# Low-rank Kalman filtering under model uncertainty

Shenglun Yi and Mattia Zorzi

Abstract—We consider a robust filtering problem where the nominal state space model is not reachable and different from the actual one. We propose a robust Kalman filter which solves a dynamic game: one player selects the least-favorable model in a given ambiguity set, while the other player designs the optimum filter for the least-favorable model. It turns out that the robust filter is governed by a low-rank risk sensitive-like Riccati equation. Finally, simulation results show the effectiveness of the proposed filter.

#### I. INTRODUCTION

Kalman filtering is widely used. However, it leads to poor performances in some applications. This is because the filter is based on a nominal model which is usually different than the actual one. To address such a weakness, robust versions based on the standard Kalman filter have been considered, [1]–[5].

Risk sensitive Kalman filters [6]–[9] address the model uncertainty by replacing the standard quadratic loss function by an exponential quadratic loss function. The latter severely penalizes large errors. Such severity is tuned by the so called risk sensitivity parameter. Later on, it has been proved that risk sensitive Kalman filtering is equivalent to a minimax game, [10]–[12]: one player (called nature) selects the least-favorable model in a given set (called ambiguity set), while the other player designs the optimum filter for the least-favorable model. Here, the ambiguity set is a ball, in the Kullback-Leibler (KL) divergence topology, about the nominal model. The radius of this ball depends on how much uncertainty the nominal model contains.

A modern formulation of risk sensitive filters is represented by robust Kalman filters where the model uncertainty is expressed incrementally, [13]–[18]. More precisely, the ambiguity set is specified at each time step: in this way the nature cannot concentrate all the uncertainty in one specific time step. These filters are built from a robust static estimation problem showing that the Bayes estimator is optimal with respect to the ambiguity set formed by the KL divergence, [19]. Interestingly, these results can be extended to ambiguity sets formed by the  $\tau$ -divergence, [20], [21]. These robust Kalman filters, however, are well defined only in the case that the dynamic game involves non-degenerate probability densities. Such a condition is guaranteed by assuming that the nominal state space model is reachable and observable. The contribution of this paper is to extend this robust Kalman filtering approach in the case that reachability is not assumed. In this case the dynamic game could involve degenerate Gaussian probability densities, i.e. their covariance matrices are low-rank. It turns out that the resulting robust Kalman filter is governed by a low-rank risk sensitive-like Riccati iteration. Although low-rank and singular Riccati iterations have been studied in the literature, e.g. [22]–[24], our iteration appears to be new.

The outline of the paper is as follows. Section II regards the low-rank robust static estimation problem showing the optimality of the Bayes estimator. The latter is then used to derive the low-rank robust Kalman filter in Section III. Some simulations are presented in Section IV which show that the proposed filter outperforms the Kalman filter in the case of model uncertainty. Finally, the conclusions and the future work are discussed in Section V.

We warn the reader that the present paper only reports some preliminary result regarding the robust estimation under model uncertainty in the case that reachability is not assumed. In particular, all the proofs and most of the technical assumptions needed therein are omitted and will be published afterwards.

Notation: The image of matrix K is denoted by Im(K). Given a symmetric matrix K: K > 0 ( $K \ge 0$ ) means that K is positive (semi) definite;  $\sigma_{max}(K)$  is the maximum eigenvalue of K. The symbol diag  $(d_1, \dots, d_n)$  denotes the diagonal matrix whose entries in the main diagonal are  $d_1 \dots d_n$ .  $x \sim \mathcal{N}(m, K)$  means that x is a Gaussian random variable with mean m and covariance matrix K.

#### **II. LOW-RANK ROBUST STATIC ESTIMATION**

We consider a robust static estimation problem where we want to estimate a random vector x, taking values in  $\mathbb{R}^n$ , given the observation vector y, taking values in  $\mathbb{R}^p$ , and whose joint probability distribution is degenerate. More precisely, let  $z := \begin{bmatrix} x^T & y^T \end{bmatrix}^T$  and  $f(z) \sim \mathcal{N}(m_z, K_z)$ denote the nominal probability density function of z where  $m_z \in \mathbb{R}^{n+p}$  and  $K_z \in \mathbb{R}^{n+p \times n+p}$  are such that

$$m_z = \begin{bmatrix} m_x \\ m_y \end{bmatrix}, \quad K_z = \begin{bmatrix} K_x & K_{xy} \\ K_{yx} & K_y \end{bmatrix}$$

We assume that  $K_z$  is such that  $\operatorname{rank}(K_z) = r + p$  with r < n and  $K_y > 0$ . Moreover, let  $\tilde{f}(z) \sim \mathcal{N}(\tilde{m}_z, \tilde{K}_z)$  be the actual probability density function where  $\operatorname{rank}(\tilde{K}_z) = r + p$ . Accordingly,

$$f(z) = \left[ (2\pi)^{r+p} \det^+ (K_z) \right]^{-1/2} \\ \times \exp\left[ -\frac{1}{2} (z - m_z)^T K_z^+ (z - m_z) \right]$$
(1)

