On the convergence of a Risk Sensitive like Filter

Mattia Zorzi, Bernard C. Levy

Abstract—In this paper, we analyze the convergence of a risk sensitive like filter where the risk sensitivity parameter is time varying. Such filter has a Kalman like structure and its gain matrix is updated according to a Riccati like iteration. We show that the iteration converges to a fixed point by using the contraction analysis.

I. INTRODUCTION

Phenomena are often modeled through a (nominal) linear model. One reason is that the corresponding filtering problem is tractable, for instance if we consider the Gauss-Markov state space model then we obtain the Kalman filter. On the other hand, linear models are rather simple and thus introduce modeling errors. This implies that the optimal filter may not perform well in the reality.

One possible strategy to face such problem is to consider the robust filtering. The pioneering works are due by Kassam, Poor and their collaborators, [11],[15]. Here, the optimal filter is selected by solving a minimax problem. One player (say, nature) selects the least favorable model in the allowable neighborhood and the other player designs the optimal filter based on that least favorable model. However, the implementation of such a filter can be very difficult because it depends on the characterization of the allowable neighborhood. To overcome that issue, it was introduced a new class of robust filters based on the minimization of risk sensitive functions, [17],[19],[2], which penalize large estimation errors. The sensitivity to large errors is tuned by the so called risk sensitivity parameter. This approach considers Gauss-Markov state space models and the resulting robust filter is a Kalman like filter where the gain matrix is updated according to a distorted version of the Riccati iteration (say, risk sensitive Riccati iteration). Unfortunately, this method is only based on the nominal model. In [13], a new robust filter has been introduced. This approach considers a minimax problem. At each time step, all possible model increments of the state space model are described by a ball about the nominal model increment. Its radius is fixed a priori and represents the incremental tolerance allowable at each time step. Therefore, the nature selects the least favorable model increment in the allowable ball. It turns out such robust filter is a risk sensitive like filter where the risk sensitivity parameter is now time varying.

An important issue for Kalman like filters is their convergence. In [3], under the assumption that the Gauss-Markov state space model is reachable and observable, it has been established that the Riccati iteration is a contraction for the Riemann metric associated to the cone of positive definite matrices, and thus the Riccati iteration asymptotically converges. The same result can be proved by using the Thompson part metric [12], [16]. In [14], a similar contraction analysis has been considered to prove the convergence of the risk sensitive Riccati iteration. Here, the problem has been formulated in the Krein space, see [7], [8]. Then, it has been shown that the N-block risk sensitive Riccati iteration is strictly contractive for the Riemann metric by choosing the risk sensitivity parameter sufficiently small.

In this paper, we consider a similar contraction analysis to prove that the risk sensitive like filter in [13] asymptotically converges for tolerance parameter values sufficiently small. More precisely, we formulate the filtering problem in the Krein space and then we show that the N-block risk sensitive Riccati iteration is strictly contractive provided that the time varying risk sensitivity parameter is smaller than a constant parameter. Moreover, it is always possible to find a lower bound of this iteration after a finite number of steps. As we will see, both the constant parameter and the lower bound allow to characterize a range of values of the tolerance parameter for which the iteration converges.

The paper is outlined as follows. In Section II we review the Thompson part metric and the contractive mappings. In Section III we review the risk sensitive like filter presented in [13]. In Section IV we construct the N-block risk sensitive Riccati iteration. In Section V we show that, after a finite number of steps, it is always possible to find a lower bound for the iteration. In Section VI we show that the mapping is a contraction. Finally, in Section VII we provide an example.

II. THOMPSON PART METRIC AND CONTRACTION MAPPINGS

Let $\mathcal{P}$ denote the cone of positive definite symmetric matrices of dimension $n$, and $\overline{\mathcal{P}}$ its closure. If $P$ is an element of $\mathcal{P}$ with eigendecomposition

$$P = U \Lambda U^T$$

where $U$ is an orthogonal matrix formed by normalized eigenvectors of $P$ and $\Lambda = \text{diag} \{\lambda_1, \ldots, \lambda_n\}$ is the diagonal eigenvalue matrix of $P$, the square-root of $P$ is defined as

$$P^{1/2} = U \Lambda^{1/2} U^T$$

where $\Lambda^{1/2}$ is diagonal, with entries $\lambda_i^{1/2}$ for $1 \leq i \leq n$. Similarly, the logarithm of $P$ is the matrix specified by

$$\log(P) = U \log(\Lambda) U^T,$$
where \( \log(\Lambda) \) is diagonal with entries \( \log(\lambda_i) \) for \( 1 \leq i \leq n \). Let \( P \) and \( Q \) be two positive definite matrices of \( \mathcal{P} \). Then \( P^{-1/2}Q P^{-1/2} \) is similar to \( P^{-1/2}Q P^{-1/2} \), so they have the same eigenvalues, and \( P^{-1/2}Q P^{-1/2} \) is positive definite. The Thompson part metric between \( P \) and \( Q \) is defined as

\[
d_T(P, Q) = \|\log(P^{-1/2}Q P^{-1/2})\|_2
\]

where \( \lambda_1(P) \geq \lambda_2(P) \geq \ldots \geq \lambda_n(P) > 0 \) are the eigenvalues of \( P \) sorted in decreasing order, and \( \|\cdot\|_2 \) denotes the spectral norm.

