On the Convergence of Degenerate Risk Sensitive Filters☆

Mattia Zorzi[a], Shenglun Yi[a]

[a] *Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy.*

Abstract

We propose a degenerate risk sensitive filter which is an extension of the risk sensitive filtering paradigm to the case in which the evolution of the covariance matrix of the prediction error can be singular. We show that the corresponding risk sensitive Riccati iteration, describing the evolution of the covariance matrix of the prediction error, converges if the risk sensitivity parameter and the eigenvalues of the initial covariance matrix are sufficiently small.

Keywords: Contraction mapping; Kalman filter; Riccati equaiton; Risk sensitive filtering.

1. Introduction

The problem of designing robust versions of the Kalman filter is an active research area since several decades, see for instance [1, 2, 3, 4, 5, 6, 7, 8]. A well known robust paradigm is represented by the so called risk sensitive filters, [9, 10, 11, 12, 13, 14, 15, 16], in which the Kalman gain minimizes an exponential quadratic loss function rather than the standard quadratic one. In plain words, risk sensitive filters seek to penalize large errors and the severity depends on the so called risk sensitivity parameter that appears in the loss function.

Risk sensitive filtering has an interesting interpretation in terms of a dynamic game. More precisely, there is a minimax game which involves two players: one "hostile" player tries to select an "actual" model far from the nominal model, while the other player designs the optimum filter by minimizing the mean square error according to this actual model, [17, 18, 19, 20].

Risk sensitive filters are characterized by the so called risk sensitive Riccati recursion, which describes the evolution of the covariance matrix of the prediction error over time. The main limitation is that such evolution must belong to the cone of the positive definite matrices. Roughly speaking, this means that it is not possible to apply the risk sensitive filter to a model for which it is possible to predict exactly certain directions of the state vector. On the other hand, such limitation helped to gain some intuition on how to prove the convergence of the iteration in the case the state space model is with constant parameters: the convergence of the risk sensitive iteration can be established drawing inspiration from [21] where the convergence of the standard Riccati iteration has been proved showing that the N-fold composition of the Riccati operator is strictly contractive over the cone of the positive definite matrices. In the risk sensitive setting the convergence is guaranteed if the risk sensitivity parameter and the eigenvalues of the initial covariance matrix are sufficiently small, [11].

In [22] it has been proposed a robust filtering paradigm in which the uncertainty is specified incrementally, i.e. the hostile player cannot concentrate the uncertainty only in few time steps but rather spread it over the entire time horizon. The mismatch budget allowed at each time step is called tolerance. This filter, as well as its extensions [23, 24, 13, 25, 26], can be understood as risk sensitive-like filters because they are characterized by a time-varying risk sensitivity parameter whose evolution cannot be chosen arbitrarily, but it is regulated by the tolerance parameter. In the case the state space model is with constant parameters and the tolerance is sufficiently small, then it is possible to prove that the corresponding risk sensitive-like Riccati iteration converges by means of the contraction analysis, [27, 28]. The general case in which the evolution of the covariance matrix of the prediction error can be singular has been recently addressed, [29, 30]. However, such idea has not been formally extended yet for the risk sensitive filters.

The aim of this paper is to introduce a degenerate risk sensitive filter which is an extension of the risk sensitive filtering paradigm to the case in which the evolution of the covariance matrix lies to the cone of the positive semidefinite matrices. Moreover, we also show that the corresponding risk sensitive Riccati iteration converges provided that the risk sensitivity parameter and the eigenvalues of the initial covariance matrix are sufficiently small. The proof of this result follows a route which is different from the ones for the existing generalized Riccati iterations, see e.g. [31, 32, 33], as well as the one used in [29] because the risk sensitivity parameter in this setting is selected by the user. Finally, we show how the convergence

☆This paper was not presented at any IFAC meeting. Corresponding author M. Zorzi Tel. +39-049-8277752. Fax +39-049-8277614.

Email addresses: `zorzimat@dei.unipd.it` (Mattia Zorzi), `shenglun@dei.unipd.it` (Shenglun Yi)

result can be helpful in practice to gain some intuition about the convergence region of the filter for a given state space model.

The outline of the paper is as follows. In Section 2 we present the degenerate risk sensitive filter. In Section 3 we introduce the conditions on the risk sensitivity parameter which guarantee the convergence of the filter, while in Section 4 the conditions on the initial covariance matrix. Section 5 presents some numerical experiments. Finally, in Section 6 we draw the conclusions.

Notation: The image of matrix K is denoted by $\mathrm{Im}(K)$. Given a symmetric matrix K: $K > 0$ ($K \geq 0$) means that K is positive (semi)definite; $\sigma_{max}(K)$ and $\sigma_{min}(K)$ are the maximum and minimum eigenvalues of K, respectively; $K^{1/2}$ is a square root matrix of K; $\Pi(K)$ denotes the projection matrix such that $\mathrm{Im}(\Pi(K)) = \mathrm{Im}(K)$. The i-th eigenvalue of a matrix A is denoted by $\sigma_i(A)$. The symbol $\mathrm{Tp}\,(A_1, A_2 \cdots, A_n)$ denotes the block upper triangular Toeplitz matrix whose first block row is $[\, A_1 \ A_2 \ \dots \ A_n \,]$. $x \sim \mathcal{N}(m, K)$ means that x is a Gaussian random vector with mean m and covariance matrix K.

2. Risk sensitive filtering

Consider the nominal linear state space model

$$\begin{cases} x_{t+1} &= A_t x_t + B_t v_t \\ y_t &= C_t x_t + D_t v_t, \quad t \geq 0 \end{cases} \tag{1}$$

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, y_t and v_t are the state vector, the observation vector and normalized white Gaussian noise, respectively. Moreover, $x_0 \sim \mathcal{N}(0, \Gamma_0)$ is independent from v_t. We assume that $D_t D_t^\top > 0$ that is all the components of the observation process are affected by a full rank p-dimensional noise. The state space model in (1) is completely described by the conditional probability density of $z_t := [\, x_{t+1}^\top \quad y_t^\top \,]^\top$ given the observations $Y_{t-1} := [\, y_0 \dots y_{t-1}\,]^\top$ which is denoted by $f_t(z_t|Y_{t-1})$, with $t \geq 0$, and it is assumed to be Gaussian. We assume that the actual conditional probability density, denoted by $\tilde{f}_t(z_t|Y_{t-1})$ is Gaussian and unknown.

We consider the problem to find a prediction of x_{t+1} from the past observations Y_t taking into account that the nominal model could not coincide with the actual one. This problem can be addressed through a risk sensitive filtering problem, [34]. The latter can be relaxed by imposing a penalty term, see [17]:

$$\hat{x}_{t+1} = \underset{g_t \in \mathcal{G}_t}{\mathrm{argmin}} \max_{\tilde{f}_t \in \mathcal{B}_t} \bar{J}_t(\tilde{f}_t, g_t) - \theta^{-1} D(\tilde{f}_t, f_t) \tag{2}$$

where

$$\bar{J}_t(\tilde{f}_t, g_t) = \int_{\mathbb{R}^{n+p}} \|x_{t+1} - g_t(y_t)\|^2 \tilde{f}_t(z_t|Y_{t-1})\, \mathrm{d}z_t \tag{3}$$

is the variance of the prediction error under the actual model $\tilde{f}_t$; $D(\tilde{f}_t, f_t)$ is the Kullback-Leibler divergence between $\tilde{f}_t$ and f_t; $\theta > 0$ is the so called risk sensitivity parameter; $\mathcal{B}_t$ is the space of Gaussian probabilities $\tilde{f}_t$ such that $D(\tilde{f}_t, f_t)$ is finite; $\mathcal{G}_t$ is the class of estimators having bounded second order moments for any $\tilde{f}_t \in \mathcal{B}_t$. The estimator defined in (2) admits the following interpretation. First, the objective function in the minimax problem (2) has two terms: the first term is the mean squared error, while the second one is a measure of the distance between the actual and the nominal probability density. Moreover, there are two players in (2): the first player minimizes the mean squared error computed with respect to $\tilde{f}_t$, while the "hostile" player, namely the maximizer, seeks to choose the worst density $\tilde{f}_t$ in $\mathcal{B}_t$. More precisely, $\tilde{f}_t$ maximizes the first term, the mean squared error, while keeping the second term, i.e. the penalty term $\theta^{-1}D(\tilde{f}_t, f_t)$, not too large, which implies that the selected $\tilde{f}_t$ can not be too much different from the nominal density f_t; such a mismatch is tuned by θ: the larger θ is, the more $\tilde{f}_t$ can be different from the nominal density f_t.