S. Yi is with the School of Automation, Beijing Institute of Technology, Beijing 100081, China; M. Zorzi is with the Department of Information Engineering, University of Padova, Via Gradenigo 6/B, 35131 Padova, Italy. Emails: 3120185460@bit.edu.cn, zorzimat@dei.unipd.it

and

$$\tilde{f}(z) = \left[ (2\pi)^{r+p} \det^+ \left( \tilde{K}_z \right) \right]^{-1/2} \\ \times \exp\left[ -\frac{1}{2} \left( z - \tilde{m}_z \right)^T \tilde{K}_z^+ \left( z - \tilde{m}_z \right) \right]$$
(2)

where  $K_z^+$  and  $\tilde{K}_z^+$  are the pseudo-inverse of  $K_z$  and  $\tilde{K}_z$ , respectively, and det<sup>+</sup> is the pseudo-determinant. Notice that the supports of  $f_z$  and  $\tilde{f}_z$  are, respectively, the r + p-dimensional affine subspaces

$$\mathcal{A} = \{ m_z + v, \quad v \in \operatorname{Im}(K_z) \}$$
$$\tilde{\mathcal{A}} = \{ \tilde{m}_z + v, \quad v \in \operatorname{Im}(\tilde{K}_z) \}.$$

To measure the mismatch between f(z) and  $\tilde{f}(z)$  we introduce the KL-divergence between these degenerate probability density functions. In order to do that, we have to impose  $\mathcal{A} = \tilde{\mathcal{A}}$ , indeed such divergence is not able to measure "deterministic" discrepancies between the nominal and the actual model. The latter condition is equivalent to impose that

$$\operatorname{Im}(K_z) = \operatorname{Im}(K_z), \ \Delta m_z \in \operatorname{Im}(K_z)$$

where  $\Delta m_z = \tilde{m}_z - m_z$ . Under the aforementioned assumption, the KL-divergence is defined as

$$D(\tilde{f}, f) = \int_{\mathcal{A}} \ln\left(\frac{\tilde{f}(z)}{f(z)}\right) \tilde{f}(z) dz.$$
(3)

Substituting (1) and (2) in (3), it is not difficult to see that

$$D(\tilde{f}, f) = \frac{1}{2} \left[ \Delta m_z^T K_z^+ \Delta m_z + \ln \det^+(K_z) - \ln \det^+(\tilde{K}_z) + \operatorname{tr}\left(K_z^+ \tilde{K}_z\right) - (r+p) \right].$$
(4)

**Lemma 1.** Let  $f(z) \sim \mathcal{N}(m_z, K_z)$  and  $\tilde{f}(z) \sim \mathcal{N}(\tilde{m}_z, \tilde{K}_z)$ be degenerate Gaussian probability density functions with the same r + p-dimensional support  $\mathcal{A}$ . Let

$$\mathcal{U} = \{ \tilde{m}_z \in \mathbb{R}^{n+p} \text{ s.t. } \tilde{m}_z - m_z \in \operatorname{Im}(K_z) \}$$
$$\mathcal{V} = \{ \tilde{K}_z \in \mathbb{R}^{n+p \times n+p} \text{ s.t. } K_z = K_z^T, \operatorname{Im}(K_z) = \operatorname{Im}(\tilde{K}_z) \}$$

Then,  $D(\tilde{f}, f)$  is strictly convex with respect to  $\tilde{m}_z \in \mathcal{U}$  and  $\tilde{K}_z \in \mathcal{V}$ . Moreover,  $D(\tilde{f}, f) \geq 0$  and equality hold if and only if  $f = \tilde{f}$ .

Assume that the nominal density f is known while the actual one  $\tilde{f}$ , having the same support of f, is not. Next, we design a robust estimator  $\hat{x} = g^0(y)$  of x according to the worst probability density  $\tilde{f}(z)$  belonging to the ambiguity set which is a ball

$$\mathcal{B} = \{ \tilde{f} \sim \mathcal{N}(\tilde{m}_z, \tilde{K}_z) \text{ s.t. } D(\tilde{f}, f) \le c \}$$

where c is the mismodeling budget hereafter called tolerance. More precisely, we aim to solve the following minimax problem

$$(\tilde{f}^0, g^0) = \arg\min_{g \in \mathcal{G}} \max_{\tilde{f} \in \mathcal{B}} J(\tilde{f}, g)$$
(5)

where

$$\begin{split} J(\tilde{f},g) &= \frac{1}{2} E_{\tilde{f}} \left[ \|H(x-g(y))\|^2 \right] \\ &= \frac{1}{2} \int_{\mathcal{A}} \|H(x-g(y))\|^2 \tilde{f}(z) dz; \end{split}$$

 $H \in \mathbb{R}^{q \times n}$  with  $q \leq r$  and full row rank;  $\mathcal{G}$  is the set of estimators for which  $E_{\tilde{f}}[||H(x - g(y))||^2]$  is bounded with respect to all the Gaussian densities in  $\mathcal{B}$ .