Let \( f(\cdot) \) be a non expansive mapping of \( \mathcal{P} \). Its contraction coefficient (or Lipschitz constant) is defined as

\[
\xi(f) = \sup_{P,Q \in \mathcal{P}, P \neq Q} \frac{d_T(f(P), f(Q))}{d_T(P, Q)}.
\]

From \( \xi(f) \) we get

\[
d_T(f(P), f(Q)) \leq \xi(f) d_T(P, Q).
\]

Moreover, if \( \xi(f) < 1 \), then \( f \) is a strictly contractive mapping.

The key result that will be used in this paper is that if \( f \) is a strict contraction of \( \mathcal{P} \) for the metric \( d_T \), by the Banach fixed point theorem [1, p. 244], there exists a unique fixed point \( P \) of \( f \) in \( \mathcal{P} \) satisfying \( P = f(P) \). Moreover, if the \( N \)-fold composition \( f^N \) of a non-expansive mapping \( f \) is strictly contractive, then \( f \) has a unique fixed point. Furthermore this fixed point can be evaluated by performing the iteration \( P_k = f(P_k) \) starting from any initial point \( P_0 \) of \( \mathcal{P} \). We will consider in particular the Riccati like map over \( \mathcal{P} \) defined by

\[
f(P) = M[P^{-1} + \Omega]^{-1}MT + W
\]

where \( P, \Omega \) and \( W \) are symmetric real positive definite matrices and \( M \) is a square real, but not necessarily invertible, matrix. For this mapping the following result was established in [12, Th. 5.3].

**Lemma 2.1:** \( f \) is a strict contraction with

\[
\xi(f) \leq \frac{\sqrt{\lambda_1(\Omega^{-1}MT W^{-1}M)}}{1 + \sqrt{1 + \lambda_1(\Omega^{-1}MT W^{-1}M)}}^2.
\]

**III. ROBUST FILTERING WITH INCREMENTAL TOLERANCE**

Consider a discrete-time stochastic process \( y_t \) described by a nominal Gauss-Markov state space model of the form

\[
x_{t+1} = Ax_t + Bu_t
\]

\[
y_t = Cx_t + Dv_t, \quad t \geq 0
\]

where the state \( x_t \in \mathbb{R}^n \), the process noise \( u_t \in \mathbb{R}^m \), and the observation noise \( v_t \in \mathbb{R}^p \). The noises \( u_t \) and \( v_t \) are assumed to be zero-mean WGN processes with normalized covariance matrices and independent, that is

\[
E[\begin{bmatrix} u_t \\ v_t \\ u_s \\ v_s \end{bmatrix}] = \begin{bmatrix} I_m \\ 0 \\ 0 \\ I_p \end{bmatrix} \delta_{t-s},
\]

where \( \delta_t \) denotes the Kronecker delta function. The initial state vector \( x_0 \) is assumed independent of the noises \( u_t \) and \( v_t \) with nominal probability density

\[
p_0(x_0) \sim \mathcal{N}(0, P_0).
\]

Since we are interested in the asymptotic behavior of the robust filter, the matrices \( A, B, C \) and \( D \) specifying the state-space nominal model are assumed to be constant. The pairs \( (A, B) \) and \( (A, C) \) are assumed to be reachable and observable, respectively. Moreover, we assume that the noises \( u_t \) and \( v_t \) affect all the components of the dynamics and observations, that is, \( BB^T \) and \( DD^T \) are positive definite. As observed in [13], this is a natural property to demand when the relative entropy is used to measure the proximity of statistical models.

Let \( z_t = \begin{bmatrix} x_{t+1}^T \\ y_t^T \end{bmatrix} \). Over a finite interval \( 0 \leq t \leq T \), the joint nominal probability density of \( X_{T+1} \) and \( Y_T \) can be expressed as

\[
p(X_{T+1}, Y_T) = p_0(x_0) \prod_{t=0}^{T} \phi_t(z_t|x_t)
\]

where, in view of \( 5 \) and \( 6 \),\n
\[
\phi_t(z_t|x_t) \sim \mathcal{N}\left( \begin{bmatrix} A \\ C \end{bmatrix} x_t, \begin{bmatrix} BB^T & 0 \\ 0 & DD^T \end{bmatrix} \right)
\]

is the nominal conditional probability density.

Assume that the true density of \( X_{T+1} \) and \( Y_T \) admits a similar Markov structure

\[
\tilde{p}(X_{T+1}, Y_T) = \tilde{p}_0(x_0) \prod_{t=0}^{T} \tilde{\phi}_t(z_t|x_t).
\]

The mismatch between the two densities can be measured through the relative entropy, [4],

\[
D(\tilde{p}, p) = \int \int \tilde{p}(X_{T+1}, Y_T) \log \left( \frac{\tilde{p}(X_{T+1}, Y_T)}{p(X_{T+1}, Y_T)} \right) dX_{T+1} dY_T.
\]

At this point, one can assume the true density \( \tilde{p} \) is such that \( D(\tilde{p}, p) \leq c \). Here, \( c \) is the tolerance parameter representing the biggest mismatch allowed. Then, the robust filter is obtained by solving a minimax problem. More precisely, one player (say, the nature) selects the least-favorable allowable density and the other player designs the optimum filter for the least favorable density. However, one potential weakness to the above filtering problem is that the nature may allocate in \( \tilde{p} \) most of the distortion budget specified by \( c \) to the moment where \( 5 \) is most susceptible to distortions.