Problem (2) admits a closed form solution which has an iterative structure similar to the Kalman filter. However, this problem has been considered only in the case that f_t is non-degenerate.

Remark 1. *It is worth noting that the minimizer and the maximizer in the objective function in (2) are not independent. Indeed, the minimizer $\hat{x}_{t+1}$ depends on $\tilde{f}_t(z_t \mid Y_{t-1})$, i.e. the maximizer:*

$$\hat{x}_{t+1} = \tilde{\mathbb{E}}\,[x_{t+1}|Y_t] = \int_{\mathbb{R}^n} x_{t+1} \tilde{f}_{t+1}\,(x_{t+1}|Y_t)\, \mathrm{d}x_{t+1}$$

where $\tilde{f}_{t+1}\,(x_{t+1} \mid Y_t)$ is given by

$$\tilde{f}_{t+1}\,(x_{t+1} \mid Y_t) = \frac{\tilde{f}_t\,(z_t \mid Y_{t-1})}{\int_{\mathbb{R}^n} \tilde{f}_t\,(z_t \mid Y_{t-1})\, \mathrm{d}x_{t+1}}.$$

On the other hand, since $\|x_{t+1} - g_t(y_t)\|^2$ is a nonnegative definite quadratic form of z_t, the density $\tilde{f}_t(z_t \mid Y_{t-1}) \in \mathcal{B}_t$ that maximizes the objective function in (2) can be expressed in the following form, see [20]:

$$\tilde{f}_t^0(z_t \mid Y_{t-1}) \propto \exp(\frac{\theta}{2}\|x_{t+1} - g_t(y_t)\|^2) f(z_t|Y_{t-1})$$

where $\propto$ means "proportional to". Thus, the maximizer, namely $\tilde{f}_t^0(z_t \mid Y_{t-1})$, depends on the estimator $g_t(y_t)$, i.e. the minimizer.

In this paper we consider the situation in which f_t can be a degenerate density. Such a circumstance is common in practice. The next example shows that it is not difficult to construct a state space model for which the corresponding density of z_t given Y_{t-1} is degenerate.

Example 1. *Consider the nominal state space model*

$$\begin{cases} x_{1,t+1} &= 1.8x_{1,t} - 1, 3x_{2,t} + u_t \\ x_{2,t+1} &= x_{1,t} - 0.5x_{2,t} + u_t \\ y_t &= x_{1,t} + x_{2,t} + w_t \end{cases}$$

where u_t and w_t are white Gaussian noises with variance equal to 1, moreover, we assume that they are independent. Such a model can be written as (1) with $x_t = [\, x_{1,t} \; x_{2,t} \,]^\top$ and the corresponding density f_t of z_t given Y_{t-1} is Gaussian and could be degenerate. Indeed, if $\tilde{f}_t(x_t|Y_{t-1}) \sim \mathcal{N}(\hat{x}_t, P_t)$, with

$$P_t = \begin{bmatrix} \mathrm{p} & \mathrm{p} \\ \mathrm{p} & \mathrm{p} \end{bmatrix}$$

and $\mathrm{p} \geq 0$, then $f_t(z_t|Y_{t-1})$ is Gaussian and degenerate because the corresponding covariance matrix is

$$\begin{bmatrix} 0.25\mathrm{p}+1 & 0.25\mathrm{p}+1 & \mathrm{p} \\ 0.25\mathrm{p}+1 & 0.25\mathrm{p}+1 & \mathrm{p} \\ \mathrm{p} & \mathrm{p} & 4\mathrm{p}+1 \end{bmatrix}.$$

Let $\mathcal{A}_t \subseteq \mathbb{R}^{n+p}$ be the support of f_t. To measure the discrepancy between f_t and $\tilde{f}_t$ through the Kullback-Leibler divergence we need to assume that $\tilde{f}_t$ has the same support of f_t. Therefore, we have

$$D(\tilde{f}_t, f_t) = \int_{\mathcal{A}_t} \tilde{f}_t(z_t|Y_{t-1}) \log\left(\frac{\tilde{f}_t(z_t|Y_{t-1})}{f_t(z_t|Y_{t-1})}\right) \mathrm{d}z_t. \quad (4)$$

In this way the uncertainty, characterizing the nominal model, does not affect the "deterministic directions" of the model.

Theorem 1. *Let $\theta > 0$ be fixed. Let f_t denote the conditional density of z_t given Y_{t-1} under the Gaussian nominal model (1). Let $\hat{x}_t$ denote the predictor at the previous step and the corresponding prediction error, under $\tilde{f}_{t-1}$, be zero mean and with covariance matrix Γ_t. Then, the optimal predictor solution to (2) is*

$$\hat{x}_{t+1} = A_t\hat{x}_t + G_t(y_t - C_t\hat{x}_t) \quad (5)$$

and the prediction gain is

$$G_t = (A_t\Gamma_t C_t^\top + B_t D_t^\top)(C_t\Gamma_t C_t^\top + D_t D_t^\top)^{-1}. \quad (6)$$

Moreover, the prediction error $x_{t+1} - \hat{x}_{t+1}$, under $\tilde{f}_t$, is zero mean and with covariance matrix

$$\Gamma_{t+1} = (P_{t+1}^+ - \theta\Pi(P_{t+1}))^+ \quad (7)$$

where

$$P_{t+1} = A_t\Gamma_t A_t^\top - G_t(C\Gamma_t C^\top + D_t D_t^\top)G_t^\top + B_t B_t^\top. \quad (8)$$

The aforementioned solution exists only if θ is such that

$$\theta < \sigma_{max}(P_{t+1})^{-1}. \quad (9)$$

Proof. First, notice that $f_t(z_t|Y_{t-1}) \sim \mathcal{N}(m_{z_t}, K_{z_t})$ where

$$\begin{aligned} m_{z_t} &= \begin{bmatrix} A_t \\ C_t \end{bmatrix} \hat{x}_t, \\ K_{z_t} &= \begin{bmatrix} A_t\Gamma_t A_t^\top + B_t B_t^\top & A_t\Gamma_t C_t^\top + B_t D_t^\top \\ C\Gamma_t A_t^\top + D_t B_t^\top & C_t\Gamma_t C_t^\top + D_t D_t^\top \end{bmatrix} \end{aligned} \quad (10)$$

and thus P_{t+1} represents the Schur complement of the block (2,2) of K_{z_t}. Since the maximizer $\tilde{f}_t$ in (2) is Gaussian, let $\tilde{f}_t(z_t|Y_{t-1}) \sim \mathcal{N}(\tilde{m}_{z_t}, \tilde{K}_{z_t})$ with

$$\tilde{m}_{z_t} = \begin{bmatrix} \tilde{m}_{x_{t+1}} \\ \tilde{m}_{y_t} \end{bmatrix}, \quad \tilde{K}_{z_t} = \begin{bmatrix} \tilde{K}_{x_{t+1}} & \tilde{K}_{x_{t+1},y_t} \\ \tilde{K}_{y_t,x_{t+1}} & \tilde{K}_{y_t} \end{bmatrix}.$$

Accordingly, the minimizer is

$$g_t^\star(y_t) = \tilde{K}_{x_{t+1},y_t}\tilde{K}_{y_t}^{-1}(y_t - \tilde{m}_{y_t}) + \tilde{m}_{x_{t+1}}.$$

