $\bar{x}$ to $0.9\bar{x}$ (which only results in a severe penalty for large values of $\bar{x}$, a condition that rarely happens) because this allows the algorithm to become less sensitive to noise.

[1] To see this, note that the vector $[\hat{x}_t^{\rm KF} \;\; x_t]^\top$ is Gaussian so that the conditional mean of $\hat{x}_t^{\rm KF}$ given $x_t = \bar{x}$ is linear in $\bar{x}$ with coefficient $c = \mathbb{E}[\hat{x}_t^{\rm KF} x_t]/\mathbb{E}[x_t^2]$. On the other hand, it is known that $\mathbb{E}[(\hat{x}_t^{\rm KF} - x_t)\hat{x}_t^{\rm KF}] = 0$ (orthogonality between the estimation error and the estimate in KF), which gives $p_t = \mathbb{E}[(\hat{x}_t^{\rm KF} - x_t)^2] = -\mathbb{E}[(\hat{x}_t^{\rm KF} - x_t)x_t] = -\mathbb{E}[\hat{x}_t^{\rm KF} x_t] + \mathbb{E}[x_t^2]$. Substituting this equation in the expression for c yields $c = 1 - \frac{p_t}{\mathbb{E}[x_t^2]}$, which provides the result on the mean. As for the variance, recall that the conditional variance of jointly Gaussian variables does not depend on the conditioning value (here $\bar{x}$) and can therefore be computed as a total variance: $\mathbb{E}\left[\left(\hat{x}_t^{\rm KF} - \left(1 - \frac{p_t}{\mathbb{E}[x_t^2]}\right)x_t\right)^2\right] = \mathbb{E}\left[\left((\hat{x}_t^{\rm KF} - x_t) + \frac{p_t}{\mathbb{E}[x_t^2]}x_t\right)^2\right] = p_t + \frac{p_t^2}{\mathbb{E}[x_t^2]} - 2\frac{p_t}{\mathbb{E}[x_t^2]}p_t = p_t\left(1 - \frac{p_t}{\mathbb{E}[x_t^2]}\right)$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3103905, IEEE Transactions on Automatic Control

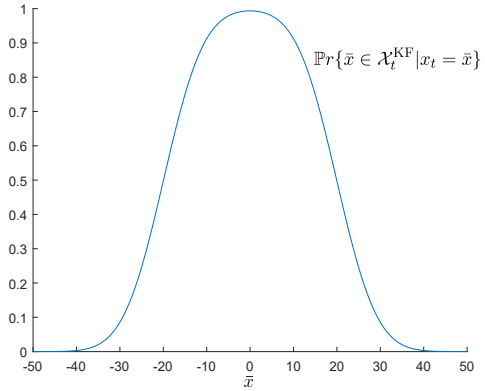


Figure 1: Violation of the state-conditional property with the Kalman filter. The plot shows the asymptotic probability with which $x_t \in \mathcal{X}_t^{\mathrm{KF}}$ for values $\bar{x}$ of x_t in the range [-50,50].

Suppose now that the state conditional Property 1 is enforced in the estimation method. Then,

$$\begin{aligned}\mathbb{E}\left[v(x_t)\left[\mathbb{1}(x_t \in \mathcal{X}_t) - 99 \cdot \mathbb{1}(x_t \notin \mathcal{X}_t)\right]\right] \\ = \mathbb{E}\left[v(\bar{x}) \cdot \mathbb{E}\left[\mathbb{1}(x_t \in \mathcal{X}_t) - 99 \cdot \mathbb{1}(x_t \notin \mathcal{X}_t) | x_t = \bar{x}\right]\right] \\ = \mathbb{E}\left[v(\bar{x}) \cdot 0\right] = 0,\end{aligned}$$

that is, the game now becomes fair. As we shall see in Section IV, a filter that constructs an estimation with the state conditional property can be built in full generality and, like KF, it admits a recursive implementation. For the problem at hand, the state-conditional filter builds an interval that, for large t, is given by $\left[\frac{1}{0.9}\hat{x}_t^{\mathrm{KF}} - \sqrt{\frac{1}{0.9}}\rho, \frac{1}{0.9}\hat{x}_t^{\mathrm{KF}} + \sqrt{\frac{1}{0.9}}\rho\right]$, where ρ is as before. Note that the center is the KF estimate rescaled by a factor $\frac{1}{0.9}$ and that the square root of the same factor $\frac{1}{0.9}$ amplifies the size of the interval. Provably, this choice ensures a fair game with the smallest possible enlargement of the size of the prediction region as compared to KF. Before closing this example, we also want to note that the game is fair regardless of the choice of the value function $v(x)$, which can also be unknown to the player. $*$

The situation described in this example where not all state values are equally important is found in many applications. For example, in mobile robotics importance is placed on avoiding collisions with fixed objects or among robots. Section II gives an example in this context. Similarly, in aerial or marine navigation, a reliable prediction of the actual position is crucial in relation to safety critical locations. Moreover, not all operating states are equally important in industrial and medical applications. For instance, in fluidized bed combustors, the rise of the coal load above a certain threshold significantly increases the generation of pollutants; in die casting machines, one is concerned about conditions possibly leading to hazardous overpressures; and, in various medical applications, certain conditions (e.g. even slight increases in lung density) can be a sign of an emerging disease. Often, critical state conditions have low marginal (prior to observations) probability, so that they are not included in $\mathcal{X}_t^{\mathrm{KF}}$. In contrast, Property 1 enforces that, whatever the value of the system state is, it is included in the set $\mathcal{X}_t$ with a probability chosen by the user irrespective of its marginal probability.

Property 1 enforces an inclusion condition irrespective of the state value. This is relevant to situations where important state values are not a priori known, which happens e.g. where different units perform their own state evaluation and the critical conditions depend on a simultaneous occurrence of certain values of the state of various units. As an example, one can think of the position of various vehicles in a navigation problem where each vehicle computes a value set for its own position and communicates it to a centralized control unit which, based on the communicated positions, delivers instructions to avoid collisions. On the other hand, in many applications prior knowledge about the important state values is available. Including such knowledge in the construction of the filter is not considered in the present paper and future work is expected to extend in this direction the theory here developed.

The set $\mathcal{X}_t^{\mathrm{SCF}}$ constructed by the State Conditional Filter introduced in this paper turns out to depend on an n-dimensional linear compression of $y_1, \ldots, y_t$, which provides a finite information state that enables a recursive implementation of the filter. It will also be shown that $\mathcal{X}_t^{\mathrm{SCF}}$ has a minimum volume property; more specifically, $\mathcal{X}_t^{\mathrm{SCF}}$ achieves minimum of the cost index $\sup_{y_1,\ldots,y_t} \mathrm{Vol}(\mathcal{X}_t)$ over all regions $\mathcal{X}_t$ that satisfy Property 1 and that are based on an n-dimensional linear compression of $y_1, \ldots, y_t$. It turns out that the compression in $\mathcal{X}_t^{\mathrm{SCF}}$ coincides with that in $\hat{x}_t^{\mathrm{KF}}$ and, moreover, $\mathrm{Vol}(\mathcal{X}_t^{\mathrm{SCF}})$ does not depend on the measurements $y_1, \ldots, y_t$. As compared to KF, securing Property 1 requires that $\mathcal{X}_t^{\mathrm{SCF}}$ has an increased volume with respect to $\mathcal{X}_t^{\mathrm{KF}}$. We shall see that the volume increase depends on how informative the measurement mechanism is and the volume of $\mathcal{X}_t^{\mathrm{SCF}}$ approaches the volume of $\mathcal{X}_t^{\mathrm{KF}}$ for systems with highly informative measurement mechanisms.

The structure of the paper is as follows. In the next Section II a numerical example in mobile robotics is presented which further motivates and clarifies the approach of this paper (Section II can be skipped without any loss of theoretical continuity). Section III presents the fundamental mathematical tools in a static context, where a geometric interpretation is also described. This section paves the way for an easy understanding of the material relating to the dynamical set-up in Section IV. The SCF is introduced in Section IV-A; the asymptotic properties of SCF are studied in Section IV-B where the relation between $\mathcal{X}_t^{\mathrm{KF}}$ and $\mathcal{X}_t^{\mathrm{SCF}}$ is also discussed. Various extensions and variations of the basic set-up, including filtering for time-varying systems and prediction algorithms, are presented in Section V. The paper ends with some concluding remarks in Section VI.

II. A NUMERICAL EXAMPLE

In this section, the SCF algorithm introduced in this paper is applied to a simulated set-up in mobile robotics.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3103905, IEEE Transactions on Automatic Control

A. Problem setting and model description

Two mobile robots A and B execute a "return to base" program, [12], [13]. During the operation, a collision between the robots has low probability by design and yet a collision is not impossible. The estimation problem consists in predicting with high probability a collision whenever one is about to occur.

The state variables of each robot are its position (p_x, p_y) and its velocity (u_x, u_y). Each robot is acceleration-controlled through an on-board state-feedback controller that is designed to drive it to its base-station. In the simulation, robot A is initially standing still at the position $(p_{xA,1}, p_{yA,1}) = (-1, 20)$ and has to go to its base-station located at $(\bar{p}_{xA}, \bar{p}_{yA}) = (-1, 0)$, while robot B is still at $(p_{xB,1}, p_{yB,1}) = (1, 20)$ and has to go to its base-station at $(\bar{p}_{xB}, \bar{p}_{yB}) = (1, 0)$. The on-board controller applies to the robot an acceleration a_t that depends on the distance $e_t = \bar{p} - p_t$ (this is a 2-dimensional vector) of the robot from its base-station according to the formula $a_t = K_p \cdot e_t + K_u \cdot \dot{e}_t$ ($\dot{e}_t$ is the time derivative of e_t). By discretization, we obtain the following equation for the movement of robot A executing its "return to base" program

$$\begin{bmatrix} p_{xA,t+1} \\ p_{yA,t+1} \\ u_{xA,t+1} \\ u_{yA,t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta & 0 \\ 0 & 1 & 0 & \Delta \\ -K_p \cdot \Delta & 0 & 1 - K_u \cdot \Delta & 0 \\ 0 & -K_p \cdot \Delta & 0 & (1 - K_u) \cdot \Delta \end{bmatrix} \begin{bmatrix} p_{xA,t} \\ p_{yA,t} \\ u_{xA,t} \\ u_{yA,t} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ K_p \bar{p}_{xA} \\ K_p \bar{p}_{yA} \end{bmatrix} \Delta + v_{A,t} \cdot \Delta, \tag{3}$$

where Δ is the discretization step and $v_{A,t}$ is the stochastic vector that accounts for the effects of modeling errors and sensor and actuator noises. In this example, $\Delta = 1/20$, $K_p = 5, K_u = 5$, and $v_{A,t}$ is white, zero mean and Gaussian with covariance matrix

$$V_A = \begin{bmatrix} 0.01^2 & 0 & 0 & 0 \\ 0 & 0.01^2 & 0 & 0 \\ 0 & 0 & 0.04^2 & 0 \\ 0 & 0 & 0 & 0.04^2 \end{bmatrix}.$$

The movement of robot B is similarly described by replacing "A" with "B" in equation (3), while the covariance matrix for robot B is only slightly different from V_A,

$$V_B = \begin{bmatrix} 0.01^2 & 0 & 0 & 0 \\ 0 & 0.01^2 & 0 & 0 \\ 0 & 0 & 0.1^2 & 0 \\ 0 & 0 & 0 & 0.04^2 \end{bmatrix}.$$

The noises affecting robot A and robot B are independent of each other. For estimating its own position, each robot uses the measurements

$$y_{A,t} = \begin{bmatrix} p_{xA,t} \\ p_{yA,t} \end{bmatrix} + w_{A,t}, \quad y_{B,t} = \begin{bmatrix} p_{xB,t} \\ p_{yB,t} \end{bmatrix} + w_{B,t},$$

where $w_{A,t}$ and $w_{B,t}$ are measurement noises with covariance matrix $W_A = W_B = 0.05^2 I$, independent of each other. The predictions are sent to a central unit which, in case a collision is foreseen, adopts suitable counter-measures to avoid it. While the above equations represent the two robots as material points, the robots are actually bodies with a non-zero dimension and collide whenever $\|(p_{xA,t}, p_{yA,t}) - (p_{xB,t}, p_{yB,t})\| < d_{\min}$. In our example, the collision distance is set to $d_{\min} = 0.2$, and this is one tenth of the relative distance that the robots would maintain if there were no noise. This makes the event of a collision rare, even though not impossible.