If the modeler exercises the effort to characterize each time component of \( 5 \) and \( 6 \), it may be more appropriate to specify separate incremental tolerances for each time step of
the conditional probability density \[7\]. In [13], it has been
assumed that \(\hat{\phi}_t\) belongs to the following ball
\[
B_t \triangleq \left\{ \hat{\phi} \text{ s.t. } \mathbb{E} \left[ \log \left( \frac{\hat{\phi}_t(z_t|x_t)}{\phi_t(z_t|x_t)} \right) \right] \mid Y_{t-1} \leq c \right\}
\]  
(9)
with
\[
\mathbb{E} \left[ \log \left( \frac{\hat{\phi}_t(z_t|x_t)}{\phi_t(z_t|x_t)} \right) \mid Y_{t-1} \right] \\
= \int \int \hat{\phi}_t(z_t|x_t) \hat{p}_t(x_t|Y_{t-1}) \log \left( \frac{\hat{\phi}_t(z_t|x_t)}{\phi_t(z_t|x_t)} \right) dz_t dx_t,
\]
where \(\hat{p}_t(x_t|Y_{t-1})\) is the conditional density of \(x_t\) based
on the least favorable model and the given observations as
prior to time \(t\). This means that the nature has the access
to the collected observations \(Y_{t-1}\) and obeys to the Markov
structure \([5]\). Accordingly, the nature is required to commit
to all least-favorable models components \(\hat{\phi}_t(z_t|x_t)\) with \(0 \leq
s \leq t - 1\) generated at earlier stages of its minmax game
with the estimating player. Using the same terminology in
[6], [5], the nature operates “under commitment”.

Let \(G_t\) denote the class of estimators with finite
second-order moments with respect to all densities \(\hat{\phi}_t(z_t|x_t)\hat{p}_t(x_t|Y_{t-1})\) such that \(\hat{\phi}_t \in B_t\). In [13], the fol-
lowing robust filter has been proposed
\[
\min_{\hat{g}_t \in G_t} \max_{\hat{\phi}_t \in B_t} \mathbb{E} \left[ \|x_{t+1} - \hat{g}_t(y_t)\|^2 \mid Y_{t-1} \right]
\]
where
\[
\mathbb{E} \left[ \|x_{t+1} - \hat{g}_t(y_t)\|^2 \mid Y_{t-1} \right] \\
= \int \int \|x_{t+1} - \hat{g}_t(y_t)\|^2 \hat{\phi}_t(z_t|x_t) \hat{p}_t(x_t|Y_{t-1}) dz_t dx_t.
\]
Let \(\hat{x}_{t+1} = \hat{g}_t(y_t)\) denote the mean-square error estimate of
\(x_{t+1}\) evaluated with respect to the least-favorable probability
density of \(z_t\). Note that \(g_t(y_t)\) is a function of \(Y_t\), it depends
not only on \(y_t\), but this dependency is suppressed to simplify
notations. In [13], it has been shown that \(\hat{x}_{t+1}\) obeys the
recursion
\[
\hat{x}_{t+1} = A\hat{x}_t + K_t \nu_t,
\]  
(10)
where
\[
\nu_t \triangleq y_t - C\hat{x}_t
\]  
(11)
is the innovations process. In \([10]\), the gain matrix
\[
K_t = A(P_t^{-1} - \theta_{t-1}I_n)^{-1}C^T(R_t^{\nu})^{-1},
\]  
(12)
where
\[
R_t^{\nu} = E[\nu_t \nu_t^T] = C(P_t^{-1} - \theta_{t-1}I_n)^{-1}C^T + DD^T
\]  
(13)
represents the variance of the innovations process, \(\theta_{t-1}\) with
\[
0 < \theta_{t-1} < (\lambda_1(P_t))^{-1}
\]  
(14)
is the unique solution to the equation
\[
\gamma(\theta_{t-1}, P_t) = c
\]  
(15)
where
\[
\gamma(\theta, P) \triangleq \frac{1}{2} \left[ \log \det(I - \theta P) + \text{tr}[(I - \theta P)^{-1}] - n \right],
\]
and if \(\hat{x}_t = x_t - \hat{x}_t\) denotes the state prediction error, its
variance matrix \(P_t = E[\hat{x}_t\hat{x}_t^T]\) obeys the distorted Riccati
iteration
\[
P_{t+1} = r_c(R_t) \triangleq A[P_t^{-1} + C^T(DD^T)^{-1}C - \theta_{t-1}I]^{-1}A^T + BB^T
\]  
(17)
with initial condition \(P_0\).