Moreover, $\bar{J}_t(\tilde{f}_t, g_t^\star) = \mathrm{tr}(\Gamma_{t+1})/2$ where $\Gamma_{t+1} = \tilde{K}_{x_{t+1}} - \tilde{K}_{x_{t+1},y_t}\tilde{K}_{y_t}^{-1}\tilde{K}_{y_t,x_{t+1}}$ is the Schur complement of the block (2,2) of $\tilde{K}_{z_t}$. Since f_t and $\tilde{f}_t$ have the same support, then it follows that $\mathrm{Im}(K_{z_t}) = \mathrm{Im}(\tilde{K}_{z_t})$ and thus $\mathrm{Im}(P_{t+1}) = \mathrm{Im}(\Gamma_{t+1})$. Therefore,

$$\begin{aligned} \bar{J}_t(\tilde{f}_t, g_t^\star) &= \frac{1}{2}\mathrm{tr}(\Gamma_{t+1}\Pi(P_{t+1})) = \frac{1}{2}\mathrm{tr}(\Gamma_{t+1}H_t^\top H_t) \\ &= \int_{\mathbb{R}^{n+p}} \|H_t(x_{t+1} - g_t(y_t))\|^2 \tilde{f}_t(z_t|Y_{t-1})\,\mathrm{d}z_t \end{aligned}$$

where $\Pi(P_{t+1}) = H_t^\top H_t$ and H_t is full row rank. We conclude that Problem (2) can be written with respect to the "reduced" random vector $z_{r,t} := [\,(H_t x_{t+1})^\top \; y_t^\top\,]^\top$ and the corresponding nominal and actual densities are non-degenerate. At this point, we can exploit the well known results about risk sensitive filtering [17, 34] in order to prove the statement. □

From Theorem 1 we obtain the degenerate risk sensitive filter outlined in Algorithm 1. In the case that $P_{t+1} > 0$, then $\Pi(P_{t+1}) = I$. Thus, (7) becomes $\Gamma_{t+1} = (P_{t+1}^{-1} - \theta I)^{-1}$. Accordingly, the minimax problem in (7) boils down to the one in [34].

Algorithm 1 Risk sensitive filter at time t

Require: y_t, $\hat{x}_t$, P_t

1: $\Gamma_t = \left(P_t^+ - \theta\Pi(P_t)\right)^+$
2: $G_t = (A_t\Gamma_t C_t^\top + B_t D_t^\top)(C_t\Gamma_t C_t^\top + D_t D_t^\top)^{-1}$
3: $\hat{x}_{t+1} = A_t\hat{x}_t + G_t(y_t - C_t\hat{x}_t)$
4: $P_{t+1} = A_t\Gamma_t A_t^\top - G_t(C_t\Gamma_t C_t^\top + D_t D_t^\top)G_t^\top + B_t B_t^\top$

In the situation that $\theta = 0$, i.e. the nominal model and the actual one coincide, we have $P_t = \Gamma_t$; in particular, the recursion of P_t becomes the usual Riccati equation and thus we obtain the standard Kalman filter.

Remark 2. *Consider the nonlinear nominal model*

$$\begin{cases} x_{t+1} &= \mathbf{f}(x_t) + B_t v_t \\ y_t &= \mathbf{h}(x_t) + D_t v_t \end{cases}$$

where $\mathbf{f}$ *and* $\mathbf{h}$ *are differentiable functions. Drawing inspiration from the Extended Kalman Filter (EKF), at time* t *we consider the linearized model around the previous estimates* $\hat{x}_t$ *and* $\hat{x}_{t|t}$*:*

$$\begin{cases} x_{t+1} &= A_t x_t - A_t \hat{x}_{t|t} + \mathbf{f}(\hat{x}_{t|t}) + B_t v_t \\ y_t &= C_t x_t - C_t \hat{x}_t + \mathbf{h}(\hat{x}_t) + D_t v_t \end{cases}$$

where

$$A_t := \left.\frac{\partial \mathbf{f}(x)}{\partial x}\right|_{x=\hat{x}_{t|t}}, \quad C_t := \left.\frac{\partial \mathbf{h}(x)}{\partial x}\right|_{x=\hat{x}_t}.$$

In [35, 36] a risk sensitive-like filter in the non-degenerate case has been proposed. Combining this idea with the one of this section, it is possible to set up a minimax game, like the one in (2), with respect to the linearized model and derive the extended risk sensitive filter in the degenerate case which is outlined in Algorithm 2.

Algorithm 2 Extended risk sensitive filter at time t

Require: y_t, $\hat{x}_t$, P_t

1: $\Gamma_t = \left(P_t^+ - \theta\Pi(P_t)\right)^+$
2: $C_t := \left.\frac{\partial \mathbf{h}(x)}{\partial x}\right|_{x=\hat{x}_t}$
3: $L_t = \Gamma_t C_t^\top (C_t \Gamma_t C_t^\top + D_t D_t^\top)^{-1}$
4: $\hat{x}_{t|t} = \hat{x}_t + L_t\,(y_t - \mathbf{h}(\hat{x}_t))$
5: $A_t := \left.\frac{\partial \mathbf{f}(x)}{\partial x}\right|_{x=\hat{x}_{t|t}}$
6: $\hat{x}_{t+1} = \mathbf{f}(\hat{x}_{t|t})$
7: $P_{t+1} = A_t(\Gamma_t^+ + C_t^\top (D_t D_t^\top)^{-1} C_t)^+ A_t^\top + B_t B_t^\top$

3. Convergence analysis

In this section we consider the asymptotic behavior of the risk sensitive filter proposed in Section 2 in the case that the nominal state space model in (1) has constant parameters, i.e. $A_t = A$, $B_t = B$, $C_t = C$ and $D_t = D$. Without loss of generality we assume that $BD^\top = 0$. Indeed, it is always possible to rewrite the filter of Section 2 in terms of $\tilde{A} = A - BD^\top(DD^\top)^{-1}C$, $\tilde{B}$ such that $\tilde{B}\tilde{B}^\top = B(I - D^\top(DD^\top)^{-1}D)B^\top$, $\tilde{C} = C$ and $\tilde{D} = D$. In view of Algorithm 1, the behavior of the filter is governed by the covariance matrix P_t which obeys the risk sensitive Riccati recursion

$$P_{t+1} = r_\theta(P_t) \tag{11}$$

where the mapping $r_\theta(\cdot)$ is defined as

$$r_\theta(P) := A\Gamma A^\top - A\Gamma C^\top (C\Gamma C^\top + R)^{-1} C\Gamma A^\top + Q \tag{12}$$

where $\Gamma = (P^+ - \theta\Pi(P))^+$, $Q = BB^\top$ and $R = DD^\top$. The main result of the next two sections is to show that, under the assumption that (A, B) is stabilizable and the corresponding reachable subsystem is observable, the Riccati recursion (11) converges to the unique solution P_∞ of the algebraic equation $P = r_\theta(P)$ provided that: i) $\theta > 0$ is sufficiently small; ii) the initial condition P_0 has eigenvalues not so large. In doing that, we need the following lemmata whose proofs are similar to the ones of the results in [29].

Lemma 1. *Let* $\bar{P}_t$*,* $t \geq 0$*, be the sequence generated by the standard Riccati equation*

$$\bar{P}_{t+1} = A\bar{P}_t A^\top - A\bar{P}_t C^\top (C\bar{P}_t C^\top + R)^{-1} C\bar{P}_t A^\top + Q \tag{13}$$

with $\bar{P}_0 = P_0$*. If the sequence* P_t*,* $t \geq 0$*, generated by (11) is well defined, i.e.* $\sigma_{max}(P_t) < \theta^{-1}$*, then we have*

$$\mathrm{Im}(P_t) = \mathrm{Im}(\bar{P}_t). \tag{14}$$

If the pair (A, B) is stabilizable and the pair (A, C) is detectable, then for any initial condition $\bar{P}_0 \geq 0$ we have that $\bar{P}_t$ converges to the unique solution $\bar{P}_\infty$ of the ARE corresponding to (13). Let U and V be two matrices whose columns form an orthonormal basis for $\mathrm{Im}(\bar{P}_\infty)$ and its complement in $\mathbb{R}^n$, respectively. In view of Lemma 1, there exists $\bar{t} \in \mathbb{N}$ such that

$$P_t^R := U^\top P_t U > 0, \quad \forall t \geq \bar{t}. \tag{15}$$

Lemma 2. *Assume that* (A, B) *and* (A, C) *are stabilizable and detectable. If* P_0 *is such that* $\mathrm{Im}(P_0) \subseteq \mathrm{Im}(U)$ *and the sequence generated by (11) is well defined, i.e.* $\sigma_{max}(P_t) < \theta^{-1}$*, then* $\mathrm{Im}(P_t) \subseteq \mathrm{Im}(U)$ *for any* $t \geq 0$*. Moreover, the subsystem* $(U^\top AU, U^\top B, CU, D)$ *is reachable.*

