Robust Kalman Filtering: Asymptotic Analysis of the Least Favorable Model

Mattia Zorzi, Bernard C. Levy

Abstract—We consider a robust filtering problem where the robust filter is designed according to the least favorable model belonging to a ball about the nominal model. In this approach, the ball radius specifies the modeling error tolerance and the least favorable model is computed by performing a Riccati-like backward recursion. We show that this recursion converges provided that the tolerance is sufficiently small.

I. INTRODUCTION

Consider the problem of estimating a state process whose state-space model is known only imperfectly. In such a situation the standard Kalman filter may perform poorly. Robust filtering seeks to find a state estimate which takes the model uncertainty into account.

In this paper, we consider the robust filtering approach proposed in [12], see also [11], [8]. The actual state-space model is assumed to belong to a ball centered about the nominal state-space model. In this approach, the ball radius specifies the modeling error tolerance and the least favorable model is computed by performing a Riccati-like backward recursion. We show that this recursion converges provided that the tolerance is sufficiently small.

II. ROBUST KALMAN FILTERING

Consider a nominal state-space model of the form

\[
\begin{align*}
x_{t+1} &= Ax_t + Bu_t \\
y_t &= Cx_t + Du_t
\end{align*}
\]

where \(x_t \in \mathbb{R}^n\) is the state process, \(y_t \in \mathbb{R}^p\) the observation process, \(v_t \in \mathbb{R}^m\) a white Gaussian noise (WGN) with unit variance, i.e. \(E[v_tv_T^*] = I\delta_{t-s}\) and \(\delta_t\) denotes the Kronecker delta function. We assume that \(v_t\) is independent of the initial state vector \(x_0 \sim \mathcal{N}(0, P_0)\), and that the pairs \((A, B)\) and \((A, C)\) are reachable and observable, respectively. Without loss of generality, we assume that \(BD^T = 0\). Indeed, if this is not the case we can always rewrite \(\textbf{(1)}\) with \(\tilde{A} = A - ABD^T(DD^T)^{-1}C\), \(\tilde{B}\) such that \(\tilde{B}B^T = B(I - D^T(DD^T)^{-1}D)B^T\), \(\tilde{C} = C\) and \(\tilde{D} = D\). The nominal model \(\textbf{(1)}\) is completely characterized by the transition probability density \(\phi_t(x_{t+1}, y_t|x_t)\) and by the probability density \(f(x_0)\) of \(x_0\). Let \(\tilde{\phi}_t(x_{t+1}, y_t|x_t)\) denote the transition probability density of the actual model. We assume that the actual and nominal densities of initial state \(x_0\) coincide, whereas \(\tilde{\phi}_t\) belongs to a ball centered about...
\[ \phi_t \text{ with radius } c > 0, \text{ hereafter called tolerance, which is specified by} \]
\[ B_t = \{ \phi_t \text{ s.t. } D_{KL}(\phi_t, \tilde{\phi}_t) \leq c \}. \]

Here \( D_{KL} \) denotes the Kullback-Leibler divergence [10] between \( \phi_t \) and \( \tilde{\phi}_t \). Note that \( D_{KL}(\phi_t, \tilde{\phi}_t) \) is finite only if matrix \( [B^T D^T]^T \) has full row rank. Accordingly, without loss of generality we assume that \( [B^T D^T]^T \) is square and invertible, so that \( m = n + p \). Indeed it is always possible to compress the column space of this matrix and remove the noises which do not affect model (1). Let \( Y_t = \{ y_s, \ s \leq t \} \) and \( g_t(Y_t) \) be an estimator of \( x_{t+1} \) given \( Y_t \). Adopting the minimax approach described in [12], a robust estimator of \( x_{t+1} \) is obtained by solving:

\[ \hat{x}_{t+1} = \text{argmin}_{\tilde{g}_t \in G_t} \max_{\phi_t \in B_t} \tilde{E}[||\tilde{x}_{t+1} - g_t(Y_t)||^2|Y_{t-1}] \]  

where \( \tilde{E} \) denotes the expectation operator taken with respect to the joint probability density of the actual model and \( G_t \) denotes the class of estimators with finite second-order moments with respect to \( \tilde{\phi}_t \in B_t \). In [12], it was shown that the robust estimator satisfies a Kalman-like recursion of the form:

\[ G_t = AV_t C^T(CV_t C^T + DD^T)^{-1} \]
\[ \hat{x}_{t+1} = A\hat{x}_t + G_t (y_t - C\hat{x}_t) \]
\[ P_{t+1} = A(V_t^{-1} + C^T(DD^T)^{-1}C)^{-1}A^T + BB^T \]
\[ V_{t+1} = (P_{t+1}^{-1} - \theta_t I)^{-1} \]

where \( \theta_t > 0 \) is the unique solution to the equation \( c = \gamma(P_{t+1}, \theta_t) \). The function \( \gamma \) is given by

\[ \gamma(\theta, P) = \frac{1}{2} \left[ \log \det(I - \theta P) + \text{tr}[(I - \theta P)^{-1}] - n \right]. \]

The initial conditions of the recursion are \( \hat{x}_0 = 0 \) and \( V_0 = P_0 \). The least favorable prediction error \( e_t = x_t - \hat{x}_t \) of the robust estimator has zero mean and covariance matrix \( V_t \).