Note that, the robust filter \([10]-[15]\) is a risk sensitive
like filter, [19], [18, Chapter 10]. In the classic formulation,
however, the risk-sensitive Riccati mapping is defined as
\[
r_{\theta}^{RS}(P) \triangleq A[P^{-1} + C^T(DD^T)^{-1}C - \theta I]^{-1}A^T + BB^T
\]
where the risk-sensitivity parameter \(\theta \geq 0\) is constant and
does not depend on \(P\). Moreover, for \(\theta = 0\) (risk-neutral
case) we obtain the Riccati mapping
\[
r(P) \triangleq A[P^{-1} + C^T(DD^T)^{-1}C]^{-1}A^T + BB^T.
\]
Finally, it is worth noting that \([14]\) implies that \(r_c(R) \in \mathcal{P}\)
for each \(P \in \mathcal{P}\), that is \(r_c(R)\) is a mapping of \(\mathcal{P}\). Such
property does not hold for the classic risk sensitive filter
because it may occur that \(r_{\theta}^{RS}(P) \notin \mathcal{P}\) even when \(P \in \mathcal{P}\),
[14].

IV. BLOCK UPDATE FILTER IN THE KREIN SPACE

The robust filter \([10]-[15]\) can be interpreted as solving a
standard least-square filtering problem with time-varying
parameters in Krein space. The Krein state-space model
consists of dynamics \([5]\) and observations \([6]\), to which we
must adjoin the new observations
\[
0 = x_t + u_t^R.
\]  
(18)
The components of noise vectors \(u_t, v_t\) and \(v_t^R\) now belong to
a Krein space and have the inner product
\[
\begin{pmatrix}
  u_t \\
  v_t \\
  v_t^R
\end{pmatrix} \cdot \begin{pmatrix}
  u_s \\
  v_s \\
  v_s^R
\end{pmatrix} = \begin{pmatrix}
  I_m & 0 & 0 \\
  0 & I_p & 0 \\
  0 & 0 & -(\theta_{t-1})^{-1}I_n
\end{pmatrix} \delta_{t-s}.
\]  
(19)
Note that, in the classical risk sensitive framework, \([7]\), \([8]\),
\(v_t^R\) with \(t \geq 0\) are identically distributed, whereas are not
in this setting. Since \(x_t\) is Gauss-Markov, the downsampling
process \(x_k^d = x_{kN}\), with \(N\) integer, is also Gauss-Markov
with state-space model
\[
x_{k+1}^d = A^N x_k^d + R_N u_k^N
\]  
(20)
\[
y_k^N = \mathcal{O} x_k^d + D_N v_k^N + \mathcal{H}_N u_k^N
\]  
(21)
\[
0 = \mathcal{O} x_k^d + v_k^R + \mathcal{L}_N u_k^N
\]  
(22)
where
\[
u_k^N \triangleq \begin{pmatrix}
  u_{kN+N-1}^T \\
  u_{kN+N-2}^T \cdots \\
  u_{kN}^T
\end{pmatrix}
\]  
\[
v_k^N \triangleq \begin{pmatrix}
  v_{kN+N-1}^T \\
  v_{kN+N-2}^T \cdots \\
  v_{kN}^T
\end{pmatrix}
\]  
\[
y_k^N \triangleq \begin{pmatrix}
  y_{kN+N-1}^T \\
  y_{kN+N-2}^T \cdots \\
  y_{kN}^T
\end{pmatrix}
\]  
\[
v_k^{RN} \triangleq \begin{pmatrix}
  (v_{kN+N-1})^T \\
  (v_{kN+N-2})^T \cdots \\
  (v_{kN})^T
\end{pmatrix}^T.
\]
In the model \((20)–(22)\)

\[
\begin{align*}
\mathcal{R}_N &\triangleq \begin{bmatrix} B & AB & \ldots & A^{N-1}B \end{bmatrix} \\
\mathcal{O}_N &\triangleq \begin{bmatrix} (CA^{N-1})^T & \ldots & (CA)^T & CT \end{bmatrix}^T \\
\mathcal{O}_N^R &\triangleq \begin{bmatrix} (A^{N-1})^T & \ldots & (A)^T & I \end{bmatrix}^T \\
\mathcal{D}_N &\triangleq I_N \otimes D.
\end{align*}
\]

(23)

Note that \(\mathcal{R}_N\) and \(\mathcal{O}_N\) denote respectively the \(N\)-block reachability and observability matrices of system \((5)–(8)\), where the blocks forming \(\mathcal{O}_N\) are written from bottom to top instead of the usual top to bottom convention. If the pairs \((A, B)\) and \((C, A)\) are reachable and observable, \(\mathcal{R}_N\) and \(\mathcal{O}_N\) have full rank for \(N \geq n\). In \((21)\) and \((22)\), if

\[
\begin{align*}
H_t &\triangleq \begin{cases} CA^{t-1}B & t \geq 1 \\
0 & \text{otherwise} \end{cases} \\
L_t &\triangleq \begin{cases} A^{t-1}B & t \geq 1 \\
0 & \text{otherwise} \end{cases}
\end{align*}
\]