We need the following matrices which can be computed from the state space matrices of the nominal model:

$$\begin{aligned}
A_R &:= U^\top AU, \quad Q_R := U^\top BB^\top U, \quad C_R := CU. \\
\mathcal{J}_N &:= \mathcal{O}_N^R - \mathcal{L}_N \mathcal{H}_N^\top [\mathcal{D}_N \mathcal{D}_N^\top + \mathcal{H}_N \mathcal{H}_N^\top]^{-1} \mathcal{O}_N \\
\mathcal{O}_N &:= \begin{bmatrix} (C_R A_R^{N-1})^\top & \dots & (C_R A_R)^\top & C_R^\top \end{bmatrix}^\top \\
\mathcal{O}_N^R &:= \begin{bmatrix} (A_R^{N-1})^\top & \dots & A_R^\top & I \end{bmatrix}^\top \\
\mathcal{D}_N &:= I_N \otimes D \\
\mathcal{H}_N &:= \mathrm{Tp}\,(0\ H_1, H_2, \cdots, H_{N-2}, H_{N-1}) \\
\mathcal{L}_N &:= \mathrm{Tp}\,(0, L_1, L_2, \cdots, L_{N-2}, L_{N-1}) \\
H_t &:= \begin{cases} C_R A_R^{t-1} Q_R^{1/2} & t \geq 1 \\ 0 & \text{otherwise} \end{cases} \\
L_t &:= \begin{cases} A_R^{t-1} Q_R^{1/2} & t \geq 1 \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

Let

$$\bar{\theta} \in (0, \sigma_{max}(\mathcal{L}_N(I + \mathcal{H}_N^\top(\mathcal{D}_N\mathcal{D}_N^\top)^{-1}\mathcal{H}_N)^{-1}\mathcal{L}_N^\top)^{-1}) \tag{16}$$

be the maximum value for which the matrix

$$\mathcal{O}_N^\top(\mathcal{D}_N\mathcal{D}_N^\top+\mathcal{H}_N\mathcal{H}_N^\top)^{-1}\mathcal{O}_N$$
$$+\mathcal{J}_N^T(\mathcal{L}_N(I+\mathcal{H}_N^\top(\mathcal{D}_N\mathcal{D}_N^\top)^{-1}\mathcal{H}_N)^{-1}\mathcal{L}_N^\top-\bar{\theta}^{-1}I)^{-1}\mathcal{J}_N$$

is positive definite. Next, we consider the subsequence

$$P_k^d := P_{kN}, \quad k\in\mathbb{N} \tag{17}$$

with $N \geq n$.

Proposition 1. *Assume that the sequence generated by (11) is well defined, i.e. $\sigma_{max}(P_t) < \theta^{-1}$. Let the pair (A,B) be stabilizable, and the reachable subsystem be also observable. Then, there always exists $\bar{\theta}>0$, defined as above, such that if $0<\theta\leq\bar{\theta}$ and $\mathrm{Im}(P_0)\subseteq\mathrm{Im}(U)$, then the subsequence (17) converges to P_∞, i.e. the unique solution of the algebraic equation $P=r_\theta(P)$.*

Proof. It is not difficult to see that $P_{t+1}^R = r_\theta^R(P_t^R)$ where

$$r_\theta^R(P_t^R) := A_R((P_t^R)^{-1}-\theta I + C_R^\top R^{-1}C_R)^{-1}A_R^\top + Q_R.$$

By Lemma 2 the pair (A_R,Q_R) is reachable and thus, by assumption, (A_R,C_R) is observable. Accordingly, by [37, Proposition 3.1] $\bar{\theta}$ defined as above exists. Moreover, by using the results of [37, Section 4] the algebraic equation $P^R=r_\theta^R(P^R)$ admits a unique solution P_∞^R and the subsequence $P_k^{Rd}=P_{kN}^R$, $k\in\mathbb{N}$, converges to P_∞^R for $0<\theta\leq\bar{\theta}$ because the N-fold composition, say $r_\theta^{Rd}(\cdot)$, of $r_\theta^R(\cdot)$, i.e. $P_{k+1}^{Rd}=r_\theta^{Rd}(P_k^{Rd})$, is a strict contraction in the matrix space of positive definite matrices equipped with the Thompson part metric δ:

$$\delta(r_\theta^{Rd}(X), r_\theta^{Rd}(Y)) \leq \eta_N\delta(X,Y), \quad X,Y>0 \tag{18}$$

for some $0<\eta_N<1$. Finally, by (15) and Lemma 2 we have that $\mathrm{Im}(P_t^d)=\mathrm{Im}(U)$ for $t\geq\bar{t}$ which implies $\mathrm{Im}(P_k^d)=\mathrm{Im}(U)$ for $k\geq\lceil\bar{t}/N\rceil$ and thus $P_k^d=UP_k^{Rd}U^\top$. We conclude that P_k^d converges to $P_\infty := UP_\infty^R U^\top$ as k approaches infinity. □

In the next proposition we remove the condition on the image of P_0.

Proposition 2. *Assume that the sequence generated by (11) is well defined, i.e. $\sigma_{max}(P_t)<\theta^{-1}$. Let the pair (A,B) be stabilizable, and the reachable subsystem be also observable. Let $0<\theta\leq\bar{\theta}$. If $V^\top P_k^d V\to 0$, then the subsequence (17) converges to P_∞.*

Proof. Let $r_\theta^d(\cdot)$ denote the N-fold composition of $r_\theta(\cdot)$. Since $r_\theta^d(\cdot)$ and $r_\theta(\cdot)$ are continuous mappings over the cone of the positive definite matrices and $V^\top P_k^d V\to 0$, we have

$$P_{k+1}^d = r_\theta^d(P_k^d) = r_\theta(UU^\top P_k^d UU^\top) + M_k$$

where $\|M_k\|\to 0$. Thus,

$$P_{k+1}^{Rd} = r_\theta^{Rd}(P_k^{Rd}) + N_k$$

where $r_\theta^{Rd}(\cdot)$ is the N-fold composition of $r_\theta^R(\cdot)$, $N_k = U^\top M_k U$ and $\|N_k\|\to 0$. Since the Thompson part metric is continuous with respect its arguments, we have

$$\begin{aligned}\delta(P_{k+1}^{Rd},P_\infty^R) &= \delta(r_\theta^{Rd}(P_k^{Rd})+N_k,P_\infty^R)\\ &=\delta(r_\theta^{Rd}(P_k^{Rd}),P_\infty^R)+\varepsilon_k = \delta(r_\theta^{Rd}(P_k^{Rd}),r_\theta^{Rd}(P_\infty^R))+\varepsilon_k\\ &\leq \eta_N\delta(P_k^{Rd},P_\infty^R)+\varepsilon_k\end{aligned} \tag{19}$$

where $\varepsilon_k\to 0$ and we exploited the contraction property of $r_\theta^{Rd}(\cdot)$ with contraction coefficient $0<\eta_N<1$. Accordingly, $\delta(P_{k+1}^{Rd},P_\infty^R)\to 0$, i.e. $P_k^{Rd}\to P_\infty^R$. Finally,

$$P_k^d = U\underbrace{U^\top P_k^d U}_{\to P_\infty^R}U^\top + V\underbrace{V^\top P_k^d V}_{\to 0}V^\top \to UP_\infty^R U^\top = P_\infty.$$

□

4. Constraints on the initial condition

Next, we show how to impose conditions on P_0 in order to guarantee that the sequence in (11) is well defined, i.e. $\sigma_{max}(P_t)<\theta^{-1}$, and $V^\top P_t V\to 0$. Notice that the latter conditions imply that $\sigma_{max}(P_k^d)<\theta^{-1}$, and $V^\top P_k^d V\to 0$. Let $(\check{A},\check{B},\check{C},\check{D})$ be the reachability standard form of (A,B,C,D) with $\check{A}=TAT^{-1}$, $\check{B}=TB$, $\check{C}=CT^{-1}$ and T is the corresponding transformation matrix. Hence, we have

$$\check{A}=\begin{bmatrix} A_{11} & A_{12}\\ 0 & A_{22}\end{bmatrix},\quad \check{B}=\begin{bmatrix} B_1\\ 0\end{bmatrix},$$
$$\check{C}=\begin{bmatrix} C_1 & C_2\end{bmatrix},\quad \check{D}=D \tag{20}$$

and the pair (A_{11},B_1) is reachable.