The following result is proved in [21, Proposition 3.5], see also [19].

**Proposition 2.1:** There exists \( c_{MAX} > 0 \) such that if \( c \in (0, c_{MAX}) \), then for any \( P_0 > 0 \) the sequence \( P_t, t \geq 0 \), generated by (4) converges to a unique solution \( P > 0 \), \( \theta_t \to \theta \) with \( \theta > 0 \), \( V_t \to V \) with \( V > 0 \) and the limit \( G_t \) of the filtering gain \( G_t \) as \( t \to \infty \) is such that \( A - GC \) is stable. Moreover, \( P \) is the unique solution of the algebraic Riccati-like equation:

\[ P = A(P^{-1} - \theta I + C^T(DD^T)^{-1}C)A^T + BB^T. \]

It is possible to show that the least favorable model obtained by solving (3) is given by [12]

\[ \xi_{t+1} = \tilde{A}_t \xi_t + \tilde{B}_t \varepsilon_t \]
\[ y_t = \tilde{C}_t \xi_t + \tilde{D}_t \varepsilon_t \]

where

\[ \tilde{A}_t = \begin{bmatrix} A & BH_t \\ 0 & A - G_tC + (B - G_tD)H_t \end{bmatrix} \]
\[ \tilde{B}_t = \begin{bmatrix} B \\ B - G_tD \end{bmatrix} L_t \]
\[ \tilde{C}_t = \begin{bmatrix} C \\ DH_t \end{bmatrix}, \tilde{D}_t = DL_t \]
\[ H_t = \tilde{K}_t(B - G_tD)^T(\Omega_{t+1}^{-1} + \theta_t I)(A - G_tC) \]
\[ \tilde{K}_t = [I - (B - G_tD)^T(\Omega_{t+1}^{-1} + \theta_t I)(B - G_tD)]^{-1} \]

and \( L_t \) is such that \( \tilde{K}_t = L_t L_t^T \). In this model \( \varepsilon_t \) is a WGN with unit variance, and \( \Omega_{t+1}^{-1} \) is computed by the backward recursion:

\[ \Omega_t^{-1} = (A - G_tC)^T(\Omega_{t+1}^{-1} + \theta_t I)^{-1} - (B - G_tD) \times \times (B - G_tD)^T]^{-1}(A - G_tC) \]

where if \( T \) denotes the simulation horizon, the initial condition is \( \Omega_T^{-1} = 0 \).

In summary, the least favorable model (7) is obtained in two steps:

1) The Riccati equation (4) for \( P_t \) is propagated forward in time over \([0, T]\) and used to compute \( G_t \) and \( \theta_t \).

2) The model \((\tilde{A}_t, \tilde{B}_t, \tilde{C}_t, \tilde{D}_t)\) is obtained by propagating (9) backward in time to evaluate \( \Omega_t^{-1} \) over interval \([0, T]\).

It is clear that the least favorable model depends on the length \( T \) of the simulation interval. Let \( \alpha, \beta > 0, \alpha < \beta < 1 \). Then, the interval \([\alpha T, \beta T]\) is contained in \([0, T]\) and in the next section we show that when \( c > 0 \) is sufficient small, then \( \Omega_t^{-1} \) converges over the interval \([\alpha T, \beta T]\) as \( T \) tends to infinity. As a consequence, the least favorable model (7) is constant over this interval.

Before establishing the convergence of the backward recursion (9), it is worth considering the limit case \( c = 0 \) when the nominal and the actual models coincide. In this case, the robust filter (4) reduces to the usual Kalman filter and \( \theta_t = 0 \) for all \( t \). Hence the limit of \( \theta_t = 0 \). By using the matrix inversion lemma, the backward recursion (9) with \( \theta_t = 0 \) can be rewritten as

\[ \Omega_t^{-1} = (A - G_tC)^T[\Omega_{t+1}^{-1} + \Omega_{t+1}^{-1} - (B - G_tD) \times \times (B - G_tD)^T]^{-1}(A - G_tC) \]

where

\[ S_t = [(B - G_tD)^T \Omega_{t+1}^{-1}(B - G_tD) - I]^{-1}. \]

Therefore, if \( \Omega_{t+1}^{-1} = 0 \) then \( \Omega_t^{-1} = 0 \). Since \( \Omega_T^{-1} = 0 \), we conclude that \( \Omega_t^{-1} = 0 \) for all \( t \in [0, T] \). Accordingly, \( H_t = 0 \) and \( L_t = I \). Substituting these expressions inside (8), it is then easy to verify that the least favorable model coincides with the nominal model.