**Theorem 1.** Let f be a Gaussian (possibly degenerate) density defined as in (1) with  $K_y > 0$ . Assume that  $\operatorname{Im}(H^T) \subseteq \operatorname{Im}(P)$  with

$$P := K_x - K_{xy} K_y^{-1} K_{yx}.$$
 (6)

Then, the least favorable Gaussian density  $\tilde{f}^0$  is with mean vector and covariance matrix

$$\tilde{m}_z^0 = m_z = \begin{bmatrix} m_x \\ m_y \end{bmatrix}, \quad \tilde{K}_z^0 = \begin{bmatrix} \tilde{K}_x & K_{xy} \\ K_{yx} & K_y \end{bmatrix}$$

so that, only the covariance of x is perturbed. Then, the optimal robust estimator is the Bayes estimator

$$g^{0}(y) = G_{0}(y - m_{y}) + m_{x}$$
(7)

with  $G_0 = K_{xy}K_y^{-1}$ . The nominal posterior covariance matrix of x given y is given by (6), while the least favorable one is:

$$\tilde{P} := \tilde{K}_x - K_{xy} K_y^{-1} K_{yx}.$$

Then, we have

$$\tilde{P} = (P^+ - \lambda^{-1} H^T H)^+.$$

Moreover, there exists a unique Lagrangian multiplier  $\lambda > \sigma_{max}(Q) > 0$ , such that  $c = D(\tilde{f}^0, f)$  where

$$Q := HH^T (HP^+H^T)^{-1}HH^T.$$

*Remark:* In the case that P > 0 then f is a non-degenerate density. In such a case, Theorem 1 still holds and: the pseudo-inverse is replaced by the inverse; moreover,  $H^T H$  can be chosen as the identity matrix. In the latter case, we recover the robust static estimation problem proposed in [19].

### III. LOW-RANK ROBUST KALMAN FILTER

We consider a nominal Gauss-Markov state space model of the form:

$$\begin{cases} x_{t+1} = A_t x_t + B_t v_t \\ y_t = C_t x_t + D_t v_t \end{cases}$$
(8)

where  $A_t \in \mathbb{R}^{n \times n}$ ,  $B_t \in \mathbb{R}^{n \times n+p}$ ,  $C_t \in \mathbb{R}^{p \times n}$ ,  $D_t \in \mathbb{R}^{p \times n+p}$ ;  $x_t$  and  $y_t$  are the state vector and the observation vector, respectively.  $v_t$  is normalized white Gaussian noise. We assume that  $D_t D_t^T > 0$  that is all the components of the observation process are affected by a full rank *p*-dimensional noise. Let  $z_t := \begin{bmatrix} x_{t+1}^T & y_t^T \end{bmatrix}^T$ , so the nominal conditional transition probability density function of the nominal state space model is  $\phi_t(z_t|x_t) \sim \mathcal{N}\left(m_{z_t|x_t}, K_{z_t|x_t}\right)$  with

$$m_{z_t|x_t} = \left[ \begin{array}{c} A_t \\ C_t \end{array} \right] x_t, \quad K_{z_t|x_t} = \left[ \begin{array}{c} B_t B_t^T & B_t D_t^T \\ D_t B_t^T & D_t D_t^T \end{array} \right].$$

Notice that  $K_{z_t|x_t}$  is not necessarily positive, i.e.  $\phi_t(z_t|x_t)$  could be degenerate. Indeed no assumption on  $B_t$  has been made. Then, we assume at time t the a priori conditional density of  $x_t$  given  $Y_{t-1} := \{y_0 \dots y_t\}$  is

$$\check{f}_t \left( x_t | Y_{t-1} \right) \sim \mathcal{N}(\hat{x}_t, \tilde{P}_t) \tag{9}$$

with  $\operatorname{rank}(\tilde{P}_t) = r_t$ . Therefore, we obtain the pseudonominal density

$$f_t\left(z_t|Y_{t-1}\right) = \int_{\check{\mathcal{A}}_t} \phi_t\left(z_t|x_t\right)\check{f}_t\left(x_t|Y_{t-1}\right) dx_t$$

where  $\check{A}_t$  denotes the support of  $\check{f}_t$ . Then, it is not difficult to see that  $f_t(z_t|Y_{t-1}) \sim \mathcal{N}(m_{z_t}, K_{z_t})$  with