B. Prediction using SCF and KF

Figure 2 shows the measurements $y_1, \ldots, y_{60}$ in a simulated run of the "return to base" program. The dots are a noisy version of the robots' positions as they move from the initial positions to the base-stations. Figure 3 shows a run where a collision occurs after 40 time steps: Figure 3a displays the measurements, while Figure 3b shows some of the robots' true positions along with the positions that are estimated by $\hat{x}_t^{\mathrm{KF}}$ and $\hat{x}_t$ (the latter is the point around which the SCF region is constructed). One can see that KF provides an "optimistic" estimate when the low probability event of a collision happens. To better appreciate the difference between KF and SCF, a Monte Carlo test was run on 10 000 low probability realizations leading to a collision at time 40,[2] and a 95% probability region was constructed both with KF and SCF. Referring e.g. to robot A, it turned out that the state of this robot was never included in $\mathcal{X}_{40}^{\mathrm{KF}}$ in the 10 000 runs, while it was included in $\mathcal{X}_{40}^{\mathrm{SCF}}$ 9 498 times, a result in line with the theoretical guarantee of 95%. The reason of this enormous discrepancy in the behavior of the two algorithms is due to the fact that a Kalman filter succeeds in correctly predicting the true system state with a user chosen probability; if this probability is, say, 99%, then on average the prediction fails in one case out of 100 only, however, the user has no control on when this happens and it may happen almost systematically in dangerous situations that occur with a probability that is no more than 1%. In contrast, with the new algorithm SCF one keeps full control on the probability

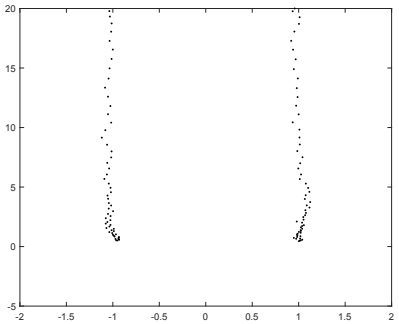
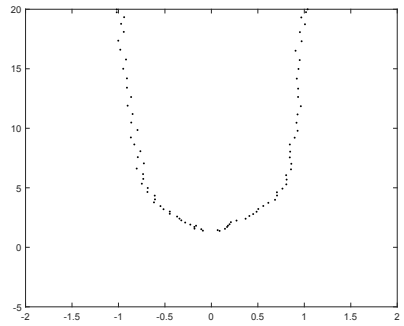


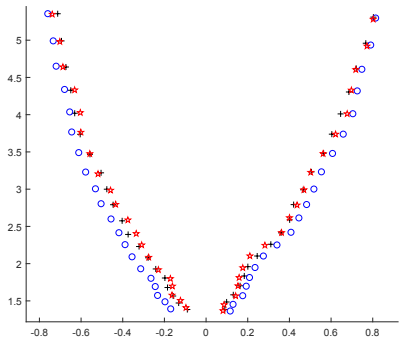
Figure 2: Measured positions of the robots for a simulated execution of the "return to base" program.

[2]The reader may be interested to know that these low probability realizations were generated backwards, i.e., from the colliding state at time 40 to the initial state x_1 by using the backward representation of the robot equations obtained according to (45).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3103905, IEEE Transactions on Automatic Control



(a) Measurements.



(b) True positions (+); SCF estimates ($\star$); KF estimates ($\circ$).

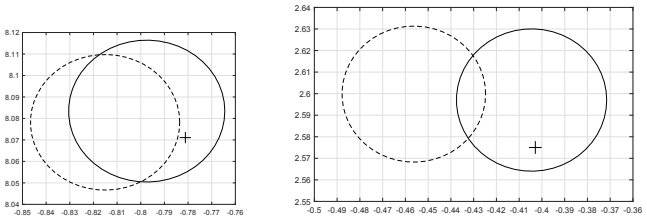
Figure 3: A rare realization with a collision.

of misprediction for all possible values of the state (including dangerous ones) at a desired level of probability.

Up to here, we have considered the estimate of all 4 state variables of robot A. Suppose instead that one wants to estimate only the position of the robot (2 variables out of 4). Estimating the position is possible according to the approach of Section V-A. Let $\zeta_t = (p_{xA,t}, p_{yA,t}) \in \mathbb{R}^2$ and consider the measurements in Figure 3. Figure 4 shows the 95% probability regions $\mathcal{Z}_{20}^{\rm SCF}$ and $\mathcal{Z}_{36}^{\rm SCF}$ for ζ_t; the KF regions $\mathcal{Z}_{20}^{\rm KF}, \mathcal{Z}_{36}^{\rm KF}$ are also displayed. Although ζ_t is included in the KF region with probability 95%, in our simulation corresponding to a rare realization leading up to a collision, ζ_t is not contained in the KF region. This fact does not happen with SCF due to the enforcement of the state conditional Property 1.

The simulations showed so far were referring to estimating the state at time t based on measurements up to and including time t. On the other hand, in a real application on collision avoidance, predicting the position ahead of time is more relevant, and Figure 5 shows the 4-step ahead prediction region $\mathcal{Z}_{40|36}^{\rm SCF}$ (based on the measurements up to time 36), compared with the region $\mathcal{Z}_{40|36}^{\rm KF}$ obtained with the Kalman predictor. Such prediction regions are constructed according to the approach of Section V-C.

To appreciate the ability of SCF to forecast a collision, we



(a) $t = 20$ (b) $t = 36$

Figure 4: The true position of robot A at time t (+), $\mathcal{Z}_t^{\rm SCF}$ (solid ellipsoid), and $\mathcal{Z}_t^{\rm KF}$ (dashed ellipsoid).

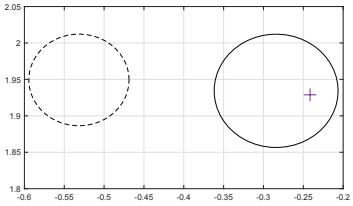


Figure 5: The true position of robot A at time 40 (+), $\mathcal{Z}_{40|36}^{\rm SCF}$ (solid ellipsoids), and $\mathcal{Z}_{40|36}^{\rm KF}$ (dashed ellipsoids).

counted how many times a warning of a possible collision at time 40 was issued at time 36 in the 10 000 Monte Carlo simulations. That is, we checked if the regions for the predicted positions of robot A and B included points at a distance smaller than d_{min}. This happened in all cases with SCF, while only in 2 out of 10 000 cases with KF. Finally, we note that satisfying the state-conditional property comes at the price of a moderate enlargement of the region $\mathcal{Z}_{40|36}^{\rm SCF}$ as compared to $\mathcal{Z}_{40|36}^{\rm KF}$.

III. The static set-up

In this section, we consider a vector x which is related to a vector y of measurements by a linear static relation, and study the problem of constructing a minimum volume region $\mathcal{X}$ for x which satisfies the state-conditional property that $\mathbb{P}r\{x \in \mathcal{X} | x = \bar{x}\} = \alpha$, $\bar{x}$-a.e. It turns out that the problem of finding the minimum volume region $\mathcal{X}$ in a Gaussian framework is tightly related to a minimum variance linear estimation problem. In the following Section III-A, the unconstrained and the constrained minimum variance linear estimation problems are considered. Some of the results in this section can be found in the literature in scattered form, see e.g. [14-17] and Remark 6 below, and are here presented in a unified geometric framework. This framework will be used in Section III-B to obtain a region $\mathcal{X}$ with minimum volume (in a sense precisely stated in Section III-B) that satisfies the state conditional property.

A. Unconstrained and constrained minimum variance linear estimation

Consider two random vectors x and ν, $x \in \mathbb{R}^n$, $\nu \in \mathbb{R}^p$. No Gaussianity assumption on x and ν is introduced at this

point, so that x and ν have generic distributions. Let A be a $p \times n$ matrix, and let

$$y = Ax + \nu. \tag{4}$$

One has to estimate x from y, which is interpreted as a measurement vector affected by the noise ν. The following condition is in order.

Condition 1. x and ν are mutually uncorrelated random vectors with finite second order moments and $\mathbb{E}[x] = 0$, $\mathbb{E}[\nu] = 0$, $\mathbb{E}[xx^\top] \succ 0$, $\mathbb{E}[\nu\nu^\top] \succ 0$.[3] $*$

Remark 1. The assumption on the covariance matrices of x and ν is without loss of generality. In fact, if $\mathbb{E}[xx^\top]$ were singular, then there would exist a linear combination of components of x that is zero with probability 1. Therefore, by a suitable change of variables, the estimation problem (4) could be rewritten as an estimation problem where one or more state components are zero, and eliminating these zero components would lead to an equivalent estimation problem with nonsingular state covariance matrix. Similarly, if $\mathbb{E}[\nu\nu^\top]$ were singular, one or more linear combinations of components of x would be perfectly known from y, and therefore these linear combinations could be eliminated from the estimation problem yielding a reduced order problem with a nonsingular noise covariance matrix. $*$

A linear estimator $\hat{x}$ of x is an n-dimensional random vector whose components are linear combinations of the components of y. The unconstrained minimum variance linear estimation problem (or Bayesian linear estimation problem) is formulated as follows:

$$\begin{aligned} \mathscr{P}_{\mathcal{U}}: \quad & \text{Find the linear estimator } \hat{x}^{\mathcal{U}} \text{ that} \\ & \text{minimizes } \mathbb{E}[(x-\hat{x})(x-\hat{x})^\top]. \end{aligned}$$

In $\mathscr{P}_{\mathcal{U}}$, "minimizes" is in the sense of the Loewner partial ordering, i.e., a matrix M is "smaller" than a matrix N ($M \preceq N$) if and only if $N - M$ is positive semidefinite. The solution to this problem can be found by considering x and y as vectors whose components are in the Hilbert space of random variables with finite second order moments, and by projecting the components of x onto the linear span $\mathbb{Y}$ of the components of y, see e.g. [14], [18]. We write this component-wise projection as $\hat{x}^{\mathcal{U}} = \mathcal{P}_{\mathbb{Y}}(x)$.[4] The existence and uniqueness of the projection is guaranteed by the Hilbert projection theorem (see e.g. [14], Chapter 3, Section 3), and it holds that

$$\mathbb{E}[(x - \hat{x}^{\mathcal{U}})y^\top] = 0. \tag{5}$$

The fact that $\hat{x}^{\mathcal{U}}$ is a solution of $\mathscr{P}_{\mathcal{U}}$, and indeed the only solution, is easily shown as follows. Consider any other candidate linear estimator $\hat{x}$, $\hat{x} \neq \hat{x}^{\mathcal{U}}$, and note that $\mathbb{E}[(x-\hat{x})(x-\hat{x})^\top] = \mathbb{E}[(x-\hat{x}^{\mathcal{U}}+\hat{x}^{\mathcal{U}}-\hat{x})(x-\hat{x}^{\mathcal{U}}+\hat{x}^{\mathcal{U}}-\hat{x})^\top] = \mathbb{E}[(x - \hat{x}^{\mathcal{U}})(x - \hat{x}^{\mathcal{U}})^\top] + \mathbb{E}[(\hat{x}^{\mathcal{U}} - \hat{x})(\hat{x}^{\mathcal{U}} - \hat{x})^\top] + \mathbb{E}[(x - \hat{x}^{\mathcal{U}})(\hat{x}^{\mathcal{U}} - \hat{x})^\top] + \mathbb{E}[(\hat{x}^{\mathcal{U}} - \hat{x})(x - \hat{x}^{\mathcal{U}})^\top] = \mathbb{E}[(x - \hat{x}^{\mathcal{U}})(x - \hat{x}^{\mathcal{U}})^\top] + \mathbb{E}[(\hat{x}^{\mathcal{U}} - \hat{x})(\hat{x}^{\mathcal{U}} - \hat{x})^\top]$, where two terms have been canceled in the last equality because $\hat{x}^{\mathcal{U}}$ and $\hat{x}$ are linearly obtained from y and the orthogonality condition (5) holds. Since $\mathbb{E}[(\hat{x}^{\mathcal{U}} - \hat{x})(\hat{x}^{\mathcal{U}} - \hat{x})^\top] \succeq 0$ and $\hat{x}^{\mathcal{U}} - \hat{x} \neq 0$, this proves that $\hat{x}^{\mathcal{U}}$ is the unique solution of $\mathscr{P}_{\mathcal{U}}$.

Denote by $\mathbb{X}$ the linear span of the components of x. The projection of $\hat{x}^{\mathcal{U}}$ onto $\mathbb{X}$, denoted by $\mathcal{P}_{\mathbb{X}}(\hat{x}^{\mathcal{U}})$, can be written as $Z^{\mathcal{U}}x$, where $Z^{\mathcal{U}}$ is a $n \times n$ matrix. Under the assumption $\mathbb{E}[xx^\top] \succ 0$, the representation $Z^{\mathcal{U}}x$ is unique. The following fact is also true.