\(\mathcal{H}_N\) and \(\mathcal{L}_N\) are block Hankel matrices defined as follows:

\[
\mathcal{H}_N = \begin{bmatrix}
0 & H_1 & H_2 & \ldots & H_{N-2} & H_{N-1} \\
0 & 0 & H_1 & H_2 & \ldots & H_{N-2} \\
& \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \ldots & \ldots & 0 & H_1 \\
0 & 0 & \ldots & \ldots & 0 & 0
\end{bmatrix}
\]

\[
\mathcal{L}_N = \begin{bmatrix}
0 & L_1 & L_2 & \ldots & L_{N-2} & L_{N-1} \\
0 & 0 & L_1 & L_2 & \ldots & L_{N-2} \\
& \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \ldots & \ldots & 0 & L_1 \\
0 & 0 & \ldots & \ldots & 0 & 0
\end{bmatrix}
\]

The Krein space inner product of the observation noise vector

\[
\begin{bmatrix}
w_k^N \\
w_k^RN
\end{bmatrix} = \begin{bmatrix} \mathcal{D}_Nw_k^N \\
w_k^RN
\end{bmatrix} + \begin{bmatrix} \mathcal{H}_N \\
\mathcal{L}_N
\end{bmatrix} \begin{bmatrix}
u_k^N
\end{bmatrix}
\]

admits the following decomposition

\[
\mathcal{K}_{\Theta_N,k} \triangleq \begin{bmatrix} \mathcal{H}_N \mathcal{L}_N^T \\
w_k^RN \end{bmatrix} \begin{bmatrix}
w_k^N \\
w_k^RN
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\mathcal{D}_N \mathcal{D}_N^T & \mathcal{D}_N \mathcal{D}_N^T + \mathcal{H}_N \mathcal{H}_N^T - \mathcal{I}_{N_n} \\
0 & -\Theta_{N,k}
\end{bmatrix} + \begin{bmatrix} \mathcal{H}_N \\
\mathcal{L}_N
\end{bmatrix} \begin{bmatrix} \mathcal{H}_N^T \\
\mathcal{L}_N^T
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\mathcal{D}_N \mathcal{D}_N^T & \mathcal{D}_N \mathcal{D}_N^T + \mathcal{H}_N \mathcal{H}_N^T - \mathcal{I}_{N_n} \\
0 & \mathcal{H}_N \mathcal{H}_N^T \end{bmatrix} \begin{bmatrix}
\mathcal{D}_N \mathcal{D}_N^T & \mathcal{D}_N \mathcal{D}_N^T + \mathcal{H}_N \mathcal{H}_N^T - \mathcal{I}_{N_n} \\
0 & \mathcal{H}_N \mathcal{H}_N^T
\end{bmatrix}
\]

\[
\Theta_{N,k} \triangleq \text{diag}(\theta_{N+2}, \ldots, \theta_{N+3}), \quad \mathcal{D}_N \triangleq I_n \otimes D.
\]

The projection of noise vector \(\hat{u}_k^N\) on the Krein subspace spanned by the observation noise vector \([w_k^N]^T (w_k^RN)^T]^T\) is then given by

\[
\hat{u}_k^N = \begin{bmatrix} G_k & G_k^R \end{bmatrix} \begin{bmatrix}
w_k^N \\
w_k^RN
\end{bmatrix}
\]

where

\[
\begin{bmatrix} G_k & G_k^R \end{bmatrix} = \begin{bmatrix} \mathcal{H}_N^T \\
\mathcal{L}_N^T
\end{bmatrix} (\Theta_{N,k})^{-1}
\]

and the residual \(\tilde{u}_k^N = u_k^N - \hat{u}_k^N\) has for inner product

\[
\mathcal{Q}_{\Theta_N,k} \triangleq \langle \tilde{u}_k^N, \hat{u}_k^N \rangle
\]

\[
= I_{Nm} - \begin{bmatrix} G_k & G_k^R \end{bmatrix} \Theta_{N,k} \begin{bmatrix} G_k^T \\
(G_k^R)^T
\end{bmatrix}
\]

\[
= [I_{Nm} + \mathcal{H}_N^T (\mathcal{D}_N \mathcal{D}_N^T)^{-1} \mathcal{H}_N - \mathcal{L}_N^T \Theta_{N,k} \mathcal{L}_N]^{-1}
\]

Then by multiplying the observation equation obtained by combining equations \((21)\) and \((22)\) by \(\mathcal{R}_N [G_k \ G_k^R]\) and subtracting it from \((20)\), we obtain the state-space equation

\[
x_{k+1} = \alpha_{N,k} x_k + \mathcal{R}_N \hat{u}_k^N + \mathcal{R}_N G_k y_k^N
\]

Then, by multiplying the observation equation obtained by combining equations \((21)\) and \((22)\) by \(\mathcal{R}_N [G_k \ G_k^R]\) and subtracting it from \((20)\), we obtain the state-space equation

\[
x_{k+1} = \alpha_{N,k} x_k + \mathcal{R}_N \hat{u}_k^N + \mathcal{R}_N G_k y_k^N
\]