$$m_{z_t} = \begin{bmatrix} A_t \\ C_t \end{bmatrix} \hat{x}_t, \quad K_{z_t} = \begin{bmatrix} K_{x_{t+1}} & K_{x_{t+1}y_t} \\ K_{y_t x_{t+1}} & K_{y_t} \end{bmatrix}$$

where the conditional covariance matrix  $K_{z_t}$  takes the parametric form

$$K_{z_t} = \begin{bmatrix} A_t \\ C_t \end{bmatrix} \tilde{P}_t \begin{bmatrix} A_t^T & C_t^T \end{bmatrix} + \begin{bmatrix} B_t \\ D_t \end{bmatrix} \begin{bmatrix} B_t^T & D_t^T \end{bmatrix}.$$
(10)

Notice that the support of  $f_t(z_t|Y_{t-1})$  is the affine subspace

$$\mathcal{A}_t := \left\{ m_{z_t} + w_t, \quad w_t \in \operatorname{Im}\left(K_{z_t}\right) \right\}.$$

Let  $\tilde{\phi}_t(z_t|x_t)$  be the actual least favorable density of  $z_t$  given  $x_t$ . Then, the marginal density is

$$\tilde{f}_t\left(z_t|Y_{t-1}\right) = \int_{\tilde{\mathcal{A}}_t} \tilde{\phi}_t\left(z_t|x_t\right) \check{f}_t\left(x_t|Y_{t-1}\right) dx_t.$$

In what follows, we assume  $\tilde{\phi}_t$  is Gaussian, accordingly the pseudo-actual density  $\tilde{f}_t(z_t|Y_{t-1}) \sim \mathcal{N}(\tilde{m}_{z_t}, \tilde{K}_{z_t})$  is Gaussian. In order to measure the mismatch between  $f_t(z_t|Y_{t-1})$  and  $\tilde{f}_t(z_t|Y_{t-1})$  using the KL-divergence we impose that  $\tilde{f}_t(z_t|Y_{t-1})$  has the same support of  $f_t(z_t|Y_{t-1})$ . Accordingly,  $\operatorname{Im}(\tilde{K}_{z_t}) = \operatorname{Im}(K_{z_t})$  and  $\Delta m_{z_t} := \tilde{m}_{z_t} - m_{z_t} \in \operatorname{Im}(K_{z_t})$ . Under the above assumptions, we assume that the actual density belongs the following ambiguity set

$$\tilde{\mathcal{B}}_t = \{ \tilde{f}_t \sim \mathcal{N}(\tilde{m}_{z_t}, \tilde{K}_{z_t}) \quad s.t. \ D(\tilde{f}_t, f_t) \le c_t \}.$$
(11)

It is worth noting that the model uncertainty in (11) is expressed incrementally. In plain words, the tolerance  $c_t$ represents the mismodeling budget allowed at time t. Then, we consider as robust one step-ahead predictor  $\hat{x}_{t+1}$  of  $x_{t+1}$ given  $Y_t$ , the solution to the following minimax game

$$\hat{x}_{t+1} = \operatorname*{argmin}_{g_t \in \mathcal{G}_t} \max_{\tilde{f}_t \in \mathcal{B}_t} \bar{J}_t(\tilde{f}_t, g_t)$$
(12)

where

$$\bar{J}_t(\tilde{f}_t, g_t) = \int_{\mathcal{A}_t} \|H_t \left( x_{t+1} - g_t \left( y_t \right) \right)\|^2 \tilde{f}_t \left( z_t | Y_{t-1} \right) \mathrm{d} z_t,$$

 $H_t \in \mathbb{R}^{q \times n}$  with  $q \leq r$  is a square root of the projection matrix having the same image of  $P_{t+1}$ , i.e.  $H_t^T H_t$  is the projection matrix such that  $\operatorname{Im}(H_t^T H_t) = \operatorname{Im}(P_{t+1})$ , where

$$P_{t+1} := K_{x_{t+1}} - K_{x_{t+1}, y_t} K_{y_t}^{-1} K_{y_t, x_{t+1}}.$$
 (13)

In the case that  $P_{t+1} > 0$ , then  $H_t$  is an orthogonal matrix and thus  $||H_t(x_{t+1}-g_t(y_t))||^2 = ||(x_{t+1}-g_t(y_t))||^2$ . Accordingly, the minimax problem in (12) boils down to the one in [13].

*Remark:* Condition (9) means that, using the terminology coined by Hansen and Sargent [25], the maximizer in (12) is operating under commitment, i.e. the maximizer is required to commit all the least favorable model components at early stages with the estimating player.