**III. CONVERGENCE OF THE BACKWARD RECURSION**

Suppose that the condition of Proposition 2.1 is satisfied. Then as \( t \to \infty \) the backward recursion (9) becomes

\[ \Omega_t^{-1} = \tilde{A}_t^T[\Omega_{t+1}^{-1} + \theta_t I]^{-1} - \tilde{B}_t \tilde{B}_t^T]^{-1} \tilde{A}_t \]

\[ \]
where the matrix $\tilde{A} := A - GC$ is stable, and $\tilde{B} := B - GD$. To ease the exposition, we assume that $T$ is finite and we study the convergence of \( [10] \) as $t$ tends to $-\infty$. This is equivalent to studying the convergence in $[\alpha T, \beta T]$ as $T$ tends to $0$. Adding $\theta I$ on both sides and defining $X_t := \Omega_t^{-1} + \theta I$ yields the equivalent recursion

$$X_t = \tilde{A}^T(X_t - \tilde{B}B^T)^{-1}\tilde{A} + \theta I$$

with terminal value $X_T = \theta I$. It has the form of a Riccati equation, but an important difference, compared to the standard case, is that in the inverse we add to $X_T^{-1}$ the negative definite matrix $-\tilde{B}B^T$. This difference makes the convergence analysis nontrivial. At this point, it is useful to introduce the following map defined for $0 < X < (BB^T)^{-1}$

$$\Theta(X) := \tilde{A}^T(X^{-1} - \tilde{B}B^T)^{-1}\tilde{A} + \theta I$$

(12)

Note that $BB^T$ is an invertible matrix since

$$\tilde{B}B^T = (B - GD)(B - GD)^T = BB^T + GDD^TG^T \geq BB^T$$

(13)

where $BB^T$ is invertible because $B \in \mathbb{R}^{n \times n + p}$ has full row-rank. Accordingly, the recursion (11) can be rewritten as

$$X_t = \Theta(X_{t+1}).$$

(14)

**Proposition 3.1:** For any $0 < X < (\tilde{B}B^T)^{-1}$, we have $\Theta(X) \geq \theta I$.

**Proof:** We have

$$\Theta(X) - \theta I = \tilde{A}^T(X^{-1} - \tilde{B}B^T)^{-1}\tilde{A}$$

(15)

where the right hand side is positive semi-definite. □

**Proposition 3.2:** The map $\Theta$ preserves the partial order of positive semi-definite matrices, so if $X_1, X_2$ are such that $0 < X_1 \leq X_2 < (BB^T)^{-1}$, we have

$$\Theta(X_1) \leq \Theta(X_2).$$

**Proof:** The first variation of $\Theta(X)$ along the direction $\delta X \in Q_n$ can be expressed as

$$\delta \Theta(X; \delta X) = \tilde{A}^T(X^{-1} - \tilde{B}B^T)^{-1}X^{-1}\delta X \times X^{-1}(X^{-1} - \tilde{B}B^T)^{-1}\tilde{A}.$$  

(16)

Thus $\delta \Theta(X; \delta X) \geq 0$ for any $\delta X \geq 0$, so the map is nondecreasing. □

Before stating the next property of $\Theta$, we prove the following lemmas.

**Lemma 3.1:** It is always possible to select $c \in (0, c_{MAX}]$ such that $\theta$ is arbitrarily small.

**Proof:** In [21, 19] it was shown that

$$\gamma(P, \theta_1) > \gamma(P, \theta_2), \quad \forall \theta_1 > \theta_2 \text{ s.t. } P \geq 0, \ P \neq 0$$

(17)

$$\gamma(P_1, \theta) \geq \gamma(P_2, \theta), \quad \forall P_1 \geq P_2$$

(18)

$$\gamma(P, 0) = 0, \quad \forall P \geq 0$$

(19)

$$\gamma(P, [0, \sigma(P)^{-1}]) = [0, \infty), \quad \forall P > 0$$

(20)

where (20) means that the image of $[0, \sigma(P)^{-1}]$ under $\gamma(P, \cdot)$ is $[0, \infty]$. Since $c \in (0, c_{MAX}]$, by Proposition 2.1, we have that $P_t \to P$, $c_t \to c$, $\theta_t \to \theta$ where $c$ and $\theta$ are related by $\sigma = \gamma(P, \theta)$. Here $P$ solves the algebraic form of Riccati equation \( (3) \), so $P \geq BB^T$. In view of (17)-(20) it follows that $\theta \leq \theta$ when $\theta$ is the unique solution of equation $c = \gamma(BB^T, \theta)$. Furthermore, the map

$$\mu : [0, \sigma(BB^T)^{-1}) \to [0, \infty)$$

$$\theta \mapsto \gamma(BB^T, \theta)$$

(21)

is injective and continuous. Accordingly, the inverse map $\mu^{-1} : [0, \infty) \to [0, \sigma(BB^T)^{-1})$ exists and is continuous, in particular $\mu^{-1}(0) = 0$. This means that we can always select $c > 0$ such that $\theta$ is arbitrarily small. Since $\theta \leq \theta$, the statement follows.