where

\[
\mathcal{K}_{\Theta_N,k} \triangleq \begin{bmatrix} \mathcal{H}_N \mathcal{L}_N^T \\
w_k^RN \end{bmatrix} \begin{bmatrix}
w_k^N \\
w_k^RN
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\mathcal{D}_N \mathcal{D}_N^T & \mathcal{D}_N \mathcal{D}_N^T + \mathcal{H}_N \mathcal{H}_N^T - \mathcal{I}_{N_n} \\
0 & -\Theta_{N,k}
\end{bmatrix} + \begin{bmatrix} \mathcal{H}_N \\
\mathcal{L}_N
\end{bmatrix} \begin{bmatrix} \mathcal{H}_N^T \\
\mathcal{L}_N^T
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\mathcal{D}_N \mathcal{D}_N^T & \mathcal{D}_N \mathcal{D}_N^T + \mathcal{H}_N \mathcal{H}_N^T - \mathcal{I}_{N_n} \\
0 & \mathcal{H}_N \mathcal{H}_N^T
\end{bmatrix} \begin{bmatrix}
\mathcal{D}_N \mathcal{D}_N^T & \mathcal{D}_N \mathcal{D}_N^T + \mathcal{H}_N \mathcal{H}_N^T - \mathcal{I}_{N_n} \\
0 & \mathcal{H}_N \mathcal{H}_N^T
\end{bmatrix}
\]

where

\[
\mathcal{L}_N \triangleq \begin{bmatrix} \mathcal{H}_N \mathcal{H}_N^T \mathcal{D}_N \mathcal{D}_N^T + \mathcal{H}_N \mathcal{H}_N^T \mathcal{I}_{N_n} \end{bmatrix}
\]

\[
\mathcal{Q}_{\Theta_N,k} \triangleq \begin{bmatrix} \mathcal{H}_N \mathcal{H}_N^T \mathcal{L}_N \end{bmatrix} \begin{bmatrix}
\mathcal{H}_N \mathcal{H}_N^T \\
\mathcal{L}_N
\end{bmatrix}
\]

(29)

V. A LOWERBOUND FOR THE ITERATION

In this section, we show it is always possible to find a lower bound, after a finite number of steps, for the sequences \(\{P_t\}\) and \(\{P_d\}\) generated by \((17)\) and \((28)\), respectively. To this aim we recall the following property regarding the risk sensitive mapping, \([9,\text{page} 379]\).

Lemma 5.1: Let \(P \in \mathcal{P}\) such that \(P^{-1} - \theta_1 I \in \mathcal{P}\) and \(\theta_1 \geq \theta_2 \geq 0\). Then,

\[
\gamma_{\theta_1}^{RS}(P) \geq \gamma_{\theta_2}^{RS}(P).
\]
Moreover, it is not difficult to show, see for instance [14], that
\[ r(P_1) \preceq r(P_2), \quad \forall P_1 \succeq P_2 \]
where \( P_1 \) and \( P_2 \) belong to \( \mathcal{P} \).

Consider the sequence
\[ \mathcal{T}_{t+1} = r_0(\mathcal{T}_t), \quad \mathcal{T}_0 = BB^T. \quad (31) \]
It is well known such a sequence is nondecreasing, see for instance [10], in particular
\[ \mathcal{T}_t \succeq \mathcal{T}_q, \quad \forall t \geq q \]
for some fixed \( q \).

**Proposition 5.1:** Consider the sequence
\[ P_{t+1} = r^R(P_t), \quad P_0 \succeq BB^T. \]
Then,
\[ P_t \succeq \Phi_q, \quad \forall t \geq q. \quad (32) \]

**Proof:** We prove by induction that
\[ P_t \succeq \Phi_t, \quad \forall t \geq 0. \quad (33) \]
Since, the sequence \( \{\Phi_t\} \) is nondecreasing, then the statement follows for a fixed value of \( q \). For \( t = 0 \), we have \( P_0 \succeq \Phi_0 \). Assume that (33) holds at time \( t \), then
\[ P_{t+1} = r^R(P_t) = r^RS_{\theta_{t-1}}(P_t) \succeq r(P_t) \succeq r(\Phi_t) = \Phi_{t+1} \]
accordingly (33) also holds at time \( t+1 \).

At this point, we are able to characterize a lower bound, after a finite number of steps, for the sequences (17) and (28).

**Proposition 5.2:** Consider the iterations (17) and (28) with an arbitrary \( P_0 \in \mathcal{P} \). Then, after a finite number of steps, say \( q+1 \), we have
\[ P_t \succeq \Phi_q, \quad t \geq q+1, \]
and after \( \tilde{q} = \left[ \frac{q+1}{N} \right] \) steps
\[ P_k^d \succeq \Phi_q, \quad k \geq \tilde{q}. \]

**Proof:** Consider the sequence generated by (17) with an arbitrary initial condition \( P_0 \in \mathcal{P} \). Note that \( r^R(P) \succeq BB^T \) for any \( P \in \mathcal{P} \). Accordingly \( P_t \succeq BB^T \) for \( t \geq 1 \). We define the sequence \( \{\tilde{P}_t\} \) with \( P_t = P_{t+1} \) and \( P_0 = P_1 \succeq BB^T \). In view of Proposition 5.1, we have that \( P_t \succeq \Phi_q \) after \( q \) steps. This implies \( P_t \succeq \Phi_q \) for \( t \geq q+1 \). Noting that \( P_k^d = P_{kN} \), the last statement follows.