To solve Problem (12) we use Theorem 1. Indeed, replacing f,  $\tilde{f}$ , g, H by  $f_t$ ,  $\tilde{f}_t$ ,  $g_t$ ,  $H_t$ , respectively, it is not difficult to see that all the assumptions are satisfied. In particular the condition on  $H_t$  is satisfied and we have

$$K_{y_t} = C_t \tilde{P}_t C_t^T + D_t D_t^T \ge D_t D_t^T > 0.$$

Then, the least favorable density is  $\tilde{f}_t^0(z_t|Y_{t-1}) \sim \mathcal{N}(m_{z_t}, \tilde{K}_{z_t})$  where

$$\tilde{K}_{z_t} = \left[ \begin{array}{cc} \tilde{K}_{x_{t+1}} & K_{x_{t+1},y_t} \\ K_{y_t,x_{t+1}} & K_{y_t} \end{array} \right]$$

Then, the nominal posterior covariance of  $x_{t+1}$  given  $Y_t$  has been defined in (13) and the the least favorable one is

$$\tilde{P}_{t+1} = \tilde{K}_{x_{t+1}} - K_{x_{t+1},y_t} K_{y_t}^{-1} K_{y_t,x_{t+1}}.$$

Moreover,

$$\tilde{P}_{t+1} = \left(P_{t+1}^+ - \lambda_{t+1}^{-1} H_t^T H_t\right)^+$$

where  $\lambda_t > \sigma_{max}(H_t H_t^T (H_t P_{t+1}^+ H_t^T)^{-1} H_t H_t^T)$  is the unique solution of the following equation

$$\gamma_t(P_{t+1}, \lambda_t) := \frac{1}{2} \left\{ \text{lndet}^+(P_{t+1}) - \text{lndet}^+(\tilde{P}_{t+1}) + \text{tr} \left[ P_{t+1}^+ \tilde{P}_{t+1} - I_{r_{t+1}} \right] \right\} = c_t.$$

Moreover, the robust estimator is

$$\hat{x}_{t+1} = A_t \hat{x}_t + G_t \left( y_t - C_t \hat{x}_t \right).$$

where  $G_t = K_{x_{t+1},y_t} K_{y_t}^{-1}$  and by equation (10), we get the parametric form of  $G_t$  and  $P_{t+1}$ 

$$G_{t} = (A_{t}\tilde{P}_{t}C_{t}^{T} + B_{t}D_{t}^{T})(C_{t}\tilde{P}_{t}C_{t}^{T} + D_{t}D_{t}^{T})^{-1}$$
$$P_{t+1} = A_{t}\tilde{P}_{t}A_{t}^{T} - G_{t}(C_{t}\tilde{P}_{t}C_{t}^{T} + D_{t}D_{t}^{T})G_{t}^{T} + B_{t}B_{t}^{T}.$$
 (14)

Algorithm 1 Low-rank robust Kalman filter at time t Require:  $y_t$ ,  $\hat{x}_t$ ,  $\tilde{P}_t$ ,  $c_t$ 1:  $G_t = (A_t \tilde{P}_t C_t + B_t D_t^T)^T (C_t \tilde{P}_t C_t^T + D_t D_t^T)^{-1}$ 2:  $\hat{x}_{t+1} = A_t \hat{x}_t + G_t (y_t - C_t \hat{x}_t)$ 3:  $P_{t+1} = A_t \tilde{P}_t A_t^T - G_t \left( C_t \tilde{P}_t C_t^T + D_t D_t^T \right) G_t^T + B_t B_t^T$ 4: Select  $H_t^T H_t$  as projection matrix with image  $\operatorname{Im}(P_{t+1})$ 5: Find  $\lambda_t$  s.t.  $\gamma_t(P_{t+1}, \lambda_t) = c_t$ 6:  $\tilde{P}_{t+1} = \left( P_{t+1}^+ - \lambda_t^{-1} H_t^T H_t \right)^+$ 

Algorithm 1 summarizes the robust Kalman filter that we obtain. It is worth noting that steps 3 and 6 represent a risk sensitive-like Riccati iteration which is well defined also in

the case that  $\tilde{P}_t$  is low-rank. Notice that  $\theta_t := \lambda_t^{-1}$  represents the time-varying risk sensitivity parameter. In the situation that  $c_t = 0$ , i.e. the actual model corresponds to the nominal model, then we have  $\theta_t = 0$  and thus  $P_t = \tilde{P}_t$ ; in particular, (14) become the usual Riccati equation and thus we obtain the standard Kalman filter.

## **IV. SIMULATION RESULTS**

We consider the linear time-invariant model

$$\begin{cases} x_{t+1} = Ax_t + Bv_t \\ y_t = Cx_t + Dv_t \end{cases}$$
(15)

where

$$A = \begin{bmatrix} 1.1 & 0.1 & 0.1 \\ 0 & 0.6364 & -0.6364 \\ 0 & 0.6364 & 0.6364 \end{bmatrix}, \ B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$
$$C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \ D = \begin{bmatrix} 0 & 1 \end{bmatrix},$$

and  $x_0 \sim \mathcal{N}\left(0, \tilde{P}_0\right)$  with

$$\tilde{P}_0 = \left[ \begin{array}{rrrr} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{array} \right].$$

Notice that the pair (A, B) is stabilizable, but not reachable. Finally, the pair (A, C) is observable.