It is worth noting that $\tilde{A}$ and $\tilde{B}$ depend on $c$ through $\theta$.

Throughout the paper we make the following assumption.

**Assumption 1:** The map

$$\gamma : [0, \theta] \to \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n}$$

$$\theta \mapsto (\tilde{A}, \tilde{B})$$

(22)

is continuous for $\tilde{\theta}$ sufficiently small.

Even though Assumption 1 may appear restrictive, it holds under mild conditions on system $(A, B, C, D)$. Indeed, for $c \in (0, c_{MAX}]$ the unique solution of (6) is $P = XY^{-1}$ where $[X^TY^T]^T$ spans the stable deflating subspace of regular matrix pencil $sL - M$ [14], where

$$L = \begin{bmatrix} A^T & 0 \\ -BB^T & I \end{bmatrix}, \quad M = \begin{bmatrix} I & C^T(DD^T)^{-1}C - \theta I \\ 0 & A \end{bmatrix}.$$  

Conditions for the continuity of such subspaces are given in [6]. Accordingly, the map $\theta \mapsto P$ is continuous over $[0, \tilde{\theta}]$ with $\tilde{\theta}$ small enough. Since the map $P \mapsto (\tilde{A}, \tilde{B})$ is continuous, we conclude that $\gamma$ is continuous for $\tilde{\theta}$ sufficiently small.

**Lemma 3.2:** For $c \in (0, c_{MAX}]$ sufficiently small, there exists $\rho \in (1, \sigma(\tilde{A})^{-1})$ such that

$$\gamma(P, \theta) > \gamma(P, \bar{\theta}), \quad \forall \theta > \bar{\theta} \text{ s.t. } P \geq 0, \ P \neq 0$$

(23)

where $\Sigma_\rho$ is the unique solution of the algebraic Lyapunov equation (ALE)

$$\Sigma_\rho = \rho^2\tilde{A}^T\Sigma_\rho\tilde{A} + \theta I.$$  

(24)

**Proof:** First, note that $\rho\tilde{A}$ is a stable matrix. Then, the solution of (24) is given by

$$\Sigma_\rho = \theta \sum_{k \geq 0} \rho^{2k}(\tilde{A})^k\tilde{A}^k$$

(25)

which is positive definite. Note that

$$\Sigma_\rho \leq \theta \sum_{k \geq 0} \rho^{2k}\sigma(\tilde{A})^{2k}I = \frac{\theta}{1 - \rho^2\sigma(\tilde{A})^2}I$$

and thus $\Sigma_\rho^{-1} \geq (1 - \rho^2\sigma(\tilde{A})^2)/\theta I$. In view of Assumption 1, for $\theta$ sufficiently small we have

$$\sigma(\tilde{A})^2 = \sigma(\tilde{A}_0)^2 + o(1)$$
where \( \tilde{A}_0 = A - G_0 C \), \( G_0 = AP(0)C^T(CP(0)C^T + D\Omega)^{-1} \) and \( P(0) \) is the unique solution of \( (6) \) with \( \theta = 0 \).

As a consequence,
\[
(1 - \rho^2)\Sigma^{-1} \geq (1 - \rho^2)\left(1 - \frac{\rho^2(\sigma(\tilde{A}_0) + o(1))^2}{\theta}\right)I. \tag{26}
\]

We can always choose \( \rho \) in the range \( (1, \sigma(\tilde{A}_0)^{-1}) \) such that \( (1 - \rho^2)(1 - \rho^2\sigma^2(\tilde{A}_0)) \) is positive. By Lemma 3.1, we can also select \( c \in [0, c_{\text{MAX}}] \) sufficiently small so that \( \theta \) is small enough that the scaled identity matrix on the right hand side of (26) upper bounds \( BB^T \).

Let \( \tilde{c} \in (0, c_{\text{MAX}}] \) be a value of \( c \) such that Lemma 3.2 is satisfied, so that (23) holds for a certain \( \rho \) and \( \theta \). Then it is useful to observe that for any \( c \in (0, \tilde{c}] \), the equation (23) still holds with the same value for \( \rho \) but with a smaller value for \( \theta \).

**Corollary 3.1:** For any \( c \in (0, \tilde{c}] \), we have \( \Sigma_{\rho} < (BB^T)^{-1} \).

**Proof:** Since (23) holds for a suitable \( \rho > 0 \), we have
\[
\Sigma_{\rho}^{-1} \geq \rho^{-2}\Sigma_{\rho}^{-1} + \tilde{B}B^T > \bar{B}\bar{B}^T
\]
which implies \( \Sigma_{\rho} < (BB^T)^{-1} \).

We are now ready to state the third property of the map \( \Theta \).