Fact 1. $Z^{\mathcal{U}}$ is nonsingular if and only if $\text{rank}(A) = n$. $*$

Proof. We start by noting that for any $h \in \mathbb{R}^n$, $h \neq 0$, the following equivalence holds:

$$\mathcal{P}_{\mathbb{X}}(\mathcal{P}_{\mathbb{Y}}(h^\top x)) = 0 \iff \mathcal{P}_{\mathbb{Y}}(h^\top x) = 0. \tag{6}$$

The implication $\mathcal{P}_{\mathbb{Y}}(h^\top x) = 0 \Rightarrow \mathcal{P}_{\mathbb{X}}(\mathcal{P}_{\mathbb{Y}}(h^\top x)) = 0$ is trivial. The opposite implication can be proved by decomposing $h^\top x$ as $h^\top x = \beta^\top y + w$, where $\beta^\top y = \mathcal{P}_{\mathbb{Y}}(h^\top x)$ and w is orthogonal to y, and by noting that[5]

$$\begin{aligned} 0 = \mathcal{P}_{\mathbb{X}}(\mathcal{P}_{\mathbb{Y}}(h^\top x)) = \mathcal{P}_{\mathbb{X}}(\mathcal{P}_{\mathbb{Y}}(\beta^\top y + w)) = \mathcal{P}_{\mathbb{X}}(\beta^\top y) \\ \Rightarrow \beta^\top y \perp \mathbb{X} \Rightarrow \beta^\top y \perp h^\top x, \end{aligned} \tag{7}$$

from which

$$\begin{aligned} \mathbb{E}[(\beta^\top y)^2] + \mathbb{E}[w^2] = \mathbb{E}[(h^\top x)(h^\top x)] = \mathbb{E}[(\beta^\top y + w)(h^\top x)] \\ \overset{\text{use (7)}}{=} \mathbb{E}[w(h^\top x)] = \mathbb{E}[w(\beta^\top y + w)] \\ = \mathbb{E}[w^2], \end{aligned}$$

which implies that $\mathbb{E}[(\beta^\top y)^2] = 0$, so that $\mathcal{P}_{\mathbb{Y}}(h^\top x) = \beta^\top y = 0$.

Using (6), Fact 1 is obtained as follows

$$\begin{aligned} Z^{\mathcal{U}} \text{ singular} &\iff \exists h \in \mathbb{R}^n, h \neq 0 : h^\top Z^{\mathcal{U}} = 0 \\ &\overset{\mathbb{E}[xx^\top] \succ 0}{\iff} h^\top Z^{\mathcal{U}} \mathbb{E}[xx^\top] Z^{\mathcal{U}} h = 0 \\ &\iff 0 = h^\top Z^{\mathcal{U}} x = h^\top \mathcal{P}_{\mathbb{X}}(\mathcal{P}_{\mathbb{Y}}(x)) \\ &\iff \mathcal{P}_{\mathbb{X}}(\mathcal{P}_{\mathbb{Y}}(h^\top x)) = 0 \overset{(6)}{\iff} \mathcal{P}_{\mathbb{Y}}(h^\top x) = 0 \\ &\iff y \perp h^\top x \iff 0 = \mathbb{E}[y \cdot x^\top h] = A\mathbb{E}[xx^\top]h \\ &\overset{E[xx^\top] \succ 0}{\iff} \text{rank}(A) < n. \end{aligned}$$

□

When A has not rank n, y conveys no information on some linear combinations of the components of x, which therefore cannot be estimated from y. In this case, these linear combinations can be discarded and the estimation problem can be reformulated as a lower dimensional problem where the

[3] The symbol "$\succ 0$" means positive definite, while "$\succeq 0$" is used for positive semidefinite.

[4] $\hat{x}^{\mathcal{U}}$ is uniquely defined up to zero probability sets. Here and throughout, we do not distinguish between random variables that differ on zero probability sets.

[5] Throughout the paper, the symbol $\perp$ refers to orthogonality in the Hilbert space of random variables with finite second order moments, that is, $\xi \perp \eta$ means $\mathbb{E}[\xi\eta] = 0$. When applied to vectors, symbol $\perp$ means that each component of one vector is orthogonal to all components of the other vector. $\xi \perp S$, where ξ is a random variable and S is a set of random variables, means that ξ is orthogonal to all random variables in S.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3103905, IEEE Transactions on Automatic Control

rank condition is satisfied. Thus, in what follows we assume that the following rank condition on A holds true.

Condition 2 (rank condition)**.** $\text{rank}(A) = n$. ∗

Given a nonsingular $n \times n$ matrix Z, the constrained minimum variance linear estimation problem is formulated as follows:

$\mathscr{P}_{\mathcal{C}}(Z)$: Find the linear estimator $\hat{x}^{\mathcal{C}}_Z$ that minimizes $\mathbb{E}[(x-\hat{x})(x-\hat{x})^\top]$ subject to the constraint $\mathcal{P}_{\mathbb{X}}(\hat{x}) = Zx$.

Under Condition 2, in view of Fact 1, $Z^{\mathcal{U}}$ is nonsingular. Since $\hat{x}^{\mathcal{U}}$ is the unique solution to the unconstrained problem $\mathscr{P}_{\mathcal{U}}$, and it also satisfies the constraint in $\mathscr{P}_{\mathcal{C}}(Z^{\mathcal{U}})$, $\hat{x}^{\mathcal{U}}$ is also the unique solution to $\mathscr{P}_{\mathcal{C}}(Z^{\mathcal{U}})$. Note that any vector[6] $\hat{x} \in \mathbb{Y}$ can be written as $\hat{x} = Ty$, where T is an $n \times p$ matrix; since $\mathbb{E}[yy^\top] \succ 0$, matrix T is unique. The next fact reformulates the constraint in $\mathscr{P}_{\mathcal{C}}(Z)$.

Fact 2. The constraint $\mathcal{P}_{\mathbb{X}}(\hat{x}) = Zx$ is equivalent to $TA = Z$. ∗

Proof. It holds that $\mathcal{P}_{\mathbb{X}}(\hat{x}) = \mathcal{P}_{\mathbb{X}}(Ty) = \mathcal{P}_{\mathbb{X}}(TAx) + \mathcal{P}_{\mathbb{X}}(T\nu) = TAx$, and $TAx = Zx \overset{E[xx^\top]\succ 0}{\iff} TA = Z$. □

We can further reformulate the objective function in $\mathscr{P}_{\mathcal{C}}(Z)$ as

$$\begin{aligned}
&\mathbb{E}[(x-\hat{x})(x-\hat{x})^\top] = \mathbb{E}[(x-Ty)(x-Ty)^\top] \\
&= \mathbb{E}[(x-TAx-T\nu)(x-TAx-T\nu)^\top] \\
&= [\text{using Fact 2}] = \mathbb{E}[(x-Zx-T\nu)(x-Zx-T\nu)^\top] \\
&= \mathbb{E}[(x-Zx)(x-Zx)^\top] + TE[\nu\nu^\top]T^\top, \qquad (8)
\end{aligned}$$

where the first term in the last expression does not depend on T, so that minimization only refers to the second term. Thus, the constrained problem $\mathscr{P}_{\mathcal{C}}(Z)$ is equivalent to the following problem $\mathscr{P}'_{\mathcal{C}}(Z)$:

$\mathscr{P}'_{\mathcal{C}}(Z)$: Find the $n \times p$ matrix T_Z that minimizes $T\mathbb{E}[\nu\nu^\top]T^\top$ subject to the constraint $TA = Z$.

This formulation is key to the following fact.

Fact 3. If $\hat{x}^{\mathcal{C}}_Z = T_Z y$ is a solution to $\mathscr{P}_{\mathcal{C}}(Z)$, then $\hat{x}^{\mathcal{C}}_{\tilde{Z}} = \tilde{Z}Z^{-1}T_Z y$ is a solution to $\mathscr{P}_{\mathcal{C}}(\tilde{Z})$. ∗

Proof. First note that $(\tilde{Z}Z^{-1}T_Z)A = (\tilde{Z}Z^{-1})(T_Z A) = \tilde{Z}Z^{-1}Z = \tilde{Z}$, so that $\tilde{Z}Z^{-1}T_Z$ satisfies the constraint in $\mathscr{P}'_{\mathcal{C}}(\tilde{Z})$. Next, consider any $\tilde{T}$ such that $\tilde{T}A = \tilde{Z}$. Then, $(Z\tilde{Z}^{-1}\tilde{T})A = Z$, so that $Z\tilde{Z}^{-1}\tilde{T}$ is a competitor of the solution T_Z to problem $\mathscr{P}'_{\mathcal{C}}(Z)$ and, hence, it holds that $(Z\tilde{Z}^{-1}\tilde{T})\mathbb{E}[\nu\nu^\top](Z\tilde{Z}^{-1}\tilde{T})^\top \succeq T_Z\mathbb{E}[\nu\nu^\top]T_Z^\top$. Hence, $\tilde{T}\mathbb{E}[\nu\nu^\top]\tilde{T}^\top = (\tilde{Z}Z^{-1})(Z\tilde{Z}^{-1}\tilde{T})\mathbb{E}[\nu\nu^\top](Z\tilde{Z}^{-1}\tilde{T})^\top(\tilde{Z}Z^{-1})^\top \succeq (\tilde{Z}Z^{-1}T_Z)\mathbb{E}[\nu\nu^\top](\tilde{Z}Z^{-1}T_Z)^\top$, showing that $\hat{x}^{\mathcal{C}}_{\tilde{Z}}$ is a solution to $\mathscr{P}_{\mathcal{C}}(\tilde{Z})$. □

[6]When applied to a vector of random variables, symbol ∈ means that each component of the vector is an element of the set that follows the ∈ symbol.

In view of Fact 3, if a unique solution to a given problem $\mathscr{P}_{\mathcal{C}}(Z)$ exists, then any other problem $\mathscr{P}_{\mathcal{C}}(\tilde{Z})$ admits a unique solution. Since $\hat{x}^{\mathcal{U}}$ is the unique solution to $\mathscr{P}_{\mathcal{C}}(Z^{\mathcal{U}})$, we conclude that every problem $\mathscr{P}_{\mathcal{C}}(Z)$ has a unique solution $\hat{x}^{\mathcal{C}}_Z = T_Z y$. Writing $\hat{x}^{\mathcal{U}} = Uy$ as a shorthand for $\hat{x}^{\mathcal{U}} = T_{Z^{\mathcal{U}}}y$, T_Z can be computed according to relation

$$T_Z = Z(Z^{\mathcal{U}})^{-1}U. \qquad (9)$$

Remark 2. The $n \times p$ matrix U compresses the vector y of measurements from dimension p to dimension n. Relation (9) shows that this compression preserves the information that is necessary to compute the optimal estimator $\hat{x}^{\mathcal{C}}_Z$. ∗

Remark 3. The matrix U depends on the state covariance matrix $\mathbb{E}[xx^\top]$ and on the noise covariance matrix $\mathbb{E}[\nu\nu^\top]$. However, matrix $\mathbb{E}[xx^\top]$ does not appear in the statement of $\mathscr{P}'_{\mathcal{C}}(Z)$ so that the solution T_Z to problem $\mathscr{P}'_{\mathcal{C}}(Z)$ does not depend on $\mathbb{E}[xx^\top]$. From this we see that the dependence on $\mathbb{E}[xx^\top]$ in U is canceled out in (9) when U is multiplied by $(Z^{\mathcal{U}})^{-1}$. ∗

Remark 4 (Geometric interpretation)**.** Figure 6 visualizes the estimators $\hat{x}^{\mathcal{U}}$ and $\hat{x}^{\mathcal{C}}_Z$ when x is a scalar random variable. $\hat{x}^{\mathcal{U}}$ is the projection of x onto $\mathbb{Y}$. The affine subspace $\mathcal{K}$, which is introduced here for the first time, is the intersection of $\mathbb{Y}$ with the affine subspace passing through Zx and orthogonal to $\mathbb{X}$. This affine subspace is the set of random variables whose projection onto $\mathbb{X}$ equals Zx, so that these random variables satisfy the constraint $\mathcal{P}_{\mathbb{X}}(\hat{x}) = Zx$ of problem $\mathscr{P}_{\mathcal{C}}(Z)$. Estimator $\hat{x}^{\mathcal{C}}_Z$ is an element of $\mathcal{K}$. The figure describes that the random variable in $\mathcal{K}$ that minimizes the distance from x, i.e., $\hat{x}^{\mathcal{C}}_Z$, is obtained as a rescaling of $\hat{x}^{\mathcal{U}}$. ∗

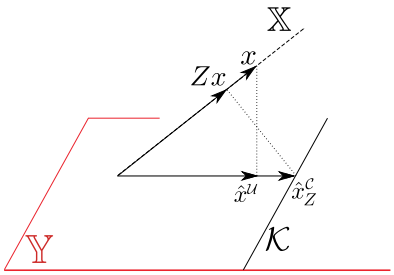


Figure 6: Geometric representation of the relation between $\hat{x}^{\mathcal{U}}$ and $\hat{x}^{\mathcal{C}}_Z$.

We next consider the special case where $Z = I$, the identity matrix.

If $Z = I$, the optimal constrained estimator is $\hat{x}^{\mathcal{C}}_I = T_I y$ where, in view of (9),

$$T_I = (Z^{\mathcal{U}})^{-1}U.$$

From relation $UA = Z^{\mathcal{U}}$ (which is obtained by an application of Fact 2), T_I can be rewritten explicitly as a function of U as follows

$$T_I = (UA)^{-1}U, \qquad (10)$$

so that the solution to $\mathscr{P}_{\mathcal{C}}(I)$ is

$$\hat{x}_I^{\mathcal{C}} = (UA)^{-1}\hat{x}^{\mathcal{U}}. \tag{11}$$

It is an interesting fact that the same matrix $(UA)^{-1}$ also relates the variance of the estimation error of $\hat{x}_I^{\mathcal{C}}$ to that of $\hat{x}^{\mathcal{U}}$, as formalized in next Fact 4.