First, we compare  $P_t$  and  $\tilde{P}_t$  by using the robust Kalman filter (RKF) of section III and the Kalman filter (KF). More precisely, we consider two values of the tolerance for RKF:  $c_1 = 5 \cdot 10^{-2}$ ,  $c_2 = 8 \cdot 10^{-2}$  and we denote the corresponding robust filters as RKF1 and RKF2, respectively. Figure 1 shows the trace of  $P_t$  over the time horizon [0, 90].



Fig. 1. Trace of  $P_t$  for KF, RKF1 with  $c = 5 \cdot 10^{-2}$ , and RKF2 with  $c = 8 \cdot 10^{-2}$ . Recall that  $\tilde{P}_t = P_t$  for KF.

Clearly  $tr(P_t)$  of KF converges to a constant value because the nominal model is stabilizable and observable, indeed it is well known that the corresponding Riccati equation converges to a unique solution. In regard to RKF1 and RKF2,  $tr(P_t)$  converges for both. Moreover, the larger cis, the more  $tr(P_t)$  is different from the one of KF. It is worth noticing that  $rank(P_t) = 2$  for KF, RKF1 and RKF2. Figure 2 shows the minimum nonnull eigenvalue of  $P_t$ : we



Fig. 2. Minimum nonnull eigenvalue of  $P_t$  as a function of t for KF, RKF1 with  $c = 5 \cdot 10^{-2}$ , and RKF2 with  $c = 8 \cdot 10^{-2}$ .

notice that  $\operatorname{rank}(P_t) \to 1$  as  $t \to \infty$  for all the filters. This is clearly expected from KF because the corresponding algebraic Riccati equation admits a unique solution with rank equal to one. It is also worth noting that the highest convergence rate to the rank one solution is given by KF, while the larger c is, the slower the convergence rate is. Figures 3 and 4 show the trace of  $\tilde{P}_t$  and its minimum nonnull eigenvalue as a function of t. Similar observations can be made also in this case. Finally, Figure 5 shows the



Fig. 3. Trace of  $\tilde{P}_t$  for KF, RKF1 with  $c = 5 \cdot 10^{-2}$ , and RKF2 with  $c = 8 \cdot 10^{-2}$ . Recall that  $\tilde{P}_t = P_t$  for KF.

risk sensitivity parameter  $\theta_t$  as a function of t. As expected, it converges to a constant value and the larger c is, the larger  $\theta_t$  is.

Next, we evaluate the performance of RKF2 and KF assuming that the actual model corresponds to (15). Let



Fig. 4. Minimum nonnull eigenvalue of  $\tilde{P}_t$  as a function of t for KF, RKF1 with  $c = 5 \cdot 10^{-2}$ , and RKF2 with  $c = 8 \cdot 10^{-2}$ .



Fig. 5. Risk sensitivity parameter  $\theta_t$  as a function of t for RKF1 with  $c = 5 \cdot 10^{-2}$ , and RKF2 with  $c = 8 \cdot 10^{-2}$ .

 $e_{t+1} := x_{t+1} - \hat{x}_{t+1}$  be the prediction error using a predictor of the form

$$\hat{x}_{t+1} = A\hat{x}_t + G'_t(y_t - C\hat{x}_t).$$
(16)

Then, it is not difficult to see that  $e_{t+1} = (A - G'_t C)e_t + (B - G'_t D)v_t$ . Thus,  $e_t$  is a process with zero mean and with covariance matrix  $V_t := \mathbb{E}[e_t e_t^T]$ . The latter is given by solving the following Lyapunov equation

$$V_{t+1} = (A - G'_t C) V_t (A - G'_t C)^T + (B - G'_t D) (B - G'_t D)^T.$$

Figure 6 shows  $tr(V_t)$  for RKF2 and KF. As expected, KF performs better than RKF2. Indeed, the former has been designed to be optimal for (15).