**VI. CONTRACTION PROPERTY OF** \( r^R \)

We now analyze under which conditions the mapping \( r^R(\cdot) \) is a strict contraction. To this aim, we have to study the properties of the Gramians \( \Omega_{\phi_{tN}} \) and \( W_{\Theta_{tN,K}} \). First, note that they depend on the positive definite matrix \( \Theta_{t,N,K} \). As noted in [14], a sufficient condition to guarantee \( \Omega_{\phi_{tN}} \) positive definite for \( 0 \leq \theta < \phi_{tN} \), and thus also \( W_{\Theta_{tN,K}} \), is that
\[ \bar{\phi}_N = \frac{1}{\lambda_1(\mathcal{L}_N(\mathcal{I}_N + \mathcal{H}_{NN}(\mathcal{D}_N\mathcal{D}_N^T)^{-1} + 1)} > 0. \]
Such a condition also guarantees that \( S_{\phi_{tN}}^{-1} \) is positive. Moreover, it was proved the following result.

**Lemma 6.1:** Let \( \phi_1 \) and \( \phi_2 \) such that \( \phi_2 \leq \phi_1 < \bar{\phi}_N \). Then,
\[ \Omega_{\phi_{1_{tN}}} \preceq \Omega_{\phi_{2_{tN}}}, \quad W_{\phi_{1_{tN}}} \preceq W_{\phi_{2_{tN}}}. \quad (35) \]
Note that
\[ \Omega_{\phi_{|\phi_0 = \Omega_N}} \]
\[ W_{\phi_{|\phi_0 = \Omega_N}} \]
which are positive definite matrices for \( N \geq n \) because the pairs \((\mathcal{C}, \mathcal{A})\) and \((\mathcal{A}, \mathcal{B})\) are observable and reachable, respectively. Accordingly, in view of Lemma 6.1 there exists a constant \( \phi_N > 0 \) such that
\[ \Omega_{\phi_{tN}} \succeq \Omega_{\phi_{tN}}, \quad W_{\phi_{tN}} \succeq W_{\phi_{tN}}, \quad \forall \theta \leq \theta < \phi_{tN}. \quad (36) \]
As noticed before, \( W_{\phi_{tN}} \) is positive definite over the range \([0, \phi_0]\). Then, we set \( \phi_{tN} = \phi_{tN} \) and check whether \( \Omega_{\phi_{tN}} \) is positive definite or not. If not, we decrease \( \phi_{tN} \) up to \( \phi_{tN} \), becomes positive semi-definite but singular. In this way, (36) holds.

Consider the sequence \( \{P_k^d\} \) generated by (28). If the sequence \( \{\Theta_{t,N,K}\} \) is such that \( \Theta_{t,N,K} \preceq \phi_{tN} \), then the Gramians \( \Omega_{\phi_{tN,K}} \) and \( W_{\phi_{tN,K}} \) are always positive definite. Accordingly, by Lemma 2.1, \( r^R(\cdot) \) is a strict contraction and the sequence \( \{P_k^d\} \) converges, and thus also \( \{P_t\} \), to a unique solution.

Let \( c_{\text{MAX}} \) be the biggest values of \( c \) such that \( \{P_t\} \) converges to a unique solution. Next we show how to recover \( c_{\text{MAX}} \) from \( \phi_{tN} \) and the lower bound \( P_{tN} \) characterized in Section V.

**Proposition 6.1:** Assuming that \( 0 < \theta < (\lambda_1(P))^{-1} \), the following facts hold:
1. \( \gamma(\cdot, P) \) is monotone increasing over \( \mathbb{R}^+ \)
2. \( \gamma(\theta, P) > 0 \) for any \( P \in \mathcal{P} \) with \( P \neq 0 \)
3. If \( P \preceq Q \) then \( \gamma(\theta, P) \geq \gamma(\theta, Q) \)

**Proof:** The first point has been proved in [13]. Regarding the second point, \( \gamma(\theta, P) \) is equal to the information divergence among the positive definite matrices \( (I - \theta P) \) and \( I \). Since \( I - \theta P \neq I \), we get \( \gamma(\theta, P) > 0 \). In order to prove the third point, we compute the first variation of \( \gamma(\theta, P) \) with respect to \( P \) in direction \( Q \in \mathcal{P} \):
\[ \delta \gamma(\theta, P; Q) = \frac{\theta}{2} \tr[-(I - \theta P)^{-1}Q + (I - \theta P)^{-1}(I - \theta P)^{-1}] = \frac{\theta}{2} \tr[Q^2(I - \theta P)^{-1} - (I + (I - \theta P)^{-1}) \times (I - \theta P)^{-1}Q^2] \geq 0. \]

**Proposition 6.2:** Let \( c \) be such that \( 0 < c < c_{\text{MAX}} \) with \( c_{\text{MAX}} \triangleq \gamma(\phi_{tN}, \Phi_q) \) and \( q \) fixed. Then, the mapping
$r_c^d(\cdot)$ is strictly contractive after $\lceil \frac{q+1}{2} \rceil$ steps. Accordingly, the sequence $\{P_k\}$ generated by (17) converges to a unique solution for any initial condition $P_0 \in \mathcal{P}$.