Fact 4.

$$\mathbb{E}[(x-\hat{x}_I^{\mathcal{C}})(x-\hat{x}_I^{\mathcal{C}})^\top] = (UA)^{-1}\mathbb{E}[(x-\hat{x}^{\mathcal{U}})(x-\hat{x}^{\mathcal{U}})^\top]. \tag{12}$$

$*$

Proof. First recall that $\mathbb{E}[(x-Uy)y^\top] = 0$ (refer to (5)) and that $\mathbb{E}[x\nu^\top] = 0$ (Condition 1), so that $\mathbb{E}[xx^\top]A^\top = \mathbb{E}[xy^\top] = \mathbb{E}[Uyy^\top] = (UA)\mathbb{E}[xx^\top]A^\top + U\mathbb{E}[\nu\nu^\top]$. Thus, $\mathbb{E}[xx^\top](A^\top U^\top) = (UA)\mathbb{E}[xx^\top](A^\top U^\top) + U\mathbb{E}[\nu\nu^\top]U^\top$, which gives

$$E[xx^\top] - (UA)\mathbb{E}[xx^\top] = U\mathbb{E}[\nu\nu^\top]U^\top(A^\top U^\top)^{-1}. \tag{13}$$

On the other hand,

$$\begin{aligned}\mathbb{E}[(x-\hat{x}^{\mathcal{U}})(x-\hat{x}^{\mathcal{U}})^\top] &= \mathbb{E}[(x-Uy)x^\top] \\ &= \mathbb{E}[xx^\top] - (UA)\mathbb{E}[xx^\top],\end{aligned} \tag{14}$$

and, substituting (13) one has $\mathbb{E}[(x-\hat{x}^{\mathcal{U}})(x-\hat{x}^{\mathcal{U}})^\top] = U\mathbb{E}[\nu\nu^\top]U^\top(A^\top U^\top)^{-1} =$ [using (10)] $= (UA)T_I\mathbb{E}[\nu\nu^\top]T_I^\top =$ [using (8) with $Z=I$] $= (UA)\mathbb{E}[(\hat{x}_I^{\mathcal{C}}-x)(\hat{x}_I^{\mathcal{C}}-x)^\top]$. □

Remark 5 (Unbiasedness)**.** When x and ν are independent, the constraint $T_IA = I$ in $\mathscr{P}'_{\mathcal{C}}(I)$ can be interpreted as a conditional unbiasedness constraint; in fact, $\mathbb{E}[\hat{x}_I^{\mathcal{C}}|x] = \mathbb{E}[T_IAx + T_I\nu|x] = T_IAx = x$ (instead $\hat{x}^{\mathcal{U}} = (UA)\hat{x}_I^{\mathcal{C}}$ – see equation (11) – is conditionally biased). $*$

Remark 6 (Related bibliography)**.** When x is a deterministic vector, the function T_Iy that has been derived in this section formally coincides with the Fisher estimator [15], [17], or the Gauss-Markov Minimum-Variance Unbiased estimator [14], or the Generalized Least Squares estimator and the Best Linear Unbiased Estimator (BLUE) [19] of x. When ν is Gaussian, $\hat{x}_I^{\mathcal{C}}$ coincides with the Maximum Likelihood estimator of x, see e.g. [15].[7]

The constrained estimation problem $\mathscr{P}_{\mathcal{C}}(Z)$ can also be interpreted as an oblique projection problem: the solution $\hat{x}_Z^{\mathcal{C}}$ is the projection of Zx onto $\mathbb{Y}$ along the space orthogonal to $\mathbb{X}$, see e.g. [24-27]. $*$

B. Construction of the minimum volume region

In this section, the following additional condition is in use.

Condition 3. x and ν are jointly Gaussian random vectors. $*$

Using the mathematical relations derived in the previous section, we construct a minimum volume region $\mathcal{X}$ for x that satisfies a static version of the state conditional property, i.e.,

$$\mathbb{P}r\{x \in \mathcal{X}|x=\bar{x}\} = \alpha,\ \bar{x}\text{-a.e.}, \tag{15}$$

[7]In the dynamical case, estimators akin to $\hat{x}_I^{\mathcal{C}}$ have been considered in [20-22] in the context of solving smoothing problems. On the other hand, the "recursive BLUE estimator" of [23] and the Fisher-type estimators of [15], Section 6.4, refer to different methods.

and such that $\mathcal{X}$ is constructed from an n-dimensional linear compression of y. This requirement allows for a finite-memory implementation when the construction is applied to the dynamical case.

Throughout, we shall denote with $\mathcal{E}_\Sigma^\alpha(\mu)$ the minimum volume ellipsoid in $\mathbb{R}^n$ that contains an α-probability mass of a Gaussian distribution $\mathcal{G}(\mu,\Sigma)$, with mean μ and covariance matrix Σ. The ellipsoid $\mathcal{E}_\Sigma^\alpha(\mu)$ can be expressed as

$$\mathcal{E}_\Sigma^\alpha(\mu) = \left\{x \in \mathbb{R}^n : (x-\mu)^\top\Sigma^{-1}(x-\mu) \le \chi^2(\alpha,n)\right\}, \tag{16}$$

where $\chi^2(\alpha,n)$ is the quantile at probability α of the Chi-square distribution with n degrees of freedom.

Let $\Pi := T_I\mathbb{E}[\nu\nu^\top]T_I^\top$, and define

$$\mathcal{X} := \mathcal{E}_\Pi^\alpha(\hat{x}_I^{\mathcal{C}}). \tag{17}$$

$\mathcal{X}$ is an ellipsoid whose center depends on the measurements, while its shape, and hence its volume, is measurement-independent. The center is obtained by a linear p-dimensional compression of y (see Remark 2) so that $\mathcal{X}$ can formally be written as $\mathcal{X} = \mathcal{S}(\mathrm{C}y)$, where $\mathrm{C} \in \mathbb{R}^{n\times p}$ is a matrix that defines a compression scheme from $y \in \mathbb{R}^p$ to $\mathrm{C}y \in \mathbb{R}^n$ (in the case of $\mathcal{X}$ one has $\mathrm{C} = U$), and $\mathcal{S}$ is a map from $z \in \mathbb{R}^n$ to a subset of $\mathbb{R}^n$. Theorem 1 below shows that $\mathcal{X}$ in (17) is minimum volume with respect to any other construction of the same form. The proof is carried out under the following technical measurability condition that any candidate map $\mathcal{S}$ is required to satisfy.

Condition 4. The set $\mathcal{S}_{XZ} = \{(x,z) \in \mathbb{R}^n \times \mathbb{R}^n : x \in \mathcal{S}(z)\}$ is Lebesgue-measurable. $*$

Theorem 1. *The region $\mathcal{X} := \mathcal{E}_\Pi^\alpha(\hat{x}_I^{\mathcal{C}})$ satisfies the conditional property* (15) *and, for any couple $(\mathrm{C},\mathcal{S})$ such that the region $\mathcal{S}(\mathrm{C}y)$ satisfies the conditional property*

$$\mathbb{P}r\{x \in \mathcal{S}(\mathrm{C}y)|x=\bar{x}\} = \alpha,\ \bar{x}\text{-a.e.}, \tag{18}$$

it holds that

$$J(\mathcal{S}(\mathrm{C}y)) \ge J(\mathcal{X}), \tag{19}$$

where $J(\cdot) := \sup_y \mathrm{Vol}(\cdot)$. $*$

Proof. First, we prove that $\mathcal{X} = \mathcal{E}_\Pi^\alpha(\hat{x}_I^{\mathcal{C}})$ satisfies the conditional property. Recall that $\hat{x}_I^{\mathcal{C}} = T_Iy$. We have, $x - \hat{x}_I^{\mathcal{C}} = x - T_Iy = x - T_IAx - T_I\nu = -T_I\nu$. Hence, $x-\hat{x}_I^{\mathcal{C}}$ distributes like $-T_I\nu$, a Gaussian variable with zero mean and variance Π (recall the definition of Π given just before equation (17)). From definition (16) it follows that $\mathbb{P}r\{\bar{x} \in \mathcal{E}_\Pi^\alpha(\hat{x}_I^{\mathcal{C}})|x=\bar{x}\} = \alpha$, $\bar{x}$-a.e., which is the conditional property.

The rest of the proof is by contradiction, and the contradiction will be obtained by assuming that (19) is false for a certain couple $(\mathcal{S},\mathrm{C})$. Under this (false) assumption, the volume of the set $\mathcal{S}_{XZ} = \{(x,z) \in \mathbb{R}^n \times \mathbb{R}^n : x \in \mathcal{S}(z)\}$ will be computed in two different ways that lead to different results. Assume first that $\mathrm{C}A$ is invertible (the case where $\mathrm{C}A$ is singular will be considered later). If $\mathrm{C}A$ is invertible, then we can assume without loss of generality that $\mathrm{C}A = I$ because the couple $(\mathrm{C}',\mathcal{S}')$ with $\mathrm{C}'A$ invertible but non-identity gives the same region as the couple $(\mathrm{C},\mathcal{S})$ where $\mathrm{C} := (\mathrm{C}'A)^{-1}\mathrm{C}'$ and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3103905, IEEE Transactions on Automatic Control

$\mathcal{S}(z) := \mathcal{S}'((\mathrm{C}'A)z)$. To obtain a contradiction, we assume that $\sup_z \mathrm{Vol}(\mathcal{S}(z)) = \mathrm{Vol}(\mathcal{X}) - \epsilon$, $\epsilon > 0$, and compute the integral of $\mathbb{1}(\mathcal{S}_{XZ})$ over subsets of $\mathbb{R}^n \times \mathbb{R}^n$ of increasing size. Precisely, for $r > 0$, define $\mathcal{B}_r = \{z \in \mathbb{R}^n : \|z\| \leq r\}$ and find a value $\varrho > 0$ such that

$$\int_{\mathbb{R}^n} \mathbb{1}(\mathcal{B}_\varrho) \mathrm{d}\mathcal{G}(0, \mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top) = 1 - \delta,$$

where δ is such that $\mathrm{Vol}(\mathcal{E}^{\alpha-\delta}_{\mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top}(0)) > \mathrm{Vol}(\mathcal{E}^{\alpha}_{\mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top}(0)) - \frac{\epsilon}{2}$. In words, the probability mass δ that is outside $\mathcal{B}_\varrho$ is small enough to preserve the value $\mathrm{Vol}(\mathcal{E}^{\alpha-\delta}_{\mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top})$ "close enough" to $\mathrm{Vol}(\mathcal{E}^{\alpha}_{\mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top})$. Now, we compute the integral

$$\int_{\mathcal{B}_r \times \mathcal{B}_{r+\varrho}} \mathbb{1}(\mathcal{S}_{XZ}) \mathrm{d}x \times \mathrm{d}z.$$

We have

$$\begin{aligned}\int_{\mathcal{B}_r \times \mathcal{B}_{r+\varrho}} \mathbb{1}(\mathcal{S}_{XZ}) \mathrm{d}x \times \mathrm{d}z &= [\text{Tonelli's theorem}] \\ &= \int_{\mathcal{B}_{r+\varrho}} \mathrm{Vol}(\mathcal{S}(z) \cap \mathcal{B}_r) \mathrm{d}z \\ &\leq \int_{\mathcal{B}_{r+\varrho}} \mathrm{Vol}(\mathcal{S}(z)) \mathrm{d}z \\ &\leq (\mathrm{Vol}(\mathcal{X}) - \epsilon) \mathrm{Vol}(\mathcal{B}_{r+\varrho}), \quad (20)\end{aligned}$$

where the false assumption has been used in the last step. By reversing the order of integration, we also have $\int_{\mathcal{B}_r \times \mathcal{B}_{r+\varrho}} \mathbb{1}(\mathcal{S}_{XZ}) \mathrm{d}x \times \mathrm{d}z = \int_{\mathcal{B}_r} \mathrm{Vol}(\mathcal{S}_x \cap \mathcal{B}_{r+\varrho}) \mathrm{d}x$, where $\mathcal{S}_x = \{z \in \mathbb{R}^n : x \in \mathcal{S}(z)\}$. Equation (18) can be written as $\mathbb{P}r\{\mathrm{C}y \in \mathcal{S}_x | x = \bar{x}\} = \alpha$, $\bar{x}$-a.e.; this means that for almost every value of x, $\mathcal{S}_x$ is an α-probability set with respect to $\mathcal{G}(x, \mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top)$, which is the conditional distribution of $\mathrm{C}y$ given x. From the definition of ϱ, we also have that for every $x \in \mathcal{B}_r$ it holds that $\int_{\mathbb{R}^n} \mathbb{1}(\mathcal{B}_{r+\varrho}) \mathrm{d}\mathcal{G}(x, \mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top) \geq 1 - \delta$; this means that $\mathcal{B}_{r+\varrho}$ is at least a $(1-\delta)$-probability set with respect to the measure $\mathcal{G}(x, \mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top)$ for every $x \in \mathcal{B}_r$. Hence, the set $\mathcal{S}_x \cap \mathcal{B}_{r+\varrho}$ is at least an $(\alpha - \delta)$-probability set with respect to $\mathcal{G}(x, \mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top)$, for almost every $x \in \mathcal{B}_r$.