Proposition 3. *Assume that the sequence in (11) is well defined, i.e. $\sigma_{max}(P_t)<\theta^{-1}$, and let $0<\theta\leq\bar{\theta}$. If $P_0<\theta^{-1}(1-\max_i|\sigma_i(A_{22})|^2)I$, then $V^\top P_t V\to 0$.*

Proof. We prove the claim in the new basis defined above. Let $\check{P}_t := TP_tT^\top$ and $\check{\Gamma}_t := T\Gamma_t T^\top$ denote the covariance matrices in the new basis. We partition $\check{P}_t$ and $\check{\Gamma}_t$ conformably with (20):

$$\check{P}_t=\begin{bmatrix}\check{P}_{11,t} & \check{P}_{12,t}\\ \check{P}_{12,t}^\top & \check{P}_{22,t}\end{bmatrix},\quad \check{\Gamma}_t=\begin{bmatrix}\check{\Gamma}_{11,t} & \check{\Gamma}_{12,t}\\ \check{\Gamma}_{12,t}^\top & \check{\Gamma}_{22,t}\end{bmatrix}.$$

Then, we have to prove that: if $P_0<\theta^{-1}(1-\max_i|\sigma_i(A_{22})|^2)I$ then $\check{P}_{22,t}\to 0$. It is not difficult to see that

$$\check{P}_{22,t+1}\leq A_{22}\check{\Gamma}_{22,t}A_{22}^\top. \tag{21}$$

Since $\check{\Gamma}_t = T\Gamma_t T^\top \geq TP_tT^\top = \check{P}_t$, we wonder whether there exists $\gamma>1$ such that

$$\check{\Gamma}_t\leq\gamma\check{P}_t. \tag{22}$$

Let λ denote a nonnull eigenvalue of $\check{P}_t$, then the corresponding eigenvector is also an eigenvector for $\check{\Gamma}_t$ and corresponding to the eigenvalue $(\lambda^{-1}-\theta)^{-1}$. Thus, in view of the fact that $\operatorname{Im}(\check{P}_t)=\operatorname{Im}(\check{\Gamma}_t)$, Condition (22) is equivalent to $(\lambda^{-1}-\theta)^{-1}\leq\gamma\lambda$ for any nonnull eigenvalue of $\check{P}_t$ and thus, recalling that P_t and $\check{P}_t$ have the same eigenvalues,

$$\gamma\geq(1-\theta\sigma_{max}(P_t))^{-1}>1 \tag{23}$$

where the last inequality follows from the assumption that $\sigma_{max}(P_t)<\theta^{-1}$. Under such condition, taking into account (21) and (22), we have

$$\check{P}_{22,t+1}\leq\gamma A_{22}\check{P}_{22,t}A_{22}^\top.$$

The latter implies $\check{P}_{22,t}\to 0$ if $(1-\theta\sigma_{max}(P_t))^{-1}\leq\gamma<\max_i|\sigma_i(A_{22})|^{-2}$. The latter condition is guaranteed by $(1-\theta\sigma_{max}(P_t))^{-1}<\max_i|\sigma_i(A_{22})|^{-2}$ that is $P_0<\theta^{-1}(1-\max_i|\sigma_i(A_{22})|^2)I$. □

Next, we characterize the condition on P_0 which guarantees that $\sigma_{max}(P_t)<\theta^{-1}$.

Lemma 3. *Let $0\leq P_1\leq P_2$ be two positive semidefinite matrices and $\sigma_{max}(P_2)<\theta^{-1}$. Then, $r_\theta(P_1)\leq r_\theta(P_2)$, i.e. $r_\theta(\cdot)$ is monotone nondecreasing over the cone of positive semidefinite matrices with respect to the partial order of symmetric matrices.*

Proof. The mapping $r_\theta(\cdot)$ is the composition of two mappings, i.e. $r_\theta(\cdot)=r\circ w_\theta(\cdot)$ where

$$\begin{aligned} r(P)&:=APA^\top-APC^\top(CPC^\top+R)^{-1}CPA^\top+Q\\ w_\theta(P)&:=(P^+-\theta\Pi(P))^+. \end{aligned}$$

$r(\cdot)$ is the Riccati mapping and it is monotone nondecreasing over the cone of positive semidefinite matrices, [38]. It remains to analyze the property of the mapping $w_\theta(\cdot)$.

First case: $P_1\leq P_2<\theta^{-1}I$ with $\operatorname{Im}(P_1)=\operatorname{Im}(P_2)$. Let $\mathcal{V}$ denote the vector space of symmetric matrices whose image coincides with the one of P_1 and P_2. Then, the first variation of $w_\theta(P)$, with $P\in\mathcal{V}$ positive semidefinite, along the direction $\delta P\in\mathcal{V}$ is

$$\delta w_\theta(P;\delta P)=w_\theta(P)P^+\delta PP^+w_\theta(P)$$

which is nonnegative for any $\delta P\geq 0$. Accordingly, $w_\theta(P_2)\geq w_\theta(P_1)$. Since $r(\cdot)$ monotone nondecreasing, we have

$$r_\theta(P_1)\leq r_\theta(P_2). \tag{24}$$

Second case: $P_1\leq P_2<\theta^{-1}I$ and the eigenvectors of P_1 corresponding to nonnull eigenvalues are also eigenvectors of P_2. Hence,

$$P_1=\sum_{j=1}^{p_1}\lambda_{1,j}v_jv_j^\top,\quad P_2=\sum_{j=1}^{p_2}\lambda_{2,j}v_jv_j^\top$$

where $0<\lambda_{1,j}\leq\lambda_{2,j}$, $p_1\leq p_2$ and v_j's form an orthonormal basis. Moreover,

$$w_\theta(P_1)=\sum_{j=1}^{p_1}\frac{1}{\lambda_{1,j}^{-1}-\theta}v_jv_j^\top\leq\sum_{j=1}^{p_2}\frac{1}{\lambda_{2,j}^{-1}-\theta}v_jv_j^\top=w_\theta(P_2).$$

Since $r(\cdot)$ monotone nondecreasing, we have

$$r_\theta(P_1)\leq r_\theta(P_2). \tag{25}$$

General case: $P_1\leq P_2<\theta^{-1}I$. It is always possible to take a matrix $P_3\geq 0$ such that $P_1\leq P_3\leq P_2$, $\operatorname{Im}(P_3)=\operatorname{Im}(P_2)$ and the eigenvectors of P_1 corresponding to nonnull eigenvalues are also eigenvectors of P_3. Indeed, let $P_2=\bar{U}D_2\bar{U}^\top$ be the "reduced" eigenvalue decomposition of P_2, i.e. $\bar{U}\in\mathbb{R}^{n\times p_2}$ such that $\bar{U}^\top\bar{U}=I_{p_2}$ and $D_2>0$ is diagonal. Since $P_1\leq P_2$ implies $\operatorname{Im}(P_1)\subseteq\operatorname{Im}(P_2)$, then we can always decompose P_1 as $P_1=\bar{U}X\bar{U}^\top$ where $X\geq 0$. Hence, $P_3:=\bar{U}[\varepsilon(X+I)-\varepsilon\Pi(X)]\bar{U}^\top$, with $\varepsilon>0$ taken sufficiently small, satisfies the aforementioned property. Therefore, we conclude that

$$r_\theta(P_1)\leq r_\theta(P_3)\leq r_\theta(P_2)$$

where the first inequality follows from the second case above, while the second inequality follows from the first case above. □