Finally, we compare the performance of RKF2 and KF using the actual model

$$\xi_{t+1} = A\xi_t + B\varepsilon_t$$
  

$$y_t = \tilde{C}\xi_t + \tilde{D}\varepsilon_t$$
(17)



Fig. 6. Scalar variance of the prediction error for KF and RKF2 with  $c=8\cdot 10^{-2}$  under the assumption that the actual model coincides with the nominal model.

where 0.10.1033 0.0663 0.0295 1.10.10 0.6364-0.63640 0 0 0 0.6364 0.63640 0 0  $\tilde{A} =$ 0 0 0 0.43650.2131 0.15030 0 0 0 0.6364-0.63640 0 0 0 0.63640.63641.1820 0 0 0 0  $\tilde{B} =$ 1.4188 -0.920 0 0 0  $\tilde{C} =$ 1 0 0 -0.0868-0.0557 -0.0247 $\tilde{D} = [-0.2821]$ 1.0956

 $\xi_t = \begin{bmatrix} x_t^T & \eta_t^T \end{bmatrix}^T$ , and  $\varepsilon_t$  is normalized white Gaussian noise. Note that  $x_t$  is the actual state, while  $\eta_t$  is a perturbation process. Model (17) is a perturbed version of (15). Indeed, the green parts in  $\tilde{A}$ ,  $\tilde{B}$ ,  $\tilde{C}$ ,  $\tilde{D}$  correspond to the nonperturbed parts of matrices A, B, C and D; the red parts in  $\tilde{A}$ ,  $\tilde{B}$ ,  $\tilde{C}$ ,  $\tilde{D}$  correspond to the perturbed parts of matrices A, B, C and D; the blue parts are the terms coupling  $x_t$  and  $y_t$  with the perturbation process  $\eta_t$ . Consider a predictor of the form (16). Let  $e_t = x_t - \hat{x}_t$  denote the prediction error of such a predictor under the actual model in (17). Since the submatrices in  $\tilde{A}$  and  $\tilde{C}$  corresponding to A and C are not perturbed, then it is not difficult to see that

$$\tilde{\xi}_{t+1} = (\tilde{A} - \tilde{G}'_t \tilde{C}) \tilde{\xi}_t + (\tilde{B} - \tilde{G}'_t \tilde{D}) \varepsilon_t$$

where  $\xi_t := [e_t^T \ \eta_t^T]^T$ ,  $\tilde{G}'_t = [(G'_t)^T \ 0]^T$ . Thus,  $e_t$  is a process with zero mean and with covariance matrix  $V_t := \mathbb{E}[e_t e_t^T]$ . The latter is given by solving the following Lyapunov equation

$$\Pi_{t+1} = (\tilde{A} - \tilde{G}_t'\tilde{C})\Pi_t (\tilde{A} - \tilde{G}_t'\tilde{C})^T + (\tilde{B} - \tilde{G}_t'\tilde{D})(\tilde{B} - \tilde{G}_t'\tilde{D})^T$$



Fig. 7. Scalar variance of the prediction error for KF and RKF2 with  $c=8\cdot 10^{-2}$  under the assumption that the actual model is different from the nominal one.

and

$$\Pi_t = \left[ \begin{array}{cc} V_t & \star \\ \star & \star \end{array} \right]$$

Figure 7 shows  $tr(V_t)$  for RKF2 and KF. In this case the former performs better than the latter.

## V. CONCLUSIONS

In this paper, we consider a robust static estimation problem in the case that the nominal density is Gaussian and possibly degenerate. We apply such a result to design a robust Kalman filter which can be used also in the case that the reachability assumption does not hold. Of course, there are many aspects that have not taken into account yet. More precisely, in our research agenda now there are the following questions that we are addressing:

- The least favorable model has been assumed to be Gaussian. However, in view of the results in [19], we believe our conclusions also holds in the case that the ambiguity set contains non-Gaussian probability densities.
- The minimax problem also provides the least favorable density  $\tilde{f}_t^0$ . Under the assumption that  $f_t$  and  $\tilde{f}_t$  are non-degenerate, Levy & Nikoukhah showed that it is possible to characterize the corresponding dynamic model in a finite simulation horizon, [13]. Our conjecture is that it is possible to adapt these arguments to our case.
- The simulation results show that the robust Kalman filter seems to converge in the case of constant parameters. Drawing inspiration from the non-degenerate case in [26], [27], we believe it is possible to prove that the risk sensitive-like Riccati iteration in Algorithm 1 (steps 3 and 6) converges provided that the tolerance is sufficiently small and the system is stabilizable and detectable.