Since $\mathcal{E}^{\alpha-\delta}_{\mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top}(x)$ is the minimum volume set containing a probability $\alpha - \delta$ with respect to the measure $\mathcal{G}(x, \mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top)$, for almost any $x \in \mathcal{B}_r$ it holds that $\mathrm{Vol}(\mathcal{S}_x \cap \mathcal{B}_{r+\varrho}) \geq \mathrm{Vol}(\mathcal{E}^{\alpha-\delta}_{\mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top}(x)) = \mathrm{Vol}(\mathcal{E}^{\alpha-\delta}_{\mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top}(0)) > \mathrm{Vol}(\mathcal{E}^{\alpha}_{\mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top}(0)) - \frac{\epsilon}{2}$, where the last inequality holds true by definition of δ. Finally, T_I is by definition the solution to problem $\mathscr{P}'_{\mathcal{C}}(I)$ in Section III-A, i.e., T_I is the minimizer of $TE[\nu\nu^\top]T^\top$ w.r.t. any T such that $TA = I$. Therefore, it holds that $\mathrm{Vol}(\mathcal{E}^{\alpha}_{\mathrm{C}E[\nu\nu^\top]\mathrm{C}^\top}(0)) \geq \mathrm{Vol}(\mathcal{E}^{\alpha}_{T_I E[\nu\nu^\top]T_I^\top}(0)) = \mathrm{Vol}(\mathcal{X})$. Putting all together, we obtain

$$\int_{\mathcal{B}_r \times \mathcal{B}_{r+\varrho}} \mathbb{1}(\mathcal{S}_{XZ}) \mathrm{d}x \times \mathrm{d}z > \left(\mathrm{Vol}(\mathcal{X}) - \frac{\epsilon}{2}\right) \mathrm{Vol}(\mathcal{B}_r). \quad (21)$$

Using (21) and (20) we now obtain

$$\left(\mathrm{Vol}(\mathcal{X}) - \frac{\epsilon}{2}\right) \mathrm{Vol}(\mathcal{B}_r) < (\mathrm{Vol}(\mathcal{X}) - \epsilon) \mathrm{Vol}(\mathcal{B}_{r+\varrho}),$$

which leads to a contradiction when we take the ratio between the right-hand side and the left-hand side and let $r \to \infty$:

$$1 \leq \lim_{r \to \infty} \frac{\mathrm{Vol}(\mathcal{X}) - \epsilon}{\mathrm{Vol}(\mathcal{X}) - \epsilon/2} \frac{\mathrm{Vol}(\mathcal{B}_{r+\varrho})}{\mathrm{Vol}(\mathcal{B}_r)} = \frac{\mathrm{Vol}(\mathcal{X}) - \epsilon}{\mathrm{Vol}(\mathcal{X}) - \epsilon/2} < 1.$$

Suppose now that $\mathrm{C}A$ is singular. We show that necessarily $\sup_y \mathrm{Vol}(\mathcal{S}(\mathrm{C}y)) = \infty$.

Consider first the easiest case where $\mathrm{C}A$ is the zero matrix, so that $z = \mathrm{C}y$ does not depend on x. In this case, $\mathrm{C}y = \mathrm{C}\nu$, and, since the random vectors x and ν are independent, the conditional probability distribution of $\mathrm{C}y$ given x is $\mathcal{G}(0, \mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top)$ independently of x. Thus, the constraint (18) implies that $\int_z \mathbb{1}(\mathcal{S}_{XZ}) \mathrm{d}\mathcal{G}(0, \mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top) = \alpha$, x-a.e. Next, let λ_x be the Lebesgue measure on $\mathbb{R}^n$, and introduce the product measure $\mu = \lambda_x \times \mathcal{G}(0, \mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top)$ on $\mathbb{R}^n \times \mathbb{R}^n$. We have $\int_{(x,z) \in \mathbb{R}^n \times \mathbb{R}^n} \mathbb{1}(\mathcal{S}_{XZ}) \mathrm{d}\mu = \int_x \left(\int_z \mathbb{1}(\mathcal{S}_{XZ}) \mathrm{d}\mathcal{G}(0, \mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top)\right) \mathrm{d}x = \alpha \cdot \int_x \mathrm{d}x = \infty$. On the other hand, assuming $\sup_y \mathrm{Vol}(\mathcal{S}(\mathrm{C}y)) = K < \infty$, we have $\int_{(x,z) \in \mathbb{R}^n \times \mathbb{R}^n} \mathbb{1}(\mathcal{S}_{XZ}) \mathrm{d}\mu = \int_z \left(\int_x \mathbb{1}(\mathcal{S}_{XZ}) \mathrm{d}x\right) \mathrm{d}\mathcal{G}(0, \mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top) \leq K \cdot \int_z \mathrm{d}\mathcal{G}(0, \mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top) = K < \infty$, which is a contradiction.

When $\mathrm{C}A$ is singular but nonzero, we decompose every value of $x \in \mathbb{R}^n$ into two orthogonal[8] components $x_k \in \ker(\mathrm{C}A)$ ($\ker(\mathrm{C}A)$ is short for the kernel of $\mathrm{C}A$) and $x_p \in \ker^\perp(\mathrm{C}A)$, such that $x = x_k + x_p$, and $\mathrm{C}Ax = \mathrm{C}Ax_p$. Introduce the product measure $\lambda_{x_k} \times \mathcal{G}(0, I) \times \mathcal{G}(0, \mathbb{E}[\nu\nu^\top])$ over the space of the triplets $(x_k, x_p, \nu) \in \ker(\mathrm{C}A) \times \ker^\perp(\mathrm{C}A) \times \mathbb{R}^p$ and let μ be the corresponding image measure over $(x, z) \in \mathbb{R}^n \times \mathbb{R}^n$ that is induced by the linear relations $x = x_k + x_p$ and $z = \mathrm{C}Ax_p + \mathrm{C}\nu$. We have

$$\begin{aligned}&\int_{(x,z) \in \mathbb{R}^n \times \mathbb{R}^n} \mathbb{1}(\{(x,z) \in \mathcal{S}_{XZ}\}) \mathrm{d}\mu \\ &\quad = \int_{(x_k, x_p, \nu) \in \ker(\mathrm{C}A) \times \ker^\perp(\mathrm{C}A) \times \mathbb{R}^p} \mathbb{1}(\{(x_k + x_p, \mathrm{C}Ax_p + \mathrm{C}\nu) \in \mathcal{S}_{XZ}\}) \mathrm{d}x_k \times \\ &\qquad \mathrm{d}\mathcal{G}(0, I) \times \mathrm{d}\mathcal{G}(0, \mathbb{E}[\nu\nu^\top]). \quad (22)\end{aligned}$$

To compute (22), we first integrate with respect to ν and then with respect to (x_k, x_p). We get $\int_{(x,z) \in \mathbb{R}^n \times \mathbb{R}^n} \mathbb{1}(\mathcal{S}_{XZ}) \mathrm{d}\mu = \int_{(x_k, x_p)} \mathbb{P}r\{\mathrm{C}y \in \mathcal{S}(x) | x = x_k + x_p\} \mathrm{d}x_k \times \mathrm{d}\mathcal{G}(0, I) = \alpha \int_{x_k} \mathrm{d}x_k = \infty$. Second, we integrate (22) first with respect to x and then with respect to z. Note that x_k and z are independent with respect to the integration measure, while x_p and z are not. So, in order to first integrate with respect to $x (= x_k + x_p)$ given $z (= \mathrm{C}Ax_p + \mathrm{C}\nu)$, we have to consider the conditional measure of x_p given z. This measure is Gaussian and does not concentrate on any subset of $\ker^\perp(\mathrm{C}A)$, due to the presence of the term $\mathrm{C}\nu$ and the assumption $\mathbb{E}[\nu\nu^\top] \succ 0$, and therefore has bounded density over $\ker^\perp(\mathrm{C}A)$. Moreover, its density can be uniformly bounded with respect to z. By denoting with $B > 0$ an essential upper bound to the conditional density function of x_p given any z, we obtain $\int_{(x,z) \in \mathbb{R}^n \times \mathbb{R}^n} \mathbb{1}(\mathcal{S}_{XZ}) \mathrm{d}\mu < \int_z \left(\int_{x_k} \int_{x_p} \mathbb{1}(\mathcal{S}_{XZ}) B \mathrm{d}x_p \mathrm{d}x_k\right) \mathrm{d}\mathcal{G}(0, \mathrm{C}\mathbb{E}[\nu\nu^\top]\mathrm{C}^\top) \leq B \cdot \sup_z \mathrm{Vol}(\mathcal{S}(z))$, so that assuming $\sup_z \mathrm{Vol}(\mathcal{S}(z)) < \infty$ leads to a contradiction. This concludes the proof. □

Theorem 1 is a fundamental stepping stone for the develop-

[8] Orthogonal in the Euclidean space $\mathbb{R}^n$.

ments in the next sections. In short, it affirms that the region constructed according to formula $\mathcal{X} := \mathcal{E}^{\alpha}_{\Pi}(\hat{x}^{\mathcal{C}}_I)$ enjoys the state conditional Property 1 and that it is minimum volume among constructions enjoying this property. The State Conditional Filter derived in Section IV is a recursive algorithm that aims at implementing the computation of the set $\mathcal{X}$ given by Theorem 1 in the context of dynamical systems, by also pursuing the objective of making the computation recursive.

C. Minimum volume region for a subset of components

This section contains useful material for extending SCF to situations where one is only interested in estimating a subset of the n components of x; it can be skipped at a first reading without missing any material that is fundamental for the comprehension of SCF in its standard form.

Without loss of generality, we assume that the components of interest are the first s components, and write $x = [\zeta^\top \xi^\top]^\top$, $\zeta \in \mathbb{R}^s$, $\xi \in \mathbb{R}^{n-s}$. Our aim here is to construct a minimum volume region $\mathcal{Z} \subseteq \mathbb{R}^s$ that satisfies a state-conditional property only for the s components of interest, i.e., $\mathbb{P}r\{\zeta \in \mathcal{Z} | \zeta = \bar{\zeta}\} = \alpha$, $\bar{\zeta}$-a.e.

Decompose A in its first columns and the remaining columns as follows $A = [A_{(1:s)}, A_{(s+1:n)}]$, and re-write (4) as

$$y = A_{(1:s)}\zeta + A_{(s+1:n)}\xi + \nu. \tag{23}$$

By the Hilbert projection theorem, and using the notation of Section III-A, ξ can be expressed as the projection of ξ onto the span of ζ, which we denote $\mathbb{Z}$, plus an "error", as follows

$$\xi = \mathcal{P}_{\mathbb{Z}}(\xi) + (\xi - \mathcal{P}_{\mathbb{Z}}(\xi)), \tag{24}$$

where the error term in parentheses is orthogonal to ζ. By substituting (24) in (23), and writing $\mathcal{P}_{\mathbb{Z}}(\xi)$ as an explicit function of ζ, i.e., $\mathcal{P}_{\mathbb{Z}}(\xi) = \mathbb{E}[\xi\zeta^\top](\mathbb{E}[\zeta\zeta^\top])^{-1}\zeta$, [14], we get

$$y = \tilde{A}\zeta + \tilde{\nu}, \tag{25}$$

where $\tilde{A} = A_{(1:s)} + A_{(s+1:n)}\mathbb{E}[\xi\zeta^\top](\mathbb{E}[\zeta\zeta^\top])^{-1}$ and $\tilde{\nu} = \nu + A_{(s+1:n)}\left(\xi - \mathbb{E}[\xi\zeta^\top](\mathbb{E}[\zeta\zeta^\top])^{-1}\zeta\right)$. Notice that ζ and $\tilde{\nu}$ are orthogonal by the very construction of $\tilde{\nu}$ and, therefore, they are also independent owing to the Gaussianity Condition 3. Therefore, Equation (25) has the same structure as the original equation (4) and the estimation of ξ can be conducted akin to the estimation of x in (4). In particular, similarly to (11) and (12) we have

$$\hat{\zeta}^{\mathcal{C}}_I = (U_{(1:s)}\tilde{A})^{-1}\hat{\zeta}^{\mathcal{U}} \tag{26}$$

$$\mathbb{E}[(\zeta - \hat{\zeta}^{\mathcal{C}}_I)(\zeta - \hat{\zeta}_I)^\top] = (U_{(1:s)}\tilde{A})^{-1}\mathbb{E}[(\zeta - \hat{\zeta}^{\mathcal{U}})(\zeta - \hat{\zeta}^{\mathcal{U}})^\top], \tag{27}$$

where $U_{(1:s)}$ is such that $U_{(1:s)}y = \hat{\zeta}^{\mathcal{U}}$. It is worth noticing that $U_{(1:s)}$ is just nothing but the first s rows of U.