For any observer gain matrix $\check{G}=[\,\check{G}_1^\top\;\check{G}_2^\top\,]^\top$, conformably partitioned with (20), it is not difficult to see that

$$\begin{aligned}\check{P}_{t+1}&=(\check{A}-\check{G}\check{C})\check{\Gamma}_t(\check{A}-\check{G}\check{C})^\top+\check{Q}+\check{G}\check{R}\check{G}^\top\\ &\quad-[(\check{A}-\check{G}\check{C})\check{\Gamma}_t\check{C}^\top-\check{G}\check{R}](\check{C}\check{\Gamma}_t\check{C}^\top+\check{R})^{-1}\\ &\quad\times[(\check{A}-\check{G}\check{C})\check{\Gamma}_t\check{C}^\top-\check{G}\check{R}]^\top\end{aligned} \tag{26}$$

where $\check{Q}=TQT^\top$ and $\check{R}=R$. The closed loop observer matrix is

$$\check{A}-\check{G}\check{C}=\begin{bmatrix}\check{A}_{11}-\check{G}_1\check{C}_1 & \check{A}_{12}-\check{G}_1\check{C}_2\\ -\check{G}_2\check{C}_1 & \check{A}_{22}-\check{G}_2\check{C}_2\end{bmatrix}. \tag{27}$$

If we take $\check{G}_2=0$, then the eigenvalues of $\check{A}-\check{G}\check{C}$ are the ones of $\check{A}_{11}-\check{G}_1\check{C}_1$ and A_{22}. Since (A_{11},C_1) is observable, it is possible to select $\check{G}_1$, with $\check{G}_2=0$, such that the matrix $\check{A}-\check{G}\check{C}$ is stable. Therefore, it is always possible to take $\check{G}$ with $\check{G}_2\neq 0$ a perturbation matrix (i.e. a matrix whose norm is sufficiently small) such that $\check{A}-\check{G}\check{C}$ is stable because the eigenvalues of a matrix are continuous functions of the entries of the matrix. Let

$$\alpha:=\max_i|\sigma_i(\check{A}-\check{G}\check{C})|<1. \tag{28}$$

For $1<\rho<\alpha^{-1}$, the matrix $\rho(\check{A}-\check{G}\check{C})$ is stable and the algebraic Lyapunov equation

$$\check{\Sigma}_\rho=\rho^2(\check{A}-\check{G}\check{C})\check{\Sigma}_\rho(\check{A}-\check{G}\check{C})^2+\check{B}\check{B}^\top+\check{G}\check{R}\check{G}^\top \tag{29}$$

admits a unique solution which is positive semidefinite. It is worth noting that condition $\rho^2\check{\Sigma}_\rho^+ - (\check{\Sigma}_\rho^+ - \theta\Pi(\check{\Sigma}_\rho))^+ \geq 0$, i.e. $\theta\Pi(\check{\Sigma}_\rho) \leq (1-\rho^{-2})\check{\Sigma}_\rho^+$, holds if and only if

$$0 < \theta \leq \beta_\rho := \frac{\rho^2 - 1}{\rho^2\sigma_{max}(\check{\Sigma}_\rho)} \tag{30}$$

where β_ρ is always strictly positive.

As we will see, the constraint on P_0 ensuring $\sigma_{max}(P_t) < \theta^{-1}$ depends on $\check{\Sigma}_\rho$: the larger the eigenvalues of $\check{\Sigma}_\rho$ are, the more freedom we have on the choice of P_0. The introduction of the observer gain allows to maximize the eigenvalues of Σ_ρ. For instance, under the assumption that the index of cyclicity of matrix A_{22}[1] is less than or equal to p (the output dimension), we can obtain $\check{\Sigma}_\rho > 0$ because it is always possible to take $\check{G}$ such that $\check{A} - \check{G}\check{C}$ is stable and the pair $(\check{A} - \check{G}\check{C}, [\check{B}\ \check{G}\check{R}^{1/2}])$ is reachable. Indeed, under this cyclicity condition it is always possible to select $\check{G}$ with $\check{G}_2$ as a perturbation matrix such that the pair $(\check{A}_{22}, \check{G}_2)$ is reachable and $\check{A} - \check{G}\check{C}$ is stable. Let $s = [\, s_1^\top \;\; s_2^\top \,]^\top$ be conformably partitioned with (20) and such that $s^\top [\, \check{A} - \check{G}\check{C} - \lambda I \;\; \check{B} \;\; \check{G}R^{1/2} \,] = 0$ where $\lambda \in \mathbb{C}$. Hence, we have

$$\begin{aligned} s_1^\top \check{A}_{11} &= \lambda s_1^\top & \qquad s_1^\top \check{A}_{12} + s_2^\top \check{A}_{22} &= \lambda s_2^\top \\ s_1^\top \check{B}_1 &= 0 & s_1^\top \check{G}_1 + s_2^\top \check{G}_2 &= 0; \end{aligned}$$

by the Popov-Belevitch-Hatus (PBH) test, we have that $s_1 = 0$ because the pair $(\check{A}_{11}, \check{B}_1)$ is reachable. Accordingly, the above conditions simplify to

$$\begin{aligned} s_2^\top \check{A}_{22} &= \lambda s_2^\top \\ s_2^\top \check{G}_2 &= 0; \end{aligned}$$

again, by the PBH test we have that $s_2 = 0$ because the pair $(\check{A}_{22}, \check{G}_2)$ is reachable. Therefore, $[\, \check{A} - \check{G}\check{C} - \lambda I \;\; \check{B} \;\; \check{G}R^{1/2} \,]$ is full row rank for any $\lambda \in \mathbb{C}$, therefore by the PBH test we conclude that the pair $(\check{A} - \check{G}\check{C}, [\check{B}\ \check{G}\check{R}^{1/2}])$ is reachable.

Proposition 4. *Let the pair (A, B) be stabilizable, and the reachable subsystem be also observable. Take the observer gain such that the closed loop observer matrix is stable. Consider the sequence generated by (11) with $0 \leq P_0 \leq \Sigma_\rho$, with $\Sigma_\rho = T^{-1}\check{\Sigma}_\rho T^{-\top}$, and $0 < \theta \leq \beta_\rho$. Then $P_t \leq \Sigma_\rho$ and $\sigma_{max}(P_{t+1}) < \theta^{-1}$. Moreover, for $P_0 = \Sigma_\rho$, the sequence is monotone nonincreasing.*

Proof. We prove the last claim by inductive reasoning. For $t = 0$ we consider the sequence $\check{P}_t$, i.e. the one obtained after the basis change in (20). More precisely, we have $\check{P}_0 = \check{\Sigma}_\rho$. By subtracting (26) for $t = 0$ from (29) we have

$$\begin{aligned} \check{\Sigma}_\rho - \check{P}_1 &= (\check{A} - \check{G}\check{C})(\rho^2\check{\Sigma}_\rho - \check{\Gamma}_0)(\check{A} - \check{G}\check{C})^\top \\ &\quad + [(\check{A} - \check{G}\check{C})\check{\Gamma}_t\check{C}^\top - \check{G}\check{R}](\check{C}\check{\Gamma}_t\check{C}^\top + \check{R})^{-1} \\ &\quad \times [(\check{A} - \check{G}\check{C})\check{\Gamma}_t\check{C}^\top - \check{G}\check{R}]^\top. \end{aligned} \tag{31}$$

[1]The cyclicity index of a matrix is the maximum number of miniblocks which can appear in a block of the Jordan form of the matrix.

Notice that

$$\rho^2\check{\Sigma}_\rho - \check{\Gamma}_0 = \rho^2\check{\Sigma}_\rho - (\check{\Sigma}_\rho^+ - \theta\Pi(\check{\Sigma}_\rho))^+$$

which is positive semidefinite by (30). Taking into account (31) it follows that

$$\check{P}_1 \leq \Sigma_\rho = \check{P}_0.$$

Notice that the latter in the original basis becomes

$$P_1 \leq P_0. \tag{32}$$

For $t > 0$ assume that $P_t \leq P_{t-1}$, then

$$P_{t+1} = r_\theta(P_t) \leq r_\theta(P_{t-1}) = P_t$$

where we exploited the monotonicity of $r_\theta(\cdot)$, see Lemma 3. Hence, we have shown the monotone nonincreasing property of the sequence which also implies, since $P_0 \leq \Sigma_\rho$, that $P_t \leq \Sigma_\rho$ for any $t \geq 0$.