**Proposition 3.3:** Consider the compact set
\[
C = \{ X \in \mathbb{Q}_n \text{ s.t. } \theta I \leq X \leq \Sigma_{\rho} \}
\]
where \( \Sigma_{\rho} \) is computed as in Lemma 3.2. If \( c \in (0, \tilde{c}] \) then \( \Theta(X) \in C \) for any \( X \in C \).

**Proof:** First, observe that \( C \) is a nonempty set. Indeed, by (25), we have \( \Sigma_{\rho} \geq \theta I \), so that \( \theta I \in C \). Since \( c \in (0, \tilde{c}] \), by Lemma 3.2 the inequality (23) holds for some \( \rho \in (1, \sigma(A)^{-1}) \), and thus
\[
\Sigma_{\rho}^{-1} - \tilde{B}B^T \geq \rho^{-2}\Sigma_{\rho}^{-1} \\
(\Sigma_{\rho}^{-1} - \tilde{B}B^T)^{-1} \leq \rho^2\Sigma_{\rho} \\
\bar{A}^T(\Sigma_{\rho}^{-1} - \tilde{B}B^T)^{-1}\bar{A} + \theta I \leq \rho^2\bar{A}^T\Sigma_{\rho}\bar{A} + \theta I \\
\Theta(\Sigma_{\rho}) \leq \Sigma_{\rho}.
\tag{27}
\]
Assume that \( X \in C \). Since \( X \leq \Sigma_{\rho} \), the nondecreasing property of \( \Theta \) and (27) imply
\[
\Theta(X) \leq \Theta(\Sigma_{\rho}) \leq \Sigma_{\rho}.
\]
Since \( X \geq \theta I \), we have
\[
\Theta(X) \geq \Theta(\theta I) \geq \theta I
\tag{28}
\]
where we exploited again the nondecreasing property of \( \Theta \) and Proposition 3.1. We conclude that \( \Theta(X) \in C \).

**Proposition 3.4:** Consider the sequence \( X_t \) satisfying the backward recursion
\[
X_t = \Theta(X_{t+1}), \quad X_T = \theta I.
\tag{29}
\]
For \( c \in (0, \tilde{c}] \), the sequence belongs to \( C \) and is nondecreasing. Thus as \( t \to -\infty \), \( X_t \) converges to \( X \in C \) which is a solution of the algebraic Riccati equation
\[
X = \bar{A}^T(X^{-1} - \tilde{B}B^T)^{-1}\bar{A} + \theta I.
\tag{30}
\]

**Proof:** We prove the first two statements by induction. We start by showing that \( X_t \in C \) for any \( t \). We know that \( X_T \in C \) because \( C \) contains \( \theta I \). Assume that \( X_{t+1} \in C \), then Proposition 3.3 implies that \( X_t = \Theta(X_{t+1}) \in C \). This proves the first claim.

Next we show that the sequence is nondecreasing. We observe that
\[
X_{T-1} = \Theta(X_T) = \Theta(\theta I) \geq \theta I = X_T
\tag{31}
\]
where we exploited the nondecreasing property of \( \Theta, \) see Propositions 3.2 and 3.1. Assume that \( X_t \geq X_{t+1} \), then
\[
X_{t-1} = \Theta(X_t) \geq \Theta(X_{t+1}) = X_t,
\tag{32}
\]
so by induction the sequence is nondecreasing.

The convergence follows from the fact that the sequence is nondecreasing and belongs to a compact set.

Since \( X_t = \Omega^{-1} + \theta I \), we have the following result.

**Convolary 3.2:** For \( c \in (0, \tilde{c}] \), the sequence \( \Omega_{t}^{-1} \) generated by (10) converges to \( \Omega^{-1} \) as \( t \to -\infty \) where \( \Omega^{-1} \) is such that \( 0 \leq \Omega^{-1} \leq \Sigma_{\rho} - \theta I \) for some \( \rho \in (1, \sigma(A)^{-1}) \) satisfying (23). Furthermore
\[
H_t \to H, \quad \bar{K}_t \to \bar{K}, \quad L_t \to L \\
\bar{A}_t \to \bar{A}, \quad \bar{B}_t \to \bar{B} \\
\bar{C}_t \to \bar{C}, \quad \bar{D}_t \to \bar{D}.
\tag{33}
\]