IV. The State Conditional Filter

Consider the process $\{x_t, y_t\}$ generated by system (1), which is repeated here for easy reference:

$$\begin{cases} x_{t+1} &= Fx_t + v_t \\ y_t &= Hx_t + w_t. \end{cases} \tag{28}$$

The State Conditional Filter (SCF) is a recursive algorithm that, given the measurements up to time t, i.e., given $y_1, \ldots, y_t$, builds a region $\mathcal{X}^{\mathrm{SCF}}_t$ that satisfies the state-conditional Property 1 for a user-chosen level of probability α. The SCF algorithm is derived in Section IV-A using the results of the previous Section III.

The following assumption is in effect throughout this section. Section V presents extensions and generalizations, including: filtering with exogenous input, time-varying and unstable systems, prediction.

Assumption 1. Matrix F is asymptotically stable, i.e., $|\lambda_i(F)| < 1$, $i = 1, \ldots, n$, where $\lambda_i(F)$ is the i-th eigenvalue of F. $\{v_t\}$ and $\{w_t\}$ are zero-mean, white Gaussian processes with constant covariance and independent of each other. $*$

In what follows, we refer to the stationary process $\{x_t, y_t\}$ generated by (28), that is, we consider the steady-state operation of the system.[9] For notational convenience, define $\Gamma := \mathbb{E}[x_t x_t^\top]$, $V := \mathbb{E}[v_t v_t^\top]$, $W := \mathbb{E}[w_t w_t^\top]$.[10]

Assumption 2. $\Gamma \succ 0$ and $W \succ 0$. $*$

Similarly to what we saw for the static set-up, the assumption on Γ is without loss of generality. In fact, if Γ were singular, it would be possible to find one or more linear combinations of the state components that are equal to zero and these linear combinations could be eliminated from the system model for the purpose of estimation. Moreover, should W be singular, one could measure the value of one or more state variables without noise directly from the system outputs (by a suitable change of variables). Therefore, these state variables could be eliminated from the system model and their effect on the state evolution could be modeled through an additional measured exogenous input to the system model (28), as discussed in Section V-E.

A. Derivation of the recursive formulae

First, the static theory of Section III-B is applied to the dynamical context at each time t to obtain a minimum volume region for the system state. This region is denoted by $\mathcal{X}^{\mathrm{SCF}}_t$. Then, we show that this region can be obtained recursively.

Start by reformulating the state equation in (28) according to its backward representation (see [17], Chapter 5, Section 4)[11]:

$$x_{t'-1} \;=\; F^b x_{t'} + v^b_{t'} \qquad t' = 2, \ldots, t \tag{29}$$

[9] The non-stationary case is considered in Section V-D.

[10] Γ can be computed using the Lyapunov equation $\Gamma = F\Gamma F^\top + V$.

[11] The process generated by (29) is equal realization by realization to the process generated by (28).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3103905, IEEE Transactions on Automatic Control

where

$$F^b = \Gamma F^\top \Gamma^{-1} \tag{30}$$

(recall that $\Gamma = \mathbb{E}[x_t x_t^\top]$) is the backward system matrix and $v_2^b, \dots, v_t^b$ is the backward state noise. The backward state noise is a zero mean Gaussian white process with covariance matrix $V^b = \Gamma - \Gamma F^\top \Gamma^{-1} F \Gamma$, and such that $\mathbb{E}[x_t (v_{t'}^b)^\top] = 0$, for every $t' \leq t$.

Using (29) together with the output equation $y_t = Hx_t + w_t$, the system equations can be written in batch form at time t as follows

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_1 \end{bmatrix} = \begin{bmatrix} H \\ HF^b \\ H(F^b)^2 \\ \vdots \\ H(F^b)^{t-1} \end{bmatrix} x_t + \begin{bmatrix} 0 & 0 & \cdots 0 \\ H & 0 & \cdots 0 \\ HF^b & H & \cdots 0 \\ \vdots & \vdots & \ddots \vdots \\ H(F^b)^{t-2} & H(F^b)^{t-3} & \cdots H \end{bmatrix} \begin{bmatrix} v_t^b \\ v_{t-1}^b \\ \vdots \\ v_2^b \end{bmatrix} + \begin{bmatrix} w_t \\ w_{t-1} \\ w_{t-2} \\ \vdots \\ w_1 \end{bmatrix} \tag{31}$$

or, in a more compact form, as

$$[y]_1^t = A_t x_t + \nu_t, \tag{32}$$

where ν_t accounts for $[v^b]_2^t$ and $[w]_1^t$. The fundamental fact is that equation (32) is in the form of (4), where $y = [y]_1^t$ (thus, p, the length of y, is now identified with $q \cdot t$), $x = x_t$, $A = A_t$, $\nu = \nu_t$. Moreover, the vectors x_t and ν_t are independent, so that Conditions 1 and 3 are satisfied. As for Condition 2, this condition is satisfied if the system is "backward observable" (i.e., (F^b, H) is observable) and $t \geq n$. If the system is not backward observable, the user might want to focus only on the observable variables and build a region for these variables only, as the other components cannot be estimated (refer to Section V-A for more details).

In this section, we enforce a backward observability assumption.

Assumption 3. The couple (F^b, H) is observable. ∗

Remark 7. Note that (F^b, H) can be observable while (F, H) is not and vice versa. For example, let $\Gamma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Here, if $F^b = \begin{bmatrix} \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{4} \end{bmatrix}$ and $H = \begin{bmatrix} 0 & 1 \end{bmatrix}$, then (F^b, H) is an observable couple, but $(F, H) = ((F^b)^\top, H)$ is not. ∗

In Section III-A we used the notation $\hat{x}_I^{\mathcal{C}}$ to indicate the solution of problem $\mathscr{P}_{\mathcal{C}}(Z)$ with $Z = I$. Here, we simplify the notation and use $\hat{x}_t$ to indicate the estimate of x_t based on the measurements $[y]_1^t$ with constraint $Z = I$. Moreover, when applied to x_t, the unconstrained minimum variance linear estimate $\hat{x}^{\mathcal{U}}$ (Bayesian estimate) of Section III-A is the estimate delivered by the Kalman filter, which is denoted by $\hat{x}_t^{\mathrm{KF}}$, [17], [28], [29]. In the present context, we will make explicit the time dependence of the matrix U that transforms $[y]_1^t$ into $\hat{x}_t^{\mathrm{KF}}$ by writing it as U_t, so that $\hat{x}_t^{\mathrm{KF}} = U_t [y]_1^t$. For easy reference, these correspondences are summed-up in Table I.

Static case (Section III)	Dynamical case
$\hat{x}_I^{\mathcal{C}}$	$\hat{x}_t$
$\hat{x}^{\mathcal{U}}$	$\hat{x}_t^{\mathrm{KF}}$
y, x, ν, U, A, $\mathcal{X}$	$[y]_1^t$, x_t, ν_t, U_t, A_t, $\mathcal{X}_t^{\mathrm{SCF}}$

Table I

As shown by equations (11) and (12) in Section III-A, matrix $U_t A_t$ relates $\hat{x}_t$ to $\hat{x}_t^{\mathrm{KF}}$ and the covariance matrix $\Pi_t = \mathbb{E}[(x_t - \hat{x}_t)(x_t - \hat{x}_t)^\top]$ to $P_t = \mathbb{E}[(x_t - \hat{x}_t^{\mathrm{KF}})(x_t - \hat{x}_t^{\mathrm{KF}})^\top]$ according to the fundamental relations:

$$\hat{x}_t = (U_t A_t)^{-1} \hat{x}_t^{\mathrm{KF}}, \tag{33}$$

$$\Pi_t = (U_t A_t)^{-1} P_t. \tag{34}$$

Matrix $(U_t A_t)$ will be called the "ratio" matrix because it sets in (33) and (34) the ratio between KF and SCF quantities. It is well known from the KF theory, [17], [29], that $\hat{x}_t^{\mathrm{KF}}$ and P_t can be recursively computed by the following updating rules

$$\hat{x}_t^{\mathrm{KF}} = F \hat{x}_{t-1}^{\mathrm{KF}} + K_t (y_t - HF \hat{x}_{t-1}^{\mathrm{KF}}) \tag{35a}$$

$$K_t = (FP_{t-1}F^\top + V) H^\top (W + H(FP_{t-1}F^\top + V)H^\top)^{-1} \tag{35b}$$

$$P_t = FP_{t-1}F^\top + V - K_t (W + H(FP_{t-1}F^\top + V)H^\top) K_t^\top \tag{35c}$$

with the initializations

$$\begin{cases} \hat{x}_0^{\mathrm{KF}} = 0 \\ P_0 = \Gamma. \end{cases} \tag{36}$$

We next derive a recursion for the matrix $U_t A_t$. For $t \geq 2$, rewrite the equation in (35a) as[12]

$$U_t [y]_1^t = FU_{t-1} [y]_1^{t-1} + K_t (y_t - HFU_{t-1} [y]_1^{t-1}).$$

This equality holds no matter what the values of vectors $[y]_1^{t-1}$ and y_t are. Hence, we can replace $[y]_1^{t-1}$ with a generic vector $\eta \in \mathbb{R}^{q \cdot (t-1)}$ and y_t with a generic vector $h \in \mathbb{R}^q$ and write

$$U_t \begin{bmatrix} h \\ \eta \end{bmatrix} = FU_{t-1}\eta + K_t (h - HFU_{t-1}\eta), \tag{37}$$

which is a relation that links U_t to U_{t-1}. Further, note that A_t can be decomposed as

$$A_t = \begin{bmatrix} H \\ A_{t-1} F^b \end{bmatrix}.$$

Hence, by a multiple application of (37) where $\begin{bmatrix} h \\ \eta \end{bmatrix}$ is identified each time with a different column of A_t we obtain

$$U_t A_t = F(U_{t-1} A_{t-1}) F^b + K_t (H - HF(U_{t-1} A_{t-1}) F^b), \tag{38}$$

[12] We start from $t = 2$ since for $t = 1$ the expression $U_{t-1} [y]_1^{t-1}$ has no meaning.

which is a recursive equation for U_tA_t, $t \geq 2$. Note now that if (38) is formally initialized with

$$U_0A_0 = 0 \tag{39}$$

it gives $U_1A_1 = K_1H$, which is correct since $U_1 = K_1$ (see (35a)) and $A_1 = H$. Hence, from now on, we consider (38) valid for $t \geq 1$ with the initialization (39). Note that U_tA_t is measurement-independent and can be computed off-line together with P_t and K_t. Finally, the region $\mathcal{X}_t^{\rm SCF}$ is obtained as (refer to (17))

$$\mathcal{X}_t^{\rm SCF} := \left\{x \in \mathbb{R}^n : (x - \hat{x}_t)^\top \Pi_t^{-1}(x - \hat{x}_t) \leq \chi^2(\alpha, n)\right\}, \tag{40}$$

where $\chi^2(\alpha, n)$ is the quantile at probability α of the Chi-square distribution with n degrees of freedom.

The SCF algorithm is summarized in the following two tables for easy reference.

Pseudocode: SCF initialization

- Set the probability level $\alpha \in (0, 1)$
- Let
  - ⋆ $U_0A_0 \leftarrow 0$
  - ⋆ $\hat{x}_0^{\rm KF} \leftarrow 0$, $P_0 \leftarrow \Gamma$ (KF initialization)
  - ⋆ $F^b \leftarrow \Gamma F^\top \Gamma^{-1}$ (backward matrix)

Pseudocode: computation of $\mathcal{X}_t^{\rm SCF}$

For $t = 1, 2, \ldots$

- Read new measurement y_t
- Update KF:
  - ⋆ $K_t \leftarrow (FP_{t-1}F^\top + V)H^\top(W + H(FP_{t-1}F^\top + V)H^\top)^{-1}$
  - ⋆ $\hat{x}_t^{\rm KF} \leftarrow F\hat{x}_{t-1}^{\rm KF} + K_t(y_t - HF\hat{x}_{t-1}^{\rm KF})$
  - ⋆ $P_t \leftarrow FP_{t-1}F^\top + V - K_t(W + H(FP_{t-1}F^\top + V)H^\top)K_t^\top$
- Update the ratio matrix:
  - ⋆ $U_tA_t \leftarrow F(U_{t-1}A_{t-1})F^b + K_t(H - HF(U_{t-1}A_{t-1})F^b)$
- If $t \geq n$
  - – Compute[13]
    - ⋆ $\hat{x}_t \leftarrow (U_tA_t)^{-1}\hat{x}_t^{\rm KF}$
    - ⋆ $\Pi_t \leftarrow (U_tA_t)^{-1}P_t$
    - ⋆ $\mathcal{X}_t^{\rm SCF} \leftarrow \left\{x \in \mathbb{R}^n : (x - \hat{x}_t)^\top \Pi_t^{-1}(x - \hat{x}_t) \leq \chi^2(\alpha, n)\right\}$, where $\chi^2(\alpha, n)$ is the quantile at probability α of the Chi-square distribution with n degrees of freedom
  - – Output: $\mathcal{X}_t^{\rm SCF}$

Notice also that in the above pseudocode the recursive part of SCF follows that of KF, therefore stability is achieved under the same conditions that give stability in KF.