To prove the first part of the statement, consider the sequence $P'_{t+1} = r_\theta(P'_t)$, $t \geq 0$, with $P'_0 = \Sigma_\rho$. In view of the previous result we have that $P'_t \leq \Sigma_\rho$. Next, we prove the claim by inductive reasoning. First, for $t = 0$ note that $P_0 \leq P'_0$. For $t > 0$, assume that $P_t \leq P'_t$ then

$$P_{t+1} = r_\theta(P_t) \leq r_\theta(P'_t) = P'_{t+1}$$

where in the last equality we exploited the monotonicity of $r_\theta(\cdot)$ proved in Lemma 3. Therefore, $P_t \leq P'_t \leq \Sigma_\rho$ which also implies, in view of (30), that $\sigma_{max}(P_t) < \theta^{-1}$. □

In view of all the results that we have found, we obtain the following convergence result of P_t of the proposed degenerate risk sensitive filter.

Theorem 2. *Let the pair (A, B) be stabilizable, and the reachable subsystem be also observable. Take the observer gain such that the closed loop observer matrix is stable. Let $0 < \theta \leq \bar{\theta}$. If P_0 is such that $P_0 < \theta^{-1}(1 - \max_i |\sigma_i(A_{22})|^2)I$ and $0 \leq P_0 \leq \Sigma_\rho$ with $0 < \theta \leq \beta_\rho$ and $1 < \rho < \alpha^{-1}$, then the sequence (11) converges to a unique solution P_∞.*

5. Numerical experiments

Theorem 2 shows that imposing some conditions on θ and P_0, then it is guaranteed the convergence of the degenerate risk sensitive filter in the case of constant parameters. Clearly, there is an interplay between these two variables. Accordingly, by Theorem 2, we can define the region

$$\begin{aligned} \mathcal{R} := \{\, (\theta, P_0) \; : \; &\exists\, \rho, \check{G} \text{ s.t. } \check{A} - \check{G}\check{C} \text{ stable}, \\ &0 < \theta \leq \bar{\theta},\, \theta \leq \beta_\rho,\, 1 < \rho < \alpha^{-1}, \\ &0 \leq P_0 \leq \Sigma_\rho,\, P_0 < \theta^{-1}(1 - \max_i |\sigma_i(A_{22})|^2)I \,\} \end{aligned}$$

where the robust filter is guaranteed that it converges. The numerical characterization of $\mathcal{R}$ is very difficult. However, in the case the index of cyclicity of the state matrix of the unreachable subsystem is less than or equal to p, it is possible to consider the following subset of $\mathcal{R}$

$$\begin{aligned}\mathcal{R}^\star := \{\, (\theta, \sigma_0^2 I) \ :\ &\exists\, \rho, \check{G} \text{ s.t. } \check{A} - \check{G}\check{C} \text{ stable}, \\ &0 < \theta \le \bar{\theta},\, \theta \le \beta_\rho,\, 1 < \rho < \alpha^{-1}, \\ &0 \le \sigma_0^2 I \le \Sigma_\rho,\, \sigma_0^2 I < \theta^{-1}(1 - \max_i |\sigma_i(A_{22})|^2) I \,\}\end{aligned}$$

where $\sigma_0^2 := \sigma_{max}(P_0)$. It is worth noting that $\mathcal{R}^\star$ is isomorphic to the set

$$\mathcal{R}_2^\star := \{\, (\theta, \sigma_0^2) \ :\ (\theta, \sigma_0^2 I) \in \mathcal{R}^\star \,\} \subseteq \mathbb{R}^2.$$

Accordingly, the characterization of $\mathcal{R}_2^\star$ means to find a region in $\mathbb{R}^2$. In what follows we propose a method in order to find $\mathcal{R}_2^\star$ numerically.

First, we fix $N \ge n$ sufficiently large (as suggested in [37]); in this way we can compute the constant $\bar{\theta}$. Let $\theta \in (0, \bar{\theta}]$. If $(\theta, \bar{\sigma}_0^2) \in \mathcal{R}_2^\star$ then

$$\{\, (\theta, \sigma_0^2) \ :\ \theta \text{ fixed},\ \sigma_0^2 \le \bar{\sigma}_0^2 \,\} \subseteq \mathcal{R}_2^\star.$$

Thus, the problem to characterize $\mathcal{R}_2^\star$ boils down to the problem to find the point $(\theta, \sigma_0^2) \in \mathcal{R}_2^\star$, with $\theta \in (0, \bar{\theta}]$ fixed, which maximizes σ_0^2. More precisely, we can consider a grid for the interval $(0, \bar{\theta}]$ of K points. Then, for any θ of this grid, the corresponding maximum for σ_0^2 is given by the following optimization problem

$$\begin{aligned} \max_{\rho, \check{G}, \sigma_0^2} \quad & \sigma_0^2 \\ \text{s.t.} \quad & \check{A} - \check{G}\check{C} \text{ stable} \\ & \Sigma_\rho > 0 \\ & 0 < \sigma_0^2 < c \\ & \sigma_0^2 \le \sigma_{min}(\Sigma_\rho) \\ & \theta \le \beta_\rho \\ & 1 < \rho < \alpha^{-1} \end{aligned} \tag{33}$$

where the constant $c := \theta^{-1}(1 - \max_i |\sigma_i(A_{22})|^2)$. In what follows, we will solve this problem through the Matlab function "fmincon" of the nonlinear optimization Toolbox.

It is worth noting that it is not guaranteed that Problem (33) admits solution. Indeed, for a given $\theta \in (0, \bar{\theta}]$ it is not guaranteed that there exists $\check{G}$ and ρ such that $\theta \le \beta_\rho$. On the other hand, Theorem 2 guarantees that there exists $0 < \tilde{\theta} \le \bar{\theta}$ such that Problem (33) admits solution for $\theta \in (0, \tilde{\theta}]$. In practice we can check whether Problem (33) admits solution for a given $\theta \in (0, \bar{\theta}]$ through a Monte Carlo experiment composed by many trials: in each trial we randomly generate $\check{G}$ such that $\check{A} - \check{G}\check{C}$ is stable and $(\check{A} - \check{G}\check{C}, [\check{B}\ \check{G}\check{R}^{1/2}])$ reachable; then we consider a grid for the interval $(1, \alpha^{-1})$ and for any ρ of this grid we compute Σ_ρ and β_ρ; if there is a ρ of this grid, say $\tilde{\rho}$, such that $\theta \le \beta_{\tilde{\rho}}$ then Problem (33) admits solution and $\check{G}$, $\tilde{\rho}$, $\sigma_0^2 = \min\{c/2, \sigma_{min}(\Sigma_{\tilde{\rho}})\}$ can be used as initial point for "fmincon"; if in all the trials it is not possible to find such $\tilde{\rho}$ then it is plausible that Problem (33) does not admit solution for such θ. Moreover, since Problem (33) is not concave, it is necessary to avoid getting stuck in a local maximum. Therefore, we will solve Problem (33) with different initial points and then we take the solution for which σ_0^2 takes the maximum value.

Next, we show the convergence region $\mathcal{R}_2^\star$ obtained with the numerical method presented above and corresponding to the state space model (1) where

$$A = \begin{bmatrix} 1.2 & 0.1 & 0.1 \\ 0 & 0.45 & -0.78 \\ 0 & 0.78 & 0.45 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$
$$C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \qquad D = \begin{bmatrix} 0 & 1 \end{bmatrix}.$$

It is worth noting that the pair (A, B) is stabilizable, but not reachable, while the pair (A, C) is observable. Hence, the hypotheses in Theorem 2 hold. Moreover, the cyclicity index of the state matrix of the unreachable subsystem is equal to the output dimension. In order to find $\mathcal{R}_2^\star$ we set $N = 50$. Figure 1 shows the region of convergence (purple area) found by means of Theorem 2. In order to assess the goodness of such region we also computed the real region of convergence (red area in Fig. 1) as follows: for θ fixed we compute the maximum value of σ_0^2 such that the sequence in (11) with $P_0 = \sigma_0^2 I$ converges (moreover, we also checked the uniqueness of the limit considering the sequence in (11) with different $P_0 \le \sigma_0^2 I$). From the figure we can see that $\mathcal{R}_2^\star$ does not coincide with the real region of convergence, however, it provides an acceptable approximation of it.