It is worth noting that the algebraic equation (30) may admit several positive definite solutions. Indeed, in the scalar case, equation (30) becomes
\[
x = \frac{a^2}{x - b^2} + \theta
\tag{34}
\]
or equivalently
\[
\tilde{b}x^2 - (1 - \tilde{a}^2 + \tilde{b}^2)\theta x + \theta = 0.
\]
For small \( \theta > 0 \), the discriminant of this equation is positive, so the equation has two positive real solutions since the coefficient \( 1 - \tilde{a}^2 - \tilde{b}^2 \theta \) is positive. For \( \tilde{a} = 0.1 \), \( \tilde{b} = 1 \) and \( \theta = 0.1 \) we obtain the two solutions \( x_1 \approx 0.99 \) and \( x_2 \approx 0.10 \). It is not difficult to see that (34) can be rewritten as a Lyapunov equation
\[
x = (a - j\tilde{b})^2 x + j^2
\tag{35}
\]
where \( j = \tilde{a}x\tilde{b} / (\tilde{b}^2 x - 1) \). Let \( f_1 := a - j\tilde{b} \) be the “feedback” matrix and \( f_1, f_2 \) denote the values corresponding to \( x_1 \) and \( x_2 \), respectively. Then we have \( f_1 \approx 8.9 \) and \( f_2 \approx 0.11 \). In view of (35), this means that \( x_1 \) is a stabilizing solution of (11) whereas \( x_2 \) corresponds to an unstable one. Accordingly, the limit of the sequence (29) is \( x_2 \). In the general case (i.e., for \( n > 1 \)) the algebraic Riccati equation (30) can be rewritten as
\[
X = (\tilde{A} - \tilde{B}J)^T X (\tilde{A} - \tilde{B}J^T) + \tilde{B}B^T - JJ^T
\]
where \( J = \bar{A}^T X B (\bar{B}X B^T - I)^{-1} \). However, the reasoning used in the scalar case cannot be applied since the matrix \( BB^T - JJ^T \) is indefinite.
Proposition 3.5: For \( c \in (0, \bar{c}) \) sufficiently small, the limit \( X \) of \( \Theta^t \) is a stabilizing solution of \( \Theta^t \) in the sense that the matrix \( \Theta^T - J \Theta^T \) is stable.

Proof: Let \( X_\theta \) be the limit of the sequence in \( \Theta^t \) where we made explicit its dependence on \( \theta \). Notice that \( \rho \) does not depend on \( \theta \). Indeed, if a certain \( \rho \) satisfies \( \Theta^T \) for a given \( \theta \), then the same \( \rho \) satisfies \( \Theta^T \) with \( \theta' \) such that \( 0 < \theta' \leq \theta \). Since \( X_\theta \in C \), we have that \( \theta I \leq X_\theta \leq \theta \sum_{k=0}^p \rho^k (A^T)^k A^k \). Let \( Q_\theta \) be such that \( X_\theta = \theta Q_\theta \). Hence \( Q_\theta \geq I \). Observe that

\[
M_\theta := \Theta^T - J \Theta^T \\
= \Theta^T [X_\theta - X_\theta B (B X_\theta B^T - I)^{-1} B^T X_\theta] X_\theta^{-1} \\
= \Theta^T (X_\theta^{-1} - B B^T)^{-1} X_\theta^{-1} \\
= \Theta^T (\theta^{-1} Q_\theta^{-1} - B B^T)^{-1} \theta^{-1} Q_\theta^{-1} \\
= \Theta^T (Q_\theta^{-1} - \theta B B^T)^{-1} Q_\theta^{-1}.
\]

(36)

For \( \theta \) sufficiently small, by Assumption 1 we have \( B B^T = B_0 B_0^T + o(1) \) where \( B_0 = B - G_0 D \) and \( G_0 \) has been defined in the proof of Lemma 4.1. Accordingly,

\[
(Q_\theta^{-1} - \theta B B^T)^{-1} = Q_\theta + o(1),
\]

which after substitution inside \( \Theta^t \) gives

\[
M_\theta = \Theta^T + o(1).
\]

(38)

The map \( \theta \mapsto \lambda(M_\theta) \) is a continuous function for \( \theta > 0 \) since the mapping from the entries of a matrix to its spectrum is continuous. Hence for \( \theta \) sufficiently small, the matrix \( M_\theta \) is stable. By Lemma 4.1 we conclude that if we select \( c \in (0, c_{MAX}) \) sufficiently small, the matrix \( M_\theta \) will be stable.

IV. PERFORMANCE ANALYSIS

We want to evaluate the performance of an arbitrary estimator

\[
\hat{e}_t' = A \hat{e}_t' + G_t (y_t - C \hat{e}_t')
\]

(39)

under the least favorable model \( \hat{e}_t' \) in steady state, i.e. with \( \hat{A}_t, \hat{B}_t, \hat{C}_t \) and \( \hat{D}_t \) constant. Note that the steady state condition is guaranteed under the assumption that \( c \in (0, \bar{c}) \). Recall that \( e_t' \) denotes the least favorable prediction error of the robust filter \( \hat{e}_t' \). Let \( e_t' = x_t - \hat{e}_t' \) be the prediction error of filter \( \hat{e}_t' \). Let \( e_t = [e_t', e_t']^T \). In [12] it was shown that the dynamics of \( e_t \) are given by

\[
e_{t+1} = F_t e_t + M_t e_t
\]

(40)

where

\[
F_t = \hat{A} - \left[ \begin{array}{cc} G_t' & 0 \\ 0 & 0 \end{array} \right] \hat{C}, \quad M_t = \hat{B} - \left[ \begin{array}{cc} G_t' & 0 \\ 0 & 0 \end{array} \right] \hat{D}
\]

and \( e_t \) is a WGN with unit variance. Then the covariance matrix \( \Pi_t \) of \( e_t \) obeys the Lyapunov equation

\[
\Pi_{t+1} = F_t \Pi_t + \Pi_t M_t^T
\]

(41)

with initial condition \( \Pi_0 = I_2 \otimes V_0 \).