B. Asymptotic state conditional filter

Asymptotically, the SCF algorithm becomes time-invariant. This fact offers the opportunity to implement a suboptimal, but simpler, filter.

To obtain the asymptotic filter, start by noting that, since F is stable, KF is asymptotically time-invariant (see e.g. [29], Chapter 4). Hence, we only need to show that U_tA_t tends to a constant matrix. Note that formula (14) is written in the present dynamical context as

$$P_t = \Gamma - (U_tA_t)\Gamma,$$

from which[14]

$$U_tA_t = I - P_t\Gamma^{-1}. \tag{41}$$

Since P_t converges to a limit matrix (due again to the asymptotic invariance of KF), the result that U_tA_t converges to a constant matrix is obtained.

Remark 8. According to formula (34), the inverse of the ratio matrix $(U_tA_t)^{-1}$ sets the enlargement to obtain Π_t from P_t (note that Π_t is necessarily no smaller than P_t since P_t is the error covariance matrix of the unconstrained minimum variance linear estimator). This enlargement makes the volume of $\mathcal{X}_t^{\rm SCF}$ larger than the volume of $\mathcal{X}_t^{\rm KF}$. Asymptotically, U_tA_t tends to $I - \bar{P}\Gamma^{-1}$ where $\bar{P} = \lim_{t\to\infty} P_t$. Hence, the enlargement becomes small in the long run provided that $\bar{P}$ is small as compared to Γ. An interpretation is that, when $\bar{P} \ll \Gamma$, KF is able to significantly reduce the process covariance to a much smaller estimation covariance and this is because the information conveyed by the measurements is rich. On the other hand, when measurements are rich, the prior used in KF becomes less important so that the need to enlarge $\mathcal{X}_t^{\rm KF}$ to generate $\mathcal{X}_t^{\rm SCF}$ is lessened. ∗

V. Extensions

Section V-A addresses the problem of estimating a subset of the components of x_t. SCF for full state prediction and prediction of a subset of state components are covered in Sections V-B and V-C, respectively. Section V-D deals with time-varying systems. Finally, measured inputs are considered in Section V-E.

A. Estimating a subset of components

In many applications, only a subset of the state components is of interest and has to be estimated. This situation is considered in this section.

Write $x_t = [\zeta_t^\top \ \xi_t^\top]^\top$ and assume (without loss of generality) that $\zeta_t \in \mathbb{R}^s$ is the subset of the state variables that we want to estimate. The aim is to construct a minimum volume region $\mathcal{Z}_t^{\rm SCF}$ such that $\mathbb{P}r\{\zeta_t \in \mathcal{Z}_t^{\rm SCF} | \zeta_t = \bar{\zeta}\} = \alpha$, $\bar{\zeta}$-a.e. This problem was addressed for the static estimation problem in Section III-C, and the result therein can be easily translated into the dynamical case. Rewrite equation (32) as

$$[y]_1^t = \tilde{A}_t\zeta_t + \tilde{\nu}_t, \tag{42}$$

where

$$\tilde{A}_t = A_{(1:s)t} + A_{(s+1:n)t}\mathbb{E}[\xi_t\zeta_t^\top]\left(\mathbb{E}[\zeta_t\zeta_t^\top]\right)^{-1}$$

and

$$\tilde{\nu}_t = \nu_t + A_{(s+1:n)t}\left(\xi_t - \mathbb{E}[\xi_t\zeta_t^\top](\mathbb{E}[\zeta_t\zeta_t^\top])^{-1}\zeta_t\right).$$

[13] For $t \geq n$, (U_tA_t) is invertible, a consequence of Assumption 3 in the light of Facts 1 and 2.

[14] This equation can also be used in place of (38) in the algorithm that computes $\mathcal{X}_t^{\rm SCF}$.

Using (26) and (27) one then has

$$\hat{\zeta}_t = (U_{(1:s)t}\tilde{A}_t)^{-1}\hat{\zeta}_t^{\mathrm{KF}},$$
$$\Pi_t^{\zeta} = (U_{(1:s)t}\tilde{A}_t)^{-1}P_t^{(s)},$$

where $\hat{\zeta}_t^{\mathrm{KF}}$ is obtained by selecting the first s components of $\hat{x}_t^{\mathrm{KF}} = U_t[y]_1^t$, and $P_t^{(s)}$ is the upper-left $s \times s$ submatrix of P_t.[15] Using now a relation akin to (41) for (42) gives the following expression for computing $U_{(1:s)t}\tilde{A}_t$:

$$U_{(1:s)t}\tilde{A}_t = I - P_t^{(s)}(\Gamma^{(s)})^{-1},$$

where $\Gamma^{(s)}$ is the upper-left $s \times s$ submatrix of Γ (covariance of ζ_t). The couple $\hat{\zeta}_t$ and Π_t^{ζ} can be used similarly to (40) to compute

$$\mathcal{Z}_t^{\mathrm{SCF}} = \{\zeta \in \mathbb{R}^s : (\zeta - \hat{\zeta}_t)^\top (\Pi_t^{\zeta})^{-1}(\zeta - \hat{\zeta}_t) \le \chi^2(\alpha, s)\}.$$

B. Prediction

In τ-step ahead prediction, one aims at building a region $\mathcal{X}_{t+\tau|t}^{\mathrm{SCF}}$ based on the measurements $[y]_1^t$ such that $\mathbb{P}r\{x_{t+\tau} \in \mathcal{X}_{t+\tau|t}^{\mathrm{SCF}} \,|\, x_{t+\tau} = \bar{x}\} = \alpha$, $\bar{x}$-a.e. This can be done by considering the equations

$$\begin{bmatrix} y_t \\ \vdots \\ y_2 \\ y_1 \end{bmatrix} = \begin{bmatrix} H(F^b)^\tau \\ \vdots \\ H(F^b)^{t-2+\tau} \\ H(F^b)^{t-1+\tau} \end{bmatrix} x_{t+\tau} + \begin{bmatrix} H(F^b)^{\tau-1} & \cdots & H & 0 & \cdots 0 \\ \vdots & \ddots & \cdots & \ddots & \ddots \vdots \\ H(F^b)^{\tau+t-3} & \cdots & \ddots & \cdots & H\ 0 \\ H(F^b)^{\tau+t-2} & \cdots H(F^b)^\tau & H(F^b)^{\tau-1} & \cdots & H \end{bmatrix} \begin{bmatrix} v_{t+\tau}^b \\ \vdots \\ v_{t+1}^b \\ v_t^b \\ \vdots \\ v_2^b \end{bmatrix} + \begin{bmatrix} w_t \\ w_{t-1} \\ w_{t-2} \\ \vdots \\ w_1 \end{bmatrix}$$

that can be written as

$$[y]_1^t = \bar{A}_{t+\tau}x_{t+\tau} + \bar{\nu}_{t+\tau}. \tag{43}$$

Equation (43) has the same structure as equation (32). Hence, applying *mutatis mutandis* the same rationale as in Section IV-A, one arrives at the conclusion that

$$\hat{x}_{t+\tau|t} = (I - P_{t+\tau|t}\Gamma^{-1})^{-1}\hat{x}_{t+\tau|t}^{\mathrm{KF}}$$
$$\Pi_{t+\tau|t} = (I - P_{t+\tau|t}\Gamma^{-1})^{-1}P_{t+\tau|t},$$

where $\hat{x}_{t+\tau|t}^{\mathrm{KF}} = F^\tau \hat{x}_t^{\mathrm{KF}}$ is the τ-step ahead Kalman prediction and $P_{t+\tau|t}$ is the corresponding prediction error covariance[16]. Finally, the SCF prediction region is computed as $\mathcal{X}_{t+\tau|t}^{\mathrm{SCF}} = \{x \in \mathbb{R}^n : (x - \hat{x}_{t+\tau|t})^\top \Pi_{t+\tau|t}^{-1}(x - \hat{x}_{t+\tau|t}) \le \chi^2(\alpha, n)\}$.

C. Prediction of a subset of components

Similarly to Section V-A, write $x_{t+\tau} = [\zeta_{t+\tau}^\top,\ \xi_{t+\tau}^\top]^\top$ where $\zeta_{t+\tau}$ is the subset of the state variables we want to predict. Following the same reasoning as in Sections V-B and V-A, one easily obtains the equations

$$\hat{\zeta}_{t+\tau|t} = (I - P_{t+\tau|t}^{(s)}(\Gamma^{(s)})^{-1})^{-1}\hat{\zeta}_{t+\tau|t}^{\mathrm{KF}}$$
$$\Pi_{t+\tau|t}^{\zeta} = (I - P_{t+\tau|t}^{(s)}(\Gamma^{(s)})^{-1})^{-1}P_{t+\tau|t}^{(s)},$$

where $\hat{\zeta}_{t+\tau|t}^{\mathrm{KF}}$ is obtained by selecting the first s components of $\hat{x}_{t+\tau|t}^{\mathrm{KF}}$ (τ-step ahead Kalman prediction), $P_{t+\tau|t}^{(s)}$ is the upper-left $s \times s$ submatrix of $P_{t+\tau|t}$ (covariance matrix of the τ-step ahead Kalman prediction error), and $\Gamma^{(s)}$ is the upper-left $s \times s$ submatrix of Γ. The SCF prediction region for $\zeta_{t+\tau}$ is then given by $\mathcal{Z}_{t+\tau|t}^{\mathrm{SCF}} = \{\zeta \in \mathbb{R}^s : (\zeta - \hat{\zeta}_{t+\tau|t})^\top (\Pi_{t+\tau|t}^{\zeta})^{-1}(\zeta - \hat{\zeta}_{t+\tau|t}) \le \chi^2(\alpha, s)\}$.

D. SCF for linear time-variant systems

Consider now the process generated by the linear time-variant system

$$\begin{cases} x_{t+1} &= F_t x_t + v_t \\ y_t &= H_t x_t + w_t, \end{cases} \tag{44}$$

where x_1 is Gaussian with $E[x_1] = 0$ and $\{v_t, w_t\}$ is a jointly Gaussian white process independent of x_1 with $E[v_t] = 0$, $E[w_t] = 0$, $W_t := E[w_t w_t^\top] \succ 0$ for any t. We also assume that $\Gamma_t := \mathbb{E}[x_t x_t^\top] \succ 0$ for any $t > 1$. Letting

$$F_{t+1}^b := \Gamma_t F_t^\top \Gamma_{t+1}^{-1}, \quad V_{t+1}^b := \Gamma_t - \Gamma_t F_t^\top \Gamma_{t+1}^{-1} F_t \Gamma_t,$$

the process generated by (44) admits the following backward representation for any t ([17], Lemma 5.4.4):

$$x_{t'-1} = F_{t'}^b x_{t'} + v_{t'}^b \qquad t' = 2, \dots, t, \tag{45}$$

where $v_2^b, \dots, v_t^b$ is a Gaussian white process such that $E[x_t(v_{t'}^b)^\top] = 0$ for every $t' \le t$ and $E[v_{t'}^b(v_{t'}^b)^\top] = V_{t'}^b$. Using the backward representation (45) in (44) yields (here, $F_{i:i+n}^b = \prod_{k=i}^{i+n} F_k^b$):

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_1 \end{bmatrix} = \begin{bmatrix} H_t \\ H_{t-1}F_t^b \\ H_{t-2}F_{t-1:t}^b \\ \vdots \\ H_1 F_{2:t}^b \end{bmatrix} x_t + \begin{bmatrix} 0 & 0 & \cdots 0 \\ H_{t-1} & 0 & \cdots 0 \\ H_{t-2}F_{t-1}^b & H_{t-2} & \cdots 0 \\ \vdots & \vdots & \ddots \vdots \\ H_1 F_{2:t-2}^b & H_1 F_{2:t-3}^b \cdots & H_1 \end{bmatrix} \begin{bmatrix} v_t^b \\ v_{t-1}^b \\ \vdots \\ v_2^b \end{bmatrix} + \begin{bmatrix} w_t \\ w_{t-1} \\ w_{t-2} \\ \vdots \\ w_1 \end{bmatrix},$$

which can be written in compact form as

$$[y]_1^t = A_t x_t + \nu_t.$$

This equation has the same structure as equation (32) and all previous technical derivations can therefore be carried over to this context. For example, $\mathcal{X}_t^{\mathrm{SCF}}$ is obtained by equations

$$\hat{x}_t = (I - P_t\Gamma_t^{-1})^{-1}\hat{x}_t^{\mathrm{KF}}$$
$$\Pi_t = (I - P_t\Gamma_t^{-1})^{-1}P_t$$
$$\mathcal{X}_t^{\mathrm{SCF}} = \{x \in \mathbb{R}^n : (x - \hat{x}_t)^\top \Pi_t^{-1}(x - \hat{x}_t) \le \chi^2(\alpha, n)\},$$

[15] Note that $\hat{\zeta}_t$ is not given by the first s components of $\hat{x}_t$ (while $\hat{\zeta}_t^{\mathrm{KF}}$ is indeed given by the first s components of $\hat{x}_t^{\mathrm{KF}}$).