**Proof:** Consider iterations (17) and (28). As showed in Proposition 6.1 we have $\theta_t < \phi_N$ for $t \geq q + 1$ and therefore $\Theta_{N,k} < \phi_N I_{Nn}$ for $k \geq \tilde{q}$. Accordingly, the Gramians, $\Omega_{\phi N,k}$ and $W_{\phi N,k}$ are positive definite for $k \geq \tilde{q}$. By Lemma 2.1 the mapping $r_c^d(\cdot)$ is strictly contractive after $\tilde{q}$ steps. Since $P_{\tilde{q}}$ is the $N$-fold composition of $r_c^R(\cdot)$, it follows that the sequence generated by (17) converges for any $P_0 \in \mathcal{P}$.

Note that, by Proposition 6.1 the map

$$q \mapsto \gamma(\phi_N, P_q)$$

is nondecreasing. Thus, we have to choose $q$ sufficiently large in order to find a bigger $c_{\text{MAX}}$.

**VII. AN EXAMPLE**

We consider the Gauss-Markov state space model earlier employed in [14]

$$A = \begin{bmatrix} 0.1 & 1 \\ 0 & 1.2 \end{bmatrix}, \quad B = I_2$$

$$C = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad D = 1$$

with $n = 2$, $m = 2$ and $p = 1$. We selected $N = 8$, in this way $N \geq n$. Note that, larger values of $N$ can be considered. We found that $\phi_N = 1.6 \cdot 10^{-2}$. In Figure 1 we depicted the smallest eigenvalue of $\Omega_{\phi N}$ over the range $\theta \in [0.8 \cdot 10^{-3}]$. We found it becomes zero for $\phi_N \cong 1.3 \times 10^{-3}$.

In Figure 2 we depicted $\gamma(\theta, P_{10})$, $\gamma(\theta, P_{20})$ and $\gamma(\theta, P_{35})$. Note that, $\gamma(\phi_N, P_{10}) \cong 2.9 \cdot 10^{-3}$, $\gamma(\phi_N, P_{20}) \cong 4.39 \cdot 10^{-2}$ and $\gamma(\phi_N, P_{35}) \cong 5.43 \cdot 10^{-2}$. As expected, it is better to choose $P_{35}$ for which we have $c_{\text{MAX}} \cong 5.43 \cdot 10^{-2}$. We conclude that the robust filter having tolerance parameter $c < 5.43 \cdot 10^{-2}$ and nominal model (37) asymptotically converges to a unique solution.

**VIII. CONCLUSION**

We analyzed the convergence of the robust filter subject to an incremental tolerance. By contraction analysis, we showed that the risk sensitive Riccati iteration is strictly contractive for tolerance parameters values sufficiently small. Accordingly, the corresponding iteration converges to a fixed point, and thus the robust filter converges.

**REFERENCES**

[1] J. P. Aubin and I. Ekeland. *Applied Nonlinear Analysis*. Wiley, New York, 1984.

[2] R. N. Banavar and J. L. Speyer. Properties of risk-sensitive filters/estimators. *IEEE Proc. Control Theory Appl.*, 145, January 1998.

[3] P. Bougerol. Kalman filtering with random coefficients and contractions. *SIAM J. Control and Optimiz.*, 31:942–959, July 1993.

[4] T. Cover and J. Thomas. *Information Theory*. Wiley, New York, 1991.

[5] L. P. Hansen and T. J. Sargent. Robust estimation and control under commitment. *Journal of Economic Theory*, pages 2–258, 2005.

[6] L. P. Hansen and T. J. Sargent. *Robustness*. Princeton University Press, Princeton, NJ, 2008.

[7] B. Hassibi, A. H. Sayed, and T. Kailath. Linear estimation in Krein spaces. I. Theory. *IEEE Trans. Automat. Control*, 41:18–33, January 1996.

[8] B. Hassibi, A. H. Sayed, and T. Kailath. Linear estimation in Krein spaces. II. Applications. *IEEE Trans. Automat. Control*, 41:34–49, January 1996.

[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] T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice Hall, Upper Saddle River, NJ, 2000.

[11] T. Kassam, S. and Lim. Robust Wiener filters. *J. Franklin Inst.*, 304:171–185, 1977.
[12] H. Lee and Y. Lim. Invariant metrics, contractions and nonlinear matrix equations. *Nonlinearity*, 2:857–878, 2008.

[13] 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.

[14] B. C. Levy and M. Zorzi. A contraction analysis of the convergence of risk-sensitive filters. *Submitted*, 2013.

[15] H. Poor. On robust Wiener filtering. *IEEE Trans. Automat. Control*, 25(3):531–536, Jun 1980.

[16] Gaubert S. and Z. Qu. The contraction rate in Thompson’s part metric of order-preserving flows on a cone—application to generalized Riccati equations. *Journal of Differential Equations*, 256(8):2902–2948, 2014.

[17] J. L. Speyer, J. Deyst, and D. H. Jacobson. Optimization of stochastic linear systems with additive measurement and process noise using exponential performance criteria. *IEEE Trans. Automat. Control*, 19:358–366, 1974.

[18] J.L. Speyer and W.H. Chung. *Stochastic Processes, Estimation, and Control*. Advances in Design and Control. Soc. Indus. Applied Math., Philadelphia, 2008.

[19] P. Whittle. *Risk-sensitive Optimal Control*. J. Wiley, Chichester, England, 1980.