From \( \Pi_t \) it is clear that the mean of the prediction error \( e_t' \) is zero. Next, we show that the covariance matrix of \( e_t' \) converges to a constant matrix and is bounded provided that \( c \) is sufficiently small. To do so, we use the following result [5, Theorem 1].

Lemma 4.1: Consider the time-varying Lyapunov equation

\[
Y_{t+1} = F_t Y_t F_t^T + \mathcal{R}_t
\]

where \( F_t \) and \( \mathcal{R}_t \) converges to \( F \) and \( \mathcal{R} \), respectively, as \( t \to \infty \) with \( F \) stable. Then \( Y_t \) converges to the unique solution \( Y \) of the Lyapunov equation:

\[
Y = FYF^T + \mathcal{R}.
\]

Proposition 4.1: Assume that the gain \( G_t' \) in \( \hat{e}_t' \) converges to a matrix \( G' \) such that \( A - G'C \) is stable. Then, for \( c \in (0, \bar{c}) \) sufficiently small the recursion \( \Pi_t \) converges to the solution II of the Lyapunov equation

\[
\Pi = F \Pi F^T + MM^T
\]

where

\[
F := \hat{A} - \left[ \begin{array}{cc} G' & 0 \\ 0 & 0 \end{array} \right], \quad M := \left[ \begin{array}{cc} \hat{B} - \left[ \begin{array}{cc} G' & 0 \\ 0 & 0 \end{array} \right] \hat{D} \end{array} \right].
\]

(42)

Proof: First, we prove that the matrix

\[
F = \left[ \begin{array}{cc} A - G'C & \left( B - G'D \right) H \\ 0 & A - GC + \left( B - GD \right) H \end{array} \right]
\]

is stable. Since \( F \) is an upper block-triangular matrix, it is sufficient to show that its two diagonal blocks are stable. The matrix \( A - G'C \) is stable by assumption. Next, by recalling that \( \hat{A} = A - GC \), \( \hat{B} = B - GD \), \( \hat{H} = \hat{K} \hat{B}^T X \hat{A} \) and \( \hat{K} = (I - \hat{B}^T X \hat{B})^{-1} \), the (2,2) block of \( F \) can be expressed as

\[
\hat{A} + \hat{B} (I - \hat{B} X \hat{B}^T)^{-1} \hat{B}^T X \hat{A}
\]

\[
= \hat{A} - \hat{B} (X \hat{B}^T - I)^{-1} \hat{B}^T X \hat{A}
\]

\[
= \hat{A} - \hat{B} J^T
\]

(43)

which has the same eigenvalues of \( \hat{A}^T - J \hat{A}^T \). By Proposition 4.1 this matrix is stable provided that \( c \) is sufficiently small. The conditions of Proposition 4.1 are satisfied since \( F_t \) converges to \( F \) with \( F \) stable and \( M_t \) converges to \( M \), and thus \( M_t M_t^T \) converges to \( MM^T \) as \( t \to \infty \). Hence \( \Pi_t \) converges to \( \Pi \).

Corollary 4.1: Under the assumption that \( c \in (0, \bar{c}) \) is sufficiently small, the prediction error \( e_t' \) of the filter \( \hat{e}_t' \) under the least favorable model (in steady state) has zero mean and bounded variance.

V. SIMULATION EXAMPLE

Consider the state-space model

\[
A = \left[ \begin{array}{cc} 0.1 & 1 \\ 0 & 1.2 \end{array} \right], \quad B = 0.01 I_2 \\
C = \left[ \begin{array}{cc} 1 & -1 \end{array} \right], \quad D = 0.04.
\]

(44)
Note that the pairs \((A, B)\) and \((A, C)\) are reachable and observable, respectively. Using the procedure of [21, Proposition 3.5], it results that the robust filter (4) converges for \(c = c_{\text{MAX}}\), with \(c_{\text{MAX}} = 0.1879\).

The minimum eigenvalue of \((1 - \rho^{-2})\Sigma_{\rho}^{-1} - \bar{B}\bar{B}^T\) is depicted in Fig. 1 as a function of \(\rho\) for \(c = c_{\text{MAX}}\). For \(c = c_{\text{MAX}}\), we see that when \(\rho = 1.382\), the minimum eigenvalue is \(4.02 \cdot 10^{-5}\), so the matrix is positive definite and \(\bar{c} = c_{\text{MAX}}\). Consider the sequence generated by (9) for \(c = c_{\text{MAX}}\). We have

\[
\Sigma_{\rho} \approx 10^2 \cdot \begin{bmatrix} 5.89 & -5.03 \\ -5.03 & 4.31 \end{bmatrix},
\]

and iteration (9) converges to

\[
\Omega^{-1} \approx 10^2 \cdot \begin{bmatrix} 4.56 & -3.90 \\ -3.90 & 3.34 \end{bmatrix}.
\]

Furthermore, the matrix \(\bar{A}^T - J\bar{B}^T\) has for eigenvalues 0.8373, 0.0892, so it is stable. Finally, Figures 2 and 3 depict the variances of the first and second component of prediction error of the Kalman filter and robust filter Kalman for the steady-state least favorable model. As expected, both variances converge to a constant value and for both components, the performance of the robust filter is approximately 1.5 dB lower than that of the Kalman filter.