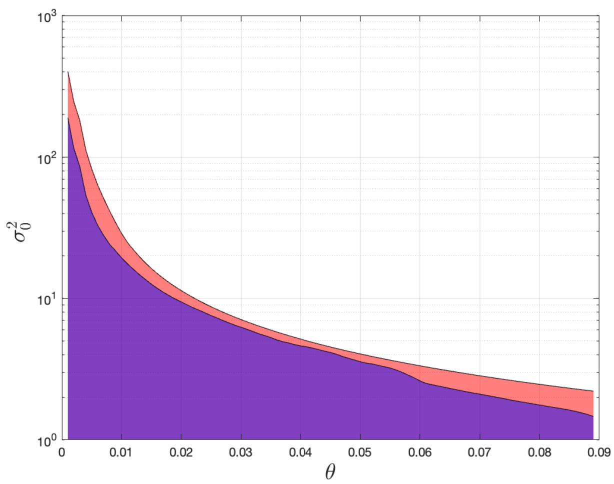

Figure 1: Region of convergence found by means of Theorem 2 (purple area) versus real region of convergence (red area).

6. Conclusions

In this paper we have proposed an extension of the risk sensitive filter which considers also the situation in which the conditional probability density is degenerate. The corresponding risk sensitive Riccati rercusion evolves on the cone of the positive semidefinite matrices, i.e. the covariance matrix could be singular. We have shown that such recursion converges to the unique solution of the corresponding algebraic equation under the assumption that the risk sensitivity parameter and the eigenvalues of the initial covariance matrix are sufficiently small. We have showed that there is an interplay between these two variables. In particular, our result can be useful to compute numerically an approximation of the region (depending on these two variables) in which the degenerate risk sensitive filter converges.

References

[1] B. Hassibi, A. H. Sayed, T. Kailath, Indefinite-quadratic estimation and control: a unified approach to H^2 and H^∞ theories, SIAM, 1999.

[2] L. El Ghaoui, G. Calafiore, Robust filtering for discrete-time systems with bounded noise and parametric uncertainty, IEEE Transactions on Automatic Control 46 (7) (2001) 1084–1089.

[3] J. L. Speyer, W. H. Chung, Stochastic processes, estimation, and control, SIAM, 2008.

[4] S. Kim, V. Deshpande, R. Bhattacharya, Robust Kalman filtering with probabilistic uncertainty in system parameters, IEEE Control Systems Letters 5 (1) (2020) 295–300.

[5] B. Li, Y. Tan, L. Zhou, R. Dong, Robust-nonsmooth Kalman filtering for stochastic sandwich systems with dead-zone, International Journal of Control, Automation and Systems 19 (1) (2021) 101–111.

[6] M. Zorzi, On the robustness of the Bayes and Wiener estimators under model uncertainty, Automatica 83 (2017) 133–140.

[7] M. Zorzi, Distributed Kalman filtering under model uncertainty, IEEE Transactions on Control of Network Systems 7 (2) (2019) 990–1001.

[8] A. Zenere, M. Zorzi, On the coupling of model predictive control and robust Kalman filtering, IET Control Theory & Applications 12 (13) (2018) 1873–1881.

[9] P. Whittle, P. R. Whittle, Risk-sensitive optimal control, vol. 2, Wiley New York, 1990.

[10] R. N. Banavar, J. L. Speyer, Properties of risk-sensitive filters/estimators, IEE Proceedings-Control Theory and Applications 145 (1) (1998) 106–112.

[11] B. C. Levy, M. Zorzi, A contraction analysis of the convergence of risk-sensitive filters, SIAM Journal on Control and Optimization 54 (4) (2016) 2154–2173.

[12] S. Gao, S. Zhao, X. Luan, F. Liu, Adaptive risk-sensitive filter for Markovian jump linear systems, Automatica 151 (2023) 110897.

[13] S. Yi, M. Zorzi, Robust fixed-lag smoothing under model perturbations, Journal of the Franklin Institute 360 (1) (2023) 458–483.

[14] R. K. Tiwari, S. Bhaumik, Risk sensitive filtering with randomly delayed measurements, Automatica 142 (2022) 110409.

[15] M. Cheng, D. Shi, T. Chen, Event-triggered risk-sensitive smoothing for linear Gaussian systems, Automatica 158 (2023) 111301.

[16] J. Huang, D. Shi, T. Chen, Robust event-triggered state estimation: A risk-sensitive approach, Automatica 99 (2019) 253–265.

[17] R. Boel, M. James, I. Petersen, Robustness and risk-sensitive filtering, IEEE Transactions on Automatic Control 47 (3) (2002) 451–461.

[18] L. P. Hansen, T. J. Sargent, Robust estimation and control under commitment, Journal of economic Theory 124 (2) (2005) 258–301.

[19] M.-G. Yoon, V. A. Ugrinovskii, I. R. Petersen, Robust finite horizon minimax filtering for discrete-time stochastic uncertain systems, Systems & Control Letters 52 (2) (2004) 99–112.

[20] B. C. Levy, R. Nikoukhah, Robust least-squares estimation with a relative entropy constraint, IEEE Transactions on Information Theory 50 (1) (2004) 89–104.

[21] P. Bougerol, Kalman filtering with random coefficients and contractions, SIAM Journal on Control and Optimization 31 (4) (1993) 942–959.

[22] B. C. Levy, R. Nikoukhah, Robust state space filtering under incremental model perturbations subject to a relative entropy tolerance, IEEE Transactions on Automatic Control 58 (3) (2012) 682–695.

[23] M. Zorzi, Robust Kalman filtering under model perturbations, IEEE Transactions on Automatic Control 62 (6) (2016) 2902–2907.

[24] S. Abadeh, V. Nguyen, D. Kuhn, P. Esfahani, Wasserstein distributionally robust Kalman filtering, in: Advances in Neural Information Processing Systems, 8474–8483, 2018.

[25] D. Boskos, J. Cortés, S. Martínez, Data-driven ambiguity sets with probabilistic guarantees for dynamic processes, IEEE Transactions on Automatic Control 66 (7) (2021) 2991–3006.

[26] B. Han, Distributionally robust Kalman filtering with volatility uncertainty, arXiv preprint arXiv:2302.05993 .

[27] M. Zorzi, B. C. Levy, On the convergence of a risk sensitive like filter, in: 54th IEEE Conference on Decision and Control (CDC), 4990–4995, 2015.

[28] M. Zorzi, B. Levy, Robust Kalman filtering: asymptotic analysis of the least favorable model, in: 57th IEEE Conference on Decision and Control (CDC), 7124–7129, 2018.

[29] S. Yi, M. Zorzi, Robust Kalman filtering under model uncertainty: The case of degenerate densities, IEEE Transactions on Automatic Control 67 (7) (2021) 3458–3471.

[30] S. Yi, M. Zorzi, Low-rank Kalman filtering under model uncertainty, in: 2020 59th IEEE Conference on Decision and Control (CDC), IEEE, 2930–2935, 2020.

[31] S. Bonnabel, R. Sepulchre, The geometry of low-rank Kalman filters, in: Matrix Information Geometry, Springer, 53–68, 2013.

[32] A. Ferrante, L. Ntogramatzidis, The generalized continuous algebraic Riccati equation and impulse-free continuous-time LQ optimal control, Automatica 50 (4) (2014) 1176–1180.

[33] A. Ferrante, L. Ntogramatzidis, The generalised discrete algebraic Riccati equation in linear-quadratic optimal control, Automatica 49 (2) (2013) 471–478.

[34] P. Whittle, Risk-sensitive optimal control, J. Wiley, Chichester, England, 1980.

[35] A. Longhini, M. Perbellini, S. Gottardi, S. Yi, H. Liu, M. Zorzi, Learning the tuned liquid damper dynamics by means of a robust EKF, in: 2021 American Control Conference (ACC), 60–65, 2021.

[36] S. Yi, M. Zorzi, On the worst case analysis of a robust EKF, IEEE Transactions on Automatic Control, under review .

[37] M. Zorzi, Convergence analysis of a family of robust Kalman filters based on the contraction principle, SIAM Journal on Control and Optimization 55 (5) (2017) 3116–3131.

[38] P. del Moral, E. Horton, A note on Riccati matrix difference equations, arXiv preprint arXiv:2107.12918, 2021.