[16] $P_{t+\tau|t}$ can be recursively computed as $P_{t+\tau|t} = FP_{t+\tau-1|t}F^\top + V$ initialized with $P_{t|t} = P_t$.

where $\hat{x}_t^{\mathrm{KF}}$ is the Bayesian estimate obtained by the Kalman filter for the present time-variant context and P_t is the corresponding error covariance.

E. SCF with measured inputs

Consider the state space system

$$\begin{cases} x_{t+1} &= F_t x_t + v_t + u_t \\ y_t &= H_t x_t + w_t, \end{cases}$$

where u_t is a deterministic measured input. In the presence of u_t, one can incorporate the input effect in $\hat{x}_t$ by using the formula

$$\hat{x}_t = (U_t A_t)^{-1} \hat{x}_t^{\mathrm{KF}} + (I - (U_t A_t)^{-1}) d_t,$$

where d_t is generated by the process $d_{t+1} = F_t d_t + u_t$ initialized with $d_0 = 0$, and $\hat{x}_t^{\mathrm{KF}}$ is the KF estimate for the case when a measured input is present. Using $\hat{x}_t$ as given above, all derivations then follow the same path as in previous sections.

VI. Concluding Remarks

In dynamical state estimation problems, not all state values have the same importance and missing to deliver an appropriate estimate has more severe consequences for certain state values than for others. In this paper, a new state estimation paradigm (SCF) has been introduced which exhibits the fundamental property that the system state is contained in the estimated region with a user-chosen probability regardless of the value of the state. This represents a conceptual departure from Kalman filtering. Among regions having this inclusion property, SCF computes the region of smallest volume.

The SCF algorithm is general-purpose and can be applied across a vast range of problems. Several variants within the same foundational paradigm centered around the inclusion property can be conceived. For instance, while the criterion of smallest volume is reasonable, it is not the only possible one and it can be replaced by volumes weighted by an "importance" function. Another direction of further investigation concerns the possibility of restricting the inclusion property to a subset of the state values. This is relevant to situations where some knowledge on what the important state values are is available in advance, which is not an uncommon situation. One aspect that has not been addressed in the present contribution and that deserves attention is that of the robustness of the SCF algorithm against possible misspecifications of the system matrices and process covariances, similarly to what has been done in [30-35] for KF. Many extensions in more structured set-ups are also possible, including distributed estimation (similarly to [36-39] in KF), with the possible presence of faults and attacks (as in [40-42]), and estimation with intermittent observations (see e.g. [43], [44]).

Acknowledgments

This work was partly supported by the Australian Research Council (ARC) under Grant DP130104028. The work of A. Carè and M.C. Campi was also partly supported by Regione Lombardia under Grant MoSoRe E81B19000840007.

References

[1] R. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME - Journal of basic engineering*, vol. 82, pp. 35–45, 1960.

[2] R. Kalman and R. Bucy, "New results in linear filtering and prediction theory," *Transactions of the ASME - Journal of basic engineering*, vol. 83, no. 1, pp. 95–108, 1961.

[3] F. Auger, M. Hilairet, J. Guerrero, E. Monmasson, T. Orlowska-Kowalska, and S. Katsura, "Industrial applications of the Kalman filter: A review," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 12, pp. 5458–5471, 2013.

[4] M. Grewal and A. Andrews, "Applications of Kalman filtering in aerospace 1960 to the present [historical perspectives]," *IEEE Control Systems Magazine*, vol. 30, no. 3, pp. 69–78, 2010.

[5] L. Gerencsér, "A modification of the extended Kalman filter algorithm with application in hydrology," in *Transactions of the Ninth Prague Conference*. Springer, 1983, pp. 241–245.

[6] S. Bittanti and S. Savaresi, "On the parametrization and design of an extended Kalman filter frequency tracker," *IEEE Transactions on Automatic Control*, vol. 45, no. 9, pp. 1718–1724, 2000.

[7] D. Dochain, "State and parameter estimation in chemical and biochemical processes: a tutorial," *Journal of Process Control*, vol. 13, no. 8, pp. 801 – 818, 2003.

[8] E. Cinquemani and M. Micheli, "State estimation in stochastic hybrid systems with sparse observations," *IEEE Transactions on Automatic Control*, vol. 51, no. 8, pp. 1337–1342, 2006.

[9] P. Frogerais, J. Bellanger, and L. Senhadji, "Various ways to compute the continuous-discrete Extended Kalman Filter," *IEEE Transactions on Automatic Control*, vol. 57, no. 4, pp. 1000–1004, 2012.

[10] D. E. Quevedo, A. Ahlen, and K. H. Johansson, "State estimation over sensor networks with correlated wireless fading channels," *IEEE Transactions on Automatic Control*, vol. 58, no. 3, pp. 581–593, 2013.

[11] G. Battistelli, L. Chisci, N. Forti, G. Pelosi, and S. Selleri, "Distributed finite-element Kalman filter for field estimation," *IEEE Transactions on Automatic Control*, vol. 62, no. 7, pp. 3309–3322, 2017.

[12] L. Parker, *Multiple Mobile Robot Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 921–941.

[13] G. Dudek and M. Jenkin, *Computational principles of mobile robotics*. Cambridge university press, 2010.

[14] D. Luenberger, *Optimization by vector space methods*. John Wiley & Sons, 1969.

[15] F. Schweppe, *Uncertain Dynamic Systems*. New York, USA: Prentice-Hall, 1973.

[16] E. Lehmann and G. Casella, *Theory of point estimation*, 2nd ed. Springer, 1998.

[17] T. Kailath, A. Sayed, and B. Hassibi, *Linear estimation*. Prentice-Hall, Inc., 2000.

[18] A. Shiryaev, *Probability (2nd ed.)*. Secaucus, New Jersey, USA: Springer-Verlag New York, Inc., 1995.

[19] D. Freedman, *Statistical models: theory and practice*. Cambridge University Press, 2009.

[20] J. Wall Jr, A. Willsky, and N. Sandell Jr, "On the fixed-interval smoothing problem," *Stochastics: An International Journal of Probability and Stochastic Processes*, vol. 5, no. 1-2, pp. 1–41, 1981.

[21] R. Nikoukhah, A. Willsky, and B. Levy, "Kalman filtering and Riccati equations for descriptor systems," *Automatic Control, IEEE Transactions on*, vol. 37, no. 9, pp. 1325–1342, 1992.

[22] K. Chou, A. Willsky, and R. Nikoukhah, "Multiscale systems, Kalman filters, and Riccati equations," *Automatic Control, IEEE Transactions on*, vol. 39, no. 3, pp. 479–492, 1994.

[23] P. Teunissen and A. Khodabandeh, "BLUE, BLUP and the Kalman filter: some new results," *Journal of Geodesy*, vol. 87, no. 5, pp. 461–473, 2013.

[24] S. Kayalar and H. Weinert, "Oblique projections: Formulas, algorithms, and error bounds," *Mathematics of Control, Signals and Systems*, vol. 2, no. 1, pp. 33–45, 1989.

[25] R. Behrens and L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Transactions on Signal Processing*, vol. 42, no. 6, pp. 1413–1424, 1994.

[26] A. Sayed and T. Kailath, "Oblique state-space estimation algorithms," in *Proceedings of the 1995 IEEE American Control Conference - ACC'95*, 1995, pp. 1969–1973.

[27] R. Behrens and L. Scharf, "Corrections to "Signal processing applications of oblique projection operators" [correspondence]," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, p. 1300, 1996.

[28] F. Schweppe, "Recursive state estimation: unknown but bounded errors and system inputs," *IEEE Transactions on Automatic Control*, vol. 13, no. 1, pp. 22–28, 1968.
[29] B. Anderson and J. Moore, "Optimal filtering," *Information and system sciences series*, 1979.
[30] S. Dey and J. Moore, "Risk-sensitive filtering and smoothing via reference probability methods," *IEEE Transactions on Automatic Control*, vol. 42, no. 11, pp. 1587–1591, 1997.
[31] I. Petersen and A. Savkin, *Robust Kalman filtering for signals and systems with large uncertainties*. Springer Science & Business Media, 1999.
[32] A. H. Sayed, "A framework for state-space estimation with uncertain models," *IEEE Transactions on Automatic Control*, vol. 46, no. 7, pp. 998–1013, 2001.
[33] B. C. Levy and R. Nikoukhah, "Robust state space filtering under incremental model perturbations subject to a relative entropy tolerance," *IEEE Transactions on Automatic Control*, vol. 58, no. 3, pp. 682–695, 2013.
[34] M. Zorzi, "Robust Kalman filtering under model perturbations," *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 2902–2907, 2017.
[35] M. Zorzi, "On the robustness of the Bayes and Wiener estimators under model uncertainty," *Automatica*, vol. 83, pp. 133 – 140, 2017.
[36] F. Dorfler, F. Pasqualetti, and F. Bullo, "Continuous-time distributed observers with discrete communication," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 296–304, 2013.
[37] G. Battistelli, L. Chisci, and C. Fantacci, "Parallel consensus on likelihoods and priors for networked nonlinear filtering," *IEEE Signal Processing Letters*, vol. 21, no. 7, pp. 787–791, 2014.
[38] F. Cacace, F. Conte, and A. Germani, "Filtering continuous-time linear systems with time-varying measurement delay," *IEEE Transactions on Automatic Control*, vol. 60, no. 5, pp. 1368–1373, 2015.
[39] S. Battilotti, F. Cacace, M. d'Angelo, and A. Germani, "Distributed Kalman filtering over sensor networks with unknown random link failures," *IEEE Control Systems Letters*, vol. 2, no. 4, pp. 587–592, 2018.
[40] A.-Y. Lu and G.-H. Yang, "Secure state estimation for multiagent systems with faulty and malicious agents," *IEEE Transactions on Automatic Control*, vol. 65, no. 8, pp. 3471–3485, 2019.
[41] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Transactions on Automatic Control*, vol. 61, no. 8, pp. 2079–2091, 2016.
[42] S. Sundaram and C. N. Hadjicostis, "Distributed function calculation via linear iterative strategies in the presence of malicious agents," *IEEE Transactions on Automatic Control*, vol. 56, no. 7, pp. 1495–1508, 2011.
[43] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, 2004.
[44] A. Fasano, A. Germani, and A. Monteriù, "A detection-estimation approach to filtering for Gaussian systems with intermittent observations," in *Proceedings of the 53rd IEEE Conference on Decision and Control*, 2014, pp. 4172–4177.

Algo Carè received the Ph.D. degree in informatics and automation engineering in 2013 from the University of Brescia, Italy, where he is currently a Research Fellow with the Department of Information Engineering. He spent two years at the University of Melbourne, VIC, Australia, as a Research Fellow in system identification with the Department of Electrical and Electronic Engineering. Dr. Carè received a two-year ERCIM Fellowship in 2016 that he spent at the Institute for Computer Science and Control (SZTAKI), Hungarian Academy of Sciences (MTA), Budapest, Hungary, and at the Multiscale Dynamics Group, National Research Institute for Mathematics and Computer Science (CWI), Amsterdam, The Netherlands. He was the recipient of the triennial Stochastic Programming Student Paper Prize awarded by the Stochastic Programming Society for the period 2013–2016. His current research interests include data-driven control and decision methods, system identification, and learning theory.

Marco Claudio Campi is Professor of Automatic Control at the *University of Brescia*, Italy. He has held visiting and teaching appointments at the *Australian National University*, Canberra, Australia; the *University of Illinois at Urbana-Champaign*, USA; the *Centre for Artificial Intelligence and Robotics*, Bangalore, India; the *University of Melbourne*, Australia; the *Kyoto University*, Japan; the *Texas A&M University*, USA; the *NASA Langley Research Center*, Hampton, Virginia. USA. Marco Campi was till recently the chair of the *Technical Committee IFAC on Modeling, Identification and Signal Processing (MISP)*, and has been in various capacities on the Editorial Board of *Automatica*, *Systems and Control Letters* and the *European Journal of Control*. He is a recipient of the "Giorgio Quazza" prize, and, in 2008, he received the IEEE CSS George S. Axelby outstanding paper award for the article *The Scenario Approach to Robust Control Design*. He has delivered plenary and semi-plenary addresses at major conferences including SYSID, MTNS, and CDC. He has been a *distinguished lecturer of the Control Systems Society* on three terms. Marco Campi is a Fellow of IEEE, and a fellow of IFAC. The research interests of Marco Campi include: *inductive methods*, *data-driven decision making*, *learning*, *system identification*, *stochastic systems*.

Erik Weyer is a Professor in the Department of Electrical and Electronic Engineering at the University of Melbourne, Australia. He received the Siv. Ing. degree in 1988 and the Ph.D. in 1993, both from the Norwegian Institute of Technology. From 1994 to 1996 he was a Research Fellow at the University of Queensland, Australia, before he joined the University of Melbourne in 1997. He has held visiting positions at the University of Brescia, Italy, the Technical University of Vienna, Austria, and Politecnico di Milano, Italy. From 2010 to 2012 he was an Associate Editor of IEEE Transactions of Automatic Control and, and he is currently and Associate Editor of Automatica. In 2014 he was awarded the IEEE Control Systems Technology Award for the development and implementation of controls for irrigation channels and water management. His research interests are in the areas of system identification and control.