VI. CONCLUSION

We have considered a robust filtering problem, where the minimum variance estimator is designed according to the least favorable model belonging to a ball about the nominal model and with a certain radius corresponding to the modeling tolerance. We showed that as long as the model tolerance does not exceed a maximum value \(\bar{c}\), the least favorable model converges to a constant model. Furthermore, as long as the tolerance is sufficiently small, the covariance matrix of the prediction error for any stable filter remains bounded when applied to the steady-state least favorable model.

REFERENCES

[1] R. N. Banavar and J. L. Speyer. Properties of risk-sensitive filters/estimators. IEEE Proc.-Control Theory Appl., 145, January 1998.
[2] R. R. Bitmead, M. R. Gevers, I. R. Petersen, and R. J. Kaye. Monotonicity and stabilizability properties of solutions of the Riccati difference equation: Propositions, lemmas, theorems, fallacious conjectures and counterexamples. Systems and Control Letters, 5:309–315, 1985.
[3] P. E. Caines and D. Q. Mayne. On the discrete time matrix Riccati equation of optimal control. Int. J. Control, 12(5):785–794, 1970.
[4] P. E. Caines and D. Q. Mayne. On the discrete time matrix Riccati equation of optimal control– a correction. Int. J. Control, 14(1):205–207, 1971.
[5] F. S. Cativelli and A. H. Sayed. Diffusion strategies for distributed Kalman filtering and smoothing. IEEE Trans. Automat. Control, 55(9):2069–2084, 2010.
[6] J. W. Demmel and B. Kagstrom. Computing stable eigendecompositions of matrix pencils. Linear Algebra and its Appl., 88/89:139–186, 1987.
[7] A. Ferrante and L. Ntogramatzidis. The generalized discrete algebraic Riccati equation in linear-quadratic optimal control. Automatica, 49:471–478, 2013.
[8] L. P. Hansen and T. J. Sargent. Robustness. Princeton University Press, Princeton, NJ, 2008.
[9] B. Hassibi, A. H. Sayed, and T. Kailath. Indefinite-Quadratic Estimation and Control– A Unified Approach to $H^2$ and $H^\infty$ Theories. Soc. Indust. Appl. Math., Philadelphia, 1999.
[10] S. Kullback. Information Theory and Statistics. J. Wiley & Sons, New York, 1959. Reprinted by Dover Publ., Mineola, NY, 1968.
[11] B. C. Levy and R. Nikoukhah. Robust least-squares estimation with a relative entropy constraint. IEEE Trans. Inform. Theory, 50(1):89–104, January 2004.
[12] B. C. Levy and R. Nikoukhah. Robust state-space filtering under incremental model perturbations subject to a relative entropy tolerance. IEEE Trans. Automat. Control, 58:682–695, March 2013.
[13] B. C. Levy and M. Zorzi. A contraction analysis of the convergence of risk-sensitive filters. SIAM J. Control and Optimiz., 54(4):2154–2173, 2016.
[14] T. Pappas, A. J. Laub, and N. R. Sandell. On the numerical solution of the discrete-time algebraic Riccati equation. IEEE Trans. Automat. Control, 25:631–641, 1980.
[15] P. Whittle. Risk-sensitive Optimal Control. J. Wiley, Chichester, England, 1980.
[16] A. Zenere and M. Zorzi. Model predictive control meets robust kalman filtering. IFAC PapersOnLine, 50(1):3774–3779, 2017.
[17] Y. Zhou. Convergence of the discrete-time matrix Riccati equation with indefinite weighting matrix. In Proc. 33rd IEEE Conference on Decision and Control (CDC), pages 1517–1518, Lake Buena Vista, FL, December 1994.
[18] M. Zorzi. Robust Kalman filtering under model perturbations. IEEE Trans. Automat. Control, 62(6):2902–2907, June 2016.
[19] M. Zorzi. Convergence analysis of a family of robust Kalman filters based on the contraction principle. SIAM J. Control and Optimiz., 55(5):3116–3131, 2017.
[20] M. Zorzi. On the robustness of the Bayes and Wiener estimators under model uncertainty. preprint, arXiv:1508.01904v2, May 2017.
[21] M. Zorzi and B. C. Levy. On the convergence of a risk sensitive like filter. In Proc. 54th IEEE Conference on Decision and Control (CDC), pages 4990–4995, Osaka, Japan, December 2015.