#### REFERENCES

- B. Hassibi, A. Sayed, and T. Kailath, Indefinite-Quadratic Estimation and Control−A Unified Approach to H<sup>2</sup> and H<sup>∞</sup> Theories. Philadelphia: Soc. Indust. Appl. Math., 1999.
- [2] J. Speyer and W. Ching, Stochastic Processes, Estimation, and Control. Advances in Design and Control, Philadelphia: Soc. Indust. Appl. Math., 2008.
- [3] L. El Ghaoui and G. Calafiore, "Robust filtering for discrete-time systems with bounded noise and parametric uncertainty," *IEEE Trans. Automat. Contr.*, vol. 46, no. 7, pp. 1084–1089, 2001.
- [4] S. Kim, V. Deshpande, and R. Bhattacharya, "Robust kalman filtering with probabilistic uncertainty in system parameters," *IEEE Contr. Syst. Lett.*, vol. 5, no. 1, pp. 295–300, 2020.
- [5] B. Li, Y. Tan, L. Zhou, and R. Dong, "Robust-nonsmooth Kalman filtering for stochastic sandwich systems with dead-zone," *Int J Control Autom Syst*, pp. 1–11, 2020.
- [6] P. Whittle, *Risk-sensitive Optimal Control*. Chichester, England: J. Wiley, 1980.
- [7] R. Banavar and J. Speyer, "Properties of risk-sensitive filters/estimators," *IEE Proc.-Control Theory Appl.*, vol. 145, no. 1, pp. 106–112, 1998.
- [8] B. Levy and M. Zorzi, "A contraction analysis of the convergence of risk-sensitive filters," *SIAM J Control Optim*, vol. 54, no. 4, pp. 2154– 2173, 2016.
- [9] J. Huang, D. Shi, and T. Chen, "Distributed robust state estimation for sensor networks: A risk-sensitive approach," in 2018 IEEE Conference on Decision and Control (CDC), pp. 6378–6383, 2018.
- [10] R. Boel, M. James, and I. Petersen, "Robustness and risk-sensitive filtering," *IEEE Trans. Automat. Contr.*, vol. 47, no. 3, pp. 451–461, 2002.
- [11] L. Hansen and T. Sargent, "Robust estimation and control under commitment," J Econ Theory, vol. 124, no. 2, pp. 258–301, 2005.
- [12] M. Yoon, V. Ugrinovskii, and I. Petersen, "Robust finite horizon minimax filtering for discrete-time stochastic uncertain systems," *Syst. Control. Lett.*, vol. 52, no. 2, pp. 99–112, 2004.
- [13] B. Levy and R. Nikoukhah, "Robust state-space filtering under incremental model perturbations subject to a relative entropy tolerance," *IEEE Trans. Automat. Contr.*, vol. 58, pp. 682–695, Mar. 2013.
- [14] M. Zorzi and B. Levy, "Robust kalman filtering: Asymptotic analysis of the least favorable model," in 57th IEEE Conference on Decision and Control (CDC), pp. 7124–7129, Dec 2018.
- [15] S. Abadeh, V. Nguyen, D. Kuhn, and P. Esfahani, "Wasserstein distributionally robust kalman filtering," in *Advances in Neural Information Processing Systems*, pp. 8474–8483, 2018.
- [16] M. Zorzi, "Distributed kalman filtering under model uncertainty," IEEE Trans. Control. Netw. Syst., vol. 7, no. 2, pp. 990–1001, 2019.
- [17] A. Emanuele, F. Gasparotto, G. Guerra, and M. Zorzi, "Robust distributed Kalman filtering: On the choice of the local tolerance," *Sensors*, vol. 20, no. 11, p. 3244, 2020.
- [18] A. Zenere and M. Zorzi, "On the coupling of model predictive control and robust Kalman filtering," *IET Control. Theory Appl.*, vol. 12, no. 13, pp. 1873–1881, 2018.
- [19] B. Levy and R. Nikoukhah, "Robust least-squares estimation with a relative entropy constraint," *IEEE Trans. Informat. Theory*, vol. 50, no. 1, pp. 89–104, 2004.
- [20] M. Zorzi, "Robust kalman filtering under model perturbations," *IEEE Trans. Automat. Control*, vol. 62, no. 6, pp. 2902–2907, 2016.
- [21] M. Zorzi, "On the robustness of the Bayes and Wiener estimators under model uncertainty," *Automatica*, vol. 83, pp. 133–140, 2017.
- [22] S. Bonnabel and R. Sepulchre, "The geometry of low-rank Kalman filters," in *Matrix Information Geometry*, pp. 53–68, Springer, 2013.
- [23] A. Ferrante and L. Ntogramatzidis, "The generalized continuous algebraic riccati equation and impulse-free continuous-time LQ optimal control," *Automatica*, vol. 50, no. 4, pp. 1176–1180, 2014.
- [24] A. Ferrante and L. Ntogramatzidis, "The generalised discrete algebraic riccati equation in linear-quadratic optimal control," *Automatica*, vol. 49, no. 2, pp. 471–478, 2013.
- [25] L. Hansen and T. Sargent, *Robustness*. Princeton, NJ: Princeton University Press, 2008.
- [26] M. Zorzi and B. C. Levy, "On the convergence of a risk sensitive like filter," in 54th IEEE Conference on Decision and Control (CDC), pp. 4990–4995, Dec 2015.
- [27] M. Zorzi, "Convergence analysis of a family of robust Kalman filters based on the contraction principle," *SIAM J Control Optim*, vol. 55, no. 5, pp. 3116–3131, 2017.